RETHINKING UNIVERSAL TIME SERIES FOUNDATION MODELS FROM UNIFIED REPRESENTATIONS TO SHARED DYNAMICS (UNITFU)

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series foundation models (TSFMs) have attracted increasing attention in cross-domain forecasting. Existing works mainly rely on large-scale pre-training and general modeling paradigms. However, the former heavily relies on the diversity and domain coverage of pre-training data. Meanwhile, the latter learns unified representations across heterogeneous domains without explicitly identifying the temporal regularities shared among them, limiting robust cross-domain generalization. This work explores cross-domain generalization by characterizing commonalities in recurring local dynamics across domains, rather than relying solely on implicitly learned temporal representations. To capture such cross-domain commonalities, we propose **Time Series Foundation Units (TFUs)** as reusable building blocks discovered from multi-scale temporal segments, each characterizing a recurring local dynamic pattern shared across domains. Building on TFUs, we develop a unified time series foundation model (**UniTFU**) that explicitly reuses and composes shared temporal regularities across domains for forecasting. Experiments on seven public datasets show that UniTFU substantially outperforms existing TSFMs in cross-domain forecasting and can efficiently adapt to target-domain dynamics by updating its reusable components with limited target data. UniTFU also achieves strong forecasting performance compared with state-of-the-art methods.

1 INTRODUCTION

Time series forecasting plays a crucial role across many domains, including weather forecasting (Wu et al., 2021), electricity monitoring (Nguyen & Hansen, 2017), and energy management (Kardakos et al., 2013). As forecasting applications expand across diverse domains, specialized models designed for individual domains (Nie et al., 2023; Liu et al., 2024a; Yao et al., 2025) struggle to scale across domains. Inspired by the success of foundation models (FMs) in natural language processing (NLP) (Zhao et al., 2026) and computer vision (CV) (Awais et al., 2025), time series foundation models (TSFMs) have emerged as a promising paradigm for cross-domain forecasting. By learning general temporal representations from large-scale time series data, TSFMs enable a single model to perform forecasting across diverse domains (Das et al., 2024; Liu et al., 2024b; Wu et al., 2026).

A major trend in current TSFM research is to build general forecasters with unified modeling capability, enabling consistent cross-domain forecasting performance (Aksu et al., 2024). Existing TSFMs generally pursue cross-domain generalization from two perspectives: 1) large-scale pre-training and 2) general modeling paradigms. TSFMs based on large-scale pre-training (Liu et al., 2025b; Jin et al., 2024) learn shared temporal patterns from large-scale, multi-domain time series corpora, enabling cross-domain generalization. However, their generalization largely depends on the diversity and domain coverage of pre-training data (Zhang et al., 2025), limiting robust performance on highly heterogeneous or underrepresented domains. TSFMs with general modeling paradigms (Liu et al., 2025a; Cheng et al., 2026) accommodate different temporal scales, variable dimensions, and forecasting tasks within a unified representation space (Woo et al., 2024), but such unified representations do not explicitly identify the temporal regularities shared across domains (Otović et al., 2022; Ning et al., 2026), thereby limiting robust cross-domain generalization. Overall, existing approaches largely obtain cross-domain generalization through data-driven representation learning

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.

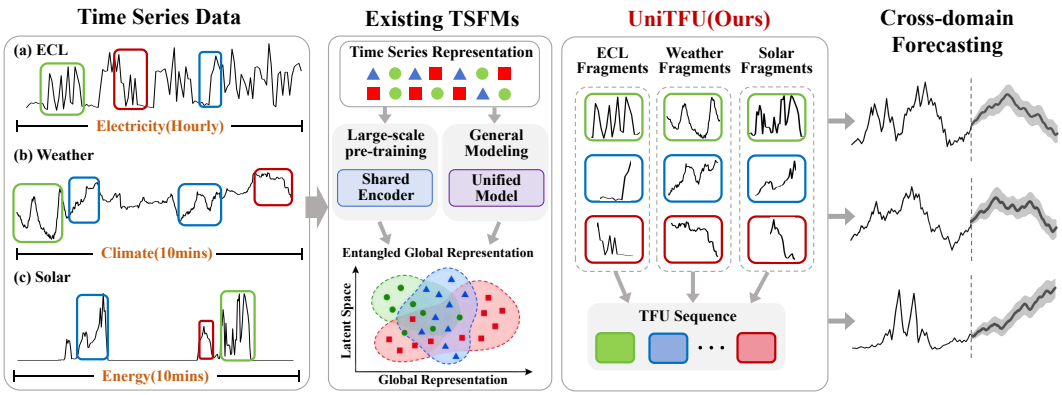


Figure 1: Conceptual comparison of existing TSFMs and UniTFU. Existing TSFMs mainly rely on implicit temporal representations, whereas UniTFU explicitly organizes recurring local dynamics across domains into shared TFUs for reuse and composition in cross-domain forecasting.

or unified modeling, while recurring local dynamics shared across domains remain insufficiently characterized and explicitly reused. This raises a key question: *How can TSFMs characterize the commonalities in temporal evolution across domains for robust cross-domain forecasting?*

In this work, we introduce a new modeling perspective that grounds TSFMs' general forecasting capability in discovering, reusing, and composing shared local dynamic structures across domains. Accordingly, we propose the **Time Series Foundation Unit (TFU)** as a reusable and composable prototype that characterizes recurring local dynamic patterns within finite temporal segments. As illustrated in Figure 1, existing TSFMs primarily rely on implicit temporal representations, whereas TFUs explicitly organize recurring local dynamics into reusable units for cross-domain forecasting. Such units may capture recurring local behaviors such as rising, falling, oscillatory, or transition patterns. Rather than being manually predefined, TFUs are automatically discovered from large-scale time series data and continuously merged and updated into a shared dictionary. Complex temporal dynamics can then be represented as compositions of these shared units, enabling explicit reuse of temporal knowledge across domains.

Building on TFUs, we develop **UniTFU**, a unified time series foundation model that leverages shared TFUs to guide temporal dependency modeling and refine forecasting across domains. Specifically, UniTFU incorporates unit-level TFUs and their aggregated scene context into the attention mechanism to guide temporal dynamics modeling. Experimental results show that UniTFU exhibits robust cross-domain generalization in forecasting and effectively adapts to target-domain dynamics with limited data. UniTFU also achieves strong performance on standard forecasting benchmarks. Our contributions are summarized as follows:

- We introduce TFUs as reusable and composable prototypes of recurring local dynamics discovered from multi-scale temporal segments. TFUs provide a unit-level perspective on shared temporal regularities across domains, offering a foundation for Universal time series forecasting.
- We develop UniTFU, a TSFM that explicitly integrates the discovery, reuse, and composition of TFUs into forecasting, enabling shared local dynamics to be leveraged across domains.
- We systematically evaluate UniTFU in cross-domain settings and on standard forecasting benchmarks, demonstrating robust cross-domain generalization and strong forecasting performance. Further analysis shows that TFUs learned from source domains can be reused in target domains and further adapted with limited target data.

2 Related Work

2.1 Time Series Foundation Models for Forecasting

Existing TSFMs mainly follow two complementary routes: large-scale pre-training and general modeling paradigms. Large-scale pretrained TSFMs learn general temporal representations from

multi-domain corpora for zero-shot forecasting and cross-domain adaptation. Representative methods include TimesFM (Das et al., 2024), Time-MoE (Shi et al., 2025), MOIRAI (Woo et al., 2024), Chronos (Ansari et al., 2024), and Sundial (Liu et al., 2025b). Recent works such as Chatime (Wang et al., 2025) and Aurora (Wu et al., 2026) further incorporate multimodal information into TSFM pre-training for heterogeneous domains. However, their generalization largely depends on pre-training diversity and domain coverage, limiting robustness on underrepresented target domains.

Another line of work develops general modeling paradigms for general-purpose time series forecasting. Early deep forecasting models, such as DLinear (Zeng et al., 2023), PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024a), and PhaseFormer (Niu et al., 2026), introduce general modeling structures for long-term and multivariate forecasting. These methods promote unified forecasting beyond task-specific architectures. Recent TSFMs further incorporate unified architectures into foundation-model design, including Timer-XL (Liu et al., 2025a), Timer (Liu et al., 2024b), and GTM (Cheng et al., 2026). Although these methods improve adaptability across diverse forecasting scenarios, their cross-domain generalization mainly relies on unified architectures, representations, or prediction procedures, which do not explicitly identify and reuse the recurring local dynamic structures shared across domains, making robust cross-domain generalization challenging.

2.2 Local Temporal Representation Learning for Time Series

Local temporal patterns have long been studied as compact representations of time series. Shapelet-based methods identify discriminative subsequences for classification (Ye & Keogh, 2009; Grabocka et al., 2014), while motif discovery searches for recurring subsequences through similarity matching (Mueen et al., 2009). Beyond direct subsequence discovery, dictionary learning represents signals through sparse combinations of learned atoms (Mairal et al., 2009), whereas symbolic approaches encode local subsequences into reusable symbolic words for classification (Schäfer, 2015). Neural approaches further introduce learned codebooks and prototype representations. VQ-based models quantize continuous latent embeddings using a learned discrete codebook (Van Den Oord et al., 2017), while prototypical networks represent each class by the mean embedding of its support examples (Snell et al., 2017). TOTEM (Talukder et al., 2024) learns a VQ-based tokenizer independently of downstream models and reuses the resulting tokenization across domains and tasks.

In this work, we introduce TFUs as reusable prototypes of recurring local dynamics and develop UniTFU. Rather than learning a discrete tokenization through a fixed-size codebook, UniTFU explicitly discovers and organizes recurring local dynamics into reusable TFUs, which are adaptively updated, expanded, and consolidated during forecasting training and composed across multiple scales to guide cross-domain forecasting.

3 Model

We propose **UniTFU**, a unified time series foundation model that enables cross-domain forecasting through reusable Time Series Foundation Units (TFUs), as illustrated in Fig. 2. UniTFU contains three core modules: (1) the **TFU Learning Module**, which discovers multi-scale local temporal structures and maps them to a shared TFU prototype dictionary; (2) **Scene Context Aggregation**, which summarizes the TFU sequence into a global context of local dynamic composition; and (3) **TFU Attention**, which incorporates the TFU sequence and scene context into a forecasting backbone. A lightweight predictor then produces the final forecast.

Problem Statement. Given a historical observation sequence with C variables, $X = \{x_1, x_2, \ldots, x_L\} \in \mathbb{R}^{L\times C}$ over L time steps, UniTFU forecasts the subsequent H steps as $\hat{Y}_{L+1:L+H} = f_\theta(X) \in \mathbb{R}^{H\times C}$, where f_θ denotes the time series foundation model. UniTFU applies sliding windows at multiple temporal scales to extract candidate temporal segments, which are subsequently represented by reusable TFUs to provide structured context for forecasting.

3.1 TFU Learning Module

The TFU learning module discovers, reuses, and composes recurring local dynamics shared across domains, organizing them as reusable prototypes into a shared adaptive dictionary. It first encodes multi-scale temporal segments and suppresses redundant candidates, then matches the retained em-

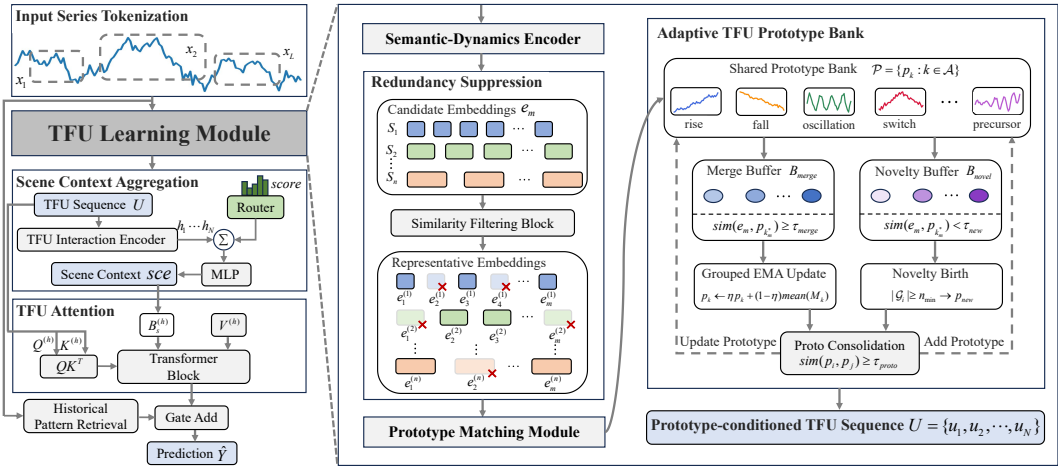


Figure 2: The overview of UniTFU. The TFU learning module encodes multi-scale temporal segments, suppresses redundant candidates, and organizes recurring local dynamics into a shared adaptive dictionary (TFU Prototype Bank) through prototype updating, novelty birth, and consolidation. The resulting TFU sequence and aggregated scene context guide temporal dynamics modeling, enabling explicit reuse and composition of shared local dynamics for cross-domain forecasting.

beddings to shared TFU prototypes. The matched prototypes are aggregated at the token positions covered by the corresponding candidates to construct the TFU prototype bank. During training, prototype updating, novelty birth, and consolidation adapt the shared dictionary to newly observed local dynamics.

Semantic Dynamics Encoder. Let $Z_m = [z_{m,1}, z_{m,2}, \ldots, z_{m,S}] \in \mathbb{R}^{S\times D}$ denote the m-th candidate segment in the token representation space extracted from consecutive tokens of a single variable, where S denotes its temporal scale in tokens. Each token represents a temporal patch through a learned linear projection with positional information. UniTFU summarizes each segment by its global activation $\bar{z}_m$, element-wise local peak response $z_m^{\max}$, and temporal offset Δz_m, which summarize average activations, peak responses, and changes between the first and last token representations. Each summary is a D-dimensional vector, yielding a $3D$-dimensional input to the segment encoder. These statistics are concatenated and projected by a lightweight encoder to obtain the segment embedding e_m:

$$\bar{z}_m = \frac{1}{S}\sum_{i=1}^{S} z_{m,i}, \qquad z_m^{\max} = \max_{1\le i\le S} z_{m,i}, \qquad \Delta z_m = z_{m,S} - z_{m,1}, \tag{1}$$
$$e_m = \mathrm{LayerNorm}(\mathrm{MLP}([\bar{z}_m; z_m^{\max}; \Delta z_m])) \in \mathbb{R}^D.$$

The segment encoder can receive forecasting gradients through the similarity-dependent token-level composition described below, while prototype assignment is discrete.

Redundancy Suppression. Overlapping windows may produce redundant candidates describing the same local temporal structure. Before prototype assignment, UniTFU filters candidates according to their temporal supports and representation similarity. For candidates extracted at temporal scale S, we denote the embedding of the m-th candidate as $e_m^{(S)}$. For two candidate segments Z_i and Z_j from the same variable and scale S, we compute the cosine similarity between their embeddings:

$$\mathrm{sim}(e_i^{(S)}, e_j^{(S)}) = \frac{(e_i^{(S)})^T \cdot e_j^{(S)}}{|e_i^{(S)}|_2 |e_j^{(S)}|_2}. \tag{2}$$

Within the same scale, a candidate is suppressed when its temporal support overlaps with another candidate and their similarity exceeds τ_{intra}. Across temporal scales, a shorter candidate covered by a longer one from the same variable is suppressed when their similarity exceeds τ_{cross}. This removes redundant candidates while preserving complementary local structures across temporal scales.

Adaptive TFU Prototype Bank. After redundancy suppression, the retained candidate embeddings are assigned to a shared TFU prototype bank $\mathcal{P} = \{p_k : k \in \mathcal{A}\}$, where $\mathcal{A} \subseteq \{1, \ldots, K\}$ indexes

active TFUs and K is the maximum dictionary capacity. For each retained embedding e_m, UniTFU selects the most similar prototype and records the corresponding matching confidence:

$$k_m^* = \arg\max_{k \in \mathcal{A}} \mathrm{sim}(e_m, p_k), \qquad q_m = \mathrm{sim}(e_m, p_{k_m^*}), \tag{3}$$

where the index k_m^* determines the TFU assigned to the candidate, while q_m measures how well the candidate is explained by the selected prototype.

To construct a temporally aligned TFU sequence, the matched prototype is written back to the token positions covered by candidate Z_m. Let $\mathcal{I}_m$ denote the token positions covered by candidate Z_m within its original variable. When multiple candidates cover the same position, their prototype representations are aggregated according to the matching confidence:

$$u_t = \frac{\sum_{Z_m : t \in \mathcal{I}_m} w_m p_{k_m^*}}{\sum_{Z_m : t \in \mathcal{I}_m} w_m}, \qquad w_m = \exp\left(\frac{q_m}{\tau_{\mathrm{temp}}}\right), \tag{4}$$

where τ_{temp} is a temperature parameter. Positions not covered by any retained candidate preserve their original representations, yielding the prototype-conditioned TFU sequence $U = \{u_1, u_2, \cdots, u_N\}$, where N is the number of token positions in U. To adapt the TFU dictionary to newly observed dynamics, UniTFU periodically maintains the prototype bank using buffered candidate embeddings. Candidates well represented by existing prototypes are routed to the merge buffer $\mathcal{B}_{\mathrm{merge}}$, while insufficiently represented candidates are routed to the novelty buffer $\mathcal{B}_{\mathrm{novel}}$:

$$\begin{aligned} \mathrm{sim}(e_m, p_{k_m^*}) \geq \tau_{\mathrm{merge}} &\Rightarrow e_m \in \mathcal{B}_{\mathrm{merge}}, \\ \mathrm{sim}(e_m, p_{k_m^*}) < \tau_{\mathrm{new}} &\Rightarrow e_m \in \mathcal{B}_{\mathrm{novel}}, \end{aligned} \tag{5}$$

where $\tau_{\mathrm{new}} < \tau_{\mathrm{merge}}$. Candidates with $\tau_{\mathrm{new}} \leq q_m < \tau_{\mathrm{merge}}$ are treated as uncertain and excluded from bank updating by default. Let $\mathcal{M}_k$ denote the merge candidates assigned to prototype k. UniTFU updates each prototype using a grouped exponential moving average:

$$p_k \leftarrow \eta p_k + (1 - \eta) \frac{1}{|\mathcal{M}_k|} \sum_{e_m \in \mathcal{M}_k} e_m, \tag{6}$$

where $\eta \in [0, 1)$ controls the contribution of the previous prototype. The novelty buffer accumulates candidates that are insufficiently covered by the current bank. During periodic bank maintenance, each unused novelty candidate e_i serves as an anchor and forms $\mathcal{G}_i = \{e_j \in \mathcal{B}_{\mathrm{novel}} \mid \mathrm{sim}(e_i, e_j) \geq \tau_{\mathrm{novel}}\}$. If $|\mathcal{G}_i| \geq n_{\min}$, the mean representation of $\mathcal{G}_i$ initializes a new prototype. Finally, prototypes satisfying $\mathrm{sim}(p_i, p_j) \geq \tau_{\mathrm{proto}}$ are consolidated to reduce dictionary redundancy. The dictionary therefore adapts both prototype representations and the number of active TFUs, while matched prototypes provide temporally aligned context for forecasting.

3.2 Scene Context Aggregation

The prototype-conditioned TFU sequence U preserves token-level local temporal structures, while forecasting also requires a global representation of how these units compose within the current input window. UniTFU first applies a lightweight interaction encoder, $\mathcal{H} = E(U) = [h_1, \ldots, h_N]$, where $h_t \in \mathbb{R}^D$, to contextualize the TFU representations. We then summarize the contextualized sequence through complementary attention, mean, and element-wise max pooling. Specifically, attention pooling uses scores $a_t = \mathrm{score}(h_t)$ with $\alpha_t = \mathrm{softmax}(a)_t$ to obtain $c_{\mathrm{attn}} = \sum_{t=1}^{N} \alpha_t h_t$, while $c_{\mathrm{mean}} = N^{-1} \sum_{t=1}^{N} h_t$ and $c_{\max} = \max_{1 \leq t \leq N} h_t$ provide complementary global summaries. The resulting scene context is:

$$sce = \mathrm{LayerNorm}(\mathrm{MLP}_{\mathrm{scene}}([c_{\mathrm{attn}}; c_{\mathrm{mean}}; c_{\max}])) \in \mathbb{R}^D. \tag{7}$$

3.3 TFU Attention

The TFU sequence U and scene context sce provide complementary structural information for forecasting. UniTFU incorporates them into each token representation z_i through gated residual conditioning:

$$\tilde{z}_i = \mathrm{LayerNorm}(z_i + \gamma_u \phi_u(u_i) + \gamma_s \phi_s(sce)), \tag{8}$$

where ϕ_u and ϕ_s are learnable projections, and γ_u and γ_s are learnable scalar gates. The TFU-aware attention blocks then process the conditioned token sequence. For attention head h, let $q_i^{(h)}$ and $k_j^{(h)}$ denote the projected query and key from the conditioned sequence. We project sce into a head-specific representation $s_h^{(h)}$ and compute $u_i^{(h)} = f_u^{(h)}([q_i^{(h)}; s_h^{(h)}])$ and $v_j^{(h)} = f_v^{(h)}([k_j^{(h)}; s_h^{(h)}])$. These r-dimensional representations define a low-rank scene-dependent relation bias that modulates the attention logits:

$$B_s^{(h)}(sce)_{ij} = \frac{(u_i^{(h)})^T v_j^{(h)}}{\sqrt{r}}, \qquad A^{(h)} = \mathrm{softmax}\left(\frac{Q^{(h)}{K^{(h)}}^T + \gamma_h B_s^{(h)}(sce)}{\sqrt{d_h}} + M_{\mathrm{p}}\right). \quad (9)$$

The head output is $O^{(h)} = A^{(h)}V^{(h)}$, where d_h is the head dimension, γ_h is a learnable head-specific gate, and M_{p} restricts token-level attention across positions, while TFU representations and scene context are constructed from the full observed input window.

Prediction. The output of TFU Attention is mapped by a lightweight prediction head to obtain the dynamic forecast Y_{d}. To exploit recurring historical patterns, UniTFU constructs Y_{h} by matching the latest query segment to historical segments within the current input window and reusing the immediately following segment of the matched candidate, with both segments required to lie entirely within the available history. The final prediction is $\hat{Y} = Y_{\mathrm{d}} + \lambda(Y_{\mathrm{h}} - Y_{\mathrm{d}})$, where $\lambda \in [0, 1]$ is a learnable gate. More details are provided in Appendix A.3.

4 EXPERIMENTS

We conduct extensive experiments to evaluate UniTFU. Section 4.1 introduces the experimental settings. Section 4.2 examines cross-domain generalization through controlled source-to-target transfer and target-domain adaptation, directly evaluating whether learned TFUs can be reused and updated across unseen domains. Section 4.3 evaluates standard supervised multivariate forecasting, while Section 4.4 compares UniTFU with publicly available pretrained TSFMs in zero-shot forecasting scenarios. Finally, Section 4.5 and Section 4.6 provide ablation studies and detailed model analyses.

4.1 EXPERIMENTAL SETTINGS

Datasets & Metrics. We evaluate UniTFU on seven forecasting datasets: the four ETT subsets, Electricity, Solar, and Weather, using mean squared error (MSE) and mean absolute error (MAE). Pre-training uses UTSD (Liu et al., 2024b) and LOTSA (Woo et al., 2024), excluding all target evaluation datasets to avoid data leakage. Dataset details are provided in Appendix A.1.

Baselines. We compare UniTFU with two categories of representative TSFMs: (1) TSFMs with general modeling paradigms, including GTM (Cheng et al., 2026), Timer-XL (Liu et al., 2025a), and Timer (Liu et al., 2024b); (2) large-scale pre-trained TSFMs, including Aurora (Wu et al., 2026), Sundial (Liu et al., 2025b), TiRex (Auer et al., 2026), Time-MoE (Shi et al., 2025), TimesFM (Das et al., 2024), Chronos (Ansari et al., 2024), and MOIRAI (Woo et al., 2024). For supervised forecasting, we additionally compare with strong task-trained forecasting models, including PhaseFormer (Niu et al., 2026), DLinear (Zeng et al., 2023), PatchTST (Nie et al., 2023), and iTransformer (Liu et al., 2024a). Detailed descriptions of the baselines are provided in Appendix A.2.

Implementation Details. We use an input length of $L = 672$. UniTFU uses input/output token lengths of 96, obtaining longer horizons through rolling prediction. Direct forecasting baselines follow the same rolling forecasting procedure. TFU learning uses scales $S = \{1, 2, 4\}$ and a maximum dictionary capacity of $K = 128$ (1024 for pre-training), with the number of active TFUs determined adaptively. We set $\tau_{\mathrm{intra}} = \tau_{\mathrm{cross}} = 0.92$, $\tau_{\mathrm{merge}} = 0.82$, $\tau_{\mathrm{new}} = \tau_{\mathrm{merge}} - 0.10$, $\tau_{\mathrm{novel}} = 0.80$, and $\tau_{\mathrm{proto}} = 0.97$. These settings are shared across datasets without dataset-specific tuning. Complete configurations and hyperparameter sensitivity analyses are provided in Appendix A.3 and Fig. 5.

4.2 CROSS-DOMAIN TIME SERIES FORECASTING

We evaluate how effectively reusable TFUs transfer to unseen target domains under direct transfer and target-domain adaptation.

Table 1: Direct transfer results of unified-paradigm TSFMs. Results are MSE/MAE averaged over four source domains. **Red**: the best, **Blue**: the 2nd best. All the results are provided in Table 9 of Appendix B.1.

Target Domain	UniTFU		w/ VQ		w/o TFU		GTM		Timer-XL		Timer	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.537**	**0.490**	**0.561**	**0.501**	0.633	0.527	0.653	0.577	0.635	0.541	0.635	0.533
ETTm1	**0.879**	**0.601**	0.919	0.617	1.097	0.661	0.947	0.645	**0.894**	**0.614**	1.104	0.669
Weather	**0.245**	**0.287**	**0.250**	**0.292**	0.267	0.302	0.326	0.351	0.277	0.313	0.269	0.300
ECL	**0.579**	**0.560**	**0.601**	**0.570**	0.605	0.575	0.652	0.592	0.612	0.592	0.612	0.580
Solar	**0.620**	**0.562**	**0.657**	**0.590**	0.700	0.600	0.851	0.653	0.862	0.691	0.706	0.602
Average	**0.572**	**0.500**	**0.598**	**0.514**	0.660	0.533	0.686	0.563	0.656	0.550	0.665	0.537

Direct Transfer. We evaluate direct transfer across ETTh1, ETTm1, Weather, ECL, and Solar under the $672 \rightarrow 96$ setting. Each model is trained on one source domain and evaluated on the other four without target-domain training samples or parameter updates. We further evaluate two controlled variants: w/o TFU removes the TFU learning module, and w/ VQ replaces the adaptive TFU dictionary with a fixed-size VQ codebook (Van Den Oord et al., 2017) using cosine-similarity assignment and grouped EMA updates. Table 1 averages results over four source domains per target. UniTFU achieves the lowest average MSE and MAE on all five targets, reducing overall MSE by 16.6% and 12.8% relative to GTM and Timer-XL, and by 13.4% and 4.3% relative to w/o TFU and w/ VQ. These results suggest that TFUs learned from source domains capture reusable local dynamics that remain effective on unseen target domains without dictionary adaptation.

Table 2: Target-domain adaptation with 10% training samples under the $672 \rightarrow 96$ setting. Results are averaged over four source domains. **Red**: the best, **Blue**: the 2nd best. All the results are provided in Table 10 of Appendix B.2.

Type	10% Few-shot						Frozen-TFU		Updated-TFU		10% Few-shot	
Target Domain	Timer		Timer-XL		GTM		UniTFU		UniTFU		UniTFU	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.509	0.485	0.491	0.477	0.526	0.490	0.482	0.467	**0.478**	**0.462**	**0.447**	**0.453**
ETTm1	0.575	0.507	0.570	0.483	0.625	0.487	**0.447**	**0.434**	**0.440**	**0.429**	0.555	0.474
Weather	**0.176**	**0.224**	**0.176**	0.228	0.179	0.230	0.190	0.248	0.186	0.244	**0.177**	**0.223**
ECL	**0.166**	0.267	0.171	**0.266**	0.188	0.285	0.235	0.334	0.217	0.316	**0.164**	**0.259**
Solar	**0.193**	0.254	**0.188**	**0.240**	0.204	**0.238**	0.259	0.329	0.233	0.291	0.202	0.241
Average	0.323	0.347	0.319	**0.339**	0.344	0.346	0.322	0.362	**0.310**	0.348	**0.309**	**0.330**

Target Domain Adaptation. We evaluate target-domain adaptation using 10% of the target domain training samples under the $672 \rightarrow 96$ setting. Alongside standard 10% few-shot results for Timer, Timer-XL, GTM, and UniTFU, we evaluate two UniTFU variants: Frozen-TFU updates only the forecasting head, while Updated-TFU additionally updates the TFU dictionary; both keep the remaining backbone fixed. As shown in Table 2, Updated-TFU improves average MSE and MAE over Frozen-TFU on all five target domains, reducing average MSE and MAE by 3.7% and 3.9%. These results indicate that source-domain TFUs provide a transferable prior for target-domain adaptation, while dictionary updates with limited target data further refine the reusable temporal structures. Updated-TFU also outperforms standard few-shot UniTFU on ETTm1, while remaining behind standard few-shot UniTFU and Timer-XL on ECL and Solar. This suggests that dictionary-level adaptation can be effective when target-domain dynamics can be accommodated by refining reusable TFUs, but dictionary updates do not always replace broader model adaptation.

4.3 MULTIVARIATE TIME SERIES FORECASTING

We follow iTransformer (Liu et al., 2024a) to evaluate UniTFU performance in standard supervised multivariate rolling forecasting. As shown in Table 3, UniTFU achieves the lowest average MSE and MAE on five of seven datasets, with the lowest overall average MSE of 0.285 and MAE of 0.330. These results demonstrate the effectiveness of UniTFU in standard multivariate forecasting.

Table 3: Multivariate forecasting results averaged over horizons $\{96, 192, 336, 720\}$ with a lookback length of 672. **Red**: the best, **Blue**: the 2nd best. Details are provided in Table 11 of Appendix C.

Type	TSFM								Forecasting Model							
Models	**UniTFU (Ours)**		GTM (2026)		Timer-XL (2025)		Timer (2024)		PhaseFormer (2026)		iTransformer (2023)		DLinear (2023)		PatchTST (2022)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.396**	**0.423**	**0.398**	0.429	0.409	0.433	0.418	0.433	0.418	**0.427**	0.419	0.440	0.436	0.451	0.407	0.432
ETTh2	**0.366**	**0.405**	0.375	0.412	0.376	0.413	0.379	0.417	**0.369**	0.413	0.395	0.424	0.520	0.494	0.372	**0.409**
ETTm1	**0.337**	**0.371**	**0.346**	**0.374**	0.369	0.394	0.347	0.378	0.360	0.383	0.387	0.409	0.356	0.379	0.363	0.394
ETTm2	0.285	0.336	0.266	0.324	0.276	0.330	**0.260**	**0.318**	0.271	0.326	0.296	0.344	0.284	0.346	**0.263**	**0.323**
ECL	**0.164**	**0.259**	0.173	0.272	0.169	0.262	0.169	0.261	0.168	**0.260**	**0.166**	**0.259**	**0.166**	0.261	0.173	0.272
Weather	**0.235**	**0.270**	0.240	0.274	0.242	0.279	0.243	0.276	**0.235**	0.278	0.239	0.278	0.240	0.291	**0.231**	**0.271**
Solar	**0.209**	**0.250**	0.231	**0.250**	0.219	**0.262**	0.220	0.266	0.216	0.273	0.219	0.290	0.221	0.282	**0.210**	0.294

4.4 PRETRAINED TIME SERIES FORECASTING

We compare pretrained UniTFU with seven pretrained TSFMs using a 672-step lookback and horizons $\{96, 192, 336, 720\}$. For UniTFU, all evaluation datasets are excluded from pre-training, while competing TSFMs are evaluated only on datasets not reported in their pre-training corpora. As shown in Table 4, UniTFU achieves the lowest average MSE on five of seven datasets and the lowest average MAE on four, with an overall MSE of 0.280 and MAE of 0.317. These results support the effectiveness of reusable TFUs for transferring shared local dynamics across downstream datasets.

Table 4: Pretrained forecasting results averaged over horizons $\{96, 192, 336, 720\}$ with a lookback length of 672. $('-')$ indicates that the dataset was included in pre-training and therefore excluded from evaluation. **Red**: the best, **Blue**: the 2nd best. Details are provided in Table 12 of Appendix D.

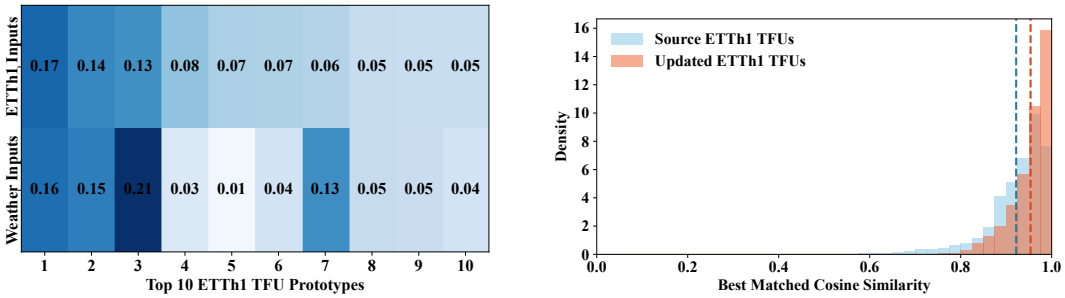
Models	**UniTFU (Ours)**		Aurora (2026)		Sundial (2025)		TiRex (2025)		Time-MoE (2025)		MOIRAI (2024)		TimesFM (2023)		Chronos (2024)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.380**	**0.403**	0.387	0.418	**0.382**	**0.406**	0.437	0.417	0.450	0.434	0.451	0.441	0.413	0.408	0.579	0.473
ETTh2	0.348	**0.379**	0.347	0.389	**0.346**	0.383	**0.343**	**0.369**	0.368	0.395	0.405	0.412	0.370	0.388	0.444	0.429
ETTm1	**0.361**	**0.370**	0.424	0.427	**0.363**	**0.372**	0.409	0.374	0.420	0.388	0.488	0.463	0.384	**0.370**	0.644	0.498
ETTm2	**0.272**	**0.314**	0.297	0.344	**0.274**	0.320	0.291	**0.316**	0.307	0.335	0.362	0.366	0.293	0.317	0.372	0.380
ECL	0.158	0.245	0.179	0.276	**0.145**	**0.242**	**0.156**	**0.240**	–	–	0.211	0.303	–	–	0.263	0.319
Weather	**0.240**	**0.258**	0.275	0.295	**0.242**	0.273	0.244	**0.260**	0.260	0.287	0.353	0.320	–	–	0.324	0.312
Solar	**0.201**	0.252	0.216	0.255	**0.203**	**0.240**	0.335	0.274	0.348	0.287	0.719	0.702	0.259	**0.238**	0.719	0.606

4.5 ABLATION STUDY

We conduct ablation experiments under the $672 \rightarrow 96$ forecasting setting using four random seeds, including the benchmark seed 2021. We consider five variants: w/o TFU and w/ VQ are defined in Section 4.2; w/o Sce excludes scene context aggregation; w/o Att replaces TFU attention with standard self-attention; w/o Res removes the historical-pattern residual. As shown in Table 5, removing individual components generally increases mean errors, supporting their complementary contributions. The w/o TFU results support the contribution of TFU learning, while the comparison with w/ VQ supports the adaptive dictionary design.

Table 5: Ablation results under the $672 \rightarrow 96$ forecasting setting. For ETT, mean and std are averaged over ETTh1, ETTh2, ETTm1, and ETTm2. All the results are provided in Table 13.

Models	UniTFU		w/o TFU		w/o Sce		w/o Att		w/o Res		w/ VQ	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT	**0.279 ± 0.0026**	**0.336 ± 0.0028**	0.287 ± 0.0023	0.342 ± 0.0025	0.283 ± 0.0026	0.339 ± 0.0028	0.283 ± 0.0046	0.339 ± 0.0045	0.285 ± 0.0034	0.341 ± 0.0034	0.284 ± 0.0022	0.340 ± 0.0027
ECL	**0.133 ± 0.0034**	**0.231 ± 0.0051**	0.138 ± 0.0019	0.236 ± 0.0021	0.134 ± 0.0026	0.232 ± 0.0028	0.138 ± 0.0033	0.236 ± 0.0037	0.136 ± 0.0028	0.234 ± 0.0037	0.136 ± 0.0013	0.233 ± 0.0019
Weather	0.159 ± 0.0030	**0.208 ± 0.0028**	0.162 ± 0.0033	0.214 ± 0.0034	0.161 ± 0.0028	0.210 ± 0.0037	**0.157 ± 0.0028**	0.209 ± 0.0024	0.158 ± 0.0029	0.210 ± 0.0030	0.161 ± 0.0005	0.212 ± 0.0010
Solar	**0.182 ± 0.0024**	**0.221 ± 0.0024**	0.192 ± 0.0022	0.227 ± 0.0022	0.195 ± 0.0025	0.231 ± 0.0022	0.198 ± 0.0036	0.233 ± 0.0037	0.188 ± 0.0028	0.224 ± 0.0029	0.187 ± 0.0038	0.224 ± 0.0018



(a) Cross-domain TFU reuse.

(b) Target-domain TFU refinement.

Figure 3: TFU analysis from ETTh1 to Weather: (a) prototype usage in both domains; (b) best-match cosine similarities of Weather candidates before and after dictionary updating.

Table 6: Efficiency analysis under the $672 \rightarrow 336$ setting with batch size 1.

Models	Weather			ECL		
	Params (M)	Latency (ms)	Memory (MB)	Params (M)	Latency (ms)	Memory (MB)
Timer	12.709	29.140	76.8	6.404	41.057	108.8
Timer-XL	12.710	100.976	65.4	6.405	156.321	880.6
GTM	35.996	85.196	187.7	33.015	174.764	691.9
UniTFU	119.137	402.177	554.2	119.137	678.949	1448.4

4.6 Model Analysis

Table 6 shows that the evaluated UniTFU configuration uses more parameters and incurs higher inference latency and peak GPU memory than the baselines. We further analyze TFU reuse, adaptation, and local shape representations from ETTh1 to Weather. Fig. 3a shows that Weather inputs reuse TFUs frequently activated in ETTh1. As shown in Fig. 3b, updating the dictionary with limited Weather data increases best-match cosine similarities, indicating closer matches between target-domain candidates and the updated TFUs. We examine the local waveform information retained by the segment encoder through nearest-neighbor retrieval, using raw temporal statistics and random retrieval as baselines. Self-matches and overlapping segments are excluded, and waveform correlation is computed after linear detrending and normalization. Fig. 4 shows that embedding-based retrieval consistently identifies segments with higher shape correlations across all three settings. Together, these analyses support cross-domain TFU reuse and adaptation, with retrieval results further supporting local shape information beyond coarse amplitude and linear trend statistics. See Appendix G for details.

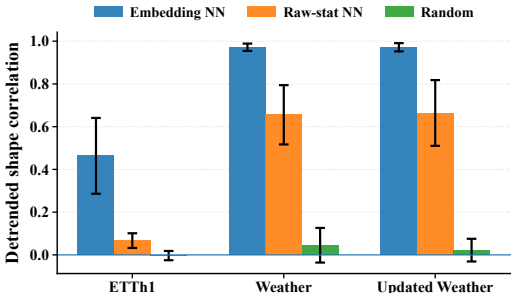


Figure 4: Local shape correlation across retrieval strategies.

5 Conclusion

We propose UniTFU, a time series foundation model that organizes recurring local dynamics into reusable and adaptable TFUs for cross-domain forecasting. Within UniTFU, comparisons with w/o TFU and a fixed-size EMA codebook support the contributions of TFU learning to cross-domain forecasting. Model analyses further support the retention of local shape information beyond coarse temporal statistics. These findings suggest that TFUs update improve average target-domain adaptation. However, TFU benefits vary across transfer settings, and the full model incurs higher inference costs than the evaluated baselines. Future work will investigate the adaptation of TFUs to a broader range of downstream time series tasks and reduce the computational cost of TFU-based forecasting.

AI USE STATEMENT

Generative AI tools were used in this work to provide feedback on experimental design and to assist in code development and debugging. We did not use generative AI tools to generate synthetic datasets, formulate mathematical claims or proofs, assist in proof writing, clean or reformat datasets, or conduct qualitative or thematic data analysis, and the remaining required-disclosure tasks are not applicable to this work. Additionally, we used generative AI tools to assist in translating, drafting, and editing portions of the manuscript for clarity and readability. We reviewed all AI-assisted work: experimental suggestions were independently evaluated by the authors, AI-assisted code was inspected and tested through the reported experiments, and AI-assisted text was verified, revised, and approved by the authors. All experimental results were produced using the authors' implementations and checked through the reported experimental procedures. We take responsibility for the final content of this work, including any text, claims, code, results, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The main text describes the formulation, architecture, and training procedure of UniTFU. The appendix provides additional information required for reproduction, including dataset descriptions, implementation and backbone configurations, hyperparameter settings, evaluation procedures, and complete experimental results. The source code will be made publicly available upon acceptance.

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. In *NeurIPS Workshop on Time Series in the Age of Large Models*, 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. *Transactions on Machine Learning Research*, 2024.

Andreas Auer, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. Tirex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. *Advances in Neural Information Processing Systems*, 38:57529–57580, 2026.

Muhammad Awais, Muzammal Naseer, Salman Khan, Rao Muhammad Anwer, Hisham Cholakkal, Mubarak Shah, Ming-Hsuan Yang, and Fahad Shahbaz Khan. Foundation models defining a new era in vision: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2245–2264, 2025. doi: 10.1109/TPAMI.2024.3506283.

HE Cheng, Xu Huang, Gangwei Jiang, Zhaoyi Li, Defu Lian, Hong Xie, Enhong Chen, Patrick Lee, et al. Gtm: A general time-series model for enhanced representation learning of time-series data. In *The Fourteenth International Conference on Learning Representations*, 2026.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 10148–10167, 2024.

Josif Grabocka, Nicolas Schilling, Martin Wistuba, and Lars Schmidt-Thieme. Learning time-series shapelets. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 392–401, 2014.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. In *International conference on learning representations*, volume 2024, pp. 23857–23880, 2024.

E. G. Kardakos, M. C. Alexiadis, S. I. Vagropoulos, C. K. Simoglou, P. N. Biskas, and A. G. Bakirtzis. Application of time series and artificial neural network models in short-term forecasting

of pv power generation. In *2013 48th International Universities' Power Engineering Conference (UPEC)*, pp. 1–6, 2013. doi: 10.1109/UPEC.2013.6714975.

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*, pp. 11116–11140, 2024a.

Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer: Generative pre-trained transformers are large time series models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 32369–32399, 2024b.

Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer-xl: Long-context transformers for unified time series forecasting. In *International Conference on Learning Representations*, volume 2025, pp. 83982–84006, 2025a.

Yong Liu, Guo Qin, Zhiyuan Shi, Zhi Chen, Caiyin Yang, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Sundial: a family of highly capable time series foundation models. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025b.

Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, pp. 689–696, 2009.

Abdullah Mueen, Eamonn Keogh, Qiang Zhu, Sydney Cash, and Brandon Westover. Exact discovery of time series motifs. In *Proceedings of the 2009 SIAM international conference on data mining*, pp. 473–484. SIAM, 2009.

Hung Nguyen and Christian K. Hansen. Short-term electricity load forecasting with time series analysis. In *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 214–221, 2017. doi: 10.1109/ICPHM.2017.7998331.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

Kanghui Ning, Zijie Pan, Yu Liu, Yushan Jiang, James Zhang, Kashif Rasul, Anderson Schneider, Lintao Ma, Yuriy Nevmyvaka, and Dongjin Song. Ts-rag: Retrieval-augmented generation based time series foundation models are stronger zero-shot forecaster. *Advances in Neural Information Processing Systems*, 38:163170–163199, 2026.

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 (ICLR)*, 2026.

Erik Otović, Marko Njirjak, Dario Jozinović, Goran Mauša, Alberto Michelini, and Ivan Stajduhar. Intra-domain and cross-domain transfer learning for time series data—how transferable are the features? *Knowledge-Based Systems*, 239:107976, 2022.

Patrick Schäfer. The boss is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, 29(6):1505–1530, 2015.

Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-moe: Billion-scale time series foundation models with mixture of experts. In *International conference on learning representations*, volume 2025, pp. 34635–34667, 2025.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Sabera J Talukder, Yisong Yue, and Georgia Gkioxari. TOTEM: TOkenized time series EMbeddings for general time series analysis. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, Lei Zhang, and Jianxin Liao. Chattime: A unified multimodal time series foundation model bridging numerical and textual data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12694–12702, 2025.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Forty-first International Conference on Machine Learning*, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430. Curran Associates, Inc., 2021.

Xingjian Wu, Jianxin Jin, Wanghui Qiu, Peng Chen, Yang Shu, Bin Yang, and Guo Guo. Aurora: Towards universal generative multimodal time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 103239–103277, 2026.

Qingren Yao, Chao-Han Huck Yang, Renhe Jiang, Yuxuan Liang, Ming Jin, and Shirui Pan. Towards neural scaling laws for time series foundation models. In *International Conference on Learning Representations*, volume 2025, pp. 83793–83820, 2025.

Lexiang Ye and Eamonn Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 947–956, 2009.

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.

Zhenwei Zhang, Jiawen Zhang, Shun Zheng, Yuantao Gu, and Jiang Bian. Does cross-domain pre-training truly help time-series foundation models? In *ICLR 2025 Workshop on Foundation Models in the Wild*, 2025.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Zican Dong, Yupeng Hou, Beichen Zhang, Yingqian Min, Junjie Zhang, Peiyu Liu, et al. A survey of large language models. *Frontiers of Computer Science*, 20(12):2012627, 2026.

A Experimental Details

A.1 Datasets

Forecasting Datasets. We evaluate UniTFU on seven widely used real-world forecasting benchmarks: ETTh1, ETTh2, ETTm1, ETTm2, Electricity (ECL), Weather, and Solar. These benchmarks differ in application domain, sampling frequency, and number of variables, enabling evaluation of forecasting performance and cross-domain generalization under heterogeneous temporal dynamics. The four ETT datasets contain seven electricity transformer variables, sampled hourly in ETTh1 and ETTh2 and every 15 minutes in ETTm1 and ETTm2. ECL records hourly electricity consumption from 321 clients. Weather contains 21 meteorological variables sampled every 10 minutes, while Solar records power generation from 137 photovoltaic plants at the same sampling frequency. Following the standard evaluation setting, we partition each dataset chronologically into training, validation, and test sets. The ETT datasets use a 12/4/4-month split, while ECL, Weather, and Solar use a $7:1:2$ split. Table 7 summarizes the dataset statistics.

Table 7: Statistics of the forecasting benchmarks.

Dataset	Domain	Variables	Frequency	Time Points
ETTh1	Electricity	7	1 hour	17,420
ETTh2	Electricity	7	1 hour	17,420
ETTm1	Electricity	7	15 min	69,680
ETTm2	Electricity	7	15 min	69,680
Electricity	Electricity	321	1 hour	26,304
Weather	Climate	21	10 min	52,696
Solar	Energy	137	10 min	52,560

Pre-training Datasets. We construct the pre-training corpus from UTSD (Liu et al., 2024b) and LOTSA (Woo et al., 2024). UTSD contains approximately one billion time points across seven domains, while LOTSA comprises over 27 billion observations across nine domains. Together, these collections span diverse application domains, sampling frequencies, and temporal dynamics, providing a heterogeneous corpus for learning reusable local temporal structures. To ensure a strict evaluation of generalization and avoid data leakage, all datasets used as target evaluation benchmarks are excluded from the pre-training corpus. This exclusion applies to UniTFU's pre-training corpus; the pre-training exposure of competing pretrained TSFMs follows the datasets documented in their original papers and released model descriptions.

A.2 Baselines

We compare UniTFU with representative TSFMs and supervised forecasting models. According to their roles in our experiments, we organize the baselines into three groups: TSFMs with general modeling paradigms, large-scale pre-trained TSFMs evaluated using released checkpoints, and task-trained forecasting models. This grouping reflects their experimental roles rather than mutually exclusive model categories.

TSFMs with general modeling paradigms. GTM (Cheng et al., 2026) uses Fourier attention to learn temporal-granularity-aware representations and combines reconstruction and autoregressive objectives through a unified pre-training strategy. Timer-XL (Liu et al., 2025a) extends next-token prediction to multivariate time series and introduces TimeAttention to capture intra-series and inter-series dependencies within a decoder-only Transformer. Timer (Liu et al., 2024b) represents time series as non-overlapping patch tokens and performs autoregressive forecasting through next-token prediction. These models serve as baselines for evaluating unified forecasting and controlled cross-domain generalization.

Large-scale pre-trained TSFMs. Aurora (Wu et al., 2026) integrates time series with multimodal information for generative forecasting across heterogeneous domains. Sundial (Liu et al., 2025b) models the conditional distribution of future patches in continuous space using TimeFlow Loss, supporting probabilistic zero-shot forecasting. TiRex (Auer et al., 2026) is an xLSTM-based pre-trained

forecasting model that uses contiguous patch masking to strengthen state tracking and support probabilistic forecasting over short and long horizons. Time-MoE (Shi et al., 2025) combines autoregressive forecasting with a sparse mixture-of-experts architecture to scale model capacity through selective expert activation. TimesFM (Das et al., 2024) is a patch-based decoder-only model pretrained for zero-shot time series forecasting. Chronos (Ansari et al., 2024) scales and quantizes continuous observations into discrete tokens and trains a language model to generate probabilistic forecasts. MOIRAI (Woo et al., 2024) accommodates heterogeneous sampling frequencies and variable dimensions through multiple patch sizes and any-variate attention, and uses a mixture distribution for probabilistic forecasting.

Task-trained forecasting models. PhaseFormer (Niu et al., 2026) uses compact phase representations and lightweight cross-phase routing for efficient forecasting. DLinear (Zeng et al., 2023) decomposes time series into trend and remainder components and predicts them using separate linear mappings. PatchTST (Nie et al., 2023) combines temporal patching with a channel-independent Transformer to model dependencies within each variable. iTransformer (Liu et al., 2024a) embeds the historical observations of each variable into a variate token and applies self-attention across variables to capture multivariate dependencies.

For reproduced baselines, we use the official implementations and follow the recommended configurations where available. All models in the supervised forecasting experiments follow the same standard rolling forecasting setting as UniTFU, using identical chronological data partitions and evaluation horizons. In controlled cross-domain experiments, the compared TSFMs are trained on the same source-domain data and evaluated on the corresponding target domains under the specified direct-transfer or adaptation setting. For pre-trained forecasting, we evaluate all competing TSFMs using their publicly released checkpoints. The lookback length is fixed to 672 for all models, and predictions are evaluated at horizons $\{96, 192, 336, 720\}$ using MSE and MAE on the same target test sets. Because the compared TSFMs are pre-trained on different corpora, pre-training overlap is determined according to the datasets documented in their original papers or released model descriptions. This comparison evaluates released pretrained models rather than controlling for pre-training data and compute.

A.3 Implementation Details

We implement UniTFU in PyTorch and conduct experiments on NVIDIA RTX 4090 GPUs. Table 8 summarizes the default model and training configurations. Unless explicitly varied in ablation or sensitivity experiments, the TFU-specific hyperparameters are shared across datasets without dataset-specific tuning. The prototype capacity is increased from 128 to 1024 for large-scale pretraining.

Training and Forecasting Details. For standard supervised forecasting, we use Adam with an initial learning rate of 1×10^{-4}. Following iTransformer (Liu et al., 2024a), we construct training targets by shifting the input window forward by one token. With an input length of $L = 672$ and a token length of 96, the input covers time steps 1-672, and the MSE is computed over outputs aligned with time steps 97-768. TFU representations and scene context are constructed from the full observed input window, while token-level attention remains causally masked. The intermediate outputs are therefore supervised on targets within the input window, whereas the final output token predicts the next 96 unobserved time steps. We train for up to 10 epochs using cosine annealing and select the checkpoint with the lowest validation MSE. The early-stopping patience is set to 10 epochs. The batch size is 32 for the four ETT datasets and Weather, 16 for Solar, and 4 for ECL. At inference, only the final output token is used for forecasting. Longer horizons are obtained by appending predicted tokens to the input without accessing future ground-truth observations. The input window is advanced until the requested horizon is covered, and the concatenated predictions are truncated to the required length when necessary. Direct forecasting baselines follow the same rolling forecasting procedure, while autoregressive-token models retain their native decoding interfaces.

Historical-Pattern Retrieval. Retrieval is performed independently for each sample within the current input window, without accessing an external database or other samples. Let $\widetilde{X} \in \mathbb{R}^{L \times C}$ denote the input after the model's input normalization. The query contains its last $Q = 96$ observations. Candidate segments have length Q and are extracted with a stride of one time step. A candidate is eligible only if its following $P = 96$ observations also lie within the current window. We select a

Table 8: Implementation settings of UniTFU. TFU-specific hyperparameters are shared across datasets unless explicitly varied in ablation or sensitivity experiments. The prototype capacity is 128 for non-pretrained experiments and 1024 for pre-training.

Category	Hyperparameter	Value
Forecasting	Input length L	672
	Input token length	96
	Output token length	96
	Supervised prediction horizons	$\{96, 192, 336, 720\}$
Backbone	Number of layers	2
	Hidden dimension D	1536
	FFN dimension	3172
	Number of attention heads	8
Scene Context	Interaction encoder layers	1
	Hidden dimension	1536
	Number of attention heads	4
	Scene-relation rank r	1
Optimization	Optimizer	Adam
	Standard initial learning rate	1×10^{-4}
	Pre-training initial learning rate	2×10^{-5}
	Pre-training effective global batch size	4096
	Maximum training epochs	10
	Forecasting objective	MSE
	Learning-rate schedule	Cosine annealing
	Early-stopping patience	10 epochs
TFU Learning	Temporal scales S	$\{1, 2, 4\}$ tokens
	τ_{intra}	0.92
	τ_{cross}	0.92
	τ_{merge}	0.82
	τ_{new}	$\tau_{\text{merge}} - 0.10 = 0.72$
	τ_{novel}	0.80
	Prototype consolidation threshold τ_{proto}	0.97
	Token-composition temperature τ_{temp}	0.50
	Overlap-ratio threshold	0.50
	Maximum cross-scale candidates per variable	16
	Local redundancy window	3
	EMA momentum η	0.95
Dictionary	Maximum dictionary capacity (non-pretrained)	128
	Maximum dictionary capacity (pre-training)	1024
	Minimum novelty support $n_{\min}$	6
	Maximum births per maintenance step	1

shared starting position across variables by minimizing the mean squared distance:

$$s^* = \underset{1 \le s \le L-Q-P+1}{\arg\min} \frac{1}{QC} \sum_{j=0}^{Q-1} \sum_{c=1}^{C} \left(\widetilde{X}_{s+j,c} - \widetilde{X}_{L-Q+1+j,c} \right)^2, \tag{10}$$

$$(Y_{\text{h}})_{r,c} = \widetilde{X}_{s^*+Q+r-1,c}, \qquad 1 \le r \le P,\ 1 \le c \le C, \qquad Y_{\text{h}} \in \mathbb{R}^{P \times C}.$$

The constraint ensures that the retrieved continuation ends no later than step L. The retrieved continuation may overlap the query, but never accesses observations beyond the current window. Retrieval and fusion are performed in the normalized space, followed by the model's inverse normalization. The historical-pattern residual is applied only to the final output token during training and inference. During rolling forecasting, retrieval is repeated on the current window, which may contain previously predicted tokens; future ground-truth observations are never appended.

TFU Learning Configuration. Candidate segments are extracted at temporal scales $\mathcal{S} = \{1, 2, 4\}$, corresponding to spans of 96, 192, and 384 observations. The maximum dictionary capacity is $K = 128$ for all non-pretrained experiments. Candidates are extracted with a stride of one token,

with at most eight candidates retained per variable at each scale before redundancy suppression. The dictionary maintains an adaptive number of active TFUs, denoted by $K_{\text{active}} \leq K$. New prototypes occupy available slots, while prototype consolidation can reduce the active count; the dictionary is not required to contain K active TFUs. Redundancy suppression uses $\tau_{\text{intra}} = \tau_{\text{cross}} = 0.92$ and an overlap-ratio threshold of 0.5, with a local redundancy window of 3 and at most 16 retained cross-scale candidates per variable. For prototype routing, candidates with best-match cosine similarity at least $\tau_{\text{merge}} = 0.82$ enter the merge buffer, while those below $\tau_{\text{new}} = 0.72$ enter the novelty buffer. The two routing thresholds are coupled by the fixed margin $\tau_{\text{new}} = \tau_{\text{merge}} - 0.10$, rather than tuned independently. Candidates with best-match similarity in $[\tau_{\text{new}}, \tau_{\text{merge}})$ do not contribute to dictionary updates. Novelty grouping uses $\tau_{\text{novel}} = 0.80$. A group must contain at least $n_{\min} = 6$ candidates to initialize a new TFU, subject to available dictionary capacity and a limit of one birth per maintenance step. Existing TFUs are refined through grouped EMA updates with momentum $\eta = 0.95$. Prototype pairs with cosine similarity at least $\tau_{\text{proto}} = 0.97$ are eligible for consolidation to reduce dictionary redundancy. This prototype-to-prototype operation is distinct from candidate routing into the merge buffer. Token-level TFU composition uses a similarity temperature of $\tau_{\text{temp}} = 0.5$.

Gradient Propagation. Nearest-prototype assignment is discrete, but the selected cosine similarities remain differentiable and determine the token-level composition weights. Forecasting gradients can therefore reach the segment encoder through these weights, with the retained candidate set and prototype indices treated as fixed during backpropagation. This path contributes zero gradient at positions covered by a single candidate or by candidates assigned to the same prototype. Encoder learning through this path thus depends on overlapping candidates with different assigned prototypes. The prototypes are maintained as non-gradient buffers using detached candidate embeddings. No straight-through estimator or additional commitment or reconstruction loss is used.

Dictionary Maintenance and Evaluation. During training, retained candidates are matched to active TFUs using cosine similarity. Dictionary maintenance enables grouped EMA updates, novelty birth, and prototype consolidation. Candidates in the merge buffer refine their matched prototypes, supported groups in the novelty buffer can initialize new prototypes, and sufficiently similar prototypes can be consolidated. All dictionary maintenance operations are disabled during validation and testing, so the learned dictionary remains fixed throughout evaluation. In direct source-to-target transfer, the source-trained dictionary is reused without modification by target evaluation samples. During target-domain adaptation, Frozen-TFU keeps both the source dictionary and the backbone fixed and updates only the forecasting head. Updated-TFU keeps the backbone fixed while adapting the forecasting head and enabling all dictionary maintenance operations on the available target-domain training data, including prototype updates, novelty birth, and prototype consolidation. Thus, Updated-TFU adapts both the prototype representations and the dictionary structure, rather than only refining a fixed set of source prototypes. Its comparison with Frozen-TFU evaluates the contribution of adaptive dictionary maintenance under the same target-data budget. The adapted dictionary is subsequently frozen for target-domain evaluation.

Pre-training Configuration. We pre-train UniTFU on the combined UTSD and LOTSA corpus after excluding all target evaluation datasets. The backbone architecture, tokenization scheme, temporal scales, and TFU-specific thresholds remain unchanged. The dictionary capacity is increased to $K = 1024$ to accommodate the more diverse pre-training corpus. Pre-training uses Adam with an MSE objective, an initial learning rate of 2×10^{-5}, an effective global batch size of 4096, and cosine annealing for at most 10 epochs. The resulting checkpoint is used in the pre-trained forecasting evaluation.

B Full Cross-Domain Forecasting Results

B.1 Direct Transfer

Table 9 presents the complete direct transfer results across ETTh1, ETTm1, Weather, ECL, and Solar under the $672 \rightarrow 96$ forecasting setting. For each target domain, we report the results of models trained separately on each of the remaining four source domains. All model parameters and, when applicable, dictionary or codebook states remain fixed during target-domain evaluation. Implementation details are provided in Appendix A.3. Averaging the results over the four source domains for each target domain gives the aggregated results in Table 1. UniTFU achieves the lowest

Table 9: Complete direct transfer results under the 672 → 96 forecasting setting. Each row reports the performance of a model trained on the specified source domain and evaluated on the target domain without adaptation. **Red**: the best, **Blue**: the 2nd best.

Target	Source	UniTFU		w/ VQ		w/o TFU		GTM		Timer-XL		Timer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	ETTm1	**0.476**	**0.460**	**0.493**	**0.473**	0.516	0.481	0.581	0.492	0.650	0.558	0.509	0.489
	Weather	**0.667**	**0.550**	**0.682**	**0.554**	0.695	0.566	0.707	0.563	0.723	0.583	0.698	0.563
	ECL	**0.358**	**0.392**	**0.369**	**0.396**	**0.369**	**0.392**	0.482	0.475	0.402	0.418	0.378	0.399
	Solar	**0.645**	**0.558**	**0.699**	**0.580**	0.951	0.670	0.841	0.776	0.766	0.605	0.954	0.682
ETTm1	ETTh1	**0.636**	**0.530**	**0.661**	**0.543**	0.783	0.556	0.857	0.595	0.748	0.562	0.794	0.568
	Weather	1.268	0.702	1.264	0.701	0.940	0.608	**0.803**	**0.554**	**0.858**	**0.605**	0.935	0.610
	ECL	**0.656**	**0.535**	**0.746**	**0.555**	0.890	0.608	0.971	0.680	0.868	0.594	0.902	0.613
	Solar	**0.956**	**0.638**	**1.003**	**0.668**	1.776	0.873	1.155	0.750	1.103	0.695	1.786	0.885
Weather	ETTh1	0.215	0.273	0.221	0.279	**0.198**	**0.255**	0.314	0.347	0.223	0.284	**0.209**	**0.269**
	ETTm1	**0.224**	**0.276**	0.227	0.278	**0.222**	**0.273**	0.307	0.336	0.240	0.296	0.230	0.283
	ECL	**0.215**	**0.270**	**0.223**	**0.272**	0.281	0.318	0.347	0.367	0.270	0.319	0.232	0.274
	Solar	**0.324**	**0.330**	**0.330**	**0.337**	0.368	0.361	0.335	0.353	0.375	0.354	0.404	0.374
ECL	ETTh1	**0.211**	**0.302**	**0.214**	**0.303**	0.220	0.306	0.348	0.426	0.259	0.353	0.217	0.305
	ETTm1	0.646	0.584	0.661	0.601	**0.385**	**0.465**	0.664	0.608	0.620	0.602	**0.397**	**0.475**
	Weather	**0.688**	**0.663**	0.697	0.670	0.825	0.750	**0.583**	**0.566**	0.807	0.734	0.836	0.755
	Solar	**0.772**	**0.692**	0.832	0.704	0.990	0.780	1.013	0.769	**0.763**	**0.680**	0.999	0.786
Solar	ETTh1	**0.790**	**0.665**	**0.808**	**0.674**	0.860	0.682	0.841	0.686	0.878	0.707	0.853	0.697
	ETTm1	0.742	0.627	0.748	0.639	**0.600**	0.564	0.669	**0.536**	0.966	0.727	**0.615**	**0.557**
	Weather	**0.396**	**0.452**	**0.403**	**0.453**	0.408	0.461	0.459	0.458	0.495	0.503	0.419	0.456
	ECL	**0.551**	**0.505**	**0.669**	**0.594**	0.932	0.691	1.434	0.932	1.108	0.827	0.938	0.697
Average		**0.572**	**0.500**	**0.598**	**0.514**	0.660	0.533	0.686	0.563	0.656	0.550	0.665	0.537

average MSE and MAE on all five target domains and the best results on both metrics in 13 of the 20 direct transfer settings. Compared with w/o TFU and w/ VQ, UniTFU reduces overall average MSE by 13.4% and 4.3%, respectively, supporting the contribution of TFU modeling and the adaptive dictionary design to direct cross-domain transfer. Performance nevertheless varies with the source domain. For example, UniTFU underperforms the baselines when transferring from Weather to ETTm1, indicating that its overall advantage does not extend to every transfer setting.

B.2 Target Domain Adaptation

Table 10 presents the complete target-domain adaptation results under the 672 → 96 forecasting setting. For each target domain, models trained separately on the remaining four source domains are adapted using 10% of the target-domain training samples. We report standard few-shot results for Timer, Timer-XL, GTM, and UniTFU, together with Frozen-TFU and Updated-TFU. Frozen-TFU updates only the forecasting head, whereas Updated-TFU additionally updates the TFU dictionary; both keep the remaining backbone fixed. All model parameters and TFU prototypes remain fixed during subsequent target-domain evaluation. Implementation details are provided in Appendix A.3. Averaging the results over the four source domains for each target domain gives the aggregated results in Table 2. Updated-TFU reduces MSE in 18 of the 20 adaptation settings and MAE in 19 compared with Frozen-TFU, yielding overall reductions of 3.7% and 3.9%, respectively. These results support the refinement of reusable temporal structures with limited target-domain data. On ETTm1, Updated-TFU outperforms standard few-shot UniTFU on both metrics for every source domain. On ECL and Solar, however, it improves over Frozen-TFU but remains behind standard few-shot UniTFU and Timer-XL for every source domain. These results indicate that the benefits of TFU dictionary adaptation vary across target domains and that some targets require broader model adaptation.

Table 10: Complete target-domain adaptation results using 10% of the target-domain training samples under the $672 \rightarrow 96$ forecasting setting. Frozen-TFU and Updated-TFU are defined in the text. Average reports the mean across all 20 adaptation settings. **Red**: the best, **Blue**: the 2nd best.

Target	Source	10% Few-shot						Frozen-TFU		Updated-TFU		10% Few-shot	
		Timer		Timer-XL		GTM		UniTFU		UniTFU		UniTFU	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	ETTm1	**0.446**	0.456	0.482	0.481	0.518	0.474	0.447	**0.448**	**0.445**	**0.445**	0.458	0.463
	Weather	0.594	0.535	0.581	0.528	**0.522**	0.503	0.566	0.513	0.557	**0.502**	**0.490**	**0.488**
	ECL	0.378	0.410	0.391	0.413	0.418	0.430	**0.355**	**0.392**	0.362	0.399	**0.340**	**0.370**
	Solar	0.616	0.537	**0.510**	**0.487**	0.644	0.552	0.559	0.513	0.546	0.500	**0.499**	**0.491**
ETTm1	ETTh1	0.557	0.542	0.628	0.500	0.704	0.500	**0.426**	**0.423**	**0.421**	**0.422**	0.523	0.464
	Weather	0.556	0.478	0.493	0.460	0.578	0.478	**0.446**	**0.435**	**0.433**	**0.426**	0.606	0.481
	ECL	0.651	0.537	0.552	0.473	0.752	0.529	**0.436**	**0.428**	**0.440**	**0.424**	0.547	0.477
	Solar	0.534	0.470	0.605	0.497	**0.465**	**0.441**	0.478	0.448	**0.464**	**0.444**	0.543	0.472
Weather	ETTh1	**0.176**	**0.227**	0.177	0.233	0.179	0.231	0.188	0.247	0.184	0.243	**0.174**	**0.230**
	ETTm1	**0.171**	**0.219**	0.178	0.230	**0.171**	0.222	0.184	0.242	0.181	0.239	**0.172**	**0.215**
	ECL	**0.172**	**0.216**	**0.172**	**0.220**	0.187	0.235	0.192	0.248	0.185	0.242	**0.183**	**0.220**
	Solar	0.183	0.233	**0.178**	**0.229**	**0.178**	0.232	0.197	0.254	0.194	0.250	**0.180**	**0.226**
ECL	ETTh1	0.173	0.282	**0.171**	**0.264**	0.183	0.280	0.217	0.315	0.211	0.308	**0.165**	**0.259**
	ETTm1	**0.167**	**0.267**	0.190	0.283	0.188	0.285	0.232	0.332	0.215	0.315	**0.164**	**0.259**
	Weather	**0.154**	**0.253**	**0.154**	**0.250**	0.187	0.284	0.233	0.333	0.217	0.317	**0.162**	0.258
	Solar	**0.168**	0.267	0.170	**0.265**	0.195	0.292	0.256	0.354	0.223	0.322	**0.163**	**0.259**
Solar	ETTh1	**0.199**	0.267	**0.180**	**0.237**	0.209	**0.241**	0.266	0.346	0.240	0.315	0.202	0.249
	ETTm1	**0.192**	0.256	**0.192**	0.252	0.203	**0.237**	0.250	0.326	0.228	0.292	**0.201**	**0.234**
	Weather	**0.187**	0.248	0.191	**0.230**	**0.189**	**0.230**	0.241	0.307	0.223	0.274	0.216	**0.244**
	ECL	0.194	0.246	**0.188**	**0.240**	0.213	0.245	0.280	0.338	0.240	0.283	**0.190**	**0.235**
Average		0.323	0.347	0.319	**0.339**	0.344	0.346	0.322	0.362	**0.310**	0.348	**0.309**	**0.330**

C Full Multivariate Time Series Forecasting Results

Table 11 presents the complete results under standard supervised multivariate rolling forecasting. The lookback length is 672, and the prediction lengths are $\{96, 192, 336, 720\}$. For each dataset, we report MSE and MAE at each prediction length and their averages over the four prediction lengths. Implementation details are provided in Appendix A.3. UniTFU achieves the lowest average MSE and MAE on ETTh1, ETTh2, ETTm1, ECL, and Solar, and the lowest average MAE on Weather. It obtains the best MSE in 13 of the 28 forecasting settings and the best MAE in 17. On ECL, UniTFU achieves the lowest MSE at all four prediction lengths, while on Solar it achieves the lowest MSE at three. However, it remains behind several baselines on ETTm2, indicating that its improvements are not uniform across datasets.

D Pretrained Time Series Forecasting

Table 12 presents the complete pretrained forecasting results with a lookback length of 672 and prediction horizons $\{96, 192, 336, 720\}$. For each dataset, we report MSE and MAE at each prediction horizon and their averages over the four horizons. Pre-training and evaluation details are provided in Appendix A.3. Datasets included in a model's pre-training data are excluded from its evaluation and marked with a dash. UniTFU achieves the lowest average MSE on five datasets and the lowest average MAE on four. In particular, it achieves the lowest MSE at all four prediction horizons on ETTh1, ETTm1, ETTm2, and Weather. Across the 28 forecasting settings, UniTFU obtains the lowest MSE in 20 settings and the lowest MAE in 16. Performance nevertheless varies across datasets and metrics: TiRex achieves lower average MSE and MAE on ETTh2 and ECL, while TimesFM and Sundial achieve lower average MAE on Solar.

Table 11: Complete multivariate rolling forecasting results with a lookback length of 672 and prediction lengths $\{96, 192, 336, 720\}$. **Red**: the best, **Blue**: the 2nd best.

Type		TSFM								Forecasting Model							
Models		**UniTFU**		GTM		Timer-XL		Timer		PhaseFormer		iTransformer		DLinear		PatchTST	
Dataset	H	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.360**	**0.396**	**0.362**	**0.397**	0.365	0.399	0.371	0.402	0.374	**0.397**	0.382	0.413	0.373	0.402	0.376	0.405
	192	**0.392**	**0.418**	**0.389**	**0.419**	0.407	0.427	0.410	0.427	0.411	0.420	0.413	0.432	0.411	0.427	0.407	0.427
	336	**0.410**	**0.430**	**0.406**	**0.434**	0.429	0.443	0.434	0.443	0.435	0.435	0.432	0.445	0.446	0.454	0.420	0.438
	720	**0.423**	**0.446**	0.433	0.465	0.436	0.461	0.458	0.459	0.451	**0.456**	0.447	0.470	0.515	0.522	**0.424**	**0.456**
	Avg.	**0.396**	**0.423**	**0.398**	0.429	0.409	0.433	0.418	0.433	0.418	**0.427**	0.419	0.440	0.436	0.451	0.407	0.432
ETTh2	96	0.297	**0.350**	**0.292**	**0.350**	0.299	0.353	**0.286**	**0.345**	0.323	0.381	0.329	0.377	0.308	0.373	**0.292**	0.351
	192	**0.360**	**0.394**	0.368	0.402	0.367	**0.400**	**0.364**	**0.400**	0.367	0.408	0.391	0.418	0.414	0.441	0.371	**0.400**
	336	**0.384**	**0.420**	0.400	0.432	0.393	0.427	0.409	0.440	**0.382**	**0.422**	0.412	0.436	0.528	0.508	0.405	0.432
	720	0.422	0.454	0.438	0.465	0.444	0.470	0.455	0.482	**0.403**	**0.442**	0.448	0.463	0.830	0.653	**0.419**	**0.453**
	Avg.	**0.366**	**0.405**	0.375	0.412	0.376	0.413	0.379	0.417	**0.369**	0.413	0.395	0.424	0.520	0.494	0.372	**0.409**
ETTm1	96	**0.272**	**0.327**	0.285	**0.334**	0.296	0.351	**0.282**	0.338	0.305	0.349	0.325	0.373	0.306	0.349	0.299	0.357
	192	**0.315**	**0.358**	**0.326**	**0.360**	0.346	0.381	0.327	0.366	0.337	0.370	0.365	0.397	0.336	0.367	0.340	0.381
	336	**0.350**	**0.382**	0.361	**0.383**	0.385	0.404	**0.360**	0.387	0.367	0.390	0.397	0.416	0.365	0.385	0.376	0.402
	720	**0.412**	**0.418**	**0.411**	**0.418**	0.449	0.441	0.418	0.420	0.429	0.424	0.460	0.449	0.417	**0.416**	0.438	0.437
	Avg.	**0.337**	**0.371**	**0.346**	**0.374**	0.369	0.394	0.347	0.378	0.360	0.383	0.387	0.409	0.356	0.379	0.363	0.394
ETTm2	96	0.182	0.265	**0.168**	**0.256**	0.184	0.270	**0.167**	**0.255**	0.177	0.265	0.189	0.276	**0.167**	0.262	0.173	0.265
	192	0.246	0.310	0.230	**0.300**	0.242	0.309	**0.226**	**0.295**	0.238	0.304	0.259	0.320	0.230	0.312	**0.229**	0.301
	336	0.304	0.351	0.287	0.340	0.297	0.343	**0.279**	**0.332**	0.291	0.339	0.321	0.360	0.300	0.362	**0.280**	**0.334**
	720	0.408	0.417	0.378	0.398	0.380	0.396	**0.369**	**0.391**	0.376	0.394	0.415	0.418	0.437	0.449	**0.368**	**0.390**
	Avg.	0.285	0.336	0.266	0.324	0.276	0.330	**0.260**	**0.318**	0.271	0.326	0.296	0.344	0.284	0.346	**0.263**	**0.323**
ECL	96	**0.132**	**0.229**	0.141	0.241	0.135	**0.230**	0.137	0.232	0.138	0.232	**0.133**	**0.230**	0.135	0.233	0.134	0.234
	192	**0.150**	**0.245**	0.160	0.259	0.153	0.247	0.153	**0.246**	**0.152**	**0.245**	**0.152**	**0.245**	0.154	0.247	0.153	0.253
	336	**0.166**	**0.263**	0.178	0.278	0.173	0.267	0.171	0.265	0.168	**0.261**	0.169	**0.263**	**0.167**	0.264	0.175	0.276
	720	**0.207**	**0.299**	0.214	0.308	0.215	0.304	0.214	0.302	0.212	0.303	**0.208**	**0.299**	0.209	**0.298**	0.229	0.325
	Avg.	**0.164**	**0.259**	0.173	0.272	0.169	0.262	0.169	0.261	0.168	**0.260**	**0.166**	**0.259**	**0.166**	0.261	0.173	0.272
Weather	96	0.158	**0.206**	0.161	0.212	0.162	0.213	**0.155**	**0.204**	0.159	0.216	0.162	0.216	0.169	0.228	**0.154**	0.207
	192	**0.203**	**0.248**	0.208	0.253	0.209	0.257	0.204	0.250	**0.203**	0.256	0.208	0.257	0.212	0.268	**0.200**	**0.249**
	336	0.256	**0.288**	0.261	0.292	0.263	0.298	0.264	0.296	**0.253**	0.295	0.259	0.296	0.257	0.307	**0.251**	**0.289**
	720	**0.324**	**0.336**	0.330	**0.340**	0.333	0.346	0.349	0.352	0.325	0.345	0.327	0.342	**0.320**	0.361	**0.320**	**0.340**
	Avg.	**0.235**	**0.270**	0.240	0.274	0.242	0.279	0.243	0.276	**0.235**	0.278	0.239	0.278	0.240	0.291	**0.231**	**0.271**
Solar	96	**0.182**	**0.220**	0.193	**0.221**	0.186	0.238	0.192	0.243	0.193	0.260	0.192	0.258	0.192	0.257	**0.185**	0.278
	192	**0.202**	**0.246**	0.220	**0.239**	0.213	0.258	0.215	0.261	0.211	0.270	0.217	0.282	0.213	0.274	**0.201**	0.288
	336	**0.215**	**0.257**	0.247	**0.257**	0.231	**0.271**	0.231	0.274	0.226	0.279	0.226	0.300	0.233	0.290	**0.217**	0.298
	720	**0.235**	**0.276**	0.262	0.283	0.244	**0.281**	0.241	0.285	**0.235**	0.282	0.240	0.320	0.245	0.306	**0.238**	0.311
	Avg.	**0.209**	**0.250**	0.231	**0.250**	0.219	**0.262**	0.220	0.266	0.216	0.273	0.219	0.290	0.221	0.282	**0.210**	0.294
1st Count		13	17	3	2	0	0	4	5	3	3	0	1	2	2	6	1

E HYPERPARAMETER SENSITIVITY

We evaluate the sensitivity of UniTFU to the main hyperparameters involved in TFU discovery and dictionary maintenance on ETTh1, Weather, and Solar under the $672 \rightarrow 96$ forecasting setting. As shown in Figure 5, we vary six factors individually: the temporal scale set $\mathcal{S}$, the intra-scale and cross-scale redundancy thresholds τ_{intra} and τ_{cross}, the candidate routing threshold τ_{merge}, the novelty grouping threshold τ_{novel}, and the prototype consolidation threshold τ_{proto}. For each factor, all other hyperparameters are fixed to the reference configuration described in Appendix A.3. For candidate routing, we maintain the fixed margin $\tau_{\text{new}} = \tau_{\text{merge}} - 0.10$ rather than tuning the two

Table 12: Complete pretrained forecasting results with a lookback length of 672 and prediction horizons $\{96, 192, 336, 720\}$. A dash indicates exclusion because the dataset is included in the corresponding model's pre-training data. **Red**: the best, **Blue**: the 2nd best.

Models		**UniTFU**		Aurora		Sundial		TiRex		Time-MoE		MOIRAI		TimesFM		Chronos	
Dataset	H	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.341**	**0.376**	**0.342**	0.386	0.343	0.379	0.379	**0.377**	0.386	0.398	0.394	0.396	0.373	0.381	0.513	0.428
	192	**0.374**	**0.398**	0.378	0.411	**0.376**	**0.400**	0.428	0.406	0.441	0.420	0.428	0.418	0.414	0.405	0.578	0.464
	336	**0.389**	**0.408**	0.401	0.426	**0.391**	**0.410**	0.463	0.424	0.476	0.442	0.447	0.428	0.438	0.418	0.612	0.488
	720	**0.416**	**0.431**	0.425	0.449	**0.418**	0.433	0.479	0.460	0.498	0.474	0.534	0.523	0.425	**0.429**	0.611	0.510
	Avg.	**0.380**	**0.403**	0.387	0.418	**0.382**	**0.406**	0.437	0.417	0.450	0.434	0.451	0.441	0.413	0.408	0.579	0.473
ETTh2	96	0.299	0.337	**0.273**	**0.334**	**0.282**	**0.334**	0.296	**0.326**	0.308	0.350	0.316	0.347	0.304	0.337	0.354	0.373
	192	0.355	0.378	**0.339**	0.377	**0.343**	**0.376**	0.353	**0.368**	0.364	0.394	0.383	0.392	0.371	0.383	0.429	0.418
	336	0.371	0.400	0.376	0.408	**0.369**	**0.398**	**0.354**	**0.379**	0.388	0.401	0.412	0.416	0.400	0.409	0.476	0.448
	720	**0.368**	**0.401**	0.401	0.436	0.390	0.422	**0.370**	**0.403**	0.410	0.436	0.508	0.494	0.405	0.422	0.516	0.476
	Avg.	0.348	**0.379**	0.347	0.389	**0.346**	0.383	**0.343**	**0.369**	0.368	0.395	0.405	0.412	0.370	0.388	0.444	0.429
ETTm1	96	**0.290**	**0.324**	0.316	0.371	**0.292**	**0.326**	0.322	0.328	0.336	0.345	0.433	0.389	0.318	0.329	0.525	0.438
	192	**0.338**	**0.356**	0.382	0.409	**0.340**	**0.358**	0.385	0.361	0.396	0.373	0.469	0.413	0.368	**0.358**	0.616	0.484
	336	**0.375**	**0.380**	0.447	0.441	**0.377**	**0.382**	0.429	0.385	0.437	0.400	0.517	0.439	0.395	**0.380**	0.678	0.516
	720	**0.442**	**0.418**	0.550	0.488	**0.444**	0.420	0.500	0.423	0.510	0.434	0.532	0.610	0.455	**0.414**	0.756	0.553
	Avg.	**0.361**	**0.370**	0.424	0.427	**0.363**	**0.372**	0.409	0.374	0.420	0.388	0.488	0.463	0.384	**0.370**	0.644	0.498
ETTm2	96	**0.177**	**0.246**	0.186	0.275	**0.179**	0.256	0.189	0.248	0.208	0.272	0.209	0.277	**0.179**	**0.244**	0.243	0.305
	192	**0.241**	**0.293**	0.255	0.321	**0.243**	0.300	0.261	0.295	0.274	0.318	0.297	0.329	0.252	**0.294**	0.323	0.355
	336	**0.294**	**0.331**	0.321	0.361	**0.296**	0.336	0.317	**0.333**	0.334	0.348	0.411	0.381	0.319	0.336	0.398	0.397
	720	**0.374**	**0.384**	0.425	0.420	**0.376**	0.388	0.395	**0.386**	0.410	0.401	0.532	0.477	0.421	0.395	0.525	0.464
	Avg.	**0.272**	**0.314**	0.297	0.344	**0.274**	0.320	0.291	**0.316**	0.307	0.335	0.362	0.366	0.293	0.317	0.372	0.380
ECL	96	0.124	**0.211**	0.140	0.240	**0.111**	**0.208**	**0.122**	**0.208**	–	–	0.168	0.267	–	–	0.209	0.273
	192	0.143	0.229	0.160	0.258	**0.127**	**0.225**	**0.141**	**0.226**	–	–	0.191	0.281	–	–	0.233	0.296
	336	0.171	0.258	0.182	0.280	**0.145**	**0.244**	**0.159**	**0.245**	–	–	0.214	0.312	–	–	0.267	0.325
	720	**0.195**	**0.280**	0.234	0.324	**0.197**	0.289	0.202	**0.282**	–	–	0.269	0.351	–	–	0.342	0.380
	Avg.	0.158	0.245	0.179	0.276	**0.145**	**0.242**	**0.156**	**0.240**	–	–	0.211	0.303	–	–	0.263	0.319
Weather	96	**0.155**	**0.190**	0.185	0.225	**0.157**	0.204	0.161	**0.192**	0.180	0.216	0.218	0.225	–	–	0.227	0.237
	192	**0.207**	**0.236**	0.233	0.270	**0.209**	0.252	**0.209**	**0.238**	0.230	0.268	0.290	0.274	–	–	0.290	0.290
	336	**0.262**	**0.278**	0.299	0.316	**0.264**	0.292	0.266	**0.280**	0.279	0.308	0.356	0.328	–	–	0.350	0.334
	720	**0.334**	**0.328**	0.384	0.370	**0.336**	0.342	0.340	**0.330**	0.349	0.357	0.547	0.454	–	–	0.429	0.388
	Avg.	**0.240**	**0.258**	0.275	0.295	**0.242**	0.273	0.244	**0.260**	0.260	0.287	0.353	0.320	–	–	0.324	0.312
Solar	96	**0.163**	**0.208**	**0.144**	**0.202**	0.165	0.210	0.300	0.252	0.309	0.271	0.686	0.673	0.213	0.209	0.686	0.690
	192	**0.187**	**0.227**	**0.182**	0.231	0.189	**0.229**	0.341	0.275	0.360	0.285	0.699	0.697	0.245	**0.229**	0.706	0.296
	336	**0.208**	0.284	0.232	0.266	**0.210**	**0.247**	0.346	0.282	0.357	0.292	0.728	0.710	0.281	**0.251**	0.723	0.707
	720	**0.245**	0.290	0.305	0.319	**0.247**	**0.274**	0.352	0.288	0.367	0.299	0.764	0.726	0.296	**0.263**	0.761	0.729
	Avg.	**0.201**	0.252	0.216	0.255	**0.203**	**0.240**	0.335	0.274	0.348	0.287	0.719	0.702	0.259	**0.238**	0.719	0.606
1st Count		20	16	4	1	3	4	1	4	0	0	0	0	0	5	0	0

thresholds independently. All configurations are trained using the same random seed, and MSE is reported. Across the three datasets, the reference configuration consistently attains the lowest MSE among the tested settings. For the temporal scale set, progressively extending $\mathcal{S}$ from $\{1\}$ to $\{1, 2, 4\}$ improves forecasting accuracy, whereas further introducing scale 6 leads to a moderate degradation. This indicates that combining local dynamics at multiple temporal scales is beneficial, while excessively long candidate segments do not provide additional gains under the current forecasting context. The threshold analyses show a similar pattern. For redundancy suppression, both τ_{intra} and τ_{cross} achieve their lowest errors at 0.92, while either more permissive or more restrictive filtering degrades performance. The effect is particularly visible for τ_{cross}, where lowering the threshold to 0.88 consistently increases the error across all three datasets. For TFU dictionary maintenance, the reference values $\tau_{\text{merge}} = 0.82$, $\tau_{\text{novel}} = 0.80$, and $\tau_{\text{proto}} = 0.97$ also yield the lowest MSE within their respective tested ranges. Moderate threshold perturbations yield relatively small MSE changes within the tested ranges. Overall, the same reference configuration performs

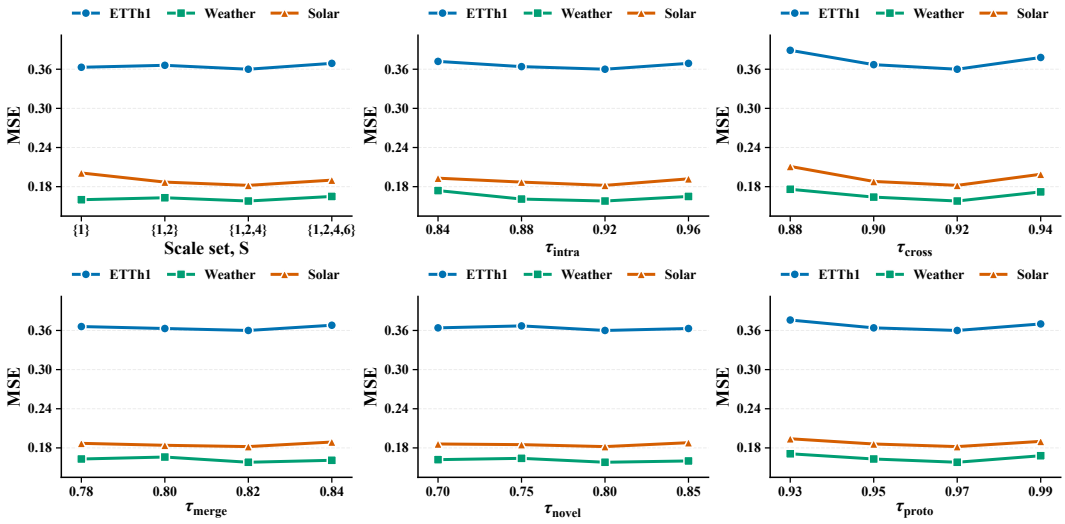


Figure 5: Hyperparameter sensitivity of UniTFU on ETTh1, Weather, and Solar under the $672 \rightarrow 96$ forecasting setting. The top row varies the temporal scale set $\mathcal{S}$, the intra-scale redundancy threshold τ_{intra}, and the cross-scale redundancy threshold τ_{cross}. The bottom row varies the candidate routing threshold τ_{merge}, the novelty grouping threshold τ_{novel}, and the prototype consolidation threshold τ_{proto}. We maintain $\tau_{\text{new}} = \tau_{\text{merge}} - 0.10$ and fix all other hyperparameters.

consistently across datasets with different temporal characteristics, supporting our use of a shared set of TFU-specific hyperparameters rather than dataset-specific tuning.

F Full Ablation Study Results

Table 13 reports the individual ablation results on all seven datasets under the $672 \rightarrow 96$ forecasting setting, using seeds $\{2021, 2023, 2025, 2027\}$. Seed 2021 is used for the benchmark experiments. Each row reports the MSE and MAE from one run. The variants are defined in Section 4.5.

VQ Codebook Replacement. The w/ VQ variant replaces the adaptive TFU dictionary with a fixed-size codebook while retaining the same multi-scale candidate encoding, redundancy suppression, token-level TFU composition, scene context aggregation, TFU attention, forecasting head, and historical-pattern residual. The codebook contains $K = 128$ codewords with the same embedding dimension as the TFU prototypes, matching the maximum capacity of the adaptive dictionary rather than its number of active TFUs. Codewords are initialized from the first non-empty batch of retained training candidates. When at least K candidates are available, K candidates are sampled without replacement; otherwise, all available candidates are retained and additional candidates are sampled with replacement to fill the codebook.

Each retained candidate is assigned to the codeword with the highest cosine similarity, without confidence-based routing. During training, assigned candidate embeddings are stored in an update buffer that retains at most the 4096 most recent assignments. At each codebook maintenance step, embeddings assigned to the same codeword are averaged, and the corresponding codeword is updated by grouped EMA using the same momentum $\eta = 0.95$ as the adaptive dictionary. Codewords without buffered assignments remain unchanged. Candidate codeword matching follows the same normalization setting as the adaptive dictionary, and the resulting codeword representations are used by the same token-level composition procedure. Unlike the adaptive TFU dictionary, the VQ codebook has a fixed size and does not perform confidence-based routing, novelty birth, or prototype consolidation. Codewords are maintained as non-gradient buffers rather than learnable parameters, and no additional VQ commitment or reconstruction loss is introduced.

Table 13: Complete ablation results under the $672 \rightarrow 96$ forecasting setting. Each row reports results for one dataset and random seed.

Dataset	Seed	UniTFU		w/o TFU		w/o Sce		w/o Att		w/o Res		w/ VQ	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	2021	**0.360**	**0.396**	0.366	0.404	0.363	0.399	0.365	0.400	0.367	0.399	0.364	0.401
	2023	0.362	0.399	0.369	0.409	**0.360**	**0.395**	0.371	0.405	0.372	0.403	0.364	0.406
	2025	**0.361**	**0.399**	0.363	0.401	0.366	0.402	0.368	0.405	0.371	0.403	0.365	0.401
	2027	0.364	0.402	0.370	0.405	**0.361**	0.400	0.373	**0.399**	0.375	0.406	0.367	0.402
ETTh2	2021	**0.297**	**0.350**	0.299	0.353	0.300	0.356	0.303	0.358	0.301	0.354	0.298	0.352
	2023	**0.294**	**0.346**	0.303	0.358	0.304	0.360	0.297	0.350	0.306	0.360	0.301	0.355
	2025	**0.299**	**0.351**	0.304	0.357	0.304	0.360	0.307	0.363	0.304	0.357	0.302	0.355
	2027	0.302	**0.354**	0.301	0.355	**0.299**	**0.354**	0.311	0.366	0.310	0.362	0.304	0.360
ETTm1	2021	**0.272**	**0.327**	0.287	0.335	0.279	0.331	0.274	0.328	0.273	0.329	0.276	0.330
	2023	0.275	0.331	0.291	0.340	0.282	0.332	**0.270**	**0.325**	0.280	0.337	0.283	0.337
	2025	**0.270**	**0.326**	0.290	0.339	0.283	0.334	0.271	**0.326**	0.279	0.336	0.282	0.333
	2027	**0.276**	0.332	0.288	0.337	0.278	**0.329**	0.284	0.336	0.282	0.339	0.285	0.336
ETTm2	2021	**0.182**	**0.265**	0.188	0.269	0.186	0.268	0.183	**0.265**	0.184	0.266	0.187	0.267
	2023	0.185	0.269	0.186	0.268	0.190	0.273	**0.181**	**0.264**	0.188	0.271	0.188	0.269
	2025	**0.180**	**0.264**	0.191	0.273	0.190	0.271	0.186	0.269	0.182	0.265	0.188	0.270
	2027	**0.185**	0.269	0.189	0.271	**0.185**	**0.266**	0.186	0.269	0.186	0.269	0.186	0.272
ECL	2021	**0.132**	**0.229**	0.135	0.233	0.133	0.231	0.136	0.234	0.135	0.232	0.134	0.230
	2023	**0.128**	**0.226**	0.139	0.238	0.136	0.235	0.140	0.239	0.133	0.229	0.136	0.233
	2025	0.134	0.230	0.139	0.236	**0.130**	**0.229**	0.141	0.238	0.138	0.236	0.135	0.234
	2027	0.136	0.238	0.138	0.235	0.135	0.234	**0.134**	**0.231**	0.139	0.237	0.137	0.234
Weather	2021	0.158	**0.206**	0.162	0.213	0.159	0.208	**0.156**	0.207	0.157	0.208	0.161	0.211
	2023	0.162	0.211	0.166	0.218	0.162	0.212	**0.154**	**0.206**	0.160	0.212	0.161	0.213
	2025	0.160	0.209	0.158	0.210	0.164	0.214	0.159	0.211	**0.154**	**0.206**	0.162	0.212
	2027	**0.155**	**0.205**	0.163	0.215	0.158	0.206	0.160	0.210	0.160	0.212	0.161	0.211
Solar	2021	**0.182**	**0.220**	0.191	0.226	0.195	0.230	0.197	0.231	0.188	0.225	0.183	0.222
	2023	**0.179**	**0.218**	0.194	0.230	0.198	0.232	0.193	0.228	0.192	0.228	0.190	0.226
	2025	**0.184**	0.223	0.193	0.227	0.192	0.228	0.201	0.236	0.186	**0.222**	**0.184**	0.223
	2027	**0.184**	0.223	0.189	0.225	0.196	0.233	0.200	0.235	0.186	**0.222**	0.190	0.225

G Additional Model Analysis

Representative TFU Semantics. We visualize representative TFUs and their matched temporal segments in the ETTh1→Weather setting. A lightweight post-hoc decoder maps latent prototypes to one-dimensional temporal curves. As shown in Fig. 6, the left column compares source-domain TFUs with matched ETTh1 and Weather segments, while the right column compares updated TFUs with their Weather matches. These examples provide a qualitative view of the local dynamics associated with individual TFUs, rather than exact reconstructions of the matched segments. The decoder is used only for visualization and does not participate in forecasting or TFU learning.

Shape Analysis Settings. We compare nearest-neighbor retrieval using learned embeddings or raw temporal statistics, together with random retrieval. The raw-statistics baseline computes the mean, maximum, and endpoint difference directly from each raw segment, whereas the semantic encoder summarizes learned token representations. Self-matches and overlapping segments are excluded. Waveform correlation is computed after linear detrending and normalization to control for level, scale, and linear trend. For each retrieval method, we first average correlations over anchors within each TFU and then compute the mean and standard deviation across TFUs. This gives each TFU equal weight rather than allowing frequently activated TFUs to dominate the reported results. The bars in Fig. 4 report these across-TFU means, and the error bars indicate across-TFU standard de-



Figure 6: Representative TFUs and matched temporal segments from ETTh1 to Weather. Left: decoded source-domain TFUs and matched segments from both domains. Right: decoded updated TFUs and matched Weather segments.

viations. The paired bootstrap 95% confidence intervals assess the mean correlation improvement of embedding-based retrieval over raw-statistics retrieval; they are distinct from the plotted error bars. The Weather setting uses the source-trained ETTh1 TFU dictionary. Updated Weather uses the dictionary updated and expanded with limited Weather training data, while retaining the same Weather evaluation data. The number of active TFUs increases from 16 to 22. Dictionary updating changes prototype assignments and TFU groups, thereby changing the weighting of anchors in the across-TFU average. Consequently, the mean correlation of raw-statistics retrieval changes slightly from 0.656 to 0.664, although its definition remains unchanged.

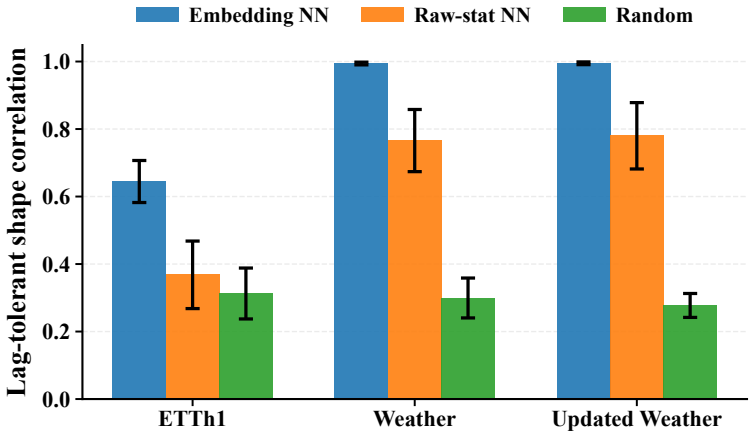


Figure 7: Waveform correlation after linear detrending and normalization, allowing temporal shifts of up to ± 8 steps. Bars report the mean of within-TFU average correlations; error bars indicate the standard deviation across TFUs.

Additional Shape Analysis. We further evaluate waveform correlation while allowing temporal shifts of up to ± 8 steps. As shown in Fig. 7, embedding-based retrieval achieves the highest mean correlations: 0.645, 0.994, and 0.995 on ETTh1, Weather, and updated Weather, respectively, com-

Table 14: Within-TFU shape consistency after linear detrending and normalization. Values report the mean $\pm$ standard deviation of within-TFU average correlations across TFUs.

Setting	Same TFU	Random
ETTh1	0.069 ± 0.078	0.007 ± 0.009
Weather	0.372 ± 0.355	0.014 ± 0.064
Updated Weather	0.456 ± 0.359	0.019 ± 0.047

pared with 0.368, 0.766, and 0.780 for raw-statistics retrieval. The advantage persists when small temporal shifts are allowed, further supporting the local shape information retained in the learned embeddings.

Within-TFU Shape Consistency. We examine whether segments assigned to the same TFU exhibit similar local shapes. Table 14 compares waveform correlations between same-TFU segments and randomly paired segments after linear detrending and normalization. Correlations are averaged within each TFU and then summarized across TFUs. Same TFU segments have higher mean correlations in all three settings, although consistency varies across TFUs and is weaker on ETTh1. On Weather, the mean correlation increases from 0.372 to 0.456 after dictionary updating, while random-pair correlations remain close to zero. Under the corresponding TFU assignments, these results indicate greater average within-TFU shape consistency after target-domain adaptation.

These results support local shape retention under the evaluated conditions, rather than complete preservation of temporal ordering or invariance to arbitrary temporal distortions.