HISTORY-GUIDED SLICE SELF-ATTENTION FOR MITIGATING FORECAST SMOOTHING IN IRREGULAR TIME SERIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Irregular multivariate time series (IMTS) often provide sparse, asynchronous observations, leaving fine-scale future dynamics weakly constrained. Although diffusion models can generate fine-scale temporal structure, mixing across distinct dynamics in existing IMTS forecasting architectures can suppress temporal variation, leading to forecast smoothing with high-frequency attenuation. Grouping based on noisy query states can amplify such dynamics mixing, motivating history-guided grouping that maintains distinct slice representations and preserves structured variation. To address this, we propose Diffusion Coordinate Transformer with history-guided Slice self-attention (DiCTS) for joint probabilistic IMTS forecasting. Its denoiser employs History-guided Slice Self-Attention (HiSSA), which derives query-to-slice assignments from encoded history and query coordinates rather than noisy query states, while aggregating sample-specific features from the current denoising states. This history-guided assignment prevents injected diffusion noise from influencing the grouping and distinguishes queries with distinct dynamics, preventing attenuation of structured variation. We demonstrate DiCTS with established IMTS benchmarks, indicating competitive probabilistic forecasting performance and point-prediction accuracy. Qualitative and spectral analyses further show that DiCTS mitigates forecast smoothing and more faithfully follows future trajectories.

1 INTRODUCTION

Time series forecasting uses past observations to predict future values in multivariate applications such as healthcare and climate monitoring, where variables may include a patient's heart rate and blood pressure or temperature and precipitation. In practice, measurements follow clinical needs, sensor availability, or variable-specific sampling schedules rather than fixed intervals shared by all variables, while missing observations leave only partial multivariate states at many timestamps. These patterns produce uneven coverage of the time–channel space, limiting the evidence available for inferring joint system dynamics. Therefore, irregular multivariate time series (IMTS) forecasting is important for leveraging real-world observations without assuming regular, simultaneous, and complete measurements.

For IMTS forecasting, recent generative models provide a flexible framework for modeling joint conditional distributions. Prior diffusion-based models can generate fine-scale temporal structure by iteratively denoising future values or their latent representations, conditioned on historical observations (Biloš et al., 2023; Chang et al., 2024; You et al., 2026). However, modeling joint predictive distributions does not by itself ensure high-fidelity temporal variation within individual trajectories. Beyond marginal predictive accuracy, these forecasts should therefore be evaluated by how faithfully they capture temporal dependence and preserve within-trajectory fluctuations (Pic et al., 2025).

Motivated by this, we focus on *forecast smoothing* within individual generated trajectories, characterized by attenuated amplitudes of temporal variation relative to the observed future (Figure 1). Prior work on probabilistic forecasting has emphasized that sample diversity alone does not ensure temporally faithful trajectories (LE GUEN & THOME, 2020). Related problems include the loss of distinct plausible future patterns under over-smoothing (Zhang et al., 2026) and the smoothing of

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.

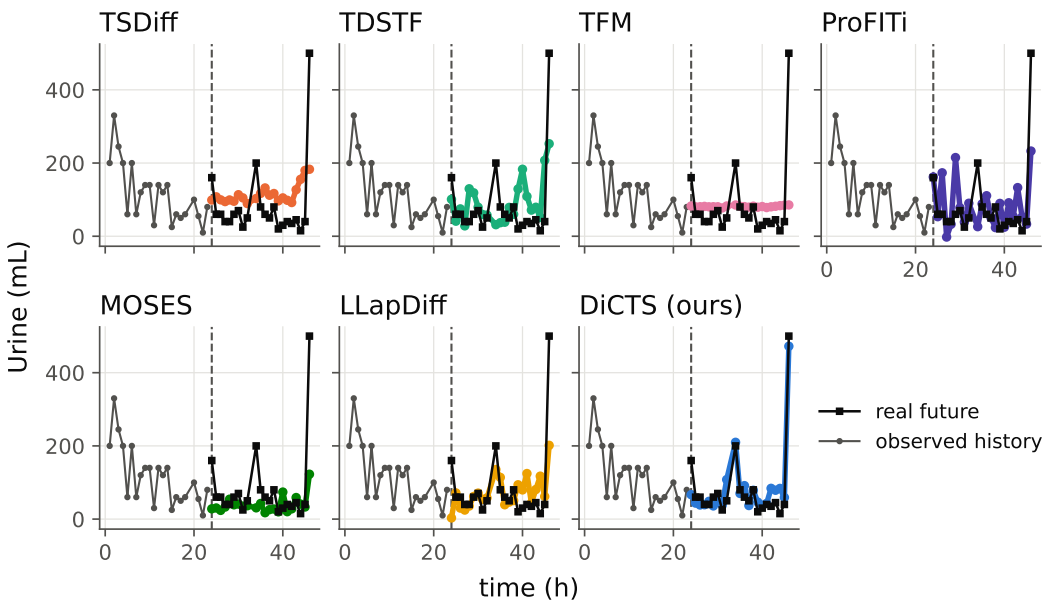


Figure 1: Illustrative example of forecast smoothing in IMTS probabilistic forecasting. The figure shows a `Urine` (mL) sequence from PhysioNet'12, with a 24 h history and a 24 h forecast horizon. Gray and black denote the observed history and future, respectively. Colored lines show model forecast trajectories. Vertical dashed lines represent the forecasting boundary. In this example, the baselines exhibit attenuated peaks and fluctuations, whereas DiCTS better captures the distinct peaks and rapid changes in the observed future.

periodic temporal structure in diffusion-based forecasting (Caldas et al., 2026). In diffusion-based IMTS forecasting, grouping queries based on noisy states may cause queries with similar noise patterns to be grouped together, even when such similarity does not reflect their underlying temporal dynamics. Aggregating their features may then obscure distinctions that are useful for generating temporally structured forecasts. This motivates a design that uses historical evidence to guide query interactions rather than relying on noisy-state similarity, while aggregating content from the current denoising states.

To this end, we propose Diffusion Coordinate Transformer with history-guided Slice self-attention (DiCTS), which jointly generates future values across query locations. Its History-guided Slice Self-Attention (HiSSA) derives query-to-slice assignments from query coordinates and historical evidence rather than noisy query states. Within each slice, HiSSA aggregates features computed from the current denoising states and then exchanges information across slices. This allows historical evidence to determine the group structure while retaining sample-specific information from the current denoising states. HiSSA aims to reduce mixing across queries with distinct temporal dynamics and help the denoiser generate individual trajectories with more faithful temporal variation.

Our contributions are as follows:

- We study forecast smoothing in diffusion-based IMTS forecasting, where temporal variation within individual generated trajectories can be attenuated, and propose DiCTS to preserve trajectory-level temporal structure during joint probabilistic forecasting.
- We introduce History-guided Slice Self-Attention (HiSSA), which derives query-to-slice assignments from historical evidence and query coordinates rather than noisy query states, while aggregating sample-specific denoising features to reduce noise-driven mixing.
- Qualitative comparisons and spectral analyses indicate that DiCTS mitigates forecast smoothing in generated trajectories and more faithfully follows ground-truth fluctuation amplitudes across frequency bands. On established IMTS benchmarks, DiCTS further achieves competitive probabilistic forecasting performance and point-prediction accuracy.

2 Related Work

2.1 Irregular Multivariate Time-Series Probabilistic Forecasting

Probabilistic IMTS forecasting models uncertainty in future values conditioned on sparse, asynchronous observations. Conditional normalizing flows provide a flexible way to model joint densities over values at queried time–channel coordinates. ProFITi (Yalavarthi et al., 2025) combines a GraFITi encoder (Yalavarthi et al., 2024) with conditional normalizing flows and invertible triangular attention to learn flexible joint predictive distributions. MOSES combines a query-separable attention encoder, correlated Gaussian-process bases, and per-query flows in a mixture model that guarantees marginalization consistency (Yalavarthi et al., 2026).

A complementary approach models trajectory evolution through continuous-time stochastic dynamics. TFM (Zhang et al., 2024b) enables simulation-free training of history-conditioned neural stochastic differential equations, using recent trajectory history to help preserve temporal couplings. It combines next-value prediction with auxiliary estimation of the time to the next observation and predictive uncertainty.

Diffusion-based methods generate probabilistic forecasts through iterative denoising conditioned on historical observations. TDSTF (Chang et al., 2024) represents observations as time–variable–value triplets and denoises future values with a history-conditioned encoder–decoder Transformer. Biloš et al. (2023) defines diffusion in function space using temporally correlated Gaussian-process or Ornstein–Uhlenbeck noise, allowing it to accommodate irregular timestamps. LLapDiff (You et al., 2026) instead performs diffusion over low-dimensional latent trajectories, using Laplace-domain mean dynamics to guide denoising and support prediction at requested timestamps.

3 Method

3.1 Problem Formulation

We formulate IMTS forecasting in a time–channel coordinate space (Yalavarthi et al., 2026; Klötergens et al., 2026a;b). Each historical observation is represented by its timestamp, channel, and observed value, while each forecast target is specified by a queried time–channel pair. Let the observed history and future query locations be

$$\mathcal{H} = \{(t_j^h, c_j^h, x_j^h)\}_{j=1}^{N_h}, \qquad \mathcal{Q} = \{(t_i^q, c_i^q)\}_{i=1}^{N_q}, \tag{1}$$

where t denotes physical time, $c \in \{1, \dots, C\}$ identifies one of C channels, and $x \in \mathbb{R}$ is the observed value at (t, c). The history contains only observed entries at their recorded timestamps, without assuming regularly spaced observations or synchronized measurements across channels. All query points occur after the observed history, $\min_i t_i^q > \max_j t_j^h$. Our goal is to learn the joint conditional distribution $p_\theta(\mathbf{x}^q \mid \mathcal{H}, \mathcal{Q})$, where $\mathbf{x}^q = (x_1^q, \dots, x_{N_q}^q) \in \mathbb{R}^{N_q}$ contains the future values corresponding to the query locations.

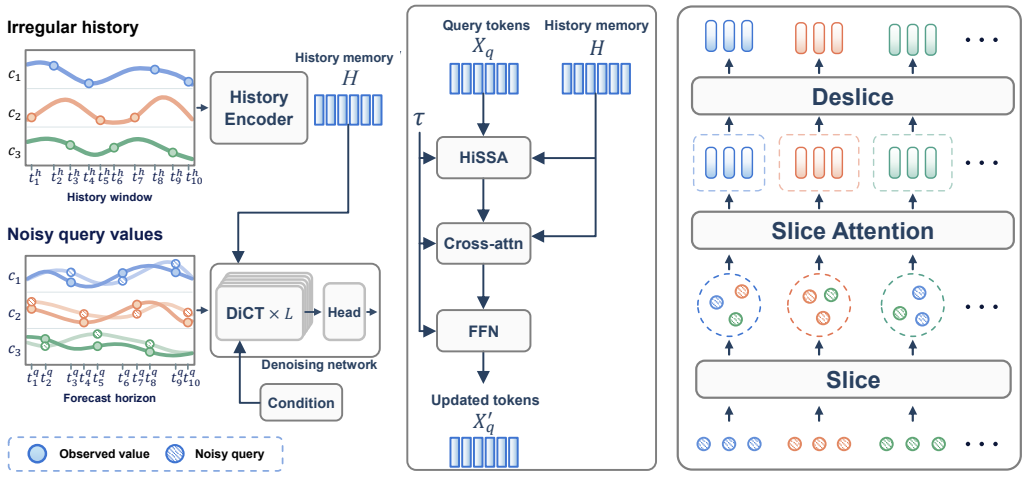
3.2 DiCTS Architecture

Adaptive layer normalization is used for conditioning in diffusion transformers (DiT) (Peebles & Xie, 2023). When history is incorporated solely through modulation parameters shared across query tokens, this pathway does not provide direct and query-dependent retrieval from individual historical tokens (Liu et al., 2024). To enable such retrieval, we retains latent history memory for cross-attention and derives adaptive-normalization parameters solely from the denoising step τ.

We encode the observed history into token-level memory and embed each query coordinate as

$$H = \mathrm{Enc}_{\mathrm{hist}}(\mathcal{H}), \qquad E_i^q = e_{\mathrm{coord}}(t_i^q, c_i^q). \tag{2}$$

The coordinate embedding e_{coord} combines a continuous-time embedding (Shukla & Marlin, 2021; Zhang et al., 2024a) with a channel embedding. Let X_τ denote the query hidden states provided to



(a) Overview (b) DiCTS (c) HiSSA

Figure 2: Overview of DiCTS. Historical observations and future query locations are represented as coordinate tokens. Historical context H and history-conditioned grouping weights guide iterative denoising to jointly generate future values across irregular times and channels. HiSSA aggregates query features using history-conditioned groups, models interactions between slices, and reconstructs features. Each denoising block combines HiSSA, history cross-attention, and a feed-forward network with τ-conditioned modulation vectors.

a DiCTS block at denoising step τ. A DiCTS block computes

$$\begin{aligned} \mathrm{mod}_i(Y,\tau) &= \mathrm{LN}_i(Y) \odot (1+\boldsymbol{\gamma}_i(\tau)) + \boldsymbol{\beta}_i(\tau), \qquad i \in \{s,c,f\}, \\ X_{s,\tau} &= X_\tau + \mathbf{g}_s(\tau) \odot \mathrm{HiSSA}(\mathrm{mod}_s(X_\tau,\tau); S, \mathcal{Q}), \\ X_{c,\tau} &= X_{s,\tau} + \mathbf{g}_c(\tau) \odot \text{Cross-Attn}(\mathrm{mod}_c(X_{s,\tau},\tau), H, H), \\ X'_\tau &= X_{c,\tau} + \mathbf{g}_f(\tau) \odot \mathrm{FFN}(\mathrm{mod}_f(X_{c,\tau},\tau)). \end{aligned} \tag{3}$$

where $\boldsymbol{\beta}_i(\tau)$, $\boldsymbol{\gamma}_i(\tau)$, and $\mathbf{g}_i(\tau)$ denote the τ-conditioned modulation vectors for shift, scale, and residual gating, respectively. The residual gates are initialized to zero, making each block an identity mapping at initialization. The indices s, c, and f correspond to the HiSSA, cross-attention, and FFN branches, respectively. Here, S denotes the slice assignments introduced in Section 3.3. History keys and values remain independent of τ, while retrieval weights can change with the evolving query states. We omit the block index for notational simplicity.

3.3 History-guided Slice Self-Attention

Recent studies have explored learned grouping mechanisms that aggregate inputs with similar representations before modeling their interactions (Wu et al., 2024; Luo et al., 2025a). However, in a diffusion-based IMTS forecasting model, grouping based on input noisy queries may cause queries with similar noise patterns to be grouped together even when they do not share underlying dynamics, potentially mixing distinct temporal structures and attenuating their variation during aggregation. To address this, we propose HiSSA, which derives query-to-slice assignments from encoded history and query coordinates rather than noisy query states, while aggregating sample-specific features from the current denoising states. This makes the grouping structure independent of injected diffusion noise, helping preserve distinct representations and structured temporal variation.

History-guided query slicing. To organize interactions among future queries, we assign them to learned slices based on their coordinates and historical context. The routing context is obtained by attending to the history memory H using query coordinate embeddings E^q:

$$\mathbf{C}^q = \text{Cross-Attn}_{\text{route}}\left(E^q, H, H\right). \tag{4}$$

The soft assignment weights over G slices are then computed as

$$\begin{aligned} \mathbf{U}_i &= \mathrm{MLP}\left([E_i^q; \mathbf{C}_i^q]\right), \\ S_{ig} &= \mathrm{softmax}_g\left(\frac{(W_S \mathbf{U}_i + \mathbf{b}_S)_g}{\delta}\right), \end{aligned} \tag{5}$$

where $\mathbf{U}_i$ represents the routing feature for query i, δ is the resulting positive temperature from $\delta = \mathrm{softplus}(\rho) + \eta_s$ with a learnable scalar ρ, and $\eta_s > 0$ is a small constant for numerical stability. S_{ig} represents the soft assignment of query i in slice g, with $\sum_{g=1}^{G} S_{ig} = 1$. Slice assignments are block- and head-specific, with these indices omitted throughout.

For fixed history and query locations, the assignments are independent of τ and $\mathbf{x}_\tau$ and can therefore be precomputed for each block and reused throughout sampling. Fixed assignments do not imply fixed interactions: slice representations and attention weights depend on the current query states.

Slice self-attention. We project the modulated query states as $F_\tau = \mathrm{mod}_s(X_\tau, \tau)W_F$ with learnable parameter W_F and define the stabilized slice mass $m_g = \max(\sum_{i=1}^{N_q} S_{ig}, \eta_m)$ with $\eta_m > 0$. Using the history-guided assignments S, we aggregate query features into slice summaries, apply self-attention across slices, and redistribute the updated summaries to queries:

$$Z_g = \frac{1}{m_g} \sum_{i=1}^{N_q} S_{ig} F_{\tau,i}, \tag{6}$$

$$\widetilde{Z}_g = \left[\text{Self-Attn}(Z)\right]_g, \tag{7}$$

$$M_i = \sum_{g=1}^{G} S_{ig} \widetilde{Z}_g. \tag{8}$$

Here, Z stacks the G slice summaries, $[\cdot]_g$ selects the g-th slice, and Self-Attn denotes self-attention with learnable query, key, and value projections. Slices with zero total assignment weight are masked out as keys. The resulting query messages are passed through a single output projection.

4 Experiments

4.1 Experimental Setup

Datasets We evaluate on five IMTS benchmarks covering climate (USHCN), healthcare (PhysioNet'12, MIMIC-III, and MIMIC-IV), and human motion (HumanActivity). These benchmarks cover diverse number of variables and observation patterns, ranging from low-dimensional climate records to high-dimensional clinical data, and from densely sampled motion signals to sparse, asynchronous clinical observations. Detailed preprocessing procedures, data splits, and forecasting settings are provided in Appendix A.

Baselines We mainly compare DiCTS against six IMTS probabilistic forecasting models, TSDiff (Biloš et al., 2023), TDSTF (Chang et al., 2024), TFM (Zhang et al., 2024b), ProFITi (Yalavarthi et al., 2025), MOSES (Yalavarthi et al., 2026), and LLapDiff (You et al., 2026). Detailed hyperparameter for baseline models are provided in Appendix B

Implementation Details All models are trained using the AdamW optimizer with an initial learning rate of 10^{-3}, a weight decay of 10^{-3}, and a batch size of 32. All experiments run for a maximum of 1,000 epochs with an early stopping patience of 30 epochs. A custom loss function proposed in the respective original paper is used. All models are trained on a Linux server with PyTorch version 2.6.0+cu124 and eight NVIDIA RTX A6000 GPUs. The detailed hyperparameter settings of DiCTS are provided in Appendix C.

4.2 Qualitative Analysis of Forecast Smoothing

Figure 3 compares the forecast trajectories of DiCTS with six IMTS probabilistic forecasting models on a PhysioNet'12 instance with four channels: Urine, HR, Temp, and NISysABP. Complete

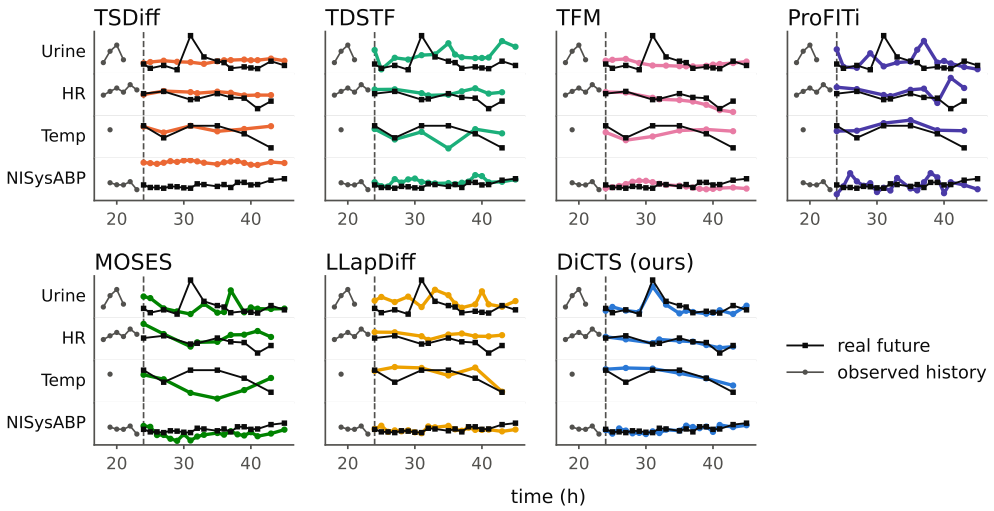


Figure 3: Qualitative comparison of forecast smoothing on PhysioNet'12. Gray and black denote observed history and ground-truth future trajectory, respectively. Colored trajectories denote model forecast trajectories. Vertical dashed lines represent the forecasting boundary. DiCTS captures the sharp Urine peak near 31 hours and its subsequent decline while following temporal changes in HR and NISysABP.

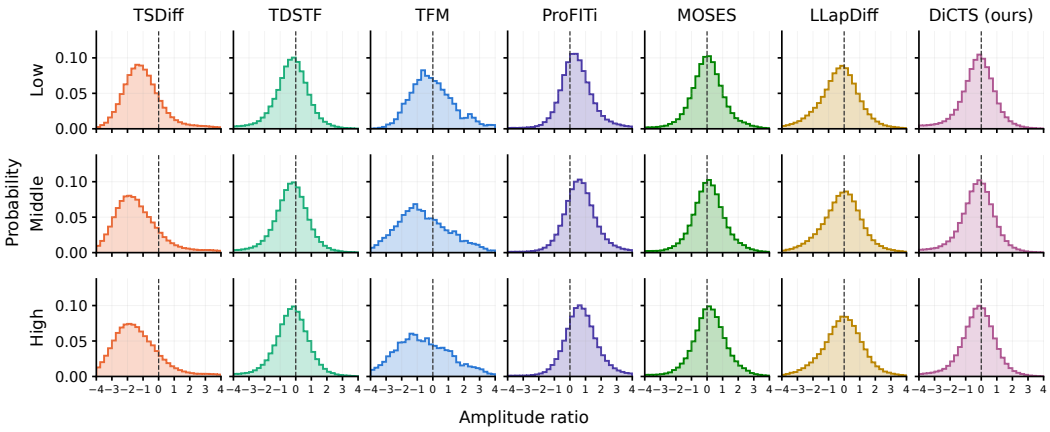
forecast trajectories for this PhysioNet'12 instance and instances from other datasets are provided in Appendix D.1. We characterize forecast smoothing as attenuated temporal fluctuations, less distinct temporal changes, or nearly flat predictions despite substantial variation in the observed future. Accordingly, we examine whether the predicted trajectories preserve the amplitude and temporal structure of the observed variation, rather than merely following broad trends.

The clearest contrast appears in Urine, where the observed future exhibits a sharp peak around 31 hours followed by a rapid decline. TSDiff and TFM produce relatively flat forecasts that miss this peak. TDSTF, MOSES, LLapDiff, and ProFITi also fail to capture the observed peak, instead producing pronounced fluctuations at other times. Forecast smoothing is also visible in HR, where TSDiff and LLapDiff produce relatively flat late forecasts that miss the pronounced decline and subsequent rebound. In NISysABP, TSDiff remains nearly flat at an elevated level, whereas TFM fails to follow the late increase. Meanwhile, the larger fluctuations produced by ProFITi in HR and NISysABP do not consistently match the observed future in location or magnitude, illustrating that greater variability alone does not guarantee faithful temporal structure.

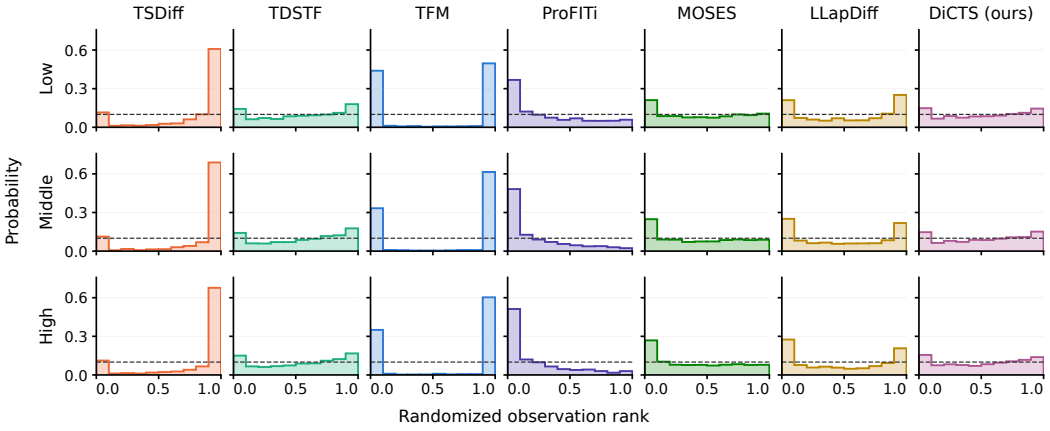
In contrast, DiCTS more closely captures the location and magnitude of the Urine peak and follows its subsequent decline. It also follows the decreases and recoveries in HR, including the late dip and rebound, and captures the overall increase together with several local fluctuations in NISysABP. For Temp, DiCTS follows the later downward trajectory, although it does not capture the initial dip after the forecasting boundary. Overall, this provides qualitative evidence of mitigated forecast smoothing, most clearly in the sharp Urine peak and the temporal changes in HR and NISysABP.

4.3 Spectral Analysis of Forecast Smoothing

To evaluate the attenuation of high-frequency fluctuations in individual forecast samples, we compare DiCTS with the baseline models using 100 joint forecast samples per instance on PhysioNet'12. Spectral analysis for other datasets are provided in Appendix D.2. Figure 4(a) plots $\log_2 r$, where $r = A_b(\widehat{x})/A_b(x)$ compares the band amplitudes of an individual forecast sample $\widehat{x}$ and the observed future x, evaluated at the same observed timestamps. Here, A_b is the square root of the mean Lomb–Scargle power within frequency band b. Ratios are computed per sample–window–channel for $A_b(x) > 0$, so that attenuation ($r < 1$) and amplification ($r > 1$) remain visible rather than being obscured by aggregation in power comparisons.



(a) Distributions of forecast-to-observation band-amplitude ratios, showing whether forecast band amplitudes are attenuated or amplified relative to observations.



(b) Rank histograms of observed band amplitudes within forecast ensembles, assessing calibration.

Figure 4: Sample-level forecast smoothing diagnostics on PhysioNet'12. Rows denote frequency bands and columns denote models. Vertical dashed lines in (a) indicate equal forecast and observed amplitudes, while horizontal dashed lines in (b) indicate the uniform-rank reference (10% per bin).

TSDiff shows a clear shift toward smaller amplitudes across all three bands, consistent with forecast smoothing. TFM also tends toward smaller amplitudes, particularly in the middle- and high-frequency bands, and its distributions are substantially broader than those of TDSTF. In contrast, TDSTF's $\log_2 r$ distributions remain relatively concentrated near zero, with a modest shift toward attenuation. ProFITi shifts toward larger amplitudes, indicating over-amplification. MOSES, LLapDiff, and DiCTS have distributions concentrated closer to zero, although this overall agreement alone does not establish ensemble calibration.

We further assess ensemble calibration using rank histograms of band amplitudes. For each eligible window–channel pair and frequency band, we compute one randomized rank of the observed amplitude among S forecast amplitudes. Unlike the ratio histograms, the rank histograms include cases with zero observed amplitude. Randomization handles ties and yields continuous ranks on $[0, 1)$. Figure 4(b) shows these ranks aggregated into ten equal-width bins, with equal total weight assigned to each channel. Under exchangeability of the observed and forecast amplitudes, each bin has an expected probability of 10%.

Table 1: Probabilistic forecasting performance on irregular multivariate time-series datasets, measured by CRPS and VS ($p = 0.5$). Lower values indicate better performance for both metrics. Values are means $\pm$ standard deviations over five seeds CPRS and VS. Bold marks the lowest mean for each dataset and metric among the reported results.

Model	Metric	HumanActivity	PhysioNet'12	USHCN	MIMIC-III	MIMIC-IV
TSDiff	CRPS	0.1714 ± 0.0199	0.3584 ± 0.0101	0.1989 ± 0.0088	0.3889 ± 0.0057	0.3219 ± 0.0017
	VS	0.0454 ± 0.0015	0.1760 ± 0.0072	0.1073 ± 0.0100	0.1526 ± 0.0025	0.1344 ± 0.0025
TDSTF	CRPS	0.1312 ± 0.0552	0.2541 ± 0.0020	0.1674 ± 0.0065	0.2603 ± 0.0028	0.1955 ± 0.0042
	VS	0.0414 ± 0.0054	0.0902 ± 0.0007	0.0880 ± 0.0080	0.1201 ± 0.0060	0.0960 ± 0.0046
TFM	CRPS	0.1175 ± 0.0021	0.3907 ± 0.0002	0.3722 ± 0.0001	0.4147 ± 0.0035	0.3068 ± 0.0019
	VS	0.0805 ± 0.0028	0.2878 ± 0.0027	0.1524 ± 0.0005	0.2412 ± 0.0050	0.2040 ± 0.0008
ProFITi	CRPS	0.0871 ± 0.0020	0.2522 ± 0.0006	0.1863 ± 0.0245	0.2655 ± 0.0013	0.1970 ± 0.0010
	VS	0.0509 ± 0.0054	0.1068 ± 0.0095	0.3478 ± 0.3507	0.1768 ± 0.0235	0.1366 ± 0.0056
MOSES	CRPS	0.0991 ± 0.0072	0.2598 ± 0.0042	0.1905 ± 0.0120	0.2988 ± 0.0046	0.2593 ± 0.0283
	VS	0.0370 ± 0.0007	0.0977 ± 0.0046	$\mathbf{0.0716 \pm 0.0079}$	0.1668 ± 0.0031	0.1661 ± 0.0058
LLapDiff	CRPS	0.1168 ± 0.0015	0.3051 ± 0.0035	0.2034 ± 0.0080	0.3775 ± 0.0068	0.2995 ± 0.0042
	VS	0.0374 ± 0.0010	0.1073 ± 0.0030	0.0916 ± 0.0066	0.2527 ± 0.0097	0.1848 ± 0.0052
DiCTS (ours)	CRPS	$\mathbf{0.0829 \pm 0.0006}$	$\mathbf{0.2487 \pm 0.0004}$	$\mathbf{0.1579 \pm 0.0023}$	$\mathbf{0.2565 \pm 0.0010}$	$\mathbf{0.1934 \pm 0.0010}$
	VS	$\mathbf{0.0321 \pm 0.0003}$	$\mathbf{0.0880 \pm 0.0014}$	0.1031 ± 0.0095	$\mathbf{0.1153 \pm 0.0020}$	$\mathbf{0.0959 \pm 0.0038}$

TSDiff concentrates rank mass near 1, consistent with the attenuation in Figure 4(a). TDSTF shows a comparatively even rank distribution, with slightly higher frequencies in the first and last bins. In contrast, TFM concentrates mass at both ends with almost no intermediate-rank mass, indicating poor ensemble calibration despite its broad log-ratio distributions. The rank histograms show higher frequencies near 0 for both ProFITi and MOSES, particularly for ProFITi. This pattern is consistent with over-amplification. LLapDiff's $\log_2 r$ distributions are centered near zero, yet its rank histograms show higher bin frequencies near both 0 and 1, indicating amplitude underestimation in some cases and overestimation in others. This pattern is consistent with insufficient ensemble spread relative to prediction errors or biases that vary across cases, showing that overall agreement in band amplitudes alone does not ensure calibration.

DiCTS combines $\log_2 r$ distributions near zero with comparatively even ranks across all three bands, although the first and last bins still show higher frequencies. These results support reduced forecast smoothing through closer agreement with observed band amplitudes rather than simply increased variability.

4.4 Irregular Multivariate Time Series Forecasting

Probabilistic forecasting. Table 1 shows that DiCTS achieves the lowest mean Continuous Ranked Probability Score (CRPS) on all five datasets. Compared with ProFITi, DiCTS consistently reduces CRPS, with the largest improvement observed on USHCN. Although the magnitude of improvement varies across datasets, these consistently lower scores support the effectiveness of DiCTS for marginal probabilistic forecasting from sparse and asynchronous histories.

To assess how well forecast trajectories preserve temporal dependence that may be attenuated by over-smoothing, we also report the Variogram Score (VS; $p = 0.5$) (Scheuerer & Hamill, 2015), computed over all observed VS compares observed pairwise absolute differences with their averages over forecast trajectories, providing a complementary assessment of within-channel temporal structure. DiCTS achieves the lowest mean VS on four of the five datasets among the reported results. These results demonstrate that DiCTS combines strong marginal forecasting performance with competitive performance in preserving temporal structure.

Deterministic forecasting. Beyond evaluating temporal structure, we examine whether DiCTS's design for mitigating forecast smoothing also benefits point-prediction accuracy. We compare DiCTS against seven deterministic models for IMTS forecasting: GraFITi (Yalavarthi et al., 2024), TimeFlow(Naour et al., 2024) Hi-Patch (Luo et al., 2025b), HyperIMTS (Li et al., 2025), ReIMTS (Li et al., 2026), APN (Liu et al., 2026b), and ASTGI (Liu et al., 2026a). We evaluate point-prediction accuracy using MAE and MSE. Table 2 shows that DiCTS achieves the lowest mean MAE on all five datasets and the lowest mean MSE on four, ranking second in MSE on Human-

Table 2: Deterministic forecasting performance on irregular multivariate time-series datasets, measured by MAE and MSE. Lower values indicate better performance. Values are means ± standard deviations over five seeds. Bold marks the lowest mean for each dataset and metric among the reported results.

Baseline	Metric	HumanActivity	PhysioNet'12	USHCN	MIMIC-III	MIMIC-IV
GraFITi	MAE	0.1161 ± 0.0017	0.3579 ± 0.0030	0.2599 ± 0.0024	0.3831 ± 0.0073	0.2899 ± 0.0007
	MSE	0.0424 ± 0.0003	0.2945 ± 0.0026	0.1900 ± 0.0051	0.4018 ± 0.0091	0.2262 ± 0.0011
TimeFlow	MAE	0.1198 ± 0.0018	0.3826 ± 0.0024	0.3074 ± 0.0239	0.4816 ± 0.0129	0.3792 ± 0.0027
	MSE	0.0448 ± 0.0021	0.3370 ± 0.0048	0.2667 ± 0.0773	0.6392 ± 0.0275	0.3712 ± 0.0084
Hi-Patch	MAE	0.1276 ± 0.0023	0.3686 ± 0.0024	0.2791 ± 0.0281	0.4158 ± 0.0040	0.2922 ± 0.0016
	MSE	0.0486 ± 0.0010	0.3037 ± 0.0021	0.2452 ± 0.0795	0.4774 ± 0.0057	0.2426 ± 0.0041
HyperIMTS	MAE	0.1128 ± 0.0007	0.3592 ± 0.0053	0.2755 ± 0.0089	0.3852 ± 0.0073	0.2860 ± 0.0033
	MSE	0.0423 ± 0.0004	0.2933 ± 0.0014	0.2331 ± 0.0647	0.4102 ± 0.0095	0.2181 ± 0.0033
ReIMTS	MAE	0.1138 ± 0.0017	0.3574 ± 0.0021	0.2678 ± 0.0073	0.5159 ± 0.0927	0.2926 ± 0.0031
	MSE	0.0427 ± 0.0003	0.2951 ± 0.0012	0.2028 ± 0.0049	0.6151 ± 0.1396	0.2306 ± 0.0026
APN	MAE	0.1173 ± 0.0002	0.3615 ± 0.0013	0.3083 ± 0.0491	0.4154 ± 0.0033	0.3092 ± 0.0023
	MSE	0.0423 ± 0.0000	0.3028 ± 0.0007	0.2448 ± 0.0448	0.4738 ± 0.0053	0.2533 ± 0.0065
ASTGI	MAE	0.1178 ± 0.0020	0.3563 ± 0.0016	0.2808 ± 0.0234	0.3903 ± 0.0052	0.2974 ± 0.0008
	MSE	$\mathbf{0.0412 \pm 0.0008}$	0.2947 ± 0.0013	0.2439 ± 0.0817	0.4246 ± 0.0052	0.2343 ± 0.0022
DiCTS (ours)	MAE	$\mathbf{0.1123 \pm 0.0009}$	$\mathbf{0.3461 \pm 0.0003}$	$\mathbf{0.2563 \pm 0.0069}$	$\mathbf{0.3545 \pm 0.0015}$	$\mathbf{0.2701 \pm 0.0012}$
	MSE	0.0418 ± 0.0009	$\mathbf{0.2857 \pm 0.0011}$	$\mathbf{0.1734 \pm 0.0139}$	$\mathbf{0.3875 \pm 0.0054}$	$\mathbf{0.2127 \pm 0.0029}$

Table 3: Ablation study on five irregular multivariate time-series benchmarks. Values are means ± sample standard deviations over five seeds.

Model	Metric	HumanActivity	PhysioNet'12	USHCN	MIMIC-III	MIMIC-IV
DiT	CRPS	0.1133 ± 0.0015	0.2778 ± 0.0014	0.1657 ± 0.0054	0.2972 ± 0.0022	0.2399 ± 0.0013
	VS	0.0358 ± 0.0005	0.0910 ± 0.0003	0.0999 ± 0.0098	0.1232 ± 0.0008	0.1193 ± 0.0015
Ours w/o HiSSA	CRPS	0.0981 ± 0.0025	0.2632 ± 0.0025	0.1617 ± 0.0009	0.2767 ± 0.0034	0.1939 ± 0.0010
	VS	0.0325 ± 0.0006	$\mathbf{0.0867 \pm 0.0002}$	0.1031 ± 0.0095	0.1153 ± 0.0020	0.0972 ± 0.0030
Ours	CRPS	$\mathbf{0.0829 \pm 0.0006}$	$\mathbf{0.2487 \pm 0.0004}$	$\mathbf{0.1579 \pm 0.0023}$	$\mathbf{0.2565 \pm 0.0010}$	$\mathbf{0.1934 \pm 0.0010}$
	VS	$\mathbf{0.0321 \pm 0.0003}$	0.0880 ± 0.0014	$\mathbf{0.0988 \pm 0.0026}$	$\mathbf{0.1132 \pm 0.0011}$	$\mathbf{0.0959 \pm 0.0038}$

Activity. These results indicate that DiCTS combines competitive point-prediction accuracy while mitigating forecast smoothing.

4.5 Ablation Study

Table 3 evaluates the contributions of the DiCT backbone and HiSSA. As a baseline, we adapt DiT to forecasting by retaining its original adaLN-Zero conditioning mechanism (Peebles & Xie, 2023) and using historical observations and the denoising step as conditioning inputs. "Ours w/o HiSSA" replaces HiSSA in DiCTS (ours) with standard query self-attention while retaining the rest of the architecture. This variant achieves lower mean CRPS and MSE than DiT on all five datasets, supporting the effectiveness of the DiCTS backbone even without HiSSA. Adding HiSSA further improves both metrics across all five datasets, supporting its advantage over standard self-attention for joint denoising.

5 Conclusion

We presented DiCTS, a coordinate-based conditional diffusion framework that jointly generates future values at arbitrary time–channel coordinates directly from irregular multivariate observations, without re-gridding or interpolation. Its HiSSA mechanism derives query-to-slice assignments from historical evidence and query coordinates independently of diffusion noise, while aggregating sample-specific denoising features within slices and exchanging information across slices. By separating the grouping criterion from the aggregated content, HiSSA helps reduce unnecessary mixing across distinct dynamics and the resulting dilution of structured temporal variation within individual generated trajectories. Experiments on IMTS benchmarks demonstrate competitive forecasting performance, while sample-wise frequency-band analysis indicates reduced forecast smoothing and closer agreement with ground-truth band-wise fluctuation amplitudes.

AI USE STATEMENT

In this work, we used generative AI tools only for polishing writing and finding related work.

We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, assisting with translation, cleaning and reformatting datasets, supporting qualitative and thematic data analysis, or interpreting results.

We have reviewed all AI-assisted work. In particular, all text generated or revised during the writing-polishing process was manually reviewed by the authors and verified to be consistent with the claims and intended meaning of the paper. Generative AI tools were used to help identify relevant literature; before incorporating any related work into the paper, the authors independently reviewed the original sources and verified their relevance and content. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Marin Biloš, Johanna Sommer, Syama Sundar Rangapuram, Tim Januschowski, and Stephan Günnemann. Neural flows: Efficient alternative to neural ODEs. In *NeurIPS*, volume 34, pp. 21325–21337, 2021.

Marin Biloš, Kashif Rasul, Anderson Schneider, Yuriy Nevmyvaka, and Stephan Günnemann. Modeling temporal data as continuous functions with stochastic process diffusion. In *ICML*, volume 202, pp. 2452–2470, 2023.

Francisco Caldas, Sahil Kumar, and Cláudia Soares. A decomposable forward process in diffusion models for time-series forecasting. *arXiv preprint arXiv:2601.21812*, 2026.

Ping Chang, Huayu Li, Stuart F. Quan, Shuyang Lu, Shu-Fen Wung, Janet Roveda, and Ao Li. A transformer-based diffusion probabilistic model for heart rate and blood pressure forecasting in intensive care unit. *Computer Methods and Programs in Biomedicine*, 246:108060, 2024.

Edward De Brouwer, Jaak Simm, Adam Arany, and Yves Moreau. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series. In *NeurIPS*, volume 32, 2019.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1):160035, 2016.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1, 2023.

Christian Klötergens, Tom Hanika, Lars Schmidt-Thieme, and Vijaya Krishna Yalavarthi. Valid and expressive copulas for irregular multivariate time series. *arXiv preprint arXiv:2605.23632*, 2026a.

Christian Klötergens, Vijaya Krishna Yalavarthi, and Lars Schmidt-Thieme. Probabilistic circuits for irregular multivariate time series forecasting. *arXiv preprint arXiv:2604.27814*, 2026b.

Vincent LE GUEN and Nicolas THOME. Probabilistic time series forecasting with shape and temporal diversity. In *NeurIPS*, volume 33, pp. 4427–4440, 2020.

Boyuan Li, Yicheng Luo, Zhen Liu, Junhao Zheng, Jianming Lv, and Qianli Ma. HyperIMTS: Hypergraph neural network for irregular multivariate time series forecasting. In *ICML*, volume 267, pp. 35502–35518, 2025.

Boyuan Li, Zhen Liu, Yicheng Luo, and Qianli Ma. Learning recursive multi-scale representations for irregular multivariate time series forecasting. In *ICLR*, 2026.

Jingwei Liu, Ling Yang, Hongyan Li, and Shenda Hong. Retrieval-augmented diffusion models for time series forecasting. In *NeurIPS*, volume 37, pp. 2766–2786, 2024.

Xvyuan Liu, Xiangfei Qiu, Hanyin Cheng, Xingjian Wu, Guo Guo, Bin Yang, and Jilin Hu. ASTGI: Adaptive spatio-temporal graph interactions for irregular multivariate time series forecasting. In *ICLR*, 2026a.

Xvyuan Liu, Xiangfei Qiu, Xingjian Wu, Zhengyu Li, Chenjuan Guo, Jilin Hu, and Bin Yang. Rethinking irregular time series forecasting: A simple yet effective baseline. *AAAI*, 40(28):23873–23881, 2026b.

Huakun Luo, Haixu Wu, Hang Zhou, Lanxiang Xing, Yichen Di, Jianmin Wang, and Mingsheng Long. Transolver++: An accurate neural solver for PDEs on million-scale geometries. In *ICML*, volume 267, pp. 41432–41449, 2025a.

Yicheng Luo, Bowen Zhang, Zhen Liu, and Qianli Ma. Hi-Patch: Hierarchical patch GNN for irregular multivariate time series. In *ICML*, 2025b.

Matthew J. Menne, Claude N. Jr. Williams, and Russell S. Vose. Long-term daily and monthly climate records from stations across the contiguous united states (u.s. historical climatology network), 2016.

Etienne Le Naour, Louis Serrano, Léon Migus, Yuan Yin, Ghislain Agoua, Nicolas Baskiotis, patrick gallinari, and Vincent Guigue. Time series continuous modeling for imputation and forecasting with implicit neural representations. *Transactions on Machine Learning Research*, 2024.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *ICCV*, pp. 4172–4182, 2023.

Romain Pic, Clément Dombry, Philippe Naveau, and Maxime Taillardat. Proper scoring rules for multivariate probabilistic forecasts based on aggregation and transformation. *NeurIPS*, 11(1): 23–58, 2025.

Michael Scheuerer and Thomas M Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321–1334, 2015.

Satya Narayan Shukla and Benjamin M Marlin. Multi-time attention networks for irregularly sampled time series. In *ICLR*, 2021.

Ikaro Silva, George Moody, Daniel J Scott, Leo A Celi, and Roger G Mark. Predicting in-hospital mortality of icu patients: The PhysioNet/Computing in cardiology challenge 2012. In *2012 Computing in Cardiology*, pp. 245–248, 2012.

Vedrana Vidulin, Mitja Lustrek, Bostjan Kaluza, Rok Piltaver, and Jana Krivec. Localization data for person activity. UCI Machine Learning Repository, 2010.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for PDEs on general geometries. In *ICML*, volume 235, pp. 53681–53705, 2024.

Vijaya Krishna Yalavarthi, Kiran Madhusudhanan, Randolf Scholz, Nourhan Ahmed, Johannes Burchert, Shayan Jawed, Stefan Born, and Lars Schmidt-Thieme. GraFITi: Graphs for forecasting irregularly sampled time series. *AAAI*, 38(15):16255–16263, 2024.

Vijaya Krishna Yalavarthi, Randolf Scholz, Stefan Born, and Lars Schmidt-Thieme. Probabilistic forecasting of irregularly sampled time series with missing values via conditional normalizing flows. *AAAI*, 39(20):21877–21885, 2025.

Vijaya Krishna Yalavarthi, Randolf Scholz, Christian Klötergens, Kiran Madhusudhanan, Stefan Born, and Lars Schmidt-Thieme. Reliable probabilistic forecasting of irregular time series through marginalization-consistent flows. In *ICLR*, 2026.

Zinuo You, Jin Zheng, and John Cartlidge. Latent Laplace diffusion for irregular multivariate time series. In *ICML*, 2026.

Weijia Zhang, Chenlong Yin, Hao Liu, Xiaofang Zhou, and Hui Xiong. Irregular multivariate time series forecasting: A transformable patching graph neural networks approach. In *ICML*, volume 235, pp. 60179–60196, 2024a.

Xi Zhang, Yuan Pu, Yuki Kawamura, Andrew Loza, Yoshua Bengio, Dennis L. Shung, and Alexander Tong. Trajectory flow matching with applications to clinical time series modelling. In *NeurIPS*, volume 37, pp. 107198–107224, 2024b.

Xingyu Zhang, Jingyao Wang, Xin Yu, Zeen Song, Jianqi Zhang, Changwen Zheng, and Wenwen Qiang. Dirichlet-guided group forecasting for alleviating over-smoothing in time series forecasting. *arXiv preprint arXiv:2606.10592*, 2026.

A Dataset Details

We use five publicly available datasets, following the preprocessing protocols of prior IMTS studies for fair comparison with state-of-the-art forecasting models (Zhang et al., 2024a; Li et al., 2025; 2026). All datasets follow the nested 90/10 splitting protocol of Li et al. (2025). It is applied to records for all datasets except Human Activity, yielding an approximate training/validation/test ratio of 8:1:1; for Human Activity it is applied to recordings before windowing. Splits preserve the original record order, except for USHCN, which uses a fixed-seed random split following the reference implementation. All sample counts below are reported in training/validation/test order. For the IMTS forecasting benchmark, the three clinical datasets use the first 36 h as context and forecast observations at the next three timestamps.

Human Activity. The Human Activity dataset from the UCI Machine Learning Repository (Vidulin et al., 2010) contains recordings of five individuals performing various activities. Four sensors, one on each ankle and one each on the belt and chest, record three-dimensional positions, yielding 12 variables. Sampling is irregular because the four sensors report asynchronously, so almost every timestamp carries a single sensor's reading, and readings are intermittently missed. The 25 recordings are split into 19/3/3 recordings before segmentation into non-overlapping 4-s windows, yielding 1,360 samples (949/193/218). Both evaluations use the first 3 s as context. Benchmark forecasting targets observations in the next 300 ms, following Li et al. (2025; 2026), with the remaining 700 ms unused. For forecast smoothing, we instead use the full remaining 1,000 ms: the 300-ms horizon contains a median of only two observations per variable, too few to characterize frequency content, compared with eight over 1,000 ms. Both settings share the same windows and samples.

PhysioNet'12. PhysioNet'12 (Silva et al., 2012) was released for an in-hospital mortality prediction challenge and contains measurements from the first 48 h of 12,000 ICU stays. Following Li et al. (2025; 2026), we aggregate measurements into 1-h bins, obtaining 11,981 samples (9,703/1,079/1,199) with 36 variables, including serum glucose and heart rate. For forecast smoothing, we instead use the first 24 h as context and the remaining 24 h as the prediction horizon. In forecast smoothing setting, we exclude five records without observations in the first 24 h, leaving 11,976 samples (9,700/1,078/1,198).

USHCN. The United States Historical Climatology Network (USHCN) (Menne et al., 2016) provides over 150 years of climate records from weather stations across the United States. Following De Brouwer et al. (2019), we use a four-year subset between 1996 and 2000 covering 1,114 stations. Daily records are subsampled to retain an average of approximately 70 observations per variable per station over this period, yielding sparse time series. Benchmark forecasting uses the first three years as context and targets observations at the next three timestamps. Excluding four stations without context observations leaves 1,110 samples (898/101/111) with five variables, including daily maximum and minimum temperatures and precipitation.

MIMIC-III. MIMIC-III (Johnson et al., 2016) contains de-identified ICU records collected at Beth Israel Deaconess Medical Center between 2001 and 2012, including vital signs, medications, and laboratory measurements. Following De Brouwer et al. (2019), we take each hospital admission's first 48 h after its first recorded measurement and aggregate them into 30-min bins, yielding 21,248 samples (17,210/1,913/2,125) with 96 variables covering laboratory tests, fluid inputs and outputs, and prescribed medications.

MIMIC-IV. MIMIC-IV (Johnson et al., 2023) builds on MIMIC-III and includes ICU admissions recorded at the same institution between 2008 and 2019. Following Biloš et al. (2021), we use the first 48 h of each admission at a 48-h window with 1-min timestamp resolution, yielding 17,874 samples (14,477/1,609/1,788) with 100 variables. Compared with MIMIC-III, this finer resolution yields sequences with substantially more time steps and greater sparsity.

Table 4: Key hyperparameters of the IMTS baselines. d, L, and H denote the hidden dimension, number of layers, and number of attention heads, respectively. A dash indicates that no additional setting is listed.

Model	**Architecture**	**Model-specific settings**
Probabilistic models		
TSDiff	GRU width: 100 Residual layers/channels: $8/8$	Diffusion steps: 100 OU noise: $\theta = 0.5$
TDSTF	$d = 128$, $L = 3$, $H = 8$ Encoder/decoder depth: $2/2$ per block	Diffusion steps: 50
TFM	$d = 64$	SDE noise: $\sigma = 0.1$ Memory: $[1, 1, 1, 0, 0]$
ProFITi	$d = 32$, $H = 2$ Encoder depth: 3	Flow depth: 10 Marginal-likelihood training
MOSES	$d = [64, 64, 64, 64, 128]$ $H = [1, 1, 2, 1, 1]$ Encoder depth: 1	Flow depth: 3 Mixture components: $[10, 10, 7, 3, 3]$
LLapDiff	$d = 256$, $L = 5$, $H = 4$ Latent channels: 16	Laplace modes: 256 Diffusion steps: 1000
Deterministic models		
GraFITi	$d = [128, 128, 128, 256, 32]$ $L = [4, 4, 1, 3, 1]$ $H = [1, 1, 8, 4, 8]$	—
TimeFlow	$d = 256$, $L = 5$ Latent width: 128	Fourier frequencies: 64 Inner updates: 3 Inner-update step size: 0.01
Hi-Patch	$d = [128, 32, 32, 32, 32]$ $L = [4, 4, 4, 4, 2]$ $H = [1, 1, 8, 1, 1]$	Patch length/stride: $[500, 6, 6, 12, 90]$
HyperIMTS	$d = [128, 256, 256, 256, 128]$ $L = [3, 1, 1, 2, 4]$ $H = [1, 8, 1, 4, 8]$	—
ReIMTS	GraFITi backbone $d = [128, 256, 32, 128, 32]$ $L = [2, 1, 1, 4, 1]$ $H = [4, 4, 4, 8, 1]$	Patch lengths: $[(2000), (24, 12), (100), (48, 24), (1440)]$
APN	$d = [56, 24, 6, 16, 24]$ Time-embedding width: $[8, 8, 32, 16, 8]$	Patches: $[300, 20, 100, 50, 360]$
ASTGI	$d = [128, 64, 64, 192, 192]$ $L = [2, 2, 3, 3, 3]$	Neighbors: $k = [96, 80, 80, 96, 96]$

B Baseline Details

We follow the hyperparameter settings specified in the original papers and official implementations of all baselines. Table 4 summarizes the key baseline settings of six IMTS probabilistic forecasting models and seven IMTS deterministic forecasting models: TSDiff (Biloš et al., 2023), TDSTF (Chang et al., 2024), TFM (Zhang et al., 2024b), ProFITi (Yalavarthi et al., 2025), MOSES (Yalavarthi et al., 2026), LLapDiff (You et al., 2026), GraFITi (Yalavarthi et al., 2024), TimeFlow(Naour et al., 2024) Hi-Patch (Luo et al., 2025b), HyperIMTS (Li et al., 2025), ReIMTS (Li et al., 2026), APN (Liu et al., 2026b), and ASTGI (Liu et al., 2026a). Here, d, L, and H denote hidden width, model depth, and attention heads, respectively. Bracketed values follow the dataset order HumanActivity, PhysioNet'12, USHCN, MIMIC-III, and MIMIC-IV; scalar values are shared.

C BACKBONE DETAILS

In this section, we describe the hyperparameter settings of DiCTS, which are shared across all datasets unless otherwise stated. Each observed history entry and each target query is treated as a single token, represented by a continuous time embedding of dimension 33 (one linear and 32 sinusoidal components) and a learnable channel embedding of dimension 128. The history encoder consists of a single Transformer layer, and the denoiser stacks four DiCTS blocks, except for USHCN, where three blocks are used. All attention layers use a hidden dimension of 128, four attention heads, a feed-forward dimension of 512, and a dropout rate of 0.1 (no dropout in the routing layer). The history-conditioned slice attention softly assigns query tokens to 32 learnable slices. The denoising step is encoded by a 128-dimensional sinusoidal embedding followed by a two-layer MLP and injected into every block via AdaLN-Zero modulation. We use sampling steps with 32.

D ADDITIONAL EXPERIMENTS

D.1 ADDITIONAL QUALITATIVE RESULTS

Figures 6–7 show forecast trajectories across all channels for the PhysioNet'12 instance in Figure 3 and one instance from each of the other datasets.

D.2 ADDITIONAL SPECTRAL ANALYSIS RESULTS

Figures 10–11 show the spectral analysis results for the HumanActivity, USHCN, MIMIC-III, including band-amplitude ratio histograms and rank histograms.

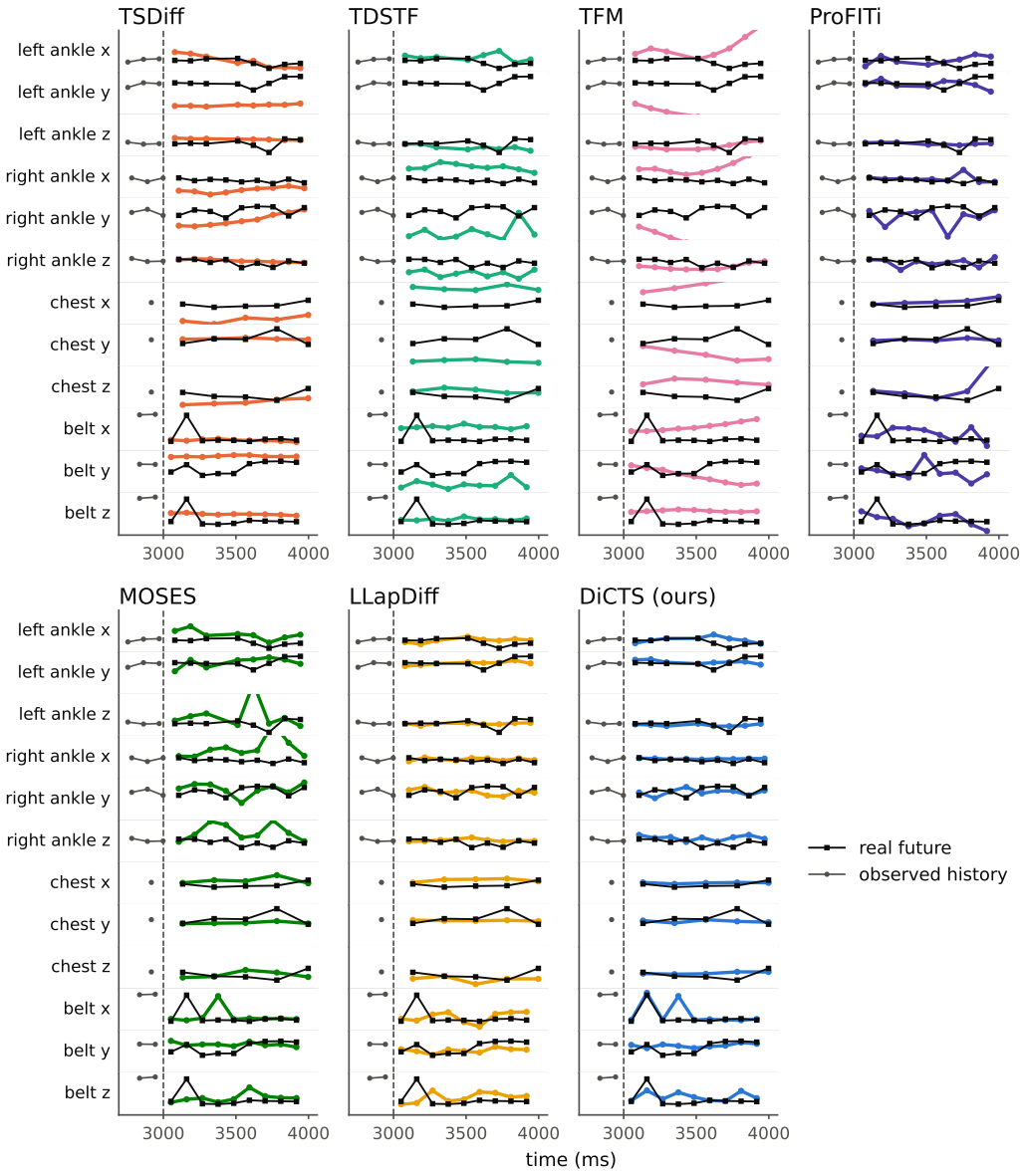


Figure 5: Forecast trajectories across all channels on HumanActivity.

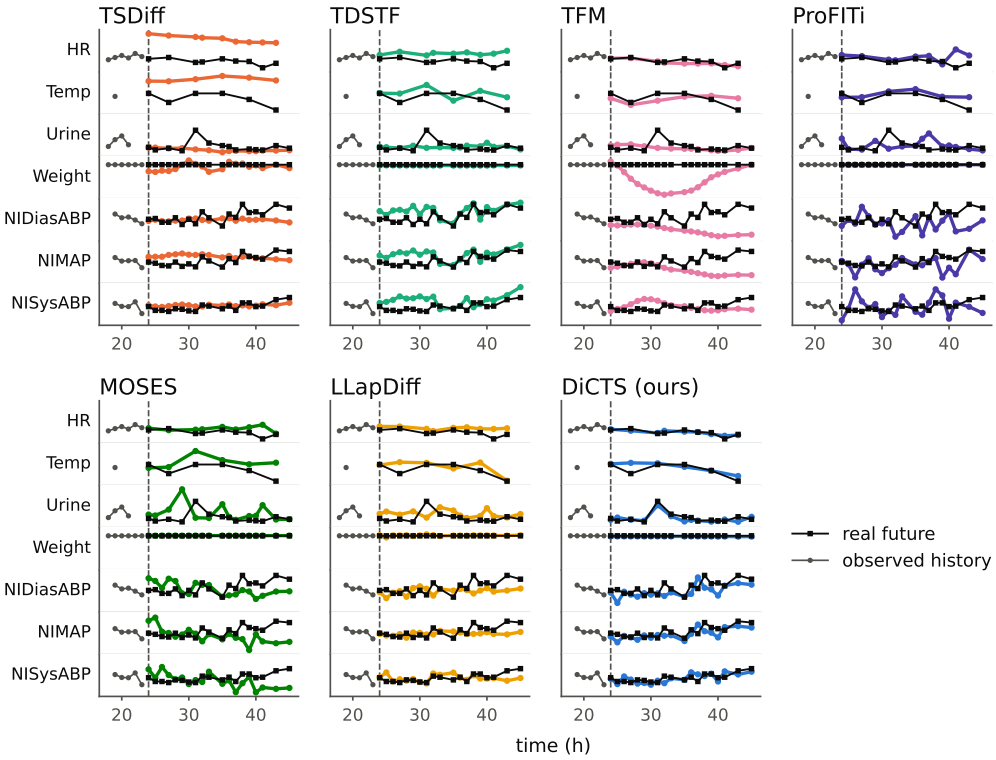


Figure 6: Forecast trajectories across all channels on PhysioNet'12.

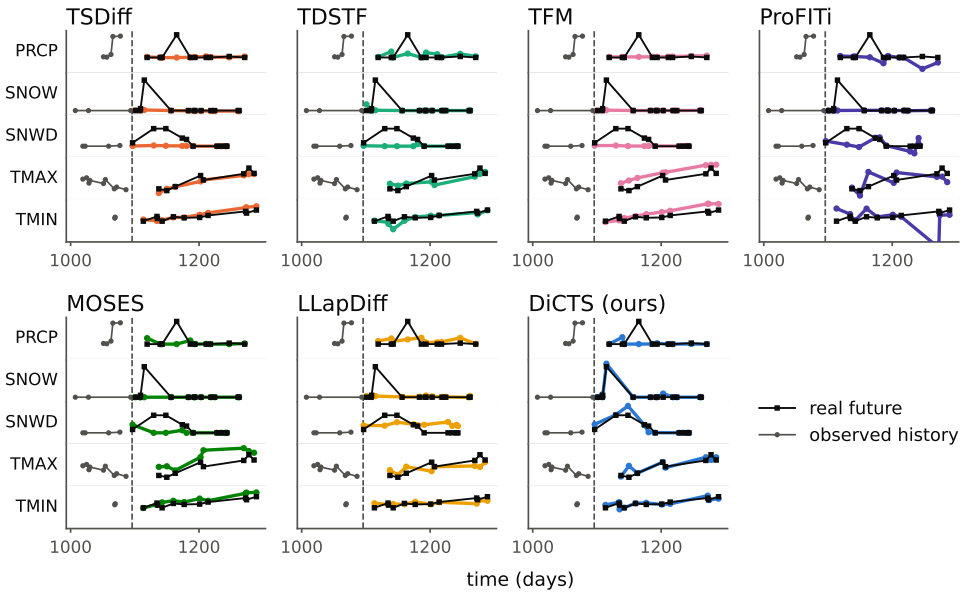


Figure 7: Forecast trajectories across all channels on USHCN.

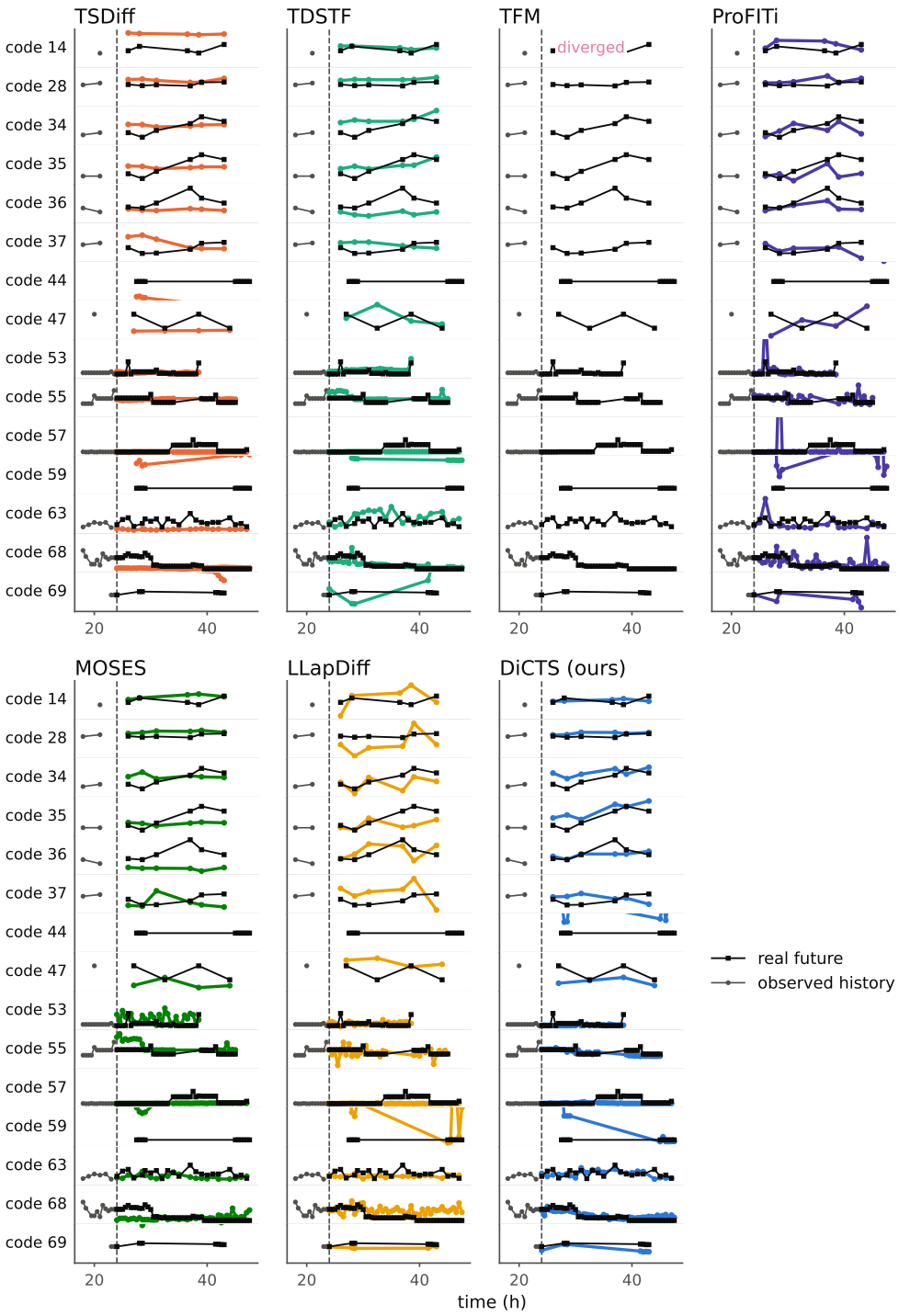


Figure 8: Forecast trajectories across all channels on MIMIC-III.

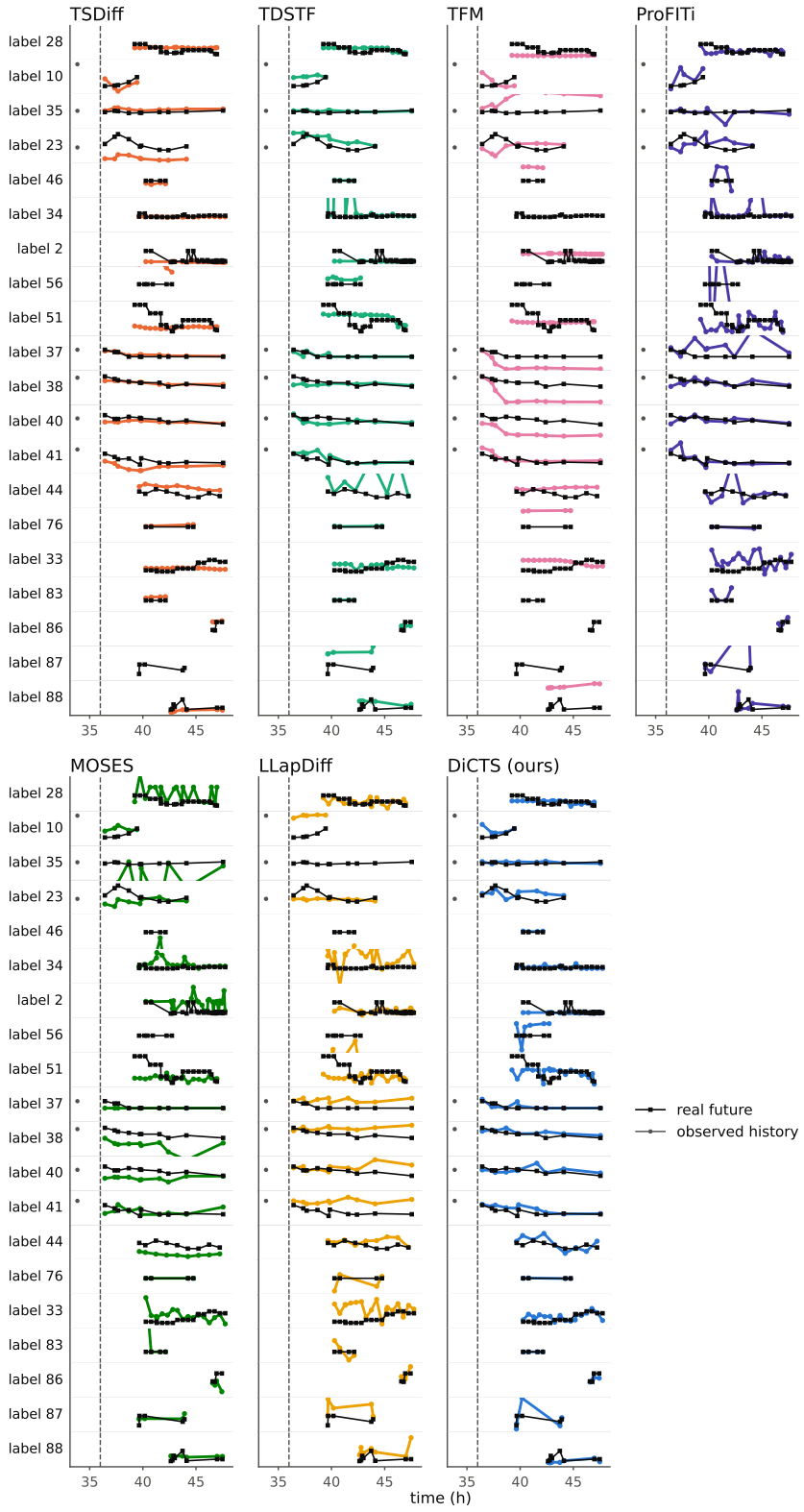
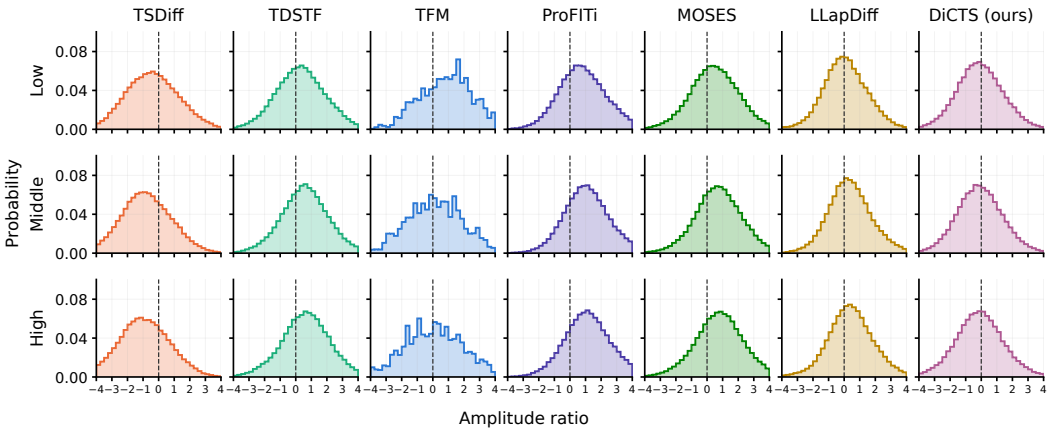
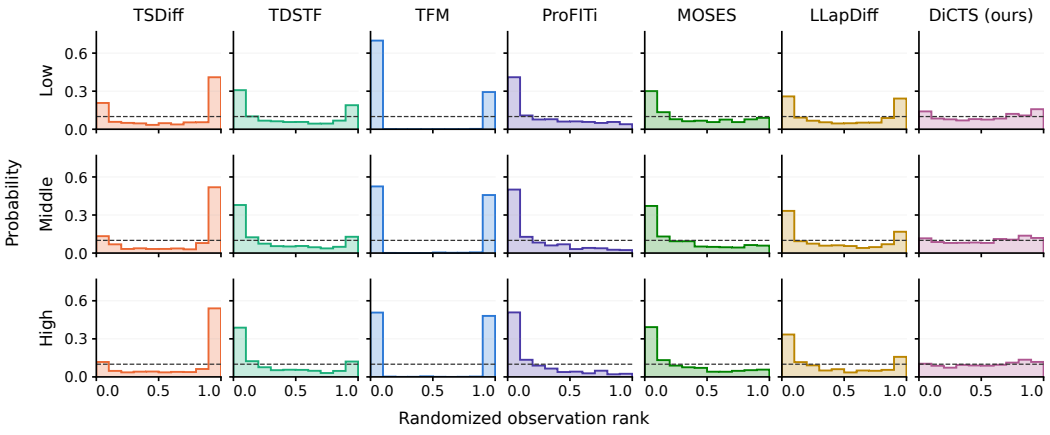


Figure 9: Forecast trajectories across all channels on MIMIC-IV.

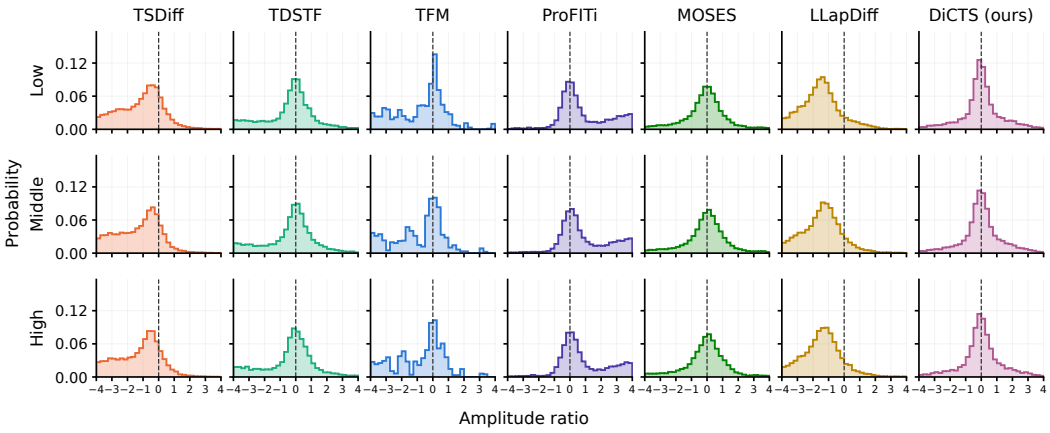


(a) Band-amplitude ratio histograms.

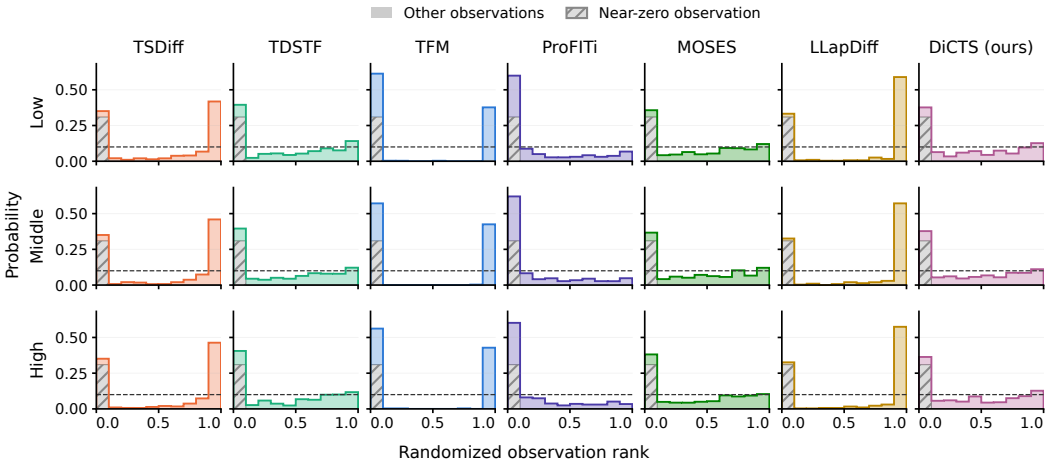


(b) Rank histograms.

Figure 10: Spectral analysis on HumanActivity.

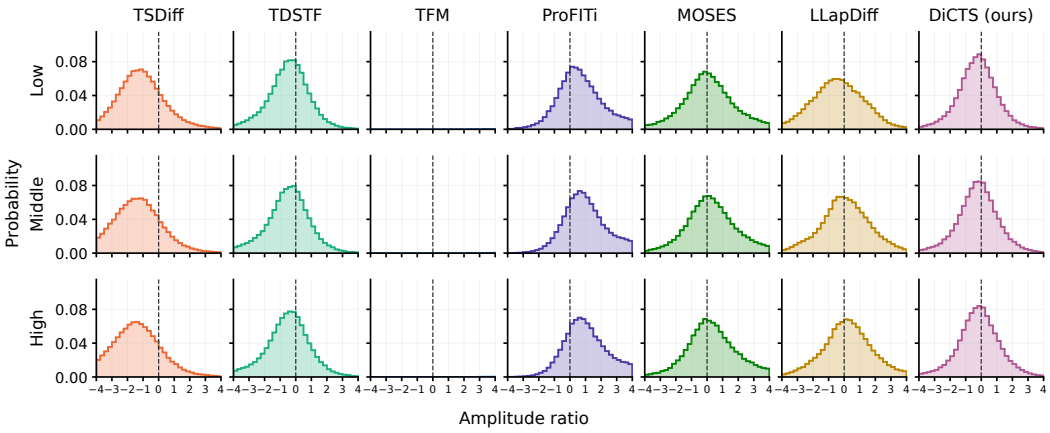


(a) Band-amplitude ratio histograms.

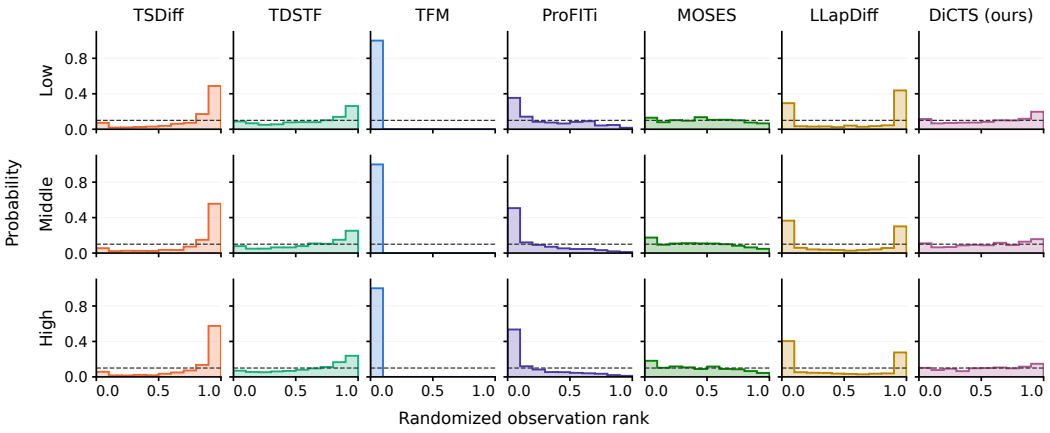


(b) Rank histograms.

Figure 11: Spectral analysis on USHCN.



(a) Band-amplitude ratio histograms.



(b) Rank histograms.

Figure 12: Spectral analysis on MIMIC-III.