STARTS: STATE-AWARE DYNAMIC SAMPLE REWEIGHTING FOR TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

In time series forecasting, a sample's training priority, reflected in its loss weight, can change as the model learns. Existing sample-aware methods typically use a single signal, such as noise, uncertainty, or current loss, without jointly considering current fitting state and fitting history. We introduce *StarTS*, a model-agnostic framework for state-aware dynamic sample reweighting in time series forecasting. Fitting Range Evaluation (FRE) measures difficulty relative to recent losses and penalizes samples beyond a loss-relative fitting boundary, while Mastery Memory (MM) tracks historical evidence of easy fitting. Together, they emphasize samples within the current fitting range with limited mastery evidence and downweight unusually difficult or consistently well-fitted samples. Across eight datasets and nine forecasting backbones, StarTS improves average MSE on every dataset, with average reductions of 0.6%–6.4% across backbones. Across four datasets and three distinct forecasting backbones, StarTS also achieves lower average MSE than two sample-aware training baselines. Code is avaliable at https://anonymous.4open.science/r/StarTs-0E38/.

1 INTRODUCTION

Time series forecasting is used in financial risk analysis, weather prediction, building energy management, and traffic operations (Idrees et al., 2019; Karevan & Suykens, 2020; Deb et al., 2017; Zheng & Huang, 2020). Its broad practical relevance has motivated increasingly capable forecasting architectures, from the decomposition-based linear model DLinear (Zeng et al., 2023) to the patch-based Transformer PatchTST (Nie et al., 2023). Despite these architectural differences, their standard training objectives average sample losses with equal weight. Forecasting windows differ in reliability, noise level, and predictability, and their fitting states evolve as the model learns. Uniform averaging ignores both sample heterogeneity and changing training needs.

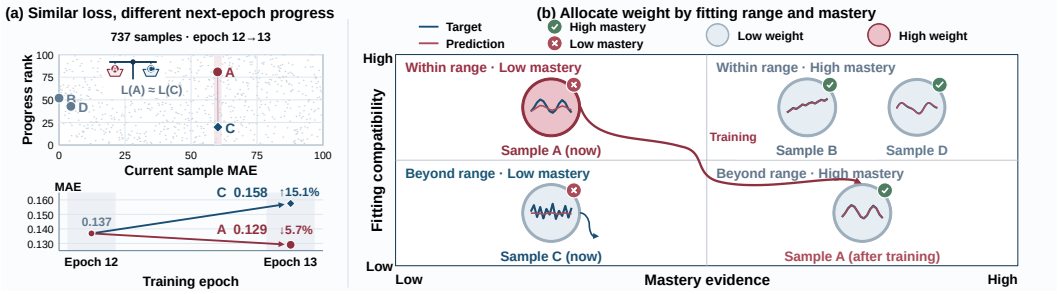

Figure 1: **Motivation for state-aware sample reweighting.** (a) Similar losses, different next-epoch changes (scatter axes: percentiles). (b) Schematic weighting by current fitting state and mastery history across training stages. Higher weights favor samples within the current fitting range that have accumulated little mastery evidence.

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.

Table 1: **Comparison of training paradigms.**

Paradigm	Representative Methods	Training Characteristics
Data-centric selection	RobustTSF (Cheng et al., 2024) Selective Learning (Fu et al., 2025)	Reliability or redundancy filtering; sample or step exclusion
Noise-aware regularization	DropoutTS (Zhong et al., 2026)	Noise-adaptive regularization; no fitting-history memory
Reference-guided valuation	Meta-Weight-Net (Shu et al., 2019) FAF (Taga et al., 2026)	Validation or reference signals; extra data or model
State-aware weighting	**StarTS (ours)**	Fitting range and mastery history; model-agnostic reweighting

Existing sample-aware training addresses these differences through three broad strategies. **Data-centric selection** removes samples or target steps judged unreliable, uncertain, or redundant, as in RobustTSF (Cheng et al., 2024), Selective Learning (Fu et al., 2025), and ST-Prune (Chen et al., 2026). **Noise-aware regularization** adapts model capacity to estimated sample noise, as in DropoutTS (Zhong et al., 2026). **Reference-guided valuation** derives weights or selections from validation feedback, predictive contribution, or an auxiliary model (Ren et al., 2018; Shu et al., 2019; Ghorbani & Zou, 2019; Yoon et al., 2020; Taga et al., 2026). Table 1 summarizes the signals and training operations that distinguish these paradigms. Overlapping forecasting windows further repeat related observations throughout training (Fu et al., 2025; Zhong et al., 2026). Yet these strategies rely on reliability, noise, uncertainty, current loss, or reference signals without jointly tracking current loss-relative difficulty and a sample's easy-fitting history. A reliable sample may be well fitted and require little further emphasis, whereas a persistently difficult one may remain beyond the model's loss-relative fitting range. Figure 1(a) shows different fitting trajectories at similar current losses, motivating training priorities that consider both current difficulty and fitting history.

We introduce *StarTS* (*state-aware dynamic sample reweighting for time series forecasting*). **A fixed training set does not imply fixed training priorities**: as fitting states evolve, training emphasis should evolve accordingly. StarTS reweights per-sample losses using two complementary modules. *Fitting Range Evaluation* (FRE) forms a loss-relative fitting boundary from moving estimates of recent loss level and variation, and increasingly downweights samples beyond the current boundary. *Mastery Memory* (MM) aggregates evidence of easy fitting across visits, distinguishing recently easy samples from those with consistently low relative loss. Together, they prioritize samples within the current fitting range with limited mastery evidence, as illustrated in Figure 1(b). StarTS reweights per-sample training losses without auxiliary models or held-out data, and applies across forecasting architectures without altering their backbones or inference procedures. StarTS also achieves lower average MSE than two sample-aware training baselines across four datasets and three forecasting backbones with distinct architectures.

Our main contributions are:

- **State-relative Loss Calibration.** We use moving loss statistics to standardize each sample's current loss against the model's recent loss distribution and derive a continuous fitting-range score that reduces emphasis on unusually difficult samples.
- **History-aware Weighting.** We combine *Fitting Range Evaluation* with per-sample *Mastery Memory* to adjust training emphasis as fitting states evolve, while preserving the forecasting architecture, base loss, and inference path.
- **Superior Performance.** StarTS improves mean MSE on all eight datasets and reduces average MSE by 0.6%–6.4% across nine backbones. It also lowers zero-shot MSE in seven of eight ETT directions and outperforms two sample-aware training baselines in average MSE.

2 Related Work

Time series forecasting. Forecasting architectures differ in how they represent temporal structure. Decomposition and linear models separate trend and seasonal components before modeling

temporal dependencies or mapping inputs directly to future values (Wu et al., 2021; Zhou et al., 2022; Zeng et al., 2023). Token-based Transformers use temporal patches or variables as attention tokens, providing distinct views of temporal and cross-variable dependence (Nie et al., 2023; Liu et al., 2024). Multiscale and dependency-oriented models capture patterns across resolutions or select patch-specific interactions (Wang et al., 2025; Hu et al., 2025). TimeRecipe (Zhao et al., 2026) evaluates forecasting architectures at the module level. These families vary in their internal representations; StarTS provides a shared sample-weighting rule across the evaluated backbones.

Sample selection and reweighting. *Difficulty-based methods* order examples through curricula or select them by current model difficulty through self-paced learning (Bengio et al., 2009; Kumar et al., 2010). Focal loss emphasizes hard examples; Co-teaching exchanges small-loss examples between two models (Lin et al., 2017; Han et al., 2018). *Gradient-based sampling* uses gradient norms to reduce update variance (Katharopoulos & Fleuret, 2018). *Reference-guided methods* use clean validation data to learn weights (Ren et al., 2018; Shu et al., 2019), estimate predictive contribution with Data Shapley (Ghorbani & Zou, 2019), or use an auxiliary model for RHO-LOSS (Mindermann et al., 2022). *History-based methods* track prediction variance, forgetting events, and confidence trajectories (Chang et al., 2017; Toneva et al., 2019; Swayamdipta et al., 2020); DIH (Zhou et al., 2020) averages difficulty over time. StarTS combines a moving loss-relative boundary with per-sample fitting-history memory to dynamically allocate training emphasis.

Selective learning for time series. RobustTSF (Cheng et al., 2024) selects samples using input trend variance and applies a robust loss to target anomalies. Selective Learning (Fu et al., 2025) masks uncertain target steps based on prediction error. ST-Prune (Chen et al., 2026) removes easy or repetitive samples using the model's learning state to accelerate training. DropoutTS (Zhong et al., 2026) adjusts dropout to frequency-estimated noise, and FAF (Taga et al., 2026) filters samples with a reference model before augmenting those retained. These methods act on target steps, whole samples, or regularization strength. StarTS assigns continuous loss weights from current relative difficulty and per-sample fitting history.

3 PRELIMINARIES

3.1 PROBLEM DEFINITION

Let $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ denote a multivariate time series forecasting dataset with N samples, where $\mathbf{x}_i \in \mathbb{R}^{L\times C}$ is a historical sequence of length L with C variables and $\mathbf{y}_i \in \mathbb{R}^{H\times C}$ is its future target of length H. A forecaster $f_\theta : \mathbb{R}^{L\times C} \to \mathbb{R}^{H\times C}$ maps each historical sequence to its future sequence. At optimization step t, the per-sample loss is

$$\ell_{i,t} = \phi(f_{\theta_t}(\mathbf{x}_i), \mathbf{y}_i)\,, \tag{1}$$

where ϕ is the base forecasting loss averaged over the predicted sequence. During standard training, these per-sample losses are averaged uniformly.

3.2 DYNAMIC SAMPLE REWEIGHTING

Example reweighting uses weighted training losses (Ren et al., 2018; Shu et al., 2019). For each sample i in a mini-batch $\mathcal{B}_t$, we write its normalized weight as $q_t(i \mid \mathcal{B}_t)$, defining the effective training distribution $q_t(\cdot \mid \mathcal{B}_t)$:

$$\mathcal{L}_t(q_t) = \sum_{i\in\mathcal{B}_t} q_t(i \mid \mathcal{B}_t)\ell_{i,t}, \qquad q_t(i \mid \mathcal{B}_t) \geq 0, \qquad \sum_{i\in\mathcal{B}_t} q_t(i \mid \mathcal{B}_t) = 1. \tag{2}$$

Uniform averaging is recovered by setting $q_t(i \mid \mathcal{B}_t) = 1/|\mathcal{B}_t|$. If q_t is computed from detached statistics, then $\nabla_{\theta_t}\mathcal{L}_t = \sum_{i\in\mathcal{B}_t} q_t(i \mid \mathcal{B}_t)\nabla_{\theta_t}\ell_{i,t}$. Thus, q_t scales each sample's contribution to the parameter update.

The weight $q_t(i \mid \mathcal{B}_t)$ denotes sample i's relative training priority within the current mini-batch. StarTS adjusts this priority using recent loss statistics and each sample's fitting history.

4 METHOD

StarTS dynamically allocates training emphasis using each sample's current loss-relative fitting state and accumulated fitting history. Figure 2 shows the training pipeline of StarTS. First, the forecasting backbone maps historical sequences to predictions, which are compared with the targets to compute per-sample losses. Second, *Fitting Range Evaluation* (FRE) standardizes these losses using recent loss statistics and penalizes unusually high relative difficulty to obtain fitting-range scores. Third, *Mastery Memory* (MM) converts the standardized losses into mastery evidence and combines it with stored memory to compute new mastery estimates. Finally, the current fitting-range scores and previous mastery values determine sample weights for the forecasting loss used to update the backbone. The new loss statistics and mastery estimates are saved after the optimization step. The following subsections detail these computations.

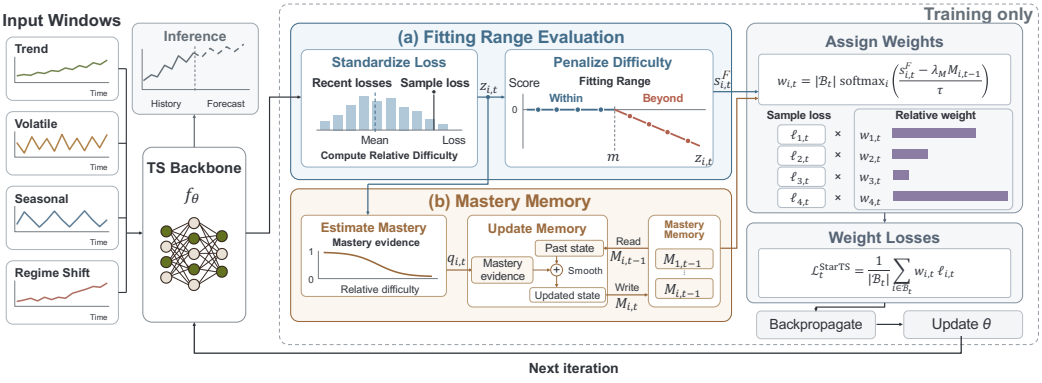

Figure 2: **Overview of StarTS.** (a) Fitting Range Evaluation evaluates each sample relative to the model's recent loss distribution. (b) Mastery Memory tracks whether each sample remains easy across repeated visits during training.

4.1 FITTING RANGE EVALUATION

The same loss value can represent different levels of difficulty across forecasting models and training stages. A fixed loss threshold may therefore fail to reflect the model's current fitting state. We design FRE to assess each sample relative to the model's recent errors and reduce the influence of unusually difficult samples.

As shown in Figure 2(a), FRE begins with the *Standardize Loss* step, which maps raw losses onto a common scale to measure each sample's relative difficulty as training progresses. Let μ_{t-1} and r_{t-1} be moving averages of the loss and squared loss, respectively, which together estimate the model's recent loss level and variability. At step t, we compute

$$\begin{aligned} \sigma_{t-1} &= \sqrt{\max(r_{t-1} - \mu_{t-1}^2, 0)}, \\ z_{i,t} &= \frac{\mathrm{sg}(\ell_{i,t}) - \mu_{t-1}}{\sigma_{t-1} + \epsilon}, \end{aligned} \tag{3}$$

where σ_{t-1} estimates recent loss variation. The standardized score $z_{i,t}$ measures relative fitting difficulty against the model's recent errors. Here, $\mathrm{sg}(\cdot)$ stops gradients through its input and $\epsilon > 0$ prevents division by zero. We update the loss statistics using exponential moving averages:

$$\begin{aligned} \mu_t &= \alpha\mu_{t-1} + (1-\alpha)\overline{\ell}_t, \\ r_t &= \alpha r_{t-1} + (1-\alpha)\overline{\ell^2}_t, \end{aligned} \tag{4}$$

where $\overline{\ell}_t$ and $\overline{\ell^2}_t$ are the batch averages of detached per-sample losses and their squares, respectively. The coefficient $\alpha \in [0, 1)$ controls how quickly the statistics adapt to new losses.

Next, *Penalize Difficulty* converts relative difficulty into a sample-level penalty using $z_{i,t} = m$ as a loss-relative fitting boundary, corresponding to the loss threshold $\mu_{t-1} + m(\sigma_{t-1} + \epsilon)$. A larger margin m allows higher relative difficulty before a penalty is applied. We convert difficulty beyond this boundary into a fitting-range score:

$$s_{i,t}^{\mathrm{F}} = -\operatorname{ReLU}(z_{i,t} - m). \tag{5}$$

Samples within the boundary receive no penalty, while those beyond it receive larger penalties as their difficulty increases.

4.2 Mastery Memory

Current relative difficulty estimates where a sample lies in the model's recent fitting range, while its training priority also depends on whether it has remained easy across visits. We introduce MM to track this history and reduce the emphasis on consistently well-fitted samples.

As shown in Figure 2(b), the *Estimate Mastery* step first assesses how easily the model currently fits each sample relative to the evolving loss distribution. It converts relative difficulty $z_{i,t}$ from FRE into current mastery evidence:

$$q_{i,t} = \operatorname{sigmoid}\left(-\frac{z_{i,t}}{\tau_{\mathrm{M}}}\right), \tag{6}$$

The evidence lies in $(0, 1)$, with higher values for easier samples. The temperature $\tau_{\mathrm{M}} > 0$ controls how sharply the score changes with difficulty.

Next, *Update Memory* smooths mastery evidence across visits by combining the previous state $M_{i,t-1}$ from the memory pool with the current evidence:

$$M_{i,t} = \rho M_{i,t-1} + (1-\rho) q_{i,t}, \qquad 0 \le \rho < 1, \tag{7}$$

The mastery estimate $M_{i,t} \in [0, 1]$ summarizes recent evidence of easy fitting and can decrease when a sample becomes harder.

The memory pool is initialized with $M_{i,0} = 0$, indexed by fixed sample identifiers, and updated only for samples in the current batch.

4.3 Dynamic Weight Assignment and Optimization

Sample weighting combines current relative fitting difficulty with historical fitting state. We favor samples within the current fitting range with limited mastery evidence.

The joint score and normalized sample weights are computed as follows:

$$\begin{aligned} a_{i,t} &= s_{i,t}^{\mathrm{F}} - \lambda_{\mathrm{M}} M_{i,t-1}, \\ q_t(i \mid \mathcal{B}_t) &= \frac{\exp(a_{i,t}/\tau)}{\sum_{j \in \mathcal{B}_t} \exp(a_{j,t}/\tau)}, \\ w_{i,t} &= |\mathcal{B}_t| q_t(i \mid \mathcal{B}_t), \qquad i \in \mathcal{B}_t, \end{aligned} \tag{8}$$

Here, $\lambda_{\mathrm{M}} \ge 0$ sets the mastery penalty and $\tau > 0$ controls weight concentration. Before optional masking, the continuous weights are positive and have batch mean one.

These weights set each sample's relative contribution to the current training loss:

$$\begin{aligned} \mathcal{S}_t &= \operatorname{TopIdx}_{|\mathcal{B}_t| - \lfloor r|\mathcal{B}_t| \rfloor}\big((w_{i,t})_{i \in \mathcal{B}_t}\big), \\ \mathcal{L}_t^{\mathrm{StarTS}} &= \frac{\sum_{i \in \mathcal{S}_t} w_{i,t} \ell_{i,t}}{\sum_{j \in \mathcal{S}_t} w_{j,t}}, \qquad 0 \le r < 1. \end{aligned} \tag{9}$$

The weights are held constant during backpropagation. Each step computes them from current losses and previous states, updates the backbone, and only then stores the new loss statistics and mastery memory. The first batch uses uniform weights to initialize the loss statistics without updating mastery memory. Continuous weighting adds $O(|\mathcal{B}_t|)$ computation per batch and $O(N)$ memory, with no change to inference.

5 EXPERIMENTS

We evaluate StarTS through four research questions:

- **RQ1: Forecasting accuracy.** How does StarTS compare with Raw across datasets and backbones, and with sample-aware training baselines? See Section 5.2.
- **RQ2: Cross-Dataset Transfer.** Does StarTS improve zero-shot transfer across ETT datasets without further training? See Section 5.3.
- **RQ3: Components, sensitivity, and efficiency.** How do FRE and MM contribute, how sensitive is StarTS, and what computational overhead does it introduce? See Section 5.4.
- **RQ4: Dynamic training emphasis.** How does StarTS weight samples with different fitting trajectories? See Section 5.5.

5.1 EXPERIMENTAL SETTINGS

Datasets and metrics. We evaluate eight datasets: ETTm1, ETTm2, ETTh1, and ETTh2 (Zhou et al., 2021), COVID19 (Dong et al., 2020), NABCPU (Lavin & Ahmad, 2015), Sunspots (Clette & Lefèvre, 2016), and ExchangeRate (Lai et al., 2018), each at four forecasting horizons. We report mean squared error (MSE), mean absolute error (MAE), and dataset means that weight the four horizons equally; full horizon-level results are in the appendix.

Baselines. To assess cross-family gains, we evaluate nine linear, Transformer, mixer, and filtering backbones: WPMixer (Murad et al., 2025), MultiPatchFormer (Naghashi et al., 2025), TimeFilter (Hu et al., 2025), TimeBase (Huang et al., 2025), OLinear (Yue et al., 2025), TimeMixer++ (Wang et al., 2025), Autoformer (Wu et al., 2021), PatchTST (Nie et al., 2023), and iTransformer (Liu et al., 2024). Selective Learning (SL) (Fu et al., 2025) and DropoutTS (DT) (Zhong et al., 2026) represent target-step selection and noise-adaptive regularization, respectively.

Implementation Details. Forecasting horizons for each dataset are listed in Appendix A.1. The look-back length is $L = 96$ for ETT4, NABCPU, and ExchangeRate, and $L = 36$ for COVID19 and Sunspots. We train each model for up to 100 epochs using MAE loss and Adam with a batch size of 32, a learning rate of 2×10^{-4}, weight decay of 5×10^{-4}, and early stopping with patience 10. Within each comparison, methods retain the same forecasting architecture, base MAE loss, validation, and inference protocols; Raw and StarTS differ only in sample weighting.

5.2 RQ1: HOW DOES STARTS AFFECT FORECASTING ACCURACY?

Table 2: Forecasting results are averaged over three runs and four horizons for each model and dataset comparison. Gray denotes +StarTS; better means are bold and the others underlined. Arrows show relative changes in mean errors; the last row averages them across eight datasets in MSE/MAE order. Full results with standard deviations are in Tables 7 and 8 in the appendix.

Dataset	Metric	Autoformer (2021)		WPMixer (2025)		MultiPatchFormer (2025)		TimeFilter (2025)		OLinear (2025)		TimeMixer++ (2025)	
		Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS
ETTm1	MSE	0.679	**0.653**↑ 3.9%	**0.391**	0.392↓ 0.1%	0.667	**0.655**↑ 1.8%	0.398	**0.397**↑ 0.2%	0.423	**0.418**↑ 1.1%	0.410	**0.410**↑ 0.1%
	MAE	0.578	**0.566**↑ 2.0%	**0.384**	0.386↓ 0.3%	**0.500**	0.501↓ 0.1%	**0.387**	0.387↓ 0.0%	0.415	**0.410**↑ 1.1%	**0.397**	0.398↓ 0.3%
ETTm2	MSE	0.324	**0.306**↑ 5.4%	**0.276**	0.278↓ 0.9%	0.306	**0.303**↑ 0.9%	**0.283**	0.285↓ 0.5%	**0.296**	0.296↓ 0.0%	0.290	**0.289**↑ 0.3%
	MAE	0.360	**0.348**↑ 3.3%	**0.317**	0.318↓ 0.3%	0.347	**0.345**↑ 0.7%	0.324	**0.323**↑ 0.2%	**0.335**	0.336↓ 0.2%	**0.328**	0.328↓ 0.1%
ETTh1	MSE	0.695	**0.612**↑ 12.0%	0.464	**0.446**↑ 3.9%	0.734	**0.728**↑ 0.8%	0.458	**0.454**↑ 1.0%	0.458	**0.455**↑ 0.7%	0.477	**0.470**↑ 1.4%
	MAE	0.608	**0.563**↑ 7.3%	0.437	**0.428**↑ 1.8%	0.542	**0.542**↑ 0.0%	0.431	**0.428**↑ 0.9%	0.447	**0.444**↑ 0.6%	0.442	**0.442**↑ 0.1%
ETTh2	MSE	0.648	**0.580**↑ 10.6%	0.379	**0.371**↑ 2.3%	0.423	**0.412**↑ 2.7%	0.388	**0.384**↑ 0.9%	**0.400**	0.401↓ 0.2%	0.404	**0.397**↑ 2.0%
	MAE	0.572	**0.529**↑ 7.6%	0.399	**0.394**↑ 1.3%	0.436	**0.429**↑ 1.5%	0.403	**0.401**↑ 0.5%	0.417	**0.417**↑ 0.0%	0.414	**0.408**↑ 1.4%
COVID19	MSE	2.193	**2.145**↑ 2.1%	0.852	**0.778**↑ 8.6%	1.431	**1.163**↑ 18.7%	0.892	**0.814**↑ 8.7%	1.958	**1.641**↑ 16.2%	0.904	**0.855**↑ 5.4%
	MAE	0.837	**0.823**↑ 1.7%	0.368	**0.362**↑ 1.6%	0.599	**0.530**↑ 11.5%	0.386	**0.377**↑ 2.2%	0.635	**0.576**↑ 9.3%	0.424	**0.408**↑ 3.8%
NABCPU	MSE	1.427	**1.334**↑ 6.5%	**1.201**	1.202↓ 0.1%	1.206	**1.204**↑ 0.2%	1.193	**1.192**↑ 0.1%	1.205	**1.203**↑ 0.2%	1.210	**1.208**↑ 0.1%
	MAE	0.590	**0.508**↑ 13.9%	0.331	**0.331**↑ 0.2%	0.312	**0.311**↑ 0.4%	**0.319**	0.319↓ 0.1%	**0.336**	0.340↓ 1.1%	**0.323**	0.326↓ 0.8%
Sunspots	MSE	1.011	**0.933**↑ 7.7%	0.728	**0.719**↑ 1.3%	0.655	**0.645**↑ 1.6%	**0.812**	0.821↓ 1.2%	1.084	**1.081**↑ 0.3%	0.706	**0.692**↑ 2.1%
	MAE	0.760	**0.727**↑ 4.4%	0.589	**0.585**↑ 0.7%	0.555	**0.552**↑ 0.7%	**0.627**	0.631↓ 0.7%	0.745	**0.744**↑ 0.1%	0.585	**0.575**↑ 1.6%
ExchangeRate	MSE	0.462	**0.448**↑ 3.1%	0.282	**0.247**↑ 12.6%	0.284	**0.270**↑ 4.8%	0.272	**0.255**↑ 6.1%	0.290	**0.273**↑ 5.8%	0.274	**0.260**↑ 4.9%
	MAE	0.505	**0.497**↑ 1.6%	0.367	**0.346**↑ 5.6%	0.379	**0.375**↑ 1.1%	0.363	**0.353**↑ 2.6%	0.379	**0.370**↑ 2.4%	0.365	**0.357**↑ 2.2%
Avg. Impv.		**6.4% / 5.2%**		**3.4% / 1.3%**		**3.9% / 2.0%**		**1.9% / 0.7%**		**3.0% / 1.5%**		**2.0% / 1.0%**	

Results. Across nine backbones, StarTS lowers the mean MSE by 0.6%–6.4%; the six models in Table 2 improve the eight-dataset averages in both MSE and MAE. Autoformer has the largest average reductions (6.4% MSE, 5.2% MAE). Among the five 2025 backbones, mean MSE reductions are 1.9%–3.9%, with larger task-specific gains: MultiPatchFormer and OLinear lower COVID19 MSE by 18.7% and 16.2%, respectively, while WPMixer lowers ExchangeRate MSE by 12.6%. Tables 7 and 8 report the full horizon-level results.

Table 3: Training-strategy comparison in forecasting accuracy (lower is better). Results are averaged over three runs; full horizon-level results appear in Appendix B.1. Within each backbone and metric, the best and second-best unrounded means are bold and underlined. Gray denotes +StarTS. The final row averages dataset-wise MSE/MAE percentage changes relative to Raw; negative values indicate higher error.

Dataset	Metric	OLinear				MultiPatchFormer				TimeMixer++			
		Raw	+StarTS	+SL	+DT	Raw	+StarTS	+SL	+DT	Raw	+StarTS	+SL	+DT
COVID19	MSE	1.958	**1.641**	1.887	1.947	1.431	1.163	**1.157**	1.427	0.904	**0.855**	0.973	0.921
	MAE	0.635	**0.576**	0.614	0.635	0.599	0.530	**0.525**	0.602	0.424	**0.408**	0.423	0.434
NABCPU	MSE	1.205	1.203	1.207	**1.200**	1.206	**1.204**	1.207	1.207	1.210	1.208	**1.204**	1.210
	MAE	**0.336**	0.340	0.364	0.337	0.312	**0.311**	0.316	0.312	**0.323**	0.326	0.339	0.332
Sunspots	MSE	1.084	**1.081**	1.133	1.105	0.655	**0.645**	0.738	0.653	0.706	**0.692**	0.788	0.998
	MAE	0.745	**0.744**	0.771	0.759	0.555	**0.552**	0.593	0.556	0.585	**0.575**	0.616	0.709
ExchangeRate	MSE	0.290	**0.273**	0.288	0.289	0.284	**0.270**	0.302	0.287	0.274	**0.260**	0.277	0.276
	MAE	0.379	**0.370**	0.381	0.379	0.379	**0.375**	0.391	0.380	0.365	**0.357**	0.368	0.366
Avg. Impv. (%)		–	5.6/2.7	-0.1/-2.3	-0.2/-0.5	–	6.3/3.4	-0.1/0.3	-0.2/-0.2	–	3.1/1.7	-4.9/-2.7	-10.9/-6.6

Table 3 shows that StarTS achieves lower average MSE than both sample-aware training baselines on all three backbones. Its average MSE reductions relative to Raw are 3.1% on TimeMixer++, 5.6% on OLinear, and 6.3% on MultiPatchFormer, while +SL and +DT increase average MSE on each backbone. StarTS improves over Raw on all twelve backbone–dataset pairs, outperforms +SL on ten and +DT on eleven, and achieves the lowest MSE on three datasets per backbone. These results are consistent with StarTS's dynamic reallocation of training emphasis using complementary current-state and history-based signals: FRE reduces the influence of samples outside the recent fitting range, while MM reduces repeated emphasis on samples that remain easy to fit. Together, these signals shift training weight toward samples within the fitting range with limited mastery evidence.

5.3 RQ2: Does StarTS Improve Cross-Dataset Transfer?

Settings. We compare Raw and StarTS in zero-shot transfer across eight ETT directions. Models retain source normalization; further settings are given in Appendix A.2.

Results. StarTS lowers target MSE in seven of eight ETT zero-shot transfer directions, with average reductions of 10.89% in MSE and 6.94% in MAE (Table 4). Gains vary by direction: ETTh2→ETTm2 achieves the largest MSE reduction of 28.89%, while ETTm1→ETTh2 increases MSE by 3.41%. These direction-dependent gains show that the benefit of sample reweighting extends to most evaluated ETT targets under zero-shot evaluation.

Table 4: Zero-shot transfer results averaged over four horizons. Parentheses average per-horizon relative improvements; the final row averages all directions. Gray denotes +StarTS, and better errors are bold. Compare absolute errors only within each direction.

Source → Target	Raw		+StarTS	
	MSE	MAE	MSE	MAE
ETTh1→ETTh2	3.392	1.282	**3.369** (↑0.58%)	**1.270** (↑0.89%)
ETTh2→ETTh1	0.426	0.460	**0.354** (↑17.09%)	**0.420** (↑8.88%)
ETTm1→ETTm2	2.311	**1.040**	**2.299** (↑0.68%)	1.041 (↓0.03%)
ETTm2→ETTm1	0.330	0.405	**0.221** (↑28.45%)	**0.322** (↑17.97%)
ETTh1→ETTm2	2.407	1.073	**2.318** (↑2.87%)	**1.044** (↑2.33%)
ETTh2→ETTm2	0.747	0.635	**0.537** (↑28.89%)	**0.516** (↑19.31%)
ETTm1→ETTh2	**3.188**	**1.225**	3.286 (↓3.41%)	1.261 (↓3.29%)
ETTm2→ETTh2	0.536	0.515	**0.460** (↑11.96%)	**0.460** (↑9.42%)
Avg. Impv.	–	–	**10.89%↑**	**6.94%↑**

5.4 RQ3: Component Contributions, Sensitivity, and Efficiency

Ablation Study. FRE and MM provide complementary gains in Figure 3. FRE-only is more effective on COVID19 and ExchangeRate, while MM-only performs better on NABCPU and Sunspots. Full achieves the lowest mean MSE on all four datasets. The higher errors of FRE-only on Sunspots and MM-only on COVID19 highlight the value of combining current difficulty with fitting history. FRE penalizes samples beyond the current fitting range, while MM downweights repeatedly easy samples. Together, they emphasize samples within the fitting range with limited mastery evidence.

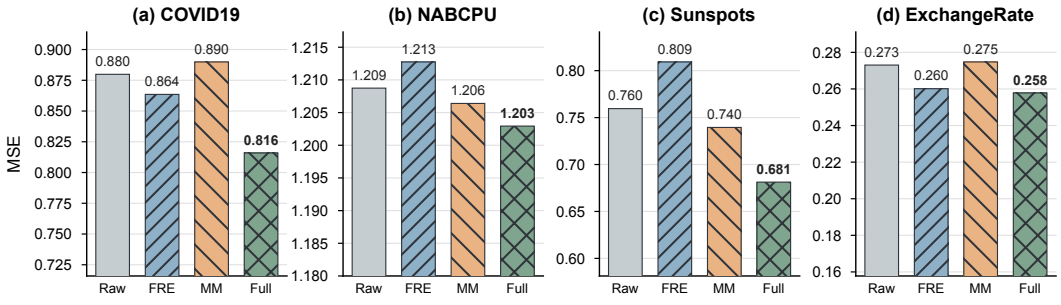

Figure 3: **TimeMixer++ component comparison by mean test MSE.**

Full improves over Raw at all 16 evaluated horizons, while the relative strengths of FRE and MM vary across horizons. Appendix Table 10 reports the detailed comparisons.

Hyperparameter Sensitivity. Figure 4 identifies the fitting-range margin m as the main source of sensitivity. Increasing m allows harder samples to retain weight, benefiting Sunspots but hurting COVID19 and ExchangeRate. This suggests that the useful fitting range depends on the dataset. MSE is less sensitive to MM strength λ_{M} and temperatures $\tau \in [0.75, 2]$, whereas $\tau = 0.5$ concentrates weights too sharply and raises errors on COVID19 and Sunspots. Thus, adaptation should focus on the fitting boundary while maintaining a balanced distribution of training weights.

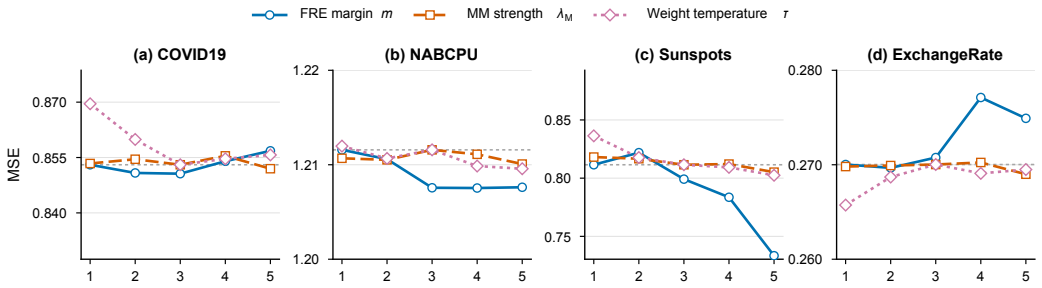

Figure 4: **Hyperparameter sensitivity.** Sweep values are listed in Appendix Table 11.

Efficiency. StarTS retains near-baseline GPU memory use, adding 0.1 MiB and 2.67 s of training time relative to Raw (Table 5). StarTS reduces GPU memory use by 17.6% and training time by 1.50 s relative to DropoutTS. SL is 1.45 s faster than StarTS, but its peak GPU memory use is 22.8 MiB higher. Its weights are computed from training losses, without an auxiliary model. The weights are used only during training, leaving the model's inference procedure unchanged. The weighting rule stores shared loss statistics and one mastery state per sample, providing a compact state for updating training priorities as the fitting range and each sample's fitting history evolve during optimization.

Table 5: Fixed-budget training efficiency.

Method	Total time (s)	Added (ms/step)	GPU memory (MiB)
Raw	31.44	–	1788.3
+StarTS	34.11	+10.67	**1788.4**
+SL	**32.66**	**+4.86**	1811.2
+DropoutTS	35.61	+16.68	2169.4

5.5 RQ4: How Does StarTS Allocate Training Emphasis Across Samples?

Settings. Figures 5 and 6 show training trajectories on Sunspots and weight preferences across COVID19 windows, respectively, using TimeMixer++.

Results. In Figure 5(a), loss drops early and stays low as the weight becomes negligible. In (b), the weight rises as forecasts improve and remains high as MAE falls by 67.6%. In (c), persistent errors receive only intermittent emphasis. These trajectories illustrate how training priorities evolve with fitting state and history. Figure 6 describes the weighting pattern: higher late-stage weights concentrate in low-variability, weak-trend, and low-volatility regions. Similar windows can receive different weights; earlier stages show similar broad preferences (Appendix Figure 7).

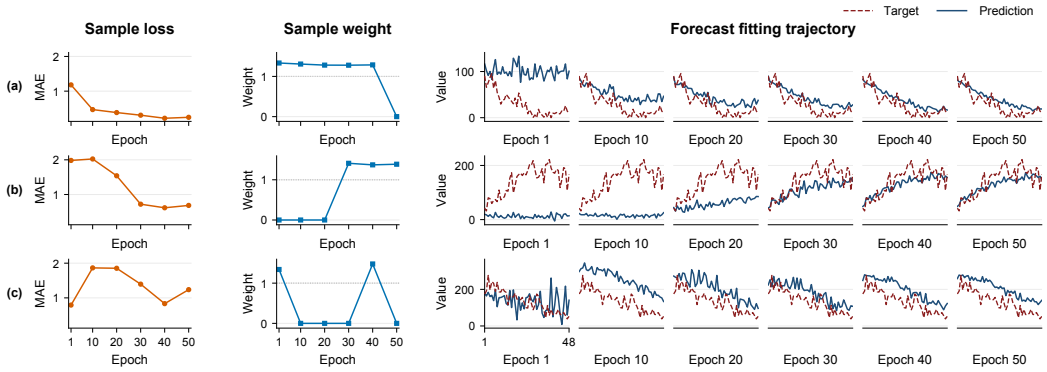

Figure 5: **Sample weights during learning.** Each row shows sample MAE, training weight, and forecasts at six stages of model training on Sunspots. The rows illustrate (a) consistently well-fitted, (b) improving in fit, and (c) persistently high-error samples.

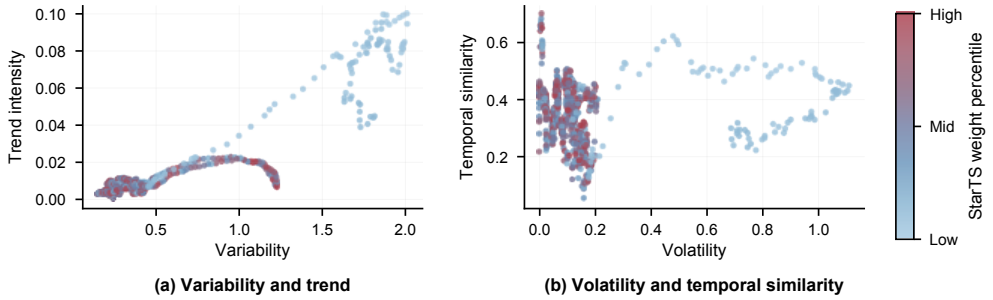

Figure 6: **Late-stage sample properties and weights.** COVID19 windows colored by weight percentile: (a) variability and trend; (b) volatility and temporal similarity.

6 Conclusion and Future Work

StarTS addresses *time-varying training priorities* through state-aware sample reweighting. Fitting Range Evaluation and Mastery Memory allocate emphasis using current loss-relative fitting state and historical evidence of easy fitting, with unchanged backbones and inference. StarTS improves mean MSE on all eight datasets and reduces average MSE by 0.6%–6.4% across nine backbones. Zero-shot transfer reduces MSE in seven of eight evaluated ETT directions, while ablations demonstrate the complementary benefits of current fitting state and mastery history. Sample-level analyses further illustrate how training priorities evolve with each sample's fitting state and history. StarTS also achieves lower average MSE than two sample-aware training baselines across four datasets. Future work will explore richer training feedback to distinguish difficult but improvable samples from noisy or irreducible cases and extend state-aware reweighting to online forecasting under distribution shifts.

AI USE STATEMENT

AI tools assisted with manuscript drafting and revision, methodological and experimental feedback, code editing and debugging, result aggregation, figure preparation, and formatting. These tools were used under author direction and are not listed as authors. The authors are responsible for verifying AI-assisted outputs against the underlying code and experiment records and take full responsibility for the final content.

REPRODUCIBILITY STATEMENT

Section 4 defines the sample-weighting rule and state updates. Section 5.1 describes the baselines, metrics, and training settings, while Appendix A.1 provides dataset characteristics, splits, input lengths, and forecasting horizons. Tables 7 and 8 in the appendix report horizon-level means and standard deviations over three runs. Section 5.4 describes the component comparisons and sensitivity sweeps, with full component results in Appendix D.

ETHICS STATEMENT

This work studies time series forecasting using existing public benchmark datasets, without recruiting human participants or collecting new personal data. The experiments evaluate forecasting methods rather than operational decisions in healthcare, finance, or infrastructure. Such uses would require domain-specific validation and human oversight. Sample reweighting can reduce the influence of rare but informative observations, so deployment should assess performance on unusual events as well as average forecasting error.

REFERENCES

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 41–48, 2009. doi: 10.1145/1553374.1553380.

Haw-Shiuan Chang, Erik Learned-Miller, and Andrew McCallum. Active bias: Training more accurate neural networks by emphasizing high variance samples. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/2f37d10131f2a483a8dd005b3d14b0d9-Abstract.html.

Wei Chen, Junle Chen, Yuqian Wu, Yuxuan Liang, and Xiaofang Zhou. Learning from complexity: Exploring dynamic sample pruning of spatio-temporal training, 2026. URL https://arxiv.org/abs/2602.19113.

Hao Cheng, Qingsong Wen, Yang Liu, and Liang Sun. RobustTSF: Towards theory and design of robust time series forecasting with anomalies. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ltZ9ianMth.

Frédéric Clette and Laure Lefèvre. The new sunspot number: Assembling all corrections. *Solar Physics*, 291(9–10):2629–2651, 2016. doi: 10.1007/s11207-016-1014-y.

Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017. doi: 10.1016/j.rser.2017.02.085.

Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534, 2020. doi: 10.1016/S1473-3099(20)30120-1.

Yisong Fu, Zezhi Shao, Chengqing Yu, Yujie Li, Zhulin An, Qi Wang, Yongjun Xu, and Fei Wang. Selective learning for deep time series forecasting, 2025. URL https://arxiv.org/abs/2510.25207.

Amirata Ghorbani and James Zou. Data shapley: Equitable valuation of data for machine learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2242–2251. PMLR, 2019. URL https://proceedings.mlr.press/v97/ghorbani19c.html.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor W. Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. TimeFilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 24893–24911. PMLR, 2025. URL https://proceedings.mlr.press/v267/hu25ac.html.

Qihe Huang, Zhengyang Zhou, Kuo Yang, Zhongchao Yi, Xu Wang, and Yang Wang. TimeBase: The power of minimalism in efficient long-term time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 26227–26246. PMLR, 2025. URL https://proceedings.mlr.press/v267/huang25az.html.

Sheikh Mohammad Idrees, M. Afshar Alam, and Parul Agarwal. A prediction approach for stock market volatility based on time series data. *IEEE Access*, 7:17287–17298, 2019. doi: 10.1109/ACCESS.2019.2895252.

Zahra Karevan and Johan A. K. Suykens. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks*, 125:1–9, 2020. doi: 10.1016/j.neunet.2019.12.030.

Angelos Katharopoulos and François Fleuret. Not all samples are created equal: Deep learning with importance sampling. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2525–2534. PMLR, 2018. URL https://proceedings.mlr.press/v80/katharopoulos18a.html.

M. Pawan Kumar, Benjamin Packer, and Daphne Koller. Self-paced learning for latent variable models. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 95–104, 2018. doi: 10.1145/3209978.3210006.

Alexander Lavin and Subutai Ahmad. Evaluating real-time anomaly detection algorithms—the Numenta anomaly benchmark. In *2015 IEEE 14th International Conference on Machine Learning and Applications*, 2015. doi: 10.1109/ICMLA.2015.141.

Yubo Liang, Zezhi Shao, Fei Wang, Zhao Zhang, Tao Sun, and Yongjun Xu. BasicTS: An open source fair multivariate time series prediction benchmark. In *International Symposium on Benchmarking, Measuring and Optimization*, pp. 87–101. Springer, 2022. doi: 10.1007/978-3-031-31180-2_6.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017. doi: 10.1109/ICCV.2017.324.

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. URL https://arxiv.org/abs/2310.06625.

Sören Mindermann, Jan M. Brauner, Muhammed T. Razzak, Mrinank Sharma, Andreas Kirsch, Winnie Xu, Benedikt Höltgen, Aidan N. Gomez, Adrien Morisot, Sebastian Farquhar, and Yarin Gal. Prioritized training on points that are learnable, worth learning, and not yet learnt. In

Proceedings of the 39th International Conference on Machine Learning, volume 162 of *Proceedings of Machine Learning Research*, pp. 15630–15649. PMLR, 2022. URL https://proceedings.mlr.press/v162/mindermann22a.html.

Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. WPMixer: Efficient multi-resolution mixing for long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 19581–19588, 2025. doi: 10.1609/aaai.v39i18.34156.

Vahid Naghashi, Mounir Boukadoum, and Abdoulaye Banire Diallo. A multiscale model for multivariate time series forecasting. *Scientific Reports*, 15:1565, 2025. doi: 10.1038/s41598-024-82417-4.

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. URL https://arxiv.org/abs/2211.14730.

Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urtasun. Learning to reweight examples for robust deep learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4334–4343. PMLR, 2018. URL https://proceedings.mlr.press/v80/ren18a.html.

Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. Meta-Weight-Net: Learning an explicit mapping for sample weighting. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/e58cc5ca94270acaceed13bc82dfedf7-Abstract.html.

Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, Yizhong Wang, Hannaneh Hajishirzi, Noah A. Smith, and Yejin Choi. Dataset cartography: Mapping and diagnosing datasets with training dynamics. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 9275–9293. Association for Computational Linguistics, 2020. doi: 10.18653/v1/2020.emnlp-main.746. URL https://aclanthology.org/2020.emnlp-main.746/.

Ege Onur Taga, Halil Alperen Gozeten, Kutay Tire, Rahul Dalvi, Reinhard Heckel, and Samet Oymak. Filter, augment, forecast: Online data selection for robust time series forecasting. In *Proceedings of the 29th International Conference on Artificial Intelligence and Statistics*, volume 300 of *Proceedings of Machine Learning Research*. PMLR, 2026. URL https://openreview.net/forum?id=PuVT8WarRl.

Mariya Toneva, Alessandro Sordoni, Remi Tachet des Combes, Adam Trischler, Yoshua Bengio, and Geoffrey J. Gordon. An empirical study of example forgetting during deep neural network learning. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1812.05159.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin. TimeMixer++: A general time series pattern machine for universal predictive analysis. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/2b187165e28fdfdc0ffb34d1bfff2b0c-Abstract-Conference.html.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Jinsung Yoon, Sercan Arik, and Tomas Pfister. Data valuation using reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 10842–10851. PMLR, 2020. URL https://proceedings.mlr.press/v119/yoon20a.html.

Wenzhen Yue, Yong Liu, Hao Wang, Haoxuan Li, Xianghua Ying, Ruohao Guo, Bowei Xing, and Ji Shi. OLinear: A linear model for time series forecasting in orthogonally transformed domain. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/

085713-2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/571c7e164fb1ffbcf2f84a63784451ec-Abstract-Conference.html.

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. doi: 10.1609/aaai.v37i9.26317.

Zhiyuan Zhao, Juntong Ni, Shangqing Xu, Haoxin Liu, Wei Jin, and B. Aditya Prakash. TimeRecipe: A time-series forecasting recipe via benchmarking module level effectiveness. In *International Conference on Learning Representations*, pp. 150745–150792, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/f46ddea413df86832418c5e04e59644f-Abstract-Conference.html.

Jianhu Zheng and Mingfang Huang. Traffic flow forecast through time series analysis based on deep learning. *IEEE Access*, 8:82562–82570, 2020. doi: 10.1109/ACCESS.2020.2990738.

Siru Zhong, Yiqiu Liu, Zhiqing Cui, Zezhi Shao, Fei Wang, Qingsong Wen, and Yuxuan Liang. DropoutTS: Sample-adaptive dropout for robust time series forecasting, 2026. URL https://arxiv.org/abs/2601.21726.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021. doi: 10.1609/aaai.v35i12.17325.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 27268–27286. PMLR, 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

Tianyi Zhou, Shengjie Wang, and Jeff A. Bilmes. Curriculum learning by dynamic instance hardness. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/62000dee5a05a6a71de3a6127a68778a-Abstract.html.

A Experimental Details

A.1 Benchmark Details

We evaluate eight forecasting datasets spanning electricity systems, public health, cloud monitoring, solar activity, and finance. Table 6 summarizes the processed datasets used in our experiments, with sampling intervals ranging from five minutes to one month.

- **ETT (Electricity Transformer Temperature).** ETT records transformer oil temperature and six power-load variables at two stations.
- **COVID19.** Eight daily series record new COVID-19 cases.
- **NABCPU.** Three series track CPU utilization in cloud monitoring.
- **Sunspots.** Monthly sunspot observations track solar activity.
- **ExchangeRate.** Daily exchange-rate series cover eight countries.

Table 6: Benchmark characteristics. Dim. counts forecast variables, excluding calendar features; Size counts time steps before constructing training samples. Split denotes the chronological train/validation/test ratio. Prediction lengths are measured in time steps.

Dataset	Dim.	Size	Split	Frequency	Prediction length	Domain
ETTh1	7	14,400	6:2:2	1 hour	$\{96, 192, 336, 720\}$	Electricity
ETTh2	7	14,400	6:2:2	1 hour	$\{96, 192, 336, 720\}$	Electricity
ETTm1	7	57,600	6:2:2	15 min	$\{96, 192, 336, 720\}$	Electricity
ETTm2	7	57,600	6:2:2	15 min	$\{96, 192, 336, 720\}$	Electricity
COVID19	8	1,143	7:1:2	1 day	$\{7, 14, 28, 60\}$	Public health
NABCPU	3	4,031	7:1:2	5 min	$\{24, 48, 96, 192\}$	Cloud CPU
Sunspots	1	3,327	7:1:2	1 month	$\{12, 24, 48, 96\}$	Solar activity
ExchangeRate	8	7,588	7:1:2	1 day	$\{96, 192, 336, 480\}$	Finance

We use the BasicTS framework (Liang et al., 2022) for our experiments. The splits preserve temporal order, with 60%/20%/20% for ETT and 70%/10%/20% for the other datasets, subject to integer rounding. The look-back length is 96 for ETT, NABCPU, and ExchangeRate, and 36 for COVID19 and Sunspots. Raw and StarTS use the same data partitions and forecasting horizons in every comparison between the methods.

For the training-strategy comparison in Table 3, results for all methods are averaged over three runs. SL uses a separately trained DLinear auxiliary scorer. Full horizon-level results are given in Appendix B.1.

Efficiency measurement. Table 5 uses TimeMixer++ on ExchangeRate with $L = H = 96$, batch size 32, and one RTX A6000. Times are medians over five 250-step runs, including data loading, callbacks, checkpoint I/O, and SL's auxiliary-estimator preparation. Added time per step is the median time difference from Raw divided by 250. Memory is the measured-phase peak allocation; SL uses the larger peak of its estimator and forecasting model.

A.2 Cross-Dataset Transfer Settings

RQ2 uses Autoformer trained with Raw or StarTS on each source dataset, with input length 96 and forecasting horizons $H \in \{96, 192, 336, 720\}$. The checkpoints retain their source normalization for zero-shot evaluation on the target test splits. The eight source–target pairs cover transfers between stations at the same sampling resolution and between hourly and 15-minute series. For each direction, MSE and MAE are averaged over four horizons. Relative improvements are computed against Raw at each horizon before averaging, giving each forecasting horizon equal weight. The final row averages these relative improvements across all eight transfer directions. Absolute errors are compared within each direction under the same source normalization.

B RESULTS OF OPEN BENCHMARKS

Table 7: Full forecasting results on ETT4. Values are means over three runs, with sample standard deviations below. Avg summarizes the four-horizon mean within each run using the same statistics. Within each model and metric, the lower unrounded mean is bold and the other underlined; exact ties are both bold. Gray cells denote +StarTS. Shaded Imp. rows report relative changes in the four-horizon means, with reductions bold and arrows indicating direction.

Dataset	H	Metric	Autoformer (2021)		PatchTST (2023)		iTransformer (2024)		WPMixer (2025)		MultiPatchFormer (2025)		TimeFilter (2025)		TimeBase (2025)		OLinear (2025)		TimeMixer++ (2025)	
			Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS
ETTm1	96	MSE	0.575 ±0.025	**0.554** ±0.051	0.328 ±0.003	**0.327** ±0.002	0.333 ±0.003	**0.331** ±0.001	**0.316** ±0.001	0.317 ±0.002	0.478 ±0.022	**0.464** ±0.008	0.332 ±0.001	**0.330** ±0.003	0.370 ±0.001	**0.367** ±0.001	0.356 ±0.000	**0.350** ±0.000	**0.339** ±0.001	0.340 ±0.001
		MAE	0.535 ±0.014	**0.521** ±0.020	**0.349** ±0.003	0.350 ±0.006	0.354 ±0.002	**0.349** ±0.001	0.343 ±0.000	**0.341** ±0.001	0.433 ±0.013	**0.428** ±0.008	**0.351** ±0.000	0.351 ±0.002	0.380 ±0.001	**0.377** ±0.001	0.378 ±0.000	**0.373** ±0.000	**0.360** ±0.001	0.363 ±0.004
	192	MSE	**0.586** ±0.048	0.591 ±0.021	0.376 ±0.002	**0.373** ±0.002	0.380 ±0.000	**0.379** ±0.001	0.372 ±0.003	**0.372** ±0.003	0.659 ±0.014	**0.638** ±0.004	0.376 ±0.000	**0.376** ±0.001	0.408 ±0.001	**0.406** ±0.000	0.398 ±0.000	**0.395** ±0.000	0.389 ±0.001	**0.388** ±0.002
		MAE	**0.530** ±0.021	0.537 ±0.015	**0.372** ±0.001	0.373 ±0.001	**0.377** ±0.001	0.380 ±0.001	**0.371** ±0.001	0.376 ±0.002	**0.497** ±0.003	0.502 ±0.004	**0.372** ±0.000	0.373 ±0.002	0.398 ±0.001	**0.396** ±0.001	0.399 ±0.000	**0.395** ±0.000	**0.383** ±0.002	0.384 ±0.002
	336	MSE	0.735 ±0.006	**0.682** ±0.023	0.409 ±0.003	**0.407** ±0.003	0.416 ±0.001	**0.414** ±0.002	**0.405** ±0.002	0.405 ±0.002	0.756 ±0.005	**0.750** ±0.007	0.410 ±0.001	**0.409** ±0.002	0.437 ±0.000	**0.436** ±0.000	0.436 ±0.000	**0.433** ±0.000	0.423 ±0.001	**0.422** ±0.001
		MAE	0.596 ±0.010	**0.574** ±0.009	0.396 ±0.001	**0.394** ±0.001	0.400 ±0.000	**0.398** ±0.002	**0.393** ±0.001	0.396 ±0.004	**0.529** ±0.001	0.530 ±0.002	**0.395** ±0.001	0.395 ±0.003	0.417 ±0.000	**0.415** ±0.000	0.422 ±0.000	**0.418** ±0.000	**0.404** ±0.001	0.404 ±0.001
	720	MSE	0.820 ±0.076	**0.785** ±0.040	**0.474** ±0.000	0.474 ±0.001	**0.473** ±0.002	0.475 ±0.003	0.472 ±0.002	**0.472** ±0.003	0.775 ±0.011	**0.769** ±0.010	0.474 ±0.001	**0.472** ±0.001	**0.499** ±0.000	0.500 ±0.001	0.503 ±0.000	**0.495** ±0.000	0.489 ±0.002	**0.489** ±0.002
		MAE	0.651 ±0.028	**0.634** ±0.018	0.432 ±0.000	**0.432** ±0.001	**0.433** ±0.001	0.438 ±0.001	0.431 ±0.000	**0.429** ±0.001	0.544 ±0.000	**0.543** ±0.002	0.430 ±0.001	**0.429** ±0.001	0.451 ±0.000	**0.450** ±0.001	0.459 ±0.000	**0.453** ±0.000	0.441 ±0.001	**0.441** ±0.001
	Avg	MSE	0.679 ±0.029	**0.653** ±0.025	0.397 ±0.001	**0.395** ±0.001	0.401 ±0.001	**0.399** ±0.001	**0.391** ±0.001	0.392 ±0.001	0.667 ±0.013	**0.655** ±0.007	0.398 ±0.000	**0.397** ±0.002	0.428 ±0.000	**0.427** ±0.000	0.423 ±0.000	**0.418** ±0.000	0.410 ±0.001	**0.410** ±0.001
		MAE	0.578 ±0.009	**0.566** ±0.012	0.387 ±0.001	**0.387** ±0.002	**0.391** ±0.001	0.391 ±0.000	**0.384** ±0.000	0.386 ±0.001	**0.500** ±0.004	0.501 ±0.000	**0.387** ±0.001	0.387 ±0.001	0.412 ±0.000	**0.410** ±0.000	0.415 ±0.000	**0.410** ±0.000	**0.397** ±0.001	0.398 ±0.001
	Imp.	MSE	–	**3.88%**↑	–	**0.31%**↑	–	**0.28%**↑	–	0.06%↓	–	**1.78%**↑	–	**0.25%**↑	–	**0.23%**↑	–	**1.13%**↑	–	**0.12%**↑
		MAE	–	**2.02%**↑	–	**0.04%**↑	–	0.04%↓	–	0.28%↓	–	0.08%↓	–	0.05%↓	–	**0.51%**↑	–	**1.14%**↑	–	0.26%↓
ETTm2	96	MSE	0.212 ±0.001	**0.208** ±0.001	**0.175** ±0.001	0.175 ±0.000	0.181 ±0.001	**0.181** ±0.001	**0.172** ±0.001	0.173 ±0.000	0.191 ±0.006	**0.187** ±0.002	0.179 ±0.000	**0.179** ±0.000	0.195 ±0.000	**0.194** ±0.000	0.192 ±0.000	**0.192** ±0.000	0.186 ±0.002	**0.184** ±0.001
		MAE	0.293 ±0.000	**0.289** ±0.001	**0.252** ±0.001	0.253 ±0.001	**0.257** ±0.000	0.257 ±0.000	**0.251** ±0.000	0.252 ±0.000	0.274 ±0.005	**0.270** ±0.001	**0.258** ±0.000	0.258 ±0.001	**0.270** ±0.000	0.271 ±0.000	**0.272** ±0.000	0.273 ±0.000	0.265 ±0.002	**0.264** ±0.001
	192	MSE	0.279 ±0.014	**0.267** ±0.001	**0.241** ±0.001	0.242 ±0.000	**0.245** ±0.000	0.248 ±0.000	**0.237** ±0.001	0.240 ±0.001	0.268 ±0.002	**0.264** ±0.003	0.244 ±0.001	0.246 ±0.002	0.257 ±0.000	**0.256** ±0.000	0.257 ±0.000	**0.257** ±0.000	0.251 ±0.001	**0.249** ±0.001
		MAE	0.336 ±0.015	**0.325** ±0.001	**0.296** ±0.001	0.298 ±0.001	**0.298** ±0.000	0.300 ±0.000	**0.294** ±0.000	0.295 ±0.001	0.326 ±0.002	**0.323** ±0.003	0.301 ±0.000	**0.300** ±0.000	**0.308** ±0.000	0.308 ±0.000	**0.312** ±0.000	0.313 ±0.000	**0.305** ±0.001	0.305 ±0.000
	336	MSE	0.336 ±0.012	**0.323** ±0.003	**0.303** ±0.001	0.307 ±0.000	**0.306** ±0.000	0.307 ±0.002	**0.296** ±0.001	0.299 ±0.001	0.335 ±0.007	**0.333** ±0.001	**0.305** ±0.000	0.306 ±0.001	0.316 ±0.000	**0.315** ±0.000	**0.317** ±0.000	0.318 ±0.000	0.311 ±0.001	**0.311** ±0.001
		MAE	0.370 ±0.011	**0.359** ±0.002	**0.336** ±0.000	0.340 ±0.001	**0.337** ±0.000	0.337 ±0.001	**0.333** ±0.000	0.334 ±0.001	0.369 ±0.005	**0.367** ±0.001	0.339 ±0.000	**0.338** ±0.000	**0.344** ±0.000	0.345 ±0.000	**0.350** ±0.000	0.351 ±0.000	**0.342** ±0.000	0.343 ±0.000
	720	MSE	0.468 ±0.022	**0.426** ±0.006	0.408 ±0.000	**0.408** ±0.004	**0.405** ±0.000	0.405 ±0.000	**0.398** ±0.001	0.402 ±0.002	0.429 ±0.001	**0.427** ±0.004	**0.406** ±0.002	0.408 ±0.003	0.416 ±0.000	**0.415** ±0.000	**0.418** ±0.000	0.419 ±0.000	0.412 ±0.001	**0.412** ±0.001
		MAE	0.441 ±0.014	**0.419** ±0.006	0.398 ±0.001	**0.396** ±0.000	0.394 ±0.000	**0.394** ±0.000	**0.392** ±0.001	0.393 ±0.000	0.419 ±0.001	**0.419** ±0.001	0.398 ±0.000	**0.397** ±0.001	**0.400** ±0.000	0.401 ±0.000	0.406 ±0.000	**0.406** ±0.000	**0.399** ±0.000	0.399 ±0.001
	Avg	MSE	0.324 ±0.010	**0.306** ±0.001	**0.282** ±0.000	0.283 ±0.001	**0.284** ±0.000	0.285 ±0.001	**0.276** ±0.000	0.278 ±0.000	0.306 ±0.004	**0.303** ±0.002	**0.283** ±0.000	0.285 ±0.001	0.296 ±0.000	**0.295** ±0.000	**0.296** ±0.000	0.296 ±0.000	0.290 ±0.001	**0.289** ±0.001
		MAE	0.360 ±0.009	**0.348** ±0.001	**0.321** ±0.000	0.322 ±0.001	**0.321** ±0.000	0.322 ±0.000	**0.317** ±0.000	0.318 ±0.001	0.347 ±0.003	**0.345** ±0.002	0.324 ±0.000	**0.323** ±0.000	**0.330** ±0.000	0.331 ±0.000	**0.335** ±0.000	0.336 ±0.000	**0.328** ±0.001	0.328 ±0.000
	Imp.	MSE	–	**5.40%**↑	–	0.44%↓	–	0.39%↓	–	0.87%↓	–	**0.94%**↑	–	0.51%↓	–	**0.28%**↑	–	0.04%↓	–	**0.29%**↑
		MAE	–	**3.30%**↑	–	0.39%↓	–	0.18%↓	–	0.29%↓	–	**0.72%**↑	–	**0.23%**↑	–	0.27%↓	–	0.18%↓	–	0.07%↓
ETTh1	96	MSE	0.603 ±0.025	**0.570** ±0.017	0.396 ±0.005	**0.396** ±0.005	**0.382** ±0.001	0.383 ±0.001	**0.383** ±0.002	0.384 ±0.004	0.736 ±0.004	**0.736** ±0.009	0.397 ±0.002	**0.394** ±0.002	0.396 ±0.000	**0.396** ±0.000	0.400 ±0.000	**0.399** ±0.000	0.420 ±0.002	**0.412** ±0.003
		MAE	0.559 ±0.009	**0.539** ±0.010	0.396 ±0.002	**0.396** ±0.004	**0.391** ±0.001	0.391 ±0.000	**0.388** ±0.001	0.389 ±0.001	0.521 ±0.002	**0.520** ±0.002	0.393 ±0.003	**0.390** ±0.000	0.394 ±0.000	**0.392** ±0.000	0.406 ±0.000	**0.405** ±0.000	0.407 ±0.002	**0.403** ±0.003
	192	MSE	0.671 ±0.024	**0.607** ±0.025	0.453 ±0.008	**0.445** ±0.005	0.439 ±0.002	**0.434** ±0.002	0.436 ±0.007	**0.432** ±0.007	0.752 ±0.010	**0.740** ±0.016	0.447 ±0.004	**0.445** ±0.002	**0.446** ±0.000	0.446 ±0.000	0.449 ±0.000	**0.448** ±0.000	0.469 ±0.002	**0.466** ±0.002
		MAE	0.596 ±0.005	**0.557** ±0.015	0.425 ±0.003	**0.424** ±0.002	0.423 ±0.001	**0.420** ±0.000	0.420 ±0.001	**0.418** ±0.002	**0.535** ±0.003	0.538 ±0.004	0.422 ±0.003	**0.421** ±0.002	**0.423** ±0.000	0.424 ±0.000	0.436 ±0.000	**0.435** ±0.000	**0.436** ±0.000	0.438 ±0.001
	336	MSE	0.699 ±0.023	**0.611** ±0.038	0.493 ±0.004	**0.486** ±0.002	0.481 ±0.003	**0.477** ±0.002	0.492 ±0.011	**0.476** ±0.001	0.730 ±0.006	**0.728** ±0.005	0.491 ±0.000	**0.490** ±0.000	0.491 ±0.000	**0.489** ±0.000	0.490 ±0.000	**0.488** ±0.000	0.513 ±0.003	**0.506** ±0.003
		MAE	0.606 ±0.007	**0.560** ±0.015	0.445 ±0.003	**0.443** ±0.001	0.444 ±0.001	**0.442** ±0.001	0.447 ±0.004	**0.440** ±0.002	**0.545** ±0.001	0.546 ±0.005	0.442 ±0.000	**0.440** ±0.001	**0.445** ±0.000	0.447 ±0.000	0.459 ±0.000	**0.457** ±0.000	0.456 ±0.001	**0.455** ±0.003
	720	MSE	0.808 ±0.075	**0.659** ±0.035	0.516 ±0.002	**0.497** ±0.005	0.479 ±0.002	**0.459** ±0.001	0.544 ±0.042	**0.491** ±0.023	0.717 ±0.003	**0.707** ±0.005	0.497 ±0.010	**0.485** ±0.003	0.477 ±0.000	**0.474** ±0.000	0.493 ±0.000	**0.485** ±0.000	0.505 ±0.005	**0.496** ±0.006
		MAE	0.670 ±0.028	**0.598** ±0.015	0.476 ±0.005	**0.466** ±0.001	0.466 ±0.001	**0.456** ±0.001	0.491 ±0.020	**0.467** ±0.013	0.568 ±0.003	**0.564** ±0.003	0.469 ±0.006	**0.459** ±0.002	**0.460** ±0.000	0.461 ±0.000	0.485 ±0.000	**0.479** ±0.000	**0.469** ±0.001	0.471 ±0.002
	Avg	MSE	0.695 ±0.035	**0.612** ±0.021	0.464 ±0.001	**0.456** ±0.004	0.445 ±0.000	**0.438** ±0.001	0.464 ±0.012	**0.446** ±0.003	0.734 ±0.004	**0.728** ±0.007	0.458 ±0.002	**0.454** ±0.000	0.453 ±0.000	**0.451** ±0.000	0.458 ±0.000	**0.455** ±0.000	0.477 ±0.002	**0.470** ±0.003
		MAE	0.608 ±0.011	**0.563** ±0.011	0.436 ±0.001	**0.432** ±0.001	0.431 ±0.000	**0.427** ±0.000	0.437 ±0.005	**0.428** ±0.003	0.542 ±0.001	**0.542** ±0.002	0.431 ±0.001	**0.428** ±0.001	**0.430** ±0.000	0.431 ±0.000	0.447 ±0.000	**0.444** ±0.000	0.442 ±0.001	**0.442** ±0.001
	Imp.	MSE	–	**12.04%**↑	–	**1.75%**↑	–	**1.63%**↑	–	**3.90%**↑	–	**0.79%**↑	–	**0.96%**↑	–	**0.27%**↑	–	**0.67%**↑	–	**1.35%**↑
		MAE	–	**7.27%**↑	–	**0.79%**↑	–	**0.85%**↑	–	**1.84%**↑	–	**0.03%**↑	–	**0.89%**↑	–	0.19%↓	–	**0.56%**↑	–	**0.05%**↑
ETTh2	96	MSE	0.553 ±0.007	**0.446** ±0.011	0.297 ±0.001	**0.294** ±0.006	0.296 ±0.002	**0.292** ±0.001	0.292 ±0.004	**0.290** ±0.002	0.342 ±0.000	**0.338** ±0.004	**0.297** ±0.001	0.299 ±0.003	**0.312** ±0.000	0.315 ±0.000	**0.307** ±0.000	0.312 ±0.000	**0.302** ±0.000	0.307 ±0.002
		MAE	0.511 ±0.002	**0.448** ±0.007	0.339 ±0.000	**0.337** ±0.001	**0.338** ±0.001	0.338 ±0.001	0.337 ±0.002	**0.335** ±0.002	0.385 ±0.001	**0.382** ±0.003	**0.339** ±0.001	0.340 ±0.001	**0.350** ±0.000	0.351 ±0.000	**0.354** ±0.000	0.357 ±0.000	**0.346** ±0.000	0.347 ±0.001
	192	MSE	0.676 ±0.033	**0.636** ±0.012	0.376 ±0.003	**0.369** ±0.003	0.373 ±0.000	**0.369** ±0.001	0.369 ±0.003	**0.363** ±0.000	0.413 ±0.003	**0.405** ±0.009	0.383 ±0.002	**0.378** ±0.006	**0.397** ±0.000	0.400 ±0.000	**0.397** ±0.000	0.401 ±0.000	0.399 ±0.010	**0.397** ±0.001
		MAE	0.582 ±0.016	**0.549** ±0.002	0.390 ±0.002	**0.388** ±0.001	0.388 ±0.001	**0.387** ±0.000	0.387 ±0.001	**0.384** ±0.001	0.425 ±0.001	**0.420** ±0.004	0.392 ±0.001	**0.391** ±0.002	**0.401** ±0.000	0.402 ±0.000	**0.408** ±0.000	0.409 ±0.000	0.404 ±0.005	**0.399** ±0.000
	336	MSE	0.678 ±0.023	**0.613** ±0.002	0.430 ±0.008	**0.411** ±0.004	0.431 ±0.001	**0.412** ±0.002	0.422 ±0.002	**0.413** ±0.005	0.458 ±0.004	**0.441** ±0.005	0.430 ±0.001	**0.427** ±0.004	**0.440** ±0.000	0.441 ±0.000	**0.443** ±0.000	0.446 ±0.001	0.448 ±0.007	**0.444** ±0.002
		MAE	0.597 ±0.006	**0.548** ±0.001	0.432 ±0.002	**0.425** ±0.002	0.430 ±0.000	**0.422** ±0.000	0.429 ±0.002	**0.424** ±0.003	0.458 ±0.002	**0.448** ±0.002	0.431 ±0.001	**0.430** ±0.002	**0.437** ±0.000	0.438 ±0.000	**0.446** ±0.000	0.447 ±0.000	0.441 ±0.004	**0.439** ±0.001
	720	MSE	0.686 ±0.055	**0.623** ±0.054	0.442 ±0.004	**0.425** ±0.001	0.435 ±0.007	**0.421** ±0.003	0.435 ±0.002	**0.418** ±0.005	0.480 ±0.001	**0.462** ±0.001	0.440 ±0.004	**0.432** ±0.005	**0.438** ±0.000	0.440 ±0.000	0.453 ±0.000	**0.443** ±0.000	0.468 ±0.007	**0.438** ±0.002
		MAE	0.599 ±0.019	**0.569** ±0.026	0.451 ±0.002	**0.442** ±0.002	0.445 ±0.004	**0.438** ±0.002	0.446 ±0.001	**0.434** ±0.002	0.476 ±0.002	**0.467** ±0.001	0.448 ±0.001	**0.443** ±0.003	**0.448** ±0.000	0.449 ±0.000	0.461 ±0.000	**0.454** ±0.000	0.463 ±0.003	**0.446** ±0.001
	Avg	MSE	0.648 ±0.020	**0.580** ±0.018	0.386 ±0.001	**0.374** ±0.003	0.384 ±0.001	**0.374** ±0.001	0.379 ±0.002	**0.371** ±0.003	0.423 ±0.001	**0.412** ±0.003	0.388 ±0.002	**0.384** ±0.002	**0.397** ±0.000	0.399 ±0.000	**0.400** ±0.000	0.401 ±0.000	0.404 ±0.002	**0.397** ±0.000
		MAE	0.572 ±0.008	**0.529** ±0.009	0.403 ±0.001	**0.398** ±0.001	0.400 ±0.001	**0.396** ±0.000	0.399 ±0.001	**0.394** ±0.002	0.436 ±0.001	**0.429** ±0.001	0.403 ±0.001	**0.401** ±0.000	**0.409** ±0.000	0.410 ±0.000	0.417 ±0.000	**0.417** ±0.000	0.414 ±0.001	**0.408** ±0.000
	Imp.	MSE	–	**10.60%**↑	–	**3.02%**↑	–	**2.58%**↑	–	**2.25%**↑	–	**2.67%**↑	–	**0.90%**↑	–	0.56%↓	–	0.21%↓	–	**1.96%**↑
		MAE	–	**7.63%**↑	–	**1.33%**↑	–	**1.01%**↑	–	**1.33%**↑	–	**1.51%**↑	–	**0.49%**↑	–	0.18%↓	–	**0.03%**↑	–	**1.39%**↑

Table 8: Full forecasting results on COVID19, NABCPU, Sunspots, and ExchangeRate. Values are means over three runs, with sample standard deviations below. Avg summarizes the four-horizon mean within each run using the same statistics. Within each model and metric, the lower unrounded mean is bold and the other underlined; exact ties are both bold. Gray cells denote +StarTS. Shaded Imp. rows report relative changes in the four-horizon means, with reductions bold and arrows indicating direction.

Dataset	H	Metric	Autoformer (2021)		PatchTST (2023)		iTransformer (2024)		WPMixer (2025)		MultiPatchFormer (2025)		TimeFilter (2025)		TimeBase (2025)		OLinear (2025)		TimeMixer++ (2025)	
			Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS	Raw	+StarTS
COVID19	7	MSE	1.151 ±0.099	**1.108** ±0.057	0.338 ±0.006	**0.333** ±0.014	0.334 ±0.003	**0.333** ±0.011	0.349 ±0.005	**0.333** ±0.005	0.693 ±0.021	**0.607** ±0.051	0.342 ±0.004	**0.329** ±0.002	0.641 ±0.008	**0.615** ±0.005	1.247 ±0.092	**0.930** ±0.077	0.380 ±0.009	**0.355** ±0.006
		MAE	0.657 ±0.046	**0.652** ±0.028	**0.219** ±0.003	0.220 ±0.002	**0.220** ±0.000	0.222 ±0.002	**0.222** ±0.001	0.225 ±0.001	0.476 ±0.012	**0.411** ±0.021	**0.229** ±0.001	0.236 ±0.001	0.369 ±0.004	**0.362** ±0.002	0.516 ±0.015	**0.445** ±0.017	0.277 ±0.006	**0.267** ±0.004
	14	MSE	**1.764** ±0.043	1.764 ±0.059	0.446 ±0.012	**0.445** ±0.029	0.442 ±0.004	**0.422** ±0.032	0.468 ±0.003	**0.424** ±0.007	1.300 ±0.481	**0.837** ±0.092	0.454 ±0.006	**0.421** ±0.001	0.841 ±0.004	**0.816** ±0.007	1.537 ±0.168	**1.205** ±0.060	0.488 ±0.003	**0.438** ±0.002
		MAE	**0.785** ±0.013	0.786 ±0.072	**0.266** ±0.003	0.269 ±0.006	0.270 ±0.001	**0.268** ±0.007	**0.275** ±0.001	0.276 ±0.001	0.592 ±0.072	**0.486** ±0.018	0.284 ±0.002	**0.282** ±0.002	0.430 ±0.004	**0.423** ±0.004	0.562 ±0.037	**0.504** ±0.012	0.337 ±0.002	**0.314** ±0.003
	28	MSE	2.538 ±0.028	**2.479** ±0.160	**0.790** ±0.005	0.794 ±0.107	0.794 ±0.021	**0.774** ±0.101	0.830 ±0.021	**0.727** ±0.029	1.507 ±0.862	**1.023** ±0.116	0.809 ±0.022	**0.736** ±0.009	1.298 ±0.040	**1.259** ±0.030	2.218 ±0.077	**1.899** ±0.077	0.896 ±0.024	**0.787** ±0.062
		MAE	0.882 ±0.034	**0.860** ±0.056	**0.378** ±0.004	0.380 ±0.019	**0.383** ±0.004	0.383 ±0.018	0.392 ±0.001	**0.381** ±0.007	0.628 ±0.107	**0.531** ±0.023	0.403 ±0.006	**0.392** ±0.003	0.543 ±0.010	**0.531** ±0.010	0.687 ±0.010	**0.632** ±0.011	0.465 ±0.010	**0.428** ±0.015
	60	MSE	3.317 ±0.434	**3.231** ±0.495	**1.630** ±0.018	1.645 ±0.135	**1.722** ±0.005	1.738 ±0.278	1.761 ±0.135	**1.630** ±0.184	2.223 ±0.223	**2.186** ±0.152	1.962 ±0.052	**1.771** ±0.136	2.359 ±0.012	**2.283** ±0.004	2.832 ±0.018	**2.528** ±0.029	1.851 ±0.121	**1.839** ±0.112
		MAE	1.025 ±0.141	**0.994** ±0.163	**0.567** ±0.003	0.570 ±0.026	**0.585** ±0.002	0.592 ±0.044	0.583 ±0.023	**0.567** ±0.030	0.701 ±0.015	**0.692** ±0.013	0.627 ±0.009	**0.597** ±0.022	0.725 ±0.006	**0.714** ±0.006	0.777 ±0.007	**0.724** ±0.006	**0.616** ±0.019	0.621 ±0.016
	Avg	MSE	2.193 ±0.096	**2.145** ±0.179	**0.801** ±0.007	0.804 ±0.071	0.823 ±0.006	**0.817** ±0.104	0.852 ±0.037	**0.778** ±0.044	1.431 ±0.318	**1.163** ±0.075	0.892 ±0.008	**0.814** ±0.036	1.285 ±0.014	**1.243** ±0.010	1.958 ±0.075	**1.641** ±0.041	0.904 ±0.028	**0.855** ±0.033
		MAE	0.837 ±0.025	**0.823** ±0.069	**0.358** ±0.001	0.360 ±0.013	**0.365** ±0.001	0.366 ±0.017	0.368 ±0.006	**0.362** ±0.008	0.599 ±0.043	**0.530** ±0.008	0.386 ±0.002	**0.377** ±0.006	0.517 ±0.003	**0.507** ±0.004	0.635 ±0.014	**0.576** ±0.007	0.424 ±0.004	**0.408** ±0.007
	Imp.	MSE	–	**2.15%↑**	–	0.39%↓	–	**0.75%↑**	–	**8.61%↑**	–	**18.69%↑**	–	**8.69%↑**	–	**3.22%↑**	–	**16.23%↑**	–	**5.44%↑**
		MAE	–	**1.70%↑**	–	0.62%↓	–	0.48%↓	–	**1.58%↑**	–	**11.53%↑**	–	**2.24%↑**	–	**1.82%↑**	–	**9.33%↑**	–	**3.82%↑**
NABCPU	24	MSE	1.245 ±0.046	**1.195** ±0.030	1.136 ±0.003	**1.135** ±0.004	**1.134** ±0.000	1.136 ±0.003	**1.142** ±0.003	1.142 ±0.003	**1.149** ±0.003	1.150 ±0.003	1.138 ±0.001	**1.134** ±0.002	1.297 ±0.003	**1.276** ±0.001	1.149 ±0.002	**1.145** ±0.001	1.148 ±0.003	**1.148** ±0.008
		MAE	0.493 ±0.105	**0.416** ±0.031	0.313 ±0.006	**0.306** ±0.008	**0.329** ±0.003	0.331 ±0.002	0.328 ±0.004	**0.327** ±0.002	**0.305** ±0.001	0.305 ±0.001	0.320 ±0.002	**0.319** ±0.002	0.425 ±0.000	**0.422** ±0.001	**0.333** ±0.002	0.338 ±0.005	0.315 ±0.003	**0.315** ±0.009
	48	MSE	1.321 ±0.147	**1.261** ±0.039	**1.150** ±0.001	1.152 ±0.001	1.155 ±0.002	**1.155** ±0.002	**1.156** ±0.002	1.158 ±0.004	1.161 ±0.002	**1.161** ±0.003	1.147 ±0.001	**1.146** ±0.003	1.322 ±0.005	**1.317** ±0.001	1.161 ±0.001	**1.157** ±0.000	**1.162** ±0.006	1.165 ±0.009
		MAE	0.544 ±0.187	**0.472** ±0.066	0.316 ±0.001	**0.316** ±0.003	**0.336** ±0.001	0.338 ±0.002	**0.327** ±0.003	0.328 ±0.003	**0.306** ±0.002	0.307 ±0.000	**0.315** ±0.000	0.316 ±0.001	0.433 ±0.000	**0.430** ±0.001	**0.332** ±0.001	0.338 ±0.002	**0.316** ±0.005	0.335 ±0.016
	96	MSE	1.454 ±0.219	**1.391** ±0.143	1.203 ±0.003	**1.202** ±0.003	**1.209** ±0.002	1.211 ±0.004	**1.210** ±0.003	1.212 ±0.004	**1.207** ±0.001	1.207 ±0.002	1.195 ±0.002	**1.195** ±0.002	1.372 ±0.005	**1.352** ±0.005	**1.212** ±0.000	1.212 ±0.000	1.217 ±0.009	**1.212** ±0.011
		MAE	0.602 ±0.221	**0.561** ±0.166	0.320 ±0.006	**0.318** ±0.005	**0.341** ±0.001	0.347 ±0.009	0.332 ±0.003	**0.328** ±0.003	0.310 ±0.002	**0.309** ±0.001	0.317 ±0.001	**0.316** ±0.001	0.440 ±0.000	**0.435** ±0.001	0.336 ±0.001	**0.335** ±0.001	0.327 ±0.007	**0.323** ±0.011
	192	MSE	1.690 ±0.356	**1.490** ±0.074	1.301 ±0.003	**1.299** ±0.003	**1.301** ±0.001	1.302 ±0.000	1.297 ±0.005	**1.296** ±0.005	1.307 ±0.004	**1.298** ±0.004	1.292 ±0.002	1.294 ±0.002	1.423 ±0.004	**1.399** ±0.003	1.300 ±0.000	**1.298** ±0.002	1.312 ±0.004	**1.309** ±0.002
		MAE	0.722 ±0.280	**0.584** ±0.042	0.332 ±0.005	**0.330** ±0.007	**0.343** ±0.003	0.344 ±0.002	**0.339** ±0.003	0.339 ±0.002	0.327 ±0.005	**0.321** ±0.003	**0.325** ±0.000	0.326 ±0.001	0.437 ±0.000	**0.434** ±0.000	**0.344** ±0.001	0.348 ±0.003	0.335 ±0.004	**0.333** ±0.003
	Avg	MSE	1.427 ±0.190	**1.334** ±0.025	1.197 ±0.001	**1.197** ±0.000	**1.200** ±0.001	1.201 ±0.002	**1.201** ±0.001	1.202 ±0.003	1.206 ±0.002	**1.204** ±0.002	1.193 ±0.001	**1.192** ±0.001	1.354 ±0.005	**1.336** ±0.000	1.205 ±0.001	**1.203** ±0.001	1.210 ±0.002	**1.208** ±0.006
		MAE	0.590 ±0.197	**0.508** ±0.055	0.320 ±0.002	**0.317** ±0.002	**0.337** ±0.001	0.340 ±0.003	0.331 ±0.001	**0.331** ±0.002	0.312 ±0.002	**0.311** ±0.001	**0.319** ±0.001	0.319 ±0.001	0.434 ±0.000	**0.431** ±0.000	**0.336** ±0.001	0.340 ±0.002	**0.323** ±0.002	0.326 ±0.008
	Imp.	MSE	–	**6.52%↑**	–	**0.02%↑**	–	0.06%↓	–	0.09%↓	–	**0.16%↑**	–	**0.06%↑**	–	**1.30%↑**	–	**0.19%↑**	–	**0.11%↑**
		MAE	–	**13.90%↑**	–	**0.87%↑**	–	0.80%↓	–	**0.18%↑**	–	**0.42%↑**	–	0.06%↓	–	**0.77%↑**	–	1.10%↓	–	0.83%↓
Sunspots	12	MSE	0.354 ±0.044	**0.319** ±0.039	**0.261** ±0.002	0.264 ±0.007	**0.272** ±0.004	0.276 ±0.019	0.258 ±0.009	**0.258** ±0.004	**0.234** ±0.014	0.238 ±0.010	**0.300** ±0.001	0.305 ±0.001	0.392 ±0.001	**0.391** ±0.001	**0.330** ±0.005	0.332 ±0.006	0.287 ±0.024	**0.278** ±0.005
		MAE	0.446 ±0.021	**0.430** ±0.037	**0.373** ±0.002	0.376 ±0.002	**0.380** ±0.002	0.382 ±0.011	**0.372** ±0.005	0.373 ±0.003	**0.354** ±0.011	0.360 ±0.006	**0.397** ±0.001	0.399 ±0.001	0.454 ±0.001	**0.454** ±0.001	**0.414** ±0.003	0.414 ±0.003	0.389 ±0.017	**0.380** ±0.005
	24	MSE	0.632 ±0.081	**0.576** ±0.069	**0.473** ±0.005	0.476 ±0.018	**0.468** ±0.007	0.469 ±0.024	0.450 ±0.013	**0.447** ±0.007	0.411 ±0.030	**0.396** ±0.012	**0.558** ±0.004	0.569 ±0.012	**0.717** ±0.002	0.721 ±0.002	0.722 ±0.031	**0.720** ±0.030	**0.482** ±0.063	0.517 ±0.021
		MAE	0.617 ±0.049	**0.578** ±0.031	**0.493** ±0.004	0.495 ±0.002	0.488 ±0.003	**0.488** ±0.011	0.480 ±0.005	**0.478** ±0.003	0.450 ±0.013	**0.447** ±0.005	**0.528** ±0.001	0.533 ±0.004	**0.611** ±0.000	0.612 ±0.000	0.610 ±0.017	**0.608** ±0.016	**0.493** ±0.039	0.506 ±0.011
	48	MSE	1.164 ±0.130	**1.113** ±0.115	1.064 ±0.032	**1.049** ±0.030	1.039 ±0.038	**1.038** ±0.018	1.087 ±0.011	**1.075** ±0.047	0.898 ±0.044	**0.893** ±0.057	**1.189** ±0.014	1.197 ±0.007	1.463 ±0.002	**1.452** ±0.003	1.599 ±0.032	**1.591** ±0.033	1.002 ±0.137	**0.881** ±0.043
		MAE	0.852 ±0.049	**0.837** ±0.027	0.736 ±0.007	**0.735** ±0.005	**0.719** ±0.017	0.731 ±0.013	0.737 ±0.004	**0.726** ±0.019	**0.663** ±0.011	0.665 ±0.017	0.784 ±0.007	**0.784** ±0.003	**0.910** ±0.000	0.911 ±0.000	0.951 ±0.015	**0.947** ±0.016	0.710 ±0.055	**0.661** ±0.016
	96	MSE	1.892 ±0.100	**1.725** ±0.271	**1.143** ±0.014	1.160 ±0.013	1.070 ±0.027	**1.057** ±0.026	1.118 ±0.042	**1.096** ±0.036	1.077 ±0.029	**1.050** ±0.070	**1.199** ±0.008	1.213 ±0.021	**1.504** ±0.001	1.511 ±0.001	1.685 ±0.022	**1.681** ±0.021	**1.055** ±0.047	1.090 ±0.128
		MAE	1.125 ±0.024	**1.062** ±0.106	**0.779** ±0.002	0.789 ±0.002	0.755 ±0.006	**0.747** ±0.006	0.766 ±0.015	**0.762** ±0.020	0.754 ±0.006	**0.735** ±0.024	**0.800** ±0.004	0.809 ±0.005	**0.917** ±0.000	0.920 ±0.000	**1.005** ±0.010	1.008 ±0.010	**0.747** ±0.017	0.754 ±0.052
	Avg	MSE	1.011 ±0.074	**0.933** ±0.113	**0.735** ±0.012	0.737 ±0.015	0.712 ±0.016	**0.710** ±0.020	0.728 ±0.011	**0.719** ±0.022	0.655 ±0.011	**0.645** ±0.031	**0.812** ±0.005	0.821 ±0.009	1.019 ±0.001	**1.019** ±0.001	1.084 ±0.022	**1.081** ±0.022	0.706 ±0.046	**0.692** ±0.029
		MAE	0.760 ±0.031	**0.727** ±0.048	**0.595** ±0.003	0.599 ±0.002	**0.585** ±0.006	0.587 ±0.004	0.589 ±0.005	**0.585** ±0.011	0.555 ±0.004	**0.552** ±0.011	**0.627** ±0.002	0.631 ±0.003	**0.723** ±0.000	0.724 ±0.000	0.745 ±0.011	**0.744** ±0.011	0.585 ±0.022	**0.575** ±0.013
	Imp.	MSE	–	**7.66%↑**	–	0.25%↓	–	**0.32%↑**	–	**1.28%↑**	–	**1.60%↑**	–	1.16%↓	–	**0.00%↑**	–	**0.29%↑**	–	**2.08%↑**
		MAE	–	**4.37%↑**	–	0.58%↓	–	0.32%↓	–	**0.70%↑**	–	**0.66%↑**	–	0.67%↓	–	0.15%↓	–	**0.07%↑**	–	**1.63%↑**
ExchangeRate	96	MSE	0.196 ±0.025	**0.195** ±0.031	0.084 ±0.003	**0.082** ±0.003	**0.080** ±0.001	0.080 ±0.001	**0.082** ±0.001	0.082 ±0.003	**0.105** ±0.001	0.107 ±0.001	0.083 ±0.001	**0.082** ±0.000	0.084 ±0.000	**0.083** ±0.000	**0.093** ±0.000	0.094 ±0.000	0.086 ±0.002	**0.085** ±0.001
		MAE	0.325 ±0.021	**0.325** ±0.027	0.204 ±0.003	**0.203** ±0.003	**0.201** ±0.001	0.202 ±0.001	**0.200** ±0.001	0.203 ±0.004	**0.238** ±0.000	0.243 ±0.001	**0.205** ±0.001	0.205 ±0.000	0.206 ±0.000	**0.206** ±0.000	**0.221** ±0.000	0.222 ±0.000	0.211 ±0.002	**0.211** ±0.001
	192	MSE	0.384 ±0.061	**0.352** ±0.059	**0.173** ±0.003	0.174 ±0.004	**0.169** ±0.001	0.163 ±0.001	0.169 ±0.005	**0.161** ±0.004	0.192 ±0.005	**0.190** ±0.005	0.170 ±0.001	**0.160** ±0.001	0.167 ±0.000	**0.162** ±0.000	0.184 ±0.000	**0.176** ±0.000	0.172 ±0.001	**0.167** ±0.002
		MAE	0.470 ±0.050	**0.451** ±0.046	**0.300** ±0.003	0.302 ±0.003	**0.291** ±0.001	0.291 ±0.001	0.296 ±0.004	**0.292** ±0.004	**0.324** ±0.004	0.324 ±0.003	0.299 ±0.001	**0.293** ±0.001	0.297 ±0.000	**0.293** ±0.000	0.315 ±0.000	**0.311** ±0.000	0.303 ±0.001	**0.300** ±0.001
	336	MSE	0.537 ±0.155	**0.480** ±0.075	0.353 ±0.011	**0.344** ±0.006	0.325 ±0.001	**0.324** ±0.003	0.346 ±0.011	**0.323** ±0.010	0.342 ±0.008	**0.328** ±0.007	0.335 ±0.012	**0.320** ±0.009	0.329 ±0.006	**0.302** ±0.000	0.358 ±0.001	**0.333** ±0.001	0.332 ±0.006	**0.307** ±0.002
		MAE	0.555 ±0.085	**0.526** ±0.052	0.434 ±0.007	**0.429** ±0.005	0.417 ±0.001	**0.417** ±0.001	0.429 ±0.006	**0.414** ±0.006	0.433 ±0.004	**0.427** ±0.005	0.423 ±0.008	**0.413** ±0.005	0.422 ±0.005	**0.402** ±0.000	0.441 ±0.000	**0.426** ±0.000	0.422 ±0.004	**0.407** ±0.001
	480	MSE	**0.731** ±0.078	0.764 ±0.128	0.520 ±0.006	**0.518** ±0.007	**0.474** ±0.001	0.478 ±0.007	0.531 ±0.011	**0.421** ±0.004	0.496 ±0.017	**0.455** ±0.009	0.500 ±0.009	**0.459** ±0.007	0.486 ±0.000	**0.460** ±0.001	0.522 ±0.001	**0.488** ±0.001	0.505 ±0.008	**0.483** ±0.012
		MAE	**0.671** ±0.035	0.687 ±0.060	0.536 ±0.003	**0.535** ±0.003	**0.507** ±0.000	0.510 ±0.005	0.541 ±0.006	**0.476** ±0.002	0.521 ±0.012	**0.506** ±0.006	0.524 ±0.006	**0.501** ±0.004	0.517 ±0.000	**0.500** ±0.001	0.540 ±0.000	**0.521** ±0.000	0.525 ±0.005	**0.511** ±0.007
	Avg	MSE	0.462 ±0.056	**0.448** ±0.052	0.282 ±0.004	**0.280** ±0.001	**0.260** ±0.001	0.261 ±0.001	0.282 ±0.001	**0.247** ±0.003	0.284 ±0.005	**0.270** ±0.004	0.272 ±0.005	**0.255** ±0.001	0.267 ±0.001	**0.252** ±0.000	0.290 ±0.000	**0.273** ±0.000	0.274 ±0.003	**0.260** ±0.004
		MAE	0.505 ±0.031	**0.497** ±0.032	0.369 ±0.003	**0.367** ±0.001	**0.354** ±0.000	0.355 ±0.002	0.367 ±0.001	**0.346** ±0.003	0.379 ±0.003	**0.375** ±0.002	0.363 ±0.003	**0.353** ±0.001	0.360 ±0.001	**0.350** ±0.000	0.379 ±0.000	**0.370** ±0.000	0.365 ±0.002	**0.357** ±0.002
	Imp.	MSE	–	**3.07%↑**	–	**0.85%↑**	–	0.36%↓	–	**12.55%↑**	–	**4.78%↑**	–	**6.15%↑**	–	**5.47%↑**	–	**5.81%↑**	–	**4.91%↑**
		MAE	–	**1.57%↑**	–	**0.40%↑**	–	0.37%↓	–	**5.60%↑**	–	**1.12%↑**	–	**2.64%↑**	–	**2.77%↑**	–	**2.44%↑**	–	**2.21%↑**

Benchmark summary. StarTS benefits multiple forecasting backbones through sample reweighting while preserving their architecture and inference procedure. Autoformer improves on all eight datasets, including four-horizon average MSE reductions of **12.04%** on ETTh1 and **10.60%** on ETTh2 relative to its own baseline. The benefit extends to PatchTST and iTransformer, whose corresponding reductions on ETTh2 are 3.02% and 2.58%. The gains also extend beyond transformer-temperature forecasting. On COVID19, MultiPatchFormer and OLinear reduce four-horizon average MSE by **18.69%** and **16.23%**, respectively, while WPMixer achieves a **12.55%** reduction on ExchangeRate. The differing gains across model–dataset pairs indicate that the benefit of reweighting depends on the forecasting task and backbone. Tables 7 and 8 show how these gains vary across forecasting horizons.

B.1 Full Training-Strategy Comparison Results

Table 9 reports three-run means and sample standard deviations for each forecasting horizon, with four-horizon averages summarized in the Avg rows. With TimeMixer++, StarTS achieves the lowest MSE at all four COVID19 horizons. On Sunspots, its largest gain over Raw occurs at $H = 48$, reducing MSE from 1.002 to 0.881.

Table 9: Full training-strategy comparison results. Values are means with sample standard deviations over three runs. Avg summarizes the four-horizon mean within each run using the same statistics. Within each backbone and metric, the best and second-best unrounded means are bold and underlined. Gray cells denote +StarTS.

Dataset	H	Metric	TimeMixer++				OLinear				MultiPatchFormer			
			Raw	+StarTS	+SL	+DT	Raw	+StarTS	+SL	+DT	Raw	+StarTS	+SL	+DT
COVID19	7	MSE	0.380 ±0.009	**0.355** ±0.006	0.359 ±0.004	0.380 ±0.017	1.247 ±0.092	**0.930** ±0.077	1.133 ±0.051	1.192 ±0.072	0.693 ±0.021	**0.607** ±0.051	0.622 ±0.021	0.719 ±0.049
		MAE	0.277 ±0.006	**0.267** ±0.004	0.269 ±0.002	0.282 ±0.011	0.516 ±0.015	**0.445** ±0.017	0.483 ±0.014	0.506 ±0.013	0.476 ±0.012	0.411 ±0.021	**0.409** ±0.002	0.483 ±0.005
	14	MSE	0.488 ±0.003	**0.438** ±0.002	0.478 ±0.014	0.530 ±0.027	1.537 ±0.168	**1.205** ±0.060	1.506 ±0.117	1.551 ±0.078	1.300 ±0.481	0.837 ±0.092	**0.734** ±0.020	1.148 ±0.567
		MAE	0.337 ±0.002	**0.314** ±0.003	0.324 ±0.006	0.357 ±0.009	0.562 ±0.037	**0.504** ±0.012	0.557 ±0.023	0.569 ±0.017	0.592 ±0.072	0.486 ±0.018	**0.456** ±0.008	0.564 ±0.085
	28	MSE	0.896 ±0.024	**0.787** ±0.062	0.955 ±0.018	0.890 ±0.058	2.218 ±0.077	**1.899** ±0.077	2.139 ±0.104	2.222 ±0.087	1.507 ±0.862	**1.023** ±0.116	1.064 ±0.181	1.655 ±0.737
		MAE	0.465 ±0.010	**0.428** ±0.015	0.453 ±0.004	0.461 ±0.010	0.687 ±0.010	**0.632** ±0.011	0.665 ±0.013	0.689 ±0.012	0.628 ±0.107	**0.531** ±0.023	0.543 ±0.024	0.651 ±0.087
	60	MSE	1.851 ±0.121	**1.839** ±0.112	2.102 ±0.039	1.883 ±0.158	2.832 ±0.018	**2.528** ±0.029	2.771 ±0.013	2.821 ±0.029	2.223 ±0.223	**2.186** ±0.152	2.209 ±0.257	2.188 ±0.121
		MAE	**0.616** ±0.019	0.621 ±0.016	0.644 ±0.006	0.638 ±0.022	0.777 ±0.007	**0.724** ±0.006	0.753 ±0.011	0.774 ±0.009	0.701 ±0.015	**0.692** ±0.013	0.692 ±0.039	0.707 ±0.019
	Avg	MSE	0.904 ±0.028	**0.855** ±0.033	0.973 ±0.014	0.921 ±0.049	1.958 ±0.075	**1.641** ±0.041	1.887 ±0.064	1.947 ±0.064	1.431 ±0.318	1.163 ±0.075	**1.157** ±0.066	1.427 ±0.346
		MAE	0.424 ±0.004	**0.408** ±0.007	0.423 ±0.002	0.434 ±0.009	0.635 ±0.014	**0.576** ±0.007	0.614 ±0.010	0.635 ±0.011	0.599 ±0.043	0.530 ±0.008	**0.525** ±0.009	0.602 ±0.047
NABCPU	24	MSE	1.148 ±0.003	1.148 ±0.008	**1.141** ±0.008	1.141 ±0.008	1.149 ±0.002	1.145 ±0.001	1.157 ±0.004	**1.143** ±0.003	**1.149** ±0.003	1.150 ±0.003	1.149 ±0.003	1.150 ±0.004
		MAE	0.315 ±0.003	**0.315** ±0.009	0.347 ±0.012	0.318 ±0.006	**0.333** ±0.002	0.338 ±0.005	0.366 ±0.005	0.335 ±0.001	**0.305** ±0.001	0.305 ±0.001	0.312 ±0.004	0.305 ±0.002
	48	MSE	**1.162** ±0.006	1.165 ±0.009	1.165 ±0.005	1.167 ±0.007	1.161 ±0.001	1.157 ±0.000	1.165 ±0.003	**1.155** ±0.002	1.161 ±0.002	**1.161** ±0.003	1.163 ±0.003	1.164 ±0.004
		MAE	**0.316** ±0.005	0.335 ±0.016	0.342 ±0.005	0.324 ±0.005	**0.332** ±0.001	0.338 ±0.002	0.369 ±0.006	0.334 ±0.001	**0.306** ±0.002	0.307 ±0.000	0.310 ±0.001	0.309 ±0.002
	96	MSE	1.217 ±0.009	**1.212** ±0.011	1.213 ±0.004	1.217 ±0.015	1.212 ±0.000	1.212 ±0.000	1.215 ±0.004	**1.208** ±0.000	**1.207** ±0.001	1.207 ±0.002	1.211 ±0.003	1.208 ±0.003
		MAE	0.327 ±0.007	**0.323** ±0.011	0.338 ±0.003	0.335 ±0.019	0.336 ±0.001	**0.335** ±0.001	0.367 ±0.003	0.337 ±0.001	0.310 ±0.002	**0.309** ±0.001	0.315 ±0.003	0.311 ±0.001
	192	MSE	1.312 ±0.004	1.309 ±0.002	**1.297** ±0.008	1.316 ±0.010	1.300 ±0.000	1.298 ±0.002	**1.293** ±0.002	1.294 ±0.000	1.307 ±0.004	**1.298** ±0.004	1.306 ±0.002	1.304 ±0.005
		MAE	0.335 ±0.004	0.333 ±0.003	**0.331** ±0.003	0.350 ±0.004	**0.344** ±0.001	0.348 ±0.003	0.355 ±0.003	0.344 ±0.001	0.327 ±0.005	**0.321** ±0.003	0.327 ±0.002	0.323 ±0.003
	Avg	MSE	1.210 ±0.002	1.208 ±0.006	**1.204** ±0.003	1.210 ±0.003	1.205 ±0.001	1.203 ±0.001	1.207 ±0.003	**1.200** ±0.001	1.206 ±0.002	**1.204** ±0.002	1.207 ±0.001	1.207 ±0.002
		MAE	**0.323** ±0.002	0.326 ±0.008	0.339 ±0.005	0.332 ±0.005	**0.336** ±0.001	0.340 ±0.002	0.364 ±0.004	0.337 ±0.001	0.312 ±0.002	**0.311** ±0.001	0.316 ±0.001	0.312 ±0.001
Sunspots	12	MSE	0.287 ±0.024	**0.278** ±0.005	0.298 ±0.009	0.329 ±0.005	**0.330** ±0.005	0.332 ±0.006	0.341 ±0.004	0.355 ±0.005	**0.234** ±0.014	0.238 ±0.010	0.268 ±0.006	0.234 ±0.006
		MAE	0.389 ±0.017	**0.380** ±0.005	0.395 ±0.003	0.417 ±0.002	**0.414** ±0.003	0.414 ±0.003	0.424 ±0.003	0.430 ±0.003	**0.354** ±0.011	0.360 ±0.006	0.377 ±0.005	0.354 ±0.004
	24	MSE	**0.482** ±0.063	0.517 ±0.021	0.557 ±0.013	0.635 ±0.031	0.722 ±0.031	**0.720** ±0.030	0.796 ±0.037	0.747 ±0.028	0.411 ±0.030	0.396 ±0.012	0.509 ±0.015	**0.389** ±0.025
		MAE	**0.493** ±0.039	0.506 ±0.011	0.520 ±0.006	0.578 ±0.011	0.610 ±0.017	**0.608** ±0.016	0.650 ±0.018	0.626 ±0.015	0.450 ±0.013	0.447 ±0.005	0.498 ±0.008	**0.447** ±0.015
	48	MSE	1.002 ±0.137	**0.881** ±0.043	1.107 ±0.027	1.467 ±0.103	1.599 ±0.032	**1.591** ±0.033	1.686 ±0.031	1.619 ±0.031	0.898 ±0.044	0.893 ±0.057	1.022 ±0.060	**0.883** ±0.054
		MAE	0.710 ±0.055	**0.661** ±0.016	0.752 ±0.006	0.902 ±0.041	0.951 ±0.015	**0.947** ±0.016	0.993 ±0.015	0.964 ±0.014	**0.663** ±0.011	0.665 ±0.017	0.713 ±0.023	0.665 ±0.021
	96	MSE	**1.055** ±0.047	1.090 ±0.128	1.189 ±0.087	1.561 ±0.078	1.685 ±0.022	**1.681** ±0.021	1.708 ±0.039	1.698 ±0.022	1.077 ±0.029	**1.050** ±0.070	1.154 ±0.036	1.106 ±0.053
		MAE	**0.747** ±0.017	0.754 ±0.052	0.798 ±0.032	0.940 ±0.029	**1.005** ±0.010	1.008 ±0.010	1.018 ±0.020	1.015 ±0.010	0.754 ±0.006	**0.735** ±0.024	0.787 ±0.012	0.758 ±0.017
	Avg	MSE	0.706 ±0.046	**0.692** ±0.029	0.788 ±0.011	0.998 ±0.038	1.084 ±0.022	**1.081** ±0.022	1.133 ±0.028	1.105 ±0.021	0.655 ±0.011	**0.645** ±0.031	0.738 ±0.007	0.653 ±0.024
		MAE	0.585 ±0.022	**0.575** ±0.013	0.616 ±0.008	0.709 ±0.015	0.745 ±0.011	**0.744** ±0.011	0.771 ±0.014	0.759 ±0.010	0.555 ±0.004	**0.552** ±0.011	0.593 ±0.004	0.556 ±0.009
ExchangeRate	96	MSE	0.086 ±0.002	**0.085** ±0.001	0.086 ±0.001	0.086 ±0.002	**0.093** ±0.000	0.094 ±0.000	0.098 ±0.000	0.094 ±0.000	**0.105** ±0.001	0.107 ±0.001	0.107 ±0.005	0.106 ±0.003
		MAE	0.211 ±0.002	**0.211** ±0.001	0.211 ±0.001	0.211 ±0.002	**0.221** ±0.000	0.222 ±0.000	0.229 ±0.000	0.222 ±0.000	**0.238** ±0.000	0.243 ±0.001	0.241 ±0.005	0.241 ±0.002
	192	MSE	0.172 ±0.001	0.167 ±0.002	**0.165** ±0.002	0.171 ±0.002	0.184 ±0.000	**0.176** ±0.000	0.180 ±0.000	0.185 ±0.000	0.192 ±0.005	0.190 ±0.005	0.195 ±0.006	**0.188** ±0.001
		MAE	0.303 ±0.001	0.300 ±0.001	**0.295** ±0.001	0.302 ±0.001	0.315 ±0.000	**0.311** ±0.000	0.313 ±0.000	0.315 ±0.000	0.324 ±0.004	0.324 ±0.003	0.330 ±0.004	**0.322** ±0.000
	336	MSE	0.332 ±0.006	**0.307** ±0.002	0.331 ±0.003	0.315 ±0.006	0.358 ±0.001	**0.333** ±0.001	0.344 ±0.001	0.357 ±0.001	0.342 ±0.008	**0.328** ±0.007	0.331 ±0.008	0.334 ±0.010
		MAE	0.422 ±0.004	**0.407** ±0.001	0.423 ±0.003	0.413 ±0.003	0.441 ±0.000	**0.426** ±0.000	0.436 ±0.001	0.440 ±0.000	0.433 ±0.004	0.427 ±0.005	**0.427** ±0.005	0.428 ±0.006
	480	MSE	0.505 ±0.008	**0.483** ±0.012	0.525 ±0.005	0.531 ±0.013	0.522 ±0.001	**0.488** ±0.001	0.530 ±0.001	0.521 ±0.001	0.496 ±0.017	**0.455** ±0.009	0.576 ±0.030	0.521 ±0.018
		MAE	0.525 ±0.005	**0.511** ±0.007	0.541 ±0.004	0.539 ±0.007	0.540 ±0.000	**0.521** ±0.000	0.547 ±0.001	0.540 ±0.000	0.521 ±0.012	**0.506** ±0.006	0.567 ±0.013	0.530 ±0.006
	Avg	MSE	0.274 ±0.003	**0.260** ±0.004	0.277 ±0.002	0.276 ±0.003	0.290 ±0.000	**0.273** ±0.000	0.288 ±0.000	0.289 ±0.000	0.284 ±0.005	**0.270** ±0.004	0.302 ±0.005	0.287 ±0.007
		MAE	0.365 ±0.002	**0.357** ±0.002	0.368 ±0.002	0.366 ±0.002	0.379 ±0.000	**0.370** ±0.000	0.381 ±0.000	0.379 ±0.000	0.379 ±0.003	**0.375** ±0.002	0.391 ±0.002	0.380 ±0.004

C Sample Properties and Weight Evolution

C.1 Definitions of Sample Properties

For training sample i, let $x_{i,t,c}$ be its standardized model input at time step $t \in \{1, \dots, L\}$ and channel $c \in \{1, \dots, C\}$. The plotted properties use these inputs without further rescaling. Write

$$\bar{x}_i = (LC)^{-1} \sum_{t,c} x_{i,t,c}, \qquad \bar{x}_{i,c} = L^{-1} \sum_{t} x_{i,t,c}, \qquad u_t = t - (L+1)/2.$$

Variability (V_i). Variability measures the overall dispersion of input values within a sample. We compute the population standard deviation pooled over time steps and channels:

$$V_i = \left[\frac{1}{LC} \sum_{t=1}^{L} \sum_{c=1}^{C} (x_{i,t,c} - \bar{x}_i)^2 \right]^{1/2}. \tag{10}$$

Trend intensity (T_i). Trend intensity measures the strength of linear change within a sample, regardless of whether the trend increases or decreases. We average the absolute channel-wise linear slopes within each window:

$$\beta_{i,c} = \frac{\sum_{t=1}^{L} u_t x_{i,t,c}}{\sum_{t=1}^{L} u_t^2},$$
$$T_i = \frac{1}{C} \sum_{c=1}^{C} |\beta_{i,c}|. \tag{11}$$

Volatility (D_i). Volatility measures the magnitude of fluctuations remaining after the linear trend is removed. We remove each channel's linear trend and average the residual standard deviations:

$$r_{i,t,c} = x_{i,t,c} - \bar{x}_{i,c} - \beta_{i,c} u_t,$$
$$D_i = \frac{1}{C} \sum_{c=1}^{C} \left[\frac{1}{L} \sum_{t=1}^{L} (r_{i,t,c} - \bar{r}_{i,c})^2 \right]^{1/2}, \tag{12}$$

where $\bar{r}_{i,c} = L^{-1} \sum_t r_{i,t,c}$. The standard deviations in V_i and D_i use population divisors LC and L, respectively.

Temporal similarity (P_i). Temporal similarity measures the signed linear association between adjacent detrended values. We center the two lagged residual sequences separately for $t \in \{1, \dots, L-1\}$:

$$r^{-}_{i,t,c} = r_{i,t,c} - (L-1)^{-1} \sum_{s=1}^{L-1} r_{i,s,c},$$
$$r^{+}_{i,t,c} = r_{i,t+1,c} - (L-1)^{-1} \sum_{s=2}^{L} r_{i,s,c}.$$

The channel-wise lag-1 Pearson correlation and sample-level temporal similarity are

$$d_{i,c} = \left[\sum_{t=1}^{L-1} (r^{-}_{i,t,c})^2 \sum_{t=1}^{L-1} (r^{+}_{i,t,c})^2 \right]^{1/2}, \tag{13}$$
$$\rho^{(1)}_{i,c} = \frac{\sum_{t=1}^{L-1} r^{-}_{i,t,c} r^{+}_{i,t,c}}{d_{i,c}},$$
$$P_i = \frac{1}{|\mathcal{C}^*_i|} \sum_{c \in \mathcal{C}^*_i} \rho^{(1)}_{i,c}, \tag{14}$$

where $\mathcal{C}^*_i = \{c : d_{i,c} > 10^{-10}\}$. Channels with negligible lagged variance are excluded; P_i is missing when no channel is valid.

C.2 Weight Evolution Across Training Stages

Early and middle checkpoints show the same broad preference for low variability, weak trends, and low volatility as the late stage (Figures 6 and 7). Individual windows change rank across stages, reflecting evolving training priorities.

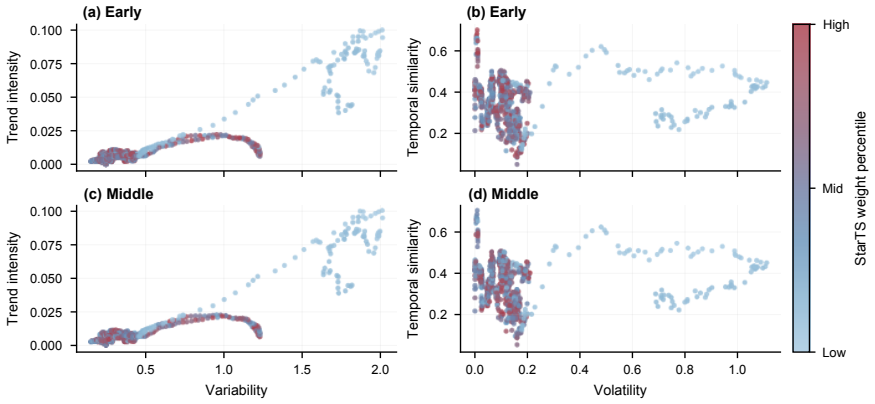

Figure 7: **Learned weights across training stages.** Early (epoch 11, top) and middle (epoch 51, bottom) weights for the same windows as Figure 6, using a shared weight-percentile color scale.

D Component Attribution and Hyperparameter Sensitivity

D.1 Component Attribution

Table 10 details the TimeMixer++ comparison in Figure 3. Each dataset–horizon setting uses the same configuration across variants, changing only the weighting rule. MM-only retains shared loss statistics for its memory update while omitting FRE from the weight.

Table 10: TimeMixer++ component results from one run (test MSE; four-horizon average). Lower is better; bold and underlining mark the best and second-best values from unrounded results.

Dataset	*H*	Raw	FRE-only	MM-only	Full (FRE+MM)
COVID19	7	0.390	**0.359**	0.393	0.361
	14	0.485	0.439	0.484	**0.435**
	28	0.917	0.756	0.893	**0.745**
	60	1.727	1.901	1.790	**1.722**
	Avg	0.880	0.864	0.890	**0.816**
NABCPU	24	1.146	1.148	1.141	**1.140**
	48	1.166	1.179	1.164	**1.161**
	96	1.209	1.208	1.206	**1.203**
	192	1.314	1.316	1.314	**1.307**
	Avg	1.209	1.213	1.206	**1.203**
Sunspots	12	0.301	0.315	**0.262**	0.281
	24	0.524	0.495	0.523	**0.494**
	48	1.160	1.159	1.190	**0.926**
	96	1.053	1.268	**0.983**	1.023
	Avg	0.760	0.809	0.740	**0.681**
ExchangeRate	96	0.087	0.087	**0.084**	0.086
	192	0.172	0.174	0.179	**0.166**
	336	0.328	**0.307**	0.329	0.307
	480	0.505	**0.472**	0.507	0.473
	Avg	0.273	0.260	0.275	**0.258**

D.2 Complete Hyperparameter Sensitivity Results

The complete results in Table 11 retain the aggregate pattern in Figure 4. MM strength produces small changes in dataset-average MSE, whereas a low weight temperature raises average error on COVID19 and Sunspots. Wider FRE margins lower error on Sunspots but increase it on COVID19 and ExchangeRate.

Table 11: **Complete TimeMixer++ hyperparameter sensitivity.** Values are test MSE; Avg weights the four horizons equally. Each one-at-a-time sweep uses the shared default $(m, \lambda_{\mathrm{M}}, \tau) = (0, 0.1, 1)$ for the other two parameters. Default rows are shaded; bold and underlined entries mark the best and second-best values within each sweep using unrounded results.

Dataset	Sweep	Value	Test MSE by horizon				Avg
COVID19	**Forecast horizon**		$H = 7$	$H = 14$	$H = 28$	$H = 60$	**Avg**
	FRE margin m	0	0.361	0.435	0.745	1.871	0.853
		0.25	**0.361**	**0.432**	0.745	1.865	0.851
		0.5	0.361	0.432	0.746	1.863	**0.851**
		1	0.365	0.438	0.751	**1.862**	0.854
		2	0.379	0.439	**0.743**	1.865	0.857
	MM strength λ_{M}	0	0.361	0.435	0.745	1.872	0.853
		0.05	0.362	0.440	0.745	1.871	0.855
		0.1	0.361	0.435	**0.745**	1.871	0.853
		0.2	0.361	0.434	0.756	1.870	0.855
		0.4	**0.361**	**0.434**	0.745	**1.868**	**0.852**
	Weight temperature τ	0.5	0.365	0.436	0.800	1.878	0.870
		0.75	0.363	0.442	0.761	1.874	0.860
		1	0.361	0.435	**0.745**	1.871	**0.853**
		1.5	**0.361**	**0.434**	0.757	1.867	0.855
		2	0.361	0.435	0.760	**1.867**	0.856
NABCPU	**Forecast horizon**		$H = 24$	$H = 48$	$H = 96$	$H = 192$	**Avg**
	FRE margin m	0	1.148	1.173	1.212	1.313	1.212
		0.25	1.144	1.171	1.216	1.311	1.211
		0.5	1.141	**1.167**	1.208	1.314	1.208
		1	**1.141**	1.171	1.208	1.310	**1.208**
		2	1.146	1.170	**1.206**	**1.308**	1.208
	MM strength λ_{M}	0	**1.144**	1.173	1.212	1.313	1.211
		0.05	1.144	1.173	1.212	1.313	1.211
		0.1	1.148	1.173	1.212	1.313	1.212
		0.2	1.147	1.173	1.212	**1.313**	1.211
		0.4	1.146	**1.172**	**1.208**	1.314	**1.210**
	Weight temperature τ	0.5	1.145	1.178	**1.209**	1.316	1.212
		0.75	1.145	1.174	1.210	1.314	1.211
		1	1.148	1.173	1.212	1.313	1.212
		1.5	1.146	1.171	1.210	**1.312**	1.210
		2	**1.144**	**1.170**	1.210	1.314	**1.210**
Sunspots	**Forecast horizon**		$H = 12$	$H = 24$	$H = 48$	$H = 96$	**Avg**
	FRE margin m	0	0.291	0.521	1.186	1.248	0.812
		0.25	0.315	0.549	1.178	1.246	0.822
		0.5	0.301	0.487	1.160	1.248	0.799
		1	**0.274**	0.466	1.130	1.264	0.784
		2	0.294	**0.444**	**1.113**	**1.082**	**0.733**
	MM strength λ_{M}	0	0.304	0.523	1.196	1.249	0.818
		0.05	0.306	0.522	1.190	1.248	0.817
		0.1	**0.291**	0.521	1.186	1.248	0.812
		0.2	0.306	0.518	1.178	1.247	0.812
		0.4	0.301	**0.511**	**1.163**	**1.245**	**0.805**
	Weight temperature τ	0.5	0.320	0.531	1.226	1.269	0.836
		0.75	0.313	0.489	1.219	1.250	0.818
		1	**0.291**	0.521	1.186	**1.248**	0.812
		1.5	0.317	0.520	1.150	1.250	0.809
		2	0.345	**0.466**	**1.134**	1.265	**0.802**
ExchangeRate	**Forecast horizon**		$H = 96$	$H = 192$	$H = 336$	$H = 480$	**Avg**
	FRE margin m	0	0.088	0.173	**0.317**	0.503	0.270
		0.25	**0.088**	0.173	0.318	**0.500**	**0.270**
		0.5	0.088	0.173	0.318	0.504	0.271
		1	0.088	**0.169**	0.344	0.507	0.277
		2	0.088	0.170	0.332	0.509	0.275
	MM strength λ_{M}	0	**0.088**	**0.173**	**0.317**	0.502	0.270
		0.05	0.088	0.173	0.317	0.502	0.270
		0.1	0.088	0.173	0.317	0.503	0.270
		0.2	0.088	0.173	0.317	0.503	0.270
		0.4	0.088	0.173	0.317	**0.498**	**0.269**
	Weight temperature τ	0.5	0.088	**0.172**	**0.314**	**0.488**	**0.266**
		0.75	**0.088**	0.173	0.316	0.499	0.269
		1	0.088	0.173	0.317	0.503	0.270
		1.5	0.088	0.173	0.318	0.497	0.269
		2	0.088	0.173	0.319	0.498	0.269

Recommended starting setting. Use $(m, \lambda_{\mathrm{M}}, \tau) = (0, 0.1, 1)$, $\alpha = \rho = 0.9$, $\tau_{\mathrm{M}} = 1$, $r = 0$, and $\epsilon = 10^{-6}$; select m on validation data when adapting to a new dataset.