PARALLAX: CORRECTING TIME SERIES FORECASTS FROM CROSS-RESOLUTION DISAGREEMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

In long-term time series forecasting, a fine-resolution forecast, when aggregated, can disagree with a forecast produced directly at a coarser resolution for the same future interval. This disagreement is available at prediction time and carries information about fine-resolution forecast errors. We present Parallax, a post-hoc, model-agnostic method that uses cross-resolution disagreement to correct the fine-resolution forecasts of frozen backbones, whether pretrained foundation models or task-specific deep forecasters trained on the target dataset. Parallax decomposes the disagreement into the aggregated fine residual and the coarse-forecast error, reducing forecast correction to residual estimation, and fits an affine corrector whose moments combine a resolution-structured covariance, built from a nested aggregation chain, with empirical moments estimated from historical forecast–target pairs. At test time, Parallax uses only the fine- and coarse-resolution forecasts, requiring neither future observations nor any updating to the backbone. Evaluated on 13 benchmark datasets with 7 task-specific models and 5 pretrained foundation models, Parallax reduces MSE by 11.8%-26.4% across task-specific backbones and by 33.2% on average across foundation models.

1 INTRODUCTION

Time series forecasting supports planning in energy, transportation, and public health (Hyndman et al., 2014). And Deep Learning (DL) has reshaped it over the past decade, ranging from task-specific forecasters (Nie et al., 2023; Liu et al., 2024; Wang et al., 2024) to foundation models (Woo et al., 2024; Liu et al., 2025). In practice, many operational decisions depend on these forecasts at fine temporal resolution, such as the next day of hourly electricity load or the next hour of five-minute traffic flow (Petropoulos et al., 2022).

A fine-resolution forecast is not the only forecast one can make for the same future interval. Consider a next-day forecast at hourly resolution. Its average over each six-hour interval need not match a forecast produced directly at six-hour resolution. There are good reasons to take the coarse forecast seriously: aggregation smooths high-frequency noise and exposes slower patterns, and forecasting an aggregated series directly can be more accurate than aggregating fine forecasts (Spiliotis et al., 2020; Kourentzes et al., 2014). This leads to the question we study: *can the disagreement between forecasts at different temporal resolutions correct a fine-resolution forecast before its target is observed?*

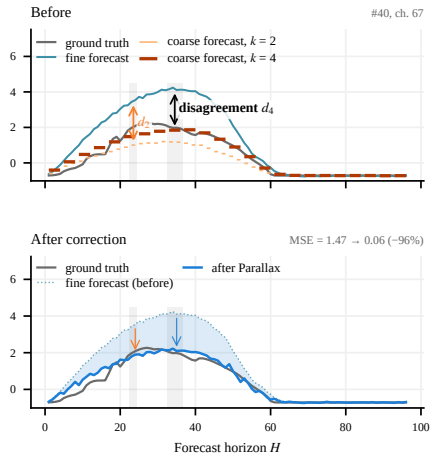


Figure 1: Parallax with OLinear (Yue et al., 2025) on sample 40, channel 67, $H = 96$, Solar-Energy data. (Top) The fine forecast disagrees with the coarse forecasts (aggregation factor $k = 2, 4$) of the same interval; the arrow marks the disagreement d_k on the shaded block. (Bottom) Parallax uses this disagreement to correct the fine forecast.

Three lines of work come close to this question. Statistical temporal hierarchy methods forecast at several aggregation levels and reconcile them so that fine forecasts aggregate exactly to coarse

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.

ones (Athanasopoulos et al., 2017; Jeon et al., 2019; Nystrup et al., 2020; Rangapuram et al., 2023). Their goal is a coherent hierarchy, so they adjust every level (the fine forecast is not the target). Multiscale deep forecasters build multiple resolutions into the architecture, through coarse-to-fine refinement or cross-scale mixing (Wang et al., 2024; Zhang et al., 2025; Murad et al., 2025; Lee et al., 2026). This requires training the forecaster and does not apply afterward to a frozen one. Post-hoc correctors work on a frozen forecaster but learn from its inputs, outputs, and past errors (Liang et al., 2026; Gao et al., 2026), not from disagreement between forecasts.

Contributions. Different from these methods, we answer the question above with Parallax, a post-hoc method that corrects a frozen fine-resolution forecast from its disagreement with coarse forecasts of the same interval, fitted from historical forecast-target pairs alone. The name Parallax comes from astronomy, where the apparent shift of a nearby star between two viewpoints of Earth's orbit can be exploited to reveal its distance (Bailer-Jones, 2015). Here, we can regard these two viewpoints as the fine- and coarse-resolution forecasts of the same future interval, and their shift as the cross-resolution disagreement. Figure 1 (top) illustrates this on a Solar-Energy window. The fine forecast overshoots the ground truth, while the coarse forecasts at higher resolutions $k = 2$ and $k = 4$ remain closer to it. Over the shaded block, the arrows indicate the resulting disagreements d_2 and d_4, defined as the gap between each coarse forecast and the corresponding block average of the fine forecast. However, unlike astronomical parallax, this disagreement does not directly determine the correction. A coarse forecast may itself be inaccurate, so the disagreement alone cannot determine which forecast is responsible for the gap. Nor does a block-level quantity such as d_4 specify how to distribute the correction across its four fine-resolution positions, or how to weight different resolutions when d_2 and d_4 differ in sign or magnitude.

Parallax addresses these questions by viewing disagreement as a noisy observation of the fine-forecast residual. Since a coarse forecast and the block average of the fine forecast target the same block mean, their difference decomposes exactly into the block-averaged fine residual and the coarse-forecast error. For example, in Figure 1, the coarse forecasts happen to lie close to the ground truth over the shaded block, so the disagreement largely captures the fine-forecast error and therefore indicates the direction of the needed correction. Parallax uses this signal to pull the fine forecast downward, reducing the MSE from 1.47 to 0.06 (bottom). However, we need not assume either forecast is correct. In general, the coarse forecast can also be inaccurate, so disagreement does not directly give the correction. Parallax therefore learns from historical forecast-target pairs how to estimate the fine-forecast residual from disagreement.

Under squared loss, the best such estimate is the conditional mean of the fine residual given the disagreement. Directly estimating this high-dimensional map is unreliable at long horizons, but we observe that the conditional mean is approximately linear across resolutions (Section 3). We therefore use an affine corrector, whose optimum depends only on the first and second moments of residuals and disagreement. These moments determine how disagreement is distributed across fine positions and how much to trust each resolution. To estimate these moments stably without fitting a dense high-dimensional covariance, we exploit the geometry of nested block averaging: each aggregation factor removes a temporal band, yielding a structured covariance with only a few band- and resolution-specific variances. We then mix this structure with empirical moments through a single weight. Candidate resolutions are restricted by calendar-aligned factors, and the nested chain and mixing weight are selected on validation data. At test time, Parallax reduces to a single closed-form linear correction of the fine and coarse forecasts.

Summarization. As mentioned above, our proposed method Parallax is a post-hoc approach that corrects a frozen fine-resolution forecast from its disagreement with coarser-resolution forecasts of the same interval. The key idea is that Parallax turns cross-resolution disagreement into an estimate of the fine-resolution residual using historical forecast-target pairs, with resolutions selected from a calendar-guided nested aggregation chain. Our contributions are summarized as follows:

Disagreement as a residual measurement: We derive an exact decomposition of cross-resolution disagreement (Equation 1) that interprets it as a noisy measurement of the fine-forecast residual, reducing forecast correction to residual estimation. Empirically, the residual-disagreement relationship is approximately affine, motivating a closed-form corrector that maps disagreement to an estimate of the fine-forecast residual using their first and second moments (Section 3.1). The residual is then used to correct the final forecasting results (Section 3.4).

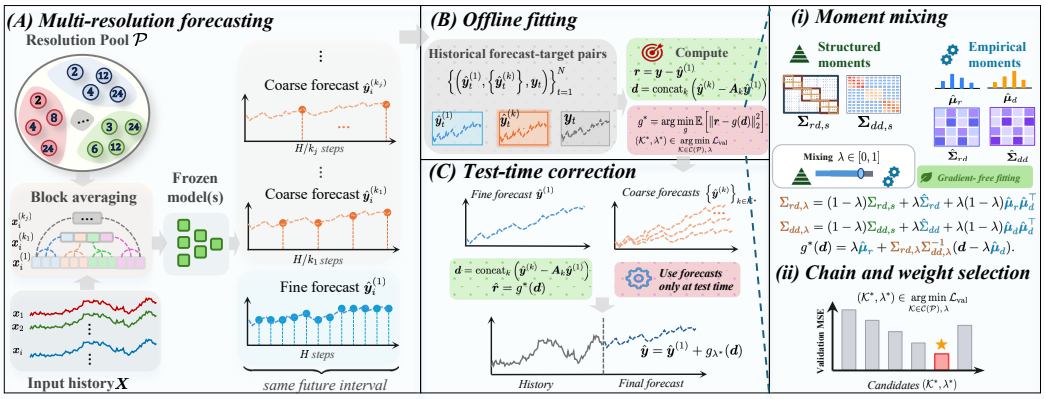


Figure 2: Overview of Parallax. **(A) Multi-resolution forecasting**: frozen forecaster(s) produce an H-step fine forecast and coarse H/k_j-step forecasts over the same future interval, for a nested chain $\mathcal{K} \in \mathcal{C}(\mathcal{P})$ of aggregation factors. **(B) Offline fitting on historical forecast-target pairs**: (i) structured and empirical moments of residuals $\boldsymbol{r}$ and cross-resolution disagreement $\boldsymbol{d}$ are mixed with weight λ into an affine corrector g_λ and (ii) the final chain $\mathcal{K}^*$ and weight λ^* are selected by a validation loss. **(C) Test-time correction**: the disagreement is computed from the forecasts alone, yielding $\hat{\boldsymbol{y}} = \hat{\boldsymbol{y}}^{(1)} + g_{\lambda^*}(\boldsymbol{d})$.

Multi-resolution moment estimator: We build a structured covariance from the nested aggregation chain: orthogonal temporal bands with one residual variance per band, and one error variance per coarse resolution, at most $2\lfloor \log_2 H \rfloor + 1$ parameters against $m(m+1)/2$ for a dense disagreement covariance. We mix it with empirical moments through a single weight λ; the mixture is a valid covariance and invertible for every $\lambda < 1$, even when the empirical covariance is singular (Section 3.2, Appendix A.4) This keeps the corrector stable when the fitting history is short relative to m: the band variances decide how block-level disagreement is distributed to fine positions, s_k decides how much each resolution is trusted, and the correction is robust across $\lambda \in [0.1, 0.95]$ (Appendix D.5). These provide stable parameter estimations for residual calculation described above.

Gradient-free, model-agnostic correction with consistent gains: Parallax needs no model weights and no gradient updates; fitting and chain selection take 44.7 s on a CPU, with 0.85 MB of test-time state (Appendix D.6). Extensive experiments on 13 datasets with 7 task-specific deep forecasters (such as OLinear and SiGMA) and 5 foundation models (such as LightGTS and Chronos-2) show that Parallax improves every backbone, reducing MSE by up to 67.3% (Section 4).

2 Problem Formulation

Time series forecasting. Let $\boldsymbol{x}_t \in \mathbb{R}^D$ contain D variables (channels) at time t. Given a lookback length L and prediction horizon H, we define the input and future ground truth at the first target time t as: $\boldsymbol{X}_t := [\boldsymbol{x}_{t-L}, \ldots, \boldsymbol{x}_{t-1}]^\top \in \mathbb{R}^{L\times D}, \boldsymbol{Y}_t := [\boldsymbol{x}_t, \ldots, \boldsymbol{x}_{t+H-1}]^\top \in \mathbb{R}^{H\times D}$. Time series forecasting learns a forecaster $f : \mathbb{R}^{L\times D} \to \mathbb{R}^{H\times D}$ that produces forecasts $\hat{\boldsymbol{Y}}_t := f(\boldsymbol{X}_t)$ before $\boldsymbol{Y}_t$ is observed. The task is univariate for $D = 1$ and multivariate for $D > 1$.

Fine/coarse forecasting. Let $(\boldsymbol{X}_t^{(k)}, \boldsymbol{Y}_t^{(k)})$ denote the time series constructed by averaging consecutive, non-overlapping blocks of k observations from the original series $(\boldsymbol{X}_t, \boldsymbol{Y}_t)$. The model $f^{(k)} : \mathbb{R}^{L^{(k)}\times D} \to \mathbb{R}^{(H/k)\times D}$ trained with $k = 1$ is named the fine forecaster, while the one with $k > 1$ is named the coarse forecaster. We focus on coarse forecasters with the set of indices $\mathcal{K} \subseteq \mathcal{P} = \{k \in \mathbb{N} : k \geq 2,\ k \mid H\}$ that ensures the forecast horizon contains an integer number of complete blocks. We define the block-averaging matrix $(\boldsymbol{A}_k)_{bh} := \frac{1}{k}\mathbf{1}\{(b-1)k < h \leq bk\}$, where $b \in \{1, \ldots, H/k\}, h \in \{1, \ldots, H\}$. This indicates $\boldsymbol{Y}_t^{(k)} = \boldsymbol{A}_k\boldsymbol{Y}_t$. For simplicity, we omit channel and origin indices and write $\boldsymbol{y} \in \mathbb{R}^H$ for the fine-resolution target and $\hat{\boldsymbol{y}}^{(k)} \in \mathbb{R}^{H/k}$ for the forecast at resolution $k \in \mathcal{K}$.

Definition 1 (Cross-resolution disagreement)**.** For aggregation factor $k \in \mathcal{K}$, the cross-resolution disagreement is defined as the difference between the coarse forecast and the corresponding ag-

gregation of the fine-resolution forecast: $\boldsymbol{d}_k := \hat{\boldsymbol{y}}^{(k)} - \boldsymbol{A}_k \hat{\boldsymbol{y}}^{(1)} \in \mathbb{R}^{H/k}$. The stacked disagreement across $\mathcal{K}$ is $\boldsymbol{d} := \mathrm{concat}_{k\in\mathcal{K}}\, \boldsymbol{d}_k \in \mathbb{R}^m$, where $m := \sum_{k\in\mathcal{K}} H/k$ is the total number of disagreement components across all coarse resolutions.

Definition 2 (Nested chain)**.** A nonempty set $\mathcal{K} = \{k_1, \ldots, k_J\} \subseteq \mathcal{P}$ with $k_1 < \cdots < k_J$ is a *nested chain* if $k_j \mid k_{j+1}$ for $1 \le j < J$. The set of all nested chains in $\mathcal{P}$ is $\mathcal{C}(\mathcal{P}) := \big\{\mathcal{K} \subseteq \mathcal{P} : \mathcal{K} \text{ is a nested chain}\big\}$. Because $\mathcal{P} = \{k \in \mathbb{N} : k \ge 2,\ k \mid H\}$, every $\mathcal{K} \in \mathcal{C}(\mathcal{P})$ satisfies $2 \le k_1$ and $k_J \mid H$. Since $k_{j+1} \ge 2k_j$, we have $2^J \le k_J \le H$, so $J \le \log_2 H$.

Cross-resolution forecast correction problem. Given frozen forecasters, we seek $g : \mathbb{R}^m \to \mathbb{R}^H$ that estimates the fine-resolution residual $\boldsymbol{r} := \boldsymbol{y} - \hat{\boldsymbol{y}}^{(1)}$ from cross-resolution disagreement $\boldsymbol{d}$. Fitted using historical forecast–target pairs, g yields the residual estimate and corrected fine-resolution forecast: $\hat{\boldsymbol{r}} := g(\boldsymbol{d}), \hat{\boldsymbol{y}} := \hat{\boldsymbol{y}}^{(1)} + \hat{\boldsymbol{r}}$. This enables correction at test time because $\boldsymbol{d}$ is observable from the forecasts without target labels. Our objective is to minimize $\mathbb{E}\|\mathbf{r} - g(\mathbf{d})\|_2^2$, equivalently the MSE of the corrected fine-resolution forecast. Over all measurable maps, the minimizer is the conditional mean $\mathbb{E}[\mathbf{r} \mid \mathbf{d}]$ (see Appendix A.1).

3 Our proposed method Parallax

Overview. Parallax, illustrated in Figure 2, corrects fine-resolution forecasts from a frozen forecaster by using cross-resolution disagreement across forecasts over the same future interval. For task-specific deep forecasters, we first train each of them independently per resolution: the fine forecaster f on the original series and each coarse forecaster $f^{(k)}$ on the block-averaged series, with their own lookback length $L^{(k)}$ and horizon H/k. Meanwhile, pretrained foundation models can process variable context lengths and forecast horizons, so they can forecast every resolution from the corresponding block-averaged history without retraining. In both cases, the forecasters are frozen before Parallax is applied to their outputs. We explain why disagreement reveals the fine-resolution residual and why the correction takes an affine form in Section 3.1 before presenting all components of Parallax in subsequent Sections 3.2 to 3.4.

3.1 From disagreement to an affine correction

A coarse forecast and an aggregated fine forecast predict the same block mean. Therefore, their difference at that block can reveal errors in both the fine and coarse forecasts. Let $\boldsymbol{\epsilon}_k := \hat{\boldsymbol{y}}^{(k)} - \boldsymbol{y}^{(k)}$ denote the coarse error at resolution k. Expanding the disagreement in Definition 1 gives

$$\boldsymbol{d}_k = \hat{\boldsymbol{y}}^{(k)} - \boldsymbol{A}_k \hat{\boldsymbol{y}}^{(1)} = \boldsymbol{A}_k\big(\boldsymbol{y} - \hat{\boldsymbol{y}}^{(1)}\big) + \big(\hat{\boldsymbol{y}}^{(k)} - \boldsymbol{A}_k \boldsymbol{y}\big) = \boldsymbol{A}_k \boldsymbol{r} + \boldsymbol{\epsilon}_k. \tag{1}$$

Concatenating across resolutions gives $\boldsymbol{d} = \boldsymbol{A}\boldsymbol{r} + \boldsymbol{\epsilon}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times H}$ concatenates the $\boldsymbol{A}_k$ row-wise and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ concatenates the $\boldsymbol{\epsilon}_k$, both in the same resolution order as $\boldsymbol{d}$. Since this disagreement also contains coarse-forecast errors, estimating $\boldsymbol{r}$ from $\boldsymbol{d}$ requires accounting for these errors and their dependence on $\boldsymbol{r}$.

To solve the cross-resolution forecast correction problem (defined in Section 2), estimating $\mathbb{E}[\mathbf{r} \mid \mathbf{d}]$ without further structure is nontrivial: $\mathbf{d}$ has m components across all coarse resolutions, leading to several hundred at long horizons H, whereas the fitting history contains only a short sequence of out-of-sample forecast–target pairs. We therefore examine whether the dependence between disagreement and the fine-resolution residual admits a simple affine approximation. For each resolution, Figure 3 relates block-level disagreement to the corresponding fine-resolution residuals. Despite substantial scatter, the binned means follow an approximately affine trend. This empirical structure motivates restricting g to the class of affine maps $\mathcal{G}_{\mathrm{aff}}$, reducing the problem to classical linear minimum mean-square error estimation (Steven et al., 1993). For finite second moments and $\boldsymbol{\Sigma}_{dd} := \mathrm{Cov}(\mathbf{d}) \succ 0$, the optimal affine corrector is

$$g^*(\boldsymbol{d}) = \boldsymbol{\mu}_r + \boldsymbol{\Sigma}_{rd}\boldsymbol{\Sigma}_{dd}^{-1}(\boldsymbol{d} - \boldsymbol{\mu}_d), \quad \boldsymbol{\mu}_r := \mathbb{E}[\mathbf{r}], \quad \boldsymbol{\mu}_d := \mathbb{E}[\mathbf{d}], \quad \boldsymbol{\Sigma}_{rd} := \mathrm{Cov}(\mathbf{r}, \mathbf{d}). \tag{2}$$

From $\boldsymbol{d} = \boldsymbol{A}\boldsymbol{r} + \boldsymbol{\epsilon}$, the covariance blocks in equation 2 satisfy $\boldsymbol{\Sigma}_{rd} = \boldsymbol{\Sigma}_{rr}\boldsymbol{A}^\top + \boldsymbol{\Sigma}_{r\epsilon}$, $\boldsymbol{\Sigma}_{dd} = \boldsymbol{A}\boldsymbol{\Sigma}_{rr}\boldsymbol{A}^\top + \boldsymbol{A}\boldsymbol{\Sigma}_{r\epsilon} + \boldsymbol{\Sigma}_{r\epsilon}^\top\boldsymbol{A}^\top + \boldsymbol{\Sigma}_{\epsilon\epsilon}$, where $\boldsymbol{\Sigma}_{r\epsilon} := \mathrm{Cov}(\mathbf{r}, \epsilon) \in \mathbb{R}^{H\times m}$ and $\boldsymbol{\Sigma}_{\epsilon\epsilon} := \mathrm{Cov}(\epsilon) \in \mathbb{R}^{m\times m}$. Moment estimation for $\boldsymbol{\Sigma}_{rr}, \boldsymbol{\Sigma}_{r\epsilon}, \boldsymbol{\Sigma}_{\epsilon\epsilon}$ are presented below.

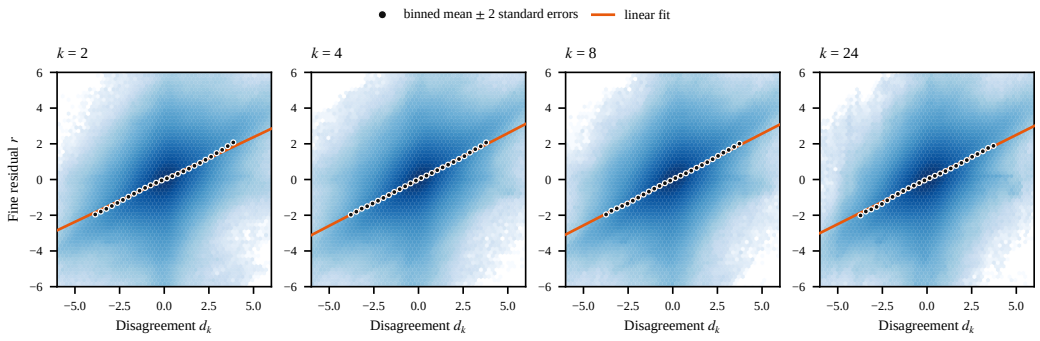


Figure 3: **Cross-resolution disagreement is informative about the fine-resolution residual** (OLinear on PEMS08, H=96). For each aggregation factor k, the figure shows the empirical relationship between disagreement and the corresponding fine-forecast residual. Black dots denote binned conditional means (± 2 standard errors), and the orange line is a linear fit. The binned means follow an approximately linear trend at each resolution, suggesting that an affine map provides a reasonable approximation to the relationship. Additional examples are provided in Appendix B.

3.2 Moment estimation and mixing

To approximate $\mathbf{\Sigma}_{rr}$, $\mathbf{\Sigma}_{\epsilon\epsilon}$, and $\mathbf{\Sigma}_{r\epsilon}$, we use a fitting set of N historical forecast-target pairs to estimate them in three steps, shown in Figure 2(B)(i): structured moments that encode the aggregation structure with few parameters (Section 3.2.1), empirical moments that impose no structure (Section 3.2.2), and a mixture of the two (Section 3.2.3).

3.2.1 Structured moments

Joint covariance of residuals and coarse errors. The residual covariance $\mathbf{\Sigma}_{rr}$ describes the variability and temporal dependence of fine-resolution forecast errors, helping translate cross-resolution disagreement into corrections at individual positions. However, estimating its $O(H^2)$ entries from limited fitting data can be noisy. We therefore leverage the geometric structure of a nested chain to construct a structured approximation using orthogonal temporal bands. These bands separate fluctuations removed by averaging from components retained at coarser resolutions. Assigning a single variance to each band and setting cross-band covariances to zero gives

$$\mathbf{\Sigma}_{rr,s} = \sum_{j=0}^{J} \tau_j^2 \boldsymbol{B}_j, \qquad \tau_0^2, \ldots, \tau_J^2 \geq 0, \tag{3}$$

where $J = |\mathcal{K}|$, $\boldsymbol{B}_j$ are the band projectors, and τ_j^2 is the residual variance shared by all directions within band j. This reduces the number of covariance parameters to $|\mathcal{K}| + 1 = O(\log H)$. We estimate the band variances from fitting residuals and provide the band construction in Appendix A.2.

To represent coarse forecast-error uncertainty, we use one variance s_k per resolution, shared across its H/k forecast positions, and set off-diagonal covariances to zero:

$$\mathbf{\Sigma}_{\epsilon\epsilon,s} := \mathrm{blockdiag}_{k\in\mathcal{K}}(s_k \boldsymbol{I}_{H/k}). \tag{4}$$

With other moments fixed, a larger s_k moderates the influence of disagreement at resolution k on the structured correction. We estimate these variances from historical coarse forecast errors using Algorithm 1. To keep the structured target simple, we set $\mathbf{\Sigma}_{r\epsilon,s} := \mathbf{0}_{H\times m}$, avoiding additional cross-covariance parameters. This model serves as a structured shrinkage target rather than an exact description of the residual–coarse-error distribution. Empirical moments can capture dependence omitted by this structure, including shared-error dependence, with their contribution controlled by the mixture weight (Section 3.2.3).

Fine-resolution correction from structured moments. The resulting residual–disagreement moments are $\mathbf{\Sigma}_{rd,s} := \mathbf{\Sigma}_{rr,s}\boldsymbol{A}^\top, \mathbf{\Sigma}_{dd,s} := \boldsymbol{A}\mathbf{\Sigma}_{rr,s}\boldsymbol{A}^\top + \mathbf{\Sigma}_{\epsilon\epsilon,s}$. With zero structured means and all $s_k > 0$, the structured correction is:

$$g_s(\boldsymbol{d}) := \mathbf{\Sigma}_{rd,s}\mathbf{\Sigma}_{dd,s}^{-1}\boldsymbol{d}. \tag{5}$$

Table 1: Comparison of forecasting results for 7 task-specific models on 13 datasets, without (w/o) and with Parallax. Horizons are $H \in \{24, 36, 48, 60\}$ for ILI, $H \in \{12, 24, 48, 96\}$ for PEMS, and $H \in \{96, 192, 336, 720\}$ otherwise. Values average horizons, with ETT and PEMS also averaged equally across their four datasets. Lower is better, bold marks the better paired result, and Δ reports percent change after Parallax, where negative means improvement. Green marks the best +Parallax value per dataset–metric. Pale yellow marks multiscale backbones. Full results are in Appendix D.1.

	ETT		PEMS		Solar-Energy		Electricity		Traffic		Weather		ILI	
Method	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
OLinear	0.360	0.376	0.095	0.187	0.215	0.217	0.160	0.249	0.451	0.247	0.238	0.260	1.816	0.816
+Parallax	**0.334**	**0.368**	**0.071**	**0.161**	**0.179**	**0.210**	**0.149**	**0.241**	**0.407**	**0.241**	**0.214**	**0.253**	**1.691**	**0.791**
Δ	**-7.11%**	**-2.26%**	**-25.56%**	**-14.13%**	**-16.67%**	**-3.35%**	**-6.42%**	**-3.18%**	**-9.78%**	**-2.36%**	**-10.14%**	**-3.03%**	**-6.92%**	**-3.01%**
DLinear	0.444	0.446	0.320	0.394	0.330	0.401	0.225	0.319	0.624	0.384	0.266	0.318	2.847	1.164
+Parallax	**0.365**	**0.399**	**0.140**	**0.249**	**0.219**	**0.298**	**0.185**	**0.289**	**0.449**	**0.308**	**0.221**	**0.282**	**2.516**	**1.103**
Δ	**-17.88%**	**-10.47%**	**-56.14%**	**-36.90%**	**-33.61%**	**-25.63%**	**-17.87%**	**-9.34%**	**-28.15%**	**-19.84%**	**-16.96%**	**-11.42%**	**-11.64%**	**-5.31%**
iTransformer	0.386	0.401	0.121	0.227	0.238	0.268	0.176	0.267	0.430	0.289	0.260	0.280	2.321	0.980
+Parallax	**0.349**	**0.382**	**0.081**	**0.179**	**0.189**	**0.243**	**0.159**	**0.253**	**0.388**	**0.271**	**0.223**	**0.263**	**1.990**	**0.916**
Δ	**-9.67%**	**-4.88%**	**-33.21%**	**-21.08%**	**-20.48%**	**-9.62%**	**-9.54%**	**-4.98%**	**-9.66%**	**-6.23%**	**-13.98%**	**-6.31%**	**-14.24%**	**-6.53%**
PatchTST	0.380	0.400	0.218	0.319	0.257	0.298	0.206	0.295	0.482	0.308	0.259	0.281	2.162	0.899
+Parallax	**0.341**	**0.378**	**0.099**	**0.207**	**0.188**	**0.254**	**0.171**	**0.270**	**0.421**	**0.292**	**0.221**	**0.265**	**1.840**	**0.859**
Δ	**-10.28%**	**-5.39%**	**-54.62%**	**-35.21%**	**-26.82%**	**-14.90%**	**-16.63%**	**-8.62%**	**-12.61%**	**-5.23%**	**-14.88%**	**-5.95%**	**-14.87%**	**-4.41%**
SiGMA	0.369	0.388	0.201	0.291	0.258	0.310	0.174	0.268	0.461	0.302	0.256	0.281	2.254	0.925
+Parallax	**0.341**	**0.376**	**0.081**	**0.179**	**0.188**	**0.250**	**0.160**	**0.255**	**0.414**	**0.285**	**0.217**	**0.262**	**1.935**	**0.863**
Δ	**-7.58%**	**-2.96%**	**-59.55%**	**-38.57%**	**-27.17%**	**-19.13%**	**-7.82%**	**-4.72%**	**-10.30%**	**-5.81%**	**-15.00%**	**-6.66%**	**-14.12%**	**-6.71%**
TimeMixer	0.375	0.393	0.190	0.287	0.216	0.294	0.185	0.274	0.510	0.304	0.245	0.274	2.422	0.958
+Parallax	**0.342**	**0.377**	**0.116**	**0.224**	**0.188**	**0.256**	**0.164**	**0.258**	**0.426**	**0.285**	**0.214**	**0.260**	**2.019**	**0.886**
Δ	**-8.96%**	**-3.88%**	**-38.79%**	**-21.98%**	**-12.94%**	**-12.92%**	**-11.55%**	**-5.80%**	**-16.44%**	**-6.25%**	**-12.88%**	**-5.16%**	**-16.67%**	**-7.54%**
Scaleformer	1.597	0.841	0.136	0.237	0.237	0.293	0.236	0.344	0.716	0.412	0.359	0.384	3.550	1.224
+Parallax	**0.522**	**0.511**	**0.106**	**0.209**	**0.183**	**0.244**	**0.207**	**0.317**	**0.610**	**0.362**	**0.226**	**0.284**	**3.248**	**1.181**
Δ	**-67.32%**	**-39.25%**	**-22.11%**	**-11.76%**	**-22.74%**	**-16.78%**	**-12.02%**	**-7.83%**	**-14.88%**	**-12.12%**	**-37.00%**	**-26.14%**	**-8.50%**	**-3.49%**

Here, the residual covariance guides how block-level disagreement is translated into corrections at individual positions, while the coarse-error covariance moderates the influence of uncertain coarse forecasts. Covariance validity is established in Appendix A.3, Proposition A.2.

3.2.2 Empirical moments

The structured target regularizes estimation, but its zero means and covariance restrictions may miss forecast bias and shared-error dependence. The empirical component complements this target by estimating the means and covariances directly from the same fitting data, without those restrictions:

$$\hat{\boldsymbol{\mu}}_r := \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{r}_i, \qquad \hat{\boldsymbol{\Sigma}}_{rd} := \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_r)(\boldsymbol{d}_i - \hat{\boldsymbol{\mu}}_d)^\top,$$
$$\hat{\boldsymbol{\mu}}_d := \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{d}_i, \qquad \hat{\boldsymbol{\Sigma}}_{dd} := \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{d}_i - \hat{\boldsymbol{\mu}}_d)(\boldsymbol{d}_i - \hat{\boldsymbol{\mu}}_d)^\top. \tag{6}$$

Fine-resolution correction from empirical moments. When $\hat{\boldsymbol{\Sigma}}_{dd} \succ 0$, substituting these moments into equation 2 gives: $g_e(\boldsymbol{d}) := \hat{\boldsymbol{\mu}}_r + \hat{\boldsymbol{\Sigma}}_{rd}\hat{\boldsymbol{\Sigma}}_{dd}^{-1}(\boldsymbol{d} - \hat{\boldsymbol{\mu}}_d)$. These estimates can be noisy with limited fitting data, and inversion may amplify estimation errors when $\hat{\boldsymbol{\Sigma}}_{dd}$ is poorly conditioned.

3.2.3 Moment mixing

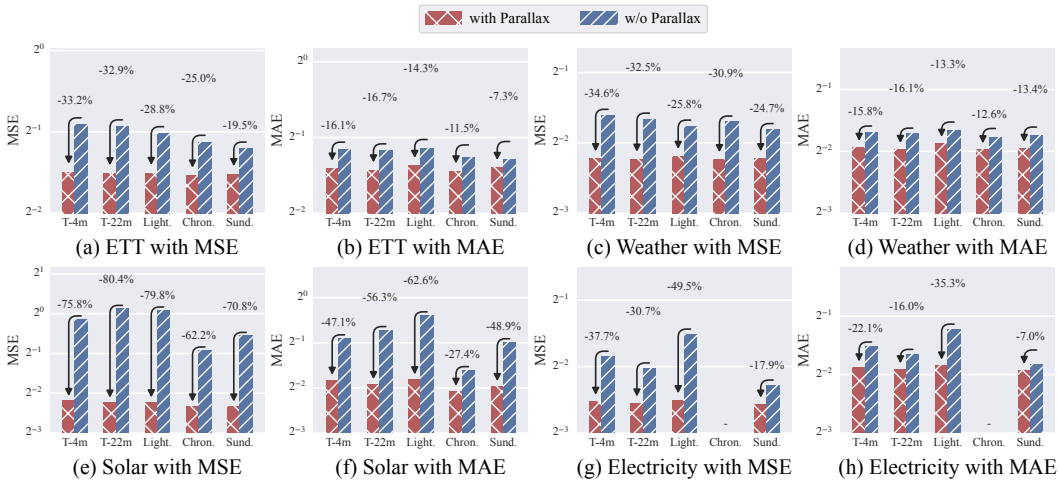
We use empirical moments to capture forecast bias and shared-error dependence omitted by the structured target, while shrinking toward that target to limit estimation noise. Specifically, we combine their joint means and uncentered second moments of $(\mathbf{r}, \mathbf{d})$ with weights $1 - \lambda$ and λ, respectively, where $\lambda \in [0, 1]$. Mixing moments rather than corrections regularizes the disagreement covariance before inversion, in the spirit of covariance shrinkage (Ledoit & Wolf, 2004). Since the structured means are zero, the mixed means are $\lambda\hat{\boldsymbol{\mu}}_r$ and $\lambda\hat{\boldsymbol{\mu}}_d$, with covariance blocks

$$\boldsymbol{\Sigma}_{rd,\lambda} = (1-\lambda)\boldsymbol{\Sigma}_{rd,s} + \lambda\hat{\boldsymbol{\Sigma}}_{rd} + \lambda(1-\lambda)\hat{\boldsymbol{\mu}}_r\hat{\boldsymbol{\mu}}_d^\top, \quad \boldsymbol{\Sigma}_{dd,\lambda} = (1-\lambda)\boldsymbol{\Sigma}_{dd,s} + \lambda\hat{\boldsymbol{\Sigma}}_{dd} + \lambda(1-\lambda)\hat{\boldsymbol{\mu}}_d\hat{\boldsymbol{\mu}}_d^\top. \tag{7}$$

Substituting these moments into equation 2 gives the Parallax corrector

$$g_\lambda(\boldsymbol{d}) := \lambda\hat{\boldsymbol{\mu}}_r + \boldsymbol{\Sigma}_{rd,\lambda}\boldsymbol{\Sigma}_{dd,\lambda}^{-1}(\boldsymbol{d} - \lambda\hat{\boldsymbol{\mu}}_d). \tag{8}$$

When all $s_k > 0$, $\boldsymbol{\Sigma}_{dd,\lambda} \succ 0$ for $\lambda < 1$, even if $\hat{\boldsymbol{\Sigma}}_{dd}$ is singular; $\lambda = 1$ requires $\hat{\boldsymbol{\Sigma}}_{dd} \succ 0$. The derivation and validity proof are given in Appendix A.4.



(a) ETT with MSE (b) ETT with MAE (c) Weather with MSE (d) Weather with MAE

(e) Solar with MSE (f) Solar with MAE (g) Electricity with MSE (h) Electricity with MAE

Figure 4: Benefit of incorporating Parallax in foundation models, shown with bars for means over forecast horizons ($H \in \{96, 192, 336, 720\}$; ETT also averaged over its four datasets; log scale). Arrows are labeled with the percent change after Parallax. T-4m/T-22m: Toto-2.0-4m/22m; Light.: LightGTS; Chron.: Chronos-2; Sund.: Sundial Base 128M. Model-dataset pairs with no documented pretraining overlap are reported. Full results are in Appendix D.

3.3 Chain and Weight Selection

The selected chain specifies which coarse forecasts are compared with the corresponding block-averaged fine forecast to form the disagreement vector. The same resolutions define the temporal bands used to model residual covariance, while λ controls the contribution of empirical moments. The candidate pool and validation selection correspond to Figure 2(A) and (B)(ii), respectively.

Calendar-guided candidate pool. Let $\mathcal{P}_0$ contain the available factors $k \geq 2$ dividing H. For a prescribed integer calendar period p_{cal} in fine steps, define $\mathcal{P}_{\text{cal}} := \{k \in \mathcal{P}_0 : k \mid \gcd(p_{\text{cal}}, H)\}$. We use $\mathcal{P} := \mathcal{P}_{\text{cal}}$ when nonempty, and $\mathcal{P} := \mathcal{P}_0$ otherwise or when no period is specified. Calendar divisibility ensures an integer number of coarse steps per cycle (De Livera et al., 2011; Athanasopoulos et al., 2017), while divisibility by H ensures complete forecast blocks.

Validation-based chain and weight selection. For each nested chain $\mathcal{K} \in \mathcal{C}(\mathcal{P})$ (Definition 2), including single-factor chains, we estimate moments on the same fitting set and select

$$(\mathcal{K}^*, \lambda^*) = \underset{\mathcal{K} \in \mathcal{C}(\mathcal{P}),\, \lambda}{\arg\min}\ \mathcal{L}_{\text{val}}, \tag{9}$$

where $\mathcal{L}_{\text{val}}$ is the loss of corrected fine-resolution forecasts on the validation set. The selected chain, weight, and fitted corrector remain fixed at deployment (see Appendix C.5 for calendar periods and fitting details).

3.4 Test-time Correction

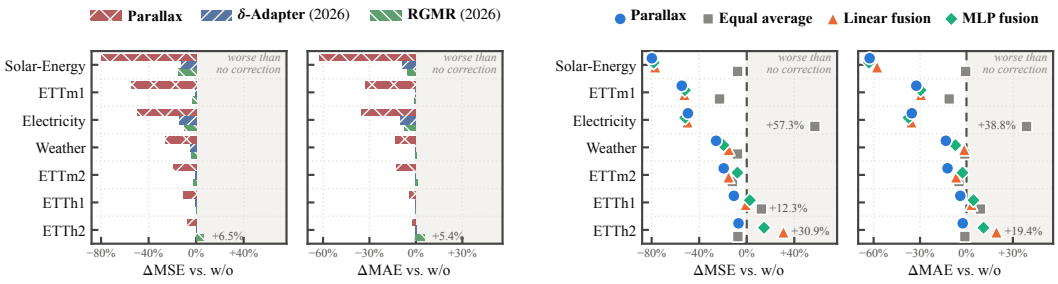
At each test origin, the frozen forecasters produce $\hat{\boldsymbol{y}}^{(1)}$ and $\hat{\boldsymbol{y}}^{(k)}$ for $k \in \mathcal{K}^*$. We form the disagreement $\boldsymbol{d}$ as in Definition 1 and obtain the corrected fine-resolution forecast (Figure 2(C)):

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{y}}^{(1)} + g_{\lambda^*}(\boldsymbol{d}). \tag{10}$$

The correction uses only these forecasts, with no target labels, gradient updates, or refitting.

4 Experiments

Datasets. We employ 13 standard forecasting benchmarks: Electricity, Weather, Traffic, Solar-Energy, ILI, four ETT datasets (ETTh1, ETTh2, ETTm1, ETTm2), and four PEMS datasets (PEMS03, PEMS04, PEMS07, PEMS08) (Wang et al., 2024; Yue et al., 2025) (Appendix C.2).



(a) Existing post-hoc forecast correctors

(b) Parallax and its variants

Figure 5: Relative change in MSE and MAE of LightGTS w.r.t. no correction (w/o), averaged over horizons, for (a) existing post-hoc forecast correctors and (b) Parallax and its variants. The shaded region marks results worse than no correction. Full results are in Tables 10 and 11.

Backbones. Parallax is a model-agnostic forecast correction method. We study 7 SOTA long-term forecasting models: OLinear (Yue et al., 2025) and DLinear (Zeng et al., 2023) (linear-based models), PatchTST (Nie et al., 2023) and iTransformer (Liu et al., 2024) (Transformer-based models), and SiGMA (Lee et al., 2026), TimeMixer (Wang et al., 2024), and Scaleformer(Shabani et al., 2023) (multiscale processing models). We also use 5 frozen foundation models: LightGTS, Chronos-2 (Ansari et al., 2025), Sundial Base 128M (Liu et al., 2025), Toto-2.0-22m and Toto-2.0-4m (Khwaja et al., 2026). Appendix C.3 provides an overview of the backbones.

Further studies. Due to space constraint, Appendices D.5 and D.6 show that Parallax is robust to the parameter λ and adds little computational overhead.

4.1 Main Results

Task-specific deep forecasters. Table 1 compares results without and with Parallax, which reduces MSE in 363 of 364 model–dataset–horizon cases and MAE in 362 (full results in Appendix D.1). Average MSE and MAE reductions are particularly large for less accurate backbones: 26.4% and 16.8% for Scaleformer (67.3% and 39.3% on ETT) and 26.0% and 17.0% for DLinear. Improvements remain substantial for the strongest backbone, OLinear (11.8% and 4.5%), and for multiscale backbones such as SiGMA (20.2% and 12.1%), confirming that cross-resolution disagreement provides information not exploited even by backbones with internal multiscale processing.

Foundation models. Figure 4 shows the results for five pretrained foundation models. Parallax reduces both MSE and MAE in all 34 retained model–dataset cases, by 33.2% and 18.7% on average, respectively. Notably, Toto-2.0-4m with Parallax outperforms the five-times larger Toto-2.0-22m without Parallax on all seven datasets in both metrics, showing that Parallax can complement model scaling with a lightweight, gradient-free post-hoc adaptation.

4.2 Comparison with Existing Post-hoc Correctors and Alternative Estimators

Existing post-hoc correctors. We compare Parallax with two post-hoc correctors: δ-**Adapter** (Liang et al., 2026), which learns bounded corrections, and **RGMR** (Gao et al., 2026), which uses residual-based refinement. Figure 5(a) shows that, despite using more complex learned correction functions, δ-Adapter and RGMR reduce MSE by only 5.5% on average. Meanwhile, Parallax achieves a 45.2% reduction and consistently gets the lowest MSE and MAE in all 28 dataset–horizon tasks (Table 10). This gap is followed by higher computational costs: δ-Adapter uses more test-time memory, while RGMR is substantially slower to fit and run (Appendix D.6).

Effects of estimation model. To test whether the gains simply come from adding coarse forecasts, we compare Parallax with **equal averaging**, **linear fusion**, and **MLP fusion** under the same backbone, resolutions, and fitting data (Appendix D.3). Figure 5(b) shows that the alternatives can fail substantially: linear fusion increases MSE by 30.9% on ETTh2, while MLP fusion also degrades both ETTh1 and ETTh2. Conversely, Parallax consistently improves both MSE and MAE across all

seven datasets. Despite its simple affine form, Parallax outperforms learned fusions, showing that how cross-resolution information is exploited matters more than estimator complexity.

4.3 Ablation study

To study the contribution of each component, we remove (1) the temporal bands in equation 3; (2) the resolution-specific variance s_k in equation 4; and (3) the empirical moments. We fix the selected chain, retain λ^* for (1–2), and set $\lambda = 0$ for (3). Table 2 shows that, with all three components, the full Parallax performs best on all four datasets. Removing s_k causes the largest degradation, while temporal bands and empirical moments provide consistent gains.

Table 2: Component ablation, MSE averaged over horizons (full results in Appendix D.4).

Dataset	ETTh1	ETTm2	Weather	PEMS04
Parallax	**0.4061**	**0.2474**	**0.2145**	**0.0716**
w/o temporal bands	0.4216	0.2524	0.2226	0.0735
w/o resolution-specific s_k	0.5170	0.3081	0.2872	0.0746
w/o empirical moments ($\lambda = 0$)	0.4071	0.2483	0.2168	0.0718
OLinear (w/o Parallax)	0.4268	0.2707	0.2385	0.0908

5 Related work

Combining information across scales and domains. Different temporal scales and frequency components reveal complementary patterns in a time series, motivating methods that bring this information together. In the statistical literature, temporal forecast reconciliation (Athanasopoulos et al., 2017) combines forecasts to ensure coherence across aggregation levels. In deep forecasting, the emphasis shifts to learning how to use this information within the forecaster itself. TimeMixer (Wang et al., 2024) mixes information across resolutions, while SiGMA (Lee et al., 2026) combines learned smooth and residual representations. WaveletMixer (Zhang et al., 2025) and WPMixer (Murad et al., 2025) decompose the same input history into wavelet components and recombine their predictions in the time domain. Scaleformer (Shabani et al., 2023) also uses forecasts across scales, iteratively refining them from coarse to fine with a shared forecasting network. These approaches build cross-scale processing into the trained forecasting pipeline. Parallax is a post-hoc, model-agnostic method for frozen task-specific deep forecasters and foundation models. It targets fine-resolution forecast correction rather than a coherent forecast hierarchy, and operates outside the forecasting architecture rather than building multiscale processing into it. It can therefore be applied as a plug-in correction even to multiscale forecasters such as Scaleformer and SiGMA.

Post-hoc forecast correction. δ-Adapter (Liang et al., 2026) trains a small neural network to make additive or multiplicative edits to a frozen forecaster's inputs or outputs. Its bounded edits are learned by minimizing forecasting loss. RGMR (Gao et al., 2026) generates forecasts from smoothed inputs and fits per-level Ridge predictors to short-context forecast errors. At inference, it blends long-context forecasts from coarse to fine, weighting residual corrections predicted from past observations and errors. This requires residual patterns learned on short contexts to transfer to the long-context forecasts being sequentially refined. Parallax instead uses the observable disagreement between coarse forecasts and the aggregated fine forecast. Its resolution-structured and empirical moments jointly model how these disagreements relate to fine-forecast error, accounting for dependence across resolutions. Fitted from historical forecast–target pairs, they determine a single correction without training an adapter network or applying residual predictors sequentially.

6 Conclusion

In this paper, we introduce Parallax, a model-agnostic post-hoc method for correcting fine-resolution time series forecasts. Parallax treats cross-resolution disagreement as a noisy measurement of the fine-resolution residual, since it decomposes into the aggregated residual and coarse forecast error. It estimates the residual with an affine corrector whose moments combine a structured covariance built from a nested aggregation chain with empirical moments from historical forecast–target pairs. Candidate aggregation factors are restricted by the forecast horizon and calendar cycle, while the chain and moment-mixing weight are selected by validation loss. Once fitted, Parallax uses forecasts alone at test time, without updating backbone weights. Extensive experiments demonstrate consistent improvements for both task-specific deep forecasters and pretrained foundation models.

AI USE STATEMENT

We used generative AI tools to improve the language and readability of paragraphs from a drafted version by the authors. The authors carefully reviewed all suggested changes and revised them where necessary to preserve the intended meaning and research content. The authors take full responsibility for the contents of this submission.

REPRODUCIBILITY STATEMENT

We use existing time series forecasting benchmarks, described in detail in Appendix C.2. Proofs and algorithm details are discussed in Appendix A. We provide the experimental settings, experimental protocol, hyperparameters, and model architectures in Appendix C. The full experimental results can be found in Appendix D.

REFERENCES

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, et al. Chronos-2: From univariate to universal forecasting. *arXiv preprint arXiv:2510.15821*, 2025.

George Athanasopoulos, Rob J Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. Forecasting with temporal hierarchies. *Eur. J. Oper. Res.*, 262(1):60–74, 2017.

Coryn AL Bailer-Jones. Estimating distances from parallaxes. *Publ. Astron. Soc. Pac.*, 127(956): 994–1009, 2015.

Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.*, 106(496):1513–1527, 2011.

Wentao Gao, Jiuyong Li, Lin Liu, Thuc Duy Le, Jixue Liu, Yanchang Zhao, and Yun Chen. Residual-guided multi-resolution refinement of foundation models: A case study in drought forecasting. In *ICML*, 2026. URL https://openreview.net/forum?id=ERCbhCrHo3.

Rob J Hyndman, George Athanasopoulos, et al. *Forecasting: principles and practice*, volume 2. OTexts Melbourne, Australia, 2014.

Jooyoung Jeon, Anastasios Panagiotelis, and Fotios Petropoulos. Probabilistic forecast reconciliation with applications to wind power and electric load. *Eur. J. Oper. Res.*, 279(2):364–379, 2019.

Emaad Khwaja, Chris Lettieri, Gerald Woo, Eden Belouadah, Marc Cenac, Guillaume Jarry, Enguerrand Paquin, Xunyi Zhao, Viktoriya Zhukov, Othmane Abou-Amal, Chenghao Liu, Ameet Talwalkar, and David Asker. Toto 2.0: Time series forecasting enters the scaling era, 2026. URL https://arxiv.org/abs/2605.20119.

Nikolaos Kourentzes, Fotios Petropoulos, and Juan R Trapero. Improving forecasting by estimating time series structural components across multiple frequencies. *Int. J. Forecast.*, 30(2):291–302, 2014.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *SIGIR*, pp. 95–104, 2018.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.*, 88(2):365–411, 2004.

Cheonwoo Lee, Dooho Lee, Doyun Choi, and Jaemin Yoo. Generalizing multi-scale time-series modeling with a single operator. In *ICML*, 2026. URL https://openreview.net/forum?id=hxNN3xhuS9.

Daojun Liang, Qi Li, Yinglong Wang, Jing Chen, Hu Zhang, Xiaoxiao Cui, Qizheng Wang, and Shuo Li. The forecast after the forecast: A post-processing shift in time series. In *ICLR*, 2026. URL https://openreview.net/forum?id=syfWdclGE1.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *ICLR*, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

Yong Liu, Guo Qin, Zhiyuan Shi, Zhi Chen, Caiyin Yang, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Sundial: A family of highly capable time series foundation models. In *ICML*, 2025. URL https://openreview.net/forum?id=LO7ciRpjI5.

Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. Wpmixer: Efficient multi-resolution mixing for long-term time series forecasting. In *AAAI*, 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. In *ICLR*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Peter Nystrup, Erik Lindström, Pierre Pinson, and Henrik Madsen. Temporal hierarchies with autocorrelation for load forecasting. *Eur. J. Oper. Res.*, 280(3):876–888, 2020.

Fotios Petropoulos, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J Bessa, Jakub Bijak, John E Boylan, et al. Forecasting: theory and practice. *Int. J. Forecast.*, 38(3):705–871, 2022.

Syama Sundar Rangapuram, Shubham Kapoor, Rajbir Singh Nirwan, Pedro Mercado, Tim Januschowski, Yuyang Wang, and Michael Bohlke-Schneider. Coherent probabilistic forecasting of temporal hierarchies. In *AISTATS*, pp. 9362–9376, 2023.

Mohammad Amin Shabani, Amir H. Abdi, Lili Meng, and Tristan Sylvain. Scaleformer: Iterative multi-scale refining transformers for time series forecasting. In *ICLR*, 2023. URL https://openreview.net/forum?id=sCrnllCtjoE.

Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *AAAI*, volume 34, pp. 914–921, 2020.

Evangelos Spiliotis, Fotios Petropoulos, Nikolaos Kourentzes, and Vassilios Assimakopoulos. Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Applied Energy*, 261:114339, 2020.

M Kay Steven et al. Fundamentals of statistical signal processing. *PTR Prentice-Hall, Englewood Cliffs, NJ*, 10(151045):148, 1993.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *ICLR*, 2024. URL https://openreview.net/forum?id=7oLshfEIC2.

Yihang Wang, Yuying Qiu, Peng Chen, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. LightGTS: A lightweight general time series forecasting model. In *ICML*, 2025. URL https://openreview.net/forum?id=Z5FJsp1U3Z.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Chen Wang, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. *TPAMI*, 48(10): 11642–11662, 2026.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *ICML*, 2024. URL https://openreview.net/forum?id=Yd8eHMY1wz.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *NeurIPS*, 2021. URL https://openreview.net/forum?id=J4gRj6d5Qm.

Wenzhen Yue, Yong Liu, Hao Wang, Haoxuan Li, Xianghua Ying, Ruohao Guo, Bowei Xing, and Ji Shi. OLinear: A linear model for time series forecasting in orthogonally transformed domain. In *NeurIPS*, 2025. URL https://openreview.net/forum?id=DAyKP1tvwI.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *AAAI*, 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 *AAAI*, 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 *AAAI*, pp. 11106–11115, 2021. URL https://ojs.aaai.org/index.php/AAAI/article/view/17325.

APPENDIX

We first provide the proofs of our propositions and technical details in Section A. We further study the residual-disagreement relation across various datasets in Section B. Experimental details, including experimental settings, data description, backbone model architectures, experimental protocol, and hyperparameters, are provided in Section C. The full details of our results, ablation studies, hyperparameter sensitivity, and efficiency analysis can be found in Section D. We discuss our limitations in Section E.

A PROOFS AND TECHNICAL DETAILS

A.1 RESIDUAL ESTIMATION AND FORECAST CORRECTION

Since $\hat{\boldsymbol{y}} := \hat{\boldsymbol{y}}^{(1)} + g(\boldsymbol{d})$ and $\boldsymbol{r} := \boldsymbol{y} - \hat{\boldsymbol{y}}^{(1)}$, for every map g,

$$\mathbb{E}\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 = \mathbb{E}\|\mathbf{r} - g(\mathbf{d})\|_2^2. \tag{11}$$

Forecast MSE is this quantity divided by H, so forecast correction and residual estimation have the same minimizers over any common class of maps. For square-integrable residuals, the minimizer over all measurable maps is $\mathbb{E}[\mathbf{r} \mid \mathbf{d}]$.

Optimal affine correction. For $g(\boldsymbol{d}) = \boldsymbol{a} + \boldsymbol{W}\boldsymbol{d}$, finite second moments and $\boldsymbol{\Sigma}_{dd} \succ 0$, minimizing the squared-error risk first over $\boldsymbol{a}$ and then over $\boldsymbol{W}$ gives

$$\boldsymbol{a} = \boldsymbol{\mu}_r - \boldsymbol{W}\boldsymbol{\mu}_d, \qquad \boldsymbol{W}\boldsymbol{\Sigma}_{dd} = \boldsymbol{\Sigma}_{rd}.$$

The risk is a convex quadratic, and $\boldsymbol{\Sigma}_{dd} \succ 0$ gives the unique solution in equation 2.

A.2 RESIDUAL COVARIANCE: CONSTRUCTION AND ESTIMATION

Band construction. Given the block-averaging matrix $\boldsymbol{A}_k$ for each $k \in \mathcal{K}$, define the orthogonal projector $\boldsymbol{P}_k := \boldsymbol{A}_k^\dagger \boldsymbol{A}_k = k\boldsymbol{A}_k^\top \boldsymbol{A}_k \in \mathbb{R}^{H\times H}$, where $\dagger$ denotes the Moore–Penrose pseudo-inverse. The projector $\boldsymbol{P}_k$ replaces each aligned block of k positions by its mean. For a nested chain $\mathcal{K} = \{k_1, \ldots, k_J\}$, let define:

$$\boldsymbol{B}_0 := \boldsymbol{I}_H - \boldsymbol{P}_{k_1}, \qquad \boldsymbol{B}_j := \boldsymbol{P}_{k_j} - \boldsymbol{P}_{k_{j+1}}\ (1 \le j < J), \qquad \boldsymbol{B}_J := \boldsymbol{P}_{k_J}. \tag{12}$$

The first band captures within-block fluctuations removed at k_1; intermediate bands capture variation retained at k_j but removed at k_{j+1}; the last captures variation retained at the coarsest resolution.

Proposition A.1 (Validity of the band decomposition)**.** *For a nested chain $\mathcal{K}$, the matrices $\boldsymbol{B}_0, \ldots, \boldsymbol{B}_J$ are mutually orthogonal projectors with positive ranks $\rho_j := \operatorname{rank}(\boldsymbol{B}_j)$ and sum to $\boldsymbol{I}_H$.*

Proof. The disjoint rows of $\boldsymbol{A}_k$ give $\boldsymbol{A}_k\boldsymbol{A}_k^\top = k^{-1}\boldsymbol{I}_{H/k}$, so $\boldsymbol{A}_k^\dagger = k\boldsymbol{A}_k^\top$ and $\boldsymbol{P}_k$ is symmetric and idempotent. For $k_\ell \mid k_j$, the projection spaces are nested, giving $\boldsymbol{P}_{k_\ell}\boldsymbol{P}_{k_j} = \boldsymbol{P}_{k_j}\boldsymbol{P}_{k_\ell} = \boldsymbol{P}_{k_j}$. Expanding the products in equation 12 shows that each $\boldsymbol{B}_j$ is symmetric and idempotent and that $\boldsymbol{B}_\ell\boldsymbol{B}_j = \boldsymbol{0}$ for $\ell \neq j$. Their sum telescopes to $\boldsymbol{I}_H$. Taking traces gives $\rho_0 = H - H/k_1$, $\rho_j = H/k_j - H/k_{j+1}$ for $1 \le j < J$, and $\rho_J = H/k_J$, which are positive. □

Band-variance estimation. We assign one variance to each band, as in equation 3. For fitting residuals $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$, estimate

$$\hat{\tau}_j^2 := \frac{1}{N\rho_j}\sum_{i=1}^{N}\|\boldsymbol{B}_j\boldsymbol{r}_i\|_2^2, \qquad j = 0, \ldots, J. \tag{13}$$

Under the zero-mean structured model $\mathbb{E}[\mathbf{r}_i\mathbf{r}_i^\top] = \sum_{\ell=0}^{J}\tau_\ell^2\boldsymbol{B}_\ell$, this estimator is unbiased because

$$\mathbb{E}\|\boldsymbol{B}_j\mathbf{r}_i\|_2^2 = \operatorname{Tr}\big(\boldsymbol{B}_j\mathbb{E}[\mathbf{r}_i\mathbf{r}_i^\top]\big) = \rho_j\tau_j^2.$$

This argument does not require independence across fitting pairs. Since the band projectors and their estimated coefficients are positive semidefinite and nonnegative, respectively, $\boldsymbol{\Sigma}_{rr,s} := \sum_{j=0}^{J} \hat{\tau}_j^2 \boldsymbol{B}_j \succeq 0$. Strict divisibility gives $k_{j+1} \geq 2k_j$ and $k_1 \geq 2$, so $2^J \leq k_J \leq H$. Thus the structured residual covariance uses $J + 1 \leq \lfloor \log_2 H \rfloor + 1$ variance parameters.

The band variances and coarse-error variances are estimated using uncentered mean squares, as shown in Algorithm 1. Consequently, systematic forecast bias is absorbed into these structured variances. The zero structured means are a modeling choice for the shrinkage target, not an assumption that the actual forecast errors are unbiased. The unstructured moments estimate forecast bias and shared-error dependence explicitly.

Algorithm 1 Structured moment estimation

Input: fitting residuals $\{\boldsymbol{r}_i\}_{i=1}^N$; coarse errors $\{\boldsymbol{\epsilon}_{i,k}\}_{i,k}$; nested chain $\mathcal{K}$
Output: $\boldsymbol{\Sigma}_{rd,s}$ and $\boldsymbol{\Sigma}_{dd,s}$

Construct $\boldsymbol{A}$ by stacking $\boldsymbol{A}_k$, $k \in \mathcal{K}$
Construct $\boldsymbol{B}_0, \ldots, \boldsymbol{B}_J$ using equation 12
Estimate $\hat{\tau}_0^2, \ldots, \hat{\tau}_J^2$ using equation 13
$\boldsymbol{\Sigma}_{rr,s} \leftarrow \sum_{j=0}^{J} \hat{\tau}_j^2 \boldsymbol{B}_j$
for $k \in \mathcal{K}$ **do**
    $s_k \leftarrow [N(H/k)]^{-1} \sum_{i=1}^{N} \|\boldsymbol{\epsilon}_{i,k}\|_2^2$
end for
$\boldsymbol{\Sigma}_{\epsilon\epsilon,s} \leftarrow \mathrm{blockdiag}_{k\in\mathcal{K}}(s_k \boldsymbol{I}_{H/k})$
$\boldsymbol{\Sigma}_{rd,s} \leftarrow \boldsymbol{\Sigma}_{rr,s}\boldsymbol{A}^\top$
$\boldsymbol{\Sigma}_{dd,s} \leftarrow \boldsymbol{A}\boldsymbol{\Sigma}_{rr,s}\boldsymbol{A}^\top + \boldsymbol{\Sigma}_{\epsilon\epsilon,s}$
return $(\boldsymbol{\Sigma}_{rd,s}, \boldsymbol{\Sigma}_{dd,s})$

The estimates ensure $s_k \geq 0$, but not necessarily $s_k > 0$. The invertibility guarantees below are conditional on the stated strict-positivity assumptions.

A.3 Validity of structured moments

Proposition A.2 (Validity of the structured joint moments)**.** *Let $\boldsymbol{\Sigma}_{rr,s} = \sum_{j=0}^{J} \tau_j^2 \boldsymbol{B}_j$ with $\tau_j^2 \geq 0$, and let $\boldsymbol{\Sigma}_{\epsilon\epsilon,s}$ be given by equation 4 with $s_k \geq 0$. Then the structured joint covariance of $(\mathbf{r}, \mathbf{d})$,*

$$\boldsymbol{\Sigma}_s := \begin{bmatrix} \boldsymbol{\Sigma}_{rr,s} & \boldsymbol{\Sigma}_{rr,s}\boldsymbol{A}^\top \\ \boldsymbol{A}\boldsymbol{\Sigma}_{rr,s} & \boldsymbol{A}\boldsymbol{\Sigma}_{rr,s}\boldsymbol{A}^\top + \boldsymbol{\Sigma}_{\epsilon\epsilon,s} \end{bmatrix}, \tag{14}$$

is positive semidefinite. If $s_k > 0$ for every $k \in \mathcal{K}$, then $\boldsymbol{\Sigma}_{dd,s} \succ 0$.

Proof. Both $\boldsymbol{\Sigma}_{rr,s}$ and $\boldsymbol{\Sigma}_{\epsilon\epsilon,s}$ are positive semidefinite. With zero structured cross-covariance and $\mathbf{d} = \boldsymbol{A}\mathbf{r} + \epsilon$, the joint covariance is $\boldsymbol{\Sigma}_s = \boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top \succeq 0$, where

$$\boldsymbol{C} := \mathrm{blockdiag}(\boldsymbol{\Sigma}_{rr,s}, \boldsymbol{\Sigma}_{\epsilon\epsilon,s}), \qquad \boldsymbol{T} := \begin{bmatrix} \boldsymbol{I}_H & \boldsymbol{0} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{bmatrix}.$$

If all $s_k > 0$, then for any nonzero $\boldsymbol{a} \in \mathbb{R}^m$,

$$\boldsymbol{a}^\top \boldsymbol{\Sigma}_{dd,s} \boldsymbol{a} \geq \boldsymbol{a}^\top \boldsymbol{\Sigma}_{\epsilon\epsilon,s} \boldsymbol{a} \geq \left(\min_{k\in\mathcal{K}} s_k\right) \|\boldsymbol{a}\|_2^2 > 0.$$

Thus $\boldsymbol{\Sigma}_{dd,s} \succ 0$. □

Within-block component. Every row of $\boldsymbol{A}$ is constant within the k_1-blocks, so $\boldsymbol{A}\boldsymbol{P}_{k_1} = \boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{B}_0 = \boldsymbol{0}$. Consequently, $\boldsymbol{\Sigma}_{rd,s}$ and $\boldsymbol{\Sigma}_{dd,s}$ do not depend on τ_0^2. Orthogonality of the bands also gives $\boldsymbol{B}_0 g_s(\boldsymbol{d}) = \boldsymbol{0}$: the structured correction is constant within each k_1-block. The unstructured moments can capture within-block corrections omitted by the structured target.

Sensitivity to s_k. Let $\boldsymbol{G} := \boldsymbol{\Sigma}_{rd,s}\boldsymbol{\Sigma}_{dd,s}^{-1}$, and let $\boldsymbol{G}_k$ and $\boldsymbol{W}_k$ denote its columns and the principal block of $\boldsymbol{\Sigma}_{dd,s}^{-1}$, respectively, associated with resolution k. With other moments fixed and $\boldsymbol{\Sigma}_{dd,s} \succ 0$, $\partial \boldsymbol{G}_k / \partial s_k = -\boldsymbol{G}_k \boldsymbol{W}_k$, giving

$$\frac{\partial}{\partial s_k}\|\boldsymbol{G}_k\|_F^2 = -2\,\mathrm{Tr}(\boldsymbol{G}_k \boldsymbol{W}_k \boldsymbol{G}_k^\top) \le 0.$$

Thus a larger s_k moderates the influence of disagreement at resolution k in this norm; individual entries and the gains for other resolutions need not decrease.

A.4 Derivation and validity of moment mixing

Let $\boldsymbol{\Sigma}_s$ and $\hat{\boldsymbol{\Sigma}}$ denote the structured and unstructured joint covariances of $(\mathbf{r}, \mathbf{d})$, and define $\hat{\boldsymbol{\mu}} := (\hat{\boldsymbol{\mu}}_r^\top, \hat{\boldsymbol{\mu}}_d^\top)^\top$. Here $\hat{\boldsymbol{\Sigma}}$ is the covariance of the empirical distribution assigning mass $1/N$ to each fitting pair $(\boldsymbol{r}_i, \boldsymbol{d}_i)$, so it is positive semidefinite and its rd and dd blocks agree with equation 6.

Since the structured mean is zero, mixing the joint means and uncentered second moments gives

$$\boldsymbol{\mu}_\lambda = \lambda\hat{\boldsymbol{\mu}}, \qquad \mathbf{M}_\lambda = (1-\lambda)\boldsymbol{\Sigma}_s + \lambda\big(\hat{\boldsymbol{\Sigma}} + \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top\big).$$

Subtracting the outer product of the mixed mean yields

$$\boldsymbol{\Sigma}_\lambda = \mathbf{M}_\lambda - \boldsymbol{\mu}_\lambda\boldsymbol{\mu}_\lambda^\top = (1-\lambda)\boldsymbol{\Sigma}_s + \lambda\hat{\boldsymbol{\Sigma}} + \lambda(1-\lambda)\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top. \tag{15}$$

The additional term accounts for the difference between the zero structured mean and the unstructured mean. Taking the rd and dd blocks gives equation 7; substituting these blocks and the mixed means into equation 2 gives equation 8.

Proposition A.3 (Validity of mixed moments)**.** *If $\boldsymbol{\Sigma}_s \succeq 0$ and $\hat{\boldsymbol{\Sigma}} \succeq 0$, then $\boldsymbol{\Sigma}_\lambda \succeq 0$ for every $\lambda \in [0,1]$. If all $s_k > 0$, then $\boldsymbol{\Sigma}_{dd,\lambda} \succ 0$ for every $\lambda \in [0,1)$, even when $\hat{\boldsymbol{\Sigma}}_{dd}$ is singular.*

Proof. Every term on the right-hand side of equation 15 is positive semidefinite for $\lambda \in [0,1]$. When all $s_k > 0$, Proposition A.2 gives $\boldsymbol{\Sigma}_{dd,s} \succ 0$. Thus, for $\lambda < 1$,

$$\boldsymbol{\Sigma}_{dd,\lambda} \succeq (1-\lambda)\boldsymbol{\Sigma}_{dd,s} \succ 0.$$

At $\lambda = 1$, the disagreement covariance is $\hat{\boldsymbol{\Sigma}}_{dd}$, so the inverse-based corrector requires $\hat{\boldsymbol{\Sigma}}_{dd} \succ 0$. □

B Residual–disagreement relation across datasets

We further examine the residual–disagreement relation on additional benchmark datasets with OLinear. In the Figure 6 and 7, each row is a dataset and each column a scale k. r and d_k represent matching scalar coordinates of fine residual and disagreement (same window, block position, and channel), pooled over the fitting split and standardized per channel. Across different datasets, the binned means consistently follow an approximately affine trend. It motivates restricting the affine relation between the fine residual and the cross-resolution disagreement.

C Experimental details

C.1 Experimental settings

For deep forecasters, we use the hyperparameter settings reported in the original papers where a matching model–dataset recipe is available. For model–dataset pairs not covered by the original paper, we follow the corresponding TSLib settings (Wang et al., 2026). We use pretrained checkpoints for foundation models. For task-specific deep forecasters, the fine-resolution model uses a lookback length of 96 time steps, except for ILI, where it uses 36. At each coarser resolution, we search over the lookback length for the corresponding aggregated horizon while requiring the forecast to cover exactly the same future interval as the fine-resolution forecast. The forecast horizons are $H \in \{96, 192, 336, 720\}$ on ETT, Electricity, Weather, Traffic, and Solar-Energy, $H \in \{12, 24, 48, 96\}$ on PEMS, and $H \in \{24, 36, 48, 60\}$ on ILI. All lengths count time steps at the dataset's sampling interval. We evaluate forecasting performance using mean squared error (MSE) and mean absolute error (MAE).

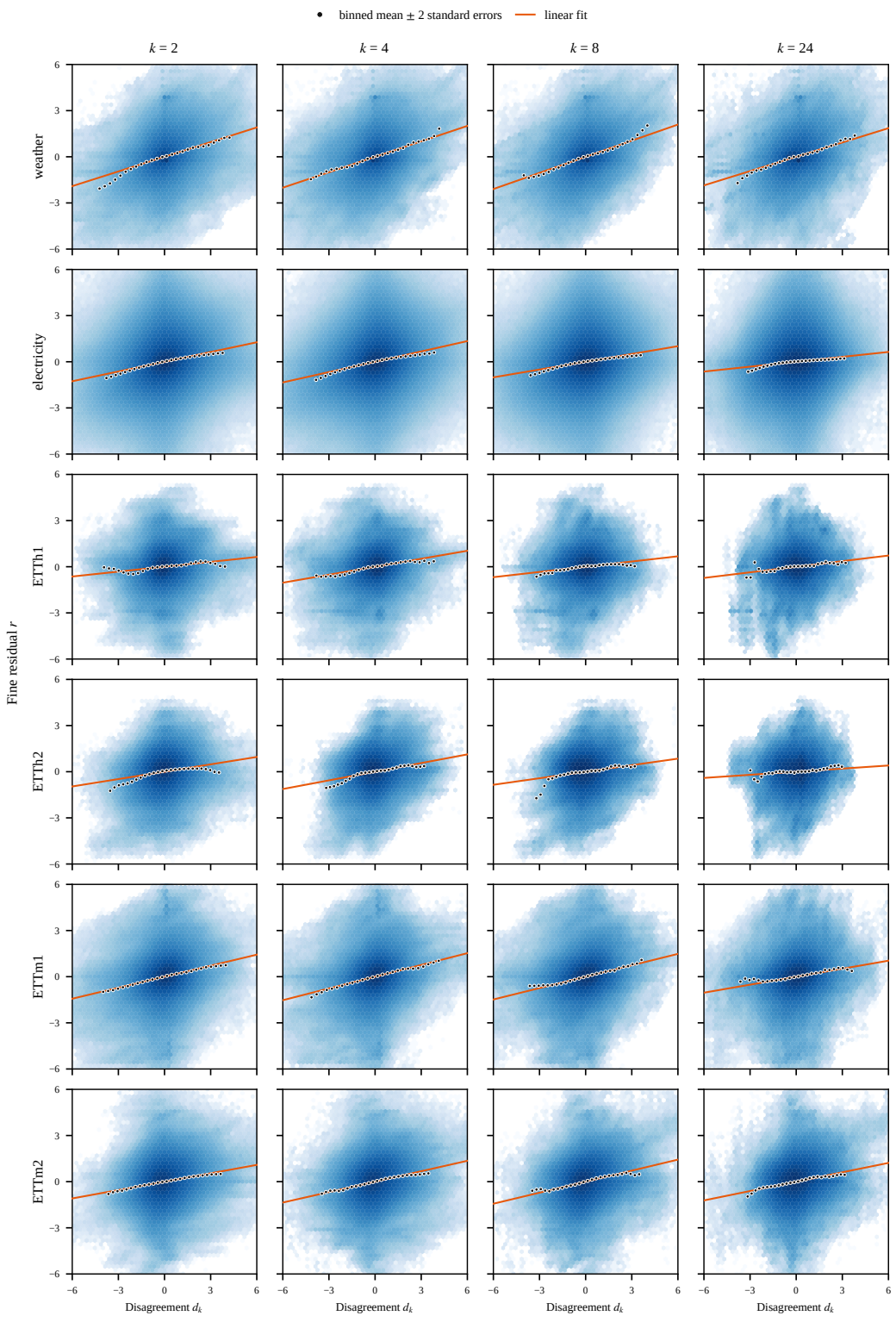


Figure 6: Fine residual r against disagreement d_k on the ETT, Electricity and Weather datasets. The binned conditional means follow an approximately affine trend at each resolution, supporting an affine approximation to the conditional mean.

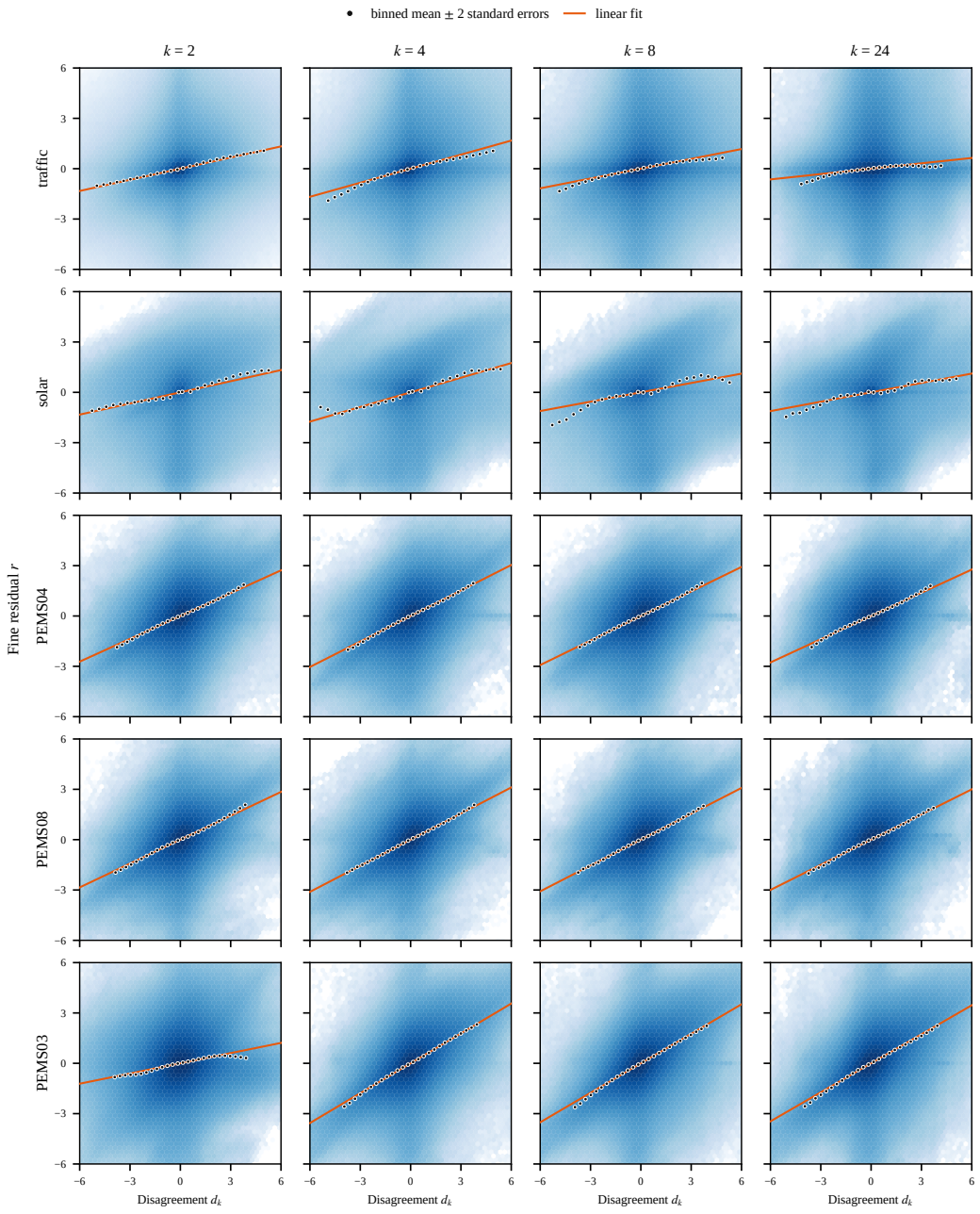


Figure 7: Fine residual r against disagreement d_k on the PEMS, Solar and Traffic datasets. The binned conditional means follow an approximately affine trend at each resolution, supporting an affine approximation to the conditional mean.

C.2 BENCHMARK DATASETS

The benchmark comprises four ETT datasets, Weather, Electricity, Traffic, Solar-Energy, ILI, and four PEMS datasets. Table 3 summarizes their dimensions, sampling intervals, sequence lengths, data splits, and prescribed calendar periods for resolution selection.

Table 3: Benchmark datasets and prescribed calendar periods.

Dataset	Channels	Length	Split	Frequency	Calendar period
ETTh1, ETTh2	7	17,420	6:2:2	1 hour	Day (24)
ETTm1, ETTm2	7	69,680	6:2:2	15 minutes	Day (96)
Weather	21	52,696	7:1:2	10 minutes	Day (144)
Electricity	321	26,304	7:1:2	1 hour	Day (24)
Traffic	862	17,544	7:1:2	1 hour	Day (24)
ILI	7	966	7:1:2	1 week	Year (52)
Solar-Energy	137	52,560	7:1:2	10 minutes	Day (144)
PEMS03	358	26,209	6:2:2	5 minutes	Day (288)
PEMS04	307	16,992	6:2:2	5 minutes	Day (288)
PEMS07	883	28,224	6:2:2	5 minutes	Day (288)
PEMS08	170	17,856	6:2:2	5 minutes	Day (288)

Notes. Length and parenthesized periods count observations; ETT statistics are per dataset. Split is train:validation:test. ILI uses an approximate 52-week year.

- **ETT.** (Zhou et al., 2021) The Electricity Transformer Temperature (ETT) benchmark consists of four datasets collected from transformers in two counties in China. Each dataset records oil temperature and six power-load variables: ETTh1 and ETTh2 are sampled hourly, and ETTm1 and ETTm2 every 15 minutes.
- **Electricity.** (Wu et al., 2021) Each of Electricity's 321 channels follows an individual customer's power consumption. Observations are available at an hourly frequency for 2012–2014.
- **Weather.** (Wu et al., 2021) Weather uses observations from the weather station at the Max Planck Institute for Biogeochemistry. It tracks 21 atmospheric variables, including temperature and humidity, at 10-minute intervals throughout 2020.
- **Traffic.** (Lai et al., 2018) Traffic represents freeway conditions in the San Francisco Bay Area through road-occupancy measurements from 862 sensors. Its hourly series span January 2015 to December 2016.
- **Solar-Energy.** (Lai et al., 2018) Solar-Energy provides 137 simulated photovoltaic power-output series for Alabama, representing 2006 at 10-minute intervals.
- **ILI.** (Wu et al., 2021) ILI tracks influenza-like illness using weekly surveillance data published by the U.S. Centers for Disease Control and Prevention. The seven series cover 2002–2021 and include the share of patient visits attributed to influenza-like illness.
- **PEMS.** (Song et al., 2020) The PEMS benchmarks use traffic-flow data from the Caltrans Performance Measurement System, aggregated at 5-minute intervals. PEMS03, PEMS04, PEMS07, and PEMS08 cover four California districts with 358, 307, 883, and 170 sensors, respectively.

Pretraining overlap. For foundation models, we aim to evaluate forecasting performance on datasets unseen during pretraining. We therefore check each model's paper and model card for pretraining exposure. When these sources indicate that a dataset was used, or may have been used, during pretraining, we exclude that model–dataset pair from evaluation and mark the corresponding cells with a dash (–) in the full foundation-model results tables. Based on the pretraining information publicly available at the time of evaluation, we found no documented overlap for the retained model–dataset pairs. Because these disclosures may be incomplete or may describe broad data sources rather than exact datasets, we cannot rule out undocumented exposure.

C.3 BACKBONE MODELS

We consider the following foundation and task-specific forecasting models.

Foundation models. We use pretrained checkpoints and keep their weights frozen. Parallax fits a separate correction from historical forecast–target pairs and applies it without access to future labels. The backbones are:

- **Toto-2.0-22m** (Khwaja et al., 2026) is the 22-million-parameter Toto 2.0 checkpoint. Its decoder-only Transformer alternates attention along time and across variables, while masked future patches allow it to predict multiple patches in parallel through a quantile output head. Unlike Toto 1.0, Toto 2.0 was pretrained exclusively on internal Datadog observability metrics and synthetic series. Its pretraining mixture contains no public time-series datasets.
- **Toto-2.0-4m** (Khwaja et al., 2026) is the smallest Toto 2.0 checkpoint, with 4.14 million parameters and 16 MB of FP32 weights. It shares the decoder-only architecture, alternating temporal and variate attention, contiguous patch masking, and quantile output head of Toto-2.0-22m, while targeting edge and CPU deployment under tight latency and memory budgets.
- **LightGTS** (Wang et al., 2025) adaptively tokenizes each series according to its cycle length and uses a flexible projection to map variable-length periodic patches into a shared embedding space. Its non-autoregressive decoder repeats the final encoder token, which summarizes the history, and predicts future patches in parallel.
- **Chronos-2** (Ansari et al., 2025) uses an encoder-only Transformer with two types of attention: one links temporal patches within each series, and the other shares information across related series. A quantile head predicts multiple future steps from these representations.
- **Sundial Base 128M** (Liu et al., 2025) encodes continuous-valued patches with a causal decoder-only Transformer. A small flow-matching network conditioned on each Transformer representation samples the next patch, and repeated patch generation produces forecast trajectories.

Task-specific models. These models are trained on the target datasets before applying Parallax.

- **SiGMA** (Lee et al., 2026) constructs multiscale representations through a learnable discrete Gaussian kernel that controls temporal scaling.
- **OLinear** (Yue et al., 2025) forecasts in an orthogonally transformed domain using linear modules for within-series and cross-series dependencies.
- **iTransformer** (Liu et al., 2024) represents each variable's lookback window as a token and applies attention across variables.
- **TimeMixer** (Wang et al., 2024) decomposes multiscale time series into seasonal and trend components and mixes information across scales using MLPs.
- **Scaleformer (Informer backbone)** (Shabani et al., 2023) applies Scaleformer's iterative multiscale refinement procedure with Informer as the underlying forecasting backbone. Throughout this paper, "Scaleformer" refers specifically to this Informer-backed configuration.
- **PatchTST** (Nie et al., 2023) treats temporal patches as Transformer tokens and shares model weights across channels processed independently.
- **DLinear** (Zeng et al., 2023) separates the input into trend and seasonal components, applies a linear projection to each, and sums their forecasts.

C.4 EXPERIMENTAL PROTOCOL

Data splits. Following common benchmarking practice (Liu et al., 2024; Yue et al., 2025), we split each dataset chronologically into training, validation, and test sets. The split ratios are 6:2:2 for ETT and PEMS, and 7:1:2 for the remaining datasets.

Forecasting settings. For task-specific deep forecasters, the fine-resolution model uses a lookback length $L = 96$, except on ILI where $L = 36$. The forecast horizons are $H \in \{96, 192, 336, 720\}$ for ETT, Electricity, Weather, Traffic, and Solar-Energy, $H \in \{12, 24, 48, 96\}$ for PEMS, and $H \in \{24, 36, 48, 60\}$ for ILI. Both L and H count time steps at the dataset's native sampling interval. At a coarser resolution k, every k consecutive observations are averaged and the model predicts H/k blocks. We search the coarse lookback length $L^{(k)}$, while aligning all resolutions at the same forecast origin t and future interval $[t, t + H)$. The coarse input therefore contains $L^{(k)}$ block means from $[t - kL^{(k)}, t)$, and the construction does not require t to be divisible by k.

Fitting Parallax. Parallax is fitted after the forecaster is frozen, using historical forecasts and their observed targets. For foundation models, we generate forecasts over the training period with the frozen pretrained checkpoint and estimate the Parallax moments from the resulting forecast–target pairs. For task-specific models, we first train the forecasters on the training set and freeze their weights. We initially estimate the Parallax moments on the first chronological half of the validation set. For the final evaluation, we re-estimate them using the other half. For both model families, the forecaster and fitted moments remain fixed during test evaluation. Test targets are used only to measure forecasting errors.

Hardware environment. All experiments were conducted with a fixed seed on a workstation running Ubuntu 24.04.3 LTS, equipped with 64 GB of system memory and two NVIDIA RTX A5000 GPUs.

C.5 Hyper-parameter Selection

Calendar pool and chain selection. Parallax first forms a pool of candidate resolutions from the divisors of the forecast horizon H. A scale factor k is eligible when $2 \leq k < H$. When a dataset has a known calendar cycle, we keep only the factors that are compatible with both that cycle and H. The cycles listed in Table 3 represent one day for the hourly and subhourly datasets and approximately one year for weekly ILI, and are fixed before validation. Table 4 lists the candidate factors for every dataset and forecast horizon.

Table 4: Forecast horizons and admissible calendar aggregation factors.

Datasets	**Horizon** H	**Admissible factors** $k \in \mathcal{P}_{\text{cal}}$
ETTh1/2, Electricity, Traffic	96, 192, 336, 720	2, 3, 4, 6, 8, 12, 24
ETTm1/2	96	2, 3, 4, 6, 8, 12, 16, 24, 32, 48
	192	2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 96
	336, 720	2, 3, 4, 6, 8, 12, 16, 24, 48
Weather, Solar-Energy	96, 192, 336	2, 3, 4, 6, 8, 12, 16, 24, 48
	720	2, 3, 4, 6, 8, 9, 12, 16, 18, 24, 36, 48, 72, 144
PEMS03/04/07/08	12	2, 3, 4, 6
	24	2, 3, 4, 6, 8, 12
	48	2, 3, 4, 6, 8, 12, 16, 24
	96	2, 3, 4, 6, 8, 12, 16, 24, 32, 48
ILI	24, 36, 48, 60	2, 4

Notes. Each row shares one candidate pool before chain selection: $k \geq 2$, $k \mid p_{\text{cal}}$, $k \mid H$, and $H/k \geq 2$.

Selecting the mixture weight λ. For each candidate resolution chain, we try $\lambda \in \{0, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 1\}$ and choose the value that gives the lowest MSE on the selection data. Foundation models use the validation set, while task-specific models use the validation halves as described above. One value is selected per forecasting task and shared across channels. We also use MSE to compare resolution chains. These selection errors guide parameter choice, and final performance is measured separately on the test set.

D Full Experimental Results

The tables report MSE and MAE on time-series forecasting benchmarks. To limit potential pretraining exposure, the foundation-model tables report only ETTh1, ETTh2, ETTm1, ETTm2, Weather, Solar-Energy, and Electricity. For each backbone, w/o and +Parallax are shown side by side at four forecast horizons, followed by Avg. and Δ. Avg. is the mean across horizons, and for either averaged error metric E,

$$\Delta := 100\,\frac{E_{\text{Parallax}} - E_{\text{w/o}}}{E_{\text{w/o}}}.$$

Negative values indicate improvement. Dashes mark model–dataset pairs excluded because the dataset appears in the model's pretraining data, and N/A denotes an undefined Δ when the w/o error is zero. Dataset details and settings are given in Appendix C.

D.1 Backbone Forecasting Results

Table 5: Part 1 of the full foundation-model forecasting results without (w/o) and with Parallax. Better MSE and MAE results are shown in bold. Avg. denotes the average across all prediction lengths, while Δ reports the relative change in the averaged error after applying Parallax. Negative Δ values indicate improvement.

Method		**Toto-2.0-22m** (2026)				**Toto-2.0-4m** (2026)			
		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.460	0.412	**0.392**	**0.392**	0.479	0.423	**0.391**	**0.398**
	192	0.512	0.442	**0.427**	**0.416**	0.528	0.452	**0.432**	**0.425**
	336	0.559	0.465	**0.475**	**0.434**	0.572	0.475	**0.458**	**0.440**
	720	0.528	0.471	**0.450**	**0.443**	0.540	0.480	**0.451**	**0.447**
	Avg.	0.515	0.447	**0.436**	**0.421**	0.530	0.458	**0.433**	**0.427**
	Δ			**-15.29%**	**-5.84%**			**-18.31%**	**-6.56%**
ETTh2	96	0.328	0.348	**0.289**	**0.330**	0.342	0.354	**0.294**	**0.337**
	192	0.409	0.397	**0.360**	**0.377**	0.425	0.404	**0.368**	**0.384**
	336	0.451	0.435	**0.389**	**0.404**	0.460	0.438	**0.395**	**0.409**
	720	0.446	0.445	**0.398**	**0.421**	0.451	0.446	**0.410**	**0.430**
	Avg.	0.409	0.406	**0.359**	**0.383**	0.419	0.410	**0.366**	**0.390**
	Δ			**-12.19%**	**-5.61%**			**-12.63%**	**-5.05%**
ETTm1	96	0.819	0.549	**0.296**	**0.335**	0.867	0.561	**0.305**	**0.341**
	192	0.834	0.567	**0.337**	**0.362**	0.859	0.570	**0.345**	**0.369**
	336	0.830	0.574	**0.377**	**0.386**	0.839	0.573	**0.380**	**0.390**
	720	0.854	0.594	**0.424**	**0.417**	0.835	0.583	**0.424**	**0.420**
	Avg.	0.834	0.571	**0.358**	**0.375**	0.850	0.572	**0.364**	**0.380**
	Δ			**-57.04%**	**-34.29%**			**-57.21%**	**-33.54%**
ETTm2	96	0.229	0.294	**0.161**	**0.239**	0.228	0.296	**0.164**	**0.242**
	192	0.293	0.333	**0.219**	**0.282**	0.289	0.333	**0.222**	**0.286**
	336	0.353	0.369	**0.265**	**0.316**	0.346	0.367	**0.284**	**0.328**
	720	0.454	0.424	**0.353**	**0.373**	0.445	0.420	**0.358**	**0.378**
	Avg.	0.332	0.355	**0.249**	**0.303**	0.327	0.354	**0.257**	**0.309**
	Δ			**-24.92%**	**-14.75%**			**-21.33%**	**-12.77%**
Weather	96	0.236	0.244	**0.140**	**0.194**	0.244	0.247	**0.140**	**0.196**
	192	0.275	0.283	**0.180**	**0.236**	0.283	0.286	**0.181**	**0.240**
	336	0.328	0.319	**0.231**	**0.274**	0.341	0.326	**0.232**	**0.276**
	720	0.424	0.373	**0.302**	**0.317**	0.445	0.380	**0.305**	**0.332**
	Avg.	0.316	0.305	**0.213**	**0.255**	0.328	0.310	**0.214**	**0.261**
	Δ			**-32.53%**	**-16.15%**			**-34.65%**	**-15.77%**
Solar-Energy	96	1.156	0.631	**0.185**	**0.228**	0.861	0.526	**0.195**	**0.264**
	192	1.092	0.600	**0.210**	**0.262**	0.899	0.531	**0.212**	**0.282**
	336	1.083	0.605	**0.226**	**0.276**	0.929	0.539	**0.225**	**0.280**
	720	1.032	0.585	**0.234**	**0.292**	0.938	0.536	**0.244**	**0.302**
	Avg.	1.091	0.605	**0.213**	**0.264**	0.907	0.533	**0.219**	**0.282**
	Δ			**-80.43%**	**-56.28%**			**-75.80%**	**-47.09%**

Continued on next page

Table 5 (continued)

Method		**Toto-2.0-22m** (2026)				**Toto-2.0-4m** (2026)			
		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.229	0.298	**0.144**	**0.242**	0.241	0.313	**0.150**	**0.250**
	192	0.224	0.299	**0.156**	**0.252**	0.252	0.328	**0.162**	**0.259**
	336	0.242	0.316	**0.172**	**0.268**	0.278	0.351	**0.175**	**0.273**
	720	0.285	0.350	**0.207**	**0.299**	0.338	0.396	**0.205**	**0.300**
	Avg.	0.245	0.316	**0.170**	**0.265**	0.277	0.347	**0.173**	**0.271**
	Δ			**-30.69%**	**-16.00%**			**-37.66%**	**-22.06%**

Table 6: Part 2 of the full foundation-model forecasting results without (w/o) and with Parallax. Better MSE and MAE results are shown in bold. Avg. denotes the average across all prediction lengths, while Δ reports the relative change in the averaged error after applying Parallax. Negative Δ values indicate improvement. Dashes mark Chronos-2 on Electricity, which is excluded because Electricity appears in its pretraining data.

Method		**LightGTS** (2025)				**Chronos-2** (2025)				**Sundial Base 128M** (2025)			
		w/o		**Parallax**		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.439	0.432	**0.380**	**0.407**	0.424	0.397	**0.376**	**0.383**	0.409	0.396	**0.371**	**0.396**
	192	0.482	0.458	**0.429**	**0.438**	0.484	0.431	**0.422**	**0.411**	0.472	0.430	**0.420**	**0.421**
	336	0.518	0.479	**0.458**	**0.455**	0.536	0.456	**0.457**	**0.430**	0.519	0.451	**0.446**	**0.434**
	720	0.506	0.491	**0.467**	**0.485**	0.526	0.471	**0.449**	**0.445**	0.503	0.461	**0.441**	**0.452**
	Avg.	0.486	0.465	**0.434**	**0.446**	0.492	0.439	**0.426**	**0.417**	0.476	0.435	**0.419**	**0.426**
	Δ			**-10.82%**	**-3.99%**			**-13.48%**	**-4.86%**			**-11.83%**	**-2.08%**
ETTh2	96	0.309	0.348	**0.294**	**0.347**	0.317	0.344	**0.286**	**0.329**	0.320	0.352	**0.304**	**0.350**
	192	0.392	0.399	**0.364**	**0.391**	0.413	0.400	**0.362**	**0.380**	0.408	0.404	**0.364**	**0.392**
	336	0.431	0.433	**0.392**	**0.414**	0.449	0.433	**0.365**	**0.394**	0.440	0.435	**0.384**	**0.412**
	720	0.435	0.445	**0.408**	**0.434**	0.443	0.442	**0.384**	**0.416**	0.438	0.445	**0.404**	**0.432**
	Avg.	0.392	0.406	**0.364**	**0.396**	0.406	0.405	**0.349**	**0.380**	0.402	0.409	**0.364**	**0.396**
	Δ			**-6.94%**	**-2.42%**			**-13.93%**	**-6.14%**			**-9.44%**	**-3.05%**
ETTm1	96	0.814	0.568	**0.293**	**0.350**	0.592	0.452	**0.288**	**0.329**	0.485	0.414	**0.284**	**0.336**
	192	0.787	0.570	**0.336**	**0.378**	0.607	0.467	**0.328**	**0.355**	0.533	0.443	**0.325**	**0.366**
	336	0.762	0.573	**0.369**	**0.395**	0.618	0.482	**0.358**	**0.376**	0.553	0.461	**0.358**	**0.387**
	720	0.754	0.578	**0.414**	**0.423**	0.646	0.508	**0.415**	**0.410**	0.601	0.488	**0.416**	**0.420**
	Avg.	0.779	0.572	**0.353**	**0.387**	0.616	0.477	**0.347**	**0.367**	0.543	0.451	**0.346**	**0.377**
	Δ			**-54.66%**	**-32.43%**			**-43.62%**	**-22.96%**			**-36.33%**	**-16.42%**
ETTm2	96	0.221	0.306	**0.163**	**0.252**	0.210	0.273	**0.161**	**0.238**	0.204	0.275	**0.167**	**0.249**
	192	0.278	0.337	**0.220**	**0.292**	0.274	0.314	**0.222**	**0.284**	0.270	0.318	**0.223**	**0.292**
	336	0.333	0.368	**0.274**	**0.327**	0.340	0.355	**0.271**	**0.321**	0.330	0.355	**0.280**	**0.332**
	720	0.430	0.419	**0.360**	**0.383**	0.448	0.412	**0.354**	**0.373**	0.427	0.409	**0.380**	**0.390**
	Avg.	0.315	0.357	**0.254**	**0.314**	0.318	0.338	**0.252**	**0.304**	0.308	0.339	**0.262**	**0.316**
	Δ			**-19.39%**	**-12.23%**			**-20.74%**	**-10.11%**			**-14.71%**	**-6.92%**
Weather	96	0.220	0.260	**0.148**	**0.214**	0.244	0.230	**0.138**	**0.190**	0.195	0.230	**0.142**	**0.195**
	192	0.264	0.296	**0.187**	**0.250**	0.273	0.270	**0.179**	**0.235**	0.254	0.279	**0.185**	**0.239**
	336	0.312	0.329	**0.237**	**0.292**	0.319	0.305	**0.229**	**0.273**	0.308	0.317	**0.234**	**0.278**
	720	0.386	0.376	**0.305**	**0.337**	0.392	0.357	**0.301**	**0.318**	0.381	0.365	**0.297**	**0.320**
	Avg.	0.295	0.315	**0.219**	**0.273**	0.307	0.291	**0.212**	**0.254**	0.285	0.298	**0.214**	**0.258**
	Δ			**-25.79%**	**-13.31%**			**-30.95%**	**-12.60%**			**-24.67%**	**-13.40%**
Solar-Energy	96	1.249	0.802	**0.198**	**0.274**	0.640	0.343	**0.172**	**0.216**	0.636	0.469	**0.174**	**0.228**
	192	1.079	0.767	**0.208**	**0.280**	0.527	0.324	**0.196**	**0.232**	0.698	0.507	**0.198**	**0.254**
	336	0.988	0.740	**0.220**	**0.288**	0.494	0.323	**0.213**	**0.246**	0.700	0.511	**0.211**	**0.267**
	720	0.923	0.728	**0.228**	**0.294**	0.451	0.319	**0.217**	**0.259**	0.689	0.520	**0.213**	**0.277**
	Avg.	1.060	0.759	**0.214**	**0.284**	0.528	0.327	**0.199**	**0.238**	0.681	0.502	**0.199**	**0.256**
	Δ			**-79.83%**	**-62.59%**			**-62.23%**	**-27.37%**			**-70.78%**	**-48.91%**
Electricity	96	0.306	0.388	**0.153**	**0.257**	–	–	–	–	0.181	0.260	**0.140**	**0.237**
	192	0.319	0.408	**0.163**	**0.264**	–	–	–	–	0.188	0.269	**0.155**	**0.250**
	336	0.348	0.434	**0.181**	**0.283**	–	–	–	–	0.205	0.286	**0.171**	**0.267**
	720	0.422	0.484	**0.207**	**0.305**	–	–	–	–	0.246	0.319	**0.208**	**0.300**

Continued on next page

Table 6 (continued)

Method		**LightGTS** (2025) w/o		**LightGTS** (2025) **Parallax**		**Chronos-2** (2025) w/o		**Chronos-2** (2025) **Parallax**		**Sundial Base 128M** (2025) w/o		**Sundial Base 128M** (2025) **Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
	Avg.	0.349	0.429	**0.176**	**0.277**	–	–	–	–	0.205	0.283	**0.168**	**0.263**
	Δ			**-49.46%**	**-35.32%**			**–**	**–**			**-17.86%**	**-7.00%**

Table 7: Part 1 of the full task-specific forecasting results without (w/o) and with Parallax. The lookback length is 96 (36 for ILI). Better MSE and MAE results are shown in bold. Avg. denotes the average across all prediction lengths, while Δ reports the relative change in the averaged error after applying Parallax. Negative Δ values indicate improvement.

Method		**OLinear** (2025) w/o		**OLinear** (2025) **Parallax**		**DLinear** (2023) w/o		**DLinear** (2023) **Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.362	0.383	**0.356**	**0.382**	0.396	0.411	**0.383**	**0.410**
	192	0.414	0.414	**0.398**	**0.412**	0.447	0.443	**0.422**	**0.434**
	336	0.462	0.440	**0.431**	**0.432**	0.489	0.467	**0.439**	**0.449**
	720	0.470	0.466	**0.437**	**0.456**	0.521	0.517	**0.496**	**0.510**
	Avg.	0.427	0.426	**0.406**	**0.421**	0.463	0.459	**0.435**	**0.451**
	Δ			**-4.93%**	**-1.22%**			**-6.06%**	**-1.85%**
ETTh2	96	0.285	0.329	**0.280**	**0.328**	0.353	0.404	**0.294**	**0.357**
	192	0.361	0.378	**0.347**	**0.378**	0.482	0.479	**0.368**	**0.406**
	336	0.406	0.414	**0.375**	**0.402**	0.594	0.541	**0.437**	**0.460**
	720	0.417	0.432	**0.397**	**0.426**	0.833	0.658	**0.582**	**0.551**
	Avg.	0.367	0.388	**0.350**	**0.383**	0.566	0.521	**0.420**	**0.443**
	Δ			**-4.79%**	**-1.32%**			**-25.69%**	**-14.87%**
ETTm1	96	0.302	0.334	**0.271**	**0.323**	0.345	0.373	**0.298**	**0.349**
	192	0.357	0.363	**0.314**	**0.349**	0.382	0.391	**0.331**	**0.365**
	336	0.388	0.386	**0.347**	**0.372**	0.413	0.413	**0.363**	**0.386**
	720	0.451	0.426	**0.405**	**0.409**	0.472	0.449	**0.421**	**0.424**
	Avg.	0.375	0.377	**0.334**	**0.363**	0.403	0.407	**0.353**	**0.381**
	Δ			**-10.75%**	**-3.75%**			**-12.31%**	**-6.37%**
ETTm2	96	0.169	0.249	**0.158**	**0.244**	0.197	0.297	**0.161**	**0.255**
	192	0.232	0.290	**0.216**	**0.283**	0.279	0.356	**0.216**	**0.297**
	336	0.292	0.329	**0.265**	**0.316**	0.381	0.423	**0.264**	**0.331**
	720	0.390	0.387	**0.351**	**0.374**	0.526	0.509	**0.362**	**0.402**
	Avg.	0.271	0.314	**0.247**	**0.304**	0.346	0.396	**0.251**	**0.321**
	Δ			**-8.63%**	**-3.03%**			**-27.44%**	**-18.91%**
PEMS03	12	0.059	0.158	**0.054**	**0.151**	0.122	0.244	**0.084**	**0.195**
	24	0.077	0.178	**0.066**	**0.166**	0.202	0.320	**0.115**	**0.227**
	48	0.105	0.210	**0.082**	**0.183**	0.335	0.429	**0.187**	**0.284**
	96	0.144	0.250	**0.101**	**0.201**	0.459	0.518	**0.246**	**0.330**
	Avg.	0.096	0.199	**0.076**	**0.175**	0.279	0.378	**0.158**	**0.259**
	Δ			**-20.97%**	**-12.05%**			**-43.46%**	**-31.47%**
PEMS04	12	0.068	0.164	**0.062**	**0.154**	0.147	0.272	**0.106**	**0.223**
	24	0.078	0.176	**0.067**	**0.160**	0.224	0.339	**0.132**	**0.249**
	48	0.095	0.197	**0.074**	**0.168**	0.355	0.437	**0.146**	**0.260**
	96	0.121	0.226	**0.084**	**0.177**	0.452	0.504	**0.157**	**0.271**
	Avg.	0.091	0.191	**0.072**	**0.164**	0.294	0.388	**0.135**	**0.251**
	Δ			**-21.19%**	**-13.70%**			**-54.10%**	**-35.43%**
PEMS07	12	0.053	0.138	**0.046**	**0.130**	0.115	0.241	**0.077**	**0.187**
	24	0.065	0.151	**0.052**	**0.138**	0.210	0.328	**0.107**	**0.222**
	48	0.083	0.171	**0.056**	**0.143**	0.396	0.456	**0.129**	**0.235**
	96	0.109	0.196	**0.063**	**0.149**	0.592	0.552	**0.155**	**0.257**
	Avg.	0.078	0.164	**0.054**	**0.140**	0.328	0.394	**0.117**	**0.225**
	Δ			**-30.01%**	**-14.80%**			**-64.33%**	**-42.89%**
PEMS08	12	0.068	0.159	**0.058**	**0.147**	0.153	0.276	**0.104**	**0.219**
	24	0.088	0.178	**0.069**	**0.159**	0.246	0.351	**0.136**	**0.250**
	48	0.123	0.204	**0.083**	**0.165**	0.437	0.469	**0.167**	**0.276**
	96	0.184	0.239	**0.115**	**0.182**	0.672	0.564	**0.197**	**0.292**

Continued on next page

Table 7 (continued)

Method		**OLinear** (2025)				**DLinear** (2023)			
		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
	Avg.	0.116	0.195	**0.081**	**0.163**	0.377	0.415	**0.151**	**0.259**
	Δ			**-29.81%**	**-16.12%**			**-60.01%**	**-37.51%**
Solar-Energy	96	0.179	0.191	**0.159**	**0.185**	0.290	0.377	**0.196**	**0.282**
	192	0.211	0.212	**0.179**	**0.206**	0.320	0.398	**0.216**	**0.298**
	336	0.231	0.229	**0.183**	**0.224**	0.352	0.415	**0.227**	**0.305**
	720	0.241	0.236	**0.196**	**0.224**	0.357	0.413	**0.236**	**0.307**
	Avg.	0.215	0.217	**0.179**	**0.210**	0.330	0.401	**0.219**	**0.298**
	Δ			**-16.67%**	**-3.35%**			**-33.61%**	**-25.63%**
Electricity	96	0.130	0.220	**0.121**	**0.212**	0.210	0.301	**0.158**	**0.264**
	192	0.151	0.240	**0.141**	**0.231**	0.210	0.305	**0.172**	**0.276**
	336	0.165	0.255	**0.157**	**0.250**	0.223	0.319	**0.187**	**0.291**
	720	0.193	0.281	**0.179**	**0.271**	0.258	0.350	**0.224**	**0.325**
	Avg.	0.160	0.249	**0.149**	**0.241**	0.225	0.319	**0.185**	**0.289**
	Δ			**-6.42%**	**-3.18%**			**-17.87%**	**-9.34%**
Traffic	96	0.407	0.226	**0.345**	**0.223**	0.648	0.396	**0.426**	**0.297**
	192	0.434	0.240	**0.399**	**0.234**	0.598	0.370	**0.433**	**0.300**
	336	0.462	0.250	**0.421**	**0.243**	0.605	0.373	**0.447**	**0.305**
	720	0.502	0.271	**0.464**	**0.263**	0.646	0.395	**0.488**	**0.329**
	Avg.	0.451	0.247	**0.407**	**0.241**	0.624	0.384	**0.449**	**0.308**
	Δ			**-9.78%**	**-2.36%**			**-28.15%**	**-19.84%**
Weather	96	0.152	0.188	**0.135**	**0.186**	0.196	0.257	**0.155**	**0.223**
	192	0.202	0.238	**0.181**	**0.236**	0.238	0.297	**0.194**	**0.262**
	336	0.259	0.280	**0.232**	**0.270**	0.283	0.334	**0.237**	**0.296**
	720	0.341	0.335	**0.309**	**0.319**	0.347	0.384	**0.298**	**0.346**
	Avg.	0.238	0.260	**0.214**	**0.253**	0.266	0.318	**0.221**	**0.282**
	Δ			**-10.14%**	**-3.03%**			**-16.96%**	**-11.42%**
ILI	24	1.843	0.831	**1.648**	**0.789**	2.871	1.183	**2.486**	**1.082**
	36	1.737	0.804	**1.602**	**0.774**	2.768	1.139	**2.398**	**1.080**
	48	1.866	0.815	**1.842**	**0.812**	2.797	1.151	**2.490**	**1.095**
	60	1.819	0.814	**1.670**	**0.791**	2.952	1.184	**2.690**	**1.153**
	Avg.	1.816	0.816	**1.691**	**0.791**	2.847	1.164	**2.516**	**1.103**
	Δ			**-6.92%**	**-3.01%**			**-11.64%**	**-5.31%**

Table 8: Part 2 of the full task-specific forecasting results without (w/o) and with Parallax. The lookback length is 96 (36 for ILI). Better MSE and MAE results are shown in bold. Avg. denotes the average across all prediction lengths, while Δ reports the relative change in the averaged error after applying Parallax. Negative Δ values indicate improvement.

Method		**iTransformer** (2024)				**PatchTST** (2023)			
		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.391	0.411	**0.368**	**0.394**	0.378	0.398	**0.363**	**0.392**
	192	0.446	0.438	**0.414**	**0.423**	0.428	0.432	**0.404**	**0.418**
	336	0.482	0.455	**0.431**	**0.436**	0.473	0.459	**0.431**	**0.438**
	720	0.499	0.489	**0.453**	**0.467**	0.532	0.511	**0.456**	**0.468**
	Avg.	0.454	0.448	**0.416**	**0.430**	0.453	0.450	**0.413**	**0.429**
	Δ			**-8.38%**	**-4.02%**			**-8.71%**	**-4.67%**
ETTh2	96	0.298	0.350	**0.292**	**0.346**	0.300	0.349	**0.283**	**0.342**
	192	0.380	0.399	**0.362**	**0.391**	0.393	0.409	**0.356**	**0.386**
	336	0.425	0.433	**0.392**	**0.418**	0.419	0.436	**0.389**	**0.414**
	720	0.440	0.454	**0.419**	**0.445**	0.451	0.462	**0.432**	**0.455**
	Avg.	0.386	0.409	**0.366**	**0.400**	0.391	0.414	**0.365**	**0.399**
	Δ			**-5.04%**	**-2.20%**			**-6.67%**	**-3.59%**
ETTm1	96	0.340	0.373	**0.285**	**0.341**	0.323	0.363	**0.279**	**0.337**
	192	0.395	0.401	**0.328**	**0.365**	0.373	0.392	**0.320**	**0.365**
	336	0.419	0.421	**0.366**	**0.390**	0.399	0.408	**0.352**	**0.384**
	720	0.492	0.461	**0.429**	**0.429**	0.457	0.445	**0.404**	**0.417**

Continued on next page

Table 8 (continued)

Method		**iTransformer** (2024)				**PatchTST** (2023)			
		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
	Avg.	0.411	0.414	**0.352**	**0.381**	0.388	0.402	**0.339**	**0.376**
	Δ			**-14.40%**	**-7.99%**			**-12.78%**	**-6.52%**
ETTm2	96	0.184	0.266	**0.162**	**0.251**	0.181	0.265	**0.163**	**0.249**
	192	0.250	0.309	**0.225**	**0.294**	0.248	0.310	**0.217**	**0.289**
	336	0.319	0.352	**0.284**	**0.333**	0.315	0.352	**0.267**	**0.324**
	720	0.418	0.410	**0.369**	**0.386**	0.411	0.408	**0.344**	**0.377**
	Avg.	0.293	0.334	**0.260**	**0.316**	0.289	0.334	**0.248**	**0.309**
	Δ			**-11.11%**	**-5.48%**			**-14.25%**	**-7.24%**
PEMS03	12	0.070	0.176	**0.058**	**0.159**	0.112	0.245	**0.069**	**0.176**
	24	0.099	0.210	**0.075**	**0.180**	0.142	0.306	**0.093**	**0.201**
	48	0.125	0.278	**0.095**	**0.199**	0.211	0.380	**0.125**	**0.227**
	96	0.164	0.275	**0.113**	**0.217**	0.269	0.370	**0.151**	**0.250**
	Avg.	0.114	0.234	**0.085**	**0.189**	0.183	0.325	**0.110**	**0.213**
	Δ			**-25.43%**	**-19.57%**			**-40.20%**	**-34.36%**
PEMS04	12	0.089	0.194	**0.071**	**0.172**	0.105	0.224	**0.086**	**0.201**
	24	0.095	0.230	**0.077**	**0.180**	0.153	0.275	**0.096**	**0.211**
	48	0.120	0.233	**0.088**	**0.191**	0.229	0.339	**0.118**	**0.235**
	96	0.150	0.262	**0.099**	**0.203**	0.291	0.389	**0.124**	**0.236**
	Avg.	0.113	0.230	**0.084**	**0.186**	0.195	0.307	**0.106**	**0.221**
	Δ			**-26.16%**	**-18.85%**			**-45.54%**	**-28.07%**
PEMS07	12	0.070	0.171	**0.052**	**0.147**	0.099	0.207	**0.062**	**0.169**
	24	0.099	0.203	**0.059**	**0.156**	0.149	0.270	**0.066**	**0.174**
	48	0.110	0.215	**0.064**	**0.160**	0.253	0.340	**0.080**	**0.196**
	96	0.139	0.245	**0.072**	**0.169**	0.346	0.404	**0.088**	**0.201**
	Avg.	0.104	0.209	**0.062**	**0.158**	0.212	0.305	**0.074**	**0.185**
	Δ			**-41.00%**	**-24.35%**			**-65.03%**	**-39.44%**
PEMS08	12	0.087	0.191	**0.069**	**0.169**	0.168	0.266	**0.077**	**0.185**
	24	0.115	0.246	**0.084**	**0.184**	0.198	0.324	**0.090**	**0.199**
	48	0.186	0.235	**0.092**	**0.180**	0.360	0.354	**0.124**	**0.224**
	96	0.221	0.267	**0.127**	**0.200**	0.408	0.417	**0.136**	**0.227**
	Avg.	0.152	0.235	**0.093**	**0.183**	0.284	0.340	**0.107**	**0.209**
	Δ			**-38.96%**	**-21.87%**			**-62.41%**	**-38.66%**
Solar-Energy	96	0.205	0.249	**0.169**	**0.225**	0.226	0.282	**0.169**	**0.242**
	192	0.238	0.267	**0.191**	**0.241**	0.258	0.302	**0.185**	**0.251**
	336	0.252	0.277	**0.196**	**0.250**	0.273	0.305	**0.194**	**0.258**
	720	0.255	0.280	**0.200**	**0.254**	0.272	0.305	**0.204**	**0.266**
	Avg.	0.238	0.268	**0.189**	**0.243**	0.257	0.298	**0.188**	**0.254**
	Δ			**-20.48%**	**-9.62%**			**-26.82%**	**-14.90%**
Electricity	96	0.148	0.240	**0.129**	**0.224**	0.182	0.275	**0.138**	**0.239**
	192	0.164	0.255	**0.148**	**0.242**	0.189	0.280	**0.157**	**0.256**
	336	0.177	0.270	**0.163**	**0.258**	0.205	0.296	**0.176**	**0.276**
	720	0.213	0.302	**0.197**	**0.289**	0.247	0.329	**0.214**	**0.307**
	Avg.	0.176	0.267	**0.159**	**0.253**	0.206	0.295	**0.171**	**0.270**
	Δ			**-9.54%**	**-4.98%**			**-16.63%**	**-8.62%**
Traffic	96	0.421	0.294	**0.360**	**0.260**	0.459	0.298	**0.385**	**0.273**
	192	0.413	0.278	**0.378**	**0.264**	0.468	0.301	**0.409**	**0.287**
	336	0.424	0.283	**0.389**	**0.268**	0.485	0.308	**0.424**	**0.291**
	720	0.460	0.301	**0.426**	**0.291**	0.516	0.326	**0.467**	**0.316**
	Avg.	0.430	0.289	**0.388**	**0.271**	0.482	0.308	**0.421**	**0.292**
	Δ			**-9.66%**	**-6.23%**			**-12.61%**	**-5.23%**
Weather	96	0.175	0.215	**0.152**	**0.200**	0.177	0.218	**0.148**	**0.201**
	192	0.223	0.257	**0.193**	**0.240**	0.225	0.260	**0.187**	**0.240**
	336	0.281	0.299	**0.241**	**0.282**	0.281	0.299	**0.243**	**0.285**
	720	0.359	0.351	**0.307**	**0.329**	0.356	0.349	**0.306**	**0.334**
	Avg.	0.260	0.280	**0.223**	**0.263**	0.259	0.281	**0.221**	**0.265**
	Δ			**-13.98%**	**-6.31%**			**-14.88%**	**-5.95%**

Continued on next page

Table 8 (continued)

Method		**iTransformer** (2024) w/o		**iTransformer** (2024) **Parallax**		**PatchTST** (2023) w/o		**PatchTST** (2023) **Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ILI	24	2.365	0.970	**1.952**	**0.881**	2.194	0.892	**1.779**	**0.816**
	36	2.264	0.967	**1.979**	**0.911**	2.393	0.924	**1.888**	**0.851**
	48	2.294	0.978	**1.958**	**0.915**	2.117	0.890	**1.833**	**0.869**
	60	2.361	1.004	**2.073**	**0.956**	1.943	**0.889**	**1.861**	0.900
	Avg.	2.321	0.980	**1.990**	**0.916**	2.162	0.899	**1.840**	**0.859**
	Δ			**-14.24%**	**-6.53%**			**-14.87%**	**-4.41%**

Table 9: Part 3 of the full task-specific forecasting results without (w/o) and with Parallax. The lookback length is 96 (36 for ILI). Better MSE and MAE results are shown in bold. Avg. denotes the average across all prediction lengths, while Δ reports the relative change in the averaged error after applying Parallax. Negative Δ values indicate improvement.

Method		**SiGMA** (2026) w/o		**SiGMA** (2026) **Parallax**		**TimeMixer** (2024) w/o		**TimeMixer** (2024) **Parallax**		**Scaleformer** (2023)† w/o		**Scaleformer** (2023)† **Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.376	0.392	**0.370**	**0.390**	0.383	0.396	**0.371**	**0.393**	0.538	0.504	**0.495**	**0.497**
	192	0.428	0.424	**0.404**	**0.415**	0.433	0.431	**0.409**	**0.418**	0.712	0.608	**0.578**	**0.555**
	336	0.470	0.444	**0.432**	**0.433**	0.495	0.451	**0.451**	**0.437**	0.945	0.709	**0.696**	**0.645**
	720	0.488	0.473	**0.458**	**0.467**	0.528	0.498	**0.469**	**0.475**	1.288	0.876	**1.100**	**0.823**
	Avg.	0.441	0.433	**0.416**	**0.426**	0.460	0.444	**0.425**	**0.431**	0.871	0.674	**0.717**	**0.630**
	Δ			**-5.62%**	**-1.61%**			**-7.57%**	**-3.03%**			**-17.60%**	**-6.53%**
ETTh2	96	0.295	0.345	**0.286**	**0.343**	0.290	0.340	**0.279**	**0.338**	0.378	0.427	**0.339**	**0.397**
	192	0.371	0.391	**0.355**	**0.386**	0.365	0.393	**0.357**	**0.391**	4.312	1.608	**0.451**	**0.478**
	336	0.414	0.426	**0.380**	**0.412**	0.423	0.435	**0.382**	**0.414**	4.306	1.668	**0.689**	**0.621**
	720	0.425	0.443	**0.415**	**0.438**	0.453	0.456	**0.410**	**0.439**	4.621	1.719	**0.765**	**0.672**
	Avg.	0.376	0.401	**0.359**	**0.395**	0.383	0.406	**0.357**	**0.396**	3.404	1.356	**0.561**	**0.542**
	Δ			**-4.48%**	**-1.60%**			**-6.81%**	**-2.56%**			**-83.52%**	**-60.03%**
ETTm1	96	0.321	0.358	**0.276**	**0.336**	0.317	0.358	**0.277**	**0.335**	0.435	0.448	**0.333**	**0.386**
	192	0.360	0.381	**0.317**	**0.362**	0.362	0.384	**0.316**	**0.359**	0.514	0.512	**0.393**	**0.419**
	336	0.391	0.403	**0.347**	**0.384**	0.388	0.401	**0.348**	**0.384**	0.570	0.528	**0.475**	**0.486**
	720	0.454	0.442	**0.407**	**0.418**	0.455	0.444	**0.407**	**0.421**	0.916	0.707	**0.562**	**0.538**
	Avg.	0.381	0.396	**0.337**	**0.375**	0.381	0.397	**0.337**	**0.375**	0.609	0.549	**0.441**	**0.457**
	Δ			**-11.73%**	**-5.28%**			**-11.49%**	**-5.61%**			**-27.57%**	**-16.71%**
ETTm2	96	0.175	0.256	**0.159**	**0.247**	0.178	0.260	**0.159**	**0.247**	0.192	0.289	**0.177**	**0.275**
	192	0.240	0.300	**0.220**	**0.287**	0.239	0.299	**0.213**	**0.285**	0.539	0.535	**0.302**	**0.387**
	336	0.295	0.335	**0.272**	**0.325**	0.297	0.337	**0.264**	**0.321**	1.441	0.900	**0.454**	**0.486**
	720	0.397	0.395	**0.354**	**0.381**	0.396	0.398	**0.354**	**0.382**	3.844	1.423	**0.541**	**0.513**
	Avg.	0.277	0.322	**0.251**	**0.310**	0.277	0.324	**0.248**	**0.309**	1.504	0.787	**0.368**	**0.415**
	Δ			**-9.20%**	**-3.62%**			**-10.75%**	**-4.57%**			**-75.50%**	**-47.20%**
PEMS03	12	0.078	0.188	**0.058**	**0.159**	0.076	0.215	**0.076**	**0.184**	0.099	0.213	**0.080**	**0.189**
	24	0.122	0.237	**0.076**	**0.179**	0.113	0.226	**0.098**	**0.212**	0.115	0.227	**0.089**	**0.200**
	48	0.205	0.306	**0.098**	**0.200**	0.191	0.292	**0.150**	**0.251**	0.144	0.252	**0.103**	**0.213**
	96	0.322	0.391	**0.116**	**0.216**	0.288	0.363	**0.171**	**0.269**	0.154	0.269	**0.121**	**0.233**
	Avg.	0.182	0.280	**0.087**	**0.189**	0.167	0.274	**0.124**	**0.229**	0.128	0.240	**0.098**	**0.209**
	Δ			**-52.16%**	**-32.77%**			**-25.95%**	**-16.41%**			**-23.38%**	**-13.16%**
PEMS04	12	0.095	0.205	**0.071**	**0.172**	0.092	0.231	**0.091**	**0.203**	0.091	0.203	**0.085**	**0.193**
	24	0.139	0.257	**0.079**	**0.183**	0.128	0.243	**0.114**	**0.229**	0.099	0.214	**0.090**	**0.198**
	48	0.231	0.329	**0.088**	**0.189**	0.213	0.315	**0.125**	**0.242**	0.114	0.234	**0.096**	**0.206**
	96	0.370	0.434	**0.100**	**0.200**	0.307	0.384	**0.143**	**0.260**	0.120	0.238	**0.101**	**0.212**
	Avg.	0.209	0.306	**0.085**	**0.186**	0.185	0.293	**0.118**	**0.234**	0.106	0.222	**0.093**	**0.202**
	Δ			**-59.51%**	**-39.23%**			**-36.06%**	**-20.33%**			**-12.17%**	**-9.13%**
PEMS07	12	0.076	0.186	**0.051**	**0.145**	0.073	0.207	**0.068**	**0.175**	0.096	0.201	**0.086**	**0.189**
	24	0.114	0.225	**0.059**	**0.154**	0.111	0.219	**0.089**	**0.203**	0.105	0.210	**0.090**	**0.192**
	48	0.199	0.299	**0.066**	**0.164**	0.237	0.328	**0.091**	**0.201**	0.125	0.230	**0.096**	**0.197**
	96	0.341	0.388	**0.076**	**0.175**	0.303	0.354	**0.112**	**0.222**	0.139	0.241	**0.103**	**0.204**
	Avg.	0.183	0.275	**0.063**	**0.159**	0.181	0.277	**0.090**	**0.200**	0.116	0.220	**0.094**	**0.196**
	Δ			**-65.53%**	**-41.92%**			**-50.37%**	**-27.64%**			**-19.14%**	**-11.24%**

Continued on next page

Table 9 (continued)

Method		**SiGMA** (2026)				**TimeMixer** (2024)				**Scaleformer** (2023)†			
		w/o		**Parallax**		w/o		**Parallax**		w/o		**Parallax**	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS08	12	0.096	0.204	**0.066**	**0.166**	0.091	0.223	**0.088**	**0.199**	0.147	0.238	**0.108**	**0.212**
	24	0.142	0.249	**0.074**	**0.174**	0.137	0.246	**0.119**	**0.229**	0.168	0.250	**0.124**	**0.225**
	48	0.247	0.331	**0.092**	**0.185**	0.265	0.343	**0.137**	**0.241**	0.207	0.281	**0.148**	**0.240**
	96	0.438	0.434	**0.130**	**0.203**	0.410	0.407	**0.187**	**0.265**	0.253	0.295	**0.174**	**0.247**
	Avg.	0.231	0.304	**0.091**	**0.182**	0.226	0.305	**0.133**	**0.233**	0.194	0.266	**0.139**	**0.231**
	Δ			**-60.67%**	**-40.23%**			**-41.23%**	**-23.41%**			**-28.49%**	**-13.11%**
Solar-Energy	96	0.228	0.303	**0.167**	**0.231**	0.189	0.269	**0.157**	**0.227**	0.205	0.267	**0.161**	**0.232**
	192	0.239	0.296	**0.186**	**0.248**	0.209	0.274	**0.184**	**0.245**	0.240	0.297	**0.178**	**0.236**
	336	0.278	0.313	**0.196**	**0.258**	0.220	0.298	**0.196**	**0.261**	0.257	0.311	**0.188**	**0.246**
	720	0.287	0.327	**0.202**	**0.265**	0.245	0.336	**0.215**	**0.292**	0.248	0.297	**0.207**	**0.262**
	Avg.	0.258	0.310	**0.188**	**0.250**	0.216	0.294	**0.188**	**0.256**	0.237	0.293	**0.183**	**0.244**
	Δ			**-27.17%**	**-19.13%**			**-12.94%**	**-12.92%**			**-22.74%**	**-16.78%**
Electricity	96	0.145	0.240	**0.131**	**0.227**	0.157	0.248	**0.133**	**0.229**	0.199	0.312	**0.183**	**0.295**
	192	0.163	0.256	**0.149**	**0.244**	0.169	0.259	**0.150**	**0.244**	0.216	0.329	**0.193**	**0.305**
	336	0.178	0.272	**0.166**	**0.262**	0.187	0.277	**0.166**	**0.262**	0.240	0.349	**0.207**	**0.318**
	720	0.211	0.304	**0.195**	**0.289**	0.227	0.312	**0.205**	**0.297**	0.289	0.388	**0.247**	**0.351**
	Avg.	0.174	0.268	**0.160**	**0.255**	0.185	0.274	**0.164**	**0.258**	0.236	0.344	**0.207**	**0.317**
	Δ			**-7.82%**	**-4.72%**			**-11.55%**	**-5.80%**			**-12.02%**	**-7.83%**
Traffic	96	0.433	0.291	**0.377**	**0.265**	0.482	0.295	**0.383**	**0.267**	0.598	0.361	**0.553**	**0.332**
	192	0.451	0.297	**0.396**	**0.275**	0.490	0.298	**0.405**	**0.274**	0.641	0.379	**0.578**	**0.346**
	336	0.465	0.302	**0.415**	**0.285**	0.509	0.302	**0.427**	**0.285**	0.751	0.427	**0.610**	**0.359**
	720	0.496	0.319	**0.466**	**0.313**	0.560	0.322	**0.490**	**0.315**	0.875	0.481	**0.696**	**0.412**
	Avg.	0.461	0.302	**0.414**	**0.285**	0.510	0.304	**0.426**	**0.285**	0.716	0.412	**0.610**	**0.362**
	Δ			**-10.30%**	**-5.81%**			**-16.44%**	**-6.25%**			**-14.88%**	**-12.12%**
Weather	96	0.160	0.203	**0.143**	**0.199**	0.163	0.209	**0.143**	**0.200**	0.184	0.254	**0.146**	**0.215**
	192	0.218	0.257	**0.188**	**0.241**	0.209	0.252	**0.184**	**0.239**	0.276	0.342	**0.190**	**0.260**
	336	0.278	0.300	**0.235**	**0.278**	0.266	0.292	**0.236**	**0.278**	0.403	0.423	**0.245**	**0.302**
	720	0.368	0.362	**0.303**	**0.329**	0.344	0.343	**0.292**	**0.321**	0.574	0.517	**0.324**	**0.357**
	Avg.	0.256	0.281	**0.217**	**0.262**	0.245	0.274	**0.214**	**0.260**	0.359	0.384	**0.226**	**0.284**
	Δ			**-15.00%**	**-6.66%**			**-12.88%**	**-5.16%**			**-37.00%**	**-26.14%**
ILI	24	2.334	0.919	**2.017**	**0.861**	2.852	0.996	**2.220**	**0.912**	3.427	1.192	**2.896**	**1.123**
	36	2.202	0.921	**1.906**	**0.857**	2.299	0.937	**1.936**	**0.873**	3.599	1.238	**3.362**	**1.235**
	48	2.255	0.933	**1.870**	**0.847**	2.281	0.946	**1.942**	**0.873**	3.591	1.241	**3.193**	**1.145**
	60	2.223	0.927	**1.948**	**0.887**	2.257	0.954	**1.978**	**0.887**	3.581	1.224	**3.540**	**1.221**
	Avg.	2.254	0.925	**1.935**	**0.863**	2.422	0.958	**2.019**	**0.886**	3.550	1.224	**3.248**	**1.181**
	Δ			**-14.12%**	**-6.71%**			**-16.67%**	**-7.54%**			**-8.50%**	**-3.49%**

† Scaleformer uses Informer (Zhou et al., 2021) as its forecasting backbone.

D.2 Comparison with Existing Post-hoc Methods

Table 10 compares Parallax with the Ada-Y variant of δ-Adapter and with RGMR using the same frozen LightGTS forecasts and evaluation tasks.

Table 10: Comparison with existing post-hoc correction methods using LightGTS on seven datasets. Avg. is the mean over the four forecast horizons. Δ is the relative change of Avg. with respect to LightGTS without correction (Table 6); negative values indicate improvement. Best results are in bold and second-best results are underlined.

		Existing post-hoc methods				*Proposed framework*	
		δ-Adapter (Ada-Y) (2026)		**RGMR** (2026)		**Parallax**	
Data	H	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**
ETTh1	96	0.438	0.431	0.433	0.427	**0.380**	**0.407**
	192	0.480	0.457	0.481	0.456	**0.429**	**0.438**
	336	0.517	0.478	0.525	0.480	**0.458**	**0.455**
	720	0.505	0.491	0.513	0.494	**0.467**	**0.485**
	Avg.	0.485	0.464	0.488	0.464	**0.434**	**0.446**
	Δ	-0.21%	-0.13%	+0.37%	-0.16%	-10.82%	-3.99%
ETTh2	96	0.309	0.348	0.304	0.351	**0.294**	**0.347**
	192	0.392	0.399	0.393	0.409	**0.364**	**0.391**
	336	0.431	0.433	0.447	0.449	**0.392**	**0.414**
	720	0.435	0.446	0.524	0.504	**0.408**	**0.434**
	Avg.	0.392	0.407	0.417	0.428	**0.364**	**0.396**
	Δ	+0.07%	+0.08%	+6.48%	+5.38%	-6.94%	-2.42%
ETTm1	96	0.807	0.565	0.779	0.560	**0.293**	**0.350**
	192	0.781	0.567	0.742	0.556	**0.336**	**0.378**
	336	0.756	0.570	0.761	0.582	**0.369**	**0.395**
	720	0.749	0.575	0.731	0.573	**0.414**	**0.423**
	Avg.	0.773	0.569	0.753	0.568	**0.353**	**0.387**
	Δ	-0.73%	-0.47%	-3.31%	-0.80%	-54.66%	-32.43%
ETTm2	96	0.220	0.306	0.221	0.314	**0.163**	**0.252**
	192	0.277	0.336	0.278	0.342	**0.220**	**0.292**
	336	0.332	0.368	0.323	0.372	**0.274**	**0.327**
	720	0.430	0.419	0.408	0.416	**0.360**	**0.383**
	Avg.	0.315	0.357	0.308	0.361	**0.254**	**0.314**
	Δ	-0.25%	+0.04%	-2.50%	+1.04%	-19.39%	-12.23%
Weather	96	0.201	0.252	0.207	0.258	**0.148**	**0.214**
	192	0.248	0.295	0.255	0.298	**0.187**	**0.250**
	336	0.296	0.328	0.302	0.328	**0.237**	**0.292**
	720	0.380	0.381	0.375	0.382	**0.305**	**0.337**
	Avg.	0.281	0.314	0.285	0.317	**0.219**	**0.273**
	Δ	-4.82%	-0.49%	-3.54%	+0.42%	-25.79%	-13.31%
Electricity	96	0.265	0.349	0.278	0.361	**0.153**	**0.257**
	192	0.274	0.365	0.289	0.377	**0.163**	**0.264**
	336	0.298	0.389	0.316	0.402	**0.181**	**0.283**
	720	0.364	0.437	0.379	0.451	**0.207**	**0.305**
	Avg.	0.300	0.385	0.315	0.398	**0.176**	**0.277**
	Δ	-13.95%	-10.15%	-9.57%	-7.17%	-49.46%	-35.32%
Solar-Energy	96	1.105	0.733	1.023	0.763	**0.198**	**0.274**
	192	0.944	0.696	0.914	0.706	**0.208**	**0.280**
	336	0.860	0.669	0.862	0.698	**0.220**	**0.288**
	720	0.804	0.665	0.826	0.699	**0.228**	**0.294**
	Avg.	0.928	0.691	0.906	0.717	**0.214**	**0.284**
	Δ	-12.43%	-8.93%	-14.51%	-5.55%	-79.83%	-62.59%

D.3 Ablation Study of Forecast Combination Variants

Table 11 tests whether other combiners of the same aligned forecasts account for the gains obtained by Parallax. Equal average is a fixed, parameter-free combination, whereas Linear fusion and MLP fusion are two learned function classes, an affine map and a multilayer perceptron, respectively. All three variants use the resolution chain $\mathcal{K}^*$ selected by Parallax for the same task, so they combine exactly the forecasts that Parallax corrects: the fine forecast $\hat{\boldsymbol{y}}^{(1)} \in \mathbb{R}^H$ and the coarse forecasts $\hat{\boldsymbol{y}}^{(k)} \in \mathbb{R}^{H/k}$, $k \in \mathcal{K}^*$. Let $K := |\mathcal{K}^*|$ and $m := \sum_{k\in\mathcal{K}^*} H/k$.

Equal average. The fine forecast and the upsampled coarse forecasts are averaged with equal weights,

$$\hat{\boldsymbol{y}} := \frac{1}{1+K}\Big(\hat{\boldsymbol{y}}^{(1)} + \sum_{k\in\mathcal{K}^*} k\boldsymbol{A}_k^\top \hat{\boldsymbol{y}}^{(k)}\Big),$$

where $k\boldsymbol{A}_k^\top$ repeats each block value k times; for example, $k\boldsymbol{A}_k^\top[10,20]^\top = [10,10,20,20]^\top$ for $k = 2$. This variant has no parameters or hyperparameters and uses neither the training nor the validation data.

Linear fusion. With the stacked input $\boldsymbol{u} := [\hat{\boldsymbol{y}}^{(1)}; \hat{\boldsymbol{y}}^{(k_1)}; \ldots; \hat{\boldsymbol{y}}^{(k_K)}] \in \mathbb{R}^{H+m}$, where $\mathcal{K}^* = \{k_1, \ldots, k_K\}$, the corrected forecast is $\hat{\boldsymbol{y}} := \hat{\boldsymbol{y}}^{(1)} + \boldsymbol{W}\boldsymbol{u} + \boldsymbol{b}$ with $\boldsymbol{W} \in \mathbb{R}^{H\times(H+m)}$ and $\boldsymbol{b} \in \mathbb{R}^H$, i.e., $(H + m + 1)H$ parameters. The layer is initialized at zero, so training starts from the uncorrected forecast $\hat{\boldsymbol{y}}^{(1)}$. The model is channel-independent: a single head is shared across channels and forecast origins, and each (origin, channel) pair is one training sample.

MLP fusion. The corrected forecast is $\hat{\boldsymbol{y}} := \hat{\boldsymbol{y}}^{(1)} + \mathrm{MLP}(\boldsymbol{u})$ with the same input $\boldsymbol{u}$. The MLP has two hidden layers of width $w = 256$,

$$\mathrm{Linear}(H+m, w) \to \mathrm{GELU} \to \mathrm{Dropout} \to \mathrm{Linear}(w, w) \to \mathrm{GELU} \to \mathrm{Dropout} \to \mathrm{Linear}(w, H),$$

with dropout rate 0.1, i.e., $(H + m + 1)w + (w + 1)w + (w + 1)H$ parameters. Its last layer is initialized at zero, and it is channel-independent with one shared head, as for Linear fusion. The width, depth, and dropout rate are fixed and not tuned.

Training of the learned fusions. Linear fusion and MLP fusion share one training protocol. Each is fitted on the full training split with early stopping on the full validation split and is not refitted afterwards; test labels are never used. We use Adam with learning rate 10^{-3}, no weight decay, and the MSE loss, without a learning-rate scheduler or gradient clipping. Training runs for at most 100 epochs.

Table 11: Ablation study of three forecast-combination variants (Equal average, Linear fusion, and MLP fusion) using LightGTS on seven datasets. Parallax is the full method. Avg. is the mean over the four forecast horizons. Δ is the relative change of Avg. with respect to LightGTS without correction (Table 6); negative values indicate improvement. Best results are in bold, and second-best results are underlined.

		Forecast-combination variants							
		Equal average		**Linear fusion**		**MLP fusion**		**Parallax**	
Data	H	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**
ETTh1	96	0.517	0.480	<u>0.383</u>	<u>0.408</u>	0.420	0.425	**0.380**	**0.407**
	192	0.546	0.499	<u>0.437</u>	<u>0.446</u>	0.445	0.447	**0.429**	**0.438**
	336	0.562	0.513	<u>0.498</u>	<u>0.484</u>	0.512	0.494	**0.458**	**0.455**
	720	<u>0.560</u>	<u>0.534</u>	0.612	0.583	0.619	0.577	**0.467**	**0.485**
	Avg.	0.546	0.507	<u>0.482</u>	<u>0.480</u>	0.499	0.486	**0.434**	**0.446**
	Δ	+12.32%	+8.99%	-0.78%	+3.28%	+2.60%	+4.54%	-10.82%	-3.99%
ETTh2	96	0.300	0.358	0.312	0.371	<u>0.299</u>	<u>0.349</u>	**0.294**	**0.347**
	192	**0.360**	<u>0.395</u>	0.406	0.433	0.382	0.414	<u>0.364</u>	**0.391**
	336	**0.385**	<u>0.417</u>	0.513	0.493	0.450	0.466	<u>0.392</u>	**0.414**
	720	**0.405**	<u>0.439</u>	0.818	0.643	0.662	0.575	<u>0.408</u>	**0.434**
	Avg.	**0.362**	<u>0.402</u>	0.512	0.485	0.449	0.451	<u>0.364</u>	**0.396**
	Δ	-7.45%	-1.05%	+30.85%	+19.40%	+14.55%	+11.00%	-6.94%	-2.42%

Continued on next page

Table 11 continued

		Forecast-combination variants							
		Equal average		**Linear fusion**		**MLP fusion**		**Parallax**	
Data	*H*	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**	**MSE↓**	**MAE↓**
ETTm1	96	0.582	0.491	0.329	0.379	0.328	0.375	**0.293**	**0.350**
	192	0.601	0.507	0.342	0.381	0.353	0.391	**0.336**	**0.378**
	336	0.602	0.510	0.371	0.402	0.377	0.404	**0.369**	**0.395**
	720	0.618	0.525	0.440	0.454	0.436	0.440	**0.414**	**0.423**
	Avg.	0.601	0.508	0.371	0.404	0.373	0.402	**0.353**	**0.387**
	Δ	-22.89%	-11.16%	-52.41%	-29.41%	-52.07%	-29.67%	-54.66%	-32.43%
ETTm2	96	0.195	0.290	0.171	0.263	0.181	0.268	**0.163**	**0.252**
	192	0.247	0.325	0.231	0.306	0.247	0.319	**0.220**	**0.292**
	336	0.294	0.350	0.290	0.351	0.296	0.346	**0.274**	**0.327**
	720	0.373	0.395	0.377	0.413	0.439	0.459	**0.360**	**0.383**
	Avg.	0.277	0.340	0.267	0.333	0.291	0.348	**0.254**	**0.314**
	Δ	-12.11%	-4.87%	-15.27%	-6.68%	-7.84%	-2.57%	-19.39%	-12.23%
Weather	96	0.214	0.267	0.188	0.255	0.164	0.228	**0.148**	**0.214**
	192	0.259	0.301	0.226	0.291	0.201	0.260	**0.187**	**0.250**
	336	0.301	0.329	0.269	0.325	0.255	0.306	**0.237**	**0.292**
	720	0.315	0.350	0.320	0.373	0.329	0.377	**0.305**	**0.337**
	Avg.	0.272	0.312	0.251	0.311	0.237	0.293	**0.219**	**0.273**
	Δ	-7.80%	-1.10%	-15.12%	-1.45%	-19.65%	-7.18%	-25.79%	-13.31%
Electricity	96	0.509	0.573	0.160	0.262	**0.140**	**0.239**	0.153	0.257
	192	0.528	0.585	0.161	0.262	**0.156**	**0.255**	0.163	0.264
	336	0.555	0.600	**0.172**	0.274	0.173	**0.274**	0.181	0.283
	720	0.601	0.623	0.207	0.307	**0.204**	**0.303**	0.207	0.305
	Avg.	0.548	0.595	0.175	0.276	**0.168**	**0.268**	0.176	0.277
	Δ	+57.30%	+38.82%	-49.79%	-35.55%	-51.78%	-37.47%	-49.46%	-35.32%
Solar-Energy	96	1.260	0.857	0.233	0.328	0.209	**0.269**	**0.198**	0.274
	192	1.080	0.797	0.245	0.323	0.232	**0.280**	**0.208**	0.280
	336	0.983	0.763	0.255	0.325	0.234	**0.279**	**0.220**	0.288
	720	0.591	0.603	0.240	0.305	0.245	0.297	**0.228**	**0.294**
	Avg.	0.978	0.755	0.243	0.321	0.230	**0.281**	**0.214**	0.284
	Δ	-7.69%	-0.51%	-77.05%	-57.75%	-78.30%	-62.94%	-79.83%	-62.59%

D.4 Full Ablation Results

Table 12 reports the full component ablation of Section 4.3 with OLinear as the backbone for each forecast horizon, on the same datasets as Table 2.

Table 12: Full ablation results (MSE) for Parallax with OLinear as backbone. Horizons are 96/192/336/720 for ETTh1, ETTm2, and Weather, and 12/24/48/96 for PEMS04. The best result among the Parallax variants is in bold.

Design	Horizon	ETTh1	ETTm2	Weather	PEMS04
Parallax	96 / 12	**0.3564**	0.1576	**0.1353**	**0.0616**
	192 / 24	**0.3989**	**0.2162**	**0.1817**	**0.0668**
	336 / 48	**0.4318**	**0.2649**	0.2322	**0.0743**
	720 / 96	0.4375	0.3511	**0.3090**	**0.0837**
	Avg.	**0.4061**	**0.2474**	**0.2145**	**0.0716**
w/o temporal bands	96 / 12	0.3620	**0.1575**	0.1358	0.0625
	192 / 24	0.4137	0.2209	0.1822	0.0686
	336 / 48	0.4410	0.2711	**0.2308**	0.0758
	720 / 96	0.4696	0.3601	0.3415	0.0873
	Avg.	0.4216	0.2524	0.2226	0.0735
w/o resolution-specific s_k	96 / 12	0.3650	0.1606	0.1531	0.0622
	192 / 24	0.4071	0.2314	0.2585	0.0680
	336 / 48	0.8602	0.3340	0.4273	0.0766
	720 / 96	**0.4356**	0.5063	0.3098	0.0918
	Avg.	0.5170	0.3081	0.2872	0.0746
w/o empirical moments ($\lambda = 0$)	96 / 12	**0.3564**	0.1606	0.1396	0.0617
	192 / 24	**0.3989**	0.2166	0.1851	0.0669
	336 / 48	0.4355	0.2670	0.2334	0.0746
	720 / 96	0.4375	**0.3492**	**0.3090**	0.0839
	Avg.	0.4071	0.2483	0.2168	0.0718
OLinear (w/o Parallax)	96 / 12	0.3620	0.1690	0.1520	0.0684
	192 / 24	0.4137	0.2320	0.2020	0.0781
	336 / 48	0.4622	0.2917	0.2586	0.0955
	720 / 96	0.4696	0.3900	0.3415	0.1212
	Avg.	0.4268	0.2707	0.2385	0.0908

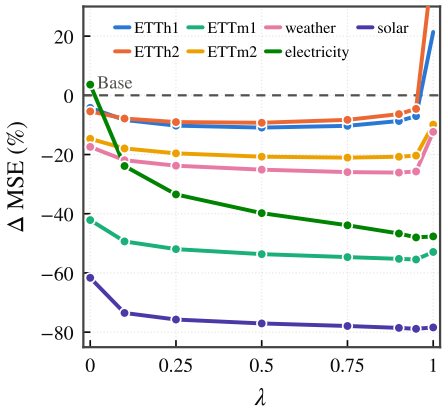
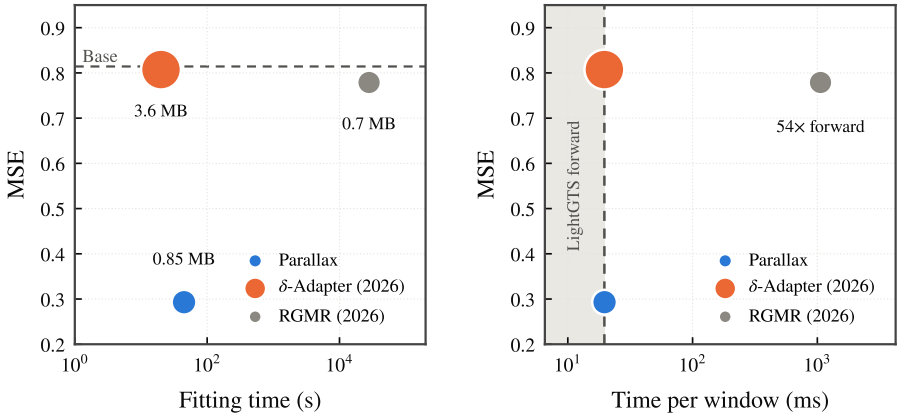


Figure 8: Test MSE change from Base across λ, averaged over horizons with a fixed chain using LightGTS.



(a) Fitting cost (ETTm1, H=96) (b) Test-time cost (ETTm1, H=96)

Figure 9: Efficiency comparison on ETTm1 ($H = 96$) with LightGTS. (a) Test MSE versus fitting time (b) Test MSE versus time per window. Labels indicate test-time memory size. Dashed lines mark Base MSE in (a) and one LightGTS forward pass in (b).

D.5 Hyperparameter sensitivity

The mixture weight λ balances the structured and empirical moments and is selected on validation data (Appendix C.5). Figure 8 shows that Parallax is robust to this choice: Every $\lambda \in [0.1, 0.95]$ improves over the backbone on all seven datasets. Both kinds of moments are needed. Using only the structured moments ($\lambda = 0$) gives up part of the gain on every dataset and all of it on Electricity. Using only the empirical moments ($\lambda = 1$) makes the correction harmful on ETTh1 and ETTh2, increasing their MSE by 21% and 45%, while keeping just 5% of the weight on the structured moments ($\lambda = 0.95$) already prevents this.

D.6 Efficiency analysis

Parallax needs no gradient-based training: it estimates moments from forecasts computed once and corrects each window with a single linear solve. Figure 9 compares its fitting and test-time cost with the existing post-hoc correctors on ETTm1 ($H = 96$) with LightGTS. Parallax fits and selects its hyperparameters in 44.7 s on a CPU, comparable to the 26.4 s of δ-Adapter on a GPU, while RGMR takes more than 100 times as long. At test time, Parallax and δ-Adapter add little to the LightGTS forward pass, whereas RGMR, which queries the backbone at every resolution for every window,

takes 54 times as long as a forward pass. Parallax keeps 0.85 MB of test-time state, compared with 3.6 MB for δ-Adapter and 0.7 MB for RGMR, and reaches a much lower MSE.

E LIMITATIONS

Parallax learns how cross-resolution disagreement relates to forecast errors from historical forecast-target pairs. If this relationship changes at test time, the fitted correction may become less reliable. Adapting Parallax to such shifts is left for future work.