PRISM: A HIERARCHICAL MULTISCALE APPROACH FOR TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series forecasting requires identifying which components of the observed context are informative about the future. Multiscale decompositions expose temporal and spectral structure of time series but the predictive relevance of these components can vary across samples and within the context window. We introduce PRISM, a forecasting framework that learns input-dependent selection over localized time-frequency supports of the context window. PRISM constructs a hierarchy of overlapping temporal regions and uses probabilistic selection to couple frequency weighting within each region with band-specific routing between sibling regions. The resulting allocation is learned end-to-end through the forecasting objective. Across eight standard long-horizon benchmarks, PRISM achieves the lowest mean squared error in 20 of 32 settings, and in 61 of 97 settings on GIFT-Eval. Controlled ablations establish the contribution of coupled hierarchical selection beyond flat aggregation and dense mixing. The learned allocations concentrate on a subset of supports, prioritize components whose deletion substantially impairs forecasting, and shift when predictive structure is relocated in synthetic experiments. These results support structured predictive-support selection as an effective inductive bias for time-series forecasting.

1 INTRODUCTION

Time-series forecasting supports decision-making across science, engineering, finance, and healthcare, where predictions guide planning, adaptation, and intervention (Box et al., 2015; Ebadi & Sahafizadeh, 2023; Zou et al., 2022). A central challenge of the time-series forecasting is to determine which parts of the observed context of a time series are informative about its future. Predictive information can be distributed unevenly across the signal, and its location and relevance can vary between inputs. Accurate forecasting therefore requires identifying the components of the time series that support a particular prediction and determining how strongly each component should contribute to it.

Multiscale time-series forecasting methods attempt to address this challenge by organizing the observed context of time series into components that expose their structure at different temporal resolutions or frequencies. This separation allows a predictor to model distinct dynamics within a time series and combine information that may be difficult to extract directly from the original signal. N-HiTS, SCINet, and TimeMixer implement this principle through hierarchical interpolation, recursive splitting, and cross-scale mixing of time-series data (Challu et al., 2023; Liu et al., 2022; Wang et al., 2024a). Spectral and time-frequency models for time-series forecasting further isolate time-series components using Fourier, wavelet, and other related decompositions (Zhou et al., 2022b; Yi et al., 2023; Murad et al., 2025; Zhang et al., 2025; Yuan & Chen, 2024). Adaptive time-series forecasting methods learn how to combine or select these representations: TimesNet weights period-specific outputs, Temporal Fusion Transformer selects input variables, and Pathformer adapts multiscale processing pathways (Wu et al., 2023; Lim et al., 2021; Chen et al., 2024).

These approaches motivate a more explicit treatment of how predictive weights are allocated across the components of context windows in time series. Decomposition of time-series context window signal into components makes its structure accessible but does not by itself determine which of the components are informative about the signal in the forecast window. Moreover, the relevance of a frequency component can vary across different parts of the context window, so a shared spectral weighting may obscure local differences in predictive content. In this work, we therefore investigate

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.

predictive-support selection: input-dependent routing of frequency bands localized within temporal segments, built on two coupled decisions–which bands stand out within each segment, and conditioned on this, where in time each band is routed.

To implement the predictive-support selection, we introduce PRISM (Predictive Routing and Selection across Multiscale representations), a forecasting framework that employs a *probabilistic selection tree*. Our model recursively partitions the time-series context window into overlapping temporal regions and decomposes each temporal region into frequency components. Learned routers compute probabilistic selection scores over frequency bands within each temporal region and balance them across temporal regions. Applying these operations throughout the temporal hierarchy defines a routing flow that weights time-frequency components aggregated by the forecasting head. The explicit routing scheme enables the analysis of predictive-support concentration, its sensitivity to individual component deletion, and its adaptation to changes in the predictive structure design (Section 4.4).

Our experiments both evaluate the model's forecasting accuracy and analyze whether the learned predictive-support allocation identifies the correct time-frequency components of time-series context window that support the forecast. We find that PRISM achieves the lowest MSE in 20 of 32 settings across eight standard long-horizon benchmarks and in 61 of 97 settings on a larger GIFT-Eval benchmark. Controlled ablations show that simpler processing of time-series components including the flat aggregation and the dense implicit mixing of the components do not recover the performance of coupled hierarchical selection, suggesting that the time-frequency components' aggregation mechanism matters beyond their availability in the representation. Support-deletion experiments further connect the learned predictive-support allocation to the model's behavior: removing high-weight components causes substantially greater degradation than removing an equal number of low-weight components. Controlled synthetic experiments further show that the predictive-support allocation shifts when predictive structure is relocated, providing evidence against a fixed preference for particular support selection. Together, these results suggest that structured predictive-support selection is an effective inductive bias for time-series forecasting and show that its learned weights reflect input-dependent relevance for forecast-window predictions. Our contributions are threefold:

- We formulate predictive-support selection in time-series forecasting as an input-dependent allocation over localized time-frequency components, explicitly modeling interactions between temporal location and spectral domains of time-series data.
- We introduce PRISM, a forecasting architecture whose probabilistic selection mechanism couples frequency weighting within temporal regions with routing across temporal regions, trained end-to-end through the forecasting objective.
- We demonstrate strong forecasting performance across standard benchmarks and GIFT-Eval, and characterize the contribution and behavior of the selection mechanism through architectural ablations, support-deletion tests, and synthetic experiments.

2 Related Work

Multiscale time-series representations as predictive supports for time-series forecasting. Decomposition-based time-series forecasting models construct time-series components that expose different aspects of temporal structure of time series. DLinear (Zeng et al., 2023) applies separate linear mappings to trend and seasonal components, while N-HiTS (Challu et al., 2023) combines multi-rate sampling with hierarchical interpolation. TimeMixer (Wang et al., 2024a) mixes decomposed time-series representations across temporal resolutions, and SCINet (Liu et al., 2022) recursively splits the time domain into branches and models cross-branch interactions. Spectral time-series forecasting methods provide a complementary type of time-series representations: FEDformer (Zhou et al., 2022b) incorporates Fourier or wavelet processing, FreTS (Yi et al., 2023) learns frequency-domain mappings of time series, and FiLM (Zhou et al., 2022a) combines Legendre memory representations with Fourier processing. WPMixer (Murad et al., 2025) and WaveletMixer (Zhang et al., 2025) combine wavelet decomposition of time series with multiresolution mixing, while D-PAD (Yuan & Chen, 2024) disentangles multi-frequency patterns fo time-series data through decomposition and reconstruction. Viewed together, these types of time-series representations provide candidate support families that may enable time-series forecasting. Their construction motivates the overarching question of how to allocate forecasting weight across time-series components localized at different times and frequencies.

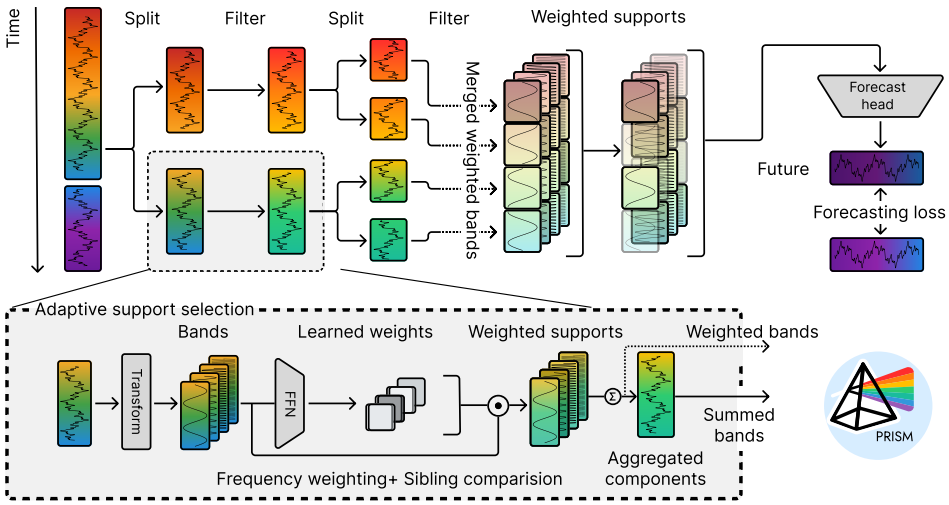

Figure 1: **Overview of PRISM.** The model splits the context window into localized time-frequency components (top), assigns input-dependent routing weights to these components through a hierarchy of sibling comparisons (bottom), and combines the routed components to form the forecast over the prediction window (right).

Adaptive selection of forecast-relevant information. Several time-series forecasting architectures explicitly adapt their time-series representations to the input data. TimesNet (Wu et al., 2023) identifies dominant periods of time series, constructs period-specific two-dimensional representations, and adaptively aggregates their outputs. Temporal Fusion Transformer (Lim et al., 2021) combines input-dependent feature selection with temporal attention, while Pathformer (Chen et al., 2024) learns adaptive pathways through multiscale processing modules. TFDNet (Luo et al., 2025) further demonstrates the utility of explicit time-frequency representations for the modeling of the local spectral structure of time-series data. These approaches establish that adaptive weighting and multiscale modeling can be viewed as two complementary aspects of time-series modeling. We thus consider both aspects in our work.

Structured selection over time-frequency supports. (Coifman & Wickerhauser, 1992) establish adapting multiscale time-series representations to signal content. They propose selecting an orthonormal basis from a structured collection of time-series data representations using additive information costs. PRISM instead learns a soft predictive support allocation over localized time-series components in time-frequency domain using the forecasting objective, which allows the components' relevance to be determined by their utility for the forecast quality, rather than an explicit representation cost.

3 METHODS

3.1 TIME-SERIES FORECASTING PROBLEM FORMULATION

Let $\{X_t\}_{t=1}^{T}$ be a multivariate time series with $X_t \in \mathbb{R}^C$, where C denotes the number of channels. Each forecasting sample consists of a context window $X_{\text{ctx}} \in \mathbb{R}^{L \times C}$ and the subsequent forecast window $Y \in \mathbb{R}^{H \times C}$, where L and H denote the context length and forecast horizon. Under squared-error loss, the optimal predictor is the conditional mean $\mathbb{E}[Y \mid X_{\text{ctx}}]$.

PRISM estimates the conditional mean via input-dependent selection over localized time–frequency components of the context window. The model constructs a hierarchy of overlapping temporal segments spanning the context window of the time series, then decomposes each of these temporal segment into frequency bands, and finally learns the routing flow that determines, for each frequency band, how much weight each temporal region receives. A forecasting head then combines the routed components to predict the time-series contents over the forecast window (Figure 1).

3.2 HIERARCHICAL TIME-FREQUENCY REPRESENTATIONS OF TIME-SERIES DATA

Temporal partitioning of time-series data. We organize the time-series data in the context window into a binary tree of overlapping temporal segments, as defined below.

Definition 1 (Binary temporal partition tree with overlaps)**.** *A binary temporal partition tree $\mathcal{T}$ of depth D consists of index intervals $\{I_l^m : l = 0, \dots, D,\ m = 1, \dots, 2^l\}$. The root is $I_0^1 = \{1, \dots, L\}$. For an internal node $I_l^m = \{a, \dots, b\}$, define $n = b - a + 1$ and $r = n/2 + o$, where the integer overlap satisfies $0 \le o < n/2$. Its children are*

$$I_{l+1}^{2m-1} = \{a, \dots, a + r - 1\}, \qquad I_{l+1}^{2m} = \{b - r + 1, \dots, b\}.$$

The children cover the parent interval and share $2o$ samples. Nodes at depth D are leaves and are not further subdivided.

Local frequency decomposition. In the temporal partition tree, an arbitrary level l contains 2^l temporal segments. Within these segments, we compute local frequency components to obtain a multiscale time-frequency representation, as follows.

Definition 2 (Local frequency feature set)**.** *For node $I_l^m \in \mathcal{T}$ and channel c, let $X_{l,c}^m = X_{\mathrm{ctx}}[I_l^m, c]$ denote the local signal. Given band-specific operators $\mathcal{D}_{l,k}^m : \mathbb{R}^{|I_l^m|} \to \mathbb{R}^{|I_l^m|}$, the local frequency feature set is*

$$\mathcal{U}_{l,c}^m = \{U_{l,k,c}^m\}_{k=1}^K, \qquad U_{l,k,c}^m = \mathcal{D}_{l,k}^m(X_{l,c}^m).$$

To form a local frequency feature set, we use the Haar discrete wavelet transform with $\mathcal{D}_{l,k}^m$ reconstructing the contribution of the k-th frequency band at the segment's original length. Unless otherwise specified, $K = 5$. Bands are indexed from low to high frequency, so that $\sum_k U_{l,k,c}^m = X_{l,c}^m$. The collection $\{U_{l,k,c}^m\}$, gathered across temporal nodes, frequency bands, and channels, forms the time-frequency representation fo the time-series context window, used for time-series forecasting. Alternative decompositions, including FFT, binomial pyramids, DoG, EMA, and D-PAD's MCD block, are evaluated in Supplementary Table A.10.

3.3 Time-frequency selection of predictive support for time-series forecasting

Time-frequency component scoring. For each time-frequency component $U_{l,k,c}^m$ of the time-series context window, we compute a fixed-dimensional statistical descriptor as follows:

$$\phi_{l,k,c}^m = \left[\mu,\ \sigma,\ a_{\max},\ \|\Delta U_{l,k,c}^m\|_1,\ \|\Delta^2 U_{l,k,c}^m\|_1,\ \frac{a_{\max}}{\sigma + \epsilon}\right],$$

where μ, σ, and $a_{\max}$ denote the time-frequency component's mean, standard deviation, and peak amplitude; $\|\Delta U\|_1$ and $\|\Delta^2 U\|_1$ denote its maximum amplitudes of the first- and second-order temporal derivatives; and $\epsilon > 0$ ensures numerical stability. These statistics summarize the location, dispersion, peak amplitude, roughness, curvature, and peak-to-dispersion ratio of the time-frequency segment of the time series. Their fixed dimensionality allows the same router architecture to operate on time-series time-frequency segments of different lengths.

For each node (l, m) of the binary temporal partition tree (depth l, position m), a two-layer MLP router $h_{\theta_{l,m}}$ maps each statistical descriptor, defined above, to a scalar score:

$$s_{l,k,c}^m = h_{\theta_{l,m}}(\phi_{l,k,c}^m).$$

Within a node, the router is shared across both children, all frequency bands, and all channels.

Normalization of component scores across frequencies and time. We convert the scores, defined above for individual time-frequency components, into corresponding routing weights through two successive normalizations. First, a softmax over frequency bands determines their relative importance within each temporal node. Second, a softmax between temporal nodes determines their relative importance separately for each frequency band.

Suppressing the channel index and conditioning on X_{ctx}, let $m_L = 2m - 1$ and $m_R = 2m$ index the children of node (l, m). The routing distributions are

$$\begin{aligned} q_\theta^{\mathrm{freq}}(k \mid l, m) &= \frac{\exp(s_{l,k}^m/\tau_f)}{\sum_{j=1}^K \exp(s_{l,j}^m/\tau_f)}, \\ q_\theta^{\mathrm{branch}}(m_b \mid l, m, k) &= \frac{\exp\Big(q_\theta^{\mathrm{freq}}(k \mid l+1, m_b)/\tau_t\Big)}{\sum_{m' \in \{m_L, m_R\}} \exp\Big(q_\theta^{\mathrm{freq}}(k \mid l+1, m')/\tau_t\Big)}, \end{aligned} \tag{1}$$

where $m_b \in \{m_L, m_R\}$ and $\tau_f, \tau_t > 0$ control the concentration of frequency and time weights, respectively. The distributions satisfy

$$\sum_{k=1}^{K} q_\theta^{\text{freq}}(k \mid l, m) = 1, \qquad \sum_{m_b \in \{m_L, m_R\}} q_\theta^{\text{branch}}(m_b \mid l, m, k) = 1.$$

This mechanism couples spectral and temporal selection rather than assigning a single temporal weight to all frequencies. Applying these normalizations throughout the time-frequency hierarchy defines the input-dependent routing flow $\mathcal{Q}_\theta(X_{\text{ctx}})$.

The routing weight of component $U_{l+1,k}^{m_b}$ is its branch probability, $w_{l+1,k}^{m_b} = q_\theta^{\text{branch}}(m_b \mid l, m, k)$, and the reweighted signal of child m_b is

$$\textstyle\sum_{k=1}^{K} w_{l+1,k}^{m_b} U_{l+1,k}^{m_b}, \tag{2}$$

which is split and decomposed at the next level. Applying these normalizations throughout the time-frequency hierarchy defines the input-dependent routing flow $\mathcal{Q}_\theta(X_{\text{ctx}}) = \{w_{l,k,c}^m\}$.

3.4 Forecasting Block

The forecasting head g_θ combines the localized time-frequency components of the time-series context window according to the learned routing flow to produce

$$\widehat{Y}_\theta(X_{\text{ctx}}) = g_\theta\big(\{U_{l,k,c}^m\}, \mathcal{Q}_\theta(X_{\text{ctx}})\big) \in \mathbb{R}^{H \times C}. \tag{3}$$

Each component has its own two-layer MLP head applied to the reweighted component, $y_{l,k,c}^m = f_{l,m,k}(w_{l,k,c}^m U_{l,k,c}^m)$; the head outputs are rescaled by a squeeze-and-excitation gate computed from their horizon averages, summed over (l, m, k), and de-normalized. All trainable modules and parameters of the model are optimized jointly using the mean squared error forecasting loss:

$$\mathcal{L}_{\text{MSE}}(\theta) = \mathbb{E}_{(X_{\text{ctx}}, Y)}\left[\|Y - \widehat{Y}_\theta(X_{\text{ctx}})\|_F^2\right]. \tag{4}$$

The routers receive supervision only through the forecasting loss; no explicit sparsity or entropy regularization is imposed. This objective admits a conditional-likelihood interpretation. Under an isotropic Gaussian observation model,

$$\text{vec}(Y) \mid X_{\text{ctx}} \sim \mathcal{N}\Big(\text{vec}(\widehat{Y}_\theta(X_{\text{ctx}})), \sigma_{\text{obs}}^2 I_{HC}\Big),$$

the expected negative log-likelihood is

$$\mathbb{E}[-\log p_\theta(Y \mid X_{\text{ctx}})] = \frac{1}{2\sigma_{\text{obs}}^2}\mathcal{L}_{\text{MSE}}(\theta) + \text{const}.$$

Under the assumption of the fixed observation variance, minimizing the mean squared error forecasting loss is thus equivalent to maximizing the conditional likelihood of the model predicting the time-series signal within the forecast window.

4 Results

We evaluate PRISM along four main sets of experiments. First, we compare its forecasting accuracy with strong recent baselines across eight standard long-horizon benchmarks. Second, we test whether its advantages persist across heterogeneous and artifact-rich datasets in GIFT-Eval. Third, we use controlled ablations to isolate the contributions of our design choices. Finally, we examine whether the learned predictive support allocations are compact, stable, and aligned with known predictive structure of the input.

4.1 Experimental setup

Datasets. We evaluate PRISM on eight standard forecasting benchmarks following prior protocols (Wu et al., 2023; Wang et al., 2024b): four ETT datasets (Zhou et al., 2021) (ETTh1, ETTh2, ETTm1, and ETTm2), Electricity, Traffic, Exchange (Lai et al., 2018), and Weather (Zeng et al.,

2023). We use the standard train/validation/test splits: 6:2:2 for ETT and 7:1:2 for the remaining datasets. We additionally evaluate our model on the GIFT-Eval benchmark (Aksu et al., 2024), which comprises approximately 100 datasets spanning diverse domains, sampling intervals from 10 seconds to one year, and context lengths from 16 to 336 time points. In contrast to the standard benchmarks, GIFT-Eval includes irregular timestamps, missing observations, transient anomalies, distribution shifts, long-term drift, and short-context settings. Conforming to the GIFT-Eval protocol, we report the MASE and additionally evaluate MSE and MAE for comparability with Section 4.2.

Training and evaluation. We evaluate the models on forecasting horizons of 96, 192, 336, and 720 timesteps. Unless otherwise specified, the components introduced in Section 3 are fixed across experiments: tree depth $D = 2$, temporal overlap $o = 8$, and temperatures $\tau_f = \tau_t = 0.2$. We fix the context length to 336, the batch size to 512, and the AdamW optimizer's learning rate to 10^{-4}. Early stopping criterion monitors the validation MSE with the patience 15 and margin $\delta = 2 \times 10^{-4}$. Dataset-specific settings, including hidden width, dropout, number of frequency bands, epoch budget, and patience, follow the released configurations and are reported for each dataset and horizon in Supplementary Table A.15. Sensitivity to the number of bands, tree depth, and overlap is reported in Supplementary Tables A.11-A.13.

To be able to estimate the statistical significance of our results in comparison with the baselines, we rerun every baseline in every setting using its published optimal hyperparameters and verify agreement with the reported results (Supplementary Table A.3). We evaluate all methods over eight random seeds and report the test MSE and MAE. The statistical significance of our model being better than the baselines is assessed by averaging the seed-level results across horizons and applying two-tailed paired t-tests ($n = 8$, $df = 7$) followed by the Benjamini-Hochberg false discovery rate correction (Supplementary Table A.4).

4.2 PRISM SHOWS BEST PERFORMANCE ACROSS STANDARD FORECASTING BENCHMARKS

To assess the performance of our model with respect to the SOTA baselines, we compare PRISM with D-PAD (Yuan & Chen, 2024), DLinear (Zeng et al., 2023), TimeMixer (Wang et al., 2024a), N-HiTS (Challu et al., 2023), iTransformer (Liu et al., 2023), PatchTST (Nie et al., 2022), and XPatch (Stitsyuk & Choi, 2025) (Table 1). The evaluation comprises eight datasets and four forecast horizons, yielding 32 settings. We found that, across these settings, PRISM achieves the lowest MSE in 20 settings and the lowest MAE in 19 settings. The closest competitor, D-PAD, achieves the lowest MSE in 10 settings and the lowest MAE in 11 settings. Ties are credited to every tied method and may therefore cause the win counts to sum to more than 32. Paired two-tailed t-tests followed by the Benjamini-Hochberg correction show that PRISM improves significantly upon the baselines at the 5% FDR level in 35 of 36 settings in the MSE and in 36 of 36 settings in the MAE, with the only exception being the Exchange dataset with the DLinear model on the MSE metric (Supplementary Tables A.4 and A.5). The significance analysis covers a partially different baseline set from Table 1, additionally including SCINet and WPMixer, for which matched eight-seed runs were available across all six datasets. Complete hyperparameter tuning protocols and selected hyperparameter configurations are provided in Supplementary Tables A.15, A.16 and A.17. We additionally evaluate PRISM against D-PAD and TimeMixer, the two strongest baselines in Table 1, using a context length of 96 (Supplementary Table A.14).

4.3 PRISM SHOWS BEST PERFORMANCE ACROSS FORECASTING CONDITIONS

We evaluate PRISM on the GIFT-Eval benchmark (Aksu et al., 2024) to assess the forecasting performance of our model across diverse types of time series, sampling frequencies, and forecast horizons. In this evaluation, we compare the performance of our model against D-PAD, DLinear, FITS, and ModernTCN using the MSE and MAE as the metrics, together with the benchmark's native mean absolute scaled error (MASE) metric. Table 2 summarizes the models' overall performances and the results across six non-exclusive dataset categories. Full results and category-assignment rationales are described in Supplementary Tables A.7 and A.6.

First, we evaluate PRISM on the entirety on the GIFT-Eval benchmark (on protocol) to be able to compare directly with the results of other models. Across 97 evaluation settings, PRISM achieves the lowest MSE in 61 (62.9%) of the settings and the lowest MAE in 65 (67.0%) of the settings, compared with 12 settings for ModernTCN and 11 settings for FITS, respectively. Under the MASE

Table 1: **Forecasting performance across future horizons (8 seeds).** (Red) best models; (purple) second best models. For datasets covered by Table A.3, D-PAD, DLinear, and TimeMixer use the more favorable result between the published value and our 8-seed reproduction, while XPatch uses the reproduced value reported there. All reproduced values are means over the same eight seeds, matching the protocol used for PRISM. Ties are counted for all tied methods, so the *Best count* rows can sum to more than the 32 evaluated settings.

Dataset	Ctx	H	PRISM		D-PAD		DLinear		TimeMixer		N-HiTS		iTransformer		PatchTST		XPatch	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	336	96	**0.355**	**0.374**	0.357	0.376	0.372	0.396	0.385	0.408	0.475	0.498	0.386	0.405	0.414	0.419	0.377	0.398
	336	192	**0.390**	0.405	0.394	**0.402**	0.405	0.416	0.429	0.432	0.492	0.519	0.441	0.436	0.413	0.429	0.431	0.427
	336	336	0.386	0.407	**0.374**	**0.406**	0.439	0.443	0.443	0.446	0.550	0.564	0.487	0.458	0.422	0.440	0.461	0.443
	336	720	0.421	0.445	**0.419**	**0.442**	0.472	0.490	0.457	0.462	0.598	0.641	0.503	0.491	0.447	0.468	0.487	0.476
ETTh2	336	96	**0.261**	**0.322**	0.270	0.327	0.289	0.353	0.276	0.339	0.328	0.364	0.297	0.349	0.274	0.337	0.280	0.334
	336	192	**0.311**	**0.359**	0.331	0.368	0.375	0.411	0.340	0.381	0.372	0.408	0.380	0.400	0.341	0.377	0.345	0.377
	336	336	**0.308**	**0.361**	0.321	0.370	0.448	0.465	0.362	0.404	0.397	0.421	0.428	0.432	0.329	0.384	0.367	0.398
	336	720	0.387	0.420	**0.369**	**0.415**	0.605	0.551	0.398	0.433	0.461	0.497	0.427	0.445	0.379	0.422	0.393	0.426
ETTm1	336	96	**0.282**	**0.326**	0.285	0.328	0.299	0.343	0.299	0.350	0.370	0.468	0.334	0.368	0.293	0.346	0.289	0.333
	336	192	**0.323**	0.352	**0.323**	**0.349**	0.335	0.365	0.335	0.374	0.436	0.488	0.377	0.391	0.333	0.370	0.331	0.358
	336	336	0.354	0.377	**0.351**	**0.372**	0.369	0.386	0.366	0.392	0.483	0.510	0.426	0.420	0.369	0.392	0.365	0.381
	336	720	0.423	0.409	**0.412**	**0.405**	0.425	0.421	0.432	0.434	0.489	0.537	0.491	0.459	0.416	0.420	0.428	0.418
ETTm2	336	96	**0.158**	**0.234**	0.162	0.247	0.167	0.260	0.172	0.258	0.184	0.262	0.180	0.264	0.166	0.256	0.164	0.248
	336	192	**0.214**	**0.267**	0.218	0.283	0.224	0.303	0.231	0.300	0.260	0.293	0.250	0.309	0.223	0.296	0.221	0.287
	336	336	**0.264**	**0.319**	0.267	0.321	0.281	0.342	0.272	0.327	0.313	0.359	0.311	0.348	0.274	0.329	0.277	0.324
	336	720	**0.352**	**0.372**	0.353	**0.372**	0.397	0.421	0.361	0.386	0.411	0.421	0.412	0.407	0.362	0.385	0.358	0.376
Traffic	336	96	0.362	0.237	**0.359**	**0.236**	0.410	0.282	0.380	0.275	0.407	0.290	0.395	0.268	0.360	0.249	0.481	0.280
	336	192	**0.376**	**0.244**	0.377	0.245	0.423	0.287	0.401	0.282	0.423	0.302	0.417	0.276	0.379	0.256	0.484	0.275
	336	336	0.401	0.257	**0.391**	**0.253**	0.436	0.296	0.413	0.290	0.446	0.321	0.433	0.283	0.392	0.264	0.500	0.279
	336	720	0.433	**0.271**	**0.413**	0.282	0.446	0.305	0.455	0.305	0.528	0.369	0.467	0.302	0.432	0.286	0.534	0.293
Electricity	336	96	**0.122**	**0.219**	0.128	0.220	0.140	0.237	0.130	0.224	0.151	0.254	0.148	0.240	0.129	0.222	0.159	0.244
	336	192	**0.145**	0.235	0.148	0.236	0.153	0.247	0.152	0.246	0.170	0.273	0.149	**0.231**	0.147	0.240	0.160	0.248
	336	336	**0.159**	**0.243**	0.163	0.253	0.160	0.259	0.169	0.263	0.200	0.291	0.178	0.253	0.163	0.259	0.182	0.267
	336	720	**0.194**	**0.281**	0.201	0.286	0.203	0.301	0.204	0.293	0.244	0.356	0.225	0.317	0.197	0.290	0.216	0.298
Exchange	336	96	**0.081**	**0.198**	0.096	0.218	**0.081**	0.203	0.094	0.218	0.092	0.208	0.086	0.206	0.095	0.217	0.084	0.203
	336	192	0.173	0.297	0.200	0.322	**0.157**	**0.293**	0.197	0.318	0.208	0.300	0.177	0.299	0.201	0.322	0.181	0.304
	336	336	0.314	0.396	0.376	0.446	**0.305**	**0.414**	0.372	0.444	0.371	0.509	0.331	0.417	0.372	0.447	0.353	0.426
	336	720	0.712	0.606	1.117	0.761	**0.643**	**0.601**	0.977	0.723	0.798	0.608	0.847	0.691	0.873	0.699	0.954	0.725
Weather	336	96	**0.140**	**0.177**	0.143	0.181	0.174	0.235	0.150	0.203	0.160	0.197	0.174	0.214	0.149	0.198	0.150	0.187
	336	192	**0.187**	**0.227**	0.189	0.229	0.217	0.276	0.195	0.246	0.207	0.265	0.221	0.254	0.194	0.241	0.193	0.230
	336	336	**0.234**	**0.266**	0.239	0.268	0.261	0.311	0.242	0.281	0.273	0.301	0.278	0.296	0.245	0.282	0.245	0.271
	336	720	0.311	0.320	**0.304**	**0.313**	0.323	0.362	0.319	0.335	0.363	0.352	0.358	0.347	0.314	0.334	0.320	0.324
Best count (MSE)			**20**	—	**10**	—	4	—	0	—	0	—	0	—	0	—	0	—
Best count (MAE)			—	**19**	—	**11**	—	2	—	0	—	0	—	1	—	0	—	0

Table 2: **Forecasting wins on GIFT-Eval.** Entries count settings with the lowest error; higher is better. Overall results cover 97 settings. Category-level results report MSE and MAE; categories overlap. Bold indicates the highest count in each column.

	Overall			Irregular		Aperiodic		Incomplete		Anomalous		Nonstationary		Long-term drift	
Model	MAE	MSE	MASE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
PRISM	**65**	**61**	**63**	**17**	**16**	**14**	**16**	**32**	**27**	**18**	**20**	**41**	**37**	**16**	**15**
D-PAD	9	11	10	6	3	6	4	3	4	3	2	2	6	1	2
DLinear	3	5	2	1	3	1	2	3	4	1	0	3	4	0	0
FITS	11	8	14	4	3	6	3	6	4	3	1	5	5	4	3
ModernTCN	9	12	8	2	5	5	7	3	8	1	3	6	5	3	4

metric, which scales the absolute forecasting error by the training-series error of a naive reference forecast, PRISM shows the best performance in 63 settings (64.9%), compared to 14 settings for FITS. Overall, the advantage of our model's performance is consistent across both scaled and unscaled error metrics. Our strong performance also extends to limited-context settings: among the 27 settings with $T_{\text{ctx}} \leq 90$, PRISM ranks first in 19 settings, compared to 6 settings for ModernTCN (Supplementary Table A.8).

We further examine the performance of our model separately in six different dataset categories: irregular, aperiodic, incomplete, anomalous, nonstationary, and long-term drifting time-series data. We find that PRISM achieves the highest MSE and MAE in every category. The largest absolute gaps in win counts for PRISM was observed for nonstationary data, where it obtained 37 MSE and 41 MAE wins, compared with 6 wins for the strongest baseline under either metric. For incomplete data, PRISM achieved 27 MSE and 32 MAE wins, compared with 8 and 6 wins for the closest baselines respectively. PRISM's advantage also persisted for irregular, weakly periodic, transient anomalous, and long-term drifting time series, indicating that the overall high performance of PRISM was not confined to a single class of time series types.

Table 3: **Concentration of the learned allocation.** Lower entropy and effective support, and higher top-3 mass, indicate stronger concentration. This analysis uses $K = 6$ frequency bands over the two temporal regions at the analyzed tree level, giving 12 candidate supports; the uniform-allocation reference (Unif.) is computed over these 12 candidates.

	ETTh1	ETTh2	ETTm1	ETTm2	Exch.	Weath.	Avg.	Unif.
Norm. entropy ↓	0.85	0.27	0.54	0.33	0.64	0.02	0.44	1.00
Eff. support ↓	8.75	3.16	5.53	4.33	7.52	1.24	5.09	12.0
Top-3 mass ↑	0.50	0.88	0.75	0.77	0.56	0.98	0.74	0.25

4.4 PRISM LEARNS SPARSE INFORMATIVE PREDICTIVE SUPPORTS

We next examine what the model selects: whether its allocation concentrates on a subset of supports, whether those supports contribute to the forecast, and whether their weights track predictive content.

The learned allocation is concentrated on a subset of time-frequency components of time series. We quantify our model's selection of the predictive time-frequency components of the time series using normalized entropy, effective support size, and the cumulative mass assigned to the three highest-weight time-frequency supports (Table 3). Across datasets and metrics, we find a substantial difference w.r.t. the uniform baselines, suggesting that our model finds a sparse forecasting support within the time-frequency decomposition of the context window data in the time series.

High-weight supports contribute more to forecasting accuracy. Concentration alone does not establish predictive relevance. We therefore remove either the highest-weight supports or an equal number of lowest-weight supports at inference time (Table 4). Removing low-weight supports increases MSE by only 1.1% on ETTh1 and 1.8% on ETTm1, whereas removing high-weight supports increases MSE by 14.1% and 65.6%, respectively. For the same deletion budget, removing prioritized supports thus produces more than an order-of-magnitude larger degradation, indicating that the learned weights identify components on which the forecast depends.

Table 4: **Deletion analysis of learned predictive-support weights.** We remove equal numbers of low- or high-weight supports and report MSE. Relative changes from the full model appear in parentheses.

Dataset	Full	Low-weight deletion	High-weight deletion
ETTh1	0.355	0.359 (+1.13%)	0.405 (+14.08%)
ETTm1	0.282	0.287 (+1.77%)	0.467 (+65.60%)

4.5 PRISM'S PREDICTIVE SUPPORTS ARE STABLE ACROSS COMPARABLE DATA

To evaluate whether PRISM captured the relevant components of the context-window data, we evaluated the importance scores (i.e. the weighting factors learned by the model to scale the time-frequency components of time series produced by the Wavelet transformation) based on their importance for the forecasting in our datasets. To this end, we used the electric transformer temperature (ETT) datasets. These datasets offer the data collected at two different electric plants (ETT1 and ETT2) collected at different sampling rates (ETTh at the rate of 1h and ETTm at the rate of 15 min). These four datasets (ETTh1, ETTh2, ETTm1, and ETTm2) provide comparable conditions with controlled variations, allowing us to evaluate whether the model learns robust and interpretable representations of time series. To make the evaluation of our model tractable, we trained the models with the tree depth equal to 2, thus only performing one binary split of the data in time. As a reference setting, we used the ETTm1 dataset with the context window of 336 time points, the forecast window of 720 time points, and 5 Wavelet bands.

Table 5: **Correlations between importance scores** learned under different settings.

Comparison	Corr
Segments (R vs L)	1.00
Seeds	0.99
Forecast length (96 vs 720)	0.99
Dataset (ETT2 vs ETT1)	0.97
Bands (5 vs 8)	0.98
Resolutions (Etth vs ETTm)	0.99

We found that the importance scores learned by our model were different, increasing towards lower frequencies, enabling the correlational analysis. We found that this distribution was stable across training seeds ($r = 0.99$, Table 5), segments of the time series (left vs. right; $r = 1.00$), forecasting horizons (96 vs. 720 time points; $r = 0.99$), and datasets (ETTm2 vs. ETTm1; $r = 0.97$) suggesting that the model has learned robust data-driven importance scores.

Based on these results, for a detailed analysis, we averaged the importance scores across forecast horizons (96, 192, 336, and 720 time points), time series segments (left vs. right) and 2 seeds. We first evaluated the trends across different numbers of Wavelet bands (5 vs. 8 bands) and found consistent linear growth of importance scores over the first 5 bands (Table 5, $r = 0.98$). Lastly, we analyzed whether the learned importance scores correspond between datasets with different sampling rates. As in the Wavelet transformation frequency doubles in every subsequent band, the fourfold difference between the sampling rates of ETTm1 and ETTh1 datasets (15 minutes and 1 hour respectively) corresponded to the shift of two Wavelet bands. Accordingly, we found that the trends align between the importance scores of the two models when shifted by two positions ($r = 0.99$). Together, these results suggest that our model, PRISM, learns robust representations that hold across datasets, sampling rates, and hyperparameters.

4.6 Ablations

To identify the roles of the design choice in our model, we ablated the selection mechanism, time-frequency hierarchy, and time-frequency component parameterizations in our model (Table 6 and Table A.2). We found that our choices of the encoder and the routing scheme are optimal. For the filter banks, we found that replacing the Haar wavelets with FFT, difference-of-Gaussians, a binomial pyramid, EMA, or D-PAD's MCD block (Yuan & Chen, 2024) changes MSE by at most 5.08%; learning the filter scale increases MSE by 8.70%. Fully shared or level-specific scorers increased the MSE by 2.9% and 2.5%, respectively (Supplementary Tables A.9 and A.10). We also confirmed that hyperparameter sweeps favor the default band count, tree depth, and overlap size configurations: $K = 5$ and $D = 2$ each attain the lowest MSE in 14 of 16 settings, and $o = 8$ in 14 of 16 settings (Supplementary Tables A.11-A.13).

Table 6: **Ablations and architecture search.** The average MSE and MAE are reported for 6 datasets over 8 seeds. The context window is 336 for all datasets and the horizons are 96, 192, 336, 720 samples. Δ is the relative change with respect to PRISM; positive values indicate worse forecasts.

Ablations	Mean MSE	Mean MAE	ΔMSE (%)	ΔMAE (%)
PRISM	**0.306**	**0.343**	–	–
Encoder:				
Tree-level 1	0.326	0.356	+6.44%	+3.64%
MLP	0.408	0.410	+33.43%	+19.35%
Routing:				
Shared wts	0.315	0.349	+2.87%	+1.71%
No shared wts	0.314	0.348	+2.52%	+1.52%
No MLPs	0.337	0.367	+10.15%	+6.84%
Learnable threshold	0.338	0.373	+10.42%	+8.80%
Filter block:				
EMA	0.321	0.352	+5.08%	+2.65%
MCD	0.321	0.359	+4.84%	+4.52%
DoG	0.317	0.351	+3.72%	+2.40%
Binomial pyr	0.316	0.349	+3.24%	+1.80%
FFT	0.318	0.352	+3.80%	+2.59%
Learnable filt	0.333	0.360	+8.70%	+4.99%

5 Discussion

In this work, we have introduced PRISM - a time-series forecasting model that projects the context-window time-series data into a set of localized time-frequency time-series components and learns how to combine them through coupled hierarchical time-frequency routing towards generating time-series forecasts for the forecast windows. Ablations show that disrupting this selection mechanism degrades the performance of our model. Additionally we show that the learned predictive supports in our model sparsify without explicit sparsity regularization; prioritize time-frequency components of time series whose deletion impairs forecasting, and consistently shift when predictive contents changes. Together, these findings suggest that the value of multiscale representations of time-series data depends on how forecast-relevant information is selected and aggregated across temporal and spectral scales.

Limitations and future directions. PRISM is formulated for point forecasting, following the standard protocol of the benchmarks used in this field, and the present study characterizes the behavior of structured routing in that setting. This constrains the mechanism to be independent of how candidate components are produced and of what the forecasting head predicts, which opens several directions, including probabilistic and quantile objectives, exogenous covariates, and event streams. Because routing weights are computed per input, they also offer an interface for uses beyond forecasting accuracy, such as monitoring distribution shift or conditioning allocation on domain priors, which all can be done in separate future work.

6 AI DISCLOSURE

In this work, we used generative AI tools to aid or polish writing, for retrieval and discovery of related work, for generating the controlled synthetic datasets used in our routing analysis, and for writing and debugging experiment code, including baseline reproduction scripts, cluster job scripts, and the scripts that aggregate per-seed results and compute the reported statistical tests. We have not used generative AI tools for proving mathematical claims or for designing the PRISM architecture, and the remaining required-disclosure tasks are not applicable to this work. Additionally, we used generative AI tools to audit consistency between the manuscript and the implementation (e.g., checking that equations, feature definitions, and hyperparameters stated in the text match the code) and to cross-check reported tables against raw per-seed training logs. We have reviewed all AI-assisted work. Every reported number was produced by running the code on our own hardware and is traceable to raw training logs. AI-assisted experiment code was reviewed by author(s), and baseline reproductions were checked against the official implementations and their published hyperparameters. Significance tests were recomputed directly from per-seed results, and AI-suggested text edits were checked against the implementation before inclusion. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

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

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

Cristian Challu, Kin G Olivares, Boris N Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 6989–6997, 2023.

Peng Chen, Yingying Zhang, Yunyao Cheng, Yang Shu, Yihang Wang, Qingsong Wen, Bin Yang, and Chenjuan Guo. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 10527–10545, 2024.

Ronald R Coifman and M Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on information theory*, 38(2):713–718, 1992.

Ali Ebadi and Ebrahim Sahafizadeh. Comparing time-series analysis approaches utilized in research papers to forecast covid-19 cases in africa: A literature review. *arXiv preprint arXiv:2310.03606*, 2023.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pp. 95–104, 2018.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4):1748–1764, 2021.

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. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

Yuxiao Luo, Songming Zhang, Ziyu Lyu, and Yuhan Hu. Tfdnet: Time–frequency enhanced decomposed network for long-term time series forecasting. *Pattern Recognition*, 162:111412, 2025.

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.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Artyom Stitsyuk and Jaesik Choi. xpatch: Dual-stream time series forecasting with exponential seasonal-trend decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20601–20609, 2025.

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. *arXiv preprint arXiv:2405.14616*, 2024a.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. 2024b.

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.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.

Xiaobing Yuan and Ling Chen. D-pad: Deep-shallow multi-frequency patterns disentangling for time series forecasting. *arXiv preprint arXiv:2403.17814*, 2024.

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

Zichi Zhang, Tuan Dung Pham, Yimeng An, Ngoc Phu Doan, Majed Alsharari, Viet-Hung Tran, Anh-Tuan Hoang, Hans Vandierendonck, and Son T Mai. Waveletmixer: a multi-resolution wavelets based mlp-mixer for multivariate long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22741–22749, 2025.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in neural information processing systems*, 35:12677–12690, 2022a.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, pp. 27268–27286. PMLR, 2022b.

Jinan Zou, Qingying Zhao, Yang Jiao, Haiyao Cao, Yanxi Liu, Qingsen Yan, Ehsan Abbasnejad, Lingqiao Liu, and Javen Qinfeng Shi. Stock market prediction via deep learning techniques: A survey. *arXiv preprint arXiv:2212.12717*, 2022.

APPENDIX

A SUPPLEMENTARY DISCUSSION

A.1 MODEL DESIGN

The performance of our model is enabled by the architectural pairing with the Haar wavelets. Haar wavelets are hierarchical filters that decompose time series into components of different frequencies sampled at different times (Supplementary Figure A.1 left). The frequency filters start from the highest frequency afforded by the discrete digital data. Each subsequent filter is tuned to a frequency that is one half of the previous one, thus selecting exponentially spaced frequencies. Similarly, the Haar's wavelets provide a hierarchical organization in the time domain. While the initial highest-frequency filters are individually applied to the smallest time segments of the data, each subsequent filter is applied to segments twice as large. This way, the joint time-frequency decomposition with the Haar's wavelets always stays at the forefront of the time-frequency resolution.

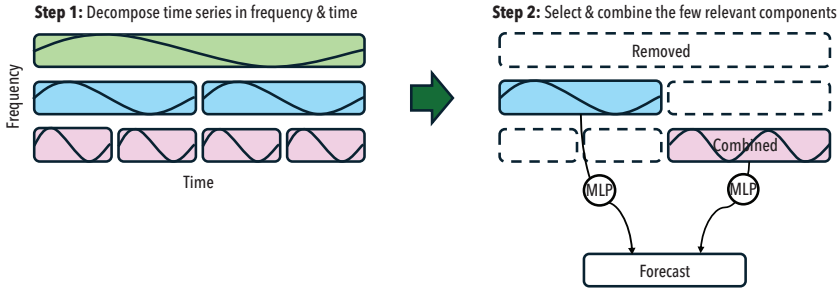

Figure A.1: **The `PRISM` model overview.**

While such a decomposition provides the most detailed view of the data, it creates an exponentially large number of time-frequency segments, many of which may be too detailed or irrelevant for forecasting. This has motivated our method as follows. Compared to the Haar's wavelets, our method, `PRISM`, has a complementary structure where time series are progressively split, through bisection, into smaller segments where each following segment is twice as short as the previous one. On each time decomposition level, the full Wavelet decomposition is applied to each temporal segment of the data, and the output Wavelet-generated components corresponding to the same frequency are concatenated in time. The process is iterated across all the levels of temporal decomposition of the original time series, thus producing all possible frequency components at all possible time decomposition levels. Thus, unlike in the original Haar's wavelets, the frequency resolution is no longer tied to the temporal resolution.

So far, such a procedure generates the number of the components that is even larger than that in the original Haar's wavelet transformation. To account for that fact, our method, `PRISM`, then only selects the time-frequency levels that are important for forecasting through a data-driven importance scoring procedure. In the limit, the model may fully suppress irrelevant time-frequency representations, thus going below the number of the components in the original Haar's wavelet transformation (Figure A.1 right). This way our approach, `PRISM`, through the architectural pairing with the Haar's wavelets, inherits its good properties and improves its performance though data-driven post-processing.

This parallel between the architecture of Haar's wavelets and the architecture of our method, `PRISM`, provides the insights into the model's optimal design choices, hyperparameters, and its learned importance scores. First, it explains why our model has achieved the best results when paired with Haar's wavelets (Table 6; Supplementary Table A.10): Other frequency features, including the FFT, have led to lesser results when paired with `PRISM` because they use different frequency banks corresponding, explicitly or implicitly through the Gabor's uncertainty principle, to different levels of temporal abstraction that are not naturally paired with the binary tree structure, used here in `PRISM`. For the Haar's wavelets, our model, `PRISM`, offers an improvement that can be explained by selecting from the (otherwise exponential) number of time-frequency components and combining them into a smaller number of task-relevant components. This allows for data-driven time-frequency tradeoff that may improve the model's forecasting performance by effectively reducing its number of

parameters and thus making it less prone to overfitting while keeping the flexibility to capture the relevant dimensions of the data.

A.2 Model hyperparameters

Second, the joint consideration of the Haar's wavelets and PRISM helps understanding the time-frequency segments selected by the model and thus inform the initial hyperparameter selection for new data and tasks. Unlike many other filter banks, Haar's wavelets start with extracting the highest frequency afforded by the data, then progressively recruiting lower frequencies but not necessarily requiring them. This design choice is substantiated because the sampling rates of time series data are typically chosen such that the data's finest resolution still contains useful information (and everything beyond it, being noise, is filtered out to save disk space). Accordingly, across the datasets in our experiments, the model retains non-negligible weight on the highest-frequency component rather than discarding it. On the other end of the spectrum, the lowest frequency that may be relevant for the time series forecasting is inversely proportional to the context window length T through Gabor's uncertainty principle. Considered together with the exponential spacing of Haar's wavelet frequencies, that suggests that the number of Haar's wavelets should not exceed $\log_2 T$. In the example of a context window of $T = 336$ time points considered in this work, the number of filters thus should not exceed $\log_2 336 \approx 8$. Aligned with this reasoning, we observed that, for the context window of $T = 336$ time points, 6 Wavelet components were consistently receiving high importance scores from PRISM, while subsequent filters got consistently suppressed. This corresponds to 6 filters leading to the model's best performance in our hyperparameter search (Supplementary Table A.11). Lastly, for the time domain decomposition, we have experimentally found that 2 levels of the temporal split were optimal across the tasks (Supplementary Table A.12). This may be explained by the model tracking the trend of the data through the differences between early and late time points. Overall, these considerations suggest that, for new datasets, a good initial performance should be expected with using Haar's filters up to the Gabor's limit over 2 temporal segments.

B Supplementary results

B.1 Comparison of training times

Besides the forecasting accuracy, another important property of time-series models is the training time. To evaluate how our model compares to other prominent models in terms of the runtime, we evaluated the training speed of several models on the ETTh1 dataset using the forecast window of 96 timepoints, one of the most popular evaluation settings among the datasets considered in this work. We followed the standard approach and recorded the times that were necessary for the training of multiple (ten) epochs to negate the effects of the training overhead. For our model, PRISM, the training of 10 epochs took 65 seconds. We found that, this way, our model has shown the two-fold improvement over its closest counterpart, the D-PAD, whose training took 141 seconds for 10 epochs. The next counterpart of our model, DLinear, was somewhat faster to train, taking 42 seconds per 10 epochs, however, this speedup came at the cost of the decreased average performance in the forecasting task (-2.95% MSE). Continuing this trend, iTransformer took 32 second per 10 epochs at the performance drop of 4.70% MSE. The fastest results were shown by PatchTST, N-HiTS, and xPatch with 12, 10, and 7 seconds per 10 epochs respectively, at the performance decrease of 1.70, 6.87, and 3.45% respectively. Overall, these results suggest that our model offers a reasonable time-accuracy tradeoff while PatchTST also fares well on this metric.

B.2 Comparison of inference compute

We have further analyzed the compute requirements for our model, PRISM, at inference and compared it to the requirements of other models. To this end, we have evaluated the forward passes of the batches of 8 samples through our model. We report the results below.

We found that PRISM's forecasting heads remained in the million-scale compute regime: the first layer of the tree (L0) uses 10 heads (level length 176) with 350 K parameters and an estimated 696 K FLOPs, and the second layer (L1) adds 20 heads (level length 96) with 496 K parameters and 983 K FLOPs, totaling 1.68 M FLOPs (Table A.1). We noticed that this number amounts to about 31%

Table A.1: Inference compute.

Model	FLOPs	Latency (ms)
PRISM (tree depth=1)	0.70 M	1.97
PRISM (tree depth=2)	1.68 M	4.10
D-PAD (num levels=3)	5.48 M	11.27
DLinear (channel-wise)	3.63 M	0.24
TimeMixer	729.31 M	1.34

of D-PAD's 5.48 M, an advantage that we attribute to PRISM's band remixing. Furthermore, we found that the low FLOPs count in PRISM is reflected in runtime: even with two tree levels (depth=2, overlap=8), the latency was only 4.10 ms, compared to 11.27 ms with D-PAD. We further performed a similar comparison with two more models, DLinear and TimeMixer. Even though DLinear has a comparable parameter count (679K), it required 3.63M FLOPs, which we have attributed to its per-channel linear mappings. Similarly, TimeMixer—also featuring a similar parameter count, amounting to 734K parameters—has incurred an estimated 729M FLOPs, which we attribute to its dense mixing/embedding layers. Overall, PRISM required substantially fewer FLOPs than all three considered models, while offering higher overall performance and built-in direct interpretability; DLinear and TimeMixer nonetheless achieve lower inference latency despite their higher FLOP counts.

To characterize how cost scales with context length, we separately profiled the use_last_layer_only variant of PRISM at $L \in \{336, 720, 1024, 2048\}$ (Table A.18). FLOPs, parameter count and peak memory grow roughly linearly up to $L = 1024$, while inference latency stays near 5 ms and rises sharply only at $L = 2048$. These measurements use a different configuration and profiling tool from Table A.1 and characterize scaling behaviour rather than the absolute cost of the full model.

B.3 Ablations

Ablation of the hierarchical selection mechanism. Table A.2 studies whether the gains of PRISM come from the proposed hierarchical selection mechanism rather than from multiscale feature construction alone. Replacing the proposed tree-structured selection with a flat multiscale variant causes the largest degradation, increasing MSE by 0.0177 (5.79%) and MAE by 0.0121 (3.53%). This shows that organizing predictive variables hierarchically is important for accurate forecasting. Removing explicit selection and relying on implicit mixing also hurts performance, increasing MSE by 4.18% and MAE by 2.07%, suggesting that dense mixing cannot fully substitute for learned predictive support allocation. Among the individual routing components, removing temporal selection, frequency selection, or branch selection all leads to consistent drops, with temporal selection producing the largest degradation among them ($+4.64\%$ MSE and $+2.42\%$ MAE). These results support the central design of PRISM: forecasting benefits not only from exposing multiscale time–frequency components, but from explicitly learning where predictive information should flow across temporal branches and frequency bands.

Table A.2: **Ablation for selection module.** We average 8 seeds for the 6 datasets (4 ETT datasets, Exchange, Weather) over 4 forecasting horizons $\{96, 192, 336, 720\}$ with input length 336. Mean errors and Δ are on the same basis as Table 6: Δ is measured against PRISM (MSE 0.306, MAE 0.343), so positive values indicate worse forecasts.

Ablation	Mean MSE	ΔMSE	ΔMSE (%)	Mean MAE	ΔMAE	ΔMAE (%)
PRISM	0.306	–	–	0.343	–	–
flat multiscale	0.324	0.0177	+5.79%	0.355	0.0121	+3.53%
implicit mixing	0.319	0.0128	+4.18%	0.350	0.0071	+2.07%
w/o branch selection	0.316	0.0098	+3.20%	0.349	0.0059	+1.72%
w/o frequency selection	0.320	0.0137	+4.48%	0.351	0.0074	+2.16%
w/o temporal selection	0.320	0.0142	+4.64%	0.351	0.0083	+2.42%

Table A.3: **Published, reproduced, and protocol-controlled baseline results at context length ($L = 336$).** Each cell reports MSE/MAE. PRISM, D-PAD, DLinear, TimeMixer, and XPatch reproduction results are 8-seed averages over the same seed set. For D-PAD, DLinear, and TimeMixer, we report both the published and reproduced values for transparency; arrows indicate whether the reproduced value is lower (↓) or higher (↑) than the published value. FITS and ModernTCN, marked by †, are additional protocol-controlled runs using context length 336, the same data splits and normalization procedure, and a single training seed. They are reported descriptively and are excluded from the multi-seed ranking and significance analyses. D-PAD published values are from Yuan & Chen (2024), and DLinear/TimeMixer published values are from Murad et al. (2025). Published XPatch values were unavailable in the referenced sources. Best value in each row in bold, second best underlined.

Dataset	H	PRISM (8 seeds)	D-PAD [1]		DLinear [2]		TimeMixer [2]		XPatch (8 seeds)	FITS†	ModernTCN†
			published	Reproduced	published	Reproduced	published	Reproduced			
ETTh1	96	**0.355/0.374**	0.357/0.376	0.369↑/0.394↑	0.375/0.399	0.372↓/0.396↓	0.385/0.408	0.389↑/0.413↑	0.377/0.398	0.375/0.396	0.375/0.396
	192	**0.390**/0.405	0.394/**0.402**	0.410↑/0.417↑	0.405/0.416	0.407↑/0.418↑	0.429/0.432	0.429/0.434↑	0.431/0.427	0.407/0.415	0.408/0.415
	336	0.386/0.407	**0.374/0.406**	0.395↑/0.418↑	0.439/0.443	0.439/0.443	0.456/0.450	0.443↓/0.446↓	0.461/0.443	0.430/0.428	0.434/0.432
	720	0.421/0.445	**0.419/0.442**	0.440↑/0.458↑	0.472/0.490	0.475↑/0.493↑	0.457/0.462	0.499↑/0.493↑	0.487/0.476	0.428/0.447	0.455/0.464
ETTh2	96	**0.261/0.322**	0.270/0.327	0.289↑/0.339↑	0.289/0.353	0.300↑/0.362↑	0.276/0.339	0.291↑/0.350↑	0.280/0.334	0.274/0.337	0.277/0.341
	192	**0.311/0.359**	0.331/0.368	0.343↑/0.377↑	0.383/0.418	0.375↓/0.411↓	0.340/0.381	0.353↑/0.392↑	0.345/0.377	0.336/0.376	0.344/0.388
	336	**0.308/0.361**	0.321/0.370	0.330↑/0.374↑	0.448/0.465	0.455↑/0.466↑	0.362/0.404	0.380↑/0.415↑	0.367/0.398	0.359/0.397	0.372/0.414
	720	0.387/0.420	**0.369/0.415**	0.406↑/0.430↑	0.605/0.551	0.694↑/0.592↑	0.398/0.433	0.437↑/0.455↑	0.393/0.426	0.386/0.423	0.407/0.439
ETTm1	96	**0.282/0.326**	0.285/0.328	0.296↑/0.341↑	0.299/0.343	0.300↑/0.344↑	0.303/0.350	0.299↓/0.350	0.289/0.333	0.305/0.346	0.295/0.346
	192	**0.323**/0.352	**0.323/0.349**	0.335↑/0.365↑	0.335/0.365	0.335/0.365	0.356/0.385	0.335↓/0.374↓	0.331/0.358	0.339/0.367	0.337/0.373
	336	0.354/0.377	**0.351/0.372**	0.364↑/0.386↑	0.369/0.386	0.372↑/0.389↑	0.366/0.392	0.371↑/0.398↑	0.365/0.381	0.374/0.386	0.373/0.395
	720	0.423/0.409	**0.412/0.405**	0.430↑/0.420↑	0.425/0.421	0.426↑/0.423↑	0.435/0.434	0.432↓/0.434	0.428/0.418	0.429/0.417	0.423/0.421
ETTm2	96	**0.158/0.234**	0.162/0.247	0.171↑/0.253↑	0.167/0.260	0.169↑/0.264↑	0.174/0.258	0.172↓/0.260↑	0.164/0.248	0.167/0.256	0.170/0.257
	192	**0.214/0.267**	0.218/0.283	0.233↑/0.293↑	0.224/0.303	0.237↑/0.316↑	0.238/0.300	0.231↓/0.300	0.221/0.287	0.222/0.293	0.223/0.294
	336	**0.264/0.319**	0.267/0.321	0.285↑/0.328↑	0.281/0.342	0.296↑/0.356↑	0.272/0.327	0.284↑/0.336↑	0.277/0.324	0.275/0.327	0.273/0.328
	720	**0.352/0.372**	0.353/**0.372**	0.367↑/0.382↑	0.397/0.421	0.422↑/0.437↑	0.368/0.389	0.361↓/0.386↓	0.358/0.376	0.368/0.383	0.372/0.389
Weather	96	**0.140/0.177**	0.143/0.181	0.147↑/0.188↑	0.176/0.237	0.174↓/0.235↓	0.158/0.204	0.150↓/0.203↓	0.150/0.187	0.145/0.194	0.149/0.204
	192	**0.187/0.227**	0.189/0.229	0.192↑/0.230↑	0.220/0.282	0.217↓/0.276↓	0.197/0.246	0.195↓/0.246	0.193/0.230	**0.187**/0.234	0.196/0.246
	336	**0.234/0.266**	0.239/0.268	0.243↑/0.272↑	0.265/0.319	0.261↓/0.311↓	0.242/0.281	0.248↑/0.286↑	0.245/0.271	0.240/0.276	0.245/0.284
	720	0.311/0.320	**0.304/0.313**	0.321↑/0.325↑	0.323/0.362	0.325↑/0.366↑	0.319/0.335	0.322↑/0.338↑	0.320/0.324	0.317/0.330	0.333/0.340
Exchange	96	**0.081/0.198**	–	0.096/0.218	**0.081**/0.203	0.084↑/0.205↑	0.119/0.247	0.094↓/0.218↓	0.084/0.203	0.087/0.207	0.167/0.303
	192	0.173/0.297	–	0.200/0.322	**0.157/0.293**	0.161↑/0.295↑	0.238/0.354	0.197↓/0.318↓	0.181/0.304	0.178/0.302	0.264/0.379
	336	0.314/**0.396**	–	0.376/0.446	**0.305**/0.414	0.315↑/0.425↑	0.417/0.472	0.372↓/0.444↓	0.353/0.426	0.324/0.415	0.439/0.498
	720	0.712/0.606	–	1.117/0.761	**0.643/0.601**	0.786↑/0.667↑	1.074/0.790	0.977↓/0.723↓	0.954/0.725	0.916/0.735	1.021/0.766

† FITS and ModernTCN are evaluated under a common context budget ($L = 336$) rather than their individually optimized published look-back settings. These runs provide a controlled same-protocol comparison rather than an exact reproduction of each method's reported best configuration.

Table A.4: **Per-dataset significance and robustness analysis between PRISM and baselines.** PRISM and each baseline are evaluated over eight matched runs ($n = 8$, $df = 7$), with seed-level performance averaged across the four forecasting horizons $\{96, 192, 336, 720\}$ at input length 336. For MSE and MAE, we report the mean performance difference (Δ) and its 95% confidence interval (CI_{95}). We additionally report two-sided paired t-test p-values with Benjamini–Hochberg FDR-corrected q-values computed over all 36 comparisons within each metric. For MSE, two-sided exact sign-test p- and q-values are reported as a complementary robustness check. The paired t-test uses the magnitudes of the eight paired differences, whereas the exact sign test uses only their directions. PRISM improves significantly at the 5% FDR level in 35 of 36 comparisons by MSE and 36 of 36 by MAE.

		MSE						MAE			
Dataset	Baseline	Δ	CI_{95}	Paired-t p	Paired-t q	Sign p	Sign q	Δ	CI_{95}	Paired-t p	Paired-t q
ETTh1	D-PAD	-0.0114	-0.0179→-0.0050	4.1e-3	5.1e-3	7.8e-3	7.8e-3	-0.0116	-0.0151→-0.0080	1.2e-4	1.7e-4
ETTh1	DLinear	-0.0328	-0.0464→-0.0192	7.0e-4	9.0e-4	8e-3	8e-3	-0.0291	-0.0391→-0.0192	2.0e-4	2.5e-4
ETTh1	SCINet	-0.2361	-0.2656→-0.2065	6.3e-9	3.8e-8	8e-3	8e-3	-0.1618	-0.1809→-0.1426	1.3e-9	4.8e-9
ETTh1	TimeMixer	-0.0494	-0.0636→-0.0353	1.5e-5	2.5e-5	8e-3	8e-3	-0.0377	-0.0479→-0.0275	1.4e-5	2.0e-5
ETTh1	WPMixer	-0.0209	-0.0350→-0.0069	8.0e-3	9.3e-3	8e-3	8e-3	-0.0144	-0.0244→-0.0043	1.1e-2	1.2e-2
ETTh1	XPatch	-0.0486	-0.0635→-0.0338	1.3e-5	2.1e-5	8e-3	8e-3	-0.0277	-0.0379→-0.0175	1.5e-4	2.1e-4
ETTh2	D-PAD	-0.0230	-0.0244→-0.0217	1.4e-9	1.7e-8	7.8e-3	7.8e-3	-0.0138	-0.0144→-0.0133	1.5e-10	6.8e-10
ETTh2	DLinear	-0.1373	-0.1498→-0.1247	3.2e-8	9.8e-8	8e-3	8e-3	-0.0918	-0.0983→-0.0853	5.2e-9	1.6e-8
ETTh2	SCINet	-3.0278	-3.6121→-2.4435	5.5e-6	1.0e-5	8e-3	8e-3	-1.0322	-1.1509→-0.9135	1.6e-7	3.1e-7
ETTh2	TimeMixer	-0.0467	-0.0511→-0.0422	3.3e-8	9.8e-8	8e-3	8e-3	-0.0371	-0.0393→-0.0349	7.3e-10	2.9e-9
ETTh2	WPMixer	-0.0241	-0.0261→-0.0220	5.1e-9	3.7e-8	8e-3	8e-3	-0.0214	-0.0233→-0.0195	1.1e-8	3.0e-8
ETTh2	XPatch	-0.0276	-0.0295→-0.0257	8.3e-10	1.5e-8	8e-3	8e-3	-0.0177	-0.0184→-0.0169	2.0e-13	7.2e-12
ETTm1	D-PAD	-0.0082	-0.0090→-0.0075	3.0e-8	9.8e-8	7.8e-3	7.8e-3	-0.0107	-0.0117→-0.0098	2.8e-8	6.1e-8
ETTm1	DLinear	-0.0101	-0.0109→-0.0093	1.0e-13	3.6e-12	8e-3	8e-3	-0.0129	-0.0140→-0.0118	8.9e-13	1.6e-11
ETTm1	SCINet	-0.3073	-0.3559→-0.2587	1.4e-6	2.7e-6	8e-3	8e-3	-0.2116	-0.2370→-0.1862	2.1e-7	3.8e-7
ETTm1	TimeMixer	-0.0109	-0.0127→-0.0090	6.2e-7	1.4e-6	8e-3	8e-3	-0.0214	-0.0229→-0.0198	6.3e-12	7.5e-11
ETTm1	WPMixer	-0.0039	-0.0048→-0.0029	1.3e-6	2.5e-6	8e-3	8e-3	-0.0098	-0.0107→-0.0089	1.3e-11	9.6e-11
ETTm1	XPatch	-0.0053	-0.0062→-0.0043	1.7e-8	6.8e-8	8e-3	8e-3	-0.0053	-0.0062→-0.0044	2.2e-8	5.2e-8
ETTm2	D-PAD	-0.0151	-0.0162→-0.0139	9.7e-9	5.0e-8	7.8e-3	7.8e-3	-0.0094	-0.0102→-0.0085	2.9e-8	6.1e-8
ETTm2	DLinear	-0.0320	-0.0394→-0.0245	1.9e-5	2.8e-5	8e-3	8e-3	-0.0385	-0.0441→-0.0329	6.4e-7	1.0e-6
ETTm2	SCINet	-1.7312	-1.9505→-1.5120	3.1e-7	7.5e-7	8e-3	8e-3	-0.6166	-0.6600→-0.5732	5.3e-9	1.6e-8
ETTm2	TimeMixer	-0.0132	-0.0152→-0.0113	1.7e-7	4.4e-7	8e-3	8e-3	-0.0162	-0.0175→-0.0149	1.2e-11	9.6e-11
ETTm2	WPMixer	-0.0093	-0.0106→-0.0080	1.4e-8	6.1e-8	8e-3	8e-3	-0.0094	-0.0105→-0.0083	1.2e-10	6.4e-10
ETTm2	XPatch	-0.0062	-0.0074→-0.0051	7.8e-8	2.2e-7	8e-3	8e-3	-0.0041	-0.0049→-0.0033	4.5e-8	8.9e-8
Exchange	D-PAD	-0.1232	-0.1315→-0.1149	3.9e-9	3.5e-8	7.8e-3	7.8e-3	-0.0573	-0.0618→-0.0528	1.2e-8	3.1e-8
Exchange	DLinear	-0.0127	-0.0369→0.0114	0.256	0.256	8e-3	8e-3	-0.0186	-0.0310→-0.0062	8.4e-3	9.5e-3
Exchange	SCINet	-0.9812	-1.7736→-0.1888	0.0221	0.0249	8e-3	8e-3	-0.2792	-0.4206→-0.1379	2.3e-3	2.8e-3
Exchange	TimeMixer	-0.0862	-0.1070→-0.0654	1.7e-5	2.6e-5	8e-3	8e-3	-0.0464	-0.0547→-0.0380	2.9e-7	5.0e-7
Exchange	WPMixer	-0.0625	-0.0820→-0.0431	9.2e-5	1.3e-4	8e-3	8e-3	-0.0325	-0.0408→-0.0242	6.1e-6	9.6e-6
Exchange	XPatch	-0.0692	-0.0816→-0.0568	7.7e-7	1.6e-6	8e-3	8e-3	-0.0353	-0.0397→-0.0308	1.1e-10	6.3e-10
Weather	D-PAD	-0.0053	-0.0070→-0.0036	1.6e-4	2.2e-4	7.8e-3	7.8e-3	-0.0033	-0.0053→-0.0013	6.1e-3	7.1e-3
Weather	DLinear	-0.0237	-0.0251→-0.0223	1.4e-9	1.3e-8	8e-3	8e-3	-0.0469	-0.0491→-0.0447	2.8e-10	1.1e-9
Weather	SCINet	-0.0524	-0.0748→-0.0300	4.4e-4	5.9e-4	8e-3	8e-3	-0.0521	-0.0702→-0.0341	1.6e-4	2.1e-4
Weather	TimeMixer	-0.0079	-0.0091→-0.0067	1.2e-6	2.5e-6	8e-3	8e-3	-0.0178	-0.0194→-0.0162	3.4e-8	6.8e-8
Weather	WPMixer	-0.0026	-0.0037→-0.0014	1.1e-3	1.2e-3	8e-3	8e-3	-0.0067	-0.0082→-0.0053	1.1e-5	1.6e-5
Weather	XPatch	-0.0063	-0.0077→-0.0049	1.5e-5	2.3e-5	8e-3	8e-3	-0.0030	-0.0046→-0.0013	4.1e-3	4.4e-3

Table A.5: **Per-dataset significance comparison between PRISM and Transformer baselines**.

Dataset	Baseline	ΔMSE	CI_{95} MSE	p_{MSE}	q_{MSE}	ΔMAE	CI_{95} MAE	p_{MAE}	q_{MAE}
ETTh1	Autoformer	-0.0815	-0.0838→-0.0792	7.117e-9	1.423e-8	-0.0668	-0.0681→-0.0656	5.042e-10	2.017e-9
ETTh1	FEDformer	-0.0362	-0.0386→-0.0337	4.326e-8	5.768e-8	-0.0432	-0.0445→-0.042	8.527e-11	2.274e-10
ETTh1	Pyraformer	-0.4977	-0.5→-0.4954	1.203e-11	4.811e-11	-0.3223	-0.3236→-0.3211	1.694e-13	1.355e-12
ETTh1	TimesNet	-0.0909	-0.0932→-0.0886	2.168e-9	5.78e-9	-0.0691	-0.0704→-0.0678	3.694e-10	1.478e-9
ETTh2	Autoformer	-0.1028	-0.1115→-0.0941	3.992e-7	5.323e-7	-0.0771	-0.0826→-0.0717	1.421e-7	2.842e-7
ETTh2	FEDformer	-0.0684	-0.0772→-0.0597	1.452e-6	1.66e-6	-0.0666	-0.0721→-0.0612	1.814e-7	2.904e-7
ETTh2	Pyraformer	-3.5314	-3.5402→-3.5227	1.45e-13	1.16e-12	-1.2134	-1.2189→-1.2079	9.638e-13	2.57e-12
ETTh2	TimesNet	-0.0911	-0.0999→-0.0824	8.545e-7	9.766e-7	-0.0762	-0.0817→-0.0708	2.131e-7	2.842e-7
ETTm1	Autoformer	-0.1678	-0.1693→-0.1662	2.555e-13	1.022e-12	-0.1265	-0.1277→-0.1254	2.262e-12	3.016e-12
ETTm1	FEDformer	-0.0348	-0.0363→-0.0332	9.069e-9	1.451e-8	-0.0548	-0.056→-0.0537	4.121e-11	1.649e-10
ETTm1	Pyraformer	-0.3414	-0.343→-0.3399	8.468e-12	2.258e-11	-0.223	-0.2242→-0.2219	1.293e-11	2.586e-11
ETTm1	TimesNet	-0.1002	-0.1017→-0.0987	1.222e-10	2.444e-10	-0.0649	-0.0661→-0.0638	1.113e-8	2.967e-8
ETTm2	Autoformer	-0.0607	-0.0632→-0.0581	3.415e-7	5.123e-7	-0.0528	-0.0551→-0.0506	1.804e-8	3.609e-8
ETTm2	FEDformer	-0.0421	-0.0446→-0.0395	8.446e-8	1.352e-7	-0.0391	-0.0414→-0.0368	3.094e-8	4.952e-8
ETTm2	Pyraformer	-1.2651	-1.2676→-1.2625	1.006e-10	2.012e-10	-0.6043	-0.6066→-0.602	1.314e-10	3.504e-10
ETTm2	TimesNet	-0.0391	-0.0417→-0.0366	2.094e-7	3.351e-7	-0.0316	-0.0338→-0.0293	5.532e-8	7.376e-8
Exchange	Autoformer	-1.0075	-1.0835→-0.9316	4.911e-8	6.548e-8	-0.5431	-0.5686→-0.5175	3.648e-8	4.952e-8
Exchange	FEDformer	-0.457	-0.533→-0.3811	1.926e-6	1.926e-6	-0.2678	-0.2934→-0.2423	2.041e-7	2.842e-7
Exchange	Pyraformer	-0.8597	-0.9357→-0.7838	1.116e-7	1.488e-7	-0.5111	-0.5366→-0.4856	4.833e-8	6.905e-8
Exchange	TimesNet	-0.3442	-0.4202→-0.2682	2.221e-5	2.221e-5	-0.1904	-0.2159→-0.1648	1.964e-6	2.619e-6
Weather	Autoformer	-0.1137	-0.1202→-0.1071	2.385e-9	5.78e-9	-0.1291	-0.1314→-0.1268	3.834e-10	1.478e-9
Weather	FEDformer	-0.0892	-0.0958→-0.0827	1.72e-8	2.293e-8	-0.1068	-0.1091→-0.1045	8.911e-10	2.037e-9
Weather	Pyraformer	-0.0838	-0.0903→-0.0772	6.282e-8	8.376e-8	-0.1136	-0.1159→-0.1113	2.174e-10	1.159e-9
Weather	TimesNet	-0.0288	-0.0353→-0.0222	3.24e-5	3.24e-5	-0.0368	-0.0391→-0.0345	1.602e-7	2.842e-7

Table A.6: **Representative GIFT datasets and artifact rationales.** The GIFT benchmark contains nearly 100 datasets. This table lists representative datasets used to define the artifact groups discussed in the main text. Groups are non-exclusive, and a dataset may appear in multiple groups. The listed datasets cover all 97 evaluation settings reported in our GIFT evaluation. Among them, 93 settings are associated with at least one artifact group, while the four *M_DENSE* settings are listed as uncategorized because we do not find sufficient dataset-level evidence for assigning them to a specific artifact group.

Artifact group	Dataset	Dataset-level rationale
Irregular	*bitbrains*	Cloud workload traces with asynchronous VM sensors.
Irregular	*bizitobs*	Business telemetry with event-driven metrics.
Irregular	*solar/10T*	Solar sensor data where outages can create irregular intervals.
Irregular	*loop_seattle*	Traffic loop data with packet loss or clock drift.
Aperiodic	*exchange-rate*	Currency exchange rates with stochastic changes and weak fixed seasonality.
Aperiodic	*bitbrains*	Cloud workload traces dominated by random events or regime changes.
Aperiodic	*bizitobs*	Event-driven business telemetry with disrupted periodicity.
Aperiodic	*solar/H*	Hourly solar irradiance with short-term fluctuations.
Aperiodic	*hospital*	Event-driven hospital records with nonstationary behavior.
Aperiodic	*covid_deaths*	Pandemic mortality series with regime changes.
Aperiodic	*us_births*	Event-driven birth-count series with nonstationary variation.
Aperiodic	*kdd_cup_2018*	Network KPI series with bursty events.
Incomplete	*solar*	Solar data with missing observations or unavailable covariates.
Incomplete	*electricity*	Electricity consumption records with partial observations.
Incomplete	*weather/jena*	Weather station data with missing values.
Incomplete	*loop_seattle*	Traffic loop data with packet loss.
Incomplete	*traffic*	Traffic sensor data with missing urban sensor readings.
Incomplete	*sz_taxi*	Urban taxi/traffic series with missing sensor readings.
Incomplete	*bizitobs*	Business telemetry with irregular ingestion logs.
Incomplete	*car_parts_with_missing*	Car-parts demand series explicitly containing missing observations.
Anomalous	*traffic/loop_seattle*	Traffic data with transient spikes or sensor faults.
Anomalous	*sz_taxi*	Urban taxi/traffic data with abnormal rush-hour spikes.
Anomalous	*m4*	Forecasting benchmark with holiday spikes and transient shocks.
Anomalous	*solar/H/short*	Short hourly solar series with cloud-passage spikes.
Anomalous	*solar/W/short*	Short weekly solar series with cloud-passage spikes.
Anomalous	*bitbrains_rnd*	Randomized cloud workload traces with burst loads.
Anomalous	*bizitobs_service*	Service-level business telemetry with burst loads.
Anomalous	*exchange-rate*	Currency exchange rates with abrupt financial jumps.
Anomalous	*crypto*	Cryptocurrency series with financial jumps.
Nonstationary	*ett1*	Electricity transformer temperature data with day/night and seasonal effects.
Nonstationary	*ett2*	Electricity transformer temperature data from another plant with distribution shifts.
Nonstationary	*electricity*	Electricity demand data with changing consumption patterns.
Nonstationary	*jena_weather*	Weather data with seasonal and distributional changes.
Nonstationary	*traffic/loop_seattle*	Traffic data with rush-hour pattern drift.
Nonstationary	*sz_taxi*	Taxi/traffic data with rush-hour pattern drift.
Nonstationary	*solar*	Solar data with daily-to-seasonal irradiance shifts.
Nonstationary	*us_births*	Birth-count series with demographic or policy drift.
Nonstationary	*hospital*	Hospital records with demographic or policy changes.
Nonstationary	*hierarchical_sales*	Hierarchical retail sales series with time-varying demand patterns across aggregation levels.
Nonstationary	*restaurant*	Restaurant demand series with time-varying consumption patterns.
Drift	*weather/jena_weather*	Weather station data with climate drift.
Drift	*temperature_rain*	Climate-related temperature and rainfall series with slow drift.
Drift	*saugeen*	Hydrological or climate series with long-term environmental drift.
Drift	*electricity/H/long*	Long-horizon electricity data with economic or demand growth trends.
Drift	*energy/ETT-long*	Long energy/ETT series with persistent trend movement.
Drift	*us_births*	Birth-count series with population or demographic phase changes.
Drift	*covid_deaths*	Pandemic mortality series with phase changes.
Drift	*traffic/METR-LA*	Traffic sensor data with infrastructure evolution or mobility drift.
Drift	*sz_taxi*	Taxi/traffic data with long-term mobility drift.
Uncategorized	*M_DENSE*	Dense multivariate time series without a clear assignment to the artifact groups above.

Table A.7: Extended results (GIFT dataset).*Note:* Extreme DLinear MASE values (e.g. BITBRAINS) come from NaN-fallback items whose near-flat history drives the seasonal-error denominator toward zero; GIFT-Eval's MASE applies no epsilon floor or item exclusion, so a few items dominate the average while MAE stays comparable. We report these values as produced by the official protocol.

Dataset	Term	PRISM			D-PAD			DLinear			FITS			ModernTCN		
		MSE	MAE	MASE	MSE	MAE	MASE	MSE	MAE	MASE	MSE	MAE	MASE	MSE	MAE	MASE
LOOP_SEATTLE/5T	short	**6.334E+01**	**4.687E+00**	**7.383E-01**	6.950E+01	5.432E+00	8.718E-01	8.143E+01	5.813E+00	9.054E-01	7.713E+01	5.492E+00	8.530E-01	6.421E+01	5.187E+00	8.256E-01
LOOP_SEATTLE/5T	medium	**8.485E+01**	**5.682E+00**	**8.985E-01**	8.770E+01	6.050E+00	9.563E-01	1.002E+02	6.463E+00	1.023E+00	9.859E+01	6.424E+00	1.011E+00	9.356E+01	6.412E+00	1.013E+00
LOOP_SEATTLE/5T	long	**9.115E+01**	**5.890E+00**	**9.284E-01**	9.141E+01	6.027E+00	9.506E-01	9.973E+01	6.275E+00	9.940E-01	9.818E+01	6.135E+00	9.666E-01	9.918E+01	6.501E+00	1.027E+00
LOOP_SEATTLE/D	short	**1.977E+01**	3.121E+00	9.368E-01	2.135E+01	3.205E+00	9.597E-01	1.978E+01	**3.097E+00**	**9.340E-01**	2.081E+01	3.226E+00	9.704E-01	2.054E+01	3.168E+00	9.535E-01
LOOP_SEATTLE/H	short	**7.418E+01**	**4.895E+00**	**9.915E-01**	7.647E+01	5.350E+00	1.085E+00	8.239E+01	5.405E+00	1.095E+00	8.199E+01	5.391E+00	1.091E+00	7.554E+01	5.170E+00	1.046E+00
LOOP_SEATTLE/H	medium	**6.770E+01**	**4.653E+00**	**9.546E-01**	7.236E+01	5.124E+00	1.051E+00	7.106E+01	5.068E+00	1.046E+00	7.095E+01	5.053E+00	1.036E+00	7.226E+01	5.088E+00	1.043E+00
LOOP_SEATTLE/H	long	**6.139E+01**	**4.474E+00**	**9.330E-01**	6.788E+01	5.055E+00	1.051E+00	6.326E+01	4.826E+00	1.018E+00	6.473E+01	4.853E+00	1.011E+00	6.787E+01	5.048E+00	1.050E+00
M_DENSE/D	short	**9.805E+03**	**4.785E+01**	**7.207E-01**	1.729E+04	7.008E+01	9.185E-01	1.618E+04	7.250E+01	1.019E+00	1.886E+04	7.686E+01	1.076E+00	2.106E+04	8.192E+01	1.137E+00
M_DENSE/H	short	**5.832E+04**	**1.081E+02**	**9.501E-01**	6.027E+04	1.203E+02	1.076E+00	6.783E+04	1.273E+02	1.096E+00	7.061E+04	1.298E+02	1.108E+00	6.384E+04	1.200E+02	1.048E+00
M_DENSE/H	medium	**3.629E+04**	**8.222E+01**	**7.210E-01**	4.643E+04	1.102E+02	9.319E-01	4.146E+04	9.888E+01	8.445E-01	4.481E+04	1.024E+02	8.593E-01	4.653E+04	1.097E+02	9.307E-01
M_DENSE/H	long	**3.551E+04**	**8.384E+01**	**7.298E-01**	3.705E+04	9.326E+01	7.881E-01	5.919E+04	1.250E+02	1.026E+00	6.253E+04	1.254E+02	1.017E+00	6.316E+04	1.300E+02	1.081E+00
SZ_TAXI/15T	short	1.796E+01	**2.863E+00**	**5.729E-01**	1.817E+01	2.927E+00	5.872E-01	1.780E+01	2.905E+00	5.832E-01	1.780E+01	2.890E+00	5.800E-01	**1.777E+01**	2.887E+00	5.788E-01
SZ_TAXI/15T	medium	**1.692E+01**	2.871E+00	5.611E-01	1.732E+01	2.908E+00	5.689E-01	1.733E+01	2.914E+00	5.690E-01	1.694E+01	**2.870E+00**	**5.604E-01**	1.805E+01	2.979E+00	5.817E-01
SZ_TAXI/15T	long	**1.801E+01**	**2.914E+00**	**5.551E-01**	2.474E+01	3.484E+00	6.656E-01	3.131E+01	3.874E+00	7.631E-01	1.965E+01	3.158E+00	5.996E-01	2.657E+01	3.642E+00	6.987E-01
SZ_TAXI/H	short	**7.699E+00**	**1.919E+00**	**5.845E-01**	8.220E+00	2.009E+00	6.194E-01	8.457E+00	2.048E+00	6.300E-01	8.011E+00	1.973E+00	6.031E-01	7.954E+00	1.967E+00	6.024E-01
BITBRAINS_FAST_STORAGE/5T	short	**1.641E+06**	**1.900E+02**	**1.403E+00**	1.673E+06	1.944E+02	1.752E+00	1.705E+06	2.005E+02	1.064E+02	1.947E+06	2.039E+02	2.136E+00	2.156E+06	2.171E+02	2.599E+00
BITBRAINS_FAST_STORAGE/5T	medium	**3.011E+06**	3.525E+02	2.111E+00	3.019E+06	**3.473E+02**	**1.538E+00**	4.191E+06	4.434E+02	4.880E+00	4.098E+06	4.314E+02	1.808E+00	4.770E+06	4.710E+02	4.581E+00
BITBRAINS_FAST_STORAGE/5T	long	5.516E+06	5.045E+02	1.819E+00	**4.435E+06**	**4.347E+02**	**1.316E+00**	5.551E+06	5.143E+02	2.852E+00	5.532E+06	5.156E+02	1.719E+00	5.830E+06	5.378E+02	4.172E+00
BITBRAINS_FAST_STORAGE/H	short	**3.004E+06**	**3.403E+02**	1.771E+00	3.565E+06	3.701E+02	**1.712E+00**	3.888E+06	4.221E+02	5.682E+01	3.292E+06	3.967E+02	1.739E+00	3.790E+06	4.100E+02	2.576E+00
BITBRAINS_RND/5T	short	**1.236E+06**	**1.341E+02**	**2.563E+00**	1.287E+06	1.395E+02	3.425E+00	1.478E+06	1.607E+02	2.714E+02	1.550E+06	1.631E+02	3.310E+00	1.441E+06	1.672E+02	4.037E+00
BITBRAINS_RND/5T	medium	**2.243E+06**	**2.489E+02**	5.782E+00	3.516E+06	2.886E+02	**5.268E+00**	2.263E+06	2.672E+02	1.612E+01	2.601E+06	2.816E+02	5.484E+00	3.404E+06	3.460E+02	7.378E+00
BITBRAINS_RND/5T	long	2.692E+06	2.637E+02	4.413E+00	**2.054E+06**	**2.245E+02**	**4.137E+00**	2.643E+06	2.898E+02	2.329E+01	2.852E+06	2.994E+02	4.489E+00	3.353E+06	3.443E+02	6.488E+00
BITBRAINS_RND/H	short	**2.071E+06**	2.399E+02	**6.674E+00**	2.073E+06	**2.372E+02**	6.766E+00	2.229E+06	2.645E+02	2.132E+02	2.138E+06	2.450E+02	6.699E+00	2.216E+06	2.677E+02	7.383E+00
bizitobs_application	short	**2.385E+06**	**6.610E+02**	6.958E+00	3.061E+06	9.999E+02	9.289E+00	7.042E+06	1.390E+03	4.843E+00	4.152E+06	9.176E+02	**4.788E+00**	2.597E+06	8.580E+02	7.214E+00
bizitobs_application	medium	5.184E+06	1.247E+03	5.468E+00	3.486E+06	1.074E+03	6.666E+00	7.305E+06	1.582E+03	3.181E+00	**2.833E+06**	**9.387E+02**	**2.081E+00**	4.346E+06	1.145E+03	2.600E+00
bizitobs_application	long	1.203E+07	2.171E+03	6.227E+00	7.529E+06	1.663E+03	6.624E+00	1.401E+07	2.338E+03	4.141E+00	**5.483E+06**	**1.445E+03**	**3.054E+00**	1.260E+07	2.175E+03	4.294E+00
BIZITOBS_L2C/5T	short	1.544E+01	**2.206E+00**	**2.285E-01**	1.671E+01	2.473E+00	2.534E-01	1.692E+01	2.422E+00	2.528E-01	1.714E+01	2.440E+00	2.519E-01	**1.474E+01**	2.243E+00	2.334E-01
BIZITOBS_L2C/5T	medium	7.802E+01	**5.718E+00**	**5.579E-01**	8.574E+01	5.980E+00	5.954E-01	1.496E+02	8.043E+00	7.703E-01	1.502E+02	8.230E+00	7.939E-01	**6.998E+01**	5.823E+00	5.973E-01
BIZITOBS_L2C/5T	long	6.944E+01	5.145E+00	5.043E-01	6.447E+01	**5.002E+00**	**4.947E-01**	6.737E+01	5.371E+00	5.369E-01	6.380E+01	5.224E+00	5.270E-01	**5.655E+01**	5.079E+00	5.295E-01
BIZITOBS_L2C/H	short	6.580E+01	5.340E+00	5.409E-01	**5.926E+01**	**5.283E+00**	**5.378E-01**	7.983E+01	5.792E+00	5.831E-01	8.277E+01	6.115E+00	6.132E-01	6.808E+01	5.473E+00	5.518E-01
BIZITOBS_L2C/H	medium	2.309E+02	1.171E+01	1.143E+00	2.473E+02	1.200E+01	1.192E+00	2.439E+02	1.216E+01	1.189E+00	2.640E+02	1.215E+01	1.180E+00	**1.921E+02**	**1.002E+01**	**1.004E+00**
BIZITOBS_L2C/H	long	**2.440E+02**	1.202E+01	1.211E+00	2.665E+02	1.255E+01	1.259E+00	2.500E+02	1.221E+01	1.228E+00	3.197E+02	1.323E+01	1.326E+00	2.484E+02	**1.162E+01**	**1.181E+00**
bizitobs_service	short	1.757E+04	**3.456E+01**	**1.541E+00**	2.157E+04	4.160E+01	1.896E+00	2.000E+04	4.506E+01	3.167E+00	1.775E+04	4.445E+01	2.650E+00	**1.626E+04**	5.208E+01	2.919E+00
bizitobs_service	medium	1.004E+05	6.719E+01	1.433E+00	3.958E+04	5.523E+01	1.728E+00	4.360E+04	4.898E+01	2.902E+00	**3.515E+04**	**3.876E+01**	**1.092E+00**	3.669E+04	5.537E+01	2.502E+00
bizitobs_service	long	**1.319E+05**	7.777E+01	**1.493E+00**	1.320E+05	7.967E+01	1.682E+00	1.376E+05	8.906E+01	1.893E+00	1.337E+05	**7.622E+01**	1.648E+00	1.397E+05	8.248E+01	2.162E+00
covid_deaths	short	2.664E+06	2.553E+02	**3.577E+01**	3.552E+06	3.415E+02	4.882E+01	4.043E+07	1.217E+03	1.382E+02	2.921E+06	3.467E+02	5.478E+01	**1.921E+06**	**2.504E+02**	4.012E+01
ELECTRICITY/15T	short	**2.301E+05**	**7.528E+01**	**1.385E+00**	2.521E+05	8.804E+01	1.660E+00	2.878E+05	8.932E+01	1.631E+00	2.941E+05	8.858E+01	1.617E+00	2.893E+05	8.895E+01	1.616E+00
ELECTRICITY/15T	medium	**2.772E+05**	**6.006E+01**	**9.694E-01**	3.959E+05	8.188E+01	1.420E+00	3.513E+05	7.364E+01	1.185E+00	3.470E+05	7.044E+01	1.114E+00	5.796E+05	9.443E+01	1.482E+00
ELECTRICITY/15T	long	**3.907E+05**	**6.680E+01**	**1.011E+00**	5.830E+05	9.076E+01	1.452E+00	6.716E+05	9.869E+01	1.472E+00	5.139E+05	8.215E+01	1.237E+00	1.514E+06	1.351E+02	1.945E+00
ELECTRICITY/D	short	2.845E+09	5.709E+03	**1.761E+00**	3.723E+09	6.920E+03	2.151E+00	2.673E+09	**5.544E+03**	1.777E+00	2.881E+09	5.900E+03	1.878E+00	**2.563E+09**	5.770E+03	1.874E+00
ELECTRICITY/H	short	**1.801E+06**	**2.046E+02**	**1.156E+00**	2.098E+06	2.621E+02	1.599E+00	2.062E+06	2.335E+02	1.298E+00	2.163E+06	2.399E+02	1.342E+00	2.187E+06	2.469E+02	1.389E+00
ELECTRICITY/H	medium	**6.188E+06**	**2.632E+02**	**1.188E+00**	6.193E+06	3.324E+02	1.664E+00	1.089E+07	4.281E+02	1.940E+00	7.991E+06	3.240E+02	1.457E+00	1.318E+07	4.197E+02	1.852E+00
ELECTRICITY/H	long	**7.871E+06**	**2.958E+02**	**1.334E+00**	8.989E+06	3.783E+02	1.914E+00	1.693E+07	4.488E+02	1.871E+00	1.130E+07	3.905E+02	1.724E+00	1.525E+07	4.879E+02	2.267E+00
ELECTRICITY/W	short	**1.359E+11**	**2.315E+04**	**5.973E+00**	1.590E+12	3.138E+05	2.000E+00	2.610E+11	5.203E+04	1.048E+00	2.206E+11	5.039E+04	2.024E+00	2.044E+11	4.809E+04	2.079E+00
ETT1/15T	short	6.057E+00	**1.264E+00**	8.348E-01	**5.683E+00**	1.277E+00	8.866E-01	6.298E+00	1.294E+00	8.358E-01	6.425E+00	1.302E+00	**8.263E-01**	6.342E+00	1.325E+00	8.470E-01
ETT1/15T	medium	1.079E+01	1.858E+00	1.188E+00	1.195E+01	1.848E+00	1.164E+00	9.072E+00	1.645E+00	1.045E+00	**8.944E+00**	**1.623E+00**	**1.034E+00**	9.599E+00	1.714E+00	1.099E+00
ETT1/15T	long	1.027E+01	1.814E+00	1.223E+00	1.101E+01	1.840E+00	1.215E+00	8.859E+00	1.721E+00	1.130E+00	**8.125E+00**	**1.632E+00**	**1.073E+00**	1.134E+01	1.975E+00	1.292E+00
ETT1/D	short	**3.803E+04**	**1.294E+02**	**1.652E+00**	4.761E+04	1.526E+02	2.050E+00	4.054E+04	1.428E+02	1.945E+00	3.805E+04	1.363E+02	1.763E+00	4.235E+04	1.377E+02	1.710E+00
ETT1/H	short	1.060E+02	5.176E+00	**8.754E-01**	**9.654E+01**	5.303E+00	9.196E-01	1.086E+02	5.302E+00	8.905E-01	1.097E+02	5.297E+00	8.938E-01	9.895E+01	**5.172E+00**	8.795E-01
ETT1/H	medium	**1.458E+02**	**7.199E+00**	**1.320E+00**	1.866E+02	8.087E+00	1.498E+00	1.563E+02	7.544E+00	1.377E+00	1.536E+02	7.532E+00	1.393E+00	1.623E+02	7.826E+00	1.436E+00
ETT1/H	long	**1.587E+02**	**7.510E+00**	**1.399E+00**	2.025E+02	9.022E+00	1.766E+00	1.590E+02	7.626E+00	1.408E+00	1.631E+02	7.746E+00	1.434E+00	1.964E+02	8.556E+00	1.542E+00
ETT1/W	short	**1.902E+06**	**9.778E+02**	**1.533E+00**	2.017E+06	1.028E+03	1.567E+00	1.983E+06	1.013E+03	1.655E+00	2.673E+06	1.160E+03	1.726E+00	2.666E+06	1.134E+03	1.687E+00
ETT2/15T	short	1.003E+01	1.993E+00	8.669E-01	**9.491E+00**	**1.948E+00**	**8.465E-01**	1.062E+01	2.098E+00	9.002E-01	1.037E+01	2.076E+00	8.929E-01	9.910E+00	2.034E+00	8.766E-01
ETT2/15T	medium	**1.098E+01**	**2.104E+00**	8.947E-01	1.394E+01	2.399E+00	1.014E+00	1.108E+01	2.118E+00	8.858E-01	1.123E+01	2.125E+00	**8.828E-01**	1.264E+01	2.288E+00	9.742E-01
ETT2/15T	long	1.423E+01	2.428E+00	1.013E+00	1.461E+01	2.466E+00	1.037E+00	1.237E+01	2.301E+00	9.571E-01	**1.193E+01**	**2.244E+00**	**9.356E-01**	1.203E+01	2.281E+00	9.964E-01
ETT2/D	short	3.227E+05	3.771E+02	2.454E+00	3.573E+05	3.990E+02	2.633E+00	2.382E+05	3.254E+02	2.112E+00	**2.187E+05**	3.037E+02	1.929E+00	2.248E+05	**2.986E+02**	**1.870E+00**
ETT2/H	short	**1.140E+02**	**6.689E+00**	**7.767E-01**	1.344E+02	7.280E+00	8.366E-01	1.159E+02	6.834E+00	7.958E-01	1.141E+02	6.756E+00	7.946E-01	1.194E+02	6.961E+00	8.164E-01
ETT2/H	medium	**2.977E+02**	**1.082E+01**	**1.177E+00**	4.439E+02	1.318E+01	1.372E+00	3.525E+02	1.207E+01	1.277E+00	3.134E+02	1.117E+01	1.185E+00	3.903E+02	1.283E+01	1.332E+00
ETT2/H	long	**2.708E+02**	**1.069E+01**	**1.233E+00**	1.902E+03	2.107E+01	1.996E+00	4.145E+02	1.331E+01	1.442E+00	3.608E+02	1.225E+01	1.327E+00	3.184E+02	1.219E+01	1.418E+00
ETT2/W	short	2.237E+07	3.303E+03	2.438E+00	2.012E+07	3.359E+03	2.458E+00	2.015E+07	3.226E+03	2.448E+00	1.657E+07	3.007E+03	2.139E+00	**1.614E+07**	**2.805E+03**	**1.981E+00**
HIERARCHICAL_SALES/D	short	**2.838E+01**	**2.340E+00**	**7.658E-01**	2.935E+01	2.526E+00	8.229E-01	2.859E+01	2.616E+00	8.786E-01	2.862E+01	2.519E+00	8.244E-01	2.844E+01	2.505E+00	8.210E-01
HIERARCHICAL_SALES/W	short	**4.717E+02**	**9.736E+00**	**8.094E-01**	4.916E+02	1.122E+01	8.946E-01	4.850E+02	1.128E+01	9.001E-01	4.818E+02	1.117E+01	9.009E-01	4.779E+02	1.110E+01	8.998E-01
HOSPITAL	short	**2.544E+03**	**1.887E+01**	**8.327E-01**	4.345E+03	2.203E+01	9.092E-01	5.412E+03	2.279E+01	8.390E-01	2.577E+03	1.918E+01	8.582E-01	2.950E+03	1.970E+01	8.768E-01
JENA_WEATHER/10T	short	9.768E+02	7.788E+00	4.366E-01	**8.500E+02**	**7.685E+00**	**4.107E-01**	1.109E+03	1.002E+01	2.153E+00	1.107E+03	9.679E+00	5.619E-01	1.144E+03	9.854E+00	5.377E-01
JENA_WEATHER/10T	medium	**1.390E+03**	**1.053E+01**	**6.521E-01**	1.711E+03	1.315E+01	7.601E-01	1.587E+03	1.447E+01	1.432E+00	1.431E+03	1.204E+01	7.398E-01	1.761E+03	1.583E+01	8.303E-01
JENA_WEATHER/10T	long	**1.364E+03**	**1.046E+01**	**6.836E-01**	1.745E+03	1.401E+01	7.912E-01	1.670E+03	1.469E+01	1.418E+00	1.412E+03	1.216E+01	7.653E-01	1.669E+03	1.432E+01	8.336E-01
JENA_WEATHER/D	short	2.939E+03	2.670E+01	2.332E+00	3.414E+03	2.834E+01	2.392E+00	7.159E+02	1.398E+01	1.745E+00	**6.498E+02**	**1.315E+01**	1.441E+00	8.971E+02	1.469E+01	**1.364E+00**
JENA_WEATHER/H	short	**1.177E+03**	**1.030E+01**	**6.420E-01**	1.262E+03	1.197E+01	6.929E-01	1.178E+03	1.138E+01	1.301E+00	1.183E+03	1.091E+01	6.707E-01	1.198E+03	1.142E+01	6.675E-01
JENA_WEATHER/H	medium	1.225E+03	**1.207E+01**	8.894E-01	**1.161E+03**	1.213E+01	**8.741E-01**	1.430E+03	1.465E+01	1.167E+00	1.415E+03	1.400E+01	1.001E+00	1.814E+03	1.837E+01	1.136E+00
JENA_WEATHER/H	long	**1.684E+03**	**1.823E+01**	**1.066E+00**	6.913E+03	3.267E+01	1.373E+00	2.506E+03	2.278E+01	1.495E+00	2.659E+03	2.186E+01	1.247E+00	6.359E+03	3.368E+01	1.654E+00
KDD_CUP_2018_WITH_MISSING/D	short	2.856E+03	**2.162E+01**	1.247E+00	3.004E+03	2.298E+01	1.312E+00	**2.826E+03**	2.189E+01	1.357E+00	2.833E+03	2.186E+01	**1.245E+00**	2.918E+03	2.230E+01	1.268E+00
KDD_CUP_2018_WITH_MISSING/H	short	**4.333E+03**	**2.398E+01**	**1.000E+00**	4.515E+03	2.414E+01	1.002E+00	4.619E+03	2.566E+01	1.578E+00	4.598E+03	2.524E+01	1.029E+00	4.454E+03	2.492E+01	1.028E+00
KDD_CUP_2018_WITH_MISSING/H	medium	**4.772E+03**	2.647E+01	1.110E+00	5.016E+03	2.601E+01	**1.081E+00**	5.371E+03	2.738E+01	1.261E+00	5.180E+03	**2.597E+01**	**1.080E+00**	4.777E+03	2.724E+01	1.196E+00
KDD_CUP_2018_WITH_MISSING/H	long	4.128E+03	2.572E+01	1.101E+00	**3.848E+03**	2.446E+01	1.057E+00	4.064E+03	2.491E+01	1.053E+00	3.947E+03	**2.422E+01**	**1.027E+00**	4.170E+03	2.657E+01	1.143E+00
M4_DAILY	short	**4.366E+05**	**1.758E+02**	**3.226E+00**	4.558E+05	1.829E+02	3.373E+00	4.444E+05	1.807E+02	3.306E+00	4.790E+05	1.854E+02	3.408E+00	4.509E+05	1.838E+02	3.366E+00
M4_HOURLY	short	2.996E+06	3.049E+02	1.858E+00	**1.556E+06**	**2.825E+02**	3.241E+00	1.977E+06	2.936E+02	2.087E+00	3.148E+06	3.419E+02	**1.520E+00**	3.503E+06	3.655E+02	1.665E+00
M4_MONTHLY	short	**1.931E+06**	**5.708E+02**	**1.031E+00**	2.060E+06	6.078E+02	1.159E+00	2.173E+06	6.936E+02	1.728E+00	2.104E+06	6.138E+02	1.153E+00	2.050E+06	6.064E+02	1.148E+00
m4_quarterly	short	**1.921E+06**	**5.800E+02**	**1.235E+00**	2.048E+06	6.508E+02	1.491E+00	1.986E+06	6.192E+02	1.346E+00	2.057E+06	6.544E+02	1.508E+00	2.013E+06	6.431E+02	1.476E+00
M4_WEEKLY	short	**3.530E+05**	**3.002E+02**	**2.314E+00**	4.202E+05	3.311E+02	2.548E+00	4.649E+05	3.501E+02	3.175E+00	4.802E+05	3.582E+02	3.296E+00	4.729E+05	3.517E+02	2.959E+00
M4_YEARLY	short	**3.159E+06**	**8.157E+02**	**3.016E+00**	3.274E+06	8.496E+02	3.142E+00	3.342E+06	8.500E+02	3.220E+00	3.335E+06	8.470E+02	3.148E+00	3.256E+06	8.275E+02	3.065E+00
SAUGEENDAY/D	short	**1.173E+03**	**1.438E+01**	**3.195E+00**	1.197E+03	1.489E+01	3.308E+00	1.330E+03	1.642E+01	3.646E+00	1.324E+03	1.722E+01	3.822E+00	1.322E+03	1.536E+01	3.410E+00
SAUGEENDAY/M	short	7.486E+02	**1.587E+01**	**9.297E-01**	7.677E+02	1.959E+01	1.147E+00	7.860E+02	1.874E+01	1.097E+00	8.617E+02	2.072E+01	1.213E+00	**6.367E+02**	1.778E+01	1.042E+00
SAUGEENDAY/W	short	**1.421E+03**	**1.920E+01**	**1.557E+00**	1.477E+03	1.982E+01	1.607E+00	1.585E+03	2.196E+01	1.779E+00	1.492E+03	2.213E+01	1.794E+00	1.424E+03	1.951E+01	1.582E+00
SOLAR/10T	short	3.330E+01	**2.778E+00**	**1.221E+00**	3.858E+01	2.955E+00	1.281E+00	**2.947E+01**	3.198E+00	1.433E+00	3.058E+01	2.952E+00	1.297E+00	2.977E+01	3.115E+00	1.364E+00
SOLAR/10T	medium	2.113E+01	**2.027E+00**	**8.834E-01**	2.184E+01	2.371E+00	1.037E+00	**1.891E+01**	2.244E+00	9.795E-01	1.896E+01	2.228E+00	9.715E-01	2.052E+01	2.655E+00	1.160E+00
SOLAR/10T	long	2.209E+01	**2.018E+00**	**8.809E-01**	2.311E+01	2.539E+00	1.110E+00	**1.956E+01**	2.355E+00	1.028E+00	1.966E+01	2.227E+00	9.715E-01	2.186E+01	2.626E+00	1.145E+00
SOLAR/D	short	1.413E+05	**2.736E+02**	**1.064E+00**	**1.338E+05**	2.846E+02	1.110E+00	1.342E+05	2.763E+02	1.079E+00	1.369E+05	2.742E+02	1.070E+00	1.355E+05	2.802E+02	1.094E+00
SOLAR/H	short	**6.313E+02**	**1.125E+01**	**8.676E-01**	7.032E+02	1.323E+01	1.019E+00	7.044E+02	1.286E+01	9.932E-01	7.023E+02	1.307E+01	1.009E+00	6.332E+02	1.436E+01	1.109E+00
SOLAR/H	medium	1.082E+03	1.531E+01	1.171E+00	8.215E+02	1.564E+01	1.199E+00	**7.425E+02**	**1.454E+01**	**1.111E+00**	7.470E+02	1.491E+01	1.136E+00	7.808E+02	1.544E+01	1.183E+00
SOLAR/H	long	**7.999E+02**	**1.488E+01**	**1.136E+00**	9.025E+02	1.712E+01	1.308E+00	8.560E+02	1.635E+01	1.246E+00	8.467E+02	1.659E+01	1.263E+00	9.789E+02	1.800E+01	1.369E+00
SOLAR/W	short	5.988E+06	2.108E+03	2.331E+00	6.553E+06	2.196E+03	2.430E+00	6.283E+06	2.154E+03	2.383E+00	5.703E+06	2.051E+03	2.270E+00	**3.971E+06**	**1.708E+03**	**1.888E+00**
TEMPERATURE_RAIN_WITH_MISSING	short	**1.994E+02**	**7.178E+00**	**1.861E+00**	2.129E+02	7.825E+00	2.109E+00	1.995E+02	7.689E+00	2.364E+00	2.030E+02	7.785E+00	2.086E+00	2.187E+02	8.247E+00	2.266E+00
US_BIRTHS/D	short	**2.317E+05**	**3.168E+02**	**4.660E-01**	2.356E+05	3.305E+02	4.860E-01	3.241E+05	4.109E+02	6.044E-01	2.772E+05	3.747E+02	5.514E-01	2.639E+05	3.544E+02	5.215E-01
US_BIRTHS/M	short	**2.324E+08**	1.226E+04	1.387E+00	2.856E+08	1.374E+04	1.555E+00	2.337E+08	1.249E+04	1.414E+00	2.677E+08	1.435E+04	1.626E+00	2.408E+08	**1.213E+04**	**1.373E+00**
US_BIRTHS/W	short	9.916E+06	2.416E+03	2.195E+00	1.282E+07	2.941E+03	2.672E+00	1.377E+07	3.076E+03	2.794E+00	9.992E+06	2.527E+03	2.296E+00	**7.350E+06**	**2.015E+03**	**1.831E+00**
car_parts_with_missing	short	**1.221E+00**	**5.090E-01**	**9.540E-01**	1.232E+00	5.221E-01	9.730E-01	1.225E+00	5.545E-01	1.120E+00	1.269E+00	5.755E-01	1.078E+00	1.283E+00	5.325E-01	1.007E+00
restaurant	short	**1.501E+02**	**7.551E+00**	**7.210E-01**	1.512E+02	7.627E+00	7.274E-01	1.520E+02	7.678E+00	7.386E-01	1.526E+02	7.662E+00	7.313E-01	1.516E+02	7.647E+00	7.299E-01
Best count (MSE)	–	61	–	–	11	–	–	5	–	–	8	–	–	12	–	–
Best count (MAE)	–	–	65	–	–	9	–	–	3	–	–	11	–	–	9	–
Best count (MASE)	–	–	–	63	–	–	10	–	–	2	–	–	14	–	–	8

Note: Win counts use the reported significant digits. Values distinguishable at the reported precision are not treated as ties; exact ties are credited to all tied methods.

Table A.8: **Short-context GIFT-Eval settings.** All GIFT-Eval settings whose context length is at most 90, scored with the benchmark's official MASE metric (lower is better). PRISM attains the lowest MASE in 19 of the 27 settings. Best value in each row is shown in bold.

Setting	T_{ctx}	PRISM	D-PAD	DLinear	FITS	ModernTCN
m4_monthly	16	**1.031**	1.159	1.728	1.153	1.148
m4_yearly	16	**3.016**	3.142	3.220	3.148	3.065
electricity/W	24	**1.973**	2.000	2.108	2.024	2.079
ett1/W	24	**1.533**	1.567	1.655	1.726	1.687
ett2/W	24	2.438	2.458	2.448	2.139	**1.981**
hierarchical_sales/W	24	**0.8094**	0.8946	0.9001	0.9009	0.8998
saugeenday/W	24	**1.557**	1.607	1.779	1.794	1.582
solar/W	24	2.331	2.430	2.383	2.270	**1.888**
us_births/W	24	2.195	2.672	2.794	2.296	**1.831**
hospital	36	**0.8327**	0.9092	0.8390	0.8582	0.8768
saugeenday/M	36	**0.9297**	1.147	1.097	1.213	1.042
us_births/M	36	1.387	1.555	1.414	1.626	**1.373**
m4_weekly	39	**2.314**	2.548	3.175	3.296	2.959
m4_daily	42	**3.226**	3.373	3.306	3.408	3.366
LOOP_SEATTLE/D	90	0.9368	0.9597	**0.9340**	0.9704	0.9535
M_DENSE/D	90	**0.7207**	0.9185	1.019	1.076	1.137
covid_deaths	90	**35.77**	44.82	138.2	54.78	40.12
electricity/D	90	**1.761**	2.151	1.777	1.878	1.874
ett1/D	90	**1.652**	2.050	1.945	1.763	1.710
ett2/D	90	2.454	2.633	2.112	1.929	**1.870**
hierarchical_sales/D	90	**0.7658**	0.8229	0.8786	0.8244	0.8210
jena_weather/D	90	2.332	2.392	1.745	1.441	**1.364**
kdd_cup_2018_with_missing/D	90	1.247	1.312	1.357	**1.245**	1.268
saugeenday/D	90	**3.195**	3.308	3.646	3.822	3.410
solar/D	90	**1.064**	1.110	1.079	1.070	1.094
temperature_rain_with_missing	90	**1.861**	2.109	2.364	2.086	2.266
us_births/D	90	**0.4660**	0.4860	0.6044	0.5514	0.5215
Wins (27 settings)		**19**	0	1	1	6

Table A.9: Extended ablation results (encoder and importance scoring ablations). Mean rows and Δ correspond to Table 6; Δ is the relative change in mean error with respect to PRISM, so positive values indicate worse forecasts.

Dataset	Context	H	PRISM		Encoder				Importance scoring							
					Tree-level 1		MLP		All shared weights		No shared weights		No MLPs		Learnable threshold	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	336	96	0.355	0.374	0.358	0.382	0.368	0.388	0.358	0.382	0.357	0.382	0.384	0.414	0.501	0.476
ETTh1	336	192	0.390	0.405	0.400	0.412	0.409	0.418	0.405	0.409	0.405	0.409	0.404	0.411	0.422	0.433
ETTh1	336	336	0.386	0.407	0.394	0.411	0.404	0.424	0.397	0.410	0.390	0.408	0.411	0.418	0.420	0.441
ETTh1	336	720	0.421	0.445	0.448	0.455	0.449	0.461	0.444	0.452	0.449	0.456	0.450	0.455	0.466	0.475
ETTh2	336	96	0.261	0.322	0.263	0.322	0.281	0.341	0.263	0.322	0.266	0.323	0.275	0.337	0.301	0.359
ETTh2	336	192	0.311	0.359	0.320	0.360	0.332	0.376	0.319	0.361	0.332	0.368	0.330	0.371	0.327	0.372
ETTh2	336	336	0.308	0.361	0.321	0.369	0.338	0.385	0.309	0.360	0.314	0.363	0.328	0.382	0.328	0.382
ETTh2	336	720	0.387	0.420	0.383	0.417	0.408	0.435	0.386	0.417	0.389	0.420	0.404	0.437	0.401	0.435
ETTm1	336	96	0.282	0.326	0.283	0.328	0.292	0.347	0.283	0.327	0.283	0.326	0.293	0.333	0.314	0.364
ETTm1	336	192	0.323	0.352	0.323	0.352	0.341	0.372	0.324	0.352	0.328	0.352	0.337	0.360	0.335	0.373
ETTm1	336	336	0.354	0.377	0.362	0.376	0.371	0.391	0.362	0.376	0.362	0.377	0.364	0.377	0.367	0.393
ETTm1	336	720	0.423	0.409	0.431	0.412	0.428	0.426	0.424	0.413	0.428	0.415	0.583	0.427	0.429	0.427
ETTm2	336	96	0.158	0.234	0.160	0.244	0.177	0.263	0.162	0.246	0.161	0.245	0.178	0.267	0.178	0.267
ETTm2	336	192	0.214	0.267	0.214	0.281	0.234	0.305	0.217	0.284	0.215	0.283	0.234	0.308	0.234	0.304
ETTm2	336	336	0.264	0.319	0.271	0.319	0.291	0.339	0.271	0.320	0.267	0.318	0.287	0.342	0.288	0.341
ETTm2	336	720	0.352	0.372	0.372	0.381	0.364	0.386	0.357	0.376	0.356	0.376	0.370	0.395	0.366	0.387
Exchange	336	96	0.081	0.198	0.085	0.205	0.187	0.320	0.089	0.212	0.085	0.205	0.087	0.210	0.089	0.212
Exchange	336	192	0.173	0.297	0.175	0.299	0.506	0.544	0.179	0.304	0.175	0.300	0.181	0.309	0.189	0.314
Exchange	336	336	0.314	0.396	0.321	0.413	0.998	0.775	0.331	0.417	0.317	0.410	0.327	0.420	0.340	0.428
Exchange	336	720	0.712	0.606	1.016	0.782	1.670	1.023	0.751	0.621	0.732	0.615	0.916	0.719	0.886	0.704
Weather	336	96	0.140	0.177	0.150	0.188	0.155	0.220	0.149	0.188	0.147	0.186	0.159	0.209	0.161	0.207
Weather	336	192	0.187	0.227	0.194	0.231	0.205	0.255	0.194	0.231	0.193	0.230	0.202	0.250	0.201	0.247
Weather	336	336	0.234	0.266	0.247	0.271	0.260	0.293	0.246	0.271	0.245	0.270	0.260	0.312	0.242	0.283
Weather	336	720	0.311	0.320	0.323	0.326	0.327	0.343	0.332	0.326	0.330	0.324	0.322	0.336	0.321	0.337
Mean			0.306	0.343	0.326	0.356	0.408	0.410	0.315	0.349	0.314	0.348	0.337	0.367	0.338	0.373
Δ (%)			–	–	+6.44%	+3.64%	+33.43%	+19.35%	+2.87%	+1.71%	+2.52%	+1.52%	+10.15%	+6.84%	+10.42%	+8.80%

Table A.10: Extended ablation results (frequency filter ablations). Mean rows and Δ correspond to Table 6; Δ is the relative change in mean error with respect to PRISM, so positive values indicate worse forecasts.

Dataset	Context	H	PRISM		EMA		MCD		DoG		Binomial pyramid		FFT		Learnable filters	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	336	96	0.355	0.374	0.365	0.389	0.377	0.403	0.358	0.382	0.352	0.380	0.372	0.398	0.388	0.404
ETTh1	336	192	0.390	0.405	0.407	0.412	0.416	0.424	0.404	0.409	0.400	0.407	0.415	0.422	0.426	0.425
ETTh1	336	336	0.386	0.407	0.404	0.415	0.392	0.411	0.390	0.407	0.385	0.405	0.412	0.425	0.406	0.419
ETTh1	336	720	0.421	0.445	0.454	0.456	0.443	0.458	0.430	0.448	0.436	0.450	0.451	0.464	0.467	0.466
ETTh2	336	96	0.261	0.322	0.270	0.326	0.271	0.327	0.271	0.327	0.262	0.321	0.265	0.323	0.281	0.337
ETTh2	336	192	0.311	0.359	0.332	0.367	0.345	0.390	0.332	0.367	0.321	0.363	0.322	0.361	0.337	0.378
ETTh2	336	336	0.308	0.361	0.324	0.369	0.317	0.376	0.323	0.368	0.308	0.360	0.308	0.359	0.340	0.379
ETTh2	336	720	0.387	0.420	0.398	0.426	0.399	0.437	0.390	0.428	0.393	0.422	0.388	0.420	0.420	0.438
ETTm1	336	96	0.282	0.326	0.287	0.327	0.284	0.330	0.284	0.329	0.285	0.329	0.292	0.334	0.292	0.332
ETTm1	336	192	0.323	0.352	0.329	0.352	0.326	0.355	0.323	0.352	0.327	0.353	0.327	0.354	0.329	0.367
ETTm1	336	336	0.354	0.377	0.368	0.381	0.363	0.379	0.362	0.377	0.361	0.375	0.365	0.379	0.371	0.383
ETTm1	336	720	0.423	0.409	0.435	0.414	0.430	0.414	0.427	0.411	0.429	0.411	0.434	0.415	0.434	0.419
ETTm2	336	96	0.158	0.234	0.159	0.243	0.173	0.263	0.160	0.244	0.161	0.245	0.162	0.247	0.165	0.247
ETTm2	336	192	0.214	0.267	0.214	0.281	0.233	0.306	0.214	0.281	0.212	0.282	0.217	0.285	0.220	0.285
ETTm2	336	336	0.264	0.319	0.273	0.320	0.286	0.341	0.271	0.318	0.270	0.319	0.272	0.320	0.283	0.326
ETTm2	336	720	0.352	0.372	0.370	0.379	0.363	0.387	0.370	0.379	0.372	0.383	0.372	0.381	0.364	0.385
Exchange	336	96	0.081	0.198	0.084	0.205	0.088	0.211	0.097	0.224	0.081	0.201	0.083	0.203	0.093	0.215
Exchange	336	192	0.173	0.297	0.177	0.301	0.180	0.308	0.212	0.331	0.173	0.299	0.173	0.298	0.197	0.321
Exchange	336	336	0.314	0.396	0.355	0.429	0.320	0.413	0.334	0.416	0.338	0.420	0.336	0.421	0.351	0.426
Exchange	336	720	0.712	0.606	0.795	0.637	0.784	0.651	0.759	0.620	0.803	0.641	0.751	0.621	0.897	0.665
Weather	336	96	0.140	0.177	0.153	0.190	0.147	0.187	0.149	0.189	0.153	0.191	0.152	0.191	0.147	0.188
Weather	336	192	0.187	0.227	0.196	0.232	0.196	0.231	0.196	0.232	0.197	0.232	0.195	0.233	0.193	0.231
Weather	336	336	0.234	0.266	0.239	0.270	0.240	0.270	0.238	0.270	0.239	0.270	0.238	0.271	0.245	0.285
Weather	336	720	0.311	0.320	0.326	0.333	0.323	0.336	0.320	0.325	0.321	0.325	0.318	0.324	0.334	0.326
Mean			0.306	0.343	0.321	0.352	0.321	0.359	0.317	0.351	0.316	0.349	0.318	0.352	0.333	0.360
Δ (%)			–	–	+5.08%	+2.65%	+4.84%	+4.52%	+3.72%	+2.40%	+3.24%	+1.80%	+3.80%	+2.59%	+8.70%	+4.99%

Table A.11: Extended hyperparameter search (number of Wavelet bands). (Red) highest counts.

Dataset	Context	H	2		5		8	
			MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	336	96	0.357	0.383	0.355	0.374	0.354	0.380
ETTh1	336	192	0.405	0.410	0.390	0.405	0.409	0.414
ETTh1	336	336	0.393	0.409	0.386	0.407	0.390	0.411
ETTh1	336	720	0.449	0.457	0.421	0.445	0.445	0.454
ETTh2	336	96	0.271	0.367	0.261	0.321	0.265	0.322
ETTh2	336	192	0.325	0.364	0.311	0.359	0.318	0.360
ETTh2	336	336	0.312	0.363	0.308	0.361	0.314	0.369
ETTh2	336	720	0.389	0.423	0.387	0.420	0.395	0.427
ETTm1	336	96	0.288	0.330	0.282	0.326	0.287	0.329
ETTm1	336	192	0.328	0.355	0.323	0.352	0.331	0.356
ETTm1	336	336	0.358	0.378	0.354	0.377	0.360	0.377
ETTm1	336	720	0.412	0.427	0.423	0.409	0.413	0.424
ETTm2	336	96	0.165	0.249	0.158	0.234	0.162	0.245
ETTm2	336	192	0.223	0.288	0.214	0.267	0.218	0.285
ETTm2	336	336	0.272	0.321	0.264	0.319	0.273	0.319
ETTm2	336	720	0.353	0.373	0.352	0.372	0.354	0.373
Best count (MSE)			1		**14**		1	
Best count (MAE)				0		**16**		2

Table A.12: Extended hyperparameter search (depth of tree decomposition). (Red) highest counts.

Dataset	Context	H	1		2		3	
			MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	336	96	0.363	0.386	0.355	0.374	0.355	0.374
ETTh1	336	192	0.416	0.416	0.390	0.405	0.407	0.411
ETTh1	336	336	0.423	0.426	0.386	0.407	0.397	0.412
ETTh1	336	720	0.495	0.482	0.421	0.445	0.446	0.461
ETTh2	336	96	0.263	0.322	0.261	0.321	0.262	0.321
ETTh2	336	192	0.321	0.361	0.311	0.359	0.316	0.359
ETTh2	336	336	0.313	0.361	0.308	0.361	0.315	0.364
ETTh2	336	720	0.387	0.417	0.387	0.420	0.386	0.419
ETTm1	336	96	0.287	0.331	0.282	0.326	0.288	0.330
ETTm1	336	192	0.327	0.355	0.323	0.352	0.330	0.355
ETTm1	336	336	0.356	0.376	0.354	0.377	0.355	0.376
ETTm1	336	720	0.412	0.427	0.423	0.409	0.413	0.423
ETTm2	336	96	0.166	0.248	0.158	0.234	0.161	0.242
ETTm2	336	192	0.221	0.286	0.214	0.267	0.219	0.284
ETTm2	336	336	0.275	0.322	0.264	0.319	0.273	0.319
ETTm2	336	720	0.355	0.374	0.352	0.372	0.353	0.373
Best count (MSE)			1		**14**		2	
Best count (MAE)				3		**14**		5

Table A.13: Extended hyperparameter search (overlap between temporal segments). (Red) highest counts.

Dataset	Context	H	0		8		24	
			MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	336	96	0.359	0.383	0.355	0.374	0.358	0.382
ETTh1	336	192	0.409	0.413	0.390	0.405	0.414	0.409
ETTh1	336	336	0.416	0.405	0.386	0.407	0.397	0.411
ETTh1	336	720	0.446	0.457	0.421	0.445	0.443	0.453
ETTh2	336	96	0.265	0.322	0.261	0.322	0.271	0.327
ETTh2	336	192	0.316	0.362	0.311	0.359	0.328	0.365
ETTh2	336	336	0.319	0.367	0.308	0.361	0.318	0.365
ETTh2	336	720	0.397	0.425	0.387	0.420	0.392	0.425
ETTm1	336	96	0.288	0.332	0.282	0.326	0.287	0.331
ETTm1	336	192	0.332	0.357	0.323	0.352	0.330	0.356
ETTm1	336	336	0.360	0.377	0.354	0.377	0.360	0.378
ETTm1	336	720	0.412	0.426	0.423	0.409	0.412	0.424
ETTm2	336	96	0.162	0.245	0.158	0.234	0.163	0.246
ETTm2	336	192	0.219	0.284	0.214	0.267	0.220	0.285
ETTm2	336	336	0.271	0.319	0.264	0.319	0.272	0.319
ETTm2	336	720	0.351	0.371	0.352	0.372	0.353	0.372
Best count (MSE)			2		**14**		1	
Best count (MAE)				5		**14**		1

Table A.14: Extended results (different context length). (Red) highest counts.

Dataset	Context	H	PRISM		D-PAD		Time Mixer	
			MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	96	0.372	0.385	0.387	0.400	0.385	0.401
ETTh1	96	192	0.432	0.415	0.453	0.436	0.440	0.428
ETTh1	96	336	0.452	0.422	0.456	0.437	0.513	0.470
ETTh1	96	720	0.463	0.447	0.469	0.463	0.501	0.480
ETTh2	96	96	0.277	0.327	0.289	0.335	0.290	0.342
ETTh2	96	192	0.347	0.373	0.359	0.381	0.382	0.401
ETTh2	96	336	0.377	0.342	0.382	0.353	0.416	0.431
ETTh2	96	720	0.401	0.423	0.408	0.429	0.419	0.440
ETTm1	96	96	0.312	0.341	0.320	0.347	0.324	0.363
ETTm1	96	192	0.368	0.370	0.375	0.375	0.369	0.389
ETTm1	96	336	0.394	0.392	0.401	0.399	0.395	0.407
ETTm1	96	720	0.459	0.437	0.479	0.438	0.459	0.446
ETTm2	96	96	0.172	0.251	0.176	0.258	0.175	0.258
ETTm2	96	192	0.237	0.294	0.243	0.301	0.239	0.306
ETTm2	96	336	0.294	0.333	0.308	0.343	0.295	0.339
ETTm2	96	720	0.392	0.395	0.397	0.402	0.393	0.396
Best count (MSE)			**16**					
Best count (MAE)				**16**				

Table A.15: **PRISM hyperparameters for the main six datasets.** Values are taken directly from the released 'quick' scripts. All settings use Haar decomposition.

Hyperparameter	ETTh1				ETTh2				ETTm1				ETTm2				Exchange				Weather			
	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720
Look-back window	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336
Batch size	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512	512
Learning rate	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4
Encoder/decoder hidden size	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	256	256	256	336
Dropout	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.5	0.5	0.5	0.5	0.25	0.25	0.25	0.0
Number of components	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	8
Training epochs	120	120	120	120	100	100	100	100	100	100	100	100	100	100	100	100	200	200	200	200	100	100	100	100
Patience	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	25	25	15	35	15	15	15	25
Tree depth	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Overlap	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Drop last	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1

Table A.16: **Baseline hyperparameter configurations in the main six-dataset benchmark (Part I: D-PAD, DLinear, SCINet).** We fixed the forecasting protocol (`seq_len=336`, multivariate forecasting, `pred_len` $\in \{96, 192, 336, 720\}$) and then used each method's released script, released horizon-specific setting, or runner default as closely as possible. For compactness, the four ETT datasets are summarized jointly using H1 = ETTh1, H2 = ETTh2, m1 = ETTm1, and m2 = ETTm2. Bracketed entries summarize the effective ranges considered across released settings and limited sweeps. Abbreviations: enc_hid = encoder hidden size, dec_hid = decoder hidden size, dec_lyr = decoder layers, indiv = individual.

Hyperparameter	D-PAD			DLinear			SCINet		
	ETT	Exchange	Weather	ETT	Exchange	Weather	ETT	Exchange	Weather
Look-back window	336	336	336	336	336	336	336	96	336
Batch size (used)	512	32	64	8	8	16	32	8	32
Batch size range	[32, 512]	[32, 512]	[32, 512]	[8, 32]	[8, 32]	[16, 16]	[16, 512]	[8, 8]	[32, 32]
Learning rate (used)	1e-4	1e-4	1e-4	1e-4	5e-4	1e-4	5e-5	5e-5	5e-4
Learning rate range	[1e-5, 1e-4]	[1e-5, 1e-4]	[1e-5, 1e-4]	[1e-4, 1e-1]	[5e-4, 5e-3]	[1e-4, 1e-4]	[1e-5, 5e-4]	[5e-5, 5e-5]	[5e-4, 5e-4]
Dropout (used)	0.25	{0,0.25,0.5}	{0,0.25,0.5}	0.05	0.05	0.05	0.5	0	0.5
Dropout range	[0, 0.5]	[0, 0.5]	[0, 0.5]	[0.05, 0.05]	[0.05, 0.05]	[0.05, 0.05]	[0, 0.5]	[0, 0]	[0.5, 0.5]
Training epochs (used)	100	100	100	10	10	10	100	20	100
Training epoch range	[100, 100]	[100, 100]	[100, 100]	[10, 10]	[10, 10]	[10, 10]	[100, 100]	[20, 20]	[100, 100]
Early stopping (used)	5	5	5	3	3	3	5	n/a	5
Early stopping range	[5, 5]	[5, 5]	[5, 5]	[3, 3]	[3, 3]	[3, 3]	[5, 5]	[n/a]	[5, 5]
Drop last	true	true	true	true	true	true	true	true	true
Method-specific hyperparameters									
Specific 1	enc_hid=336; levels=2	enc_hid=336; levels=2	enc_hid=336; levels=2	indiv=False; linear backbone	same	same	H1 hidden_size=4/4/1/1; H2=0.5/0.5/1/4; m1,m2=0.5/0.5/4/4 H1 levels=3/3/4/5;	hidden_size=0.125; levels=3	hidden_size=1; levels=3
Specific 2	RIN=1; dec_hid=336 K_IMP=6;	RIN=1; dec_hid=336 K_IMP=6;	RIN=1; dec_hid=336 K_IMP=6;	indiv option exists; disabled here	same	same	H2=4/4/4/5; m1,m2=4/4/5/5 H1/H2 stacks=1;	normalize=3; dec_lyr=2 lastWeight=0.5;	stacks=1; kernel=5 dropout=0.5;
Specific 3	num_heads=1	num_heads=1	num_heads=1	–	–	–	m1,m2 stacks=2/2/1/2	stacks=1	stacks=1

Table A.17: **Baseline hyperparameter configurations in the main six-dataset benchmark (Part II: TimeMixer, WPMixer, XPatch).** We fixed the forecasting protocol (`seq_len=336`, multivariate forecasting, `pred_len` $\in \{96, 192, 336, 720\}$) and then used each method's released script, released horizon-specific setting, or runner default as closely as possible. For compactness, the four ETT datasets are summarized jointly using H1 = ETTh1, H2 = ETTh2, m1 = ETTm1, and m2 = ETTm2. Bracketed entries summarize the effective ranges considered across released settings and limited sweeps. Abbreviations: emb = embedding dropout, DS_L = down-sampling layers, DS_W = down-sampling window, DS_M = down-sampling method, tf = tfactor, df = dfactor.

Hyperparameter	TimeMixer			WPMixer			XPatch		
	ETT	Exchange	Weather	ETT	Exchange	Weather	ETT	Exchange	Weather
Look-back window	336	336	336	336	336	336	336	336	336
Batch size (used)	16	8	128	128	32	32	128	32	256
Batch size range	[16, 128, 512]	[8, 512]	[128, 512]	[128, 512]	[32, 512]	[32, 512]	[32, 256, 512]	[32, 512]	[256, 512]
Learning rate (used)	1e-2	1e-2	1e-2	1.47e-3	5e-4	1e-3	1e-4	1e-4	1e-4
Learning rate range	[1e-5, 1e-2]	[1e-5, 1e-2]	[1e-5, 1e-2]	[2.96e-4, 4.52e-3]	[2e-4, 1e-3]	[1e-3, 3e-3]	[1e-5, 1e-4]	[1e-5, 1e-4]	[1e-5, 1e-4]
Dropout (used)	0.0	0.25	0.25	0.2	0.2	0.2	n/a	n/a	n/a
Dropout range	[0.0, 0.5]	[0.0, 0.5]	[0.0, 0.5]	[0.1, 0.4]	[0.2, 0.2]	[0.2, 0.2]	[not exposed]	[not exposed]	[not exposed]
Training epochs (used)	100	100	20	10	20	20	20	10	100
Training epoch range	[10, 100]	[100, 100]	[20, 20]	[10, 10]	[20, 20]	[20, 20]	[20, 20]	[10, 10]	[100, 100]
Early stopping (used)	15	15	10	5	5	5	5	3	10
Early stopping range	[10, 15]	[15, 15]	[10, 10]	[5, 5]	[5, 5]	[5, 5]	[5, 5]	[3, 3]	[10, 10]
Drop last	true	true	true	train=true; val/test=false	train=true; val/test=false	train=true; val/test=false	true	true	true
Method-specific hyperparameters									
Specific 1	e_layers=2; d_model=16	e_layers=2; d_model=16	e_layers=2; d_model=16	wavelet/level vary by horizon; patch_len/stride vary mildly	wavelet=coif4/coif4 sym4/sym4; level=1 tf=5;	wavelet=sym3/coif4 coif4/sym2; level=1	patch_len=16; stride=8	patch_len=16; stride=8	patch_len=16; stride=8
Specific 2	d_ff=32; DS_L=3	d_ff=32; DS_L=3	d_ff=32; DS_L=3	d_model [128,256]; tf/df vary by horizon	df=5; d_model=128 patch_len=16;	d_model [128,256]; emb=0.0/0.1/0.1/0.0 tf=3/5/7/7;	ma_type=ema; alpha=0.3; beta=0.3 revin=1;	ma_type=reg; alpha=0.3; beta=0.3 revin=1;	alpha=0.3; beta=0.3 revin=1;
Specific 3	DS_W=2; DS_M=avg	DS_W=2; DS_M=avg	DS_W=2; DS_M=avg	ETT emb [0.0,0.2]	stride=8; emb=[0.05,0.05]	df=5/3/7/5; emb [0.0,0.1]	lradj=type3; padding=end	lradj=type1; padding=end	lradj=sigmoid; padding=end

Table A.18: **Scaling of PRISM to longer context lengths.** We report accuracy (when checkpoints are available), FLOPs, parameter count, inference latency, training step time, and peak memory usage on ETTh1 with 96 horizons.

seq_len	MSE ↓	MAE ↓	FLOPs (G)	Params (M)	Infer. latency (ms/batch)	Train step (ms/batch)	Infer. peak mem (MB)	Train peak mem (MB)
336	0.355	0.374	0.00252371	0.352399	4.04	6.46	235.47	389.13
720	0.367	0.399	0.00424403	0.598159	4.60	9.02	448.97	641.39
1024	0.360	0.393	0.00560595	0.792719	5.33	11.61	620.41	830.26
2048	0.379	0.415	0.01019347	1.448079	17.08	20.30	1187.81	1547.39