COPUADAPT: AN ATTENTIONAL COPULA ADAPTER FOR IMPROVED JOINT-DISTRIBUTION SAMPLING IN PROBABILISTIC TIME-SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time-series Foundation Models (TSFMs) have demonstrated strong forecasting performance and generalization across diverse domains. For multivariate forecasting, many TSFMs adopt channel-independent modelling. These approaches, however, can struggle to capture cross-variable dependencies in joint forecast samples. We introduce *CopuAdapt*, a novel attentional copula adapter that augments frozen TSFMs with a learned multivariate dependence model without requiring foundation-model retraining. CopuAdapt constructs empirical marginal distributions from the backbone's forecast samples or quantiles and couples them through an attentional copula architecture operating in probability space. To capture regime-aware dependence, we employ a sparse attention mechanism that infers a directed acyclic graph of conditional independences for each forecasting window, supported by a JEPA-inspired auxiliary objective that learns forecast-relevant temporal memory representations. By leveraging these structural details to dynamically re-order marginal sampling, CopuAdapt strictly preserves the backbone's empirical marginals. Consequently, CopuAdapt, evaluated across the TSLib datasets, exactly preserves the Continuous Ranked Probability Score (CRPS) of the underlying TSFM while significantly improving joint-sampling sensitive metrics, including the Variogram Score (VS), and RBF-MMD, with a median improvement of 4 to 18% across metrics. Code for this paper can be found at https://anonymous.4open.science/r/CopuAdapt-6C85.

1 INTRODUCTION

Time-series Foundation Models (TSFMs) leverage massive, heterogeneous datasets to learn highly transferable representations, offering robust performance across diverse downstream tasks. A primary advantage of these models is the ability to use a frozen, pre-trained backbone directly, avoiding the immense computational cost and instability associated with retraining massive architectures from scratch. To maximise data efficiency during pre-training, many TSFMs adopt a univariate or channel-independent formulation. While highly scalable, this design creates a fundamental mismatch with many real-world forecasting tasks that require coherent multivariate scenarios. For example, managing energy grids requires understanding how wind generation, temperature, and load demand vary together, which independent marginal forecasts cannot capture.

Adapting these powerful univariate backbones to multivariate settings remains challenging. Existing adaptation strategies typically entangle the learning of multivariate dependencies with the marginal forecasting process itself (Cheng et al., 2025; Zetlin et al., 2026; Benechehab et al., 2025). This coupling often degrades or alters the foundation model's carefully calibrated marginals, sacrificing one of the primary benefits of using a frozen backbone. To resolve this, we propose explicitly separating contextual representation learning from marginal forecasting. By isolating the modelling of cross-variable dependence from the generation of marginal forecasts, we can augment powerful foundation models with multivariate capabilities without compromising their core predictive strengths.

Leveraging this principle, we introduce *CopuAdapt*, a parameter-efficient copula adapter designed to be trained for specific frozen TSFMs. CopuAdapt extracts contextual regime representations directly from the observed lookback window, entirely independent of the backbone's generative

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.

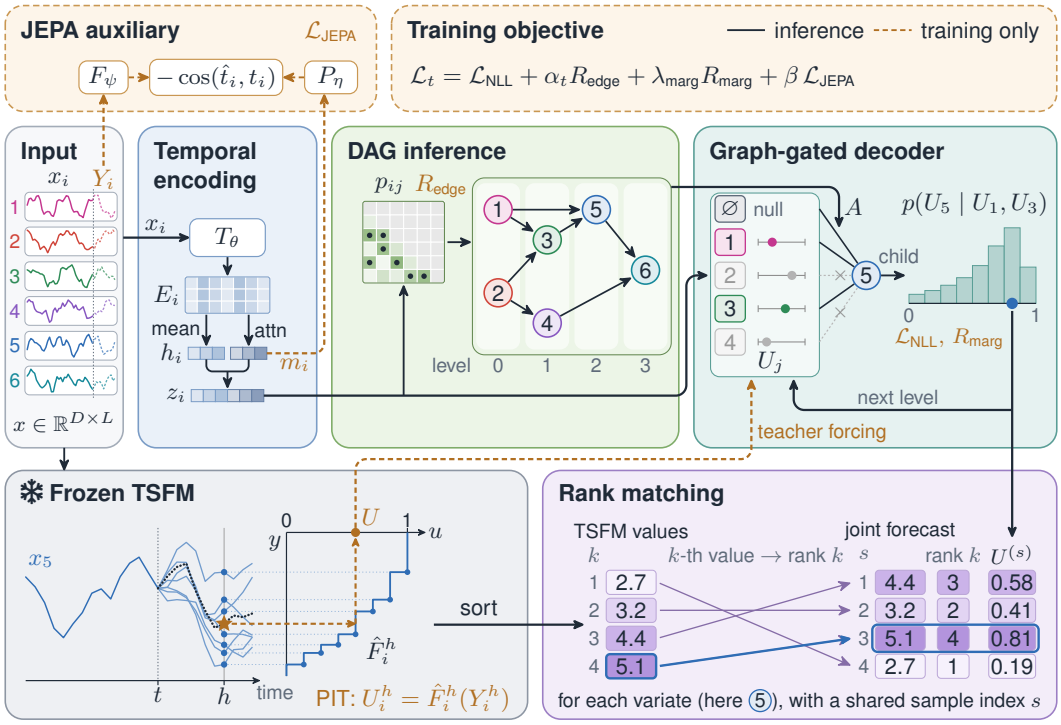


Figure 1: CopuAdapt architecture. The sampled DAG gates the decoder, so each child sees only its parents' PIT values U_j, level by level. For each variate, rank matching reorders the TSFM's forecasts by the ranks of the samples, so its marginal is kept exactly. Dashed parts are training only.

pathway. To couple these contextual representations with the backbone's outputs, we employ a non-parametric attentional copula (Drouin et al., 2022; Ashok et al., 2023) operating in probability space, where forecast values are represented by their positions within their marginal distributions. Instead of relying on dense, unstructured mixing, CopuAdapt infers a sparse, context-dependent Directed Acyclic Graph (DAG) for each forecasting window. This graph dictates the conditional independence structure of the multivariate distribution, allowing the adapter to dynamically reorder the backbone's forecast samples. Crucially, sampling proceeds along this DAG by topological level, so variables whose parents have already been sampled can be generated in parallel.

A key property of this formulation is that it provides a valid copula-based coupling while strictly preserving the foundation model's empirical marginals at each variable and forecast time. Empirically, CopuAdapt enhances multivariate forecasting metrics across multiple TSFM backbones while perfectly preserving the empirical marginals of the underlying models.

To summarise, our contributions are as follows:

1. **Preserving Strong Marginals in Multivariate Adaptation.** By explicitly separating contextual representation learning from marginal forecasting, CopuAdapt allows practitioners to augment a specific frozen univariate backbone with multivariate dependence without altering or degrading its carefully calibrated univariate predictions.
2. **Interpretable and Dynamic Dependency Structures.** Rather than relying on black-box channel mixing, CopuAdapt extracts regime-aware features to dynamically infer a sparse, context-dependent Directed Acyclic Graph (DAG) for each forecasting window, offering insight into how variables interact under different temporal conditions.
3. **Efficient Multivariate Generation.** By factorising the copula according to the learned DAG, sampling proceeds by topological level. This reduces the sequential bottleneck typically associated with autoregressive multivariate sampling when the learned graph has fewer topological levels, enabling fast, parallel generation across conditionally independent variables.

4. **State-of-the-Art Joint Forecasting Performance.** We demonstrate across multiple benchmark datasets and foundation-model backbones that CopuAdapt significantly improves joint-sampling metrics, such as Variogram Score and RBF-MMD, achieving state-of-the-art performance while requiring only a fraction of the trainable parameters of full fine-tuning.

2 Background Work

Time-series Foundation Models (TSFMs). The field of time-series foundation models has shown promise in generalizable forecasting across domains. This progress has been spurred by the development of benchmarks that test forecasting performance and zero-shot generalization of these models (Shchur et al., 2025; Aksu et al., 2024; Wang et al., 2026). Most contemporary TSFMs are either strictly univariate or channel-independent (Liu et al., 2025b; Das et al., 2024; Ansari et al., 2024; Auer et al., 2025; Liu et al., 2025a), a design choice that eases training on heterogeneous datasets by treating individual variates as separate training series (Ansari et al., 2024; Das et al., 2024). Recently, however, models explicitly considering multivariate dependencies have been designed to improve performance on multivariate forecasting tasks (Ansari et al., 2025; Woo et al., 2024; Liu et al., 2026; Podest et al., 2026). Despite this shift, the majority of these models still rely on simple mixing or cross-attention to represent multivariate information, rather than designing explicit joint sampling mechanisms.

Additionally, these models vary significantly in their output types. While some focus on point forecasts (Das et al., 2024), those that output probabilistic forecasts generally fall into three categories: parametric forecasts (Woo et al., 2024), quantile forecasts (Liu et al., 2025a; Auer et al., 2025; Podest et al., 2026; Ansari et al., 2025), and samples from the forecast distribution (Ansari et al., 2024; Liu et al., 2025b). Of these, sample-based forecasters are particularly useful for downstream tasks, as forecast trajectories can be directly used to estimate exceedance times, threshold-crossing probabilities, and other path-dependent quantities (Perez-Diaz et al., 2025).

Evaluation Metrics. Traditionally, point-forecasting models are evaluated via mean squared error (MSE), mean absolute error (MAE), and mean absolute scaled error (MASE) (Wang et al., 2026). For probabilistic forecasters, scores such as CRPS assess distribution spread as well as forecast accuracy; however, these are often univariate (Aksu et al., 2024; Ansari et al., 2024) and averaged across variates for multivariate prediction settings. Evaluation metrics sensitive to dependence structure, such as the Variogram Score (Scheuerer & Hamill", "2015"), or joint-distribution fidelity, such as maximum mean discrepancy (MMD) (Redhead et al., 2026), are rarely used to evaluate TSFMs or adapters, despite their suitability for evaluating joint forecasts.

TSFM adapters. Due to the multitude of TSFMs which are univariate or make a channel-independence assumption, recent work has developed adapters which aim to incorporate multivariate dependencies into TSFMs; however, the majority of these focus on point forecasting (Cheng et al., 2025; Zetlin et al., 2026), whereas those which facilitate probabilistic forecasting use latent space mixing (Benechehab et al., 2025), without explicitly facilitating joint-sampling. These adapters have been designed with the aim of capturing multivariate dependencies but are often evaluated using CRPS or quantile loss, which alone do not establish joint-distribution fidelity (Redhead et al., 2026).

3 Problem Formulation

To describe CopuAdapt, we formulate multivariate time series forecasting. A multivariate time series is a sequence $\{\boldsymbol{x}_t\}_{t\in\mathbb{N}}$ of observations $\boldsymbol{x}_t \in \mathbb{R}^D$, where each dimension corresponds to a variate and D denotes the number of variates. We write $\boldsymbol{x}_{t:t'}^{(i)} \in \mathbb{R}^{t'-t+1}$ for the observations of variate i over the interval $[t, t']$ and $\boldsymbol{x}_{t:t'}$ for the corresponding multivariate window containing all variates. Our goal is, given a lookback window $\boldsymbol{x}_{1:t}$ and forecast horizon H, to predict the joint conditional distribution of the next H observations

$$P(\boldsymbol{x}_{t+1:t+H} \mid \boldsymbol{x}_{1:t}). \tag{1}$$

This is unlike many contemporary TSFMs, which instead adopt a univariate formulation (Das et al., 2024; Ansari et al., 2024; Liu et al., 2025b). More formally, for each variate i, these models learn a

predictive distribution

$$p_i = P\Big(\boldsymbol{x}^{(i)}_{t+1:t+H} \mid \boldsymbol{x}^{(i)}_{1:t}\Big), \tag{2}$$

conditioned solely on that variate's history. Applying such models independently across all variates yields the approximation

$$P(\boldsymbol{x}_{t+1:t+H} \mid \boldsymbol{x}_{1:t}) \approx \prod_{i=1}^{D} p_i, \tag{3}$$

which corresponds to an assumption of variate/channel independence, while allowing temporal dependence within each variate. While this formulation dramatically increases the amount of data available for pre-training and enables scalable learning across heterogeneous domains, it neglects the cross-variable dependencies present in many forecasting problems.

A central observation motivating this work is that strong marginal forecasting and strong multivariate forecasting are not the same problem. Modern TSFMs provide powerful predictive marginals, but typically do not provide an explicit mechanism for joining those marginals into a coherent joint predictive distribution. Consequently, rather than replacing the forecasting behaviour learned by a pretrained foundation model, we seek to preserve its predictive marginals and recover only the missing multivariate structure. Here, we preserve each variable's marginal at each forecast time, while allowing temporal dependence to change. To do this we use *copulas* (Drouin et al., 2022; Ashok et al., 2023).

Copulas: allow us to represent a multivariate distribution by its marginals F_i and a copula, which models the dependence structure between variates. Sklar's theorem (Sklar, 1959) establishes that any multivariate distribution can be represented in this way. More formally, let $F^{(i)}_{t'}$ denote the cumulative distribution function associated with $p(x^{(i)}_{t'}|\boldsymbol{x}_{1:t})$, where $t + H \geq t' > t$. Then, we can model the joint cumulative distribution as

$$P(\boldsymbol{X}_{t+1:t+H} \leq \boldsymbol{x}_{t+1:t+H} \mid \boldsymbol{x}_{1:t}) = C\Big(F^{(1)}_{t+1}(x^{(1)}_{t+1}), \ldots, F^{(D)}_{t+H}(x^{(D)}_{t+H}) \mid \boldsymbol{x}_{1:t}\Big), \tag{4}$$

where $\boldsymbol{X}$ denotes future random variables (with coordinate-wise inequality) and C is a copula, a joint CDF with uniform marginals conditioned on the full multivariate lookback. This decomposition enables the forecasting backbone to focus on predicting univariate time series while the copula can model the dependencies across variates and forecast times.

To learn a copula C, which models *only* dependence, we first remove the influence of the marginals $F^{(i)}_{t'}$. To do this, the probability integral transform (PIT) is used:

$$u^{(i)}_{t'} = F^{(i)}_{t'}(x^{(i)}_{t'}), \tag{5}$$

and collecting these values into $\boldsymbol{u} \in [0,1]^{D\times H}$, the copula can be learned entirely in probability space. PIT values express relative positions within marginals, placing differently scaled variables on a common probability scale. For continuous, correctly specified conditional marginals, each coordinate is uniform on $[0,1]$ while retaining joint dependence; empirical transforms need not yield exact uniformity. Operating in this transformed domain separates dependence modelling from marginal forecasting and permits the adapter to remain agnostic to the particular foundation model used to supply the marginals.

In CopuAdapt, the predictive marginals are inherited directly from the frozen channel-independent TSFM, so each variate's marginal forecasts remain conditioned only on its own observed history. The true data-generating process may exhibit cross-variable effects on the marginals themselves, and CopuAdapt does not attempt to correct such marginal misspecification. Instead, it focuses exclusively on recovering the missing multivariate structure while preserving the forecasting behaviour learned during foundation-model pre-training. This separation is deliberate. Methods that combine variables before applying a frozen forecasting backbone may alter both the marginal forecasts and the dependence structure, making it difficult to preserve the behaviour of the pretrained model. In contrast, CopuAdapt treats the predictive marginals as fixed and augments them with a learned dependence model. This produces joint forecasts without retraining or modifying the underlying foundation model. The following section instantiates this dependence model using backbone-independent context representations, sparse interaction discovery, and a DAG-conditioned copula decoder operating directly in probability space.

4 THE COPUADAPT ARCHITECTURE

CopuAdapt is a parameter-efficient, post-hoc module designed to couple independent univariate forecasts into coherent multivariate trajectories. Only the adapter is trained. As illustrated in Figure 1, the system comprises five primary components: a frozen pretrained TSFM supplying marginal forecasts, temporal encoding with auxiliary representation learning, inference of a dependency graph, graph-conditioned copula generation, and non-parametric rank matching for exact marginal preservation.

4.1 TEMPORAL ENCODING

Let $x \in \mathbb{R}^{D\times L}$ denote the observed history of length L and $Y \in \mathbb{R}^{D\times H}$ the future over horizon H. A separate trainable temporal convolutional neural encoder T_θ extracts univariate temporal representations. Each history x_i is standardized by subtracting its mean and dividing by its standard deviation, giving $\tilde{x}_i$. The encoder processes variates independently with shared parameters θ, producing features $E_i^\tau \in \mathbb{R}^d$ at each time τ. Averaging these features over time and applying layer normalization (LN) gives the summary h_i:

$$E_i = T_\theta(\tilde{x}_i), \quad h_i = \mathrm{LN}\left(\frac{1}{L}\sum_{\tau=1}^{L} E_i^\tau\right) \tag{6}$$

A learned query also extracts an attention-pooled temporal memory m_i by weighting the history positions:

$$a_i^\tau = \frac{\exp(q_{\text{mem}}^\top W_K E_i^\tau/\sqrt{d})}{\sum_s \exp(q_{\text{mem}}^\top W_K E_i^s/\sqrt{d})}, \quad m_i = \mathrm{LN}\left(\sum_\tau a_i^\tau W_V E_i^\tau\right) \tag{7}$$

Here, q_{mem} is a learned query, W_K, W_V are learned key and value projections, and a_i^τ weights time τ; both sums cover the history. These summaries are concatenated into a window-specific memory $z_i = [h_i; m_i]$, which conditions the dependency graph and copula decoder.

4.2 JEPA-INSPIRED AUXILIARY REPRESENTATION LEARNING

An auxiliary objective inspired by Joint-Embedding Predictive Architectures (JEPA) (Assran et al., 2023) encourages historical memory m_i to retain forecast-relevant information. A predictor produces $\hat{t}_i = P_\eta(m_i)$ to match the target encoder's learned summary $t_i = F_\psi(\tilde{Y}_i) \in \mathbb{R}^d$ of the observed future. Here, $\tilde{Y}_i$ standardizes Y_i using its own mean and standard deviation. Both feedforward networks have trainable parameters ψ, η shared across variates. Let sim denote cosine similarity. The loss averages its negative over a batch of B windows and D variates:

$$\mathcal{L}_{\text{JEPA}} = -\frac{1}{BD}\sum_{b=1}^{B}\sum_{i=1}^{D}\mathrm{sim}(\hat{t}_i^b, t_i^b) \tag{8}$$

This loss supplements the copula objective, updating both networks and the temporal encoder through m_i. No stop-gradient or momentum teacher is used. The observed future, F_ψ, and P_η are needed only during training.

4.3 DAG INFERENCE FOR CONDITIONAL DEPENDENCIES

To model cross-channel dependencies, a residual feedforward network shared across variates yields $f_i = \text{Refine}_\theta(z_i) \in \mathbb{R}^d$. These features determine a graph that selects which other variates' forecast values each variate can condition on. Its binary adjacency matrix $A \in \{0,1\}^{D\times D}$ has $A_{ij} = 1$ when j is a parent of i. An edge logit ℓ_{ij} determines its inclusion probability p_{ij}:

$$\ell_{ij} = \frac{(W_Q^G f_i)^\top (W_K^G f_j)}{\sqrt{d}} + b_{ij}, \quad p_{ij} = \mathbf{1}_{[j<i]}\sigma(\ell_{ij}) \tag{9}$$

Here, W_Q^G, W_K^G are learned query and key projections, b_{ij} is a learned bias for each variate pair, and σ is the logistic function. The indicator $\mathbf{1}_{[j<i]}$ enforces a strictly lower-triangular adjacency matrix

under the dataset's fixed variate ordering, guaranteeing acyclicity. Conditioned on the history x, eligible edges are sampled independently:

$$q_\theta(A \mid x) = \prod_i \prod_{j<i} \text{Bernoulli}(A_{ij} \mid p_{ij}) \tag{10}$$

Training uses a straight-through estimator (STE) to learn binary edge decisions, with binary gates in the forward pass and gradients from a soft sigmoid approximation (Appendix B). Each generated trajectory receives its own sampled graph, shared across all forecast steps and decoder layers. Sampling follows the graph's topological levels: variates whose parents have already been generated can be sampled in parallel. Graphs with fewer levels require fewer sequential sampling stages, providing the basis for the inference-speed improvements examined in Figure 2.

4.4 Copula Generation and Training Signal

Futures are transformed into Probability Integral Transform (PIT) coordinates $U_i^h = \hat{F}_i^h(Y_i^h)$, using empirical supports constructed from frozen-backbone forecasts and observed history. The graph-gated decoder processes child queries $r_i^h = f_i + e_h^Q$ and parent memory tokens $v_j^t = f_j + E_u(U_j^t) + e_t^{KV}$. Here, E_u denotes a learned embedding function that projects the continuous PIT coordinate into the dense feature space. For each attention head, forward attention weights are restricted to open parents and an always-open learned null token:

$$\alpha_i^{h,r} = \frac{g_{i,r} \exp(s_i^{h,r})}{\sum_{r'} g_{i,r'} \exp(s_i^{h,r'})}, \quad g_{i,(j,t)} = A_{ij}, \quad g_{i,0} = 1 \tag{11}$$

The decoder produces M histogram-bin probabilities π_i^h bounded in the $[0, 1]$ domain. Its conditional density factorizes as:

$$q_\theta(U \mid x, A) = \prod_{i=1}^{D} \prod_{h=1}^{H} M \pi^h_{i,\lfloor MU_i^h \rfloor}(x, U_{\text{Pa}_A(i),:}, A) \tag{12}$$

With teacher-forced parent PITs, the predictive objective is the expected conditional negative log-likelihood (NLL):

$$\mathcal{L}_{\text{NLL}} = \mathbb{E}_{(x,Y),A\sim q_\theta(A|x)} \left[-\frac{1}{DH} \log q_\theta(U \mid x, A) \right] \tag{13}$$

where $\mathbb{E}$ denotes the expectation computed over the joint data distribution of histories and futures, as well as the inferred distribution of acyclic graphs. Let $\bar{\pi}_i^h = \frac{1}{B} \sum_b \pi_i^{bh}$ denote batch-averaged probabilities. A marginal-uniformity regularizer and an expected edge penalty are defined as:

$$R_{\text{marg}} = \frac{1}{DH} \sum_{i,h} D_{\text{KL}}(\bar{\pi}_i^h \parallel \text{Unif}_M), \quad R_{\text{edge}} = \frac{1}{BD} \sum_{b,i} \sum_{j<i} p_{ij}^b \tag{14}$$

The final composite sparse-phase objective is:

$$\mathcal{L}_t = \mathcal{L}_{\text{NLL}} + \alpha_t R_{\text{edge}} + \lambda_{\text{marg}} R_{\text{marg}} + \beta \mathcal{L}_{\text{JEPA}} \tag{15}$$

Training initially fits a dense ordered reference. Sparse training adapts α_t to target an excess training NLL of at most δ relative to that reference, choosing the sparsest non-degenerate checkpoint satisfying $\mathcal{L}_{\text{val}} \leq \mathcal{L}_{\text{dense,val}} + \delta$.

4.5 Sample Reshuffling and Rank Matching

To preserve univariate forecasting performance, let $\{\hat{y}^{(s)}\}_{s=1}^S$ be the raw generated ensemble and $\{R^{(s)}\}_{s=1}^S$ a fixed reference ensemble from the frozen backbone. For each coordinate (i, h), let σ_i^h sort the adapter's raw predictions and $R_{i,(k)}^h$ denote the sorted reference values. The non-parametric rank matching assigns:

$$\tilde{y}_i^h(\sigma_i^h(k)) = R_{i,(k)}^h, \quad k = 1, \ldots, S \tag{16}$$

Table 1: Geometric means of adapter-to-TSFM ratios after arithmetic horizon averaging. Lower is better; 1 is the channel-independent baseline. Each dataset has equal weight. Overall first geometric-means the available backbones within each dataset. Parenthesized counts are datasets; Chronos excludes wide datasets. VS denotes Variogram Score and MMD denotes RBF-MMD-RFF. † marks an adapter with retained sparse-selection fallbacks; bold includes exact ties. Per dataset results can be found in Appendix. C. * indicates a performance that is best among adapters but worse than using the TSFM without adaptation.

Dataset	Metric	CopuAdapt† (ours)	AdaPTS	ICA
Moirai2 (8)	VS	**0.92685**	3.1887	1.7711
	CRPS	**1**	2.3277	1.4326
	MMD	**0.75241**	22.489	2.3703
Sundial (8)	VS	**0.99196**	3.886	1.9912
	CRPS	**1**	2.7006	1.545
	MMD	**0.95741**	83.595	2.966
Chronos (7)	VS	**1.0189***	2.9024	1.3977
	CRPS	**1**	2.5071	1.3364
	MMD	**1.4116***	38.801	2.0755
TiRex (8)	VS	**0.96079**	3.3044	1.8931
	CRPS	**1**	2.4488	1.569
	MMD	**0.72405**	52.538	5.8092
Overall (8)	VS	**0.97345**	3.4281	1.8836
	CRPS	**1**	2.5256	1.534
	MMD	**0.90845**	47.115	3.6232

Table 2: Across-horizon parameter vs. backbone comparison on the traffic dataset.

Horizon	CopuAdapt Params	Chronos	Sundial	Moirai2	TiRex
24	941,622	2.04%	0.73%	8.27%	2.67%
32	943,670	2.04%	0.74%	8.29%	2.67%
96	960,054	2.08%	0.75%	8.43%	2.72%
192	984,630	2.13%	0.77%	8.65%	2.79%

With deterministic tie-breaking, the empirical marginal is preserved exactly:

$$\frac{1}{S}\sum_{s}\mathbf{1}_{[\tilde{y}_i^{h(s)}\le a]} = \frac{1}{S}\sum_{s}\mathbf{1}_{[R_i^{h(s)}\le a]}, \quad \forall a, i, h \tag{17}$$

This reordering ensures coordinate-wise empirical marginals remain identical to the underlying TSFM, transferring only the rank dependence structure.

5 Evaluation

5.1 Foundation Model Experiments

Here we validate our proposed adaptor on the TSLib benchmark (Wang et al., 2026), demonstrating the preservation of point-wise marginals demonstrated by identical CRPS with the utilised foundation model baseline, as well as the improvement of joint-sampling demonstrated by improved Variogram Score and RBF-MMD across a majority of tested TSFMs. These experiments were run across 4 foundation model baselines Sundial Liu et al. (2025b), Moirai2 Liu et al. (2025a), Chronos Ansari et al. (2024) and TiRex Auer et al. (2025).

In these experiments we benchmark the CRPS, Variogram Score (VS), and RBF-Kernel Maximum Mean Discrepancy (MMD). This gives us insight into the quality of the marginals of each variate,

Table 3: Percentage of edges retained on the ETTh2 dataset using Sundial (lower indicates greater sparsity). The results demonstrate a clear sparsity benefit when utilizing JEPA in CopuAdapt.

Horizon	With JEPA	Without JEPA
24	13.96%	99.99%
32	3.84%	99.98%
96	6.99%	5.02%
192	4.20%	30.62%

the correlations between variates, and the overall estimation of the ground truth distribution. Our proposed CopuAdapt is benchmarked on various state-of-the-art TSFMs against their normal inference, an ICA baseline where ICA is applied prior to forecasting and then the inverse is applied to the TSFM outputs, and AdaPTS an adapter that focuses on latent space mixing prior to the TSFM and on the TSFM forecast outputs.

The results in Tab.1 demonstrate the superiority of CopuAdapt across Moirai2, Sundial, and TiRex, with slightly worsened performance on Chronos where the original model is already trained on some of the TSLib benchmarks used for evaluation. CopuAdapt demonstrates a preservation of marginals compared to the average marginal degradation shown by AdaPTS or the application of ICA. As well as this CopuAdapt boasts an average improvement of up to 7.3% in Variogram Score and 27.6% in RBF-MMD. This illustrates the benefits of applying our proposed adaptor for fine-tuning in multivariate tasks where joint-sampling is necessary. The effect of this sampling improvement is more pronounced on TSFMs that output quantiles such as TiRex and Moirai2 as opposed to Sundial, which outputs samples and sees a modest improvement in VS and MMD, and Chronos. This highlights the additional benefit provided by the ability of our CopuAdapt architecture to produce output samples from models previously restricted to quantile outputs.

Additionally, when looking at the parameter efficiency of CopuAdapt we first analyse the scaling of the adapter with dimension and forecasting horizon.

With architectural width and depth fixed:

$$P_{\text{CopuAdapt}}(D, H, L) = \Theta(D^2 + H) \tag{18}$$

Thus, parameter count is quadratic in dimension, linear in horizon, and independent of context length. This comes from the following components: learned graph-bias matrix D^2, two horizon position embeddings $2Hd$, and future projection (training only) $2Hd$. Thus we benchmark our adapter size as a percentage of TSFM parameters on Traffic, our highest dimensional dataset, getting an upper-bound for the TSLib datasets. The results of this are shown in Table.2. These demonstrate CopuAdapt's parameter efficiency ranging from 8% of Moirai2-Small's parameters to 0.73% of Sundial Base's parameters.

5.2 Ablation Studies

5.2.1 JEPA embedding

Here we compare the performance of our CopuAdapt adapter with and without the JEPA embeddings during training to demonstrate the benefit of regime aware conditioning. Despite similar accuracy performance, we can see from Table.3 that utilisation of the JEPA temporal memory significantly improves the representation available to our Adapter resulting in the ability to learn much sparser graphs, with the non-JEPA scenario regressing to a nearly fully connected graph in half of the scenarios tested.

5.2.2 Topological parallelism inference time

Here we show the improvement in inference speed due to our parallel execution over a DAG compared to variate-wise autoregression. Initially our sparse inference adds some computational overhead due to processing interim DAG states and attention, however as sparsity increases so does the benefit of our parallelised inference. This can be seen in Fig.2. This demonstrates that the negative

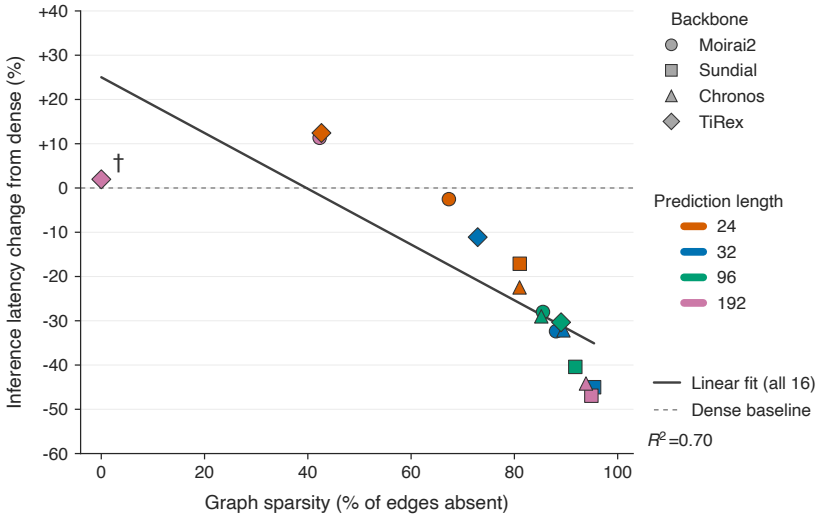


Figure 2: Impact of DAG sparsity on inference latency. The plot illustrates the relationship between graph sparsity and relative inference time compared to a standard variate-wise autoregressive baseline. The horizontal dashed line denotes the break-even point where inference speed matches the baseline. Data points above the dashed line reflect the initial computational overhead of processing interim DAG states and attention mechanisms at lower sparsity levels. The solid trend-line demonstrates a clear negative correlation: as sparsity increases, inference latency decreases, highlighting the growing efficiency gains of topological parallel execution.

effect of the increased overhead on inference speed where the graph is not sufficiently sparse (resulting in a similar depth of inference passes with reduced edges) is 10% however the improvement in inference speed as sparsity increases is up to -50%. Thus illustrating the benefits of our parallelised DAG sampling when a sufficiently sparse DAG is learned.

6 Conclusion and Future Work

Conclusion. In this paper we present CopuAdapt a novel attentional copula adapter framework for parameter efficient fine-tuning of time-series foundation models on multi-variate joint-distribution sampling. The structure of CopuAdapt allows for learning useful cross-variate structure, aided by a temporal memory, to better capture correlations and auto-correlations, improving joint-sampling. Across the TSLib datasets our strict marginal reshuffling approach results in a preservation of the TSFM's point-wise marginals maintaining identical CRPS in all cases while improving the model on joint-distribution sensitive metrics. This provides a novel adapter suited to converting univariate models for better joint-forecasting while also having the ability to adapt models from quantile-based forecasting into outputting forecast distribution samples.

Future Work. While the proposed adapter improves joint-sampling metrics, and the DAG parallel sampling improves inference speed, the use of auto-regressive inference still contributes to slower inference time and the possibility of compounding errors. Thus a copula based adapter that performs direct multi-step forecasting could bypass these limitations. As well as this, future works can investigate the use of copula structures in the design of time-series foundation models to facilitate native multi-variate forecasting and joint-distribution sampling. Finally, usage of sparse attention to create conditional-independence DAGs is only one way of learning cross-variate structure in time-series, applying time-series causal discovery methods could lead to a more interpretable DAG capable of generalisation across backbones without the need for retraining.

AI USE STATEMENT

In this work, we used generative AI tools for boilerplate code of our experiments and architecture. We have not used generative AI tools for designing the core research questions or architecture, or for any mathematical claims, and usage for proofs or dataset construction are not applicable to this work. Additionally, we used generative AI tools for initial design of our architecture figure and for refining paper writing. We have reviewed all AI-assisted work, manually sourcing references and creating the bibliography to avoid risk of hallucinated citations, manually checking code correctness and redesigning or rewriting AI generated content that did not meet the expected standard for a scientific conference. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To ensure reproducibility of this work our code is open-sourced at https://anonymous.4open.science/r/CopuAdapt-6C85 with a detailed read-me, seeds, and details around hyperparameter selection and tuning.

REFERENCES

Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. Gift-eval: A benchmark for general time series forecasting model evaluation. *arXiv*, 2024. doi: 10.48550/arxiv.2410.10393.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the Language of Time Series, November 2024. URL http://arxiv.org/abs/2403.07815. arXiv:2403.07815 [cs.LG].

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From Univariate to Universal Forecasting, October 2025. URL http://arxiv.org/abs/2510.15821. arXiv:2510.15821 [cs.LG].

Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Nicolas Chapados, and Alexandre Drouin. TACTiS-2: Better, Faster, Simpler Attentional Copulas for Multivariate Time Series. October 2023. URL https://openreview.net/forum?id=xtOydkE1Ku.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture, 2023. URL https://arxiv.org/abs/2301.08243.

Andreas Auer, Patrick Podest, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. TiRex: Zero-Shot Forecasting Across Long and Short Horizons with Enhanced In-Context Learning. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://arxiv.org/abs/2505.23719.

Abdelhakim Benechehab, Vasilii Feofanov, Giuseppe Paolo, Albert Thomas, Maurizio Filippone, and Balázs Kégl. AdaPTS: Adapting Univariate Foundation Models to Probabilistic Multivariate Time Series Forecasting. June 2025. URL https://openreview.net/forum?id=yeICCRy3lE.

Hanyin Cheng, Xingjian Wu, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. CoRA: Boosting Time Series Foundation Models for Multivariate Forecasting through Correlation-aware Adapter. October 2025. URL https://openreview.net/forum?id=JRlNrcTllN.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting, April 2024. URL http://arxiv.org/abs/2310.10688. arXiv:2310.10688 [cs.CL].

Alexandre Drouin, Étienne Marcotte, and Nicolas Chapados. TACTiS: Transformer-Attentional Copulas for Time Series. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 5447–5493. PMLR, June 2022. URL https://proceedings.mlr.press/v162/drouin22a.html.

Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. Moirai 2.0: When Less Is More for Time Series Forecasting, November 2025a. URL https://arxiv.org/abs/2511.11698v3.

Yiding Liu, Yifan Hu, Hongjie Xia, Peiyuan Liu, Hongzhou Chen, Xilin Dai, Zewei Dong, and Jiang-Ming Yang. Falcon-x: A time series foundation model for heterogeneous multivariate modeling. *arXiv preprint arXiv:2605.27286*, 2026.

Yong Liu, Guo Qin, Zhiyuan Shi, Zhi Chen, Caiyin Yang, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Sundial: A Family of Highly Capable Time Series Foundation Models. June 2025b. URL https://openreview.net/forum?id=LO7ciRpjI5.

Alvaro Perez-Diaz, James C. Loach, Danielle E. Toutoungi, and Lee Middleton. Foundation model forecasts: Form and function. *arXiv preprint arXiv:2510.19345*, 2025.

Patrick Podest, Marco Pichler, Elias Bürger, Levente Zólyomi, Bernhard Voggenberger, Wilhelm Berghammer, Daniel Klotz, Sebastian Böck, Günter Klambauer, and Sepp Hochreiter. Tirex-2: Generalizing tirex to multivariate data and streaming, 2026. URL https://arxiv.org/abs/2607.01204.

Benjamin R. Redhead, Thomas L. Lee, Peng Gu, Víctor Elvira, and Amos Storkey. Signature-Kernel Based Evaluation Metrics for Robust Probabilistic and Tail-Event Forecasting, February 2026. URL http://arxiv.org/abs/2602.10182. arXiv:2602.10182 [cs.LG].

"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". doi: "10.1175/MWR-D-14-00269.1".

Oleksandr Shchur, Abdul Fatir Ansari, Caner Turkmen, Lorenzo Stella, Nick Erickson, Pablo Guerron, Michael Bohlke-Schneider, and Yuyang Wang. fev-bench: A realistic benchmark for time series forecasting. *arXiv*, 2025. doi: 10.48550/arxiv.2509.26468.

Abe Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Chen Wang, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(10):11642–11662, 2026. doi: 10.1109/TPAMI.2026.3690845.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 53140–53164. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/woo24a.html.

Zara Zetlin, Kayhan Moharreri, and Maria Safi. TopoPrimer: The Missing Topological Context in Forecasting Models, May 2026. URL http://arxiv.org/abs/2605.15035. arXiv:2605.15035 [cs.LG].

A FORECASTING SETUP

Across all datasets except ILLNESS we condition on a context of length 96 and evaluate horizons $H \in \{24, 32, 96, 192\}$. ILLNESS is recorded at weekly resolution and comprises substantially fewer observations than the remaining datasets; we therefore use a shorter context of length 48 together with the shorter horizons $H \in \{16, 24, 32, 48\}$. Contexts and horizons are held fixed across all methods, so that every baseline receives identical inputs and is evaluated against identical targets.

Table 4: Context lengths and forecast horizons.

Dataset	Context	Horizons
ILLNESS	48	16, 24, 32, 48
All others	96	24, 32, 96, 192

B DAG INFERENCE: STRAIGHT-THROUGH EDGE SAMPLING

During training, DAG inference uses a straight-through estimator to learn the binary adjacency matrix A. To sample its edges, we add independent standard logistic noise ϵ_{ij} to each eligible edge logit ℓ_{ij}. Thresholding the perturbed logit yields a Bernoulli sample with inclusion probability $\sigma(\ell_{ij})$, while applying a sigmoid gives the soft gate $\bar{A}_{ij}$:

$$\epsilon_{ij} \sim \text{Logistic}(0, 1), \quad \bar{A}_{ij} = \sigma(\ell_{ij} + \epsilon_{ij}) \tag{19}$$

$$A_{ij} = \mathbf{1}_{[j<i]}\mathbf{1}_{[\ell_{ij}+\epsilon_{ij}>0]}, \quad A^{\text{ST}}_{ij} = A_{ij} + \mathbf{1}_{[j<i]}(\bar{A}_{ij} - \text{sg}(\bar{A}_{ij})) \tag{20}$$

Here, sg returns its input while blocking gradients. Thus, A^{ST}_{ij} equals the binary gate A_{ij} in the forward pass, but its gradient with respect to an eligible edge logit is taken from the soft gate $\bar{A}_{ij}$. The indicator $\mathbf{1}_{[j<i]}$ keeps all other edges and their gradients zero.

C EXPERIMENTAL RESULTS

Here we show the per dataset and median performance results for our TSLib benchmark experiments.

Table 5: Electricity: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	**41.058**	**202**	0.0047272
	CopuAdapt	41.094	**202**	**0.0035091**
	AdaPTS	314.55	691.95	0.26897
	ICA	385.4	864.21	0.38792
Sundial	Independent	**18.234**	**141.2**	0.00048736
	CopuAdapt	18.236	**141.2**	**0.00042089**
	AdaPTS	228.42	620.73	0.25951
	ICA	351.49	838.08	0.48731
TiRex	Independent	17.713	**128.24**	0.00016387
	CopuAdapt	**17.708**	**128.24**	**0.00011097**
	AdaPTS	166.38	477.56	0.13186
	ICA	352.62	835.64	0.4435

Table 6: ETTh1: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	0.5873	**1.4629**	0.015143
	CopuAdapt	**0.55251**	**1.4629**	**0.0079845**
	AdaPTS	1.2653	2.4322	0.19856
	ICA	0.77957	1.7238	0.034836
Sundial	Independent	0.51795	**1.3161**	0.0021124
	CopuAdapt†	**0.51639**	**1.3161**	**0.0020884**
	AdaPTS	1.1465	2.4108	0.20219
	ICA	0.5355	1.3416	0.0033068
Chronos	Independent	0.61436	**1.4744**	0.0063363
	CopuAdapt	**0.6113**	**1.4744**	**0.0061508**
	AdaPTS	1.2005	2.6191	0.24066
	ICA	0.68224	1.6929	0.017812
TiRex	Independent	0.50685	**1.2555**	0.0031368
	CopuAdapt	**0.49161**	**1.2555**	**0.001445**
	AdaPTS	1.0578	2.1288	0.16653
	ICA	0.51197	1.2679	0.005717

Table 7: ETTh2: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	0.57717	2.0503	0.0081061
	CopuAdapt	**0.49681**	2.0503	**0.0055708**
	AdaPTS	1.1257	3.622	0.19759
	ICA	0.60484	**2.0503**	0.005795
Sundial	Independent	0.46011	**1.9867**	0.0035569
	CopuAdapt	**0.45661**	**1.9867**	0.0034243
	AdaPTS	1.0251	3.8628	0.2268
	ICA	0.47422	2.0179	**0.0023029**
Chronos	Independent	**0.48193**	**1.9667**	**0.0034771**
	CopuAdapt	0.48201	**1.9667**	0.0039993
	AdaPTS	1.0451	3.9417	0.23487
	ICA	0.50494	2.154	0.0080135
TiRex	Independent	0.46764	**1.7908**	0.002837
	CopuAdapt†	**0.43464**	**1.7908**	**0.0015234**
	AdaPTS	0.99152	3.4609	0.17927
	ICA	0.4727	1.8753	0.0064862

Table 8: ETTm1: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	0.66875	1.8849	0.04134
	CopuAdapt†	**0.62574**	1.8849	0.032671
	AdaPTS	1.2762	2.7209	0.23875
	ICA	0.67598	**1.6797**	**0.028604**
Sundial	Independent	0.44912	**1.2701**	**0.0090131**
	CopuAdapt†	**0.44612**	**1.2701**	0.009041
	AdaPTS	1.1665	2.6126	0.28267
	ICA	0.51863	1.3933	0.010844
Chronos	Independent	**0.64719**	**1.6802**	0.018042
	CopuAdapt†	0.64878	**1.6802**	**0.017406**
	AdaPTS	1.3134	3.0909	0.34334
	ICA	0.74388	2.0051	0.030631
TiRex	Independent	0.60621	**1.5666**	0.023831
	CopuAdapt†	**0.59301**	**1.5666**	**0.01809**
	AdaPTS	1.1582	2.5239	0.21919
	ICA	0.60905	1.569	0.018639

Table 9: ETTm2: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	0.44829	1.8404	0.013871
	CopuAdapt†	**0.39872**	1.8404	0.012411
	AdaPTS	0.96366	3.4443	0.20051
	ICA	0.44085	**1.6959**	**0.0052713**
Sundial	Independent	0.34923	**1.639**	0.0021941
	CopuAdapt	**0.34546**	**1.639**	0.0020995
	AdaPTS	0.9224	3.6847	0.23093
	ICA	0.3476	1.642	**0.0011623**
Chronos	Independent	**0.38829**	**1.7079**	0.0075573
	CopuAdapt	0.3992	**1.7079**	0.010178
	AdaPTS	0.93474	3.7771	0.23809
	ICA	0.39542	1.8765	**0.0050298**
TiRex	Independent	0.39661	**1.601**	0.0044024
	CopuAdapt†	**0.36603**	**1.601**	**0.0042486**
	AdaPTS	0.90648	3.3546	0.17899
	ICA	0.39305	1.6823	0.0075446

Table 10: Exchange: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	0.0017499	**0.013105**	0.0032268
	CopuAdapt†	**0.001521**	**0.013105**	**0.0031412**
	AdaPTS	0.017616	0.072638	0.79307
	ICA	0.0017275	0.013553	0.0088485
Sundial	Independent	0.0016968	0.015402	0.0065083
	CopuAdapt†	0.0016729	0.015402	0.0065198
	AdaPTS	0.017352	0.076984	0.82523
	ICA	**0.0016091**	**0.01467**	**0.0058655**
Chronos	Independent	**0.0017086**	**0.013588**	**0.0092566**
	CopuAdapt	0.0018806	**0.013588**	0.038727
	AdaPTS	0.017199	0.078918	0.83968
	ICA	0.0019236	0.014396	0.012727
TiRex	Independent	0.0016315	**0.012615**	0.005684
	CopuAdapt†	0.0015487	**0.012615**	**0.0055584**
	AdaPTS	0.017179	0.070226	0.7566
	ICA	**0.0015226**	0.0127	0.0061723

Table 11: Illness: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	6376.7	**23390**	0.064873
	CopuAdapt	**5963.9**	**23390**	**0.038821**
	AdaPTS	17931	52880	0.37198
	ICA	40052	92278	1.0975
Sundial	Independent	4512.3	**22028**	0.064832
	CopuAdapt	**4413.4**	**22028**	**0.059646**
	AdaPTS	16756	58126	0.39086
	ICA	40120	92543	1.1087
Chronos	Independent	**5450.4**	**21074**	**0.038739**
	CopuAdapt†	5502.4	**21074**	0.038749
	AdaPTS	19812	62832	0.49248
	ICA	40430	92978	1.1006
TiRex	Independent	6127.5	**22236**	0.060123
	CopuAdapt†	**5759.9**	**22236**	**0.03787**
	AdaPTS	16362	47752	0.24299
	ICA	40222	92575	1.0911

Table 12: Weather: arithmetic mean, with equal horizon weights. Lower is better; bold includes exact ties within each backbone. Independent is channel-independent. † marks a retained sparse-selection fallback at one or more horizons.

Backbone	Method	Variogram	CRPS	RBF-MMD-RFF
Moirai2	Independent	13.303	**16.087**	0.0082155
	CopuAdapt†	**13.213**	**16.087**	0.0076159
	AdaPTS	38.039	40.71	0.25119
	ICA	16.001	16.899	**0.0049731**
Sundial	Independent	12.058	**14.958**	0.00093571
	CopuAdapt	**12.057**	**14.958**	0.00090849
	AdaPTS	38.962	44.317	0.27214
	ICA	14.97	17.965	**0.00056292**
Chronos	Independent	17.016	19.258	0.0037138
	CopuAdapt	17.016	19.258	0.0068311
	AdaPTS	38.971	47.81	0.27723
	ICA	**15.589**	**19.023**	**0.0021547**
TiRex	Independent	**13.941**	**15.152**	0.0023218
	CopuAdapt	13.944	**15.152**	**0.0023202**
	AdaPTS	38.829	39.777	0.2482
	ICA	18.553	18.308	0.010109

Table 13: Medians of adapter-to-Independent ratios across datasets, after median horizon aggregation. Each ratio divides the adapter's horizon median by the matching Independent horizon median. Lower is better; 1 is the channel-independent baseline. Each dataset has equal weight. Overall first takes the median of available backbone ratios within each dataset, then the median across datasets. Even-sized medians average the two central values. Parenthesized counts are datasets; Chronos excludes wide datasets. VS denotes Variogram and MMD denotes RBF-MMD-RFF. † marks an adapter with retained sparse-selection fallbacks; bold includes exact ties.

Backbone	Metric	Independent	CopuAdapt†	AdaPTS	ICA
Moirai2 (8)	VS	1	**0.93585**	2.4510	1.1274
	CRPS	**1**	**1**	2.0578	1.0323
	MMD	1	**0.74181**	30.0800	1.6333
Sundial (8)	VS	1	**0.99244**	2.9076	1.0981
	CRPS	**1**	**1**	2.4532	1.0593
	MMD	1	**0.96141**	96.3580	1.0543
Chronos (7)	VS	**1**	1.00470	2.2854	1.1170
	CRPS	**1**	**1**	2.2872	1.1060
	MMD	**1**	1.21370	39.3220	1.6683
TiRex (8)	VS	1	**0.96619**	2.4548	1.0206
	CRPS	**1**	**1**	2.1118	1.0573
	MMD	1	**0.72377**	68.0940	2.2821
Overall (8)	VS	1	**0.97763**	2.5858	1.0869
	CRPS	**1**	**1**	2.2946	1.0698
	MMD	1	**0.81288**	60.8890	1.4274