TINYCYCLE: TIME-SERIES FORECASTING IN 30 PARAMETERS WITH CLOSED-FORM FITTING

Anonymous authors
Paper under double-blind review

ABSTRACT

In long-term time-series forecasting, lightweight models can compare to those of complex models, but aggregate metrics reveal little about the information these models recover or the sources of their performance differences. Based on phase recurrence across cycles, we decompose both the true future and model predictions into Structure and Dynamics. We then use orthogonality to decompose MSE. Diagnostics across multiple models show that models recover Structure substantially better than Dynamics, and that modeling Structure can account for most of their predictions. Motivated by this observation, we propose TinyCycle, which directly models the Structure of the data without modeling complex temporal mappings. Across experiments, TinyCycle recovers most of the forecasting performance of mainstream time-series forecasting models with no more than 31 parameters using closed-form fitting.

Code is available at https://anonymous.4open.science/r/TinyCycle

1 INTRODUCTION

Long-term time series forecasting is used for energy scheduling, traffic management, weather analysis, and financial analysis (Zhu et al., 2025; Cao et al., 2025; Bodnar et al., 2025; Park et al., 2025). To extract patterns from historical observations that can inform the future, deep models employ mechanisms such as attention, sequence patching, and multiscale representations (Wu et al., 2021; Zhou et al., 2021; Nie et al., 2022; Wu et al., 2022), and time series foundation models (TSFMs) further support zero-shot forecasting through large-scale pretraining (Ansari et al., 2025; Liu et al., 2025a; Khwaja et al., 2026). Meanwhile, linear models such as DLinear (Zeng et al., 2023) and RLinear (Li et al., 2023) remain competitive, and SparseTSF (Lin et al., 2024b), TimeBase (Huang et al., 2025), and MixLinear (Ma et al., 2026) further compress models by exploiting periodicity or low-dimensional Structure. This relationship between model size and forecasting error raises a question: in long-term forecasting, what information underlies the competitiveness of lightweight models, and what information accounts for performance differences between models?

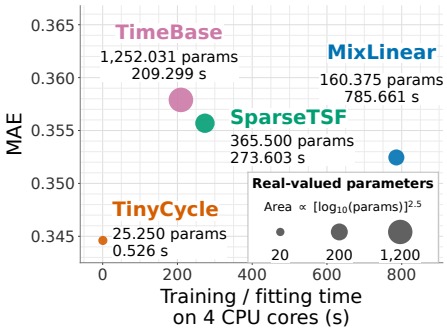


Figure 1: Comparison of TinyCycle and lightweight models.

Answering this question requires distinguishing the Structure that models recover, the variations they fail to recover, and the sources of error gaps between models. Overall MSE or MAE compresses these factors into a single value, making it difficult to distinguish improvements in the mean level and repeated profile from improvements in local variations. Previous studies have examined smoothing in forecasting outputs (Liu et al., 2025b; Ao et al.), but reduced overall variation does not establish that all types of Structure are attenuated equally. Our preliminary experiments find that forecasting outputs preserve periodic frequencies and overall variation to different degrees, and that associations between historical observations and forecasts also exhibit repeated phase Structure(shown in Figure 2). These

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.

findings motivate an analysis based on repetition across cycles to determine what models retain and whether the retained components are predicted accurately.

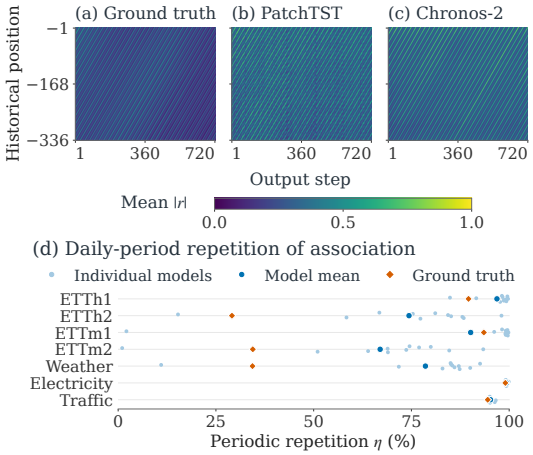


Figure 2: Demo of pre-experiment. The top row shows the correlation map on ETTh1. Each cell is the mean absolute Pearson correlation over channels for which the correlation is defined. The bottom row shows the periodic recurrence ratio η_P for seven datasets with daily periods.

We decompose both the true future and model forecasts into Structure and Dynamics. Structure consists of the window's mean level and a profile repeated across cycles, obtained by averaging values at the same phase across cycles and repeating the resulting profile. Dynamics describe changes in Structure over time, including level and amplitude variations across cycles, local shifts, and phase changes. Periodically repeated spikes belong to Structure, whereas slowly varying shifts across cycles can belong to Dynamics; the distinction is therefore based on phase repetition. Under the same period and scoring window, the errors of the two components are orthogonal, and total MSE decomposes exactly into the sum of Structure and Dynamic MSE, allowing direct analysis of the sources of model errors and gaps between models.

Our diagnostics show substantially better recovery of Structure than of Dynamics. Relative to predicting a zero component, Structure recovery scores above 80% and Dynamic recovery scores below 10%. Forecasting outputs also concentrate more on Structure, while Dynamics still contribute substantial error. Across forecasting intervals from the near to the distant future, advantages in modeling Structure often determine models' relative performance. These results indicate that, in the tasks examined, limitations in modeling Dynamics account for the main errors, whereas the ability to model Structure dominates differences between models.

This observation motivates allocating a limited number of parameters to modeling Structure. We propose TinyCycle, which divides historical observations into cycle blocks, directly models the Structure of the data, and forecasts recursively. TinyCycle does not need to model temporal relationships directly and can instead rely on linear combinations. It therefore requires only a small number of parameters; for a fixed number of historical cycles Q, its parameter count is independent of the forecast horizon, and it admits a closed-form least-squares solution.

As shown in Figure 1, TinyCycle achieves an average MAE of 0.345 with an average of 25.25 real-valued parameters, at a lower fitting cost than three lightweight baselines with matched periods. These results show that explicitly sharing periodic Structure can achieve competitive accuracy with a very small model. Our main contributions are as follows:

❶ **A forecasting diagnostic framework based on phase repetition.** We apply the same Structure–Dynamics decomposition to the true future and model outputs, using orthogonality to decompose total MSE into the two component errors and connect output Structure, component recovery, and error gaps between models.

❷ **An analysis of the distinct roles of Structure recovery and Dynamic error.** Experiments with multiple models on seven datasets show substantially better recovery of Structure than of Dynamics. Dynamics leave substantial residual error, whereas Structure accounts for more of the MSE improvement for most model pairs in the distant interval. These findings help explain the competitiveness of lightweight models and the bottlenecks in current forecasting methods.

❸ **TinyCycle, a periodic recursive model with closed-form fitting.** By combining shared cycle coefficients, age-weighted soft sharing, and a decaying boundary correction, TinyCycle performs long-term forecasting with at most 31 real-valued parameters, achieves lower average MSE and MAE than three lightweight baselines with matched periods across 32 tasks, and substantially reduces fitting cost.

2 Related Work

Time series perspectives. STL uses local regression to separate trend, seasonality, and remainder (RB, 1990), while singular spectrum analysis (SSA) extracts trend and oscillations through trajectory-matrix decomposition (Golyandina & Korobeynikov, 2014). Interpretable N-BEATS uses polynomial and Fourier bases for trend and seasonality, respectively (Oreshkin et al., 2019). TimesNet models within- and across-cycle variations through period-based two-dimensional representations (Wu et al., 2022), and TimeMixer mixes seasonal and trend information across scales (Wang et al., 2024). Our diagnostics instead apply the same phase-based orthogonal projection to true and predicted futures, separating Structure (mean level and shared profile) from its Dynamic residual and decomposing MSE to attribute forecasting improvements.

Traditional and deep time series forecasting. ARIMA models dependencies using differencing, autoregressive terms, and moving-average terms, while exponential smoothing tracks level, trend, and seasonality (Hyndman et al., 2014). Prophet adds known event effects to an additive model when external information is available (Taylor & Letham, 2018). LSTM learns dependencies through gated memory (Graves, 2012), and DeepAR shares an autoregressive recurrent network across series for probabilistic forecasting (Salinas et al., 2020). Autoformer combines decomposition with autocorrelation (Wu et al., 2021), FEDformer with frequency-domain representations (Zhou et al., 2022), while PatchTST uses patches and channel-independent modeling with parameters shared across channels (Nie et al., 2022).

Lightweight time series forecasting. DLinear forecasts decomposed trend and seasonality linearly, highlighting their importance at long horizons as the influence of recent states diminishes (Zeng et al., 2023). RLinear combines reversible instance normalization with temporal linear mapping (Li et al., 2023). SparseTSF learns across cycles at matching phases (Lin et al., 2024b), while TimeBase extracts basic temporal components for segment-level forecasting (Huang et al., 2025). MixLinear combines segment-wise temporal mappings with low-rank frequency transformations (Ma et al., 2026). CycleNet learns periodic templates and adds predicted residuals to them (Lin et al., 2024a).

Time series foundation models. Pretrained on heterogeneous data, first-generation Chronos quantizes scaled values into discrete tokens for probabilistic forecasting with a language modeling objective (Ansari et al., 2024), Moirai combines masked encoding, patch projections, and any-variate attention (Woo et al., 2024), and TimesFM uses a patch-based decoder for zero-shot forecasting (Das et al., 2023). Chronos-2 supports multivariate and covariate-conditioned forecasting (Ansari et al., 2025). Moirai 2.0 uses decoder-based quantile forecasting (Liu et al., 2025a), while Toto 2.0 studies model, data, and compute scaling (Khwaja et al., 2026).

3 Preliminaries

Time series forecasting Given historical observations $X = \{x_1, x_2, \ldots, x_H\}$ of length H with c channels, time series forecasting typically aims to predict the values $Y = \{y_1, y_2, \ldots, y_L\}$ of the same channels over the next L steps. As the forecasting distance increases, historical observations become less explanatory of the distant future, while the accumulation of unobserved factors increases forecasting uncertainty. This resembles a chaotic problem, in which errors grow through a butterfly effect.

Preliminary experiment: what current models capture To construct an effective decomposition of model outputs, we first examine the outputs of current models. The composition of real data differs from the portion that historical observations can explain. We therefore examine correlations of historical observations with the true future and with model outputs to analyze what models actually capture.

Figure 2(a–c) compares correlation maps between historical observations and the true future or model outputs. Repeated bands indicate that these associations have phase Structure. Figure 2(d) further uses the periodic recurrence ratio to measure the variation in the correlation maps that can be represented by shared phases, with a comparison against the true future. Statistical definitions and sensitivity results are provided in Appendix C.1. Comparing (a) with (b,c) shows that model outputs substantially strengthen correlations between periodic phases, a finding further supported by (d).

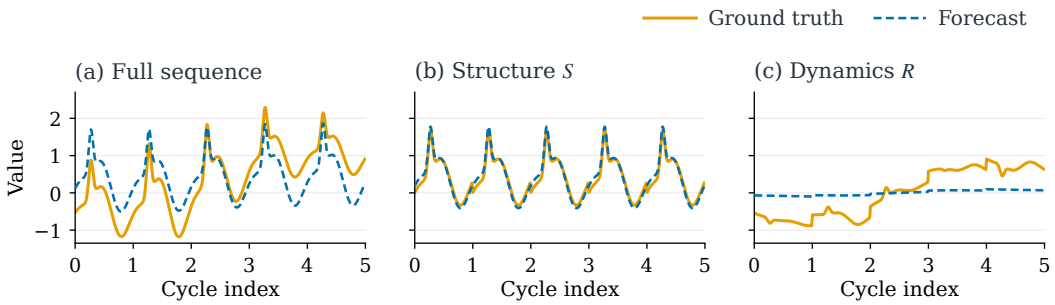


Figure 3: An analytically constructed illustration of forecast decomposition. The trajectory and its forecast are each projected into Structure and Dynamics. A visual comparison with moving-average (MA) decomposition of the same trajectory is shown in Figure 9.

4 Structure–Dynamics Decomposition

Time series are commonly decomposed into trends, seasonality, and abrupt events, an approach widely used in model design. However, this decomposition concerns the time series themselves and does not explain the interpretable relationship between inputs and outputs in forecasting and can not explain the forecast results.

The aforementioned observations show clear information sharing across cycles in forecasting outputs, suggesting that forecasts can be constructed from repeated Structure. As illustrated in Figure 3, we refer to the shared level and repeated profile within a window as **Structure**, and to the part that changes over time relative to this Structure as **Dynamics**. This approach decomposes both ground truth and forecasts, allowing us to compare them and analyze the sources of forecasting performance.

4.1 Decomposition of Structure

Structure. Among all trajectories that repeat the same profile, the mean minimizes the total squared deviation from the original trajectory (Liu et al., 2025b); see Appendix A.1 for the derivation. We therefore average values at the same position across cycles to obtain a shared profile. For example, for hourly electricity consumption, averaging observations at the same hour across days yields a 24-point mean daily profile. Repeating this profile day by day to the length of the original window produces the Structure. Let Z denote the trajectory to be decomposed and P the number of sampling points in one cycle. The phase at step h is $\phi(h) = (h-1) \bmod P$. Let $\mathcal{I}_p = \{h : \phi(h) = p\}$ denote the set of time steps at phase p. For each channel j, we compute

$$(\Pi_P Z)_{h,j} = \frac{1}{|\mathcal{I}_{\phi(h)}|} \sum_{s \in \mathcal{I}_{\phi(h)}} Z_{s,j}, \qquad S = \Pi_P Y, \quad \widehat{S} = \Pi_P \widehat{Y}. \tag{1}$$

Here, Π_P denotes the averaging and repetition operation described above. Mean profiles are computed separately for the true future Y and model forecast $\widehat{Y}$, yielding Structures S and $\widehat{S}$, respectively. Structure preserves the window's mean level, the relative highs and lows at positions within a cycle, and recurring peaks and troughs.

Dynamics. We define $R = Y - S$ and $\widehat{R} = \widehat{Y} - \widehat{S}$ as the temporal variation that Structure cannot explain. This includes local differences in peaks and troughs relative to the shared profile, turning points, phase changes, and level and amplitude modulation across cycles. The predictive value of Dynamics is measured by the error reduction obtained by retaining this component. For a window comprising complete cycles, Structure corresponds to the DC component and periodic harmonics, namely the fundamental periodic frequency and its integer multiples, while the remaining frequencies correspond to R. Complex spectral error further distinguishes amplitude and phase discrepancies; the calculations are given in Appendix A.6.

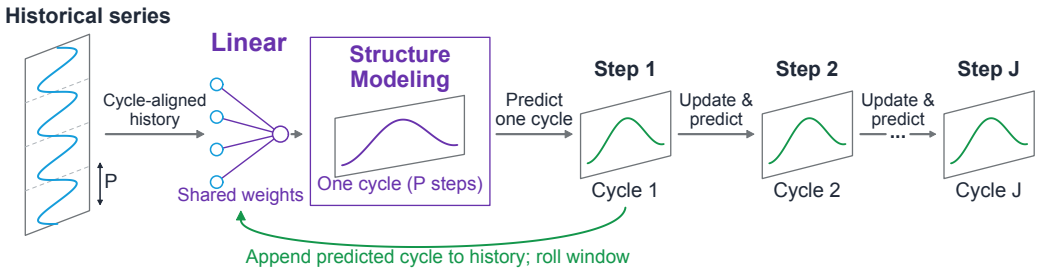


Figure 4: Cycle-by-cycle forecasting with TinyCycle. The historical sequence is divided into aligned blocks of length P. Linear weights the same position across historical cycles using shared coefficients to model the Structure of a single cycle, which is then extended into the future recursively.

4.2 Decomposition of Error Sources

Mean Squared Error We compare the Structure and Dynamics of forecasts with those of the true future separately. Let $e_S = \widehat{S} - S$ and $e_R = \widehat{R} - R$. The former reflects prediction errors in the window's mean level and repeated profile, whereas the latter reflects whether variations departing from the repeated profile are predicted accurately. Because these two types of error are orthogonal, total MSE is exactly the sum of their MSEs:

$$\mathrm{MSE}(\widehat{Y}, Y) = \underbrace{\frac{\|e_S\|_F^2}{Lc}}_{E_S} + \underbrace{\frac{\|e_R\|_F^2}{Lc}}_{E_R}. \tag{2}$$

Here, L is the forecast horizon and c is the number of channels. If model b has a lower total MSE than model a, we use $E_S^{(a)} - E_S^{(b)}$ and $E_R^{(a)} - E_R^{(b)}$ to distinguish how much of the error reduction comes from more accurate predictions of Structure and how much comes from more accurate predictions of Dynamics.

Mean Absolute Error MAE cannot be added directly in this way because the two component errors at the same time point may have opposite signs and cancel each other. Starting from the forecast of model a, we first replace its Structure with that of model b and then replace its Dynamics with those of model b, recording the MAE reduction at each step. We then reverse the replacement order and repeat the procedure. Averaging the MAE reductions for a given component across the two orders yields its contribution to the MAE gap between models. The two contributions sum to the total MAE gap; the formulas are given in Appendix A.4.

5 TinyCycle

TinyCycle represents Structure using historical cycle blocks and recursively extends it through shared coefficients. Figure 4 illustrates the alignment of historical cycles and the recursive prediction of future cycles.

Constructing Structure Given a period length P and a number of historical cycles Q, we divide the most recent $H = QP$ observations into $Z_0, Z_{-1}, \ldots, Z_{-(Q-1)} \in \mathbb{R}^{P\times c}$, where Z_0 is the most recent block and c is the number of channels. Each historical block has a coefficient w_i, and $w = (w_0, \ldots, w_{Q-1})^\top$ is shared across all channels and positions within a cycle. For example, predicting electricity consumption at a particular hour of the next day amounts to taking a weighted sum of observations at the same hour on previous days. The block length is selected on the validation set and is not restricted to a calendar-day period.

Forecasting Initialized with the historical blocks, the model sequentially generates $J = \lceil L/P \rceil$ future cycles:

$$\widetilde{Z}_r = \sum_{i=0}^{Q-1} w_i \widetilde{Z}_{r-1-i}, \qquad \widetilde{Z}_{-i} = Z_{-i}, \quad r = 1, \ldots, J. \tag{3}$$

After generating each new block, the model adds it to the most recent Q cycles, removes the oldest block, and predicts the next block using the same coefficients. Through recursion, TinyCycle extends stable periodic states in the history into the distant future. When the weights sum to 1, a repeated periodic profile remains unchanged after one recursion step. If the recursion matrix has a simple eigenvalue at 1 and all other eigenvalues have modulus less than 1, transient components that deviate from this profile also gradually decay. The trajectory thus returns from Structure plus Dynamics to stable Structure, giving TinyCycle some capacity to model Dynamics. The derivation is provided in Appendix B.1.

Boundary correction The boundary correction carries the local deviation of the final observation into near-term forecasts. Let $d_X = x_H - Q^{-1}\sum_{i=0}^{Q-1} Z_{-i}[P,:]$ denote the difference between the last observation and the mean at the final position of the historical cycles. The correction added to the output is

$$b_h = g\, d_X \exp[-h/(\tau P)], \qquad h = 1, \ldots, L. \tag{4}$$

$g \in [0,1]$ controls how much of the deviation is transferred, and $\tau > 0$ controls the decay rate. This term is added only to the forecasting output and does not participate in cycle recursion. It approaches zero as the forecasting distance increases and therefore does not change the long-term limit of the recursion.

Training Let $u_{t,j} = (x_{t-P,j}, \ldots, x_{t-QP,j})^\top$. Historical observations at the same position within a cycle are used to predict training time point t:

$$\widehat{w} = \arg\min_{w} \sum_{t \in \mathcal{T}_{\text{fit}}} \sum_{j=1}^{c} \frac{(x_{t,j} - u_{t,j}^\top w)^2}{s_j^2} + \lambda\rho\, \Omega_Q(w). \tag{5}$$

Here, $\mathcal{T}_{\text{fit}}$ is the set of training time points, s_j is the channel scale over the training segment, λ controls sharing strength, and ρ matches the scales of the fitting and penalty terms. The objective is quadratic in w and requires only solving a linear system of size Q. Scale definitions and solution details are provided in Appendix B.2.

Training encourages the coefficients of older historical cycles to be closer to one another. Ordering cycles from the most recent to the oldest, we set $a_i = ((i+1)/Q)^2$ and define the penalty as

$$\Omega_Q(w) = \frac{1}{\kappa_Q} \sum_{0 \le i < k < Q} a_i a_k (w_i - w_k)^2. \tag{6}$$

The older the two cycles, the more strongly differences between their coefficients are penalized. The coefficients need not be equal, and older coefficients are not required to be smaller. κ_Q normalizes the penalty scale; for $Q = 1$, we set $\Omega_1 = 0$. Definitions are given in Appendix B.2. With $\widehat{w}$ fixed, we fit g in closed form using multistep forecasting residuals on training windows, allowing the boundary correction to supplement near-term deviations. Both stages use only training data and require no gradient iterations. The complete settings and calculations are provided in Appendix B.2.

6 Experiments

We evaluate and analyze SparseTSF, GTR (Cao et al., 2026), PhaseFormer (Niu et al., 2026), Chronos-2, Toto2-1B, PatchTST, iTransformer (Liu et al., 2024), TimeMixer, DLinear, and Moirai-2 on eight datasets: the four ETT datasets, Electricity, Weather, Traffic, and Exchange(Zhou et al., 2021; Wu et al., 2021). Exchange serves as a counterexample with extremely weak periodicity and is excluded from some analyses over a common set of datasets; its Structure component represents the window level, as discussed in Appendix G. Unless otherwise specified, the following experiments use these datasets and models. Analyses on additional datasets are included in the appendix.

6.1 Is TinyCycle Effective at Modeling Structure?

This experiment evaluates TinyCycle's accuracy and computational cost at four forecast horizons: 96, 192, 336, and 720. The number of historical cycles for TinyCycle is $Q = \lfloor 720/P \rfloor$. TinyCycle uses 672 input steps on ETTm1 and ETTm2 because 720 cannot be evenly divided by 96, while 720 input steps on the others. Other models use 720-step inputs. Detailed settings are provided in Appendix B.4.

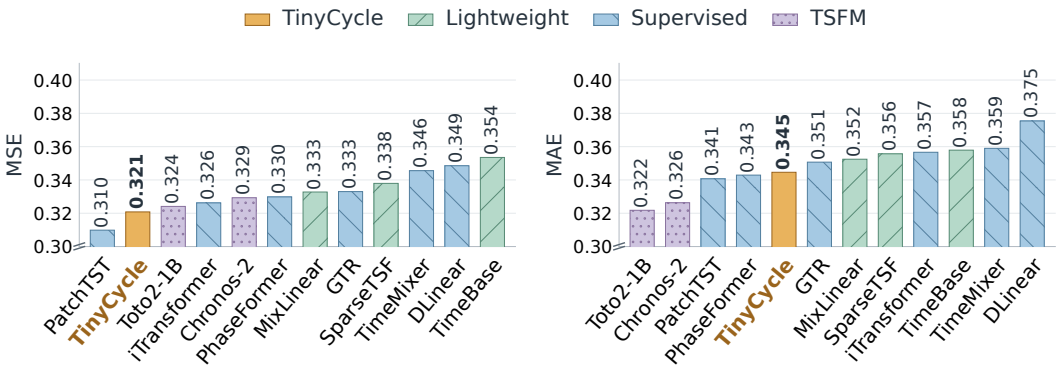


Figure 5: Average MSE and MAE of twelve models across all 32 tasks; lower is better for both metrics. TinyCycle uses the extended-lookback results in Table 4.

Table 1: Model efficiency. Red and blue indicate the lowest and second-lowest values, respectively, among the four models for each metric in each row, ranked using unrounded values.

Dataset	TinyCycle			MixLinear			SparseTSF			TimeBase		
	Params	Train	Test	Params	Train	Test	Params	Train	Test	Params	Train	Test
ETTh1	31.000	0.049	0.004	153.000	10.717	0.075	445.000	3.939	0.051	284.000	1.533	0.020
ETTh2	31.000	0.043	0.004	198.000	6.256	0.076	445.000	4.596	0.051	264.500	1.209	0.020
ETTm1	8.000	0.037	0.016	156.000	60.822	0.647	127.000	37.273	0.519	253.750	17.800	0.167
ETTm2	8.000	0.031	0.015	128.000	42.453	0.616	127.000	43.326	0.527	258.750	22.608	0.182
Weather	31.000	0.094	0.028	130.000	154.974	0.975	445.000	53.748	0.672	5964.000	58.403	0.341
Electricity	31.000	0.953	0.153	190.000	2605.510	13.008	445.000	636.191	8.900	284.000	623.054	4.847
Traffic	31.000	2.965	0.282	198.000	3398.915	23.737	445.000	1407.466	16.180	435.250	948.395	8.453
Exchange	31.000	0.034	0.002	130.000	5.640	0.037	445.000	2.285	0.026	2272.000	1.387	0.013
Average	25.250	0.526	0.063	160.375	785.661	4.896	365.500	273.603	3.366	1252.031	209.299	1.755

Accuracy comparison. Figure 5 retains the average MSE and MAE of TinyCycle and eleven fixed baselines, covering different model types and zero-shot TSFMs. Per-task results for TinyCycle are reported in Table 4. SparseTSF, TimeBase, and MixLinear use the same modeling periods as TinyCycle. TinyCycle uses deterministic closed-form fitting. TinyCycle achieves lower MSE and MAE than other light models and most of complex models. These results indicate that TinyCycle's modeling of Structure can recover some of the modeling capacity of complex models. Nevertheless, performance differences remain relative to more complex models.

Parameters and computational cost. Table 1 uses the extended-lookback configuration described above. Efficiency is measured on an Intel Core i5-13490F CPU with 2×16 GB of DDR5-5600 memory, using four fixed physical performance cores and four computation threads. TinyCycle uses approximately 15.7% as many parameters, 0.25% of the training time, and 3.6% of the test time as the second-best model for each metric. By focusing on modeling sequence Structure and reducing reliance on complex mappings, TinyCycle reduces model size and computational cost while maintaining comparable forecasting performance.

6.2 How Do Models Recover Structure and Dynamics?

This section analyses of forecasting distance, performance attribution, and counterexamples. Within the Structure–Dynamics decomposition framework, this section examines the recovery of Structure and Dynamics in model forecasts to analyze the sources and limitations of existing models' performance and explain TinyCycle's effectiveness. Using the error of predicting zero for each Structure or Dynamic component as a reference, we define the recovery score as

$$\mathrm{Rec}(Z) = 1 - \frac{\sum_{o,h,j} (\widehat{Z}_{o,h,j} - Z_{o,h,j})^2}{\sum_{o,h,j} Z_{o,h,j}^2}, \qquad Z \in \{S, R\}, \tag{7}$$

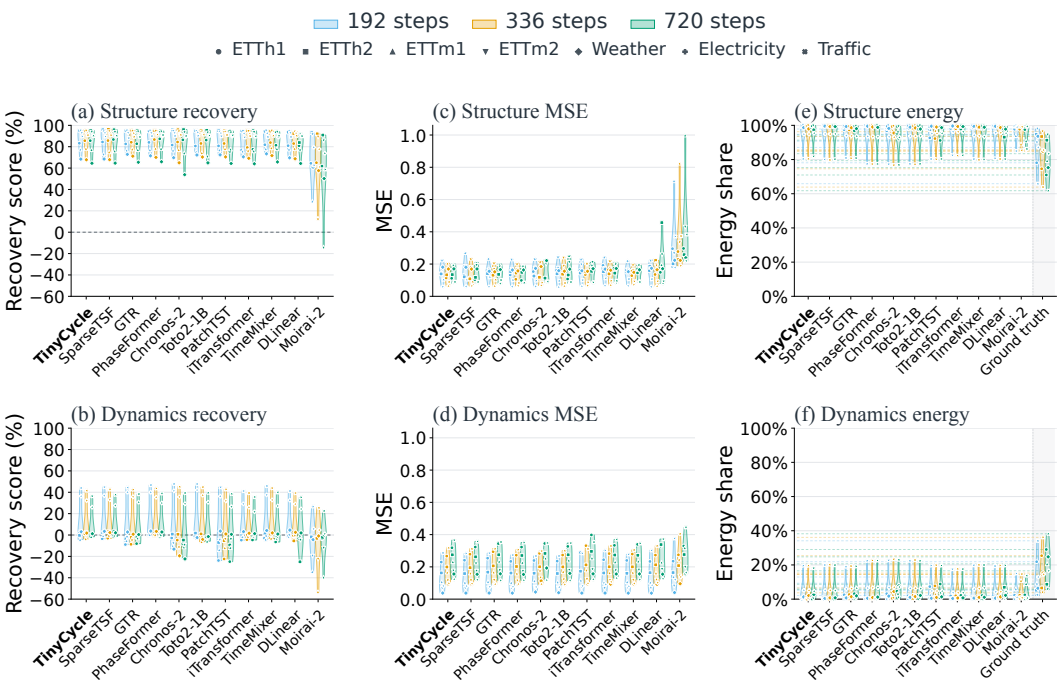


Figure 6: Six-panel distributions of Structure and Dynamics, with the same eleven models in every panel. The top and bottom rows show Structure and Dynamics, respectively. The left column (a,b) shows recovery scores, the middle column (c,d) component MSE, and the right column (e,f) energy shares. Blue, orange, and green consistently denote the 192-, 336-, and 720-step prefixes, while marker shapes distinguish the seven datasets.

where o, h, j denote the forecast origin, forecast step, and channel, respectively. For each length, the score is computed over its own range $h = 1, \ldots, H$. A higher score indicates better recovery relative to predicting a zero component.

Mismatch between modeling emphasis and error distribution Figure 6(a,b) shows the distributions of Structure and Dynamic recovery scores. Models generally achieve much higher recovery scores for Structure than for Dynamics. At the 192-, 336-, and 720-step prefixes, the average Structure recovery score across the eleven models exceeds 80%, whereas the Dynamic recovery score remains below 10%. Figure 6(e,f) compares the per-dataset forecast and ground-truth energy shares of Structure and Dynamics, respectively. The results show that model outputs are more concentrated on Structure, while their Dynamics are substantially reduced relative to the ground truth. Figure 6(c,d) shows that, despite accounting for a smaller share of the real data, Dynamics contribute an absolute error comparable to that of Structure.

Dynamic error increases with forecast horizon, while Structure error changes little Figure 6(c,d) shows that Dynamic MSE increases between both pairs of adjacent lengths for every model–dataset combination. In contrast, Structure MSE changes relatively little. As actual error increases with the forecast horizon while Structure error shows no substantial change, Structure can be modeled over the long term in a relatively simple way, whereas Dynamics account for most of the growth in long-term error. These results show that Structure is the main focus of the models, explaining why TinyCycle remains competitive with only a small number of cycle coefficients.

6.3 Sources of Performance Differences Between Models

This section further compares the sources of performance differences between models to better explain TinyCycle's advantages over some models and its performance bottlenecks.

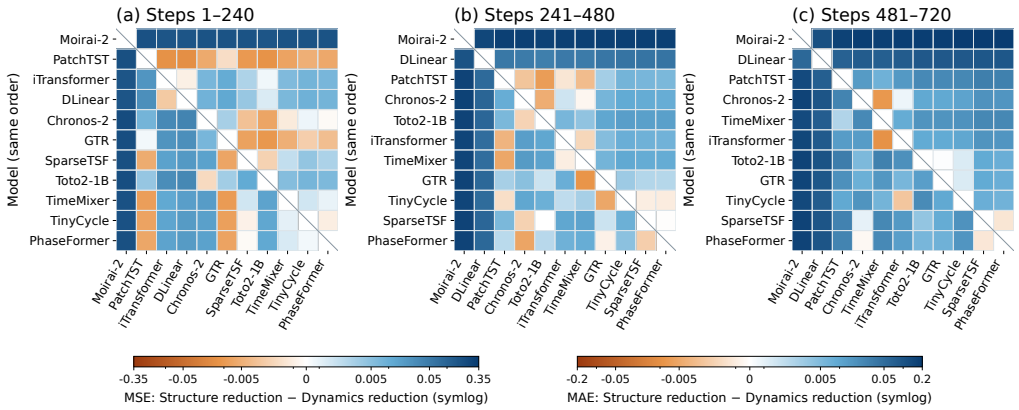


Figure 8: Comparison of component error reductions between models. In each panel, both axes use the same model order, sorted by decreasing total MSE. The upper and lower triangles show MSE and MAE comparisons, respectively. Blue indicates a greater reduction in Structure, orange a greater reduction in Dynamics, and white equal reductions.

Relative gaps in Structure remain stable, while Dynamic gaps are smaller in the distant interval. We first decompose each full 720-step forecast and its ground truth over the entire window, and then score the segments spanning steps 1–192, 193–336, and 337–720. Let $E^{(q)}_{d,m}(I)$ denote the mean squared error over all origins and channels for dataset d and model m in interval I, where $q \in \{\text{full}, \text{body}, \text{dyn}\}$ denotes the full forecast, Structure, and Dynamics, respectively. The mean excess relative MSE in Figure 7 is

$$G^{(q)}(I) = \frac{100\%}{7 \times 11} \sum_{d=1}^{7} \sum_{m=1}^{11} \left[\frac{E^{(q)}_{d,m}(I)}{\min_{m'} E^{(q)}_{d,m'}(I)} - 1 \right]. \tag{8}$$

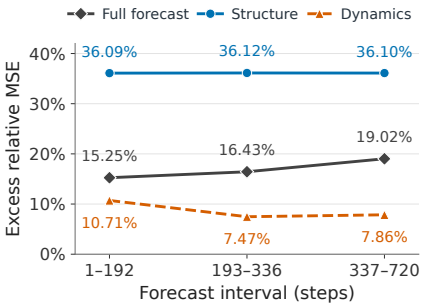


Figure 7: Mean excess relative MSE in three forecasting intervals, computed using Equation (8); lower is better.

The best model is determined separately for each dataset, interval, and metric. Figure 7 shows how relative differences between models vary across forecasting intervals. The mean excess relative MSE for Structure remains essentially stable. The corresponding values for Dynamics are lower. In all three intervals, the average gap from the respective best level is greater for Structure than for Dynamics.

Structure dominates performance differences between models. Figure 8 directly compares the reductions in Structure and Dynamic errors when one model outperforms another. Across the three intervals, Structure MSE decreases more than Dynamic accounting for 61.82%, 81.82%, and 96.36% of all model pairs. Some model pairs still achieve most of their error reduction through Dynamics in the near interval. Together with the preceding findings, this indicates that Dynamic recovery has a greater effect in the near term than in the distant term, whereas in the distant term, the effect of recovering Structure largely overshadows that of recovering Dynamics.

7 Conclusion

We decompose the true future and forecasting outputs into Structure and Dynamics based on phase repetition and use orthogonality to analyze the sources of MSE. We demonstrate that Structure is the main focus of current models. Motivated by this observation, TinyCycle recursively extends historical Structure through shared cycle coefficients and uses closed-form fitting, achieving performance comparable to state-of-the-art models with extremely few parameters and very low computational cost.

AI USE STATEMENT

We used large language models (LLMs) for code generation, experimental data analysis, figure generation, manuscript polishing, and translation. We also used LLMs to assist with the proofs in Appendix A and to organize the data presented in the appendices. The experiments in Appendix E.2 were designed by an LLM, and we verified their results. All other LLM-generated content was reviewed by the human authors. We take full responsibility for all content of this paper.

REPRODUCIBILITY STATEMENT

TinyCycle uses deterministic closed-form fitting. With the same data, preprocessing, hyperparameters, and numerical precision, reproduced forecasting metrics are expected to agree up to floating-point rounding error. Fitting and inference efficiency depend on the hardware and software environment, including the processor, numerical libraries, and thread settings, so execution times may vary across devices and environments.

REFERENCES

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

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815*, 2024.

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.

Xiang Ao, Yinyu Tan, Mengru Chen, Zhenghao Xie, and Yijun Cao. Timesl: A generic hierarchical framework for mitigating spectral bias and gradient starvation in time series forecasting. *Available at SSRN 6561123*.

Cristian Bodnar, Wessel P Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A Weyn, Haiyu Dong, et al. A foundation model for the earth system. *Nature*, 641(8065):1180–1187, 2025.

Fanpu Cao, Lu Dai, Jindong Han, and Hui Xiong. Enhancing multivariate time series forecasting with global temporal retrieval. *arXiv preprint arXiv:2602.10847*, 2026.

Lingxiao Cao, Bin Wang, Guiyuan Jiang, Yanwei Yu, and Junyu Dong. Spatiotemporal-aware trend-seasonality decomposition network for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 11463–11471, 2025.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688*, 2023.

Nina Golyandina and Anton Korobeynikov. Basic singular spectrum analysis and forecasting with r. *Computational Statistics & Data Analysis*, 71:934–954, 2014.

Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.

Qihe Huang, Zhengyang Zhou, Kuo Yang, Zhongchao Yi, Xu Wang, and Yang Wang. Timebase: The power of minimalism in efficient long-term time series forecasting. In *ICML*, 2025.

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

Emaad Khwaja, Chris Lettieri, Gerald Woo, Eden Belouadah, Marc Cenac, Guillaume Jarry, Enguerrand Paquin, Xunyi Zhao, Viktoriya Zhukov, Othmane Abou-Amal, et al. Toto 2.0: Time series forecasting enters the scaling era. *arXiv preprint arXiv:2605.20119*, 2026.

Zhe Li, Shiyi Qi, Yiduo Li, and Zenglin Xu. Revisiting long-term time series forecasting: An investigation on linear mapping. *arXiv preprint arXiv:2305.10721*, 2023.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. *Advances in Neural Information Processing Systems*, 37:106315–106345, 2024a.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: Modeling long-term time series forecasting with 1k parameters. *arXiv preprint arXiv:2405.00946*, 2024b.

Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. Moirai 2.0: When less is more for time series forecasting. *arXiv preprint arXiv:2511.11698*, 2025a.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 11116–11140, 2024.

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. *arXiv preprint arXiv:2502.00816*, 2025b.

Aitian Ma, Dongsheng Luo, and Mo Sha. Mixlinear: Extreme low resource multivariate time series forecasting with $0.1k$ parameters. In *International Conference on Learning Representations*, volume 2026, pp. 77764–77789, 2026.

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

Yiming Niu, Jinliang Deng, and Yongxin Tong. Phaseformer: From patches to phases for efficient and effective time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 8262–8285, 2026.

Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.

Young-Jin Park, François Germain, Jing Liu, Ye Wang, Toshiaki Koike-Akino, Gordon Wichern, Navid Azizan, Christopher Laughman, and Ankush Chakrabarty. Probabilistic forecasting for building energy systems using time-series foundation models. *Energy and Buildings*, pp. 116446, 2025.

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S Jensen, Zhenli Sheng, et al. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *arXiv preprint arXiv:2403.20150*, 2024.

CLEVELAND RB. Stl: A seasonal-trend decomposition procedure based on loess. *J Off Stat*, 6: 3–73, 1990.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International journal of forecasting*, 36(3): 1181–1191, 2020.

Sean J Taylor and Benjamin Letham. Forecasting at scale. *The American Statistician*, 72(1):37–45, 2018.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 38626–38652, 2024.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. *arXiv preprint arXiv:2402.02592*, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

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

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

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

Zhuohang Zhu, Haodong Chen, Qiang Qu, and Vera Chung. Fincast: A foundation model for financial time-series forecasting. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pp. 4539–4549, 2025.

A Theoretical Proofs for the Structure–Dynamics Decomposition

A.1 Least-Squares Solution for the Shared Profile

For a given period P, the shared profile assigns the same value to all positions with the same phase, without restricting the shape across different phases. For a trajectory Z, finding its closest shared representation is equivalent to

$$\min_{s\in\mathbb{R}^{P\times c}} \sum_{j=1}^{c}\sum_{p=0}^{P-1}\sum_{h\in\mathcal{I}_p}(Z_{h,j}-s_{p,j})^2. \tag{9}$$

Each observed phase corresponds to a strictly convex univariate quadratic objective. Setting its derivative to zero gives

$$s_{p,j}=\frac{1}{|\mathcal{I}_p|}\sum_{h\in\mathcal{I}_p} Z_{h,j}.$$

This is the phase mean in Eq. (1). Equal weighting of time points gives equal contributions to observations at the same phase, without requiring independent or Gaussian perturbations, or equal means and amplitudes across cycles.

For a window containing complete cycles, $L=KP$, arrange each channel as a $K\times P$ matrix. Let U denote the operator that circularly shifts the entire window by one period. Then

$$\Pi_P=\frac{1}{K}\sum_{r=0}^{K-1}U^r. \tag{10}$$

The right-hand side averages the K observations at each phase, each appearing once, and thus extracts the shared component satisfying $UZ=Z$. This representation preserves phase information and recurring spikes; circular shifts are used only to characterize the subspace on a finite window. For windows that do not contain an integer number of cycles, Eq. (9) is solved directly using the actual number of observations at each phase.

The arithmetic mean provides the shared fit under squared distance and yields a linear orthogonal decomposition into Structure and Dynamics. The phase median provides the shared fit under absolute distance, but generally yields neither a linear projection nor additive squared errors. We use the same definition of Structure for MSE and MAE, and compute MAE contributions through symmetric replacement evaluated with the actual loss. The optimality of the shared representation specifies the decomposition; the predictive value of each component is still determined by its prediction error.

A.2 Orthogonal Error Decomposition and Gains from Dynamic Recovery

After grouping positions with the same phase, the corresponding block of Π_P is $\mathbf{1}\mathbf{1}^\top/n_p$, where $n_p = |\mathcal{I}_p|$. Each block is symmetric and idempotent, so

$$\Pi_P^\top = \Pi_P, \qquad \Pi_P^2 = \Pi_P, \qquad \Pi_P(I - \Pi_P) = 0.$$

Let $e = \widehat{Y} - Y$. The Structure error $e_S = \Pi_P e$ and the Dynamic error $e_R = (I - \Pi_P)e$ then satisfy

$$\langle e_S, e_R \rangle = \langle e, \Pi_P(I - \Pi_P)e \rangle = 0.$$

Expanding $\|e_S + e_R\|_F^2$ gives Eq. (2). The proof allows different numbers of observations across phases and therefore also applies to windows that do not contain an integer number of cycles. Setting $e = Y$ or $e = \widehat{Y}$ similarly gives a decomposition of the squared energy of the true or predicted trajectory.

Trajectories that are constant over the window form a subspace of the Structure subspace. The profile obtained by subtracting the window mean from the Structure sums to zero over the window and is therefore orthogonal to the constant term. Structure MSE can thus be further decomposed into window-level error and demeaned-profile error; the latter retains deviations in profile magnitude.

Corollary: component recovery scores and gains from Dynamic prediction. Let $\langle\cdot\rangle$ denote an equally weighted average over the same error elements, and let $E_R = \langle(\widehat{R} - R)^2\rangle$. The scoring range coincides with the phase-projection window that defines Structure and Dynamics. When $\langle R^2\rangle > 0$, the Dynamic recovery score in Eq. (7) is $\mathrm{Rec}(R) = 1 - E_R/\langle R^2\rangle$, with the squared error of a zero prediction for the Dynamic component as the reference.

Setting the model's Dynamic component to zero gives the prediction $\widehat{S}$. By orthogonality between the Structure error and the true Dynamic component,

$$\mathrm{MSE}(\widehat{S}, Y) = \langle(\widehat{S} - S)^2\rangle + \langle R^2\rangle.$$

Subtracting the orthogonal error decomposition of the full prediction gives

$$\mathrm{MSE}(\widehat{S}, Y) - \mathrm{MSE}(\widehat{Y}, Y) = \langle R^2\rangle - E_R = \langle R^2\rangle\,\mathrm{Rec}(R). \tag{11}$$

Thus, a positive Dynamic recovery score is equivalent to a reduction in total MSE from retaining the model's Dynamic component. The score is normalized by the true Dynamic energy; the relative improvement in total error attributable to Dynamics also depends on the chosen total-error reference. When the true Dynamic energy is zero, the score is undefined, but the first two differences in the above equation remain defined.

A.3 Five-Term Orthogonal Decomposition of Level, Profile, and Dynamics

For a window containing complete cycles, $L = KP$, let $k = 0, \ldots, K-1$ and $p = 0, \ldots, P-1$, and write the sequence in each channel as

$$x_{k,p} = \mu + u_p + \ell_k + a_k q_p + r_{k,p}. \tag{12}$$

Here, μ is the window mean, u_p is the demeaned average phase profile, and $\ell_k = P^{-1}\sum_p x_{k,p} - \mu$ is the deviation of each cycle mean from μ. The true future's u_p is normalized to zero mean and unit root mean square to obtain q_p. The predicted and true sequences share this direction; when the template energy is zero, we set $q = 0$. This direction is used for component-error analysis after prediction is complete.

Let $v_{k,p} = x_{k,p} - \mu - u_p - \ell_k$. When $q \neq 0$, define

$$a_k = \frac{1}{P}\sum_p v_{k,p} q_p, \qquad r_{k,p} = v_{k,p} - a_k q_p;$$

When $q = 0$, set $a_k = 0$ and $r_{k,p} = v_{k,p}$. By the definitions of the means,

$$\sum_p u_p = 0, \quad \sum_k \ell_k = 0, \quad \sum_p v_{k,p} = 0, \quad \sum_k v_{k,p} = 0.$$

The identity $\sum_k v_{k,p} = 0$ implies $\sum_k a_k = 0$. Together with $\sum_p q_p = 0$, this ensures that every row and column of r sums to zero. The projection definition also gives $\sum_p r_{k,p} q_p = 0$.

These identities ensure pairwise orthogonality among the constant term, phase profile, cycle level, amplitude modulation, and residual waveform. Because the predicted and true sequences use the same q, the differences between their corresponding components also satisfy these constraints, so the five error terms are pairwise orthogonal. Structure comprises the first two terms and Dynamics the last three, giving

$$E_R = E_\ell + E_{aq} + E_r.$$

The residual-waveform error E_r is a subcomponent of the Dynamic error and retains residuals arising from phase changes and within-cycle trends.

A.4 MAE Attribution through Symmetric Component Replacement

Fix two models a, b and consider the performance improvement from a to b. Let $S_0 = \widehat{S}^{(a)}$, $S_1 = \widehat{S}^{(b)}$, $R_0 = \widehat{R}^{(a)}$, and $R_1 = \widehat{R}^{(b)}$, and define $F_{ij} = \mathrm{MAE}(S_i + R_j, Y)$. The four hybrid outputs use only model predictions; the true future is used for evaluation. Compute the loss changes for the two orders—replacing Structure first and replacing Dynamics first—and average the two orders equally to obtain

$$G_S = \tfrac{1}{2}[(F_{00} - F_{10}) + (F_{01} - F_{11})], \tag{13}$$

$$G_R = \tfrac{1}{2}[(F_{00} - F_{01}) + (F_{10} - F_{11})], \tag{14}$$

$$G_S + G_R = F_{00} - F_{11}. \tag{15}$$

The last equality follows by adding the first two equations and canceling the intermediate losses. This symmetric average allocates the interaction between the two components to both; an individual contribution can be negative or exceed the total improvement. If the two models and the ground truth use the same projection and are scored within its window, replacing MAE in F_{ij} with MSE gives $E_S^{(i)} + E_R^{(j)}$ by orthogonality. The two contributions then reduce to the respective differences in component errors.

Within Structure, we can further distinguish the window level from the demeaned profile. Consider four replacement orders: level, profile, Dynamics; profile, level, Dynamics; Dynamics, level, profile; and Dynamics, profile, level. In each order, the two Structure components are adjacent, and the sum of their loss changes equals the change from replacing Structure as a whole. Equally averaging the four orders therefore preserves G_S as the sum of the level and profile contributions.

A.5 Allocation of Actual MAE and Error Cancellation

For a fixed prediction, define

$$A_S = \langle |e_S| \rangle, \qquad A_R = \langle |e_R| \rangle, \qquad A = \langle |e_S + e_R| \rangle.$$

Starting from zero error, adding the Structure error first increases the loss by A_S, whereas adding it second increases the loss by $A - A_R$. Equally averaging the two orders gives the allocation of the actual total MAE:

$$C_S^{\mathrm{MAE}} = \tfrac{1}{2}(A_S + A - A_R), \qquad C_R^{\mathrm{MAE}} = \tfrac{1}{2}(A_R + A - A_S). \tag{16}$$

The two terms sum to A. The inequality $|e_R| \leq |e_S + e_R| + |e_S|$, together with its counterpart obtained by exchanging the two errors, gives $A_R \leq A + A_S$ and $A_S \leq A + A_R$, so both allocations are nonnegative. These allocations differ from the standalone component MAEs A_S, A_R.

The triangle inequality also gives the amount of error cancellation, $K = A_S + A_R - A \geq 0$. The above allocations can be written as $C_S^{\mathrm{MAE}} = A_S - K/2$ and $C_R^{\mathrm{MAE}} = A_R - K/2$, symmetrically distributing the cancellation. For any two models B, T, when $A_B > 0$, define

$$\begin{aligned} \Delta_S &= 100\frac{A_{S,B} - A_{S,T}}{A_B}, \qquad \Delta_R = 100\frac{A_{R,B} - A_{R,T}}{A_B}, \\ \Delta_K &= 100\frac{K_T - K_B}{A_B}, \qquad \Delta = 100\frac{A_B - A_T}{A_B} = \Delta_S + \Delta_R + \Delta_K. \end{aligned} \tag{17}$$

The last equality follows by subtracting the identities $A_m = A_{S,m} + A_{R,m} - K_m$ for the two models. Δ_S and Δ_R measure improvements in the component errors themselves, while $\Delta_K > 0$ indicates greater error cancellation for model T. All three terms retain their signs and use model B's total MAE as their common denominator. Equation (15) allocates performance gains through hybrid outputs, whereas this equation directly retains the changes in the two component errors and their cancellation.

A.6 Structure Error in the Complex Spectrum

On a window containing complete cycles, $L = KP$, define the orthonormally normalized discrete Fourier transform $\mathcal{F}$. For the Fourier basis at frequency index k, the multiplier in Eq. (10) is

$$\frac{1}{K}\sum_{r=0}^{K-1} e^{2\pi \mathrm{i} kr/K}.$$

When k is a multiple of K, this geometric sum equals 1; at all other indices, it equals 0. The frequency support of the phase-repetition subspace is therefore $\mathcal{H}_P = \{0, K, 2K, \ldots, (P-1)K\}$. By Parseval's identity,

$$E_S = \frac{1}{Lc}\sum_{j=1}^{c}\sum_{k\in\mathcal{H}_P} |\mathcal{F}(\widehat{Y})_{k,j} - \mathcal{F}(Y)_{k,j}|^2. \tag{18}$$

Summing over the complementary set gives E_R. This expression uses the full two-sided spectrum of the original trajectory. The time-domain projection remains valid for windows that do not contain an integer number of cycles, but cannot be directly replaced by the above discrete harmonic indices.

For a single nonzero frequency, denote the true and predicted coefficients by $Ae^{\mathrm{i}\theta}$ and $\widehat{A}e^{\mathrm{i}\widehat{\theta}}$. Expanding the squared complex modulus and using $\widehat{A}^2 + A^2 - 2\widehat{A}A\cos(\widehat{\theta} - \theta) = (\widehat{A} - A)^2 + 2\widehat{A}A[1 - \cos(\widehat{\theta} - \theta)]$ gives

$$|\widehat{A}e^{\mathrm{i}\widehat{\theta}} - Ae^{\mathrm{i}\theta}|^2 = (\widehat{A} - A)^2 + 2\widehat{A}A[1 - \cos(\widehat{\theta} - \theta)]. \tag{19}$$

Magnitude and phase deviations jointly determine the complex spectral error. When an amplitude is zero, the complex difference is used directly; the DC term is treated as a signed mean error.

A.7 MSE, MAE, and the Variance of Absolute Errors

If MSE and MAE for task j are computed with equal weights over the same error elements, then

$$V_j = \mathrm{MSE}_j - \mathrm{MAE}_j^2 = \mathrm{Var}(|e_j|), \tag{20}$$

$$\Delta\overline{\mathrm{MSE}} = \overline{\Delta(\mathrm{MAE}_j^2)} + \overline{\Delta V_j}. \tag{21}$$

The first equality follows from $e_j^2 = |e_j|^2$ and the variance identity $\mathrm{Var}(|e_j|) = \langle |e_j|^2\rangle - \langle |e_j|\rangle^2$. Taking differences between two models in the same direction and averaging with the same task weights gives the second equality.

The mean of the within-task variances of absolute errors differs from macro-averaged MSE minus the square of macro-averaged MAE. Under equal task weighting, they satisfy

$$\overline{\mathrm{MSE}} - \left(\overline{\mathrm{MAE}}\right)^2 = \overline{V} + \mathrm{Var}_j(\mathrm{MAE}_j).$$

The second term on the right arises from differences in MAE across tasks and cannot be counted as within-task error variance.

B THEORY AND IMPLEMENTATION OF TINYCYCLE

B.1 LONG-TERM BEHAVIOR OF THE RECURRENCE UNDER SPECTRAL CONDITIONS

Age-weighted soft sharing constrains neither the sum of the weights nor the eigenvalues. The following conclusions about the limit of the recurrence assume that the given coefficients satisfy the spectral conditions.

Fix a channel and a position within a block, and write $z_r = \widetilde{Z}_r[p, c]$. The recurrence in the main text can be expressed as

$$v_r = A(w)v_{r-1}, \qquad v_r = (z_r, z_{r-1}, \ldots, z_{r-Q+1})^\top, \tag{22}$$

where the companion matrix is

$$A(w) = \begin{pmatrix} w_0 & w_1 & \cdots & w_{Q-2} & w_{Q-1} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{23}$$

For $Q = 1$, set $A(w) = (w_0)$. Thus, $z_r = e_1^\top A(w)^r v_0$, and all forecast horizons share the same dynamical operator.

Theorem 1 (Periodic limit under spectral conditions)**.** *Assume that $\sum_{i=0}^{Q-1} w_i = 1$, that $A(w)$ has a simple eigenvalue at 1, and that all other eigenvalues have modulus less than 1. Then the base predictions of TinyCycle satisfy*

$$\widetilde{Z}_r \longrightarrow S_\infty(X), \qquad S_\infty(X) = \sum_{j=0}^{Q-1} \pi_j Z_{-j}, \tag{24}$$

where

$$\pi_j = \frac{\displaystyle\sum_{i=j}^{Q-1} w_i}{\displaystyle\sum_{i=0}^{Q-1} (i+1) w_i}, \qquad j = 0, \ldots, Q-1. \tag{25}$$

There exist $q \in (0, 1)$ and $K < \infty$ such that

$$\|\widetilde{Z}_r - S_\infty(X)\|_F \le K q^r \left(\sum_{j=0}^{Q-1} \|Z_{-j}\|_F^2 \right)^{1/2}. \tag{26}$$

Proof. The constraint on the sum of the weights gives $A\mathbf{1} = \mathbf{1}$. Let $\pi^\top A = \pi^\top$ and $\pi^\top \mathbf{1} = 1$. Solving for the left eigenvector of the companion matrix entry by entry gives Eq. (25). The denominator equals the derivative of the characteristic polynomial of A at 1 and is nonzero by the simple-root condition.

Define $\Pi = \mathbf{1}\pi^\top$ and $A_\perp = A - \Pi$. Then

$$\Pi^2 = \Pi, \qquad A_\perp \Pi = \Pi A_\perp = 0, \qquad \rho(A_\perp) < 1.$$

Hence, for $r \ge 1$, we have $A^r = \Pi + A_\perp^r$ and

$$z_r = \underbrace{\pi^\top v_0}_{\text{steady state}} + \underbrace{e_1^\top A_\perp^r v_0}_{\text{transient component}}. \tag{27}$$

For any $q \in (\rho(A_\perp), 1)$, there exists $K < \infty$ such that $\|A_\perp^r\|_2 \le K q^r$. Applying this bound to each channel and each position within a block, and then summing, gives Eqs. (24) and (26). □

One attainable set of sufficient conditions is $w_i > 0$ and $\sum_i w_i = 1$. In this case, A is a primitive stochastic matrix: its unit eigenvalue is simple, and all other eigenvalues are strictly contractive. More generally, let

$$p_w(z) = z^Q - \sum_{i=0}^{Q-1} w_i z^{Q-1-i}. \tag{28}$$

When the weights sum to 1, requiring all roots of $p_w(z)/(z-1)$ to satisfy $|z| < 1$ gives the spectral conditions in Theorem 1. These conditions allow negative weights and can therefore represent damped extrapolation.

Finite-window projection and the recurrence limit. $S_\infty(X)$ is the limiting profile determined by the history X and the recurrence coefficients, whereas $\widehat{S}$ in Eq. (1) is the phase projection over a finite prediction window. When the diagnostic period equals the block length and the window contains J complete predicted blocks, writing $\widetilde{Z}_r = S_\infty(X) + D_r$ gives the projected profile

$$\frac{1}{J}\sum_{r=1}^{J} \widetilde{Z}_r = S_\infty(X) + \frac{1}{J}\sum_{r=1}^{J} D_r. \tag{29}$$

Thus, Structure over a finite window may still include the average contribution of transients. Under the conditions of Theorem 1, as the window moves farther into the future, this contribution vanishes with the decay of D_r. A weight sum of 1 ensures that repeating profiles are fixed points, while the spectral conditions further ensure convergence of the recurrence to such a profile.

B.2 Age-Weighted Soft Sharing and Two-Stage Fitting

Age weights and the shared center. For the age weights in Eq. (6), write $a = (a_0, \ldots, a_{Q-1})^\top$ and $A_a = \sum_i a_i$. The corresponding difference-penalty matrix is

$$D_Q = A_a \operatorname{diag}(a) - aa^\top, \qquad R_Q = D_Q/\kappa_Q,$$
$$\kappa_Q = \max\left(\frac{\operatorname{tr}(D_Q)}{Q}, 10^{-12}\right), \tag{30}$$

Then $w^\top R_Q w = \Omega_Q(w)$. For $Q > 1$, the mean diagonal entry of the normalized matrix is 1; for $Q = 1$, we directly set $R_1 = 0$.

This form is also equivalent to drawing the coefficients toward a center jointly determined by them:

$$\sum_{i<k} a_i a_k (w_i - w_k)^2 = A_a \min_m \sum_i a_i (w_i - m)^2,$$
$$m^\star = \frac{\sum_i a_i w_i}{A_a}. \tag{31}$$

The shared center $m^\star$ is eliminated algebraically and does not need to be stored separately. The regularization term is zero for a constant coefficient vector. The penalty therefore constrains differences without prespecifying the common center as zero or $1/Q$.

Training time points and cycle coefficients. Let $x_{t,j}$ be the observation standardized using training-set statistics, and let s_j be the unbiased standard deviation of channel j in that training segment, with a lower bound of 10^{-4}. Using the regression vector in Eq. (5), denote the set of training time points by $\mathcal{T}$; the sampling rule is given in Appendix B.4. The regression statistics and scale are

$$G = \sum_{t\in\mathcal{T},j} s_j^{-2} u_{t,j} u_{t,j}^\top, \qquad b = \sum_{t\in\mathcal{T},j} s_j^{-2} u_{t,j} x_{t,j},$$
$$\rho = \max\left(\frac{\operatorname{tr}(G)}{Q}, 10^{-12}\right). \tag{32}$$

The solution to Eq. (5) satisfies

$$(G + \lambda\rho R_Q)\widehat{w} = b. \tag{33}$$

The implementation uses double precision and a pseudoinverse when the linear system is singular or its numerical solution is unreliable. Neither the window-mean correction nor the boundary correction enters the objective at this stage. For fixed P, Q, λ, fitting the cycle coefficients is independent of the forecast horizon.

Mean correction and the boundary coefficient. Both the mean and the boundary features use only the selected $H = QP$ steps of history. Let $\mu_X = H^{-1} \sum_{h=1}^{H} x_h$, and write the rth base predicted block as $\widetilde{Z}_r = \sum_{i=0}^{Q-1} \kappa_{r,i} Z_{-i}$, where $\kappa_{r,i}$ are the weights obtained by unrolling the recurrence. With mean correction enabled, the predicted block becomes

$$\widetilde{Z}_r + \left(1 - \sum_{i=0}^{Q-1} \kappa_{r,i}\right) \mu_X = \mu_X + \sum_{i=0}^{Q-1} \kappa_{r,i}(Z_{-i} - \mu_X). \tag{34}$$

This is equivalent to subtracting the input mean, applying the same recurrence, and adding the mean back; it neither changes channel scales nor adds coefficients. If the fitted weights satisfy $\sum_i w_i = 1$ exactly, the unrolled weights sum to 1 at every step and this correction is zero. If the spectral radius of the recurrence is less than 1, the base prediction tends to zero, whereas the prediction with mean correction tends to μ_X.

Let $f_0(X)$ denote the prediction including the optional mean correction but excluding the boundary term. Denote the part of Eq. (4) that does not contain g by the boundary feature $\Phi(X)_h$. With $\widehat{w}$ fixed, fit the scalar gain over the set $\mathcal{W}$ of complete training windows:

$$\widehat{g} = \mathrm{clip}_{[0,1]} \frac{\sum_{(X,Y)\in\mathcal{W}} \langle Y - f_0(X), \Phi(X) \rangle_F}{\sum_{(X,Y)\in\mathcal{W}} \|\Phi(X)\|_F^2} \tag{35}$$

Here, $\mathrm{clip}_{[0,1]}$ clips the result to $[0, 1]$; if the denominator does not exceed 10^{-30}, we set $\widehat{g} = 0$. This stage directly uses the multistep residuals of the standardized windows, without further weighting by s_j. The final output is $f_0(X) + \widehat{g}\Phi(X)$, and every model stores g, even if its fitted value is zero. The boundary term is not fed back into subsequent blocks, and

$$\|\widehat{g}\Phi(X)_h\|_2 \le \|d_X\|_2 \exp\left(-\frac{h}{\tau P}\right) \longrightarrow 0, \tag{36}$$

It therefore supplements near-term offsets without changing the limit determined by the base prediction and mean correction.

B.3 Computation and Acceleration

Let $J = \lceil L/P \rceil$. Direct blockwise recurrence costs $O(JQPc)$ operations. The implementation first unrolls the recurrence into a kernel $\kappa \in \mathbb{R}^{J\times Q}$ acting on historical blocks, and then generates the outputs for all channels simultaneously. Constructing this kernel recursively from Q basis vectors costs $O(JQ^2)$ operations. Applying it to a single window costs $O(JQPc)$, with an additional $O(Hc + Lc)$ for the mean and boundary corrections. All windows in the same batch share this kernel.

With $M = |\mathcal{T}|$ training time points in the first stage, accumulating the $Q \times Q$ matrix G costs $O(McQ^2)$ operations. Constructing the sharing penalty costs $O(Q^2)$, and solving the regularized linear system costs $O(Q^3)$. The second stage generates predictions for $|\mathcal{W}|$ complete training windows, then accumulates the two inner products in Eq. (35) and solves for one scalar.

For a batch size of B_{batch}, the working memory for the prediction kernel and batched prediction is $O(Q^2 + JQ + B_{\mathrm{batch}}(H + L)c)$.

B.4 Training Sampling and Shared Configurations

The original shared configurations of TinyCycle use the search space in Table 2, with one hyperparameter configuration selected jointly across the four forecast horizons for each dataset. The original number of historical cycles satisfies $1 \le Q \le \min(29, \lfloor 720/P \rfloor)$.

Training sampling. The training sequence is indexed from zero and has length T_{train}. For a fixed P, let $r_P = P\lfloor 720/P \rfloor$. The candidate time points for fitting the cycle coefficients are $t = r_P, \ldots, T_{\mathrm{train}} - 1$; the candidate origins for fitting the boundary gain are $t = r_P, \ldots, T_{\mathrm{train}} - L$, with each window using history $[t - QP, t)$ and future $[t, t + L)$. Each set is sampled at evenly spaced positions, rounded to the nearest integer, and deduplicated. When there are more than 4096 positions, at most 4096 are retained; otherwise, all are retained. This gives the time-point set $\mathcal{T}$ and

Table 2: Search space and fitting settings for the original shared configurations of TinyCycle.

Setting	Values or rule
Forecast horizon L	$\{96, 192, 336, 720\}$
Period length P	$\{4, 8, 24, 96\}$
Sharing strength λ	$0, 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000, 3000, 10000$
Age weights	$a_i = ((i+1)/Q)^2$, indexed from recent to distant
Window-mean correction η	0 or 1
Boundary decay scale τ	$\{1, 3, 7, 14, 28\}$ cycles
Boundary gain g	Fitted on the training set and constrained to $[0, 1]$
Weight sum and signs	No constraint on the weight sum or nonnegativity
Stored fitted coefficients	Q cycle coefficients and one boundary gain, totaling $Q + 1 \leq 30$

complete-window set $\mathcal{W}$, respectively, both retaining all channels. Different values of Q for the same period length share candidate training targets, and the target values in both stages lie within the training interval.

Standardization parameters are estimated only from the training interval, and the validation set is used only for hyperparameter selection. After selection, fitting still uses only the training set; validation data are not incorporated into training.

C Additional Material for the Preliminary Experiments

C.1 Statistical Protocol and Results for Periodic Association

Degree of periodic repetition. For an output $Z \in \{Y, \widehat{Y}\}$, define

$$C^Z_{i,h} = \frac{1}{|\mathcal{C}|} \sum_{j \in \mathcal{C}} |\mathrm{Corr}_o(X_{o,i,j}, Z_{o,h,j})| \,, \qquad \widetilde{C}^Z_{i,h} = C^Z_{i,h} - \frac{1}{L} \sum_{s=1}^{L} C^Z_{i,s}, \tag{37}$$

Here, o is the forecast origin, i is the historical position, h is the output step, and $\mathcal{C}$ is the set of channels for which the correlation is defined. The absolute value is taken within each channel before averaging across channels, preventing cancellation between opposite directions. For each input position, apply the same phase projection along the output steps to compute

$$\eta_P(Z) = \frac{\|\widetilde{C}^Z \Pi_P\|_F^2}{\|\widetilde{C}^Z\|_F^2}. \tag{38}$$

Here, Π_P acts on the output axis of the correlation matrix. The row means have been removed, so a constant background does not contribute to the numerator.

The most recent 336 historical positions are used as a common reference, and the daily period P is specified by the sampling interval. The proportion is computed for each model before equally weighted aggregation over models and datasets; repeated references to the same true future are not counted as additional samples. The signed sensitivity analysis retains the sign of Pearson correlation before averaging across channels to examine the effect of rectifying alternating positive and negative correlations by taking absolute values. For the sensitivity analysis with window-mean removal, the means of the model's actual input window and the 720-step output are subtracted separately for each origin and channel, after which the most recent 336 historical positions are retained to recompute the correlation matrix. For the sensitivity analysis removing linear variation from correlation maps, a line is fitted to and subtracted from each row along the output steps before computing Eq. (38). This examines the effect of slowly varying backgrounds in the maps; it does not amount to refitting models on detrended original time series. Under this protocol, the mean proportions for the models and the true future are 89.54% and 84.08%, respectively.

For the original absolute-correlation maps, the mean proportions for the models and the true future are 85.93% and 67.79%, respectively; the model proportion exceeds one half in 73/77 settings. After demeaning using the actual input window, the mean proportions for the models and the paired true future are 78.50% and 79.76%, respectively; when the history is also demeaned over the common 336 steps, they are 79.01% and 80.83%. Under the signed protocol, the mean proportions for the models

Table 3: Fraction of periodic recurrence in history–output association maps over all valid test forecast origins (%).

Dataset	P	Raw absolute correlation True future	Raw absolute correlation Model	Model Signed	Model Demeaned	Model Periodic amplitude
ETTh1	24	89.60	96.87	98.64	94.16	0.131
ETTh2	24	29.10	74.38	73.83	65.53	0.062
ETTm1	96	93.53	90.15	90.48	86.75	0.148
ETTm2	96	34.43	67.00	66.95	51.46	0.060
Weather	144	34.35	78.58	84.92	57.41	0.070
Electricity	24	99.02	99.28	99.74	99.06	0.203
Traffic	24	94.52	95.23	98.91	95.09	0.177

and the true future in the original windows are 87.64% and 70.04%, respectively; after window-mean removal using the actual input, they are 88.00% and 88.20%.

C.2 Visual Comparison with Moving-Average Decomposition

Figure 9 uses the same true analytical trajectory as Figure 3, comprising five cycles with 96 points per cycle, and decomposes it into Structure and Dynamics using Eq. (1). The moving-average (MA) decomposition extracts the trend with a centered 25-point average, repeating each endpoint 12 times for padding; the seasonal component is the original trajectory minus the trend (Wu et al., 2021). In both decompositions, adding the two components recovers the original trajectory.

Structure retains recurring spikes, while level changes across cycles enter the Dynamic component. Because the 25-point MA window is shorter than the 96-point period, the trend still retains broad periodic fluctuations, while spikes appear mainly in the seasonal component. The two decompositions partition the trajectory according to sharing across cycles and the local smoothing scale, respectively.

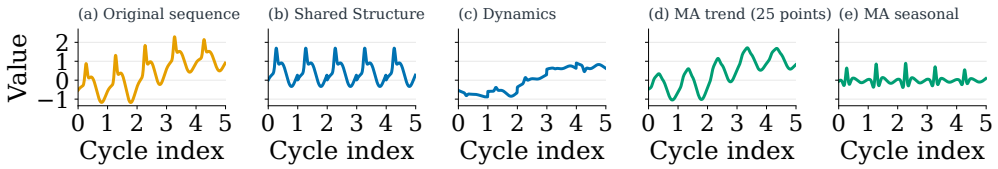


Figure 9: Comparison of the Structure–Dynamics and MA decompositions of the same analytical trajectory. From left to right, (a) is the true trajectory in Figure 3; (b–c) show Structure and Dynamics; and (d–e) show the trend and seasonal component obtained with a 25-point moving average. All panels use the same axis scales and display five consecutive cycles.

D Complete Experimental Results

Efficiency by forecast horizon. Table 5 expands the results in Table 1 into eight datasets and four forecast horizons, using the same model configurations and four-core CPU timing. Params counts the stored real-valued scalars, including the boundary gain for TinyCycle and counting each complex-valued parameter in MixLinear as two real-valued scalars. Train measures the time for fitting the selected configuration or completing all training epochs, excluding data loading, standardization, initialization, validation, and search. Test includes forecasting at all valid test origins and channels, error aggregation, and compressed storage of the results for each origin, as well as construction of the prediction cache for TinyCycle; one-time library compilation and loading are measured separately. TinyCycle's Test time is the median of three measurements for each task, whereas the other Test times and all Train times retain a single complete measurement. Average is the equally weighted average over the eight datasets at the same forecast horizon, and Overall is the equally weighted average over all 32 tasks; averaging the four horizons equally for each dataset gives the corresponding row in Table 1.

Table 4: Per-task MSE and MAE. Red and blue indicate the best and second-best values within TinyCycle, MixLinear, SparseTSF, and TimeBase, respectively. Underlining marks the top three ranks among all thirteen models.

Dataset	H	TinyCycle		MixLinear		SparseTSF		TimeBase		PatchTST		iTransformer		TimeMixer		DLinear		GTR		PhaseFormer		Chronos-2		Moirai-2		Toto2-1B	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.353	0.393	0.352	0.383	0.362	0.389	0.367	0.383	0.379	0.410	0.395	0.426	0.408	0.429	0.395	0.419	0.381	0.409	0.360	0.386	0.362	0.370	0.376	0.376	0.374	0.370
	192	0.386	0.416	0.397	0.425	0.402	0.411	0.399	0.406	0.417	0.435	0.429	0.449	0.435	0.444	0.439	0.450	0.420	0.434	0.406	0.411	0.407	0.401	0.426	0.406	0.420	0.400
	336	0.406	0.433	0.412	0.434	0.432	0.427	0.441	0.433	0.444	0.455	0.493	0.493	0.464	0.462	0.467	0.470	0.445	0.449	0.443	0.436	0.436	0.421	0.471	0.425	0.448	0.417
	720	0.431	0.468	0.423	0.456	0.426	0.446	0.461	0.465	0.468	0.484	0.674	0.598	0.463	0.477	0.512	0.524	0.478	0.480	0.421	0.443	0.451	0.449	0.509	0.446	0.434	0.430
	Avg.	0.394	0.427	0.396	0.425	0.406	0.418	0.417	0.422	0.427	0.446	0.498	0.492	0.443	0.453	0.453	0.466	0.431	0.443	0.407	0.419	0.414	0.410	0.446	0.413	0.419	0.404
ETTh2	96	0.283	0.336	0.286	0.341	0.284	0.342	0.300	0.354	0.278	0.341	0.306	0.362	0.310	0.374	0.323	0.382	0.278	0.346	0.283	0.344	0.290	0.322	0.295	0.327	0.288	0.318
	192	0.342	0.375	0.341	0.378	0.339	0.377	0.349	0.389	0.345	0.383	0.403	0.419	0.353	0.398	0.408	0.432	0.370	0.401	0.344	0.383	0.364	0.372	0.365	0.372	0.365	0.368
	336	0.369	0.400	0.358	0.396	0.360	0.398	0.372	0.412	0.370	0.407	0.460	0.458	0.381	0.418	0.495	0.488	0.385	0.421	0.367	0.405	0.379	0.395	0.403	0.404	0.400	0.398
	720	0.391	0.427	0.381	0.424	0.379	0.423	0.402	0.448	0.403	0.440	0.433	0.463	0.433	0.456	0.833	0.646	0.411	0.446	0.419	0.450	0.396	0.423	0.427	0.430	0.407	0.418
	Avg.	0.346	0.385	0.342	0.385	0.340	0.385	0.356	0.401	0.349	0.393	0.401	0.426	0.369	0.412	0.515	0.487	0.361	0.403	0.353	0.395	0.357	0.378	0.372	0.384	0.365	0.376
ETTm1	96	0.307	0.351	0.325	0.358	0.366	0.393	0.321	0.353	0.297	0.352	0.319	0.371	0.326	0.369	0.315	0.359	0.293	0.349	0.298	0.344	0.304	0.319	0.331	0.330	0.303	0.313
	192	0.334	0.366	0.381	0.400	0.377	0.397	0.339	0.364	0.337	0.376	0.349	0.389	0.358	0.389	0.347	0.381	0.336	0.371	0.331	0.364	0.367	0.357	0.393	0.366	0.374	0.351
	336	0.362	0.383	0.397	0.407	0.398	0.408	0.368	0.381	0.366	0.396	0.382	0.410	0.383	0.406	0.373	0.394	0.369	0.391	0.359	0.381	0.404	0.383	0.443	0.396	0.424	0.379
	720	0.415	0.415	0.413	0.410	0.439	0.427	0.422	0.410	0.422	0.422	0.448	0.448	0.474	0.453	0.424	0.425	0.435	0.428	0.417	0.413	0.480	0.429	0.524	0.438	0.494	0.420
	Avg.	0.355	0.379	0.379	0.394	0.395	0.406	0.363	0.377	0.356	0.387	0.375	0.405	0.385	0.404	0.365	0.390	0.358	0.385	0.351	0.376	0.389	0.372	0.423	0.383	0.399	0.366
ETTm2	96	0.172	0.258	0.200	0.278	0.211	0.292	0.224	0.292	0.165	0.256	0.182	0.276	0.169	0.261	0.178	0.273	0.170	0.260	0.170	0.260	0.169	0.235	0.188	0.251	0.169	0.233
	192	0.224	0.294	0.251	0.317	0.252	0.318	0.262	0.319	0.217	0.294	0.249	0.320	0.231	0.304	0.282	0.359	0.232	0.303	0.233	0.301	0.243	0.287	0.280	0.309	0.239	0.281
	336	0.274	0.327	0.294	0.345	0.295	0.346	0.304	0.348	0.270	0.329	0.303	0.355	0.283	0.338	0.296	0.359	0.286	0.338	0.281	0.334	0.309	0.332	0.393	0.369	0.302	0.324
	720	0.362	0.381	0.366	0.385	0.365	0.389	0.375	0.396	0.352	0.381	0.376	0.402	0.358	0.386	0.414	0.433	0.360	0.392	0.353	0.381	0.393	0.388	0.507	0.435	0.389	0.380
	Avg.	0.258	0.315	0.278	0.331	0.281	0.336	0.291	0.338	0.251	0.315	0.278	0.338	0.260	0.322	0.293	0.356	0.262	0.323	0.259	0.319	0.278	0.310	0.342	0.341	0.275	0.305
Weather	96	0.181	0.241	0.199	0.260	0.179	0.235	0.158	0.216	0.149	0.199	0.159	0.212	0.163	0.223	0.174	0.242	0.146	0.200	0.148	0.194	0.154	0.175	0.173	0.202	0.151	0.180
	192	0.222	0.277	0.231	0.284	0.220	0.267	0.199	0.254	0.193	0.243	0.203	0.252	0.201	0.255	0.215	0.277	0.198	0.248	0.193	0.237	0.215	0.229	0.220	0.247	0.198	0.228
	336	0.266	0.313	0.282	0.320	0.264	0.301	0.248	0.293	0.240	0.281	0.253	0.291	0.258	0.300	0.262	0.319	0.244	0.285	0.246	0.281	0.290	0.278	0.285	0.293	0.254	0.272
	720	0.323	0.361	0.330	0.347	0.324	0.342	0.315	0.338	0.312	0.334	0.317	0.337	0.329	0.348	0.319	0.359	0.323	0.339	0.317	0.333	0.414	0.332			0.328	0.324
	Avg.	0.248	0.298	0.261	0.303	0.247	0.286	0.230	0.275	0.224	0.264	0.233	0.273	0.238	0.282	0.243	0.299	0.228	0.268	0.226	0.261	0.268	0.254			0.233	0.251
Electricity	96	0.137	0.233	0.139	0.234	0.139	0.234	0.140	0.234	0.130	0.224	0.134	0.229	0.135	0.237	0.133	0.230	0.127	0.222	0.129	0.221	0.126	0.210	0.126	0.212	0.125	0.209
	192	0.151	0.245	0.154	0.249	0.151	0.245	0.155	0.247	0.147	0.241	0.153	0.247	0.156	0.256	0.147	0.244	0.145	0.240	0.146	0.236	0.145	0.228	0.144	0.230	0.141	0.225
	336	0.167	0.262	0.171	0.265	0.166	0.261	0.170	0.262	0.162	0.257	0.167	0.263	0.170	0.271	0.162	0.261	0.161	0.258	0.165	0.257	0.165	0.248			0.159	0.244
	720	0.205	0.298	0.210	0.299	0.205	0.295	0.211	0.296	0.197	0.288	0.188	0.284	0.215	0.311	0.198	0.292	0.195	0.291	0.200	0.288	0.208	0.285			0.198	0.279
	Avg.	0.165	0.260	0.168	0.262	0.165	0.259	0.169	0.260	0.159	0.253	0.161	0.256	0.169	0.269	0.160	0.257	0.157	0.253	0.160	0.251	0.161	0.243			0.156	0.239
Traffic	96	0.391	0.269	0.391	0.270	0.389	0.266	0.399	0.275	0.369	0.255	0.345	0.250	0.462	0.285	0.422	0.327	0.370	0.255	0.366	0.235	0.347	0.216	0.348	0.221	0.347	0.215
	192	0.402	0.273	0.404	0.276	0.398	0.270	0.407	0.276	0.383	0.261	0.360	0.259	0.473	0.296	0.434	0.332	0.397	0.282	0.383	0.242	0.374	0.229	0.370	0.232	0.374	0.227
	336	0.416	0.280	0.416	0.282	0.411	0.276	0.418	0.281	0.397	0.268	0.369	0.266	0.498	0.296	0.449	0.338	0.409	0.286	0.394	0.250	0.396	0.239			0.393	0.237
	720	0.455	0.302	0.452	0.301	0.447	0.296	0.455	0.299	0.434	0.289	0.394	0.280	0.506	0.313	0.491	0.359	0.455	0.312	0.435	0.272	0.440	0.264			0.436	0.261
	Avg.	0.416	0.281	0.416	0.282	0.411	0.277	0.420	0.283	0.396	0.268	0.367	0.264	0.485	0.298	0.449	0.339	0.408	0.284	0.395	0.250	0.390	0.237			0.387	0.235
Exchange	96	0.086	0.205	0.099	0.225	0.139	0.273	0.134	0.265	0.083	0.211	0.096	0.244	0.092	0.219	0.096	0.243	0.085	0.208	0.095	0.221	0.084	0.199	0.089	0.209	0.087	0.204
	192	0.180	0.303	0.191	0.316	0.246	0.364	0.244	0.353	0.168	0.307	0.169	0.325	0.205	0.326	0.169	0.310	0.200	0.327	0.246	0.365	0.176	0.295	0.190	0.307	0.181	0.302
	336	0.334	0.421	0.343	0.426	0.459	0.505	0.456	0.486	0.362	0.413	0.312	0.399	0.370	0.442	0.344	0.438	0.395	0.449	0.427	0.482	0.329	0.411	0.339	0.421	0.321	0.410
	720	0.938	0.724	1.060	0.788	0.991	0.768	1.497	0.926	0.659	0.670	0.619	0.635	0.996	0.746	0.637	0.648	1.158	0.805	1.178	0.823	0.919	0.719	0.828	0.678	0.845	0.680
	Avg.	0.385	0.413	0.423	0.439	0.458	0.478	0.583	0.508	0.318	0.400	0.299	0.401	0.416	0.433	0.312	0.410	0.460	0.447	0.486	0.473	0.377	0.406	0.362	0.404	0.358	0.399
Average		0.321	0.345	0.333	0.352	0.338	0.356	0.354	0.358	0.310	0.341	0.326	0.357	0.346	0.359	0.349	0.375	0.333	0.351	0.330	0.343	0.329	0.326			0.324	0.322
w/o Exchange		0.312	0.335	0.320	0.340	0.321	0.338	0.321	0.337	0.309	0.332	0.330	0.350	0.336	0.348	0.354	0.371	0.315	0.337	0.307	0.324	0.322	0.315			0.319	0.311

D.1 Model Performance under Hyperparameter Search

Table 6 reports the parameter search results. For TinyCycle, we fix $Q = \lfloor 720/P \rfloor$ for each candidate period $P \in \{4, 8, 24, 96\}$, using the longest lookback window of complete cycles that does not exceed 720 steps. The candidate ranges for the sharing strength, mean-correction switch, and boundary-decay scale follow Appendix B.4, giving 640 candidate shared configurations per dataset. The validation MSE for each forecast horizon is divided by the minimum validation MSE for that horizon in the new candidate space, and the normalized values are averaged equally over the four horizons to select a single set of hyperparameters. Standardization, cycle coefficients, and the boundary gain for each horizon are fitted using only the training set; the validation set is not merged into the training set for refitting after hyperparameter selection. For $P = 4, 8$, the first training sample origin is moved to 720 to accommodate the extended lookback window; all other fitting rules remain unchanged. The selected configurations are $P = 24, Q = 30$ for ETTh1, $P = 8, Q = 90$ for Weather, and $P = 4, Q = 180$ for the other six datasets; the actual input length is 720 steps for all eight datasets.

The three baselines retain the periods and training hyperparameters from the original code experiments, without matching TinyCycle's period, and all use 720-step inputs. SparseTSF uses its linear version and, like TimeBase, retains fixed configurations; for MixLinear, the original candidates are selected by validation MSE on ETTm2, Weather, and Exchange, while fixed configurations are retained on the other datasets. Exchange retains the extended configurations specified in the original experiment, and all three baselines select trained weights by validation MSE. Parameter counts are measured in real-valued scalars: TinyCycle has Q cycle coefficients and one boundary gain, each complex-valued parameter in MixLinear counts as two real-valued scalars, and structural hyperparameters are excluded.

D.2 Hyperparameter Sensitivity

We examine the distribution of test errors over the hyperparameter search spaces on ETTh1, ETTh2, ETTm1, ETTm2, and Weather, with forecast horizon $L = 720$. TinyCycle uses the 640 configurations per dataset defined in Appendix D.1: $P \in \{4, 8, 24, 96\}$, $Q = \lfloor 720/P \rfloor$, 16 sharing strengths,

Table 5: Efficiency by forecast horizon for Table 1. TinyCycle uses the extended lookback with its original hyperparameters fixed. Params counts real-valued parameters; Train and Test are in seconds per task. Red and blue indicate the lowest and second-lowest values, respectively, for each metric in each row, ranked using unrounded values. Ties receive the same color.

Dataset	H	TinyCycle			MixLinear			SparseTSF			TimeBase		
		Params	Train	Test	Params	Train	Test	Params	Train	Test	Params	Train	Test
ETTh1	96	31.000	0.035	0.003	69.000	9.323	0.068	145.000	4.274	0.049	214.000	2.147	0.015
	192	31.000	0.032	0.004	113.000	8.399	0.073	265.000	3.525	0.050	242.000	1.358	0.017
	336	31.000	0.041	0.004	137.000	12.829	0.071	445.000	3.881	0.053	284.000	1.393	0.020
	720	31.000	0.089	0.006	293.000	12.316	0.088	925.000	4.074	0.054	396.000	1.232	0.028
ETTh2	96	31.000	0.034	0.003	141.000	3.970	0.065	145.000	3.224	0.049	214.000	1.239	0.015
	192	31.000	0.034	0.004	169.000	5.248	0.073	265.000	4.784	0.050	164.000	1.627	0.017
	336	31.000	0.035	0.004	189.000	6.342	0.077	445.000	3.113	0.051	284.000	0.907	0.020
	720	31.000	0.068	0.006	293.000	9.465	0.088	925.000	7.262	0.054	396.000	1.061	0.028
ETTm1	96	8.000	0.021	0.010	139.000	48.730	0.555	105.000	22.818	0.502	181.000	15.816	0.124
	192	8.000	0.022	0.014	149.000	67.546	0.662	113.000	59.085	0.507	222.000	12.537	0.164
	336	8.000	0.025	0.015	157.000	67.665	0.625	129.000	35.914	0.521	264.000	20.608	0.174
	720	8.000	0.082	0.023	179.000	59.348	0.747	161.000	31.275	0.547	348.000	22.239	0.205
ETTm2	96	8.000	0.021	0.010	111.000	32.824	0.520	105.000	22.811	0.502	201.000	10.864	0.137
	192	8.000	0.025	0.012	121.000	29.631	0.636	113.000	56.537	0.521	222.000	24.356	0.152
	336	8.000	0.033	0.015	129.000	58.615	0.607	129.000	33.466	0.530	264.000	21.959	0.209
	720	8.000	0.046	0.023	151.000	48.743	0.703	161.000	60.488	0.555	348.000	33.252	0.231
Weather	96	31.000	0.037	0.015	69.000	96.714	0.752	145.000	49.838	0.568	4494.000	57.298	0.239
	192	31.000	0.062	0.019	97.000	97.794	0.852	265.000	45.832	0.608	5082.000	56.134	0.261
	336	31.000	0.076	0.027	133.000	214.793	0.958	445.000	41.415	0.656	5964.000	39.377	0.328
	720	31.000	0.202	0.052	221.000	210.597	1.336	925.000	77.908	0.857	8316.000	80.803	0.534
Electricity	96	31.000	0.576	0.058	141.000	2292.974	10.543	145.000	344.700	7.032	214.000	429.415	2.892
	192	31.000	0.802	0.102	153.000	1712.410	9.960	265.000	620.412	7.619	242.000	491.471	2.969
	336	31.000	0.944	0.157	189.000	2319.988	10.691	445.000	724.993	9.267	284.000	884.476	5.224
	720	31.000	1.491	0.296	277.000	4096.670	20.838	925.000	854.657	11.684	396.000	686.853	8.302
Traffic	96	31.000	1.846	0.097	141.000	2316.290	16.723	145.000	1586.338	13.280	214.000	312.620	4.751
	192	31.000	2.395	0.170	153.000	4444.128	16.276	265.000	996.346	13.678	320.000	544.448	4.988
	336	31.000	2.597	0.309	205.000	3333.356	27.664	445.000	1323.529	17.069	689.000	1223.196	10.234
	720	31.000	5.022	0.553	293.000	3501.886	34.286	925.000	1723.653	20.692	518.000	1713.316	13.841
Exchange	96	31.000	0.026	0.002	69.000	8.466	0.037	145.000	3.628	0.028	1712.000	1.036	0.012
	192	31.000	0.029	0.002	97.000	5.439	0.038	265.000	3.165	0.028	1936.000	2.009	0.012
	336	31.000	0.033	0.002	133.000	4.169	0.037	445.000	1.121	0.027	2272.000	1.469	0.013
	720	31.000	0.049	0.003	221.000	4.487	0.037	925.000	1.226	0.022	3168.000	1.035	0.014
Average	96	25.250	0.324	0.025	110.000	601.161	3.658	135.000	254.704	2.751	930.500	103.804	1.023
	192	25.250	0.425	0.041	131.500	796.324	3.571	227.000	223.711	2.883	1053.750	141.743	1.073
	336	25.250	0.473	0.067	159.000	752.220	5.091	366.000	270.929	3.522	1288.125	274.173	2.028
	720	25.250	0.881	0.120	241.000	992.939	7.265	734.000	345.068	4.308	1735.750	317.474	2.898
Overall	All	25.250	0.526	0.063	160.375	785.661	4.896	365.500	273.603	3.366	1252.031	209.299	1.755

two mean-correction settings, and five boundary-decay scales. For SparseTSF, TimeBase, and MixLinear, we vary the period over $\{2, 4, 6, 12, 24, 48, 96, 144\}$ and the nominal learning rate over $\{0.001, 0.01, 0.1\}$, giving 24 configurations per model and dataset. The baselines use 720-step inputs and retain their other training settings with seed 2023. TinyCycle uses 672 steps for $P = 96$ and 720 otherwise. For $P = 96$, the baselines copy 48 phase-aligned values from the input history to complete the cycle blocks and crop the forecast to 720 steps. MixLinear's low-frequency cutoff is capped by the available number of historical blocks. Cycle coefficients and boundary gains are fitted on training data, while the baselines select checkpoints by validation MSE. Every configuration is evaluated over all valid test origins and channels and contributes equally to its model's distribution. These distributions describe sensitivity within each model's specified search space, with different candidate grids and search budgets.

Figure 10 shows that TinyCycle is sensitive to hyperparameter selection. Its test MSE interquartile ranges are $[0.394, 0.816]$ on ETTh2 and $[0.420, 0.637]$ on ETTm1, compared with $[0.382, 0.393]$ and $[0.419, 0.428]$ for SparseTSF. SparseTSF has the lowest median MSE on all five datasets in these grids. TinyCycle relies on modeling stable, cycle-aligned Structure, and its period, sharing strength, mean correction, and boundary decay affect how that Structure is extrapolated. The observed spread makes validation-based hyperparameter selection important for its long-horizon accuracy.

Table 6: TinyCycle with hyperparameters reselected on validation data after extending the lookback, compared with baselines in their original configurations. L is the forecast horizon and N_θ is the number of real-valued parameters. Avg. is the equally weighted mean over four horizons for each dataset; Average is the equally weighted mean over all 32 tasks. Boldface marks the lowest value for each metric in each row, based on unrounded values.

Dataset	L	TinyCycle			SparseTSF			TimeBase			MixLinear		
		MSE	MAE	N_θ	MSE	MAE	N_θ	MSE	MAE	N_θ	MSE	MAE	N_θ
ETTh1	96	0.353	0.393	**31**	0.361	0.388	145	0.361	0.384	214	**0.352**	**0.383**	69
	192	**0.386**	0.416	**31**	0.404	**0.412**	265	0.410	0.415	242	0.396	0.423	113
	336	0.406	0.433	**31**	0.434	0.428	445	**0.406**	**0.416**	284	0.411	0.434	137
	720	0.431	0.468	**31**	0.426	**0.447**	925	0.462	0.461	396	**0.425**	0.458	293
	Avg.	**0.394**	0.427	**31**	0.406	**0.419**	445	0.410	0.419	284	0.396	0.425	153
ETTh2	96	**0.272**	**0.335**	181	0.293	0.345	145	0.295	0.352	214	0.282	0.340	**141**
	192	**0.331**	**0.373**	181	0.340	0.378	265	0.347	0.388	**164**	0.341	0.377	169
	336	**0.355**	**0.394**	**181**	0.356	0.397	445	0.359	0.410	284	0.357	0.396	189
	720	0.382	**0.423**	**181**	**0.381**	0.423	925	0.400	0.449	396	0.387	0.427	293
	Avg.	**0.335**	**0.382**	**181**	0.342	0.386	445	0.350	0.400	264.5	0.342	0.385	198
ETTm1	96	**0.310**	**0.354**	**181**	0.314	0.354	4325	0.318	0.360	3714	0.338	0.378	1037
	192	**0.338**	0.371	**181**	0.339	**0.368**	8645	0.360	0.389	4628	0.360	0.388	1343
	336	**0.367**	0.388	**181**	0.367	**0.386**	15125	0.374	0.396	5384	0.390	0.407	1745
	720	0.419	0.420	**181**	**0.417**	**0.412**	32405	0.429	0.430	4940	0.437	0.434	2741
	Avg.	**0.358**	0.383	**181**	0.359	**0.380**	15125	0.370	0.394	4666.5	0.381	0.402	1716.5
ETTm2	96	0.167	0.257	**181**	**0.164**	**0.254**	4325	0.166	0.257	4124	0.179	0.271	245
	192	0.222	0.294	**181**	**0.217**	**0.291**	8645	0.219	0.293	4628	0.232	0.306	397
	336	0.274	0.329	**181**	**0.269**	**0.326**	15125	0.271	0.328	5384	0.284	0.340	625
	720	0.359	0.382	**181**	**0.351**	**0.378**	32405	0.353	0.382	4940	0.366	0.390	1121
	Avg.	0.255	0.315	**181**	**0.250**	**0.312**	15125	0.252	0.315	4769	0.265	0.327	597
Weather	96	0.171	0.229	**91**	0.170	0.224	4325	**0.145**	**0.200**	26334	0.179	0.237	245
	192	0.213	0.267	**91**	0.214	0.261	8645	**0.186**	**0.241**	29862	0.222	0.273	397
	336	0.258	0.305	**91**	0.258	0.294	15125	**0.237**	**0.282**	35154	0.269	0.309	625
	720	0.318	0.355	**91**	0.321	0.340	32405	**0.308**	**0.334**	49266	0.330	0.351	1121
	Avg.	0.240	0.289	**91**	0.241	0.280	15125	**0.219**	**0.264**	35154	0.250	0.293	597
Electricity	96	**0.135**	**0.232**	181	0.139	0.233	145	0.140	0.234	214	0.139	0.233	**141**
	192	**0.149**	0.246	181	0.151	**0.245**	265	0.155	0.247	242	0.154	0.248	**153**
	336	**0.164**	0.263	**181**	0.166	**0.261**	445	0.170	0.263	284	0.172	0.267	189
	720	**0.200**	0.299	**181**	0.205	**0.294**	925	0.215	0.300	396	0.210	0.300	277
	Avg.	**0.162**	0.260	**181**	0.165	**0.258**	445	0.170	0.261	284	0.169	0.262	190
Traffic	96	0.391	0.270	181	**0.389**	**0.267**	145	0.396	0.269	214	0.392	0.269	**141**
	192	0.402	0.274	181	**0.399**	**0.271**	265	0.404	0.274	320	0.404	0.276	**153**
	336	0.416	0.281	**181**	**0.411**	**0.275**	445	0.418	0.281	689	0.416	0.283	205
	720	0.456	0.305	**181**	**0.447**	**0.296**	925	0.460	0.305	518	0.451	0.301	293
	Avg.	0.416	0.282	**181**	**0.412**	**0.277**	445	0.420	0.282	435.25	0.416	0.282	198
Exchange	96	0.085	0.203	**181**	0.111	0.241	4325	0.102	0.227	10032	**0.084**	**0.203**	245
	192	**0.176**	**0.298**	**181**	0.208	0.330	8645	0.215	0.331	11376	0.177	0.301	397
	336	0.327	0.413	**181**	0.368	0.441	15125	0.432	0.468	13392	**0.321**	**0.407**	625
	720	**0.962**	**0.726**	**181**	1.100	0.807	32405	1.308	0.833	18768	1.052	0.788	1121
	Avg.	**0.387**	**0.410**	**181**	0.447	0.455	15125	0.514	0.465	13392	0.409	0.425	597
Average		**0.319**	**0.344**	**151**	0.328	0.346	7785	0.338	0.350	7406.156	0.328	0.350	530.812

Search cost. Closed-form fitting keeps this search inexpensive in absolute terms. On an ARM-based Mac using two CPU threads, the core computation for TinyCycle's 640-configuration search, jointly covering horizons 96, 192, 336, and 720, takes 5.16–9.64 seconds per dataset across these five datasets. This includes fitting, validation scoring, and result writing, but excludes data loading, standardization, final configuration selection, and test evaluation. The shared training statistics and closed-form coefficient and boundary-gain solutions reduce the cost of evaluating the grid. For the main-result configurations in Table 5, fitting TinyCycle at horizon 720 takes 0.046–0.202 seconds across these datasets, less than the mean time per training epoch of each baseline in that table on the same dataset and four-core CPU. Thus, TinyCycle requires careful hyperparameter selection, but the measured core search computation remains below ten seconds per dataset.

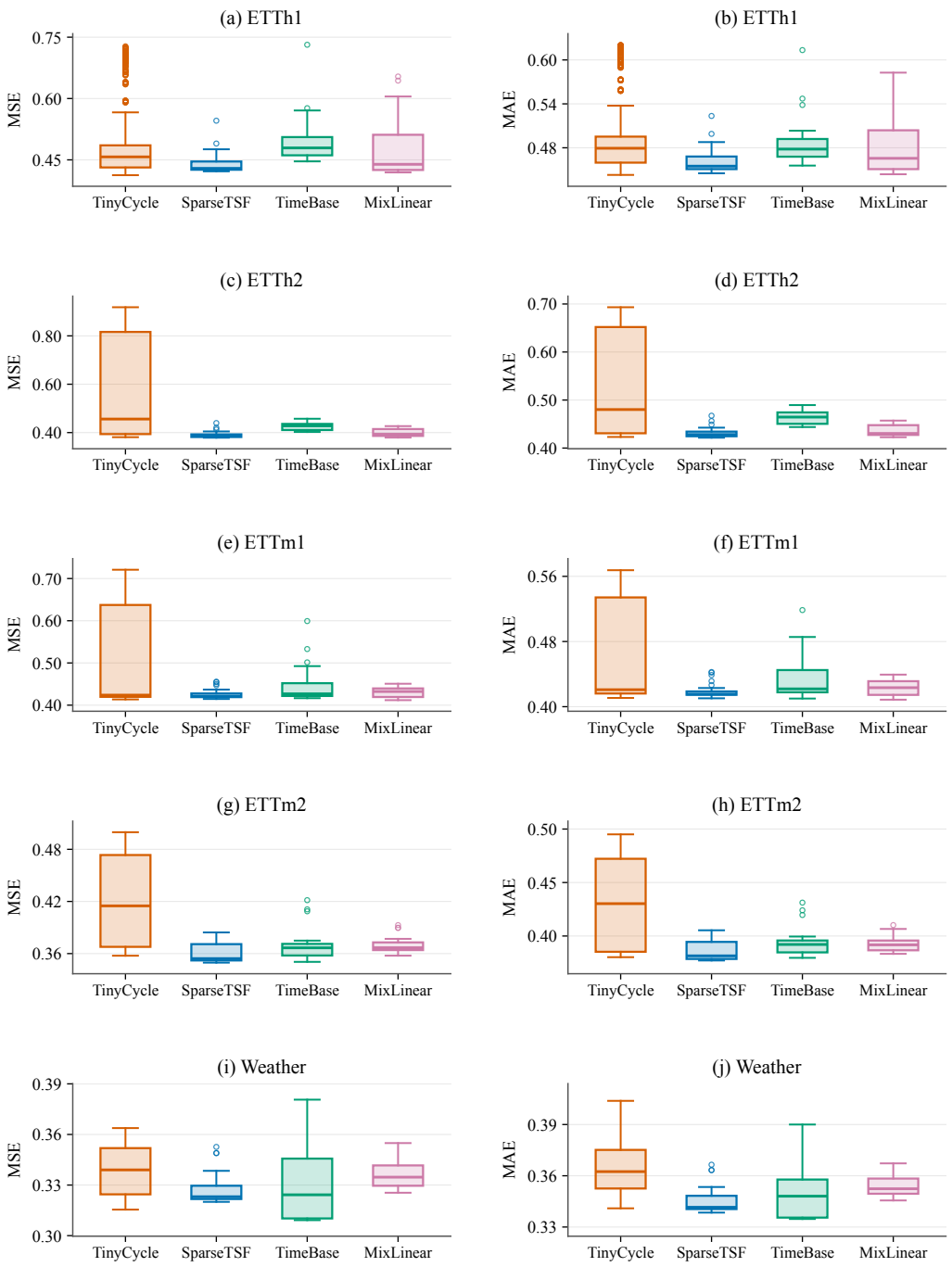


Figure 10: Hyperparameter sensitivity at forecast horizon 720. Each row corresponds to one dataset; the left and right columns report test MSE and MAE, respectively. TinyCycle contributes 640 configurations per dataset, and SparseTSF, TimeBase, and MixLinear contribute 24 each. Boxes show the median and interquartile range, whiskers extend to the most extreme observations within 1.5 interquartile ranges, and points show all remaining observations. Full error ranges are retained. The distributions represent the respective hyperparameter grids rather than repeated random seeds.

E SUPPLEMENTARY EXPERIMENTS

E.1 DIAGNOSTIC SETTINGS AND RECOVERY RESULTS FOR STRUCTURE AND DYNAMICS

Prefix decomposition and aggregation. The 192-, 336-, and 720-step prefixes are taken from the same H720 forecast, using the set of origins from the complete task, and phase means are computed independently within each prefix; incomplete cycles are averaged using the actual number of occurrences of each phase. These prefixes are not separate forecasting tasks with shorter horizons. Recovery scores are normalized within each setting according to Eq. (7) and then averaged equally over models and datasets; a score is undefined when its denominator is zero.

The actual MSE is decomposed orthogonally according to Eq. (2), and MAE is allocated symmetrically according to Eq. (16). The allocated values sum to the actual total MAE, unlike the two independently computed component MAEs. Error shares are computed by first averaging the component and total errors over models, then averaging equally over datasets, and finally taking their ratio; they are neither averages of per-setting shares nor shares of improvement between models.

The energy share of the predicted Structure is

$$q_S(H) = \frac{\sum_{o,h,j} \widehat{S}^2_{o,h,j}}{\sum_{o,h,j} \widehat{Y}^2_{o,h,j}}, \qquad q_R(H) = 1 - q_S(H), \tag{39}$$

where the sums are restricted to the current prefix. In the tables, component and total energies are first averaged over models before computing their ratio for each dataset; cross-dataset summaries likewise average the energies equally before taking their ratio. Each ground-truth trajectory is counted once per dataset, and predictions are not averaged beforehand to form an ensemble. Table 7 presents macro-averaged summaries, and Table 8 combines per-dataset recovery, energy, and actual errors. Dynamic columns are omitted only for complementary shares; recovery scores retain both components.

Table 7: Summary of eleven models at three prefix lengths (%). Recovery scores are reported separately for Structure and Dynamics. The remaining columns show Structure shares; the corresponding Dynamic shares equal 100 minus each value.

H	Structure recovery	Dynamic recovery	MSE share	MAE share	Predicted energy share
192	84.30	9.05	53.47	56.29	96.89
336	84.17	7.65	46.19	50.03	96.67
720	81.98	5.82	41.89	46.88	96.56

Profile recovery after removing the level. To evaluate the repeated profile separately, let

$$C_{o,h,j} = S_{o,h,j} - \mu_{o,j}, \qquad \widehat{C}_{o,h,j} = \widehat{S}_{o,h,j} - \widehat{\mu}_{o,j}, \tag{40}$$

where $\mu, \widehat{\mu}$ are the window means of the ground truth and prediction, respectively, within the current prefix, and the same recovery score is applied to C. The demeaned profile has nonzero energy on all seven datasets. Over the full 720 steps, the macro-averaged recovery scores for Structure, profile, and Dynamics are 81.98%, 66.54%, and 5.82%, respectively. In all 77/77 settings, Dynamic recovery is lower than both Structure and profile recovery, with a minimum gap of 0.013, or approximately 1.28 percentage points, relative to the latter. Table 9 retains all per-setting scores, including negative values.

E.2 SOURCES OF PERFORMANCE GAINS ACROSS MODELS

We separately examine the gains of lightweight models over the historical mean, the improvements of supervised models over lightweight references, and the direct component errors and cancellation terms of TSFMs; each analysis retains its own reference and normalization denominator.

E.2.1 GAINS OF LIGHTWEIGHT AND SUPERVISED MODELS

Lightweight models relative to the historical mean. TinyCycle uses its original frozen configuration before the lookback extension, and SparseTSF retains its matched-period configuration, with

Table 8: Recovery, energy, and errors of eleven models by dataset and prefix length. Recovery scores and energy shares are percentages. For energy and errors, only Structure shares are shown; Dynamic shares equal 100 minus the Structure shares. Profile recovery removes the level of each window. Appendix E.1 specifies the aggregation order.

		Recovery score			Structure energy%		MSE		MAE	
Dataset	H	Structure	Profile	Dynamics	Predicted	True	Total MSE	Structure%	Total MAE	Structure%
ETTh1	192	80.13	72.37	-4.36	97.83	78.10	0.409	40.45	0.421	44.31
ETTh1	336	82.17	75.72	-4.48	97.57	74.59	0.424	33.38	0.436	40.35
ETTh1	720	81.14	79.74	-5.83	96.99	70.96	0.484	30.34	0.482	40.29
ETTh2	192	94.84	51.72	-0.76	99.69	94.78	0.315	48.18	0.365	49.80
ETTh2	336	94.75	56.34	-0.82	99.61	93.47	0.355	42.73	0.396	45.96
ETTh2	720	93.71	59.77	-4.07	99.40	91.31	0.460	38.83	0.463	42.97
ETTm1	192	72.26	62.79	-7.02	98.60	90.71	0.391	71.69	0.396	66.24
ETTm1	336	71.49	63.27	-8.62	97.81	85.48	0.446	60.70	0.421	57.31
ETTm1	720	68.18	62.32	-6.95	97.56	79.35	0.524	53.35	0.457	51.38
ETTm2	192	94.13	32.25	-2.86	99.80	97.86	0.251	72.28	0.312	67.55
ETTm2	336	93.72	40.79	-3.07	99.61	96.29	0.312	61.24	0.350	58.88
ETTm2	720	92.55	47.95	-2.82	99.56	94.40	0.403	55.01	0.405	54.24
Weather	192	70.00	18.01	0.78	99.17	93.62	0.210	81.62	0.256	82.42
Weather	336	68.16	25.76	-0.82	98.51	84.76	0.263	63.72	0.295	65.22
Weather	720	62.53	33.26	-1.14	97.89	75.27	0.339	52.99	0.345	57.47
Electricity	192	91.71	94.80	34.33	94.95	87.07	0.156	45.95	0.253	47.79
Electricity	336	91.00	94.76	30.32	94.90	85.54	0.177	43.32	0.272	46.25
Electricity	720	88.45	94.07	25.00	94.91	83.43	0.222	43.67	0.307	46.56
Traffic	192	87.02	88.06	43.20	80.14	65.87	0.420	30.60	0.277	39.78
Traffic	336	87.94	89.02	41.00	79.94	63.88	0.437	26.55	0.287	36.91
Traffic	720	87.33	88.67	36.55	80.34	61.77	0.482	24.39	0.311	35.35

Table 9: Complete component recovery scores for eleven models at $H = 720$ (%). Profile denotes the demeaned profile. All negative values are retained; a zero score indicates the same error as predicting a zero component.

Model	ETTh1			ETTh2			ETTm1			ETTm2			Weather			Electricity			Traffic		
	Structure	Profile	Dynamics	Structure	Profile	Dynamics	Structure	Profile	Dynamics	Structure	Profile	Dynamics	Structure	Profile	Dynamics	Structure	Profile	Dynamics	Structure	Profile	Dynamics
TinyCycle	85.67	88.74	0.09	95.26	64.08	-1.60	77.21	69.19	-1.48	93.69	53.88	0.75	64.54	26.23	1.20	90.16	95.73	26.04	90.31	91.92	36.36
SparseTSF	86.72	86.33	-1.36	95.87	73.55	3.13	76.05	65.35	-0.17	93.54	47.33	1.62	64.66	38.63	2.08	89.90	95.98	27.61	90.37	91.99	37.76
GTR	82.77	84.40	-8.05	95.22	59.97	-1.71	77.21	71.08	-2.79	93.86	60.21	-0.87	65.45	41.99	0.44	90.54	96.39	30.64	89.17	90.79	38.31
PhaseFormer	87.35	85.68	-1.28	94.78	35.80	-0.01	78.36	69.25	0.49	93.90	51.58	2.48	65.93	32.42	2.82	90.04	95.96	29.92	90.95	92.41	38.89
Chronos-2	86.38	86.76	-8.28	96.03	76.21	-4.72	74.22	66.90	-11.14	93.28	56.21	-9.51	53.81	40.63	-22.39	89.23	96.04	29.21	90.11	91.70	39.37
Toto2-1B	86.19	84.81	-2.59	95.35	75.10	-1.70	71.27	61.17	-5.77	92.93	54.58	-1.60	64.87	38.68	-1.45	90.12	96.16	30.66	90.47	92.15	39.53
PatchTST	75.71	85.75	-24.90	94.55	71.94	-9.39	77.76	70.37	-11.74	92.90	48.94	-4.14	64.31	38.10	0.64	89.78	95.52	26.56	90.69	92.25	38.57
iTransformer	77.59	78.44	-4.65	94.94	48.42	-1.08	76.05	69.23	-1.11	93.40	41.66	-0.32	64.00	34.02	1.27	89.97	95.34	21.80	89.63	91.24	35.61
TimeMixer	81.36	83.93	-6.52	94.00	61.45	-0.47	78.06	71.05	-3.16	93.63	54.74	-2.27	65.74	40.48	0.89	88.91	96.20	27.17	90.41	92.52	39.88
DLinear	80.90	81.08	-3.17	83.92	64.03	-25.05	77.64	69.52	-0.22	91.32	55.38	-4.84	64.32	33.32	1.95	90.50	95.94	24.34	89.88	91.82	35.22
Moirai-2	61.89	31.20	-3.46	90.90	26.89	-2.19	-13.87	2.48	-39.39	85.56	2.96	-12.28	50.21	1.34	0.07	73.81	75.55	1.08	58.70	56.52	22.57

the mean of the most recent 720 historical steps as the reference. MAE contributions are attributed by symmetric replacement according to Eq. (15). Structure contributions and total improvements are averaged separately over the eight datasets before their ratio is computed. Table 10 shows that Structure accounts for 92.83% and 92.26% of the MSE improvements and 94.31% and 94.51% of the MAE improvements, respectively, exceeding half on every dataset. Table 11 retains the results after removing the Dynamic components of the lightweight models; Dynamics still provide considerable value on datasets such as Traffic.

Component errors and scoring windows. The five error terms are decomposed according to Eq. (12). The common direction is the demeaned phase profile of the true future, normalized to unit root mean square and set to zero when its energy is zero; it is used only for post hoc analysis. The residual waveform is a subcomponent of Dynamics. Table 12 combines per-dataset errors and macro-averages over seven datasets, listing Exchange separately and excluding it from the summary of datasets with natural daily cycles.

Associations with dataset characteristics. The component improvement of model M over a fixed reference B is $g_C = [E_C(B) - E_C(M)]/E(B)$, $C \in \{S, R\}$, with the reference total MSE as the denominator; signs and pairs with worse total error are retained. Descriptive metrics are defined

Table 10: Error reductions of lightweight models relative to the baseline that predicts the mean of the 720-step input.

Dataset	Model	MSE Total gain	MSE Structure gain	MSE share (%)	MAE Total gain	MAE Structure gain	MAE share (%)
ETTh1	TinyCycle	0.304	0.304	99.90	0.159	0.159	100.55
ETTh1	SparseTSF	0.308	0.312	101.41	0.178	0.182	101.83
ETTh2	TinyCycle	0.024	0.028	118.06	0.028	0.029	104.77
ETTh2	SparseTSF	0.054	0.046	84.34	0.039	0.035	90.91
ETTm1	TinyCycle	0.289	0.292	101.28	0.145	0.148	101.96
ETTm1	SparseTSF	0.279	0.279	100.05	0.144	0.144	100.05
ETTm2	TinyCycle	0.064	0.063	98.72	0.054	0.053	98.12
ETTm2	SparseTSF	0.059	0.057	96.18	0.045	0.045	98.79
Weather	TinyCycle	0.045	0.043	95.81	0.023	0.021	91.31
Weather	SparseTSF	0.047	0.044	93.01	0.045	0.042	93.02
Electricity	TinyCycle	0.691	0.648	93.72	0.479	0.452	94.32
Electricity	SparseTSF	0.692	0.646	93.35	0.483	0.454	93.96
Traffic	TinyCycle	0.975	0.766	78.58	0.502	0.438	87.33
Traffic	SparseTSF	0.984	0.767	77.95	0.509	0.442	86.86
Exchange	TinyCycle	0.600	0.633	105.48	0.208	0.206	98.98
Exchange	SparseTSF	0.525	0.569	108.44	0.153	0.165	108.00
Eight-dataset mean	TinyCycle	0.374	0.347	92.83	0.200	0.188	94.31
Eight-dataset mean	SparseTSF	0.368	0.340	92.26	0.199	0.189	94.51

Table 11: Per-dataset errors and relative increases after removing the Dynamic component of lightweight models. Table 12 reports the original per-dataset MSE and MAE; each relative increase uses the corresponding original model error as its denominator. The original macro-average MSE/MAE across eight datasets is 0.440/0.422 for TinyCycle and 0.446/0.422 for SparseTSF.

Dataset	Model	Structure only MSE	MSE increase (%)	Structure only MAE	MAE increase (%)
ETTh1	TinyCycle	0.430	0.07	0.464	-0.43
ETTh1	SparseTSF	0.422	-1.02	0.442	-0.90
ETTh2	TinyCycle	0.405	-1.06	0.431	-0.34
ETTh2	SparseTSF	0.388	2.23	0.426	0.87
ETTm1	TinyCycle	0.421	-0.87	0.421	-0.58
ETTm1	SparseTSF	0.434	-0.03	0.424	-0.08
ETTm2	TinyCycle	0.354	0.23	0.376	0.22
ETTm2	SparseTSF	0.360	0.64	0.384	0.03
Weather	TinyCycle	0.328	0.58	0.367	0.82
Weather	SparseTSF	0.327	1.01	0.345	0.89
Electricity	TinyCycle	0.249	21.09	0.330	10.38
Electricity	SparseTSF	0.251	22.40	0.328	11.29
Traffic	TinyCycle	0.664	45.85	0.402	32.66
Traffic	SparseTSF	0.664	48.53	0.401	35.70
Exchange	TinyCycle	0.883	-3.59	0.713	-0.05
Exchange	SparseTSF	0.946	-4.48	0.750	-2.41
Eight-dataset mean	TinyCycle	0.467	6.09	0.438	3.77
Eight-dataset mean	SparseTSF	0.474	6.40	0.438	3.63

consistently in Appendix E.2.2; the ground-truth level share in this analysis uses the scoring windows above. Correlations treat the seven common datasets as observations, excluding Exchange.

Table 13 combines the two fixed references, TinyCycle and SparseTSF. The directions of the associations differ across supervised models. For PhaseFormer, periodic gain and Dynamic improvement remain positively associated in leave-one-out analyses, but the association weakens substantially when Electricity and Traffic are removed together and can reverse with SparseTSF as the reference (Table 14). After Exchange is included, PhaseFormer's macro-averaged MSE and MAE over the eight datasets are both higher than those of the two lightweight references.

E.2.2 Pointwise Gains of TSFMs and Dataset Characteristics

We use PhaseFormer as the primary reference and additionally check the results against TinyCycle, SparseTSF, and GTR. Following Eq. (17), MAE improvement is divided into changes in Structure, Dynamics, and cancellation, all normalized by the reference total MAE, with both positive and negative values retained. In Table 15, both component errors of Toto2-1B increase on ETTm2, yet greater cancellation reduces the total MAE by 0.22%; overall gains therefore do not require simultaneous improvements in both components.

Table 12: Per-dataset errors and components for fixed models, followed by equally weighted macro-averages over the seven datasets with natural daily periods. Exchange is reported only at the dataset level. Macro-average MAE is not shown.

Dataset	Model	MSE	MAE	Structure MSE	Dynamics MSE	Remaining waveform
ETTh1	TinyCycle	0.430	0.466	0.112	0.318	0.106
ETTh1	SparseTSF	0.426	0.446	0.103	0.323	0.108
ETTh1	PatchTST	0.587	0.541	0.189	0.398	0.136
ETTh1	iTransformer	0.508	0.509	0.175	0.333	0.119
ETTh1	TimeMixer	0.485	0.486	0.145	0.339	0.112
ETTh1	DLinear	0.477	0.499	0.149	0.329	0.108
ETTh1	GTR	0.478	0.480	0.134	0.344	0.124
ETTh1	PhaseFormer	0.421	0.443	0.099	0.323	0.108
ETTh2	TinyCycle	0.410	0.433	0.135	0.275	0.065
ETTh2	SparseTSF	0.379	0.423	0.117	0.262	0.065
ETTh2	PatchTST	0.451	0.471	0.155	0.296	0.069
ETTh2	iTransformer	0.417	0.445	0.144	0.273	0.069
ETTh2	TimeMixer	0.442	0.460	0.170	0.272	0.068
ETTh2	DLinear	0.795	0.635	0.457	0.338	0.066
ETTh2	GTR	0.411	0.446	0.136	0.275	0.076
ETTh2	PhaseFormer	0.419	0.450	0.148	0.271	0.065
ETTm1	TinyCycle	0.425	0.424	0.199	0.226	0.085
ETTm1	SparseTSF	0.435	0.425	0.212	0.223	0.084
ETTm1	PatchTST	0.443	0.444	0.194	0.249	0.099
ETTm1	iTransformer	0.434	0.435	0.209	0.225	0.087
ETTm1	TimeMixer	0.420	0.425	0.191	0.229	0.087
ETTm1	DLinear	0.418	0.416	0.195	0.223	0.085
ETTm1	GTR	0.428	0.425	0.199	0.229	0.091
ETTm1	PhaseFormer	0.412	0.410	0.190	0.221	0.084
ETTm2	TinyCycle	0.353	0.375	0.184	0.170	0.055
ETTm2	SparseTSF	0.358	0.384	0.189	0.168	0.056
ETTm2	PatchTST	0.385	0.399	0.208	0.178	0.059
ETTm2	iTransformer	0.363	0.388	0.192	0.171	0.057
ETTm2	TimeMixer	0.360	0.385	0.185	0.175	0.057
ETTm2	DLinear	0.429	0.445	0.250	0.179	0.055
ETTm2	GTR	0.353	0.386	0.180	0.173	0.059
ETTm2	PhaseFormer	0.345	0.376	0.179	0.166	0.055
Weather	TinyCycle	0.326	0.364	0.170	0.156	0.054
Weather	SparseTSF	0.324	0.342	0.170	0.154	0.053
Weather	PatchTST	0.328	0.337	0.171	0.157	0.053
Weather	iTransformer	0.328	0.341	0.173	0.156	0.053
Weather	TimeMixer	0.321	0.337	0.164	0.156	0.053
Weather	DLinear	0.326	0.366	0.171	0.155	0.053
Weather	GTR	0.323	0.339	0.166	0.157	0.054
Weather	PhaseFormer	0.317	0.333	0.163	0.153	0.053
Electricity	TinyCycle	0.206	0.299	0.083	0.123	0.048
Electricity	SparseTSF	0.205	0.295	0.085	0.121	0.047
Electricity	PatchTST	0.208	0.300	0.086	0.122	0.048
Electricity	iTransformer	0.214	0.306	0.084	0.130	0.053
Electricity	TimeMixer	0.214	0.300	0.093	0.121	0.048
Electricity	DLinear	0.206	0.304	0.080	0.126	0.049
Electricity	GTR	0.195	0.291	0.079	0.116	0.046
Electricity	PhaseFormer	0.200	0.288	0.084	0.117	0.046
Traffic	TinyCycle	0.456	0.303	0.090	0.366	0.217
Traffic	SparseTSF	0.447	0.296	0.089	0.358	0.214
Traffic	PatchTST	0.439	0.291	0.086	0.353	0.211
Traffic	iTransformer	0.466	0.327	0.096	0.370	0.222
Traffic	TimeMixer	0.434	0.285	0.089	0.345	0.207
Traffic	DLinear	0.466	0.316	0.094	0.372	0.219
Traffic	GTR	0.455	0.312	0.101	0.354	0.214
Traffic	PhaseFormer	0.435	0.272	0.084	0.351	0.209
Exchange	TinyCycle	0.916	0.713	0.635	0.280	—
Exchange	SparseTSF	0.991	0.768	0.699	0.292	—
Exchange	PatchTST	0.990	0.731	0.677	0.313	—
Exchange	iTransformer	1.003	0.761	0.732	0.271	—
Exchange	TimeMixer	1.141	0.778	0.798	0.343	—
Exchange	DLinear	1.075	0.788	0.781	0.294	—
Exchange	GTR	1.158	0.805	0.815	0.343	—
Exchange	PhaseFormer	1.178	0.823	0.880	0.298	—
Seven-dataset mean	TinyCycle	0.372	—	0.139	0.233	0.090
Seven-dataset mean	SparseTSF	0.368	—	0.138	0.230	0.090
Seven-dataset mean	PatchTST	0.406	—	0.156	0.250	0.096
Seven-dataset mean	iTransformer	0.390	—	0.153	0.237	0.094
Seven-dataset mean	TimeMixer	0.382	—	0.148	0.234	0.090
Seven-dataset mean	DLinear	0.445	—	0.199	0.246	0.091
Seven-dataset mean	GTR	0.378	—	0.142	0.235	0.095
Seven-dataset mean	PhaseFormer	0.364	—	0.135	0.229	0.089

Common descriptive metrics for datasets. Training periodic gain, $1 - E_{\text{phase}}/E_{\text{level}}$, is computed within non-overlapping 720-step training blocks by predicting bidirectionally between the first and second halves. The phase profile and half-window mean are used as the respective predictions, and MSE is averaged over the two directions. Phases are aligned by their original temporal positions, and incomplete cycles are averaged using the actual phase counts. The ground-truth level share is $\langle \mu_Y^2 \rangle / \langle Y^2 \rangle$, where μ_Y is the mean over the scoring window. Energies are aggregated over origins and channels before their ratio is computed; this analysis uses the complete H720 window. Scale drift is the absolute logarithm of the ratio between training and test population standard deviations for

Table 13: Spearman correlations across seven datasets between dataset characteristics and the Dynamic MSE reductions of supervised models. TinyCycle and SparseTSF are each held fixed as the reference; reductions are divided by the reference total MSE on the corresponding dataset.

	TinyCycle reference			SparseTSF reference		
Model	Periodic gain	Level share	Scale drift	Periodic gain	Level share	Scale drift
PatchTST	+0.393	-0.393	-0.250	+0.357	-0.357	-0.071
iTransformer	-0.464	+0.607	+0.607	-0.393	+0.393	+0.714
TimeMixer	+0.536	-0.500	-0.429	+0.643	-0.571	-0.357
GTR	+0.571	-0.464	-0.464	+0.536	-0.429	-0.321
PhaseFormer	+0.643	-0.536	-0.607	+0.536	-0.464	-0.429
DLinear	+0.571	-0.393	-0.107	+0.143	-0.036	+0.250

Table 14: Sensitivity of correlations with the Dynamic improvement of PhaseFormer. The main correlations include seven datasets. The leave-one-out range omits one dataset at a time, and w/o et excludes both Electricity and Traffic. The Dynamic-denominator column divides the improvement by the reference Dynamic MSE; all other columns divide it by the reference total MSE.

Measure	Reference	ρ	Leave-one-out range	w/o et	Dynamic denominator
Periodic gain	TinyCycle	+0.643	$[+0.429, +0.771]$	+0.100	+0.679
Periodic gain	SparseTSF	+0.536	$[+0.257, +0.829]$	-0.300	+0.536
Level share	TinyCycle	-0.536	$[-0.829, -0.257]$	+0.300	-0.500
Level share	SparseTSF	-0.464	$[-0.714, -0.200]$	+0.400	-0.464
Scale drift	TinyCycle	-0.607	$[-0.943, -0.429]$	0.000	-0.536
Scale drift	SparseTSF	-0.429	$[-0.600, -0.086]$	+0.600	-0.429

each valid channel, followed by the mean or median over channels. The test trajectory is not counted repeatedly, and both standard deviations must exceed 10^{-12}. True future energy and test drift are used for post hoc description, and Weather retains the original missing-value placeholders.

Table 15: Per-dataset MAE reductions relative to PhaseFormer, in percentage points of the reference total MAE. Δ_S and Δ_R are the direct error reductions in Structure and Dynamics, and Δ_K is the increase in cancellation; $\Delta = \Delta_S + \Delta_R + \Delta_K$. Each entry covers all test forecast origins, forecast steps, and channels at $H = 720$.

Dataset	TSFM	Δ_S	Δ_R	Δ_K	Δ
ETTh1	Chronos-2	-2.98	-1.32	+2.91	-1.40
ETTh1	Toto2-1B	+0.08	+0.97	+1.93	+2.98
ETTh2	Chronos-2	+9.02	-2.55	-0.47	+5.99
ETTh2	Toto2-1B	+5.71	-0.95	+2.20	+6.96
ETTm1	Chronos-2	-4.52	-3.32	+3.98	-3.86
ETTm1	Toto2-1B	-4.49	-1.12	+4.10	-1.52
ETTm2	Chronos-2	-1.54	-4.16	+4.04	-1.66
ETTm2	Toto2-1B	-2.92	-1.19	+4.33	+0.22
Weather	Chronos-2	+0.56	-3.88	+3.40	+0.08
Weather	Toto2-1B	+1.94	-1.55	+2.33	+2.72
Electricity	Chronos-2	-1.95	+0.47	+2.22	+0.74
Electricity	Toto2-1B	+1.41	+0.99	+0.74	+3.15
Traffic	Chronos-2	-1.68	+1.52	+3.38	+3.21
Traffic	Toto2-1B	+0.58	+1.58	+2.01	+4.17

F Analysis on Additional Datasets

This section extends the analyses of Structure and Dynamic recovery, error variation, and model gaps to TFB datasets (Qiu et al., 2024). Appendix F.5 reports TinyCycle's settings and forecasting results on GIFT-Eval. The TFB figures follow the corresponding formats in the main text.

Table 16: Descriptive measures for seven datasets. P is the decomposition period. Training periodic gain is obtained from bidirectional half-window predictions in nonoverlapping 720-step training blocks. The level share uses the true $H = 720$ future. Scale drift uses the continuous test trajectory without duplicate counting; the last two columns give the mean and median over valid channels.

Dataset	P	Periodic gain	Level share	Scale mean	Scale median
ETTh1	24	0.194	0.411	0.524	0.531
ETTh2	24	0.040	0.900	0.666	0.486
ETTm1	96	0.258	0.419	0.527	0.537
ETTm2	96	0.020	0.928	0.667	0.485
Weather	144	0.077	0.682	0.906	0.527
Electricity	24	0.734	0.187	0.130	0.083
Traffic	24	0.599	0.066	0.148	0.101

Table 17: Sensitivity of correlations between descriptive measures and Dynamic improvement. The main correlations use PhaseFormer as the reference (seven points). Leave-one-out ranges omit one dataset at a time (six points). Merged ETT averages the related h/m datasets into ETT1 and ETT2 (five points). Excluding E/T removes both Electricity and Traffic (five points). The reference range holds TinyCycle, SparseTSF, GTR, and PhaseFormer fixed as references in turn.

Measure	TSFM	ρ	Leave-one-out range	Merged ETT	Excluding E/T	Reference range
Training periodicity	Chronos-2	+0.786	$[+0.657, +0.886]$	+0.800	+0.500	$[+0.607, +0.893]$
Training periodicity	Toto2-1B	+0.714	$[+0.600, +0.829]$	+0.800	+0.300	$[+0.536, +0.821]$
Level share	Chronos-2	-0.893	$[-1.000, -0.829]$	-0.900	-0.700	$[-0.964, -0.786]$
Level share	Toto2-1B	-0.821	$[-0.943, -0.714]$	-0.900	-0.500	$[-0.893, -0.714]$
Scale drift	Chronos-2	-0.893	$[-0.886, -0.829]$	-0.900	-0.800	$[-0.893, -0.571]$
Scale drift	Toto2-1B	-0.929	$[-0.943, -0.886]$	-0.900	-0.900	$[-0.929, -0.607]$

Scale drift is aggregated using the channel mean. Using the channel median instead gives correlations of -0.429 and -0.500 for the two models, respectively. Weather retains missing-value placeholders in the training segment, so scale drift is also affected by preprocessing.

F.1 Dataset Scope and Evaluation Settings

We compare six models, TinyCycle, DLinear, PatchTST, iTransformer, Chronos-2, and Toto2-1B, including only datasets for which all six models have complete predictions at all four forecast horizons. The common scope comprises 20 datasets: 13 long-term and 7 short-term datasets. The long-term datasets include ETTh1, ETTh2, ETTm1, ETTm2, Exchange, and Weather, together with the additional AQShunyi, AQWan, CzeLan, ZafNoo, Wind, PEMS04, and PEMS08. The short-term datasets are Covid-19, FRED-MD, NASDAQ, NN5, NYSE, Wike2000, and ILI. Electricity, Traffic, METR-LA, Solar, and PEMS-BAY have missing results because some runs exceeded the preset time budget and are therefore excluded from the common panel; this section does not cover all TFB datasets.

Each dataset uses the same TFB preprocessing and temporal split: ETT uses a 12/4/4-month split, while the other datasets retain their respective 60/20/20 or 70/10/20 splits. Channel-wise standardizers and model coefficients are fitted using only the training set. The three supervised baselines use fixed configurations and are trained for at most 10 epochs, with early-stopping patience of 3 based on validation MSE; the two foundation models perform zero-shot forecasting using their median outputs. TinyCycle selects hyperparameters on the validation set, shares a configuration across the four forecast horizons of each dataset, and is then fitted on the training set and frozen. The input limit is 720 steps for long-term tasks, 48 for short-term tasks, and 104 for ILI. TinyCycle's actual input length is the product of the selected number of cycles and period length and does not exceed the corresponding limit. Tuning budgets differ across models; this section compares forecasts and component properties under these frozen configurations.

The component analysis uses all valid consecutive test origins, channels, and forecast steps from each dataset's longest forecast (720 steps for long-term tasks and 60 for short-term tasks), retaining the final partial batch in full. All errors are computed on the channel-wise scale standardized using the training set. Elements are first averaged within each setting, followed by equal weighting over datasets and models, with long-term and short-term results reported separately. Recovery scores

follow Eq. (7), including negative values. Macro-level energy and MSE shares average the numerators and denominators before taking their ratio, while scatter points in the figures retain per-dataset shares.

Diagnostic periods are determined by the original protocol and training time axis, separately from TinyCycle's modeling periods selected on the validation set. The diagnostic periods for the seven additional long-term datasets are 24, 24, 48, 48, 96, 288, and 288, respectively. Q2 re-estimates phase Structure separately within the 192-, 336-, and 720-step prefixes. For the relative gaps in Q3, Structure is first estimated over the complete 720 steps, followed by selection of steps 1–192, 193–336, and 337–720. The model-pair heatmaps instead use independent decomposition within three 240-step segments. The first two analyses thus differ in the order of projection and segmentation. The 288-step period of PEMS04/08 exceeds both the 192-step prefix and the 240-step segment, causing phase decomposition within these short windows to degenerate. The common-prefix and independent 240-step analyses therefore use 11 long-term datasets, including 5 additional datasets. The complete 720-step analysis and the Q3 curves with fixed full-window Structure retain all 13 datasets, including 7 additional datasets. All prefixes and segments come from the same longest forecast and are not independently trained shorter-horizon forecasting tasks.

F.2 Q2: Recovery and Errors of Structure and Dynamics

Stronger Structure recovery with substantial differences across datasets. Figure 11 presents the three-prefix results on the five additional long-term datasets. Over the complete 720 steps, the seven additional datasets, including the two PEMS datasets, have average Structure recovery of 64.13% and Dynamic recovery of 11.97%; Structure accounts for 90.58% of prediction energy, and Dynamics contribute 66.63% of total MSE. Structure recovery exceeds Dynamic recovery in 40 of the 42 model–dataset settings. Across all 13 long-term datasets, the corresponding recovery scores are 71.93% and 2.87%, the predicted Structure energy share is 95.54%, and the Dynamic MSE share is 59.77%. These results support the observation that predictions are more concentrated in Structure and Dynamic recovery is generally weaker.

This pattern does not require ground-truth Dynamic energy to be small on every dataset. The ground-truth Dynamic shares on AQShunyi, AQWan, and Wind are 54.78%, 58.87%, and 63.23%, respectively. CzeLan achieves average Structure and Dynamic recovery of 95.65% and 64.50%, respectively, exhibiting strong Dynamic recovery. This phenomenon persists when test origins are divided into ten consecutive time blocks or channels are removed one at a time and cannot be attributed to a few samples. Wind is generally harder to recover, with average Structure and Dynamic recovery of only 6.21% and -0.43%; iTransformer and Toto2-1B are the two settings in which Structure recovery falls below Dynamic recovery. Thus, the observation that Structure recovery is usually stronger generalizes, while the recovery thresholds and ground-truth Dynamic energy shares within the particular scope of the main text should not be generalized uniformly.

Dynamic error grows with prefix length, while Structure-error variation depends on the dataset. Across the 11 common long-term datasets, Dynamic MSE increases successively over the 192-, 336-, and 720-step prefixes in all 66 model–dataset combinations; this also holds for all 30 combinations on the 5 additional datasets. Structure MSE, however, is not universally stable. Table 18 first averages the six models within each dataset and then compares the first and last prefixes. Exchange changes the most, increasing from 0.144 to 0.674, or 369.4%. Its diagnostic period is 1, so Structure represents only the window level, and this change reflects level error. Among the additional datasets, Wind, AQWan, ZafNoo, and AQShunyi decrease by 33.3%, 28.1%, 24.1%, and 23.9%, respectively, while CzeLan increases by 15.6%. These changes reflect both re-estimation of Structure within each prefix and expansion of the scoring window.

The macro-averaged Structure MSE on the 5 additional datasets decreases from 0.296 to 0.216, a reduction of 26.79%, while Dynamic MSE increases from 0.380 to 0.577. Removing any one additional dataset still yields a Structure reduction of 20.64%–29.48%. Changes in Structure error therefore cannot be attributed solely to a small number of outliers. On a broader range of datasets, the different directions of change in Structure and Dynamics should be retained, rather than treating small changes in absolute Structure error as a universal pattern.

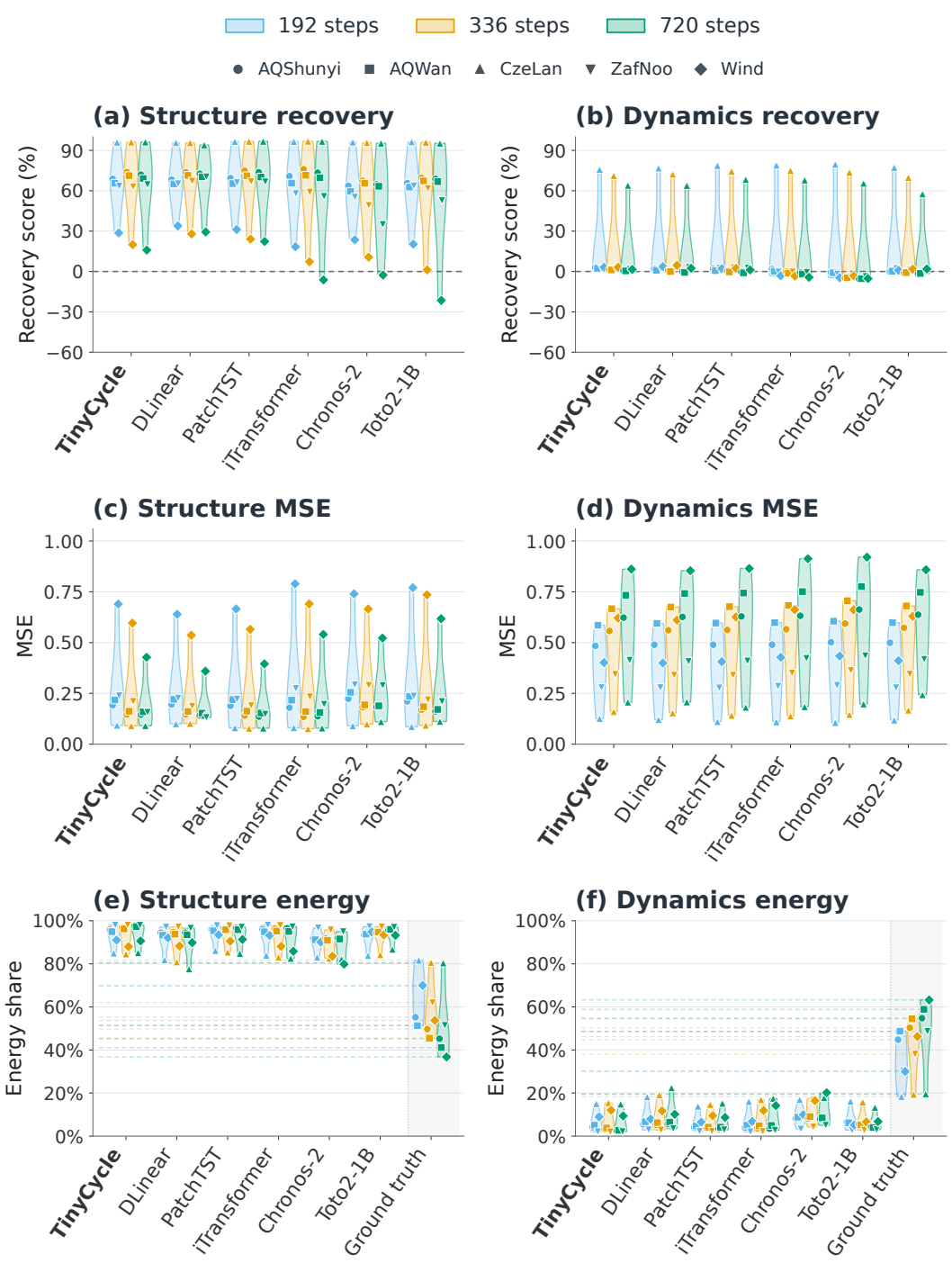


Figure 11: Structure and Dynamic analysis on five additional long-term datasets. The panels, colors for the three prefixes, marker shapes, violins, and ground-truth energy references follow Figure 6, using the six models in this section, with five per-dataset points retained in each violin. The vertical axes are continuous, and negative recovery values are retained in full.

Table 18: Changes in Structure MSE on the 11 common long-term datasets. Each entry is the mean MSE over six models, and the first five rows are additional datasets. The last two columns give the absolute and percentage changes from 192 to 720 steps. Percentages are computed from unrounded model means, rather than averaging per-model percentages.

Dataset	192 steps	336 steps	720 steps	Absolute change	Relative change
Wind	0.716	0.631	0.477	-0.239	-33.3%
AQWan	0.228	0.171	0.164	-0.064	-28.1%
ZafNoo	0.249	0.221	0.189	-0.060	-24.1%
AQShunyi	0.198	0.153	0.151	-0.047	-23.9%
CzeLan	0.088	0.089	0.102	+0.014	+15.6%
Exchange	0.144	0.261	0.674	+0.531	+369.4%
ETTh1	0.146	0.123	0.123	-0.023	-15.8%
ETTh2	0.146	0.144	0.176	+0.030	+20.4%
ETTm1	0.231	0.212	0.211	-0.020	-8.7%
ETTm2	0.170	0.175	0.200	+0.031	+18.2%
Weather	0.169	0.165	0.175	+0.006	+3.6%

F.3 Q3: Smaller Relative Dynamic Gaps and a Larger Role of Structure in Model Gaps

Relative Dynamic gaps are smaller than Structure gaps. Figure 12 follows Eq. (8) in Section 6.3, with the scope changed to the seven additional long-term datasets and six models; the minimum for each metric is selected among these six models.

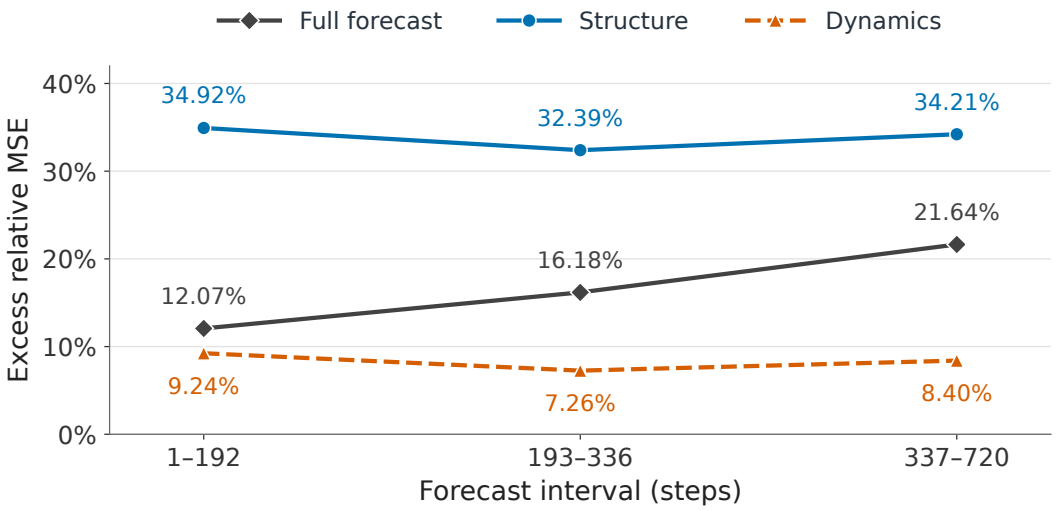


Figure 12: Average excess relative MSE over three segments for six models on seven additional long-term datasets. The curves, segment boundaries, and 0–42% vertical axis follow Figure 7; the projection scope is specified in Appendix F.1.

Structure gaps in the three segments are 34.92%, 32.39%, and 34.21%, whereas Dynamic gaps are only 9.24%, 7.26%, and 8.40%. Dynamic gaps are smaller than Structure gaps in all 21 dataset–segment comparisons over the seven additional datasets. Across all 13 datasets, this holds in 38 of 39 comparisons; the sole exception is the first segment on Exchange, where the Structure gap is 15.64% and the Dynamic gap is 20.07%. Macro-averaged Structure gaps across all 13 datasets are 28.52%, 27.18%, and 28.16%, and Dynamic gaps are 8.76%, 7.03%, and 7.56%. For both the full and additional dataset scopes, the macro-averaged direction is preserved in all three segments when any single dataset is removed.

These results support **smaller relative Dynamic gaps** within the same segment, without requiring them to shrink monotonically with forecast distance. For example, CzeLan's Dynamic gap increases from 6.14% to 15.84%, but remains below the Structure gap of approximately 30.17% in all three segments. Smaller relative gaps also do not imply smaller absolute Dynamic errors or higher recovery

scores; the interpretation conditions for fixed full-window Structure and segment-wise scoring follow Section 6.3.

Structure explains more of the MSE gaps between models. Figure 13 follows the model-pair comparison protocol in Section 6.3, using independent decompositions within three 240-step segments on the five additional datasets. Results are first averaged equally over datasets, followed by comparison of all 15 pairs among the six models.

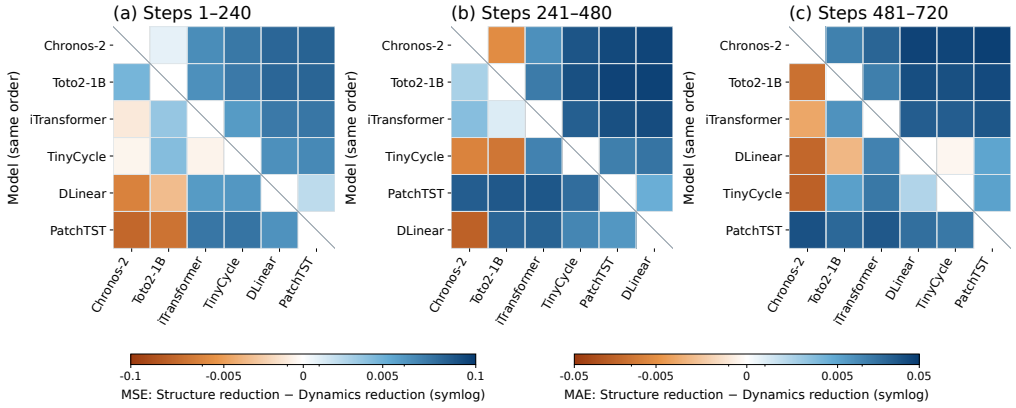


Figure 13: Reductions in component errors between model pairs on five additional long-term datasets. The upper-right triangle shows MSE and the lower-left triangle MAE, with colors interpreted as in Figure 8. Models are ordered by decreasing total MSE within each segment, and the two metrics are oriented independently. The symmetric logarithmic color scales shared across the three panels are linear within ± 0.005, with ranges of ± 0.10 for MSE and ± 0.05 for MAE.

The numbers of model pairs with larger Structure MSE reductions in the three segments are 15/15, 14/15, and 14/15, respectively. Dividing the sum of Structure reductions over all model pairs by the sum of total MSE reductions gives Structure shares of 78.96%, 81.74%, and 82.28%. On the 11 common datasets, the corresponding pair counts are 11/15, 13/15, and 13/15, with Structure shares of 78.18%, 77.44%, and 90.40%. These results support Structure as the larger contributor to MSE gaps between models in most cases. The corresponding MAE pair counts on the 5 additional datasets are 8/15, 12/15, and 10/15, still favoring Structure in most pairs but without the same monotonic pattern. MSE is additive under independent decomposition, while component MAEs are not; the MAE pair counts are therefore not interpreted as contribution shares of total MAE.

F.4 Short-Term Datasets and the Influence of Local Samples

Figure 14 retains the same six-panel format for six short-term datasets, changing the prefixes to 24, 36, and 60 steps. NASDAQ and NYSE use $P = 1$, for which Structure is only the window level; FRED-MD uses $P = 12$, and Covid-19, NN5, and Wike2000 use $P = 7$. ILI's $P = 52$ exceeds the first two prefixes, so it is excluded from the common-prefix figure but retained in the complete 60-step statistics.

Across the seven short-term datasets over the complete 60 steps, average Structure and Dynamic recovery are 72.66% and -41.93%, respectively, with Structure recovery higher in all 42 model–dataset settings. The predicted Structure energy share is 98.55%, and the Dynamic MSE share is 58.81%. The extremely low Dynamic recovery value in the figure comes from DLinear on Covid-19, indicating that its Dynamic predictions are worse than the zero-Dynamics reference. Error magnitudes vary greatly across short-term datasets, so they are not pooled with long-term datasets for averaging.

Among the 36 combinations on the six common datasets, Dynamic MSE increases successively over the three prefixes in 30. The six nonmonotonic combinations are concentrated in five models on Wike2000 and TinyCycle on NN5. The Wike2000 results are sensitive to one channel: its ground-truth peak on November 4, 2013 is 2,291,780, whereas its training-set maximum is 7,914. After removing this channel solely for diagnostic purposes, all five combinations recover the monotonic direction. On NN5, TinyCycle's 36-step Dynamic MSE exceeds its 60-step value by 0.001, and removing the single

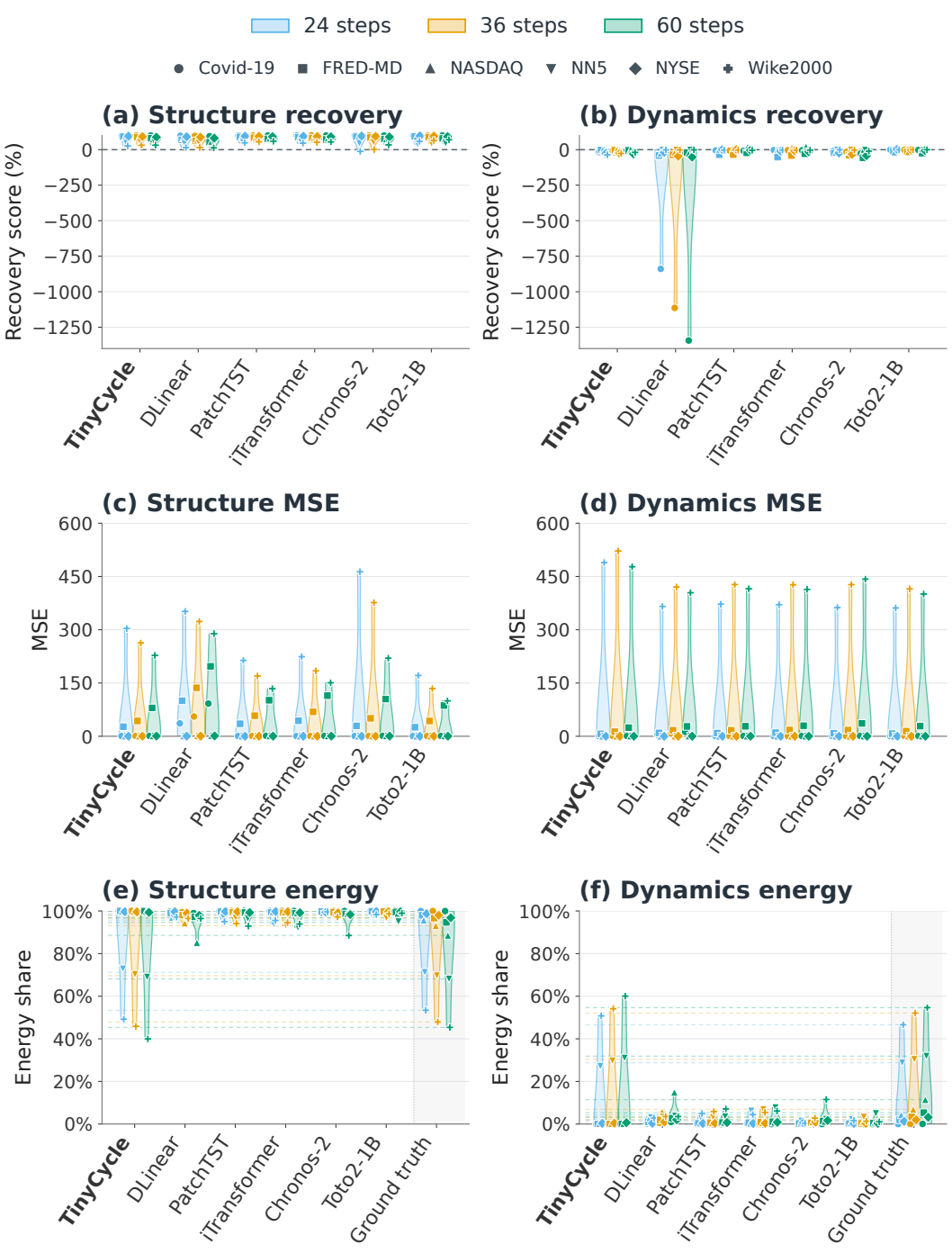


Figure 14: Structure and Dynamic analysis on six short-term datasets. Panels, marker shapes, and violins are defined as in Figure 11, with blue, orange, and green now indicating the 24-, 36-, and 60-step prefixes. To retain all observations, the recovery axis extends from -1400% to 100%, and the MSE axis from 0 to 600, both continuously; the energy axis remains 0–100%. Extreme points are neither removed nor clipped.

most unfavorable origin out of 100 test origins, or one relevant channel, can reverse the direction. Both phenomena can be traced to local samples; the main figures and summaries in this section retain all original data. These cases differ from the widespread changes in Structure error on the additional long-term datasets, strong Dynamic recovery on CzeLan, and TinyCycle's performance differences at distant horizons on the additional datasets.

F.5 TinyCycle Settings and Results on GIFT-Eval

GIFT-Eval (Aksu et al., 2024) covers multiple sampling frequencies and forecast horizons. We use all 97 task configurations, comprising 55 short-term, 21 medium-term, and 21 long-term configurations, and evaluate 371,330 univariate test windows under each setting. Testing follows the official series and rolling origins without subsampling windows. Figure 15 presents the supervised and context-based self-supervised settings and compares them with models from the same snapshot of public results.

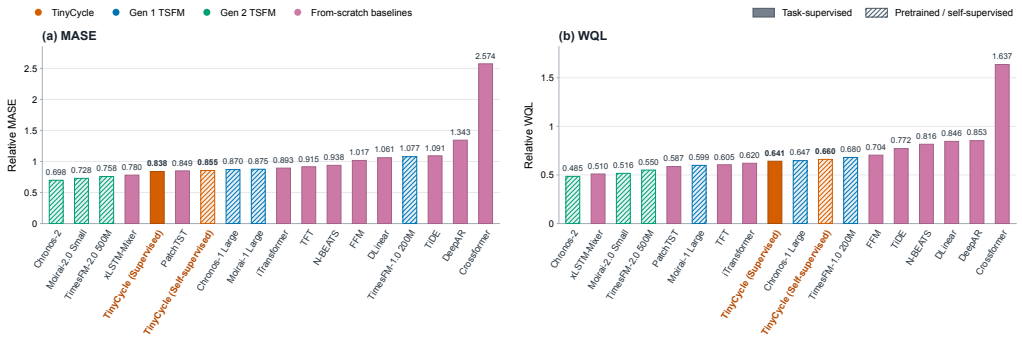


Figure 15: Relative MASE and relative WQL over all 97 GIFT-Eval configurations. Both TinyCycle variants use the adaptive-order extension described in this section; solid bars indicate supervised training on the target task, and hatched bars indicate pretrained foundation models or context-based self-supervised adaptation. Metrics are first normalized by SeasonalNaive within each configuration and then aggregated by the geometric mean. Each panel is sorted independently, and lower values are better.

Model and fitting settings. We use an adaptive-order extension of TinyCycle that shares autoregressive coefficients across periodic phases, has no bias, and constrains the coefficients to sum to 1. The candidate orders are $\{1, 2, 4, 8, 16, 30, 48, 64, 96, 128\}$, restricted by the available history length. Candidate periods are 1 and the seasonal period provided by GluonTS for the sampling frequency, regularization strength is selected from $\{0, 0.01, 1\}$, and at most the most recent 4096 historical steps are used. Coefficients are fitted by regularized least squares with the sum constraint. Forecasting recursively uses previously generated values, checks the stability of recursive characteristic roots other than the unit root, and shrinks the coefficients toward uniform weights when necessary. This extension uses at most 128 autoregressive coefficients.

The supervised version fits coefficients shared within each configuration on the official training set and uses one forecast window at the end of the official validation set to select the period, order, and regularization strength. The selection metric is MAE normalized by the historical seasonal-difference scale and then averaged equally over series. At most 2048 series are selected deterministically from each of the training and validation sets, and training fits use at most 50,000 time points. After hyperparameter selection, the coefficients fitted on the training set are retained, and configurations and weights are frozen before testing.

The context-based self-supervised version uses no externally pretrained weights and does not access the official training or validation splits of the target task. At each forecast origin, it independently uses the currently observed history of each univariate series. If the history length is $L \geq 8$, the last $\min(H, \lfloor L/5 \rfloor)$ steps are held out for internal validation, where H is the forecast horizon. Candidates are fitted on the earlier prefix and selected using this holdout, then refitted on the entire known history, with at most 2048 time points used in each fit. If fewer than 8 historical steps are available, the last

value is repeated. Coefficients are re-estimated for each context and are not retained across forecast origins.

Probabilistic forecasts and evaluation metrics. Both settings estimate predictive standard deviations from a robust scale of one-step innovation residuals and seasonal autoregressive error propagation, producing nine Gaussian quantiles at $0.1, 0.2, \ldots, 0.9$. An uncertainty multiplier is calibrated on the supervised validation segment or the internal context holdout and constrained to $[0.1, 10]$; the supervised version additionally stores this scalar. Missing inputs are filled by causal forward filling, and missing test labels are masked by the official evaluator. We report mean absolute scaled error (MASE) and weighted quantile loss (WQL) averaged over the nine quantiles. In the figure, each configuration's metric is first divided by the corresponding SeasonalNaive metric for that configuration, and the geometric mean is then taken over the 97 ratios. Lower values are better, and 1 indicates performance equal to SeasonalNaive.

Results. The supervised version achieves relative MASE and relative WQL of 0.838 and 0.641, respectively, and the context-based self-supervised version achieves 0.855 and 0.660, respectively, all below SeasonalNaive. In MASE, the supervised version outperforms all supervised models except xLSTM-Mixer, as well as first-generation TSFMs, suggesting that Structure modeling has some generality. Because TinyCycle is not designed specifically for quantile forecasting, its WQL performance is relatively weaker.

G Counterexample Analysis

The Structure–Dynamics decomposition provides a common coordinate system for comparing forecast components, while the recoverability of each component depends on the data and model. Using Exchange and the counterexamples in Appendix F, this section explains the meaning of the decomposition under weak periodicity, how TinyCycle models such data, and why individual observations deviate from the overall results.

G.1 Structure and Dynamics in Data with Extremely Weak Periodicity

The diagnostic period for Exchange is $P = 1$. In this case, Eq. (1) remains a valid orthogonal projection, with all time points assigned to the same phase. For each channel j, Structure and Dynamics are

$$S_{h,j} = \bar{y}_j = \frac{1}{L}\sum_{s=1}^{L} y_{s,j}, \qquad R_{h,j} = y_{h,j} - \bar{y}_j. \tag{41}$$

Applying the same procedure to predictions, Structure error measures the difference in mean levels between the predicted and true windows, whereas Dynamic error measures the difference between the two demeaned trajectories. Because Dynamics have zero mean within the scoring window, total MSE still satisfies Eq. (2). Extremely weak periodicity therefore changes the meaning of the Structure term without invalidating the decomposition. The periodic profile after removing the window mean remains well defined and is identically zero, but its energy-normalized recovery score is undefined because the denominator is zero.

This interpretation preserves the distinction among level, drift, and turning points in exchange-rate trajectories. The window mean level enters Structure, while within-window trend changes, turning points, and local fluctuations enter Dynamics. Dynamics are therefore not equivalent to high-frequency noise, nor are they unpredictable by definition. High Structure energy likewise does not establish strong periodicity, because it includes the squared energy of the mean level. The true future mean is used for post hoc evaluation and is unavailable to TinyCycle during forecasting. Whether a prediction accurately recovers this level still depends on how well the history explains the future mean. In the six-model TFB analysis, the Structure MSE on Exchange increases from 0.144 for the 192-step prefix to 0.674 for the 720-step prefix, a 369.4% increase (Table 18). Because the mean is re-estimated separately in each prefix, this result indicates substantial differences in the models' level-estimation errors for future windows of different lengths; patterns of Structure-error variation on periodic data cannot be applied directly to Exchange.

G.2 How TinyCycle Models Exchange

The component analysis of TinyCycle in this section uses the original frozen outputs before the lookback extension. Mean correction is enabled for Exchange and takes the form in Eq. (34): the input-window mean is first subtracted, recursion is applied to historical blocks centered on that mean, and the mean is then added back. Thus, even without a clear periodic profile, the model has a path for propagating the local level, without relying entirely on recursive weights whose sum is not constrained to 1 to preserve the absolute input level.

The boundary term further carries into the forecast the deviation of the last observation from the mean at the final position of the historical blocks. The decay scale on Exchange is $\tau P = 672$ steps, and the gains fitted on the training set are 1 for all four TFB forecast horizons, allowing this correction to persist over longer forecast intervals. It is added only to the output and does not enter subsequent recursion. These mechanisms explain how TinyCycle can still produce useful forecasts on weakly periodic data. Under the original configuration, TinyCycle's average MSE/MAE across the four horizons on Exchange is 0.376/0.407, lower than those of the three lightweight baselines; at 720 steps, its MSE is 0.916, higher than DLinear's 0.637 and iTransformer's 0.619 (the two baseline results are in Table 4). Level propagation and a shared combination of historical blocks retain some forecasting ability, while trajectory changes over long horizons that are not described by these mechanisms can still produce substantial errors.

Existing component analyses further locate this forecasting ability. Under the original configuration, TinyCycle's Structure and Dynamic MSE on the complete 720-step Exchange forecast are 0.635 and 0.280, respectively. Relative to the 720-step historical-mean baseline, total MSE decreases by 0.600, with Structure decreasing by 0.633 and Dynamics offsetting part of the improvement (Tables 10 and 12). Removing Dynamics from the same prediction reduces total MSE from 0.916 to 0.883 (Table 11). This comparison shows that this configuration's gains over the historical mean arise from more accurate estimation of the future level, while the generated nonconstant variations provide no net improvement.

G.3 Sources of the TFB Counterexamples

Relative gaps in the first segment on Exchange. Among the 39 dataset–segment comparisons across all 13 long-term datasets, the first segment on Exchange is the only case in which the relative Dynamic gap exceeds the Structure gap, at 20.07% and 15.64%, respectively. The immediate reason is evident in the values before normalization: the six-model average Structure MSE is 0.674 and its minimum is 0.583, while the average Dynamic MSE is 0.433 and its minimum is 0.361. The absolute Dynamic excess over its own minimum is 0.072, smaller than the Structure excess of 0.091, but division by the smaller minimum error reverses their relative ordering.

Here, projection is first applied over the complete 720 steps, followed by scoring over the first 192 steps: Structure compares errors in the full-window mean, while Dynamics compare trajectory deviations from that mean in the first segment. The latter need not have zero mean within the shorter interval. The two segment MSEs therefore cannot be added directly, and the relative gaps above cannot determine which component contributes more to total error. This differs from an analysis that re-centers within the 192-step prefix.

Gains from Dynamics in individual model pairs. The model-pair comparisons on the five additional datasets use independent projections within three 240-step segments, allowing total MSE reductions to be decomposed directly. In the middle segment, Toto2-1B reduces Structure and Dynamic MSE relative to Chronos-2 by 0.007 and 0.015, respectively; in the final segment, TinyCycle reduces them relative to DLinear by 0.001 and 0.002, respectively. These two pairs are the exceptions to the 14/15 pairs with larger Structure reductions in their respective segments. Both components improve, but Dynamics improve more, showing that the relative advantage of a particular model pair can arise from Dynamics. A larger Structure contribution in most model pairs does not determine the source of the performance difference in every pair.

Recoverable Dynamics and difficult-to-recover Structure. The six-model average Dynamic recovery on CzeLan reaches 64.50%, providing an example of relatively strong Dynamic recoverability. Each model's Dynamic recovery exceeds 15% in all ten consecutive time blocks and remains high

when channels are removed one at a time. The phenomenon persists under these diagnostics and cannot be attributed to a single time interval or variable. Statistically, this means that the shared profile with $P = 48$ does not retain all predictable variation, and models also recover some deviations from that profile. The current diagnostics locate the distribution of this variation but cannot yet attribute it to a particular external driving mechanism. In contrast, Structure itself is difficult to recover on Wind: Toto2-1B has Structure recovery of -21.42% and Dynamic recovery of 1.89%, and their ordering remains unchanged after removing the most unfavorable 5% of origins. The corresponding values for iTransformer are -6.22% and -4.34%, both worse than their respective zero-component references. Here, the reversal in recovery ordering primarily indicates inaccurate Structure prediction and cannot generally be interpreted as strong Dynamic prediction.

Local events and sample sensitivity in short-term counterexamples. On Wike2000, Dynamic MSE decreases from 36 to 60 steps for five models, with the counterexamples concentrated in the same channel. This channel has a ground-truth peak of 2,291,780 on November 4, 2013, while its training-set maximum is only 7,914. With the training scale held fixed, this event produces a large squared error and is scored at multiple origins because consecutive forecast windows overlap. The two prefixes also re-estimate Structure separately and use different averaging denominators, so the influence of a few large errors on the prefix averages can reverse the direction of monotonicity. After removing this channel solely for diagnostic purposes, Dynamic MSE again increases with prefix length for all five models. For TinyCycle on NN5, the difference is only 0.001, and removing the single most unfavorable origin out of 100, or one relevant channel, can reverse the direction; this channel does not exhibit the large scale jump seen in Wike2000. The former case identifies a concentrated event, while the latter reflects sample sensitivity of a small difference. Neither provides a reason to remove these observations from the reported statistics.

Projection scope and phase weights. Some changes in Structure error arise from the evaluation definition. When each prefix is projected independently, the average profile is re-estimated as window length changes, so Structure MSE can increase or decrease. When full-window Structure is fixed before segmentation, the error of the same profile in each segment depends on how often each phase appears. The diagnostic period of PEMS04/08 is 288, whereas the three segments have lengths of 192, 144, and 384, which are not integer multiples of a complete cycle. Structure MSE and the best-model denominator can therefore vary between segments. When phase weights are identical across segments, equal fixed-Structure errors follow by definition. These two types of comparison should be interpreted separately. Stable relative Structure gaps also do not imply that TinyCycle has an advantage at distant horizons on more datasets: in the final segment over the seven additional datasets, it ranks 4/6 in MSE and 5/6 in MAE. After removing the 5% of origins most unfavorable to TinyCycle on each dataset, its MAE still exceeds that of the best baseline selected using the complete data (Appendix F).

H Ablation Study of TinyCycle

H.1 Experimental Settings and Ablation Variants

Except for the ablated components, the experimental settings match those in the main text.

Recursive feedback and output corrections. The *w/o AR* variant removes autoregressive feedback: it repeats the first cycle obtained by weighting the observed history throughout the forecast horizon, without feeding generated cycles into subsequent inputs. When retained, the boundary term still decays with the actual forecast lead according to Eq. (4). The *w/o mean correction* variant disables the input-window mean compensation in Eq. (34), while *w/o boundary correction* sets $g = 0$ and removes this parameter. The full model and all removal combinations of these three components form eight variants. We also report *w/o AR* with the original w, g frozen, distinguishing a prediction intervention with fixed parameters from an ablation that refits the boundary gain.

Learning cycle coefficients. The *w/o soft sharing* variant sets $\lambda = 0$. The *w/o age weighting* variant uses uniform age weights $a_i = 1$, still normalizes the mean diagonal entry of the penalty matrix to one, and retains the original λ. Hard sharing constrains all historical cycle coefficients w_i

to a single learned scalar. Fixed uniform coefficients set $w_i = 1/Q$. Both retain recursive feedback and refit one global boundary gain, giving two and one learned parameters, respectively.

Sharing across channels and phases. The *w/o channel sharing* variant learns Q coefficients per channel. The *w/o phase sharing* variant groups training targets by their absolute row index $t \bmod P$, learns Q coefficients per phase, and uses the corresponding absolute phase at prediction time. Both use the original training observations and loss weighted by channel scale, with the regularization term scaled by the mean diagonal entry of each group's Gram matrix. Each variant still fits only one global g, yielding $cQ + 1$ and $PQ + 1$ parameters, respectively. The phase-specific variant thus also introduces a dependence on absolute phase alignment.

H.2 Overall Results and Component Effects

Table 19 summarizes all 15 variants. The full model has a mean MSE of 0.321 and MAE of 0.345, with an average of 25.25 learned parameters. For *w/o AR* with a refitted boundary gain, MSE and MAE increase by 22.89% and 7.88%, respectively. The corresponding increases with frozen parameters are 23.71% and 8.13%. The *w/o boundary correction* variant increases MSE and MAE by 18.40% and 11.14%. Under the current configurations, both cycle recursion and boundary correction therefore help reduce average forecasting error.

Table 19: Overall comparison of TinyCycle component and sharing ablations. Full TinyCycle is the reference; all variants cover the same eight datasets and four forecast horizons.

Variant	MSE	MAE	ΔMSE (%)	ΔMAE (%)	Mean params.
Full TinyCycle	0.321	0.345	+0.00	+0.00	25.25
w/o AR (refit g)	0.394	0.372	+22.89	+7.88	25.25
w/o boundary correction	0.380	0.383	+18.40	+11.14	24.25
w/o mean correction	0.335	0.367	+4.57	+6.52	25.25
w/o AR & boundary	0.445	0.405	+38.74	+17.64	24.25
w/o AR & mean	0.389	0.380	+21.25	+10.30	25.25
w/o mean & boundary	0.362	0.386	+12.76	+11.98	24.25
w/o AR, mean & boundary	0.417	0.399	+30.03	+15.66	24.25
w/o soft sharing	0.326	0.346	+1.56	+0.54	25.25
w/o age weighting	0.322	0.347	+0.46	+0.60	25.25
Hard-shared coefficients	0.364	0.373	+13.42	+8.16	2.00
Fixed uniform coefficients	0.364	0.370	+13.59	+7.41	1.00
w/o channel sharing	0.321	0.345	+0.03	+0.05	4610.75
w/o phase sharing	0.330	0.349	+2.86	+1.25	709.00
w/o AR (frozen w, g)	0.397	0.373	+23.71	+8.13	25.25

Note: MSE and MAE are averaged over all valid test origins, forecast steps, and channels within each task, then macro-averaged with equal weight across the 32 tasks. $\Delta = 100(E_{\text{variant}}/E_{\text{full}} - 1)$ is computed from unrounded macro-averages; positive values indicate higher error. Mean params. is the equally weighted mean number of independent learnable real scalars in each task's model. The frozen control counts its inherited model parameters; expanded copies of coefficients are not counted again. The refitted non-AR variant refits the global boundary gain g for its modified predictions using training data only; the frozen variant retains the full model's w, g. Other variants retaining boundary correction also fit g on training data for their respective predictions. Here, w/o means without, AR denotes autoregressive feedback, and mean/boundary refer to the corresponding corrections. w/o age weighting sets $a_i = 1$ while retaining soft-sharing regularization.

Mean correction is already disabled in the original configurations for ETTh1, ETTm1, Weather, and Electricity, so *w/o mean correction* changes only the other 16 tasks. On these tasks, MSE and MAE increase by 8.35% and 12.89%, compared with 4.57% and 6.52% over all 32 tasks. The effects of multiple components cannot be obtained by adding their individual increases. For example, *w/o mean & boundary* has an MSE of 0.362, lower than the 0.380 of *w/o boundary correction*, but a slightly higher MAE. These combinations reflect interactions between the corrections and the base prediction.

The *w/o soft sharing* and *w/o age weighting* variants increase overall MSE by 1.56% and 0.46%, respectively. ETTm1, Weather, and Electricity, whose original configurations have $\lambda = 0$, remain unchanged in both comparisons. Among the other 20 tasks, uniform age weights lower MSE on 13 and raise it on seven. The average advantage of age weighting therefore does not imply a benefit on every task. Constraining historical coefficients to one learned scalar or fixing all of them to $1/Q$

raises MSE by 13.42% and 13.59%, respectively, indicating that allowing different coefficients for different historical cycles improves average accuracy on the current tasks.

For *w/o channel sharing*, the average parameter count rises to 4610.75, while MSE and MAE increase by only 0.03% and 0.05%. MSE improves on 16 tasks and worsens on 16. For *w/o phase sharing*, the average parameter count is 709, and MSE and MAE increase by 2.86% and 1.25%. With the same training data and fixed hyperparameters, neither finer-grained coefficient setting improves overall error. The full model's sharing scheme achieves similar or lower average error with fewer parameters.

H.3 Results by Dataset and Forecast Horizon

Table 20 averages the four forecast horizons equally for each dataset, while Table 21 averages the eight datasets equally for each horizon. Both retain all 15 variants. The effect of *w/o AR* is larger on Traffic and Electricity, where mean MSE increases by 98.05% and 53.03%, respectively. The *w/o boundary correction* variant raises mean MSE on Exchange by 110.11% but lowers it on ETTh1 by 0.89%. ETTm1 and ETTm2 predict only one cycle when $L = P = 96$, so removing recursive feedback does not change their predictions. Differences emerge at longer horizons. These results correspond to four separate forecasting tasks, rather than segments of a single long forecast.

Table 20: Dataset-level means for TinyCycle ablations. Each entry equally averages the four forecast horizons $L \in \{96, 192, 336, 720\}$ for that dataset. All 15 variants are shown in two groups of dataset columns.

Variant	ETTh1		ETTh2		ETTm1		ETTm2	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Full TinyCycle	0.394	0.427	0.346	0.385	0.355	0.379	0.258	0.315
w/o AR (refit g)	0.404	0.417	0.380	0.398	0.355	0.377	0.261	0.316
w/o boundary correction	0.391	0.426	0.346	0.390	0.359	0.379	0.295	0.342
w/o mean correction	0.394	0.427	0.496	0.486	0.355	0.379	0.302	0.366
w/o AR & boundary	0.402	0.417	0.380	0.399	0.360	0.377	0.293	0.341
w/o AR & mean	0.404	0.417	0.391	0.417	0.355	0.377	0.278	0.345
w/o mean & boundary	0.391	0.426	0.495	0.489	0.359	0.379	0.336	0.387
w/o AR, mean & boundary	0.402	0.417	0.391	0.417	0.360	0.377	0.308	0.365
w/o soft sharing	0.420	0.429	0.345	0.386	0.355	0.379	0.269	0.320
w/o age weighting	0.412	0.442	0.340	0.385	0.355	0.379	0.257	0.317
Hard-shared coefficients	0.414	0.443	0.342	0.387	0.372	0.390	0.259	0.319
Fixed uniform coefficients	0.411	0.434	0.341	0.386	0.382	0.390	0.262	0.321
w/o channel sharing	0.403	0.436	0.337	0.383	0.360	0.381	0.257	0.315
w/o phase sharing	0.397	0.430	0.347	0.386	0.394	0.400	0.261	0.317
w/o AR (frozen w, g)	0.412	0.422	0.397	0.403	0.355	0.377	0.262	0.316

Variant	Weather		Electricity		Traffic		Exchange	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Full TinyCycle	0.248	0.298	0.165	0.260	0.416	0.281	0.385	0.413
w/o AR (refit g)	0.292	0.328	0.253	0.309	0.824	0.422	0.385	0.408
w/o boundary correction	0.257	0.306	0.166	0.260	0.417	0.281	0.808	0.680
w/o mean correction	0.248	0.298	0.165	0.260	0.415	0.280	0.309	0.442
w/o AR & boundary	0.292	0.328	0.253	0.309	0.824	0.422	0.757	0.650
w/o AR & mean	0.292	0.328	0.253	0.309	0.824	0.422	0.316	0.426
w/o mean & boundary	0.257	0.306	0.166	0.260	0.416	0.280	0.475	0.560
w/o AR, mean & boundary	0.292	0.328	0.253	0.309	0.824	0.422	0.508	0.553
w/o soft sharing	0.248	0.298	0.165	0.260	0.416	0.280	0.388	0.421
w/o age weighting	0.248	0.298	0.165	0.260	0.416	0.281	0.386	0.413
Hard-shared coefficients	0.280	0.331	0.226	0.309	0.632	0.390	0.386	0.414
Fixed uniform coefficients	0.285	0.325	0.227	0.307	0.633	0.388	0.375	0.410
w/o channel sharing	0.225	0.285	0.162	0.257	0.428	0.284	0.394	0.418
w/o phase sharing	0.276	0.308	0.166	0.259	0.415	0.278	0.385	0.413
w/o AR (frozen w, g)	0.289	0.324	0.253	0.309	0.824	0.423	0.383	0.408

Note: For each dataset, errors are first averaged over all test elements within each of its four tasks, then macro-averaged equally across the four horizons. Tasks are not weighted by their numbers of origins or channels. Values are rounded to three decimal places and do not establish equality of the unrounded values. The non-AR variant with refitted g and the intervention with frozen w, g are listed separately. Variant names follow Table 19.

Table 21: Forecast-horizon means for TinyCycle ablations. Each entry equally averages the eight datasets at the corresponding forecast horizon L.

Variant	$L = 96$		$L = 192$		$L = 336$		$L = 720$	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Full TinyCycle	0.239	0.286	0.280	0.319	0.324	0.352	0.440	0.422
w/o AR (refit g)	0.305	0.314	0.349	0.346	0.405	0.382	0.518	0.446
w/o boundary correction	0.301	0.336	0.341	0.363	0.385	0.390	0.492	0.443
w/o mean correction	0.261	0.315	0.304	0.346	0.347	0.376	0.430	0.432
w/o AR & boundary	0.355	0.357	0.399	0.383	0.459	0.416	0.568	0.465
w/o AR & mean	0.318	0.336	0.356	0.360	0.403	0.388	0.479	0.436
w/o mean & boundary	0.292	0.339	0.331	0.366	0.371	0.393	0.454	0.446
w/o AR, mean & boundary	0.345	0.357	0.381	0.379	0.431	0.407	0.512	0.451
w/o soft sharing	0.245	0.289	0.286	0.321	0.329	0.353	0.443	0.423
w/o age weighting	0.243	0.291	0.282	0.321	0.324	0.353	0.440	0.422
Hard-shared coefficients	0.289	0.322	0.326	0.350	0.365	0.378	0.476	0.441
Fixed uniform coefficients	0.290	0.320	0.327	0.348	0.366	0.376	0.475	0.437
w/o channel sharing	0.237	0.285	0.278	0.318	0.323	0.353	0.446	0.423
w/o phase sharing	0.246	0.290	0.288	0.323	0.332	0.356	0.453	0.427
w/o AR (frozen w, g)	0.306	0.314	0.352	0.346	0.410	0.384	0.521	0.447

Note: MSE/MAE are first averaged over all valid test origins, forecast steps, and channels within each task, then macro-averaged equally across the eight datasets. L denotes the forecast horizon, distinct from the history length. Parameter settings and the meanings of refitting and freezing are as in Table 19. Values are rounded to three decimal places.

I Limitations and Future Work

Our findings on differences in the recovery of Structure and Dynamics characterize most evaluated tasks, but a small number of counterexamples indicate that this empirical trend does not hold universally. TinyCycle remains effective on these datasets, yet predictive accuracy does not imply recovery of the underlying data-generating mechanisms. Motivated by the observation that existing models generally recover Structure more effectively than Dynamics, TinyCycle focuses on modeling Structure. However, the Structure–Dynamics decomposition provides a perspective for analyzing predictive behavior and error sources; its components do not necessarily correspond to the underlying data-generating mechanisms, and TinyCycle likewise leaves this identification problem unresolved. Forecasting research should extend beyond improving accuracy to examine which predictable patterns models capture and how these patterns relate to the underlying mechanisms. Through the Structure–Dynamics decomposition, our work offers a framework for testing falsifiable hypotheses about the sources of predictive performance. Future work could further examine the scope of this decomposition, identify the sources of predictability in Structure and Dynamics, and develop modeling approaches and evaluation frameworks that connect predictive behavior to data-generating mechanisms.