PRISM: A TRAINING-FREE VISUAL FORECASTER FOR TIME SERIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Vision-based time series forecasting provides a promising route for reusing pretrained visual models for temporal modeling. Existing approaches have explored increasingly rich visual representations, multimodal augmentation, learned temporal–visual alignment, and adaptation of pretrained vision models. However, input-format compatibility does not necessarily ensure structural compatibility with pretrained visual representations. A key issue therefore remains underexplored: how to organize complementary temporal factors in a way that is compatible with the model-specific geometry learned during visual pretraining? We term the resulting mismatch between temporal factor organization and the pretrained RGB geometry Time-Series-to-RGB (TS2RGB) misalignment. To address this problem, we propose **PRISM**, a training-free forecasting model built on a frozen MAE encoder–decoder. Temporal Chromatic Rendering (TCR) organizes complementary temporal structures along mutually orthogonal RGB directions derived from the spatially constant color sensitivity of the frozen MAE patch-embedding layer, while Topology-Guided Visual Token Propagation (TVTP) injects dynamic cross-variable dependencies through spatially corresponding latent tokens. PRISM introduces no additional trainable parameters and performs forecasting without downstream weight updates. Across eight multivariate benchmarks and four forecasting horizons, PRISM ranks first in 43/64 dataset–horizon–metric comparisons and 12/16 dataset–metric averages, with real-token analyses showing 63.5% and 14.9% reductions in cross-component coupling at the patch-embedding interface and final encoder representation, respectively.

1 INTRODUCTION

Long-term time series forecasting (LTSF) is fundamental to applications in energy, meteorology, transportation, and public health (Wu et al., 2021; Zhou et al., 2021). Recent advances increasingly move beyond dataset-specific training toward large-scale pretraining on heterogeneous time series, enabling transferable temporal priors and zero-shot or few-shot forecasting across datasets and domains (Woo et al., 2024; Ansari et al., 2024; Shi et al., 2025; Liu et al., 2025b; He et al., 2026; Lu et al., 2025). Beyond learning transferable priors exclusively from numerical time-series corpora, pretrained vision models provide another source of reusable knowledge for temporal forecasting. Images and time series are both dense and continuous observations, and pretrained visual representations encode useful priors related to local continuity, recurring structure, and multi-scale variation. This motivates converting numerical sequences into image-like inputs and reusing pretrained visual backbones for temporal prediction (Chen et al., 2024; Zhong et al., 2025; Shen et al., 2026; Wu et al., 2026). However, making a time series acceptable to a vision model at the input level does not necessarily make it compatible with the representation geometry learned during visual pretraining. As illustrated in Figure 1, natural images are natively represented through heterogeneous RGB channels, whereas existing time-series methods first construct an artificial visual interface using periodic structural rendering or value-driven visual rendering before feeding the resulting representation into a frozen vision model. Although both inputs are formally represented in RGB space, their channel structures arise from fundamentally different principles. This motivates a more fundamental question: **how should heterogeneous temporal factors be organized within the input representation space of a frozen vision model so that their structure is compatible with the geometry learned during visual pretraining?**

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.

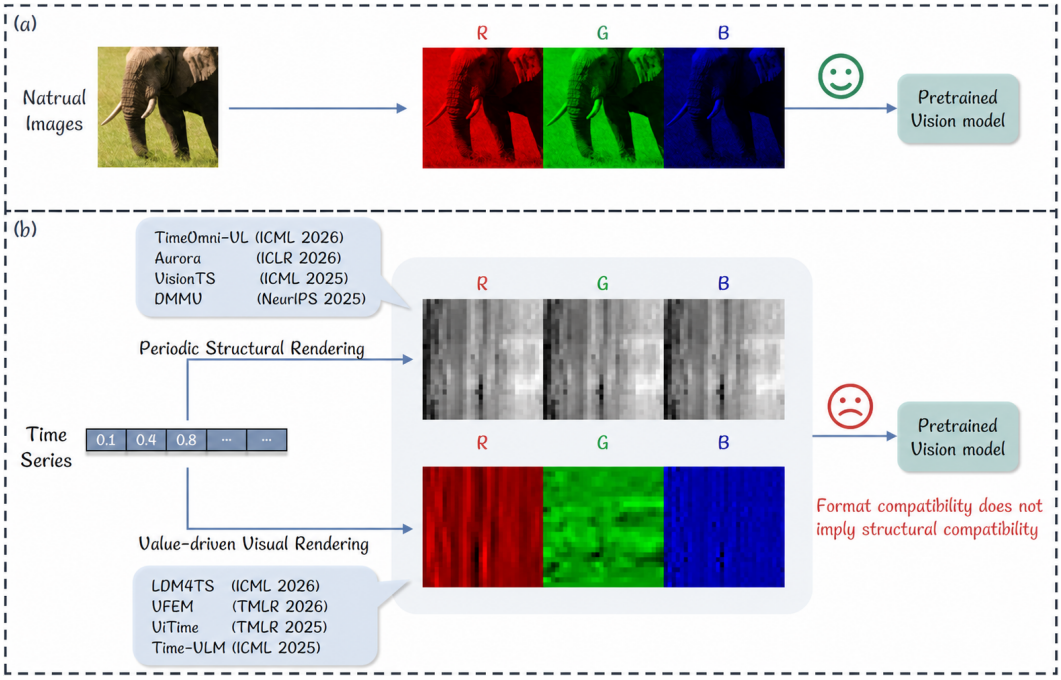


Figure 1: **Natural Image Input vs. Time-Series Visualized Input.** (a) Natural images preserve native heterogeneous RGB structures compatible with pretrained vision models. (b) Time-series visualizations construct artificial RGB representations through periodic or value-driven rendering, whose channel organization may not align with the pretrained RGB sensitivity geometry.

Existing vision-based forecasting methods can be broadly grouped into two representative rendering paradigms, as illustrated in Figure 1(b). The first is **periodic structural rendering**, which folds one-dimensional sequences according to temporal periodicity to construct two-dimensional visual representations. This paradigm has been adopted or extended by VisionTS, DMMV, Aurora, and TimeOmni-VL, which exploit periodic organization as a primary interface between time series and visual models (Chen et al., 2024; Shen et al., 2026; Wu et al., 2026; Guan et al., 2026). The second is **value-driven visual rendering**, which constructs visual patterns from temporal values, multivariate structures, or derived transformations rather than relying primarily on periodic folding. Representative methods include LDM4TS, VFEM, ViTime, and Time-VLM, which explore multi-view transformations, cross-variable visual structures, value-based image spaces, and multimodal visual encodings (Ruan et al., 2025; Wang et al., 2025a; Yang et al., 2024; Zhong et al., 2025). Despite their different rendering strategies, both paradigms ultimately construct artificial visual representations for frozen vision models.

We argue that this format compatibility leaves an overlooked structural issue. Period-folded grayscale rendering provides the clearest example: replicating the same temporal image across channels enforces $R = G = B$, restricting the representation to the achromatic direction of RGB space. Value-driven encodings can produce richer channel patterns, but using multiple RGB channels does not by itself determine whether distinct temporal factors are assigned to directions that are meaningful to a frozen visual patch embedding. A frozen vision model is not an isotropic RGB container: visual pretraining already induces a model-specific representation geometry at its input interface. Meanwhile, periodic organization of a time series already exposes complementary temporal factors. The folded values preserve temporal level, within-period variations describe local phase evolution, and changes across periods capture inter-period dynamics (Wu et al., 2022; Niu et al., 2026; Ma et al., 2026; Cao et al., 2026). The central challenge is therefore not merely to enrich RGB inputs, but to organize these temporal factors consistently with the pretrained visual geometry, instantiated here by model-native RGB directions derived from the frozen patch embedding. We refer to this mismatch as **Time-Series-to-RGB** (TS2RGB) misalignment.

To address this problem, we propose PRISM, a training-free model built on a frozen MAE encoder–decoder that reorganizes temporal information without introducing additional trainable parameters. Its **Temporal Chromatic Rendering** (TCR) derives an orthogonal coordinate system directly from the spatially constant RGB sensitivity of the frozen MAE patch-embedding layer. The direction most consistent with achromatic variation preserves the original temporal level, while the remaining orthogonal directions encode within-period phase variation and inter-period dynamics. TCR therefore follows a preserve-and-augment principle: it retains the original level information while exposing complementary temporal structures along patch-embedding-derived visual directions. Unlike approaches that learn a new temporal–visual representation space or adapt the visual model to time-series data, TCR reorganizes temporal information at the input interface while leaving the frozen MAE encoder–decoder unchanged.

For multivariate forecasting, we further introduce **Topology-Guided Visual Token Propagation** (TVTP). Channel-independent visual encoding processes variables separately and can therefore miss cross-variable dependencies in the visual latent space. TVTP constructs a dynamic topology from the observed multivariate history and propagates relational information only between spatially corresponding latent tokens. Its residual update uses a fixed propagation strength and can optionally be modulated by correlation-derived confidence. Together, TCR and TVTP organize temporal and inter-variable structure around the frozen MAE reconstruction pipeline without downstream weight adaptation.

Our contributions are summarized as follows:

- **TS2RGB misalignment.** We identify an overlooked structural compatibility issue in vision-based time series forecasting: RGB-format compatibility does not ensure that complementary temporal factors align with pretrained visual geometry. For frozen MAE, this mismatch is characterized through patch-embedding-induced RGB sensitivity and termed TS2RGB misalignment.
- **Prism.** We propose PRISM, a training-free forecasting model built on a frozen MAE encoder–decoder. TCR organizes complementary temporal structures along patch-embedding-derived RGB directions, while TVTP introduces dynamic cross-variable context through spatially corresponding latent-token propagation.
- **Comprehensive empirical validation.** On eight multivariate benchmarks and four forecasting horizons, PRISM ranks first in **43/64** dataset–horizon–metric comparisons and **12/16** dataset–metric averages, with additional controlled and representation analyses supporting the proposed structural design.

2 RELATED WORK

Time Series Forecasting. Recent time series forecasting research increasingly leverages knowledge acquired through large-scale pretraining. Time-series foundation models such as MOIRAI, Chronos, Timer, Time-MoE, Timer-XL, and GTM learn transferable temporal priors from large collections of numerical sequences (Woo et al., 2024; Ansari et al., 2024; Liu et al., 2024b; Shi et al., 2025; Liu et al., 2025b; He et al., 2026). In parallel, pretrained language models have been adapted to forecasting through reprogramming, alignment, or semantic augmentation (Jin et al., 2024; Liu et al., 2025a; 2026a). Recent efforts further explore in-context prediction and test-time model reuse to improve generalization without task-specific retraining (Lu et al., 2025; Kayaalp et al., 2026). Beyond language, recent studies exploit pretrained vision and multimodal models to transfer visual and cross-modal knowledge into temporal prediction (Chen et al., 2024; Zhong et al., 2025; Wu et al., 2026; Guan et al., 2026).

Vision-based Time Series Forecasting. Recent studies increasingly transfer pretrained visual knowledge to temporal forecasting. VisionTS (Chen et al., 2024) establishes a direct reconstruction paradigm by converting periodically folded sequences into grayscale images for a frozen MAE. Subsequent methods enrich or adapt this interface through complementary visual views, latent diffusion, multivariate colorization, or visual–temporal fusion (Shen et al., 2026; Ruan et al., 2025; Yang et al., 2024; Shen et al., 2025; Wang et al., 2025a; Zhong et al., 2025). TimeArtist (Ma et al., 2025a) further learns semantic-level correspondences between temporal fluctuations and visual concepts in a shared discrete latent space, while OccamVTS (Lyu et al., 2026) highlights the importance of rela-

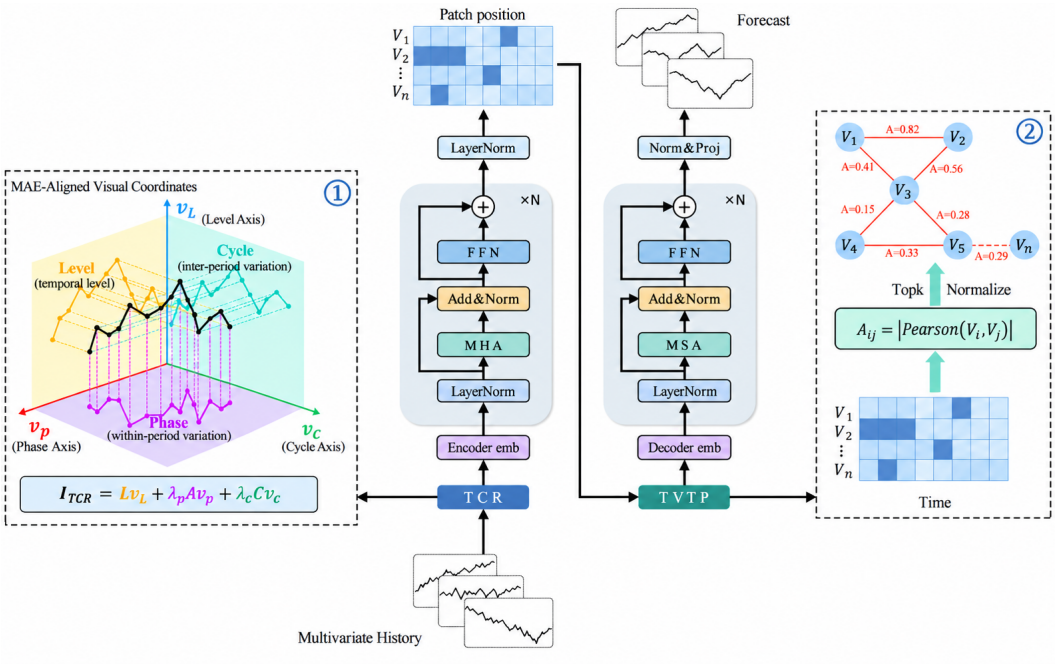


Figure 2: Overview of PRISM. ① TCR organizes complementary temporal structures along MAE-aligned visual coordinates. ② TVTP propagates dynamic cross-variable relations between spatially corresponding latent tokens.

tively low-level visual representations for forecasting. Complementary to these approaches, PRISM focuses on the RGB sensitivity geometry of the frozen MAE patch-embedding layer and organizes complementary temporal structures directly along directions derived from this local input geometry.

Multivariate Dependency Modeling. Multivariate forecasting increasingly emphasizes adaptive cross-variable dependency modeling. Recent methods capture informative interactions through dynamic graphs, variable-aware states, local cross-variate modeling, and correlation-aware adaptation (Hu et al., 2025; Ma et al., 2025b; Li et al., 2026; Liu et al., 2026b; Xu et al., 2026; Cheng et al., 2026; Wang et al., 2026). These studies highlight the importance of explicit cross-variable structure, but such dependencies are mainly modeled in numerical representation spaces. For visual forecasting, independently encoded variables likewise lack explicit interaction in the visual latent space. As a complementary extension to TCR, TVTP transfers dynamic variable dependencies into spatially corresponding visual tokens through lightweight residual propagation, without introducing an additional trainable graph or feature-fusion module.

3 METHOD

3.1 PROBLEM STATEMENT

We consider the multivariate time series forecasting problem, a standard setting in long-term forecasting (Nie et al., 2022; Liu et al., 2024a; Wu et al., 2022). Given a historical observation window $\mathbf{X} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}] \in \mathbb{R}^{L \times n}$ containing n variables over L time steps, the objective is to predict their future values for the subsequent H steps, i.e., $\widehat{\mathbf{Y}} = f_\theta(\widetilde{\mathbf{X}}) \in \mathbb{R}^{H \times n}$, where f_θ denotes the forecasting model, L is the look-back length, and H is the forecasting horizon. Here, $\widetilde{\mathbf{X}} = [\widetilde{\mathbf{x}}^{(1)}, \ldots, \widetilde{\mathbf{x}}^{(n)}]$ denotes the normalized historical window used as model input.

3.2 FRAMEWORK OVERVIEW

As illustrated in Fig. 2, PRISM augments a frozen MAE at two complementary stages. TCR first reorganizes the historical signal before visual encoding by separating temporal level, within-period

phase variation, and inter-period dynamics, and assigning them to visual coordinates derived from the frozen MAE patch-embedding layer. This construction determines how temporal structure is presented to the pretrained visual backbone without modifying its parameters. After MAE encoding, TVTP complements this within-variable organization by introducing dynamic cross-variable context between spatially corresponding visible tokens. The resulting latent representations are then passed through the standard MAE reconstruction path. Throughout the pipeline, the frozen MAE encoder–decoder remains unchanged, and no additional trainable modules are introduced.

3.3 Temporal Geometry Decomposition

The first step is to expose temporal structures that can be assigned distinct visual roles. Period folding naturally separates two complementary directions: progression within a period and variation across successive periods. Rather than treating the folded matrix only as an image-like arrangement, TCR preserves these temporal semantics explicitly. For the normalized history $\widetilde{\mathbf{x}}^{(i)}$ of variable i, we fold the sequence according to periodicity P and construct

$$\begin{aligned} \mathbf{L}^{(i)} &= \mathrm{Fold}_P\left(\widetilde{\mathbf{x}}^{(i)}\right), \\ \mathbf{A}^{(i)}_{p,q} &= \mathbf{L}^{(i)}_{p,q} - \mathbf{L}^{(i)}_{p-\Delta,q}, \\ \mathbf{C}^{(i)}_{p,q} &= \mathbf{L}^{(i)}_{p,q} - \mathbf{L}^{(i)}_{p,q-1}. \end{aligned} \tag{1}$$

Here, $\mathbf{L}^{(i)} \in \mathbb{R}^{P\times Q}$ represents the temporal level, while $\mathbf{A}^{(i)}$ captures the within-period phase variation, and $\mathbf{C}^{(i)}$ captures the inter-period (cycle) variation at the same phase. Together, these three maps describe the signal value itself and two complementary directions of temporal change. The phase displacement is $\Delta = \max(1, \mathrm{round}(P/P_{\mathrm{ref}}))$, where P_{ref} controls the reference temporal resolution and keeps the phase difference comparable across sampling granularities. Differences are computed before visual resizing, with zero-valued boundaries and no circular wrap-around. Histories not divisible by P are padded in a replicate on the left before folding. The three maps are then independently resized with bilinear interpolation to obtain $\mathbf{L}^{(i)}_r$, $\mathbf{A}^{(i)}_r$, and $\mathbf{C}^{(i)}_r$; the same interpolation mode is used for output resampling. This ordering ensures that phase and cycle variations are defined in the original temporal raster rather than induced by image resizing.

3.4 MAE Patch-Embedding Coordinate Construction

The temporal decomposition specifies what information should be represented; the next question is how these components should be organized in RGB space. Instead of assigning them to arbitrary color channels, TCR derives a coordinate system from the spatially constant color sensitivity of the frozen MAE patch-embedding layer. Let $\mathbf{W} \in \mathbb{R}^{D\times 3\times K\times K}$ denote the pretrained patch-embedding kernel. We spatially average the kernel and construct

$$\begin{aligned} \mathbf{W}_{\mathrm{DC}}[d,c] &= \frac{1}{K^2}\sum_{h=1}^{K}\sum_{w=1}^{K}\mathbf{W}[d,c,h,w], \\ \mathbf{G} &= \mathbf{W}_{\mathrm{DC}}^{\top}\mathbf{W}_{\mathrm{DC}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\top}. \end{aligned} \tag{2}$$

For any RGB direction $\mathbf{v}$, $\mathbf{v}^{\top}\mathbf{G}\mathbf{v} = \|\mathbf{W}_{\mathrm{DC}}\mathbf{v}\|_2^2$. Thus, the eigenvectors of $\mathbf{G}$ define orthogonal directions under the spatially constant patch-embedding response:

$$(\mathbf{W}_{\mathrm{DC}}\mathbf{u}_i)^{\top}(\mathbf{W}_{\mathrm{DC}}\mathbf{u}_j) = \mathbf{u}_i^{\top}\mathbf{G}\mathbf{u}_j = 0, \qquad i \neq j. \tag{3}$$

This construction provides a compact, model-specific RGB geometry derived directly from the frozen MAE. The property concerns the local patch-embedding sensitivity and does not imply that arbitrary token representations remain orthogonal throughout the full Transformer.

The temporal level is assigned to the eigen-axis most aligned with the canonical achromatic direction $\mathbf{g}_{\mathrm{ach}} = [1,1,1]^{\top}/\sqrt{3}$:

$$\mathbf{u}_L = \arg\max_{\mathbf{u}_j} \left|\mathbf{u}_j^{\top}\mathbf{g}_{\mathrm{ach}}\right|. \tag{4}$$

For the remaining two eigen-directions, we use fixed chromatic references

$$\mathbf{r}_p = \mathrm{norm}([1, -\tfrac{1}{2}, -\tfrac{1}{2}]^{\top}), \qquad \mathbf{r}_c = \mathrm{norm}([0, \tfrac{\sqrt{3}}{2}, -\tfrac{\sqrt{3}}{2}]^{\top}).$$

The phase–cycle assignment minimizes the total sign-invariant axis-angle deviation from these references, after which signs are oriented toward the corresponding references. This removes permutation and sign ambiguity and makes the coordinate construction deterministic. We finally define $\mathbf{v}_L = s_L\mathbf{u}_L$, $\mathbf{v}_p = s_p\mathbf{u}_p$, and $\mathbf{v}_c = s_c\mathbf{u}_c$, with $s_L = \sqrt{3}$ and $s_p = s_c = \sqrt{3/2}$. The scaling affects only the component amplitudes and preserves the directions of the derived axes. The resulting basis is extracted once from the frozen MAE and remains fixed across datasets and forecasting horizons.

3.5 Temporal Chromatic Rendering

With the temporal factors and their visual coordinates defined, TCR combines them into a single RGB representation. Because the temporal level and finite differences naturally have different numerical scales, directly combining them could make the contribution of phase and cycle depend mainly on derivative magnitude. TCR therefore first aligns the variation amplitudes with the level map. Let $\sigma_{\epsilon_{\mathrm{TCR}}}(\mathbf{M}) = \sqrt{\mathrm{Var}(\mathbf{M}) + \epsilon_{\mathrm{TCR}}}$. We define

$$\begin{aligned}\widetilde{\mathbf{A}}^{(i)} &= \mathbf{A}_r^{(i)} \frac{\sigma_{\epsilon_{\mathrm{TCR}}}\left(\mathbf{L}_r^{(i)}\right)}{\sigma_{\epsilon_{\mathrm{TCR}}}\left(\mathbf{A}_r^{(i)}\right)}, \\ \widetilde{\mathbf{C}}^{(i)} &= \mathbf{C}_r^{(i)} \frac{\sigma_{\epsilon_{\mathrm{TCR}}}\left(\mathbf{L}_r^{(i)}\right)}{\sigma_{\epsilon_{\mathrm{TCR}}}\left(\mathbf{C}_r^{(i)}\right)}.\end{aligned} \tag{5}$$

TCR then renders the three structures as

$$\mathbf{I}_{\mathrm{TCR}}^{(i)} = \mathbf{L}_r^{(i)}\mathbf{v}_L + \lambda_p\widetilde{\mathbf{A}}^{(i)}\mathbf{v}_p + \lambda_c\widetilde{\mathbf{C}}^{(i)}\mathbf{v}_c. \tag{6}$$

Here, λ_p and λ_c control the contributions of the two variation components. Because the three rendering axes are mutually orthogonal, phase and cycle do not alter the level coefficient. Hence, the temporal level remains explicitly recoverable while complementary dynamics are introduced along separate visual directions. Each variable is rendered independently and processed by the same frozen MAE. The historical region occupies visible patch columns, while a zero-valued future region occupies the masked columns. A common deterministic forecasting mask is shared across variables, so the same visible-token index corresponds to the same two-dimensional patch location. This spatial consistency becomes important when relational information is introduced after encoding.

3.6 Topology-Guided Visual Token Propagation

TCR organizes temporal structure within each variable, but independently encoded visual streams still lack explicit information about how variables co-evolve. TVTP addresses this complementary limitation in the latent space. For each forecasting window, it constructs a dynamic variable topology from the normalized history using absolute Pearson correlation. Self-connections are removed and the strongest $k_{\mathrm{eff}} = \min(k, n-1)$ neighbors are retained:

$$\begin{aligned}\rho_{ij} &= \left|\mathrm{Pearson}\left(\widetilde{\mathbf{x}}^{(i)}, \widetilde{\mathbf{x}}^{(j)}\right)\right|, \qquad \rho_{ii} = 0, \\ \mathcal{N}_i &= \mathrm{TopK}_j^{k_{\mathrm{eff}}}\left(\rho_{ij}\right), \\ S_{ij} &= \frac{\rho_{ij}\mathbf{1}_{j\in\mathcal{N}_i}}{\sum_{r\in\mathcal{N}_i}\rho_{ir} + \epsilon_{\mathrm{rel}}}, \\ \gamma_i &= \max_{j\in\mathcal{N}_i}\rho_{ij}.\end{aligned} \tag{7}$$

Here, $\mathbf{S}$ is the sparse normalized propagation topology and γ_i measures the strongest retained dependency. Because the topology is recomputed from each input window, the relational structure can adapt to the current multivariate state rather than relying on a fixed global graph. Let $\mathbf{Z}_{i,\ell} \in \mathbb{R}^D$ denote the visible token of variable i at patch position ℓ. Because all variables share the same forecasting mask, tokens with the same ℓ correspond to the same spatial location in the visual canvas.

This allows cross-variable information to be propagated without mixing unrelated patch positions:

$$\begin{aligned} \mathbf{Z}^{\mathrm{rel}}_{i,\ell} &= \sum_j S_{ij}\mathbf{Z}_{j,\ell}, \\ \eta_i(g_{\mathrm{gate}}) &= (1-g_{\mathrm{gate}}) + g_{\mathrm{gate}}\gamma_i, \qquad g_{\mathrm{gate}} \in \{0,1\}, \\ \alpha_i &= \mathrm{clip}\left(\beta\eta_i(g_{\mathrm{gate}}), 0, 1\right), \\ \mathbf{Z}'_{i,\ell} &= \mathbf{Z}_{i,\ell} + \alpha_i\left(\mathbf{Z}^{\mathrm{rel}}_{i,\ell} - \mathbf{Z}_{i,\ell}\right). \end{aligned} \tag{8}$$

Here, β controls the residual propagation strength, while g_{gate} enables optional correlation-based confidence modulation: $g_{\mathrm{gate}} = 1$ yields $\alpha_i = \beta\gamma_i$, whereas $g_{\mathrm{gate}} = 0$ yields $\alpha_i = \beta$. Propagation is restricted to visible patch tokens and leaves the class token unchanged. TVTP therefore introduces cross-variable structure directly into the existing MAE latent representation without requiring an additional trainable graph or fusion module.

3.7 Reconstruction and Forecasting

The corrected visible tokens follow the standard frozen MAE decoding path: they are projected to the decoder space, mask tokens are inserted according to the stored restoration indices, and the original patch ordering is restored before reconstruction. Let $\widehat{\mathbf{I}}^{(i)} = \mathcal{D}_{\mathrm{MAE}}(\mathbf{Z}'^{(i)})$ denote the reconstructed RGB canvas. The temporal-level component is recovered by projection onto $\mathbf{v}_L$:

$$\widehat{\mathbf{L}}^{(i)} = \frac{\left\langle \widehat{\mathbf{I}}^{(i)}, \mathbf{v}_L \right\rangle}{\|\mathbf{v}_L\|_2^2}. \tag{9}$$

The recovered level map is resized back to the temporal raster using the same interpolation scheme, unfolded according to the original periodic ordering, and restricted to the target forecasting horizon. The resulting sequence is then transformed back to the original numerical scale to obtain the forecast $\widehat{\mathbf{y}}^{(i)}$ for each variable. Applying this procedure independently to all variables produces the complete multivariate prediction $\widehat{\mathbf{Y}} = [\widehat{\mathbf{y}}^{(1)}, \ldots, \widehat{\mathbf{y}}^{(n)}] \in \mathbb{R}^{H\times n}$. Thus, the reconstructed visual representation is converted back into a standard numerical time-series forecast without modifying the pretrained MAE reconstruction mechanism.

4 Experiments

4.1 Experimental Setup

Datasets. PRISM is evaluated on eight multivariate forecasting benchmarks from diverse domains, including energy infrastructure, meteorological observations, and renewable-energy systems. Specifically, we use **ETTh1**, **ETTh2**, **ETTm1**, **ETTm2**, **Weather** (Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a), **SPSS** (Chen & Xu, 2022), **Wind** (Li et al., 2022), and **SDWPF** (Zhou et al., 2024). These datasets cover electricity transformer measurements, atmospheric variables, photovoltaic generation, and wind-power dynamics, providing a diverse evaluation across different temporal resolutions and application scenarios.

Baselines. We compare PRISM with pretrained forecasting models, including LightGTS$_{\text{Tiny}}$ (Wang et al., 2025b), Timer-XL (Liu et al., 2025b), Timer (Liu et al., 2024b), MOIRAI$_{\text{Base}}$ (Woo et al., 2024), and MOMENT$_{\text{Large}}$ (Goswami et al., 2024). We additionally compare against 10% few-shot baselines, including SE-LLM (Liu et al., 2026a), TFPS (Sun et al., 2026), DUET (Qiu et al., 2025), CALF (Liu et al., 2025a), and iTransformer (Liu et al., 2024a). VisionTS (Chen et al., 2024) is evaluated separately as the matched visual predecessor in the controlled analysis provided in the Appendix.

Experimental Settings. PRISM is evaluated in a training-free zero-shot setting with a fully frozen MAE and no downstream weight updates. Non-trainable configurations are selected on the validation split and fixed before testing. We evaluate $H \in \{96, 192, 336, 720\}$ using MSE and MAE; full implementation and configuration details are provided in the Appendix.

Table 1: Average forecasting performance over prediction horizons $H \in \{96, 192, 336, 720\}$. MSE and MAE are averaged over the four prediction horizons; lower values are better. Pretrained baselines are evaluated in the zero-shot setting, while few-shot baselines use 10% of the downstream training data. The best and second-best results are highlighted in red and blue, respectively.

	Zero-shot Pretrain												Few-shot (10%)									
Dataset	**Prism** (Ours)		**LightGTS$_{\text{Tiny}}$** (2025b)		**Timer-XL** (2025b)		**Timer** (2024b)		**MOIRAI$_{\text{Base}}$** (2024)		**MOMENT$_{\text{Large}}$** (2024)		**SE-LLM** (2026a)		**TFPS** (2026)		**DUET** (2025)		**CALF** (2025a)		**iTrans.** (2024a)	
Metric	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
ETTh1	**0.388**	**0.413**	0.467	0.479	**0.412**	**0.418**	0.562	0.484	0.425	0.441	0.725	0.599	0.587	0.521	0.691	0.596	0.702	0.614	0.623	0.548	0.641	0.566
ETTh2	**0.328**	**0.374**	0.406	0.430	**0.357**	**0.391**	0.370	0.400	0.364	0.397	0.404	0.437	0.381	0.425	0.449	0.470	0.421	0.448	0.400	0.440	0.406	0.440
ETTm1	**0.368**	**0.370**	0.765	0.587	**0.455**	**0.425**	0.768	0.569	0.614	0.498	0.701	0.559	0.592	0.497	0.624	0.519	0.490	0.474	0.471	0.459	0.504	0.479
ETTm2	**0.279**	**0.321**	0.359	0.393	0.290	0.337	0.315	0.356	0.339	0.360	0.359	0.394	0.293	0.342	0.331	0.377	**0.283**	**0.335**	0.286	0.337	0.299	0.351
SPSS	**0.169**	**0.224**	0.781	0.709	0.465	0.467	1.357	0.865	0.225	0.326	0.841	0.664	0.224	0.316	0.305	0.391	0.272	0.356	0.230	0.325	**0.201**	**0.303**
SDWPF	0.872	0.575	1.008	0.707	**0.775**	**0.540**	0.979	0.653	**0.779**	**0.563**	1.033	0.699	0.867	0.615	1.137	0.741	0.997	0.674	0.973	0.667	0.933	0.652
Weather	**0.262**	**0.288**	0.356	0.379	**0.264**	**0.299**	0.333	0.358	0.382	0.329	0.418	0.416	0.274	0.302	0.342	0.343	0.320	0.349	0.327	0.351	0.319	0.346
Wind	1.252	0.802	1.404	0.869	1.198	**0.781**	1.467	0.890	**1.119**	**0.754**	1.476	0.908	**1.184**	0.796	1.728	0.980	1.548	0.905	1.487	0.880	1.439	0.874
Wins	**12**		0		**2**		0		**2**		0		0		0		0		0		0	

4.2 Main Results

Table 1 reports the forecasting results on eight datasets across $H \in \{96, 192, 336, 720\}$, corresponding to 64 dataset–horizon–metric comparisons when MSE and MAE are counted separately. Despite using no downstream weight adaptation, PRISM ranks first in **43/64** comparisons and **12/16** dataset–metric averages, outperforming both zero-shot pretrained models and 10% few-shot baselines across a broad range of forecasting settings. The improvements span different temporal resolutions, prediction horizons, and application domains, showing that the proposed temporal rendering and cross-variable interaction are not limited to a particular dataset or forecasting regime. Complete horizon-level results and additional controlled analyses are provided in the Appendix.

4.3 Model Analysis

Alignment with MAE Patch-Embedding RGB Sensitivity Geometry. Figure 3 examines how the visual coordinate system affects the response of the frozen MAE patch embedding. Using the same ETTh1 sample and identical temporal components, we compare canonical RGB coordinates with the proposed MAE-aligned rendering. The upper panels show that changing only the rendering basis produces clearly different spatial activation patterns, indicating that the pretrained patch embedding is sensitive to how temporal information is organized across RGB channels. The lower panels show that canonical RGB responses remain coupled under the spatially averaged patch-embedding operator, with cosine similarities reaching -0.581 and -0.637, whereas the MAE-aligned directions are orthogonal by construction under $\mathbf{W}_{\text{DC}}$.

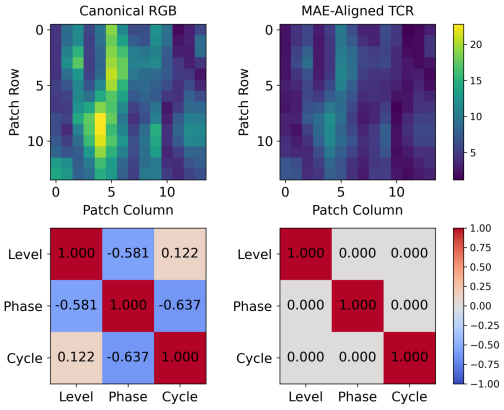


Figure 3: Visual-space analysis on ETTh1. Top: patch-embedding responses under canonical RGB and MAE-aligned TCR. Bottom: cosine similarities of the corresponding coordinate responses.

Ablation Study. Table 2 evaluates TCR and TVTP on four representative $H = 720$ settings. TVTP provides clear improvements relative to the corresponding TCR-only configuration on several datasets, with particularly visible gains on SDWPF and Weather. The contribution associated with TCR is more dataset-dependent, consistent with the fact that the usefulness of phase and cycle cues varies across temporal processes. Overall, the complete PRISM obtains the lowest four-dataset average MSE and MAE in this comparison. A more controlled factor-level analysis of TCR is provided in the Appendix.

Table 2: Ablation study of TCR and TVTP under zero-shot forecasting with prediction horizon $H = 720$. PRISM is compared with variants removing TCR, TVTP, or both components. Each entry reports MSE and MAE; lower values are better. The best and second-best results are highlighted in red and blue, respectively, with ties sharing the same highlight.

Variant	TCR	TVTP	ETTh2		SPSS		SDWPF		Weather		Avg	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Prism	✓	✓	**0.382**	**0.418**	**0.209**	**0.259**	**0.938**	**0.623**	**0.323**	**0.334**	**0.463**	**0.409**
w/o TCR	–	✓	**0.388**	**0.422**	**0.211**	**0.260**	**0.945**	**0.623**	**0.323**	**0.334**	**0.467**	**0.410**
w/o TVTP	✓	–	**0.382**	**0.418**	0.213	0.261	0.966	**0.626**	0.332	**0.338**	0.473	0.411
w/o Both	–	–	**0.388**	**0.422**	0.214	0.261	0.966	**0.626**	**0.330**	**0.338**	0.474	0.412

Computational Efficiency. As shown in Table 3, all methods are profiled with a fixed input length of 672, prediction horizon $H = 96$, and batch size of 32 on a single NVIDIA RTX 4090 GPU. Inference latency is averaged over three runs under the same profiling procedure. Under this matched setting, PRISM achieves the lowest inference latency and FLOPs among the profiled pretrained baselines, with an inference latency of 0.151 s and a computational cost of 145.49 GFLOPs. Although LightGTS$_{\text{Tiny}}$ uses fewer parameters and less peak memory, PRISM maintains substantially lower computational cost than the larger profiled foundation-model baselines without introducing additional trainable modules.

Table 3: Computational efficiency comparison. Computational efficiency under a matched input length of 672, prediction horizon $H = 96$, and batch size of 32 on a single NVIDIA RTX 4090 GPU. Inference latency is averaged over three runs. Total parameters, inference time, peak GPU memory, and FLOPs are reported; lower values are better.

Model	Params	Time (s)	Peak Mem.	FLOPs
PRISM	111.91M	**0.151**	1831.64 MiB	**145.49G**
MOIRAI$_{\text{Base}}$	91.36M	1.732	1517.64 MiB	47.23T
LightGTS$_{\text{Tiny}}$	**1.34M**	**0.294**	**58.63 MiB**	**428.13G**
Timer	**67.40M**	0.719	**339.30 MiB**	9.25T
Timer-XL	84.14M	0.337	653.81 MiB	22.98T

Additional Analyses. Beyond the analytical coordinate construction, we further examine whether the proposed structural organization remains effective on real time-series inputs and under reduced configuration freedom. On ETTh1, SPSS, and Weather, TCR reduces the average cross-component coupling by **63.5%** after the actual MAE patch embedding and by **14.9%** at the final encoder representation, showing that the effect persists beyond the $\mathbf{W}_{\text{DC}}$ construction itself. We also evaluate a single globally fixed structural configuration across all 32 forecasting settings, which remains within only **0.98% MSE** and **0.46% MAE** of the validation-selected configuration on average. Under matched context lengths, PRISM further improves in **7/9** MSE and **8/9** MAE comparisons. Together, these results indicate that the observed gains are not solely attributable to analytical orthogonality, per-setting configuration selection, or access to longer historical contexts. Full depth-wise and controlled analyses are provided in the Appendix.

5 CONCLUSION

Overall, PRISM demonstrates that effective reuse of pretrained visual models for time series forecasting depends not only on input-format compatibility, but also on how temporal and relational structures are organized within the visual representation space. By combining backbone-aligned temporal rendering with topology-guided latent interaction, PRISM provides a training-free model that consistently improves forecasting performance without modifying the frozen MAE backbone. The results further suggest that pretrained visual models can serve as effective forecasting priors when temporal information is organized with respect to their native representation geometry, rather than being treated merely as generic image processors. These findings highlight structural compatibility as an important principle for transferring pretrained visual reconstruction priors to temporal forecasting and offer a useful direction for future vision-based time series models built on frozen pretrained backbones.

AI USE STATEMENT

In this work, the authors used the ChatGPT web interface with GPT-5.6 Sol as the only generative AI tool. No autonomous AI agents or other generative AI models were used.

ChatGPT was used as an assistive tool in refining hypotheses and conceptual formulations, providing feedback on research methodology and experimental design, supporting method implementation through code generation and editing, suggesting experimental parameters, and assisting in the interpretation of experimental results. It was also used for literature discovery and summarization, identification of potentially relevant research gaps, drafting and revising selected portions of the manuscript, improving readability, and formatting references.

The research questions, core hypotheses, conceptual framework, and overall methodology were developed and determined by the authors. The authors designed the model and code framework, conducted the literature review to establish the research context and assess novelty, performed all experiments, examined and verified the experimental results, and made the final scientific interpretations and conclusions. The manuscript was written, reviewed, and revised by the authors, and the final selection of references was determined by the authors.

All AI-assisted outputs were subject to human review and verification. Code generated or modified with the assistance of ChatGPT was manually reviewed, integrated, tested, and validated before use. Literature and references suggested by ChatGPT were checked against the original publications before inclusion. Research ideas and potential research gaps identified with AI assistance were independently assessed through manual literature review. Suggestions concerning methodology, experimental design, experimental parameters, and result interpretation were evaluated by the authors and adopted only after human assessment. AI-assisted manuscript text was reviewed and revised before inclusion.

Generative AI was not used to generate synthetic datasets, clean or reformat datasets, autonomously conduct experiments, fabricate or modify experimental results, or perform qualitative or thematic data analysis. It was not used for translation in this work.

The authors have reviewed all AI-assisted work and take full responsibility for the final content of this submission, including its text, claims, methodology, code, experimental results, interpretations, references, and other artifacts produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

The experiments are designed to be fully reproducible. PRISM is evaluated under a training-free zero-shot inference protocol using a frozen pretrained MAE-Base backbone, without downstream gradient updates or task-specific weight training. Non-trainable inference configurations are selected using the validation split for each dataset–horizon setting and fixed before final test evaluation; test labels are not used for configuration selection. The implementation includes the complete Temporal Chromatic Rendering (TCR) and Topology-Guided Visual Token Propagation (TVTP) modules, together with the preprocessing, forecasting, and evaluation pipeline.

For all eight benchmarks, dataset-specific scripts specify the complete configurations for $H \in \{96, 192, 336, 720\}$, including the look-back length, temporal periodicity, TCR coefficients, reference-bin configuration, topology size, propagation strength, and confidence-gating indicator. The pretrained MAE and all PRISM components remain unchanged during inference. Input normalization is performed independently for each forecasting window using statistics from the observed historical sequence together with the fixed normalization constant used by the implementation, and the predicted horizon is transformed back using the corresponding window statistics. Performance is evaluated using MSE and MAE. A fixed random seed is used throughout the implementation.

The anonymous supplementary material will include the complete source code, environment dependencies, benchmark configurations, dataset-processing pipeline, pretrained-backbone setup, and executable scripts for reproducing every reported forecasting setting. Additional implementation details, hyperparameter configurations, dataset descriptions, and complete experimental results are provided in the Appendix.

REFERENCES

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

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

Mouxiang Chen, Lefei Shen, Zhuo Li, Xiaoyun Joy Wang, Jianling Sun, and Chenghao Liu. Visionts: Visual masked autoencoders are free-lunch zero-shot time series forecasters. *arXiv preprint arXiv:2408.17253*, 2024.

Yongbao Chen and Junjie Xu. Solar and wind power data from the chinese state grid renewable energy generation forecasting competition. *Scientific data*, 9(1):577, 2022.

Hanyin Cheng, Xingjian Wu, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Guo Guo. Cora: Boosting time series foundation models for multivariate forecasting through correlation-aware adapter. In *International Conference on Learning Representations*, volume 2026, pp. 106657–106680, 2026.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. Moment: A family of open time-series foundation models. *arXiv preprint arXiv:2402.03885*, 2024.

Tong Guan, Sheng Pan, Johan Barthelemy, Zhao Li, Yujun Cai, Cesare Alippi, Ming Jin, and Shirui Pan. Timeomni-vl: Unified models for time series understanding and generation. *arXiv preprint arXiv:2602.17149*, 2026.

Cheng He, Xu Huang, Gangwei Jiang, Zhaoyi Li, Defu Lian, Hong Xie, Enhong Chen, Patrick PC Lee, et al. Gtm: A general time-series model for enhanced representation learning of time-series data. In *International Conference on Learning Representations*, volume 2026, pp. 7280–7309, 2026.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 15979–15988. IEEE, 2022.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. Timefilter: Patch-specific spatial-temporal graph filtration for time series forecasting. *arXiv preprint arXiv:2501.13041*, 2025.

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.

Mert Kayaalp, Caner Turkmen, Oleksandr Shchur, Pedro Mercado, Abdul Fatir Ansari, Michael Bohlke-Schneider, and Bernie Wang. Test-time efficient pretrained model portfolios for time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 16739–16772, 2026.

Hao Li, Lu Zhang, Liu Chong, Yankai Chen, Pengyang Wang, and Yingjie Zhou. Seesawnet: Towards non-stationary time series forecasting with balanced modeling of common and specific dependencies. *arXiv preprint arXiv:2605.14551*, 2026.

Yan Li, Xinjiang Lu, Yaqing Wang, and Dejing Dou. Generative time series forecasting with diffusion, denoise, and disentanglement. *Advances in Neural Information Processing Systems*, 35: 23009–23022, 2022.

Hao Liu, Chun Yang, Xiaobin Zhu, et al. Semantic-enhanced time-series forecasting via large language models. In *International Conference on Learning Representations*, volume 2026, pp. 101404–101425, 2026a.

Kun Liu, Renjun Jia, Ruifeng Yang, Xirui Zeng, Yuqi Liang, and Cen Chen. Are global dependencies necessary? scalable time series forecasting via local cross-variate modeling. In *International Conference on Learning Representations*, volume 2026, pp. 102794–102817, 2026b.

Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18915–18923, 2025a.

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

Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer: Generative pre-trained transformers are large time series models. *arXiv preprint arXiv:2402.02368*, 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, 2025b.

Jiecheng Lu, Yan Sun, and Shihao Yang. In-context time series predictor. In *International Conference on Learning Representations*, volume 2025, pp. 79882–79908, 2025.

Sisuo Lyu, Siru Zhong, Weilin Ruan, Qingxiang Liu, Qingsong Wen, Hui Xiong, and Yuxuan Liang. Occamvts: Distilling vision models to 1% parameters for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 24216–24225, 2026.

Jiaming Ma, Qihe Huang, Haofeng Ma, Guanjun Wang, Sheng Huang, Zhengyang Zhou, Pengkun Wang, Xu Wang, Binwu Wang, and Yang Wang. Phat: Modeling period heterogeneity for multivariate time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 135659–135685, 2026.

Xiangkai Ma, Han Zhang, Wenzhong Li, and Sanglu Lu. Temporal-visual semantic alignment: A unified architecture for transferring spatial priors from vision models to zero-shot temporal tasks. *arXiv preprint arXiv:2511.19856*, 2025a.

Xiaowen Ma, Zhenliang Ni, Shuai Xiao, and Xinghao Chen. Timepro: Efficient multivariate long-term time series forecasting with variable-and time-aware hyper-state. *arXiv preprint arXiv:2505.20774*, 2025b.

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

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

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD conference on knowledge discovery and data mining v. 1*, pp. 1185–1196, 2025.

Weilin Ruan, Siru Zhong, Haomin Wen, and Yuxuan Liang. Vision-enhanced time series forecasting via latent diffusion models. *arXiv preprint arXiv:2502.14887*, 2025.

ChengAo Shen, Wenchao Yu, Ziming Zhao, Dongjin Song, Wei Cheng, Haifeng Chen, and Jingchao Ni. Multi-modal view enhanced large vision models for long-term time series forecasting. *Advances in Neural Information Processing Systems*, 38:93269–93296, 2026.

Lefei Shen, Mouxiang Chen, Xu Liu, Han Fu, Xiaoxue Ren, Jianling Sun, Zhuo Li, and Chenghao Liu. Visionts++: Cross-modal time series foundation model with continual pre-trained vision backbones. *arXiv preprint arXiv:2508.04379*, 2025.

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.

Yanru Sun, Zongxia Xie, Emadeldeen Eldele, Dongyue Chen, Qinghua Hu, and Min Wu. Learning pattern-specific experts for time series forecasting under patch-level distribution shift. *Advances in Neural Information Processing Systems*, 38:91810–91844, 2026.

Huiqiang Wang, Jieming Shi, and Li Qing. Fact: Fine-grained across-variable convolution for multivariate time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026.

Yanlong Wang, Hang Yu, Jian Xu, Fei Ma, Hongkang Zhang, Tongtong Feng, Zijian Zhang, Shao-Lun Huang, Danny Dongning Sun, and Xiao-Ping Zhang. Vfem: Visual feature empowered multivariate time series forecasting with cross-modal fusion. *arXiv preprint arXiv:2510.03244*, 2025a.

Yihang Wang, Yuying Qiu, Peng Chen, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. Lightgts: A lightweight general time series forecasting model. *arXiv preprint arXiv:2506.06005*, 2025b.

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

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

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

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.

Jiyuan Xu, Wenyu Zhang, Xin Jing, Jiahao Nie, Shuai Chen, and Shuai Zhang. Cpiri: Channel permutation-invariant relational interaction for multivariate time series forecasting. In *International Conference on Learning Representations*, volume 2026, pp. 145879–145900, 2026.

Luoxiao Yang, Yun Wang, Xinqi Fan, Israel Cohen, Jingdong Chen, and Zijun Zhang. Vitime: Foundation model for time series forecasting powered by vision intelligence. *arXiv preprint arXiv:2407.07311*, 2024.

Siru Zhong, Weilin Ruan, Ming Jin, Huan Li, Qingsong Wen, and Yuxuan Liang. Time-vlm: Exploring multimodal vision-language models for augmented time series forecasting. *arXiv preprint arXiv:2502.04395*, 2025.

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

Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Jian Tang, Jiantao Su, Yu Li, Ji Liu, Junfu Lyu, Yanjun Ma, and Dejing Dou. Sdwpf: A dataset for spatial dynamic wind power forecasting over a large turbine array. *Scientific Data*, 11(1):649, 2024.

APPENDIX

A EXPERIMENTAL DETAILS AND REPRODUCIBILITY

This section provides the experimental details used to reproduce the results reported in the main paper. The evaluation covers eight multivariate forecasting benchmarks and four prediction horizons $H \in \{96, 192, 336, 720\}$. The visual backbone is a frozen MAE-Base ViT-B/16 (He et al., 2022). Neither TCR nor TVTP introduces additional model parameters, and no downstream gradient update is performed during forecasting.

A.1 DATASETS

We evaluate PRISM on eight multivariate forecasting benchmarks: ETTh1, ETTh2, ETTm1, ETTm2, Weather, SPSS, Wind, and SDWPF (Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a; Chen & Xu, 2022; Li et al., 2022; Zhou et al., 2024). For the four ETT datasets, we follow the standard chronological split of 12/4/4 months for training, validation, and testing. Weather, SPSS, Wind, and SDWPF are chronologically split with a ratio of 7:1:2. Table 4 summarizes the dataset information.

Table 4: Summary of the eight multivariate forecasting benchmarks. P denotes the periodicity used for temporal folding in PRISM. All benchmarks are evaluated at $H \in \{96, 192, 336, 720\}$.

Dataset	Domain	Sampling	P	Reference
ETTh1	Electricity	1 hour	24	(Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a)
ETTh2	Electricity	1 hour	24	(Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a)
ETTm1	Electricity	15 min	96	(Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a)
ETTm2	Electricity	15 min	96	(Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a)
Weather	Meteorology	10 min	144	(Wu et al., 2021; Nie et al., 2022; Liu et al., 2024a)
SPSS	Solar power	15 min	96	(Chen & Xu, 2022)
Wind	Wind power	15 min	96	(Li et al., 2022)
SDWPF	Wind power	10 min	144	(Zhou et al., 2024)

A.2 BASELINES

The comparison contains two groups of forecasting methods. The zero-shot pretrained group includes Timer (Liu et al., 2024b), Timer-XL (Liu et al., 2025b), $\text{MOIRAI}_{\text{Base}}$ (Woo et al., 2024), $\text{LightGTS}_{\text{Tiny}}$ (Wang et al., 2025b), and $\text{MOMENT}_{\text{Large}}$ (Goswami et al., 2024). The few-shot group, trained with 10% of the downstream training data, contains SE-LLM (Liu et al., 2026a), TFPS (Sun et al., 2026), DUET (Qiu et al., 2025), CALF (Liu et al., 2025a), and iTransformer (Liu et al., 2024a). For the complete numerical comparison, the evaluation setting associated with each baseline is kept consistent with the main-paper protocol.

A.3 ZERO-SHOT INFERENCE AND VALIDATION-BASED CONFIGURATION

PRISM operates on the observed multivariate history without updating the pretrained MAE. Each forecasting window is normalized using statistics from its own observed history together with the fixed normalization constant, and is then folded according to periodicity P. TCR decomposes the folded sequence into the temporal level, within-period phase variation, and inter-period variation. These components are rendered along visual directions derived once from the frozen MAE patch embedding. The resulting coordinate system is fixed across downstream datasets and samples.

For multivariate forecasting, TVTP constructs a window-dependent variable topology from the normalized historical observations using absolute Pearson correlation. For each variable, the effective neighborhood size is $k_{\text{eff}} = \min(k, n-1)$, and information is propagated only between spatially corresponding visible MAE tokens. The update is residual and controlled by the propagation strength β; when the gate is enabled, the strength is additionally modulated by the strongest retained correlation. The corrected visible tokens then follow the standard MAE decoder path, which inserts learned mask tokens at the masked future positions before reconstruction.

Table 5: Selected and frozen PRISM configurations for all dataset–horizon settings. L denotes the input length and P the folding periodicity. λ_p and λ_c control phase and cycle rendering, P_{ref} determines the phase displacement, k is the nominal TVTP neighborhood size, and β is the residual propagation strength. Gate indicates whether correlation-based confidence modulation is enabled (1: enabled; 0: disabled). The effective neighborhood size is $k_{\mathrm{eff}} = \min(k, n-1)$ for n variables. When $\beta = 0$, TVTP propagation is inactive regardless of the Gate value. All configurations are selected using validation MSE and fixed before final test evaluation.

Dataset	H	L	P	λ_p	λ_c	P_{ref}	k	β	Gate
ETTh1	96	2880	24	0.00	0.00	32	1	0.04	0
ETTh1	192	2880	24	0.00	0.00	32	1	0.04	0
ETTh1	336	2304	24	0.00	0.00	16	1	0.04	1
ETTh1	720	2304	24	0.10	0.05	16	2	0.00	1
ETTh2	96	1728	24	0.09	0.00	32	3	0.05	1
ETTh2	192	1728	24	0.04	0.00	16	3	0.08	1
ETTh2	336	1728	24	0.05	0.02	16	2	0.01	1
ETTh2	720	1152	24	0.10	0.00	16	1	0.04	1
ETTm1	96	2304	96	0.00	0.00	16	2	0.06	0
ETTm1	192	2304	96	0.00	0.00	16	3	0.08	0
ETTm1	336	4032	96	0.00	0.00	16	2	0.06	0
ETTm1	720	4032	96	0.00	0.00	16	2	0.06	0
ETTm2	96	2304	96	0.00	0.00	16	2	0.06	0
ETTm2	192	2304	96	0.00	0.00	24	3	0.07	0
ETTm2	336	4032	96	0.00	0.00	16	2	0.06	0
ETTm2	720	4032	96	0.00	0.00	16	2	0.06	0
Weather	96	4032	144	0.01	0.05	48	4	0.08	0
Weather	192	4032	144	0.01	0.05	48	4	0.08	0
Weather	336	4032	144	0.03	0.04	24	4	0.08	1
Weather	720	4032	144	0.10	0.00	32	4	0.08	0
SPSS	96	2304	96	0.07	0.06	32	1	0.05	0
SPSS	192	2304	96	0.00	0.04	32	3	0.08	0
SPSS	336	2304	96	0.00	0.04	32	3	0.08	0
SPSS	720	4032	96	0.10	0.04	48	2	0.08	1
Wind	96	2304	96	0.00	0.00	32	1	0.08	0
Wind	192	2304	96	0.07	0.00	48	3	0.08	0
Wind	336	4032	96	0.00	0.04	32	3	0.08	0
Wind	720	4032	96	0.00	0.00	32	1	0.08	0
SDWPF	96	4032	144	0.00	0.00	32	1	0.08	0
SDWPF	192	4032	144	0.00	0.00	32	1	0.08	0
SDWPF	336	4032	144	0.00	0.00	24	3	0.07	0
SDWPF	720	4032	144	0.00	0.00	32	1	0.08	0

All forecasting experiments use MSE and MAE as evaluation metrics. In this paper, zero-shot forecasting denotes inference without downstream weight updates or task-specific training; it does not imply that all non-trainable inference hyperparameters are fixed globally across datasets. For each dataset–horizon setting, multiple configurations are evaluated on the validation split, and the configuration with the lowest validation MSE is selected. The selected configuration is then frozen before testing, with no test-label-based model selection.

A.4 Selected and Frozen Hyperparameter Configuration

Table 5 reports the final configuration used for each dataset–horizon pair, including the input length L, folding periodicity P, TCR strengths λ_p and λ_c, reference resolution P_{ref}, nominal TVTP neighborhood size k, propagation strength β, and confidence-gating status. Gate$= 1$ enables correlation-based confidence modulation, whereas Gate$= 0$ uses the fixed residual strength $\alpha_i = \beta$; the effective neighborhood size is $k_{\mathrm{eff}} = \min(k, n-1)$. These configurations correspond to the lowest validation MSE obtained across repeated experimental trials and remain fixed during the final test evaluation.

A.5 Efficiency Measurement

Computational efficiency is measured under a unified profiling setting for all compared methods. All models use a fixed input length of 672, prediction horizon $H = 96$, and batch size of 32, and are evaluated on a single NVIDIA RTX 4090 GPU. Inference latency is averaged over three runs under the same profiling procedure. Peak GPU memory and FLOPs are measured under the same input configuration for every method. The reported comparison therefore reflects differences in computational cost under matched input size, prediction horizon, batch size, and hardware.

B Complete Forecasting Results

Table 6 expands the average results reported in the main paper and provides the complete horizon-level comparison over all eight benchmarks. Each entry reports MSE and MAE, where lower values indicate better forecasting accuracy. Pretrained baselines are evaluated in the zero-shot setting, whereas the few-shot baselines follow the 10% few-shot setting used in the main experiments. The complete table shows that PRISM's performance is not confined to a single prediction range: it ranks first across short, medium, and long horizons on a broad subset of the benchmarks, while the matched VisionTS comparison isolates consistent gains from the proposed structural interface.

C Controlled Analysis of PRISM

The following controlled experiments examine the contribution of PRISM's structural organization while keeping the frozen MAE backbone unchanged.

C.1 Matched Grayscale Control

To isolate the effect of PRISM's structural organization, we construct a matched grayscale control using the same frozen MAE forecasting pipeline. When both TCR and TVTP are disabled, the implementation explicitly switches to the periodic grayscale interface used by VisionTS (Chen et al., 2024): the folded temporal map is replicated across the three RGB channels, and the reconstructed output is reduced by channel averaging. Accordingly, throughout the controlled analysis, the "w/o Both" variant denotes this matched VisionTS grayscale fallback rather than the MAE-aligned level-only representation.

Table 7 compares full PRISM with this matched grayscale control. Averaged over the four prediction horizons, PRISM achieves lower MSE and MAE on all eight benchmarks, corresponding to improvements in all **16/16** dataset–metric averages. This comparison therefore isolates the benefit of the proposed temporal organization and cross-variable propagation while keeping the underlying MAE backbone fixed.

To verify that the fallback faithfully reproduces the intended VisionTS predecessor rather than a separately implemented approximation, we additionally rerun three representative $H = 720$ settings under the same frozen MAE pipeline. As shown in Table 8, the grayscale fallback reproduces the stored VisionTS reference to within 6.2×10^{-5} MSE and 1.8×10^{-5} MAE, confirming implementation-level parity.

C.2 Globally Fixed Structural Configuration

The main results use validation-selected non-trainable configurations. To test whether the performance depends on narrow per-setting structural tuning, we additionally evaluate one shared structural configuration, $(\lambda_p, \lambda_c, P_{\mathrm{ref}}, k, \beta, \mathrm{Gate}) = (0.06, 0, 16, 3, 0.08, 0)$, across all 32 dataset–horizon settings. The benchmark-specific look-back length L and folding period P remain unchanged from the main protocol.

Table 9 summarizes the resulting comparison against the validation-selected PRISM and the matched VisionTS grayscale control. Across all 32 settings, PRISM-Fixed has only 0.98% higher MSE and 0.46% higher MAE on average than the validation-selected PRISM. Despite using a single shared structural configuration, it still outperforms the matched VisionTS control in **24/32** MSE and **19/32**

Table 6: Forecasting performance across datasets and prediction horizons $H \in \{96, 192, 336, 720\}$. Pretrained baselines are evaluated in the zero-shot setting, while few-shot baselines use 10% of the downstream training data. Each entry reports MSE and MAE; lower values are better. Avg denotes the average over the four prediction horizons. The best and second-best results are highlighted in red and blue, respectively.

		Zero-shot Pretrain												Few-shot (10%)									
Dataset	Horizon	Prism (Ours)		LightGTS$_{\text{Tiny}}$ (2025b)		Timer-XL (2025b)		Timer (2024b)		MOIRAI$_{\text{Base}}$ (2024)		MOMENT$_{\text{Large}}$ (2024)		SE-LLM (2026a)		TFPS (2026)		DUET (2025)		CALF (2025a)		iTrans. (2024a)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.351**	**0.382**	0.383	0.419	**0.363**	**0.387**	0.454	0.435	0.397	0.419	0.722	0.586	0.535	0.489	0.629	0.561	0.513	0.513	0.499	0.478	0.553	0.515
	192	**0.390**	**0.410**	0.465	0.472	**0.406**	**0.410**	0.522	0.466	0.423	0.436	0.726	0.593	0.581	0.512	0.689	0.605	0.600	0.568	0.548	0.508	0.605	0.556
	336	**0.405**	**0.422**	0.499	0.498	**0.433**	**0.426**	0.559	0.484	0.436	0.446	0.720	0.597	0.606	0.530	0.717	0.596	0.669	0.604	0.600	0.542	0.648	0.567
	720	**0.404**	**0.438**	0.522	0.527	0.446	**0.450**	0.714	0.549	**0.444**	0.464	0.731	0.621	0.624	0.554	0.729	0.621	1.027	0.770	0.845	0.664	0.759	0.626
	Avg	**0.388**	**0.413**	0.467	0.479	**0.412**	**0.418**	0.562	0.484	0.425	0.441	0.725	0.599	0.587	0.521	0.691	0.596	0.702	0.614	0.623	0.548	0.641	0.566
ETTh2	96	**0.269**	**0.328**	0.309	0.365	0.291	**0.345**	0.316	0.359	0.296	0.349	0.383	0.422	0.292	0.362	0.389	0.431	**0.275**	0.363	0.317	0.385	0.359	0.409
	192	**0.322**	**0.367**	0.393	0.420	**0.353**	**0.384**	0.374	0.398	0.358	0.388	0.397	0.431	0.357	0.407	0.410	0.447	0.383	0.428	0.364	0.415	0.394	0.430
	336	**0.339**	**0.381**	0.438	0.449	0.382	**0.407**	**0.380**	0.410	0.384	0.410	0.403	0.436	0.402	0.441	0.431	0.460	0.417	0.451	0.391	0.439	0.404	0.441
	720	**0.382**	**0.418**	0.483	0.487	**0.400**	**0.427**	0.409	0.434	0.419	0.442	0.432	0.460	0.473	0.489	0.564	0.540	0.610	0.551	0.529	0.521	0.465	0.480
	Avg	**0.328**	**0.374**	0.406	0.430	**0.357**	**0.391**	0.370	0.400	0.364	0.397	0.404	0.437	0.381	0.425	0.449	0.470	0.421	0.448	0.400	0.440	0.406	0.440
ETTm1	96	**0.335**	**0.346**	0.735	0.557	**0.357**	**0.371**	0.698	0.530	0.556	0.472	0.689	0.551	0.492	0.455	0.528	0.487	0.407	0.426	0.409	0.424	0.425	0.435
	192	**0.354**	**0.359**	0.761	0.584	**0.419**	**0.405**	0.744	0.556	0.598	0.492	0.693	0.554	0.546	0.479	0.621	0.519	0.455	0.451	0.454	0.451	0.471	0.461
	336	**0.371**	**0.372**	0.772	0.594	**0.475**	**0.436**	0.801	0.582	0.619	0.500	0.702	0.559	0.607	0.504	0.666	0.529	0.500	0.478	0.488	0.470	0.523	0.492
	720	**0.410**	**0.404**	0.791	0.613	0.568	**0.486**	0.829	0.606	0.682	0.528	0.718	0.570	0.723	0.549	0.680	0.539	0.597	0.539	**0.534**	0.491	0.598	0.529
	Avg	**0.368**	**0.370**	0.765	0.587	**0.455**	**0.425**	0.768	0.569	0.614	0.498	0.701	0.559	0.592	0.497	0.624	0.519	0.490	0.474	0.471	0.459	0.504	0.479
ETTm2	96	0.226	**0.282**	0.245	0.327	**0.199**	**0.279**	0.225	0.300	0.215	0.294	0.310	0.366	0.205	0.285	0.264	0.341	**0.197**	**0.279**	**0.197**	**0.279**	0.218	0.301
	192	0.259	**0.304**	0.338	0.382	0.256	0.317	0.286	0.339	0.278	0.335	0.336	0.380	**0.253**	0.317	0.304	0.365	**0.250**	**0.311**	0.254	0.318	0.269	0.333
	336	**0.291**	**0.328**	0.402	0.416	0.307	0.349	0.335	0.369	0.334	0.369	0.367	0.397	0.307	0.352	0.338	0.379	**0.298**	**0.347**	0.307	0.351	0.313	0.361
	720	**0.341**	**0.370**	0.452	0.447	0.399	0.404	0.414	0.416	0.527	0.442	0.424	0.433	0.408	0.412	0.416	0.423	0.388	0.404	**0.386**	**0.401**	0.395	0.408
	Avg	**0.279**	**0.321**	0.359	0.393	0.290	0.337	0.315	0.356	0.339	0.360	0.359	0.394	0.293	0.342	0.331	0.377	**0.283**	**0.335**	0.286	0.337	0.299	0.351
SPSS	96	**0.126**	**0.187**	0.694	0.648	0.377	0.401	1.788	1.002	0.164	**0.265**	0.868	0.647	**0.156**	0.271	0.206	0.326	0.224	0.333	0.194	0.296	0.158	0.277
	192	**0.157**	**0.213**	0.771	0.704	0.454	0.454	1.381	0.880	0.214	0.317	0.838	0.647	0.219	0.316	0.306	0.389	0.286	0.364	0.232	0.328	**0.192**	**0.294**
	336	**0.185**	**0.235**	0.804	0.726	0.457	0.468	1.179	0.800	0.237	0.344	0.850	0.669	0.237	0.325	0.331	0.412	0.268	0.363	0.235	0.331	**0.205**	**0.304**
	720	**0.209**	**0.259**	0.856	0.758	0.574	0.547	1.078	0.779	0.284	0.379	0.809	0.693	0.283	0.353	0.378	0.437	0.310	0.366	0.257	0.343	**0.250**	**0.336**
	Avg	**0.169**	**0.224**	0.781	0.709	0.465	0.467	1.357	0.865	0.225	0.326	0.841	0.664	0.224	0.316	0.305	0.391	0.272	0.356	0.230	0.325	**0.201**	**0.303**
SDWPF	96	0.824	0.541	0.817	0.604	**0.614**	**0.458**	0.961	0.637	**0.622**	**0.470**	1.085	0.721	0.687	0.526	0.968	0.681	0.807	0.585	0.797	0.581	0.777	0.575
	192	0.850	0.557	0.934	0.676	**0.726**	**0.519**	0.937	0.632	**0.754**	**0.543**	1.035	0.700	0.794	0.587	1.070	0.710	0.954	0.658	1.021	0.684	0.932	0.643
	336	0.877	**0.580**	1.079	0.743	**0.841**	**0.570**	0.992	0.657	**0.834**	0.598	1.037	0.700	0.891	0.634	1.186	0.763	1.110	0.713	1.019	0.693	0.982	0.675
	720	0.938	**0.623**	1.202	0.805	**0.918**	**0.614**	1.026	0.685	**0.904**	0.641	0.975	0.675	1.096	0.714	1.323	0.810	1.117	0.739	1.055	0.711	1.041	0.716
	Avg	0.872	0.575	1.008	0.707	**0.775**	**0.540**	0.979	0.653	**0.779**	**0.563**	1.033	0.699	0.867	0.615	1.137	0.741	0.997	0.674	0.973	0.667	0.933	0.652
Weather	96	0.214	0.253	0.231	0.293	**0.168**	**0.223**	0.265	0.315	0.213	0.241	0.424	0.422	**0.178**	**0.225**	0.207	0.261	0.234	0.289	0.224	0.281	0.224	0.281
	192	0.238	0.271	0.310	0.351	**0.218**	**0.270**	0.285	0.327	0.299	0.300	0.419	0.417	**0.228**	**0.268**	0.286	0.321	0.293	0.333	0.274	0.318	0.287	0.326
	336	**0.274**	**0.296**	0.407	0.416	**0.290**	**0.315**	0.358	0.370	0.382	0.353	0.431	0.423	0.315	0.328	0.348	0.358	0.333	0.354	0.362	0.372	0.354	0.368
	720	**0.323**	**0.334**	0.475	0.456	0.379	0.388	0.426	0.419	0.637	0.424	0.400	0.404	**0.376**	**0.387**	0.529	0.432	0.419	0.421	0.448	0.432	0.412	0.410
	Avg	**0.262**	**0.288**	0.356	0.379	**0.264**	**0.299**	0.333	0.358	0.382	0.329	0.418	0.416	0.274	0.302	0.342	0.343	0.320	0.349	0.327	0.351	0.319	0.346
Wind	96	1.189	0.754	1.166	0.763	**0.940**	**0.665**	1.350	0.844	**0.893**	**0.645**	1.538	0.924	0.983	0.706	1.568	0.926	1.341	0.837	1.328	0.815	1.292	0.812
	192	1.252	0.799	1.325	0.841	**1.128**	**0.753**	1.457	0.879	**1.081**	**0.734**	1.468	0.903	1.195	0.795	1.695	0.976	1.509	0.897	1.598	0.910	1.472	0.886
	336	1.320	0.831	1.499	0.908	1.349	0.834	1.572	0.922	**1.220**	**0.801**	1.488	0.915	**1.238**	**0.825**	1.814	1.012	1.631	0.936	1.505	0.890	1.515	0.905
	720	**1.247**	**0.825**	1.625	0.964	1.375	0.870	1.490	0.914	**1.281**	**0.837**	1.410	0.889	1.319	0.858	1.836	1.004	1.710	0.952	1.517	0.903	1.476	0.893
	Avg	1.252	0.802	1.404	0.869	1.198	**0.781**	1.467	0.890	**1.119**	**0.754**	1.476	0.908	**1.184**	0.796	1.728	0.980	1.548	0.905	1.487	0.880	1.439	0.874

Table 7: Controlled comparison between PRISM and the matched periodic grayscale VisionTS interface, averaged over prediction horizons $H \in \{96, 192, 336, 720\}$. Both methods use the same frozen MAE forecasting pipeline; lower values are better.

Variant	ETTh1		ETTh2		ETTm1		ETTm2		SPSS		SDWPF		Weather		Wind	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PRISM	**0.3878**	**0.4130**	**0.3282**	**0.3735**	**0.3677**	**0.3703**	**0.2793**	**0.3209**	**0.1693**	**0.2235**	**0.8725**	**0.5754**	**0.2622**	**0.2885**	**1.2521**	**0.8022**
VisionTS	0.3895	0.4141	0.3331	0.3746	0.3735	0.3716	0.2814	0.3213	0.1728	0.2247	0.8994	0.5773	0.2684	0.2921	1.2845	0.8108

MAE comparisons. These results demonstrate that a shared structural configuration maintains performance close to the validation-selected configuration, supporting the robustness of PRISM to structural hyperparameter choices.

C.3 MATCHED CONTEXT-LENGTH CONTROL

We further test whether the gains over VisionTS can be explained by access to longer historical windows. Table 10 compares PRISM and VisionTS using identical look-back lengths at $H = 720$

Table 8: VisionTS parity check under the same MAE forecasting pipeline. The rerun disables TCR augmentation and TVTP; Δ is the absolute difference from the stored VisionTS reference.

Setting	Rerun MSE	Ref. MSE	Δ MSE	Rerun MAE	Ref. MAE	Δ MAE
ETTh1-720	0.406159	0.406158	1.2e-06	0.440971	0.440968	3.0e-06
SPSS-720	0.214221	0.214221	0.0e+00	0.260986	0.260986	0.0e+00
SDWPF-720	0.966228	0.966289	6.1e-05	0.625811	0.625828	1.7e-05

Table 9: Globally fixed structural-parameter control. PRISM-Fixed uses $(\lambda_p, \lambda_c, P_{\mathrm{ref}}, k, \beta, \mathrm{Gate}) = (0.06, 0, 16, 3, 0.08, 0)$ for all 32 settings, while the benchmark-specific look-back length L and folding period P follow the main protocol. Lower values are better.

Dataset	Fixed MSE	Fixed MAE	Selected MSE	Selected MAE	VisionTS MSE	VisionTS MAE
ETTh1	0.3918	0.4148	0.3878	0.4130	0.3895	0.4141
ETTh2	0.3302	0.3770	0.3282	0.3735	0.3331	0.3746
ETTm1	0.3718	0.3708	0.3677	0.3703	0.3735	0.3716
ETTm2	0.2817	0.3229	0.2793	0.3209	0.2814	0.3213
SDWPF	0.8848	0.5775	0.8725	0.5754	0.8994	0.5773
SPSS	0.1714	0.2246	0.1693	0.2235	0.1728	0.2247
Weather	0.2646	0.2899	0.2622	0.2885	0.2684	0.2921
Wind	1.2633	0.8047	1.2521	0.8022	1.2845	0.8108

on ETTh1, ETTm1, and Weather. Across the nine matched comparisons, PRISM improves MSE in **7/9** cases and MAE in **8/9**, with average relative gains of **0.62%** and **0.65%**, respectively. PRISM improves on ETTm1 and Weather at all three context scales, while ETTh1 shows a stronger dependence on sufficient history length. Overall, the gains persist across multiple context scales and cannot be attributed to long look-back windows alone.

C.4 TEMPORAL-FACTOR ANALYSIS OF TCR

To examine the role of complementary temporal information, we start from the backbone-aligned level representation and progressively introduce the within-period phase and inter-period cycle components. Results on representative settings are reported in Table 11. The individual phase and cycle components are not uniformly dominant on every dataset, which is expected because their usefulness depends on the local temporal structure. Their combination is substantially more stable: full TCR improves over the level-only representation in 4 of 5 MSE comparisons and all 5 MAE comparisons, with mean relative improvements of 0.83% and 0.42%, respectively.

This behavior supports the preserve-and-augment design of TCR. The temporal level remains explicitly recoverable, while phase and cycle provide complementary information whose relative importance varies with the temporal process. Accordingly, TCR provides a structured visual interface that preserves the level coefficient during rendering while exposing complementary temporal factors whose contributions vary across forecasting settings.

C.5 EFFECTIVENESS OF TVTP

We isolate the effect of cross-variable relational propagation using matched controlled reruns in which the TCR representation and all frozen settings are held fixed and only TVTP is enabled or disabled. Table 12 summarizes all 32 dataset–horizon comparisons. TVTP improves MSE in 31 settings with one tie and no degradation, and improves MAE in 29 settings with one tie and two small degradations. The mean relative improvements are 1.57% and 0.46% for MSE and MAE, respectively. A configuration with $\beta = 0$ is a deactivated control by definition and is treated as a tie rather than as evidence of active propagation.

The matched long-horizon comparison in Table 13 shows that the benefit is especially clear on datasets with stronger multivariate coupling: MSE improves by 3.17% on SDWPF, 2.31% on Weather, 1.39% on Wind, 1.34% on SPSS, and 1.19% on ETTm1. These results establish TVTP

Table 10: Matched context-length control at $H = 720$. PRISM and VisionTS use the same look-back length L in every paired comparison; positive Δ indicates an improvement by PRISM.

Dataset	L	VisionTS MSE	PRISM MSE	Δ MSE	VisionTS MAE	PRISM MAE
ETTh1	672	0.4269	0.4343	-1.73%	0.4397	0.4408
ETTh1	1440	0.4214	0.4229	-0.35%	0.4440	0.4424
ETTh1	2880	0.4062	0.4057	+0.11%	0.4410	0.4383
ETTm1	768	0.4550	0.4511	+0.86%	0.4200	0.4188
ETTm1	1536	0.4364	0.4333	+0.71%	0.4123	0.4106
ETTm1	2304	0.4164	0.4148	+0.38%	0.4054	0.4048
Weather	720	0.3682	0.3606	+2.08%	0.3560	0.3503
Weather	2016	0.3447	0.3378	+2.00%	0.3463	0.3408
Weather	4032	0.3295	0.3244	+1.56%	0.3375	0.3338

Table 11: Temporal-factor analysis of TCR. Each cell reports MSE/MAE for controlled factor variants under the same forecasting setting.

Setting	Level only	+ Phase	+ Cycle	Full TCR
ETTh1-720	0.4076/0.4403	0.4057/0.4384	0.4078/0.4403	0.4054/0.4384
ETTh2-720	0.3845/0.4178	0.3812/0.4166	0.3859/0.4191	0.3815/0.4174
ETTm1-96	0.3342/0.3455	0.3348/0.3456	0.3363/0.3452	0.3362/0.3454
SPSS-720	0.2133/0.2612	0.2105/0.2598	0.2103/0.2611	0.2090/0.2602
Weather-192	0.2385/0.2723	0.2389/0.2720	0.2347/0.2696	0.2351/0.2692

as a highly consistent source of cross-variable improvement rather than an artifact of independent hyperparameter selection.

C.6 Confidence-Gating Control

Confidence gating is an optional modulation of TVTP rather than its primary mechanism. We compare Gate-on and Gate-off variants on the five settings whose validation-selected configurations have Gate$= 1$ and active TVTP ($\beta > 0$), while holding all other parameters fixed. Table 14 shows that gating improves MAE in all five settings, with an average gain of 0.19%, while its mean MSE effect is essentially neutral. The gate therefore acts as a small robustness correction when beneficial, whereas the dominant TVTP gains come from topology-guided token propagation itself.

D Hyperparameter Robustness

Sensitivity analyses are conducted on representative active settings while holding all remaining parameters fixed. TCR sweeps use ETTh1, ETTh2, SPSS, and Weather, whereas the TVTP sweeps use ETTh1, SPSS, and SDWPF at $H = 720$, all with active propagation. The vertical axis reports relative degradation from the best value within each local sweep, allowing datasets with different error scales to be compared on a common scale.

D.1 Sensitivity of TCR Parameters

Figure 4 reports the MSE sensitivity to λ_p, λ_c, and P_{ref}. Across the four representative settings, the average MSE sweep spans are approximately 0.79%, 0.82%, and 0.40%, respectively. Even the largest observed spread is below 1% for λ_p and P_{ref}, while the largest λ_c spread is 1.75%. The corresponding MAE trends in Figure 5 are similarly stable, with average spans of approximately 0.52%, 0.38%, and 0.33%.

D.2 Sensitivity of TVTP Parameters

Figure 6 evaluates the neighborhood size k and propagation strength β on ETTh1, SPSS, and SDWPF at $H = 720$ using active TVTP settings. The neighborhood-size sweep is highly stable: the

Table 12: Effect of TVTP over all 32 dataset–horizon settings using matched controlled reruns. Each comparison holds the TCR representation and all frozen settings fixed and changes only whether TVTP is enabled.

Metric	Win	Tie	Loss	Mean Δ
MSE	31/32	1/32	0/32	1.57%
MAE	29/32	1/32	2/32	0.46%

Table 13: Controlled TVTP ablation at the longest forecasting horizon $H = 720$ across all eight benchmarks. TCR-only and TCR+TVTP are matched controlled reruns under the same frozen configuration; lower values are better. Δ denotes relative improvement from TCR-only to TCR+TVTP.

Dataset	TCR-only MSE	TCR+TVTP MSE	Δ	TCR-only MAE	TCR+TVTP MAE	Δ
ETTh1	0.4058	0.4057	+0.02%	0.4384	0.4383	+0.02%
ETTh2	0.3851	0.3851	+0.00%	0.4204	0.4204	+0.00%
ETTm1	0.4198	0.4148	+1.19%	0.4056	0.4048	+0.20%
ETTm2	0.3432	0.3410	+0.63%	0.3700	0.3696	+0.12%
Weather	0.3321	0.3244	+2.31%	0.3380	0.3338	+1.23%
SPSS	0.2130	0.2102	+1.34%	0.2612	0.2608	+0.15%
Wind	1.2765	1.2588	+1.39%	0.8328	0.8275	+0.64%
SDWPF	0.9833	0.9522	+3.17%	0.6306	0.6261	+0.71%

average MSE and MAE spans are only **0.29%** and **0.38%**, with maxima below **0.68%**. In contrast, the preferred propagation strength is more dataset-dependent. The average β sweep spans are **2.10%** MSE and **0.47%** MAE, with the largest MSE spread of **3.98%** on SDWPF. This contrast is informative: TVTP does not depend on a narrowly chosen neighbor count, while the amount of relational correction naturally tracks the strength of dataset-specific cross-variable structure.

E Computational Efficiency

TCR and TVTP introduce no additional trainable modules and leave the pretrained MAE backbone unchanged. Detailed efficiency results are reported in Table 3 of the main paper, where PRISM has an inference time of 0.151 s and requires 145.49 GFLOPs.

F Additional Representation Analysis

F.1 Patch-Embedding Response Geometry

The MAE-aligned coordinates are constructed from $\mathbf{G} = \mathbf{W}_{\mathrm{DC}}^{\top}\mathbf{W}_{\mathrm{DC}}$ and therefore depend on the frozen MAE patch embedding rather than on a downstream dataset. For two distinct eigen-directions $\mathbf{u}_i$ and $\mathbf{u}_j$,

$$(\mathbf{W}_{\mathrm{DC}}\mathbf{u}_i)^{\top}(\mathbf{W}_{\mathrm{DC}}\mathbf{u}_j) = \mathbf{u}_i^{\top}\mathbf{G}\mathbf{u}_j = 0.$$

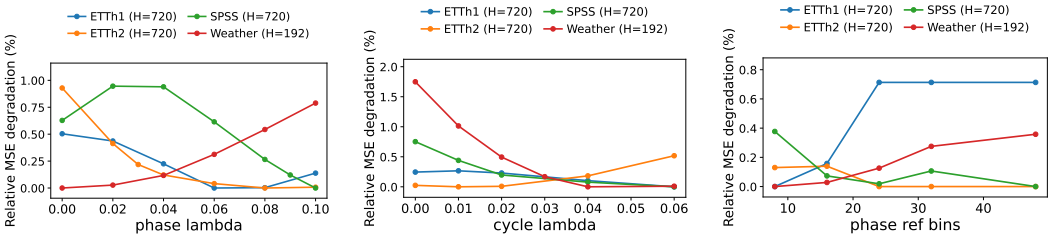
Thus, orthogonality of the spatially constant patch-embedding responses is an analytical property of the coordinate construction. Table 15 numerically verifies this identity and contrasts it with canonical RGB coordinates, whose responses need not be orthogonal under the same operator. Under canonical RGB coordinates, the Level–Phase and Phase–Cycle cosine similarities are -0.581 and -0.637, respectively, whereas the absolute off-diagonal cosine similarities under the MAE-aligned coordinates are below 10^{-6}. The table should therefore be interpreted as a verification of the constructed coordinate geometry rather than as an independent empirical test on downstream time-series samples.

F.2 Empirical Real-Token Geometry

The analytical result above concerns spatially constant RGB directions. To test whether the same organization affects representations induced by actual time-series inputs, we construct level, phase, and cycle components from real windows of ETTh1, SPSS, and Weather and measure the mean

Table 14: Confidence-gating control on the five settings whose validation-selected configurations have Gate$= 1$ and active TVTP ($\beta > 0$). All other settings are fixed; positive Δ indicates improvement from enabling the gate.

Setting	Gate-off MSE	Gate-on MSE	Δ MSE	Gate-off MAE	Gate-on MAE	Δ MAE
ETTh1-96	0.3553	0.3554	-0.04%	0.3831	0.3823	+0.20%
ETTh2-96	0.2711	0.2711	+0.01%	0.3286	0.3282	+0.11%
ETTh2-192	0.3240	0.3238	+0.08%	0.3683	0.3673	+0.27%
ETTh2-336	0.3419	0.3418	+0.03%	0.3821	0.3820	+0.04%
SPSS-720	0.2101	0.2102	-0.01%	0.2617	0.2608	+0.36%



(a) Phase strength λ_p (b) Cycle strength λ_c (c) Reference resolution P_{ref}

Figure 4: MSE sensitivity of TCR parameters. The curves report relative degradation from the best value within each local sweep. Performance remains stable across broad ranges of the rendering hyperparameters.

absolute pairwise cosine similarity of their responses under canonical RGB and MAE-aligned TCR. Measurements are taken after the patch embedding and after encoder layers 3, 6, 9, and 12.

Table 16 shows a substantial reduction in cross-component coupling at the visual interface: the average cosine decreases from 0.0578 to 0.0211, corresponding to a **63.5% reduction**. Transformer blocks subsequently remix the components, as expected for contextual processing, but the advantage is again pronounced at the final encoder representation, where the mean coupling decreases from 0.3197 to 0.2721 (**14.9% reduction**). The depth-wise curves in Fig. 8 therefore provide empirical evidence that the proposed coordinate system changes real MAE representations beyond the analytical $\mathbf{W}_{\text{DC}}$ construction.

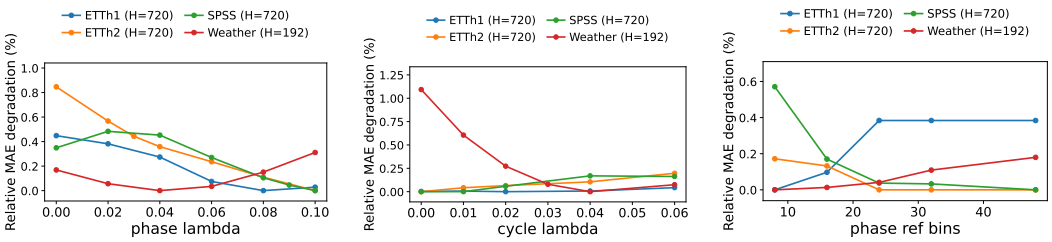
F.3 Additional Activation Examples

Figure 9 provides additional activation examples on SPSS and Weather. The upper panels are dataset-dependent and visualize the actual frozen patch-embedding activation magnitude produced by the corresponding rendered time-series inputs. The accompanying cosine matrices characterize the fixed coordinate-axis geometry defined by $\mathbf{W}_{\text{DC}}$ and are therefore dataset-independent; they are included as a reference for the coordinate construction rather than as separate dataset-specific orthogonality measurements.

G Limitations and Scope

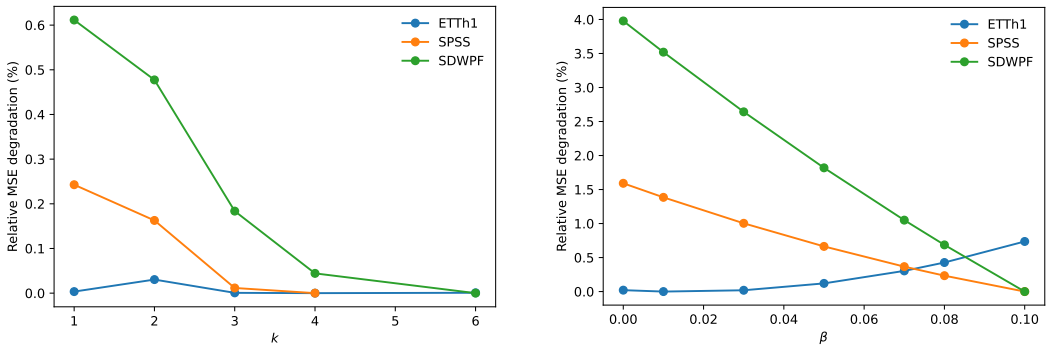
The current study focuses on a frozen MAE-Base masked-reconstruction backbone. The proposed coordinate construction is derived from the spatially constant color sensitivity of its patch-embedding layer, and the forecasting pipeline uses the pretrained MAE decoder for masked future reconstruction. Accordingly, the present results do not establish direct applicability to encoder-only vision models such as standard ViTs, CLIP, or DINO-style representations, nor do they establish identical gains across visual pretraining objectives or backbone scales.

TCR relies on periodic folding to expose within-period and inter-period structure. Its benefit can therefore vary with the strength and stability of periodic patterns, as reflected by the dataset-



(a) Phase strength λ_p (b) Cycle strength λ_c (c) Reference resolution P_{ref}

Figure 5: MAE sensitivity of TCR parameters. The trends are consistent with the MSE analysis and show limited dependence on any single hyperparameter value.



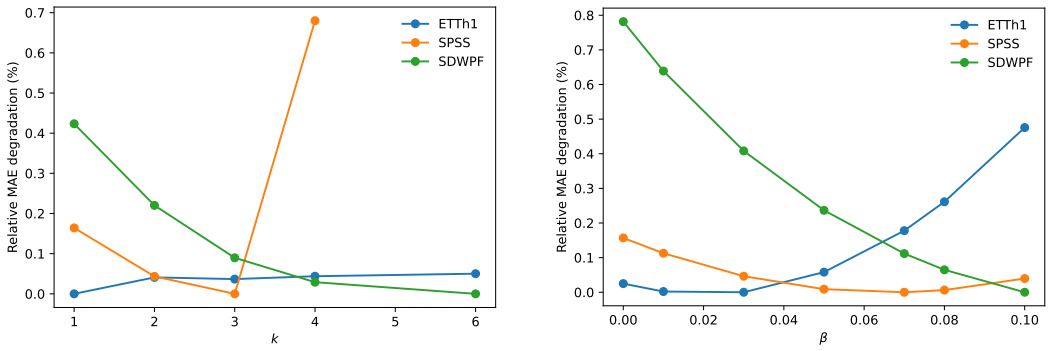
(a) Neighborhood size k (b) Propagation strength β

Figure 6: MSE sensitivity of TVTP on active $H = 720$ settings from ETTh1, SPSS, and SDWPF. Neighborhood size is highly stable, while propagation strength shows more dataset-dependent behavior.

dependent factor analysis. The current benchmark suite does not constitute a systematic evaluation of strongly aperiodic time series.

The main benchmark uses validation-selected non-trainable configurations for each dataset–horizon setting. The additional globally fixed structural control, however, incurs only 0.98% higher MSE and 0.46% higher MAE than the selected model on average across all 32 settings, showing that the method does not depend on narrowly tuned structural parameters. The look-back length L and folding period P still follow the benchmark-specific protocol; a fully uniform configuration across all datasets and sampling frequencies is not evaluated.

Finally, each variable is rendered as an independent full-resolution MAE input, so the visual encoding cost grows approximately linearly with the number of variables. TVTP additionally forms pairwise variable affinities, whose construction scales as $\mathcal{O}(n^2)$ before sparse top-k propagation. The current experiments do not comprehensively evaluate extremely high-dimensional multivariate systems, and improving this scaling behavior is an important direction for future work.



(a) Neighborhood size k (b) Propagation strength β

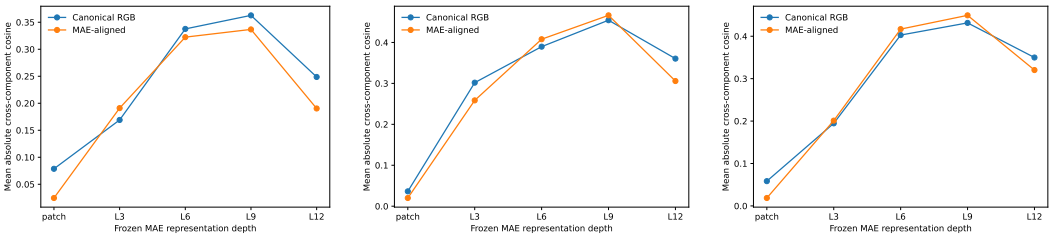
Figure 7: MAE sensitivity of TVTP on active $H = 720$ settings from ETTh1, SPSS, and SDWPF. MAE remains stable across both sparse-neighborhood and propagation-strength sweeps.

Table 15: Pairwise cosine similarity between the responses of the three rendering directions under the spatially averaged patch-embedding operator $\mathbf{W}_{\mathrm{DC}}$. Canonical RGB responses exhibit cross-direction coupling, whereas MAE-aligned responses are numerically orthogonal under this operator.

Direction pair	**Canonical RGB**	**MAE-aligned**
Level–Phase	-0.581	$< 10^{-6}$
Level–Cycle	0.122	$< 10^{-6}$
Phase–Cycle	-0.637	$< 10^{-6}$

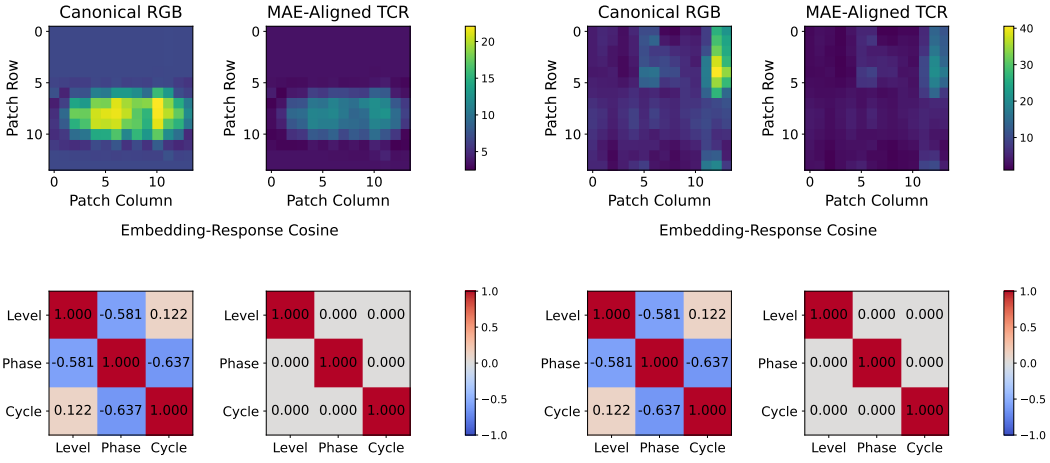
Table 16: Empirical cross-component coupling on real time-series-derived inputs, averaged over ETTh1, SPSS, and Weather and over the three component pairs. Lower mean absolute cosine similarity indicates stronger separation.

Representation stage	Canonical RGB	MAE-aligned TCR	Reduction
Patch embedding	0.0578	0.0211	+63.5%
L3	0.2218	0.2169	+2.2%
L6	0.3767	0.3824	-1.5%
L9	0.4162	0.4174	-0.3%
L12	0.3197	0.2721	+14.9%



(a) ETTh1 (b) SPSS (c) Weather

Figure 8: Real-token cross-component geometry across MAE depth. Curves compare the mean absolute cosine coupling of level, phase, and cycle responses under canonical RGB and MAE-aligned TCR. TCR strongly reduces coupling at the patch interface and again at the final encoder representation, while intermediate Transformer blocks partially remix the components.



(a) SPSS, $H = 720$ (b) Weather, $H = 192$

Figure 9: Additional activation examples on SPSS and Weather. The activation maps are input-dependent, while the coordinate-response cosine geometry is defined by the fixed spatially constant operator $\mathbf{W}_{\mathrm{DC}}$; the MAE-aligned coordinate responses are orthogonal by construction.