# RETHINKING FEDERATED TIME SERIES PRETRAINING THROUGH SHARED LATENT DYNAMICS

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

## ABSTRACT

Pretraining a time series model typically relies on centralizing large and diverse collections of data, an assumption incompatible with privacy-sensitive domains such as power grids, healthcare, finance, and distributed sensor networks. Beyond privacy, Federated learning (FL) addresses client heterogeneity by combining personalized local models with a shared global model, placing the choice of shared representation at the center of the model design. Existing FL approaches for time series make it encode observation-level temporal patterns, which vary substantially across heterogeneous clients. We instead identify latent state transitions as a more transferable object for federation and propose FEDSTATE, a federated pretraining framework that keeps client models private and shares only a latent transition operator. The operator is indexed by elapsed physical time rather than patch position and is learned entirely in latent state space, allowing clients with different sampling intervals, channel counts, value scales, and local architectures to learn shared dynamics without aligning their observations. Only the transition operator is exchanged, a small fraction of the model parameters and of the per-round communication of full-model federation. Extensive experiments across eight benchmarks demonstrate that FEDSTATE outperforms full-model federation, advanced federated time series pretraining methods, and centralized models in both deterministic and probabilistic forecasting, including zero-shot transfer to unseen domains without target-domain adaptation.

## 1 INTRODUCTION

Recent advances in general-purpose time-series forecasting have been driven by pretraining on large and diverse collections of series (Ansari et al., 2024; Woo et al., 2024; Das et al., 2024). However, such pretraining typically assumes centralized access to data. This assumption is often untenable in some data-sensitive domains such as power grids, healthcare, finance, and distributed sensing, where records remain siloed because of privacy, regulation, ownership, or operational constraints. Federated learning (McMahan et al., 2017) provides a natural alternative. Beyond keeping raw data local, it addresses heterogeneity by pairing client-specific personalized models with a shared global model (Li et al., 2021b; Collins et al., 2021). Federated pretraining therefore requires identifying the temporal structure that can be shared across highly heterogeneous clients.

Existing approaches to federated time-series pretraining typically share models that encode observation-level temporal patterns. This includes full-model aggregation (McMahan et al., 2017), federated masked pretraining with heterogeneity-aware alignment (Chen et al., 2025), and federated dataset learning that reduces dataset-specific bias (Chen et al., 2026a). Despite their different designs, the exchanged parameters remain tied to client-specific observations or representations. Across domains, waveforms, scales, channel semantics, and temporal patterns can differ substantially, leading to poorly aligned client updates. Existing remedies regularize local optimization (Li et al., 2020; Karimireddy et al., 2020) or personalize model components (Li et al., 2021b), but retain observation-dependent representations as the shared object. Figure 1 illustrates the resulting mismatch. Cross-domain displacement fields agree far more closely in latent state space than in observation space, which suggests that latent transition dynamics provide a more suitable shared representation across heterogeneous clients than domain-specific observation patterns.



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.

A second source of heterogeneity is temporal discretization. Real-world time-series collections span sampling periods from minutes to months, yet many sequence models, including patch-based forecasters (Nie et al., 2023) and pretrained forecasters (Das et al., 2024), parameterize temporal evolution by patch or token position. The same positional offset can therefore correspond to different elapsed times across clients, introducing heterogeneity that arises from the sampling grid itself. As a result, the same model coordinate may be trained to represent different physical horizons on different clients, making positional dynamics difficult to share consistently. A shared dynamics model must therefore assign consistent temporal semantics across clients with different sampling periods.

Based on these two observations, we propose FEDSTATE, a federated pretraining framework that separates client-specific observation modeling from shared temporal dynamics. Each client retains its encoder and decoder, while only an interval-conditioned latent transition operator is federated. The private observation models accommodate differences in channels and value scales, while physical-time conditioning gives the shared operator consistent temporal semantics across sampling grids. We train the operator entirely in latent state space and further introduce heterogeneity-aware aggregation that reweights client updates according to their transition errors. Because only the transition operator is exchanged, each client uploads about a fifth of what full-model federation sends per round. Across eight benchmarks, FEDSTATE consistently outperforms full-model federation at that budget, while the learned transition operator transfers zero-shot to unseen domains. Our contributions can be summarized as four-fold:

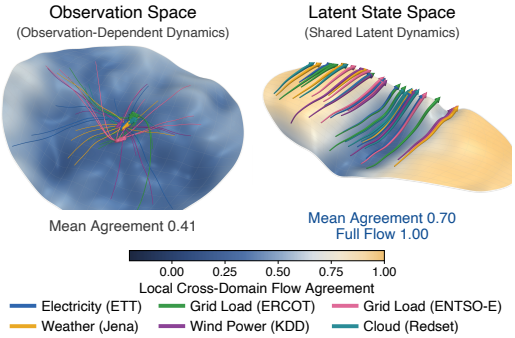



Figure 1: Cross-domain displacement fields in value space (left) and latent state space (right). Relief and color indicate local agreement, fibers show domain-specific trajectories, and values below report mean agreement.

- We characterize the shared representation as a central design factor in federated time-series pretraining and identify latent state dynamics as more transferable across heterogeneous clients than observation-level patterns.
- We propose FEDSTATE, a federated pretraining framework that keeps each client's observation model private and federates only an interval-conditioned latent transition operator, shared by clients with different channels, scales, sampling periods, and architectures.
- We introduce three mechanisms for sharing the operator across heterogeneous clients. Physical-time conditioning aligns temporal semantics across sampling grids, latent-state supervision decouples the shared parameters from observation-level information, and heterogeneity-aware aggregation reweights client updates by transition error.
- Extensive experiments across eight benchmarks demonstrate that FEDSTATE outperforms full-model federation, advanced federated pretraining methods, and centralized models in both deterministic and probabilistic forecasting, including zero-shot transfer.

## 2 RELATED WORK

**Time Series Model Pretraining.** Existing time-series pretraining methods can be broadly divided into direct forecasting and representation learning. General-purpose forecasters are pretrained on large and heterogeneous time-series collections to support transfer across datasets, domains, and forecasting horizons (Ansari et al., 2024; Woo et al., 2024; Das et al., 2024; Rasul et al., 2023; Shi et al., 2025; Wang et al., 2026). Representation-oriented methods instead learn reusable temporal features through masked reconstruction (Dong et al., 2023; He et al., 2026), contrastive objectives (Yue et al., 2022; Eldele et al., 2021; Woo et al., 2022), or latent prediction with slowly updated targets (Grill et al., 2020; Bardes et al., 2022; Assran et al., 2023). These objectives impose supervision at different levels, from observed values to latent representations, while targeting transferable temporal structure. Related work on latent world models further models state transitions directly in representation space (Hafner et al., 2019). Despite differences in prediction target and representation, existing pretraining methods largely assume centralized access to the pretraining data.

**Federated Learning for Heterogeneous Time Series.** Federated learning enables collaborative training without centralizing client data, but heterogeneous local distributions can substantially degrade local-update optimization (Glasgow et al., 2022). Existing approaches address this heterogeneity through proximal regularization (Li et al., 2020), drift correction (Karimireddy et al., 2020), server-side momentum (Hsu et al., 2019), or client-specific model components (Li et al., 2021b; Collins et al., 2021; Li et al., 2021a). Recent work extends these principles to time-series pretraining. FFTS (Chen et al., 2025) improves aggregation across heterogeneous clients, FeDaL (Chen et al., 2026a) reduces dataset-specific bias in the learned representation, and FedTRL (Chen et al., 2026b) promotes domain-invariant representations through prototype alignment and domain-aware aggregation. Despite their different mechanisms, these methods retain shared parameters tied to client observations or representations. FEDSTATE instead keeps observation mappings local and federates only the latent transition operator, isolating cross-client temporal dynamics from client-specific representations. Table 1 summarizes the resulting differences.

Table 1: Comparison of federated time series pretraining paradigms. ZS denotes zero-shot capability.

| | | | Heterogeneity | | | |
|---|---|---|---|---|---|---|
| **Paradigm** | **Shared** | **Target** | **Dist.** | **Arch.** | **Grid** | **ZS** |
| Centralized | – | Values | ✓ | – | varies | ✓ |
| Full-model FL | Full | Task | ✓ | × | × | × |
| Personalized FL | Backbone | Task | ✓ | × | × | × |
| FFTS | Full | Masked | ✓ | × | × | × |
| FeDaL | Full | Masked | ✓ | × | × | ✓ |
| FedTRL | Encoder | Invariant repr. | ✓ | × | × | ✓ |
| FEDSTATE (Ours) | Operator | Latent state | ✓ | ✓ | ✓ | ✓ |

## 3 METHODOLOGY

**Problem Formulation.** Consider $N$ clients. Client $i$ holds a private dataset $\mathcal{D}_i$ of $n_i$ pairs $(x, y)$ of a context $x \in \mathbb{R}^{L_i \times C_i}$ and its future $y \in \mathbb{R}^{H_i \times C_i}$, recorded with sampling period $\tau_i$. The sampling period, context length $L_i$, prediction horizon $H_i$, and channel dimension $C_i$ may all differ across clients. Each client's forecaster combines parameters $\phi$ shared by all clients with private parameters $\vartheta_i$. Raw observations and $\vartheta_i$ never leave the client, and every message exchanged during training is a function of $\phi$ alone. Federated pretraining seeks

$$\min_{\phi} \sum_{i=1}^{N} w_i R_i(\phi), \qquad R_i(\phi) = \min_{\vartheta_i} \mathcal{L}_i(\phi, \vartheta_i), \tag{1}$$

where $\mathcal{L}_i$ is the local pretraining objective of client $i$ and $w_i$ its aggregation weight. The shared parameters $\phi$ must therefore represent temporal structure that remains commensurable across heterogeneous observation spaces and sampling grids.

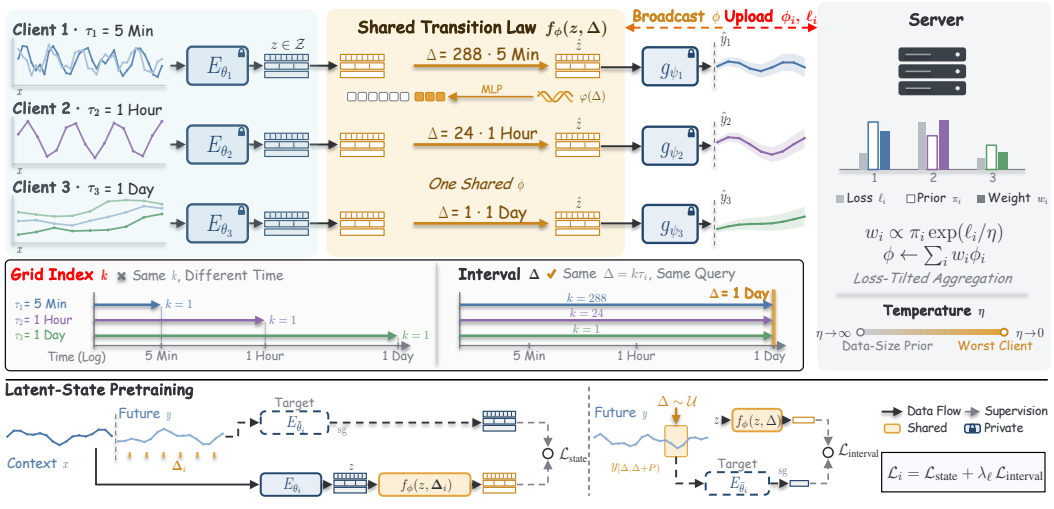



Figure 2: Overview of FEDSTATE. $\Delta$ is the elapsed physical time between two states, and the interval-conditioned operator $f_\phi(z, \Delta)$ is the only federated object.

**Overview.** Figure 2 illustrates our FEDSTATE. It decouples client-specific observation models from shared temporal dynamics. Clients keep their encoders and decoders private and federate only an interval-conditioned latent transition operator. The operator is trained from latent-state predictions over multiple physical intervals, while the server reweights client updates by their transition errors. This allows heterogeneous clients to learn a common dynamics model without sharing observation spaces or sampling grids. Concretely, the private parameters $\vartheta_i = (\theta_i, \psi_i)$ define an encoder and a decoder, and the shared parameters $\phi$ define a transition operator $f_\phi$, the *transition law*:

$$\underbrace{\mathbb{R}^{L_i \times C_i} \xrightarrow{E_{\theta_i}} \mathcal{Z}}_{\text{private to client } i} \xrightarrow[\text{federated}]{f_\phi(\cdot,\Delta)} \mathcal{Z} \underbrace{\xrightarrow[\text{fast scale}]{g_{\psi_i}} \mathbb{R}^{H_i \times C_i}}_{\text{private to client } i}, \tag{2}$$

where $\mathcal{Z} = \prod_{s \in \mathcal{S}} \mathcal{Z}_s$ collects latent states at the temporal scales $\mathcal{S} = \{\text{fast}, \text{mid}, \text{slow}\}$. The encoder reads one fast state per patch of the window, one mid state per block of consecutive patches, and a single slow state for the whole window, so the three scales track short transients, intermediate phases, and slow drifts. The client index appears only at the two ends of Eq. (2). The map in the middle is one function for every client, and it is the only object communicated during training.

**Interval-Conditioned Latent Dynamics.** The key to sharing latent dynamics across clients is to give the transition operator a client-independent notion of time. We therefore parameterize the transition law by the elapsed physical time $\Delta > 0$ rather than by sequence position. For client $i$, an offset of $k$ observations is the interval $\Delta = k\tau_i$, so clients with different sampling periods query the same operator at the physical intervals their grids represent. We encode $\Delta$ as

$$e(\Delta) = [\sin(\omega_m t), \cos(\omega_m t)]_{m=1}^{M}, \qquad t = \log(1 + \Delta), \tag{3}$$

where $\{\omega_m\}_{m=1}^{M}$ are logarithmically spaced frequencies. The transition operator is implemented as a causal Transformer over past latent states, with the query token at interval $\Delta$ conditioned by an MLP embedding of $e(\Delta)$. A grid-indexed operator instead ties its $k$-th slot to an interval that changes with the client's sampling period. To formalize the distinction, let $\zeta_i(\Delta) \in \mathcal{Z}$ denote the latent state of client $i$'s series an interval $\Delta$ after the context with state $z$, and let $\Phi_i(z, \Delta) = \mathbb{E}[\zeta_i(\Delta) \mid z]$ be its *transport kernel*, the Bayes-optimal prediction of that state. Appendix A proves the following.

**Theorem 1** (Excess Risk of Grid Indexing)**.** *Under finite second moments, the best shared interval-indexed operator exceeds the clients' Bayes risks by the dispersion of their kernels at* equal *intervals $\Delta$, while a grid-indexed operator, whose $k$-th slot serves $k\tau_i$ on client $i$, incurs their dispersion at the* different *intervals $\{\Phi_i(\cdot, k\tau_i)\}_i$. If every client has the same kernel $\Phi$, the first excess is zero and the second strictly positive as soon as two clients with different periods have $\Phi(\cdot, k\tau_i) \neq \Phi(\cdot, k\tau_j)$ on a set of positive measure where both their states occur.*

**Latent-State Pretraining Objective.** To train the shared transition operator while keeping supervision local to each client, we define the following pretraining objective:

$$\mathcal{L}_i = \mathcal{L}_{\text{state}} + \lambda_\ell \mathcal{L}_{\text{interval}} + \mathcal{R}_i, \tag{4}$$

where $\lambda_\ell > 0$ weights the interval term and $\mathcal{R}_i$ contains an observation-space grounding term for the decoder and a variance regularizer over latent states (Appendix B.2). Given a context $x$, the online encoder produces multi-scale states $z = (z^s)_{s \in \mathcal{S}} = E_{\theta_i}(x)$, while an exponential-moving-average target encoder $E_{\bar{\theta}_i}$ provides targets under the stop-gradient operator $\text{sg}[\cdot]$. A superscript $s$ selects the scale-$s$ component of a state. Both terms compare states with the element-wise Huber loss (Huber, 1964), at threshold $\beta_s$ for scale $s$ and $\beta_\ell$ for the interval term: for $u, v \in \mathbb{R}^d$,

$$\mathcal{H}_\beta(u, v) = \frac{1}{d} \sum_{j=1}^{d} h_\beta(u_j - v_j), \qquad h_\beta(r) = \begin{cases} r^2/(2\beta), & |r| < \beta, \\ |r| - \beta/2, & \text{otherwise.} \end{cases} \tag{5}$$

We first supervise predictions at the future window's native physical intervals $\mathbf{\Delta}_i = (\Delta_{i,1}, \ldots, \Delta_{i,K_i})$, the $K_i$ intervals at which the target encoder reads the states of $y$, and $f_\phi(z, \mathbf{\Delta}_i)$ stacks the operator's output at each of them:

$$\mathcal{L}_{\text{state}} = \sum_{s \in \mathcal{S}} \mathcal{H}_{\beta_s}\left(f_\phi(z, \mathbf{\Delta}_i)^s,\, \text{sg}\left[E_{\bar{\theta}_i}(y)^s\right]\right). \tag{6}$$

To supervise transitions beyond the native grid, we additionally sample an elapsed interval $\Delta$ uniformly from the future window, $\Delta \sim \mathcal{U}$, and predict the fast-scale state of the patch of length $P$ that begins at that interval:

$$\mathcal{L}_{\text{interval}} = \mathbb{E}_{\Delta\sim\mathcal{U}}\, \mathcal{H}_{\beta_\ell}\Big(f_\phi(z,\Delta)^{\text{fast}},\, \text{sg}\Big[E_{\bar{\theta}_i}\big(y_{[\Delta,\Delta+P)}\big)^{\text{fast}}\Big]\Big). \tag{7}$$

**Heterogeneity-Aware Federated Optimization.** In each communication round, client $i$ receives the transition parameters $\phi$, performs local updates on $\mathcal{L}_i$, and returns the updated parameters with its latent prediction loss $\ell_i$. The server adapts the aggregation weights to these transition errors. Let $\pi_i \propto n_i$ denote the standard data-size prior. Data-size weighting optimizes the average client objective but does not account for clients with persistently large transition errors. We therefore define the aggregation weights over the probability simplex $\mathcal{P}_N$ through

$$\max_{w\in\mathcal{P}_N} \sum_i w_i\ell_i - \eta\,\text{KL}(w\,\|\,\pi), \qquad \text{solved by} \qquad w_i = \frac{\pi_i \exp(\ell_i/\eta)}{\sum_j \pi_j \exp(\ell_j/\eta)}. \tag{8}$$

The temperature $\eta > 0$ controls the degree of reweighting. As $\eta \to \infty$ the update recovers the data-size prior, whereas smaller $\eta$ places greater weight on clients with larger transition error. Proposition 1 states the robust objective this solves. The implementation of FEDSTATE in Algorithm 1.

# 4 EXPERIMENTS

We evaluate FEDSTATE from four perspectives: in-domain representation learning, in-federation forecasting, out-of-federation transfer of the pretrained transition operator, and framework analysis (Sections 4.1–4.4). Implementation details are provided in Appendix B.2.

## 4.1 IN-DOMAIN REPRESENTATION LEARNING

**Setup.** We evaluate in-domain representation learning on TSLib (Wu et al., 2023) and BuildingsBench (Emami et al., 2023), where pretraining and downstream fine-tuning use the same client set. On TSLib, each of the eight datasets is treated as one client with a look-back of 512, and performance is measured by MSE and MAE averaged over horizons $\{96, 192, 336, 720\}$. On BuildingsBench, the ten institutions covering 1,909 buildings form the clients, with a $168\to24$ forecasting window and CV-RMSE reported as the median across buildings in each collection. We compare FEDSTATE with PatchTST (Nie et al., 2023), FFTS (Chen et al., 2025), SimMTM (Dong et al., 2023), TimeDART (Wang et al., 2025), CoST (Woo et al., 2022), TimeMAE (Cheng et al., 2026), FeDaL (Chen et al., 2026a), centralized pooled training (ALL MIXED), and training from scratch. All methods federate for 60 rounds under the same settings, with details Appendix C.

Table 2: Results of in-domain representation learning on TSLib and BuildingBench. **Bold** and Underline: the best and second best. Full results are in Table 30.

| | | Ours | | Federated Self-Supervised | | | | | | | | | | Fed. Agg. | | Centralized | | No FL | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | FEDSTATE | | TimeDART | | SimMTM | | PatchTST | | FFTS | | CoST | | TimeMAE | | FeDaL | | All Mixed | | Scratch | | No Pretrain | |
| | Dataset | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| TSLib | ETTh1 | **0.422** | **0.435** | 0.506 | 0.478 | 0.472 | 0.468 | 0.503 | 0.476 | 0.487 | 0.473 | 0.498 | 0.478 | 0.528 | 0.493 | 0.470 | 0.465 | 0.496 | 0.470 | 0.458 | 0.459 | 0.484 | 0.470 |
| | ETTh2 | **0.339** | **0.386** | 0.413 | 0.427 | 0.444 | 0.452 | 0.429 | 0.434 | 0.427 | 0.438 | 0.426 | 0.439 | 0.429 | 0.441 | 0.386 | 0.416 | 0.401 | 0.423 | 0.375 | 0.411 | 0.406 | 0.428 |
| | ETTm1 | 0.360 | **0.383** | 0.356 | 0.388 | 0.361 | 0.393 | 0.355 | 0.390 | 0.358 | 0.391 | 0.358 | 0.390 | 0.357 | 0.392 | **0.352** | 0.389 | 0.358 | 0.389 | 0.361 | 0.386 | 0.357 | 0.390 |
| | ETTm2 | **0.254** | **0.313** | 0.277 | 0.338 | 0.284 | 0.340 | 0.279 | 0.335 | 0.289 | 0.342 | 0.275 | 0.333 | 0.293 | 0.346 | 0.274 | 0.334 | 0.277 | 0.336 | 0.260 | 0.320 | 0.282 | 0.337 |
| | ECL | **0.165** | **0.258** | 0.177 | 0.276 | 0.178 | 0.277 | 0.177 | 0.277 | 0.178 | 0.278 | 0.178 | 0.276 | 0.178 | 0.278 | 0.193 | 0.297 | 0.177 | 0.276 | 0.170 | 0.265 | 0.178 | 0.277 |
| | Traffic | **0.406** | **0.279** | 0.458 | 0.341 | 0.454 | 0.335 | 0.462 | 0.343 | 0.462 | 0.342 | 0.454 | 0.337 | 0.453 | 0.335 | 0.486 | 0.361 | 0.457 | 0.340 | 0.414 | 0.292 | 0.455 | 0.338 |
| | Weather | 0.235 | 0.271 | 0.233 | 0.269 | 0.232 | 0.269 | 0.233 | 0.269 | 0.235 | 0.272 | 0.232 | 0.270 | 0.234 | 0.271 | 0.231 | 0.269 | **0.230** | **0.268** | 0.233 | 0.270 | 0.232 | 0.270 |
| | Exchange | **0.389** | **0.416** | 0.501 | 0.472 | 0.519 | 0.483 | 0.506 | 0.467 | 0.504 | 0.477 | 0.517 | 0.479 | 0.496 | 0.482 | 0.451 | 0.450 | 0.480 | 0.463 | 0.390 | 0.421 | 0.499 | 0.470 |
| | Collection | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | | CV-RMSE | |
| BuildingBench | BDG-2 | 15.92 | | 17.73 | | 18.03 | | 17.79 | | 17.83 | | 17.90 | | 18.03 | | 19.56 | | 19.15 | | **15.78** | | 18.07 | |
| | Electricity | **11.20** | | 12.58 | | 13.83 | | 13.39 | | 13.57 | | 12.01 | | 13.44 | | 15.52 | | 19.48 | | 11.55 | | 12.74 | |
| | LCL | **74.43** | | 77.18 | | 77.44 | | 77.43 | | 77.04 | | 76.67 | | 77.44 | | 77.97 | | 81.92 | | 75.06 | | 77.08 | |
| | IDEAL | **85.52** | | 87.71 | | 88.20 | | 87.65 | | 87.89 | | 87.74 | | 87.38 | | 89.35 | | 92.58 | | 86.06 | | 86.99 | |
| | Borealis | 58.94 | | **58.03** | | 59.45 | | 60.88 | | 59.04 | | 59.52 | | 58.45 | | 59.01 | | 63.11 | | 62.28 | | 58.08 | |
| | SMART | **72.92** | | 83.35 | | 85.42 | | 82.34 | | 82.23 | | 81.48 | | 87.65 | | 82.09 | | 77.16 | | 80.41 | | 81.55 | |
| | Sceaux | 73.93 | | 75.74 | | 78.53 | | 77.20 | | 77.01 | | 71.55 | | 77.11 | | 70.65 | | **69.56** | | 82.60 | | 73.01 | |
| | 1st | **17** | | 1 | | 0 | | 0 | | 0 | | 0 | | 0 | | 1 | | 3 | | 1 | | 0 | |

**Results.** Table 2 shows that FEDSTATE achieves the strongest overall in-domain performance across TSLib and BuildingsBench, leading most evaluation settings with particularly clear gains

on ETTh1, ETTh2, and several building collections. Under the same evaluation setting, federated self-supervised baselines that exchange the full encoder remain close to centralized pooled training and non-federated pretraining. These results demonstrate that the main advantage comes from the shared latent dynamics rather than from changing the pretraining objective alone.

### 4.2 IN-FEDERATION FORECASTING

**Setup.** We evaluate federated pretraining on three benchmarks. **(1) TSLib** (Wu et al., 2023) treats seven datasets (ETT-series, ECL, Weather, and Exchange) as clients with a look-back of 96. **(2) TIME** (Qiao et al., 2026) contains 46 clients spanning six orders of magnitude in sampling period and 91 evaluation configurations. **(3) GIFT-Eval** (Aksu et al., 2024) contains 52 clients from 97 configurations. TIME and GIFT-Eval use look-backs of up to 512. Because full-model federation requires a common context length, its baselines use the 34 TIME and 38 GIFT-Eval clients with context length 512, whereas FEDSTATE federates all clients. We report MSE and MAE on TSLib, seven point metrics on TIME, and $\text{MASE}_n$ on GIFT-Eval for deterministic forecasting, together with CRPS, $\text{CRPS}_n$, and $\text{WQL}_n$ for probabilistic forecasting, where $n$ denotes normalization by the seasonal naive forecast. Baselines cover full-model FL (FedAvg (McMahan et al., 2017), FedAvgM (Hsu et al., 2019), FedProx (Li et al., 2020), SCAFFOLD (Karimireddy et al., 2020)), personalized FL (FedBN (Li et al., 2021b), FedRep (Collins et al., 2021)), federated time-series pretraining on TSLib (FFTS (Chen et al., 2025), FeDaL (Chen et al., 2026a)), centralized and pooled-data training, and training from scratch. Details in Appendix B.3.

Table 3: Results of in-federation forecasting on TSLib. Full results in Table 14.

| | | | Ours | Federated | | | | | | Own arch. | | Centralized | | | | | No FL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dataset | Metric | FEDSTATE | FedAvg | FedAvgM | FedProx | FedBN | FedRep | SCAFFOLD | FFTS | FeDaL | DLinear | PatchTST | iTrans. | TimesNet | Pooled | Scratch |
| TSLib | ETTh1 | MSE | **0.424** | 0.464 | 0.450 | 0.465 | 0.461 | 0.470 | 0.472 | 0.446 | 0.464 | 0.454 | 0.453 | 0.448 | 0.502 | 0.455 | 0.473 |
| | | MAE | **0.431** | 0.452 | 0.436 | 0.451 | 0.451 | 0.458 | 0.463 | 0.449 | 0.451 | 0.448 | 0.437 | 0.442 | 0.481 | 0.449 | 0.462 |
| | ETTh2 | MSE | **0.353** | 0.374 | 0.375 | 0.386 | 0.376 | 0.367 | 0.384 | 0.380 | 0.384 | 0.480 | 0.379 | 0.383 | 0.420 | 0.423 | 0.385 |
| | | MAE | **0.395** | 0.402 | 0.398 | 0.410 | 0.406 | 0.399 | 0.409 | 0.405 | 0.407 | 0.467 | 0.402 | 0.406 | 0.429 | 0.426 | 0.409 |
| | ETTm1 | MSE | **0.367** | 0.386 | 0.413 | 0.395 | 0.387 | 0.387 | 0.391 | 0.392 | 0.385 | 0.405 | 0.383 | 0.409 | 0.407 | 0.397 | 0.399 |
| | | MAE | **0.389** | 0.398 | 0.406 | 0.403 | 0.401 | 0.401 | 0.400 | 0.404 | 0.397 | 0.410 | 0.397 | 0.410 | 0.412 | 0.403 | 0.404 |
| | ETTm2 | MSE | **0.273** | 0.285 | 0.285 | 0.289 | 0.286 | 0.287 | 0.284 | 0.281 | 0.292 | 0.349 | 0.284 | 0.296 | 0.297 | 0.330 | 0.284 |
| | | MAE | **0.326** | 0.330 | 0.327 | 0.333 | 0.330 | 0.331 | 0.330 | **0.326** | 0.338 | 0.395 | 0.331 | 0.336 | 0.333 | 0.357 | 0.329 |
| | ECL | MSE | 0.187 | 0.180 | 0.201 | 0.182 | 0.180 | 0.180 | 0.180 | 0.215 | 0.215 | 0.198 | 0.181 | 0.180 | 0.215 | **0.178** | **0.178** |
| | | MAE | 0.271 | 0.266 | 0.280 | 0.267 | 0.265 | 0.265 | 0.265 | 0.317 | 0.298 | 0.285 | 0.269 | 0.269 | 0.314 | **0.263** | **0.263** |
| | Weather | MSE | **0.235** | 0.255 | 0.273 | 0.257 | 0.254 | 0.255 | 0.261 | 0.256 | 0.263 | 0.268 | 0.258 | 0.262 | 0.255 | 0.264 | 0.261 |
| | | MAE | **0.272** | 0.275 | 0.290 | 0.277 | 0.275 | 0.275 | 0.280 | 0.279 | 0.283 | 0.322 | 0.281 | 0.283 | 0.283 | 0.283 | 0.280 |
| | Exchange | MSE | 0.364 | 0.398 | 0.355 | 0.367 | 0.400 | 0.362 | 0.368 | 0.365 | 0.333 | **0.318** | 0.365 | 0.355 | 0.412 | 0.488 | 0.363 |
| | | MAE | 0.404 | 0.417 | 0.398 | 0.406 | 0.418 | 0.403 | 0.407 | 0.406 | 0.369 | **0.389** | 0.406 | 0.402 | 0.440 | 0.460 | 0.406 |
| | 1st | | **10** | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 2 |

**Deterministic Forecasting.** Table 3 shows that FEDSTATE achieves the best on TSLib, reducing MSE by 2.4% over the strongest federated baseline and 2.1% over PatchTST, at 0.36M uploaded parameters per round against 1.13M for full-model federation. The same architecture trained on pooled data reaches 0.369 MSE and personalized variants sharing observation-level parameters remain behind, so the improvement follows from the federated object, not from data access. On TIME, where sampling periods span six orders of magnitude, FEDSTATE improves six of the seven point metrics over FL baseline, by up to 3.1% on $\text{MAPE}_n$ (Table 5). Under the same setting, FFTS reaches a median $\text{MASE}_n$ of 1.083 against 0.821 for the law while uploading 6.48M parameters per round.

**Probabilistic Forecasting.** Table 4 shows that FEDSTATE achieves the best overall probabilistic performance on TSLib. The FL baselines over-disperse, covering 95 to 96% of the targets at a nominal 80%, while FEDSTATE reaches mean CRPS 0.346 against 0.493 for the closest of them at a coverage of 0.82. This advantage generalizes to TIME (Table 5) and GIFT-Eval (Table 6), where FEDSTATE consistently improves probabilistic performance over the federated baselines.

### 4.3 FOUNDATION MODEL PRETRAINING

**Setup.** We evaluate transfer beyond the pretraining federation under full-/few-shot and zero-shot settings. **(1) Full- and few-shot.** Following FFTS (Chen et al., 2025), all methods are pretrained on the nine-client Monash corpus for 60 rounds and fine-tuned on ETT-series and Weather using 100%, 10%, or 5% of the training data with a look-back of 512. Baselines include FFTS, the federated methods from Section 4.2, and training from scratch. **(2) Zero-shot.** All methods are pretrained once on a unified corpus of 94 clients, 8,160 series, and 34.0M observations from TIME,

Table 4: Probabilistic forecasting on TSLib, with Cov80 ranked by its distance to the nominal 0.80.

| | | Ours | Federated | | | | | | Centralized | No FL |
|---|---|---|---|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **FEDSTATE** | **FedAvg** | **FedAvgM** | **FedProx** | **FedBN** | **FedRep** | **SCAFFOLD** | **Pooled** | **Scratch** |
| ETTh1 | CRPS | **0.375** | 0.651 | 0.628 | 0.652 | 0.651 | 0.650 | 0.641 | 0.384 | 0.642 |
| | Cov80 | 0.910 | 0.988 | 0.986 | 0.988 | 0.989 | 0.988 | 0.987 | **0.909** | 0.987 |
| ETTh2 | CRPS | **0.342** | 0.643 | 0.616 | 0.642 | 0.648 | 0.647 | 0.633 | 0.363 | 0.632 |
| | Cov80 | **0.781** | 0.997 | 0.995 | 0.996 | 0.997 | 0.997 | 0.996 | 0.777 | 0.996 |
| ETTm1 | CRPS | **0.333** | 0.533 | 0.499 | 0.503 | 0.511 | 0.520 | 0.493 | 0.338 | 0.494 |
| | Cov80 | **0.869** | 0.981 | 0.969 | 0.974 | 0.976 | 0.979 | 0.972 | 0.878 | 0.971 |
| ETTm2 | CRPS | **0.279** | 0.525 | 0.431 | 0.471 | 0.525 | 0.499 | 0.479 | 0.304 | 0.444 |
| | Cov80 | **0.790** | 0.991 | 0.974 | 0.979 | 0.991 | 0.985 | 0.982 | 0.789 | 0.977 |
| ECL | CRPS | 0.227 | 0.223 | 0.236 | 0.224 | 0.223 | 0.223 | 0.222 | **0.221** | **0.221** |
| | Cov80 | 0.777 | **0.783** | 0.778 | 0.780 | 0.782 | **0.783** | 0.781 | 0.775 | 0.782 |
| Weather | CRPS | **0.246** | 0.406 | 0.406 | 0.426 | 0.407 | 0.417 | 0.415 | 0.250 | 0.399 |
| | Cov80 | **0.853** | 0.982 | 0.979 | 0.984 | 0.982 | 0.984 | 0.981 | 0.856 | 0.980 |
| Exchange | CRPS | 0.760 | 0.659 | **0.637** | 0.654 | 0.659 | 0.656 | 0.657 | 0.694 | 0.658 |
| | Cov80 | **0.791** | 0.998 | 0.999 | 0.999 | 0.997 | 0.999 | 0.999 | 0.768 | 0.999 |
| 1st | | **10** | 1 | 1 | 0 | 0 | 1 | 0 | 2 | 1 |

Table 5: Results on TIME, on the 87 configurations every listed method completed. Comm./round is the number of parameters each client uploads per round. **Bold** and Underline: the best and second best. Full results are in Table 15.

| **Model** | **Comm./round** | **MSE**$_n$ | **MAE**$_n$ | **RMSE**$_n$ | **MAPE**$_n$ | **sMAPE**$_n$ | **MASE**$_n$ | **ND**$_n$ | **CRPS**$_n$ |
|---|---|---|---|---|---|---|---|---|---|
| Scratch | — | 0.678 | 0.858 | 0.845 | 0.811 | 0.898 | 0.867 | 0.837 | 0.795 |
| FedAvgM | 1.88M | 0.624 | 0.816 | 0.794 | 0.766 | 0.862 | **0.820** | 0.793 | 0.751 |
| FedBN | 1.87M | 0.637 | 0.821 | 0.810 | 0.765 | 0.880 | 0.832 | 0.799 | 0.750 |
| FedProx | 1.88M | 0.634 | 0.829 | 0.805 | 0.772 | 0.878 | 0.839 | 0.802 | 0.757 |
| FedRep | 0.99M | 0.622 | 0.820 | 0.805 | 0.770 | 0.879 | 0.836 | 0.789 | 0.757 |
| FFTS | 6.48M | 0.905 | 0.987 | 0.953 | 0.880 | 0.996 | 1.083 | 0.932 | 1.989 |
| FeDaL‡ | 0.61M | 0.049 | 0.376 | 0.351 | 0.554 | 1.026 | 1.028 | 0.888 | 2.021 |
| **FEDSTATE** | **0.36M** | **0.610** | **0.813** | **0.788** | **0.741** | **0.850** | 0.821 | **0.775** | **0.735** |

Table 6: Results on GIFT-Eval, marked as in Table 5. † FFTS and FeDaL cover 80 of the 97 configurations. Full results in Table 16.

| **Model** | **Comm./round** | **MASE**$_n$ | **WQL**$_n$ |
|---|---|---|---|
| FedAvgM | 1.92M | **0.915** | 0.753 |
| FedBN | 1.91M | 0.928 | 0.728 |
| FedProx | 1.92M | 0.956 | 0.757 |
| FedRep | 0.99M | 0.955 | 0.755 |
| FFTS† | 6.48M | 0.914 | 0.766 |
| FeDaL† | 0.61M | 1.044 | 0.855 |
| **FEDSTATE** | **0.37M** | 0.919 | **0.695** |

GIFT-Eval, and Monash-18, where each client is a dataset–sampling-frequency pair. We remove all clients overlapping the downstream evaluation sets. Zero-shot evaluation covers TSLib under the paired protocol of Shi et al. (2025), WEATHER-5K (Han et al., 2026), 15 unseen weather stations, LiveHouse-TS (Wen et al., 2026), and BuildingsBench (Emami et al., 2023). Details in Appendix B.

**Full- and Few-Shot Forecasting.** Table 7 shows that FEDSTATE achieves the lowest average MAE in all three data regimes and the lowest average MSE with the full and the 10% training split, reducing MSE by 1.2% and 4.7% relative to the strongest federated baseline, and wins 20 of the 30 dataset-level cells against 4 for FFTS. The benefit of the law grows as local data shrink. Relative to training from scratch, MSE falls by 3% with the full split and by 21% with 5% of it, since the law carries dynamics learned across the nine Monash clients and a client with few samples only needs to fit its private observation mapping. These gains come at 0.36M uploaded parameters per round, against 1.73M for full-model federation and 1.44M for FFTS.

Table 7: Full- and few-shot transfer results. **Bold** and Underline: the best and second best.

| | | Ours | | Federated | | | | | | | | | | Own arch. | | | | No FL | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Comm./round* | | **FEDSTATE** 0.36M | | **FedAvg** 2.92M | | **FedAvgM** 2.92M | | **FedProx** 2.92M | | **FedBN** 2.92M | | **FedRep** 2.92M | | **FFTS** 5.63M | | **FeDaL** 0.61M | | **Scratch** — | |
| **Split** | **Dataset** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| Full | ETTh1 | **0.406** | **0.423** | 0.475 | 0.459 | 0.429 | 0.438 | 0.432 | 0.446 | 0.493 | 0.466 | 0.459 | 0.462 | 0.411 | 0.438 | 0.455 | 0.452 | 0.430 | 0.444 |
| | ETTh2 | **0.343** | **0.389** | 0.449 | 0.441 | 0.359 | 0.400 | 0.388 | 0.417 | 0.420 | 0.432 | 0.390 | 0.414 | 0.361 | 0.401 | 0.378 | 0.410 | 0.376 | 0.412 |
| | ETTm1 | 0.361 | **0.380** | 0.370 | 0.391 | 0.363 | 0.388 | 0.364 | 0.387 | 0.364 | 0.389 | 0.386 | 0.400 | **0.360** | 0.392 | 0.356 | 0.390 | **0.360** | 0.385 |
| | ETTm2 | **0.254** | **0.313** | 0.272 | 0.328 | 0.265 | 0.325 | 0.271 | 0.324 | 0.272 | 0.326 | 0.267 | 0.323 | 0.268 | 0.327 | 0.270 | 0.331 | 0.260 | 0.318 |
| | Weather | 0.241 | 0.278 | 0.249 | 0.280 | 0.233 | 0.271 | 0.241 | 0.276 | 0.251 | 0.282 | 0.248 | 0.279 | **0.226** | **0.267** | 0.232 | 0.270 | 0.232 | 0.268 |
| | Avg | **0.321** | **0.357** | 0.363 | 0.380 | 0.330 | 0.365 | 0.339 | 0.370 | 0.360 | 0.379 | 0.350 | 0.376 | 0.325 | 0.365 | 0.338 | 0.371 | 0.332 | 0.365 |
| 10% | ETTh1 | 0.449 | 0.460 | 0.549 | 0.524 | **0.429** | **0.448** | 0.928 | 0.664 | 0.569 | 0.535 | 0.470 | 0.477 | 0.569 | 0.513 | 1.117 | 0.770 | 0.672 | 0.571 |
| | ETTh2 | **0.366** | **0.404** | 0.402 | 0.430 | 0.368 | **0.404** | 0.415 | 0.438 | 0.413 | 0.437 | 0.378 | 0.410 | 0.389 | 0.422 | 0.615 | 0.527 | 0.439 | 0.451 |
| | ETTm1 | **0.405** | **0.413** | 0.889 | 0.612 | 0.477 | 0.451 | 0.478 | 0.456 | 0.915 | 0.627 | 0.642 | 0.538 | 0.646 | 0.495 | 0.788 | 0.579 | 0.446 | 0.444 |
| | ETTm2 | **0.274** | **0.324** | 0.354 | 0.375 | 0.288 | 0.338 | 0.341 | 0.371 | 0.349 | 0.372 | 0.329 | 0.364 | 0.294 | 0.342 | 0.374 | 0.381 | 0.292 | 0.340 |
| | Weather | 0.240 | **0.275** | 0.316 | 0.319 | 0.257 | 0.292 | 0.273 | 0.298 | 0.315 | 0.318 | 0.285 | 0.309 | **0.239** | 0.280 | 0.279 | 0.302 | 0.249 | 0.286 |
| | Avg | **0.347** | **0.375** | 0.502 | 0.452 | 0.364 | 0.387 | 0.487 | 0.445 | 0.512 | 0.458 | 0.421 | 0.420 | 0.427 | 0.410 | 0.634 | 0.512 | 0.420 | 0.418 |
| 5% | ETTh1 | 0.452 | 0.447 | 0.545 | 0.509 | **0.419** | **0.436** | 0.545 | 0.514 | 0.653 | 0.563 | 0.456 | 0.452 | 0.653 | 0.541 | 1.531 | 0.881 | 0.772 | 0.594 |
| | ETTh2 | **0.340** | **0.382** | 0.366 | 0.397 | 0.345 | 0.384 | 0.377 | 0.408 | 0.368 | 0.401 | 0.341 | **0.382** | 0.360 | 0.403 | 0.432 | 0.444 | 0.463 | 0.461 |
| | ETTm1 | 0.460 | 0.446 | 0.568 | 0.508 | **0.447** | **0.441** | 0.485 | 0.459 | 0.572 | 0.507 | 0.540 | 0.498 | 0.508 | 0.474 | 0.992 | 0.639 | 0.470 | 0.456 |
| | ETTm2 | **0.293** | **0.337** | 0.333 | 0.367 | 0.297 | 0.340 | 0.355 | 0.380 | 0.355 | 0.380 | 0.316 | 0.359 | 0.300 | 0.342 | 0.412 | 0.418 | 0.311 | 0.353 |
| | Weather | **0.265** | **0.300** | 0.318 | 0.327 | 0.280 | 0.320 | 0.303 | 0.322 | 0.319 | 0.329 | 0.288 | 0.321 | 0.270 | 0.308 | 0.324 | 0.329 | 0.284 | 0.318 |
| | Avg | 0.362 | **0.383** | 0.426 | 0.422 | **0.357** | 0.384 | 0.413 | 0.416 | 0.453 | 0.436 | 0.388 | 0.402 | 0.418 | 0.414 | 0.738 | 0.542 | 0.460 | 0.436 |
| 1st | | **20** | | 0 | | 7 | | 0 | | 0 | | 1 | | 4 | | 0 | | 1 | |

**Zero-Shot Forecasting.** Table 8 shows that FEDSTATE transfers consistently across the six zero-shot benchmarks without target-domain adaptation. It outperforms all full-model federated baselines in every evaluation setting and remains competitive with architecture-specific federated pretraining, with particularly clear gains on unseen weather stations and LiveHouse-TS. BuildingsBench

is the main exception, where architecture-specific methods and the seasonal-naive baseline remain stronger, indicating a limitation under this domain shift toward building-level energy consumption.

Table 8: Results of zero-shot forecasting on six benchmarks, one pretraining on the unified corpus per column and no adaptation at any target. Full results are in Tables 18 to 22.

| | | Ours | Federated | | | | | Own arch. | | Centralized | Naive |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Benchmark** | **Metric** | **FEDSTATE** | **FedAvg** | **FedAvgM** | **FedProx** | **FedBN** | **FedRep** | **FFTS** | **FeDaL** | **Pooled** | **S-naive** |
| TSLib | MSE | **0.407** | 0.413 | 0.469 | 0.428 | 0.408 | 0.511 | 0.467 | 0.540 | 0.757 | — |
| | MAE | **0.418** | 0.436 | 0.469 | 0.444 | 0.429 | 0.492 | 0.463 | 0.476 | 0.572 | — |
| WEATHER-5K$_{512}$ | MAE | **17.109** | 20.815 | 20.854 | 20.654 | 20.793 | 20.806 | 17.213 | 17.251 | 23.805 | 20.594 |
| WEATHER-5K$_{48}$ | MAE | 18.027 | 20.066 | 20.016 | 19.951 | 20.048 | 19.984 | **17.864** | 17.934 | 20.904 | 20.594 |
| Unseen stations | MASE | **1.078** | 1.481 | 1.477 | 1.470 | 1.486 | 1.482 | 1.190 | 1.270 | 1.764 | 1.286 |
| | MAE | **0.540** | 0.695 | 0.693 | 0.692 | 0.698 | 0.695 | 0.570 | 0.602 | 0.803 | — |
| LiveHouse-TS | MASE | **3.179** | 3.937 | 3.982 | 4.121 | 3.939 | 4.146 | 4.774 | 3.844 | 3.600 | — |
| BuildingsBench | CVRMSE | 51.4 | 53.6 | 55.8 | 52.2 | 56.3 | 57.7 | 39.4 | 32.6 | 61.4 | **29.4** |
| 1st | | **6** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

### 4.4 FRAMEWORK ANALYSES

**Choice of the Federated Component.** Figure 3(a) isolates the effect of the shared component on TIME. Sharing only the transition law matches larger shared components while reducing the per-client upload from 1.86M to 0.35M parameters per round. As shown in Figure 3(b), under the matched-budget FFTS transfer protocol, latent-state prediction reduces MSE by 14.7% and yields substantially lower worst-client error than its observation-space counterpart.

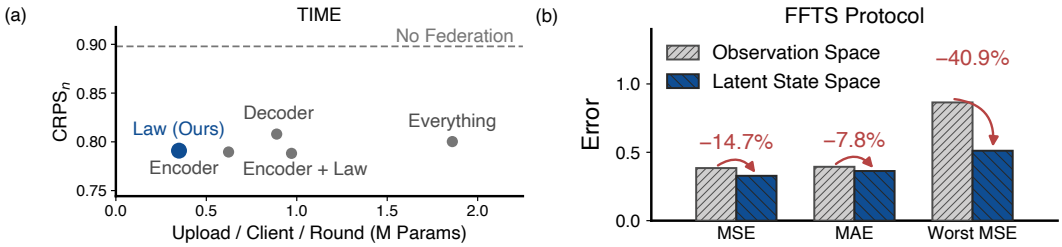



Figure 3: **(a)** Accuracy against upload on TIME for each aggregated component, and **(b)** the latent state space against an observation space of equal size under the FFTS-protocol transfer.

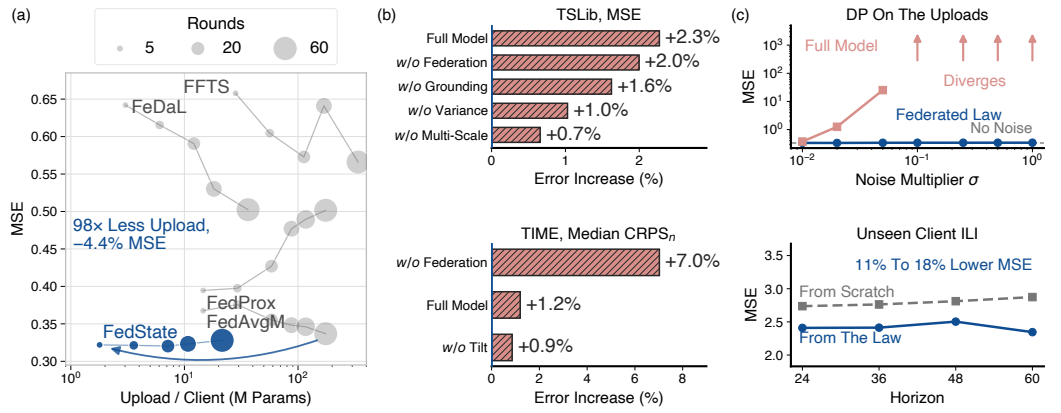



Figure 4: **(a)** MSE against the parameters a client has uploaded on the 10% FFTS settings; bubble area grows with the round. **(b)** The error each removed component costs. **(c)** The law under central differential privacy on the uploads (top) and on the unseen client ILI (bottom).

**Communication Budget.** Figure 4(a) compares forecasting error against cumulative client upload. FEDSTATE reaches 0.322 MSE after five rounds with only 1.8M uploaded parameters, outperforming FedAvgM after 60 rounds (0.337 MSE with 175M parameters). More generally, FEDSTATE maintains a lower error profile than full-model federation across communication budgets, while reducing per-round traffic from 11.7MB to 1.43MB per client.

**Ablation Studies.** Figure 4(b) removes one component at a time and reads the loss where that component acts. The shared law lowers the mean MSE by 2.0% against the identical backbone trained from scratch. Removing the variance regularizer gives back 1.0% of that and removing the observation-space grounding term 1.6%, so both terms of $\mathcal{R}_i$ contribute to the shared representation. The worst-client tilt moves the mean MSE by 0.002, so it buys no accuracy on seven clients that a single federation already fits, and its effect appears where it is designed to act. On the 46 TIME clients, removing it raises the median $\text{CRPS}_n$ of one federation by 0.9%.

**Robustness and Privacy.** Figure 4(c) tests the law under partial participation and noise. With half of the clients sampled per round it still improves on training from scratch, 0.334 against 0.337 mean MSE and 0.330 at full participation, and an unseen client, ILI, falls from 2.796 to 2.418. Under central differential privacy on the uploads it stays within 0.008 MSE of its noiseless value up to $\sigma$=1 ($\varepsilon$=67), where full-model federation of the same backbone has already diverged. Gradient inversion and membership inference attacks on the uploads are reported in Table 32.

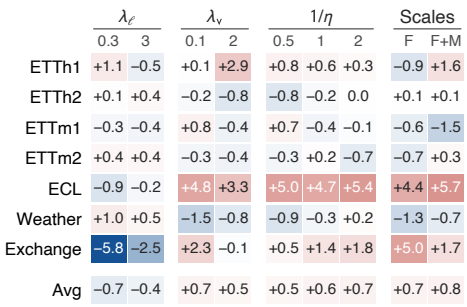



Figure 5: Change in mean MSE on TSLib when one hyperparameter departs from its default. Pink is an increase, blue a decrease.

**Hyperparameter Sensitivity.** Figure 5 varies each of the four hyperparameters around its default under the identical protocol, the interval weight $\lambda_\ell$, the variance weight $\lambda_v$, the tilt strength $1/\eta$, and the set of latent scales. No setting moves the mean MSE by more than 0.8%, the largest departure being the removal of the mid and slow scales, and two thirds of the per-dataset changes remain within 1%. The three scales and the default weights are therefore a plateau rather than a tuned point.

Table 9: Results with different client encoders on TSLib. Full results are in Table 29.

| | Transformer | | | | TSMixer | | | | | | Mamba | | | | | | Mixed | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Scratch | | FEDSTATE | | Scratch | | FEDSTATE | | +Law | | Scratch | | FEDSTATE | | +Law | | FEDSTATE | |
| **Dataset** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| ETTh1 | 0.473 | 0.462 | **0.445** | **0.442** | 0.461 | 0.453 | **0.435** | **0.432** | 0.462 | 0.452 | 0.478 | 0.467 | **0.443** | **0.439** | 0.463 | 0.457 | 0.463 | 0.451 |
| ETTh2 | 0.385 | 0.409 | **0.375** | **0.398** | 0.382 | 0.407 | **0.372** | **0.397** | 0.377 | 0.406 | 0.387 | 0.413 | **0.378** | **0.401** | 0.391 | 0.416 | 0.368 | 0.397 |
| ETTm1 | **0.399** | **0.404** | 0.399 | 0.405 | **0.387** | **0.397** | 0.389 | 0.398 | 0.389 | 0.399 | 0.399 | 0.406 | **0.393** | **0.403** | 0.395 | 0.404 | 0.398 | 0.406 |
| ETTm2 | 0.284 | 0.329 | **0.283** | **0.326** | 0.286 | 0.332 | **0.281** | **0.327** | 0.283 | 0.329 | 0.288 | 0.334 | **0.282** | **0.326** | 0.287 | 0.332 | 0.288 | 0.333 |
| ECL | **0.178** | **0.263** | 0.181 | 0.266 | **0.182** | **0.267** | 0.184 | 0.268 | 0.184 | 0.268 | **0.176** | **0.261** | 0.178 | 0.263 | 0.176 | 0.262 | 0.183 | 0.268 |
| Weather | **0.261** | **0.280** | 0.261 | 0.280 | 0.260 | **0.279** | **0.260** | 0.279 | 0.261 | 0.279 | 0.260 | 0.280 | **0.258** | **0.278** | 0.260 | 0.279 | 0.269 | 0.285 |
| Exchange | 0.363 | 0.406 | **0.354** | **0.398** | 0.365 | 0.406 | **0.358** | **0.399** | 0.362 | 0.404 | 0.375 | 0.412 | **0.369** | **0.407** | 0.376 | 0.411 | 0.428 | 0.430 |
| **Avg** | 0.341 | 0.368 | **0.334** | **0.363** | 0.337 | 0.367 | **0.331** | **0.360** | 0.337 | 0.366 | 0.344 | 0.371 | **0.334** | **0.363** | 0.341 | 0.370 | 0.348 | 0.371 |
| **Gain (%)** | – | – | +2.0 | +1.5 | – | – | +1.9 | +1.7 | +0.3 | +0.1 | – | – | +2.6 | +2.2 | +0.7 | +0.4 | – | – |

**Heterogeneous Model Transfer.** Table 9 shows that FEDSTATE consistently improves Transformer-, TSMixer- (Chen et al., 2023), and Mamba-based (Gu & Dao, 2023) clients, reducing mean MSE by 2.0%, 1.9%, and 2.6% relative to training the same encoder from scratch. Since only the transition operator is shared, heterogeneous architectures can participate in a common federation, and an operator learned with Transformer clients transfers directly to TSMixer and Mamba clients. This transfer further extends to unseen frozen encoders, including Chronos-Bolt, Chronos-2, and Time-MoE, where the pretrained operator consistently improves downstream forecasting (as shown in Table 10).

Table 10: Frozen pretrained encoders as private TSLib clients. + Law adds the federated transition operator.

| | Local head | | + Law | | |
|---|---|---|---|---|---|
| **Private encoder** | **MSE** | **MAE** | **MSE** | **MAE** | **ΔMSE (%)** |
| Attention (2.5M, trained) | 0.361 | 0.382 | 0.353 | 0.375 | −2.2 |
| Chronos-Bolt (205M) | 0.374 | 0.392 | 0.362 | 0.385 | −3.2 |
| Chronos-2 (120M) | 0.371 | 0.389 | 0.365 | 0.384 | −1.8 |
| TimeMoE-50M | 0.363 | 0.385 | 0.355 | 0.378 | −2.0 |

## 5 CONCLUSION

This paper propose FEDSTATE, a federated time-series pretraining framework that keeps observation models local and shares only latent dynamics. Interval conditioning aligns temporal semantics across sampling grids, while heterogeneity-aware aggregation accounts for unequal client difficulty. Across eight benchmarks, FEDSTATE outperforms full-model federation with substantially lower communication and transfers effectively to unseen domains without target adaptation.

### AI USE STATEMENT

We used generative AI tools to check the mathematical proofs, to check the consistency of notation, to assist in implementing and revising code, and to polish the language of the paper. They were not used for any other research task, including developing the method, designing experiments, or interpreting results. The authors reviewed all AI-assisted content and take full responsibility for this work.

### REPRODUCIBILITY STATEMENT

Appendix B lists every benchmark with its clients, protocol, and metrics (Table 11) and every hyperparameter (Table 13), Appendix C gives the per-setting results behind every main-text table, and Appendix A carries every proof. Benchmark data, splits, and baseline scores come from the benchmarks' public releases, and code and records will be released.

## REFERENCES

Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. Gift-eval: A benchmark for general time series forecasting model evaluation. *arXiv preprint arXiv:2410.10393*, 2024.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting, 2025. URL https://arxiv.org/abs/2510.15821.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15619–15629. IEEE, 2023.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*, 2022.

Shengchao Chen, Guodong Long, Jing Jiang, and Chengqi Zhang. Federated foundation models on heterogeneous time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15839–15847, 2025.

Shengchao Chen, Guodong Long, Michael Blumenstein, and Jing Jiang. Fedal: Federated dataset learning for general time series foundation models. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=HK6t5x5gJq.

Shengchao Chen, Guodong Long, Dikai Liu, and Jing Jiang. Bi-level heterogeneous learning for time series foundation models: A federated learning approach. *arXiv preprint arXiv:2604.06727*, 2026b.

Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O Arik, and Tomas Pfister. Tsmixer: An all-mlp architecture for time series forecasting. *arXiv preprint arXiv:2303.06053*, 2023.

Mingyue Cheng, Qi Liu, Zhiding Liu, Hao Zhang, Rujiao Zhang, and Enhong Chen. Timemae: Self-supervised representations of time series with decoupled masked autoencoders. In *Proceedings of the Nineteenth ACM International Conference on Web Search and Data Mining*, 2026.

Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *International conference on machine learning*, pp. 2089–2099. PMLR, 2021.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Forty-first International Conference on Machine Learning*, 2024.

Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. Simmtm: A simple pre-training framework for masked time-series modeling. *Advances in Neural Information Processing Systems*, 36:29996–30025, 2023.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan. Time-series representation learning via temporal and contextual contrasting. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 2352–2359, 2021.

Patrick Emami, Abhijeet Sahu, and Peter Graf. Buildingsbench: A large-scale dataset of 900k buildings and benchmark for short-term load forecasting. *Advances in Neural Information Processing Systems*, 36:19823–19857, 2023.

Margalit R Glasgow, Honglin Yuan, and Tengyu Ma. Sharp bounds for federated averaging (local sgd) and continuous perspective. In *International Conference on Artificial Intelligence and Statistics*, pp. 9050–9090. PMLR, 2022.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019.

Tao Han, Zhibin Wen, Zhenghao Chen, Dazhao Du, Song Guo, and Lei Bai. Benchmarking physics-informed time-series models for operational global station weather forecasting, 2026. URL `https://arxiv.org/abs/2406.14399`.

Cheng He, Xu Huang, Gangwei Jiang, Zhaoyi Li, Defu Lian, Hong Xie, Enhong Chen, Patrick PC Lee, et al. Gtm: A general time-series model for enhanced representation learning of time-series data. In *International Conference on Learning Representations*, volume 2026, pp. 7280–7309, 2026.

Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv preprint arXiv:1909.06335*, 2019.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pp. 5132–5143. PMLR, 2020.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.

Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *International conference on machine learning*, pp. 6357–6368. PMLR, 2021a.

Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. In *International Conference on Learning Representations*, 2021b.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024.

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.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

Zhongzheng Qiao, Sheng Pan, Anni Wang, Viktoriya Zhukova, Yong Liu, Xudong Jiang, Qingsong Wen, Mingsheng Long, Ming Jin, and Chenghao Liu. It's time: Towards the next generation of time series forecasting benchmarks. *arXiv preprint arXiv:2602.12147*, 2026.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, et al. Lag-llama: Towards foundation models for probabilistic time series forecasting. *arXiv preprint arXiv:2310.08278*, 2023.

Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019.

Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-moe: Billion-scale time series foundation models with mixture of experts. In *International Conference on Learning Representations*, 2025.

Daoyu Wang, Mingyue Cheng, Zhiding Liu, and Qi Liu. Timedart: A diffusion autoregressive transformer for self-supervised time series representation. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

Wenxuan Wang, Kai Wu, Dan Wang, Xiaoyu Zhang, et al. Synthetic series-symbol data generation for time series foundation models. *Advances in Neural Information Processing Systems*, 38: 10491–10554, 2026.

Haomin Wen, Ziyu Zhou, Qingxiang Liu, Siru Zhong, and Yuxuan Liang. Livehouse-ts: An open-world living benchmark for time series foundation models. *arXiv preprint arXiv:2608.17299*, 2026.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *International Conference on Learning Representations*, 2022.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53140–53164, 2024.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8980–8987, 2022.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

**Roadmap.** Appendix A states and proves Theorem 1 and Corollary 1, and shows that the implemented aggregation solves the tilted objective exactly (Proposition 1). Appendix B describes the benchmarks and baselines, the implementation, and the fine-tuning and calibration procedure. Appendix C reports the full results of in-federation forecasting (Appendix C.1) and zero-shot transfer, including pretraining on LOTSA (Appendix C.2), followed by further analyses (Appendix C.3) and case studies (Appendix C.4).

# A Theorems and Proofs

## A.1 Setting and Statement of Theorem 1

**Setting.** States live in $\mathcal{Z} = \mathbb{R}^d$ with its Borel $\sigma$-algebra. Client $i \in [N]$ has sampling period $\tau_i > 0$, horizon $H_i \in \mathbb{N}$, and interval grid $G_i := \{\tau_i, 2\tau_i, \dots, H_i\tau_i\}$. On a probability space of its own, client $i$ observes a random state $z$ (the encoder's reading of the context) with law $P_i^z$ and, for every $\Delta \in G_i$, a random target state $\zeta_i(\Delta) \in \mathcal{Z}$ (the target encoder's reading of the future segment that begins at offset $\Delta$, the two sides of Eq. (7)). Per training step the client draws $\Delta$ from an offset distribution $\mu_i$ on $G_i$ and writes $T_i := \{\Delta \in G_i : \mu_i(\{\Delta\}) > 0\}$. The implemented $\mu_i$ first draws a horizon from the declared grid and then an integer offset uniformly within it, and every statement below holds for an arbitrary $\mu_i$. Let $\mathcal{G} := \bigcup_i T_i$, the finite set of intervals that some client samples. A shared interval-indexed predictor is a family $f = (f(\cdot, \Delta))_{\Delta \in \mathcal{G}}$ of Borel maps $\mathcal{Z} \to \mathcal{Z}$. Client $i$'s risk and the federation's weighted risk, for weights $w$ in the simplex with every $w_i > 0$, are

$$R_i(f) := \sum_{\Delta \in T_i} \mu_i(\{\Delta\})\, \mathbb{E}\left\|f(z, \Delta) - \zeta_i(\Delta)\right\|^2 \in [0, \infty], \qquad R_w(f) := \sum_{i=1}^{N} w_i\, R_i(f). \tag{9}$$

Both are sums of finitely many terms in $[0, \infty]$, since every grid is finite.

**Assumption 1.** For every client $i$ and every $\Delta \in G_i$, $\mathbb{E}\|\zeta_i(\Delta)\|^2 < \infty$.

**Transport Kernels.** Under Assumption 1, $\zeta_i(\Delta)$ is integrable, so its conditional expectation given $z$ exists, and by the Doob–Dynkin lemma it equals $\Phi_i(z, \Delta)$ almost surely for a Borel map $\Phi_i(\cdot, \Delta) : \mathcal{Z} \to \mathcal{Z}$, unique up to $P_i^z$-null sets. We call $\Phi_i$ client $i$'s *transport kernel*. By the conditional Jensen inequality,

$$\mathbb{E}\,\|\Phi_i(z, \Delta)\|^2 \le \mathbb{E}\,\|\zeta_i(\Delta)\|^2 < \infty. \tag{10}$$

No relation between the kernels of different clients is assumed. Their agreement is what the theorem measures.

**Mixing Measures.** For $\Delta \in \mathcal{G}$ let $A(\Delta) := \{i : \Delta \in T_i\}$, the nonempty set of clients that sample $\Delta$, let $\alpha_i(\Delta) := w_i\, \mu_i(\{\Delta\}) > 0$ for $i \in A(\Delta)$, and let $m_\Delta := \sum_{i \in A(\Delta)} \alpha_i(\Delta)\, P_i^z$, a finite Borel measure on $\mathcal{Z}$. Each $\alpha_i(\Delta) P_i^z$ is dominated by $m_\Delta$, so by the Radon–Nikodym theorem it has a Borel density $r_i(\cdot, \Delta) = d(\alpha_i(\Delta) P_i^z)/dm_\Delta$, and $\sum_{i \in A(\Delta)} r_i(\cdot, \Delta) = 1$ holds $m_\Delta$-almost everywhere by uniqueness of Radon–Nikodym derivatives. The set $B_\Delta$ where some $r_i$ leaves $[0, 1]$ or the sum differs from 1 is Borel and $m_\Delta$-null, and redefining the densities on $B_\Delta$ as a fixed probability vector changes no integral against $m_\Delta$, so we fix versions with $r_i(\cdot, \Delta) \in [0, 1]$ and $\sum_{i \in A(\Delta)} r_i(\cdot, \Delta) \equiv 1$ everywhere. The *barycentric kernel*

$$\bar{\Phi}(z, \Delta) := \sum_{i \in A(\Delta)} r_i(z, \Delta)\, \Phi_i(z, \Delta) \tag{11}$$

involves only kernels at intervals in their own client's grid, and is Borel in $z$ as a finite sum of products of Borel functions. Changing $\Phi_i(\cdot, \Delta)$ on a $P_i^z$-null set $\mathcal{N}$ changes $\bar{\Phi}(\cdot, \Delta)$ only on an $m_\Delta$-null set, since $\int_{\mathcal{N}} r_i\, dm_\Delta = \alpha_i(\Delta) P_i^z(\mathcal{N}) = 0$ forces $r_i(\cdot, \Delta) = 0$ $m_\Delta$-almost everywhere on $\mathcal{N}$.

**Slot Indexing.** Slot indexing is the grid indexing of Theorem 1. Fix an integer slot budget $K \ge 1$ with $K \le \min_i H_i$, so that slot indices $k \in \{1, \dots, K\}$ satisfy $k\tau_i \in G_i$ for every client. A shared

slot-indexed predictor is a family $g = (g_k)_{k \le K}$ of Borel maps $\mathcal{Z} \to \mathcal{Z}$, where slot $k$ serves the interval $k\tau_i$ on client $i$ and every slot carries weight $1/K$:

$$R_i^{\mathrm{slot}}(g) := \frac{1}{K} \sum_{k=1}^{K} \mathbb{E} \left\| g_k(z) - \zeta_i(k\tau_i) \right\|^2, \qquad R_w^{\mathrm{slot}}(g) := \sum_i w_i \, R_i^{\mathrm{slot}}(g). \tag{12}$$

Here the mixing measure is $m := \sum_i w_i P_i^z$ for every slot, with Borel densities $\rho_i := d(w_i P_i^z)/dm$, versions chosen as above with $\rho_i \in [0,1]$ and $\sum_i \rho_i \equiv 1$, and the slot barycenter is $\bar{\Phi}_k(z) := \sum_i \rho_i(z) \, \Phi_i(z, k\tau_i)$.

**Theorem 1 (Excess risk induced by grid indexing).** *Let Assumption 1 hold. For each client let $R_i^\star := \inf_f R_i(f)$ and $R_i^{\mathrm{slot}\star} := \inf_g R_i^{\mathrm{slot}}(g)$ be its Bayes risks, the infima taken over all Borel families.*

*(a)* ***Interval indexing.*** *$R_i^\star = \sum_{\Delta \in T_i} \mu_i(\{\Delta\}) \, \mathbb{E}\|\Phi_i(z,\Delta) - \zeta_i(\Delta)\|^2 < \infty$, and for every Borel family $f$,*

$$R_w(f) - \sum_i w_i R_i^\star = \sum_{\Delta \in \mathcal{G}} \int \left\| f(z,\Delta) - \bar{\Phi}(z,\Delta) \right\|^2 dm_\Delta(z) \; + \; D, \tag{13}$$

*both sides in $[0,\infty]$, where the* interval dispersion

$$D := \sum_{\Delta \in \mathcal{G}} \int \sum_{i \in A(\Delta)} r_i(z,\Delta) \left\| \Phi_i(z,\Delta) - \bar{\Phi}(z,\Delta) \right\|^2 dm_\Delta(z) \tag{14}$$

*is finite and does not depend on $f$. Consequently $\min_f R_w(f) - \sum_i w_i R_i^\star = D$, the minimum is attained exactly by the families with $f(\cdot,\Delta) = \bar{\Phi}(\cdot,\Delta)$ $m_\Delta$-almost everywhere for every $\Delta \in \mathcal{G}$, and $R_w(f) - \min R_w$ equals the first sum in Eq. (13). Moreover $D = 0$ if and only if, for every $\Delta \in \mathcal{G}$ and every pair $i \neq j$ in $A(\Delta)$, $\Phi_i(\cdot,\Delta) = \Phi_j(\cdot,\Delta)$ holds $m_\Delta$-almost everywhere on $\{r_i(\cdot,\Delta) > 0\} \cap \{r_j(\cdot,\Delta) > 0\}$.*

*(b)* ***Slot indexing.*** *$R_i^{\mathrm{slot}\star} = \frac{1}{K} \sum_{k \le K} \mathbb{E}\|\Phi_i(z,k\tau_i) - \zeta_i(k\tau_i)\|^2 < \infty$, and*

$$\min_g R_w^{\mathrm{slot}}(g) - \sum_i w_i R_i^{\mathrm{slot}\star} = D^{\mathrm{slot}} := \frac{1}{K} \sum_{k=1}^{K} \int \sum_i \rho_i \left\| \Phi_i(\cdot,k\tau_i) - \bar{\Phi}_k \right\|^2 dm, \tag{15}$$

*attained exactly by $g_k = \bar{\Phi}_k$ $m$-almost everywhere: the slot-indexed excess is the dispersion of the kernels evaluated at the* different *physical intervals $\{k\tau_i\}_i$.*

**Corollary 1** (The dichotomy)**.** *Let Assumption 1 hold and suppose every client obeys the same dynamics: there are Borel maps $\Phi(\cdot,\Delta) : \mathcal{Z} \to \mathcal{Z}$, one for each $\Delta \in \bigcup_i G_i$, with $\Phi_i(\cdot,\Delta) = \Phi(\cdot,\Delta)$ $P_i^z$-almost everywhere for every $i$ and every $\Delta \in G_i$. Then*

*(i) $D = 0$, and the single family $f = (\Phi(\cdot,\Delta))_{\Delta \in \mathcal{G}}$ attains every client's Bayes risk at once: $R_i(\Phi) = R_i^\star$ for every $i$;*

*(ii) the slot-indexed excess is $D^{\mathrm{slot}} = \frac{1}{K} \sum_{k=1}^{K} \int \sum_i \rho_i(z) \left\| \Phi(z,k\tau_i) - \sum_j \rho_j(z) \, \Phi(z,k\tau_j) \right\|^2 dm(z)$, and $D^{\mathrm{slot}} > 0$ whenever there exist a slot $k \le K$, clients $i \neq j$, and a Borel set $E \subseteq \{\rho_i > 0\} \cap \{\rho_j > 0\}$ with $m(E) > 0$ on which $\Phi(\cdot,k\tau_i) \neq \Phi(\cdot,k\tau_j)$. This requires $\tau_i \neq \tau_j$, since equal periods give equal arguments.*

### A.2 Proof of Theorem 1 and Corollary 1

**Lemma 1** (Pythagorean decomposition)**.** *Let $\zeta$ be a random vector in $\mathcal{Z}$ with $\mathbb{E}\|\zeta\|^2 < \infty$, $z$ a random vector in $\mathcal{Z}$, and $\Phi$ a Borel map with $\Phi(z) = \mathbb{E}[\zeta \mid z]$ almost surely. Then for every*

*Borel map $h : \mathcal{Z} \to \mathcal{Z}$,*

$$\mathbb{E}\|h(z) - \zeta\|^2 = \mathbb{E}\|h(z) - \Phi(z)\|^2 + \mathbb{E}\|\Phi(z) - \zeta\|^2 \quad \textit{in } [0, \infty],$$

*and the last term is finite.*

*Proof.* By Eq. (10), $\mathbb{E}\|\Phi(z)\|^2 \le \mathbb{E}\|\zeta\|^2 < \infty$, so $\mathbb{E}\|\Phi(z) - \zeta\|^2 \le 2\mathbb{E}\|\Phi(z)\|^2 + 2\mathbb{E}\|\zeta\|^2 < \infty$.

*Case* $\mathbb{E}\|h(z)-\Phi(z)\|^2 < \infty$. Expand $\|h-\zeta\|^2 = \|h-\Phi\|^2+2\langle h-\Phi, \Phi-\zeta\rangle+\|\Phi-\zeta\|^2$ pointwise, writing $h = h(z)$ and $\Phi = \Phi(z)$. The cross term is integrable by Cauchy–Schwarz, $\mathbb{E}|\langle h - \Phi, \Phi - \zeta\rangle| \le (\mathbb{E}\|h-\Phi\|^2)^{1/2}(\mathbb{E}\|\Phi-\zeta\|^2)^{1/2} < \infty$. Its expectation is zero: with $V := h(z) - \Phi(z)$, which is $\sigma(z)$-measurable, truncate by $V_n := V\,\mathbf{1}\{\|V\| \le n\}$. Then $\mathbb{E}\langle V_n, \Phi(z)-\zeta\rangle = \mathbb{E}\langle V_n, \mathbb{E}[\Phi(z)-\zeta \mid z]\rangle = 0$, because $V_n$ is bounded and $\sigma(z)$-measurable and $\mathbb{E}[\zeta \mid z] = \Phi(z)$ almost surely. Moreover $|\langle V_n, \Phi - \zeta\rangle| \le |\langle V, \Phi - \zeta\rangle|$, which is integrable, and $\langle V_n, \Phi - \zeta\rangle \to \langle V, \Phi - \zeta\rangle$ pointwise, so dominated convergence gives $\mathbb{E}\langle V, \Phi(z) - \zeta\rangle = 0$. Taking expectations of the expansion proves the identity.

*Case* $\mathbb{E}\|h(z) - \Phi(z)\|^2 = \infty$. The right side is $+\infty$. By the pointwise inequality $\|a\|^2 \le 2\|a + b\|^2 + 2\|b\|^2$ with $a = h - \Phi$ and $b = \Phi - \zeta$, we have $\|h - \Phi\|^2 \le 2\|h - \zeta\|^2 + 2\|\Phi - \zeta\|^2$. Taking expectations, $\infty \le 2\mathbb{E}\|h - \zeta\|^2 + 2\mathbb{E}\|\Phi - \zeta\|^2$, and the last term is finite, so $\mathbb{E}\|h(z) - \zeta\|^2 = \infty$ as well. □

**Lemma 2** (Barycentric decomposition)**.** *Let $v, \phi_1, \dots, \phi_N \in \mathcal{Z}$ and let $\rho_1, \dots, \rho_N \ge 0$ with $\sum_i \rho_i = 1$. With $\bar\phi := \sum_i \rho_i \phi_i$,*

$$\sum_i \rho_i \|v - \phi_i\|^2 = \|v - \bar\phi\|^2 + \sum_i \rho_i \|\phi_i - \bar\phi\|^2, \qquad \sum_i \rho_i \|\phi_i - \bar\phi\|^2 \le \sum_i \rho_i \|\phi_i\|^2.$$

*The dispersion $\sum_i \rho_i \|\phi_i - \bar\phi\|^2$ is zero if and only if $\phi_i = \phi_j$ for every pair $i, j$ with $\rho_i, \rho_j > 0$.*

*Proof.* For each $i$, $\|v - \phi_i\|^2 = \|v - \bar\phi\|^2 + 2\langle v - \bar\phi, \bar\phi - \phi_i\rangle + \|\bar\phi - \phi_i\|^2$. Multiply by $\rho_i$ and sum. The first terms give $\|v - \bar\phi\|^2$ because $\sum_i \rho_i = 1$. The cross terms give $2\langle v - \bar\phi, \sum_i \rho_i(\bar\phi - \phi_i)\rangle = 2\langle v - \bar\phi, \bar\phi - \bar\phi\rangle = 0$. This proves the identity. Taking $v = 0$ in the identity gives $\sum_i \rho_i\|\phi_i\|^2 = \|\bar\phi\|^2 + \sum_i \rho_i\|\phi_i - \bar\phi\|^2$, which proves the inequality. For the last claim: if all $\phi_i$ with $\rho_i > 0$ equal a common $\phi$, then $\bar\phi = \phi$ and every term of the dispersion vanishes. Conversely, a zero dispersion is a sum of nonnegative terms, so $\phi_i = \bar\phi$ for every $i$ with $\rho_i > 0$, and any two such $\phi_i, \phi_j$ are equal. □

*Proof of Theorem 1(a).* Fix $i$ and $\Delta \in T_i$, and define $e_i(\Delta)$ as in Eq. (16), which Lemma 1 shows to be finite. Lemma 1 with $\zeta = \zeta_i(\Delta)$, $\Phi = \Phi_i(\cdot, \Delta)$, and $h = f(\cdot, \Delta)$ gives

$$\mathbb{E}\|f(z, \Delta) - \zeta_i(\Delta)\|^2 = \mathbb{E}\|f(z, \Delta) - \Phi_i(z, \Delta)\|^2 + e_i(\Delta), \qquad e_i(\Delta) := \mathbb{E}\|\Phi_i(z, \Delta) - \zeta_i(\Delta)\|^2. \tag{16}$$

Multiplying by $\mu_i(\{\Delta\})$ and summing over the finitely many $\Delta \in T_i$ gives $R_i(f) = \sum_{\Delta \in T_i} \mu_i(\{\Delta\})\,\mathbb{E}\|f(z, \Delta) - \Phi_i(z, \Delta)\|^2 + \sum_{\Delta \in T_i} \mu_i(\{\Delta\}) e_i(\Delta)$. The first sum is nonnegative and vanishes at the admissible Borel family $f^{(i)}$ with $f^{(i)}(\cdot, \Delta) := \Phi_i(\cdot, \Delta)$ for $\Delta \in T_i$ and $f^{(i)}(\cdot, \Delta) := 0$ for $\Delta \in \mathcal{G} \setminus T_i$, whose values off $T_i$ do not enter $R_i$. Hence $R_i^\star = \sum_{\Delta \in T_i} \mu_i(\{\Delta\}) e_i(\Delta)$, which is finite.

Subtract $R_i^\star$ from $R_i(f)$ (legitimate, since $R_i^\star < \infty$), multiply by $w_i$, and sum over $i$. Grouping the finitely many terms by interval, client $i$ contributes at $\Delta$ exactly when $i \in A(\Delta)$, with coefficient $w_i \mu_i(\{\Delta\}) = \alpha_i(\Delta)$. This gives

$$R_w(f) - \sum_i w_i R_i^\star = \sum_{\Delta \in \mathcal{G}} \sum_{i \in A(\Delta)} \alpha_i(\Delta) \int \|f(z, \Delta) - \Phi_i(z, \Delta)\|^2 \, dP_i^z(z).$$

For a nonnegative Borel function $u$, the density $r_i(\cdot, \Delta)$ gives $\alpha_i(\Delta) \int u \, dP_i^z = \int u \, r_i(\cdot, \Delta) \, dm_\Delta$, first for simple $u$ and then by monotone convergence. Applying it with $u = \|f(\cdot, \Delta) - \Phi_i(\cdot, \Delta)\|^2$

and exchanging the finite sum over $i$ with the integral,

$$R_w(f) - \sum_i w_i R_i^\star = \sum_{\Delta \in \mathcal{G}} \int \sum_{i \in A(\Delta)} r_i(z, \Delta) \, \|f(z, \Delta) - \Phi_i(z, \Delta)\|^2 \, dm_\Delta(z). \tag{17}$$

For each $\Delta$ and each $z$, apply Lemma 2 to the clients in $A(\Delta)$ with $v = f(z, \Delta)$, $\phi_i = \Phi_i(z, \Delta)$, and $\rho_i = r_i(z, \Delta)$. Then $\phi = \bar{\Phi}(z, \Delta)$, and the integrand of Eq. (17) equals $\|f(z, \Delta) - \bar{\Phi}(z, \Delta)\|^2 + \sum_{i \in A(\Delta)} r_i(z, \Delta) \|\Phi_i(z, \Delta) - \bar{\Phi}(z, \Delta)\|^2$. Both summands are nonnegative and Borel, so integrating against $m_\Delta$ and summing over $\Delta$ gives Eq. (13), as an identity in $[0, \infty]$.

The dispersion $D$ is finite. By the inequality of Lemma 2, the change of measure above with $u = \|\Phi_i(\cdot, \Delta)\|^2$, and Eq. (10),

$$\begin{aligned} \int \sum_{i \in A(\Delta)} r_i \|\Phi_i - \bar{\Phi}\|^2 \, dm_\Delta &\le \int \sum_{i \in A(\Delta)} r_i \|\Phi_i\|^2 \, dm_\Delta = \sum_{i \in A(\Delta)} \alpha_i(\Delta) \, \mathbb{E}\|\Phi_i(z, \Delta)\|^2 \\ &\le \sum_{i \in A(\Delta)} \alpha_i(\Delta) \, \mathbb{E}\|\zeta_i(\Delta)\|^2 < \infty. \end{aligned}$$

Summing over the finitely many $\Delta$ gives $D < \infty$. Nothing in $D$ involves $f$.

The first sum in Eq. (13) is nonnegative. It vanishes if and only if $f(\cdot, \Delta) = \bar{\Phi}(\cdot, \Delta)$ $m_\Delta$-almost everywhere for every $\Delta \in \mathcal{G}$, because a nonnegative function has zero integral if and only if it vanishes almost everywhere. The family $f = \bar{\Phi}$ is Borel by Eq. (11), so it is admissible. By Eq. (13), $R_w(\bar{\Phi}) = \sum_i w_i R_i^\star + D < \infty$, and every other $f$ has $R_w(f) \ge R_w(\bar{\Phi})$. Hence the minimum is attained, equals $\sum_i w_i R_i^\star + D$, and is attained exactly by the stated families. Subtracting the finite number $\min R_w$ from Eq. (13) gives $R_w(f) - \min R_w$ as the first sum.

Finally, $D$ is a finite sum of integrals of nonnegative functions, so $D = 0$ if and only if, for every $\Delta$, the dispersion $\sum_{i \in A(\Delta)} r_i \|\Phi_i - \bar{\Phi}\|^2$ vanishes $m_\Delta$-almost everywhere. Suppose first that $D = 0$ and fix $\Delta$ and $i \ne j$ in $A(\Delta)$. For $m_\Delta$-almost every $z \in \{r_i > 0\} \cap \{r_j > 0\}$ the dispersion is zero, and Lemma 2 then gives $\Phi_i(z, \Delta) = \Phi_j(z, \Delta)$. Conversely, suppose the pairwise agreement holds for every pair. There are finitely many pairs, so the union of the exceptional null sets is a single $m_\Delta$-null set $\mathcal{N}$. For $z \notin \mathcal{N}$, all kernels with $i \in A(\Delta)$ and $r_i(z, \Delta) > 0$ coincide, and Lemma 2 gives zero dispersion at $z$. Hence $D = 0$. □

*Proof of Theorem 1(b).* The argument of part (a) applies slot by slot, with $(m, \rho_i, \bar{\Phi}_k, g_k)$ in place of $(m_\Delta, r_i, \bar{\Phi}(\cdot, \Delta), f(\cdot, \Delta))$ and the uniform slot weight $1/K$ in place of $\mu_i(\{\Delta\})$. Every slot is admissible: $k\tau_i \in G_i$ for each $k \le K$ and each $i$, so Assumption 1 holds at $k\tau_i$ and Lemma 1 gives Eq. (16) with $\zeta = \zeta_i(k\tau_i)$, $\Phi = \Phi_i(\cdot, k\tau_i)$, and $h = g_k$; averaging over $k$ gives $R_i^{\mathrm{slot}\star} = \frac{1}{K} \sum_{k \le K} e_i(k\tau_i) < \infty$, attained at $g_k = \Phi_i(\cdot, k\tau_i)$. Subtracting it, weighting by $w_i$, summing, and changing measure with $w_i \int u \, dP_i^z = \int u \, \rho_i \, dm$ leaves $\frac{1}{K} \sum_{k \le K} \int \sum_i \rho_i \|g_k - \Phi_i(\cdot, k\tau_i)\|^2 \, dm$, which Lemma 2 splits pointwise into $\|g_k - \bar{\Phi}_k\|^2$ plus the dispersion; its inequality with Eq. (10) bounds the dispersion's integral by $\sum_i w_i \mathbb{E}\|\zeta_i(k\tau_i)\|^2 < \infty$. Part (a) then yields Eq. (15) with the minimum attained exactly at $g_k = \bar{\Phi}_k$ $m$-almost everywhere. □

*Proof of Corollary 1.* (i) Fix $\Delta \in \mathcal{G}$. Each $\Phi_i(\cdot, \Delta)$, $i \in A(\Delta)$, equals $\Phi(\cdot, \Delta)$ off a $P_i^z$-null set, on which $r_i(\cdot, \Delta) = 0$ holds $m_\Delta$-almost everywhere by the null-set argument below Eq. (11). Hence, outside the $m_\Delta$-null set formed by these finitely many exceptions, every kernel with $i \in A(\Delta)$ and $r_i(z, \Delta) > 0$ equals $\Phi(z, \Delta)$. The pairwise-agreement condition of Theorem 1(a) therefore holds, and $D = 0$. For each $i$, Eq. (16) with $f = \Phi$ gives $\mathbb{E}\|\Phi(z, \Delta) - \Phi_i(z, \Delta)\|^2 = 0$ for every $\Delta \in T_i$, so $R_i(\Phi) = R_i^\star$.

(ii) The same null-set argument, with $w_i P_i^z$ and $\rho_i$ in place of $\alpha_i(\Delta) P_i^z$ and $r_i$, shows that $\Phi_i(z, k\tau_i)$ may be replaced by $\Phi(z, k\tau_i)$ $m$-almost everywhere wherever $\rho_i(z) > 0$. This turns Eq. (15) into the stated formula. Now take $k, i, j, E$ as in the statement. For every $z \in E$, the family $\{\Phi(z, k\tau_l)\}_l$ has two members with positive weight, $\rho_i(z), \rho_j(z) > 0$, that differ. By Lemma 2, the pointwise dispersion at slot $k$ is then strictly positive at every $z \in E$. A nonnegative function that is strictly positive on a set of positive measure has a strictly positive integral, so the slot-$k$ term, and with it $D^{\mathrm{slot}}$, is strictly positive. Finally, if $\tau_i = \tau_j$ then $\Phi(\cdot, k\tau_i) = \Phi(\cdot, k\tau_j)$ identically, so the hypothesis forces $\tau_i \ne \tau_j$. □

*Remark* 1 (Loss used in training). Theorem 1 is stated for the squared loss. Training uses the Huber loss of Eq. (5), which coincides with the squared loss below its threshold.

### A.3 Exactness of the Implemented Aggregation

**Proposition 1** (Exactness of the implemented aggregation)**.** *Fix a round with client losses $\ell \in \mathbb{R}^N$, a prior $\pi$ with every $\pi_i > 0$ and $\sum_i \pi_i = 1$, and a temperature $\eta > 0$. Let $\mathcal{P}_N$ denote the probability simplex.*

- *(a) The problem $\max_{w \in \mathcal{P}_N} J_\ell(w)$, with $J_\ell(w) := \sum_i w_i \ell_i - \eta\,\mathrm{KL}(w \,\|\, \pi)$ and $\mathrm{KL}(w\|\pi) = \sum_i w_i \log(w_i/\pi_i)$ (with $0 \log 0 = 0$), has the unique solution $w_i^\star = \pi_i e^{\ell_i/\eta} / \sum_j \pi_j e^{\ell_j/\eta}$, and its optimal value is $\eta \log \sum_j \pi_j e^{\ell_j/\eta}$.*
- *(b) The implemented update standardizes the losses, $s_i = (\ell_i - \bar{\ell})/(\sigma + \epsilon)$, where $\bar{\ell}$ is the round mean, $\sigma \geq 0$ the population (ddof-0) standard deviation, and $\epsilon > 0$ ($10^{-8}$ in the implementation). It clips the logits, $c_i = \mathrm{clip}(\beta s_i, -4, 4)$ with $\beta > 0$, and sets $w_i \propto \pi_i e^{c_i}$. Let $\eta_r := (\sigma + \epsilon)/\beta$ and $\tilde{\ell}_i := \bar{\ell} + \eta_r c_i$, the loss $\ell_i$ capped at $\bar{\ell} \pm 4\eta_r$. Then the implemented weights are exactly the unique maximizer of $J_{\tilde{\ell}}$ at temperature $\eta_r$. When no logit is clipped, $\tilde{\ell} = \ell$ and they are exactly the maximizer of $J_\ell$ at $\eta_r$.*
- *(c) If $N \geq 2$ and $\sigma > 0$, then $\max_i |s_i| < \sqrt{N-1}$. Hence no logit is clipped whenever $\beta\sqrt{N-1} \leq 4$, which at $\beta = 1$, the value used in all experiments, covers every federation of at most 17 clients. In any federation, client $i$ is clipped only when $|\ell_i - \bar{\ell}| > 4(\sigma + \epsilon)/\beta$.*

*Proof.* (a) Let $Z := \sum_j \pi_j e^{\ell_j/\eta} > 0$ and $w_i^\star := \pi_i e^{\ell_i/\eta}/Z$, which lies in the interior of $\mathcal{P}_N$. For any $w \in \mathcal{P}_N$, the terms with $w_i = 0$ contribute zero to every sum below. For the others, $\log(w_i/\pi_i) = \log(w_i/w_i^\star) + \ell_i/\eta - \log Z$. Multiplying by $w_i$ and summing,

$$\mathrm{KL}(w\|\pi) = \mathrm{KL}(w\|w^\star) + \frac{1}{\eta}\sum_i w_i \ell_i - \log Z, \quad \text{that is,} \quad J_\ell(w) = \eta \log Z - \eta\,\mathrm{KL}(w\|w^\star).$$

By Gibbs' inequality, $\mathrm{KL}(w\|w^\star) \geq 0$ with equality if and only if $w = w^\star$. Therefore $J_\ell(w) \leq \eta \log Z$, with equality if and only if $w = w^\star$. This proves existence, uniqueness, the closed form, and the optimal value.

(b) By definition $c_i = (\tilde{\ell}_i - \bar{\ell})/\eta_r$, so $\pi_i e^{c_i} = \pi_i e^{\tilde{\ell}_i/\eta_r} \cdot e^{-\bar{\ell}/\eta_r}$, and the second factor does not depend on $i$. It cancels in the normalization, so $w_i = \pi_i e^{\tilde{\ell}_i/\eta_r} / \sum_j \pi_j e^{\tilde{\ell}_j/\eta_r}$. By (a) applied to the losses $\tilde{\ell}$ at temperature $\eta_r > 0$, these are exactly the unique maximizing weights. Since $\mathrm{clip}(x, -4, 4) = x$ for $|x| \leq 4$, an unclipped logit has $c_i = \beta s_i$ and $\tilde{\ell}_i = \bar{\ell} + \eta_r \beta s_i = \ell_i$, and $\tilde{\ell}_i = \bar{\ell} \pm 4\eta_r$ otherwise, which is the stated cap.

(c) Let $u_i := (\ell_i - \bar{\ell})/\sigma$. By definition of the mean and of the population standard deviation, $\sum_i u_i = 0$ and $\sum_i u_i^2 = N$. Fix $i$. Then $u_i = -\sum_{j \neq i} u_j$, and Cauchy–Schwarz over the $N - 1$ indices $j \neq i$ gives

$$u_i^2 \leq (N-1)\sum_{j \neq i} u_j^2 = (N-1)(N - u_i^2), \quad \text{so} \quad N u_i^2 \leq N(N-1), \quad \text{that is,} \quad |u_i| \leq \sqrt{N-1}.$$

Since $s_i = u_i\,\sigma/(\sigma + \epsilon)$ and $0 < \sigma/(\sigma + \epsilon) < 1$, we get $|s_i| < \sqrt{N-1}$. If $\beta\sqrt{N-1} \leq 4$, then $|\beta s_i| < 4$ for every $i$, so no logit is clipped and (b) gives the maximizer of $J_\ell$. At $\beta = 1$ the condition $\sqrt{N-1} \leq 4$ reads $N \leq 17$. In general, $|\beta s_i| > 4$ is equivalent to $|\ell_i - \bar{\ell}| > 4(\sigma + \epsilon)/\beta$. □

## B Experimental Setup

### B.1 Benchmarks and Baselines

Table 11 lists every benchmark with the federation it is pretrained on, its evaluation, and its metrics.

Table 11: Benchmarks of Section 4. Deterministic and probabilistic metrics are separated by a semicolon.

| Benchmark | Pretrained on | Evaluation | Metrics |
|---|---|---|---|
| *In-domain representation learning (Section 4.1)* | | | |
| TSLib | TSLib, 8 clients | Fine-tuning, four horizons | MSE, MAE |
| BuildingsBench | BuildingsBench, 10 clients | Fine-tuning, 1,909 buildings, 168→24 | CVRMSE |
| *In-federation forecasting (Section 4.2)* | | | |
| TSLib | TSLib, 7 clients | Fine-tuning, four horizons | MSE, MAE; CRPS, Cov80 |
| TIME | TIME, 46 clients | Fine-tuning, 91 configurations | Seven point metrics; $\text{CRPS}_n$ |
| GIFT-Eval | GIFT-Eval, 52 clients | Fine-tuning, 97 configurations | $\text{MASE}_n$; $\text{WQL}_n$ |
| *Transfer beyond the federation (Section 4.3)* | | | |
| FFTS protocol | Monash, 9 clients | Fine-tuning on ETT and Weather at $100/10/5\%$ | MSE, MAE |
| TSLib zero-shot | Unified, 94 clients | No adaptation, paired 512→96 to 3072→720 | MSE, MAE |
| WEATHER-5K | Unified, 94 clients | No adaptation, 5,672 stations, inputs 48 and 512 | Overall MAE |
| Unseen stations | Unified, 94 clients | No adaptation, 15 stations at 512 | MASE, MAE; CRPS |
| LiveHouse-TS | Unified, 94 clients | No adaptation, 240 rolling origins | MASE |
| BuildingsBench | Unified, 94 clients | No adaptation, 4,020 buildings, 168→24 | CVRMSE |

**Unseen Weather Stations.** The unseen stations are the 15 ground-based observation stations of RW-Bench (Chen et al., 2026b), spread over five continents and the Pacific, from the tropics to Antarctica (Table 12). Each station records ten meteorological variables every hour from January 2022 to August 2025: air pressure, air temperature, relative humidity, precipitation, wind direction, wind speed, maximum wind direction, maximum wind speed, maximum temperature, and minimum temperature. The records are kept as operated, with outliers retained and missing values filled with zero, and neither the stations nor the period overlap with any pretraining corpus, so the benchmark measures zero-shot transfer to raw operational data. Every method forecasts 96, 192, 336, and 720 hours ahead from an input of 512 hours.

Table 12: The 15 stations of RW-Bench. Every station records ten meteorological variables hourly from January 2022 to August 2025.

| Station | Location | Station | Location | Station | Location |
|---|---|---|---|---|---|
| RW-1 | Yuzhno-Sakhalinsk | RW-6 | Cairo | RW-11 | Edmonton |
| RW-2 | Tokyo | RW-7 | Nairobi Dagoretti | RW-12 | Oxford |
| RW-3 | Beijing | RW-8 | Nauru | RW-13 | Zurich |
| RW-4 | Mumbai Santacruz | RW-9 | Honolulu | RW-14 | McMurdo |
| RW-5 | Ho Chi Minh City | RW-10 | Los Angeles | RW-15 | Novolazarevskaya |

**Baselines.** The baselines fall into five groups: federated methods, self-supervised pretraining objectives, centralized forecasters, pretrained encoders, and reference baselines.

- **FedAvg** (McMahan et al., 2017) trains a single global model by having each client take local SGD steps and averaging the resulting full-model updates at the server, weighted by client sample counts.
- **FedAvgM** (Hsu et al., 2019) augments FedAvg with server-side momentum on the aggregated update to stabilize convergence under non-identical client data distributions.
- **FedProx** (Li et al., 2020) adds a proximal term to each client's local objective that penalizes deviation from the current global model, limiting client drift in heterogeneous federations.
- **SCAFFOLD** (Karimireddy et al., 2020) uses server- and client-side control variates to correct the drift of local updates, reducing the variance introduced by non-IID clients.
- **FedBN** (Li et al., 2021b) keeps each client's batch-normalization parameters local while federating all remaining weights, accommodating feature-distribution shift across clients.
- **FedRep** (Collins et al., 2021) federates a shared representation backbone while each client learns a personalized prediction head on top of it.
- **FFTS** (Chen et al., 2025) is a federated masked-reconstruction pretraining method whose adaptive trend module is aggregated separately from the backbone.

- **FeDaL** (Chen et al., 2026a) pretrains a time series foundation model federatedly with masked patch reconstruction and handles cross-client heterogeneity on the server side.
- **SimMTM** (Dong et al., 2023) pretrains by masked time-series modeling, reconstructing each masked series from a weighted aggregation of several masked neighbors.
- **TimeDART** (Wang et al., 2025) combines an autoregressive Transformer over patches with a diffusion denoising decoder in self-supervised pretraining.
- **CoST** (Woo et al., 2022) learns disentangled seasonal and trend representations with contrastive objectives in the time and frequency domains.
- **TimeMAE** (Cheng et al., 2026) masks decoupled sub-series windows and pretrains with masked codeword classification and masked representation regression.
- **DLinear** (Zeng et al., 2023) decomposes each series into trend and seasonal components and forecasts each with a single linear layer, a deliberately minimal architecture.
- **PatchTST** (Nie et al., 2023) segments each channel into subseries patches and feeds them to a channel-independent Transformer encoder for long-horizon forecasting.
- **iTransformer** (Liu et al., 2024) inverts the standard attention layout, embedding each variate's whole series as a token so that attention operates across variates rather than across time.
- **TimesNet** (Wu et al., 2023) folds a series into 2D tensors along its dominant periods and models intra- and inter-period variations with inception-style convolutions.
- **Chronos-2** (Ansari et al., 2025) is the current generation of the Chronos family of pretrained time series models (Ansari et al., 2024), trained on a large pooled corpus; we use it as a frozen client encoder and as a zero-shot forecaster in the case study.
- **ChronosBolt** is an efficiency-oriented variant in the Chronos family (Ansari et al., 2024) that forecasts patches of future values directly rather than autoregressively sampling tokens; we use it as a frozen client encoder.
- **Time-MoE** (Shi et al., 2025) is a sparse mixture-of-experts time series foundation model pretrained on Time-300B, used here as a frozen client encoder.
- **Seasonal Naive** forecasts each future point as the value observed exactly one seasonal period earlier, and serves as the normalizer for MASE and the relative metrics.
- **Scratch** trains the identical local model architecture on each client or task alone with no communication, isolating the contribution of federation from that of the architecture.

### B.2 Implementation Details

The backbone is a three-layer decoder-only Transformer ($d_{\text{model}} = 128$, 64-dimensional states at three temporal scales with mid states pooling blocks of 4 patches, patch length 16, stride 8), of which the transition law is 0.36M parameters. Clients run 50 local Adam steps per round and the server aggregates with momentum 0.9. Every other hyperparameter is fixed across benchmarks (Table 13). Under the FFTS protocol, 5% of an ETTh training split is too short for a 720-step horizon, so the ETTh rows at 5% in Table 7 average three horizons. All experiments run on two NVIDIA RTX 5500 GPUs (24 GB each) in bfloat16. Every federated method uses the same quantile head, from whose nine levels CRPS and the coverage of the 80% interval are computed. Appendix B.3 describes fine-tuning and calibration.

**Regularizers and Target Encoder.** The target encoder $E_{\bar{\theta}_i}$ is an exponential moving average of $E_{\theta_i}$ with rate 0.996. The term $\mathcal{R}_i$ of Eq. (4) is

$$\mathcal{R}_i = \lambda_{\text{g}}\, \mathcal{H}_{\beta_g}\Big(g_{\psi_i}\big(f_\phi(z, \boldsymbol{\Delta}_i)^{\text{fast}}\big),\ \text{clip}\big(\tfrac{y-\mu_x}{\sigma_x}\big)\Big) + \lambda_{\text{v}}\, \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\frac{1}{d}\sum_{j=1}^{d}\big[\,1 - \text{std}_j(z^s)\,\big]_+. \tag{18}$$

The grounding term decodes the predicted fast states and compares them with the future normalized by the context's own mean $\mu_x$ and standard deviation $\sigma_x$, clipped to $[-10, 10]$. This decoder serves as the point forecaster. The variance term is a hinge on the batch standard deviation $\text{std}_j$ of every state dimension, with $[\cdot]_+ = \max(\cdot, 0)$ (Bardes et al., 2022).

**Algorithm 1** Implementation workflow of FEDSTATE

**Require:** clients $1, \dots, N$ with private data $\mathcal{D}_i$ and sampling period $\tau_i$, rounds $R$, local steps $S$, interval weight $\lambda_\ell$, temperature $\eta$, data-size prior $\pi_i \propto n_i$, server momentum $\mu$
**Ensure:** shared law $\phi^R$; client $i$ keeps its private $\vartheta_i = (\theta_i, \psi_i)$

**Initialize** $\phi^0$ and the momentum buffer $m^0 \leftarrow 0$; client $i$ initializes $\vartheta_i$ and its target encoder $\bar{\theta}_i \leftarrow \theta_i$
**for** $r = 1, \dots, R$ **do**
  **Broadcast** $\phi^{r-1}$ to every client
  **for** each client $i$ **in parallel do**
    **Local pretraining:** $\phi_i \leftarrow \phi^{r-1}$
    **for** $s = 1, \dots, S$ **do**
      draw a batch of context and future pairs $(x, y) \sim \mathcal{D}_i$
      $z \leftarrow E_{\theta_i}(x)$, the multi-scale states of the context
      $\boldsymbol{\Delta}_i \leftarrow (\Delta_{i,1}, \dots, \Delta_{i,K_i})$, the future window's physical intervals $k\tau_i$
      $\mathcal{L}_{\text{state}} \leftarrow \sum_{s \in \mathcal{S}} \mathcal{H}_{\beta_s}\big(f_{\phi_i}(z, \boldsymbol{\Delta}_i)^s, \text{sg}[E_{\bar{\theta}_i}(y)^s]\big)$ ▷ Eq. (6)
      sample $\Delta \sim \mathcal{U}$ and read $\mathcal{L}_{\text{interval}}$ on the patch $y_{[\Delta, \Delta+P)}$ ▷ Eq. (7)
      $\mathcal{L}_i \leftarrow \mathcal{L}_{\text{state}} + \lambda_\ell \mathcal{L}_{\text{interval}} + \mathcal{R}_i$ ▷ Eq. (4)
      update $(\vartheta_i, \phi_i)$ with Adam on $\mathcal{L}_i$; $\bar{\theta}_i \leftarrow 0.996\,\bar{\theta}_i + 0.004\,\theta_i$
    **end for**
    **Upload** $\phi_i$ and the latent prediction loss $\ell_i$ ▷ $\vartheta_i$ and raw data stay local
  **end for**
  **Tilt:** $w_i \leftarrow \pi_i \exp(\ell_i/\eta) / \sum_j \pi_j \exp(\ell_j/\eta)$ ▷ Eq. (8)
  **Aggregate:** $m^r \leftarrow \mu\, m^{r-1} + \sum_i w_i(\phi_i - \phi^{r-1})$; $\phi^r \leftarrow \phi^{r-1} + m^r$
**end for**
**Deployment:** a participating client loads $\phi^R$ and fine-tunes on its own split; a client outside the federation maps its data with the average of the private encoders and uses $\phi^R$ unchanged; quantiles are calibrated as in Appendix B.3

Table 13: Hyperparameters of each federation. All zero-shot benchmarks are evaluated on the unified-corpus federation and add no hyperparameter of their own.

| | In-Domain (Law) | In-Domain (Objectives) | TSLib | TIME | GIFT-Eval | FFTS protocol | Unified corpus |
|---|---|---|---|---|---|---|---|
| *Federated pretraining* | | | | | | | |
| Clients | 8 | 8 | 7 | 46 / 34 | 52 / 38 | 9 | 71 / 48 |
| Rounds × local steps | $60 \times 50$ | $60 \times 50$ | $60 \times 50$ | $60 \times 50$ | $60 \times 50$ | $60 \times 50$ | $60 \times 50$ |
| Context / horizon | 512/720 | 512/720 | 96/720 | $\le$ 512/864 | $\le$ 512/904 | 512/720 | $\le$ 3072/720 |
| Batch size per client | 32 | 32 | 16–32 | 16 | 16 | 128 | 32 |
| Learning rate | $4 \cdot 10^{-4}$ | $10^{-4}$ | $4 \cdot 10^{-4}$ | $4 \cdot 10^{-4}$ | $4 \cdot 10^{-4}$ | $4 \cdot 10^{-4}$ | $4 \cdot 10^{-4}$ |
| Server momentum | 0.9 | 0 | 0.9 | 0.9 | 0.9 | 0 | 0 |
| Worst-client tilt | off | — | off | on | on | on | off |
| *Downstream* | | | | | | | |
| Fine-tuning | 1,000 steps | 1,000 steps | TSLib protocol | 300 steps | 300 steps, lr $5 \cdot 10^{-5}$ | 10 epochs, patience 3 | zero-shot |
| Calibration split | — | — | validation split | last 15% of training | last 15% of training | — | — |

### B.3 Fine-Tuning and Calibration

Each participating client initializes its downstream model from its own federated encoder and decoder together with the final law and fine-tunes it on its training split. A client outside the federation, and every zero-shot benchmark, instead maps data into the state space with the average of the private encoders, computed once after the last round. The point forecast adds a private linear level branch to the law path and a linear bypass, with per-step weights fitted on the validation split. The reported FedState averages the forecasts of three federations on TSLib and six on TIME and GIFT-Eval. Probabilistic forecasts are calibrated in the manner of conformalized quantile regression (Romano et al., 2019). Each client scales its predictive quantiles by a multiplier that attains nominal coverage on a separate calibration split. Only the multiplier is shared, and the server pools the multipliers in log space with shrinkage toward the pooled value.

## C Additional Results

### C.1 In-Federation Forecasting

Table 14 reports the per-horizon TSLib results underlying Tables 3 and 4. Figure 6 decomposes the TSLib gain by dataset and horizon. Tables 15 and 16 report the full TIME and GIFT-Eval results.

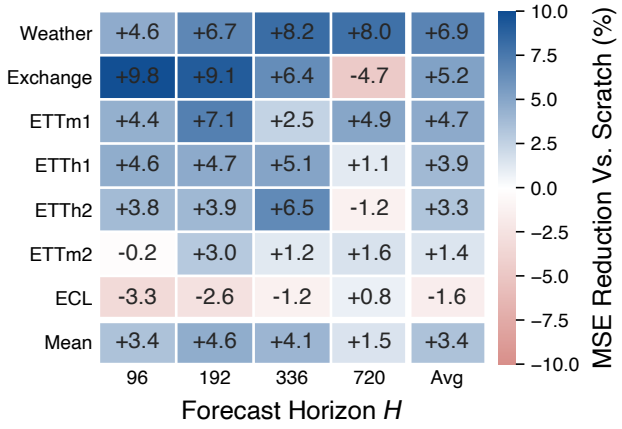



Figure 6: Relative reduction in MSE of the federated law over the same backbone trained from scratch, per dataset and horizon. Blue is a gain, pink a loss.

### C.2 Zero-Shot Transfer

Every method in the zero-shot evaluation of Section 4.3 is federated once on the corpus of Table 17 for 60 rounds of 50 local steps, at a batch of 32 and a learning rate of $4 \cdot 10^{-4}$, with aggregation weighted by the number of training windows per client. Each client uses the longest context that its training and validation splits can fill, which admits 71 clients when contexts up to 3072 are allowed and 48 when the context is fixed at 512. Each pretrained model is then applied to the six benchmarks without fine-tuning, calibration, or any other adaptation: WEATHER-5K per horizon at inputs 48 and 512 (Table 18), the 15 unseen stations per horizon (Table 19), BuildingsBench (Table 20), LiveHouse-TS per horizon (Table 21), and TSLib per dataset under the paired protocol (Table 22). On BuildingsBench every method is evaluated at one forecast origin per week, 180,790 windows, with the 168-step input repeated to the 512 steps the pretrained models take.

**Pretraining on LOTSA.** To examine whether the advantage persists on a much larger corpus, we pretrain every method on LOTSA (Woo et al., 2024). We audit all 170 LOTSA subsets (861 GiB) against the evaluation benchmarks and keep only the 54 subsets (48.1 GiB) that share neither identity nor source with any of them, which excludes every BuildingsBench evaluation set together with its

Table 14: Full results on TSLib, test MSE and MAE. The 1st row counts the cells in which a method is strictly best over the 28 dataset settings and their mean.

| | | FEDSTATE | | DLinear | | PatchTST | | iTrans. | | TimesNet | | Pooled | | FFTS | | FeDaL | | Scratch | | FedAvgM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dataset** | *H* | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| ETTh1 | 96 | **0.378** | **0.397** | 0.385 | **0.397** | 0.388 | 0.398 | 0.384 | 0.402 | 0.426 | 0.436 | 0.389 | 0.405 | 0.382 | 0.404 | 0.393 | 0.408 | 0.399 | 0.409 | 0.392 | 0.399 |
| | 192 | **0.414** | **0.418** | 0.438 | 0.431 | 0.438 | 0.427 | 0.437 | 0.432 | 0.489 | 0.471 | 0.456 | 0.444 | 0.425 | 0.432 | 0.463 | 0.449 | 0.460 | 0.452 | 0.440 | 0.426 |
| | 336 | **0.439** | **0.435** | 0.489 | 0.465 | 0.474 | 0.446 | 0.478 | 0.452 | 0.521 | 0.490 | 0.494 | 0.469 | 0.460 | 0.454 | 0.503 | 0.466 | 0.504 | 0.477 | 0.483 | 0.448 |
| | 720 | **0.466** | 0.473 | 0.504 | 0.500 | 0.511 | 0.479 | 0.493 | 0.482 | 0.573 | 0.527 | 0.482 | 0.478 | 0.516 | 0.504 | 0.495 | 0.480 | 0.529 | 0.509 | 0.483 | **0.471** |
| ETTh2 | 96 | **0.270** | **0.336** | 0.320 | 0.371 | 0.290 | 0.339 | 0.306 | 0.353 | 0.320 | 0.365 | 0.310 | 0.355 | 0.292 | 0.339 | 0.309 | 0.354 | 0.297 | 0.350 | 0.289 | 0.338 |
| | 192 | **0.328** | **0.377** | 0.399 | 0.411 | 0.370 | 0.389 | 0.382 | 0.400 | 0.443 | 0.432 | 0.382 | 0.405 | 0.375 | 0.395 | 0.383 | 0.399 | 0.382 | 0.397 | 0.375 | 0.390 |
| | 336 | **0.389** | **0.420** | 0.477 | 0.475 | 0.426 | 0.432 | 0.426 | 0.432 | 0.458 | 0.454 | 0.529 | 0.477 | 0.422 | 0.436 | 0.420 | 0.431 | 0.431 | 0.439 | 0.415 | 0.425 |
| | 720 | 0.423 | 0.448 | 0.726 | 0.611 | 0.430 | 0.446 | **0.419** | **0.439** | 0.459 | 0.464 | 0.470 | 0.467 | 0.432 | 0.449 | 0.424 | 0.443 | 0.432 | 0.448 | 0.422 | 0.440 |
| ETTm1 | 96 | **0.304** | **0.353** | 0.344 | 0.371 | 0.322 | 0.360 | 0.338 | 0.372 | 0.330 | 0.370 | 0.340 | 0.365 | 0.333 | 0.367 | 0.319 | 0.358 | 0.334 | 0.367 | 0.353 | 0.374 |
| | 192 | **0.342** | **0.375** | 0.381 | 0.392 | 0.363 | 0.382 | 0.386 | 0.395 | 0.383 | 0.399 | 0.389 | 0.390 | 0.373 | 0.393 | 0.360 | 0.380 | 0.385 | 0.390 | 0.390 | 0.391 |
| | 336 | **0.383** | **0.399** | 0.418 | 0.423 | 0.393 | 0.405 | 0.418 | 0.416 | 0.416 | 0.418 | 0.390 | 0.402 | 0.405 | 0.413 | 0.397 | 0.405 | 0.402 | 0.410 | 0.423 | 0.412 |
| | 720 | **0.440** | **0.429** | 0.474 | 0.454 | 0.454 | 0.439 | 0.495 | 0.458 | 0.499 | 0.459 | 0.471 | 0.454 | 0.458 | 0.444 | 0.464 | 0.445 | 0.473 | 0.448 | 0.485 | 0.448 |
| ETTm2 | 96 | 0.178 | 0.266 | 0.187 | 0.283 | 0.176 | 0.261 | 0.186 | 0.267 | 0.191 | 0.267 | 0.201 | 0.277 | **0.175** | **0.259** | 0.180 | 0.263 | 0.179 | 0.264 | 0.182 | 0.265 |
| | 192 | **0.243** | 0.307 | 0.275 | 0.352 | **0.243** | 0.306 | 0.261 | 0.316 | 0.257 | 0.310 | 0.272 | 0.326 | **0.243** | **0.302** | 0.250 | 0.311 | 0.246 | 0.307 | 0.246 | 0.304 |
| | 336 | **0.296** | **0.341** | 0.383 | 0.425 | 0.310 | 0.350 | 0.320 | 0.354 | 0.314 | 0.344 | 0.354 | 0.372 | 0.302 | 0.343 | 0.309 | 0.350 | 0.306 | 0.345 | 0.306 | 0.342 |
| | 720 | **0.374** | **0.392** | 0.552 | 0.521 | 0.408 | 0.406 | 0.417 | 0.408 | 0.427 | 0.412 | 0.493 | 0.452 | 0.404 | 0.401 | 0.428 | 0.428 | 0.406 | 0.401 | 0.407 | 0.398 |
| ECL | 96 | 0.178 | 0.262 | 0.195 | 0.278 | 0.170 | 0.259 | **0.166** | 0.256 | 0.182 | 0.286 | **0.166** | **0.252** | 0.205 | 0.308 | 0.193 | 0.276 | 0.168 | **0.252** | 0.194 | 0.272 |
| | 192 | 0.183 | 0.268 | 0.194 | 0.280 | 0.179 | 0.267 | 0.178 | 0.266 | 0.206 | 0.306 | 0.176 | 0.261 | 0.214 | 0.316 | 0.197 | 0.282 | **0.175** | **0.260** | 0.197 | 0.277 |
| | 336 | 0.199 | 0.284 | 0.207 | 0.297 | 0.195 | 0.283 | 0.195 | 0.284 | 0.257 | 0.349 | 0.193 | 0.278 | 0.227 | 0.326 | 0.213 | 0.299 | **0.192** | **0.277** | 0.212 | 0.292 |
| | 720 | 0.240 | 0.318 | 0.242 | 0.329 | 0.236 | 0.316 | 0.236 | 0.316 | 0.299 | 0.372 | **0.233** | **0.312** | 0.269 | 0.357 | 0.257 | 0.334 | **0.233** | **0.312** | 0.254 | 0.325 |
| Weather | 96 | **0.164** | 0.216 | 0.197 | 0.259 | 0.179 | 0.221 | 0.180 | 0.221 | 0.166 | **0.215** | 0.181 | 0.221 | 0.174 | 0.217 | 0.180 | 0.220 | 0.176 | **0.215** | 0.195 | 0.233 |
| | 192 | **0.197** | **0.244** | 0.236 | 0.295 | 0.223 | 0.259 | 0.227 | 0.260 | 0.223 | 0.264 | 0.223 | 0.258 | 0.223 | 0.258 | 0.228 | 0.262 | 0.224 | 0.256 | 0.240 | 0.270 |
| | 336 | **0.252** | **0.288** | 0.291 | 0.347 | 0.277 | 0.298 | 0.282 | 0.300 | 0.278 | 0.302 | 0.278 | 0.296 | 0.275 | 0.297 | 0.283 | 0.300 | 0.283 | 0.298 | 0.292 | 0.306 |
| | 720 | **0.325** | **0.340** | 0.347 | 0.386 | 0.353 | 0.347 | 0.357 | 0.350 | 0.353 | 0.350 | 0.375 | 0.356 | 0.351 | 0.345 | 0.359 | 0.351 | 0.362 | 0.350 | 0.364 | 0.353 |
| Exchange | 96 | 0.085 | 0.203 | 0.087 | 0.208 | 0.085 | 0.205 | 0.086 | 0.207 | 0.108 | 0.238 | 0.123 | 0.246 | 0.085 | 0.203 | **0.060** | **0.164** | 0.090 | 0.209 | 0.083 | 0.199 |
| | 192 | 0.177 | 0.296 | 0.156 | 0.294 | 0.179 | 0.302 | 0.175 | 0.299 | 0.210 | 0.333 | 0.202 | 0.320 | 0.175 | 0.298 | **0.148** | **0.261** | 0.182 | 0.305 | 0.174 | 0.295 |
| | 336 | 0.337 | 0.420 | **0.276** | 0.393 | 0.331 | 0.417 | 0.322 | 0.411 | 0.377 | 0.451 | 0.425 | 0.474 | 0.351 | 0.429 | 0.307 | **0.385** | 0.337 | 0.421 | 0.324 | 0.410 |
| | 720 | 0.859 | 0.698 | **0.750** | **0.659** | 0.865 | 0.700 | 0.837 | 0.690 | 0.954 | 0.738 | 1.202 | 0.799 | 0.849 | 0.694 | 0.816 | 0.668 | 0.844 | 0.690 | 0.838 | 0.689 |
| Mean | | **0.3165** | **0.3572** | 0.3546 | 0.3895 | 0.3310 | 0.3621 | 0.3352 | 0.3656 | 0.3614 | 0.3865 | 0.3643 | 0.3790 | 0.3355 | 0.3708 | 0.3335 | 0.3633 | 0.3368 | 0.3664 | 0.3378 | 0.3640 |
| 1st | | **33** | | 3 | | 0 | | 2 | | 0 | | 0 | | 3 | | 5 | | 4 | | 1 | |

Table 15: Full results on TIME, $\text{MASE}_n$ and $\text{CRPS}_n$ on each dataset. FFTS federates the context-unified subset, so the datasets outside it are marked with a dash.

| **Dataset** | FEDSTATE | | FedAvgM | | FedProx | | FedBN | | FedRep | | FFTS | | Scratch | | FeDaL | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** | **$\text{MASE}_n$** | **$\text{CRPS}_n$** |
| Australia_Solar | **0.839** | 0.855 | 0.857 | 0.870 | 0.848 | 0.853 | 0.844 | 0.847 | 0.845 | **0.843** | 1.491 | 1.409 | 0.875 | 0.875 | 5.192 | 1.876 |
| Auto_Production_SF | **1.010** | **1.084** | 1.174 | 1.409 | 1.170 | 1.402 | 1.167 | 1.243 | 1.173 | 1.257 | — | — | 1.162 | 1.383 | — | — |
| CPHL | 0.834 | 0.783 | **0.820** | **0.763** | 0.842 | 0.807 | 0.870 | 0.802 | 0.870 | 0.806 | 0.882 | 3.007 | 0.880 | 0.821 | 0.437 | 3.957 |
| Coastal_T_S | **0.808** | **0.646** | 0.823 | 0.659 | 0.836 | 0.679 | 0.822 | 0.666 | 0.834 | 0.661 | 1.085 | 1.297 | 0.862 | 0.708 | 2.703 | 1.482 |
| Commodity_Import | **1.072** | **1.193** | 1.224 | 1.311 | 1.144 | 1.246 | 1.299 | 1.567 | 1.234 | 1.497 | 1.345 | 5.516 | 1.198 | 1.274 | 0.619 | 0.343 |
| Crypto | **1.017** | **0.965** | 1.572 | 1.359 | 1.163 | 1.081 | 1.536 | 1.372 | 1.330 | 1.248 | 2.180 | 2.065 | 1.787 | 1.551 | 0.505 | 1.224 |
| ECDC_COVID | **1.233** | **1.171** | 1.558 | 1.226 | 1.771 | 1.543 | 1.376 | 1.409 | 1.477 | 1.492 | 1.325 | 1.700 | 1.910 | 3.065 | 1.724 | 0.335 |
| EWELD_Load | **0.737** | 0.566 | 0.765 | **0.554** | 0.765 | 0.605 | 0.750 | 0.556 | 0.761 | 0.556 | 1.683 | 1.497 | 0.788 | 0.715 | 4.358 | 2.485 |
| Finland_Traffic | 0.727 | **0.608** | 0.722 | 0.614 | 0.766 | 0.624 | 0.720 | 0.617 | 0.760 | 0.629 | 0.687 | 0.762 | 0.804 | 0.668 | 0.669 | 0.790 |
| Global_Price | **0.913** | **0.988** | 1.170 | 1.273 | 1.209 | 1.320 | 1.359 | 1.460 | 1.411 | 1.532 | 2.503 | 2.690 | 1.374 | 1.509 | 1.175 | 2.576 |
| JOLTS | 1.640 | 1.945 | **1.440** | 1.774 | 1.465 | 1.796 | 1.444 | 1.777 | 1.441 | **1.768** | 2.025 | 2.540 | 1.463 | 1.824 | 3.293 | 3.690 |
| Job_Claims | **1.351** | **1.303** | 1.431 | 1.431 | 1.755 | 1.614 | 1.424 | 1.325 | 1.481 | 1.412 | — | — | 1.430 | 1.442 | — | — |
| MetroPT-3 | **0.945** | **0.822** | 1.157 | 0.868 | 0.981 | 0.859 | 1.183 | 0.891 | 1.095 | 0.861 | 1.258 | 15.484 | 1.160 | 0.901 | NaN | 2.635 |
| NE_China_Wind | **0.840** | **0.722** | 1.007 | 0.784 | 0.933 | 0.827 | 0.982 | 0.827 | 0.949 | 0.806 | 0.896 | 1.642 | 0.887 | 0.788 | 0.617 | 3.583 |
| Oil_Price | 1.514 | 1.437 | 1.297 | 1.233 | 1.786 | 1.676 | 1.287 | **1.214** | **1.268** | 1.215 | 1.904 | 5.658 | 2.001 | 1.885 | 1.148 | 0.952 |
| Online_Retail_2_UCI | **0.645** | **0.460** | 0.671 | 0.559 | 0.665 | 0.479 | 0.672 | 0.478 | 0.662 | 0.474 | — | — | 0.664 | 0.567 | — | — |
| OpenElectricity_NEM | 0.704 | 0.713 | **0.674** | **0.670** | 0.724 | 0.757 | 0.705 | 0.687 | 0.693 | 0.694 | 1.520 | 1.271 | 0.725 | 0.756 | 1.490 | 0.301 |
| Port_Activity | 0.701 | 0.398 | 0.744 | 0.426 | 0.730 | 0.424 | 0.738 | 0.426 | 0.739 | 0.427 | 0.600 | 0.241 | 0.755 | 0.441 | 0.770 | 0.270 |
| SG_Carpark | 0.809 | **0.747** | 0.907 | 0.817 | 0.903 | 0.809 | 0.850 | 0.776 | 0.866 | 0.789 | 0.789 | 0.809 | 1.031 | 0.901 | 0.712 | 1.016 |
| SG_PM25 | **0.754** | **0.731** | 0.766 | 0.747 | 0.762 | 0.740 | 0.763 | 0.733 | 0.760 | 0.735 | 0.778 | 0.811 | 0.779 | 0.752 | 0.376 | 0.448 |
| SG_Weather | **0.985** | 0.644 | 1.009 | 0.653 | 1.031 | 0.656 | 1.008 | 0.645 | 1.035 | **0.640** | 1.051 | 0.893 | 1.141 | 0.703 | 0.636 | 0.124 |
| Smart_Manufacturing | 0.837 | **0.877** | 0.842 | 0.888 | 0.883 | 0.918 | **0.831** | 0.908 | 0.840 | 0.899 | 0.978 | 0.948 | 0.967 | 0.938 | 1.720 | 0.834 |
| Supply_Chain_Customer | **0.350** | **0.268** | 0.355 | 0.272 | 0.358 | 0.274 | 0.355 | 0.271 | 0.353 | 0.270 | 0.670 | 0.591 | 0.377 | 0.291 | 0.676 | 0.675 |
| Supply_Chain_Location | **0.405** | **0.279** | 0.408 | 0.285 | 0.408 | 0.286 | 0.410 | 0.286 | 0.409 | 0.285 | 0.652 | 0.559 | 0.413 | 0.290 | 0.519 | 0.568 |
| US_Labor | **1.018** | **1.230** | 1.139 | 1.488 | 1.121 | 1.350 | 1.162 | 1.377 | 1.249 | 1.476 | 3.463 | 1.452 | 1.155 | 1.509 | 7.069 | 1.022 |
| US_Term_Structure | 1.289 | 1.153 | **1.100** | 1.103 | 1.316 | 1.195 | 1.108 | 1.075 | 1.137 | **1.069** | 2.264 | 9.142 | 1.685 | 1.453 | 11.386 | 1.100 |
| Uncertainty_1M | 1.268 | 1.598 | 1.208 | 1.525 | 1.046 | **1.327** | 1.085 | 1.335 | 1.195 | 1.526 | 0.582 | 9.435 | 1.956 | 2.467 | 0.677 | 8.996 |
| Vehicle_Sales | **1.172** | **1.547** | 1.193 | 1.743 | 1.260 | 1.737 | 1.300 | 2.005 | 1.265 | 1.631 | 3.046 | 4.993 | 1.415 | 2.219 | 0.921 | 0.634 |
| Vehicle_Supply | 1.717 | 1.981 | 1.824 | 2.481 | 1.684 | 1.954 | 1.741 | 2.001 | **1.643** | **1.935** | 2.898 | 4.079 | 1.729 | 2.246 | 2.140 | 1.405 |
| WUI_Global | 1.321 | 1.695 | **1.156** | **1.471** | 1.231 | 1.765 | 1.254 | 1.627 | 1.236 | 1.588 | 0.813 | 4.360 | 1.198 | 1.634 | 0.217 | 7.291 |
| Water_Quality_Darwin | **0.786** | **0.596** | 0.804 | 0.625 | 0.806 | 0.612 | 0.798 | 0.604 | 0.794 | 0.603 | 1.593 | 3.781 | 0.826 | 0.658 | 1.965 | 1.067 |
| azure2019_U | **0.768** | **0.369** | 0.867 | 0.424 | 0.857 | 0.416 | 0.864 | 0.423 | 0.857 | 0.415 | 0.820 | 1.243 | 0.860 | 0.416 | 0.628 | 1.163 |
| current_velocity | **0.750** | **0.746** | 0.782 | 0.765 | 0.764 | 0.769 | 0.775 | 0.766 | 0.757 | 0.760 | 1.152 | 64.739 | 0.778 | 0.796 | 1.265 | 15.607 |
| epf_electricity_price | **0.732** | **0.694** | 0.734 | 0.699 | 0.760 | 0.708 | 0.752 | 0.704 | 0.748 | 0.701 | 0.735 | 0.766 | 0.772 | 0.730 | 0.300 | 0.451 |
| Median | **0.814** | **0.731** | 0.825 | 0.751 | 0.842 | 0.757 | 0.838 | 0.750 | 0.837 | 0.757 | 1.014 | — | 0.868 | 0.795 | 0.980 | 1.876 |
| 1st | **22** | **22** | 5 | 4 | 0 | 1 | 1 | 1 | 2 | 5 | 4 | 1 | | | | |

Table 16: Full results on GIFT-Eval, $\text{MASE}_n$ and $\text{WQL}_n$ on each of the 97 configurations. FFTS is scored on the 38 configurations of the context-unified subset.

| Configuration | FEDSTATE | | FedAvgM | | FedProx | | FedBN | | FedRep | | FFTS | | FeDaL | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\text{MASE}_n$ | $\text{WQL}_n$ | $\text{MASE}_n$ | $\text{WQL}_n$ | $\text{MASE}_n$ | $\text{WQL}_n$ | $\text{MASE}_n$ | $\text{WQL}_n$ | $\text{MASE}_n$ | $\text{WQL}_n$ | $\text{MASE}_n$ | $\text{WQL}_n$ | $\text{MASE}_n$ | $\text{WQL}_n$ |
| bitbrains_fast_storage/5T/long | 1.074 | 0.596 | 1.122 | 0.619 | 1.251 | 0.614 | 1.134 | **0.577** | **1.069** | **0.577** | 0.969 | 0.790 | 1.168 | 0.858 |
| bitbrains_fast_storage/5T/medium | **1.041** | 0.576 | 1.090 | 0.628 | 1.211 | 0.621 | 1.100 | **0.564** | 1.051 | 0.583 | 0.992 | 0.735 | 1.168 | 0.778 |
| bitbrains_fast_storage/5T/short | 1.052 | 0.427 | 1.084 | 0.444 | 1.083 | **0.424** | **0.999** | 0.455 | 1.008 | 0.428 | 0.903 | 0.539 | 1.164 | 0.647 |
| bitbrains_fast_storage/H/short | 1.145 | 0.936 | 1.139 | 0.903 | 1.137 | 0.931 | 1.122 | 1.122 | **1.118** | **0.854** | 1.248 | 1.317 | 1.463 | 1.257 |
| bitbrains_rnd/5T/long | **1.085** | 0.582 | 1.106 | 0.612 | 1.170 | 0.591 | 1.126 | 0.604 | **1.085** | **0.568** | 1.084 | 0.665 | 1.128 | 0.688 |
| bitbrains_rnd/5T/medium | **1.069** | **0.596** | 1.085 | 0.630 | 1.127 | 0.625 | 1.086 | 0.631 | 1.070 | 0.597 | 1.065 | 0.643 | 1.108 | 0.736 |
| bitbrains_rnd/5T/short | 1.162 | 0.479 | 1.172 | 0.509 | 1.184 | 0.489 | 1.136 | 0.486 | **1.130** | **0.474** | 1.100 | 0.605 | 1.292 | 0.686 |
| bitbrains_rnd/H/short | 1.058 | 0.747 | **1.048** | 0.761 | 1.065 | 0.708 | 1.051 | **0.669** | 1.051 | 0.708 | 1.112 | 0.774 | 1.172 | 0.872 |
| bizitobs_application/10S/long | 1.038 | 1.101 | **1.010** | 1.113 | 1.043 | 0.955 | 1.045 | **0.933** | 1.118 | 1.052 | 1.703 | 1.361 | 1.570 | 1.400 |
| bizitobs_application/10S/medium | **0.985** | 1.110 | 1.029 | 1.170 | 1.034 | 1.169 | 1.006 | **1.081** | 1.048 | 1.123 | 1.915 | 1.304 | 1.647 | 1.309 |
| bizitobs_application/10S/short | 0.802 | 0.634 | 0.792 | 0.598 | 0.800 | 0.572 | **0.758** | **0.495** | 0.844 | 0.618 | 2.217 | 0.925 | 1.793 | 1.032 |
| bizitobs_l2c/5T/long | **0.701** | 0.782 | 0.788 | 0.899 | 0.743 | **0.780** | 0.775 | 0.823 | 0.765 | 0.843 | 0.694 | 0.831 | 0.813 | 1.007 |
| bizitobs_l2c/5T/medium | 0.667 | 0.759 | 0.714 | 0.834 | 0.717 | 0.779 | **0.649** | **0.694** | 0.697 | 0.764 | 0.707 | 0.852 | 0.810 | 0.946 |
| bizitobs_l2c/5T/short | 0.354 | 0.367 | 0.365 | 0.398 | 0.372 | 0.402 | **0.329** | **0.335** | 0.337 | 0.344 | 0.321 | 0.356 | 0.462 | 0.508 |
| bizitobs_l2c/H/long | 0.767 | 0.530 | **0.708** | **0.503** | 0.787 | 0.551 | 0.712 | **0.503** | 0.763 | 0.529 | — | — | — | — |
| bizitobs_l2c/H/medium | **0.390** | **0.329** | 0.398 | 0.334 | 0.408 | 0.340 | 0.406 | 0.337 | 0.394 | 0.330 | 0.375 | 0.334 | 0.387 | 0.340 |
| bizitobs_l2c/H/short | 0.479 | **0.476** | 0.482 | 0.491 | **0.478** | 0.491 | 0.486 | 0.520 | 0.483 | 0.485 | 0.436 | 0.486 | 0.469 | 0.520 |
| bizitobs_service/10S/long | **1.088** | **0.992** | 1.142 | 1.056 | 1.163 | 1.045 | 1.133 | 1.020 | 1.167 | 1.015 | 1.568 | 1.190 | 1.606 | 1.173 |
| bizitobs_service/10S/medium | 1.097 | 0.879 | 1.127 | 0.959 | **1.079** | **0.845** | 1.098 | 0.902 | 1.104 | 0.922 | 1.531 | 1.120 | 1.639 | 1.160 |
| bizitobs_service/10S/short | 0.929 | 0.552 | 0.989 | 0.599 | 1.044 | 0.587 | 0.886 | 0.542 | **0.880** | **0.503** | 1.326 | 0.924 | 1.789 | 1.145 |
| car_parts/M/short | 0.968 | 0.770 | 0.963 | 0.759 | **0.956** | **0.738** | 0.959 | 0.748 | 0.967 | 0.768 | — | — | — | — |
| covid_deaths/D/short | 2.375 | 2.848 | **2.059** | **2.395** | 2.065 | 2.396 | 2.072 | 2.415 | 2.373 | 2.843 | — | — | — | — |
| electricity/15T/long | 0.868 | 0.789 | **0.867** | 0.792 | 0.912 | 0.822 | 0.878 | 0.796 | 0.874 | **0.788** | 1.057 | 1.155 | 1.196 | 1.350 |
| electricity/15T/medium | **0.897** | 0.824 | 0.905 | 0.822 | 0.943 | 0.856 | 0.902 | 0.824 | 0.905 | **0.814** | 1.009 | 1.117 | 1.128 | 1.254 |
| electricity/15T/short | 0.797 | 0.767 | 0.793 | 0.771 | **0.787** | 0.762 | 0.789 | **0.748** | 0.797 | 0.759 | 0.900 | 1.048 | 1.065 | 1.232 |
| electricity/D/short | 0.779 | 0.589 | 0.776 | 0.603 | **0.770** | 0.581 | 0.780 | **0.578** | 0.771 | 0.585 | 0.810 | 0.741 | 0.948 | 0.860 |
| electricity/H/long | 0.882 | 0.688 | **0.866** | 0.688 | 0.977 | 0.764 | 0.928 | 0.735 | **0.875** | **0.680** | 1.014 | 0.947 | 1.211 | 1.166 |
| electricity/H/medium | 0.862 | **0.671** | **0.859** | 0.685 | 0.912 | 0.711 | 0.867 | 0.679 | 0.872 | 0.681 | 1.043 | 1.026 | 1.285 | 1.282 |
| electricity/H/short | 0.816 | **0.770** | **0.807** | 0.777 | 0.818 | 0.783 | 0.809 | 0.786 | 0.819 | 0.784 | 0.901 | 0.952 | 1.262 | 1.403 |
| electricity/W/short | 1.073 | 1.144 | **1.014** | **1.064** | 1.052 | 1.140 | 1.038 | 1.105 | 1.073 | 1.148 | — | — | — | — |
| ett1/15T/long | **1.013** | **0.919** | 1.041 | 0.982 | 1.038 | 0.952 | 1.025 | 0.943 | 1.045 | 0.965 | 0.957 | 0.886 | 1.038 | 0.958 |
| ett1/15T/medium | **1.045** | **1.019** | 1.066 | 1.111 | 1.143 | 1.151 | 1.123 | 1.090 | 1.081 | 1.101 | 0.964 | 0.980 | 1.053 | 1.059 |
| ett1/15T/short | **0.848** | **0.848** | 0.858 | 0.901 | 0.885 | 0.900 | 0.916 | 0.969 | 0.955 | 0.959 | 0.901 | 1.003 | 1.018 | 1.080 |
| ett1/D/short | **1.906** | **1.405** | 2.190 | 1.447 | 2.243 | 1.808 | 2.211 | 1.806 | 2.209 | 1.676 | 1.320 | 1.035 | 1.122 | 1.066 |
| ett1/H/long | **1.095** | **0.747** | 1.246 | 0.888 | 1.421 | 0.929 | 1.283 | 0.931 | 1.175 | 0.819 | 0.994 | 0.713 | 1.060 | 0.782 |
| ett1/H/medium | **0.990** | **0.782** | 1.080 | 0.902 | 1.232 | 0.911 | 1.094 | 0.820 | 1.089 | 0.892 | 0.887 | 0.767 | 0.968 | 0.852 |
| ett1/H/short | 1.010 | 1.020 | **1.006** | **1.003** | 1.100 | 1.009 | 1.131 | 1.067 | 1.151 | 1.117 | 0.917 | 0.956 | 1.145 | 1.210 |
| ett1/W/short | 0.785 | 0.973 | 0.765 | 0.909 | **0.748** | **0.902** | 0.763 | 0.910 | 0.780 | 0.969 | — | — | — | — |
| ett2/15T/long | 0.953 | **0.722** | **0.948** | 0.727 | 0.969 | 0.757 | 0.953 | 0.728 | 1.004 | 0.788 | 0.938 | 0.756 | 0.964 | 0.782 |
| ett2/15T/medium | 0.957 | **0.815** | **0.954** | 0.822 | 1.021 | 0.875 | 1.014 | 0.866 | 0.975 | 0.818 | 0.872 | 0.805 | 0.908 | 0.839 |
| ett2/15T/short | **0.800** | 0.814 | 0.821 | 0.815 | 0.851 | 0.862 | 0.836 | 0.827 | 0.806 | **0.783** | 0.772 | 0.793 | 0.847 | 0.868 |
| ett2/D/short | 3.387 | 1.567 | 3.580 | 1.621 | 3.796 | 1.864 | 3.039 | **1.488** | **3.028** | 1.544 | 2.482 | 1.537 | 3.535 | 1.703 |
| ett2/H/long | 1.419 | 0.773 | **1.355** | **0.711** | 1.380 | 0.794 | 1.446 | 0.808 | 1.545 | 0.816 | 1.018 | 0.652 | 1.126 | 0.726 |
| ett2/H/medium | **1.085** | **0.717** | 1.170 | 0.761 | 1.132 | 0.727 | 1.172 | 0.766 | 1.234 | 0.808 | 0.885 | 0.684 | 0.981 | 0.764 |
| ett2/H/short | 0.921 | 0.835 | **0.913** | **0.825** | 0.934 | 0.844 | 0.955 | 0.849 | 0.949 | 0.866 | 0.868 | 0.895 | 1.005 | 1.050 |
| ett2/W/short | 2.178 | 1.098 | **1.852** | **0.983** | 2.203 | 1.105 | 2.062 | 1.054 | 2.190 | 1.096 | — | — | — | — |
| hierarchical_sales/D/short | **0.720** | **0.366** | 0.756 | 0.385 | 0.731 | 0.371 | 0.725 | 0.368 | **0.720** | 0.367 | 0.675 | 0.360 | 0.705 | 0.372 |
| hierarchical_sales/W/short | 0.853 | 0.588 | 0.879 | 0.610 | 0.877 | 0.607 | 0.868 | 0.604 | **0.852** | **0.587** | — | — | — | — |
| hospital/M/short | 0.979 | 1.189 | 0.915 | **1.067** | **0.908** | 1.069 | 0.915 | 1.077 | 0.973 | 1.178 | — | — | — | — |
| jena_weather/10T/long | 0.987 | 0.325 | **0.976** | 0.328 | 1.016 | **0.324** | 1.041 | 0.331 | 1.077 | 0.342 | 0.908 | 0.311 | 0.946 | 0.340 |
| jena_weather/10T/medium | **0.943** | **0.342** | 0.956 | 0.360 | 0.991 | 0.388 | 1.079 | 0.384 | 1.057 | 0.361 | 0.996 | 0.334 | 1.051 | 0.378 |
| jena_weather/10T/short | **0.605** | 0.375 | 0.625 | 0.385 | 0.646 | 0.397 | 0.607 | **0.347** | 0.620 | 0.393 | 0.880 | 0.387 | 1.141 | 0.503 |
| jena_weather/D/short | 0.989 | 0.512 | 0.951 | **0.497** | **0.948** | 0.500 | 0.967 | 0.502 | 0.982 | 0.512 | — | — | — | — |
| jena_weather/H/long | **0.894** | **0.204** | 0.925 | 0.222 | 0.991 | 0.260 | 0.914 | 0.210 | 0.964 | 0.239 | 0.822 | 0.217 | 0.856 | 0.246 |
| jena_weather/H/medium | 1.032 | 0.213 | 1.043 | 0.221 | 1.095 | 0.229 | **0.945** | **0.207** | 1.135 | 0.232 | 1.042 | 0.225 | 1.078 | 0.256 |
| jena_weather/H/short | 0.932 | 0.428 | **0.915** | **0.423** | 0.961 | 0.427 | 0.954 | **0.423** | 0.992 | 0.440 | 1.014 | 0.424 | 1.159 | 0.515 |
| kdd_cup_2018/D/short | **0.746** | **0.755** | 0.794 | 0.772 | 0.787 | 0.776 | 0.767 | 0.768 | 0.748 | **0.755** | — | — | — | — |
| kdd_cup_2018/H/long | **0.771** | **0.483** | 0.776 | 0.485 | 0.836 | 0.526 | 0.842 | 0.542 | 0.817 | 0.511 | 0.767 | 0.587 | 0.799 | 0.611 |
| kdd_cup_2018/H/medium | 0.791 | **0.596** | **0.790** | 0.599 | 0.817 | 0.625 | 0.802 | 0.605 | 0.804 | 0.620 | 0.745 | 0.690 | 0.783 | 0.719 |
| kdd_cup_2018/H/short | 0.758 | **0.770** | **0.747** | 0.775 | 0.780 | 0.804 | 0.802 | 0.821 | 0.772 | 0.785 | 0.711 | 0.854 | 0.780 | 0.934 |
| loop_seattle/5T/long | **0.882** | **0.826** | 0.892 | 0.831 | 0.927 | 0.849 | 0.905 | 0.837 | 0.895 | 0.831 | 0.867 | 0.868 | 0.897 | 0.898 |
| loop_seattle/5T/medium | **0.891** | **0.830** | 0.907 | 0.847 | 0.926 | 0.857 | 0.910 | 0.847 | 0.905 | 0.845 | 0.912 | 0.917 | 0.947 | 0.953 |
| loop_seattle/5T/short | **0.802** | **0.669** | 0.828 | 0.711 | 0.838 | 0.710 | 0.831 | 0.699 | 0.841 | 0.706 | 0.818 | 0.729 | 0.926 | 0.836 |
| loop_seattle/D/short | 0.716 | 0.559 | 0.620 | 0.519 | **0.583** | **0.501** | 0.602 | 0.509 | 0.711 | 0.557 | — | — | — | — |
| loop_seattle/H/long | **0.677** | **0.414** | 0.685 | **0.414** | 0.743 | 0.445 | 0.697 | 0.420 | 0.682 | 0.417 | 0.723 | 0.460 | 0.810 | 0.518 |
| loop_seattle/H/medium | 0.768 | 0.520 | **0.765** | **0.513** | 0.818 | 0.542 | 0.778 | 0.532 | 0.770 | 0.529 | 0.751 | 0.528 | 0.826 | 0.586 |
| loop_seattle/H/short | 0.811 | 0.760 | 0.812 | 0.758 | 0.811 | 0.758 | **0.801** | **0.751** | 0.808 | 0.758 | 0.821 | 0.792 | 0.979 | 0.954 |
| m4_daily/D/short | **1.069** | **1.031** | 1.123 | 1.091 | 1.127 | 1.085 | 1.164 | 1.103 | 1.206 | 1.150 | 1.335 | 1.531 | 1.989 | 2.190 |
| m4_hourly/H/short | **0.950** | 0.731 | 1.004 | 0.731 | 1.055 | **0.700** | 1.082 | 0.718 | 1.085 | 0.702 | 3.109 | 1.213 | 4.381 | 1.858 |
| m4_monthly/M/short | 0.887 | **0.879** | **0.838** | 0.912 | 0.850 | 0.913 | 0.863 | 0.946 | 0.894 | 0.917 | 1.045 | 1.214 | 1.576 | 1.826 |
| m4_quarterly/Q/short | 2.427 | 1.783 | **1.708** | **1.372** | 1.892 | 1.476 | 1.909 | 1.479 | 2.433 | 1.785 | — | — | — | — |
| m4_weekly/W/short | 1.036 | 0.895 | 1.013 | 0.861 | 0.978 | 0.866 | **0.955** | **0.840** | 1.041 | 0.912 | 1.388 | 1.207 | 2.227 | 1.564 |
| m4_yearly/A/short | 2.559 | 2.355 | **1.932** | **1.784** | 2.139 | 1.981 | 2.088 | 1.925 | 2.557 | 2.352 | — | — | — | — |
| m_dense/D/short | **0.467** | **0.334** | 0.528 | 0.413 | 0.505 | 0.387 | 0.484 | 0.367 | 0.472 | 0.356 | 0.575 | 0.522 | 0.765 | 0.682 |
| m_dense/H/long | 0.664 | 0.398 | 0.659 | **0.393** | 0.704 | 0.429 | 0.657 | 0.402 | **0.652** | 0.397 | 0.771 | 0.559 | 0.896 | 0.663 |
| m_dense/H/medium | 0.588 | 0.414 | **0.581** | **0.406** | 0.644 | 0.452 | 0.594 | 0.417 | 0.594 | 0.421 | 0.687 | 0.583 | 0.828 | 0.720 |
| m_dense/H/short | 0.694 | 0.644 | **0.680** | **0.626** | 0.701 | 0.649 | 0.706 | 0.661 | 0.696 | 0.655 | 0.668 | 0.739 | 0.897 | 1.037 |
| restaurant/D/short | 0.818 | 0.483 | 0.780 | 0.463 | **0.761** | **0.454** | 0.768 | 0.458 | 0.823 | 0.485 | — | — | — | — |
| saugeen/D/short | 1.261 | 1.051 | 1.309 | 1.100 | **1.230** | **1.029** | 1.258 | 1.060 | 1.241 | 1.046 | 0.900 | 0.722 | 0.990 | 0.799 |
| saugeen/M/short | 0.923 | 1.023 | 0.919 | **1.020** | 1.035 | 1.158 | 0.990 | 1.109 | **0.917** | 1.028 | 0.930 | 0.993 | 0.882 | 0.937 |
| saugeen/W/short | 0.826 | **0.817** | 0.869 | 0.867 | 0.909 | 0.912 | 0.898 | 0.900 | **0.818** | 0.823 | 0.660 | 0.633 | 0.763 | 0.737 |
| solar/10T/long | 1.252 | 0.759 | **1.251** | **0.753** | 1.437 | 0.861 | 1.376 | 0.838 | 1.268 | 0.759 | 1.346 | 0.764 | 1.429 | 0.804 |
| solar/10T/medium | 1.117 | 0.737 | 1.121 | 0.743 | 1.176 | 0.767 | **1.106** | **0.728** | 1.131 | 0.742 | 1.280 | 0.826 | 1.409 | 0.887 |
| solar/10T/short | **1.261** | **0.899** | 1.289 | 0.971 | 1.372 | 0.963 | 1.393 | 0.978 | 1.351 | 0.961 | 1.260 | 1.004 | 1.412 | 1.076 |
| solar/D/short | **0.881** | 0.586 | 0.899 | **0.582** | 0.914 | 0.586 | 0.908 | 0.586 | 0.893 | 0.588 | — | — | — | — |
| solar/H/long | **1.212** | **0.498** | 1.222 | 0.501 | 1.429 | 0.563 | 1.264 | 0.514 | 1.252 | 0.507 | 1.155 | 0.496 | 1.456 | 0.628 |
| solar/H/medium | 1.267 | 0.533 | **1.240** | **0.526** | 1.479 | 0.602 | 1.303 | 0.554 | 1.310 | 0.553 | 1.279 | 0.565 | 1.626 | 0.725 |
| solar/H/short | 1.139 | 0.793 | **1.126** | **0.781** | 1.213 | 0.842 | 1.216 | 0.849 | 1.150 | 0.805 | 1.149 | 0.839 | 1.592 | 1.170 |
| solar/W/short | 0.679 | 0.658 | 0.794 | 0.766 | 0.786 | 0.750 | 0.721 | 0.694 | **0.673** | **0.656** | — | — | — | — |
| sz_taxi/15T/long | **0.950** | **0.599** | 0.966 | 0.608 | 1.015 | 0.638 | 1.012 | 0.641 | 1.007 | 0.639 | 0.999 | 0.660 | 1.005 | 0.662 |
| sz_taxi/15T/medium | **0.861** | **0.605** | 0.867 | 0.620 | 0.926 | 0.650 | 0.902 | 0.640 | 0.896 | 0.639 | 0.899 | 0.672 | 0.923 | 0.688 |
| sz_taxi/15T/short | **0.760** | **0.694** | 0.773 | 0.709 | 0.773 | 0.708 | 0.770 | 0.706 | 0.762 | 0.697 | 0.746 | 0.718 | 0.778 | 0.746 |
| sz_taxi/H/short | **0.798** | **0.670** | 0.804 | 0.679 | 0.816 | 0.686 | **0.798** | 0.672 | 0.805 | 0.674 | 0.831 | 0.703 | 0.883 | 0.745 |
| temperature_rain/D/short | 0.804 | 0.524 | 0.845 | 0.531 | 0.825 | 0.502 | 0.806 | 0.491 | **0.792** | **0.488** | 0.724 | 0.544 | 0.805 | 0.584 |
| us_births/D/short | **0.277** | **0.263** | 0.287 | 0.282 | **0.277** | 0.272 | 0.283 | 0.274 | 0.278 | 0.265 | 0.230 | 0.228 | 0.321 | 0.318 |
| us_births/M/short | 2.556 | 2.333 | 2.095 | 1.932 | **2.020** | **1.879** | 2.115 | 1.949 | 2.517 | 2.306 | — | — | — | — |
| us_births/W/short | 0.850 | 1.005 | 0.804 | 0.951 | 0.729 | 0.873 | **0.720** | **0.855** | 0.864 | 1.035 | 0.647 | 0.782 | 0.792 | 0.958 |
| 1st | **37** | **37** | 28 | 22 | 13 | 11 | 10 | 19 | 13 | 13 | | | | |

Table 17: The unified federated corpus, where a client is one (dataset, sampling frequency) pair; the right panel counts clients per frequency. Candidates overlapping any evaluation set are excluded.

| Source | Clients | Series | Observations |
|---|---|---|---|
| TIME | 45 | 2,182 | 13,502,756 |
| GIFT-Eval | 35 | 3,868 | 15,535,638 |
| Monash-18 | 14 | 2,110 | 4,958,900 |
| **Total** | **94** | **8,160** | **33,997,294** |

| Freq. | Cl. | Freq. | Cl. |
|---|---|---|---|
| native | 23 | W | 4 |
| H | 16 | 10T | 2 |
| D | 14 | 20T | 2 |
| 5T | 11 | B | 2 |
| M | 10 | 30T | 1 |
| 15T | 8 | Q | 1 |

Table 18: WEATHER-5K zero-shot per horizon, Overall MAE over the 5,672 stations at the benchmark's input of 48 and at 512.

| | | Ours | Federated | | | | | Own arch. | | Centralized | Naive |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Input | $H$ | FEDSTATE | FedAvg | FedAvgM | FedProx | FedBN | FedRep | FFTS | FeDaL | Pooled | S-naive |
| 512 | 24 | 16.754 | 20.339 | 20.440 | 20.206 | 20.304 | 20.454 | 15.673 | **15.078** | 23.085 | 20.594 |
| | 72 | **17.113** | 20.813 | 20.869 | 20.585 | 20.851 | 20.850 | 17.269 | 17.363 | 23.277 | 20.594 |
| | 120 | **17.227** | 20.982 | 20.994 | 20.862 | 20.940 | 20.887 | 17.787 | 18.055 | 24.436 | 20.594 |
| | 168 | **17.342** | 21.126 | 21.114 | 20.961 | 21.077 | 21.032 | 18.122 | 18.510 | 24.422 | 20.594 |
| 48 | 24 | 16.477 | 18.457 | 18.659 | 18.584 | 18.571 | 18.505 | 15.872 | **15.592** | 19.436 | 20.594 |
| | 72 | 17.973 | 20.040 | 19.941 | 19.817 | 19.969 | 19.910 | **17.807** | 17.911 | 20.696 | 20.594 |
| | 120 | **18.618** | 20.651 | 20.539 | 20.481 | 20.609 | 20.558 | 18.628 | 18.819 | 21.499 | 20.594 |
| | 168 | **19.040** | 21.117 | 20.926 | 20.921 | 21.045 | 20.963 | 19.151 | 19.413 | 21.984 | 20.594 |

Table 19: The 15 unseen stations, zero-shot from an input of 512: MASE and MAE per horizon, their mean, and CRPS. The seasonal-naive floor is not ranked.

| | Ours | | Federated | | | | | | | | | | Own arch. | | | | Centralized | | Naive |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FEDSTATE | | FedAvg | | FedAvgM | | FedProx | | FedBN | | FedRep | | FFTS | | FeDaL | | Pooled | | S-naive |
| $H$ | MASE | MAE | MASE | MAE | MASE | MAE | MASE | MAE | MASE | MAE | MASE | MAE | MASE | MAE | MASE | MAE | MASE | MAE | MASE |
| 96 | **1.035** | **0.526** | 1.419 | 0.676 | 1.425 | 0.677 | 1.411 | 0.674 | 1.428 | 0.681 | 1.424 | 0.678 | 1.086 | 0.536 | 1.098 | 0.542 | 1.662 | 0.767 | 1.286 |
| 192 | **1.053** | **0.533** | 1.449 | 0.686 | 1.446 | 0.684 | 1.435 | 0.682 | 1.450 | 0.688 | 1.453 | 0.687 | 1.141 | 0.555 | 1.199 | 0.579 | 1.674 | 0.770 | 1.286 |
| 336 | **1.078** | **0.541** | 1.484 | 0.697 | 1.474 | 0.692 | 1.471 | 0.693 | 1.486 | 0.698 | 1.484 | 0.696 | 1.201 | 0.576 | 1.323 | 0.623 | 1.775 | 0.807 | 1.286 |
| 720 | **1.146** | **0.560** | 1.574 | 0.721 | 1.565 | 0.716 | 1.564 | 0.717 | 1.582 | 0.724 | 1.567 | 0.718 | 1.333 | 0.614 | 1.463 | 0.663 | 1.944 | 0.867 | 1.286 |
| Mean | **1.078** | **0.540** | 1.481 | 0.695 | 1.477 | 0.693 | 1.470 | 0.692 | 1.486 | 0.698 | 1.482 | 0.695 | 1.190 | 0.570 | 1.270 | 0.602 | 1.764 | 0.803 | 1.286 |
| **CRPS** | **0.521** | | 0.549 | | 0.548 | | 0.547 | | 0.551 | | 0.549 | | 1.191 | | — | | 0.631 | | — |

Table 20: BuildingsBench zero-shot, per-building median CVRMSE (%) over 4,020 buildings and 180,790 windows, overall and per building type.

| | | Ours | Federated | | | | | Own arch. | | Centralized | Naive |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Buildings | FEDSTATE | FedAvg | FedAvgM | FedProx | FedBN | FedRep | FFTS | FeDaL | Pooled | S-naive |
| Median | All buildings | 51.4 | 53.6 | 55.8 | 52.2 | 56.3 | 57.7 | 39.4 | **32.6** | 61.4 | 29.4 |
| | Commercial | 31.1 | 31.0 | 33.7 | 30.6 | 33.1 | 34.7 | 21.9 | **14.9** | 36.8 | 12.4 |
| | Residential | 88.2 | 97.3 | 98.5 | 95.3 | 97.3 | 99.6 | 78.9 | **78.3** | 103.4 | 94.1 |

Table 21: LiveHouse-TS zero-shot, MASE per horizon; $H$=24 is the setting of Table 8.

| | Ours | Federated | | | | | Own arch. | | Centralized |
|---|---|---|---|---|---|---|---|---|---|
| $H$ | FEDSTATE | FedAvg | FedAvgM | FedProx | FedBN | FedRep | FFTS | FeDaL | Pooled |
| 12 | **2.607** | 3.426 | 3.496 | 3.658 | 3.426 | 3.688 | 4.471 | 3.336 | — |
| 24 | **3.179** | 3.937 | 3.982 | 4.121 | 3.939 | 4.146 | 4.774 | 3.844 | 3.600 |
| 48 | **5.143** | 5.563 | 5.640 | 5.726 | 5.563 | 5.742 | 6.171 | 5.552 | — |

Table 22: TSLib zero-shot per dataset under the paired protocol of Shi et al. (2025), test MSE and MAE.

| | | Ours | | Federated | | | | | | | | | | Own arch. | | | | Centralized | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | FEDSTATE | | FedAvg | | FedAvgM | | FedProx | | FedBN | | FedRep | | FFTS | | FeDaL | | Pooled | |
| Dataset | $H$ | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ETTh1 | 96 | **0.353** | **0.378** | 0.407 | 0.425 | 0.414 | 0.426 | 0.402 | 0.425 | 0.410 | 0.423 | 0.510 | 0.475 | 0.523 | 0.481 | 0.444 | 0.445 | 0.386 | 0.410 |
| | 192 | **0.390** | **0.418** | 0.440 | 0.459 | 0.459 | 0.471 | 0.430 | 0.453 | 0.433 | 0.447 | 0.556 | 0.523 | 0.556 | 0.512 | 0.528 | 0.489 | 0.763 | 0.618 |
| | 336 | 0.487 | 0.500 | 0.498 | 0.512 | 0.563 | 0.559 | 0.514 | 0.522 | **0.480** | **0.494** | 0.670 | 0.612 | 0.618 | 0.564 | 0.672 | 0.557 | 0.760 | 0.651 |
| | 720 | 0.598 | 0.587 | 0.600 | 0.600 | 0.675 | 0.647 | 0.693 | 0.643 | **0.569** | **0.575** | 0.797 | 0.701 | 0.739 | 0.657 | 0.656 | **0.575** | 1.718 | 0.996 |
| ETTh2 | 96 | **0.238** | **0.304** | 0.284 | 0.349 | 0.284 | 0.352 | 0.306 | 0.362 | 0.288 | 0.350 | 0.291 | 0.360 | 0.294 | 0.359 | 0.328 | 0.379 | 0.297 | 0.354 |
| | 192 | **0.308** | **0.361** | 0.345 | 0.396 | 0.356 | 0.407 | 0.357 | 0.403 | 0.345 | 0.394 | 0.360 | 0.411 | 0.351 | 0.397 | 0.368 | 0.406 | 0.453 | 0.463 |
| | 336 | 0.483 | 0.497 | 0.438 | 0.477 | 0.548 | 0.550 | 0.478 | 0.498 | 0.433 | 0.473 | 0.544 | 0.548 | 0.439 | 0.478 | **0.398** | **0.428** | 0.949 | 0.707 |
| | 720 | 0.833 | 0.657 | 0.701 | 0.623 | 0.953 | 0.730 | 0.791 | 0.655 | 0.700 | 0.620 | 0.939 | 0.731 | 0.726 | 0.635 | **0.446** | **0.465** | 1.963 | 1.031 |
| ETTm1 | 96 | 0.457 | **0.415** | 0.497 | 0.453 | 0.523 | 0.467 | **0.455** | 0.441 | 0.500 | 0.450 | 0.590 | 0.496 | 0.582 | 0.489 | 0.648 | 0.499 | 0.468 | 0.425 |
| | 192 | **0.463** | **0.428** | 0.511 | 0.467 | 0.534 | 0.478 | 0.473 | 0.449 | 0.509 | 0.458 | 0.590 | 0.505 | 0.588 | 0.496 | 0.724 | 0.542 | 0.911 | 0.621 |
| | 336 | **0.486** | **0.449** | 0.517 | 0.473 | 0.539 | 0.490 | 0.499 | 0.468 | 0.519 | 0.464 | 0.608 | 0.523 | 0.604 | 0.509 | 0.835 | 0.595 | 0.792 | 0.576 |
| | 720 | **0.518** | **0.478** | 0.537 | 0.500 | 0.548 | 0.510 | 0.541 | 0.501 | 0.532 | 0.488 | 0.649 | 0.559 | 0.641 | 0.547 | 1.029 | 0.670 | 0.937 | 0.642 |
| ETTm2 | 96 | **0.160** | **0.248** | 0.196 | 0.285 | 0.202 | 0.293 | 0.199 | 0.287 | 0.197 | 0.284 | 0.212 | 0.303 | 0.213 | 0.301 | 0.251 | 0.322 | 0.190 | 0.277 |
| | 192 | **0.232** | **0.308** | 0.259 | 0.333 | 0.275 | 0.347 | 0.255 | 0.330 | 0.257 | 0.330 | 0.289 | 0.357 | 0.267 | 0.336 | 0.310 | 0.360 | 0.394 | 0.424 |
| | 336 | **0.265** | **0.340** | 0.291 | 0.361 | 0.305 | 0.376 | 0.305 | 0.370 | 0.290 | 0.359 | 0.319 | 0.385 | 0.297 | 0.363 | 0.370 | 0.395 | 0.396 | 0.427 |
| | 720 | **0.316** | **0.369** | 0.349 | 0.398 | 0.361 | 0.408 | 0.362 | 0.409 | 0.348 | 0.397 | 0.376 | 0.418 | 0.365 | 0.406 | 0.492 | 0.464 | 0.517 | 0.500 |
| Weather | 96 | **0.163** | **0.222** | 0.195 | 0.255 | 0.203 | 0.266 | 0.205 | 0.264 | 0.194 | 0.254 | 0.207 | 0.271 | 0.206 | 0.269 | 0.221 | 0.265 | 0.174 | **0.222** |
| | 192 | **0.223** | **0.277** | 0.244 | 0.293 | 0.259 | 0.312 | 0.252 | 0.301 | 0.243 | 0.291 | 0.268 | 0.320 | 0.247 | 0.298 | 0.277 | 0.309 | 0.456 | 0.429 |
| | 336 | **0.284** | **0.323** | 0.291 | 0.328 | 0.321 | 0.353 | 0.302 | 0.339 | 0.288 | 0.324 | 0.323 | 0.358 | 0.287 | 0.327 | 0.331 | 0.348 | 0.567 | 0.495 |
| | 720 | **0.305** | **0.339** | 0.321 | 0.351 | 0.342 | 0.371 | 0.343 | 0.369 | 0.320 | 0.349 | 0.337 | 0.371 | 0.314 | 0.351 | 0.403 | 0.393 | 0.546 | 0.490 |
| Exchange | 96 | 0.123 | 0.242 | 0.187 | 0.317 | 0.203 | 0.331 | 0.190 | 0.323 | 0.183 | 0.314 | 0.235 | 0.359 | 0.250 | 0.370 | 0.358 | 0.423 | **0.104** | **0.229** |
| | 192 | **0.358** | **0.445** | 0.401 | 0.487 | 0.481 | 0.538 | 0.438 | 0.507 | 0.398 | 0.484 | 0.569 | 0.585 | 0.524 | 0.551 | 0.540 | 0.542 | 0.911 | 0.698 |
| | 336 | 0.650 | 0.618 | 0.596 | 0.597 | 0.732 | 0.676 | 0.617 | 0.604 | **0.578** | **0.586** | 0.853 | 0.729 | 0.723 | 0.662 | 0.782 | 0.663 | 1.870 | 1.033 |
| | 720 | 1.075 | 0.829 | 0.806 | 0.713 | 1.167 | 0.897 | 0.863 | 0.733 | **0.765** | **0.692** | 1.183 | 0.918 | 0.859 | 0.755 | 1.542 | 0.897 | 1.635 | 1.015 |

pretraining corpus Buildings-900K. From this pool we select 24 subsets across energy, transport, cloud, web, sales, and health, capping any single subset at 15% and any single sampling period at 50% of the points so that no single source dominates (Table 23), and split each subset into clients by partitioning its series. Of the 307 clients built, the 185 that contain a series of at least 1,232 steps, one input of 512 plus one horizon of 720, participate, with 375,072 series and $9.4 \times 10^8$ points in total, 27.7× the unified corpus. Every method is trained for 60 rounds of 50 local steps at a batch of 32, with 48 clients sampled per round. Table 24 shows that FEDSTATE achieves the best result on three of the four benchmarks while uploading 30.9% of the full model per round. It lowers the WEATHER-5K Overall MAE to 20.084 at an input of 512 and to 19.032 at 48, against 20.887 and 19.873 for the strongest baselines, the unseen-station MASE to 1.403 against 1.479, and the BuildingsBench median CVRMSE to 50.1 against 54.8. This advantage is consistent across settings: FEDSTATE is best in all eight WEATHER-5K settings and in every per-horizon MASE and MAE of the unseen stations (Table 25).

Table 23: The 24 LOTSA subsets and the points allocated to each; 375,072 series and 942M points remain after client admission.

| Subset | Domain | Period | Points (M) | Subset | Domain | Period | Points (M) |
|---|---|---|---|---|---|---|---|
| wind_farms_with_missing | Energy | 1 min | 153.0 | SHMETRO | Transport | 15 min | 4.8 |
| Q-TRAFFIC | Transport | 15 min | 153.0 | project_tycho | Health | 1 week | 1.3 |
| extended_web_traffic_with_missing | Web | 1 day | 153.0 | uber_tlc_hourly | Transport | 1 hour | 1.1 |
| borg_cluster_data_2011 | Cloud | 5 min | 153.0 | HZMETRO | Transport | 15 min | 0.36 |
| favorita_sales | Sales | 1 day | 132.7 | cdc_fluview_ilinet | Health | 1 week | 0.30 |
| azure_vm_traces_2017 | Cloud | 5 min | 108.1 | cdc_fluview_who_nrevss | Health | 1 week | 0.16 |
| m5 | Sales | 1 day | 55.6 | covid_mobility | Mobility | 1 day | 0.14 |
| kdd2022 | Energy | 10 min | 44.8 | tourism_monthly | Tourism | 1 month | 0.10 |
| wiki-rolling_nips | Web | 1 day | 38.8 | favorita_transactions | Sales | 1 day | 0.08 |
| solar_power | Energy | 4 s | 7.0 | sunspot_with_missing | Nature | 1 day | 0.07 |
| wind_power | Energy | 4 s | 7.0 | uber_tlc_daily | Transport | 1 day | 0.05 |
| BEIJING_SUBWAY_30MIN | Transport | 30 min | 5.4 | vehicle_trips_with_missing | Transport | 1 day | 0.03 |

Table 24: Zero-shot results after federated pretraining on LOTSA, with the same backbone and budget for every method. Upload is relative to the full model, and Pooled trains one model on the pooled corpus with the same number of sample windows. Full results are in Table 25.

| | | Ours | Federated | | | | | Centralized |
|---|---|---|---|---|---|---|---|---|
| **Benchmark** | **Metric** | **FEDSTATE** | **FedAvg** | **FedAvgM** | **FedProx** | **FedBN** | **FedRep** | **Pooled** |
| Upload | — | 30.9% | 100% | 100% | 100% | ≈99% | ≈70% | — |
| WEATHER-5K$_{512}$ | MAE | **20.084** | 21.066 | 20.971 | 21.271 | 21.137 | 20.887 | 26.930 |
| WEATHER-5K$_{48}$ | MAE | **19.032** | 20.364 | 19.873 | 20.393 | 20.144 | 19.941 | 22.130 |
| Unseen stations | MASE | **1.403** | 1.497 | 1.480 | 1.510 | 1.496 | 1.479 | 1.940 |
| | MAE | **0.656** | 0.702 | 0.694 | 0.706 | 0.702 | 0.694 | 0.870 |
| LiveHouse-TS | MASE | 3.876 | 3.469 | **3.346** | 3.498 | 3.465 | 3.681 | 3.545 |
| BuildingsBench | CVRMSE | **50.1** | 56.0 | 57.1 | 56.8 | 54.8 | 57.5 | 66.9 |
| 1st | | **5** | 0 | 1 | 0 | 0 | 0 | 0 |

Table 25: LOTSA zero-shot per setting: WEATHER-5K per input and horizon, the unseen stations per horizon, LiveHouse-TS per dataset at $H{=}24$, and BuildingsBench per-building median CVRMSE (%).

| | | | Ours | Federated | | | | | Centralized |
|---|---|---|---|---|---|---|---|---|---|
| **Benchmark** | **Metric** | **Setting** | **FEDSTATE** | **FedAvg** | **FedAvgM** | **FedProx** | **FedBN** | **FedRep** | **Pooled** |
| WEATHER-5K | MAE | $512 \to 24$ | **19.733** | 20.742 | 20.760 | 20.726 | 20.802 | 20.620 | 29.373 |
| | | $512 \to 72$ | **20.154** | 21.045 | 20.985 | 21.274 | 21.143 | 20.928 | 26.619 |
| | | $512 \to 120$ | **20.176** | 21.184 | 21.045 | 21.474 | 21.239 | 20.944 | 25.949 |
| | | $512 \to 168$ | **20.270** | 21.292 | 21.092 | 21.608 | 21.363 | 21.057 | 25.778 |
| | | $48 \to 24$ | **17.572** | 18.838 | 18.430 | 18.962 | 18.570 | 18.407 | 21.430 |
| | | $48 \to 72$ | **18.909** | 20.271 | 19.736 | 20.328 | 20.062 | 19.888 | 21.921 |
| | | $48 \to 120$ | **19.611** | 20.964 | 20.420 | 20.913 | 20.750 | 20.541 | 22.382 |
| | | $48 \to 168$ | **20.036** | 21.383 | 20.906 | 21.367 | 21.193 | 20.930 | 22.787 |
| Unseen stations | MASE | $H{=}96$ | **1.345** | 1.434 | 1.424 | 1.455 | 1.436 | 1.428 | 1.908 |
| | | $H{=}192$ | **1.370** | 1.466 | 1.448 | 1.477 | 1.467 | 1.449 | 1.897 |
| | | $H{=}336$ | **1.402** | 1.499 | 1.481 | 1.511 | 1.498 | 1.478 | 1.930 |
| | | $H{=}720$ | **1.493** | 1.591 | 1.566 | 1.599 | 1.583 | 1.563 | 2.023 |
| | | Mean | **1.403** | 1.497 | 1.480 | 1.510 | 1.496 | 1.479 | 1.940 |
| | MAE | $H{=}96$ | **0.639** | 0.683 | 0.678 | 0.689 | 0.683 | 0.679 | 0.864 |
| | | $H{=}192$ | **0.647** | 0.693 | 0.685 | 0.697 | 0.693 | 0.685 | 0.857 |
| | | $H{=}336$ | **0.657** | 0.703 | 0.695 | 0.707 | 0.703 | 0.694 | 0.868 |
| | | $H{=}720$ | **0.680** | 0.728 | 0.718 | 0.731 | 0.727 | 0.717 | 0.893 |
| | | Mean | **0.656** | 0.702 | 0.694 | 0.706 | 0.702 | 0.694 | 0.870 |
| | CRPS | Mean | **0.521** | 0.556 | 0.548 | 0.558 | 0.555 | 0.548 | 0.689 |
| LiveHouse-TS | MASE | Binance BTC/USDT | 1.456 | **1.235** | 1.275 | 1.247 | 1.243 | 1.419 | 1.295 |
| | | CoinGecko Bitcoin | 1.268 | **1.067** | 1.096 | 1.083 | 1.072 | 1.240 | 1.105 |
| | | NASA POWER Shanghai | 6.192 | 5.459 | **5.385** | 5.507 | 5.462 | 5.765 | 5.462 |
| | | Shanghai weather | 2.667 | 2.360 | **2.317** | 2.347 | 2.362 | 2.398 | 2.486 |
| | | Shanghai air quality | 1.593 | 1.449 | **1.420** | 1.449 | 1.449 | 1.532 | 1.486 |
| | | KSFO observations | 1.969 | 1.451 | **1.425** | 1.490 | 1.461 | 1.837 | 1.597 |
| | | NDBC buoy 46013 | 6.195 | 5.537 | **5.407** | 5.569 | 5.549 | 5.969 | 5.670 |
| | | SF water level | 16.906 | 17.852 | 16.173 | 17.781 | 17.555 | 16.297 | **15.583** |
| | | Potomac River | 7.904 | 6.639 | **6.079** | 6.848 | 6.647 | 7.576 | 7.888 |
| | | Mean of 240 tasks | 3.876 | 3.469 | **3.346** | 3.498 | 3.465 | 3.681 | 3.545 |
| BuildingsBench | CVRMSE | All buildings | **50.1** | 56.0 | 57.1 | 56.8 | 54.8 | 57.5 | 66.9 |
| | | BDG-2 | **27.8** | 30.4 | 30.8 | 31.3 | 29.4 | 31.3 | 37.9 |
| | | Electricity | 30.1 | 36.0 | 38.7 | 34.3 | **29.5** | 38.5 | 37.4 |
| | | Borealis | **66.8** | 74.9 | 71.5 | 74.9 | 71.9 | 72.0 | 83.4 |
| | | IDEAL | **91.0** | 105.6 | 104.5 | 104.8 | 101.2 | 104.2 | 114.4 |
| | | LCL | **87.1** | 99.3 | 98.8 | 100.5 | 98.2 | 98.9 | 108.9 |
| | | SMART | **86.0** | 98.2 | 97.8 | 97.6 | 93.0 | 96.6 | 111.0 |
| | | Sceaux | **80.8** | 90.9 | 93.8 | 89.6 | 85.6 | 90.2 | 98.3 |
| | | Commercial | **29.5** | 34.6 | 36.3 | 33.6 | **29.5** | 36.4 | 37.6 |
| | | Residential | **87.5** | 100.4 | 99.7 | 101.1 | 98.3 | 99.6 | 110.1 |
| 1st | | | **28** | 2 | 7 | 0 | 2 | 0 | 1 |

### C.3 ANALYSES

Table 26 reports the per-setting results underlying Figure 5, Table 29 those underlying Table 9, and Tables 27 and 28 those underlying Figure 4(c).

**Scaling and Robustness.** A compact configuration suffices. The default law of 0.36M parameters performs on par with a law 55 times wider, the default encoder with an encoder 37 times wider, and pretraining on an eighth of the data reaches the accuracy of the full corpus (Figure 7). With half

Table 26: Hyperparameter sensitivity on TSLib. ΔMSE is relative to the default, and $\lambda_\ell$ is varied on the model with interval conditioning.

| Hyperparameter | Value | MSE | MAE | ΔMSE |
|---|---|---|---|---|
| $\lambda_\ell$ | 0.3 | **0.3316** | **0.3615** | −0.0022 |
| | 1.0 (default) | 0.3338 | 0.3630 | — |
| | 3.0 | 0.3324 | 0.3626 | −0.0014 |
| $\lambda_v$ | 0.1 | 0.3324 | 0.3616 | +0.0022 |
| | 0.5 (default) | **0.3302** | **0.3611** | — |
| | 2.0 | 0.3318 | 0.3620 | +0.0016 |
| Tilt $1/\eta$ | 0 (default, no tilt) | **0.3302** | **0.3611** | — |
| | 0.5 | 0.3318 | 0.3620 | +0.0016 |
| | 1.0 | 0.3323 | 0.3624 | +0.0021 |
| | 2.0 | 0.3326 | 0.3623 | +0.0023 |
| Latent scales $\mathcal{S}$ | fast | 0.3324 | 0.3620 | +0.0022 |
| | fast, mid | 0.3328 | 0.3621 | +0.0026 |
| | fast, mid, slow (default) | **0.3302** | **0.3611** | — |

Table 27: Partial participation on TSLib.

| Method | ETTh1 | ETTh2 | ETTm1 | ETTm2 | ECL | Weather | Exchange | All |
|---|---|---|---|---|---|---|---|---|
| Scratch | 0.473 | 0.385 | **0.399** | 0.284 | **0.192** | 0.261 | 0.363 | 0.337 |
| FEDSTATE, 50% participation | 0.460 | 0.376 | 0.403 | **0.282** | 0.203 | **0.259** | 0.358 | 0.334 |
| FEDSTATE, full participation | **0.445** | **0.375** | **0.399** | 0.283 | 0.194 | 0.261 | **0.354** | 0.330 |

Table 28: Results on the unseen client ILI. ΔMSE is the MSE with the law minus that of training from scratch.

| | | Scratch | | | | + Federated Law | | | | ΔMSE |
|---|---|---|---|---|---|---|---|---|---|---|
| Client | $H$ | MSE | MAE | CRPS | Cov80 | MSE | MAE | CRPS | Cov80 | |
| ILI | 24 | 2.737 | 1.158 | 1.097 | 0.882 | 2.409 | 1.049 | 1.068 | 0.885 | −0.328 |
| | 36 | 2.763 | 1.170 | 1.032 | 0.909 | 2.413 | 1.045 | 0.989 | 0.913 | −0.350 |
| | 48 | 2.812 | 1.168 | 1.054 | 0.902 | 2.504 | 1.046 | 1.042 | 0.899 | −0.308 |
| | 60 | 2.874 | 1.189 | 1.084 | 0.894 | 2.345 | 1.021 | 1.079 | 0.886 | −0.529 |
| | Mean | 2.796 | 1.171 | 1.067 | 0.897 | 2.418 | 1.040 | 1.044 | 0.896 | −0.379 |

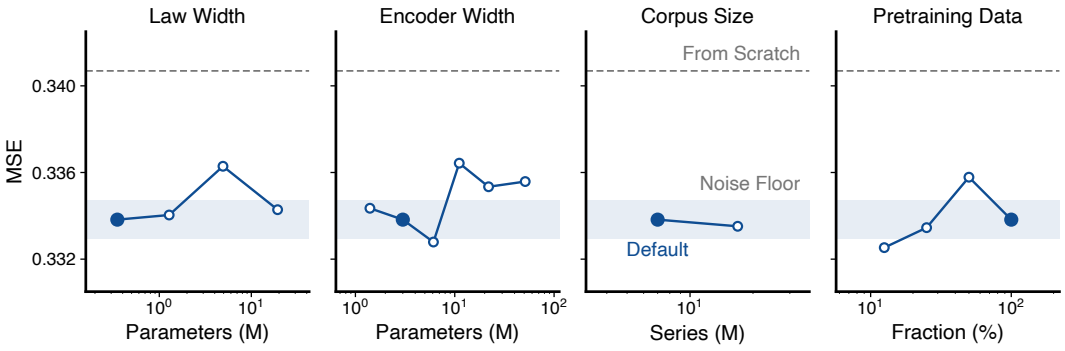



Figure 7: Scaling on TSLib, mean MSE against the width of the law and of the private encoder, the corpus size, and the fraction of pretraining data.

Table 29: Per-horizon results with different client encoders on TSLib, test MSE and MAE.

| | | Transformer | | | | TSMixer | | | | | | Mamba | | | | | | Mixed | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Scratch | | FEDSTATE | | Scratch | | FEDSTATE | | +Law | | Scratch | | FEDSTATE | | +Law | | FEDSTATE | |
| **Dataset** | *H* | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| ETTh1 | 96 | 0.399 | 0.409 | **0.385** | **0.399** | 0.397 | 0.409 | **0.382** | **0.397** | 0.386 | 0.407 | 0.401 | 0.417 | **0.381** | **0.400** | 0.391 | 0.415 | 0.379 | 0.396 |
| | 192 | 0.460 | 0.452 | **0.435** | **0.434** | 0.445 | 0.440 | **0.431** | **0.428** | 0.447 | 0.439 | 0.461 | 0.454 | **0.440** | **0.431** | 0.440 | 0.442 | 0.459 | 0.448 |
| | 336 | 0.504 | 0.477 | **0.472** | **0.453** | 0.502 | 0.474 | **0.466** | **0.442** | 0.490 | 0.467 | 0.498 | 0.473 | **0.454** | **0.442** | 0.486 | 0.465 | 0.502 | 0.468 |
| | 720 | 0.529 | 0.509 | **0.490** | **0.482** | 0.498 | 0.490 | **0.463** | **0.463** | 0.525 | 0.497 | 0.551 | 0.524 | **0.498** | **0.484** | 0.534 | 0.507 | 0.511 | 0.494 |
| ETTh2 | 96 | 0.297 | 0.350 | **0.288** | **0.337** | 0.296 | 0.349 | **0.287** | **0.335** | 0.294 | 0.346 | 0.310 | 0.357 | **0.297** | **0.347** | 0.316 | 0.363 | 0.277 | 0.332 |
| | 192 | 0.382 | 0.397 | **0.374** | **0.389** | 0.372 | 0.396 | 0.371 | **0.387** | **0.369** | 0.398 | 0.381 | 0.402 | **0.373** | **0.388** | 0.385 | 0.407 | 0.370 | 0.392 |
| | 336 | 0.431 | 0.439 | **0.416** | **0.428** | 0.430 | 0.437 | **0.413** | **0.425** | 0.417 | 0.433 | 0.429 | 0.440 | **0.418** | **0.427** | 0.435 | 0.443 | 0.405 | 0.422 |
| | 720 | 0.432 | 0.448 | **0.422** | **0.440** | 0.429 | 0.447 | **0.418** | **0.439** | 0.429 | 0.448 | 0.428 | 0.451 | **0.423** | **0.442** | 0.429 | 0.451 | 0.418 | 0.442 |
| ETTm1 | 96 | **0.334** | **0.367** | 0.337 | 0.370 | **0.316** | 0.354 | 0.317 | **0.353** | 0.322 | 0.357 | 0.334 | 0.366 | **0.329** | **0.364** | 0.332 | 0.366 | 0.342 | 0.366 |
| | 192 | 0.385 | **0.390** | **0.384** | 0.394 | **0.363** | **0.381** | 0.367 | 0.383 | 0.373 | 0.387 | 0.373 | 0.388 | 0.371 | 0.389 | **0.367** | **0.387** | 0.379 | 0.391 |
| | 336 | **0.402** | 0.410 | 0.406 | **0.408** | 0.403 | **0.409** | 0.407 | 0.411 | **0.402** | 0.409 | 0.412 | 0.415 | **0.408** | **0.413** | 0.410 | 0.413 | 0.404 | 0.412 |
| | 720 | 0.473 | **0.448** | **0.469** | 0.448 | 0.465 | 0.444 | 0.465 | 0.446 | **0.460** | **0.443** | 0.478 | 0.453 | **0.466** | **0.446** | 0.472 | 0.451 | 0.469 | 0.455 |
| ETTm2 | 96 | 0.179 | 0.264 | **0.178** | **0.261** | 0.178 | 0.261 | **0.177** | **0.260** | 0.178 | 0.262 | 0.182 | 0.265 | **0.180** | **0.263** | 0.180 | 0.264 | 0.178 | 0.261 |
| | 192 | **0.246** | 0.307 | 0.247 | **0.305** | 0.249 | 0.310 | **0.242** | **0.302** | 0.248 | 0.309 | 0.251 | 0.316 | **0.246** | **0.305** | 0.250 | 0.313 | 0.249 | 0.310 |
| | 336 | 0.306 | 0.345 | **0.304** | **0.342** | 0.310 | 0.350 | **0.305** | **0.347** | 0.307 | 0.347 | 0.310 | 0.349 | **0.301** | **0.339** | 0.305 | 0.343 | 0.319 | 0.355 |
| | 720 | 0.406 | 0.401 | **0.403** | **0.396** | 0.407 | 0.406 | 0.400 | 0.398 | **0.399** | **0.397** | 0.410 | 0.406 | **0.400** | **0.397** | 0.411 | 0.406 | 0.407 | 0.407 |
| ECL | 96 | **0.168** | **0.252** | 0.172 | 0.256 | **0.172** | **0.256** | 0.174 | 0.257 | 0.174 | 0.258 | **0.164** | **0.249** | 0.168 | 0.253 | 0.165 | 0.250 | 0.173 | 0.257 |
| | 192 | **0.175** | **0.260** | 0.178 | 0.262 | **0.179** | **0.264** | 0.181 | 0.265 | 0.181 | 0.265 | **0.173** | **0.259** | 0.176 | 0.261 | 0.174 | 0.259 | 0.181 | 0.265 |
| | 336 | **0.192** | **0.277** | 0.194 | 0.279 | **0.195** | **0.280** | 0.197 | 0.281 | 0.197 | 0.282 | 0.190 | 0.276 | 0.192 | 0.277 | **0.190** | **0.276** | 0.196 | 0.281 |
| Weather | 96 | **0.176** | **0.215** | 0.177 | 0.215 | **0.176** | **0.214** | 0.178 | 0.216 | 0.176 | 0.215 | 0.176 | 0.214 | 0.175 | 0.214 | **0.175** | **0.213** | 0.188 | 0.223 |
| | 192 | **0.224** | **0.256** | 0.228 | 0.259 | **0.223** | **0.256** | 0.225 | 0.256 | 0.225 | 0.257 | 0.225 | 0.258 | **0.221** | **0.255** | 0.225 | 0.258 | 0.225 | 0.258 |
| | 336 | 0.283 | 0.298 | **0.281** | **0.297** | 0.282 | 0.298 | **0.280** | **0.297** | 0.281 | 0.298 | 0.282 | 0.299 | **0.280** | **0.297** | 0.280 | 0.298 | 0.292 | 0.304 |
| | 720 | 0.362 | 0.350 | **0.359** | **0.349** | 0.360 | 0.349 | **0.358** | **0.348** | 0.360 | 0.348 | 0.359 | 0.349 | **0.357** | **0.347** | 0.358 | 0.348 | 0.372 | 0.354 |
| Exchange | 96 | 0.090 | 0.209 | **0.084** | **0.200** | 0.085 | 0.203 | **0.082** | **0.198** | 0.085 | 0.202 | 0.084 | 0.203 | **0.082** | **0.198** | 0.083 | 0.201 | 0.094 | 0.210 |
| | 192 | 0.182 | 0.305 | **0.173** | **0.296** | 0.177 | 0.299 | 0.178 | 0.297 | **0.174** | **0.296** | **0.185** | **0.306** | 0.189 | 0.308 | 0.188 | 0.307 | 0.220 | 0.328 |
| | 336 | 0.337 | 0.421 | **0.320** | **0.408** | 0.341 | 0.424 | **0.332** | **0.416** | 0.342 | 0.423 | 0.345 | 0.426 | **0.327** | **0.413** | 0.345 | 0.425 | 0.332 | 0.416 |
| | 720 | 0.844 | 0.690 | **0.838** | **0.688** | 0.858 | 0.697 | **0.841** | **0.687** | 0.847 | 0.694 | 0.885 | 0.712 | **0.879** | **0.708** | 0.890 | 0.710 | 1.064 | 0.767 |
| **Avg** | | 0.341 | 0.368 | **0.334** | **0.363** | 0.337 | 0.367 | **0.331** | **0.360** | 0.337 | 0.366 | 0.344 | 0.371 | **0.334** | **0.363** | 0.341 | 0.370 | 0.348 | 0.371 |
| **Gain (%)** | | – | – | +2.0 | +1.5 | – | – | +1.9 | +1.7 | +0.3 | +0.1 | – | – | +2.6 | +2.2 | +0.7 | +0.4 | – | – |

of the clients sampled per round, the law still improves on training from scratch (Table 27), and an unseen client, ILI, lowers its MSE by $13.5\%$ when initialized from the law (Table 28).

**In-Domain Representation Learning.** Table 30 reports the per-setting results underlying Table 2. Every method is fine-tuned for $1{,}000$ steps with all parameters trainable, and Traffic and ECL are fine-tuned for three epochs instead of ten. Every objective is trained on the shared backbone. SimMTM and CoST compute their losses across a batch, and CoST keeps its momentum queue local to each client. FeDaL keeps its own server-side aggregation and its published hyperparameters.

**Noise Tolerance of the Uploads.** Table 31 applies central differential privacy to the uploads of the law and of the full model under the same clipping norm and noise multiplier, so that the cost of a given $\varepsilon$ can be compared across shared objects of different sizes. The comparison measures how much noise each shared object tolerates. The law's mean MSE stays within $0.0082$ of its noiseless value up to $\sigma{=}1$ ($\varepsilon{=}67$), while the full model under the same clipping norm already degrades at $\sigma{=}0.01$ and diverges from $\sigma{=}0.1$ onward.

**Domain Identity of the Transition.** Figure 8 embeds windows from ten domains in three ways. A nearest-neighbor classifier identifies a window's domain with $78\%$ accuracy from its log-spectrum, $83\%$ from its value increments, and $16\%$ from its latent transition direction, where chance is $17\%$. The transition direction therefore encodes no domain identity.

### C.4 CASE STUDIES

Figures 9 and 10 show probabilistic forecasts on TSLib and zero-shot forecasts on the unseen stations, and Figure 11 compares the methods of the FFTS protocol on one few-shot case.

Table 30: Full results of in-domain representation learning on TSLib, MSE/MAE on each setting.

| | | Ours | | Federated Self-Supervised | | | | | | | | | | | | Fed. Aggregation | | Centralized | | No FL | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | FEDSTATE | | TimeDART | | SimMTM | | PatchTST | | FFTS | | CoST | | TimeMAE | | FeDaL | | All Mixed | | Scratch | | No Pretrain | |
| Dataset | $H$ | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ETTh1 | 96 | **0.376** | **0.402** | 0.398 | 0.420 | 0.424 | 0.433 | 0.432 | 0.434 | 0.430 | 0.435 | 0.401 | 0.421 | 0.430 | 0.436 | 0.391 | 0.416 | 0.399 | 0.418 | 0.405 | 0.426 | 0.390 | 0.413 |
| | 192 | **0.406** | **0.421** | 0.480 | 0.463 | 0.471 | 0.463 | 0.502 | 0.464 | 0.486 | 0.467 | 0.462 | 0.458 | 0.498 | 0.474 | 0.434 | 0.443 | 0.468 | 0.453 | 0.442 | 0.447 | 0.437 | 0.442 |
| | 336 | **0.433** | **0.437** | 0.510 | 0.487 | 0.476 | 0.473 | 0.536 | 0.491 | 0.506 | 0.481 | 0.496 | 0.480 | 0.556 | 0.512 | 0.463 | 0.464 | 0.506 | 0.481 | 0.439 | 0.446 | 0.514 | 0.487 |
| | 720 | **0.474** | **0.481** | 0.637 | 0.542 | 0.515 | 0.503 | 0.545 | 0.515 | 0.524 | 0.510 | 0.635 | 0.553 | 0.630 | 0.551 | 0.591 | 0.537 | 0.613 | 0.529 | 0.549 | 0.517 | 0.597 | 0.538 |
| ETTh2 | 96 | **0.266** | **0.332** | 0.310 | 0.361 | 0.319 | 0.371 | 0.327 | 0.369 | 0.319 | 0.367 | 0.338 | 0.378 | 0.326 | 0.374 | 0.309 | 0.361 | 0.301 | 0.355 | 0.289 | 0.352 | 0.313 | 0.365 |
| | 192 | **0.336** | **0.378** | 0.447 | 0.437 | 0.448 | 0.447 | 0.419 | 0.420 | 0.407 | 0.419 | 0.446 | 0.440 | 0.430 | 0.433 | 0.401 | 0.416 | 0.409 | 0.424 | 0.385 | 0.406 | 0.409 | 0.420 |
| | 336 | **0.362** | **0.403** | 0.422 | 0.443 | 0.488 | 0.480 | 0.459 | 0.455 | 0.470 | 0.468 | 0.435 | 0.448 | 0.460 | 0.460 | 0.393 | 0.425 | 0.413 | 0.434 | 0.395 | 0.429 | 0.423 | 0.440 |
| | 720 | **0.392** | **0.430** | 0.474 | 0.469 | 0.521 | 0.510 | 0.513 | 0.492 | 0.510 | 0.497 | 0.487 | 0.490 | 0.499 | 0.499 | 0.441 | 0.461 | 0.481 | 0.479 | 0.430 | 0.457 | 0.479 | 0.488 |
| ETTm1 | 96 | 0.310 | 0.351 | 0.295 | **0.347** | 0.301 | 0.353 | 0.299 | 0.353 | 0.299 | 0.353 | 0.293 | 0.352 | 0.291 | 0.350 | 0.291 | 0.349 | 0.297 | 0.348 | 0.313 | 0.359 | **0.287** | 0.347 |
| | 192 | 0.338 | 0.376 | 0.327 | 0.373 | 0.331 | 0.375 | 0.331 | 0.376 | 0.329 | 0.375 | 0.329 | 0.373 | 0.331 | 0.375 | 0.336 | 0.377 | **0.327** | 0.374 | 0.341 | **0.371** | 0.331 | 0.375 |
| | 336 | 0.371 | **0.389** | 0.363 | 0.398 | 0.373 | 0.401 | **0.357** | 0.396 | 0.367 | 0.399 | 0.359 | 0.395 | 0.376 | 0.403 | 0.363 | 0.398 | 0.367 | 0.399 | 0.370 | 0.391 | 0.373 | 0.401 |
| | 720 | 0.423 | **0.415** | 0.437 | 0.435 | 0.438 | 0.442 | 0.431 | 0.437 | 0.437 | 0.439 | 0.451 | 0.440 | 0.431 | 0.439 | **0.418** | 0.430 | 0.440 | 0.435 | 0.421 | 0.421 | 0.437 | 0.436 |
| ETTm2 | 96 | **0.166** | **0.256** | 0.181 | 0.277 | 0.178 | 0.271 | 0.179 | 0.270 | 0.185 | 0.276 | 0.182 | 0.272 | 0.184 | 0.279 | 0.180 | 0.273 | 0.178 | 0.273 | 0.168 | 0.259 | 0.178 | 0.269 |
| | 192 | **0.218** | **0.290** | 0.238 | 0.310 | 0.245 | 0.319 | 0.243 | 0.313 | 0.249 | 0.316 | 0.240 | 0.311 | 0.257 | 0.323 | 0.243 | 0.314 | 0.242 | 0.309 | 0.240 | 0.307 | 0.245 | 0.315 |
| | 336 | 0.271 | **0.324** | 0.301 | 0.354 | 0.314 | 0.358 | 0.302 | 0.347 | 0.325 | 0.362 | 0.296 | 0.349 | 0.309 | 0.355 | 0.296 | 0.350 | 0.297 | 0.352 | **0.268** | 0.327 | 0.303 | 0.350 |
| | 720 | **0.359** | **0.381** | 0.388 | 0.410 | 0.399 | 0.412 | 0.391 | 0.409 | 0.397 | 0.413 | 0.384 | 0.402 | 0.422 | 0.427 | 0.377 | 0.400 | 0.389 | 0.410 | 0.363 | 0.386 | 0.403 | 0.413 |
| ECL | 96 | **0.133** | **0.229** | 0.150 | 0.252 | 0.149 | 0.251 | 0.149 | 0.253 | 0.150 | 0.254 | 0.148 | 0.250 | 0.146 | 0.248 | 0.163 | 0.272 | 0.150 | 0.252 | 0.138 | 0.236 | 0.153 | 0.256 |
| | 192 | **0.151** | **0.246** | 0.163 | 0.265 | 0.165 | 0.267 | 0.163 | 0.265 | 0.164 | 0.265 | 0.161 | 0.261 | 0.167 | 0.271 | 0.180 | 0.285 | 0.163 | 0.265 | 0.155 | 0.251 | 0.160 | 0.258 |
| | 336 | **0.166** | **0.261** | 0.179 | 0.278 | 0.182 | 0.281 | 0.179 | 0.281 | 0.179 | 0.280 | 0.179 | 0.280 | 0.178 | 0.277 | 0.194 | 0.299 | 0.179 | 0.277 | 0.171 | 0.268 | 0.177 | 0.278 |
| | 720 | **0.208** | **0.297** | 0.217 | 0.310 | 0.217 | 0.308 | 0.217 | 0.311 | 0.218 | 0.311 | 0.222 | 0.312 | 0.221 | 0.314 | 0.233 | 0.330 | 0.217 | 0.310 | 0.214 | 0.304 | 0.222 | 0.316 |
| Traffic | 96 | **0.379** | **0.267** | 0.440 | 0.336 | 0.433 | 0.326 | 0.438 | 0.330 | 0.436 | 0.326 | 0.431 | 0.326 | 0.425 | 0.318 | 0.465 | 0.350 | 0.439 | 0.335 | 0.384 | 0.278 | 0.433 | 0.328 |
| | 192 | **0.392** | **0.269** | 0.443 | 0.332 | 0.451 | 0.341 | 0.457 | 0.345 | 0.459 | 0.347 | 0.440 | 0.328 | 0.447 | 0.336 | 0.480 | 0.361 | 0.443 | 0.332 | 0.402 | 0.285 | 0.454 | 0.344 |
| | 336 | **0.407** | **0.282** | 0.462 | 0.347 | 0.450 | 0.331 | 0.448 | 0.330 | 0.448 | 0.330 | 0.456 | 0.340 | 0.454 | 0.338 | 0.484 | 0.362 | 0.460 | 0.344 | 0.413 | 0.292 | 0.451 | 0.336 |
| | 720 | **0.446** | **0.299** | 0.487 | 0.350 | 0.481 | 0.344 | 0.505 | 0.366 | 0.507 | 0.367 | 0.489 | 0.353 | 0.485 | 0.349 | 0.516 | 0.373 | 0.486 | 0.349 | 0.456 | 0.312 | 0.481 | 0.347 |
| Weather | 96 | 0.150 | 0.201 | 0.153 | 0.202 | 0.152 | 0.202 | 0.150 | 0.200 | 0.153 | 0.203 | 0.155 | 0.207 | 0.152 | 0.203 | 0.153 | 0.205 | 0.151 | 0.201 | **0.149** | **0.198** | 0.156 | 0.207 |
| | 192 | 0.212 | 0.257 | 0.200 | 0.247 | 0.202 | 0.249 | 0.197 | 0.246 | 0.197 | 0.247 | 0.201 | 0.249 | 0.200 | 0.249 | 0.197 | 0.246 | 0.197 | 0.245 | **0.196** | **0.244** | 0.197 | 0.246 |
| | 336 | 0.255 | 0.288 | 0.255 | 0.290 | 0.250 | 0.287 | 0.254 | 0.288 | 0.257 | 0.290 | **0.248** | 0.285 | 0.253 | 0.288 | 0.250 | **0.284** | 0.251 | 0.288 | 0.259 | 0.292 | 0.250 | 0.287 |
| | 720 | 0.324 | 0.338 | 0.323 | 0.337 | 0.324 | 0.340 | 0.333 | 0.342 | 0.333 | 0.347 | 0.323 | 0.341 | 0.331 | 0.343 | 0.325 | 0.338 | **0.321** | **0.337** | 0.329 | 0.344 | 0.325 | 0.340 |
| Exchange | 96 | **0.089** | **0.208** | 0.105 | 0.234 | 0.117 | 0.248 | 0.109 | 0.234 | 0.124 | 0.251 | 0.112 | 0.239 | 0.120 | 0.248 | 0.109 | 0.235 | 0.105 | 0.231 | 0.092 | 0.215 | 0.104 | 0.234 |
| | 192 | 0.196 | 0.315 | 0.236 | 0.353 | 0.259 | 0.367 | 0.220 | 0.337 | 0.256 | 0.360 | 0.250 | 0.362 | 0.267 | 0.378 | 0.236 | 0.353 | 0.235 | 0.352 | **0.192** | **0.314** | 0.251 | 0.359 |
| | 336 | **0.342** | **0.423** | 0.457 | 0.495 | 0.508 | 0.521 | 0.439 | 0.481 | 0.462 | 0.494 | 0.511 | 0.517 | 0.521 | 0.524 | 0.403 | 0.465 | 0.455 | 0.497 | 0.358 | 0.436 | 0.447 | 0.487 |
| | 720 | 0.928 | 0.718 | 1.206 | 0.806 | 1.192 | 0.795 | 1.258 | 0.819 | 1.174 | 0.801 | 1.194 | 0.798 | 1.075 | 0.781 | 1.058 | 0.748 | 1.126 | 0.771 | **0.917** | **0.716** | 1.193 | 0.800 |
| *BuildingsBench, per-building median CVRMSE (%)* | | | | | | | | | | | | | | | | | | | | | | | |
| **Collection** | 24 | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | | **CV-RMSE** | |
| BDG-2 | 24 | 15.92 | | 17.73 | | 18.03 | | 17.79 | | 17.83 | | 17.90 | | 18.03 | | 19.56 | | 19.15 | | **15.78** | | 18.07 | |
| Electricity | 24 | **11.20** | | 12.58 | | 13.83 | | 13.39 | | 13.57 | | 12.01 | | 13.44 | | 15.52 | | 19.48 | | 11.55 | | 12.74 | |
| LCL | 24 | **74.43** | | 77.18 | | 77.44 | | 77.43 | | 77.04 | | 76.67 | | 77.44 | | 77.97 | | 81.92 | | 75.06 | | 77.08 | |
| IDEAL | 24 | **85.52** | | 87.71 | | 88.20 | | 87.65 | | 87.89 | | 87.74 | | 87.38 | | 89.35 | | 92.58 | | 86.06 | | 86.99 | |
| Borealis | 24 | 58.94 | | **58.03** | | 59.45 | | 60.88 | | 59.04 | | 59.52 | | 58.45 | | 59.01 | | 63.11 | | 62.28 | | 58.08 | |
| SMART | 24 | **72.92** | | 83.35 | | 85.42 | | 82.34 | | 82.23 | | 81.48 | | 87.65 | | 82.09 | | 77.16 | | 80.41 | | 81.55 | |
| Sceaux | 24 | 73.93 | | 75.74 | | 78.53 | | 77.20 | | 77.01 | | 71.55 | | 77.11 | | 70.65 | | **69.56** | | 82.60 | | 73.01 | |
| 1st | | **45** | | 1 | | 0 | | 1 | | 0 | | 1 | | 0 | | 2 | | 3 | | 10 | | 1 | |

Table 31: Central differential privacy on TSLib, with the same clipping norm and noise multiplier $\sigma$ for both methods. *Diverged* marks a mean MSE above $10^3$.

| **Shared object** | $\sigma$ | $\varepsilon$ | **MSE** | **MAE** | **CRPS** | **Cov80** | **ΔMSE vs. no DP** |
|---|---|---|---|---|---|---|---|
| Transition law (FEDSTATE) | 0 | $\infty$ | **0.3338** | **0.3628** | 0.5179 | 0.964 | — |
| | 0.01 | 330115.1 | 0.3360 | 0.3640 | 0.5444 | 0.987 | +0.0022 |
| | 0.02 | 82615.1 | 0.3349 | 0.3636 | 0.5373 | 0.987 | +0.0011 |
| | 0.05 | 13315.1 | 0.3396 | 0.3649 | 0.5344 | 0.985 | +0.0058 |
| | 0.1 | 3415.1 | 0.3407 | 0.3655 | 0.5374 | 0.986 | +0.0069 |
| | 0.25 | 633.6 | 0.3410 | 0.3656 | 0.5388 | 0.986 | +0.0071 |
| | 0.5 | 194.4 | 0.3412 | 0.3652 | 0.5273 | 0.984 | +0.0074 |
| | 1 | 67.2 | 0.3421 | 0.3660 | 0.5691 | 0.991 | +0.0082 |
| Full model (FedAvgM) | 0 | $\infty$ | 0.3408 | 0.3654 | **0.5028** | 0.961 | — |
| | 0.01 | 330115.1 | 0.3729 | 0.3899 | 0.5260 | 0.981 | +0.0321 |
| | 0.02 | 82615.1 | 1.2557 | 0.7089 | 0.6867 | 0.926 | +0.9149 |
| | 0.05 | 13315.1 | 25.2158 | 2.8874 | 2.4579 | **0.715** | +24.8750 |
| | $\geq 0.1$ | $\leq 3415.1$ | *diverged* | | | | |

Table 32: Attacks on the uploads, one block per dataset. Leak is the random-guess reconstruction error over the error after gradient inversion (a value of 1.00 means nothing is recovered); the leaked attacker additionally has access to the client's encoder; MIA is the AUC of a loss-threshold membership test.

| | Raw gradients | | Adam updates | | MIA AUC | |
|---|---|---|---|---|---|---|
| **Uploaded object** | **Server** | **Leaked** | **Server** | **Leaked** | **Raw** | **Adam** |
| **ECL** | | | | | | |
| Transition law (ours) | 0.99 | 1.21 | 1.01 | 0.97 | 0.50 | 0.51 |
| Shared representation | 1.31 | 1.32 | 0.97 | 0.97 | 0.52 | 0.47 |
| Full model | 2.46 | 2.46 | 0.97 | 0.97 | 0.51 | 0.47 |
| **ETTh1** | | | | | | |
| Transition law (ours) | 1.01 | 1.21 | 1.01 | 0.98 | 0.52 | 0.52 |
| Shared representation | 1.19 | 1.22 | 0.98 | 0.98 | 0.52 | 0.51 |
| Full model | 1.41 | 1.38 | 0.98 | 0.98 | 0.52 | 0.51 |

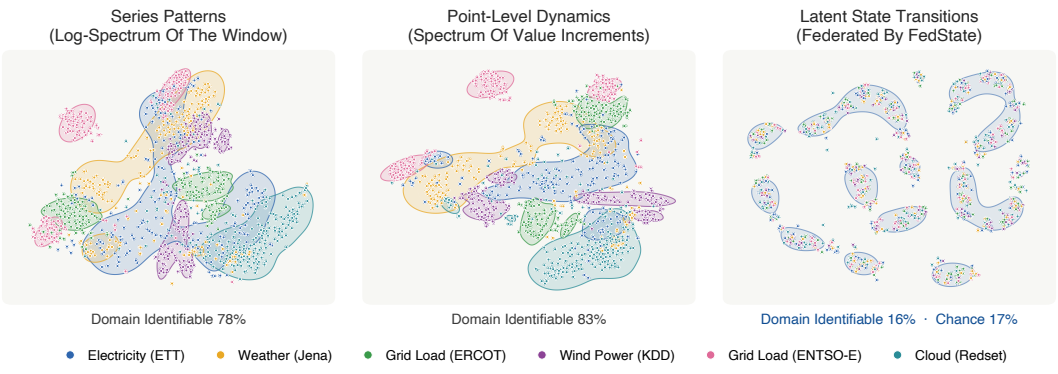



Figure 8: t-SNE of the same windows under the three embeddings. Color is the domain.

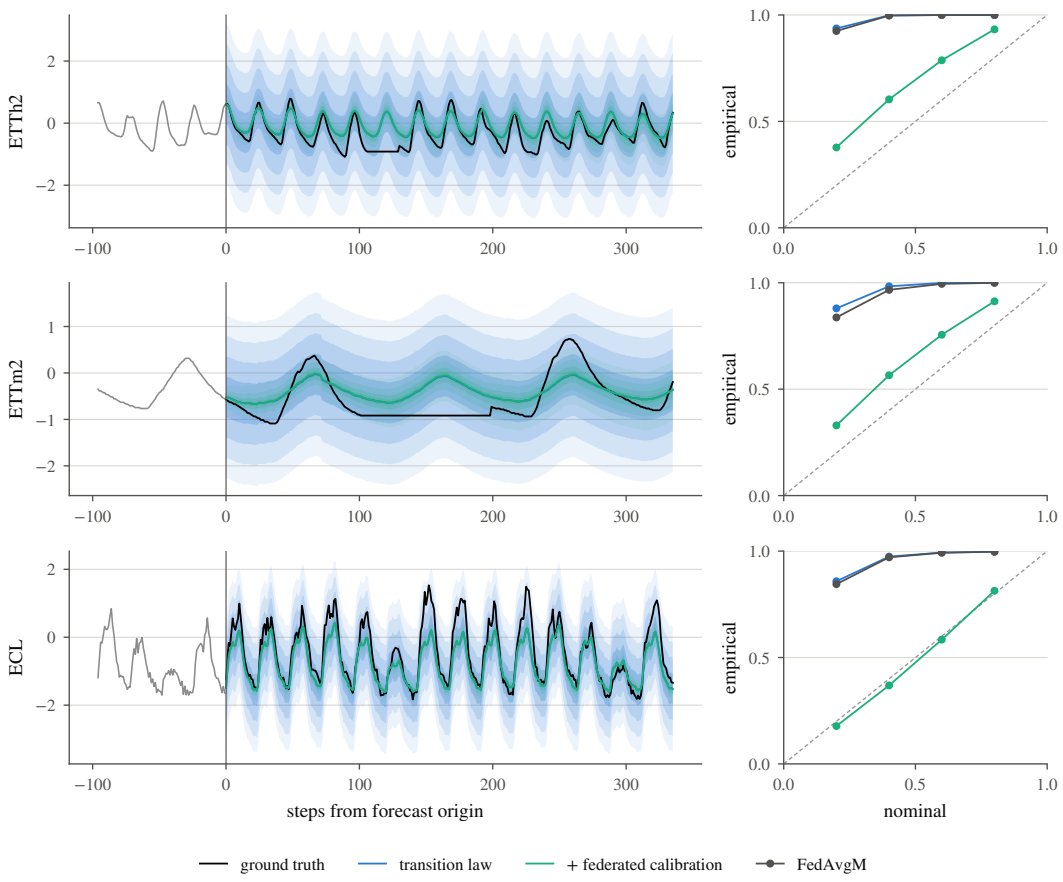



Figure 9: Probabilistic case study on TSLib at $H=336$. Left: the predictive intervals of the federated law before and after federated calibration. Right: the reliability of both.

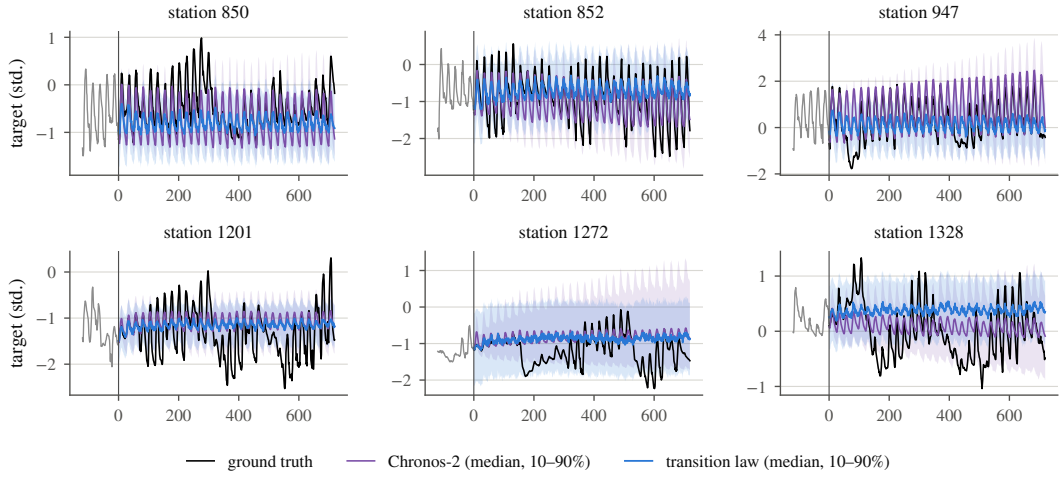



Figure 10: Zero-shot probabilistic forecasts on six unseen stations, target variable, 30 days ahead from a 512-step context. The band is the 80% interval and the black line the ground truth.

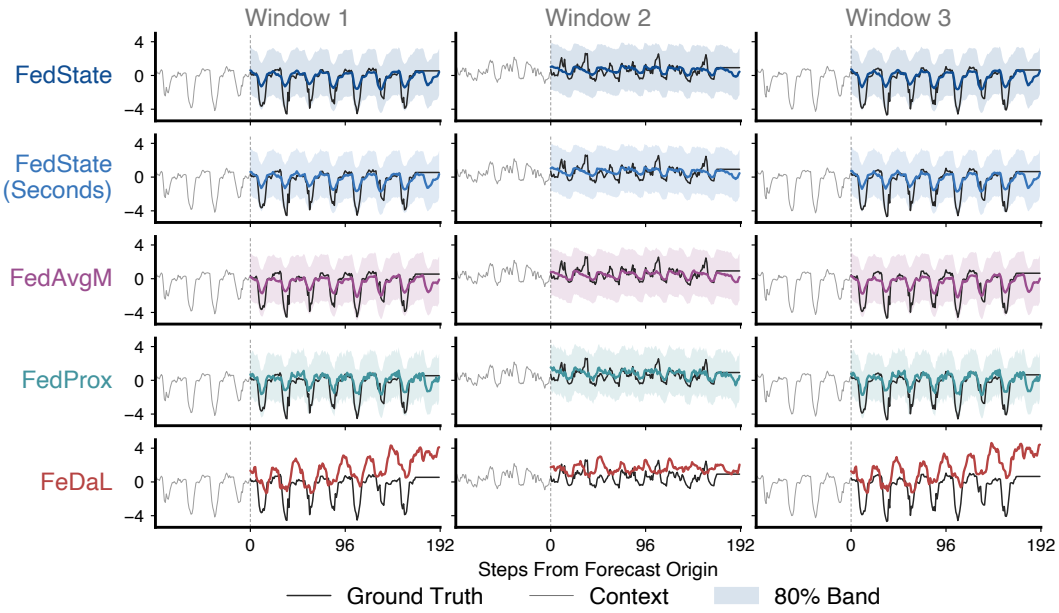



Figure 11: ETTh1 at $H{=}192$ under the FFTS protocol with $5\%$ of the training split, one row per method on three test windows. Bands are $80\%$ intervals; FeDaL produces point forecasts only.