A^{2}TTA: Anchored-and-Agile Test-Time Adaptation for Evolving Traffic Sensor Networks

Du Yin
du.yin@unsw.edu.au
University of New South Wales
Sydney, NSW, Australia
HKUST(GZ)
China

Xiachong Lin
dawn.lin@student.unsw.edu.au
University of New South Wales
Sydney, NSW, Australia

Yue Tan
yue.tan@griffith.edu.au
Griffith University
Brisbane, Queensland, Australia

Jinliang Deng
dengjinliang@ust.hk
Beihang University
Beijing, China

Estrid He
estrid.he@rmit.edu.au
RMIT University
Melbourne, VIC, Australia

Hao Xue
hao.xue1@unsw.edu.au
University of New South Wales
Sydney, NSW, Australia
HKUST(GZ)
China

Flora Salim
flora.salim@unsw.edu.au
University of New South Wales
Sydney, NSW, Australia

Abstract

Traffic forecasting is important for efficient traffic management and route planning in smart cities. Existing traffic forecasting studies typically assume fixed sensor graphs, overlooking the continuous evolution of real-world traffic networks, e.g., ongoing road network construction and evolving human mobility patterns. These dynamic changes can substantially degrade conventional forecasting models, motivating test-time adaptation (TTA) to efficiently adapt pretrained models during deployment. However, applying TTA to evolving traffic sensor networks remains challenging in two aspects. First, topology expansion introduces new sensors and connections, continuously reshaping the sensor graph. Second, temporal shifts vary in time scale and stability, requiring differentiated adaptation to long-term and short-term shifts. In this study, we address these challenges by proposing **A^{2}TTA**, an **A**nchored-and-**A**gile **T**est-**T**ime **A**daptation framework for evolving traffic sensor networks, which transforms topology-induced forecasting errors into an expandable output calibration problem and separates temporal adaptation into persistent global correction and agile context-specific specialization. By jointly addressing topology evolution and multi-scale temporal shifts, A^{2}TTA enables efficient and robust adaptation to continuously evolving traffic environments. Extensive experiments on ten real-world traffic networks demonstrate that A^{2}TTA consistently improves forecasting performance across different backbones, datasets, and prediction horizons. Our code is available in https://anonymous.4open.science/r/A2TTA.

CCS Concepts

• **Information systems → Sensor networks**.

Keywords

Traffic Forecasting, Evolving Traffic Sensor, Test Time Adaptation

ACM Reference Format:
Du Yin, Xiachong Lin, Yue Tan, Jinliang Deng, Estrid He, Hao Xue, and Flora Salim. 2027. A^{2}TTA: Anchored-and-Agile Test-Time Adaptation for Evolving Traffic Sensor Networks. In . ACM, New York, NY, USA, 19 pages.

1 Introduction

Traffic forecasting is vital for building smart cities and intelligent transportation systems, helping alleviate congestion and improve mobility by supporting efficient traffic management and route planning [4, 20, 29, 34, 35]. Usually, existing traffic forecasting studies are built upon fixed sensor graphs, where nodes represent traffic sensors deployed at specific locations and edges characterize the spatial connectivity of road networks. To model complex spatio-temporal traffic dynamics and accurately predict future traffic states, e.g., traffic flow, speed, and demand, various data-driven methods have been developed, typically following a workflow of traffic data collection, graph construction, spatio-temporal modeling, and future-state prediction. This pipeline has substantially broadened the range of downstream applications that traffic forecasting can support, such as congestion management, route planning, and traffic control, etc.

Traditional methods generally assume a static traffic network, following a train-once-and-deploy paradigm where the trained model is expected to remain effective in a long period [9–11, 21]. Unfortunately, this assumption rarely holds in practice, as real-world traffic

Conference'17, Washington, DC, USA

sensor networks continuously evolve over time. On the one hand, road networks undergo continuous construction and renewal, leading to the deployment of new sensors and the removal of outdated ones. Meanwhile, ongoing urbanization reshapes road layouts and connectivity, further altering the underlying sensor graph. On the other hand, human mobility patterns also evolve over time due to changes in commuting behavior, population distribution, and transportation policies, resulting in shifts in traffic dynamics even when the road topology remains unchanged. In these cases, conventional train-once-and-deploy forecasting models may suffer from substantial performance degradation, as both the underlying sensor graph and traffic dynamics progressively shift away from those observed during training. To handle the above dynamic changes, a straightforward solution is to regularly train the model. However, retraining can be expensive and may overwrite useful knowledge with short-lived patterns. Under this circumstance, a promising solution is Test-Time Adaptation (TTA), which updates a pretrained model from data observed during deployment [8, 26], rather than retraining the model from scratch. As a result, TTA maintains adaptation efficiency while effectively reusing the knowledge encoded in the pretrained backbone.

Despite the strong potential of TTA, existing vanilla TTA solutions fail to effectively handle the evolving traffic forecasting problem due to the following two challenges. One key challenge is ***Challenge 1 - Topology Expansion***. Unlike conventional distribution shifts that occur in a fixed input space, evolving traffic networks exhibit continuous expansion of both the node set and the graph structure. Specifically, the deployment of new sensors introduces previously unseen nodes and connections into the graph, which alters the local neighborhoods of existing nodes and reshapes local graph structures and spatial dependencies. It is hard for conventional TTA methods to deal with such topology-level shift since they are typically designed to adapt to shifts over a fixed set of nodes and a predefined graph topology, lacking the ability to capture dynamic structural changes.

In addition to topology expansion, evolving human mobility patterns over time introduce another key challenge, termed ***Challenge 2 - Temporal Distribution Shift***. Due to periodic and irregular fluctuations in human behavior, the input-output relationships and spatio-temporal patterns encountered at test time can deviate from those learned during training. For example, long-lasting changes in population, road infrastructure, or transportation policies may induce long-term shifts to the overall traffic characteristics over years. In contrast, temporary events or unexpected incidents may cause short-term, context-specific shifts that are highly dependent on the specific temporal and environmental conditions. In Figure 1(a, b), we exhibit this evolving deployment setting and show the persistent drift and transient deviations alongside topology expansion, respectively. Also, labels become available for adaptation only after the forecasting horizon has elapsed. In this case, this challenge lies not only in adapting to temporal shifts, but in identifying distinct types of temporal shifts with varying time scales and degrees of stability, and applying appropriate persistent corrections.adaptation strategies accordingly.

To address the above challenges, we propose **A**nchored-and-**A**gile **T**est-**T**ime **A**daptation (**A^{2}TTA** for short) for evolving traffic sensor networks. As illustrated in Figure 3, A^{2}TTA couples two

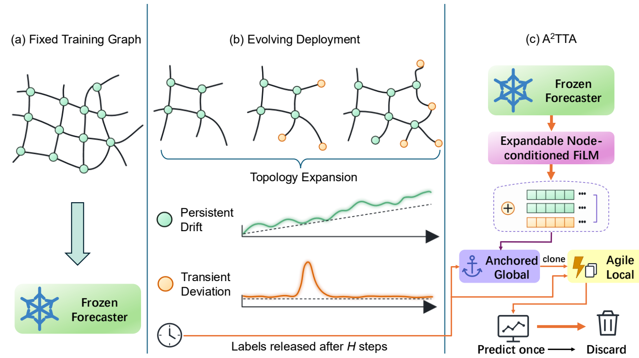

Figure 1: Motivation and design overview of A^{2}TTA. (a) The conventional fixed-graph assumption trains and deploys a forecaster on one unchanged sensor graph. (b) Deployment instead brings topology expansion together with persistent drift and transient deviations, while labels are released only after the forecasting horizon H. (c) A^{2}TTA keeps the forecasting backbone frozen, appends node embeddings to an expandable node-conditioned FiLM calibrator, and routes delayed feedback to an anchored global state and a disposable local clone.

complementary designs. To handle ***Challenge 1***, we attach an expandable and node-conditioned FiLM calibrator to the frozen backbone, which corrects the base forecast by conditioning on recent node observations, temporal statistics, and an expandable node embedding. To address ***Challenge 2***, we maintain two complementary calibrator states that operate at different temporal scales. For persistent long-term shifts, an anchored global calibrator is continually updated using all causally released samples retained in the feedback pool and regularized toward its warm-up state. Meanwhile, for context-specific short-term deviations, an agile local clone is specialized using released samples weighted by temporal phase, traffic-pattern similarity, and recency. To prevent transient contexts from contaminating the persistent global state, the clone is used for a single prediction and then discarded. Through these complementary designs, A^{2}TTA enables topology-aware and temporally adaptive forecasting, providing an efficient and robust solution for forecasting in continuously evolving traffic sensor networks.

Our main contributions are summarized as follows:

- We formulate traffic forecasting on evolving sensor networks as a causal delayed-feedback adaptation problem, jointly considering topology expansion and multi-scale temporal distribution shifts.
- We propose A^{2}TTA, a lightweight and backbone-agnostic framework that converts node-heterogeneous forecasting errors into an expandable FiLM-based output calibration problem while keeping the forecasting backbone frozen.
- We develop an anchored-and-agile adaptation mechanism that combines a persistent global calibrator for long-term drift with a disposable, context-specialized local clone for transient deviations. Extensive experiments on EvoXXLTraffic and TFNSW demonstrate consistent improvements across backbones, datasets, and forecasting horizons.

Table 1: EvoXXLTraffic scale and sensor growth.

District	Span	#Yr	$N_{\text{first}} \rightarrow N_{\text{last}}$	Timesteps	Obs.	Size
D03	2001–2025	25	174⟶1859	2,611,872	2.56B	10.2 GB
D04	2001–2025	25	1266⟶4110	2,559,744	6.63B	26.5 GB
D05	2005–2025	21	6⟶572	2,185,344	0.50B	2.0 GB
D06	2005–2025	21	13⟶746	2,138,976	0.84B	3.4 GB
D07	2001–2025	25	3215⟶4888	2,629,728	11.28B	45.1 GB
D08	2001–2025	25	262⟶2074	2,612,736	3.49B	14.0 GB
D10	2006–2025	20	45⟶1396	2,051,712	1.70B	6.8 GB
D11	1999–2025	27	424⟶1440	2,788,128	3.23B	12.9 GB
D12	2002–2025	24	1783⟶2587	2,524,608	5.73B	22.9 GB
Total	1999–2025	–	–	22,102,848	**35.97B**	**144 GB**

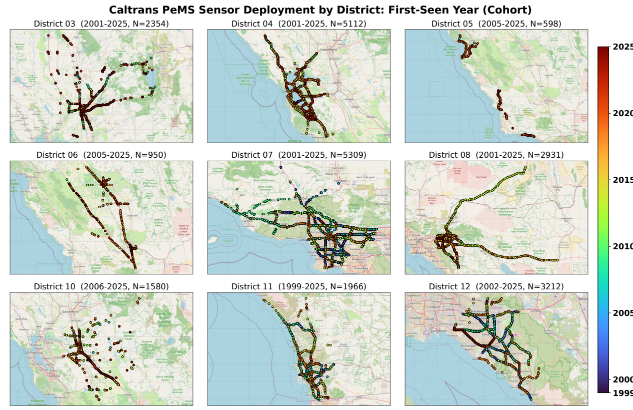

Figure 2: Sensor deployment across nine districts.

2 Related Work

Traffic forecasting has progressed from statistical time-series models [1, 27] to deep temporal and graph architectures. Representative spatio-temporal graph neural networks (STGNNs) combine recurrent, convolutional, or diffusion operators with road-network structure [13, 34, 36], while later methods learn adaptive graphs or attention-based spatial dependencies [29, 30, 37]. Recent architectures such as STID and STAEFormer further improve forecasting efficiency and accuracy [15, 20]. Although these methods can model complex and even adaptive spatial dependencies, they are generally developed and evaluated under a fixed sensor universe. Our setting instead considers multi-year deployment in which sensors are added, graph topology changes, and traffic distributions drift simultaneously.

Beyond fixed-graph forecasting, online and continual approaches address network evolution through replay, pattern memories, localized updates, or expandable prompts [4, 6, 24, 25]. Retrieval and test-time calibration methods further adapt predictions using historical patterns or streaming statistics [5, 35]. However, existing approaches typically update substantial forecasting components, rely on explicit replay or pattern memories, or focus on temporal drift over a fixed node set. A^{2}TTA takes a different route: it freezes the forecasting backbone, introduces an expandable node-conditioned FiLM calibrator, and separates causal adaptation into a persistent global state and a disposable context-specialized local clone. Both states use only labels released after the complete forecasting horizon has elapsed. A more complete review is provided in Appendix A.2.

3 Dataset and Empirical Motivation

3.1 EvoXXLTraffic Construction

EvoXXLTraffic organizes five-minute records from the California Department of Transportation Performance Measurement System (PeMS) [3, 33] into yearly graph snapshots for nine districts (D03–D08 and D10–D12). For each district and year, the union of observed sensors forms $\mathcal{V}_y$, while persistent IDs align flow matrices and identify sensor additions and removals. Missing values use sensor-wise forward fill, backward fill, and a zero fallback. Sensor coordinates define $\mathbf{A}_y$ through Haversine distances and a thresholded Gaussian kernel ($\epsilon = 0.1$). TFNSW is a long-span external test network.

3.2 Sensor Network Growth

Deployment cohorts in Figure 2 extend and densify road corridors, with sensor counts increasing by up to 95×. Table 1 gives the district-level scale. The structural diagnostics in Appendix A.1 show non-monotonic changes in average degree and graph density, motivating adaptation to both topology and traffic distributions.

4 Problem Formulation

Following EvoXXLTraffic [33], we model each district as yearly graph snapshots whose sensor set, adjacency matrix, and traffic observations may change.

4.1 Evolving Sensor Graph

For district d and year y, let $\mathbf{X}_y^d$, $\mathbf{A}_y^d$, and $\mathcal{V}_y^d$ denote the traffic tensor, adjacency matrix, and active sensor set:

$$\mathcal{D}_{evo}^d = \left\{ \left(\mathbf{X}_y^d, \mathbf{A}_y^d, \mathcal{V}_y^d \right) \right\}_{y=y_0}^{y_1}, \quad \mathbf{X}_y^d \in \mathbb{R}^{T_y^d \times N_y^d \times C}, \quad N_y^d = |\mathcal{V}_y^d|. \tag{1}$$

Here, T_y^d, N_y^d, and C are the numbers of time steps, sensors, and channels. Both $\mathcal{V}_y^d$ and $\mathbf{A}_y^d$ may change between years.

New sensors in year y are

$$\Delta \mathcal{V}_y^d = \mathcal{V}_y^d \setminus \mathcal{V}_{y-1}^d. \tag{2}$$

The intersection contains retained sensors, while $\mathcal{V}_{y-1}^d \setminus \mathcal{V}_y^d$ contains inactive ones. Sensors in $\Delta \mathcal{V}_y^d$ appear in year y's training, validation, and test partitions but in no earlier snapshot. Thus, "new sensor" denotes a topology-expansion cohort rather than zero-label cold start.

4.2 Forecasting Task

For clarity, we omit the district superscript d below and focus on univariate traffic flow ($C = 1$). Within yearly snapshot y, let $\mathbf{s}_t \in \mathbb{R}^{N_y}$ denote the traffic state at time t. The input and target windows are

$$\begin{aligned} \mathbf{X}_t &= [\mathbf{s}_{t-L+1}, \dots, \mathbf{s}_t] \in \mathbb{R}^{N_y \times L}, \\ \mathbf{Y}_t &= [\mathbf{s}_{t+1}, \dots, \mathbf{s}_{t+H}] \in \mathbb{R}^{N_y \times H}. \end{aligned} \tag{3}$$

For online batch $\mathcal{B}_b$ starting at window τ_b, target $\mathbf{Y}_i$ is available only if $i + H \leq \tau_b$. The main evaluation uses $|\mathcal{B}_b| = 1$, so every prediction precedes the next test window.

Yearly training and deployment lifecycle. For each year, the host method produces f_{θ_y} from the chronological training partition and its standard validation procedure. The baseline and A^{2}TTA share this checkpoint and seed. A^{2}TTA freezes θ_y, expands and warms up the carried calibrator on the same training data, and updates only that calibrator during testing with released targets. The global calibrator is then carried to the next year and expanded for new nodes.

5 Methodology

5.1 Overview: From Evolving-Graph Challenges to A^{2}TTA

Figure 3 illustrates A^{2}TTA. Given the current graph snapshot $\mathcal{G}_y = (\mathcal{V}_y, \mathbf{A}_y)$ and input window $\mathbf{X}_t$, a frozen backbone produces a base forecast. An expandable node-conditioned FiLM calibrator then corrects node- and horizon-specific bias without modifying the backbone. A delayed-feedback pool exposes each target only after its full forecasting horizon has elapsed. Released samples update a persistent global state for shared drift and a disposable local clone for the current context. The clone is used once and discarded, separating persistent from transient shifts without explicitly classifying them.

5.2 Expandable Node-Conditioned Calibration

Frozen backbone interface. For a yearly graph snapshot $\mathcal{G}_y = (\mathcal{V}_y, \mathbf{A}_y)$, let f_{θ_y} be the matched year-specific checkpoint produced by the host forecasting method. It maps the current input window to an H-step base forecast:

$$\widehat{\mathbf{Y}}_t^{\text{base}} = f_{\theta_y}(\mathbf{X}_t, \mathbf{A}_y). \tag{4}$$

Once loaded by A^{2}TTA, θ_y remains frozen during calibrator warm-up and online adaptation. The corresponding uncalibrated baseline uses the same checkpoint, so their difference isolates output calibration rather than backbone training. A^{2}TTA consumes only the forecast and does not alter architecture-specific internal layers.

Node-conditioned FiLM. For each window-node pair (t, n), we concatenate the base forecast standardized by exponential-moving statistics, the normalized input history, four temporal summaries (last value, mean, standard deviation, and slope), and a learnable node embedding:

$$\mathbf{z}_{t,n} = \left[\widetilde{\mathbf{y}}_{t,n}^{\text{base}} \parallel \mathbf{x}_{t,n} \parallel \text{Stat}(\mathbf{x}_{t,n}) \parallel \mathbf{e}_n\right]. \tag{5}$$

We use a shared two-layer multilayer perceptron (MLP). It maps $\mathbf{z}_{t,n}$ to horizon-specific modulation vectors:

$$\begin{aligned} [\mathbf{a}_{t,n}, \mathbf{b}_{t,n}] &= h_\phi(\mathbf{z}_{t,n}), \\ \boldsymbol{\gamma}_{t,n} &= 1 + \tfrac{1}{2}\tanh \mathbf{a}_{t,n}, \qquad \boldsymbol{\beta}_{t,n} = \sigma_{\text{base}}\mathbf{b}_{t,n}. \end{aligned} \tag{6}$$

The resulting FiLM calibrator produces

$$\widehat{\mathbf{y}}_{t,n} = g_\phi\left(\widehat{\mathbf{y}}_{t,n}^{\text{base}}, \mathbf{x}_{t,n}, n\right) = \boldsymbol{\gamma}_{t,n} \odot \widehat{\mathbf{y}}_{t,n}^{\text{base}} + \boldsymbol{\beta}_{t,n}. \tag{7}$$

Here, σ_{base} is the running standard deviation of the base forecasts. Zero initialization of the MLP output head gives $\boldsymbol{\gamma}_{t,n} = \mathbf{1}$ and $\boldsymbol{\beta}_{t,n} = \mathbf{0}$, so the initial calibrator is the identity mapping.

Expansion and warm-up. As h_ϕ is shared across nodes, its parameterization is independent of the sensor-set cardinality. This property allows A^{2}TTA to accommodate topology growth without altering the shared correction function: when new sensors enter snapshot y. Only their embedding rows are added while existing embeddings and shared FiLM parameters are preserved. The matched backbone encodes the current topology through $\mathbf{A}_y$ and remains frozen, whereas the global calibrator inherited from year $y-1$ preserves adaptation knowledge across graph snapshots. Before online deployment in year y, the expanded calibrator is warm-started on the chronological training partition. The validation MAE is generated without gradient updates, and test targets remain inaccessible at this stage. The resulting state $\phi_{y,0}$ combines inherited calibration knowledge and serves as the proximal reference for subsequent updates, which use only causally released test labels.

5.3 Causal Delayed-Feedback Pool

A forecast at chronological index i covers $i+1{:}i+H$, so its complete target $\mathbf{Y}_i$ becomes available at $i + H$. After prediction, the corresponding record enters a pending queue $\mathcal{Q}$. At the start of batch b, whose first index is τ_b, records satisfying $i + H \leq \tau_b$ are released into a bounded first-in, first-out (FIFO) pool:

$$\mathcal{P}_b = \text{Tail}_M(\mathcal{P}_{b-1} \cup \{\mathcal{R}_i \in \mathcal{Q} \mid i + H \leq \tau_b\}), \tag{8}$$

where M is the pool capacity. Each $\mathcal{R}_i$ stores the input window, cached base forecast, observed target, node indices, and chronological index. Released records leave $\mathcal{Q}$. Both adaptation branches read only $\mathcal{P}_b$, so pending targets cannot leak into adaptation.

5.4 Anchored Global Adaptation

The persistent global state is updated from every released record retained in $\mathcal{P}_b$ once the pool is sufficiently populated:

$$\begin{aligned} \mathcal{L}_{\text{g}}(\phi; \mathcal{P}_b) = {} & \frac{1}{|\mathcal{P}_b| N_y H} \sum_{i \in \mathcal{P}_b} \sum_{n \in \mathcal{V}_y} \sum_{h=1}^{H} \left| \widehat{y}_{i,n,h}^{\phi} - y_{i,n,h} \right| \\ & + \lambda_{\text{c}} \mathcal{L}_{\text{con}} + \lambda_{\text{p}} \|\phi - \phi_{y,0}\|_2^2, \end{aligned} \tag{9}$$

where $\widehat{\mathbf{Y}}_i^{\phi} = g_\phi(\widehat{\mathbf{Y}}_i^{\text{base}}, \mathbf{X}_i)$, and $\mathcal{L}_{\text{con}}$ is the mean absolute difference between predictions under two weak perturbations of the same input. Starting from ϕ_{b-1}^{g}, we take K_{g} AdamW steps on eq. (9) to obtain ϕ_b^{g}. The full-pool term favors patterns shared across contexts, while the proximal term limits departure from $\phi_{y,0}$. This reduces persistent overreaction to transient samples. The updated state is retained across batches and carried to the next yearly graph snapshot.

5.5 Agile Local Specialization

The persistent state may underfit a short-lived traffic regime. Before predicting batch b, A^{2}TTA creates a disposable local state $\phi_b^{\text{l}} \leftarrow \text{copy}(\phi_b^{\text{g}})$ and updates only this clone. Online evaluation processes one chronological window at a time, so $|\mathcal{B}_b| = 1$; we retain batch

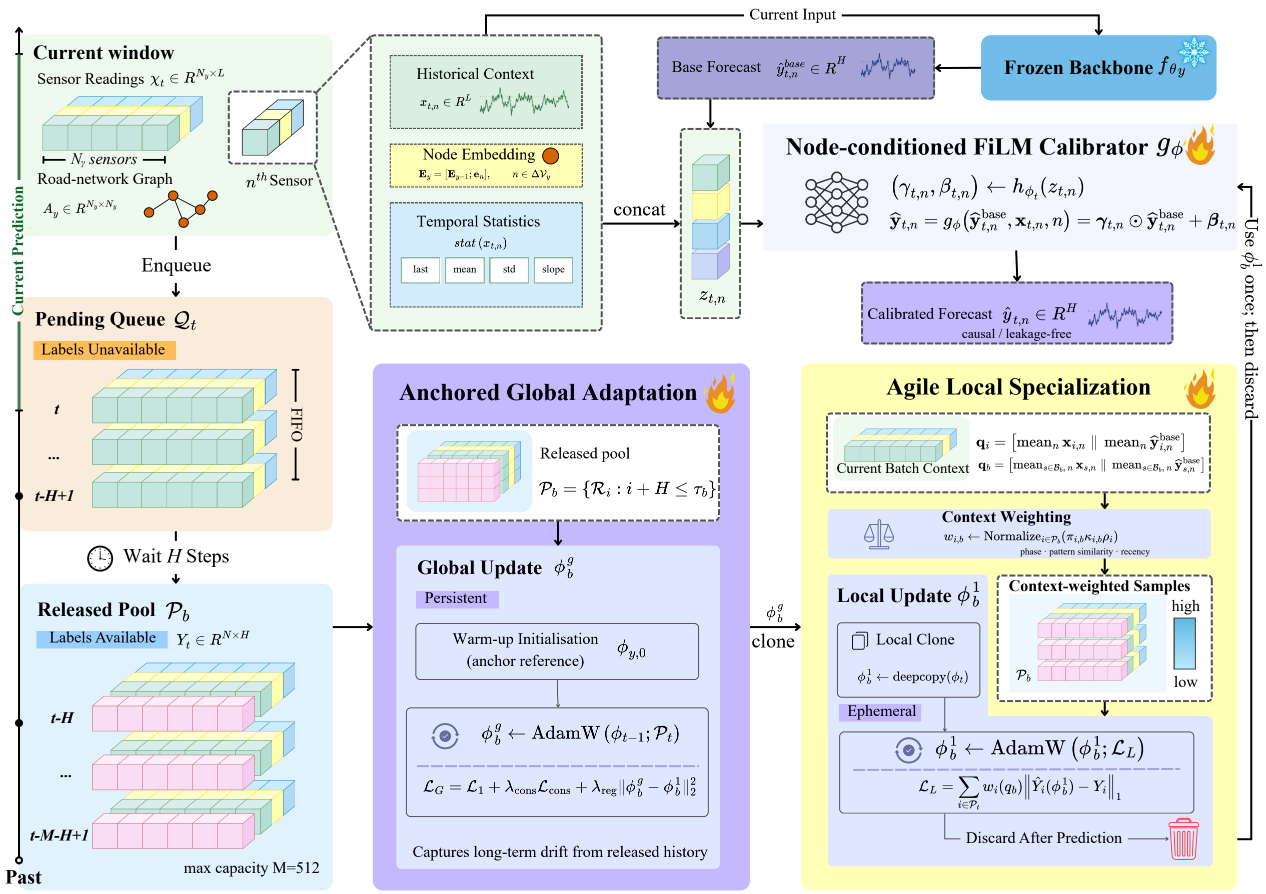

Figure 3: Overview of A^{2}TTA. An expandable node-conditioned FiLM calibrator corrects heterogeneous forecast bias as the graph expands. A pending and released feedback pipeline enforces causal label access. A persistent global state and a disposable local clone handle long-term and context-specific temporal shifts, respectively.

notation for generality. Each released window i is summarized by

$$\mathbf{q}_i = \left[\, \mathrm{mean}_n\, \mathbf{x}_{i,n} \,\|\, \mathrm{mean}_n\, \widehat{\mathbf{y}}^{\mathrm{base}}_{i,n} \right],$$
$$\mathbf{q}_b = \left[\, \mathrm{mean}_{s\in\mathcal{B}_b,n}\, \mathbf{x}_{s,n} \,\|\, \mathrm{mean}_{s\in\mathcal{B}_b,n}\, \widehat{\mathbf{y}}^{\mathrm{base}}_{s,n} \right]. \tag{10}$$

Its relevance to the current batch combines temporal phase, pattern similarity, and recency:

$$\begin{aligned}
\pi_{i,b} &= e^{-d_{\mathrm{tod}}(i,b)/\omega_{\mathrm{tod}}}\left(0.7 + 0.3\,\mathbf{1}[\mathrm{dow}_i = \mathrm{dow}_b]\right),\\
\kappa_{i,b} &= \max\{0,\, \cos(\mathbf{q}_i, \mathbf{q}_b)\} + 10^{-3},\\
\rho_i &= 0.5 + 0.5 r_i,\\
a_{i,b} &= \pi_{i,b}\kappa_{i,b}\rho_i,\\
\widetilde{w}_{i,b} &= |\mathcal{P}_b|\ \mathrm{softmax}_{i\in\mathcal{P}_b}\left(\frac{\log(\max\{a_{i,b}, \epsilon\})}{\tau}\right),\\
\bar{w}_{i,b} &= \mathrm{clip}\left(\widetilde{w}_{i,b}, c^{-1}, c\right),\\
w_{i,b} &= \frac{\bar{w}_{i,b}}{\frac{1}{|\mathcal{P}_b|}\sum_{j\in\mathcal{P}_b}\bar{w}_{j,b}}, \qquad i \in \mathcal{P}_b.
\end{aligned} \tag{11}$$

where d_{tod} is the circular time-of-day distance and S_{day} is the number of sampling steps per day. We set

$$\omega_{\mathrm{tod}} = \max(S_{\mathrm{day}}/12, 1). \tag{12}$$

Thus, $\omega_{\mathrm{tod}} = 24$ for five-minute PEMS data and 2 for hourly TFNSW data. The term $r_i \in [0,1]$ is normalized recency within the pool. The day-of-week indicator rewards phase agreement. We use $\epsilon = 10^{-6}$, temperature $\tau = 1$, and clipping constant $c = 5$. The clipped weights are rescaled to have unit mean. Their effective sample size is

$$\mathrm{ESS}(\mathbf{w}_b) = \frac{\left(\sum_{i\in\mathcal{P}_b} w_{i,b}\right)^2}{\sum_{i\in\mathcal{P}_b} w_{i,b}^2}. \tag{13}$$

If $\mathrm{ESS}(\mathbf{w}_b) < 0.2|\mathcal{P}_b|$, we use uniform weights, $w_{i,b} = 1$. The same window weight is applied to every node in record i. The local clone then minimizes

$$\mathcal{L}_1(\phi^1_b; \mathcal{P}_b) = \frac{1}{|\mathcal{P}_b| N_y H}\sum_{i\in\mathcal{P}_b}\sum_{n\in\mathcal{V}_y} w_{i,b}\left\|\widehat{\mathbf{y}}^1_{i,n} - \mathbf{y}_{i,n}\right\|_1, \tag{14}$$

Algorithm 1 Causal adaptation with global and local states in A^{2}TTA

Require: Frozen f_{θ_y}, warm-started $g_{\phi_{y,0}}$, stream $\{\mathcal{B}_b\}$, delay H, capacity M, global-update interval Δ, minimum pool size $m = \max\{8, \lfloor 0.1M \rfloor\}$
Ensure: Causal forecasts $\{\widehat{\mathbf{Y}}_b\}$
$\mathcal{Q} \leftarrow \emptyset$; $\mathcal{P} \leftarrow \emptyset$; $\phi^{\mathrm{g}} \leftarrow \phi_{y,0}$
for each batch $\mathcal{B}_b = (\mathbf{X}_b, \mathbf{A}_y)$ starting at index τ_b **do**
  Move all records with $i + H \le \tau_b$ from $\mathcal{Q}$ to $\mathcal{P}$; retain the latest M
  if $|\mathcal{P}| \ge m$ and $\tau_b \bmod \Delta = 0$ **then**
    for $k = 1, \ldots, K_{\mathrm{g}}$ **do**
      $\phi^{\mathrm{g}} \leftarrow \mathrm{AdamW}\Big(\phi^{\mathrm{g}}, \nabla_{\phi^{\mathrm{g}}} \mathcal{L}_{\mathrm{g}}\Big)$ using all of $\mathcal{P}$
    end for
  end if
  $\widehat{\mathbf{Y}}_b^{\mathrm{base}} \leftarrow f_{\theta_y}(\mathbf{X}_b, \mathbf{A}_y)$
  if $|\mathcal{P}| \ge m$ **then**
    $\phi_b^{\mathrm{l}} \leftarrow \mathrm{copy}(\phi^{\mathrm{g}})$; compute $\{w_{i,b}\}_{i \in \mathcal{P}}$ by eq. (11)
    for $k = 1, \ldots, K_{\mathrm{l}}$ **do**
      $\phi_b^{\mathrm{l}} \leftarrow \mathrm{AdamW}\Big(\phi_b^{\mathrm{l}}, \nabla_{\phi_b^{\mathrm{l}}} \mathcal{L}_{\mathrm{l}}\Big)$
    end for
    $\widehat{\mathbf{Y}}_b \leftarrow g_{\phi_b^{\mathrm{l}}}(\widehat{\mathbf{Y}}_b^{\mathrm{base}}, \mathbf{X}_b)$; discard ϕ_b^{l}
  else
    $\widehat{\mathbf{Y}}_b \leftarrow g_{\phi^{\mathrm{g}}}(\widehat{\mathbf{Y}}_b^{\mathrm{base}}, \mathbf{X}_b)$
  end if
  Emit $\widehat{\mathbf{Y}}_b$ and enqueue each current window i in $\mathcal{Q}$ with release time $i + H$
end for

After K_{l} AdamW steps, the clone predicts the current batch:

$$\widehat{\mathbf{Y}}_b = g_{\phi_b^{\mathrm{l}}}(\widehat{\mathbf{Y}}_b^{\mathrm{base}}, \mathbf{X}_b). \tag{15}$$

The clone is discarded after use, so local changes do not enter the next batch. Before the released pool is sufficiently populated, A^{2}TTA predicts directly with the global state.

5.6 Unified Online Procedure

Algorithm 1 summarizes the causal order. Eligible labels are released before adaptation, the global state is updated before the local clone is created, and current windows are enqueued only after their forecasts are emitted.

6 Experiments

6.1 Experimental Settings

Each year uses a chronological 60%/20%/20% train/validation/test split. PEMS windows contain 12 five-minute steps and TFNSW windows contain 12 hourly steps. All methods share the graph snapshots, splits, test windows, and metric code. Normalization uses training statistics only; validation and test targets do not enter preprocessing.

Online methods process one chronological window at a time, match nodes by metadata sensor IDs, and release each target after 12 steps. A^{2}TTA keeps 512 released windows and updates its global state every 64 windows. Static methods receive the same windows without test labels. We report cumulative mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) at horizons 3, 6, and 12. Avg averages the cumulative scores from horizons 1 to 12, with Avg-MAE as the primary metric. Results use five paired seeds unless noted otherwise.

A^{2}TTA(OLAN) and A^{2}TTA(STAE) load and freeze the exact per-year Online-AN and STAEFormer checkpoints, respectively. Other than learning rate, their adaptation settings are shared. Appendix A.7 lists the learning rates and sensitivity results.

6.2 Baselines

We compare nine static forecasters (DCRNN, ASTGNN, TGCN, GWN, STID, STNorm, iTrans, DLinear, and STAEFormer); Pretrain, Retrain, Online-NN, and Online-AN; four evolving-graph methods (TrafficStream, PECPM, STKEC, and EAC); STRAP and ST-TTC; and zero-shot (ZS) or yearly fine-tuned (FT) versions of Chronos-2, TimesFM 2.5, and Moirai-2.0.

Each matched A^{2}TTA/backbone pair shares checkpoint seeds, yearly cohorts, and preprocessing. New-sensor results use the metadata set difference $\mathcal{V}_y \setminus \mathcal{V}_{y-1}$ at every graph-growing transition with the same five paired seeds. This isolates calibration from checkpoint, cohort, and node-order differences.

6.3 Main Results

Overall performance. Results for three representative networks appear in Table 2. A^{2}TTA(STAE) has the lowest MAE, RMSE, and MAPE on all three, reducing Avg-MAE over frozen STAEFormer by 9.4%, 13.8%, and 29.4%. On TFNSW, A^{2}TTA(OLAN) reduces MAE from 130.61 to 86.78 (33.6%) and RMSE from 235.35 to 176.97 (24.8%).

Comparison with time-series foundation models. TSFMs are reference rows and are not included in the non-foundation ranking. A^{2}TTA(STAE) outperforms every zero-shot and fine-tuned TSFM cell in Table 2, while updating only its calibrator after label release. Appendix A reports the other seven PEMS districts and horizons 3, 6, and 12 for all ten networks (tables 3 to 6).

Forecast horizons. Across all 12 steps in Figure 4, A^{2}TTA(STAE) has the lowest error on PEMS03 and TFNSW. Its mean per-step MAE reductions over frozen STAEFormer are 8.2% and 27.3%. Appendix A gives results for all sensors and the newly added subset on all ten datasets (figs. 12 and 13).

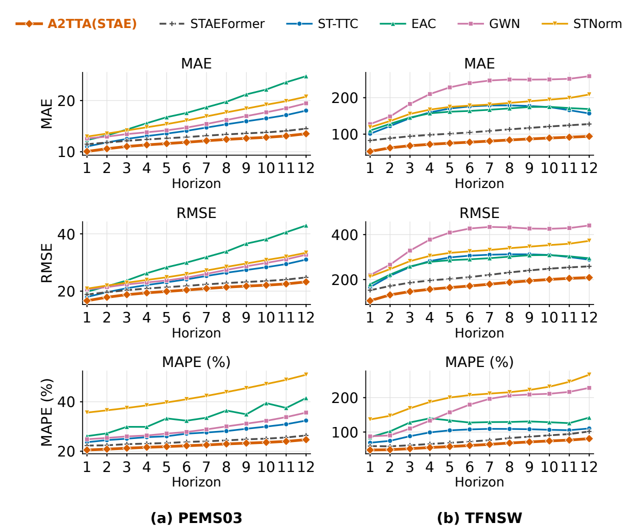

Figure 4: All-sensor per-step errors on PEMS03 and TFNSW. Lower is better.

Table 2: Average results on PEMS03, PEMS04, and TFNSW (part 1 of 3). Each row is one baseline, with MAE, RMSE, and MAPE reported for each dataset. Avg is the mean cumulative-horizon score over prediction lengths from 1 to 12. Chronos-2 uses multivariate graph-grouped inference, and its fine-tuned variant uses per-year low-rank adaptation (LoRA). Values are mean ± standard deviation when available; deterministic zero-shot runs are shown as ±0.00. MAPE is in percent. Bold and underlining mark the best and second-best non-foundation methods; foundation-model rows are not ranked.

Category	Baseline	PEMS03			PEMS04			TFNSW		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Static STGNN Backbones	DCRNN	$14.99_{\pm0.18}$	$24.90_{\pm0.19}$	$28.69_{\pm0.86}$	$22.22_{\pm0.27}$	$34.58_{\pm0.32}$	$20.53_{\pm0.40}$	$179.87_{\pm14.50}$	$303.16_{\pm19.20}$	$211.72_{\pm42.72}$
	ASTGNN	$15.15_{\pm0.10}$	$25.15_{\pm0.17}$	$26.56_{\pm0.43}$	$22.45_{\pm0.09}$	$34.96_{\pm0.12}$	$19.29_{\pm0.42}$	$140.15_{\pm0.82}$	$247.22_{\pm1.02}$	$125.28_{\pm3.02}$
	TGCN	$15.88_{\pm0.10}$	$30.12_{\pm0.40}$	$28.95_{\pm0.89}$	$23.60_{\pm0.37}$	$38.82_{\pm0.55}$	$20.15_{\pm0.28}$	$208.33_{\pm1.48}$	$390.75_{\pm3.25}$	$177.95_{\pm3.89}$
	GWN	$15.33_{\pm0.26}$	$25.15_{\pm0.39}$	$28.01_{\pm0.71}$	$23.41_{\pm0.27}$	$35.69_{\pm0.22}$	$19.94_{\pm0.55}$	$170.89_{\pm3.29}$	$313.61_{\pm5.86}$	$138.49_{\pm10.57}$
Naïve Schemes	Pretrain	$31.37_{\pm1.11}$	$41.64_{\pm0.99}$	$50.04_{\pm2.80}$	$153.33_{\pm6.62}$	$186.37_{\pm7.58}$	$82.25_{\pm2.99}$	$262.63_{\pm10.87}$	$383.38_{\pm12.29}$	$430.83_{\pm35.95}$
	Retrain	$14.16_{\pm0.15}$	$23.46_{\pm0.14}$	$26.29_{\pm0.58}$	$21.64_{\pm0.20}$	$33.58_{\pm0.22}$	$19.15_{\pm0.30}$	$138.04_{\pm0.62}$	$247.58_{\pm1.00}$	$104.00_{\pm3.59}$
	Online-NN	$17.47_{\pm0.16}$	$28.29_{\pm0.08}$	$29.62_{\pm1.04}$	$38.46_{\pm0.27}$	$51.85_{\pm0.46}$	$26.93_{\pm0.71}$	$174.36_{\pm3.31}$	$318.17_{\pm5.87}$	$129.82_{\pm8.24}$
	Online-AN	$13.80_{\pm0.05}$	$22.89_{\pm0.11}$	$24.96_{\pm0.21}$	$21.01_{\pm0.26}$	$32.73_{\pm0.31}$	$17.82_{\pm0.31}$	$130.61_{\pm0.96}$	$235.35_{\pm1.95}$	$99.93_{\pm3.54}$
Evolving Graph	TrafficStream	$14.32_{\pm0.20}$	$23.60_{\pm0.28}$	$27.63_{\pm1.67}$	$21.43_{\pm0.53}$	$33.35_{\pm0.60}$	$18.83_{\pm0.59}$	$145.42_{\pm2.41}$	$259.40_{\pm4.67}$	$106.92_{\pm8.01}$
	PECPM	$14.90_{\pm0.81}$	$24.32_{\pm0.98}$	$27.78_{\pm2.30}$	$25.16_{\pm3.18}$	$39.17_{\pm4.12}$	$22.29_{\pm3.69}$	$174.19_{\pm22.07}$	$318.61_{\pm45.67}$	$106.24_{\pm4.44}$
	STKEC	$14.47_{\pm0.22}$	$23.89_{\pm0.34}$	$27.58_{\pm0.64}$	$22.78_{\pm1.14}$	$34.78_{\pm1.18}$	$21.54_{\pm1.53}$	$147.86_{\pm1.05}$	$259.82_{\pm2.21}$	$148.01_{\pm7.69}$
	EAC	$15.91_{\pm0.48}$	$26.40_{\pm0.88}$	$30.12_{\pm0.81}$	$34.02_{\pm9.03}$	$48.20_{\pm9.87}$	$39.19_{\pm17.54}$	$137.90_{\pm1.42}$	$245.35_{\pm1.93}$	$119.37_{\pm14.85}$
Retrieval TTC	STRAP	$14.15_{\pm0.13}$	$23.37_{\pm0.22}$	$26.12_{\pm0.31}$	$21.93_{\pm0.67}$	$33.78_{\pm0.64}$	$19.49_{\pm0.57}$	$135.57_{\pm1.80}$	$243.47_{\pm4.15}$	$104.21_{\pm3.09}$
	ST-TTC	$13.98_{\pm0.14}$	$23.18_{\pm0.13}$	$26.15_{\pm0.57}$	$21.33_{\pm0.20}$	$33.21_{\pm0.22}$	$19.04_{\pm0.29}$	$144.41_{\pm0.72}$	$257.96_{\pm1.08}$	$106.33_{\pm3.01}$
Static Forecasting Backbones	STID	$15.14_{\pm0.14}$	$27.10_{\pm0.38}$	$26.44_{\pm0.62}$	$22.43_{\pm0.18}$	$35.63_{\pm0.31}$	$18.51_{\pm0.26}$	$153.63_{\pm6.20}$	$307.19_{\pm10.11}$	$112.25_{\pm11.68}$
	STNorm	$18.20_{\pm0.47}$	$28.58_{\pm0.65}$	$52.94_{\pm2.57}$	$23.99_{\pm0.62}$	$36.65_{\pm0.80}$	$27.98_{\pm1.51}$	$148.78_{\pm2.33}$	$275.25_{\pm3.16}$	$182.00_{\pm6.10}$
	iTrans	$14.21_{\pm0.13}$	$23.42_{\pm0.14}$	$26.79_{\pm0.71}$	$21.22_{\pm0.11}$	$33.04_{\pm0.17}$	$19.36_{\pm0.25}$	$144.04_{\pm0.82}$	$257.27_{\pm2.16}$	$113.97_{\pm6.78}$
	DLinear	$15.73_{\pm0.06}$	$26.23_{\pm0.04}$	$28.55_{\pm0.98}$	$23.58_{\pm0.18}$	$36.54_{\pm0.22}$	$19.60_{\pm0.33}$	$182.20_{\pm1.03}$	$318.13_{\pm1.10}$	$187.17_{\pm5.55}$
	STAEFormer	$12.26_{\pm0.11}$	$\underline{20.58}_{\pm0.12}$	$\underline{23.06}_{\pm0.64}$	$19.38_{\pm0.94}$	$30.64_{\pm1.20}$	$16.97_{\pm0.95}$	$96.12_{\pm1.67}$	$190.33_{\pm2.83}$	$66.43_{\pm2.35}$
Times-series Forecasting Foundation Model	Chronos-2 (ZS)	$15.27_{\pm0.00}$	$26.98_{\pm0.00}$	$25.83_{\pm0.00}$	$21.58_{\pm0.00}$	$35.22_{\pm0.00}$	$17.62_{\pm0.00}$	$227.74_{\pm0.00}$	$440.18_{\pm0.00}$	$120.67_{\pm0.00}$
	Chronos-2 (FT)	$13.31_{\pm0.01}$	$23.66_{\pm0.03}$	$22.80_{\pm0.03}$	$18.94_{\pm0.04}$	$31.89_{\pm0.12}$	$15.25_{\pm0.01}$	$160.20_{\pm0.68}$	$323.47_{\pm2.06}$	$78.26_{\pm0.28}$
	TimesFM2.5 (ZS)	$15.91_{\pm0.00}$	$27.86_{\pm0.00}$	$26.71_{\pm0.00}$	$23.49_{\pm0.00}$	$37.69_{\pm0.00}$	$19.32_{\pm0.00}$	$266.77_{\pm0.00}$	$487.06_{\pm0.00}$	$160.45_{\pm0.00}$
	TimesFM2.5 (FT)	$14.16_{\pm0.01}$	$25.36_{\pm0.04}$	$23.79_{\pm0.01}$	$20.92_{\pm0.02}$	$35.20_{\pm0.02}$	$16.65_{\pm0.02}$	$215.22_{\pm0.33}$	$420.65_{\pm0.52}$	$124.25_{\pm0.83}$
	Moirai-2.0 (ZS)	$16.19_{\pm0.00}$	$28.00_{\pm0.00}$	$26.15_{\pm0.00}$	$24.52_{\pm0.00}$	$38.87_{\pm0.00}$	$19.11_{\pm0.00}$	$294.05_{\pm0.00}$	$545.09_{\pm0.00}$	$137.97_{\pm0.00}$
	Moirai-2.0 (FT)	$13.95_{\pm0.01}$	$24.51_{\pm0.03}$	$24.18_{\pm0.01}$	$20.68_{\pm0.03}$	$34.42_{\pm0.09}$	$16.78_{\pm0.02}$	$142.36_{\pm0.28}$	$303.78_{\pm0.69}$	$62.69_{\pm0.22}$
Ours	A2TTA(STAE)	$\mathbf{11.10}_{\pm0.04}$	$\mathbf{18.87}_{\pm0.07}$	$\mathbf{21.46}_{\pm0.18}$	$\mathbf{16.70}_{\pm0.11}$	$\mathbf{27.46}_{\pm0.15}$	$\mathbf{14.89}_{\pm0.05}$	$\mathbf{67.89}_{\pm0.32}$	$\mathbf{147.15}_{\pm0.64}$	$\mathbf{54.66}_{\pm0.33}$
	A2TTA(OLAN)	$\underline{12.19}_{\pm0.04}$	$20.72_{\pm0.07}$	$23.43_{\pm0.27}$	$\underline{18.15}_{\pm0.05}$	$\underline{29.59}_{\pm0.07}$	$\underline{16.41}_{\pm0.28}$	$\underline{86.78}_{\pm0.24}$	$\underline{176.97}_{\pm0.84}$	$\underline{66.01}_{\pm1.49}$

6.4 Delayed-Label Sensitivity

Figure 5 uses all ten networks. For PEMS12, Avg-MAE rises by 0.68% at 524 steps, 0.95% at 1,036 steps, and 0.65% without online labels. Across ten networks, the median increases at 524 steps are 0.62% for A^{2}TTA(OLAN) and 0.63% for A^{2}TTA(STAE); without online labels, they are 0.44% and 0.80%. The sweep changes only the release delay under the one-window protocol.

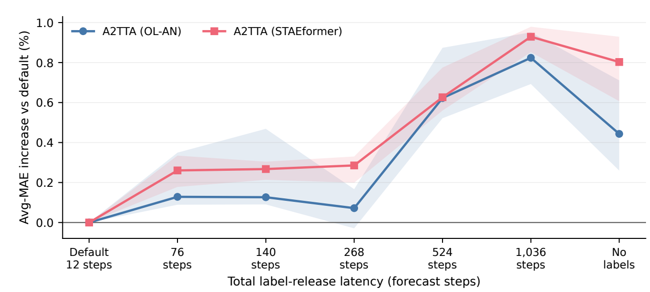

Figure 5: Delayed-label sensitivity; bands show the interquartile range.

6.5 Ablation Study

Full A^{2}TTA has the lowest Avg-MAE in seven of eight settings in Figure 6. On Online-AN and STAEFormer, respectively, reverting to the frozen backbone raises error by 6.5% and 6.7%, replacing FiLM with a static affine transform by 3.7% and 4.7%, and removing the local clone by 1.2% and 1.6%. Freezing FiLM after warm-up is worse in seven settings but better on STAEFormer with PEMS06. Appendix A.8 defines the variants and gives the full results.

6.6 Newly Added Sensors / High-Drift Periods

For the metadata-defined new-sensor cohort, A^{2}TTA(STAE) is best in five of six panels in Figure 7: all metrics on PEMS07 and MAE and MAPE on TFNSW. A^{2}TTA(OLAN) has lower TFNSW RMSE (177.20 versus 178.38). A^{2}TTA(STAE) reduces MAE over frozen

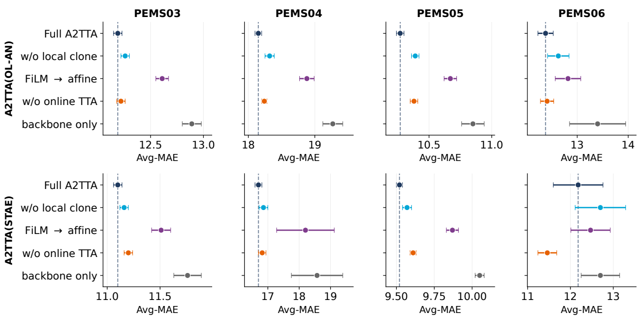

Figure 6: Component knockouts on four PEMS datasets and two frozen backbones.

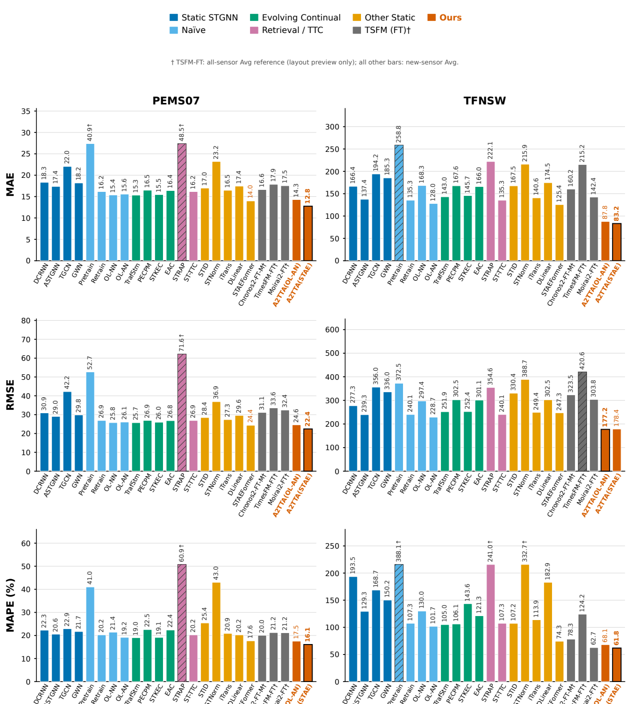

Figure 7: New-sensor performance.

STAEFormer by 8.8% on PEMS07 and 33.7% on TFNSW. Appendix A.3.2 gives all ten datasets.

High-drift periods. Figure 8 groups paired dataset-year results by drift severity. A^{2}TTA improves both backbones in every quartile, with all 95% bootstrap intervals above zero. Gains peak in the highest sensor-churn quartile at 9.0% over Online-AN and 7.9% over STAEFormer. This legacy-batched diagnostic shows an association, not a causal effect; Appendix A.4 gives the other metrics and aggregation.

6.7 Case Study

Figure 9 examines a high-variance PEMS06-2015 new sensor selected without using either method's errors. Over seeds 51 and

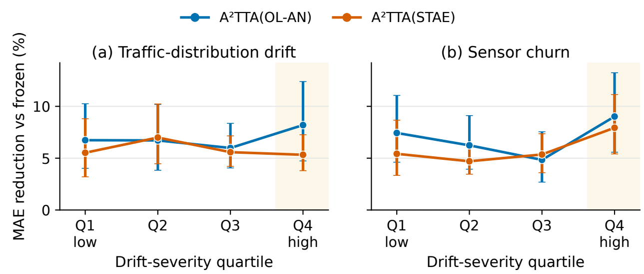

Figure 8: Performance across drift-severity quartiles.

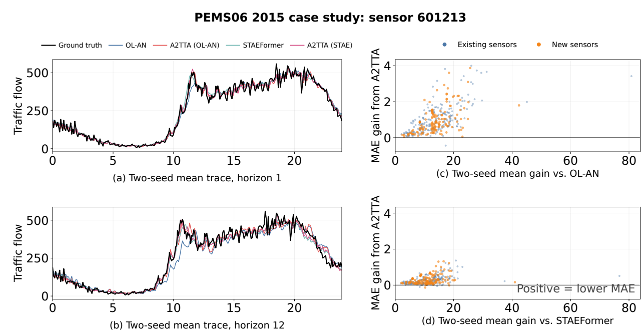

Figure 9: Two-seed case study on PEMS06-2015.

52, A^{2}TTA tracks the morning rise more closely, especially at 60 minutes. Across 366 sensors, it beats Online-AN on 99.7% and STAEFormer on 98.6%, reducing mean per-sensor MAE by 7.29% and 2.62%. Appendix A.9 gives selection and cohort details.

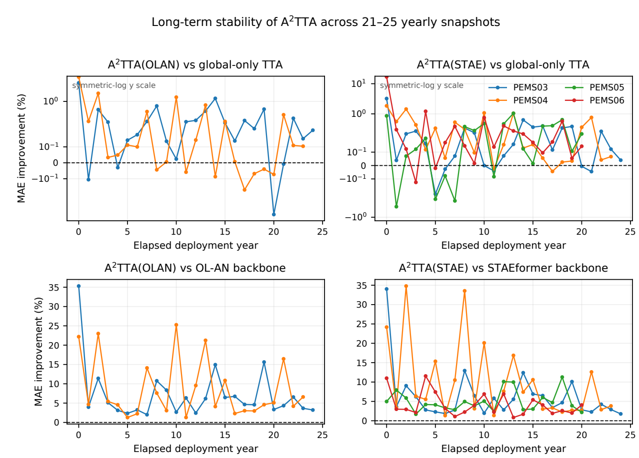

Figure 10: Long-term five-seed stability.

6.8 Long-Term Adaptation

The legacy-batched five-seed analysis in Figure 10 covers six streams over 21 to 25 years. Gains over the frozen backbone are positive every year; gains over global-only TTA are smaller and less uniform. Appendix A.5 gives a pre-specified single-seed check on all ten datasets.

Deployment cost. We profile 233 yearly snapshots on an H200. A^{2}TTA(STAE) takes 11.60 ms per window versus 10.48 ms for frozen STAEFormer and raises peak allocated memory from 10,021 to 10,552 MiB. It updates 33.2K calibrator parameters on average versus 2.06M backbone parameters. A^{2}TTA(OLAN) takes 5.35 ms versus 4.30 ms. Appendix A.6 gives the three-repeat protocol and additional measurements.

7 Limitations

A^{2}TTA assumes that labels eventually arrive and uses a labeled training partition to warm up the calibrator for each yearly snapshot. The main setting therefore does not cover fully label-free deployment. The new-sensor cohort contains sensors observed in the current year's training partition and should not be interpreted as zero-shot node commissioning. A warm-up-free control in Appendix A.6 measures this dependence explicitly. The evaluation is limited to univariate traffic flow, yearly graph changes, and two forecasting backbones. Absolute latency and memory also depend on the deployment hardware and graph size.

8 Conclusion

We studied forecasting on evolving sensor graphs as causal delayed-feedback adaptation and introduced A^{2}TTA, an output calibrator that keeps each matched yearly frozen forecaster. Its expandable FiLM module supports node growth, while persistent global and disposable local states address long-lived and context-specific shifts. The anchored global state learns persistent shifts from released labels, while a disposable local clone specializes to the current context and is discarded after prediction. Under one-windows chronological evaluation on nine EvoXXLTraffic districts and TFNSW, A^{2}TTA improves both matched host forecasters on every dataset. Relative to the corresponding backbones, mean MAE reductions at horizons 3, 6, and 12 range from 9.7% to 13.2%. A^{2}TTA also improves the metadata-defined new-sensor cohorts. With STAEFormer, it updates 33.2K calibrator parameters on average while leaving the 2.06M-parameter backbone unchanged, adding 1.12 ms per test window.

Acknowledgments

This work was supported by the ARC Centre of Excellence for Automated Decision-Making and Society (CE200100005). We acknowledge the resources and services provided by the National Computational Infrastructure (NCI), which is supported by the Australian Government. This research is also partially supported by the ARC Training Centre for Whole Life Design of Carbon Neutral Infrastructure (IC230100015).

References

[1] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time series analysis: forecasting and control.* John Wiley & Sons.

[2] Pinlong Cai, Yunpeng Wang, Guangquan Lu, Peng Chen, Chuan Ding, and Jianping Sun. 2016. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies* 62 (2016), 21–34.

[3] Chao Chen, Karl Petty, Alexander Skabardonis, Pravin Varaiya, and Zhanfeng Jia. 2001. Freeway performance measurement system: mining loop detector data. *Transportation research record* 1748 (2001), 96–102.

[4] Wei Chen and Yuxuan Liang. 2025. Expand and compress: Exploring tuning principles for continual spatio-temporal graph forecasting. In *International Conference on Learning Representations*, Vol. 2025. 81631–81656.

[5] Wei Chen and Yuxuan Liang. 2025. Learning with calibration: Exploring test-time computing of spatio-temporal forecasting. *Advances in Neural Information Processing Systems* 38 (2025), 155895–155929.

[6] Xu Chen, Junshan Wang, and Kunqing Xie. 2021. 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, IJCAI-21*, Zhi-Hua Zhou (Ed.). International Joint Conferences on Artificial Intelligence Organization, 3620–3626. doi:10.24963/ijcai.2021/498 Main Track.

[7] Zheng Dong, Renhe Jiang, Haotian Gao, Hangchen Liu, Jinliang Deng, Qingsong Wen, and Xuan Song. 2024. Heterogeneity-informed meta-parameter learning for spatiotemporal time series forecasting. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining.* 631–641.

[8] Pengxin Guo, Pengrong Jin, Ziyue Li, Lei Bai, and Yu Zhang. 2024. Online Test-Time Adaptation of Spatial-Temporal Traffic Flow Forecasting. *arXiv preprint arXiv:2401.04148* (2024).

[9] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 922–929.

[10] Jiawei Jiang, Chengkai Han, Wayne Xin Zhao, and Jingyuan Wang. 2023. Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 4365–4373.

[11] Renhe Jiang, Du Yin, Zhaonan Wang, Yizhuo Wang, Jiewen Deng, Hangchen Liu, Zekun Cai, Jinliang Deng, Xuan Song, and Ryosuke Shibasaki. 2021. Dl-traff: Survey and benchmark of deep learning models for urban traffic prediction. In *Proceedings of the 30th ACM international conference on information & knowledge management.* 4515–4525.

[12] Selvaraj Vasantha Kumar. 2017. Traffic flow prediction using Kalman filtering technique. *Procedia Engineering* 187 (2017), 582–587.

[13] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations.*

[14] Aoyu Liu and Yaying Zhang. 2026. A General Spatio-Temporal Backbone with Scalable Contextual Pattern Bank for Urban Continual Forecasting. In *The Fourteenth International Conference on Learning Representations.*

[15] Hangchen Liu, Zheng Dong, Renhe Jiang, Jiewen Deng, Jinliang Deng, Quanjun Chen, and Xuan Song. 2023. Spatio-temporal adaptive embedding makes vanilla transformer sota for traffic forecasting. In *Proceedings of the 32nd ACM international conference on information and knowledge management.* 4125–4129.

[16] Tengfei Lyu, Weijia Zhang, Jinliang Deng, and Hao Liu. 2025. Autostf: Decoupled neural architecture search for cost-effective automated spatio-temporal forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1.* 985–996.

[17] Haoyuan Ma, Mintao Zhou, Xiaodong Ouyang, Du Yin, Renhe Jiang, and Xuan Song. 2022. Forecasting Regional Multimodal Transportation Demand with Graph Neural Networks: An Open Dataset. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC).* IEEE, 3263–3268. doi:10.1109/ITSC55140.2022.9922512

[18] Minbo Ma, Kai Tang, Huan Li, Fei Teng, Dalin Zhang, and Tianrui Li. 2025. Beyond fixed variables: Expanding-variate time series forecasting via flat scheme and spatio-temporal focal learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2.* 2054–2065.

[19] Florin Schimbinschi, Luis Moreira-Matias, Vinh Xuan Nguyen, and James Bailey. 2017. Topology-regularized universal vector autoregression for traffic forecasting in large urban areas. *Expert Systems with Applications* 82 (2017), 301–316.

[20] Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu. 2022. 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.* 4454–4458.

[21] Zezhi Shao, Zhao Zhang, Wei Wei, Fei Wang, Yongjun Xu, Xin Cao, and Christian S Jensen. 2022. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *Proceedings of the VLDB Endowment* 15 (2022), 2733–2746.

[22] Shun-Yao Shih, Fan-Keng Sun, and Hung-yi Lee. 2019. Temporal pattern attention for multivariate time series forecasting. *Machine Learning* 108, 8 (2019), 1421–1441.

[23] Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, et al. 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499* 12, 1 (2016).

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

[25] Binwu Wang, Yudong Zhang, Xu Wang, Pengkun Wang, Zhengyang Zhou, Lei Bai, and Yang Wang. 2023. 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*. 2223–2232.

[26] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. 2021. Tent: Fully Test-Time Adaptation by Entropy Minimization. In *International Conference on Learning Representations*.

[27] Billy M Williams and Lester A Hoel. 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of transportation engineering* 129, 6 (2003), 664–672.

[28] Chun-Hsin Wu, Jan-Ming Ho, and Der-Tsai Lee. 2004. Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems* 5, 4 (2004), 276–281.

[29] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. 1907–1913.

[30] Mingxing Xu, Wenrui Dai, Chunmiao Liu, Xing Gao, Weiyao Lin, Guo-Jun Qi, and Hongkai Xiong. 2020. Spatial-temporal transformer networks for traffic flow forecasting. *arXiv preprint arXiv:2001.02908* (2020).

[31] Du Yin, Jinliang Deng, Shuang Ao, Zechen Li, Hao Xue, Arian Prabowo, Renhe Jiang, Xuan Song, and Flora Salim. 2024. Enhancing Spatio-temporal Quantile Forecasting with Curriculum Learning: Lessons Learned. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*. 42–53.

[32] Du Yin, Renhe Jiang, Jiewen Deng, Yongkang Li, Yi Xie, Zhongyi Wang, Yifan Zhou, Xuan Song, and Jedi S Shang. 2023. MTMGNN: Multi-time multi-graph neural network for metro passenger flow prediction. *GeoInformatica* 27, 1 (2023), 77–105.

[33] Du Yin, Hao Xue, Arian Prabowo, Shuang Ao, and Flora Salim. 2026. From XXLTraffic to EvoXXLTraffic: Scaling Traffic Forecasting to Sensor-Evolving Networks. *arXiv preprint arXiv:2605.29768* (2026).

[34] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3634–3640.

[35] Haoyu Zhang, Hao Miao, Xinke Jiang, Yuchen Fang, and Yifan Zhang. 2025. Strap: Spatio-temporal pattern retrieval for out-of-distribution generalization. *Advances in Neural Information Processing Systems* 38 (2025), 118006–118041.

[36] Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. 2019. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems* 21, 9 (2019), 3848–3858.

[37] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 1234–1241.

A Appendix

A.1 Structural Evolution Details

Node counts alone do not characterize this evolution. Figure 11 therefore tracks average node degree and graph density for the yearly snapshots. Both quantities change non-monotonically, because sensor additions and removals alter local neighborhoods and the set of spatial connections. Successive snapshots consequently differ in both dimension and connectivity, rather than being simple zero-padded extensions of the preceding graph.

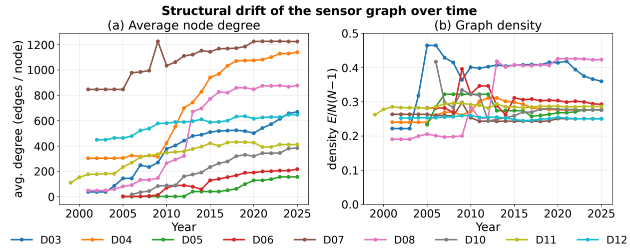

Figure 11: Structural evolution of the sensor graphs. Average node degree and graph density vary across yearly snapshots, showing that graph evolution extends beyond node accumulation.

A.2 Extended Related Work

A.2.1 Spatio-Temporal Traffic Forecasting. Traffic forecasting, including traffic flow, speed, and congestion status, has always been one of the most important issues in computational urban and intelligent transportation systems. Early traffic forecasting methods primarily relied on statistical modeling and traditional machine learning methods, such as Historical Average (HA), Autoregressive Integrated Moving Average (ARIMA) [1], Seasonal ARIMA [27], Kalman filtering [12], Vector Autoregression (VAR) [19], Support Vector Regression (SVR) [28], and k-Nearest Neighbors (kNN) [2]. These methods typically treat traffic sequences as one-dimensional or low-dimensional time series and make predictions by modeling trends, periodicity, and short-term autocorrelation in historical observations. Among them, HA uses historical averages over the same period as predictions, making it simple and efficient but difficult to adapt to sudden changes. ARIMA and its variants can characterize linear temporal dependence and periodic patterns but usually rely on stationarity assumptions. Kalman filtering is suitable for online state estimation and short-term forecasting, while SVR and kNN improve predictive capability through nonlinear mapping or similar-pattern matching. Nevertheless, these methods mostly rely on manual feature design and struggle to characterize complex nonlinear temporal variations and spatial dependencies within road networks simultaneously.

Deep traffic forecasting has used convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), Transformers, and, more recently, state-space models. Early temporal models used long short-term memory (LSTM) or gated recurrent unit (GRU) networks to capture long-range dependencies. CNNs [22] and temporal convolutional networks (TCNs) [23] later improved parallelism while retaining local temporal structure. Graph models then made road-network dependencies explicit. Representative methods include T-GCN [36], STGCN [34], and DCRNN [13]. T-GCN places graph convolution inside GRU units, STGCN combines graph and temporal convolutions, and DCRNN models directed propagation with diffusion convolution and an encoder-decoder architecture. GNNs have also been applied to regional multimodal demand forecasting: Ma et al. [17] release an NYC taxi-bike dataset and study multi-source, multi-graph, and meta-information-enhanced STGNNs. Our setting instead concerns sensor graphs whose nodes and connectivity change over time.

Traffic forecasting models have since developed toward adaptive graph learning, attention mechanisms, and large-scale sequence modeling. Graph WaveNet [29], for example, reduces reliance on predefined adjacency matrices by learning adaptive spatial dependencies. ASTGCN [9], GMAN [37], and STTN [30] introduce spatio-temporal attention to dynamically model relationships among sensors and time steps. Transformer-based models are now widely used because of their long-sequence modeling capability, while state-space models offer linear-complexity alternatives for large-scale forecasting. This development has produced many strong models, including MTMGNN [32], STQCL [31], STAEFormer [15], STID [20], HIMNet [7], PDFormer [10], and AutoSTF [16].

Despite their progress on static benchmarks, most existing methods assume that the sensor set and road-network topology remain unchanged between training and testing. In real deployments, however, sensor networks expand over time, new sensors are commissioned, and connectivity and traffic patterns drift. A deployed forecaster therefore faces not merely a longer time series but a continuously evolving traffic graph. The resulting node growth, topology change, and temporal distribution shift challenge models built under a fixed-graph assumption.

A.2.2 Online and Continual Traffic Forecasting. Long-term traffic-system deployment has consequently motivated online and continual forecasting. New sensors are repeatedly added, while traffic patterns at existing nodes evolve over time. An effective model must therefore acquire new spatio-temporal patterns without catastrophically forgetting previously learned knowledge.

TrafficStream [6] was among the first methods to formalize network expansion and pattern evolution as a streaming continual-learning problem. It uses local subgraphs to integrate neighborhood information for new sensors, Jensen–Shannon divergence to detect drift at existing nodes, and historical replay with parameter smoothing to mitigate catastrophic forgetting. Following an explicit-memory approach, PECPM [25] maintains a library of representative spatio-temporal patterns and updates only new or conflicting nodes when a new graph arrives. STKEC [24] selects influential old nodes to construct local subgraphs, combines them with a memory bank for long-term periodic patterns, and preserves prior knowledge through parameter-distance regularization. To reduce update cost, EAC [4] uses a continual prompt pool to represent knowledge from different stages and follows an expansion–compression strategy that avoids large-scale updates of the whole model. STBP [14] freezes a general spatio-temporal backbone and expands only a scalable contextual pattern library, balancing stability and plasticity. STEV [18] treats new sensors as a variable-expansion problem and uses a flattening scheme to improve robustness to varying input scales. STRAP [35] constructs an external pattern library and retrieves similar patterns as prompts at inference time. ST-TTC [5] performs test-time calibration with frequency-domain alignment and streaming memory queues, but primarily addresses temporal drift within a fixed node set.

Unlike methods that update substantial model components or depend on large replay memories, we adapt only a residual calibrator attached to a frozen forecaster. Records remain pending until their complete forecast horizon has elapsed. Released records update the global calibrator, and the local clone weights the retained pool by temporal phase, traffic-pattern similarity, and recency. This design handles node growth, topology change, and traffic drift without updating the deployed forecasting backbone.

A.3 Complete Main Results

Average performance on the remaining districts. Tables 3 and 4 add PEMS05 to PEMS08 and PEMS10 to PEMS12 to the representative results in Table 2. Across all ten networks, A^{2}TTA(STAE) is the best non-foundation method in all 30 Avg dataset-metric comparisons. Averaged equally across datasets, it reduces MAE, RMSE, and MAPE over STAEFormer by 11.0%, 8.5%, and 9.7%, respectively. A^{2}TTA (OLAN) improves Online-AN by 12.8%, 9.7%, and 12.2% on the same metrics.

Table conventions. In Tables 3 and 4, each row is one baseline and each dataset reports MAE, RMSE, and MAPE in percent. Avg denotes the mean of cumulative-horizon scores over prediction lengths 1–12. Entries are mean±standard deviation (SD) where available and deterministic zero-shot references are single-run and shown with ±0.00. Chronos-2 uses multivariate graph-grouped inference, with per-year low-rank adaptation (LoRA) for its FT variant. Bold and underlined entries mark the best and second-best non-foundation methods, respectively. Foundation-model rows are excluded from highlighting.

Per-horizon performance and TSFM comparison. For compactness, the preceding tables average cumulative errors over prediction lengths 1 to 12. Tables 5 and 6 report horizons 3, 6, and 12, where A^{2}TTA-S and A^{2}TTA-O denote A^{2}TTA(STAE) and A^{2}TTA(OLAN), respectively. A^{2}TTA-S is the best non-foundation method in 119 of the 120 reported cells. The exception is PEMS06 MAE at horizon 3, where A^{2}TTA-O is lower. Its MAE reductions over STAEFormer average 12.1%, 10.8%, and 9.7% at horizons 3, 6, and 12. A^{2}TTA-O improves Online-AN by 13.2%, 12.7%, and 12.3%.

Across the three horizons and Avg rows, A^{2}TTA-S also outperforms the strongest TSFM reference in 113 of 120 cells, including all 30 Avg cells. Fine-tuning remains important for the foundation models: Chronos-2, TimesFM 2.5, and Moirai-2.0 improve over their zero-shot counterparts in all 30 Avg dataset-metric comparisons, with mean relative reductions of 13.5%, 10.9%, and 15.4%, respectively.

Per-horizon table conventions. In Tables 5 and 6, A^{2}TTA-S and A^{2}TTA-O are evaluated over five seeds, and MAPE is reported in percent. Compact headers pair zero-shot and fine-tuned variants: Chro2-Z/F, TimesF-Z/T, Moira1-Z/T, and Moira2-Z/T. Deterministic zero-shot references are single-run and shown with ±0.00, whereas fine-tuned references report mean±SD over three seeds. Chronos-2 uses multivariate graph-grouped inference and per-year LoRA fine-tuning for its FT variant. TSFM columns are excluded from best/second-best highlighting. Pretrain fixes the first-year model, Retrain starts from scratch each year, Online-NN updates new sensors, and Online-AN updates all active sensors.

A.3.1 Complete Per-Horizon Results. The main text uses PEMS03 and TFNSW for a readable per-horizon comparison. Here, we report the complete 12-step profiles for all nine EvoXXLTraffic districts and TFNSW. The curves use per-step rather than cumulative errors. The

Table 3: Average results on PEMS05–08 (mean±standard deviation).

Category	Baseline	PEMS05			PEMS06			PEMS07			PEMS08		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Static STGNN Backbones	DCRNN	11.69±0.07	18.34±0.09	26.98±0.74	14.87±0.31	23.29±0.43	24.66±0.77	19.39±0.28	32.62±0.30	31.09±1.27	15.50±0.13	25.20±0.19	23.31±0.41
	ASTGNN	11.69±0.04	18.52±0.08	25.31±0.60	15.52±0.16	24.04±0.28	23.41±0.46	19.56±0.05	33.05±0.13	29.95±0.26	15.22±0.05	25.23±0.10	20.85±0.39
	TGCN	11.94±0.10	19.43±0.16	26.83±0.59	15.52±0.46	25.38±0.56	24.06±0.40	20.62±0.41	37.75±0.75	32.43±1.02	15.86±0.17	27.85±0.40	22.13±0.71
	GWN	12.57±0.13	19.49±0.18	27.19±0.86	16.16±0.54	24.84±0.65	24.76±0.81	19.51±0.99	32.22±1.35	31.81±2.36	15.69±0.16	25.27±0.29	21.90±0.74
Naïve Schemes	Pretrain	24.94±0.99	34.77±1.06	116.10±7.29	78.66±1.39	97.96±2.03	79.40±1.33	40.30±2.05	54.10±1.93	49.24±2.54	52.25±2.71	62.21±3.13	57.09±1.58
	Retrain	11.31±0.04	18.03±0.07	24.42±0.41	14.37±0.34	22.39±0.39	22.54±0.33	18.02±0.27	30.73±0.23	27.64±1.35	14.34±0.07	23.85±0.09	20.58±0.25
	Online-NN	15.98±0.52	26.35±0.67	29.06±0.76	27.64±0.65	38.29±1.20	32.16±0.38	22.72±0.59	35.82±0.68	35.08±1.56	17.30±0.05	26.86±0.13	24.09±0.35
	Online-AN	11.10±0.07	17.63±0.13	23.49±0.44	14.02±0.21	21.68±0.28	21.93±0.52	17.67±0.31	30.13±0.32	26.68±1.27	13.92±0.10	23.20±0.15	19.28±0.22
Evolving Graph	TrafficStream	11.32±0.13	18.04±0.14	24.56±0.93	14.46±0.34	22.41±0.47	22.43±0.49	17.54±0.31	30.04±0.29	26.67±2.19	14.04±0.13	23.35±0.18	20.96±1.33
	PECPM	11.67±0.31	18.50±0.44	25.13±0.85	18.02±3.19	29.03±6.07	26.88±5.50	19.23±1.47	32.52±2.46	27.85±1.72	15.58±0.77	25.25±1.04	22.26±1.64
	STKEC	11.48±0.16	18.26±0.21	25.11±0.78	14.90±0.24	23.19±0.34	24.14±0.58	17.07±1.26	29.33±1.84	25.66±3.04	14.47±0.38	23.89±0.50	21.55±1.34
	EAC	16.35±3.53	23.77±3.62	61.09±37.33	21.10±4.19	31.23±4.99	36.63±16.01	18.26±0.27	30.41±0.37	28.73±0.39	14.76±0.28	23.76±0.37	26.61±1.73
Retrieval TTC	STRAP	11.33±0.04	17.97±0.04	25.25±0.88	14.44±0.46	22.53±0.49	22.60±0.66	17.81±0.16	30.37±0.18	26.88±0.66	14.15±0.03	23.54±0.04	19.96±0.41
	ST-TTC	11.22±0.04	17.89±0.07	24.33±0.41	14.25±0.34	22.21±0.38	22.42±0.33	17.74±0.27	30.37±0.23	27.40±1.33	14.16±0.07	23.63±0.08	20.45±0.25
Static Forecasting Backbones	STID	11.64±0.06	18.69±0.08	24.51±0.52	15.39±0.23	24.34±0.40	22.72±0.38	19.27±0.19	32.22±0.21	34.13±0.84	15.12±0.07	25.69±0.17	20.33±0.76
	STNorm	12.57±0.10	19.00±0.17	35.42±0.64	16.70±0.35	24.97±0.45	32.58±1.07	25.17±1.15	39.57±1.56	58.70±3.03	18.33±0.31	28.00±0.33	43.33±2.01
	iTrans	11.42±0.06	18.04±0.07	24.92±0.37	15.18±0.41	23.39±0.46	24.00±0.96	18.64±0.09	31.19±0.17	30.76±1.13	14.60±0.08	23.99±0.13	22.07±0.99
	DLinear	12.16±0.05	19.40±0.06	26.15±0.70	16.40±0.13	25.55±0.15	24.49±0.37	19.85±0.01	33.90±0.07	29.94±0.50	15.58±0.04	26.07±0.04	21.16±0.42
	STAEFormer	10.16±0.07	16.07±0.10	22.83±0.44	13.13±0.60	19.88±0.68	20.28±0.85	15.32±0.28	26.93±0.26	21.39±0.65	12.52±0.03	20.88±0.09	17.66±0.23
Times-series Forecasting Foundation Model	Chronos-2 (ZS)	12.67±0.00	20.54±0.00	25.64±0.00	15.59±0.00	25.65±0.00	22.27±0.00	18.67±0.00	34.18±0.00	22.66±0.00	16.19±0.00	27.58±0.00	20.06±0.00
	Chronos-2 (FT)	11.12±0.01	18.61±0.02	21.67±0.02	13.66±0.01	23.12±0.04	19.42±0.04	16.63±0.04	31.15±0.13	20.04±0.03	14.22±0.02	25.08±0.08	17.49±0.03
	TimesFM2.5 (ZS)	13.11±0.00	21.07±0.00	26.75±0.00	16.08±0.00	26.22±0.00	22.96±0.00	19.55±0.00	35.34±0.00	23.68±0.00	16.87±0.00	28.38±0.00	20.96±0.00
	TimesFM2.5 (FT)	11.85±0.01	19.92±0.01	22.50±0.02	14.28±0.01	24.21±0.01	20.06±0.01	17.86±0.02	33.57±0.03	21.24±0.02	14.90±0.01	26.18±0.02	18.26±0.00
	Moirai-2.0 (ZS)	13.26±0.00	21.14±0.00	25.93±0.00	16.33±0.00	26.23±0.00	22.46±0.00	20.58±0.00	36.42±0.00	23.57±0.00	17.64±0.00	29.31±0.00	20.70±0.00
	Moirai-2.0 (FT)	11.66±0.00	19.31±0.00	23.30±0.01	14.28±0.01	23.75±0.04	20.71±0.01	17.53±0.01	32.42±0.05	21.17±0.03	14.99±0.02	25.97±0.06	18.63±0.04
Ours	A2TTA(STAE)	**9.52**±0.02	**15.25**±0.04	**20.54**±0.10	**12.18**±0.58	**18.66**±0.79	**18.80**±0.45	**13.77**±0.07	**25.05**±0.08	**19.07**±0.20	**11.76**±0.03	**19.91**±0.06	**16.51**±0.07
	A2TTA(OLAN)	10.27±0.03	16.55±0.06	21.14±0.03	12.38±0.15	19.62±0.27	19.51±0.27	15.22±0.06	27.30±0.08	20.90±0.39	12.97±0.04	21.96±0.07	17.47±0.10

Table 4: Average results on PEMS10–12 (mean±standard deviation).

Category	Baseline	PEMS10			PEMS11			PEMS12		
		MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)
Static STGNN Backbones	DCRNN	12.26±0.16	20.13±0.26	31.65±0.35	19.69±0.30	33.66±0.34	31.28±0.86	16.16±0.17	28.46±0.25	30.90±0.78
	ASTGNN	12.45±0.05	20.80±0.06	30.22±0.32	20.00±0.16	34.32±0.27	26.69±0.39	15.98±0.10	28.49±0.18	27.32±0.53
	TGCN	12.43±0.08	21.49±0.14	30.86±0.22	21.35±0.31	40.11±0.53	30.37±0.32	17.13±0.08	33.43±0.19	30.68±0.38
	GWN	12.54±0.11	20.45±0.13	30.65±0.69	20.29±0.44	34.35±0.80	27.55±0.81	16.38±0.23	28.47±0.35	27.32±0.41
Naïve Schemes	Pretrain	81.39±3.16	84.52±3.24	755.06±28.01	85.43±3.52	116.95±3.38	82.12±18.72	16.41±0.21	27.96±0.24	28.48±0.81
	Retrain	11.85±0.04	19.66±0.03	30.30±0.22	18.77±0.13	32.21±0.15	27.67±0.81	14.91±0.05	26.54±0.06	26.88±0.52
	Online-NN	12.35±0.23	20.58±0.45	31.79±1.11	21.54±1.59	36.45±2.86	30.83±1.25	18.04±0.57	32.01±1.14	29.42±0.55
	Online-AN	11.60±0.09	19.22±0.17	29.80±0.34	18.29±0.13	31.40±0.13	25.39±0.75	14.54±0.06	25.97±0.07	25.48±0.28
Evolving Graph	TrafficStream	11.93±0.08	19.77±0.14	32.18±0.70	18.67±0.35	31.90±0.47	28.29±1.63	14.74±0.07	26.32±0.10	26.88±0.62
	PECPM	12.61±1.16	21.29±2.68	30.41±1.69	25.89±2.53	44.81±4.91	36.90±6.66	16.18±1.49	28.57±3.19	29.30±1.38
	STKEC	11.92±0.07	19.80±0.11	31.06±0.42	19.02±0.23	32.38±0.23	30.78±1.62	14.71±0.08	26.28±0.12	26.66±0.60
	EAC	13.60±1.76	21.08±1.43	59.52±40.27	25.76±4.14	41.08±5.52	47.34±12.72	14.98±0.30	26.18±0.40	29.71±2.65
Retrieval TTC	STRAP	11.79±0.09	19.58±0.10	30.34±0.58	18.60±0.15	31.88±0.20	28.85±1.00	14.88±0.14	26.48±0.17	27.09±1.22
	ST-TTC	11.76±0.04	19.53±0.04	30.14±0.22	18.52±0.13	31.83±0.15	27.53±0.80	14.74±0.05	26.28±0.05	26.77±0.51
Static Forecasting Backbones	STID	12.53±0.08	20.54±0.09	33.11±0.25	20.02±0.32	33.89±0.36	32.17±1.75	16.07±0.07	28.14±0.19	33.77±0.88
	STNorm	13.75±0.32	21.81±0.40	41.57±1.79	21.77±0.34	35.62±0.53	48.00±2.68	19.49±0.39	32.19±0.51	58.93±2.28
	iTrans	11.83±0.07	19.63±0.10	30.28±0.74	19.24±0.28	32.79±0.31	30.95±2.77	15.27±0.11	26.92±0.15	29.32±1.41
	DLinear	12.86±0.04	21.36±0.05	31.07±0.38	20.62±0.12	35.52±0.13	27.66±0.61	16.23±0.02	29.24±0.05	27.74±0.47
	STAEFormer	10.34±0.03	17.20±0.05	27.83±0.45	17.05±0.12	28.82±0.20	24.54±0.69	12.91±0.11	23.13±0.12	23.94±0.74
Times-series Forecasting Foundation Model	Chronos-2 (ZS)	13.30±0.00	22.54±0.00	30.51±0.00	20.04±0.00	37.05±0.00	25.36±0.00	16.67±0.00	30.85±0.00	26.63±0.00
	Chronos-2 (FT)	11.85±0.01	20.64±0.04	27.84±0.07	17.19±0.02	32.41±0.08	21.93±0.04	14.66±0.01	27.68±0.07	22.84±0.04
	TimesFM2.5 (ZS)	13.63±0.00	22.92±0.00	31.47±0.00	20.91±0.00	38.14±0.00	26.41±0.00	17.48±0.00	31.92±0.00	27.93±0.00
	TimesFM2.5 (FT)	12.27±0.01	21.11±0.01	28.62±0.03	18.32±0.02	34.43±0.07	23.11±0.03	15.69±0.02	29.74±0.05	24.31±0.02
	Moirai-2.0 (ZS)	14.18±0.00	23.64±0.00	30.73±0.00	21.24±0.00	38.56±0.00	25.78±0.00	17.90±0.00	32.32±0.00	27.26±0.00
	Moirai-2.0 (FT)	12.43±0.01	21.32±0.03	29.22±0.02	17.71±0.01	32.78±0.04	23.23±0.03	15.32±0.01	28.62±0.02	24.38±0.02
Ours	A2TTA(STAE)	**9.86**±0.03	**16.58**±0.03	**26.41**±0.04	**14.40**±0.07	**25.74**±0.10	**21.44**±0.25	**11.95**±0.05	**21.71**±0.08	**21.98**±0.24
	A2TTA(OLAN)	10.88±0.04	18.28±0.08	27.73±0.07	15.60±0.08	28.06±0.13	23.73±0.38	13.28±0.01	24.05±0.02	23.36±0.11

A^{2}TTA curves follow the one-window delayed-feedback protocol used by the main runs.

All sensors. As shown in fig. 12, A^{2}TTA(STAE) has the lowest error among the six plotted methods in all $10 \times 3 \times 12 = 360$ combinations. Averaged over dataset and horizon, it improves on frozen STAEFormer by 9.7% in MAE, 7.3% in RMSE, and 9.3% in MAPE.

A.3.2 Detailed Results on Newly Added Sensors.

Table 5: Main experimental results, part 1/2 (PEMS03, PEMS04, PEMS05, PEMS06, PEMS07, PEMS08, PEMS10, PEMS11; mean±std over seeds per dataset). Bold: best, underline: second best. A2TTA-S and A2TTA-O are measured over 5 seeds. MAPE values are reported in percent. Chronos-2 (ZS) and Chronos-2 (FT) use multivariate graph-grouped inference; FT uses per-year LoRA fine-tuning. TSFM columns are excluded from highlighting. Compact headers denote zero-shot/fine-tuned pairs: Chro2-Z/F for Chronos-2, TimesF-Z/T for TimesFM2.5, Moira1-Z/T for Moirai-MoE, and Moira2-Z/T for Moirai-2.0. Their deterministic zero-shot (ZS) references are single-run and shown with ±0.00 for visual consistency; fine-tuned (FT) references are reported as mean±std over 3 seeds.

		Static STGNN Backbones				Naïve Schemes				Evolving Graph				Retrieval TTC		Static Forecasting Backbones					TSFMs								Ours	
Metric	Len	DCRNN	ASTGNN	TGCN	GWN	Pretrain	Retrain	OL-NN	OL-AN	TrafStm	PECPM	STKEC	EAC	STRAP	ST-TTC	STID	STNorm	iTrans	DLinear	STAE	Chro2-Z	Chro2-F	TimesF-Z	TimesF-T	Moira1-Z	Moira1-T	Moira2-Z	Moira2-T	A2TTA-S	A2TTA-O
PEMS03																														
MAE	3	14.42±0.21	13.70±0.12	15.14±0.10	14.22±0.29	30.61±1.22	13.09±0.16	16.25±0.19	12.76±0.05	13.22±0.11	13.88±0.85	13.42±0.27	14.18±0.31	13.10±0.12	12.89±0.15	14.28±0.15	16.46±0.42	13.16±0.13	14.17±0.07	11.81±0.11	12.53±0.00	11.73±0.01	13.20±0.00	12.13±0.01	14.50±0.00	16.03±0.64	13.66±0.00	12.16±0.00	**10.57±0.03**	11.18±0.03
	6	14.67±0.18	14.89±0.10	15.63±0.10	15.08±0.23	31.21±1.19	14.01±0.14	17.32±0.16	13.68±0.05	14.17±0.17	14.76±0.81	14.28±0.20	15.64±0.48	14.00±0.12	13.82±0.13	14.89±0.14	17.86±0.45	14.05±0.14	15.45±0.06	12.22±0.10	14.74±0.00	13.09±0.01	15.55±0.00	13.87±0.01	17.22±0.00	18.39±0.59	15.87±0.00	13.70±0.01	**11.08±0.03**	12.10±0.04
	12	16.27±0.18	17.57±0.08	17.27±0.11	17.24±0.30	32.66±0.99	15.92±0.14	19.41±0.20	15.46±0.09	16.10±0.35	16.57±0.78	16.24±0.18	18.65±0.79	15.86±0.15	15.74±0.13	16.71±0.13	21.02±0.59	15.95±0.13	18.34±0.04	12.98±0.13	19.88±0.00	15.90±0.02	20.21±0.00	17.38±0.02	22.10±0.00	22.80±0.59	20.35±0.00	16.86±0.01	**11.93±0.05**	13.74±0.07
	Avg	14.99±0.18	15.15±0.10	15.88±0.10	15.33±0.26	31.37±1.11	14.16±0.15	17.47±0.16	13.80±0.05	14.32±0.20	14.90±0.81	14.47±0.22	15.91±0.48	14.15±0.13	13.98±0.14	15.14±0.14	18.20±0.47	14.21±0.13	15.73±0.06	12.26±0.11	15.27±0.00	13.31±0.01	15.91±0.00	14.16±0.01	17.59±0.00	18.68±0.59	16.19±0.00	13.95±0.01	**11.10±0.04**	12.19±0.04
RMSE	3	23.64±0.22	22.30±0.20	28.88±0.42	23.01±0.41	39.61±1.17	21.27±0.15	25.71±0.14	20.83±0.06	21.39±0.10	22.19±1.02	21.71±0.38	22.90±0.58	21.23±0.22	20.98±0.13	25.67±0.42	25.72±0.59	21.37±0.14	23.03±0.04	19.61±0.15	21.37±0.00	20.23±0.02	22.46±0.00	20.93±0.03	25.10±0.00	27.89±1.02	23.21±0.00	20.77±0.02	**17.79±0.06**	18.79±0.05
	6	24.33±0.17	24.71±0.16	29.72±0.40	24.70±0.33	41.30±1.09	23.20±0.12	28.00±0.10	22.70±0.10	23.35±0.22	24.08±0.95	23.56±0.28	25.90±0.88	23.13±0.21	22.93±0.11	26.68±0.38	27.99±0.61	23.14±0.15	25.74±0.04	20.52±0.12	25.87±0.00	23.16±0.04	27.14±0.00	24.74±0.04	30.33±0.00	32.52±0.89	27.41±0.00	23.96±0.03	**18.88±0.07**	20.62±0.07
	12	27.56±0.21	29.82±0.14	32.46±0.37	28.79±0.51	44.99±0.98	26.94±0.16	32.37±0.35	26.14±0.20	27.11±0.58	27.75±0.94	27.50±0.37	31.87±1.46	26.81±0.26	26.70±0.15	29.78±0.35	33.28±0.82	26.74±0.15	31.45±0.04	22.05±0.13	36.36±0.00	29.28±0.07	36.42±0.00	32.31±0.09	39.29±0.00	40.97±0.93	35.98±0.00	30.56±0.04	**20.50±0.08**	23.65±0.12
	Avg	24.90±0.19	25.15±0.17	30.12±0.40	25.15±0.39	41.64±0.99	23.46±0.14	28.29±0.08	22.89±0.11	23.60±0.28	24.32±0.98	23.89±0.34	26.40±0.88	23.37±0.22	23.18±0.13	27.10±0.38	28.58±0.65	23.42±0.14	26.23±0.04	20.58±0.12	26.98±0.00	23.66±0.03	27.86±0.00	25.36±0.04	30.92±0.00	32.99±0.89	28.00±0.00	24.51±0.03	**18.87±0.07**	20.72±0.07
MAPE (%)	3	28.08±0.88	24.72±0.40	28.13±0.82	26.65±0.67	50.48±2.93	25.11±0.64	28.43±0.78	24.05±0.22	26.49±1.13	26.87±2.27	26.48±0.70	27.89±1.36	25.03±0.27	24.96±0.64	25.31±0.62	49.62±2.31	25.56±0.78	26.28±0.91	22.52±0.64	22.59±0.00	21.07±0.03	23.53±0.00	21.74±0.02	23.11±0.00	23.82±0.36	23.23±0.00	22.00±0.01	**20.86±0.18**	22.34±0.26
	6	28.30±0.86	26.15±0.42	28.46±0.91	27.65±0.69	49.81±2.94	26.03±0.57	29.35±0.93	24.75±0.17	27.29±1.48	27.58±2.27	27.28±0.67	29.56±0.89	25.82±0.29	25.89±0.57	26.15±0.61	52.24±2.52	26.53±0.74	27.99±0.96	22.95±0.64	25.06±0.00	22.52±0.03	26.19±0.00	23.46±0.02	25.83±0.00	26.96±0.32	25.68±0.00	23.83±0.01	**21.37±0.18**	23.27±0.26
	12	30.21±0.81	29.78±0.50	30.89±0.96	30.44±0.82	49.61±2.26	28.34±0.49	31.64±1.66	26.62±0.25	29.70±2.58	29.44±2.39	29.57±0.77	34.17±1.00	28.11±0.57	28.20±0.48	28.48±0.63	58.49±3.04	28.91±0.60	32.53±1.08	24.00±0.63	31.43±0.00	25.67±0.02	31.92±0.00	27.12±0.01	31.32±0.00	32.00±0.24	30.84±0.00	27.77±0.03	**22.44±0.20**	25.20±0.30
	Avg	28.69±0.86	26.56±0.43	28.95±0.89	28.01±0.71	50.04±2.80	26.29±0.58	29.62±1.04	24.96±0.21	27.63±1.67	27.78±2.30	27.58±0.64	30.12±0.81	26.12±0.31	26.15±0.57	26.44±0.62	52.94±2.57	26.79±0.71	28.55±0.98	23.06±0.64	25.83±0.00	22.80±0.03	26.71±0.00	23.79±0.01	26.24±0.00	27.23±0.31	26.15±0.00	24.18±0.01	**21.46±0.18**	23.43±0.27
PEMS04																														
MAE	3	21.15±0.32	20.23±0.12	22.27±0.38	22.00±0.32	153.31±6.85	19.92±0.23	36.81±0.33	19.34±0.22	19.74±0.48	23.62±3.33	21.20±1.34	31.75±9.85	20.31±0.70	19.59±0.23	21.19±0.21	22.37±0.62	19.59±0.15	21.27±0.23	18.62±0.99	17.60±0.00	16.52±0.03	19.32±0.00	17.78±0.02	21.81±0.00	26.08±0.33	20.63±0.00	17.86±0.03	**15.81±0.07**	16.56±0.04
	6	21.75±0.26	22.06±0.11	23.22±0.37	22.97±0.28	153.21±6.57	21.35±0.19	38.21±0.32	20.78±0.26	21.18±0.50	24.90±3.16	22.41±1.06	33.61±9.31	21.63±0.64	21.04±0.19	21.96±0.17	23.63±0.63	20.94±0.11	23.12±0.17	19.33±0.95	20.75±0.00	18.59±0.04	22.92±0.00	20.46±0.02	26.14±0.00	30.01±0.17	24.04±0.00	20.27±0.03	**16.65±0.11**	18.01±0.04
	12	24.38±0.25	26.11±0.06	25.97±0.36	26.06±0.26	153.51±6.60	24.45±0.19	41.12±0.17	23.68±0.33	24.16±0.63	27.74±2.99	25.55±1.01	37.67±7.70	24.64±0.65	24.17±0.18	24.86±0.16	26.74±0.68	23.91±0.11	27.45±0.13	20.61±0.94	28.27±0.00	22.84±0.07	30.08±0.00	25.86±0.01	33.86±0.00	39.47±0.65	30.91±0.00	25.23±0.02	**18.06±0.18**	20.62±0.08
	Avg	22.22±0.27	22.45±0.09	23.60±0.37	23.41±0.27	153.33±6.62	21.64±0.20	38.46±0.27	21.01±0.26	21.43±0.53	25.16±3.18	22.78±1.14	34.02±9.03	21.93±0.67	21.33±0.20	22.43±0.18	23.99±0.62	21.22±0.11	23.58±0.18	19.38±0.94	21.58±0.00	18.94±0.04	23.49±0.00	20.92±0.02	26.72±0.00	31.16±0.33	24.52±0.00	20.68±0.03	**16.70±0.11**	18.15±0.05
RMSE	3	32.49±0.37	31.23±0.11	36.57±0.57	33.27±0.26	185.75±7.85	30.57±0.22	48.99±0.57	29.87±0.23	30.41±0.49	36.55±4.40	31.87±1.33	44.57±11.20	30.86±0.67	30.17±0.22	33.49±0.37	34.23±0.76	30.24±0.20	32.51±0.25	29.26±1.26	28.79±0.00	27.45±0.07	30.99±0.00	29.58±0.02	34.99±0.00	41.11±0.57	32.99±0.00	29.40±0.11	**25.94±0.11**	27.03±0.04
	6	33.93±0.28	34.42±0.10	38.28±0.55	35.06±0.24	186.05±7.49	33.18±0.20	51.49±0.51	32.43±0.30	33.00±0.54	38.80±4.09	34.23±1.05	47.58±10.31	33.36±0.61	32.80±0.20	34.98±0.29	36.16±0.79	32.65±0.16	35.90±0.21	30.61±1.20	33.78±0.00	31.20±0.10	36.72±0.00	34.39±0.01	41.80±0.00	47.11±0.31	38.04±0.00	33.63±0.09	**27.44±0.14**	29.43±0.07
	12	38.46±0.30	40.98±0.17	42.75±0.53	40.07±0.30	187.60±7.68	38.43±0.24	56.38±0.30	37.23±0.45	37.99±0.82	43.47±3.76	39.70±1.19	53.93±7.70	38.50±0.67	38.08±0.25	39.67±0.26	40.75±0.93	37.57±0.19	43.10±0.21	32.76±1.20	46.16±0.00	39.07±0.21	48.32±0.00	43.96±0.03	53.81±0.00	64.62±2.50	48.84±0.00	42.50±0.06	**29.73±0.25**	33.51±0.13
	Avg	34.58±0.32	34.96±0.12	38.82±0.55	35.69±0.22	186.37±7.58	33.58±0.22	51.85±0.46	32.73±0.31	33.35±0.60	39.17±4.12	34.78±1.18	48.20±9.87	33.78±0.64	33.21±0.22	35.63±0.31	36.65±0.80	33.04±0.17	36.54±0.22	30.64±1.20	35.22±0.00	31.89±0.12	37.69±0.00	35.20±0.02	42.74±0.00	49.27±0.58	38.87±0.00	34.42±0.09	**27.46±0.15**	29.59±0.07
MAPE (%)	3	19.95±0.41	17.57±0.43	19.35±0.26	18.96±0.52	82.38±3.00	17.87±0.35	25.59±0.52	16.67±0.26	17.63±0.50	21.16±3.83	20.61±1.82	37.87±18.81	18.37±0.71	17.74±0.35	17.69±0.30	26.18±1.42	18.27±0.21	17.68±0.31	16.44±0.94	14.71±0.00	13.71±0.01	16.27±0.00	14.69±0.02	16.45±0.00	17.91±0.07	16.23±0.00	14.90±0.01	**14.27±0.03**	15.20±0.22
	6	20.17±0.42	18.89±0.42	19.76±0.28	19.60±0.54	82.26±3.00	18.86±0.28	26.67±0.66	17.58±0.32	18.53±0.54	22.08±3.77	21.21±1.50	38.74±17.46	19.14±0.60	18.75±0.28	18.18±0.26	27.56±1.51	19.11±0.24	19.09±0.32	16.86±0.94	16.93±0.00	15.00±0.01	18.84±0.00	16.33±0.02	19.03±0.00	20.80±0.03	18.75±0.00	16.48±0.02	**14.81±0.04**	16.25±0.28
	12	21.91±0.40	22.29±0.42	21.86±0.33	21.79±0.59	82.01±3.01	21.34±0.25	29.13±1.08	19.79±0.43	20.92±0.83	24.23±3.51	23.44±1.32	42.12±15.90	21.54±0.39	21.24±0.24	20.14±0.21	31.05±1.68	21.23±0.34	22.99±0.46	17.90±1.02	22.62±0.00	17.78±0.01	24.24±0.00	19.78±0.02	23.86±0.00	26.17±0.16	23.60±0.00	19.89±0.03	**15.90±0.09**	18.36±0.38
	Avg	20.53±0.40	19.29±0.42	20.15±0.28	19.94±0.55	82.25±2.99	19.15±0.30	26.93±0.71	17.82±0.31	18.83±0.59	22.29±3.69	21.54±1.53	39.19±17.54	19.49±0.57	19.04±0.29	18.51±0.26	27.98±1.51	19.36±0.25	19.60±0.33	16.97±0.95	17.62±0.00	15.25±0.01	19.32±0.00	16.65±0.02	19.38±0.00	21.25±0.07	19.11±0.00	16.78±0.02	**14.89±0.05**	16.41±0.28
PEMS05																														
MAE	3	11.22±0.09	10.77±0.03	11.31±0.11	11.71±0.12	24.31±1.08	10.54±0.04	14.56±0.51	10.39±0.07	10.54±0.12	10.93±0.34	10.73±0.17	15.58±3.35	10.58±0.05	10.45±0.04	11.03±0.07	11.98±0.14	10.69±0.07	11.00±0.06	9.88±0.06	10.72±0.00	10.01±0.01	11.20±0.00	10.44±0.01	12.13±0.00	13.76±0.38	11.56±0.00	10.39±0.00	**9.20±0.02**	9.63±0.01
	6	11.49±0.06	11.55±0.04	11.79±0.09	12.34±0.13	24.77±0.99	11.20±0.04	15.76±0.54	11.00±0.07	11.21±0.12	11.55±0.31	11.36±0.14	16.21±3.54	11.21±0.03	11.11±0.04	11.48±0.05	12.44±0.11	11.30±0.06	11.96±0.05	10.12±0.07	12.23±0.00	10.94±0.01	12.84±0.00	11.62±0.01	14.04±0.00	15.48±0.25	12.97±0.00	11.46±0.00	**9.50±0.02**	10.21±0.03
	12	12.63±0.05	13.21±0.05	13.04±0.08	14.10±0.18	26.01±1.05	12.57±0.06	18.26±0.54	12.24±0.11	12.57±0.16	12.87±0.27	12.72±0.17	17.64±3.79	12.55±0.05	12.48±0.06	12.72±0.05	13.56±0.08	12.62±0.05	14.06±0.06	10.60±0.08	15.95±0.00	12.93±0.02	16.14±0.00	14.09±0.01	17.55±0.00	18.96±0.26	16.18±0.00	13.73±0.00	**10.01±0.04**	11.28±0.05
	Avg	11.69±0.07	11.69±0.04	11.94±0.10	12.57±0.13	24.94±0.99	11.31±0.04	15.98±0.52	11.10±0.07	11.32±0.13	11.67±0.31	11.48±0.16	16.35±3.53	11.33±0.04	11.22±0.04	11.64±0.06	12.57±0.10	11.42±0.06	12.16±0.05	10.16±0.07	12.67±0.00	11.12±0.01	13.11±0.00	11.85±0.01	14.34±0.00	15.75±0.28	13.26±0.00	11.66±0.00	**9.52±0.02**	10.27±0.03
RMSE	3	17.36±0.11	16.92±0.07	18.29±0.18	17.96±0.15	33.60±1.18	16.59±0.07	23.55±0.72	16.33±0.09	16.61±0.11	17.14±0.48	16.84±0.20	22.45±3.39	16.58±0.06	16.46±0.06	17.64±0.10	18.03±0.22	16.69±0.08	17.33±0.06	15.54±0.09	17.14±0.00	16.31±0.01	17.78±0.00	16.98±0.00	19.54±0.00	21.76±0.54	18.39±0.00	16.77±0.01	**14.67±0.03**	15.41±0.03
	6	18.03±0.07	18.31±0.07	19.20±0.16	19.12±0.18	34.52±1.07	17.84±0.06	25.94±0.75	17.49±0.13	17.87±0.12	18.33±0.43	18.07±0.18	23.58±3.68	17.79±0.05	17.71±0.06	18.43±0.07	18.81±0.16	17.86±0.07	19.09±0.05	16.02±0.10	19.78±0.00	18.19±0.01	20.61±0.00	19.44±0.01	22.68±0.00	24.70±0.40	20.68±0.00	18.85±0.01	**15.24±0.04**	16.48±0.05
	12	20.16±0.08	21.07±0.10	21.40±0.15	22.14±0.27	36.72±1.19	20.30±0.12	30.76±0.61	19.68±0.22	20.32±0.22	20.68±0.38	20.57±0.27	25.93±3.97	20.19±0.09	20.17±0.12	20.58±0.06	20.61±0.16	20.20±0.08	22.73±0.08	16.89±0.13	26.27±0.00	22.43±0.04	26.26±0.00	24.56±0.02	28.35±0.00	30.31±0.41	25.89±0.00	23.50±0.02	**16.13±0.06**	18.33±0.10
	Avg	18.34±0.09	18.52±0.08	19.43±0.16	19.49±0.18	34.77±1.06	18.03±0.07	26.35±0.67	17.63±0.13	18.04±0.14	18.50±0.44	18.26±0.21	23.77±3.62	17.97±0.04	17.89±0.07	18.69±0.08	19.00±0.17	18.04±0.07	19.40±0.06	16.07±0.10	20.54±0.00	18.61±0.02	21.07±0.00	19.92±0.01	23.12±0.00	25.07±0.41	21.14±0.00	19.31±0.00	**15.25±0.04**	16.55±0.06
MAPE (%)	3	26.28±0.86	23.31±0.62	25.81±0.68	25.60±0.88	114.78±6.27	23.06±0.42	27.15±0.88	22.28±0.43	23.18±0.88	23.88±0.92	23.69±0.81	59.98±37.51	23.90±0.94	22.97±0.42	23.40±0.55	33.91±0.52	23.53±0.44	23.50±0.68	22.26±0.46	21.96±0.00	20.12±0.01	23.23±0.00	20.62±0.02	22.62±0.00	23.96±0.28	22.69±0.00	21.26±0.01	**19.94±0.11**	20.14±0.06
	6	26.59±0.75	24.92±0.62	26.46±0.61	26.73±0.88	115.48±6.90	24.15±0.42	28.67±0.81	23.24±0.42	24.29±0.95	24.84±0.87	24.82±0.74	60.44±37.27	24.93±0.92	24.05±0.42	24.20±0.51	35.01±0.59	24.65±0.36	25.54±0.69	22.67±0.37	24.72±0.00	21.39±0.01	26.15±0.00	22.15±0.01	25.53±0.00	27.63±0.14	25.33±0.00	22.97±0.01	**20.45±0.09**	21.01±0.02
	12	28.62±0.63	28.72±0.60	28.84±0.49	30.08±0.80	118.51±8.78	26.77±0.42	32.35±0.78	25.61±0.55	26.89±1.00	27.37±0.73	27.67±0.80	63.79±37.30	27.61±0.75	26.67±0.42	26.50±0.51	38.11±0.99	27.28±0.38	30.73±0.73	23.86±0.49	32.01±0.00	24.30±0.03	32.54±0.00	25.58±0.01	31.30±0.00	33.24±0.06	31.29±0.00	26.70±0.01	**21.51±0.11**	22.78±0.04
	Avg	26.98±0.74	25.31±0.60	26.83±0.59	27.19±0.86	116.10±7.29	24.42±0.41	29.06±0.76	23.49±0.44	24.56±0.93	25.13±0.85	25.11±0.78	61.09±37.33	25.25±0.88	24.33±0.41	24.51±0.52	35.42±0.64	24.92±0.37	26.15±0.70	22.83±0.44	25.64±0.00	21.67±0.02	26.75±0.00	22.50±0.02	25.96±0.00	27.92±0.16	25.93±0.00	23.30±0.01	**20.54±0.10**	21.14±0.03
PEMS06																														
MAE	3	14.29±0.34	14.25±0.15	14.75±0.46	14.91±0.58	78.18±1.49	13.39±0.36	26.65±0.40	13.10±0.23	13.47±0.33	17.23±3.32	13.86±0.26	19.38±4.55	13.53±0.45	13.28±0.36	14.65±0.27	15.70±0.34	14.19±0.45	14.86±0.15	12.73±0.61	13.03±0.00	12.20±0.01	13.60±0.00	12.46±0.01	14.70±0.00	17.57±0.43	14.03±0.00	12.68±0.01	11.73±0.57	**11.50±0.14**
	6	14.60±0.30	15.36±0.17	15.33±0.46	15.92±0.53	78.77±1.35	14.24±0.34	27.53±0.65	13.91±0.21	14.33±0.33	17.90±3.21	14.75±0.23	20.83±4.24	14.30±0.46	14.13±0.34	15.20±0.21	16.52±0.33	15.02±0.42	16.16±0.13	13.09±0.61	15.11±0.00	13.47±0.01	15.79±0.00	14.02±0.01	17.33±0.00	19.73±0.46	16.01±0.00	14.08±0.01	**12.15±0.58**	12.31±0.16
	12	16.11±0.30	17.62±0.22	16.90±0.44	18.24±0.51	79.36±1.36	15.99±0.31	29.23±1.05	15.54±0.26	16.09±0.35	19.37±2.99	16.67±0.21	23.88±3.70	15.99±0.52	15.88±0.32	16.77±0.20	18.34±0.40	16.83±0.37	18.94±0.11	13.78±0.56	19.88±0.00	16.06±0.01	20.01±0.00	17.21±0.01	22.01±0.00	24.25±0.31	20.28±0.00	16.91±0.01	**12.88±0.58**	13.77±0.18
	Avg	14.87±0.31	15.52±0.16	15.52±0.46	16.16±0.54	78.66±1.39	14.37±0.34	27.64±0.65	14.02±0.21	14.46±0.34	18.02±3.19	14.90±0.24	21.10±4.19	14.44±0.46	14.25±0.34	15.39±0.23	16.70±0.35	15.18±0.41	16.40±0.13	13.13±0.60	15.59±0.00	13.66±0.01	16.08±0.00	14.28±0.01	17.67±0.00	20.04±0.40	16.33±0.00	14.28±0.01	**12.18±0.58**	12.38±0.15
RMSE	3	22.13±0.48	21.74±0.27	23.99±0.57	22.66±0.70	96.67±2.18	20.57±0.42	36.24±0.70	20.00±0.18	20.57±0.42	27.60±6.52	21.19±0.35	28.19±5.72	20.82±0.45	20.39±0.41	23.12±0.45	23.42±0.45	21.56±0.48	22.68±0.15	19.15±0.71	20.74±0.00	19.77±0.05	21.65±0.00	20.03±0.02	23.69±0.00	27.96±0.62	22.24±0.00	20.34±0.02	**17.89±0.81**	17.96±0.25
	6	22.84±0.42	23.84±0.29	25.08±0.56	24.49±0.62	98.00±2.04	22.23±0.39	38.11±1.19	21.54±0.29	22.27±0.46	28.90±6.09	22.98±0.33	30.83±5.09	22.35±0.47	22.06±0.39	24.03±0.42	24.72±0.43	23.16±0.48	25.24±0.14	19.84±0.70	24.87±0.00	22.65±0.03	25.77±0.00	23.64±0.01	28.35±0.00	32.03±0.80	25.81±0.00	23.30±0.04	**18.64±0.79**	19.55±0.28
	12	25.62±0.41	27.71±0.34	27.79±0.53	28.42±0.62	100.02±1.93	25.34±0.36	41.47±2.03	24.41±0.45	25.39±0.54	31.36±5.41	26.49±0.34	36.07±4.02	25.35±0.62	25.18±0.37	26.62±0.35	27.54±0.47	26.36±0.43	30.13±0.17	21.03±0.61	33.64±0.00	28.54±0.08	33.36±0.00	30.70±0.03	36.75±0.00	44.91±8.74	33.00±0.00	29.32±0.08	**19.85±0.76**	22.20±0.32
	Avg	23.29±0.43	24.04±0.28	25.38±0.56	24.84±0.65	97.96±2.03	22.39±0.39	38.29±1.20	21.68±0.28	22.41±0.47	29.03±6.07	23.19±0.34	31.23±4.99	22.53±0.49	22.21±0.38	24.34±0.40	24.97±0.45	23.39±0.46	25.55±0.15	19.88±0.68	25.65±0.00	23.12±0.04	26.22±0.00	24.21±0.01	28.85±0.00	32.90±1.12	26.23±0.00	23.75±0.04	**18.66±0.79**	19.62±0.27
MAPE (%)	3	24.10±0.81	21.76±0.38	23.21±0.43	23.31±0.78	79.39±1.35	21.38±0.30	31.34±0.34	20.89±0.38	21.45±0.51	25.89±5.52	23.18±0.54	34.95±16.21	21.52±0.65	21.25±0.30	21.75±0.39	30.95±1.10	22.76±0.99	22.34±0.37	19.81±0.84	19.36±0.00	17.99±0.04	20.16±0.00	18.36±0.01	19.80±0.00	21.95±0.35	19.90±0.00	18.89±0.01	**18.31±0.45**	18.52±0.20
	6	24.33±0.75	23.07±0.47	23.76±0.39	24.42±0.81	79.52±1.43	22.31±0.32	32.02±0.39	21.73±0.47	22.24±0.49	26.69±5.49	23.91±0.56	36.41±15.97	22.36±0.64	22.19±0.32	22.48±0.39	32.23±1.07	23.77±0.94	24.02±0.37	20.23±0.87	21.61±0.00	19.19±0.04	22.53±0.00	19.79±0.01	22.19±0.00	24.49±0.32	22.05±0.00	20.42±0.01	**18.75±0.46**	19.38±0.26
	12	25.99±0.73	26.28±0.63	25.75±0.39	27.32±0.86	79.07±1.37	24.56±0.38	33.58±0.89	23.69±0.83	24.14±0.51	28.62±5.53	25.98±0.56	39.24±15.64	24.53±0.73	24.43±0.38	24.45±0.40	35.38±1.06	26.08±0.96	28.21±0.42	21.03±0.84	27.27±0.00	21.83±0.05	27.51±0.00	22.86±0.01	26.87±0.00	28.81±0.24	26.63±0.00	23.73±0.02	**19.61±0.46**	21.11±0.39
	Avg	24.66±0.77	23.41±0.46	24.06±0.40	24.76±0.81	79.40±1.33	22.54±0.33	32.16±0.38	21.93±0.52	22.43±0.49	26.88±5.50	24.14±0.58	36.63±16.01	22.60±0.66	22.42±0.33	22.72±0.38	32.58±1.07	24.00±0.96	24.49±0.37	20.28±0.85	22.27±0.00	19.42±0.04	22.96±0.00	20.06±0.01	22.54±0.00	24.70±0.29	22.46±0.00	20.71±0.01	**18.80±0.45**	19.51±0.27
PEMS07																														
MAE	3	18.59±0.29	17.59±0.06	19.66±0.44	18.38±1.08	40.22±2.16	16.53±0.25	21.47±0.50	16.22±0.28	16.12±0.30	17.90±1.53	15.62±1.26	17.21±0.25	16.37±0.15	16.24±0.26	18.02±0.25	23.09±0.91	17.17±0.11	17.63±0.04	14.67±0.34	15.19±0.00	14.38±0.02	15.97±0.00	14.96±0.03	18.20±0.00	20.28±0.24	17.22±0.00	14.98±0.02	**13.00±0.06**	13.77±0.04
	6	18.93±0.27	19.22±0.05	20.25±0.40	19.21±0.96	40.32±1.94	17.83±0.27	22.58±0.58	17.53±0.32	17.40±0.31	19.09±1.46	16.93±1.25	18.14±0.23	17.64±0.15	17.55±0.27	18.94±0.17	24.77±1.12	18.44±0.08	19.49±0.01	15.29±0.29	18.03±0.00	16.36±0.03	19.10±0.00	17.51±0.02	21.91±0.00	23.40±0.16	20.19±0.00	17.24±0.01	**13.76±0.07**	15.12±0.06
	12	21.19±0.27	22.81±0.08	22.49±0.38	21.54±0.96	40.46±2.10	20.40±0.30	24.70±0.74	19.94±0.36	19.79±0.36	21.32±1.38	19.34±1.26	19.99±0.38	20.09±0.22	20.14±0.30	21.55±0.15	28.54±1.56	21.02±0.10	23.49±0.03	16.32±0.21	24.47±0.00	20.20±0.06	25.20±0.00	22.33±0.02	28.49±0.00	29.63±0.12	26.07±0.00	21.56±0.01	**14.91±0.10**	17.46±0.09
	Avg	19.39±0.28	19.56±0.05	20.62±0.41	19.51±0.99	40.30±2.05	18.02±0.27	22.72±0.59	17.67±0.31	17.54±0.31	19.23±1.47	17.07±1.26	18.26±0.27	17.81±0.16	17.74±0.27	19.27±0.19	25.17±1.15	18.64±0.09	19.85±0.01	15.32±0.28	18.67±0.00	16.63±0.04	19.55±0.00	17.86±0.02	22.38±0.00	23.89±0.16	20.58±0.00	17.53±0.01	**13.77±0.07**	15.22±0.06
RMSE	3	30.83±0.32	29.38±0.14	36.00±0.80	30.15±1.49	52.97±2.13	27.82±0.20	33.25±0.58	27.41±0.29	27.28±0.25	29.96±2.69	26.60±1.78	28.38±0.35	27.62±0.15	27.46±0.21	29.94±0.28	36.09±1.22	28.42±0.19	29.63±0.04	25.56±0.32	27.55±0.00	26.41±0.05	28.60±0.00	27.46±0.05	32.69±0.00	36.91±0.74	30.43±0.00	27.18±0.07	**23.51±0.08**	24.67±0.05
	6	31.89±0.28	32.54±0.12	37.21±0.75	31.78±1.30	54.08±1.79	30.46±0.22	35.59±0.67	29.94±0.33	29.84±0.28	32.30±2.45	29.14±1.82	30.27±0.30	30.13±0.17	30.10±0.22	31.73±0.17	38.85±1.49	30.90±0.16	33.33±0.06	26.92±0.27	32.91±0.00	30.53±0.10	34.44±0.00	32.80±0.03	39.61±0.00	42.56±0.17	35.69±0.00	31.79±0.06	**25.09±0.08**	27.21±0.09
	12	36.27±0.30	39.02±0.14	41.01±0.68	35.84±1.30	55.87±1.97	35.29±0.28	39.86±0.86	34.34±0.39	34.29±0.38	36.50±2.15	33.55±1.93	33.61±0.60	34.67±0.25	34.96±0.28	36.28±0.20	45.39±2.19	35.56±0.17	40.75±0.12	28.97±0.18	45.31±0.00	38.69±0.25	46.03±0.00	42.92±0.05	52.40±0.00	59.72±6.83	46.50±0.00	40.60±0.03	**27.33±0.11**	31.28±0.14
	Avg	32.62±0.30	33.05±0.13	37.75±0.75	32.22±1.35	54.10±1.93	30.73±0.23	35.82±0.68	30.13±0.32	30.04±0.29	32.52±2.46	29.33±1.84	30.41±0.37	30.37±0.18	30.37±0.23	32.22±0.21	39.57±1.56	31.19±0.17	33.90±0.07	26.93±0.26	34.18±0.00	31.15±0.13	35.34±0.00	33.57±0.03	40.57±0.00	44.62±1.28	36.42±0.00	32.42±0.05	**25.05±0.08**	27.30±0.08
MAPE (%)	3	30.48±1.24	27.98±0.33	31.59±1.03	30.73±2.59	50.26±2.43	26.51±1.30	34.13±1.18	25.57±1.32	25.51±2.27	26.75±1.59	24.19±3.12	27.34±0.96	25.71±0.67	26.27±1.27	32.44±0.94	55.33±2.34	29.41±1.13	27.44±0.37	20.97±0.86	19.33±0.00	18.12±0.03	20.30±0.00	18.88±0.03	20.43±0.00	21.13±0.06	20.34±0.00	18.90±0.03	**18.51±0.27**	19.46±0.37
	6	30.62±1.27	29.47±0.24	31.86±1.01	31.46±2.35	49.24±2.48	27.39±1.34	34.81±1.48	26.44±1.29	26.43±2.20	27.64±1.71	25.38±3.00	28.43±0.54	26.65±0.63	27.16±1.32	33.59±0.86	58.00±2.92	30.46±1.12	29.34±0.46	21.31±0.65	21.88±0.00	19.74±0.04	23.14±0.00	20.91±0.02	23.33±0.00	24.52±0.11	23.12±0.00	20.85±0.03	**18.99±0.18**	20.67±0.39
	12	32.67±1.31	33.46±0.27	34.54±1.08	33.85±2.10	47.91±3.13	29.62±1.46	36.82±2.23	28.63±1.18	28.65±2.11	29.78±1.91	28.11±3.10	31.10±1.22	28.92±0.73	29.40±1.43	37.31±0.74	64.27±4.24	33.08±1.16	34.30±0.78	22.17±0.41	28.40±0.00	23.16±0.04	29.16±0.00	24.96±0.02	28.97±0.00	29.96±0.09	28.69±0.00	24.87±0.05	**20.02±0.14**	23.20±0.40
	Avg	31.09±1.27	29.95±0.26	32.43±1.02	31.81±2.36	49.24±2.54	27.64±1.35	35.08±1.56	26.68±1.27	26.67±2.19	27.85±1.72	25.66±3.04	28.73±0.39	26.88±0.66	27.40±1.33	34.13±0.84	58.70±3.03	30.76±1.13	29.94±0.50	21.39±0.65	22.66±0.00	20.04±0.03	23.68±0.00	21.24±0.02	23.73±0.00	24.80±0.09	23.57±0.00	21.17±0.03	**19.07±0.20**	20.90±0.39
PEMS08																														
MAE	3	14.86±0.14	13.63±0.05	15.01±0.19	14.62±0.19	52.44±2.51	13.13±0.06	16.14±0.04	12.76±0.09	12.85±0.12	14.58±0.83	13.32±0.44	13.86±0.28	12.98±0.04	12.95±0.06	14.23±0.08	17.19±0.35	13.42±0.09	13.79±0.05	11.99±0.04	13.26±0.00	12.46±0.01	13.94±0.00	12.77±0.02	15.65±0.00	17.49±0.13	14.90±0.00	13.02±0.02	**11.20±0.02**	11.87±0.03
	6	15.17±0.13	14.97±0.05	15.60±0.16	15.39±0.16	52.24±2.61	14.18±0.07	17.16±0.06	13.80±0.10	13.91±0.12	15.47±0.78	14.31±0.33	14.66±0.28	14.02±0.03	14.00±0.06	14.83±0.06	18.07±0.30	14.44±0.08	15.28±0.04	12.47±0.03	15.62±0.00	14.00±0.02	16.49±0.00	14.62±0.01	18.77±0.00	19.98±0.13	17.30±0.00	14.75±0.02	**11.74±0.03**	12.88±0.04
	12	16.88±0.13	17.84±0.05	17.43±0.16	17.61±0.20	52.07±3.05	16.27±0.09	19.13±0.11	15.75±0.11	15.90±0.17	17.19±0.67	16.32±0.36	16.25±0.27	16.00±0.03	16.10±0.08	16.83±0.06	20.27±0.37	16.48±0.09	18.53±0.04	13.32±0.03	21.09±0.00	17.01±0.04	21.51±0.00	18.18±0.01	24.43±0.00	25.47±0.26	22.15±0.00	18.10±0.03	**12.63±0.04**	14.67±0.06
	Avg	15.50±0.13	15.22±0.05	15.86±0.17	15.69±0.16	52.25±2.71	14.34±0.07	17.30±0.05	13.92±0.10	14.04±0.13	15.58±0.77	14.47±0.38	14.76±0.28	14.15±0.03	14.16±0.07	15.12±0.07	18.33±0.31	14.60±0.08	15.58±0.04	12.52±0.03	16.19±0.00	14.22±0.02	16.87±0.00	14.90±0.01	19.21±0.00	20.52±0.16	17.64±0.00	14.99±0.02	**11.76±0.03**	12.97±0.04
RMSE	3	23.73±0.20	22.36±0.08	26.25±0.43	23.28±0.29	61.63±2.94	21.49±0.07	24.62±0.06	21.01±0.12	21.09±0.13	23.27±1.20	21.64±0.51	21.99±0.33	21.29±0.07	21.26±0.05	24.03±0.21	26.14±0.40	21.79±0.12	22.74±0.04	19.83±0.09	22.30±0.00	21.40±0.03	23.20±0.00	21.87±0.03	26.33±0.00	29.66±0.56	24.72±0.00	22.10±0.05	**18.81±0.05**	19.93±0.04
	6	24.70±0.18	24.88±0.10	27.48±0.38	24.83±0.26	62.11±3.02	23.63±0.09	26.65±0.12	23.05±0.15	23.19±0.16	25.08±1.05	23.66±0.40	23.65±0.34	23.35±0.04	23.41±0.08	25.24±0.16	27.58±0.31	23.77±0.13	25.65±0.03	20.82±0.08	26.57±0.00	24.61±0.06	27.72±0.00	25.66±0.02	31.97±0.00	33.94±0.36	28.76±0.00	25.56±0.05	**19.91±0.06**	21.88±0.07
	12	28.02±0.18	29.80±0.13	30.67±0.37	28.68±0.40	63.24±3.50	27.52±0.13	30.31±0.23	26.53±0.20	26.81±0.29	28.33±0.80	27.41±0.62	26.49±0.45	27.00±0.05	27.31±0.12	28.72±0.14	31.18±0.47	27.42±0.17	31.37±0.05	22.44±0.10	36.33±0.00	30.88±0.15	36.53±0.00	32.74±0.02	41.51±0.00	47.49±1.38	37.04±0.00	31.97±0.08	**21.57±0.09**	25.03±0.11
	Avg	25.20±0.19	25.23±0.10	27.85±0.40	25.27±0.29	62.21±3.13	23.85±0.09	26.86±0.13	23.20±0.15	23.35±0.18	25.25±1.04	23.89±0.50	23.76±0.37	23.54±0.04	23.63±0.08	25.69±0.17	28.00±0.33	23.99±0.13	26.07±0.04	20.88±0.09	27.58±0.00	25.08±0.08	28.38±0.00	26.18±0.02	32.55±0.00	35.46±0.39	29.31±0.00	25.97±0.06	**19.91±0.06**	21.96±0.07
MAPE (%)	3	22.82±0.42	19.07±0.41	21.25±0.69	20.76±0.73	57.52±1.13	19.25±0.27	22.79±0.31	18.05±0.22	19.62±1.41	21.16±1.57	20.30±1.59	26.19±1.99	18.63±0.35	19.13±0.27	19.38±0.76	41.36±1.88	20.82±1.06	18.96±0.43	17.16±0.26	17.01±0.00	15.85±0.02	17.93±0.00	16.36±0.01	18.02±0.00	18.90±0.04	17.86±0.00	16.69±0.04	**16.02±0.07**	16.31±0.10
	6	22.94±0.42	20.47±0.38	21.71±0.71	21.51±0.73	56.94±1.54	20.33±0.24	23.83±0.33	19.02±0.21	20.70±1.34	22.07±1.65	21.30±1.42	26.35±1.97	19.69±0.40	20.20±0.23	19.98±0.75	42.87±1.97	21.81±1.00	20.66±0.40	17.53±0.25	19.33±0.00	17.22±0.03	20.46±0.00	17.97±0.00	20.63±0.00	21.71±0.05	20.31±0.00	18.33±0.04	**16.42±0.07**	17.28±0.09
	12	24.58±0.45	23.92±0.38	24.00±0.73	24.07±0.76	56.54±2.39	22.74±0.27	26.26±0.49	21.37±0.32	23.18±1.24	24.14±1.73	23.62±0.94	27.80±2.10	22.19±0.54	22.61±0.27	22.15±0.79	46.66±2.26	24.16±0.88	24.95±0.43	18.49±0.24	25.32±0.00	20.17±0.04	25.89±0.00	21.27±0.00	25.69±0.00	26.68±0.07	25.19±0.00	21.82±0.04	**17.33±0.09**	19.33±0.10
	Avg	23.31±0.41	20.85±0.39	22.13±0.71	21.90±0.74	57.09±1.58	20.58±0.25	24.09±0.35	19.28±0.22	20.96±1.33	22.26±1.64	21.55±1.34	26.61±1.73	19.96±0.41	20.45±0.25	20.33±0.76	43.33±2.01	22.07±0.99	21.16±0.42	17.66±0.23	20.06±0.00	17.49±0.03	20.96±0.00	18.26±0.00	21.00±0.00	22.04±0.05	20.70±0.00	18.63±0.04	**16.51±0.07**	17.47±0.10
PEMS10																														
MAE	3	11.81±0.19	11.46±0.05	11.82±0.09	11.87±0.11	81.21±3.23	11.09±0.04	11.49±0.20	10.86±0.08	11.14±0.06	11.90±1.24	11.15±0.08	12.84±1.70	11.05±0.08	11.01±0.04	11.85±0.09	13.13±0.29	11.13±0.07	11.73±0.04	10.09±0.03	11.46±0.00	10.79±0.01	11.82±0.00	11.04±0.01	13.03±0.00	14.44±0.12	12.56±0.00	11.24±0.00	**9.59±0.03**	10.23±0.03
	6	12.06±0.14	12.32±0.05	12.29±0.08	12.38±0.11	81.43±3.27	11.74±0.04	12.23±0.23	11.50±0.10	11.82±0.08	12.51±1.18	11.81±0.07	13.48±1.73	11.68±0.09	11.65±0.04	12.36±0.07	13.62±0.32	11.74±0.08	12.69±0.04	10.32±0.03	12.93±0.00	11.74±0.01	13.38±0.00	12.10±0.01	15.01±0.00	16.15±0.13	13.93±0.00	12.30±0.00	**9.84±0.03**	10.81±0.04
	12	13.19±0.15	14.05±0.05	13.48±0.08	13.70±0.14	81.53±3.26	13.06±0.06	13.70±0.29	12.76±0.12	13.18±0.13	13.72±1.04	13.14±0.08	14.87±1.91	12.97±0.10	12.97±0.05	13.73±0.09	14.81±0.35	12.95±0.07	14.67±0.04	10.73±0.04	16.37±0.00	13.51±0.01	16.49±0.00	14.17±0.01	18.64±0.00	19.70±0.22	16.97±0.00	14.30±0.01	**10.26±0.03**	11.91±0.06
	Avg	12.26±0.16	12.45±0.05	12.43±0.08	12.54±0.11	81.39±3.16	11.85±0.04	12.35±0.23	11.60±0.09	11.93±0.08	12.61±1.16	11.92±0.07	13.60±1.76	11.79±0.09	11.76±0.04	12.53±0.08	13.75±0.32	11.83±0.07	12.86±0.04	10.34±0.03	13.30±0.00	11.85±0.01	13.63±0.00	12.27±0.01	15.31±0.00	16.45±0.14	14.18±0.00	12.43±0.01	**9.86±0.03**	10.88±0.04
RMSE	3	19.17±0.29	19.03±0.09	20.39±0.16	19.21±0.12	83.86±3.27	18.21±0.02	18.96±0.38	17.84±0.14	18.31±0.08	20.00±2.91	18.35±0.11	19.71±1.41	18.17±0.10	18.08±0.02	19.33±0.09	20.78±0.37	18.33±0.09	19.35±0.05	16.69±0.07	19.29±0.00	18.45±0.03	19.75±0.00	18.67±0.01	21.83±0.00	23.87±0.10	20.99±0.00	19.00±0.02	**16.07±0.04**	17.08±0.05
	6	19.79±0.24	20.59±0.06	21.24±0.14	20.20±0.12	84.50±3.37	19.49±0.04	20.39±0.45	19.06±0.18	19.58±0.14	21.13±2.71	19.63±0.10	20.94±1.41	19.40±0.10	19.35±0.04	20.30±0.07	21.60±0.41	19.49±0.10	21.11±0.05	17.17±0.05	21.87±0.00	20.39±0.04	22.48±0.00	20.80±0.01	25.46±0.00	28.17±2.25	23.25±0.00	21.08±0.02	**16.57±0.03**	18.16±0.08
	12	21.96±0.25	23.58±0.05	23.35±0.13	22.52±0.20	85.41±3.41	21.94±0.07	23.10±0.55	21.39±0.21	22.03±0.24	23.31±2.32	22.08±0.12	23.29±1.57	21.80±0.14	21.80±0.07	22.59±0.11	23.56±0.44	21.66±0.11	24.54±0.05	17.96±0.05	27.96±0.00	24.06±0.05	27.93±0.00	24.87±0.01	32.24±0.00	33.91±1.29	28.30±0.00	24.93±0.04	**17.36±0.03**	20.18±0.11
	Avg	20.13±0.26	20.80±0.06	21.49±0.14	20.45±0.13	84.52±3.24	19.66±0.03	20.58±0.45	19.22±0.17	19.77±0.14	21.29±2.68	19.80±0.11	21.08±1.43	19.58±0.10	19.53±0.04	20.54±0.09	21.81±0.40	19.63±0.10	21.36±0.05	17.20±0.05	22.54±0.00	20.64±0.04	22.92±0.00	21.11±0.01	25.99±0.00	28.09±1.21	23.64±0.00	21.32±0.03	**16.58±0.03**	18.28±0.08
MAPE (%)	3	30.74±0.34	28.38±0.32	29.73±0.29	29.49±0.74	756.76±29.47	29.11±0.29	30.04±0.94	28.70±0.37	30.52±0.75	29.20±1.71	29.69±0.29	58.30±38.83	29.13±0.55	28.95±0.28	31.70±0.31	40.22±1.74	29.02±0.61	28.77±0.41	27.39±0.42	27.81±0.00	26.36±0.07	28.80±0.00	27.03±0.03	28.16±0.00	29.19±0.16	28.44±0.00	27.36±0.03	**25.98±0.03**	26.87±0.08
	6	31.29±0.33	29.84±0.32	30.52±0.23	30.31±0.69	755.22±29.03	30.00±0.23	31.38±0.98	29.47±0.34	31.73±0.67	30.11±1.70	30.78±0.50	58.55±40.02	30.04±0.59	29.84±0.22	32.71±0.23	41.21±1.76	30.02±0.71	30.57±0.38	27.74±0.45	29.81±0.00	27.60±0.07	30.98±0.00	28.35±0.03	30.54±0.00	31.69±0.10	30.28±0.00	28.91±0.03	**26.34±0.05**	27.55±0.07
	12	33.45±0.41	33.35±0.34	32.93±0.18	32.78±0.66	752.24±26.19	32.42±0.18	34.68±1.59	31.77±0.42	35.13±0.79	32.53±1.65	33.28±0.51	62.45±42.54	32.43±0.61	32.25±0.17	35.68±0.25	43.96±1.88	32.40±0.94	34.98±0.32	28.55±0.44	35.14±0.00	30.28±0.06	35.86±0.00	31.18±0.03	35.72±0.00	36.23±0.13	34.51±0.00	32.29±0.03	**27.08±0.07**	29.19±0.10
	Avg	31.65±0.35	30.22±0.32	30.86±0.22	30.65±0.69	755.06±28.01	30.30±0.22	31.79±1.11	29.80±0.34	32.18±0.70	30.41±1.69	31.06±0.42	59.52±40.27	30.34±0.58	30.14±0.22	33.11±0.25	41.57±1.79	30.28±0.74	31.07±0.38	27.83±0.45	30.51±0.00	27.84±0.07	31.47±0.00	28.62±0.03	30.97±0.00	32.02±0.11	30.73±0.00	29.22±0.02	**26.41±0.04**	27.73±0.07
PEMS11																														
MAE	3	18.69±0.36	17.88±0.20	20.19±0.31	18.90±0.41	83.38±3.81	17.03±0.13	19.74±1.56	16.62±0.13	16.93±0.27	24.48±2.64	17.24±0.28	23.71±3.81	16.91±0.12	16.76±0.12	18.57±0.32	20.18±0.43	17.54±0.29	18.32±0.15	16.33±0.16	15.80±0.00	14.70±0.01	16.84±0.00	15.25±0.02	18.69±0.00	20.45±0.14	17.40±0.00	15.09±0.00	**13.40±0.06**	13.97±0.07
	6	19.19±0.28	19.63±0.17	20.93±0.31	19.91±0.39	85.10±3.52	18.50±0.12	21.22±1.61	18.07±0.12	18.43±0.34	25.70±2.54	18.77±0.21	25.31±4.21	18.36±0.15	18.26±0.12	19.61±0.32	21.46±0.36	18.98±0.29	20.20±0.11	16.94±0.14	19.21±0.00	16.84±0.02	20.39±0.00	17.88±0.02	22.74±0.00	24.00±0.10	20.82±0.00	17.34±0.01	**14.32±0.07**	15.44±0.08
	12	21.83±0.24	23.48±0.12	23.55±0.32	22.77±0.55	88.77±3.23	21.61±0.15	24.45±1.70	20.92±0.13	21.44±0.47	28.19±2.36	21.86±0.16	29.22±4.63	21.32±0.24	21.38±0.15	22.64±0.34	24.40±0.67	21.99±0.27	24.43±0.09	18.28±0.12	27.09±0.00	21.20±0.04	27.29±0.00	23.15±0.03	29.56±0.00	30.42±0.16	27.38±0.00	21.94±0.03	**15.96±0.09**	18.12±0.10
	Avg	19.69±0.30	20.00±0.16	21.35±0.31	20.29±0.44	85.43±3.52	18.77±0.13	21.54±1.59	18.29±0.13	18.67±0.35	25.89±2.53	19.02±0.23	25.76±4.14	18.60±0.15	18.52±0.13	20.02±0.32	21.77±0.34	19.24±0.28	20.62±0.12	17.05±0.12	20.04±0.00	17.19±0.02	20.91±0.00	18.32±0.02	23.16±0.00	24.37±0.09	21.24±0.00	17.71±0.01	**14.40±0.07**	15.60±0.08
RMSE	3	31.04±0.45	29.94±0.35	37.69±0.55	31.36±0.74	112.94±3.71	28.37±0.17	32.65±2.94	27.81±0.14	28.11±0.36	42.15±5.33	28.52±0.27	37.06±5.06	28.19±0.16	27.97±0.16	30.85±0.38	32.56±0.56	29.15±0.31	30.60±0.15	27.05±0.36	28.35±0.00	26.78±0.04	30.08±0.00	27.62±0.04	33.70±0.00	36.68±0.41	31.28±0.00	27.08±0.01	**23.33±0.09**	24.48±0.12
	6	32.77±0.30	33.71±0.26	39.44±0.53	33.73±0.69	116.32±3.41	31.76±0.14	35.91±2.87	31.05±0.11	31.50±0.44	44.47±4.94	31.98±0.20	40.37±5.70	31.47±0.20	31.38±0.13	33.20±0.37	35.12±0.52	32.37±0.31	34.80±0.12	28.65±0.25	35.45±0.00	31.71±0.06	37.19±0.00	33.61±0.08	41.63±0.00	43.90±0.27	37.82±0.00	32.06±0.04	**25.63±0.10**	27.84±0.12
	12	38.65±0.27	41.33±0.17	44.47±0.51	39.49±1.05	123.41±3.10	38.27±0.17	42.48±2.88	36.98±0.16	37.79±0.69	49.10±4.27	38.39±0.19	47.69±6.19	37.68±0.32	37.94±0.17	39.19±0.37	40.62±1.15	38.55±0.30	43.44±0.13	31.70±0.19	51.41±0.00	41.40±0.19	50.75±0.00	44.93±0.10	54.25±0.00	56.11±0.11	50.35±0.00	41.93±0.10	**29.38±0.12**	33.49±0.16
	Avg	33.66±0.34	34.32±0.27	40.11±0.53	34.35±0.80	116.95±3.38	32.21±0.15	36.45±2.86	31.40±0.13	31.90±0.47	44.81±4.91	32.38±0.23	41.08±5.52	31.88±0.20	31.83±0.15	33.89±0.36	35.62±0.53	32.79±0.31	35.52±0.13	28.82±0.20	37.05±0.00	32.41±0.08	38.14±0.00	34.43±0.07	42.24±0.00	44.37±0.19	38.56±0.00	32.78±0.04	**25.74±0.10**	28.06±0.13
MAPE (%)	3	30.78±0.93	24.39±0.34	29.45±0.64	26.32±0.78	77.32±15.78	26.11±0.80	28.92±0.90	23.82±0.70	26.53±1.24	35.21±6.88	28.98±1.96	45.25±14.22	27.10±0.83	25.97±0.79	30.64±1.81	45.43±2.56	29.58±2.86	24.98±0.73	23.94±0.63	21.66±0.00	20.03±0.04	22.80±0.00	20.81±0.03	22.44±0.00	23.02±0.01	22.43±0.00	20.95±0.01	**20.75±0.22**	22.48±0.39
	6	30.74±0.87	26.12±0.37	29.65±0.33	27.14±0.80	80.97±18.12	27.25±0.78	30.36±1.20	25.01±0.79	27.76±1.43	36.53±6.71	30.24±1.63	46.71±12.65	28.41±1.01	27.11±0.78	31.58±1.76	47.41±2.67	30.58±2.79	26.94±0.61	24.36±0.69	24.49±0.00	21.62±0.04	25.85±0.00	22.74±0.04	25.41±0.00	26.42±0.04	25.28±0.00	22.88±0.02	**21.31±0.25**	23.52±0.39
	12	32.84±0.78	30.75±0.53	32.76±0.27	29.87±0.87	89.84±23.97	30.48±0.87	34.14±1.85	28.09±0.80	31.40±2.51	39.89±6.43	33.90±1.15	51.27±11.76	31.92±1.30	30.33±0.87	35.10±1.52	52.32±3.03	33.38±2.70	32.38±0.59	25.71±0.78	31.70±0.00	25.05±0.03	32.22±0.00	26.79±0.04	30.99±0.00	31.71±0.03	31.06±0.00	26.96±0.05	**22.59±0.30**	25.80±0.38
	Avg	31.28±0.86	26.69±0.39	30.37±0.32	27.55±0.81	82.12±18.72	27.67±0.81	30.83±1.25	25.39±0.75	28.29±1.63	36.90±6.66	30.78±1.62	47.34±12.72	28.85±1.00	27.53±0.80	32.17±1.75	48.00±2.68	30.95±2.77	27.66±0.61	24.54±0.69	25.36±0.00	21.93±0.04	26.41±0.00	23.11±0.03	25.78±0.00	26.67±0.03	25.78±0.00	23.23±0.03	**21.44±0.25**	23.73±0.38

Table 6: Per-horizon results on PEMS12 and TFNSW (mean±SD).

Metric	Len	Static STGNN Backbones: DCRNN	ASTGNN	TGCN	GWN	Naïve Schemes: Pretrain	Retrain	OL-NN	OL-AN
PEMS12									
MAE	3	15.43±0.22	14.23±0.12	16.26±0.09	15.16±0.28	15.17±0.18	13.59±0.05	16.25±0.47	13.24±0.05
	6	15.76±0.16	15.67±0.09	16.80±0.08	16.11±0.25	16.28±0.20	14.75±0.05	17.78±0.56	14.40±0.06
	12	17.77±0.14	18.87±0.10	18.83±0.06	18.46±0.20	18.34±0.30	17.02±0.06	20.91±0.77	16.58±0.09
	Avg	16.16±0.17	15.98±0.10	17.13±0.08	16.38±0.23	16.41±0.21	14.91±0.05	18.04±0.57	14.54±0.06
RMSE	3	26.80±0.31	25.07±0.18	31.84±0.21	26.13±0.45	25.43±0.16	23.87±0.08	28.37±0.92	23.41±0.03
	6	27.76±0.23	28.00±0.16	32.92±0.19	28.03±0.39	27.77±0.22	26.29±0.06	31.54±1.11	25.78±0.06
	12	31.86±0.20	34.05±0.23	36.42±0.17	32.38±0.27	31.84±0.41	30.73±0.07	37.79±1.53	29.93±0.15
	Avg	28.46±0.25	28.49±0.18	33.43±0.19	28.47±0.35	27.96±0.24	26.54±0.06	32.01±1.14	25.97±0.07
MAPE (%)	3	30.24±0.93	24.84±0.51	29.71±0.50	25.72±0.40	27.23±0.80	25.35±0.64	27.74±0.49	24.09±0.35
	6	30.30±0.77	26.69±0.51	29.80±0.42	26.82±0.39	28.24±0.93	26.56±0.55	29.07±0.52	25.13±0.23
	12	32.77±0.67	31.71±0.65	33.40±0.39	30.26±0.53	30.60±1.19	29.49±0.46	32.30±0.73	27.98±0.40
	Avg	30.90±0.78	27.32±0.53	30.68±0.38	27.32±0.41	28.48±0.81	26.88±0.52	29.42±0.55	25.48±0.28
TFNSW									
MAE	3	168.71±15.14	122.66±1.26	195.29±1.44	143.12±2.43	257.73±11.31	120.27±0.46	148.65±3.87	114.38±1.09
	6	181.43±14.38	145.13±0.79	211.00±1.58	175.19±3.28	261.85±11.96	142.34±0.66	180.04±3.54	134.84±1.26
	12	191.66±13.98	158.22±0.48	219.72±1.46	203.53±4.54	269.70±12.03	156.34±0.69	199.28±4.63	148.00±1.26
	Avg	179.87±14.50	140.15±0.82	208.33±1.48	170.89±3.29	262.63±10.87	138.04±0.62	174.36±3.31	130.61±0.96
RMSE	3	280.25±20.45	217.63±1.09	371.81±3.45	262.49±4.16	370.25±14.96	215.03±0.78	269.38±4.49	205.28±2.18
	6	308.03±18.86	256.45±1.03	394.80±3.35	326.67±6.57	384.60±12.76	257.56±1.26	333.40±6.29	244.87±2.33
	12	327.96±18.20	282.09±0.90	407.82±2.97	370.26±8.00	402.93±11.34	283.78±1.17	365.29±9.55	269.48±2.24
	Avg	303.16±19.20	247.22±1.02	390.75±3.25	313.61±5.86	383.38±12.29	247.58±1.00	318.17±5.87	235.35±1.95
MAPE (%)	3	203.98±42.96	101.54±2.40	167.79±4.05	109.87±6.56	419.02±18.83	93.75±3.87	121.57±11.39	87.14±5.01
	6	210.50±43.26	130.32±3.10	176.75±3.86	137.59±9.85	427.56±35.35	104.87±3.92	131.86±9.09	102.09±3.95
	12	221.79±41.52	146.31±4.51	188.49±3.75	174.63±17.60	438.74±48.59	115.98±3.12	137.54±6.98	114.71±2.98
	Avg	211.72±42.72	125.28±3.02	177.95±3.89	138.49±10.57	430.83±35.95	104.00±3.59	129.82±8.24	99.93±3.54

Metric	Len	Evolving Graph: TrafStm	PECPM	STKEC	EAC	Retrieval TTC: STRAP	ST-TTC	Static Forecasting Backbones: STID	STNorm	iTrans	DLinear	STAE
PEMS12												
MAE	3	13.42±0.06	14.97±1.63	13.42±0.10	13.96±0.25	13.59±0.15	13.41±0.05	15.00±0.09	17.97±0.53	13.92±0.13	14.31±0.02	12.39±0.11
	6	14.60±0.06	16.02±1.50	14.57±0.08	14.84±0.28	14.72±0.14	14.57±0.05	15.76±0.07	19.17±0.41	15.06±0.10	15.90±0.02	12.87±0.10
	12	16.81±0.13	18.11±1.30	16.75±0.10	16.63±0.44	16.95±0.14	16.87±0.07	18.07±0.06	22.03±0.30	17.48±0.08	19.37±0.03	13.70±0.12
	Avg	14.74±0.07	16.18±1.49	14.71±0.08	14.98±0.30	14.88±0.14	14.74±0.05	16.07±0.07	19.49±0.39	15.27±0.11	16.23±0.02	12.91±0.11
RMSE	3	23.68±0.09	26.16±3.54	23.69±0.15	24.14±0.31	23.87±0.18	23.58±0.05	26.16±0.22	29.50±0.83	24.26±0.18	25.36±0.03	22.03±0.12
	6	26.12±0.09	28.34±3.19	26.08±0.12	26.01±0.36	26.23±0.17	26.03±0.04	27.64±0.19	31.61±0.57	26.59±0.14	28.70±0.05	23.09±0.11
	12	30.40±0.22	32.39±2.68	30.26±0.19	29.33±0.71	30.58±0.21	30.50±0.08	31.73±0.20	36.73±0.45	31.16±0.13	35.51±0.08	24.77±0.15
	Avg	26.32±0.10	28.57±3.19	26.28±0.12	26.18±0.40	26.48±0.17	26.28±0.05	28.14±0.19	32.19±0.51	26.92±0.15	29.24±0.05	23.13±0.12
MAPE (%)	3	25.52±0.76	28.04±1.67	24.99±0.83	28.88±3.23	25.59±1.28	25.24±0.63	31.60±0.74	55.15±2.04	27.92±1.57	24.78±0.36	23.40±0.78
	6	26.46±0.63	29.04±1.40	26.32±0.61	29.31±2.63	26.69±1.16	26.45±0.53	33.05±0.86	58.10±2.24	28.96±1.38	27.01±0.44	23.81±0.72
	12	29.56±0.75	31.53±0.96	29.50±0.47	31.58±2.32	29.76±1.26	29.39±0.45	37.84±1.06	65.26±2.85	31.77±1.26	32.90±0.72	24.90±0.70
	Avg	26.88±0.62	29.30±1.38	26.66±0.60	29.71±2.65	27.09±1.22	26.77±0.51	33.77±0.88	58.93±2.28	29.32±1.41	27.74±0.47	23.94±0.74
TFNSW												
MAE	3	126.95±2.36	160.08±23.60	129.61±0.97	124.47±2.51	118.46±1.42	129.39±0.47	146.99±8.64	132.28±2.29	127.68±0.63	155.32±0.84	88.89±1.68
	6	150.05±2.33	177.30±21.89	152.15±1.41	139.81±1.76	139.74±2.06	147.98±0.77	151.13±5.35	150.92±2.57	148.02±0.71	191.54±1.27	95.46±1.73
	12	164.32±2.37	189.11±20.37	167.11±1.75	154.35±1.24	153.70±2.03	160.78±0.84	162.81±4.96	170.66±2.26	161.09±1.47	205.55±1.20	107.50±1.60
	Avg	145.42±2.41	174.19±22.07	147.86±1.05	137.90±1.42	135.57±1.80	144.41±0.72	153.63±6.20	148.78±2.33	144.04±0.82	182.20±1.03	96.12±1.67
RMSE	3	224.54±4.17	292.97±49.14	224.34±1.36	219.33±3.97	211.67±3.60	229.39±0.90	291.53±13.80	241.64±3.36	226.79±1.65	277.14±1.12	171.62±2.96
	6	270.67±4.71	326.33±44.92	271.09±2.95	250.85±2.25	253.32±4.77	265.94±1.21	303.05±8.85	280.63±3.49	266.95±2.18	335.83±1.27	189.36±3.02
	12	296.76±4.74	346.80±41.06	297.74±2.86	277.73±1.60	279.24±4.41	290.96±1.24	330.87±7.88	318.01±2.67	290.09±2.68	358.40±1.12	219.12±2.63
	Avg	259.40±4.67	318.61±45.67	259.82±2.21	245.35±1.93	243.47±4.15	257.96±1.08	307.19±10.11	275.25±3.16	257.27±2.16	318.13±1.10	190.33±2.83
MAPE (%)	3	94.78±7.81	94.19±7.03	134.83±7.19	109.08±12.15	93.03±3.88	99.28±3.48	112.69±14.09	162.98±5.57	98.77±5.85	154.16±5.94	59.94±3.07
	6	108.40±8.04	108.65±5.53	147.27±10.71	122.40±16.15	105.70±3.72	106.79±3.16	110.90±12.12	183.80±8.11	114.90±9.27	194.19±6.66	64.60±2.35
	12	120.47±9.27	119.62±3.65	162.82±10.53	132.49±17.32	117.09±2.11	115.50±2.33	111.46±8.46	206.11±5.57	130.80±6.23	218.83±4.82	76.98±1.65
	Avg	106.92±8.01	106.24±4.44	148.01±7.69	119.37±14.85	104.21±3.09	106.33±3.01	112.25±11.68	182.00±6.10	113.97±6.78	187.17±5.55	66.43±2.35

Metric	Len	TSFMs: Chro2-Z	Chro2-F	TimesF-Z	TimesF-T	Moira1-Z	Moira1-T	Moira2-Z	Moira2-T	Ours: A2TTA-S	A2TTA-O
PEMS12											
MAE	3	13.51±0.00	12.73±0.00	14.31±0.00	13.24±0.03	15.91±0.00	17.32±0.62	14.88±0.00	13.21±0.01	**11.37±0.04**	12.11±0.02
	6	16.05±0.00	14.41±0.01	17.06±0.00	15.37±0.03	19.07±0.00	20.17±0.58	17.51±0.00	15.06±0.01	**11.94±0.05**	13.20±0.01
	12	21.94±0.00	17.73±0.04	22.48±0.00	19.51±0.03	24.73±0.00	25.32±0.63	22.80±0.00	18.67±0.01	**12.81±0.07**	15.09±0.02
	Avg	16.67±0.00	14.66±0.01	17.48±0.00	15.69±0.02	19.49±0.00	20.49±0.59	17.90±0.00	15.32±0.01	**11.95±0.05**	13.28±0.01
RMSE	3	24.46±0.00	23.44±0.02	25.67±0.00	24.31±0.05	29.08±0.00	32.10±1.13	26.66±0.00	24.10±0.03	**20.53±0.06**	21.80±0.03
	6	29.57±0.00	27.08±0.04	31.07±0.00	29.01±0.05	35.26±0.00	37.46±1.02	31.59±0.00	28.07±0.03	**21.73±0.07**	23.97±0.02
	12	41.53±0.00	34.49±0.19	41.80±0.00	38.08±0.07	46.15±0.00	47.62±0.88	41.70±0.00	35.73±0.01	**23.45±0.11**	27.45±0.05
	Avg	30.85±0.00	27.68±0.07	31.92±0.00	29.74±0.05	36.00±0.00	38.04±1.02	32.32±0.00	28.62±0.02	**21.71±0.08**	24.05±0.02
MAPE (%)	3	22.43±0.00	20.71±0.03	23.78±0.00	21.67±0.02	23.72±0.00	24.09±0.26	23.33±0.00	21.79±0.02	**21.39±0.23**	21.95±0.18
	6	25.58±0.00	22.46±0.04	27.22±0.00	23.88±0.03	27.23±0.00	28.45±0.18	26.61±0.00	23.97±0.02	**21.87±0.24**	23.15±0.12
	12	33.94±0.00	26.38±0.07	34.72±0.00	28.54±0.03	34.24±0.00	35.19±0.14	33.62±0.00	28.62±0.03	**22.95±0.23**	25.61±0.11
	Avg	26.63±0.00	22.84±0.04	27.93±0.00	24.31±0.02	27.76±0.00	28.81±0.19	27.26±0.00	24.38±0.02	**21.98±0.24**	23.36±0.11
TFNSW											
MAE	3	160.32±0.00	128.86±0.38	190.66±0.00	165.69±0.60	255.65±0.00	288.08±1.27	204.29±0.00	114.89±0.23	**61.18±0.30**	76.26±0.21
	6	245.53±0.00	171.51±0.85	282.31±0.00	234.82±0.32	324.03±0.00	362.17±0.58	301.81±0.00	151.95±0.30	**68.18±0.34**	88.38±0.24
	12	288.66±0.00	187.39±0.77	345.79±0.00	256.73±0.39	359.26±0.00	371.62±0.68	399.37±0.00	167.54±0.34	**78.02±0.42**	101.46±0.59
	Avg	227.74±0.00	160.20±0.68	266.77±0.00	215.22±0.33	306.03±0.00	334.59±0.89	294.05±0.00	142.36±0.28	**67.89±0.32**	86.78±0.24
RMSE	3	327.55±0.00	268.64±1.59	369.29±0.00	336.90±0.43	462.40±0.00	561.94±2.60	417.01±0.00	252.28±0.75	**129.32±0.65**	154.06±0.79
	6	473.73±0.00	348.32±2.49	519.48±0.00	458.77±1.05	570.16±0.00	649.78±1.17	565.48±0.00	326.33±0.66	**147.94±0.81**	180.12±0.90
	12	546.05±0.00	372.36±2.11	608.15±0.00	493.26±0.47	648.32±0.00	652.88±0.85	692.51±0.00	352.26±0.80	**173.98±0.61**	211.44±1.27
	Avg	440.18±0.00	323.47±2.06	487.06±0.00	420.65±0.52	544.51±0.00	609.59±1.67	545.09±0.00	303.78±0.69	**147.15±0.64**	176.97±0.84
MAPE (%)	3	60.65±0.00	54.11±0.13	84.30±0.00	73.71±1.57	103.01±0.00	54.54±0.87	68.13±0.00	43.39±0.01	**49.30±0.51**	56.66±1.98
	6	117.43±0.00	82.43±0.32	163.00±0.00	128.42±0.85	164.37±0.00	164.10±1.19	131.57±0.00	65.49±0.27	**53.55±0.32**	66.51±1.39
	12	182.97±0.00	99.59±0.44	242.48±0.00	171.51±0.37	207.96±0.00	206.18±1.08	226.64±0.00	80.90±0.33	**63.01±0.27**	77.64±1.30
	Avg	120.67±0.00	78.26±0.28	160.45±0.00	124.25±0.83	155.80±0.00	144.74±1.00	137.97±0.00	62.69±0.22	**54.66±0.33**	66.01±1.49

Newly added sensors. Figure 13 isolates nodes added at graph-growing transitions. A^{2}TTA(STAE) is best in all 360 plotted combinations and reduces MAE, RMSE, and MAPE relative to frozen STAEFormer by 9.1%, 7.4%, and 8.8% on average, respectively.

Average-horizon baseline comparison. While the per-step plot focuses on six representative methods, Figures 7 and 14 broadens the Avg comparison to the full baseline set. Together, the figures cover all nine EvoXXLTraffic districts and TFNSW. A^{2}TTA(STAE) ranks first in 29 of the 30 comparable panels. The only exception is TFNSW RMSE, where A^{2}TTA(OLAN) records 177.20 and A^{2}TTA(STAE) records 178.38. Averaged equally across datasets, A^{2}TTA(STAE) reduces MAE, RMSE, and MAPE by 10.2%, 8.3%, and 9.2% relative to frozen STAEFormer, and by 8.8%, 8.1%, and 7.0% relative to A^{2}TTA(OLAN). Its MAE reduction over frozen STAEFormer ranges from 4.2% on PEMS10 to 33.7% on TFNSW, including 12.6% on PEMS04 and 9.9% on PEMS11. Daggered TSFM-FT bars are all-sensor references and are excluded from these new-sensor rankings.

A.4 Drift-Severity Analysis Details

Metrics and aggregation. This auxiliary diagnostic uses legacy batched outputs. It joins drift measurements to paired yearly errors from the five-seed analysis and to one pre-specified seed for the all-network breadth check. For consecutive years $y-1$ and y, traffic-distribution drift is the 1-Wasserstein distance between deterministic samples of the two raw traffic arrays, normalized by their average standard deviation. Graph-density drift is $|\log(\delta_y/\delta_{y-1})|$, where δ_y is the nonzero adjacency density. For TFNSW, sensor churn is computed from exact yearly sensor IDs as $(|\mathcal{V}_y \setminus \mathcal{V}_{y-1}| + |\mathcal{V}_{y-1} \setminus \mathcal{V}_y|)/|\mathcal{V}_y \cup \mathcal{V}_{y-1}|$. The corresponding PEMS analysis uses the conservative net-count lower bound $|N_y - N_{y-1}|/\max(N_y, N_{y-1})$, because the processed experimental arrays used for this diagnostic do not retain the full station-ID transition record.

For a backbone b, the relative gain in dataset d and year y is $100(\mathrm{MAE}_{b,d,y} - \mathrm{MAE}_{\mathrm{A^2TTA}(b),d,y})/\mathrm{MAE}_{b,d,y}$. The primary layer first averages the errors within each dataset-year over the five paired seeds and contains 47 Online-AN and 87 STAEFormer yearly cells. Quartiles are formed separately for each backbone and drift signal. Their intervals are nonparametric 95% bootstrap intervals over the resulting yearly cells, rather than seed-level intervals. The breadth layer uses one pre-specified seed and covers 223 adjacent-year cells per backbone across all ten networks. It is a breadth check rather than an additional inference layer and therefore has no error bars.

Detailed observations. Full A^{2}TTA improves every yearly cell in the paired five-seed layer, so all primary quartile means and confidence intervals in Figure 15 remain above zero. Normalized Wasserstein drift has no reliable monotonic association with gain ($\rho = 0.163$, $p = .275$ for Online-AN and $\rho = 0.123$, $p = .255$ for STAEFormer). The gain remains positive across the observed severity range. Structural drift is more informative. For STAEFormer, graph-density change is positively associated with gain ($\rho = 0.228$, $p = .034$), while sensor churn gives $\rho = 0.209$ ($p = .052$). The primary Online-AN churn association is weaker, but its highest-churn quartile still has the largest mean gain (9.0%). In the all-network breadth layer, churn is positively associated with gain for both Online-AN ($\rho = 0.263$) and STAEFormer ($\rho = 0.241$), with both $p < .001$. The highest-churn quartile reaches 13.5% and 11.4%, respectively. These diagnostics show positive gains across the observed temporal shifts and a larger benefit during severe sensor turnover.

A.5 Long-Term Adaptation Breadth

This legacy-batched, pre-specified single-seed check covers ten datasets (fig. 16). Mean MAE improvements over frozen backbones range from 3.7% to 26.8%. The figure also reports yearly win rates against global-only TTA and the corresponding backbone. Timing repeats are not independent seeds.

A.6 Protocol Controls and Deployment Cost

Protocol checks. The result files behind every main-table A^{2}TTA cell record an evaluation batch size of one, training-only input normalization, metadata-based sensor identities, a 12-step label delay, and a 64-window global-update interval. The matched backbone row uses the same per-year checkpoint and seed. For the new-sensor comparison, all 21 comparable methods were evaluated on the same metadata-defined transition cohorts over ten networks and five seeds.

Dependence on yearly warm-up. We also initialize the FiLM calibrator as the identity and skip its three warm-up epochs on PEMS05, PEMS08, and TFNSW. Across five paired seeds, removing warm-up

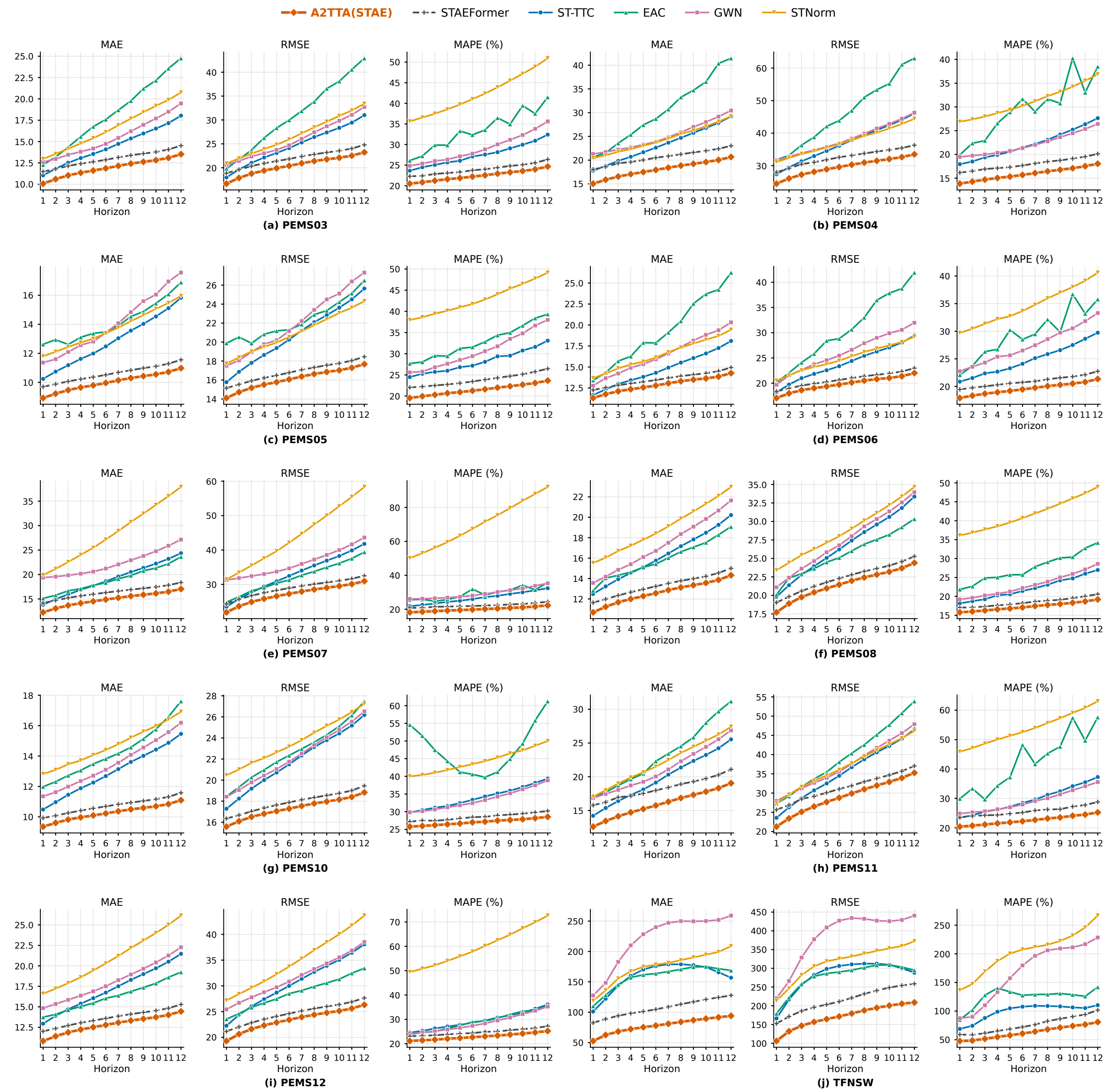

Figure 12: All-sensor per-step results on ten networks.

increases Avg-MAE by 1.0%, 1.1%, and 2.4% for the STAEFormer variant and by 1.7%, 1.4%, and 11.4% for the Online-AN variant, respectively. A^{2}TTA can therefore enter the causal stream without calibrator warm-up, but the labeled training partition provides a measurable and sometimes substantial benefit. We retain three warm-up epochs in the main protocol and do not present it as label-free deployment.

Controlled cost measurement. We profile the frozen backbone, warm-up-only calibration, global-only TTA, and full A^{2}TTA on one H200 GPU over 233 natural yearly graph snapshots. Each snapshot is repeated three times before aggregation. Full A^{2}TTA averages 5.35 ms per window with Online-AN and 11.60 ms with STAEFormer; the frozen backbones require 4.30 and 10.48 ms, respectively. For STAEFormer, peak allocated GPU memory rises from 10,021 to 10,552 MiB. The calibrator has an average of 33.2K

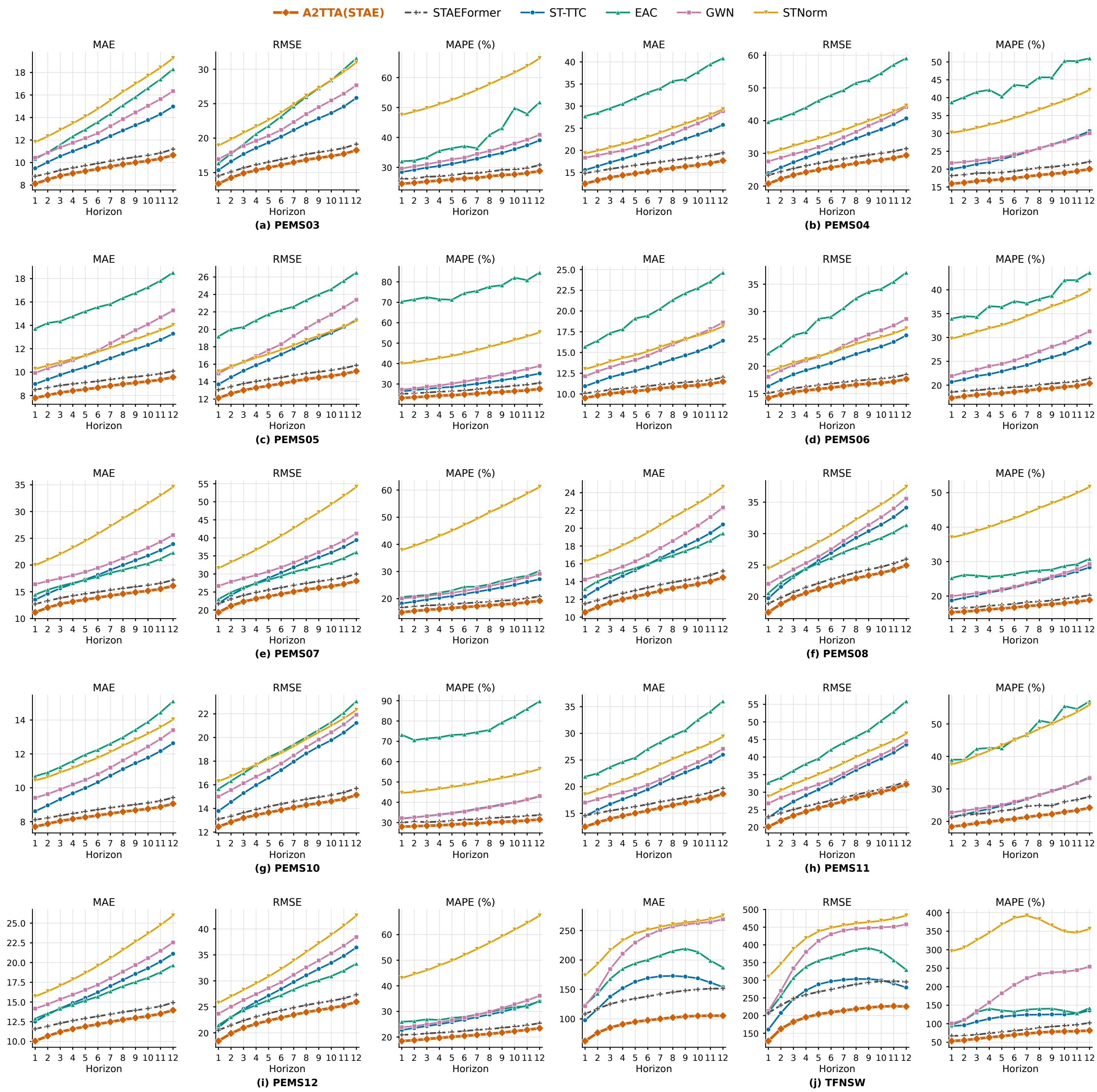

Figure 13: New-sensor per-step results on ten networks.

updated parameters, including the node table, equal to 1.6% of the 2.06M-parameter STAEFormer backbone. Yearly online adaptation averages 2.1 s for Online-AN and 2.0 s for STAEFormer after warm-up.

A.7 Hyperparameter Sensitivity

The Online-AN variant uses 3×10^{-3} on PEMS03, PEMS04, PEMS06, PEMS11, and TFNSW and 10^{-3} elsewhere. The STAEFormer variant uses 3×10^{-3} on TFNSW and 10^{-3} elsewhere.

Figure 17 reports a one-at-a-time singleton-window sweep on PEMS03, PEMS05, and PEMS06 with seeds 51 to 53. The non-learning-rate defaults are three adaptation steps, a feedback pool

Figure 14: New-sensor Avg errors on the remaining eight datasets.

of 512 windows, an update interval of 64 windows, three warm-up epochs, and three local steps. The learning-rate anchor is 10^{-3} for A^{2}TTA(STAE) on all three datasets and for A^{2}TTA(OLAN) on PEMS05; it is 3×10^{-3} for A^{2}TTA(OLAN) on PEMS03 and PEMS06.

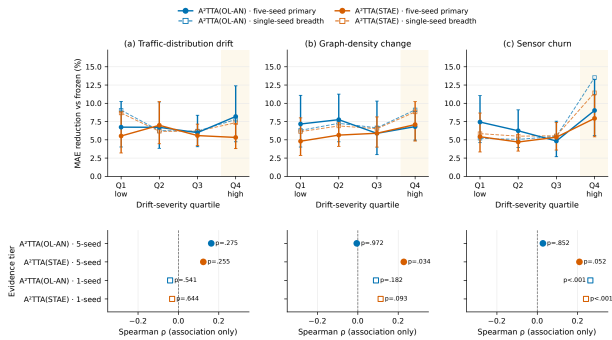

Figure 15: Drift-severity diagnostics.

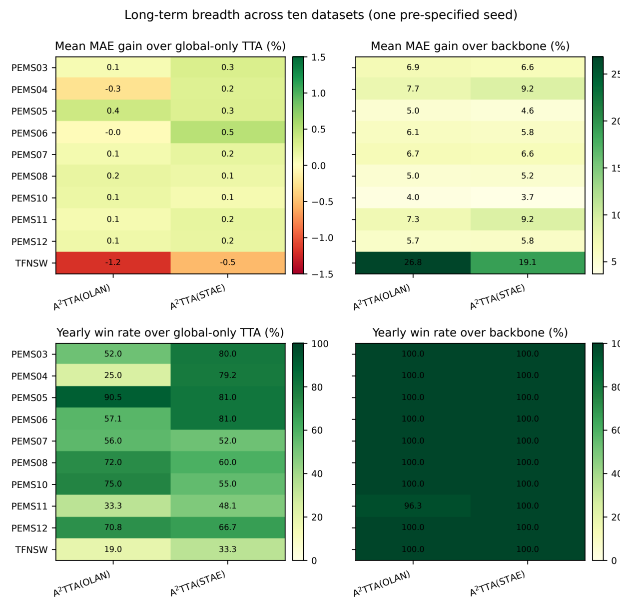

Figure 16: Long-term single-seed breadth check.

Each dataset-backbone pair is normalized by its own paper configuration before averaging, so an aggregate learning-rate point that is a default for only some pairs need not equal zero. Bands show variation across datasets.

The learning-rate changes in Avg-MAE for A^{2}TTA(OLAN) and A^{2}TTA(STAE), respectively, are +0.51% and +0.39% at 10^{-4}, +0.11% and −0.12% at 3×10^{-4}, −0.05% and −1.30% at 3×10^{-3}, and +2.75% and +1.59% at 10^{-2}. With one adaptation step, the changes are +0.35% and −1.98%; with a pool of 128 windows, they are +1.83% and −1.25%. Across the other settings, the largest absolute mean change is 2.31%. Negative values denote lower Avg-MAE. The default configuration therefore offers a reasonable accuracy-computation balance without requiring a separate exhaustive search for each stream.

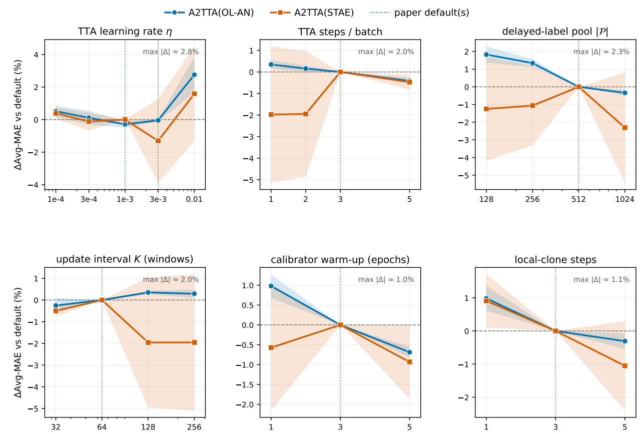

Figure 17: One-at-a-time hyperparameter sensitivity.

A.8 Ablation Study Details

The ablation in Figure 6 starts from the same full stack for both backbones: a frozen forecaster, node-conditioned FiLM, global updates from matured labels, and a disposable context-weighted local clone. "w/o local clone" retains the global update but removes per-window specialization. "FiLM → affine" keeps adaptation and the clone but replaces input-conditioned modulation with a static scale and shift. "w/o online TTA" warm-starts FiLM and then freezes it during the test stream, while "backbone only" removes the calibration stack. All variants process one window at a time, release labels after 12 windows, and trigger global adaptation every 64 windows. Avg-MAE is the mean of the cumulative scores for horizons 1 to 12. Values are means and standard deviations over seeds 51–55, with years averaged within each seed.

Reverting to the frozen backbone increases absolute Avg-MAE by 0.58–1.12 for Online-AN and 0.52–1.87 for STAEFormer, or 6.5% and 6.7% on average relative to the full model. Replacing FiLM with affine costs 0.41–0.72 and 0.29–1.50, respectively, while removing the local clone costs 0.07–0.25 and 0.05–0.53. Freezing FiLM after warm-up raises Avg-MAE by 0.03–0.11 for Online-AN and by 0.09–0.13 for STAEFormer on PEMS03 to PEMS05. STAEFormer on PEMS06 is the exception: the warm-up-only variant reaches 11.46±0.22, compared with 12.18±0.58 for full A^{2}TTA, and is better in four of five seeds. The rerun therefore supports conditional FiLM and local refinement consistently, while online updating is helpful in most, but not all, settings.

A.9 Case Study Details

The case in Figure 9 uses only cohort membership and traffic variance for selection, not relative forecast errors. We take the first sensor in the new/high-variance cohort by metadata order (sensor ID 601213; graph index 14) and its highest-volatility day, then average predictions from seeds 51 and 52. The rerun evaluates one window at a time and releases labels after 12 steps. Panels (a) and (b) cover all 288 five-minute windows of the selected day. At horizon 1, A^{2}TTA reduces trace MAE from 21.51 to 20.00 with Online-AN and from 20.75 to 18.43 with STAEFormer. At horizon 12, the corresponding changes are 34.63 to 25.54 and 23.85 to 21.37.

Panels (c) and (d) report $\mathrm{MAE}_{\mathrm{backbone}} - \mathrm{MAE}_{\mathrm{A^2TTA}}$ for the 366 sensors with valid paired errors, comprising 207 existing and 159 new sensors; positive values indicate improvement. With Online-AN, 206 of 207 existing sensors and all 159 new sensors improve. With STAEFormer, the counts are 204 of 207 and 157 of 159. Mean per-sensor MAE falls by 7.29% relative to Online-AN and 2.62% relative to STAEFormer. The reductions remain similar across existing and new sensors: 7.24% versus 7.36% for Online-AN, and 2.41% versus 2.90% for STAEFormer. This case is illustrative; the multi-dataset, multi-seed experiments provide the aggregate evidence.

A.10 Mechanism Visualization

Figure 18 uses a separate legacy-batched seed-51 PEMS06-2015 A^{2}TTA(STAE) diagnostic. It inspects learned representations and does not contribute to the accuracy tables. We record internal quantities only after delayed feedback is available. For every retained window-sensor pair, we extract the penultimate FiLM features from the persistent global calibrator and its context-specialized disposable clone. The two feature sets are concatenated before fitting one shared Uniform Manifold Approximation and Projection (UMAP). Separate fits would make their coordinate systems incomparable. For the heatmaps, each cell first averages over sensors and then over chronological windows whose forecast target begins in that clock hour.

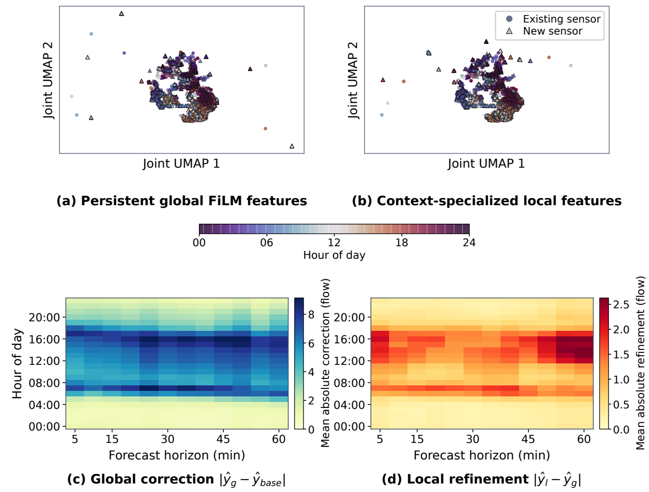

Figure 18: Global and local calibration mechanisms on PEMS06-2015.

The shared projection retains a clear time-of-day organization in both representations, while the local clone makes comparatively small paired movements within this structure. The correction maps give the same two-stage picture in output space: averaged over the captured stream, the global calibrator changes the frozen forecast by 4.40 flow units in absolute value, whereas the local clone adds a 0.90-unit refinement, or 20.4% of the global correction magnitude. Both corrections concentrate in the more active daytime regimes and vary across forecast horizons. Thus, the local clone acts as a targeted adjustment to a stable global adaptation rather than replacing it. This single-run visualization is mechanistic and illustrative. The ablations and multi-seed experiments provide the quantitative evidence.