CROSSFOREST: PHASE-SPECIALIZED LEARNERS FOR LONG-HORIZON TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series forecasting supports applications such as traffic planning, energy management and weather forecasting. Forecast accuracy often declines as the prediction horizon increases. To take advantage of the higher accuracy often seen at shorter prediction lengths while preserving the full forecast horizon, we propose CROSSFOREST, a framework that can be applied to different forecasting backbones. It splits a long forecast into phase-specific tasks aligned to a common time origin. For N phases, it routes each timestamp by its residue modulo N to a phase learner with its own parameters. Each learner predicts only its assigned phase: H/N values spaced N time steps apart. Cross-interleaving assembles these outputs into the complete H-step forecast. Adjacent sliding windows share most phase tasks, so the training set keeps one copy of each distinct task and retains complete stride-one coverage without repeated computation. We call the combination of globally aligned phase assignment and duplicate-free task construction the Global-Unique protocol. Our experiments show that WPMixer and PatchTST with CROSSFOREST outperform the baselines on multiple forecasting tasks, while CROSSFOREST improves eight of the nine baseline backbones.

1 INTRODUCTION

Long-term time-series forecasting (LTSF) supports decisions in energy, transportation, finance, and environmental monitoring. As the prediction horizon increases, models must forecast farther into the future from observed data while accounting for distribution drift, multiple periodicities, and non-stationary long-term relationships (Kim et al., 2022; Liu et al., 2022b; 2025). Forecast accuracy usually declines as the horizon increases, making performance at longer horizons an ongoing challenge.

One line of LTSF research improves representations within forecasting models. Autoformer and FEDformer decompose temporal patterns (Wu et al., 2021; Zhou et al., 2022); PatchTST encodes local patches and processes each variable independently (Nie et al., 2023); iTransformer models correlations between variables (Liu et al., 2024); and TimeMixer uses multiscale mixing (Wang et al., 2024b). Another line of work organizes forecasting through sub-series decomposition. SutraNets converts the original sequence into lower-frequency sub-series and generates complete forecasts through conditional autoregression across time and sub-series (Bergsma et al., 2023). SparseTSF uses cross-period downsampling, applies a predictor with shared parameters to the sub-series, and reconstructs the complete forecast sequence (Lin et al., 2024).

SutraNets and SparseTSF provide a basis for organizing long-sequence forecasting through sub-series decomposition. Motivated by the higher accuracy that forecasting models often achieve at shorter horizons, we ask whether shortening each learner's output sequence can improve accuracy while preserving the full forecast horizon. To study this question, we use phase learners with separate parameters so that different phases can learn their own forecasting mappings, and introduce the Global-Unique protocol. This protocol fixes each learner's phase assignment through global alignment and removes phase tasks repeated across adjacent sliding windows, avoiding repeated computation while retaining coverage of all distinct supervision pairs.

The resulting framework, CROSSFOREST, can wrap an existing forecasting backbone. Let L be the input length and H the forecast length, with the phase count N dividing both. For each timestamp,

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.

the framework takes its time-step index relative to a common time origin modulo N to select the corresponding phase learner. Each learner receives L/N observations sampled every N time steps and predicts H/N values for the same phase. Interleaving the phase outputs reconstructs the complete H-step forecast. Each learner produces a shorter output sequence, while the reconstructed forecast retains the full horizon on the original time axis.

Our contributions are as follows:

- We propose CROSSFOREST, a framework with a separate set of model parameters for each temporal phase. Each learner uses only the history of its assigned phase to learn its own forecasting mapping. The framework supports different forecasting backbones and reconstructs complete forecasts by interleaving the phase outputs.
- We introduce the Global-Unique protocol, which combines fixed global phase assignments with duplicate-free training-task construction. When the phase count N divides both the input and forecast lengths, the protocol reduces the WN phase tasks expanded from W consecutive stride-one windows in a single continuous sequence to $W + N - 1$ distinct tasks, preserving coverage of distinct supervision pairs.
- We evaluate CROSSFOREST across multiple datasets and forecasting backbones. Comparisons of separate versus shared learner parameters, and of Global-Unique versus a protocol with window-local phase assignments and repeated tasks, examine how phase parameterization and training-task organization affect forecasting performance.

2 RELATED WORK

Long-term time-series forecasting. Transformer-based forecasters have progressively introduced sparse attention, series decomposition, frequency-domain computation, patching, and alternative tokenization schemes (Zhou et al., 2021; Wu et al., 2021; Zhou et al., 2022; Nie et al., 2023; Liu et al., 2024). Other work shows that carefully designed linear or MLP models remain competitive (Zeng et al., 2023; Wang et al., 2024b), while TimesNet represents temporal variation in two dimensions (Wu et al., 2023). Recent models explicitly address multiscale and non-stationary structure (Wang et al., 2024a; Liu et al., 2025). Rather than replacing these representations, CROSSFOREST changes how the temporal prediction problem is partitioned and can therefore use several of them as its base architecture.

Channel independence. Channel-independent forecasting applies a shared model separately to each univariate channel, reducing cross-variable interference and improving data efficiency (Nie et al., 2023). DLinear also exposes shared and channel-specific parameterizations (Zeng et al., 2023). CROSSFOREST adapts the separation principle to temporal phases, but deliberately assigns *unshared* parameters to the phases. The two forms of independence are orthogonal: a channel-independent backbone may still be used inside each phase learner.

Temporal decomposition. Temporal decomposition models sequence structure at different sampling rates. SCINet recursively splits odd- and even-indexed samples and exchanges information between branches (Liu et al., 2022a); LightTS combines continuous and interval sampling within a single forecasting network (Zhang et al., 2022). SutraNets splits a univariate sequence into lower-frequency sub-series and uses conditional autoregression across time and sub-series to generate coherent high-frequency forecasts (Bergsma et al., 2023). SparseTSF uses cross-period downsampling and a lightweight predictor with shared parameters to learn cross-period trends (Lin et al., 2024). Models such as FEDformer, TimesNet, and TimeMixer retain interactions between temporal components when modeling frequency or periodic structure (Zhou et al., 2022; Wu et al., 2023; Wang et al., 2024b). CROSSFOREST uses globally aligned polyphase sampling and assigns separate parameters to each phase. Each learner forecasts from its own phase history, without conditioning on observations or generated values from other phases. The phase outputs are interleaved to reconstruct the complete forecast, with exactly one learner responsible for each output position.

Ensemble learning. Bagging and random forests train diverse estimators and aggregate predictions to reduce variance (Breiman, 1996; 2001). Deep ensembles extend this principle to neural

predictors (Lakshminarayanan et al., 2017). CROSSFOREST is forest-inspired but structurally different. Its experts are specialized by absolute temporal phase, and each forecast position has exactly one responsible expert. This also distinguishes the method from mixture-of-experts systems that learn a soft or hard gate over multiple predictions (Shazeer et al., 2017).

3 CROSSFOREST

3.1 FORECASTING SETUP

Let $\tau_t = \tau_0 + t\Delta$ denote a timestamp on a regular grid. Here, τ_0 is the dataset-level origin, Δ is the sampling interval, and $t \in \mathbb{Z}$ is the corresponding global slot index. Both τ_0 and Δ are fixed before training, and t is not reset at the validation or test boundary. Let $\boldsymbol{x}_t \in \mathbb{R}^C$ be the C-variate observation at slot t. A forecasting window that begins at slot s contains

$$\begin{aligned} \boldsymbol{X}_s &= [\boldsymbol{x}_s, \boldsymbol{x}_{s+1}, \ldots, \boldsymbol{x}_{s+L-1}] \in \mathbb{R}^{L\times C}, \\ \boldsymbol{Y}_s &= [\boldsymbol{x}_{s+L}, \ldots, \boldsymbol{x}_{s+L+H-1}] \in \mathbb{R}^{H\times C}. \end{aligned}$$

A base forecaster learns $f_\theta : \boldsymbol{X}_s \mapsto \widehat{\boldsymbol{Y}}_s$ by minimizing mean squared error (MSE). We assume that the phase count N divides both L and H. For a decoder-based backbone, N must also divide the decoder label length. Setting $N = 1$ removes phase factorization and provides a direct control within the wrapper.

3.2 GLOBAL-UNIQUE PROTOCOL

3.2.1 GLOBALLY ALIGNED POLYPHASE ROUTING

Define the absolute phase of timestamp slot t as $\phi(t) = t \bmod N$ and the starting phase of a window as $\rho_s = \phi(s)$. For phase $r \in \{0, \ldots, N-1\}$, the first offsets of that phase in the history and forecast intervals are

$$\begin{aligned} a_r(s) &= (r - \rho_s) \bmod N, \\ b_r(s) &= (r - \phi(s+L)) \bmod N. \end{aligned}$$

Because N divides L, $b_r(s) = a_r(s)$. We keep both symbols so that the forecast-time indexing remains explicit. The phase-specific history and target are

$$\begin{aligned} \boldsymbol{X}_s^{(r)} &= [\boldsymbol{x}_{s+a_r(s)+jN}]_{j=0}^{L/N-1}, \\ \boldsymbol{Y}_s^{(r)} &= [\boldsymbol{x}_{s+L+b_r(s)+jN}]_{j=0}^{H/N-1}. \end{aligned} \tag{1}$$

Since N divides L, the history and forecast begin with the same phase ρ_s. Known future covariates, when available, follow the same indices. Each phase has a learner with its own parameters:

$$\widehat{\boldsymbol{Y}}_s^{(r)} = f_{\theta_r}(\boldsymbol{X}_s^{(r)}), \qquad \theta_r \cap \theta_q = \varnothing \quad (r \neq q), \tag{2}$$

where $r, q \in \{0, \ldots, N-1\}$ index temporal phases, and the parameter sets are disjoint whenever $r \neq q$. Every f_{θ_r} has the same selected backbone architecture but a separate parameter set.

At inference time, the phase outputs are restored in chronological order. For forecast position $k \in \{0, \ldots, H-1\}$, let r_k denote the index of the phase learner responsible for that position and j_k its index within that learner's output sequence. Both indices use zero-based indexing and are defined as follows:

$$r_k = \phi(s+L+k) = (\rho_s + k) \bmod N, \qquad j_k = \frac{k - b_{r_k}(s)}{N} = \left\lfloor \frac{k}{N} \right\rfloor. \tag{3}$$

Then

$$\widehat{\boldsymbol{x}}_{s+L+k} = \widehat{\boldsymbol{Y}}_s^{(r_k)}[j_k]. \tag{4}$$

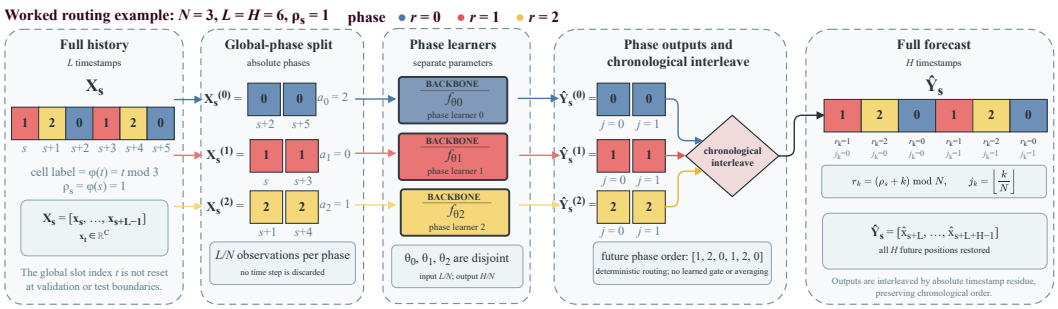

Figure 1: Overview of CROSSFOREST with three phases. Absolute timestamp residues define fixed assignments across stride-one windows. Learners with separate parameters predict phase-specific subsequences, and deterministic cross-interleaving restores every future time step. Diagram labels: Full history (L timestamps) $\rightarrow$ Global-phase split $\rightarrow$ Learners $f_{\theta_0}, f_{\theta_1}, f_{\theta_2}$ $\rightarrow$ Chronological interleave $\rightarrow$ Full forecast (H timestamps); absolute phases; all positions restored.

Nonzero starting phase. Let $N = 3$, $L = H = 6$, and $\rho_s = 1$. Then $(a_0, a_1, a_2) = (2, 0, 1)$, and the three histories are

$$\boldsymbol{X}_s^{(0)} = [\boldsymbol{x}_{s+2}, \boldsymbol{x}_{s+5}], \quad \boldsymbol{X}_s^{(1)} = [\boldsymbol{x}_s, \boldsymbol{x}_{s+3}], \quad \boldsymbol{X}_s^{(2)} = [\boldsymbol{x}_{s+1}, \boldsymbol{x}_{s+4}]. \tag{5}$$

The future learner order is $[1, 2, 0, 1, 2, 0]$, so interleaving produces

$$[\widehat{\boldsymbol{Y}}_s^{(1)}[0], \widehat{\boldsymbol{Y}}_s^{(2)}[0], \widehat{\boldsymbol{Y}}_s^{(0)}[0], \widehat{\boldsymbol{Y}}_s^{(1)}[1], \widehat{\boldsymbol{Y}}_s^{(2)}[1], \widehat{\boldsymbol{Y}}_s^{(0)}[1]]. \tag{6}$$

This example also covers a validation or test split whose first window begins at a nonzero phase.

The routing and interleaving equations permute the input and output time axes. Taken together, the phase decomposition retains every time step. Its inductive bias comes from keeping information separate across phase learners.

Global alignment is required. If each validation or test split labels its first timestamp as phase zero, a timestamp may be assigned to a different parameter set than it was during training. The global slot index keeps each learner's role fixed across windows and data splits. Window-local alignment has the same problem: labeling the first element of every sliding window as phase zero would send the same absolute timestamp to different learners as the window moves. Absolute routing keeps the role of learner r consistent across training, validation, and test splits, including splits that begin at a nonzero index.

3.2.2 DUPLICATE-FREE STRIDE-ONE TRAINING

A naive implementation creates N phase tasks for each of W stride-one windows, but most of these tasks are exact duplicates. When the window moves from s to $s + 1$, only phase $\phi(s)$ loses one historical value and gains one future value. Every other phase-specific input and target pair stays unchanged.

We index the phase-specific input–target pair of phase r in window w by (w, r); specifically, (w, r) denotes $(\boldsymbol{X}_{s_w}^{(r)}, \boldsymbol{Y}_{s_w}^{(r)})$, where s_w is the absolute start of window w. Thus, $(0, r)$ denotes the phase-r task of the first window.

The set of distinct tasks is

$$\mathcal{U} = \{(0, r) : 0 \le r < N\} \cup \{(w, \phi(s_w - 1)) : 1 \le w < W\}, \tag{7}$$

Hence, $|\mathcal{U}| = W + N - 1$ rather than WN.

With $\Theta = \{\theta_0, \ldots, \theta_{N-1}\}$, define the loss for one task as

$$\ell_{w,r}(\Theta) = \frac{1}{(H/N)C} \left\| f_{\theta_r}(\boldsymbol{X}_{s_w}^{(r)}) - \boldsymbol{Y}_{s_w}^{(r)} \right\|_F^2. \tag{8}$$

Here, $\| \cdot \|_F$ is the Frobenius norm, whose square sums the squared matrix entries. The loss is therefore the mean squared error over the predicted time points and output variables.

A phase-pure batch contains tasks for only one phase r and updates only the corresponding learner's parameters θ_r. Parameters of the other phase learners are not updated by that batch. This is parameter separation rather than statistical independence of the extracted data. A single optimizer may still manage all parameter groups. Early stopping uses the validation error after reconstruction of the full forecast.

For each backbone, CrossForest instantiates N parameter-disjoint learners with the same architecture. Each learner receives its phase-specific subsequence and predicts the corresponding phase-specific target.

4 EXPERIMENTS

We evaluate CROSSFOREST on a range of long-term time-series forecasting tasks and compare its forecasting performance with baseline models. We then examine its applicability across backbone architectures and how performance varies with the numbers of input channels and phase learners. Further experiments compare phase learners with separate and shared parameters, as well as forecasting performance under the Global-Unique and WindowLocal-Repeated training protocols.

Datasets. We extensively include nine real-world multivariate datasets in our experiments, spanning energy, transportation, finance, weather, and solar power. The collection comprises Electricity (ECL), four ETT subsets (ETTh1, ETTh2, ETTm1, and ETTm2), Exchange, Traffic, Weather, and Solar-Energy. Electricity and Traffic use fixed seven-channel Last7 subsets in the cross-backbone and phase-count comparisons. Detailed dataset descriptions and chronological partitions are provided in Appendix B.

4.1 MAIN RESULTS

We evaluate the forecasting performance of CROSSFOREST with long-term time-series forecasting backbones. Specifically, we compare two phase-factorized configurations, CF-WPMixer with $N = 4$ and CF-PatchTST with $N = 8$, against their unfactorized $N = 1$ counterparts and other forecasting models.

Baselines. We use nine forecasting models as baselines. Transformer- and patch-based methods include Autoformer (Wu et al., 2021), PatchTST (Nie et al., 2023), Crossformer (Zhang & Yan, 2023), iTransformer (Liu et al., 2024), and MultiPatchFormer (Naghashi et al., 2025). MLP- and mixing-based methods include TimeMixer (Wang et al., 2024b) and WPMixer (Murad et al., 2025), while DLinear (Zeng et al., 2023) and TimesNet (Wu et al., 2023) represent linear and temporal-variation methods, respectively. We evaluate all baselines with $N = 1$, without phase factorization.

Forecasting performance. Table 1 summarizes the forecasting results; lower MSE and MAE are better. Across the 12 combinations of dataset and metric, the two CROSSFOREST configurations take 14 of the 24 top-two positions (58.3%), with eight for CF-WPMixer and six for CF-PatchTST. CF-WPMixer achieves the lowest MSE and MAE on ETT and the lowest MAE on Weather. CF-PatchTST achieves the lowest MSE and MAE on both Exchange and Traffic. Overall, phase factorization improves forecasting performance across a range of settings.

4.2 BACKBONE GENERALITY

To assess the applicability and performance gains of CROSSFOREST across forecasting backbones, we apply the same global-phase routing and cross-interleaving procedure to the nine models in Table 1. These models span different forecasting architectures, allowing us to examine whether the benefits of CROSSFOREST depend on a particular backbone design.

Evaluation protocol. For each backbone, we compare phase-factorized variants with the same model in the CROSSFOREST wrapper at $N = 1$ under a common evaluation setup. Detailed experimental settings and the selection rule are provided in Appendix B.4.

Table 1: Multivariate forecasting results on the test set, averaged over $H \in \{96, 192, 336, 720\}$. ETT (Avg) also averages across ETTh1, ETTh2, ETTm1, and ETTm2. CF-WPMixer and CF-PatchTST use $N = 4$ and $N = 8$, respectively. All models use the same fixed seven-channel Last7 subsets for Electricity and Traffic. Lower values are better. The best results are marked in red boldface, and the second-best results are underlined in blue. Values are rounded to three decimal places; rankings are determined from the unrounded values.

Dataset	CF-WPMixer (Ours)		CF-PatchTST (Ours)		WPMixer (2025)		MultiPatch Former (2025)		TimeMixer (2024)		iTransformer (2024)		PatchTST (2023)		Crossformer (2023)		TimesNet (2023)		DLinear (2023)		Autoformer (2021)	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity (Last7)	0.155	0.260	0.158	0.268	0.166	0.268	0.173	0.286	0.162	0.272	0.170	0.273	0.191	0.299	0.163	0.271	0.454	0.431	**0.151**	**0.253**	0.652	0.547
ETT (Avg)	**0.345**	**0.382**	0.353	0.387	0.364	0.394	0.373	0.403	0.388	0.407	0.369	0.402	0.391	0.412	0.933	0.677	0.483	0.461	0.473	0.444	0.635	0.554
Exchange	0.405	0.425	**0.393**	**0.425**	0.573	0.484	0.444	0.455	0.522	0.459	0.402	0.430	0.497	0.463	1.113	0.761	1.413	0.769	0.557	0.517	2.963	1.261
Traffic (Last7)	0.291	0.268	**0.282**	**0.267**	0.290	0.278	0.289	0.283	0.291	0.283	0.321	0.321	0.286	0.277	0.316	0.296	0.369	0.354	0.293	0.276	0.398	0.380
Weather	0.222	**0.261**	0.233	0.269	0.230	0.270	0.225	0.264	0.247	0.283	0.242	0.277	0.241	0.279	0.268	0.316	0.320	0.326	**0.220**	0.277	0.563	0.538
Solar-Energy	0.206	0.265	0.209	0.267	0.201	0.259	**0.183**	0.245	0.198	0.259	0.209	0.271	0.192	0.251	0.192	**0.230**	0.237	0.258	0.231	0.294	0.759	0.589

Table 2: Generality of CROSSFOREST across forecasting backbones. Phase counts are selected using validation MSE. The G_m^σ columns report $G_m^\sigma(b, N_b)$ for each backbone and its selected phase count: the average relative reduction in metric m on split σ against $N = 1$, expressed as a percentage. Positive values indicate lower errors. The table also reports test win/loss counts (W/L). Blue and red cells denote positive and negative reductions, respectively, with intensity increasing with the absolute reduction on a common scale. The final row summarizes all 324 matched tasks. Detailed evaluation settings are provided in Appendix B.4.

Backbone	N_b	$G^{\text{val}}_{\text{MSE}}$	$G^{\text{test}}_{\text{MSE}}$	$G^{\text{test}}_{\text{MAE}}$	MSE W/L	MAE W/L
WPMixer	4	+3.33%	+4.79%	+3.04%	27/9	29/7
MultiPatchFormer	4	+1.18%	+0.98%	+0.32%	23/13	24/12
TimeMixer	2	+5.96%	+3.16%	+2.09%	28/8	29/7
iTransformer	2	+3.01%	+2.60%	+2.56%	28/8	28/8
PatchTST	8	+5.17%	+6.25%	+4.29%	29/7	30/6
Crossformer	2	+4.13%	+9.74%	+6.50%	23/13	24/12
TimesNet	8	+15.49%	+19.35%	+9.08%	33/3	28/8
DLinear	2	+0.14%	−0.45%	−0.21%	12/24	14/22
Autoformer	4	+11.82%	+24.16%	+15.14%	34/2	32/4
All backbones	–	+5.58%	+7.84%	+4.76%	237/87	238/86

Table 2 reports average relative reductions in validation MSE and test MSE and MAE compared with the $N = 1$ control. For data split $\sigma \in \{\text{val}, \text{test}\}$, these reductions are calculated as follows:

$$G_m^\sigma(b, N) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} 100 \frac{E_{b,1,g}^{m,\sigma} - E_{b,N,g}^{m,\sigma}}{E_{b,1,g}^{m,\sigma}}, \tag{9}$$

where b denotes the forecasting backbone, N is the evaluated phase count, and $m \in \{\text{MSE}, \text{MAE}\}$. Here, $E_{b,N,g}^{m,\sigma}$ denotes the error measured by metric m on split σ for backbone b with phase count N on task g. Table 2 reports $G_m^\sigma(b, N_b)$, where N_b is the phase count selected using validation MSE for backbone b. The set $\mathcal{G}$ contains all 36 combinations of dataset and forecast horizon, each weighted equally. Positive values indicate lower forecasting errors with CROSSFOREST.

Cross-backbone results. In this evaluation, validation-selected CROSSFOREST configurations reduce test MSE and MAE for eight of the nine architectures. Across 324 matched comparisons, MSE improves in 237 cases and MAE in 238, with macro-averaged reductions of 7.84% and 4.76%, respectively. Autoformer and TimesNet show the largest gains; mixing, patch-based, variate-token, and cross-dimensional models also improve. Gains vary by backbone: MultiPatchFormer improves modestly, and DLinear is the only model for which both test metrics worsen. These results support the applicability of CROSSFOREST across forecasting architectures, with performance gains depending on the backbone.

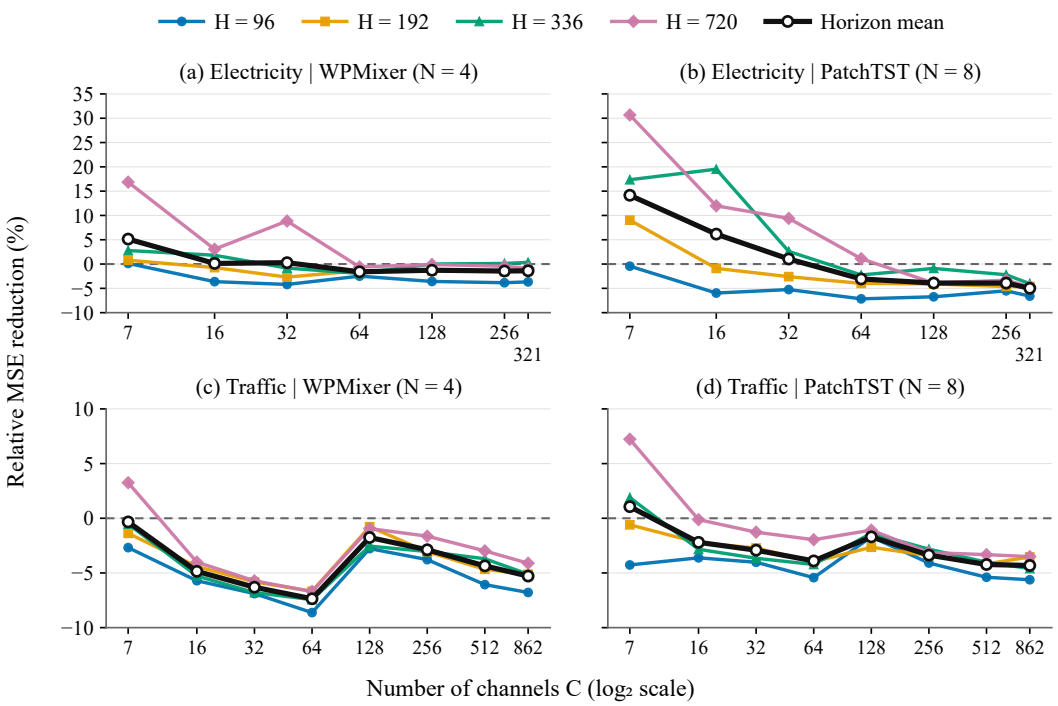

Figure 2: Relative reductions in MSE for CROSSFOREST against the $N = 1$ control at different channel counts. Colored lines show results for the four forecast horizons, and the thicker black line shows their mean. The horizontal dashed line marks zero reduction; positive values indicate lower forecasting errors. The horizontal axis uses a base-2 logarithmic scale. Vertical scales are shared within each dataset and differ between Electricity and Traffic.

4.3 SENSITIVITY ANALYSIS

4.3.1 CHANNEL-COUNT SENSITIVITY

We fix $N = 4$ for WPMixer and $N = 8$ for PatchTST to examine how the gains of these configurations change as the channel count increases. Figure 2 shows how the forecasting errors of CROSSFOREST change relative to the $N = 1$ control as channel count varies. On Electricity, both backbones benefit from phase factorization with seven channels, but average gains generally weaken as the channel set expands and become negative at larger channel counts. PatchTST's average MSE reduction falls from 14.15% with seven channels to -4.98% with all channels, while WPMixer's falls from 5.14% to -1.38%. On Traffic, PatchTST achieves a positive average MSE reduction only with seven channels, and WPMixer has negative average reductions at every evaluated channel count. These results show that gains on the seven-channel subsets do not necessarily extend to larger channel sets. Detailed protocols and tabulated results are provided in Appendix C.1.

4.3.2 PHASE-COUNT SENSITIVITY

We vary the number of phase learners over $N \in \{1, 2, 4, 8\}$ while holding the other experimental settings fixed to examine the sensitivity of WPMixer and PatchTST to phase factorization. Figure 3 shows relative reductions in test MSE and MAE against the $N = 1$ control, averaged across 36 combinations of dataset and forecast horizon. Detailed settings, calculation methods, and numerical results are provided in Appendix C.2.

The two models show different trends. WPMixer's gains peak at $N = 4$, with MSE and MAE reductions of 4.79% and 3.04%. At $N = 8$, these reductions fall to 1.37% and 0.44%. PatchTST's gains increase as N rises from 2 to 8: its MSE reduction grows from 1.09% to 6.25%, while MAE changes from a 0.14% increase to a 4.29% reduction.

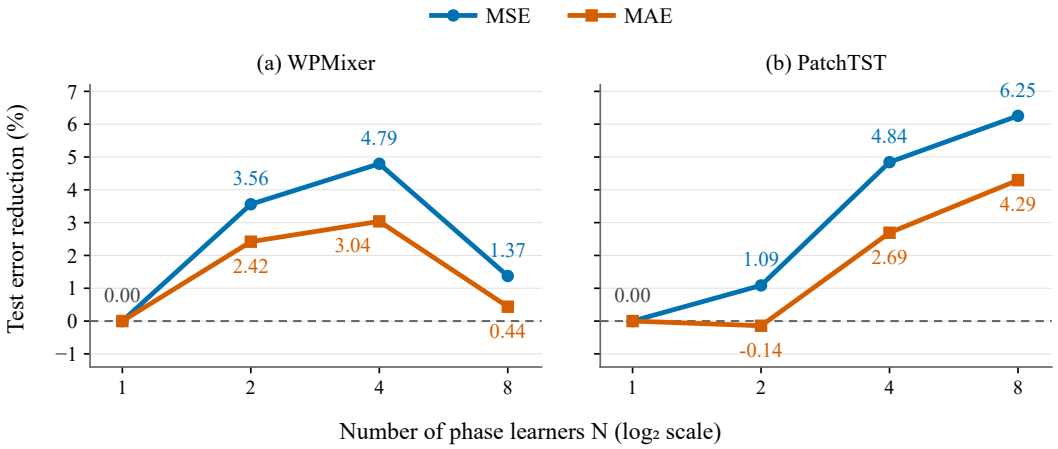

Figure 3: Forest-size (phase-count) sensitivity of WPMixer and PatchTST. Each panel shows macro-averaged test MSE and MAE reductions relative to the same backbone at $N = 1$ over 36 matched dataset–horizon tasks. Positive values indicate lower error. The dashed horizontal line marks zero reduction; both panels share the same vertical scale. The horizontal axis uses a base-2 logarithmic scale.

These curves show that the effect of adding phase learners depends on the backbone. Within the evaluated range, WPMixer performs best at $N = 4$, while PatchTST's gains continue to grow up to the largest tested value, $N = 8$.

4.4 Ablation Studies

4.4.1 Independent versus shared phase learners

We compare independent and shared phase learners to test whether separate parameters for each phase improve forecasting accuracy. Both variants use the same globally aligned phase inputs, per-phase sequence lengths, interleaving rule, and duplicate-free training. Table 3 reports results for PatchTST with $N = 8$ and WPMixer with $N = 4$. Appendix D.1 provides detailed settings and additional analysis.

Table 3: Forecasting errors with independent and shared phase learners, aggregated from Table 21. For each dataset, MSE and MAE are averaged separately over $H \in \{96, 192, 336, 720\}$. The four ETT subsets are listed separately; Electricity and Traffic use the fixed Last7 subsets. Lower values are better. Bold marks the lower error in each comparison of independent and shared learners for the same backbone and metric. Comparisons use unrounded values; displayed values are rounded to three decimal places.

Dataset	CF-PatchTST ($N = 8$)				CF-WPMixer ($N = 4$)			
	Shared		Independent		Shared		Independent	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity (Last7)	0.159	0.269	**0.158**	**0.268**	**0.153**	**0.257**	0.155	0.260
ETTh1	0.448	0.450	**0.436**	**0.444**	0.419	0.434	**0.417**	**0.432**
ETTh2	0.383	0.413	**0.361**	**0.398**	0.360	0.405	**0.350**	**0.396**
ETTm1	0.377	0.403	**0.355**	**0.386**	**0.352**	0.381	0.353	**0.379**
ETTm2	0.279	0.333	**0.260**	**0.321**	0.268	0.322	**0.259**	**0.321**
Exchange	0.435	0.448	**0.393**	**0.425**	0.451	0.441	**0.404**	**0.425**
Traffic (Last7)	**0.275**	**0.257**	0.282	0.267	**0.282**	**0.262**	0.291	0.268
Weather	0.237	0.272	**0.233**	**0.269**	0.223	0.263	**0.222**	**0.261**
Solar-Energy	**0.200**	**0.263**	0.209	0.267	**0.200**	**0.262**	0.206	0.265

Independent learners achieve lower errors in 25 of the 36 comparisons across datasets, backbones, and metrics in Table 3. For PatchTST, both MSE and MAE improve on seven of nine datasets. For WPMixer, MSE improves on five datasets and MAE on six. Overall, phase learners with independent parameters achieve better forecasting accuracy.

4.4.2 GLOBAL-UNIQUE VERSUS WINDOWLOCAL-REPEATED

We compare Global-Unique with WindowLocal-Repeated to test whether combining fixed global phase assignments with duplicate-free training improves forecasting accuracy. Global-Unique fixes each learner's phase across windows and includes one copy of each distinct phase task in the training set. WindowLocal-Repeated assigns phases from the start of each window and retains repeated tasks. Both protocols use independent phase learners, with identical data splits and backbone hyperparameters. We use PatchTST with $N = 8$ and WPMixer with $N = 4$. Table 4 reports relative error reductions averaged over four forecast horizons for each dataset. Appendix D.2 gives detailed settings and additional analysis.

Table 4: Relative error reductions (%) of Global-Unique compared with WindowLocal-Repeated, averaged over $H \in \{96, 192, 336, 720\}$ for each dataset. Positive values (blue) favor Global-Unique; negative values (red) favor WindowLocal-Repeated. Darker colors indicate larger absolute differences on a shared color scale.

Dataset	CF-PatchTST ($N = 8$)		CF-WPMixer ($N = 4$)	
	MSE	MAE	MSE	MAE
Electricity (Last7)	+0.17%	−0.06%	−1.48%	−2.08%
ETTh1	+0.57%	+0.86%	−0.84%	−0.64%
ETTh2	+6.84%	+4.06%	+1.60%	+2.49%
ETTm1	+4.96%	+3.93%	−0.53%	+0.48%
ETTm2	+5.21%	+3.35%	+2.00%	+0.18%
Exchange	+6.88%	+3.38%	+11.91%	+5.52%
Traffic (Last7)	−2.04%	−2.91%	−3.30%	−2.41%
Weather	+0.51%	+0.10%	+1.79%	+1.66%
Solar-Energy	−5.50%	−3.49%	−2.76%	−0.74%
Backbone average	+1.95%	+1.02%	+0.93%	+0.50%

Table 4 shows that Global-Unique reduces both MSE and MAE on average for both backbones. The mean relative reductions are 1.95% in MSE and 1.02% in MAE for PatchTST, compared with 0.93% and 0.50% for WPMixer. Global-Unique thus improves overall forecasting accuracy, with larger gains for PatchTST.

5 CONCLUSION

We introduced CROSSFOREST, which assigns a learner with independent parameters to each temporal phase and interleaves the learners' predictions into a complete forecast. Its Global-Unique protocol combines globally aligned phase assignments with duplicate-free training-task construction. It preserves all distinct phase supervision pairs from stride-one sliding windows while avoiding repeated construction of phase tasks shared by overlapping windows. On nine datasets, WPMixer and PatchTST with CROSSFOREST outperform the baselines on several tasks, and eight of nine backbones show lower test MSE and MAE with validation-selected phase counts. Across 324 matched comparisons, the average reductions are 7.84% for MSE and 4.76% for MAE. The gains vary with the backbone, phase count, and channel setting. In particular, CROSSFOREST performs better in low-channel settings; the Global-Unique protocol and independently parameterized phase learners further improve predictive accuracy. These findings support the design of globally aligned assignments, duplicate-free task construction, and phase specialization.

AI USE STATEMENT

In this work, we used generative AI tools to assist with writing portions of the experimental code. We have not used generative AI tools for experimental design or the development of the core research ideas. Additionally, we used generative AI tools for English translation, language polishing, and formatting. We have reviewed all AI-assisted work, including the translated and polished English text and the generated code. We take responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the aid of generative AI.

REFERENCES

Shane Bergsma, Tim Zeyl, and Lei Guo. SutraNets: Sub-series autoregressive networks for long-sequence, probabilistic forecasting. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 30518–30533. Curran Associates, Inc., 2023. doi: 10.52202/075280-1330.

Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, August 1996. ISSN 1573-0565. doi: 10.1007/BF00058655.

Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6405–6416, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 978-1-5108-6096-4.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. SparseTSF: Modeling long-term time series forecasting with 1k parameters. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24, Vienna, Austria, 2024. JMLR.org.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. SCINet: Time series modeling and forecasting with sample convolution and interaction. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, Nips '22, Red Hook, NY, USA, 2022a. Curran Associates Inc. ISBN 978-1-7138-7108-8.

Peiyuan Liu, Beiliang Wu, Yifan Hu, Naiqi Li, Tao Dai, Jigang Bao, and Shu-tao Xia. TimeBridge: Non-stationarity matters for long-term time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25, Vancouver, Canada, 2025. JMLR.org.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, Nips '22, Red Hook, NY, USA, 2022b. Curran Associates Inc. ISBN 978-1-7138-7108-8.

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 *The Twelfth International Conference on Learning Representations*, 2024.

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.

Vahid Naghashi, Mounir Boukadoum, and Abdoulaye Banire Diallo. A multiscale model for multivariate time series forecasting. *Scientific Reports*, 15(1):1565, January 2025. ISSN 2045-2322. 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 *The Eleventh International Conference on Learning Representations*, 2023.

Noam Shazeer, *Azalia Mirhoseini, *Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

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. *arXiv preprint arXiv:2410.16032*, 2024a.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and JUN ZHOU. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024b.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, Nips '21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 978-1-7138-4539-3.

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

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i9.26317.

Tianping Zhang, Yizhuo Zhang, Wei Cao, Jiang Bian, Xiaohan Yi, Shun Zheng, and Jian Li. Less Is More: Fast Multivariate Time Series Forecasting with Light Sampling-oriented MLP Structures, July 2022.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*, 2023.

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 *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proc. 39th International Conference on Machine Learning (ICML 2022)*, Baltimore, Maryland, 2022.

A PHASE DECOMPOSITION PROPERTIES

Exact phase decomposition. Given the start phase ρ_s, each index $i \in \{0, \dots, L-1\}$ determines a unique residue $r = (\rho_s + i) \bmod N$ and a unique within-phase index $j = \lfloor i/N \rfloor$. The mapping from $\boldsymbol{X}_s$ to $\{\boldsymbol{X}_s^{(r)}\}_{r=0}^{N-1}$ is therefore bijective. The same holds for $\boldsymbol{Y}_s$, with Equation 4 giving the inverse mapping. Since interleaving only permutes entries, the MSE over a complete window decomposes exactly:

$$\frac{\|\widehat{\boldsymbol{Y}}_s - \boldsymbol{Y}_s\|_F^2}{HC} = \frac{1}{N} \sum_{r=0}^{N-1} \frac{\|\widehat{\boldsymbol{Y}}_s^{(r)} - \boldsymbol{Y}_s^{(r)}\|_F^2}{(H/N)C}. \tag{10}$$

This identity gives exact loss accounting across the phase learners.

Unique task count. The first window produces N distinct phase tasks. When the window of length L advances by one time step, only the phase subsequence containing the outgoing timestamp changes; the other $N-1$ remain unchanged. Since H is also divisible by N, the same holds for the target sequence. Each subsequent window therefore adds one new input and target pair, for a total of $N + (W-1)$ distinct tasks across W windows.

Gradient separation. The loss for a task in phase r depends only on θ_r, so $\nabla_{\theta_q} \ell_{w,r}(\Theta) = 0$ for $q \neq r$. This property holds without phase-pure batching. Grouping tasks by phase makes routing explicit and avoids running learners that have no examples in the batch.

Time-origin shifts. The choice of global time origin affects only the phase labels, not how the data are grouped. If all time indices are shifted by δ sampling time steps, phase r is relabeled as $(r+\delta) \bmod N$. Reassigning the corresponding learner parameters to the new labels keeps each data point with its original learner, so the predictions remain unchanged.

Phase labels and learner parameters must be adjusted together. Restarting the index at zero at the beginning of the validation or test split may change the phase labels. Keeping the learner indices used during training may then route data to the wrong learner. The implementation should therefore preserve consistent global phases or adjust the mapping between phase labels and learners accordingly.

B Experimental protocol

B.1 Dataset preparation

All datasets are split chronologically, with dataset statistics and target partitions listed in Table 5. ETT uses fixed splits covering only the first 14,400 hourly observations or the first 57,600 observations sampled at 15-minute intervals. Other datasets use a 70%/10%/20% split, with boundaries rounded down to integers. Standardization parameters are fitted on the training split and reused for validation and testing. ElectricityLast7 and TrafficLast7 retain all rows and the last seven source variables, including the target; their evaluations cover only these fixed subsets. The loaders do not fill missing values or adjust sampling intervals. Solar-Energy has no timestamps, so its loader supplies four time-feature channels filled with zeros and containing no calendar information.

B.2 Evaluation windows

For a validation or test split D with target interval $[a_D, b_D)$, a window starting at s must satisfy $s + L \geq a_D$ and $s + L + H \leq b_D$. Historical inputs may include observations from the preceding split, but all forecast targets must lie within the designated target interval. We retain each sample's absolute start index s to keep phase labels consistent across splits.

B.3 Training settings

All screening experiments use an input length of $L = 512$, forecast horizons $H \in \{96, 192, 336, 720\}$, and phase counts $N \in \{1, 2, 4, 8\}$. We train with Adam and MSE loss and set the gradient clipping threshold to 1.0. Tables 6 and 7 list the batch sizes used and the main default settings for each model.

DLinear applies separate seasonal and trend projections to each channel. For $N = 1, 2, 4, 8$, DLinear, Autoformer, and TimeMixer use moving-average kernel lengths of 25, 13, 7, and 3, respectively. TimeMixer processes channels independently and uses normalization and moving-average decomposition. Three average-pooling stages with a window size of two produce four temporal resolutions. Its frequency-selection setting of five is inactive under moving-average decomposition and applies only to the discrete Fourier transform decomposition branch. Crossformer uses two encoder scale blocks and three decoder blocks; the decoder depth is determined by the implementation rather than the supplied decoder-layer setting. The attention-head setting has no effect on TimeMixer or TimesNet.

Table 5: Dataset metadata and chronological target partitions. Intervals are zero-based and half open. Validation and test prepend the preceding $L = 512$ observations as historical context without overlapping target partitions.

Dataset	Observations	Variables	Cadence	Train / validation / test target partition
ETTh1, ETTh2	17,420	7	1 hour	$[0, 8640)$ / $[8640, 11520)$ / $[11520, 14400)$
ETTm1, ETTm2	69,680	7	15 min	$[0, 34560)$ / $[34560, 46080)$ / $[46080, 57600)$
Electricity	26,304	321	1 hour	$[0, 18412)$ / $[18412, 21044)$ / $[21044, 26304)$
Traffic	17,544	862	1 hour	$[0, 12280)$ / $[12280, 14036)$ / $[14036, 17544)$
Exchange	7,588	8	1 day	$[0, 5311)$ / $[5311, 6071)$ / $[6071, 7588)$
Weather	52,696	21	10 min	$[0, 36887)$ / $[36887, 42157)$ / $[42157, 52696)$
Solar-Energy	52,560	137	10 min	$[0, 36792)$ / $[36792, 42048)$ / $[42048, 52560)$

Table 6: Effective screening batch sizes on the unified nine-dataset grid.

Backbone	ETT	ElectricityLast7	TrafficLast7	Exchange	Weather	Solar-Energy
WPMixer	16	8	8	8	8	2
MultiPatchFormer	16	8	8	8	8	4
TimeMixer	16	16	8	32	16	16
iTransformer	32	16	16	32	32	16
PatchTST	16	8	4	32	32	16
Crossformer	16	8	8	8	8	4
TimesNet	32	16	16	16	16	8
DLinear	8	16	8	32	32	16
Autoformer	16	16	16	16	16	8

Table 7: Principal screening defaults.

Backbone	d_{model}	d_{ff}	Enc. layers	Heads	Learning rate	Epochs / patience	Additional setting
WPMixer	32	64	1	8	10^{-3}	10 / 3	patch 16, stride 8; wavelet db2, level 1
MultiPatchFormer	32	64	1	4	5×10^{-4}	10 / 3	patches 8/16/24/32; strides 8/8/7/6
TimeMixer	16 / 32	32 / 64	2 / 3	–	10^{-3}	20 / 5	Traffic uses 32/64; ETT uses 2 blocks, others 3
iTransformer	128 / 512	128 / 512	2–4	8	10^{-4}–10^{-3}	10 / 3	dataset-specific capacity
DLinear	–	–	–	–	5×10^{-3} / 5×10^{-4}	10 / 3	individual mode; $H < 720$ / $H = 720$
PatchTST	512	2048	2	8	10^{-4}	10 / 3	patch 16, stride 8
Autoformer	256	512	2	8	10^{-4}	10 / 3	one decoder layer
TimesNet	128	256	2	–	10^{-4}	10 / 3	top-k 5, six kernels
Crossformer	128	256	2	8	10^{-4}	10 / 3	three decoder blocks; segment 12, window 2

Each training batch contains tasks from a single phase, with tasks shuffled without replacement. During evaluation, we run all phase learners sequentially and interleave their outputs to reconstruct the complete forecast, on which we compute the evaluation metrics.

During screening, we compute and report validation and test MSE after each epoch by averaging batch MSE values. Only validation loss determines checkpoint selection and early stopping.

Execution environment. Experiments run with Python 3.11.14, PyTorch 2.2.2+cu121 built with CUDA 12.1, and cuDNN 8.9.2. The machine has two NVIDIA RTX 6000 Ada Generation GPUs, each with 49,140 MiB of memory, and uses driver 580.159.03. The system CUDA toolkit is version 13.0.

Table 8: Dataset-specific iTransformer screening defaults. All runs use eight heads, 10 epochs, and patience 3.

Scope	d_{model}	d_{ff}	Layers	LR	Batch
ETT*	128	128	2	10^{-4}	32
ElectricityLast7	512	512	3	5×10^{-4}	16
TrafficLast7	512	512	4	10^{-3}	16
Exchange	512	512	2	10^{-4}	32
Weather	512	512	3	10^{-4}	32
Solar-Energy	512	512	2	10^{-4}	16

B.4 PHASE-COUNT SELECTION

For each backbone, we compare CROSSFOREST configurations at $N \in \{2, 4, 8\}$ with the $N = 1$ control in the same framework. All configurations use $L = 512$ and are evaluated on nine datasets at four forecast horizons, $H \in \{96, 192, 336, 720\}$. This common task set $\mathcal{G}$ gives 36 matched comparisons per backbone. Electricity and Traffic use the same fixed seven-channel Last7 subsets for all phase counts.

We select one phase count N_b for each backbone using only validation data. Following the notation in Equation 9, $E^{\text{MSE,val}}_{b,N,g}$ denotes the minimum validation MSE recorded in the training log for backbone b, phase count N, and task g. This is the minimum recorded Vali Loss. This value is used only for candidate screening. We choose the phase count with the largest mean relative reduction across tasks compared with $N = 1$:

$$\begin{aligned} N_b &= \underset{N\in\{2,4,8\}}{\arg\max}\; G^{\text{val}}_{\text{MSE}}(b, N) \\ &= \underset{N\in\{2,4,8\}}{\arg\max}\; \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} 100\frac{E^{\text{MSE,val}}_{b,1,g} - E^{\text{MSE,val}}_{b,N,g}}{E^{\text{MSE,val}}_{b,1,g}}. \end{aligned} \tag{11}$$

All four horizons contribute to the choice of N_b.

B.5 COMPLETE RESULTS

Tables 10–18 provide the detailed validation and test metrics underlying Tables 9 and 2. For each backbone, phase count, and forecast horizon, we compute the relative validation-MSE reduction from $N = 1$ for each dataset, then average these reductions equally across the nine datasets to obtain Table 9. Positive values indicate lower error. Averaging these values over the four horizons gives $G^{\text{val}}_{\text{MSE}}(b, N)$ for each candidate phase count. The phase count with the largest mean is selected as N_b, and this mean is the validation reduction reported in Table 2.

The test reductions in Table 2 are computed from the corresponding test metrics at N_b and $N = 1$ in Tables 10–18, with relative reductions averaged equally across the 36 tasks per backbone. W/L counts tasks with lower or higher test error than the $N = 1$ control.

Table 9: Validation-MSE reduction (%) by horizon and phase count on the same nine-dataset grid as Table 2. Each value macro-averages nine matched dataset-level reductions; Electricity and Traffic use Last7. Blue and red cells indicate positive and negative reductions, respectively. Color intensity increases with the absolute reduction on the same scale as Table 2.

Backbone	H	$N=2$	$N=4$	$N=8$	Backbone	H	$N=2$	$N=4$	$N=8$
WPMixer	96	+0.221%	−0.067%	−4.694%	MultiPatchFormer	96	+1.030%	−0.414%	−4.219%
	192	+1.691%	+1.189%	−1.607%		192	+2.737%	+2.744%	+0.250%
	336	+3.743%	+4.355%	+2.214%		336	+0.813%	+1.123%	−0.505%
	720	+5.656%	+7.850%	+4.650%		720	−0.685%	+1.261%	+0.171%
TimeMixer	96	+6.394%	+5.005%	+1.798%	iTransformer	96	+2.192%	−0.402%	−2.772%
	192	+6.241%	+5.891%	+2.543%		192	+2.561%	+2.071%	+0.193%
	336	+6.469%	+6.477%	+3.079%		336	+5.653%	+4.477%	+3.442%
	720	+4.742%	+3.047%	+8.065%		720	+1.620%	+3.590%	+7.150%
PatchTST	96	+1.699%	+4.344%	+3.038%	Crossformer	96	−4.174%	−2.747%	−6.688%
	192	+1.519%	+4.559%	+5.070%		192	−2.412%	−2.754%	−5.843%
	336	+0.090%	+4.113%	+5.389%		336	+13.828%	+13.963%	+11.409%
	720	+6.193%	+4.994%	+7.175%		720	+9.264%	+7.983%	+6.037%
TimesNet	96	+11.638%	+11.298%	+11.831%	DLinear	96	−1.024%	−2.982%	−8.488%
	192	+16.436%	+19.556%	+21.156%		192	+0.436%	−2.671%	−5.191%
	336	+14.652%	+18.119%	+21.480%		336	+0.461%	−0.389%	−3.252%
	720	+5.642%	+2.245%	+7.503%		720	+0.689%	+0.273%	+1.845%
Autoformer	96	+4.040%	+5.576%	+0.712%					
	192	+8.355%	+15.532%	+10.304%					
	336	+13.633%	+16.273%	+13.937%					
	720	+7.060%	+9.916%	+5.645%					

Table 10: Combined screening validation and test metrics for WPMixer. $N = 4$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: WPMixer

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	**0.680**	0.381	0.405	0.682	0.376	0.405	0.681	**0.363**	**0.393**	0.708	0.378	0.403
	192	1.028	0.417	0.431	0.948	0.415	0.427	**0.944**	**0.409**	**0.423**	0.954	0.410	**0.423**
	336	1.420	0.463	0.462	1.306	0.436	0.444	1.199	0.436	0.443	**1.193**	**0.428**	**0.436**
	720	1.634	0.564	0.519	1.455	0.470	0.472	1.443	**0.458**	**0.468**	**1.440**	0.462	0.471
ETTh2	96	0.234	0.289	0.353	0.227	0.280	0.347	**0.219**	**0.278**	**0.343**	0.222	0.286	0.348
	192	0.333	0.355	0.396	0.303	**0.336**	0.389	0.288	0.337	**0.382**	**0.287**	0.354	0.391
	336	0.410	0.376	**0.413**	0.407	0.379	0.428	**0.375**	**0.371**	**0.413**	0.379	0.377	0.414
	720	0.809	0.486	0.490	0.702	0.437	0.461	**0.617**	0.415	0.447	0.624	**0.409**	**0.444**
ETTm1	96	0.368	0.292	0.346	**0.366**	**0.289**	**0.344**	0.372	0.292	0.345	0.394	0.306	0.353
	192	0.505	0.341	0.378	**0.490**	0.337	0.371	**0.490**	**0.331**	**0.367**	0.507	0.338	0.371
	336	0.672	0.370	0.398	0.646	0.372	0.392	**0.638**	**0.364**	**0.385**	0.654	0.370	0.388
	720	1.030	0.428	0.433	0.980	0.438	0.428	0.963	0.426	0.418	**0.955**	**0.425**	**0.417**
ETTm2	96	0.122	**0.168**	0.260	**0.120**	0.175	**0.256**	**0.120**	0.170	0.259	0.122	0.173	0.263
	192	0.166	0.232	**0.297**	0.162	0.230	0.300	0.162	0.227	0.301	**0.160**	**0.226**	0.299
	336	0.218	0.288	0.334	0.210	0.279	0.333	0.202	0.278	0.334	**0.201**	**0.275**	**0.332**
	720	0.290	0.366	0.391	0.278	0.363	**0.389**	0.274	0.361	**0.389**	**0.272**	**0.360**	0.390
ElectricityLast7	96	**0.086**	0.134	0.231	0.087	**0.129**	**0.230**	0.090	0.134	0.237	0.103	0.150	0.256
	192	0.099	0.148	0.248	**0.096**	**0.144**	**0.243**	0.099	0.146	0.248	0.112	0.163	0.268
	336	0.109	0.163	0.267	0.108	0.161	**0.261**	**0.107**	**0.159**	0.263	0.119	0.173	0.279
	720	0.140	0.219	0.325	0.121	0.191	0.299	**0.119**	**0.182**	**0.293**	0.130	0.195	0.308
TrafficLast7	96	**0.337**	0.262	0.254	0.339	**0.259**	**0.247**	0.350	0.269	0.253	0.368	0.288	0.279
	192	**0.337**	0.277	0.266	0.342	**0.274**	**0.257**	0.354	0.281	0.262	0.373	0.302	0.287
	336	**0.352**	0.294	0.280	0.353	**0.289**	**0.269**	0.363	0.296	0.271	0.381	0.318	0.297
	720	0.351	0.329	0.310	**0.333**	**0.315**	0.289	0.341	0.318	**0.286**	0.365	0.342	0.312
Exchange	96	0.150	0.109	0.233	0.148	0.098	0.219	**0.146**	**0.092**	**0.212**	0.151	0.095	0.216
	192	**0.226**	0.208	0.326	0.238	0.203	0.322	0.249	0.200	0.317	0.246	**0.191**	**0.312**
	336	0.432	0.532	0.514	**0.378**	0.374	0.447	0.426	0.367	0.438	0.411	**0.354**	**0.433**
	720	**1.512**	1.443	0.861	1.594	1.319	0.824	1.575	**0.959**	**0.732**	1.753	0.968	0.741
Weather	96	**0.377**	0.145	0.200	0.383	**0.143**	**0.195**	0.384	0.145	0.197	0.398	0.151	0.203
	192	**0.441**	0.190	0.241	0.445	**0.186**	**0.234**	0.444	0.187	0.237	0.457	0.193	0.243
	336	0.536	0.253	0.293	0.514	0.241	0.279	**0.513**	**0.238**	**0.278**	0.526	0.243	0.281
	720	0.635	0.333	0.344	0.610	0.321	0.336	**0.604**	0.317	0.333	0.622	**0.316**	**0.332**
Solar-Energy	96	**0.159**	**0.186**	**0.247**	0.162	0.189	0.250	0.160	0.194	0.251	0.163	0.203	0.259
	192	**0.170**	**0.201**	**0.259**	0.174	**0.201**	0.261	0.172	0.208	0.265	0.175	0.220	0.272
	336	**0.178**	**0.203**	**0.259**	**0.178**	0.206	0.265	0.181	0.211	0.271	0.184	0.230	0.282
	720	0.176	0.213	**0.270**	0.174	0.218	0.276	**0.167**	**0.211**	0.272	**0.167**	0.230	0.281

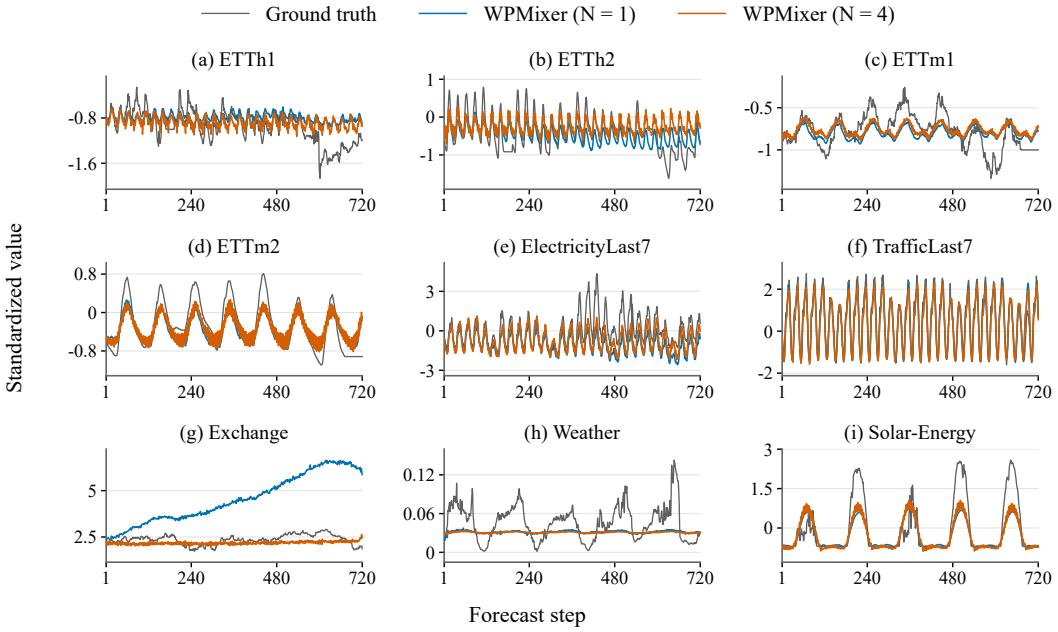

Figure 4: WPMixer forecasts at $N = 1$ and $N = 4$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 11: Combined screening validation and test metrics for MultiPatchFormer. $N = 4$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: MultiPatchFormer

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	0.772	0.399	0.425	0.748	0.379	0.411	**0.737**	0.377	0.406	0.747	**0.375**	**0.402**
	192	1.115	0.432	0.449	1.050	0.437	0.449	1.001	0.411	0.430	**0.963**	**0.402**	**0.419**
	336	1.436	0.446	0.458	1.334	0.453	0.464	1.242	0.469	0.471	**1.183**	**0.417**	**0.431**
	720	1.798	0.541	0.519	1.679	0.515	0.515	1.488	0.503	0.505	**1.424**	**0.444**	**0.467**
ETTh2	96	0.232	0.344	0.380	0.224	0.317	0.363	**0.217**	0.289	0.348	0.228	**0.285**	**0.347**
	192	0.305	0.393	0.410	**0.285**	0.378	0.404	0.286	0.361	0.395	0.294	**0.350**	**0.387**
	336	0.389	0.399	0.424	0.385	0.399	0.430	**0.377**	0.382	0.418	0.387	**0.379**	**0.414**
	720	0.677	0.465	0.472	0.671	**0.415**	**0.446**	0.651	0.419	0.452	**0.643**	0.419	0.451
ETTm1	96	**0.374**	0.297	0.355	0.376	**0.296**	**0.353**	0.383	0.301	0.355	0.405	0.311	0.360
	192	0.509	0.344	0.386	0.502	0.340	0.382	**0.500**	**0.336**	**0.376**	0.520	0.343	0.378
	336	**0.638**	0.379	0.404	0.647	0.373	0.400	0.647	**0.367**	0.395	0.664	0.371	**0.394**
	720	0.968	0.430	0.437	0.968	0.426	0.429	**0.953**	**0.418**	0.426	0.966	0.425	**0.424**
ETTm2	96	0.126	0.186	0.270	0.126	0.172	**0.262**	**0.123**	**0.171**	**0.262**	0.124	0.176	0.265
	192	0.170	0.247	0.313	0.169	0.242	0.313	0.166	0.232	0.307	**0.162**	**0.231**	**0.304**
	336	0.214	0.293	0.347	0.214	0.289	0.345	0.207	0.286	0.342	**0.204**	**0.280**	**0.337**
	720	0.293	0.379	0.403	0.276	0.365	0.395	0.272	0.367	0.394	**0.271**	**0.361**	**0.390**
ElectricityLast7	96	0.096	0.139	0.247	0.093	**0.138**	0.243	**0.091**	**0.138**	0.242	0.092	0.139	**0.241**
	192	0.117	0.169	0.279	0.112	0.164	0.272	0.109	0.162	0.270	**0.104**	**0.156**	**0.260**
	336	0.126	0.182	0.295	0.121	0.178	0.289	0.121	0.177	0.288	**0.117**	**0.172**	**0.280**
	720	0.136	0.204	0.324	0.135	0.202	0.319	0.134	0.202	0.319	**0.133**	**0.201**	**0.316**
TrafficLast7	96	**0.327**	**0.257**	**0.253**	0.331	**0.257**	**0.253**	0.343	0.264	0.257	0.353	0.274	0.258
	192	**0.334**	**0.279**	**0.277**	0.343	0.284	0.283	0.353	0.290	0.284	0.361	0.294	0.279
	336	**0.350**	**0.294**	**0.287**	0.352	0.305	0.299	0.369	0.315	0.305	0.374	0.319	0.301
	720	0.340	**0.327**	**0.315**	**0.333**	0.331	0.316	0.349	0.344	0.324	0.364	0.349	0.323
Exchange	96	0.151	0.107	0.235	**0.140**	0.100	0.227	0.155	**0.096**	**0.220**	0.163	0.104	0.228
	192	0.276	**0.190**	**0.319**	**0.227**	0.207	0.334	**0.227**	0.228	0.348	0.263	0.207	0.329
	336	0.419	0.407	0.477	0.401	0.393	0.468	**0.397**	**0.375**	**0.453**	0.426	0.381	0.454
	720	**1.667**	1.069	0.788	1.996	1.129	0.813	1.907	**0.940**	**0.738**	1.998	0.973	0.745
Weather	96	**0.387**	**0.150**	**0.202**	0.388	0.152	0.204	0.389	0.152	**0.202**	0.401	0.158	0.208
	192	**0.451**	**0.194**	**0.240**	0.459	0.196	0.245	0.455	**0.194**	0.242	0.464	0.200	0.247
	336	0.527	0.244	0.283	0.524	**0.242**	**0.280**	**0.522**	0.243	**0.280**	0.533	0.247	0.284
	720	0.619	0.312	0.331	0.622	0.315	0.335	**0.617**	**0.311**	**0.330**	0.626	0.315	0.334
Solar-Energy	96	**0.130**	**0.166**	**0.225**	0.137	0.173	0.236	0.146	0.182	0.243	0.158	0.196	0.254
	192	**0.144**	**0.178**	**0.241**	0.154	0.187	0.253	0.164	0.199	0.264	0.175	0.215	0.272
	336	**0.155**	**0.185**	**0.250**	0.167	0.195	0.261	0.175	0.212	0.277	0.187	0.230	0.284
	720	**0.161**	**0.200**	**0.262**	0.164	0.203	0.265	0.166	0.215	0.281	0.171	0.234	0.291

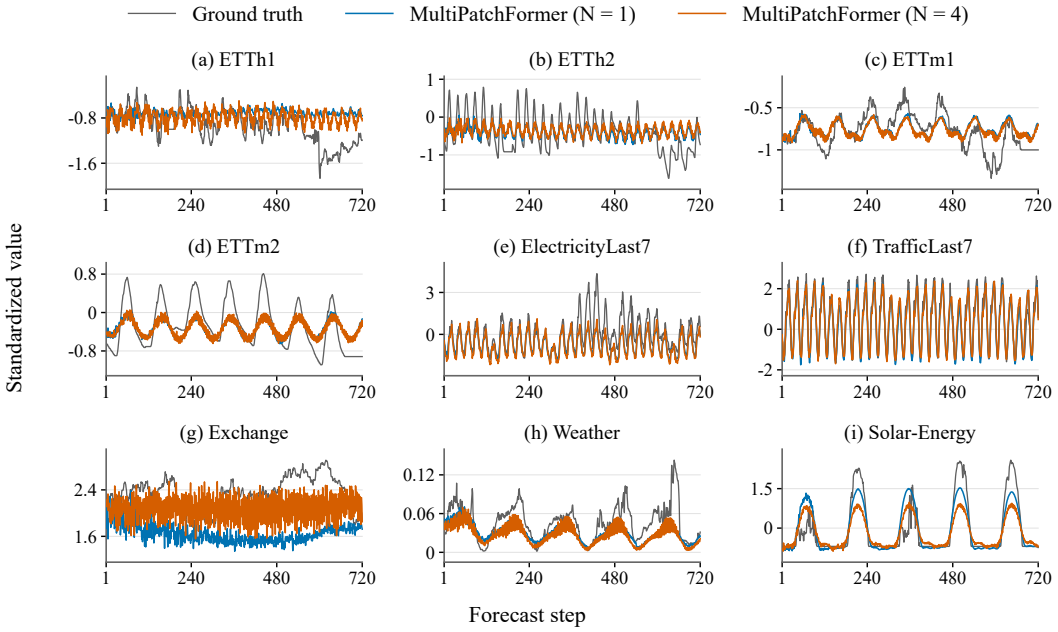

Figure 5: MultiPatchFormer forecasts at $N = 1$ and $N = 4$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 12: Combined screening validation and test metrics for TimeMixer. $N = 2$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: TimeMixer

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	0.778	0.392	0.415	**0.699**	**0.374**	**0.403**	0.710	0.376	0.404	0.735	0.375	**0.403**
	192	1.193	0.436	0.448	1.035	0.421	0.435	0.984	0.419	0.435	**0.974**	**0.405**	**0.421**
	336	1.572	0.544	0.513	1.412	0.462	0.466	1.216	0.444	0.452	**1.205**	**0.428**	**0.437**
	720	1.946	0.655	0.555	1.582	0.493	0.485	1.467	0.496	0.493	**1.442**	**0.458**	**0.468**
ETTh2	96	0.252	0.301	0.354	**0.224**	0.287	**0.349**	**0.224**	**0.286**	0.350	0.225	0.289	**0.349**
	192	0.337	0.363	0.401	0.311	0.351	0.394	0.295	0.354	0.390	**0.294**	**0.349**	**0.388**
	336	0.456	0.391	0.418	0.398	0.376	0.416	**0.378**	0.382	0.420	0.385	**0.375**	**0.414**
	720	0.736	0.438	0.458	0.699	0.440	0.461	**0.641**	0.416	0.447	**0.641**	**0.411**	**0.443**
ETTm1	96	0.403	0.315	0.362	**0.379**	**0.301**	**0.354**	0.385	0.303	0.358	0.397	0.309	0.359
	192	0.541	0.366	0.391	**0.505**	**0.337**	0.377	0.509	0.340	0.378	0.511	0.341	**0.375**
	336	0.725	0.397	0.416	0.662	0.373	0.399	0.665	**0.366**	0.395	**0.657**	0.372	**0.391**
	720	1.062	0.490	0.465	0.996	0.440	0.438	0.973	0.442	0.432	**0.963**	**0.426**	**0.420**
ETTm2	96	0.124	0.178	**0.261**	**0.123**	0.175	0.263	0.124	0.176	0.268	0.124	**0.174**	0.265
	192	0.176	0.244	0.309	0.166	0.234	0.302	**0.163**	0.228	0.302	0.164	**0.227**	**0.301**
	336	0.231	0.306	0.347	0.214	0.291	0.341	0.209	**0.274**	**0.331**	**0.206**	0.276	0.333
	720	0.291	0.389	0.400	**0.276**	0.370	0.392	0.277	0.362	0.389	0.277	**0.358**	**0.388**
ElectricityLast7	96	0.091	0.135	0.242	**0.088**	**0.132**	**0.234**	**0.088**	0.134	0.236	0.091	0.137	0.240
	192	0.104	0.154	0.263	0.100	0.150	0.251	**0.098**	**0.148**	**0.250**	0.102	0.152	0.255
	336	0.110	0.165	0.274	0.110	0.166	0.273	**0.109**	**0.164**	**0.269**	0.113	0.166	0.274
	720	0.125	0.194	0.311	**0.122**	0.194	0.308	0.124	**0.191**	**0.305**	0.129	0.194	0.309
TrafficLast7	96	**0.328**	**0.258**	0.253	0.331	0.263	0.261	0.337	0.263	0.256	0.339	0.265	**0.250**
	192	0.329	0.274	0.267	**0.326**	**0.272**	0.265	0.339	0.276	0.264	0.342	0.277	**0.261**
	336	**0.341**	0.301	0.289	0.342	0.295	0.284	0.342	**0.287**	**0.272**	0.351	0.293	0.273
	720	0.331	0.331	0.321	**0.322**	0.321	0.311	**0.322**	**0.313**	0.295	0.329	0.314	**0.288**
Exchange	96	0.190	**0.099**	**0.222**	**0.153**	0.118	0.245	0.157	0.100	**0.222**	0.182	0.119	0.247
	192	0.258	**0.195**	**0.313**	**0.235**	0.227	0.342	0.248	0.200	0.318	0.303	0.234	0.349
	336	0.427	0.381	0.449	**0.381**	0.408	0.461	0.454	**0.368**	**0.444**	0.535	0.424	0.480
	720	1.501	1.411	0.852	1.546	1.289	0.842	1.968	**0.955**	**0.733**	**1.198**	1.068	0.789
Weather	96	0.400	0.153	0.206	0.385	**0.146**	**0.197**	**0.384**	0.148	0.201	0.396	0.154	0.206
	192	0.477	0.208	0.256	**0.446**	**0.193**	**0.244**	**0.446**	**0.193**	**0.244**	0.458	0.196	0.245
	336	0.578	0.278	0.310	0.515	**0.244**	**0.282**	**0.512**	**0.244**	0.283	0.526	0.246	0.283
	720	0.640	0.350	0.359	**0.602**	0.320	0.336	**0.602**	0.325	0.340	0.618	**0.316**	**0.332**
Solar-Energy	96	0.155	0.186	0.236	**0.148**	**0.180**	**0.233**	0.156	0.187	0.246	0.161	0.199	0.254
	192	0.166	0.194	0.262	**0.162**	**0.192**	**0.250**	0.170	0.200	0.259	0.174	0.215	0.271
	336	**0.166**	0.205	**0.259**	0.170	0.202	0.261	0.175	**0.201**	0.262	0.183	0.226	0.282
	720	**0.169**	**0.208**	0.277	**0.169**	0.209	**0.268**	0.172	0.209	0.274	0.170	0.229	0.285

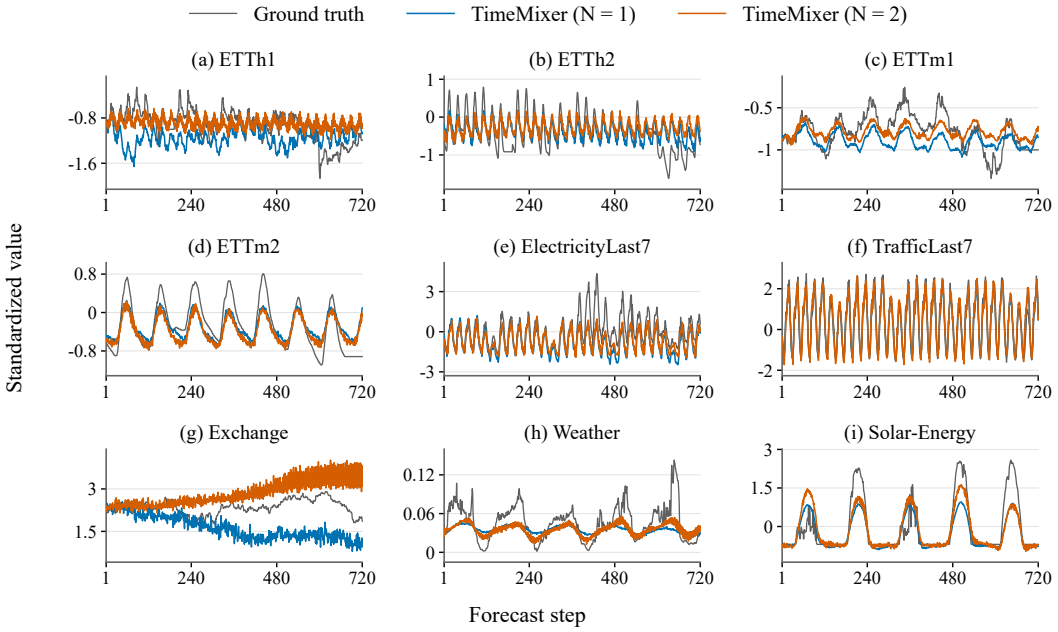

Figure 6: TimeMixer forecasts at $N = 1$ and $N = 2$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 13: Combined screening validation and test metrics for iTransformer. $N = 2$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: iTransformer

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	**0.699**	0.398	**0.422**	0.723	**0.397**	0.423	0.779	0.409	0.428	0.839	0.414	0.429
	192	**0.943**	0.431	0.445	0.958	**0.429**	**0.444**	1.008	0.441	0.451	1.053	0.443	0.451
	336	1.192	**0.445**	**0.457**	**1.186**	0.448	0.460	1.239	0.465	0.469	1.271	0.460	0.468
	720	1.471	0.519	0.519	**1.451**	0.518	0.516	1.505	0.523	0.516	1.572	**0.516**	**0.515**
ETTh2	96	**0.227**	0.297	**0.356**	0.233	**0.295**	0.357	0.245	0.302	0.365	0.254	0.322	0.378
	192	**0.300**	0.377	0.405	0.308	**0.360**	**0.398**	0.315	0.364	0.402	0.324	0.375	0.413
	336	**0.389**	0.419	0.436	0.405	0.397	0.425	0.408	**0.385**	**0.423**	0.425	0.393	0.431
	720	**0.650**	0.444	0.465	0.654	**0.426**	**0.456**	0.661	**0.426**	0.457	0.688	0.436	0.463
ETTm1	96	**0.377**	**0.308**	0.363	0.379	0.309	**0.362**	0.396	0.316	0.364	0.418	0.328	0.370
	192	**0.496**	0.349	0.385	0.497	**0.345**	**0.382**	0.515	0.350	0.384	0.537	0.360	0.390
	336	0.646	0.377	0.403	**0.645**	**0.374**	**0.400**	0.664	0.381	0.402	0.685	0.391	0.409
	720	0.964	0.438	0.440	**0.951**	**0.433**	**0.434**	0.966	0.438	0.436	0.985	0.446	0.441
ETTm2	96	0.126	0.179	0.272	**0.125**	**0.176**	**0.269**	0.127	0.181	0.273	0.129	0.185	0.276
	192	0.168	0.245	0.315	0.169	**0.237**	0.312	**0.167**	0.239	0.312	0.169	0.238	**0.311**
	336	0.208	0.294	0.347	0.208	**0.286**	**0.342**	**0.207**	0.289	0.344	0.211	0.289	0.343
	720	0.278	0.375	0.397	**0.275**	0.376	**0.394**	**0.275**	**0.373**	**0.394**	0.276	0.375	0.395
ElectricityLast7	96	0.087	0.138	0.244	**0.086**	**0.134**	0.238	**0.086**	**0.134**	0.237	0.087	**0.134**	**0.236**
	192	0.097	0.156	0.261	**0.094**	0.153	0.257	**0.094**	0.151	0.254	**0.094**	**0.148**	**0.249**
	336	0.107	0.173	0.278	0.104	0.171	0.275	0.101	0.168	0.274	**0.099**	**0.161**	**0.266**
	720	0.116	0.212	0.310	0.109	0.203	0.305	0.108	0.185	0.296	**0.107**	**0.178**	**0.290**
TrafficLast7	96	0.409	0.341	0.355	**0.339**	**0.260**	0.251	0.343	0.262	0.250	0.348	0.263	**0.245**
	192	0.356	0.288	0.288	**0.342**	0.271	0.264	0.344	**0.268**	0.255	0.346	0.270	**0.250**
	336	0.364	0.294	0.289	**0.353**	0.282	0.267	0.357	**0.280**	**0.261**	0.364	0.293	0.271
	720	0.387	0.359	0.351	0.342	0.302	0.289	0.342	0.300	0.277	**0.334**	**0.296**	**0.268**
Exchange	96	0.168	**0.104**	**0.231**	**0.161**	0.112	0.240	0.166	0.123	0.252	**0.161**	0.109	0.235
	192	0.330	**0.190**	**0.311**	0.278	0.214	0.334	0.267	0.234	0.353	**0.261**	0.231	0.349
	336	0.852	**0.372**	**0.442**	**0.458**	0.380	0.451	0.514	0.397	0.461	**0.458**	0.400	0.465
	720	2.677	**0.940**	**0.733**	2.962	0.964	0.748	2.372	1.012	0.764	**1.218**	1.022	0.768
Weather	96	0.410	**0.162**	**0.215**	**0.408**	0.163	**0.215**	0.418	0.165	0.217	0.419	0.171	0.221
	192	0.490	0.212	0.258	0.472	**0.204**	**0.252**	**0.470**	0.205	0.253	0.484	0.211	0.256
	336	0.538	0.262	0.293	**0.532**	**0.252**	**0.289**	0.536	0.256	0.290	0.543	0.257	0.290
	720	0.631	0.333	0.341	0.623	**0.321**	**0.337**	**0.618**	0.326	0.338	0.627	0.322	**0.337**
Solar-Energy	96	0.155	0.186	0.255	0.152	0.183	0.247	**0.149**	**0.178**	**0.237**	0.150	0.181	**0.237**
	192	0.169	0.205	0.267	0.165	0.193	0.253	**0.162**	**0.192**	**0.249**	0.165	0.198	0.250
	336	0.182	0.216	0.276	0.180	**0.201**	0.261	**0.176**	0.202	**0.258**	0.179	0.215	0.264
	720	0.175	0.229	0.286	**0.169**	**0.209**	**0.268**	0.170	0.216	0.270	0.171	0.231	0.281

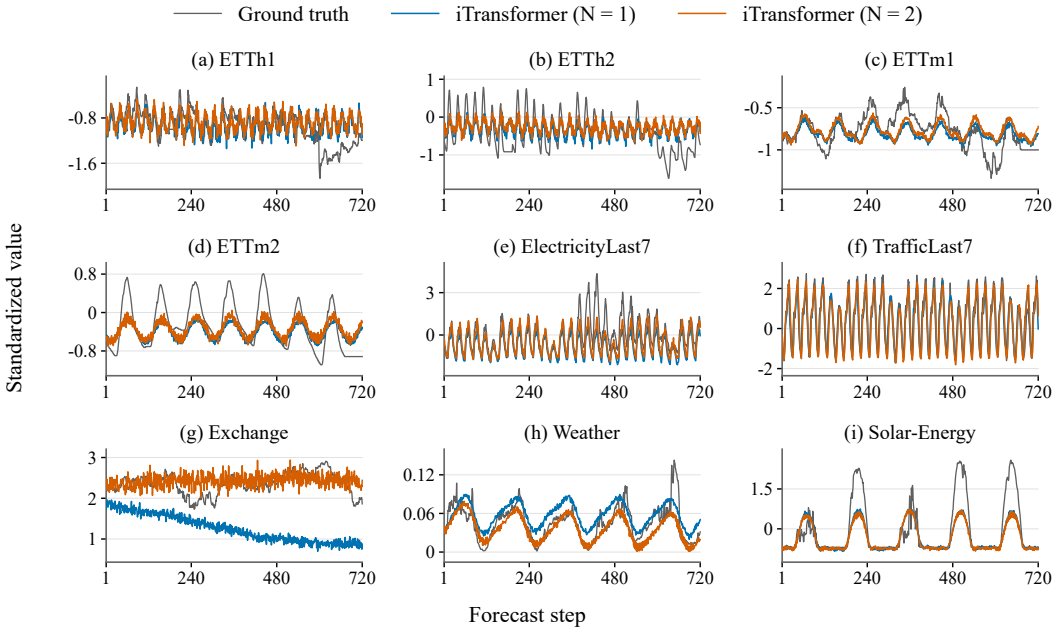

Figure 7: iTransformer forecasts at $N = 1$ and $N = 2$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 14: Combined screening validation and test metrics for PatchTST. $N = 8$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: PatchTST

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	0.854	0.398	0.427	0.821	0.383	0.412	0.742	0.386	0.411	**0.733**	**0.376**	**0.402**
	192	1.278	0.437	0.452	1.247	0.448	0.460	1.077	**0.419**	**0.436**	**1.033**	0.434	0.437
	336	1.556	0.539	0.497	1.487	0.476	0.478	1.380	0.462	0.466	**1.286**	**0.450**	**0.449**
	720	1.824	0.592	0.531	1.661	0.576	0.527	1.689	0.609	0.543	**1.557**	**0.485**	**0.487**
ETTh2	96	0.233	0.330	0.375	0.234	0.309	0.370	**0.213**	0.294	0.350	0.217	**0.287**	**0.345**
	192	0.306	0.392	0.412	0.312	0.408	0.428	0.292	0.377	0.402	**0.283**	**0.356**	**0.388**
	336	0.406	0.454	0.456	0.405	0.417	0.438	0.385	0.420	0.432	**0.373**	**0.385**	**0.413**
	720	0.737	0.460	0.467	0.712	0.507	0.494	0.669	0.440	0.461	**0.649**	**0.414**	**0.444**
ETTm1	96	0.412	0.316	0.366	0.409	0.307	0.358	**0.385**	**0.300**	0.356	**0.385**	0.301	**0.355**
	192	0.592	0.361	0.395	0.563	0.367	0.403	0.512	0.341	0.380	**0.501**	**0.335**	**0.374**
	336	0.726	0.394	0.414	0.715	0.385	0.410	0.675	0.379	0.403	**0.649**	**0.366**	**0.392**
	720	1.047	0.453	0.453	1.003	0.446	0.447	0.979	0.443	0.445	**0.963**	**0.419**	**0.424**
ETTm2	96	0.131	0.193	0.279	0.130	0.191	0.281	0.126	0.174	0.262	**0.124**	**0.170**	**0.259**
	192	0.173	0.236	0.307	0.171	0.236	0.311	0.174	0.236	0.310	**0.165**	**0.229**	**0.300**
	336	0.230	0.289	0.341	0.217	0.290	0.345	0.214	0.285	0.343	**0.208**	**0.281**	**0.336**
	720	0.300	0.407	0.410	0.290	0.384	0.403	0.290	0.376	0.399	**0.278**	**0.361**	**0.388**
ElectricityLast7	96	0.092	0.135	**0.241**	**0.089**	**0.132**	0.242	**0.089**	0.134	0.242	**0.089**	0.135	0.242
	192	0.112	0.163	0.273	0.102	0.151	0.260	**0.100**	**0.148**	0.257	**0.100**	**0.148**	**0.255**
	336	0.137	0.195	0.314	0.131	0.180	0.290	0.116	0.167	0.279	**0.110**	**0.161**	**0.271**
	720	0.235	0.271	0.368	0.172	0.243	0.364	0.142	0.217	0.340	**0.125**	**0.188**	**0.304**
TrafficLast7	96	**0.327**	**0.252**	**0.243**	0.331	0.259	0.256	0.334	0.254	0.248	0.341	0.262	0.254
	192	**0.328**	0.272	0.264	0.337	0.272	0.265	0.336	**0.266**	**0.258**	0.342	0.273	0.262
	336	**0.344**	0.292	0.285	0.351	0.288	0.277	0.351	**0.285**	0.275	0.351	0.286	**0.271**
	720	0.366	0.330	0.317	0.356	0.324	0.312	0.370	0.319	0.303	**0.330**	**0.306**	**0.283**
Exchange	96	0.175	0.099	0.225	0.149	0.107	0.233	**0.147**	**0.094**	**0.218**	0.153	0.095	0.219
	192	0.261	0.210	0.332	0.243	0.216	0.340	**0.241**	0.213	0.333	0.245	**0.197**	**0.320**
	336	**0.420**	0.389	0.462	0.458	0.429	0.488	0.429	**0.359**	**0.440**	0.435	0.361	**0.440**
	720	1.573	1.291	0.834	**1.368**	1.387	0.867	1.792	1.024	0.761	1.997	**0.919**	**0.719**
Weather	96	0.405	0.158	0.212	**0.397**	**0.150**	**0.202**	0.398	**0.150**	**0.202**	0.406	0.158	0.209
	192	0.470	0.209	0.258	0.468	0.200	0.249	**0.461**	**0.194**	**0.244**	0.470	0.202	0.248
	336	0.561	0.260	0.300	0.542	0.251	0.289	**0.532**	**0.247**	**0.284**	0.540	0.251	0.286
	720	0.651	0.336	0.347	0.628	0.323	0.338	**0.619**	**0.321**	**0.335**	0.627	**0.321**	**0.335**
Solar-Energy	96	**0.143**	**0.174**	**0.236**	0.155	0.178	0.247	0.160	0.182	0.249	0.167	0.195	0.257
	192	**0.156**	**0.192**	**0.253**	0.167	**0.192**	0.263	0.172	0.196	0.262	0.178	0.207	0.266
	336	**0.162**	**0.194**	**0.257**	0.174	0.197	0.267	0.177	0.200	0.264	0.186	0.215	0.271
	720	**0.157**	**0.205**	**0.258**	0.173	0.206	0.276	0.175	0.208	0.272	0.175	0.220	0.276

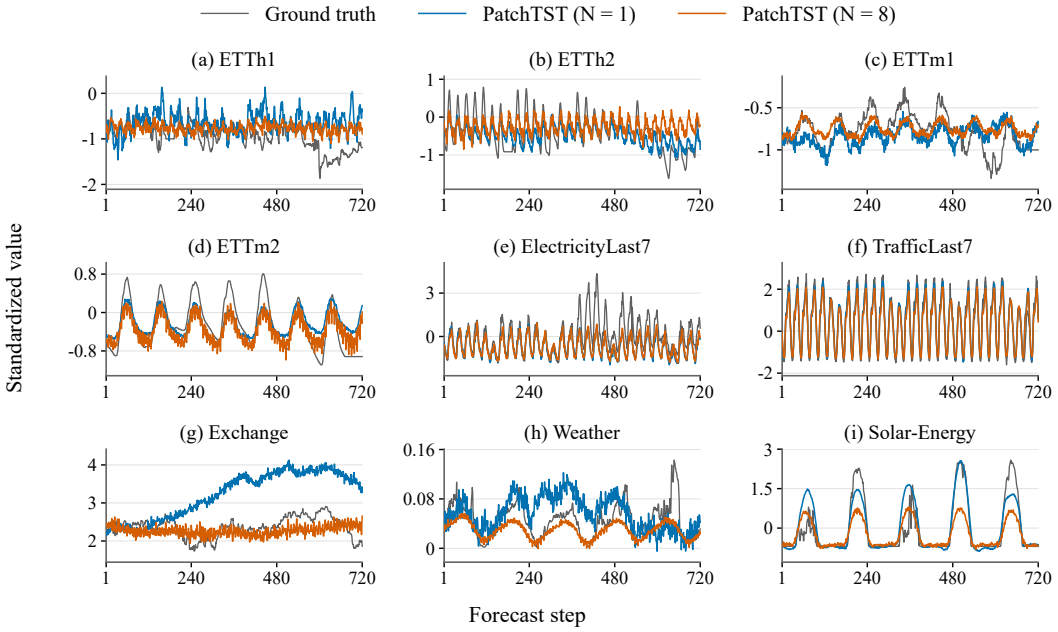

Figure 8: PatchTST forecasts at $N = 1$ and $N = 8$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 15: Combined screening validation and test metrics for Crossformer. $N = 2$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: Crossformer

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	0.702	0.403	0.440	**0.685**	0.379	0.415	0.693	**0.374**	**0.413**	0.715	0.379	0.418
	192	0.902	0.431	0.452	**0.886**	0.490	0.503	**0.886**	0.418	0.445	**0.886**	**0.407**	**0.438**
	336	1.128	1.168	0.869	1.055	0.450	0.471	1.049	0.500	0.510	**1.041**	**0.430**	**0.459**
	720	1.544	0.971	0.777	1.174	0.508	0.516	**1.155**	0.739	0.672	1.165	**0.505**	**0.515**
ETTh2	96	0.393	**0.662**	**0.578**	0.356	0.712	0.590	0.357	0.752	0.593	**0.348**	0.742	0.583
	192	0.683	1.933	1.063	0.479	1.112	0.753	**0.432**	**0.832**	**0.641**	0.445	0.844	**0.641**
	336	0.655	1.514	0.929	0.571	1.217	0.776	**0.501**	0.888	0.673	0.514	**0.761**	**0.619**
	720	0.838	2.641	1.403	0.715	1.607	0.930	**0.673**	1.104	0.773	0.689	**0.974**	**0.721**
ETTm1	96	0.424	0.319	0.369	**0.410**	**0.308**	**0.363**	0.413	0.310	0.364	0.431	0.311	0.368
	192	0.555	0.362	0.394	0.537	0.364	0.402	**0.522**	0.354	0.396	0.524	**0.344**	**0.389**
	336	0.712	0.413	0.434	0.665	0.403	0.427	**0.657**	**0.385**	**0.420**	0.658	0.388	**0.420**
	720	1.073	0.666	0.606	0.926	0.542	0.521	0.895	**0.475**	**0.476**	**0.893**	0.476	0.478
ETTm2	96	0.173	**0.274**	**0.352**	**0.160**	0.293	0.378	0.177	0.320	0.386	0.167	0.321	0.375
	192	0.247	**0.429**	0.478	0.276	0.534	0.515	0.266	0.490	0.489	**0.233**	0.470	**0.461**
	336	0.480	1.679	0.923	0.349	**0.650**	0.607	0.331	0.651	**0.554**	**0.325**	0.709	0.574
	720	0.528	1.063	0.771	0.541	1.217	0.859	0.557	1.321	0.835	**0.456**	**1.062**	**0.708**
ElectricityLast7	96	**0.083**	**0.125**	**0.227**	0.085	0.127	0.228	0.090	0.133	0.238	0.093	0.138	0.244
	192	**0.096**	0.144	0.250	0.098	**0.143**	**0.245**	0.101	0.148	0.253	0.107	0.156	0.263
	336	0.118	0.168	0.281	0.113	0.165	0.276	**0.112**	**0.163**	**0.273**	0.125	0.180	0.291
	720	**0.120**	0.214	0.327	0.127	0.191	**0.308**	0.131	**0.190**	0.310	0.159	0.233	0.349
TrafficLast7	96	0.335	0.288	0.280	**0.327**	**0.264**	**0.261**	0.348	0.292	0.292	0.354	0.299	0.302
	192	**0.327**	**0.277**	**0.271**	0.329	**0.277**	**0.271**	0.366	0.326	0.319	0.394	0.353	0.343
	336	0.355	0.335	0.306	**0.339**	**0.299**	**0.286**	0.389	0.361	0.343	0.424	0.392	0.369
	720	0.349	0.365	0.326	**0.327**	**0.346**	**0.323**	0.401	0.424	0.384	0.424	0.437	0.396
Exchange	96	**0.446**	0.305	0.418	0.739	0.320	0.409	0.560	**0.255**	**0.378**	0.663	0.313	0.410
	192	**0.733**	**0.419**	**0.500**	1.075	0.481	0.541	1.119	0.504	0.538	1.235	0.589	0.582
	336	4.143	2.032	1.097	**2.014**	**0.911**	0.782	2.155	1.003	0.780	2.036	0.944	**0.765**
	720	4.070	1.695	1.029	**2.767**	1.503	1.005	3.049	**1.444**	**0.975**	3.226	1.716	1.083
Weather	96	0.359	**0.144**	0.213	**0.353**	0.148	0.214	0.359	**0.144**	**0.211**	0.373	0.150	0.217
	192	0.422	0.204	0.279	0.428	0.192	0.265	**0.418**	**0.187**	**0.256**	0.433	0.192	0.261
	336	0.511	0.265	0.330	**0.493**	**0.240**	0.305	0.497	0.248	0.313	0.499	**0.240**	**0.303**
	720	0.705	0.457	0.441	0.677	0.401	0.416	0.603	0.322	0.374	**0.595**	**0.310**	**0.358**
Solar-Energy	96	0.143	**0.161**	**0.208**	**0.138**	0.166	0.214	0.141	0.166	0.232	0.146	0.190	0.260
	192	0.162	0.200	0.232	0.152	0.186	**0.224**	**0.148**	**0.180**	0.242	0.156	0.199	0.266
	336	0.174	0.211	0.252	0.161	0.195	**0.240**	**0.156**	**0.192**	0.250	0.165	0.211	0.272
	720	**0.155**	**0.195**	**0.226**	0.160	0.203	0.252	**0.155**	0.208	0.259	0.157	0.213	0.274

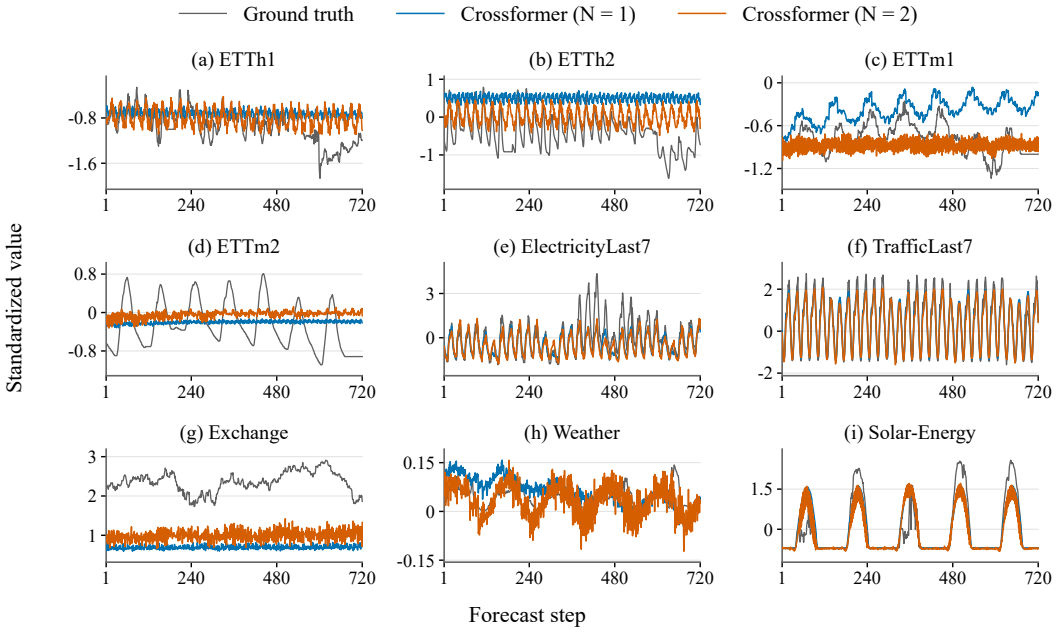

Figure 9: Crossformer forecasts at $N = 1$ and $N = 2$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 16: Combined screening validation and test metrics for TimesNet. $N = 8$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: TimesNet

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	1.093	0.477	0.479	0.931	0.449	0.460	**0.929**	**0.439**	**0.449**	0.949	0.444	0.451
	192	1.291	0.572	0.520	1.243	0.478	0.483	**1.189**	0.491	0.485	1.194	**0.460**	**0.462**
	336	1.631	0.627	0.551	1.578	0.574	0.532	1.510	**0.491**	**0.489**	**1.383**	0.501	0.492
	720	1.819	0.744	0.608	1.690	0.655	0.564	1.609	0.593	0.538	**1.590**	**0.552**	**0.514**
ETTh2	96	0.296	0.376	0.416	**0.288**	**0.375**	**0.415**	0.303	0.386	0.418	0.304	0.380	0.418
	192	0.598	0.423	0.434	**0.357**	0.421	0.436	0.391	0.404	0.432	0.393	**0.388**	**0.426**
	336	0.610	0.472	0.465	0.522	0.428	0.456	0.502	0.400	0.440	**0.477**	**0.379**	**0.426**
	720	0.830	0.606	0.536	0.779	0.457	0.475	0.735	0.445	0.469	**0.724**	**0.444**	**0.466**
ETTm1	96	0.584	0.374	0.403	0.468	**0.331**	**0.376**	**0.464**	0.337	0.381	**0.464**	0.341	0.383
	192	0.833	0.495	0.464	0.641	0.400	0.416	0.597	**0.375**	**0.403**	**0.577**	0.379	0.404
	336	0.859	0.547	0.485	0.777	0.491	0.458	0.726	**0.407**	**0.421**	**0.717**	0.408	**0.421**
	720	1.313	0.645	0.532	1.078	0.616	0.521	1.021	**0.462**	**0.456**	**1.014**	0.481	0.467
ETTm2	96	0.155	0.212	0.293	**0.142**	**0.196**	**0.283**	0.144	0.204	0.292	0.146	0.204	0.291
	192	0.245	0.291	0.341	0.206	0.265	**0.329**	**0.191**	0.272	0.337	0.193	**0.264**	0.332
	336	0.277	0.332	**0.370**	0.250	0.344	0.374	0.251	0.341	0.378	**0.245**	**0.331**	0.372
	720	0.353	0.534	0.477	0.325	0.462	0.436	0.321	**0.423**	**0.423**	**0.310**	0.428	0.426
ElectricityLast7	96	0.114	0.178	0.289	0.108	0.167	0.281	**0.103**	**0.156**	**0.267**	0.104	0.157	**0.267**
	192	0.141	0.251	0.354	0.118	0.183	0.293	**0.114**	0.175	0.285	**0.114**	**0.171**	**0.281**
	336	0.173	0.493	0.478	0.132	0.205	0.318	0.123	0.192	0.303	**0.120**	**0.181**	**0.292**
	720	0.137	0.893	0.603	0.147	0.336	0.411	0.138	0.236	0.342	**0.131**	**0.208**	**0.325**
TrafficLast7	96	0.373	0.307	0.309	**0.372**	**0.293**	**0.299**	0.381	0.302	0.310	0.394	0.318	0.318
	192	0.392	0.328	0.329	**0.384**	0.348	0.340	0.391	**0.323**	**0.325**	0.395	0.334	0.331
	336	0.428	0.417	0.382	**0.410**	0.388	0.368	0.411	**0.356**	**0.347**	0.420	0.359	**0.347**
	720	0.424	0.422	0.395	0.435	0.471	0.425	0.432	0.428	0.393	**0.420**	**0.402**	**0.370**
Exchange	96	0.655	0.290	0.397	0.399	**0.231**	**0.357**	0.399	0.283	0.394	**0.290**	0.251	0.375
	192	1.333	0.561	0.550	0.924	0.475	0.517	0.741	0.459	0.504	**0.460**	**0.405**	**0.480**
	336	1.759	1.415	0.854	1.067	0.825	0.677	1.143	0.907	0.714	**0.748**	**0.680**	**0.626**
	720	**1.615**	3.385	1.273	1.712	2.828	1.234	2.754	1.686	0.978	2.244	**1.427**	**0.915**
Weather	96	0.416	0.187	0.237	**0.396**	0.178	**0.228**	0.403	**0.171**	0.229	0.412	0.176	0.230
	192	0.527	0.282	0.305	0.486	0.261	0.296	**0.465**	**0.222**	0.272	0.469	**0.222**	**0.270**
	336	0.635	0.389	0.355	0.614	0.364	0.358	**0.535**	**0.276**	0.312	0.537	0.277	**0.310**
	720	0.854	0.422	0.405	0.724	0.405	0.393	0.638	0.363	0.363	**0.632**	**0.353**	**0.358**
Solar-Energy	96	0.169	0.209	**0.247**	0.153	0.193	0.257	**0.150**	**0.185**	0.252	0.152	0.195	0.264
	192	0.173	0.232	0.261	**0.159**	0.203	**0.253**	0.161	**0.197**	0.262	0.170	0.215	0.280
	336	0.234	0.246	**0.258**	0.178	0.214	0.285	**0.165**	**0.205**	0.265	0.181	0.227	0.290
	720	0.196	0.259	**0.266**	0.173	0.231	0.290	0.169	**0.230**	0.281	**0.167**	0.231	0.286

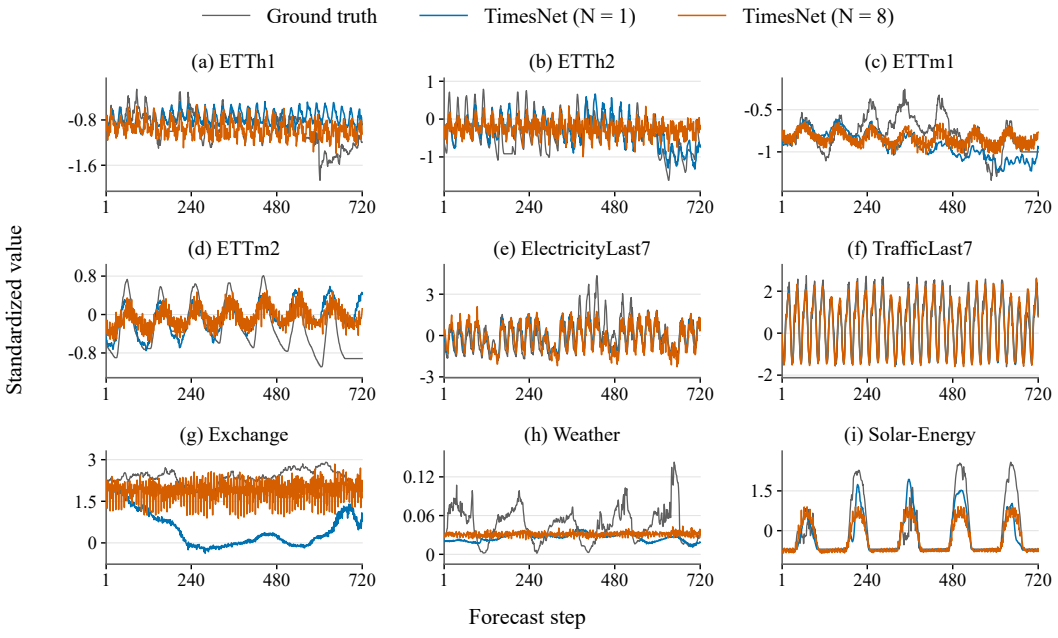

Figure 10: TimesNet forecasts at $N = 1$ and $N = 8$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 17: Combined screening validation and test metrics for DLinear. $N = 2$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: DLinear

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	**0.695**	0.387	0.407	0.697	**0.386**	**0.405**	0.709	0.387	**0.405**	0.726	0.388	**0.405**
	192	**0.938**	0.442	0.447	0.943	**0.422**	0.429	0.948	0.423	0.430	0.957	0.424	**0.428**
	336	**1.155**	0.460	0.452	1.163	0.456	0.450	1.162	**0.455**	0.450	1.158	0.456	**0.448**
	720	1.269	0.496	**0.492**	1.262	0.493	0.493	1.229	**0.484**	0.494	**1.173**	0.487	0.499
ETTh2	96	0.279	0.406	0.420	0.281	0.399	0.412	**0.275**	**0.374**	**0.405**	0.294	0.409	0.419
	192	0.392	0.556	0.494	**0.382**	**0.523**	**0.479**	0.395	0.576	0.502	0.399	0.580	0.503
	336	**0.478**	**0.555**	**0.506**	0.505	0.694	0.563	0.511	0.724	0.575	0.515	0.735	0.581
	720	0.795	1.102	0.734	0.811	1.134	0.748	**0.784**	1.145	0.737	**0.784**	**1.080**	**0.725**
ETTm1	96	**0.374**	**0.292**	**0.341**	0.376	0.295	**0.341**	0.382	0.298	0.344	0.393	0.305	0.349
	192	**0.483**	**0.336**	**0.365**	0.487	0.337	**0.365**	0.489	0.338	0.366	0.498	0.342	0.368
	336	**0.606**	**0.369**	**0.383**	0.612	0.370	0.385	0.615	0.371	0.385	0.621	0.374	0.387
	720	**0.836**	0.431	0.421	0.856	**0.426**	**0.416**	0.858	0.428	0.417	0.869	0.430	0.420
ETTm2	96	**0.124**	**0.190**	**0.278**	0.127	0.192	0.279	0.132	0.209	0.294	0.137	0.218	0.301
	192	0.195	**0.304**	**0.354**	**0.190**	0.325	0.364	0.194	0.337	0.373	0.199	0.343	0.378
	336	0.278	0.519	0.456	**0.261**	**0.445**	**0.429**	0.265	0.471	0.447	0.281	0.521	0.467
	720	0.380	0.714	0.552	**0.362**	**0.644**	**0.525**	0.380	0.714	0.545	0.396	0.747	0.569
ElectricityLast7	96	**0.086**	**0.130**	**0.229**	**0.086**	0.131	0.230	0.088	0.132	0.230	0.089	0.134	0.231
	192	**0.095**	**0.144**	**0.242**	**0.095**	0.145	0.243	0.096	0.145	0.243	0.097	0.146	0.243
	336	**0.103**	**0.157**	**0.258**	**0.103**	**0.157**	0.259	0.104	0.158	0.259	0.104	0.158	0.259
	720	**0.110**	**0.174**	**0.284**	0.111	**0.174**	0.286	0.114	0.177	0.293	0.124	0.187	0.303
TrafficLast7	96	0.389	**0.270**	0.257	**0.381**	0.271	**0.256**	0.384	0.276	0.257	0.383	0.281	0.257
	192	0.392	**0.282**	0.265	**0.383**	0.283	**0.264**	0.385	0.287	0.265	0.385	0.291	**0.264**
	336	0.407	**0.296**	**0.275**	**0.399**	0.299	0.278	**0.399**	0.303	0.278	**0.399**	0.305	0.276
	720	0.373	**0.326**	**0.305**	**0.357**	0.330	0.310	0.365	0.338	0.322	0.371	0.348	0.329
Exchange	96	**0.215**	**0.103**	**0.240**	0.227	0.105	0.241	0.242	0.108	0.247	0.300	0.127	0.276
	192	**0.470**	0.263	0.386	0.477	**0.246**	**0.385**	0.564	0.306	0.427	0.619	0.403	0.502
	336	1.333	0.599	0.601	**1.274**	0.579	0.575	1.314	**0.457**	**0.512**	1.524	0.490	0.543
	720	4.109	1.260	0.840	4.073	1.489	0.921	4.022	1.476	0.938	**2.805**	**0.681**	**0.679**
Weather	96	**0.365**	**0.142**	**0.203**	0.367	0.143	0.206	0.372	0.145	0.209	0.381	0.151	0.218
	192	**0.423**	**0.186**	**0.251**	0.426	0.187	0.252	0.429	0.189	0.253	0.436	0.195	0.261
	336	**0.491**	**0.238**	**0.295**	0.494	0.239	0.296	0.494	0.240	0.297	0.501	0.245	0.303
	720	0.587	0.314	0.358	**0.581**	**0.311**	**0.354**	0.584	0.315	0.357	0.593	0.323	0.367
Solar-Energy	96	**0.166**	**0.207**	**0.274**	0.167	0.208	0.277	0.169	0.211	0.280	0.174	0.218	0.288
	192	**0.179**	**0.227**	**0.291**	**0.179**	0.229	0.292	0.181	0.230	0.294	0.184	0.235	0.299
	336	**0.187**	**0.241**	**0.302**	**0.187**	0.242	0.303	0.188	0.243	0.303	0.190	0.246	0.308
	720	**0.173**	**0.250**	**0.309**	**0.173**	**0.250**	**0.309**	0.174	0.252	0.312	0.179	0.265	0.340

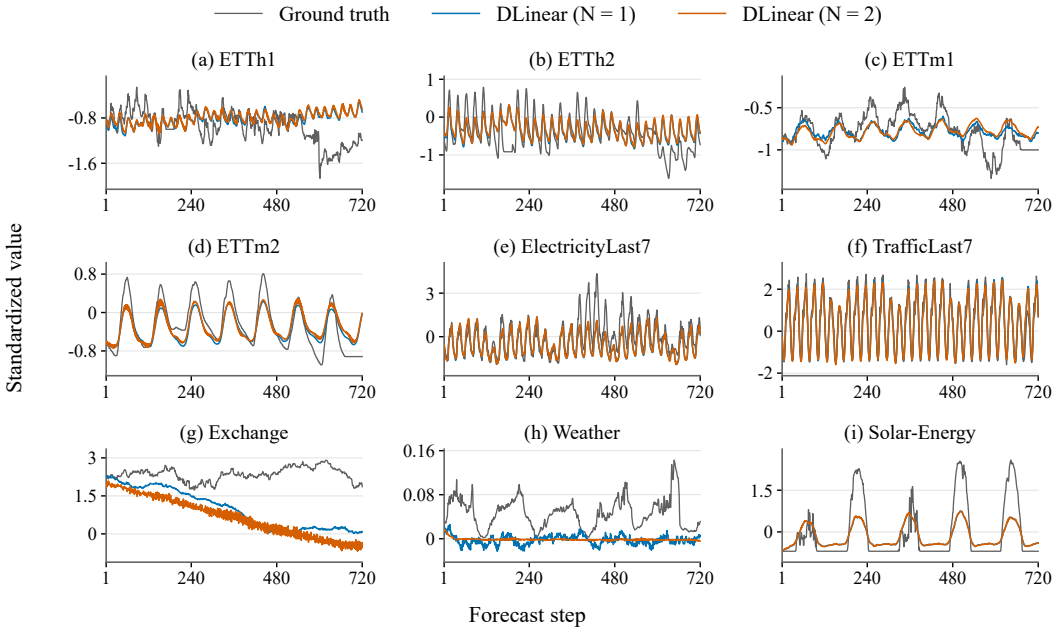

Figure 11: DLinear forecasts at $N = 1$ and $N = 2$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

Table 18: Combined screening validation and test metrics for Autoformer. $N = 4$ was selected from the validation values; test values were not used for selection. $\mathrm{MSE}_{\mathrm{val}}$ denotes the logged validation MSE. All values are rounded to three decimal places. Bold values are the lowest within each dataset–horizon metric across phase counts; ties after rounding are all marked.

Model: Autoformer

Dataset	H	N=1			N=2			N=4			N=8		
		$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE	$\mathrm{MSE}_{\mathrm{val}}$	MSE	MAE
ETTh1	96	1.061	0.533	0.523	1.084	0.540	0.505	1.045	0.452	0.468	**1.024**	**0.425**	**0.455**
	192	1.308	0.537	0.516	1.173	0.470	0.484	1.142	0.459	0.474	**1.141**	**0.450**	**0.472**
	336	1.417	0.626	0.585	1.416	0.661	0.604	1.290	0.452	**0.474**	**1.285**	**0.451**	0.475
	720	1.689	0.709	0.639	1.666	0.707	0.628	**1.519**	**0.505**	**0.516**	1.530	0.526	0.523
ETTh2	96	0.444	0.639	0.597	0.413	0.429	0.479	0.382	0.370	0.426	**0.378**	**0.368**	**0.424**
	192	0.601	0.527	0.529	0.549	0.593	0.567	**0.433**	**0.383**	**0.433**	0.438	0.394	0.441
	336	0.632	0.764	0.671	0.587	0.786	0.652	0.516	**0.390**	**0.442**	**0.509**	0.398	0.447
	720	1.030	2.050	1.035	0.884	1.236	0.829	**0.723**	**0.465**	0.497	0.730	0.468	**0.494**
ETTm1	96	0.811	0.555	0.518	0.820	**0.505**	**0.489**	**0.804**	0.532	**0.489**	0.816	0.565	0.505
	192	**0.887**	0.547	0.517	0.904	**0.496**	**0.486**	0.888	0.523	0.504	**0.887**	0.563	0.511
	336	1.037	0.644	0.540	0.983	0.536	0.503	0.963	**0.486**	**0.479**	**0.943**	0.524	0.502
	720	1.233	0.547	0.527	1.191	0.525	0.499	1.184	0.528	0.508	**1.176**	**0.504**	**0.494**
ETTm2	96	0.211	0.277	0.364	0.217	0.297	0.371	0.215	0.286	0.359	**0.208**	**0.275**	**0.354**
	192	0.250	0.326	0.388	0.239	0.317	0.379	**0.236**	**0.310**	**0.369**	0.239	0.311	0.372
	336	0.330	0.411	0.438	0.276	0.357	0.402	**0.265**	0.345	**0.389**	**0.265**	**0.343**	0.390
	720	0.454	0.462	0.475	0.382	0.450	0.457	0.341	**0.417**	**0.438**	**0.337**	0.421	0.440
ElectricityLast7	96	0.225	0.349	0.425	**0.167**	0.291	0.379	0.188	0.310	0.398	0.203	**0.278**	**0.372**
	192	0.304	0.501	0.511	0.307	0.503	0.504	**0.213**	0.337	0.412	0.248	**0.325**	**0.407**
	336	0.444	0.755	0.592	**0.240**	0.415	0.475	0.253	0.341	0.419	0.259	**0.334**	**0.414**
	720	0.225	1.003	0.661	**0.205**	0.440	0.490	0.260	0.373	0.448	0.244	**0.328**	**0.417**
TrafficLast7	96	0.422	0.384	0.371	0.376	0.351	0.353	**0.374**	**0.334**	**0.339**	0.406	0.378	0.374
	192	0.401	0.363	0.356	**0.372**	0.370	0.358	0.389	**0.331**	**0.331**	0.400	0.368	0.363
	336	0.439	0.388	0.371	0.378	**0.348**	**0.342**	**0.376**	0.350	**0.342**	0.387	0.370	0.353
	720	0.497	0.459	0.420	0.371	0.393	0.376	**0.370**	0.375	0.362	0.375	**0.365**	**0.346**
Exchange	96	0.443	0.802	0.710	0.413	0.540	0.566	0.416	0.513	0.542	**0.396**	**0.503**	**0.537**
	192	0.874	0.915	0.767	0.614	**0.616**	**0.609**	0.571	0.644	0.616	**0.537**	0.658	0.622
	336	0.950	8.692	2.633	**0.665**	2.184	1.274	0.798	**0.917**	**0.756**	0.772	0.928	0.764
	720	**1.184**	**1.441**	**0.934**	1.405	1.549	0.960	1.366	1.581	0.978	1.321	1.622	0.994
Weather	96	0.569	0.376	0.444	0.593	0.335	0.394	0.587	0.335	0.391	**0.567**	**0.310**	**0.379**
	192	0.691	0.491	0.504	0.653	0.379	0.427	0.648	0.361	0.408	**0.646**	**0.347**	**0.398**
	336	0.715	0.701	0.602	0.722	0.413	0.443	**0.676**	0.397	0.428	0.688	**0.390**	**0.425**
	720	0.815	0.685	0.602	0.752	0.430	**0.437**	0.727	0.438	0.452	**0.710**	**0.420**	0.441
Solar-Energy	96	0.559	0.631	**0.528**	0.581	0.678	0.536	**0.530**	**0.558**	0.558	0.766	0.723	0.658
	192	0.811	0.816	**0.611**	0.712	0.736	0.616	**0.647**	**0.684**	0.617	0.933	0.738	0.676
	336	0.906	0.823	**0.638**	0.859	0.847	0.661	**0.786**	0.796	0.680	0.981	**0.759**	0.675
	720	0.670	0.763	0.579	0.632	0.684	0.588	**0.568**	**0.620**	**0.570**	0.907	0.829	0.675

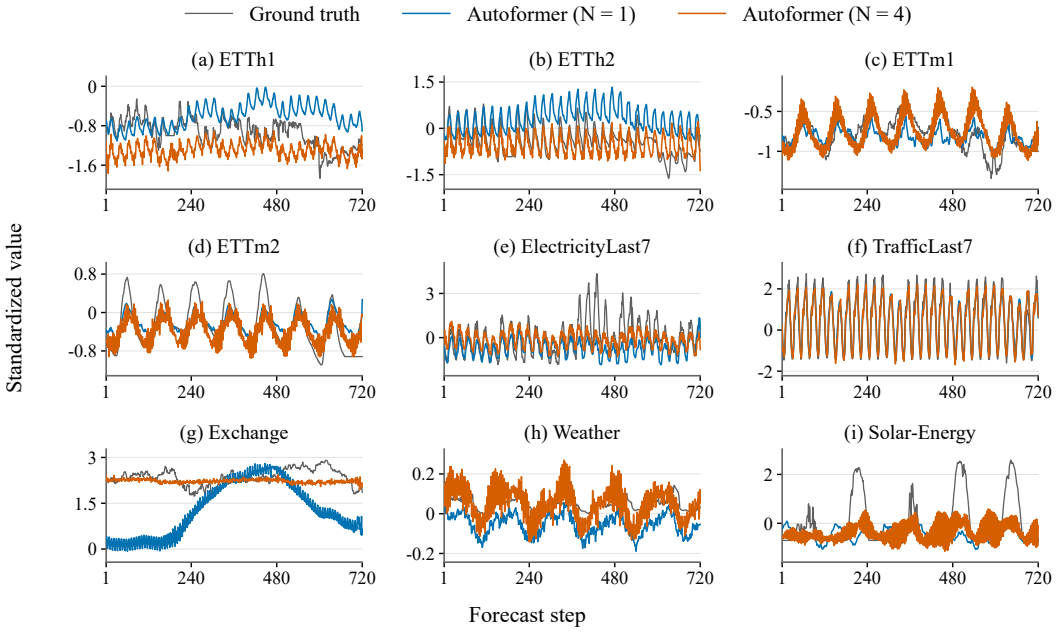

Figure 12: Autoformer forecasts at $N = 1$ and $N = 4$ over 720 steps. Each panel shows the first test window and last output channel of one dataset, with ground truth. Values use training-set standardization; Electricity and Traffic use Last7.

C Additional sensitivity results

C.1 Channel-count sensitivity

C.1.1 Experimental settings

We compare WPMixer ($N_b = 4$) and PatchTST ($N_b = 8$) with their respective $N = 1$ controls, keeping the phase counts selected by the validation rule in Appendix B.4 fixed as the channel count varies. Starting with the seven-channel Last7 subsets, we expand the channel count to $C \in \{7, 16, 32, 64, 128, 256, 321\}$ on Electricity and $C \in \{7, 16, 32, 64, 128, 256, 512, 862\}$ on Traffic. Both configurations in each comparison use the same variables, data split, optimization settings, input length ($L = 512$), and forecast horizon.

WPMixer uses $d_{\text{model}} = 32$, $d_{\text{ff}} = 64$, one encoder layer, eight heads, a learning rate of 10^{-3}, and a batch size of 8. PatchTST uses $d_{\text{model}} = 512$, $d_{\text{ff}} = 2048$, two encoder layers, eight heads, and a learning rate of 10^{-4}. Its patch length is 16 and its stride is 8, with batch sizes of 8 for Electricity and 4 for Traffic. Both models train for at most 10 epochs with an early-stopping patience of 3.

C.1.2 Evaluation metrics

Let $E^m_{d,b,N,C,H}$ denote the test error of backbone b on dataset d with phase count N, channel count C, and forecast horizon H, where $m \in \{\text{MSE}, \text{MAE}\}$. We define $g_m(d, b, C, H)$ as the relative reduction against the $N = 1$ control and $\Delta_m(d, b, C)$ as its arithmetic mean over the four forecast horizons:

$$g_m(d, b, C, H) = 100 \frac{E^m_{d,b,1,C,H} - E^m_{d,b,N_b,C,H}}{E^m_{d,b,1,C,H}},$$
$$\Delta_m(d, b, C) = \frac{1}{4} \sum_{H \in \{96,192,336,720\}} g_m(d, b, C, H). \tag{12}$$

Positive values indicate lower errors with CrossForest. Both configurations in each comparison use the same channel set. MSE wins and losses count forecast horizons with $g_{\text{MSE}} > 0$ and $g_{\text{MSE}} < 0$, respectively; there are no ties. Table 19 reports each configuration's MSE and MAE at every forecast horizon and their arithmetic means over the four horizons. Figure 2 shows the corresponding relative MSE reductions. We average the relative reductions and absolute errors separately across horizons.

Table 19: Test errors and relative reductions by channel count and forecast horizon. Each block compares CrossForest (CF), using the selected phase count N_b, with the same backbone's $N = 1$ control. For dataset d, backbone b, channel count C, and forecast horizon H, the $\text{MSE}_{N=1}$ and $\text{MAE}_{N=1}$ columns report $E^m_{d,b,1,C,H}$, and the MSE_{CF} and MAE_{CF} columns report $E^m_{d,b,N_b,C,H}$, with m denoting MSE or MAE as appropriate. The MSE and MAE columns under "Relative reduction" report percentage reductions: $g_m(d, b, C, H)$ from Equation (12) for each forecast horizon and $\Delta_m(d, b, C)$ for Avg. rows. In the error columns, Avg. is the arithmetic mean over the four horizons. All averages are computed before rounding. Bold marks the lower error for each metric in each matched comparison. Positive reductions indicate lower CrossForest error. Blue and red denote positive and negative reductions, respectively. Color intensity increases with the absolute reduction.

			Errors				Relative reduction (%)	
Dataset	C	H	$\text{MSE}_{N=1}$	$\text{MAE}_{N=1}$	MSE_{CF}	MAE_{CF}	MSE	MAE
Electricity / WPMixer ($N = 4$)								
		96	0.1342	**0.2310**	**0.1341**	0.2365	0.1276	-2.3880
		192	0.1476	0.2481	**0.1464**	**0.2480**	0.7935	0.0333
Electricity	7	336	0.1631	0.2665	**0.1585**	**0.2635**	2.7834	1.1426
		720	0.2186	0.3250	**0.1818**	**0.2929**	16.8552	9.8789
		Avg.	0.1659	0.2677	**0.1552**	**0.2602**	5.1399	2.1667

Continued on next page

Table 19 (continued)

Dataset	C	H	Errors				Relative reduction (%)	
			$\mathrm{MSE}_{N=1}$	$\mathrm{MAE}_{N=1}$	$\mathrm{MSE}_{\mathrm{CF}}$	$\mathrm{MAE}_{\mathrm{CF}}$	MSE	MAE
		96	**0.1034**	**0.2056**	0.1072	0.2132	-3.6177	-3.7100
		192	**0.1168**	**0.2203**	0.1176	0.2240	-0.7272	-1.6738
Electricity	16	336	0.1307	**0.2377**	**0.1284**	0.2388	1.8087	-0.4476
		720	0.1609	0.2772	**0.1560**	**0.2709**	3.0535	2.2623
		Avg.	0.1279	**0.2352**	**0.1273**	0.2367	0.1293	-0.8923
		96	**0.1504**	**0.2400**	0.1567	0.2469	-4.1970	-2.8712
		192	**0.1646**	**0.2531**	0.1690	0.2587	-2.6870	-2.2057
Electricity	32	336	**0.1811**	**0.2715**	0.1826	0.2744	-0.8245	-1.0484
		720	0.2395	0.3218	**0.2183**	**0.3080**	8.8486	4.2807
		Avg.	0.1839	**0.2716**	**0.1817**	0.2720	0.2850	-0.4611
		96	**0.1391**	**0.2269**	0.1425	0.2340	-2.4852	-3.1398
		192	**0.1554**	**0.2422**	0.1575	0.2472	-1.3678	-2.0689
Electricity	64	336	**0.1730**	**0.2612**	0.1764	0.2646	-1.9267	-1.2865
		720	**0.2219**	**0.2999**	0.2232	0.3004	-0.5991	-0.1797
		Avg.	**0.1723**	**0.2575**	0.1749	0.2615	-1.5947	-1.6687
		96	**0.1420**	**0.2336**	0.1470	0.2409	-3.5768	-3.1241
		192	**0.1611**	**0.2502**	0.1634	0.2547	-1.4083	-1.7936
Electricity	128	336	0.1813	**0.2701**	0.1813	0.2718	0.0041	-0.6130
		720	**0.2292**	0.3087	0.2294	**0.3086**	-0.0885	0.0192
		Avg.	**0.1784**	**0.2657**	0.1803	0.2690	-1.2674	-1.3779
		96	**0.1329**	**0.2265**	0.1380	0.2343	-3.8408	-3.4269
		192	**0.1511**	**0.2425**	0.1535	0.2476	-1.5548	-2.1054
Electricity	256	336	0.1697	**0.2607**	**0.1695**	0.2637	0.1240	-1.1576
		720	**0.2084**	**0.2957**	0.2093	0.2971	-0.4594	-0.4548
		Avg.	**0.1655**	**0.2564**	0.1676	0.2607	-1.4328	-1.7862
		96	**0.1312**	**0.2246**	0.1360	0.2323	-3.6630	-3.4062
		192	**0.1488**	**0.2407**	0.1508	0.2453	-1.3437	-1.9282
Electricity	321	336	0.1669	**0.2594**	**0.1663**	0.2613	0.3413	-0.7461
		720	**0.2035**	**0.2933**	0.2053	0.2945	-0.8730	-0.4261
		Avg.	**0.1626**	**0.2545**	0.1646	0.2584	-1.3846	-1.6267
Electricity / PatchTST ($N = 8$)								
		96	**0.1347**	**0.2407**	0.1353	0.2415	-0.4377	-0.3711
		192	0.1632	0.2731	**0.1484**	**0.2552**	9.0318	6.5619
Electricity	7	336	0.1952	0.3141	**0.1614**	**0.2709**	17.3467	13.7554
		720	0.2710	0.3678	**0.1879**	**0.3036**	30.6684	17.4566
		Avg.	0.1910	0.2989	**0.1583**	**0.2678**	14.1523	9.3507
		96	**0.1041**	**0.2102**	0.1103	0.2211	-5.9614	-5.1980
		192	**0.1215**	**0.2304**	0.1226	0.2343	-0.8876	-1.6864
Electricity	16	336	0.1658	0.2930	**0.1335**	**0.2488**	19.5212	15.0848
		720	0.1852	0.3020	**0.1630**	**0.2821**	11.9777	6.6147
		Avg.	0.1441	0.2589	**0.1323**	**0.2466**	6.1625	3.7038
		96	**0.1531**	**0.2423**	0.1611	0.2546	-5.2476	-5.0776
		192	**0.1695**	**0.2592**	0.1739	0.2676	-2.5958	-3.2401
Electricity	32	336	0.1921	0.2873	**0.1871**	**0.2828**	2.6033	1.5825
		720	0.2463	0.3473	**0.2231**	**0.3158**	9.3978	9.0600
		Avg.	0.1902	0.2840	**0.1863**	**0.2802**	1.0394	0.5812
		96	**0.1369**	**0.2279**	0.1467	0.2426	-7.1597	-6.4272
		192	**0.1555**	**0.2463**	0.1617	0.2561	-3.9794	-4.0081
Electricity	64	336	**0.1758**	**0.2651**	0.1797	0.2725	-2.2435	-2.8116
		720	0.2289	0.3090	**0.2264**	**0.3075**	1.1062	0.4832
		Avg.	**0.1743**	**0.2621**	0.1786	0.2697	-3.0691	-3.1909
		96	**0.1417**	**0.2345**	0.1512	0.2492	-6.7348	-6.2736
		192	**0.1611**	**0.2535**	0.1677	0.2635	-4.0848	-3.9445
Electricity	128	336	**0.1840**	**0.2746**	0.1857	0.2801	-0.8965	-2.0021
		720	**0.2241**	**0.3059**	0.2328	0.3154	-3.9098	-3.1149
		Avg.	**0.1777**	**0.2671**	0.1844	0.2771	-3.9065	-3.8338
		96	**0.1351**	**0.2346**	0.1426	0.2441	-5.5060	-4.0367
		192	**0.1512**	**0.2452**	0.1582	0.2581	-4.6246	-5.2593
Electricity	256	336	**0.1710**	**0.2659**	0.1747	0.2739	-2.2005	-3.0393
		720	**0.2074**	**0.2956**	0.2143	0.3059	-3.3268	-3.5004
		Avg.	**0.1662**	**0.2603**	0.1725	0.2705	-3.9145	-3.9589
		96	**0.1319**	**0.2299**	0.1406	0.2421	-6.5919	-5.3093
		192	**0.1486**	**0.2439**	0.1555	0.2557	-4.6264	-4.8095
Electricity	321	336	**0.1648**	**0.2618**	0.1715	0.2714	-4.0337	-3.6575
		720	**0.2010**	**0.2919**	0.2104	0.3033	-4.6694	-3.9123
		Avg.	**0.1616**	**0.2569**	0.1695	0.2681	-4.9803	-4.4222

Continued on next page

Table 19 (continued)

Dataset	C	H	Errors $\mathrm{MSE}_{N=1}$	Errors $\mathrm{MAE}_{N=1}$	Errors $\mathrm{MSE}_{\mathrm{CF}}$	Errors $\mathrm{MAE}_{\mathrm{CF}}$	Relative reduction (%) MSE	Relative reduction (%) MAE
Traffic / WPMixer ($N = 4$)								
		96	**0.2616**	0.2544	0.2686	**0.2534**	-2.6849	0.4134
		192	**0.2768**	0.2661	0.2806	**0.2619**	-1.3783	1.5964
Traffic	7	336	**0.2945**	0.2804	0.2960	**0.2712**	-0.5232	3.2776
		720	0.3286	0.3097	**0.3179**	**0.2856**	3.2534	7.7786
		Avg.	**0.2904**	0.2777	0.2908	**0.2680**	-0.3333	3.2665
		96	**0.3231**	**0.2474**	0.3416	0.2589	-5.7110	-4.6697
		192	**0.3384**	**0.2570**	0.3533	0.2644	-4.4084	-2.8970
Traffic	16	336	**0.3453**	**0.2625**	0.3635	0.2715	-5.2782	-3.4171
		720	**0.3821**	**0.2836**	0.3974	0.2875	-4.0120	-1.3574
		Avg.	**0.3472**	**0.2626**	0.3639	0.2706	-4.8524	-3.0853
		96	**0.3235**	**0.2695**	0.3458	0.2857	-6.8880	-6.0186
		192	**0.3354**	**0.2751**	0.3550	0.2900	-5.8247	-5.4231
Traffic	32	336	**0.3380**	**0.2770**	0.3609	0.2944	-6.7892	-6.2772
		720	**0.3667**	**0.2978**	0.3876	0.3091	-5.7175	-3.7919
		Avg.	**0.3409**	**0.2799**	0.3623	0.2948	-6.3048	-5.3777
		96	**0.3540**	**0.2728**	0.3846	0.2931	-8.6241	-7.4296
		192	**0.3760**	**0.2811**	0.4011	0.2995	-6.6764	-6.5423
Traffic	64	336	**0.3912**	**0.2876**	0.4203	0.3082	-7.4422	-7.1457
		720	**0.4440**	**0.3164**	0.4737	0.3338	-6.6973	-5.5157
		Avg.	**0.3913**	**0.2895**	0.4199	0.3087	-7.3600	-6.6583
		96	**0.6989**	**0.2992**	0.7183	0.3112	-2.7726	-4.0057
		192	**0.7296**	**0.3169**	0.7353	0.3175	-0.7760	-0.1690
Traffic	128	336	**0.7321**	**0.3104**	0.7508	0.3254	-2.5480	-4.8512
		720	**0.7979**	**0.3320**	0.8054	0.3464	-0.9399	-4.3399
		Avg.	**0.7396**	**0.3146**	0.7524	0.3251	-1.7591	-3.3414
		96	**0.5103**	**0.2790**	0.5296	0.2944	-3.7811	-5.4915
		192	**0.5278**	**0.2875**	0.5439	0.3000	-3.0635	-4.3611
Traffic	256	336	**0.5398**	**0.2923**	0.5562	0.3067	-3.0381	-4.9346
		720	**0.5910**	**0.3138**	0.6006	0.3267	-1.6365	-4.1109
		Avg.	**0.5422**	**0.2932**	0.5576	0.3070	-2.8798	-4.7245
		96	**0.4186**	**0.2665**	0.4440	0.2863	-6.0668	-7.4130
		192	**0.4366**	**0.2738**	0.4569	0.2916	-4.6622	-6.4899
Traffic	512	336	**0.4508**	**0.2828**	0.4675	0.2975	-3.7011	-5.1979
		720	**0.4919**	**0.3016**	0.5065	0.3169	-2.9818	-5.0545
		Avg.	**0.4495**	**0.2812**	0.4688	0.2980	-4.3530	-6.0388
		96	**0.3761**	**0.2626**	0.4017	0.2827	-6.7845	-7.6666
		192	**0.3931**	**0.2703**	0.4132	0.2877	-5.1286	-6.4336
Traffic	862	336	**0.4025**	**0.2761**	0.4231	0.2936	-5.1134	-6.3295
		720	**0.4425**	**0.2981**	0.4607	0.3136	-4.1114	-5.2177
		Avg.	**0.4036**	**0.2768**	0.4247	0.2944	-5.2845	-6.4118
Traffic / PatchTST ($N = 8$)								
		96	**0.2516**	**0.2432**	0.2623	0.2537	-4.2686	-4.3002
		192	**0.2716**	0.2638	0.2732	**0.2617**	-0.5985	0.7631
Traffic	7	336	0.2919	0.2853	**0.2864**	**0.2705**	1.8877	5.1687
		720	0.3304	0.3173	**0.3065**	**0.2828**	7.2287	10.8701
		Avg.	0.2864	0.2774	**0.2821**	**0.2672**	1.0623	3.1254
		96	**0.3171**	**0.2371**	0.3286	0.2528	-3.6153	-6.6216
		192	**0.3353**	**0.2527**	0.3428	0.2604	-2.2300	-3.0587
Traffic	16	336	**0.3422**	**0.2590**	0.3520	0.2654	-2.8362	-2.4866
		720	**0.3869**	0.2879	0.3874	**0.2806**	-0.1228	2.5307
		Avg.	**0.3454**	**0.2592**	0.3527	0.2648	-2.2011	-2.4091
		96	**0.3189**	**0.2677**	0.3317	0.2777	-4.0227	-3.7613
		192	**0.3339**	**0.2782**	0.3430	0.2837	-2.7209	-1.9880
Traffic	32	336	**0.3360**	**0.2764**	0.3482	0.2868	-3.6566	-3.7473
		720	**0.3715**	0.3016	0.3762	**0.3010**	-1.2710	0.2030
		Avg.	**0.3401**	**0.2810**	0.3498	0.2873	-2.9178	-2.3234
		96	**0.3473**	**0.2664**	0.3661	0.2832	-5.4285	-6.2927
		192	**0.3720**	**0.2743**	0.3869	0.2920	-3.9800	-6.4625
Traffic	64	336	**0.3892**	**0.2875**	0.4057	0.2988	-4.2266	-3.9403
		720	**0.4489**	**0.3160**	0.4577	0.3215	-1.9597	-1.7457
		Avg.	**0.3894**	**0.2861**	0.4041	0.2989	-3.8987	-4.6103

Continued on next page

Table 19 (continued)

Dataset	C	H	Errors				Relative reduction (%)	
			$\mathrm{MSE}_{N=1}$	$\mathrm{MAE}_{N=1}$	$\mathrm{MSE}_{\mathrm{CF}}$	$\mathrm{MAE}_{\mathrm{CF}}$	MSE	MAE
Traffic	128	96	**0.6727**	**0.2903**	0.6842	0.2988	-1.7010	-2.9503
		192	**0.6864**	**0.2963**	0.7045	0.3073	-2.6468	-3.7058
		336	**0.7087**	**0.3107**	0.7184	0.3137	-1.3712	-0.9599
		720	**0.7658**	**0.3324**	0.7741	0.3343	-1.0864	-0.5861
		Avg.	**0.7084**	**0.3074**	0.7203	0.3135	-1.7013	-2.0505
Traffic	256	96	**0.4854**	**0.2675**	0.5053	0.2826	-4.0912	-5.6409
		192	**0.5054**	**0.2743**	0.5227	0.2905	-3.4203	-5.9057
		336	**0.5199**	**0.2842**	0.5346	0.2961	-2.8232	-4.2005
		720	**0.5616**	**0.3033**	0.5792	0.3158	-3.1335	-4.1305
		Avg.	**0.5181**	**0.2823**	0.5354	0.2963	-3.3670	-4.9694
Traffic	512	96	**0.4016**	**0.2569**	0.4232	0.2750	-5.3939	-7.0331
		192	**0.4218**	**0.2660**	0.4394	0.2824	-4.1701	-6.1938
		336	**0.4330**	**0.2766**	0.4504	0.2875	-4.0206	-3.9520
		720	**0.4751**	**0.2951**	0.4909	0.3066	-3.3252	-3.9108
		Avg.	**0.4329**	**0.2736**	0.4510	0.2879	-4.2275	-5.2724
Traffic	862	96	**0.3628**	**0.2548**	0.3832	0.2719	-5.6203	-6.7217
		192	**0.3842**	**0.2632**	0.3978	0.2791	-3.5410	-6.0632
		336	**0.3903**	**0.2691**	0.4082	0.2842	-4.5994	-5.6139
		720	**0.4313**	**0.2892**	0.4464	0.3039	-3.5071	-5.0857
		Avg.	**0.3921**	**0.2691**	0.4089	0.2848	-4.3169	-5.8711

C.1.3 AGGREGATE RESULTS

Across the 120 matched comparisons, the mean MSE reduction is -1.64% and the median is -2.91%; CrossForest improves MSE in 21 comparisons. For MAE, the mean reduction is -2.14% and the median is -3.41%, with improvements in 23 comparisons. On Electricity, WPMixer achieves a mean MSE reduction of -0.02%, improving in 10 of 28 comparisons. PatchTST has a mean MSE reduction of $+0.78\%$ and a median of -2.96%, improving in 8 of 28 comparisons. On Traffic, WPMixer and PatchTST have mean MSE reductions of -4.14% and -2.70%, respectively, with improvements in 1 of 32 and 2 of 32 comparisons.

The mean MSE reductions across the 30 comparisons at each forecast horizon are -4.54%, -2.51%, -0.79%, and $+1.27\%$ for $H = 96, 192, 336, 720$, respectively. At $H = 720$, only 9 of 30 comparisons show improvement, and the median reduction is -1.01%. A few large gains account for the positive mean.

C.2 PHASE-COUNT SENSITIVITY

We assess sensitivity to the number of phase learners N using the WPMixer and PatchTST results in Tables 10 and 14, respectively. We compare $N \in \{1, 2, 4, 8\}$ with fixed dataset splits, forecast horizons, input length ($L = 512$), and optimization hyperparameters. The nine datasets and four horizons give 36 matched tasks per backbone and phase count. Electricity and Traffic use Last7 throughout.

Following Equation 9, we average the relative error reductions over the 36 matched tasks for backbone b, phase count N, metric $m \in \{\mathrm{MSE}, \mathrm{MAE}\}$, and data split $\sigma \in \{\mathrm{val}, \mathrm{test}\}$:

$$G_m^{\sigma}(b, N) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} 100 \frac{E_{b,1,g}^{m,\sigma} - E_{b,N,g}^{m,\sigma}}{E_{b,1,g}^{m,\sigma}}, \qquad |\mathcal{G}| = 36, \tag{13}$$

where $E_{b,N,g}^{m,\sigma}$ is the error for task g, measured by metric m on split σ; validation error is the minimum logged validation MSE. Tasks are weighted equally, and positive values indicate lower error than the matched $N = 1$ control.

Table 20 summarizes the relative reductions in validation and test errors at each phase count for each backbone, together with the numbers of test tasks showing improvement or deterioration.

Table 20: Sensitivity to the number of phase learners across nine datasets and four forecast horizons. For each backbone and phase count, the G_m^σ columns report $G_m^\sigma(b, N)$, the macro-averaged relative reduction from the matched $N = 1$ control on the validation or test split. All values are percentages, and positive values indicate lower error. Blue and red shading indicate positive and negative reductions, respectively; darker shading denotes a larger absolute reduction. Zero reductions are unshaded. W/L gives the number of test tasks with lower or higher error than the $N = 1$ control. Results use $L = 512$ and the fixed Last7 subsets for Electricity and Traffic.

Backbone	N	$G_{\mathrm{MSE}}^{\mathrm{val}}$	$G_{\mathrm{MSE}}^{\mathrm{test}}$	$G_{\mathrm{MAE}}^{\mathrm{test}}$	MSE W/L	MAE W/L
WPMixer	1	0.00%	0.00%	0.00%	–	–
WPMixer	2	+2.83%	+3.56%	+2.42%	28/8	30/6
WPMixer	4	+3.33%	+4.79%	+3.04%	27/9	29/7
WPMixer	8	+0.14%	+1.37%	+0.44%	20/16	19/17
PatchTST	1	0.00%	0.00%	0.00%	–	–
PatchTST	2	+2.38%	+1.09%	−0.14%	22/14	18/18
PatchTST	4	+4.50%	+4.84%	+2.69%	29/7	26/10
PatchTST	8	+5.17%	+6.25%	+4.29%	29/7	30/6

D Ablation details

D.1 Phase-parameter sharing

The independent variant uses a separate learner for each phase; the shared variant uses one learner across all phases. Both use identical globally aligned phase inputs, input and output lengths for each phase, and interleaving rules. We evaluate PatchTST at $N = 8$ and WPMixer at $N = 4$, with $L = 512$ and $H \in \{96, 192, 336, 720\}$. The shared variant retains the same duplicate-free task construction and batches containing tasks from a single phase. It uses no communication module or prediction filter. Table 21 reports all 72 paired comparisons, and Table 3 reports the mean errors over the four horizons for each dataset and backbone.

Dataset-level averages show different responses to parameter sharing. Both backbones perform better with independent learners on ETT, Exchange, and Weather. Relative to the shared variant's mean MSE, independent learners reduce the error on ETT by 5.01% for PatchTST and 1.36% for WPMixer, and on Exchange by 9.65% and 10.24%, respectively. On Traffic and Solar-Energy, shared learners produce lower mean MSE and MAE for both backbones. Electricity shows a backbone-specific pattern: PatchTST slightly favors independent learners, whereas WPMixer performs better with shared parameters.

Table 21: Shared and independent phase learners evaluated on nine datasets. The table reports all 72 paired results for PatchTST with $N = 8$ and WPMixer with $N = 4$, using $L = 512$ and $H \in \{96, 192, 336, 720\}$. MSE and MAE are test errors in training-set-standardized units. Lower values indicate better performance; bold marks the lower unrounded value in each pair. Values are rounded to three decimals. Independent-learner test results are taken from Table 10 for WPMixer ($N = 4$) and Table 14 for PatchTST ($N = 8$).

Dataset	H	CF-PatchTST ($N = 8$)				CF-WPMixer ($N = 4$)			
		Shared		Independent		Shared		Independent	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.397	0.415	**0.376**	**0.402**	0.371	0.400	**0.363**	**0.393**
	192	**0.418**	**0.429**	0.434	0.437	0.409	0.426	**0.409**	**0.423**
	336	**0.437**	**0.444**	0.450	0.449	**0.421**	**0.433**	0.436	0.443
	720	0.539	0.512	**0.485**	**0.487**	0.476	0.478	**0.458**	**0.468**
ETTh2	96	0.301	0.360	**0.287**	**0.345**	0.279	0.351	**0.278**	**0.343**
	192	0.383	0.407	**0.356**	**0.388**	0.348	0.387	**0.337**	**0.382**
	336	0.409	0.426	**0.385**	**0.413**	0.374	0.420	**0.371**	**0.413**
	720	0.440	0.459	**0.414**	**0.444**	0.438	0.462	**0.415**	**0.447**
ETTm1	96	0.316	0.366	**0.301**	**0.355**	**0.289**	**0.344**	0.292	0.345
	192	0.352	0.387	**0.335**	**0.374**	**0.325**	**0.366**	0.331	0.367
	336	0.391	0.413	**0.366**	**0.392**	**0.358**	0.386	0.364	**0.385**
	720	0.449	0.447	**0.419**	**0.424**	0.435	0.428	**0.426**	**0.418**
ETTm2	96	0.193	0.274	**0.170**	**0.259**	0.173	**0.258**	**0.170**	0.259
	192	0.253	0.316	**0.229**	**0.300**	0.230	**0.300**	**0.227**	0.301
	336	0.289	0.341	**0.281**	**0.336**	0.293	0.338	**0.278**	**0.334**
	720	0.380	0.403	**0.361**	**0.388**	0.374	0.391	**0.361**	**0.389**
ElectricityLast7	96	**0.131**	**0.237**	0.135	0.242	**0.128**	**0.228**	0.134	0.237
	192	**0.146**	**0.252**	0.148	0.255	**0.142**	**0.240**	0.146	0.248
	336	0.163	0.272	**0.161**	**0.271**	**0.154**	**0.257**	0.159	0.263
	720	0.195	0.315	**0.188**	**0.304**	0.187	0.302	**0.182**	**0.293**
TrafficLast7	96	**0.251**	**0.239**	0.262	0.254	**0.259**	**0.246**	0.269	0.253
	192	**0.267**	**0.253**	0.273	0.262	**0.269**	**0.253**	0.281	0.262
	336	**0.282**	**0.262**	0.286	0.271	**0.287**	**0.266**	0.296	0.271
	720	**0.301**	**0.275**	0.306	0.283	**0.311**	**0.282**	0.318	0.286
Exchange	96	0.118	0.244	**0.095**	**0.219**	0.093	0.213	**0.092**	**0.212**
	192	0.206	0.329	**0.197**	**0.320**	0.207	0.323	**0.200**	**0.317**
	336	0.460	0.501	**0.361**	**0.440**	0.397	0.460	**0.367**	**0.438**
	720	0.956	**0.718**	**0.919**	0.719	1.106	0.768	**0.959**	**0.732**
Weather	96	0.158	0.209	**0.158**	**0.209**	0.147	0.198	**0.145**	**0.197**
	192	0.206	0.252	**0.202**	**0.248**	0.189	0.242	**0.187**	**0.237**
	336	**0.251**	0.287	0.251	**0.286**	0.240	0.279	**0.238**	**0.278**
	720	0.332	0.341	**0.321**	**0.335**	**0.317**	**0.333**	0.317	0.333
Solar-Energy	96	**0.187**	**0.249**	0.195	0.257	**0.189**	0.251	0.194	**0.251**
	192	**0.198**	**0.258**	0.207	0.266	**0.195**	**0.258**	0.208	0.265
	336	**0.205**	0.271	0.215	**0.271**	**0.200**	**0.263**	0.211	0.271
	720	**0.210**	**0.275**	0.220	0.276	0.215	0.277	**0.211**	**0.272**

D.2 GLOBAL-UNIQUE ABLATION

We compare the Global-Unique and WindowLocal-Repeated protocols defined in Section 4.4.2. Both use independent phase learners, identical data splits and backbone hyperparameters, an input length of $L = 512$, and forecast horizons $H \in \{96, 192, 336, 720\}$. Neither uses a communication module or prediction filter. We use the configurations selected by validation in Tables 10 and 14: WPMixer with $N = 4$ and PatchTST with $N = 8$, respectively.

For each backbone and metric m, we compute the relative error reduction from WindowLocal-Repeated to Global-Unique for each task, then average over the 36 matched tasks:

$$\Delta_m^{\mathrm{GU}}(b) = \frac{1}{|\mathcal{G}|} \sum_{g \in \mathcal{G}} 100 \frac{E_{b,\mathrm{WLR},g}^{m,\mathrm{test}} - E_{b,\mathrm{GU},g}^{m,\mathrm{test}}}{E_{b,\mathrm{WLR},g}^{m,\mathrm{test}}}, \tag{14}$$

Here, GU and WLR denote Global-Unique and WindowLocal-Repeated, respectively. $E_{b,p,g}^{m,\mathrm{test}}$ is the test error measured by metric m for backbone b under protocol p on task g. The task set $\mathcal{G}$ contains 36 combinations of dataset and forecast horizon, each weighted equally. Positive values indicate lower error with Global-Unique. Table 4 reports the relative reductions averaged over the four horizons for each dataset, and Table 22 lists all 72 paired results.

Table 22: Global-Unique (GU) and WindowLocal-Repeated (WLR) compared on nine datasets. The table reports all 72 WLR runs with their corresponding GU baselines. PatchTST uses $N = 8$ and WPMixer uses $N = 4$, with $L = 512$ and $H \in \{96, 192, 336, 720\}$ throughout. MSE and MAE are test errors computed after standardization using training-set statistics. Lower values are better. Bold marks the lower error in each pair, determined before rounding; displayed values are rounded to three decimals. GU test results are taken from Table 10 for WPMixer ($N = 4$) and Table 14 for PatchTST ($N = 8$).

Dataset	H	CF-PatchTST ($N = 8$)				CF-WPMixer ($N = 4$)			
		GU		WLR		GU		WLR	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.376**	**0.402**	0.399	0.418	**0.363**	**0.393**	0.371	0.399
	192	0.434	**0.437**	**0.428**	0.437	0.409	0.423	**0.400**	**0.417**
	336	0.450	**0.449**	**0.446**	0.454	0.436	0.443	**0.426**	**0.436**
	720	0.485	0.487	**0.480**	**0.479**	0.458	0.468	**0.454**	**0.463**
ETTh2	96	**0.287**	**0.345**	0.319	0.368	**0.278**	**0.343**	0.286	0.357
	192	**0.356**	**0.388**	0.390	0.414	0.337	**0.382**	**0.332**	0.387
	336	**0.385**	**0.413**	0.414	0.427	**0.371**	**0.413**	0.377	0.424
	720	**0.414**	**0.444**	0.422	0.447	**0.415**	**0.447**	0.431	0.457
ETTm1	96	**0.301**	**0.355**	0.315	0.367	0.292	0.345	**0.286**	**0.343**
	192	**0.335**	**0.374**	0.356	0.391	0.331	**0.367**	**0.330**	0.369
	336	**0.366**	**0.392**	0.383	0.407	**0.364**	**0.385**	0.366	0.387
	720	**0.419**	**0.424**	0.443	0.444	0.426	**0.418**	**0.426**	0.425
ETTm2	96	**0.170**	**0.259**	0.187	0.275	**0.170**	0.259	0.175	**0.259**
	192	**0.229**	**0.300**	0.241	0.310	**0.227**	0.301	0.231	**0.300**
	336	**0.281**	**0.336**	0.291	0.344	**0.278**	**0.334**	0.283	0.337
	720	**0.361**	**0.388**	0.374	0.396	**0.361**	**0.389**	0.368	0.391
ElectricityLast7	96	0.135	0.242	**0.131**	**0.237**	0.134	0.237	**0.128**	**0.228**
	192	0.148	0.255	**0.146**	**0.253**	0.146	0.248	**0.141**	**0.240**
	336	**0.161**	**0.271**	0.164	0.272	0.159	0.263	**0.155**	**0.257**
	720	**0.188**	**0.304**	0.197	0.311	**0.182**	**0.293**	0.191	0.297
TrafficLast7	96	0.262	0.254	**0.253**	**0.243**	0.269	0.253	**0.258**	**0.245**
	192	0.273	0.262	**0.268**	**0.254**	0.281	0.262	**0.270**	**0.255**
	336	0.286	0.271	**0.283**	**0.266**	0.296	0.271	**0.288**	**0.266**
	720	0.306	0.283	**0.302**	**0.276**	0.318	0.286	**0.311**	**0.281**
Exchange	96	**0.095**	**0.219**	0.104	0.230	**0.092**	**0.212**	0.102	0.222
	192	0.197	0.320	**0.195**	**0.318**	**0.200**	**0.317**	0.211	0.326
	336	**0.361**	**0.440**	0.424	0.481	**0.367**	**0.438**	0.474	0.489
	720	**0.919**	**0.719**	0.965	0.725	**0.959**	**0.732**	1.061	0.765
Weather	96	0.158	0.209	**0.156**	**0.207**	**0.145**	**0.197**	0.146	0.199
	192	0.202	0.248	**0.199**	**0.245**	**0.187**	**0.237**	0.189	0.239
	336	**0.251**	**0.286**	0.254	0.288	**0.238**	**0.278**	0.245	0.286
	720	**0.321**	**0.335**	0.331	0.340	**0.317**	**0.333**	0.324	0.339
Solar-Energy	96	0.195	0.257	**0.189**	**0.248**	0.194	**0.251**	**0.190**	0.252
	192	0.207	0.266	**0.196**	**0.255**	0.208	0.265	**0.195**	**0.255**
	336	0.215	0.271	**0.201**	**0.263**	0.211	0.271	**0.201**	**0.265**
	720	0.220	0.276	**0.207**	**0.267**	**0.211**	**0.272**	0.217	0.279