FACO: TRANSFERRING CROSS-CITY LEARNING EXPERIENCE VIA SPATIOTEMPORAL RELAY FUNCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-city spatiotemporal forecasting aims to transfer predictive experience from data-rich cities to targets with limited observations. Existing methods typically assume continued access to source data, models, or computation, which may fail when historical sources and future targets are temporally separated. We study **Forecasting spatiotemporal observations Across Cities Offline (FACO)**, where source-side learning is completed before the target becomes available and source-private observations, forecasting systems, and source-side computation are inaccessible during target adaptation. This setting raises three challenges: preserving predictive experience beyond source availability, maintaining consistent knowledge across heterogeneous cities and forecasting systems, and identifying trustworthy historical dynamics from limited target evidence. We develop FACO as an offline cross-city forecasting framework that aligns heterogeneous forecasting systems in predictive-functional coordinates, distills source-conditioned temporal evolution, preserves transferable dynamics through a frozen cross-city relay function, and adapts the relay to a future target using functional trust and closed-loop evolution. The relay is realized through a **Spatiotemporal Relay Function (SRF)**, which enables source-derived dynamics to remain usable after source exit. Experiments across heterogeneous cities and forecasting architectures show that FACO achieves the strongest overall performance among the evaluated approaches, reducing MAE by 29.7% over STGP on Innsbruck under the three-day target adaptation setting. These results support representing cross-city predictive experience as source-independent dynamics that remain reusable beyond the availability and architecture of their source systems.

1 INTRODUCTION

Traffic forecasting is a fundamental component of intelligent transportation systems. Modern models capture temporal dependencies and city-specific spatial structure through graph neural networks, temporal convolution, attention mechanisms, and related architectures (Li et al., 2017; Yu et al., 2017; Wu et al., 2019; Shao et al., 2022). As observations remain unevenly distributed across cities, transferring knowledge from mature systems to data-scarce cities is important for forecasting with limited local data.

Cross-city transfer learning and federated forecasting are established paradigms for knowledge reuse. RegionTrans learns transferable source-target regions (Wang et al., 2018), STAN performs spatio-temporal adaptation (Fang et al., 2022), and TSJT exploits target-relevant source knowledge during joint optimization (Zhang et al., 2025). Federated approaches rely on repeated coordination, with FedAvg aggregating iterative client updates (McMahan et al., 2017) and pFedCTP and FedGTP extending this paradigm to personalized and graph-structured traffic forecasting (Zhang et al., 2024; Yang et al., 2024). Recent studies further explore federated cross-city transfer, invariant structural knowledge, traffic foundation models, zero-shot graph forecasting, masked pre-training, prompting, and test-time adaptation (Zeng et al., 2026; Yu et al., 2026; Zhou & Chen, 2026; Xu et al., 2026a; Chen et al., 2026; Xu et al., 2026b; Guo et al., 2026; Zhang et al., 2026; Yin et al., 2026). Across these lines, knowledge reuse generally relies on continued source participation during collaboration or on retaining a source-trained model, representation, or transferable state that remains available when the target is adapted. A broader discussion of related work is provided in Appendix A.

This dependence on continued source participation or retained source-side systems can fail in real deployments. Historical collaborators may participate only temporarily, institutional access may

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

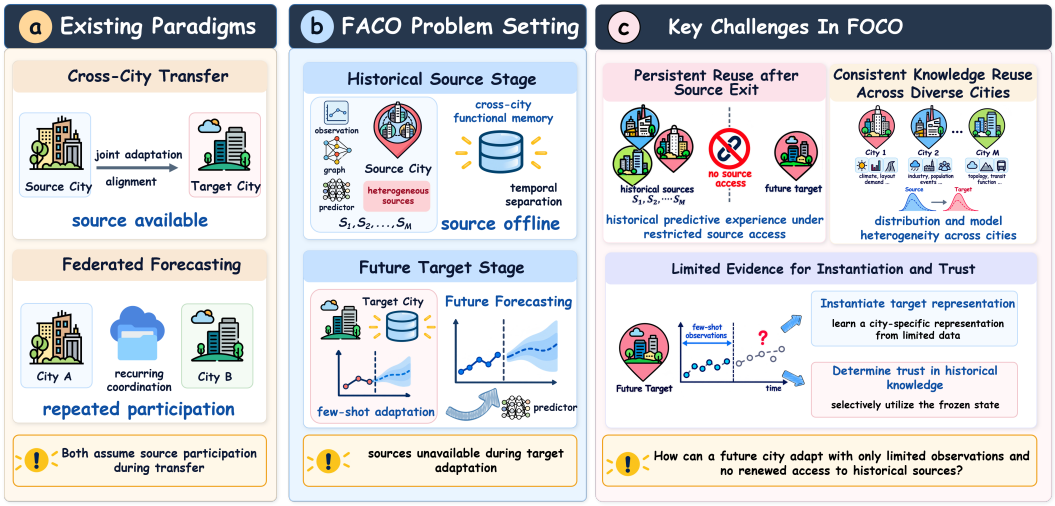


Figure 1: Forecasting spatiotemporal observations Across Cities Offline (FACO).

terminate, or forecasting platforms may leave before a target city is ready. We study this regime as **Forecasting spatiotemporal observations Across Cities Offline (FACO)**. FACO considers historical sources and a future target separated in time, with source-private traffic data, forecasting systems, and source-side computation unavailable during target adaptation. The target therefore adapts using its own forecasting system, local context, and limited observations. Figure 1 illustrates this information pattern.

FACO introduces three challenges. First, **historical predictive experience must remain usable after source exit and under restricted source access**. Second, **cross-city and model heterogeneity complicates persistent reuse**. Cities differ in traffic distributions, sensing structures, semantic context, and forecasting architectures, so transferred knowledge must retain consistent meaning across heterogeneous systems. Third, **limited target observations provide incomplete evidence for instantiating and trusting historical knowledge**. Few-shot adaptation must establish a target-specific representation while determining when historical dynamics remain relevant.

To address these challenges, we develop the **FACO framework**. FACO represents heterogeneous forecasting systems in shared **predictive-functional coordinates** combining a canonical temporal state, calibrated semantic context, and local predictive behavior. Historical sources identify and distill **source-conditioned temporal evolution** in these coordinates, while transferable dynamics are materialized into a frozen cross-city relay before source exit. At the core of this relay is a **Spatiotemporal Relay Function (SRF)**, which preserves source-derived temporal dynamics in a form independent of the original forecasting systems. When a future target arrives, limited local observations instantiate its predictive-functional representation and determine relay participation. **Functional trust** regulates the relayed dynamics according to historical functional support and agreement with the target predictive condition, while **closed-loop temporal evolution** recursively propagates the resulting dynamics across future horizons. This enables historical predictive experience to be reused without renewed source access.

Our contributions are summarized as follows.

- To our knowledge, we are the first to formulate and study **Forecasting spatiotemporal observations Across Cities Offline (FACO)**, characterized by temporal separation between historical sources and a future target, restricted access to source-private information during target adaptation, and few-shot adaptation from limited target observations. FACO formalizes cross-city knowledge reuse after historical source systems become unavailable.
- We develop the **FACO framework**, which aligns heterogeneous forecasting systems in shared predictive-functional coordinates, distills source-conditioned temporal evolution, preserves transferable dynamics through a frozen cross-city relay function, and adapts them to future targets through functional trust and closed-loop evolution. A **Spatiotemporal Relay Function (SRF)** provides the core mechanism for retaining source-derived temporal dynamics beyond source exit.
- We construct a FACO benchmark and evaluate heterogeneous cities and forecasting systems. FACO achieves the strongest overall forecasting performance among the evaluated approaches

and yields substantial improvements under limited target observations, including a 29.7% MAE reduction over STGP on Innsbruck in the three-day setting. The results support representing transferable cross-city knowledge as source-independent predictive dynamics reusable beyond the lifetime and architecture of their source systems.

2 Problem Definition

We formalize **Forecasting spatiotemporal observations Across Cities Offline (FACO)** under temporal separation and restricted information access. Let $\mathcal{C}_s$ denote the set of historical source cities and let T denote a future target city. Each source environment is

$$\mathcal{E}_s = \big(\mathcal{D}_s, G_s, B_s, S_s\big), \qquad s \in \mathcal{C}_s, \tag{1}$$

where $\mathcal{D}_s$ is the local traffic history, G_s the local graph, B_s the city-local forecasting system, and S_s the available urban context. Source cities may differ in sensing structure, forecasting architecture, parameterization, and latent representation, while their private data and local systems remain confined to the corresponding source environments.

The future target is not assumed to coexist with the historical sources. Before source exit, the only source-derived information that may remain available for future reuse is a bounded artifact,

$$a_s = \Gamma_s\left(\mathcal{E}_s\right), \qquad a_s \in \mathcal{A}_{\mathrm{rel}}, \tag{2}$$

$$\Omega = \mathrm{Freeze}\left(\{a_s\}_{s\in\mathcal{C}_s}\right), \tag{3}$$

where $\mathcal{A}_{\mathrm{rel}}$ denotes the admissible release space and Ω is the frozen cross-city state available after source participation ends. Once the sources have exited, their raw observations, private forecasting systems, local graphs, and source-side computation cannot be directly accessed or queried during target adaptation and inference.

When the future target city becomes available, its accessible information is

$$\mathcal{I}_T = \left\{\mathcal{D}_T^K, G_T, B_T, S_T, \Omega\right\}, \tag{4}$$

where $\mathcal{D}_T^K$ denotes the limited target-local observations available for few-shot adaptation. The target forecasting system B_T is local to the target city and need not share architecture or representation with any historical source system.

An admissible target forecasting rule uses only $\mathcal{I}_T$,

$$\widehat{Y}_T = F_T\left(\mathcal{I}_T\right), \tag{5}$$

and aims to minimize the expected target forecasting risk

$$\mathfrak{R}_T(F_T) = \mathbb{E}\left[\ell\left(Y_T, \widehat{Y}_T\right)\right]. \tag{6}$$

Here, F_T may be any forecasting procedure that satisfies the FACO temporal and information constraints.

FACO specifies the deployment constraints while leaving the forms of a_s, Ω, and F_T method-dependent. In the following section, **Spatiotemporal Relay Function (SRF)** provides one realization through a frozen predictive-functional relay and target-specific few-shot instantiation.

3 Spatiotemporal Relay Function for FACO

The FACO framework is organized into four modules: canonical predictive-functional coordinates, source-conditioned temporal evolution distillation, a frozen cross-city relay function, and trust-conditioned target adaptation. Figure 2 summarizes this sequence.

3.1 Canonical Predictive-Functional Coordinates

FACO represents each city in shared predictive-functional coordinates comprising a canonical temporal state, aligned semantic context, and predictive behavior from the local forecasting system. The same construction applies to all sources and the future target, placing heterogeneous cities and forecasting architectures in a common functional space.

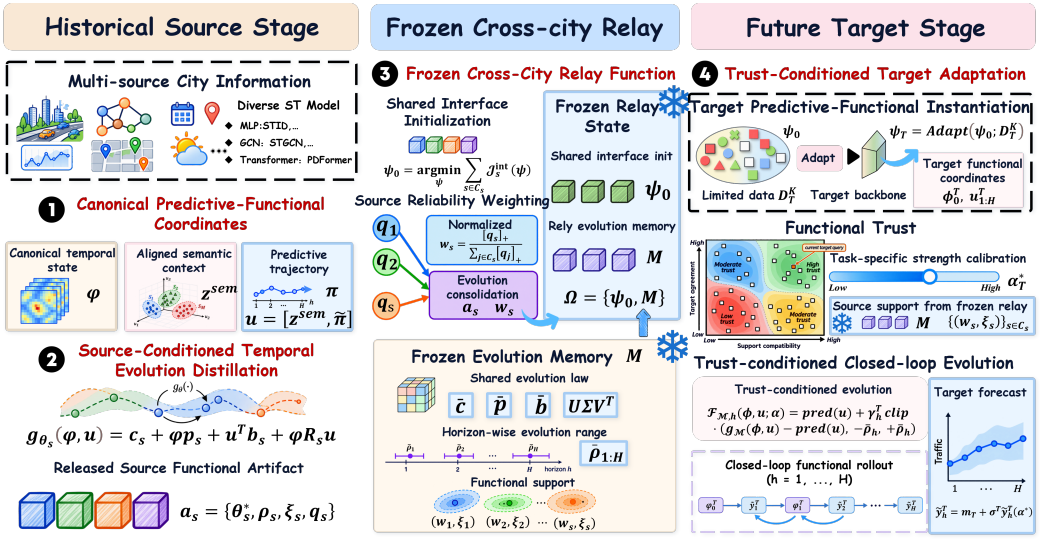


Figure 2: Overview of SRF for FACO. Historical sources learn conditional temporal evolution in shared predictive-functional coordinates and release finite artifacts. A frozen cross-city relay consolidates these artifacts before source exit; a future target instantiates the relay from local context and few-shot observations, then generates forecasts through trust-conditioned closed-loop evolution.

For city $v \in \mathcal{C}_s \cup \{T\}$, node n, and forecast origin t, let H denote the forecasting horizon and let $h \in \{1, \dots, H\}$ be its horizon index. Let $y^v_{t,n}$ denote the traffic observation and let $\mathbf{x}^v_{t,n} \in \mathbb{R}^L$ collect the most recent L observations, where L is the history length. We normalize this history using a local anchor and scale and project it onto a deterministic orthonormal temporal basis $Q \in \mathbb{R}^{L\times L}$,

$$m^v_{t,n} = y^v_{t,n}, \qquad \sigma^v_{t,n} = \max\left(\frac{1}{L}\sum_{\ell=1}^{L} x^v_{t,n,\ell}, 1\right), \\ \boldsymbol{\phi}^v_{t,n,0} = \left(\frac{\mathbf{x}^v_{t,n} - m^v_{t,n}\mathbf{1}}{\sigma^v_{t,n}}\right) Q^\top. \tag{7}$$

Here, $m^v_{t,n}$ and $\sigma^v_{t,n}$ are the local anchor and scale, and $\mathbf{1} \in \mathbb{R}^L$ is the all-ones vector. Because Q is fixed and shared across cities, $\boldsymbol{\phi}^v_{t,n,0}$ describes recent temporal behavior in a common temporal coordinate system.

The urban context S_v is organized under a shared semantic schema containing calendar information, static node attributes, and causally available exogenous context, yielding a raw descriptor $\mathbf{z}^{v,\mathrm{raw}}_{t,n,h}$. To absorb city-specific scale and response differences while preserving this common semantic structure, FACO applies a lightweight element-wise alignment,

$$\mathbf{z}^{v,\mathrm{sem}}_{t,n,h} = \mathcal{A}_{\eta_v}\left(\mathbf{z}^{v,\mathrm{raw}}_{t,n,h}\right), \qquad \mathcal{A}_{\eta_v}(\mathbf{r}) = \mathbf{g}_v \odot \mathbf{r} + \mathbf{b}_v, \qquad \eta_v = \{\mathbf{g}_v, \mathbf{b}_v\}, \tag{8}$$

Here, $\mathcal{A}_{\eta_v}$ denotes the city-specific semantic alignment map, with element-wise gain $\mathbf{g}_v$ and bias $\mathbf{b}_v$; $\mathbf{r}$ is a generic semantic input and $\odot$ denotes element-wise multiplication.

The predictive interface Ψ_{ω_v} maps the behavior of the local forecasting system into an H-step trajectory in the shared functional space,

$$\boldsymbol{\pi}^v_{t,n} = \Psi_{\omega_v}\left(B_v, G_v, \mathbf{x}^v_{t,n}, \mathbf{z}^{v,\mathrm{sem}}_{t,n,1:H}\right), \qquad \psi_v = (\eta_v, \omega_v), \tag{9}$$

where G_v is the city-local graph or spatial structure defined in the FACO environment. We denote ψ_v as the city-specific predictive-functional interface state, combining semantic alignment η_v and predictive mapping ω_v.

Its horizon-h prediction is expressed in the same local coordinates as the temporal state and combined with the aligned semantic context,

$$\widetilde{\pi}^v_{t,n,h} = \frac{\pi^v_{t,n,h} - m^v_{t,n}}{\sigma^v_{t,n}}, \qquad \mathbf{u}^v_{t,n,h} = \left[\mathbf{z}^{v,\mathrm{sem}}_{t,n,h}, \widetilde{\pi}^v_{t,n,h}\right]. \tag{10}$$

The predictive component of $\mathbf{u}^v_{t,n,h}$ is denoted by $\mathrm{pred}(\mathbf{u}^v_{t,n,h}) = \widetilde{\pi}^v_{t,n,h}$. We refer to $\left(\boldsymbol{\phi}^v_{t,n,0}, \mathbf{u}^v_{t,n,1:H}\right)$ as the predictive-functional coordinates of a forecasting window. The canonical temporal state provides a common description of recent dynamics, semantic alignment places contextual variables in comparable coordinates, and the predictive interface exposes local forecasting behavior in the shared functional space. These coordinates form the space in which FACO learns and relays conditional temporal evolution. Full construction details are given in Appendix B.1.

3.2 Source-Conditioned Temporal Evolution Distillation

Given the shared predictive-functional coordinates, each historical source learns how the canonical temporal state evolves under semantic and predictive conditions. For source city s, FACO parameterizes the conditional evolution law as

$$g_{\theta_s}(\boldsymbol{\phi}, \mathbf{u}) = c_s + \boldsymbol{\phi}\mathbf{p}_s + \mathbf{u}^\top \mathbf{b}_s + \boldsymbol{\phi} R_s \mathbf{u}, \tag{11}$$

where θ_s collects the parameters of the source evolution law, c_s is a transition bias, $\mathbf{p}_s$ and $\mathbf{b}_s$ describe the contributions of the temporal state and conditioning variables, and R_s captures their interaction. This parameterization expresses the learned dynamics directly in the shared predictive-functional coordinates.

At horizon h, source evolution is expressed relative to the predictive component of $\mathbf{u}$,

$$\begin{aligned} \widetilde{y}^s_{t,n,h} &= \mathcal{E}_{\theta_s,h}\left(\boldsymbol{\phi}^s_{t,n,h-1}, \mathbf{u}^s_{t,n,h}\right) \\ &= \mathrm{pred}(\mathbf{u}^s_{t,n,h}) + \mathrm{clip}\left(g_{\theta_s}\left(\boldsymbol{\phi}^s_{t,n,h-1}, \mathbf{u}^s_{t,n,h}\right) - \mathrm{pred}(\mathbf{u}^s_{t,n,h}), -\rho^{\mathrm{fit}}_{s,h}, \rho^{\mathrm{fit}}_{s,h}\right). \end{aligned} \tag{12}$$

Here, $\widetilde{y}^s_{t,n,h}$ is the evolved normalized value, $\mathcal{E}_{\theta_s,h}$ is the horizon-h source transition, $\mathrm{clip}(x, a, b) = \min(\max(x, a), b)$ truncates x to $[a, b]$, and $\rho^{\mathrm{fit}}_{s,h}$ is the horizon-specific predictive-relative range used during source fitting. The transition combines the current temporal state, aligned semantic context, and local predictive behavior within the source-supported fitting range.

Evolution proceeds recursively across the forecasting horizon. Let S_ϕ and $\mathbf{e}_\phi$ denote the deterministic shift and injection operators induced by the shared temporal basis Q. Starting from $\boldsymbol{\phi}^s_{t,n,0}$, the canonical state is updated by

$$\boldsymbol{\phi}^s_{t,n,h} = \boldsymbol{\phi}^s_{t,n,h-1} S_\phi + \widetilde{y}^s_{t,n,h} \mathbf{e}_\phi^\top, \qquad h = 1, \ldots, H. \tag{13}$$

Each evolved value is incorporated into the state governing the next horizon, producing closed-loop temporal evolution across the forecasting trajectory.

Let $\mathcal{W}_s$ denote the chronological source forecasting windows used for fitting. The source evolution parameters are learned with trajectory-level supervision,

$$\theta^\star_s = \arg\min_{\theta_s} \sum_{(t,n)\in\mathcal{W}_s} \mathcal{L}_s\left(\mathrm{Rollout}\left(\mathcal{E}_{\theta_s,1:H}; \boldsymbol{\phi}^s_{t,n,0}, \mathbf{u}^s_{t,n,1:H}\right), \widetilde{\mathbf{y}}^{s,*}_{t,n,1:H}\right). \tag{14}$$

Here, Rollout recursively applies the transition and state update across H horizons, $\widetilde{\mathbf{y}}^{s,*}_{t,n,1:H}$ is the normalized ground-truth source trajectory, and $\mathcal{L}_s$ is the source trajectory loss.

After fitting, held-out source windows provide the horizon-wise supported evolution range $\boldsymbol{\rho}_s = [\rho_{s,1}, \ldots, \rho_{s,H}]$. Source s then contributes the bounded functional artifact

$$a_s = \{\theta^\star_s, \boldsymbol{\rho}_s, \xi_s, q_s\}, \tag{15}$$

where ξ_s summarizes the supported region of the shared functional space and q_s records source-side reliability. These finite source releases are consolidated into the frozen cross-city relay in Sec. 3.3. Detailed parameterization, range estimation, and rollout construction are given in Appendix B.2.

3.3 Frozen Cross-City Relay Function

Because all source evolution laws are expressed in the same predictive-functional coordinates, FACO consolidates their transferable dynamics into a source-independent relay before the future target appears. The relay preserves three forms of source-side evidence: predictive reliability, functional support, and conditional temporal evolution.

For source s, let q_s denote the predictive value of its learned evolution law on held-out source windows relative to the source predictive trajectory. The relay weight is

$$w_s = \begin{cases} \dfrac{[q_s]_+}{\sum_{j\in\mathcal{C}_s}[q_j]_+}, & \sum_{j\in\mathcal{C}_s}[q_j]_+ > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad [x]_+ = \max(x, 0). \tag{16}$$

When all source reliabilities are nonpositive, every w_s is zero and cross-city evolution is disabled.

FACO also records the functional region represented by each source. For a source forecasting window, define

$$\mathbf{d}^s_{t,n} = \left[\boldsymbol{\phi}^s_{t,n,0}, \frac{1}{H}\sum_{h=1}^{H}\mathbf{u}^s_{t,n,h}\right], \qquad \xi_s = \{\boldsymbol{\mu}^{\mathrm{sup}}_s, \boldsymbol{\sigma}^{\mathrm{sup}}_s\}, \tag{17}$$

where $\boldsymbol{\mu}^{\mathrm{sup}}_s$ and $\boldsymbol{\sigma}^{\mathrm{sup}}_s$ are the coordinate-wise mean and stabilized standard deviation of the source descriptors, respectively. The support summary ξ_s is retained for target-side trust estimation.

The shared coordinates make corresponding source evolution parameters directly comparable. FACO forms their reliability-weighted aggregate as

$$\begin{aligned} \bar{c} &= \sum_{s\in\mathcal{C}_s} w_s c_s, \qquad \bar{\mathbf{p}} = \sum_{s\in\mathcal{C}_s} w_s \mathbf{p}_s, \qquad \bar{\mathbf{b}} = \sum_{s\in\mathcal{C}_s} w_s \mathbf{b}_s, \\ \bar{R}_{\mathrm{full}} &= \sum_{s\in\mathcal{C}_s} w_s R_s, \quad \bar{\rho}_h = \sum_{s\in\mathcal{C}_s} w_s \rho_{s,h}, \qquad h = 1, \ldots, H. \end{aligned} \tag{18}$$

The dominant interaction structure is retained as

$$\bar{R} = U\Sigma V^\top \approx \bar{R}_{\mathrm{full}}, \qquad \bar{\boldsymbol{\rho}} = [\bar{\rho}_1, \ldots, \bar{\rho}_H]. \tag{19}$$

Here, U, Σ, and V are the retained singular factors. These quantities define the frozen cross-city evolution state

$$\mathcal{M} = \left\{\bar{c}, \bar{\mathbf{p}}, \bar{\mathbf{b}}, \bar{R}, \bar{\boldsymbol{\rho}}, \{(w_s, \xi_s)\}_{s\in\mathcal{C}_s}\right\}. \tag{20}$$

The predictive-functional interface is shared through source-side joint learning. Let $\mathcal{J}^{\mathrm{int}}_s(\psi)$ denote the interface-learning objective contributed by source s. The shared initialization is

$$\psi_0 = \arg\min_{\psi} \sum_{s\in\mathcal{C}_s} \mathcal{J}^{\mathrm{int}}_s(\psi). \tag{21}$$

In FACO, the shared artifact a_0 in the FACO formulation is instantiated by ψ_0. The complete frozen relay state is therefore

$$\Omega = \{\psi_0, \mathcal{M}\}. \tag{22}$$

Here, ψ_0 provides the source-trained initialization of the common semantic-alignment and predictive-interface family, while $\mathcal{M}$ carries the relayed conditional temporal dynamics and their support evidence. The reliability weights w_s govern evolution consolidation in $\mathcal{M}$; the shared interface ψ_0 is learned jointly across source cities. Both components are frozen before target arrival. Detailed reliability estimation, support construction, low-rank relay compression, and source-side interface learning are provided in Appendix B.3.

3.4 Trust-Conditioned Target Adaptation

When the future target city T becomes available, FACO keeps the frozen relay state $\Omega = \{\psi_0, \mathcal{M}\}$ fixed and uses the limited target observations $\mathcal{D}^K_T$ to instantiate the target predictive-functional interface,

$$\psi_T = \mathrm{Adapt}\left(\psi_0; \mathcal{D}^K_T\right), \qquad \psi_T = (\eta_T, \omega_T). \tag{23}$$

Here, Adapt denotes few-shot specialization of the predictive-functional interface initialized by ψ_0 using target-local observations; the frozen evolution state $\mathcal{M}$ remains unchanged. The resulting ψ_T constructs the target predictive-functional coordinates $\left(\boldsymbol{\phi}^T_{t,n,0}, \mathbf{u}^T_{t,n,1:H}\right)$ according to Sec. 3.1.

When cross-city evolution is active, FACO forms the target functional descriptor $\mathbf{d}^T_{t,n} = [\boldsymbol{\phi}^T_{t,n,0}, H^{-1}\sum_{h=1}^{H}\mathbf{u}^T_{t,n,h}]$. Let d^T_{sup} denote its standardized distance to the nearest active source support recorded in $\mathcal{M}$. The frozen evolution memory defines

$$g_{\mathcal{M}}(\boldsymbol{\phi}, \mathbf{u}) = \bar{c} + \boldsymbol{\phi}\bar{\mathbf{p}} + \mathbf{u}^\top\bar{\mathbf{b}} + \boldsymbol{\phi}\bar{R}\mathbf{u}. \tag{24}$$

Let $r^T_{t,n,1:H}$ denote the relay responses obtained by evaluating $g_{\mathcal{M}}$ along the target predictive trajectory. Their trajectory-level disagreement with the target predictive condition is

$$d^T_{\mathrm{agr}} = \frac{1}{H}\sum_{h=1}^{H} \left| r^T_{t,n,h} - \mathrm{pred}\left(\mathbf{u}^T_{t,n,h}\right)\right|. \tag{25}$$

For an active relay, its participation at horizon h is governed by

$$\gamma^T_{t,n,h}(\alpha) = \alpha\, g^{\mathrm{sup}}\left(d^T_{\mathrm{sup}}\right) g^{\mathrm{agr}}\left(d^T_{\mathrm{agr}}\right) g^{\mathrm{hor}}(h), \qquad \alpha \in [0,1], \tag{26}$$

where g^{sup}, g^{agr}, and g^{hor} are bounded in $[0,1]$ and measure historical functional support, trajectory-level agreement, and horizon-dependent confidence, respectively. The scalar α controls the overall target-specific participation of the frozen relay. If cross-city evolution is disabled, FACO sets $\gamma^T_{t,n,h}(\alpha) = 0$.

FACO then applies the trust-conditioned evolution operator

$$\mathcal{F}^T_{\mathcal{M},t,n,h}\left(\boldsymbol{\phi}, \mathbf{u}; \alpha\right) = \mathrm{pred}(\mathbf{u}) + \gamma^T_{t,n,h}(\alpha)\,\mathrm{clip}\left(g_{\mathcal{M}}(\boldsymbol{\phi}, \mathbf{u}) - \mathrm{pred}(\mathbf{u}), -\bar{\rho}_h, \bar{\rho}_h\right). \tag{27}$$

The target trajectory is generated recursively,

$$\widetilde{y}^T_{t,n,h}(\alpha) = \mathcal{F}^T_{\mathcal{M},t,n,h}\left(\boldsymbol{\phi}^T_{t,n,h-1}(\alpha), \mathbf{u}^T_{t,n,h}; \alpha\right), \tag{28}$$

$$\boldsymbol{\phi}^T_{t,n,h}(\alpha) = \boldsymbol{\phi}^T_{t,n,h-1}(\alpha) S_\phi + \widetilde{y}^T_{t,n,h}(\alpha)\mathbf{e}_\phi^\top, \qquad h = 1, \ldots, H. \tag{29}$$

The recursion starts from $\boldsymbol{\phi}^T_{t,n,0}(\alpha) = \boldsymbol{\phi}^T_{t,n,0}$. Here, $\widetilde{y}^T_{t,n,h}(\alpha)$ is the evolved normalized target value. Thus relay participation at one horizon also enters the state governing subsequent evolution.

Finally, the target-local observations determine the overall relay participation,

$$\alpha^\star_T = \arg\min_{\alpha \in \mathcal{A}_\alpha} \mathcal{J}_T\left(\alpha; \mathcal{D}^K_T\right), \qquad \mathcal{A}_\alpha \subset [0,1], \tag{30}$$

where $\mathcal{A}_\alpha$ is a finite target-side candidate set and $\mathcal{J}_T$ evaluates complete closed-loop trajectories on calibration windows constructed from $\mathcal{D}^K_T$. Hence the limited target observations instantiate both ψ_T and $\alpha^\star_T$, while the historical evolution memory $\mathcal{M}$ remains fixed. Detailed target adaptation, trust construction, and calibration are given in Appendix B.4.

4 Experiments

4.1 Experimental Protocol and Evaluation Settings

We construct the FACO benchmark from the IUTF traffic–weather resource (Lin et al., 2025), augmented with calendar and daylight-saving variables, Geofabrik POI semantics, GHSL urban context, and city-local spatial supports. The benchmark contains 18 eligible cities, including 16 target cities and two source-only cities (Kassel and Utrecht). Under leave-one-city-out (LOCO) evaluation, each target uses the remaining 17 cities as historical sources. Traffic is aggregated to 15-minute intervals, with the latest $L = 12$ observations used to predict the next $H = 12$.

We evaluate one-day and three-day target adaptation settings. The available target prefix is split chronologically for adaptation and validation, while subsequent observations are reserved for testing. The frozen relay is constructed before target availability, and future weather is excluded. We compare twelve baselines under the same protocol, with city-local backbones drawn from STGCN, STID, and PDFormer (Jiang et al., 2023a). We report city-level MAE and WAPE and their macro averages across target cities. Detailed data construction, city splits, baseline implementations, and hyperparameter settings are provided in Appendices C.1–C.3.

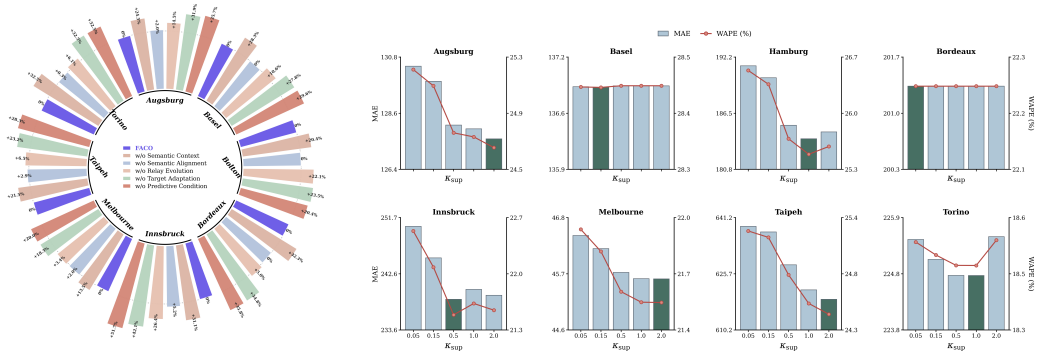
4.2 Main Results

Table 1 reports the main results on the 16-city FACO benchmark under the three-day target adaptation setting. FACO achieves the best MAE and WAPE in every target city, reducing the macro-averaged MAE and WAPE by approximately 14.1% and 14.6% relative to the best competing result for each target, and by 25.2% and 23.8% relative to Local-only forecasting. The gains are consistent across heterogeneous cities, including Marseille ($581.5/0.272$ to $385.8/0.181$), Innsbruck ($364.4/0.328$ to $240.1/0.216$), Bolton ($104.7/0.100$ to $65.1/0.062$), and Melbourne ($54.7/0.258$ to $45.6/0.215$). Although the closest competitor varies across cities among STGP, ST-MPPT, TPB, MTPB, SERT, and ST-GFSL, FACO remains lower in every column, indicating consistent benefits from the source-conditioned functional relay and target-side adaptation across the full heterogeneous benchmark.

Table 1: Main results on the 16-city FACO benchmark with a three-day target adaptation prefix. Each block contains eight target cities, and each city reports MAE and WAPE. Lower is better; the best result for each city and metric is bold and the second-best is underlined.

Method	Augs.		Basel		Bern		Bolton		Bord.		Bremen		Const.		Hamb.	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
CMSVC (2025)	215.8	0.415	182.9	0.375	197.4	0.452	73.6	0.070	252.6	0.278	254.0	0.265	118.3	0.280	258.1	0.355
CrossTReS (2022)	155.6	0.299	197.7	0.406	174.3	0.399	77.4	0.074	276.1	0.304	270.0	0.282	127.4	0.301	258.2	0.355
DASTNet (2022)	281.8	0.542	282.8	0.581	302.0	0.692	124.5	0.119	476.0	0.523	430.9	0.450	199.4	0.471	354.8	0.488
FEPCross (2024)	368.1	0.708	193.0	0.396	193.4	0.443	82.0	0.079	255.4	0.281	278.2	0.291	136.5	0.323	260.4	0.358
MTPB (2025)	231.9	0.446	172.8	0.355	200.0	0.458	72.4	0.069	240.1	0.264	248.1	0.259	120.1	0.284	235.0	0.323
pFedCTP (2024)	352.8	0.678	279.3	0.573	319.3	0.732	157.4	0.151	466.5	0.513	374.8	0.392	203.4	0.481	380.0	0.523
SERT (2025)	326.4	0.628	169.7	0.348	187.3	0.429	69.7	0.067	240.3	0.264	252.7	0.264	120.4	0.284	244.3	0.336
ST-GFSL (2022)	156.8	0.302	192.4	0.395	171.8	0.394	75.8	0.073	274.6	0.302	267.6	0.280	124.7	0.295	263.6	0.363
ST-MPPT (2026)	199.9	0.384	170.7	0.350	180.8	0.414	66.4	0.064	235.5	0.259	247.2	0.258	118.5	0.280	238.6	0.328
STGP(2024)	145.8	0.280	179.2	0.368	202.0	0.463	68.0	0.065	243.7	0.268	248.9	0.260	117.9	0.279	233.8	0.322
TPB (2023)	337.0	0.648	168.4	0.346	188.1	0.431	68.5	0.066	239.5	0.263	249.7	0.261	120.6	0.285	234.0	0.322
TransGTR (2023)	188.8	0.363	189.7	0.389	222.0	0.509	82.2	0.079	281.5	0.310	281.4	0.294	131.3	0.310	265.3	0.365
A2TTA (2026)	900.7	1.732	201.7	0.414	187.7	0.430	93.4	0.089	242.5	0.267	247.2	0.258	120.4	0.284	299.4	0.412
Local-only	173.7	0.336	177.4	0.368	179.4	0.415	104.7	0.100	273.4	0.302	273.9	0.288	123.6	0.293	239.1	0.331
FACO (ours)	**128.0**	**0.248**	**136.9**	**0.284**	**151.8**	**0.351**	**65.1**	**0.062**	**201.3**	**0.222**	**213.7**	**0.225**	**92.9**	**0.220**	**183.9**	**0.255**

Method	Inns.		Mars.		Melb.		Speyer		Taipeh		Torino		Wolfs.		Zurich	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
CMSVC (2025)	406.7	0.362	460.5	0.217	50.3	0.236	173.1	0.248	677.8	0.267	284.5	0.233	370.8	0.409	244.1	0.234
CrossTReS (2022)	377.8	0.337	490.6	0.231	52.3	0.246	184.6	0.265	679.5	0.268	300.3	0.246	343.0	0.378	251.2	0.241
DASTNet (2022)	538.1	0.479	687.5	0.323	85.4	0.401	306.2	0.439	1058.5	0.417	524.2	0.429	476.8	0.526	428.7	0.411
FEPCross (2024)	421.7	0.376	490.4	0.231	55.4	0.260	191.2	0.274	709.3	0.279	289.8	0.237	343.3	0.379	256.7	0.246
MTPB (2025)	359.0	0.320	448.9	0.211	48.9	0.230	167.2	0.240	663.9	0.262	276.4	0.226	338.8	0.374	231.7	0.222
pFedCTP (2024)	449.4	0.400	734.1	0.345	83.6	0.393	268.2	0.385	1329.1	0.524	414.9	0.340	460.5	0.508	449.3	0.431
SERT (2025)	373.5	0.333	459.5	0.216	49.1	0.231	164.7	0.236	664.0	0.262	270.9	0.222	333.2	0.367	237.0	0.227
ST-GFSL (2022)	367.3	0.327	487.8	0.229	51.8	0.244	184.6	0.265	679.9	0.268	296.2	0.242	344.3	0.380	247.9	0.238
ST-MPPT (2026)	370.7	0.330	454.2	0.214	48.4	0.227	165.6	0.238	643.5	0.254	271.6	0.222	308.5	0.340	232.6	0.223
STGP(2024)	341.4	0.304	454.7	0.214	49.6	0.233	165.8	0.238	649.5	0.256	278.9	0.228	333.5	0.368	234.2	0.225
TPB (2023)	366.6	0.327	454.8	0.214	49.0	0.230	167.8	0.241	666.8	0.263	279.5	0.229	327.0	0.361	233.1	0.224
TransGTR (2023)	409.7	0.365	474.2	0.223	57.2	0.269	190.4	0.273	775.6	0.306	328.0	0.268	303.8	0.335	283.2	0.272
A2TTA (2026)	444.3	0.396	460.5	0.217	49.6	0.233	161.7	0.232	705.0	0.278	274.5	0.225	336.5	0.371	231.8	0.222
Local-only	364.4	0.328	581.5	0.272	54.7	0.258	179.2	0.258	796.9	0.315	297.9	0.245	344.4	0.382	242.8	0.234
FACO (ours)	**240.1**	**0.216**	**385.8**	**0.181**	**45.6**	**0.215**	**135.3**	**0.195**	**621.3**	**0.246**	**224.8**	**0.185**	**282.1**	**0.313**	**189.1**	**0.182**



(a) Ablation study.

(b) Gate-temperature sensitivity across the eight cities.

Figure 3: Ablation and hyperparameter sensitivity under the frozen protocol. The radial panel reports city-level MAE changes relative to FACO, while the sensitivity panel reports MAE and WAPE over the five gate-temperature settings.

4.3 Ablation Study and Hyperparameter Sensitivity

Figure 3 shows that removing major FACO components consistently worsens performance, with particularly clear degradation when predictive conditioning or target adaptation is disabled. This confirms the importance of target-specific predictive-functional instantiation and relayed temporal evolution. Sensitivity analysis of the support-gate temperature κ_{sup} shows stable performance across a broad range of values, indicating that FACO is relatively robust to this hyperparameter across heterogeneous target cities.

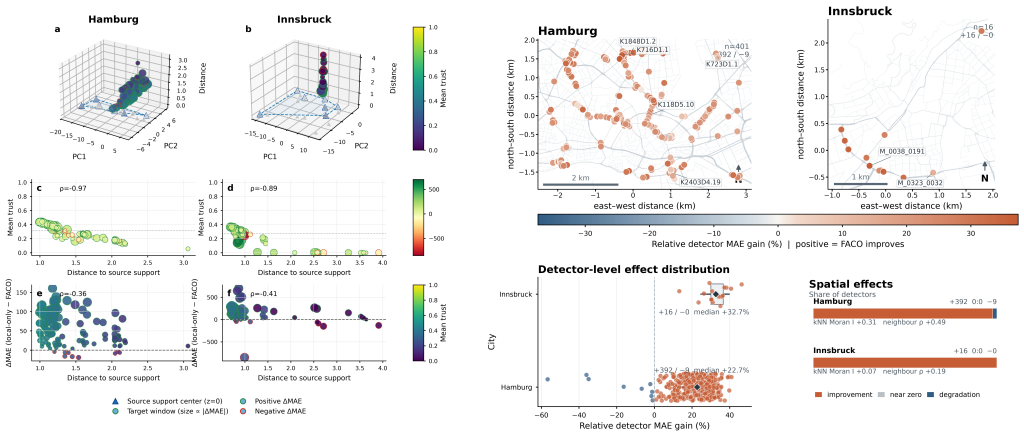


Figure 4: Functional support and spatial effects of FACO. Left: target test windows and source support in functional space, with support distance related to trust and MAE change. Right: detector locations colored by their relative MAE gain over local-only, followed by the detector-level effect distributions. Results use three-day target adaptation.

4.4 Case Study

Figure 4 provides window-level and detector-level views of FACO in Hamburg and Innsbruck. In the left block, each target forecasting window is represented by its first two functional coordinates, its distance from the source-supported region, and its functional trust. Functional trust decreases as support distance increases, with Spearman correlations of $\rho = -0.97$ for Hamburg and $\rho = -0.89$ for Innsbruck. The lower-left panels further show that windows closer to the source support tend to achieve larger MAE reductions, with correlations of $\rho = -0.36$ in Hamburg and $\rho = -0.41$ in Innsbruck. Higher trust and larger gains are therefore concentrated near historically supported functional regions.

The right block resolves the same effect at detector level. In Hamburg, FACO improves 392 of 401 valid detectors, with a median relative MAE gain of 22.7%. The nine deteriorating detectors are spatially concentrated around locations including Reeperbahn and Osakaallee, and the detector-level spatial summary shows positive local coherence, with a k-nearest-neighbour Moran statistic of 0.31 and a neighbour-effect correlation of 0.49. Innsbruck shows improvement for all 16 detectors, with a median relative MAE gain of 32.7%. Its spatial coherence statistics are smaller (0.07 and 0.19) because the detector set is compact and nearly uniformly improved. Overall, the results show that proximity to historical functional support is associated with higher relay trust and larger forecasting gains, which are also reflected in the spatial distribution of detector-level improvements.

5 Limitations

FACO currently assumes that a future target provides a small amount of local observation for target-specific instantiation and trust calibration. Fully zero-shot deployment, in which no target observations are available before forecasting, is therefore outside the current setting. Extending FACO to infer target predictive-functional coordinates and relay trust without any target-side supervision is an important direction for future work.

6 Conclusion

We introduced **FACO**, an offline cross-city forecasting setting and framework for reusing predictive experience after historical source systems have exited. FACO aligns heterogeneous forecasting systems in predictive-functional coordinates, distills source-conditioned temporal evolution, preserves transferable dynamics through a frozen cross-city relay function with SRF as its core mechanism, and adapts the relay to future targets through functional trust and closed-loop evolution. Experiments across 16 target cities and heterogeneous forecasting backbones show consistent improvements over the evaluated approaches under limited target observations. These results support the view that cross-city predictive experience can be represented as source-independent functional dynamics that remain reusable beyond the availability and architecture of their source systems.

AI USE STATEMENT

this work, we used generative AI tools for language editing, document organization, LaTeX troubleshooting, and suggestions for code and formatting. We did not use generative AI tools to generate experimental data, fabricate or alter results, make unverified scientific claims, or replace the authors decisions about the problem formulation, methodology, or evaluation. Other required disclosure tasks are not applicable to this work.reviewed all AI-assisted text and code. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

work uses city-level traffic observations and public urban and contextual data. It does not involve human participants, clinical interventions, or the collection of personally identifiable information. The analysis is conducted on aggregate mobility observations and is not intended to infer or expose individual trajectories. We follow the applicable data-use and licensing requirements and recognize that any real-world deployment should address local privacy, security, and fairness requirements.

REPRODUCIBILITY STATEMENT

main text specifies the FACO deployment protocol, source and target chronology, predictive-functional relay, target adaptation procedure, evaluation metrics, and experimental comparisons. The appendices provide the data construction procedure, city splits, forecasting windows, baseline implementations, hyperparameter settings, detailed model construction, and supporting formal results. The implementation, configuration files, preprocessing and split definitions will be provided as anonymous supplementary material and released publicly after acceptance. These materials are intended to make the reported results and comparisons independently reproducible.

REFERENCES

Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33:17804–17815, 2020.

Xiaojun Bi, Wudi Li, Yiwen Sun, Siyuan Qi, and Zheng Wang. Leveraging multiple source cities in selective transfer learning for traffic prediction with limited data. *IEEE Transactions on Intelligent Transportation Systems*, 2025.

Yujun Chen, Shihao Tu, Wenyue Ding, Yicheng Lu, Qingkai Ren, Yangjie Zheng, and Yang Yang. Unified spatio-temporal tokens are bases for generalizable traffic forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 508–519, 2026.

C. Corbane, D. Ehrlich, A. Florczyk, S. Freire, T. Kemper, L. Maffenini, M. Melchiorri, M. Pesaresi, P. Politis, F. Sabo, M. Schiavina, and L. Zanchetta. *GHSL data package 2019 – Public release GHS P2019*. Publications Office, 2019. doi: doi/10.2760/290498.

Zheng Fang, Qingqing Long, Guojie Song, and Kunqing Xie. Spatial-temporal graph ode networks for traffic flow forecasting. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 364–373, 2021.

Ziquan Fang, Dongen Wu, Lu Pan, Lu Chen, and Yunjun Gao. When transfer learning meets cross-city urban flow prediction: Spatio-temporal adaptation matters. In *IJCAI*, volume 22, pp. 2030–2036, 2022.

Yan Ge, Zhengyang Zhou, Qihe Huang, Yuxuan Liang, and Yang Wang. Enabling arbitrary inference in spatio-temporal dynamic systems: A physics-inspired perspective. In *International Conference on Learning Representations*, volume 2026, pp. 108261–108281, 2026.

Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 922–929, 2019.

Xianwei Guo, Zhiyong Yu, Jiangtao Wang, Fangwan Huang, Xing Chen, Dingqi Yang, and Ke Xu. Cross-city few-shot spatiotemporal graph forecasting via masked pre-training and prompt tuning. *Neural Networks*, pp. 109553, 2026.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pp. 2790–2799. PMLR, 2019.

Junfeng Hu, Xu Liu, Zhencheng Fan, Yifang Yin, Shili Xiang, Savitha Ramasamy, and Roger Zimmermann. Prompt-based spatio-temporal graph transfer learning, 2024. URL `https://arxiv.org/abs/2405.12452`.

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

Renhe Jiang, Zhaonan Wang, Jiawei Yong, Puneet Jeph, Quanjun Chen, Yasumasa Kobayashi, Xuan Song, Shintaro Fukushima, and Toyotaro Suzumura. Spatio-temporal meta-graph learning for traffic forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 8078–8086, 2023b.

Guangyin Jin, Yuxuan Liang, Yuchen Fang, Zezhi Shao, Jincai Huang, Junbo Zhang, and Yu Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *IEEE transactions on knowledge and data engineering*, 36(10):5388–5408, 2023a.

Yilun Jin, Kai Chen, and Qiang Yang. Selective cross-city transfer learning for traffic prediction via source city region re-weighting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 731–741, 2022.

Yilun Jin, Kai Chen, and Qiang Yang. Transferable graph structure learning for graph-based traffic forecasting across cities. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 1032–1043, 2023b.

Weiyang Kong, Ziyu Guo, and Yubao Liu. Spatio-temporal pivotal graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8627–8635, 2024.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*, 2017.

Zhonghang Li, Lianghao Xia, Yong Xu, and Chao Huang. Flashst: A simple and universal prompt-tuning framework for traffic prediction. *arXiv preprint arXiv:2405.17898*, 2024.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020a.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. *Advances in neural information processing systems*, 33:6755–6766, 2020b.

Xuhui Lin, Qiuchen Lu, Long Chen, Jack Chin Pang Cheng, Jiayi Yan, Jingke Hong, and Pengjun Zhao. Iutf dataset: Enabling cross-border resource for analysing the impact of rainfall on urban transportation. *Scientific Data*, 13(1):22, 2025.

Jiayue Liu, Zhongchao Yi, Zhengyang Zhou, Qihe Huang, Kuo Yang, Xu Wang, and Yang Wang. Synevo: A neuro-inspired spatiotemporal evolutional framework for cross-domain adaptation. *arXiv preprint arXiv:2505.16080*, 2025a.

Zhanyu Liu, Guanjie Zheng, and Yanwei Yu. Cross-city few-shot traffic forecasting via traffic pattern bank. In *Proceedings of the 32nd ACM international conference on information and knowledge management*, pp. 1451–1460, 2023.

Zhanyu Liu, Jianrong Ding, and Guanjie Zheng. Frequency enhanced pre-training for cross-city few-shot traffic forecasting. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 35–52. Springer, 2024.

Zhanyu Liu, Guanjie Zheng, and Yanwei Yu. Multi-scale traffic pattern bank for cross-city few-shot traffic forecasting. *ACM Transactions on Knowledge Discovery from Data*, 19(4):1–24, 2025b.

Zhidan Liu, Zhengze Sun, Junru Zhang, Bolin Zhang, and Panrong Tong. Regional knowledge transfer for urban traffic flow prediction via satellite imagery assisted contrastive domain adaptation. *IEEE Transactions on Intelligent Transportation Systems*, 26(11):19120–19134, 2025c.

Allister Loder, Lukas Ambühl, Monica Menendez, and Kay W Axhausen. Understanding traffic capacity of urban networks. *Scientific reports*, 9(1):16283, 2019.

Bin Lu, Xiaoying Gan, Weinan Zhang, Huaxiu Yao, Luoyi Fu, and Xinbing Wang. Spatio-temporal graph few-shot learning with cross-city knowledge transfer. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 1162–1172, 2022.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Bethany Lusch, J Nathan Kutz, and Steven L Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature communications*, 9(1):4950, 2018.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. Pmlr, 2017.

Zezhi Shao, Zhao Zhang, Fei Wang, Wei Wei, and Yongjun Xu. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pp. 4454–4458, 2022.

Yihong Tang, Ao Qu, Andy HF Chow, William HK Lam, Sze Chun Wong, and Wei Ma. Domain adversarial spatial-temporal network: A transferable framework for short-term traffic forecasting across cities. In *Proceedings of the 31st ACM international conference on information & knowledge management*, pp. 1905–1915, 2022.

Leye Wang, Xu Geng, Xiaojuan Ma, Feng Liu, and Qiang Yang. Cross-city transfer learning for deep spatio-temporal prediction. *arXiv preprint arXiv:1802.00386*, 2018.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 753–763, 2020.

Ronghui Xu, Jihao Chen, Jindong Tian, Chenjuan Guo, and Bin Yang. Most: A foundation model for multi-modality spatio-temporal traffic prediction. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1716–1727, 2026a.

Yue Xu, Wenying Duan, Xiaoxi He, and Shuai Ma. Orbit-adaptive zero-shot forecasting on spatio-temporal graph. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 5845–5855, 2026b.

Linghua Yang, Wantong Chen, Xiaoxi He, Shuyue Wei, Yi Xu, Zimu Zhou, and Yongxin Tong. Fedgtp: exploiting inter-client spatial dependency in federated graph-based traffic prediction. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 6105–6116, 2024.

Du Yin, Xiachong Lin, Yue Tan, Jinliang Deng, Estrid He, Hao Xue, and Flora D. Salim. A2tta: Anchored-and-agile test-time adaptation for evolving traffic sensor networks, 2026. URL https://arxiv.org/abs/2607.25875.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

Xian Yu, Yinxin Bao, Weiwei Jiang, and Quan Shi. Federated spatial-temporal synchronous graph convolutional network for cross-city traffic prediction. *IEEE Transactions on Consumer Electronics*, 2026.

Zhihao Zeng, Ziquan Fang, Yuting Huang, Yujie Fang, and Lu Chen. Fedtt: A federated spatio-temporal learning framework for cross-city traffic knowledge transfer. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 6409–6420, 2026.

Qianru Zhang, Xinyi Gao, Alexander Zhou, Reynold Cheng, Siu-Ming Yiu, and Hongzhi Yin. Efficient prompt learning for traffic forecasting: Q. zhang et al. *The VLDB Journal*, 35(5):45, 2026.

Yu Zhang, Hua Lu, Ning Liu, Yonghui Xu, Qingzhong Li, and Lizhen Cui. Personalized federated learning for cross-city traffic prediction. In *IJCAI*, pp. 5526–5534, 2024.

Yudong Zhang, Xu Wang, Xuan Yu, Zhaoyang Sun, Kai Wang, and Yang Wang. Drawing informative gradients from sources: A one-stage transfer learning framework for cross-city spatiotemporal forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 1147–1155, 2025.

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

Qianqian Zhou and Nan Chen. A transferable framework for cross-city forecasting via invariant structural knowledge learning. *Engineering Applications of Artificial Intelligence*, 164:113161, 2026.

Appendix

Table of Contents

A Extended Related Work 15

B Detailed Construction of the FACO Framework 17

B.1 Detailed Construction of Canonical Predictive-Functional Coordinates 17

B.2 Detailed Source-Conditioned Temporal Evolution Distillation 18

B.3 Detailed Construction of the Frozen Cross-City Relay Function 19

B.4 Detailed Trust-Conditioned Target Adaptation 21

B.5 Algorithmic Summary of FACO 24

C Experimental Details 24

C.1 Datasets and Data Processing 24

C.2 Task Protocol and City Splits 26

C.3 Baselines and Implementation Details 26

D Additional Experimental Results 27

D.1 One-Day Target Adaptation Benchmark 28

D.2 One-Day Baseline Augmentation with FACO 28

D.3 Cross-Backbone Target Robustness 31

D.4 Hyperparameter Sensitivity 30

A Related Work

We review the literature most closely related to FACO from four perspectives: spatio-temporal traffic forecasting, cross-city knowledge transfer, federated and parameter-efficient adaptation, and function-space learning. Together, these lines clarify the distinction between existing knowledge-reuse paradigms and the temporally separated, restricted-access setting considered in FACO.

Spatio-temporal traffic forecasting. Deep traffic forecasting has evolved from recurrent and convolutional temporal modeling toward graph-based and architecture-diverse spatio-temporal predictors. DCRNN models traffic propagation as diffusion over a directed graph and couples graph convolution with recurrent sequence modeling (Li et al., 2017). STGCN replaces recurrent computation with localized graph convolution and temporal convolution (Yu et al., 2017), while Graph WaveNet combines dilated temporal convolution with an adaptive adjacency mechanism that can learn latent spatial dependencies beyond a predefined road graph (Wu et al., 2019). STID demonstrates that strong traffic forecasting performance can also be obtained with comparatively simple architectures by explicitly encoding spatial identity and temporal patterns (Shao et al., 2022). Collectively, these methods show that local traffic dynamics can be modeled effectively under very different architectural choices. Additional representative architectures include attention-based spatial-temporal convolution, adaptive graph recurrent modeling, graph-based multivariate forecasting, continuous-time graph dynamics, multi-attention prediction, and meta-graph learning (Guo et al., 2019; Bai et al., 2020; Wu et al., 2020; Fang et al., 2021; Zheng et al., 2020; Jiang et al., 2023b). Recent surveys organize this broader family of spatio-temporal graph models and its traffic-flow applications (Jin et al., 2023a; Kong et al., 2024). Recent traffic foundation models also study multimodal spatio-temporal prediction and unified tokens for generalizable traffic forecasting (Xu et al., 2026a; Chen et al., 2026). At the same time, their latent states, node embeddings, graph operators, and parameterizations are defined relative to the sensing system on which they are trained. A hidden representation learned for one detector graph therefore has no immediate canonical meaning for another city with a different node set, topology, or forecasting backbone. This architecture- and graph-dependence is one reason why cross-city transfer requires a representation that can remain interpretable outside a particular local model.

Cross-city transfer learning. A substantial body of work addresses data scarcity by transferring information from mature source cities to target cities with fewer observations. RegionTrans learns an inter-city region matching function and uses matched source regions to transfer region-level representations to the target (Wang et al., 2018). This establishes an early formulation of cross-city transfer around explicit spatial correspondence. STAN focuses on the fact that inter-city relationships may vary over time and combines spatial adversarial adaptation with temporal attentive adaptation to identify transferable spatio-temporal features (Fang et al., 2022). CrossTReS further considers that source regions are not equally useful. It learns source-region weights through source-target meta-learning so that regions that better support target fine-tuning contribute more strongly (Jin et al., 2022). TransGTR shifts attention from transferring only predictor representations to transferring graph construction itself. It learns city-agnostic node features and a transferable structure generator so that spatial dependencies can be reconstructed in a target city rather than inherited from a fixed source graph (Jin et al., 2023b). These methods progressively relax assumptions about direct spatial correspondence, but their transfer mechanisms are still formed in a process where source-derived representations, graphs, or training signals are available while the target is being optimized. Recent work also studies invariant structural knowledge for transferable cross-city forecasting (Zhou & Chen, 2026).

Recent studies further investigate source selection, joint optimization, and data-efficient target adaptation. TSJT formulates cross-city transfer as one-stage joint training and selectively refines source gradients according to their effect on the target, thereby emphasizing source information that is beneficial to target optimization (Zhang et al., 2025). ST-MPPT combines masked pre-training on source cities with prompt-based target adaptation. Its source stage learns spatio-temporal representations through decoupled masked pre-training, while a prompt network dynamically adapts the pre-trained encoders to different target distributions under few-shot observations (Guo et al., 2026). SynEVO further studies neuro-inspired spatiotemporal evolution for cross-domain adaptation (Liu et al., 2025a). Orbit-adaptive zero-shot forecasting provides another transfer setting on spatio-temporal graphs (Xu et al., 2026b). These methods demonstrate that cross-city knowledge can be selected or specialized more carefully than conventional fine-tuning. Their formulation nevertheless assumes that a source-trained representation, source-derived optimization signal, or pre-trained

encoder is directly available to the target-side adaptation procedure. This differs from settings in which historical source systems have already left and future target adaptation must proceed from a bounded state materialized in advance.

Federated traffic forecasting. Federated learning addresses distributed prediction under data-locality requirements by coordinating model updates instead of centralizing raw observations (McMahan et al., 2017). In cross-city traffic forecasting, pFedCTP studies personalization under both spatial and temporal heterogeneity. It decomposes traffic representation learning to capture space-independent patterns and interpolates global and local parameters to mitigate client drift across heterogeneous cities (Zhang et al., 2024). FedGTP further incorporates inter-client spatial relationships into federated graph-based traffic prediction (Yang et al., 2024). These methods are particularly relevant when multiple cities need to collaborate without sharing raw traffic records. Their collaboration model, however, is generally iterative. Clients remain participants in a repeated server-client optimization process and can continue to provide model updates across communication rounds. A temporally separated deployment imposes a different requirement because a historical city may no longer be available when a future participant appears, ruling out renewed client updates or source-side computation.

Efficient adaptation under limited observations. Parameter-efficient adaptation studies how a pre-trained model can be specialized with a small number of target-side parameters. General adapter methods introduce lightweight task-specific modules while retaining most of the pre-trained backbone (Houlsby et al., 2019). In traffic forecasting, FlashST proposes a lightweight spatio-temporal prompt network together with distribution mapping to adapt pre-trained predictors to downstream datasets under distribution shift (Li et al., 2024). SimpleST similarly uses model-agnostic prompt learning to steer frozen spatio-temporal predictors toward shifted traffic distributions while updating only a small fraction of model parameters (Zhang et al., 2026). ST-MPPT also adopts prompt-based specialization, but couples it with source-side masked pre-training for cross-city few-shot forecasting (Guo et al., 2026). These studies show that limited target labels can be sufficient for efficient specialization when a suitable pre-trained representation or backbone is already available. The transferable state in this family is therefore usually tied to the existence of a specific pre-trained model, encoder family, or parameterization that can be retained and adapted at the target.

Function-space learning and temporal dynamics. Operator learning provides a complementary view of transfer by modeling relations between functions rather than only transferring finite-dimensional parameter vectors. DeepONet learns nonlinear operators through a branch-trunk decomposition over input functions and evaluation locations (Lu et al., 2021). The Fourier Neural Operator instead parameterizes mappings between functions through spectral transformations and can evaluate the learned relation on different discretizations (Li et al., 2020a). Physics-inspired operator-learning perspectives further consider arbitrary inference in spatio-temporal dynamic systems (Ge et al., 2026). Neural-operator formulations include general maps between function spaces and multipole graph operators for parametric equations (Kovachki et al., 2023; Li et al., 2020b), while deep linear embeddings provide another view of reusable nonlinear dynamics (Lusch et al., 2018). More broadly, state-space and dynamical modeling approaches represent future behavior through repeated evolution of a state, suggesting that reusable temporal knowledge can be expressed as a transition law rather than as a particular predictor parameterization. This perspective is attractive for heterogeneous cross-city forecasting because the transferred object need not coincide with the latent state of any one forecasting architecture. Nevertheless, standard operator-learning formulations usually assume a coherent function space and observation semantics, whereas cross-city systems may differ simultaneously in graph structure, semantic covariates, local predictive models, and data availability.

Discussion. The literature therefore provides several complementary forms of reusable knowledge. Cross-city transfer methods reuse spatial correspondences, representations, graph structures, or source optimization signals; federated methods reuse information through repeated distributed updates; efficient adaptation specializes pre-trained models with limited target data; and operator learning motivates reusable mappings in a common function space. These lines leave a distinct regime only partially explored. Historical source systems and a future target may not coexist in time, private source data and forecasting systems may be unavailable during target adaptation, heterogeneous local predictors may expose incompatible internal representations, and the target may provide only a small amount of local evidence. Under this combination of temporal separation and

information constraints, the transferable object must remain finite and usable after source exit while retaining a consistent meaning across heterogeneous forecasting systems.

B Detailed Construction of the FACO Framework

B.1 Detailed Construction of Canonical Predictive-Functional Coordinates

This section provides the complete construction of the predictive-functional coordinates used in Sec. 3.1. The construction is identical for historical source cities and the future target.

Canonical temporal state. For city $v \in \mathcal{C}_s \cup \{T\}$, node n, and forecast origin t, let H denote the forecasting horizon and let $h \in \{1, \dots, H\}$ be its horizon index. Let $y^v_{t,n}$ denote the traffic observation. The causal traffic history of length L is

$$\mathbf{x}^v_{t,n} = \left[y^v_{t-L+1,n}, \dots, y^v_{t,n}\right]. \tag{31}$$

We use the latest observation as a local anchor and the mean historical magnitude as a robust scale,

$$m^v_{t,n} = y^v_{t,n}, \qquad \sigma^v_{t,n} = \max\left(\frac{1}{L}\sum_{\ell=1}^{L} x^v_{t,n,\ell}, 1\right). \tag{32}$$

The normalized history is projected onto a deterministic orthonormal temporal basis $Q \in \mathbb{R}^{L\times L}$,

$$\boldsymbol{\phi}^v_{t,n,0} = \left(\frac{\mathbf{x}^v_{t,n} - m^v_{t,n}\mathbf{1}}{\sigma^v_{t,n}}\right) Q^\top, \tag{33}$$

where $\mathbf{1} \in \mathbb{R}^L$ is the all-ones vector. The basis is fixed before cross-city learning and shared by all cities. Its leading directions capture coarse temporal structure such as level, recent-history contrast, and latest-value contrast, while the remaining directions are completed from deterministic cosine components and orthonormalized.

Shared semantic schema and city-specific alignment. The urban context S_v is organized under a shared semantic schema,

$$\mathbf{z}^{v,\mathrm{raw}}_{t,n,h} = \left[\mathbf{z}^{v,\mathrm{cal}}_{t,h}, \mathbf{z}^{v,\mathrm{static}}_{n}, \overline{\mathbf{z}}^{v,\mathrm{exo}}_{t,n}\right], \tag{34}$$

where $\mathbf{z}^{v,\mathrm{cal}}_{t,h}$ contains calendar and temporal descriptors, $\mathbf{z}^{v,\mathrm{static}}_{n}$ contains static node or urban attributes, and $\overline{\mathbf{z}}^{v,\mathrm{exo}}_{t,n}$ summarizes exogenous information causally available at the forecast origin. The semantic meaning and ordering of these variables are shared across cities, although their empirical scales and local responses may differ.

FACO applies a lightweight semantic alignment map,

$$\mathbf{z}^{v,\mathrm{sem}}_{t,n,h} = \mathcal{A}_{\eta_v}\left(\mathbf{z}^{v,\mathrm{raw}}_{t,n,h}\right), \tag{35}$$

with the element-wise affine parameterization

$$\mathcal{A}_{\eta_v}(\mathbf{r}) = \mathbf{g}_v \odot \mathbf{r} + \mathbf{b}_v, \qquad \eta_v = \{\mathbf{g}_v, \mathbf{b}_v\}, \tag{36}$$

where $\mathbf{r}$ is a generic semantic input and $\odot$ denotes element-wise multiplication. The alignment preserves the common semantic coordinates while allowing each city to adapt feature scale and response with a small number of parameters.

Predictive-functional interface. The predictive interface maps the behavior of the local forecasting system into the shared predictive-functional space. Given the local forecaster B_v, city-local graph G_v, recent history, and aligned semantic context, Ψ_{ω_v} produces an H-step predictive trajectory,

$$\boldsymbol{\pi}^v_{t,n} = \Psi_{\omega_v}\left(B_v, G_v, \mathbf{x}^v_{t,n}, \mathbf{z}^{v,\mathrm{sem}}_{t,n,1:H}\right) = \left[\pi^v_{t,n,1}, \dots, \pi^v_{t,n,H}\right]. \tag{37}$$

Together with the semantic alignment parameters, the predictive interface forms the city-specific predictive-functional interface state

$$\psi_v = (\eta_v, \omega_v). \tag{38}$$

Historical source interfaces are estimated using source-side information. For the future target, ψ_T is instantiated from the frozen shared initialization using the limited target observations, as described in Sec. 3.4.

To express local predictive behavior in the same anchor-scale coordinates as the temporal state, each horizon prediction is normalized as

$$\widetilde{\pi}^{v}_{t,n,h} = \frac{\pi^{v}_{t,n,h} - m^{v}_{t,n}}{\sigma^{v}_{t,n}}, \qquad h = 1, \dots, H. \tag{39}$$

The aligned semantic context and normalized predictive behavior are concatenated into the horizon-specific condition

$$\mathbf{u}^{v}_{t,n,h} = \left[\mathbf{z}^{v,\mathrm{sem}}_{t,n,h}, \widetilde{\pi}^{v}_{t,n,h}\right]. \tag{40}$$

We denote its predictive component by

$$\mathrm{pred}\left(\mathbf{u}^{v}_{t,n,h}\right) = \widetilde{\pi}^{v}_{t,n,h}. \tag{41}$$

Final predictive-functional coordinates. A forecasting window is represented by

$$\left(\boldsymbol{\phi}^{v}_{t,n,0}, \mathbf{u}^{v}_{t,n,1:H}\right). \tag{42}$$

The temporal component is expressed in a fixed basis, the semantic component retains a common schema with lightweight city-specific alignment, and the predictive component records the behavior of the local forecasting system in normalized output coordinates. The same representation can therefore be constructed for heterogeneous source systems and a future target city.

B.2 Detailed Source-Conditioned Temporal Evolution Distillation

This section expands the source-side temporal evolution introduced in Sec. 3.2. Let $\mathcal{W}_s$ denote the chronological source forecasting windows used for fitting. All source learning is performed using source-local windows expressed in the predictive-functional coordinates defined in Sec. 3.1.

Conditional evolution parameterization. For source city s, the conditional evolution function is

$$g_{\theta_s}(\boldsymbol{\phi}, \mathbf{u}) = c_s + \boldsymbol{\phi}\mathbf{p}_s + \mathbf{u}^{\top}\mathbf{b}_s + \boldsymbol{\phi}R_s\mathbf{u}. \tag{43}$$

Here, $\boldsymbol{\phi} \in \mathbb{R}^{1\times L}$ is the canonical temporal state and $\mathbf{u} \in \mathbb{R}^{D_u}$ is the horizon-specific semantic-predictive condition, where $D_u = \dim(\mathbf{u})$. The coefficients satisfy $c_s \in \mathbb{R}$, $\mathbf{p}_s \in \mathbb{R}^{L}$, $\mathbf{b}_s \in \mathbb{R}^{D_u}$, and $R_s \in \mathbb{R}^{L\times D_u}$, so every term in g_{θ_s} is scalar. The parameters c_s, $\mathbf{p}_s$, and $\mathbf{b}_s$ capture the transition bias and first-order contributions of the two inputs. Their interaction is represented by

$$R_s = U_s V_s^{\top}, \qquad \mathrm{rank}(R_s) \ll \min(L, D_u), \tag{44}$$

where U_s and V_s are low-rank interaction factors. This keeps the state–condition interaction compact. The source evolution parameters are

$$\theta_s = \{c_s, \mathbf{p}_s, \mathbf{b}_s, U_s, V_s\}. \tag{45}$$

Predictive-relative transition and fitting range. The predictive component of the horizon condition is

$$\mathrm{pred}\left(\mathbf{u}^{s}_{t,n,h}\right) = \widetilde{\pi}^{s}_{t,n,h}. \tag{46}$$

The corresponding source ground-truth value at horizon h is represented in the same anchor-scale coordinates,

$$\widetilde{y}^{s,*}_{t,n,h} = \frac{y^{s}_{t+h,n} - m^{s}_{t,n}}{\sigma^{s}_{t,n}}, \qquad h = 1, \dots, H. \tag{47}$$

Before source fitting, the horizon-wise fitting range is estimated from $\mathcal{W}_s$ as

$$\rho^{\mathrm{fit}}_{s,h} = \mathrm{clip}\left(Q_{0.95}\left(\left\{\left|\widetilde{y}^{s,*}_{t,n,h} - \widetilde{\pi}^{s}_{t,n,h}\right|\right\}_{(t,n)\in\mathcal{W}_s}\right), \rho_{\min}, \rho_{\max}\right), \tag{48}$$

where $Q_{0.95}$ denotes the empirical 95th percentile and $\rho_{\min}, \rho_{\max} > 0$ are numerical bounds. The source transition is then

$$\mathcal{E}_{\theta_s,h}(\boldsymbol{\phi}, \mathbf{u}) = \mathrm{pred}(\mathbf{u}) + \mathrm{clip}\left(g_{\theta_s}(\boldsymbol{\phi}, \mathbf{u}) - \mathrm{pred}(\mathbf{u}), -\rho^{\mathrm{fit}}_{s,h}, \rho^{\mathrm{fit}}_{s,h}\right). \tag{49}$$

Thus the source law is learned in a predictive-relative coordinate whose admissible fitting range is determined from source-local observations.

Canonical closed-loop update. Let $\mathbf{r}^s_{t,n,h} \in \mathbb{R}^L$ denote the normalized temporal history associated with horizon h. In the original temporal coordinates, one forecasting step removes the oldest entry and appends the newly evolved value. Denoting this deterministic shift by $S \in \mathbb{R}^{L\times L}$ and the newest-position injection vector by $\mathbf{e} \in \mathbb{R}^L$, the update is

$$\mathbf{r}^s_{t,n,h} = \mathbf{r}^s_{t,n,h-1} S + \widetilde{y}^s_{t,n,h} \mathbf{e}^\top. \tag{50}$$

Because the canonical state satisfies $\boldsymbol{\phi} = \mathbf{r}Q^\top$, the corresponding operators in the shared temporal basis are

$$S_\phi = QSQ^\top, \qquad \mathbf{e}_\phi = Q\mathbf{e}. \tag{51}$$

The recursive canonical update is therefore

$$\boldsymbol{\phi}^s_{t,n,h} = \boldsymbol{\phi}^s_{t,n,h-1} S_\phi + \widetilde{y}^s_{t,n,h} \mathbf{e}^\top_\phi, \qquad h = 1, \dots, H. \tag{52}$$

Combining Eqs. equation 49 and equation 52 yields the complete closed-loop source evolution.

Trajectory-level source learning. For a source fitting window $(t, n) \in \mathcal{W}_s$, define the predicted normalized trajectory by

$$\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{evo}}_{t,n,1:H} = \mathrm{Rollout}\left(\mathcal{E}_{\theta_s,1:H}; \boldsymbol{\phi}^s_{t,n,0}, \mathbf{u}^s_{t,n,1:H}\right). \tag{53}$$

The source evolution parameters are obtained through

$$\theta^\star_s = \arg\min_{\theta_s} \sum_{(t,n)\in\mathcal{W}_s} \mathcal{L}_s\left(\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{evo}}_{t,n,1:H}, \widetilde{\mathbf{y}}^{s,*}_{t,n,1:H}\right). \tag{54}$$

Here, $\mathcal{L}_s$ is the source trajectory loss. The recursive rollout supervises the complete evolved trajectory with horizon-to-horizon state coupling.

Held-out supported evolution range. After fitting $\theta^\star_s$, let $\mathcal{W}^{\mathrm{val}}_s$ denote held-out source windows used to characterize the learned law. For each horizon, define the realized predictive-relative innovation

$$\Delta^s_{t,n,h} = \left|\widehat{\widetilde{y}}^{\,s,\mathrm{evo}}_{t,n,h} - \widetilde{\pi}^s_{t,n,h}\right|, \qquad (t,n) \in \mathcal{W}^{\mathrm{val}}_s. \tag{55}$$

The released horizon-wise support radius is

$$\rho_{s,h} = \mathrm{clip}\left(Q_{0.95}\left(\left\{\Delta^s_{t,n,h}\right\}_{(t,n)\in\mathcal{W}^{\mathrm{val}}_s}\right), \rho_{\min}, \rho_{\max}\right), \qquad h = 1, \dots, H. \tag{56}$$

The fitting radius $\rho^{\mathrm{fit}}_{s,h}$ controls source-side optimization, while $\rho_{s,h}$ summarizes the predictive-relative range exhibited by the learned evolution law on held-out source windows and is retained for cross-city relay construction.

Bounded source release. After source-local learning, the quantities needed for later cross-city consolidation are summarized as

$$a_s = \left\{\theta^\star_s, \boldsymbol{\rho}_s, \xi_s, q_s\right\}, \qquad \boldsymbol{\rho}_s = \left[\rho_{s,1}, \dots, \rho_{s,H}\right]. \tag{57}$$

The state $\theta^\star_s$ specifies the source conditional temporal law and $\boldsymbol{\rho}_s$ records the held-out horizon-wise supported evolution range. The support summary ξ_s and reliability quantity q_s provide source-side evidence used during relay construction. Their construction and cross-city aggregation are detailed in Appendix B.3 and Sec. 3.3.

B.3 Detailed Construction of the Frozen Cross-City Relay Function

This section expands the source-side relay construction introduced in Sec. 3.3. The frozen state contains a reliability-weighted cross-city evolution memory and a shared predictive-functional interface initialization learned from historical source cities before target arrival.

Source-side reliability. Each source evolution law is evaluated on chronological held-out source windows. Let $\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{evo}}_i$ denote the closed-loop trajectory generated by the learned source evolution for held-out window i, and let $\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{pred}}_i$ denote the normalized trajectory exposed by the corresponding source predictive interface. We measure the relative predictive gain as

$$\delta_{s,i} = \frac{\mathcal{L}_s\left(\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{pred}}_i, \widetilde{\mathbf{y}}^{s,*}_i\right) - \mathcal{L}_s\left(\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{evo}}_i, \widetilde{\mathbf{y}}^{s,*}_i\right)}{\mathcal{L}_s\left(\widehat{\widetilde{\mathbf{y}}}^{\,s,\mathrm{pred}}_i, \widetilde{\mathbf{y}}^{s,*}_i\right) + \varepsilon_{\mathrm{rel}}}, \tag{58}$$

where $\varepsilon_{\mathrm{rel}} > 0$ prevents numerical instability when the reference loss is small. The reliability statistic q_s is a conservative lower-confidence estimate of the held-out relative-gain distribution. In implementation, temporally contiguous blocks are resampled to preserve local temporal dependence. The relay weight is

$$w_s = \begin{cases} \dfrac{[q_s]_+}{\sum_{j\in\mathcal{C}_s}[q_j]_+}, & \sum_{j\in\mathcal{C}_s}[q_j]_+ > 0, \\ 0, & \text{otherwise}, \end{cases} \qquad [x]_+ = \max(x, 0). \tag{59}$$

When all source reliabilities are nonpositive, every w_s is zero and the cross-city evolution memory is disabled.

Functional support summary. For each source forecasting window, FACO constructs the functional descriptor

$$\mathbf{d}^s_{t,n} = \left[\boldsymbol{\phi}^s_{t,n,0}, \frac{1}{H}\sum_{h=1}^{H}\mathbf{u}^s_{t,n,h}\right]. \tag{60}$$

Let

$$\mathcal{Z}^{\mathrm{sup}}_s = \left\{\mathbf{d}^s_{t,n} : (t,n) \in \mathcal{W}_s\right\} \tag{61}$$

denote the collection of source descriptors. Its coordinate-wise center and scale are

$$\boldsymbol{\mu}^{\mathrm{sup}}_s = \mathrm{Mean}\left(\mathcal{Z}^{\mathrm{sup}}_s\right), \qquad \boldsymbol{\sigma}^{\mathrm{sup}}_s = \max\left(\mathrm{Std}\left(\mathcal{Z}^{\mathrm{sup}}_s\right), \varepsilon_{\mathrm{sup}}\right), \tag{62}$$

where the scalar floor is applied element-wise and $\varepsilon_{\mathrm{sup}} > 0$ supplies a minimum scale for stable standardization. The source support summary is

$$\xi_s = \{\boldsymbol{\mu}^{\mathrm{sup}}_s, \boldsymbol{\sigma}^{\mathrm{sup}}_s\}. \tag{63}$$

The horizon-wise supported evolution range from Appendix B.2 is retained as

$$\boldsymbol{\rho}_s = [\rho_{s,1}, \ldots, \rho_{s,H}]. \tag{64}$$

Thus q_s, ξ_s, and $\boldsymbol{\rho}_s$ summarize source reliability, functional coverage, and predictive-relative evolution magnitude.

Reliability-weighted evolution consolidation. Because every g_{θ_s} acts on the same canonical state and condition coordinates, corresponding source coefficients have a common functional meaning. The scalar and linear terms are consolidated as

$$\bar{c} = \sum_{s\in\mathcal{C}_s} w_s c_s, \qquad \bar{\mathbf{p}} = \sum_{s\in\mathcal{C}_s} w_s \mathbf{p}_s, \qquad \bar{\mathbf{b}} = \sum_{s\in\mathcal{C}_s} w_s \mathbf{b}_s. \tag{65}$$

The full reliability-weighted interaction matrix is

$$\bar{R}_{\mathrm{full}} = \sum_{s\in\mathcal{C}_s} w_s R_s. \tag{66}$$

Let $\bar{R}_{\mathrm{full}} = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^\top$ be its singular value decomposition. Retaining the leading r_{relay} components gives

$$\bar{R} = U\Sigma V^\top \approx \bar{R}_{\mathrm{full}}, \tag{67}$$

where r_{relay} is the retained relay rank, $U = \widetilde{U}_{:,1:r_{\mathrm{relay}}}$, $\Sigma = \widetilde{\Sigma}_{1:r_{\mathrm{relay}},1:r_{\mathrm{relay}}}$, and $V = \widetilde{V}_{:,1:r_{\mathrm{relay}}}$. The supported predictive-relative range is consolidated horizon by horizon,

$$\bar{\rho}_h = \sum_{s\in\mathcal{C}_s} w_s \rho_{s,h}, \qquad h = 1, \ldots, H, \qquad \bar{\boldsymbol{\rho}} = [\bar{\rho}_1, \ldots, \bar{\rho}_H]. \tag{68}$$

The resulting frozen evolution memory is

$$\mathcal{M} = \left\{\bar{c}, \bar{\mathbf{p}}, \bar{\mathbf{b}}, \bar{R}, \bar{\boldsymbol{\rho}}, \{(w_s, \xi_s)\}_{s\in\mathcal{C}_s}\right\}. \tag{69}$$

In implementation, $\bar{R}$ is stored through its compact factors (U, Σ, V).

Shared predictive-functional interface initialization. The shared interface initialization is learned jointly across historical source cities and is distinct from the reliability-weighted aggregation used for $\mathcal{M}$. Let

$$\mathcal{J}_s^{\mathrm{int}}(\psi) = \mathbb{E}_{(t,n)\sim\mathcal{W}_s}\left[\ell_{\mathrm{int}}\left(\boldsymbol{\pi}_{t,n}^{s,\psi}, \mathbf{y}_{t,n,1:H}^{s,*}\right)\right] \tag{70}$$

denote the interface-learning contribution of source s, where ℓ_{int} is the source-side interface forecasting loss, $\boldsymbol{\pi}_{t,n}^{s,\psi}$ is the trajectory obtained by the semantic alignment and predictive interface of Sec. 3.1 under $\psi = (\eta, \omega)$, and $\mathbf{y}_{t,n,1:H}^{s,*}$ is the corresponding source ground-truth trajectory in the interface output space. The source-trained initialization is

$$\psi_0 = \arg\min_{\psi} \sum_{s\in\mathcal{C}_s} \mathcal{J}_s^{\mathrm{int}}(\psi). \tag{71}$$

This shared state instantiates the shared source-side artifact a_0 in the FACO formulation and provides the initialization later specialized by the future target. The reliability weights w_s are specific to the consolidation of conditional temporal evolution in $\mathcal{M}$ and are not used in this interface objective.

Frozen relay state. The complete frozen cross-city state is

$$\Omega = \{\psi_0, \mathcal{M}\}. \tag{72}$$

Once Ω has been formed, both ψ_0 and $\mathcal{M}$ remain fixed until target adaptation. The target receives a finite source-derived state containing a shared predictive-functional initialization, reusable conditional temporal dynamics, and the source-support evidence required for trust-conditioned deployment.

B.4 Detailed Trust-Conditioned Target Adaptation

This section gives the detailed target-side construction corresponding to Sec. 3.4. The frozen relay state $\Omega = \{\psi_0, \mathcal{M}\}$ remains fixed throughout target adaptation and inference.

Target predictive-functional instantiation. Let $\mathcal{W}_T^K$ denote the target-local forecasting windows constructed from the limited observations $\mathcal{D}_T^K$, and let $\mathcal{W}_T^{\mathrm{adp}} \subseteq \mathcal{W}_T^K$ denote the windows used for interface adaptation. The target interface is

$$\psi_T = \mathrm{Adapt}\left(\psi_0; \mathcal{D}_T^K\right), \qquad \psi_T = (\eta_T, \omega_T), \tag{73}$$

where Adapt performs few-shot optimization of the target-local interface objective

$$\mathcal{J}_T^{\mathrm{int}}\left(\psi; \mathcal{W}_T^{\mathrm{adp}}\right) = \sum_{(t,n)\in\mathcal{W}_T^{\mathrm{adp}}} \mathcal{L}_T^{\mathrm{int}}\left(\boldsymbol{\pi}_{t,n}^{T,\psi}, \mathbf{y}_{t,n,1:H}^{T,*}\right). \tag{74}$$

Here, $\boldsymbol{\pi}_{t,n}^{T,\psi}$ is the target predictive trajectory produced by the interface state ψ, $\mathbf{y}_{t,n,1:H}^{T,*}$ is the corresponding target ground-truth trajectory in the interface output space, and $\mathcal{L}_T^{\mathrm{int}}$ is the target interface loss. The optimization is initialized at ψ_0, and its resulting state is denoted by ψ_T. Adaptation updates only the lightweight target-side interface parameters; the frozen cross-city evolution state $\mathcal{M}$ is not updated.

Using ψ_T, the target forms the predictive trajectory in the same functional coordinates as the historical sources,

$$\boldsymbol{\pi}_{t,n}^{T} = \Psi_{\omega_T}\left(B_T, G_T, \mathbf{x}_{t,n}^{T}, \mathbf{z}_{t,n,1:H}^{T,\mathrm{sem}}\right), \tag{75}$$

with

$$\widetilde{\pi}_{t,n,h}^{T} = \frac{\pi_{t,n,h}^{T} - m_{t,n}^{T}}{\sigma_{t,n}^{T}}, \qquad \mathbf{u}_{t,n,h}^{T} = \left[\mathbf{z}_{t,n,h}^{T,\mathrm{sem}}, \widetilde{\pi}_{t,n,h}^{T}\right]. \tag{76}$$

Together with $\boldsymbol{\phi}_{t,n,0}^{T}$, these quantities place the target in the predictive-functional coordinates defined in Sec. 3.1.

Algorithm 1: FACO framework

Input: historical source windows $\{\mathcal{W}_s, \mathcal{W}_s^{\mathrm{val}}\}_{s\in\mathcal{C}_s}$; future target observations $\mathcal{D}_T^K$; local systems $\{B_v, G_v, S_v\}$; shared temporal basis Q; candidate set $\mathcal{A}_\alpha$
Output: frozen relay state $\Omega = \{\psi_0, \mathcal{M}\}$; target interface ψ_T; target forecast $\widehat{\mathbf{y}}_{t,n,1:H}^T$

`// Phase 1: Canonical predictive-functional coordinates`
foreach $s \in \mathcal{C}_s$ **do**
  Construct $\boldsymbol{\phi}_{t,n,0}^s$ from the normalized source history;
  Construct $\mathbf{z}_{t,n,h}^{s,\mathrm{sem}}$ and $\boldsymbol{\pi}_{t,n}^s$ through the source predictive-functional interface;
  Form $\mathbf{u}_{t,n,h}^s = [\mathbf{z}_{t,n,h}^{s,\mathrm{sem}}, \widetilde{\pi}_{t,n,h}^s]$;
`// Phase 2: Source-Conditioned Temporal Evolution Distillation`
foreach $s \in \mathcal{C}_s$ **do**
  Estimate the fitting range $\boldsymbol{\rho}_s^{\mathrm{fit}} = [\rho_{s,1}^{\mathrm{fit}}, \ldots, \rho_{s,H}^{\mathrm{fit}}]$;
  Fit $\theta_s^\star$ by closed-loop trajectory supervision using $\mathcal{E}_{\theta_s,h}$ and the recursive state update;
  Evaluate the fitted source law on $\mathcal{W}_s^{\mathrm{val}}$ to obtain $\boldsymbol{\rho}_s = [\rho_{s,1}, \ldots, \rho_{s,H}]$;
  Form $\mathbf{d}_{t,n}^s = [\boldsymbol{\phi}_{t,n,0}^s, H^{-1}\sum_{h=1}^H \mathbf{u}_{t,n,h}^s]$ and $\xi_s = \{\boldsymbol{\mu}_s^{\mathrm{sup}}, \boldsymbol{\sigma}_s^{\mathrm{sup}}\}$;
  Compute source reliability q_s;
  Release $a_s = \{\theta_s^\star, \boldsymbol{\rho}_s, \xi_s, q_s\}$;
`// Phase 3: Frozen Cross-City Relay Function`
Compute $w_s = [q_s]_+ / \sum_{j\in\mathcal{C}_s}[q_j]_+$ when the denominator is positive; otherwise set all $w_s = 0$;
Aggregate $\bar{c}, \bar{\mathbf{p}}, \bar{\mathbf{b}}, \bar{R}_{\mathrm{full}}, \bar{\boldsymbol{\rho}}$ with $\{w_s\}$;
Retain $\bar{R} = U\Sigma V^\top \approx \bar{R}_{\mathrm{full}}$;
Form $\mathcal{M} = \{\bar{c}, \bar{\mathbf{p}}, \bar{\mathbf{b}}, \bar{R}, \bar{\boldsymbol{\rho}}, \{(w_s, \xi_s)\}_{s\in\mathcal{C}_s}\}$;
Learn the shared interface initialization $\psi_0 = \arg\min_\psi \sum_{s\in\mathcal{C}_s} \mathcal{J}_s^{\mathrm{int}}(\psi)$;
Freeze $\Omega = \{\psi_0, \mathcal{M}\}$;
`// Phase 4: Trust-Conditioned Target Adaptation`
Construct target-local windows $\mathcal{W}_T^K$ from $\mathcal{D}_T^K$;
Adapt $\psi_T = \mathrm{Adapt}(\psi_0; \mathcal{D}_T^K)$ using $\mathcal{W}_T^{\mathrm{adp}} \subseteq \mathcal{W}_T^K$;
foreach *target window* (t, n) **do**
  Construct $(\boldsymbol{\phi}_{t,n,0}^T, \mathbf{u}_{t,n,1:H}^T)$ using ψ_T;
  if *cross-city evolution is active* **then**
    Form $\mathbf{d}_{t,n}^T = [\boldsymbol{\phi}_{t,n,0}^T, H^{-1}\sum_{h=1}^H \mathbf{u}_{t,n,h}^T]$ and compute d_{sup}^T;
    Roll out the predictive reference state $\boldsymbol{\phi}_{t,n,h}^{T,\mathrm{pred}}$, evaluate $r_{t,n,h}^T$, and compute d_{agr}^T;
foreach $\alpha \in \mathcal{A}_\alpha$ **do**
  foreach $(t, n) \in \mathcal{W}_T^{\mathrm{cal}}$ **do**
    Compute $\gamma_{t,n,h}^T(\alpha) = \alpha\, g^{\mathrm{sup}}(d_{\mathrm{sup}}^T) g^{\mathrm{agr}}(d_{\mathrm{agr}}^T) g^{\mathrm{hor}}(h)$ for $h = 1, \ldots, H$;
    If the relay is inactive, set $\gamma_{t,n,h}^T(\alpha) = 0$;
    Generate $\widetilde{\mathbf{y}}_{t,n,1:H}^T(\alpha)$ by the trust-conditioned closed-loop operator $\mathcal{F}_{\mathcal{M},t,n,h}^T$;
  Evaluate $\mathcal{J}_T(\alpha; \mathcal{D}_T^K)$ on $\mathcal{W}_T^{\mathrm{cal}}$;
Select $\alpha_T^\star = \arg\min_{\alpha\in\mathcal{A}_\alpha} \mathcal{J}_T(\alpha; \mathcal{D}_T^K)$;
foreach *target inference window* (t, n) **do**
  Generate $\widetilde{\mathbf{y}}_{t,n,1:H}^T(\alpha_T^\star)$ by the same trust-conditioned closed-loop evolution;
  Reconstruct $\widehat{y}_{t,n,h}^T = m_{t,n}^T + \sigma_{t,n}^T \widetilde{y}_{t,n,h}^T(\alpha_T^\star)$, $h = 1, \ldots, H$;
return $\Omega, \psi_T, \widehat{\mathbf{y}}_{t,n,1:H}^T$;

Historical functional support. When cross-city evolution is active, FACO constructs

$$\mathbf{d}_{t,n}^T = \left[\boldsymbol{\phi}_{t,n,0}^T, \frac{1}{H}\sum_{h=1}^H \mathbf{u}_{t,n,h}^T\right]. \tag{77}$$

Let $D_{\mathrm{f}} = \dim(\mathbf{d}_{t,n}^T)$. Its standardized distance to source s is

$$d_s\left(\mathbf{d}_{t,n}^T\right) = \left[\frac{1}{D_{\mathrm{f}}}\left\|\frac{\mathbf{d}_{t,n}^T - \boldsymbol{\mu}_s^{\mathrm{sup}}}{\boldsymbol{\sigma}_s^{\mathrm{sup}}}\right\|_2^2\right]^{1/2}, \tag{78}$$

where $\boldsymbol{\sigma}_s^{\mathrm{sup}}$ is the stabilized support scale defined in Appendix B.3. The target support distance is

$$d_{\mathrm{sup}}^T = \min_{s:w_s>0} d_s\left(\mathbf{d}_{t,n}^T\right). \tag{79}$$

A smaller d_{sup}^T indicates stronger historical support for the current target condition.

Trajectory-level relay agreement. The frozen evolution memory defines

$$g_{\mathcal{M}}(\boldsymbol{\phi}, \mathbf{u}) = \bar{c} + \boldsymbol{\phi}\bar{\mathbf{p}} + \mathbf{u}^\top\bar{\mathbf{b}} + \boldsymbol{\phi}\bar{R}\mathbf{u}. \tag{80}$$

To evaluate agreement with the target predictive trajectory, define the predictive reference state by

$$\boldsymbol{\phi}_{t,n,0}^{T,\mathrm{pred}} = \boldsymbol{\phi}_{t,n,0}^T, \tag{81}$$

$$\boldsymbol{\phi}_{t,n,h}^{T,\mathrm{pred}} = \boldsymbol{\phi}_{t,n,h-1}^{T,\mathrm{pred}} S_\phi + \mathrm{pred}\left(\mathbf{u}_{t,n,h}^T\right)\mathbf{e}_\phi^\top. \tag{82}$$

The corresponding relay response is

$$r_{t,n,h}^T = g_{\mathcal{M}}\left(\boldsymbol{\phi}_{t,n,h-1}^{T,\mathrm{pred}}, \mathbf{u}_{t,n,h}^T\right), \tag{83}$$

and the trajectory-level disagreement is

$$d_{\mathrm{agr}}^T = \frac{1}{H}\sum_{h=1}^{H}\left|r_{t,n,h}^T - \mathrm{pred}\left(\mathbf{u}_{t,n,h}^T\right)\right|. \tag{84}$$

Functional trust. For an active relay, FACO combines target-level relay participation with historical support, trajectory-level agreement, and horizon confidence,

$$\gamma_{t,n,h}^T(\alpha) = \alpha\, g^{\mathrm{sup}}\left(d_{\mathrm{sup}}^T\right) g^{\mathrm{agr}}\left(d_{\mathrm{agr}}^T\right) g^{\mathrm{hor}}(h), \qquad \alpha \in [0,1]. \tag{85}$$

Here, g^{sup}, g^{agr}, and g^{hor} take values in $[0,1]$; g^{sup} decreases with weaker historical support, g^{agr} decreases with stronger relay–target disagreement, and g^{hor} encodes horizon-dependent confidence. If cross-city evolution is disabled, FACO sets $\gamma_{t,n,h}^T(\alpha) = 0$. Their concrete parameter settings are specified with the experimental protocol.

Trust-conditioned closed-loop evolution. Given $\gamma_{t,n,h}^T(\alpha)$, the target evolution operator is

$$\mathcal{F}_{\mathcal{M},t,n,h}^T(\boldsymbol{\phi}, \mathbf{u}; \alpha) = \mathrm{pred}(\mathbf{u}) + \gamma_{t,n,h}^T(\alpha)\,\mathrm{clip}\left(g_{\mathcal{M}}(\boldsymbol{\phi}, \mathbf{u}) - \mathrm{pred}(\mathbf{u}), -\bar{\rho}_h, \bar{\rho}_h\right). \tag{86}$$

For a candidate α, the target rollout is generated recursively,

$$\widetilde{y}_{t,n,h}^T(\alpha) = \mathcal{F}_{\mathcal{M},t,n,h}^T\left(\boldsymbol{\phi}_{t,n,h-1}^T(\alpha), \mathbf{u}_{t,n,h}^T; \alpha\right), \tag{87}$$

$$\boldsymbol{\phi}_{t,n,h}^T(\alpha) = \boldsymbol{\phi}_{t,n,h-1}^T(\alpha) S_\phi + \widetilde{y}_{t,n,h}^T(\alpha)\mathbf{e}_\phi^\top, \qquad h = 1, \ldots, H. \tag{88}$$

The recursion starts from $\boldsymbol{\phi}_{t,n,0}^T(\alpha) = \boldsymbol{\phi}_{t,n,0}^T$. Because the evolved value enters the next canonical state, relay participation propagates through the complete forecasting trajectory.

Target-side relay calibration. Let $\mathcal{W}_T^{\mathrm{cal}} \subseteq \mathcal{W}_T^K$ denote the target-local windows used to calibrate relay participation, and let $\mathcal{A}_\alpha \subset [0,1]$ be a fixed finite candidate set. FACO selects

$$\alpha_T^\star = \arg\min_{\alpha\in\mathcal{A}_\alpha} \sum_{(t,n)\in\mathcal{W}_T^{\mathrm{cal}}} \mathcal{L}_T\left(\mathrm{Rollout}\left(\mathcal{F}_{\mathcal{M},t,n,1:H}^T; \boldsymbol{\phi}_{t,n,0}^T, \mathbf{u}_{t,n,1:H}^T, \alpha\right), \widetilde{\mathbf{y}}_{t,n,1:H}^{T,*}\right), \tag{89}$$

where $\widetilde{\mathbf{y}}_{t,n,1:H}^{T,*}$ is the normalized target ground-truth trajectory and $\mathcal{L}_T$ is the target calibration loss. The summed trajectory loss is the objective $\mathcal{J}_T(\alpha; \mathcal{D}_T^K)$ used in Sec. 3.4. The complete recursive trajectory is evaluated for each candidate α, while $\mathcal{M}$ remains fixed.

Physical-scale reconstruction. After selecting $\alpha_T^\star$, the normalized trajectory is mapped back to the target traffic scale,

$$\widehat{y}_{t,n,h}^{T} = m_{t,n}^{T} + \sigma_{t,n}^{T} \widetilde{y}_{t,n,h}^{T}(\alpha_T^\star), \qquad h = 1, \dots, H. \tag{90}$$

The resulting forecast uses only the frozen relay state and information available locally to the future target.

B.5 Algorithmic Summary of FACO

Algorithm 1 summarizes FACO in the same order as Secs. 3.1–3.4. All symbols follow the definitions in the main text and the preceding appendix sections.

C Experimental Details

This section describes the data construction, city roles, evaluation protocol, comparison methods, implementation settings, and evaluation metrics used in the experiments.

C.1 Datasets and Data Processing

Our experiments use the IUTF traffic–weather resource introduced by Lin et al. (Lin et al., 2025). Its traffic records originate from the UTD19 collection (Loder et al., 2019)[1], and the released city folders contain detector flow, hourly weather, spatial metadata, and aligned calendar records. We retain 18 cities according to temporal coverage and data availability. Their observation spans are reported in Table 2, while their source and target roles are defined in Section C.2.

Traffic observations. The released traffic observations have a five-minute source cadence. We aggregate flow by detector and UTC 15-minute bucket to obtain the 15-minute traffic representation. The observation mask is stored with the flow array and is propagated to window construction, scaling, optimization, and evaluation. The city manifests preserve gaps in the temporal index, and the dense modeling timeline is the first-to-last UTC 15-minute grid for each city. Nodes are retained after the role-specific training-window observation filter, which keeps a detector when its observed-bin fraction reaches 0.8.

Weather and temporal alignment. Each city uses the hourly weather records released with the traffic–weather resource. Detector coordinates are assigned to the nearest weather grid by the haversine distance, and the hourly record is aligned to each 15-minute traffic timestamp with a backward as-of join. We use four continuous weather variables—total precipitation, large-scale rain rate, a precipitation indicator, and weather age in minutes—together with the categorical precipitation type. The two non-negative precipitation variables are transformed with $\log(1 + x)$ before zero-aware robust scaling. Forecasting uses the weather information available at each forecast origin.

POI semantics from Geofabrik. Geofabrik historical country or region extracts provide the OSM PBF input for POI construction[2]. The released metadata record the source URL, snapshot date, parser, PBF digest, and the city-specific sensor-centered radius. The PBF files were parsed with `pyrosm`; each detector receives counts of OSM elements within an exact 1-km Euclidean radius. The resulting detector-level POI table contains eleven categories: food, retail, transport, education, healthcare, finance, government and civic, recreation, tourism and accommodation, religion, and parking, fuel, and automotive. The model applies $\log(1+x)$ to these counts before the shared static-feature scaler. Geofabrik is used here for POI extraction. The spatial support used by the forecasting backbones is constructed separately from detector coordinates.

Detector coordinates and spatial support. Detector longitude and latitude are read from the city-specific detector metadata. We construct a symmetric weighted 8-nearest-neighbor graph from haversine distances, add self-loops, and convert the adjacency into order-three Chebyshev supports. This city-local support is retained with the detector identifiers and coordinates for the corresponding backbone. Each city keeps its own detector ordering, while cross-city transfer uses the shared functional interface.

[1] UTD19 project portal: `https://utd19.ethz.ch/`

[2] Geofabrik download portal: `https://download.geofabrik.de/`

Table 2: City-level statistics for the 18 cities used in the experiments. Nodes are the detectors retained after joining traffic, POI, GHSL, and coordinate tables. Observed bins count non-empty UTC 15-minute buckets in the released city summary. The POI snapshot column gives the year of the historical Geofabrik extract recorded in each city manifest.

City	Country	Nodes	Observed 15-min bins	UTC span	POI snapshot
Augsburg	Germany	713	1,919	2017-05-05–2017-05-25	2017
Basel	Switzerland	77	673	2016-10-23–2016-10-30	2016
Bern	Switzerland	755	673	2016-10-23–2016-10-30	2016
Bolton	United Kingdom	166	497	2017-11-13–2017-11-18	2017
Bordeaux	France	425	672	2016-11-20–2016-11-27	2016
Bremen	Germany	548	1,344	2016-09-18–2016-10-02	2016
Constance	Germany	122	672	2017-02-12–2017-02-19	2017
Hamburg	Germany	418	10,029	2016-08-26–2016-12-09	2016
Innsbruck	Austria	16	2,880	2017-03-31–2017-04-30	2017
Kassel	Germany	505	391	2016-08-28–2016-09-01	2016
Marseille	France	167	2,881	2017-05-31–2017-06-30	2017
Melbourne	Australia	954	1,440	2018-02-11–2018-02-26	2018
Speyer	Germany	184	1,344	2016-09-18–2016-10-02	2016
Taipeh	China	445	1,324	2017-09-17–2017-10-01	2017
Torino	Italy	399	2,016	2016-09-25–2016-10-16	2016
Utrecht	Netherlands	1,027	384	2017-06-11–2017-06-15	2017
Wolfsburg	Germany	133	1,344	2016-09-18–2016-10-02	2016
Zurich	Switzerland	1,020	672	2015-10-25–2015-11-01	2015

Table 3: External variables and spatial inputs used by the data processing pipeline.

Group	Variables	Model representation
Traffic	flow sum, detector identifier, UTC timestamp, observation mask	15-minute flow history and future target with validity mask.
POI semantics	11 detector-centered OSM count categories	$\log(1 + x)$ static features from the 1-km Geofabrik extraction.
GHSL context	15 population, built-environment, settlement-class, and coverage fields	Joined detector-level static context with source-side robust scaling.
Calendar and DST	hour, weekday, weekend, holiday, UTC offset, DST offset, DST indicator, fold	Ten-dimensional temporal feature matrix with periodic encodings.
Weather	total precipitation, large-scale rain rate, precipitation indicator, weather age, precipitation type	Four continuous channels and one categorical channel aligned from the nearest weather grid.
Spatial support	detector coordinates	Symmetric weighted 8-nearest-neighbor adjacency and order-three Chebyshev supports.

Urban-scale context. GHSL layers provide the static population and built-environment context (Corbane et al., 2019). The processed detector table contains fifteen GHSL fields, including the GHSL epoch, population coverage and population sums at 250 m, 500 m, and 1 km, built-surface coverage and built fractions at the same radii, the settlement-model class, and its valid-coverage indicator. These fields are joined by detector identifier and standardized with the source-side static-feature statistics. The corresponding GHSL data portal is provided in a footnote.[3]

Calendar, holidays, and daylight saving time. The preprocessed calendar table records local and UTC timestamps, local date and hour, weekday, weekend status, public-holiday status, holiday name, UTC offset, daylight-saving offset, daylight-saving indicator, fold, country and subdivision codes, timezone, and the holiday-library version. The temporal covariates comprise ten channels: sinusoidal hour and weekday coordinates, weekend and public-holiday indicators, UTC offset, daylight-saving offset, daylight-saving indicator, and fold. Holiday dates follow the country and subdivision calendars generated with the Python `holidays` package[4]. Calendar and daylight-saving variables are available at future timestamps; future weather is not.

[3]GHSL data portal: `https://human-settlement.emergency.copernicus.eu/datasets.php`

[4]Python `holidays` documentation: `https://holidays.readthedocs.io/`

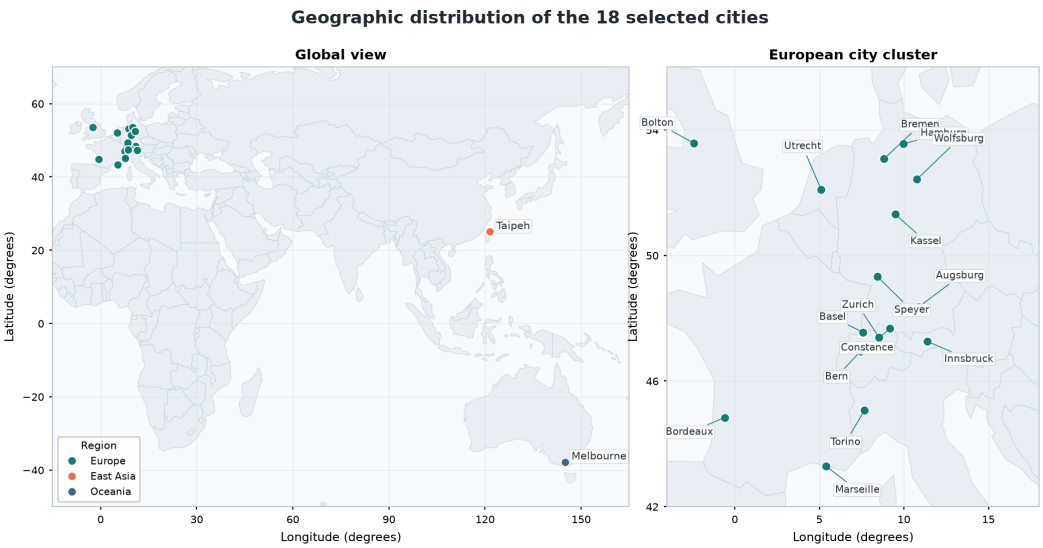


Figure 5: Geographic distribution of the 18 selected cities. Points use the mean detector coordinates from the city-level detector tables. The right panel enlarges the European cluster; Taipeh and Melbourne remain visible in the global panel.

C.2 Task Protocol and City Splits

We evaluate the forecasting task with a leave-one-city-out (LOCO) protocol. Each target city is held out from source-side function construction, and the remaining eligible cities provide the historical source pool. The target receives the source-derived functional artifact together with a limited set of local observations and then predicts the subsequent interval.

City roles. The experimental city pool contains 18 eligible cities, of which 16 are target cities. Kassel and Utrecht are reserved as source-only cities because their available temporal coverage does not satisfy the target-side adaptation and evaluation requirement. The complete city list and the corresponding observation spans are reported in Table 2.

For a target city c, the source set is $\mathcal{S}_c = \mathcal{E} \setminus \{c\}$, where $\mathcal{E}$ denotes the 18-city eligible pool. Thus, every target run uses 17 source cities, including the two source-only cities. The held-out target records enter after source-side construction. The limited local observations support adaptation, while observations after the adaptation period are reserved for final evaluation.

Source-side chronology. Each source timeline is divided into four contiguous segments with boundaries at 70%, 80%, 90%, and 100% of the available timeline. The first segment supports source fitting, the second supports early stopping, the third supports source quality assessment, and the final segment is used by the outer LOCO audit. Source base predictions for fitting follow a forward-chaining construction, so each held origin is predicted from information available before that origin.

Target-side few-shot adaptation. We evaluate one-day and three-day target-data settings at 15-minute resolution. In each setting, the available target observations are divided chronologically into a training segment followed by a validation segment. All model selection decisions use the validation segment, while final evaluation uses the observations that follow the adaptation period.

Forecasting windows. At each forecast origin, the forecaster uses the latest $L = 12$ traffic observations at 15-minute resolution and predicts the following $H = 12$ observations. The observation mask introduced in the data representation identifies the available entries used to form forecasting examples and compute the reported metrics. The chronological source and target partitions above determine which observations contribute to fitting, adaptation, selection, and final assessment.

C.3 Baselines and Implementation Details

Baselines. We adapt twelve cross-city transfer and federated forecasting methods to the FACO task setting. Each method retains its original transfer mechanism and is evaluated with the same

city split, target observations, forecasting horizon, and evaluation protocol. We also include the **A2TTA** anchored-and-agile test-time adaptation baseline for evolving traffic sensor networks (Yin et al., 2026).

Domain alignment and selective regional transfer. **DASTNet** (Tang et al., 2022) pre-trains on multiple source networks and learns domain-invariant node representations through graph representation learning and adversarial multi-domain alignment. **CrossTReS** (Jin et al., 2022) estimates the relevance of source regions to the target city and reweights them for selective transfer. **SERT** (Liu et al., 2025c) forms source–target region pairs from satellite imagery and POI information, then aligns matched regions through contrastive domain adaptation. **CMSVC** (Bi et al., 2025) selects useful regions from multiple source cities, preserves their spatial relations in a virtual city, and transfers the resulting multi-source representation to the target.

Pre-training, prompting, and structural transfer. **FEPCross** (Liu et al., 2024) learns traffic representations jointly in the time, amplitude, and phase domains through masked reconstruction and contrastive pre-training, followed by few-shot target adaptation. **TransGTR** (Jin et al., 2023b) distills city-independent node features and transfers a learned graph structure generator together with the forecasting model. **ST-MPPT** (Guo et al., 2026) learns spatial and temporal representations through decoupled masked pre-training and uses input-dependent prompts for target adaptation. **STGP** (Hu et al., 2024) uses a task-agnostic spatio-temporal graph network and a two-stage prompting procedure to encode domain and task information.

Pattern retrieval and graph meta-learning. **TPB** (Liu et al., 2023) clusters source traffic representations into a traffic-pattern bank that is queried by target observations to support graph construction and forecasting. **MTPB** (Liu et al., 2025b) extends this design with source-side spatio-temporal pre-training and pattern banks at multiple temporal scales. **ST-GFSL** (Lu et al., 2022) learns node-level meta-knowledge to generate city-specific parameters and reconstructs graph structure during meta-learning.

Federated transfer. **pFedCTP** (Zhang et al., 2024) shares spatially independent traffic patterns across source-city clients and uses layer-wise interpolation between global and local parameters to construct a personalized target model. Recent federated frameworks explicitly target cross-city traffic knowledge transfer and synchronous spatio-temporal graph learning across cities (Zeng et al., 2026; Yu et al., 2026).

Implementation details. The city-specific forecasting models are drawn from three architecture families: STGCN, STID, and PDFormer. Target observations are divided chronologically, with 80% used for model fitting and 20% for validation. All experiments are trained on two NVIDIA GeForce RTX 4080 SUPER GPUs.

The cross-city evolution model is fitted from source-only data using the three forward-chaining out-of-fold partitions specified above. Table 4 collects the remaining optimization, relay, and adaptation settings; Appendix D.4 defines the one-factor sensitivity protocol.

Evaluation metrics. Let Ω_c denote the observed target entries used to evaluate city c, with ground truth y_i and prediction $\widehat{y}_i$ for each $i \in \Omega_c$. We report weighted absolute percentage error (WAPE), mean absolute error (MAE), and root mean squared error (RMSE):

$$\mathrm{WAPE}_c = \frac{\sum_{i\in\Omega_c} |\widehat{y}_i - y_i|}{\sum_{i\in\Omega_c} |y_i|}, \tag{91}$$

$$\mathrm{MAE}_c = \frac{1}{|\Omega_c|} \sum_{i\in\Omega_c} |\widehat{y}_i - y_i|, \tag{92}$$

$$\mathrm{RMSE}_c = \sqrt{\frac{1}{|\Omega_c|} \sum_{i\in\Omega_c} (\widehat{y}_i - y_i)^2}. \tag{93}$$

D Additional Experimental Results

This section provides complementary empirical evidence beyond the main experimental results. We first examine FACO under a stricter one-day target observation budget, then evaluate how FACO augments representative forecasting baselines under the same setting. We next test whether the

Table 4: Key hyperparameter settings of FACO.

Symbol	Setting	Value
κ_{sup}	Support-gate temperature	1.0
λ_{hor}	Horizon-dependent trust decay	0.5
r_{relay}	Retained relay rank	4
λ_{pred}	Predictive-interface adaptation learning rate	2×10^{-4}
λ_{sem}	Semantic-interface adaptation learning rate	5×10^{-3}
B_{pred}	Predictive-interface adaptation budget (epochs $\times$ steps)	2×40
B_{sem}	Semantic-interface adaptation budget (epochs $\times$ steps)	2×40
seed	Random seed	42
L	Input history length	12
H	Forecast horizon	12
D_{adapt}	Main target-adaptation prefix	1/3 days
η_{forecast}	Forecasting-model learning rate	10^{-3}
ω_{wd}	Forecasting-model weight decay	10^{-5}
E_{forecast}	Maximum forecasting epochs	100
P_{early}	Early-stopping patience	8 epochs
S_{src}	Source evolution optimization steps	80
S_{enh}	Enhanced evolution optimization steps	120
α	Target relay-strength candidates	$\{0, 0.125, 0.25, 0.5, 0.75, 1\}$

transferred interface remains usable when the target backbone is changed, and finally report one-at-a-time sensitivity sweeps for the relay and target-adaptation hyperparameters. The discussion of the first three subsections is integrated with the corresponding tables; the hyperparameter subsection is retained as a direct record of the measured settings.

D.1 One-Day Target Adaptation Benchmark

Table 5 reports all 16 target cities using the one-day adaptation setting. The table should be read city by city rather than as an aggregate claim: every city is evaluated with the same one-day target prefix, while the traffic scale and the number of observed detectors vary substantially across locations. FACO is the lowest-MAE and lowest-WAPE entry for every city in this table. The separation is especially visible in Innsbruck, where FACO reaches MAE/WAPE values of 305.9/0.271 versus 792.4/0.703 for Local-only, and in Taipeh, where it reaches 713.6/0.279 versus 1,362.6/0.532. These two cases differ markedly in absolute error scale, yet both show that the transferred interface remains useful when the target supplies only a short local prefix. The smaller margin in Bolton (93.1 versus 119.1 MAE and 0.089 versus 0.114 WAPE) is also informative: the method does not depend on an unusually difficult city to produce a gain, but improves across the full heterogeneous set.

D.2 Baseline Augmentation with FACO under One-Day and Three-Day Adaptation

This subsection reports paired baseline and baseline+FACO results under both one-day and three-day target adaptation. The first table covers the one-day prefix for eight target cities and six baselines; the subsequent three-day results table (Table 7) reports the same cities and methods. Each cell is written as standalone baseline (baseline + FACO), and the final row summarizes the mean relative reduction across the six baselines for the corresponding city and metric.

Under one-day adaptation, FACO reduces both MAE and WAPE for every baseline–city pair. The city-level mean MAE reduction ranges from 36.5% in Melbourne to 52.7% in Torino, and the WAPE reduction ranges from 36.2% to 52.5%. The largest average gains occur in Torino and Bordeaux, at 52.7/52.5% and 50.4/50.1% for MAE/WAPE, respectively. The three-day table shows the same direction for every paired result, with mean MAE reductions from 21.6% in Taipeh to 35.3% in Augsburg and mean WAPE reductions from 21.4% to 34.9%. The duration comparison is systematic: for example, the Torino mean MAE/WAPE reduction changes from 52.7/52.5% in the one-day setting to 27.6/27.2% in the three-day setting, while Bordeaux changes from 50.4/50.1% to 28.7/28.5%. The paired tables quantify the improvements contributed by FACO across six distinct

Table 5: Results on the 16-city FACO benchmark with a one-day target adaptation prefix. Each block contains eight target cities, and each city reports MAE and WAPE. Lower is better; the best result for each city and metric is bold and the second-best is underlined.

Method	Augs.		Basel		Bern		Bolton		Bord.		Bremen		Const.		Hamb.	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
CMSVC (2025)	203.3	0.398	212.8	0.411	239.6	0.519	105.4	0.101	339.6	0.364	398.8	0.410	170.0	0.389	383.3	0.533
CrossTReS (2022)	224.6	0.440	272.1	0.526	224.7	0.486	97.3	0.093	373.4	0.401	385.4	0.396	228.2	0.522	388.2	0.540
DASTNet (2022)	327.1	0.641	371.6	0.718	404.2	0.875	148.6	0.143	592.4	0.636	673.7	0.692	303.5	0.695	394.4	0.549
FEPCross (2024)	244.4	0.479	248.3	0.480	286.4	0.620	105.8	0.102	360.3	0.387	528.7	0.543	271.7	0.622	380.3	0.529
MTPB (2025)	195.9	0.384	210.3	0.406	246.6	0.534	96.5	0.093	315.9	0.339	357.5	0.367	167.9	0.384	366.8	0.510
pFedCTP (2024)	349.6	0.685	382.7	0.740	421.3	0.912	190.2	0.183	619.4	0.665	653.6	0.671	327.8	0.750	414.7	0.577
SERT (2025)	192.0	0.376	197.3	0.381	250.0	0.541	99.1	0.095	321.4	0.345	324.9	0.334	164.8	0.377	389.6	0.542
ST-GFSL (2022)	219.5	0.430	275.0	0.531	224.2	0.485	99.1	0.095	403.6	0.433	384.5	0.395	242.9	0.556	377.7	0.526
ST-MPPT (2026)	199.6	0.391	211.3	0.408	227.7	0.493	94.9	0.091	303.4	0.326	334.0	0.343	175.8	0.402	384.4	0.535
STGP(2024)	205.1	0.402	206.7	0.400	222.4	0.481	105.4	0.101	299.8	0.322	339.6	0.349	178.5	0.409	329.3	0.458
TPB (2023)	190.4	0.373	199.8	0.386	234.2	0.507	94.6	0.091	303.5	0.326	375.4	0.386	170.5	0.390	365.2	0.508
TransGTR (2023)	267.2	0.524	261.5	0.505	365.8	0.792	209.5	0.201	551.5	0.592	610.1	0.627	264.0	0.604	381.1	0.530
A2TTA (TrafficStream, 2026)	239.7	0.470	251.3	0.486	231.6	0.501	110.1	0.106	313.2	0.336	391.8	0.402	276.9	0.634	435.1	0.605
Local-only	303.5	0.596	320.5	0.623	366.1	0.795	119.1	0.114	588.6	0.635	673.3	0.695	331.2	0.762	355.1	0.492
FACO (ours)	**165.0**	**0.324**	**148.6**	**0.289**	**209.0**	**0.454**	**93.1**	**0.089**	**227.5**	**0.245**	**278.4**	**0.287**	**162.2**	**0.373**	**195.3**	**0.271**

Method	Inns.		Mars.		Melb.		Speyer		Taipeh		Torino		Wolfs.		Zurich	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
CMSVC (2025)	463.8	0.412	634.1	0.298	63.7	0.297	280.2	0.397	884.9	0.344	420.8	0.337	526.1	0.558	301.8	0.286
CrossTReS (2022)	463.9	0.412	616.5	0.290	72.9	0.340	304.8	0.432	945.0	0.367	547.6	0.439	581.3	0.617	363.9	0.345
DASTNet (2022)	642.5	0.571	1027.1	0.483	109.4	0.510	471.9	0.669	1344.9	0.523	733.1	0.587	832.9	0.884	573.7	0.543
FEPCross (2024)	501.7	0.446	661.6	0.311	69.7	0.325	349.5	0.496	994.3	0.387	438.3	0.351	616.7	0.655	367.6	0.348
MTPB (2025)	469.4	0.417	624.5	0.294	63.2	0.295	294.4	0.417	896.9	0.349	388.5	0.311	480.0	0.509	294.4	0.279
pFedCTP (2024)	663.6	0.589	1103.2	0.519	112.8	0.527	479.4	0.680	1422.2	0.553	855.2	0.685	869.0	0.922	584.4	0.553
SERT (2025)	467.5	0.415	581.0	0.273	63.0	0.294	259.1	0.367	933.0	0.363	434.5	0.348	510.7	0.542	308.9	0.293
ST-GFSL (2022)	460.3	0.409	608.9	0.286	73.2	0.342	304.4	0.432	920.2	0.358	569.7	0.456	515.0	0.547	358.6	0.340
ST-MPPT (2026)	465.0	0.413	607.9	0.286	63.5	0.296	286.9	0.407	860.2	0.334	411.5	0.330	456.9	0.485	299.9	0.284
STGP(2024)	442.9	0.393	664.6	0.313	62.1	0.290	318.2	0.451	900.4	0.350	431.4	0.346	531.9	0.564	300.8	0.285
TPB (2023)	440.6	0.391	600.5	0.282	63.1	0.294	295.7	0.419	866.8	0.337	395.7	0.317	497.3	0.528	309.8	0.293
TransGTR (2023)	660.2	0.586	644.3	0.303	68.5	0.320	391.7	0.556	1008.4	0.392	806.7	0.646	825.8	0.876	531.4	0.503
A2TTA (TrafficStream, 2026)	728.0	0.647	599.6	0.282	62.1	0.290	263.6	0.374	1074.4	0.418	462.4	0.370	964.2	1.023	335.5	0.318
Local-only	792.4	0.703	1008.6	0.474	103.2	0.484	467.3	0.666	1362.6	0.532	783.0	0.630	785.5	0.838	606.1	0.577
FACO (ours)	**305.9**	**0.271**	**411.1**	**0.193**	**57.6**	**0.270**	**150.4**	**0.214**	**713.6**	**0.279**	**370.3**	**0.298**	**351.1**	**0.375**	**291.2**	**0.273**

Table 6: One-day target adaptation results for six baselines and their FACO-augmented counterparts across eight cities in seven countries. Each cell reports MAE or WAPE as baseline (baseline + FACO); the final row reports the per-city mean relative gain across the six methods.

Method	Augsburg		Basel		Bordeaux		Innsbruck	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
DASTNet (+FACO)	327.1(151.5)	0.641(0.298)	371.6(171.0)	0.718(0.332)	592.4(225.5)	0.636(0.243)	642.5(282.9)	0.571(0.251)
FEPCross (+FACO)	244.4(218.2)	0.479(0.429)	248.3(149.2)	0.480(0.290)	360.3(197.3)	0.387(0.213)	501.7(282.5)	0.446(0.251)
pFedCTP (+FACO)	349.6(155.9)	0.685(0.306)	382.7(178.0)	0.740(0.346)	619.4(233.5)	0.665(0.252)	663.6(265.1)	0.589(0.235)
TransGTR (+FACO)	267.2(149.5)	0.524(0.294)	261.5(180.8)	0.505(0.351)	551.5(228.2)	0.592(0.246)	660.2(266.9)	0.586(0.237)
TPB (+FACO)	190.4(125.0)	0.373(0.246)	199.8(135.6)	0.386(0.263)	303.5(190.0)	0.326(0.205)	440.6(279.5)	0.391(0.248)
ST-MPPT (+FACO)	199.6(124.6)	0.391(0.245)	211.3(136.2)	0.408(0.265)	303.4(191.3)	0.326(0.206)	465.0(274.1)	0.413(0.243)
Avg. gain (%)	+39.3%	+39.2%	+41.0%	+40.7%	+50.4%	+50.1%	+49.5%	+49.6%

Method	Melbourne		Taipeh		Torino		Hamburg	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
DASTNet (+FACO)	109.4(51.0)	0.510(0.239)	1344.9(658.1)	0.523(0.257)	733.1(281.5)	0.587(0.226)	394.4(195.9)	0.549(0.272)
FEPCross (+FACO)	69.7(48.2)	0.325(0.226)	994.3(590.8)	0.387(0.231)	438.3(236.2)	0.351(0.190)	380.3(193.7)	0.529(0.269)
pFedCTP (+FACO)	112.8(51.6)	0.527(0.242)	1422.2(654.9)	0.553(0.256)	855.2(348.3)	0.685(0.280)	414.7(201.1)	0.577(0.279)
TransGTR (+FACO)	68.5(50.9)	0.320(0.239)	1008.4(641.4)	0.392(0.251)	806.7(276.7)	0.646(0.223)	381.1(205.6)	0.530(0.285)
TPB (+FACO)	63.1(46.0)	0.294(0.216)	866.8(595.3)	0.337(0.233)	395.7(234.4)	0.317(0.189)	365.2(196.7)	0.508(0.273)
ST-MPPT (+FACO)	63.5(45.8)	0.296(0.215)	860.2(610.5)	0.334(0.239)	411.5(236.5)	0.330(0.190)	384.4(197.0)	0.535(0.273)
Avg. gain (%)	+36.5%	+36.2%	+40.4%	+40.1%	+52.7%	+52.5%	+48.6%	+48.8%

baseline mechanisms, with the strongest relative gains appearing when the target adaptation prefix is shortest.

D.3 Cross-Backbone Target Robustness

In the primary setup, one of three backbone families is randomly assigned to each target city. Here we keep source-city predictors and transferred memory fixed, then replace only the target backbone

Table 7: Three-day target adaptation results for six baselines and their FACO-augmented counterparts across eight cities in seven countries. Each cell reports MAE or WAPE as baseline (baseline + FACO); the final row reports the per-city mean relative gain across the six methods.

Method	Augsburg		Basel		Bordeaux		Innsbruck	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
DASTNet (+FACO)	281.8(146.1)	0.542(0.283)	282.8(164.4)	0.581(0.341)	476.0(217.6)	0.523(0.240)	538.1(274.5)	0.479(0.247)
FEPCross (+FACO)	368.1(235.6)	0.708(0.456)	193.0(150.9)	0.396(0.313)	255.4(207.0)	0.281(0.228)	421.7(284.9)	0.376(0.256)
pFedCTP (+FACO)	352.8(146.7)	0.678(0.284)	279.3(169.4)	0.573(0.351)	466.5(252.8)	0.513(0.279)	449.4(261.9)	0.400(0.235)
TransGTR (+FACO)	188.8(142.8)	0.363(0.276)	189.7(175.2)	0.389(0.364)	281.5(219.4)	0.310(0.242)	409.7(248.0)	0.365(0.223)
TPB (+FACO)	337.0(240.3)	0.648(0.465)	168.4(146.5)	0.346(0.304)	239.5(200.2)	0.263(0.221)	366.6(283.9)	0.327(0.255)
ST-MPPT (+FACO)	199.9(167.7)	0.384(0.325)	170.7(145.7)	0.350(0.302)	235.5(200.3)	0.259(0.221)	370.7(285.4)	0.330(0.257)
Avg. gain (%)	+35.3%	+34.9%	+23.0%	+22.2%	+28.7%	+28.5%	+34.7%	+34.1%

Method	Melbourne		Taipeh		Torino		Hamburg	
	MAE	WAPE	MAE	WAPE	MAE	WAPE	MAE	WAPE
DASTNet (+FACO)	85.4(48.1)	0.401(0.227)	1058.5(694.9)	0.417(0.275)	524.2(234.3)	0.429(0.193)	354.8(191.6)	0.488(0.265)
FEPCross (+FACO)	55.4(46.4)	0.260(0.219)	709.3(613.7)	0.279(0.243)	289.8(237.5)	0.237(0.195)	260.4(190.7)	0.358(0.264)
pFedCTP (+FACO)	83.6(52.0)	0.393(0.246)	1329.1(692.2)	0.524(0.274)	414.9(279.0)	0.340(0.230)	380.0(196.8)	0.523(0.273)
TransGTR (+FACO)	57.2(47.3)	0.269(0.223)	775.6(679.7)	0.306(0.269)	328.0(237.0)	0.268(0.195)	265.3(191.8)	0.365(0.266)
TPB (+FACO)	49.0(44.6)	0.230(0.211)	666.8(587.1)	0.263(0.232)	279.5(233.1)	0.229(0.192)	234.0(184.2)	0.322(0.255)
ST-MPPT (+FACO)	48.4(44.1)	0.227(0.208)	643.5(580.4)	0.254(0.229)	271.6(229.8)	0.222(0.189)	238.6(183.9)	0.328(0.255)
Avg. gain (%)	+22.2%	+21.7%	+21.6%	+21.4%	+27.6%	+27.2%	+32.1%	+31.7%

with each of the other two families. This yields two alternative target configurations per city. Reading each reference together with its two alternatives shows that the effect of the target backbone is local rather than one-directional. For example, both alternatives improve on the assigned backbone in Augsburg, while the Taipeh–STID variant lowers MAE by 27.87 and WAPE by 0.0110. In contrast, the Bern–STGCN variant raises MAE by 19.70 and WAPE by 0.0456, and both alternatives are slightly worse for Hamburg. These paired changes make the role of target-side architecture explicit without changing the transferred source artifact. Target instantiation can shift the final error in either direction, while the cross-backbone results provide a direct robustness check of the relay interface.

D.4 Hyperparameter Sensitivity

We study the sensitivity of FACO to the hyperparameters governing target-side interface adaptation, functional trust, and relay compression. Recall that the target predictive-functional interface is instantiated as

$$\psi_T = \mathrm{Adapt}\left(\psi_0; \mathcal{D}_T^K\right), \qquad \psi_T = (\eta_T, \omega_T),$$

where η_T parameterizes target-side semantic alignment and ω_T parameterizes the predictive interface. We denote the learning rates used to specialize these two components by λ_{sem} and λ_{pred}, respectively, and their corresponding optimization budgets by B_{sem} and B_{pred}. The budgets are defined as

$$B_{\mathrm{sem}} = E_{\mathrm{sem}} S_{\mathrm{sem}}, \qquad B_{\mathrm{pred}} = E_{\mathrm{pred}} S_{\mathrm{pred}}, \tag{94}$$

where E and S denote the maximum numbers of adaptation epochs and update steps per epoch. These quantities control the optimization performed within $\mathrm{Adapt}(\psi_0; \mathcal{D}_T^K)$ while leaving the frozen relay memory $\mathcal{M}$ unchanged.

The support component of functional trust is controlled by a support threshold τ and a temperature parameter $\kappa_{\mathrm{sup}} > 0$:

$$g^{\mathrm{sup}}(d) = \frac{1}{1 + \left(\frac{\max(d-\tau, 0)}{\kappa_{\mathrm{sup}}}\right)^2}. \tag{95}$$

Here, κ_{sup} controls the smoothness of the transition from well-supported to weakly supported target states. Smaller values produce a sharper decrease in relay participation once $d > \tau$, whereas larger values yield a smoother attenuation.

Horizon-dependent relay confidence is parameterized by $\lambda_{\mathrm{hor}} \geq 0$:

$$g^{\mathrm{hor}}(h) = \exp\left(-\lambda_{\mathrm{hor}} \frac{h-1}{\max(H-1, 1)}\right). \tag{96}$$

Larger λ_{hor} values induce a faster reduction in relay confidence toward longer forecasting horizons. Both g^{sup} and g^{hor} contribute to the functional trust $\gamma_{t,n,h}^T(\alpha)$.

Table 8: Cross-backbone target results under the three-day adaptation setting. Each city shows its assigned-backbone reference once across two alternative target-backbone rows. Deltas are variant minus reference; negative values indicate lower error.

Target (assigned backbone)	Reference MAE / WAPE / RMSE	Alternative backbone	Variant MAE / WAPE / RMSE	Δ MAE / WAPE / RMSE
Augsburg (STGCN)	127.97/0.2476/340.21	STID	118.35/0.2290/327.43	(−9.62)/(−0.0186)/(−12.78)
		PDFormer	120.68/0.2335/329.03	(−7.29)/(−0.0141)/(−11.19)
Basel (STID)	136.89/0.2840/233.71	STGCN	137.96/0.2863/225.39	(+1.07)/(+0.0022)/(−8.31)
		PDFormer	138.50/0.2874/231.76	(+1.61)/(+0.0033)/(−1.95)
Bern (PDFormer)	151.76/0.3508/1437.11	STGCN	171.47/0.3964/1602.43	(+19.70)/(+0.0456)/(+165.32)
		STID	149.96/0.3467/1446.25	(−1.80)/(−0.0042)/(+9.14)
Bolton (STGCN)	87.01/0.0833/199.23	STID	86.04/0.0823/197.33	(−0.97)/(−0.0009)/(−1.90)
		PDFormer	75.50/0.0723/174.41	(−11.50)/(−0.0110)/(−24.82)
Bordeaux (STID)	201.30/0.2221/321.12	STGCN	196.79/0.2172/322.98	(−4.51)/(−0.0050)/(+1.86)
		PDFormer	198.52/0.2191/323.13	(−2.78)/(−0.0031)/(+2.01)
Bremen (PDFormer)	213.73/0.2245/455.62	STGCN	220.53/0.2317/464.16	(+6.80)/(+0.0071)/(+8.55)
		STID	209.20/0.2198/444.38	(−4.54)/(−0.0048)/(−11.23)
Constance (STGCN)	92.87/0.2205/149.67	STID	91.65/0.2176/147.89	(−1.22)/(−0.0029)/(−1.78)
		PDFormer	89.92/0.2135/146.95	(−2.95)/(−0.0070)/(−2.73)
Hamburg (STID)	183.87/0.2547/353.04	STGCN	189.93/0.2631/360.35	(+6.06)/(+0.0084)/(+7.31)
		PDFormer	186.46/0.2583/356.73	(+2.59)/(+0.0036)/(+3.69)
Innsbruck (PDFormer)	240.11/0.2158/398.73	STGCN	230.37/0.2071/379.69	(−9.74)/(−0.0088)/(−19.04)
		STID	242.67/0.2181/421.88	(+2.57)/(+0.0023)/(+23.14)
Marseille (PDFormer)	385.84/0.1808/688.77	STGCN	403.37/0.1890/718.30	(+17.53)/(+0.0082)/(+29.53)
		STID	382.96/0.1795/690.55	(−2.88)/(−0.0013)/(+1.79)
Melbourne (STGCN)	45.58/0.2152/69.49	STID	46.17/0.2180/70.83	(+0.59)/(+0.0028)/(+1.34)
		PDFormer	45.43/0.2145/69.20	(−0.15)/(−0.0007)/(−0.29)
Speyer (STID)	135.25/0.1950/217.44	STGCN	137.09/0.1977/219.07	(+1.84)/(+0.0027)/(+1.63)
		PDFormer	132.84/0.1916/215.12	(−2.41)/(−0.0035)/(−2.32)
Taipeh (PDFormer)	621.26/0.2456/2342.97	STGCN	621.77/0.2458/2422.80	(+0.51)/(+0.0002)/(+79.82)
		STID	593.39/0.2346/2254.61	(−27.87)/(−0.0110)/(−88.37)
Torino (STGCN)	224.80/0.1850/404.62	STID	220.91/0.1818/390.83	(−3.89)/(−0.0032)/(−13.79)
		PDFormer	226.97/0.1868/399.67	(+2.16)/(+0.0018)/(−4.95)
Wolfsburg (PDFormer)	282.13/0.3131/452.12	STGCN	280.40/0.3112/445.94	(−1.73)/(−0.0019)/(−6.18)
		STID	271.89/0.3017/432.08	(−10.23)/(−0.0114)/(−20.04)
Zurich (STGCN)	189.06/0.1821/284.04	STID	194.11/0.1870/290.22	(+5.05)/(+0.0049)/(+6.18)
		PDFormer	191.66/0.1846/285.91	(+2.60)/(+0.0025)/(+1.87)

Finally, the retained relay rank r_{relay} controls the dimensionality of the consolidated interaction term. Given the singular value decomposition of $\bar{R}_{\text{full}}$, FACO retains the leading r_{relay} components:

$$\bar{R} = U_{:,1:r_{\text{relay}}} \Sigma_{1:r_{\text{relay}},1:r_{\text{relay}}} V^{\top}_{:,1:r_{\text{relay}}} \approx \bar{R}_{\text{full}}. \tag{97}$$

The following experiments vary one hyperparameter at a time while keeping the remaining configuration fixed, and report the resulting MAE and WAPE.

D.5 More Case Studies

We provide two additional case studies to visualize how FACO behaves under different target-adaptation budgets. Each figure contains three target cities selected by their full-test relative WAPE gain and three traffic-history regimes within each city. The black curve denotes observed traffic, the dashed orange curve denotes target-local-only forecasting, and the blue curve denotes FACO. Shaded regions show the 10–90

Under the one-day adaptation budget, the selected cities span a wide range of full-test gains: Speyer represents the largest-gain case (+67.8%), Torino the median-gain case (+52.7%), and Bolton the weakest-gain case (+12.9%). Across low, middle, and high pre-history traffic regimes, FACO generally follows the observed trajectory more closely than target-local-only forecasting. The contrast is most visible in the low-flow and middle-flow windows, where the local-only trajectory can remain biased in level or trend across the 12-step horizon. The weakest-gain row still shows a consistent trajectory-level alignment, indicating that the case-level benefit is not restricted to the city with the largest aggregate improvement.

With three days of target observations, the selected examples are Innsbruck (+34.1%) for the largest-gain case, Basel (+22.8%) for the median-gain case, and Bern (+15.4%) for the weakest-gain case. FACO continues to track the observed trajectories across the three traffic regimes, while the local-only forecasts show larger level or trend deviations in several panels. Comparing the two adaptation budgets, the plots show that additional target observations improve the target-local reference in some

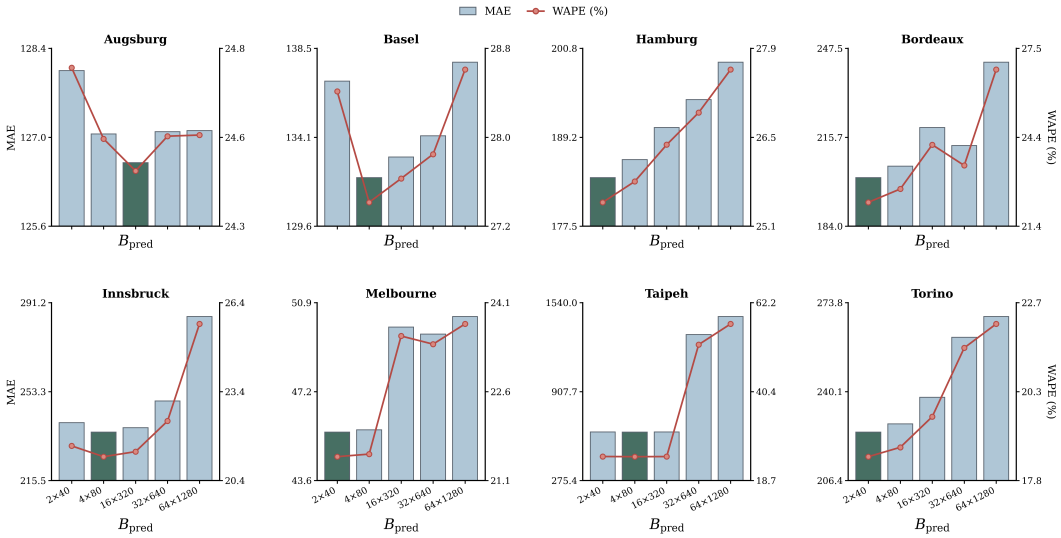


Figure 6: Sensitivity to the predictive-interface adaptation budget B_{pred}.

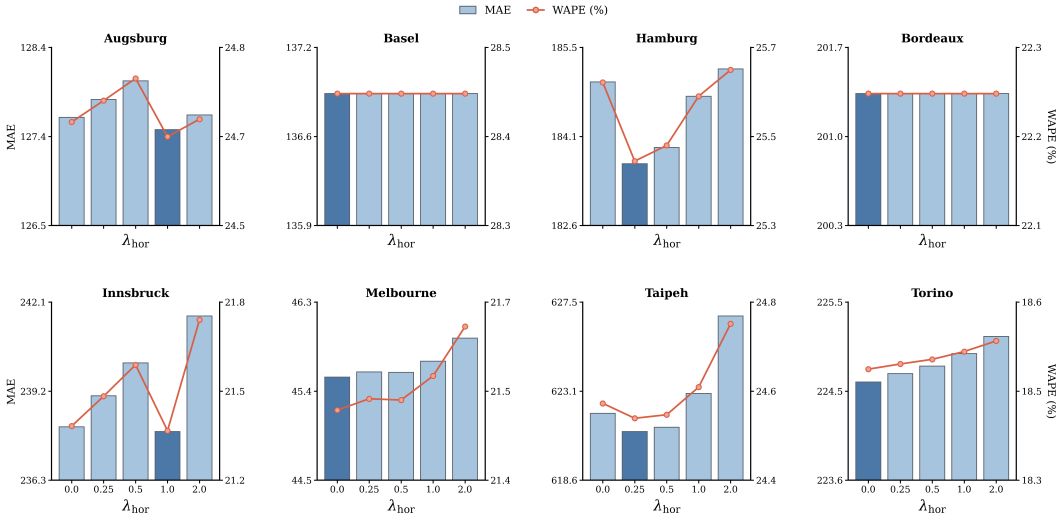


Figure 7: Sensitivity to the horizon-dependent confidence decay λ_{hor}.

regimes, but the FACO trajectory remains consistently aligned with the observed evolution across both budgets. These examples illustrate how the released cross-city dynamics complement target-local forecasting under different traffic conditions.

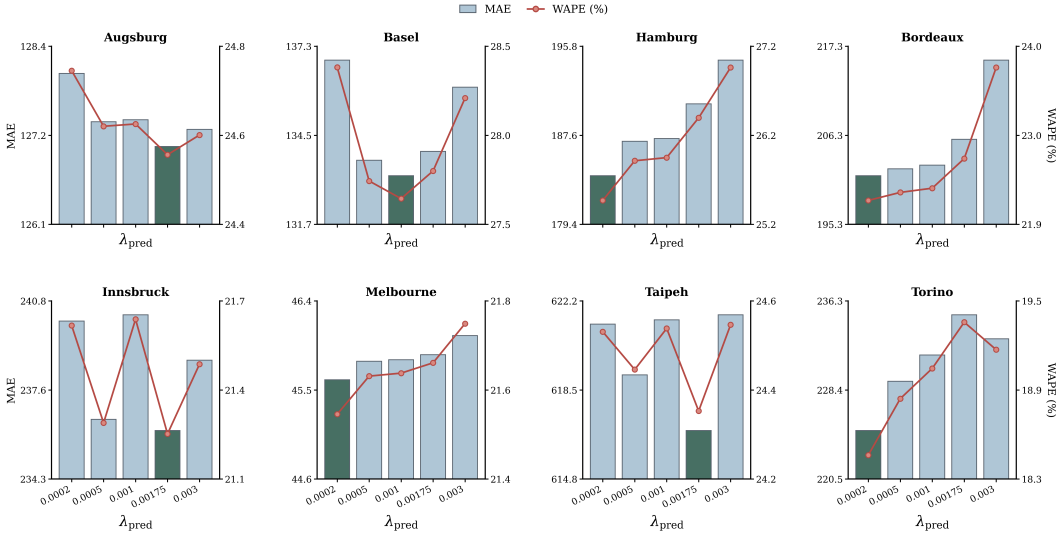


Figure 8: Sensitivity to the predictive-interface adaptation learning rate λ_{pred}.

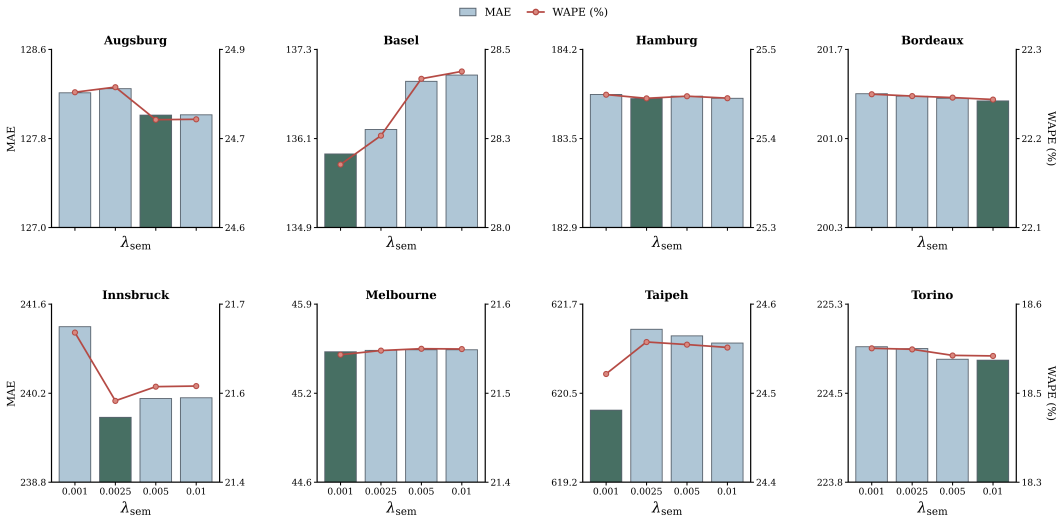


Figure 9: Sensitivity to the semantic-interface adaptation learning rate λ_{sem}.

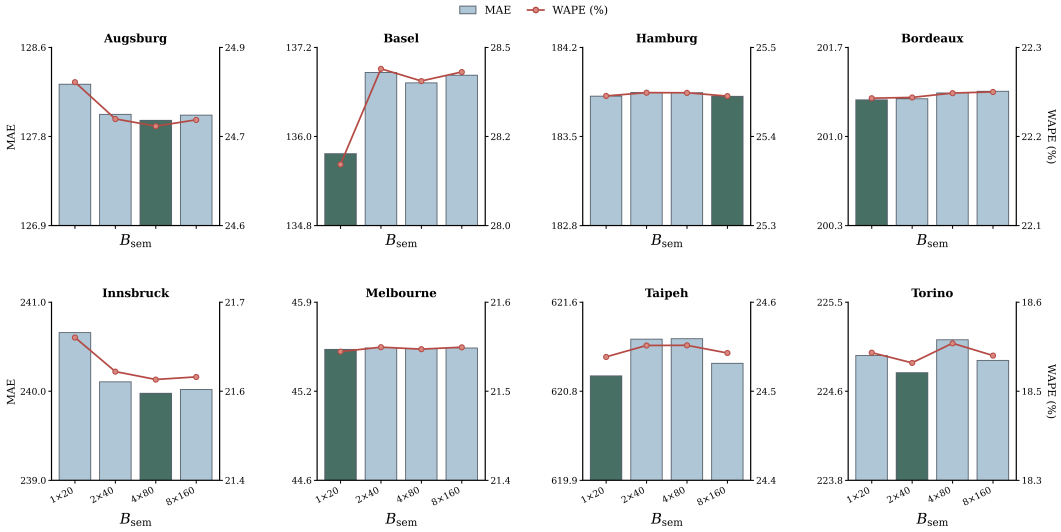


Figure 10: Sensitivity to the semantic-interface adaptation budget B_{sem}.

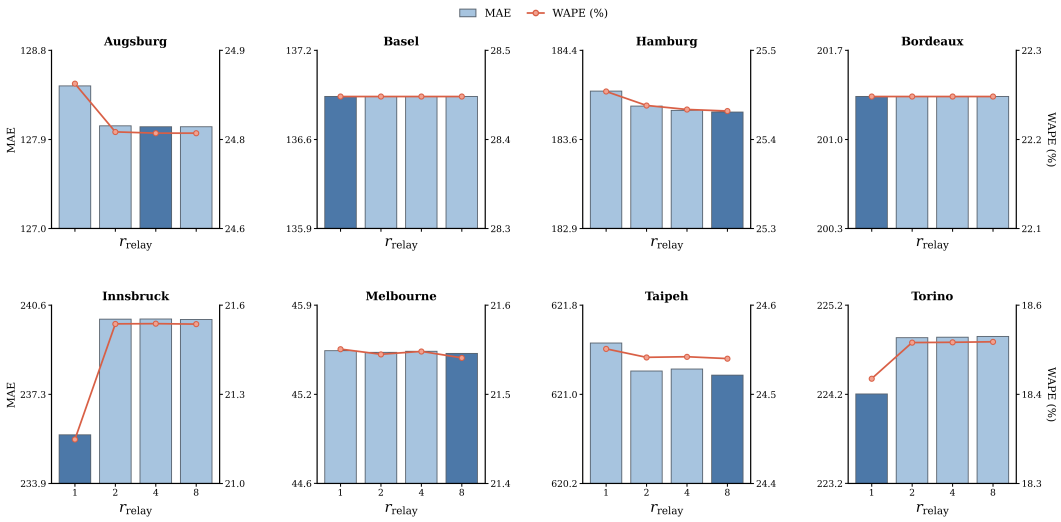


Figure 11: Sensitivity to the retained relay rank r_{relay}.

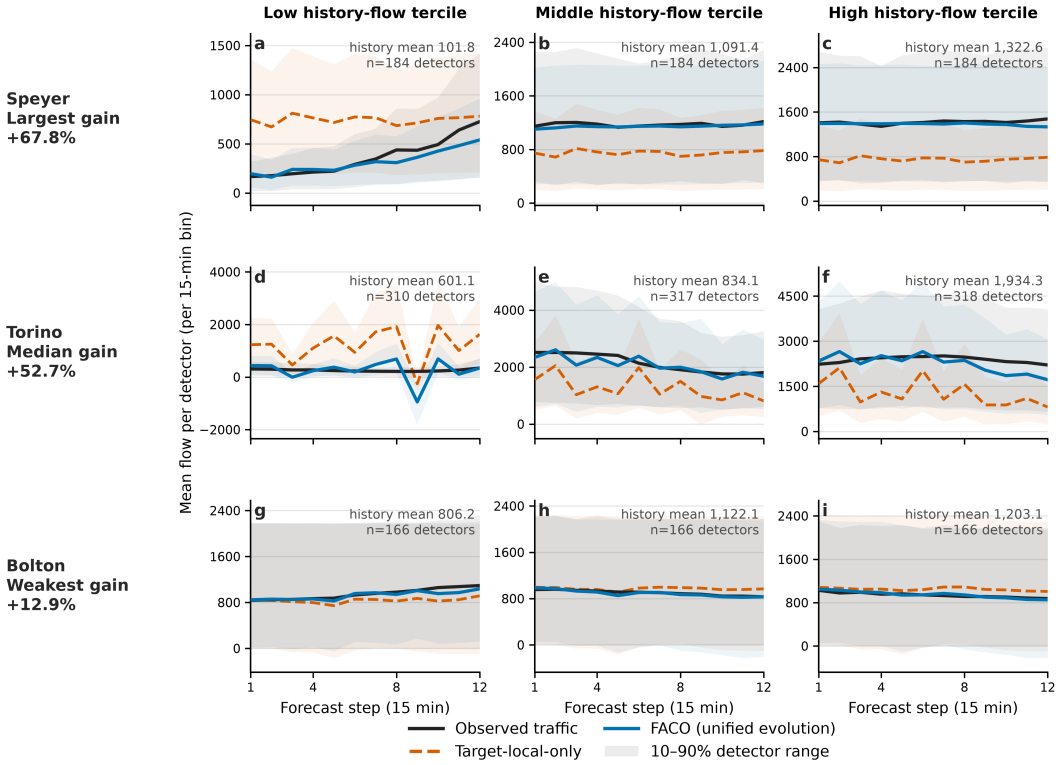


Figure 12: Case studies under one-day target adaptation. The rows show the largest-gain, median-gain, and weakest-gain cities, while the columns correspond to low, middle, and high pre-history traffic regimes. FACO remains closer to the observed trajectories than target-local-only forecasting across the selected regimes, including the weakest-gain city.

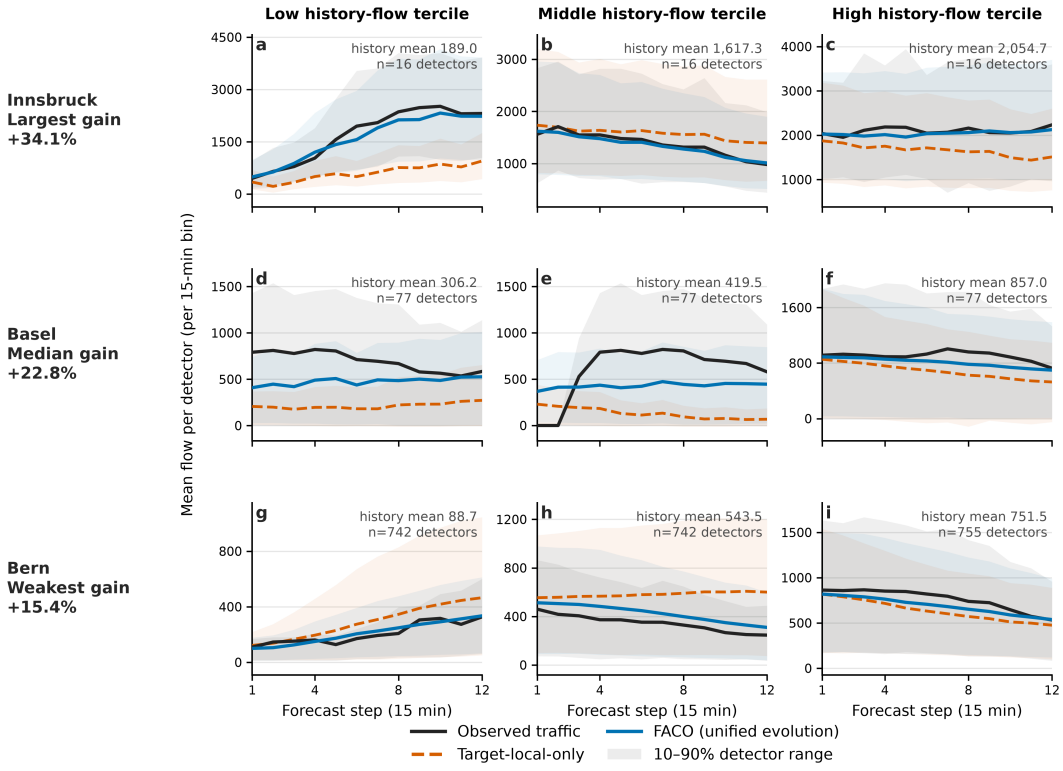


Figure 13: Case studies under three-day target adaptation. The same city and traffic-regime organization is used as in Fig. 12; the longer target prefix provides a direct comparison with the one-day setting.