ZERO-SHOT MULTIVARIATE TIME SERIES FORECASTING USING TABULAR PRIOR FITTED NETWORKS

Mayuka Jayawardhana[1], **Nihal Sharma**[2], **Kazem Meidani**[2], **Bayan Bruss**[2], **Tom Goldstein**[1] **& Doron Bergman**[2] *
[1]University of Maryland, [2]Capital One
{mayukaj,tomg}@umd.edu
{nihal.sharma,mohammadkazem.meidani}@capitalone.com
{bayan.bruss,doron.bergman}@capitalone.com

ABSTRACT

Tabular foundation models, particularly Prior-data Fitted Networks like TabPFN have emerged as the leading contender in a myriad of tasks ranging from data imputation to label prediction on the tabular data format surpassing the historical successes of tree-based models. This has led to investigations on their applicability to forecasting time series data which can be formulated as a tabular problem. While recent work to this end has displayed positive results, most works have limited their treatment of multivariate time series problems to several independent univariate time series forecasting subproblems, thus ignoring any inter-channel interactions. Overcoming this limitation, we introduce a generally applicable framework for multivariate time series forecasting using tabular foundation models. We achieve this by recasting the multivariate time series forecasting problem as a *series* of scalar regression problems which can then be solved *zero-shot* by any tabular foundation model with regression capabilities. We present results of our method using the TabPFN-TS backbone and compare performance with the current state of the art tabular methods.

Track: Research

1 INTRODUCTION

Multivariate time series (MTS) are ubiquitous in real-world applications, presenting themselves as critical regression, classification, or forecasting tasks across domains such as finance, meteorology, and engineering Lim & Zohren (2021); Ismail Fawaz et al. (2019); Tsay (2005); Duchon & Hale (2012). Unlike their univariate counterparts, MTS data exhibit added complexity due to the inclusion of both temporal (inter-sample) and spatial (intra-sample) dependencies. While it is reasonable to assume that best possible forecasting necessitates modeling the joint probability distribution of these evolving variables, prior work has often sidestepped the complexity of intra-sample dependence among covariates. Instead, many approaches simply decompose MTS into independent univariate problems, a strategy often referred to as Channel Independence (CI) Han et al. (2024); Wu et al. (2021). Reformulating the problem is particularly appealing in regimes where training data is not abundant — modeling both Channel Dependent (intra-sample) and CI interactions generally would need models with higher capacity and thus, more data Han et al. (2024).

This CI decomposition of MTS problems is also appealing since it makes MTS tasks organically applicable to be solved using a class models that are trained to regress over a scalar target variable using data provided in the context (a.k.a In-Context Learning). Particularly, Prior-data Fitted Networks (PFNs) are trained to be in-context learners for general tabular data: given all information in a table apart from the label of one of the rows, these networks learn to estimate this scalar target label. While the training data here assumes that every row of a table is independently generated given a fixed underlying dependency structure, prior works have suggested the use of PFNs for univariate time series forecasting by treating the future value of the time series as the target label and have been

*Corresponding author

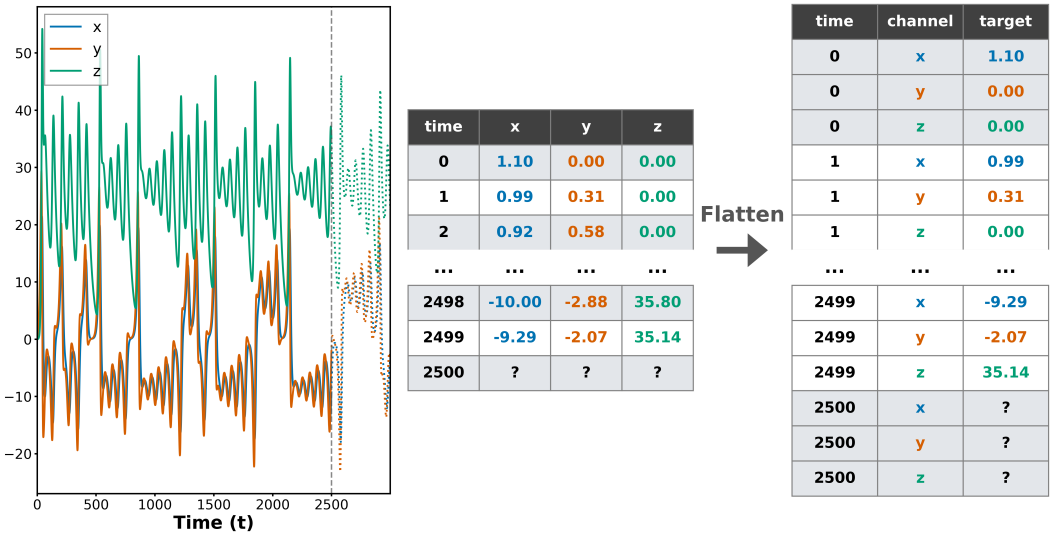

Figure 1: **Reformulating multivariate time series forecasting as a scalar regression problem**: A pictorial and tabular representation of a Lorenz system (a time series with three covariates x, y, z introduced in Lorenz (1963)) is presented on the left parts of the figure. The time-indexed table is "rolled out" by introducing a channel indicator column (as seen in the table on the right). We note that this transformation expands the size of the table by a factor equal to the number of covariates.

found to be competitive with purely temporal models. This observation has led to two streams of work for MTS forecasting: one that involves data and architectural modifications to adapt the PFN-like training to account for the temporal dependence (Dooley et al. (2023); Moroshan et al. (2025) for univariate forecasting, Taga et al. (2025); Ansari et al. (2025) for MTS problems), and a second, that includes zero-shot methods which reformulate the univariate time series forecasting problem to appear as a tabular regression task by augmenting the time series data to fit the input data format of tabular foundation models Hoo et al. (2025).

In this work, we propose a new framework that reformulates MTS forecasting tasks to fit the mold of the input data of tabular foundation models and can be used to extract zero-shot predictions of future samples. Building on the existing technique of univariate time series data augmentation, our method explicitly models the intra-sample dependencies among covariates without requiring architectural modifications or retraining. We achieve this by serializing the multivariate structure into a "rolled out" tabular format: at each time step, the multivariate vector is flattened into several rows, where the features consist of the timestamp, the covariate index, and the value (see Figure 1). The regression target is structured to be the subsequent value in this flattened sequence. This method effectively transforms the spatial correlations between variables into a sequential dependency that the PFN can interpret. Our method trades off the zero-shot applicability of off-the-shelf tabular foundation models against increased computational costs, which arise from the inflated context lengths inherent in unrolling the MTS data.

We empirically evaluate our approach by benchmarking it against the univariate decomposition baseline established by TabPFN-TS Hoo et al. (2025), as well as other specialized architectures trained on time series priors, including TimePFN Taga et al. (2025), TempoPFN Moroshan et al. (2025), and Chronos-2 Ansari et al. (2025). The remainder of this paper is organized as follows: Section 2 details our methodology; Section 3 presents our experimental results and comparative analysis; and Section 4 provides the conclusion and directions for future work.

2 METHODOLOGY

In this section we discuss our lightweight approach for multivariate time series forecasting using tabular PFN regressors for scalar targets. TabPFN-TS Hoo et al. (2025) demonstrates that standard tabular PFNs can solve time series forecasting problems by treating them as regression problems

- time is the independent variable, and the time series value is the dependent variable. Additionally, augmenting the time series data with simple auto generated temporal features, as additional independent variables, is sufficient to make the performance of TabPFN-TS in time series forecasting competitive. However, TabPFN-TS does not have a mechanism for unconstrained multivariate forecasting that models cross-channel dependencies. We demonstrate that a simple reformulation strategy enables efficient multivariate time series forecasting that can utilize cross-channel context using tabular PFNs.

2.1 Univariate time series forecasting as a regression problem

Time series forecasting can be stated as predicting the distribution of a horizon H of next few observations, $x_{t:t+H} = \{x_t, x_{t+2}, \ldots, x_{t+H-1}\}$, given all past observations $x_{<t} = x_{0:t} = \{x_0, x_1, \ldots, x_{t-1}\}$: $P(x_{t:t+H}|x_{<t})$. The joint probability of a time series sequence $\{x_0, \ldots, x_T\}$ can be decomposed into a product of conditional next-step predictions: $P(x_0, \ldots, x_T) = \prod_{t=0}^{T} P(x_t|x_{<t})$. Therefore, the fundamental task is solving the next-step prediction $P(x_t|x_{<t})$ for all t. We can frame $P(x_t|x_{<t})$ as a standard conditional density estimation (regression) problem: $P(x_t|x_{<t}) = P(u|t; (t_i, u_i)|_{i=1}^{t-1})$.

By defining the feature vector to include time as a explicit feature, a regression model can leverage t as the sample feature to predict target $y = x_t$. Tabular PFNs are trained to solve regression problems in this form.

2.2 Multivariate time series forecasting as a univariate regression problem

Consider the simplest multivariate time series: A two variate time series forecasting problem can be stated as $P(x_t, y_t|x_{<t}, y_{<t})$.

Existing tabular PFN regressors are trained to predict only one target variable. As re-training a foundational model that can natively predict multiple targets requires considerable resources and entails engineering challenges, we employ the following simple reformulation of the problem.

We can introduce an indicator variable $\eta \in \{x, y\}$ to denote which component is being predicted. Features for each sample are now: t, η and we predict the single target variable u. The prediction problem can now be reformulated as,

$$P(x_t|x_{<t}, y_{<t}) = P(u|\eta = x, t; (\eta = x, s; u_s)|_{s=1}^{t-1}, (\eta = y, s; u_s)|_{s=1}^{t-1}) \tag{1a}$$

$$P(y_t|x_t, x_{<t}, y_{<t}) = P(u|\eta = y, t; (\eta = x, s; u_s)|_{s=1}^{t}, (\eta = y, s; u_s)|_{s=1}^{t-1}) \tag{1b}$$

Predicting the different components autoregressivly draws predictions from the joint multivariate distribution of next steps in the multivariate timeseries, but the joint distribution $P(x_t, y_t|x_{<t}, y_{<t})$ can also be drawn from directly, as we do in our implementation.

More generally we can simply use the following notation to describe the multivariate timeseries, $P(u_{a,t}|u_{b,<t}|_{b=1}^{d}) = P(u_|a, t; (b, t_i, u_{b,i})|_{i=1,b=1}^{i=t-1,b=d})$

While our method introduces a position-dependent indicator column (η), we do not experiment with shuffling its order (as detailed in Appendix B). This formulation is easily extended to more variates by treating η as a categorical feature which encodes which variate that row treats as the prediction target, and both the different channels and the full forecasting horizon can all be predicted simultaneously using any tabular foundation model.

In practice, before combining the channels to form a unified sequence we normalize each channel's sequence independently (z-score normalization) before passing it to TabPFN. The predictions are then mapped back to the original channel scale, by applying the inverse to the z-score normalization before calculating metrics. We find that this independent normalization helps when different variates are of different scales, which is a common occurrence (see Appendix E for more details).

Thus, this simple feature engineering approach enables the use of off-the shelf tabular foundational models for multivariate forecasting. In what follows we label our approach as **TabPFN-TS-MV**.

3 EXPERIMENTS

We evaluate our approach using the Gift-Eval time series benchmark Aksu et al. (2024). Gift-Eval spans 23 datasets with diverse characteristics, totaling over 144,000 time series and 177 million data points across seven application domains and ten sampling frequencies. It includes both univariate and multivariate forecasting settings and covers prediction horizons ranging from short- to long-term. In total, combining all valid datasets, frequencies, and horizons, GIFT-Eval provides 97 distinct benchmarking tasks. As our method differs from TabPFN-TS on multivariate time series we only compare the multivariate datasets. We note that our method is in principle sensitive to the order and choice of variates used for forecasting. We do not perform ablations on these as the order of variates is inherently permuted by TabPFN-TS during inference and the effects of censoring the participation of some variates would be dataset-specific and not fundamental to the applicability of our method.

3.1 POINT FORECAST ACCURACY

Following standard practice we evaluate point forecast accuracy using the Mean Absolute Scaled Error (MASE) and observe that we outperform the standard TabPFN-TS model on 60% of multivariate datasets in Gift-Eval 2. Note that these results were generated using the same number of past time steps for both approaches. In practice the context length limits of TabPFN V2 restricts our approach and independently predicting each variate allows for the use of more time steps. However, newer tabular foundation models with expanded context capabilities, such as TabPFN 2.5 Grinsztajn et al. (2025), will alleviate this limitation. Further comparison of our approach against SOTA time series models is provided in Appendix D.

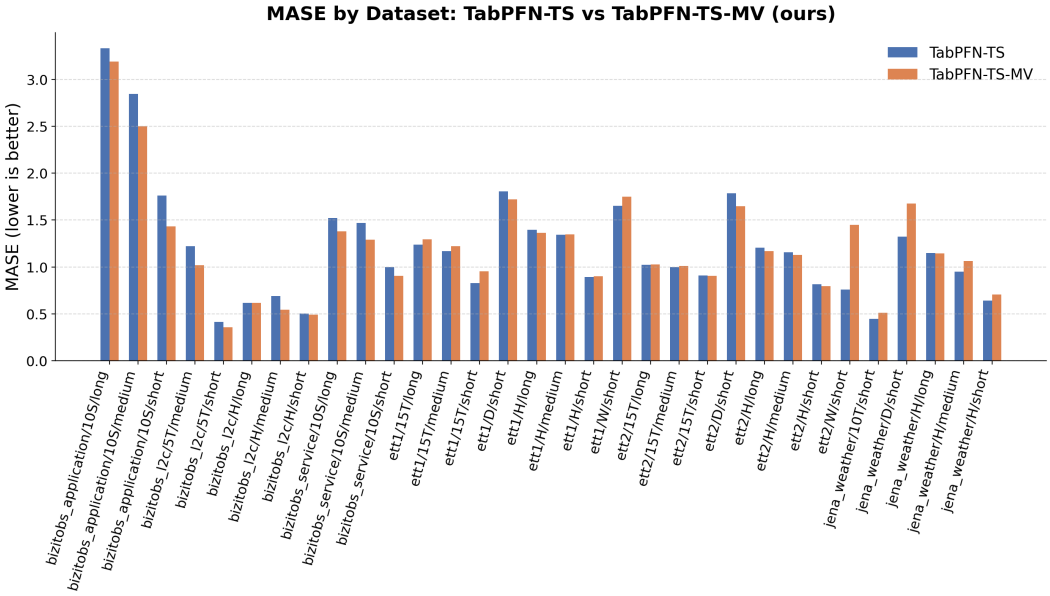

Figure 2: **Performance of TabPFN-TS compared to ours on multivariate datasets in the gift-eval benchmark** our approach lowers MASE on 60% of the datasets. We refer the reader to Table 1 in the appendix for detailed results.

4 DISCUSSION AND FUTURE WORK

In this work, we introduced a framework for multivariate time series forecasting that can be applied in a zero-shot manner to PFN-based regressors by reformulating the task into a series of scalar regression problems. In contrast to the existing channel-independent decomposition approach, our method allows for inter-channel interaction that could aid in improving performance of forecasting at the cost of increased context length. While this increased cost is not significant for the datasets in the

Gift-Eval benchmark, we note that this might be an important consideration as the number of variates in the time-series of interest grows; the context length scales linearly in the number of variates and is also limited by maximum context length that can be used for the base tabular foundation model. Our approach is similar in spirit to MORPHEUS Patil et al. (2025), which flattens multivariate time series into a tokenized sequence of discretized inputs for a transformer model. On the other hand, our framework is model-agnostic and requires no retraining or architectural modifications.

While we find the empirical performance of our method to be mostly positive, we observe the channel-independent to work better under some datasets. Determining precisely when a CI or CD method is superior remains an open question in the field. We hypothesize that probabilistic forecasting using quantile regression within our framework may require better calibration, and exploring further mitigation strategies to address this deficit serves as an important direction for future research.

REFERENCES

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

Abdul Fatir Ansari, 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. *arXiv preprint arXiv:2510.15821*, 2025. URL https://arxiv.org/abs/2510.15821.

Samuel Dooley, Gurnoor Singh Khurana, Chirag Mohapatra, Siddartha V Naidu, and Colin White. Forecastpfn: Synthetically-trained zero-shot forecasting. In *Advances in Neural Information Processing Systems*, 2023.

Claude Duchon and Robert Hale. *Time Series Analysis in Meteorology and Climatology: An Introduction*. Advancing Weather and Climate Science. John Wiley & Sons, Chichester, UK, 2012. ISBN 978-0-470-97199-4.

Léo Grinsztajn, Klemens Flöge, Oscar Key, Felix Birkel, Brendan Roof, Phil Jund, Benjamin Jäger, Adrian Hayler, Dominik Safaric, Felix Jablonski Simone Alessi, Mihir Manium, Rosen Yu, Anurag Garg, Jake Robertson, Shi Bin (Liam) Hoo, Vladyslav Moroshan, Magnus Bühler, Lennart Purucker, Clara Cornu, Lilly Charlotte Wehrhahn, Alessandro Bonetto, Sauraj Gambhir, Noah Hollmann, and Frank Hutter. Tabpfn-2.5: Advancing the state of the art in tabular foundation models, 2025.

Lu Han, Han-Jia Ye, and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

Shi Bin Hoo, Samuel Müller, David Salinas, and Frank Hutter. From tables to time: How tabpfn-v2 outperforms specialized time series forecasting models, 2025. URL https://arxiv.org/abs/2501.02945.

Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, March 2019. ISSN 1573-756X. doi: 10.1007/s10618-019-00619-1. URL http://dx.doi.org/10.1007/s10618-019-00619-1.

Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200209, 02 2021. ISSN 1364-503X. doi: 10.1098/rsta.2020.0209. URL https://doi.org/10.1098/rsta.2020.0209.

Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130 – 141, 1963. doi: 10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.

CO;2. URL https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963_020_0130_dnf_2_0_co_2.xml.

Max Planck Institute for Biogeochemistry. Jena climate dataset, 2017. URL https://www.bgc-jena.mpg.de/wetter/.

Vladyslav Moroshan, Julien Siems, Arber Zela, Timur Carstensen, and Frank Hutter. Tempopfn: Synthetic pre-training of linear rnns for zero-shot time series forecasting, 2025.

Prathamesh Patil, Amit Varshney, Manoj Cherukumalli, Harsh Deshpande, Leonard Eun, Dushyant Sahoo, and Naren Chittar. MORPHEUS : A foundation model for multivariate time series forecasting. In *1st ICML Workshop on Foundation Models for Structured Data*, 2025. URL https://openreview.net/forum?id=3q3dI6hwal.

Ege Onur Taga, Muhammed Emrullah Ildiz, and Samet Oymak. Timepfn: Effective multivariate time series forecasting with synthetic data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20761–20769, 2025.

Ruey S. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, NJ, 2. ed. edition, 2005. ISBN 978-0-471-69074-0. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+483463442&sourceid=fbw_bibsonomy.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, 2021.

A Probabilistic Forecast Accuracy

We further evaluate probabilistic forecast accuracy via the Weighted Quantile Loss (WQL). While our standardization strategies prove beneficial (Appendix E), our approach demonstrates a performance deficit compared to TabPFN-TS (Table 1). Future work is required to isolate the drivers of this behavior and investigate potential mitigation strategies.

B Sensitivity of TabPFN-TS-MV to column permutations.

Our approach leverages the TabPFN-TS framework that introduces temporal feature engineering to TabPFN v2 Hoo et al. (2025). In our experiments, we maintain a fixed column order for input data including the position of the channel indicator column. Since our method enforces an ordering across columns when the time series is rolled out, it is natural that the performance of our method is sensitive to column permutations. However, TabPFN v2 inherently ensures robustness to this ordering choice: it's inference process is an ensemble method that automatically performs row and column permutations and subsampling. Thus, we do not perform any ablations on the column ordering in this work.

C Detailed Results

Dataset	MASE		WQL	
	TabPFN-TS	TabPFN-TS-MV (ours)	TabPFN-TS	TabPFN-TS-MV (ours)
bizitobs_application/10S/long	3.3314	**3.1884**	**0.0483**	0.0490
bizitobs_application/10S/medium	2.8469	**2.5012**	**0.0395**	0.0415
bizitobs_application/10S/short	1.7608	**1.4309**	0.0178	**0.0171**
bizitobs_l2c/5T/medium	1.2231	**1.0165**	0.5009	**0.4470**
bizitobs_l2c/5T/short	0.4127	**0.3548**	0.1046	**0.0973**
bizitobs_l2c/H/long	0.6166	**0.6165**	**0.2880**	0.3027
bizitobs_l2c/H/medium	0.6899	**0.5444**	0.3120	**0.2815**
bizitobs_l2c/H/short	0.5022	**0.4904**	0.2242	**0.2241**
bizitobs_service/10S/long	1.5209	**1.3807**	**0.0524**	0.0527
bizitobs_service/10S/medium	1.4672	**1.2911**	**0.0365**	0.0409
bizitobs_service/10S/short	0.9972	**0.9042**	0.0181	**0.0169**
ett1/15T/long	**1.2373**	1.2947	**0.2883**	0.3208
ett1/15T/medium	**1.1681**	1.2228	**0.2766**	0.3071
ett1/15T/short	**0.8265**	0.9547	**0.1845**	0.2301
ett1/D/short	1.8064	**1.7220**	**0.2882**	0.3112
ett1/H/long	1.3978	**1.3638**	0.2825	**0.2782**
ett1/H/medium	**1.3425**	1.3480	0.2799	**0.2700**
ett1/H/short	**0.8913**	0.9027	**0.1998**	0.2033
ett1/W/short	**1.6502**	1.7472	**0.2824**	0.2948
ett2/15T/long	**1.0210**	1.0254	0.1050	**0.1022**
ett2/15T/medium	**0.9968**	1.0091	0.1033	**0.1032**
ett2/15T/short	0.9074	**0.9028**	**0.0779**	0.0804
ett2/D/short	1.7877	**1.6473**	0.1306	**0.1243**
ett2/H/long	1.2061	**1.1680**	0.1247	**0.1184**
ett2/H/medium	1.1583	**1.1288**	0.1192	**0.1133**
ett2/H/short	0.8164	**0.7941**	0.0720	**0.0693**
ett2/W/short	**0.7571**	1.4492	**0.0982**	0.1507
jena_weather/10T/short	**0.4473**	0.5111	**0.0690**	0.0701
jena_weather/D/short	**1.3235**	1.6777	**0.0508**	0.0786
jena_weather/H/long	1.1469	**1.1449**	**0.0607**	0.0714
jena_weather/H/medium	**0.9479**	1.0637	**0.0619**	0.0642
jena_weather/H/short	**0.6395**	0.7051	**0.0481**	0.0543
Average	1.2139	**1.2032**	**0.1514**	0.1558

Table 1: **Detailed performance comparison with TabPFN-TS.** We evaluate our approach using the multivariate datsets in the Gift-Eval time series benchmark Aksu et al. (2024). Following standard practice we evaluate point forecast accuracy using the Mean Absolute Scaled Error (MASE) and probabilistic forecast accuracy via the Weighted Quantile Loss (WQL). Lower is better for both metrics. Standard deviations are omitted, as variance across random seeds for TabPFN was negligible.

D Comparison with SOTA Time Series Models

We compare our approach against current state-of-the-art (SOTA) foundational time series models below. Consistent with our main results, average metrics are computed using the multivariate datasets from Gift-Eval. Notably, Chronos-2 exhibited significantly degraded performance on the Jena-Weather dataset, likely due to its high variate count (21). Consequently, we also report average metrics excluding this specific dataset to provide a more robust comparison.

	MASE				
	Chronos 2	TempoPFN	TimePFN	TabPFN-TS	TabPFN-TS-MV (ours)
Average	6.8068	1.2838	2.8429	1.2386	**1.2255**
Average (w/o jena)	**1.0355**	1.32823	3.00798	1.2718	1.2370
Average Rank	**1.6452**	2.8387	4.871	2.9677	2.6774

Table 2: **MASE comparison with SOTA Models (Lower is better):** Average (w/o jena) denotes the average MASE across the multivariate datasets in the Gift-Eval Benchmark excluding the Jena Weather dataset Max Planck Institute for Biogeochemistry (2017). We denote this metric due the poor performance of Chronos-2 specifically for this dataset.

	WQL				
	Chronos 2	TempoPFN	TimePFN	TabPFN-TS	TabPFN-TS-MV (ours)
Average	0.2577	0.1511	0.3354	**0.1541**	0.1586
Average (w/o jena)	**0.1392**	0.1648	0.3659	0.1687	0.1721
Average Rank	**1.5484**	2.6129	4.871	2.9032	3.0645

Table 3: **WQL comparison with SOTA Models (Lower is better):** Average (w/o jena) denotes the average WQL across the multivariate datasets in the Gift-Eval Benchmark excluding the Jena Weather dataset Max Planck Institute for Biogeochemistry (2017). We denote this metric due the poor performance of Chronos-2 specifically for this dataset.

A detailed per dataset MASE comparison is provided below (see Figure 3). A separate detailed comparison is provided for TimePFN, which requires a fixed context window of 96 time steps. We restrict our model's context to match this constraint for fairness (see Figure 4).

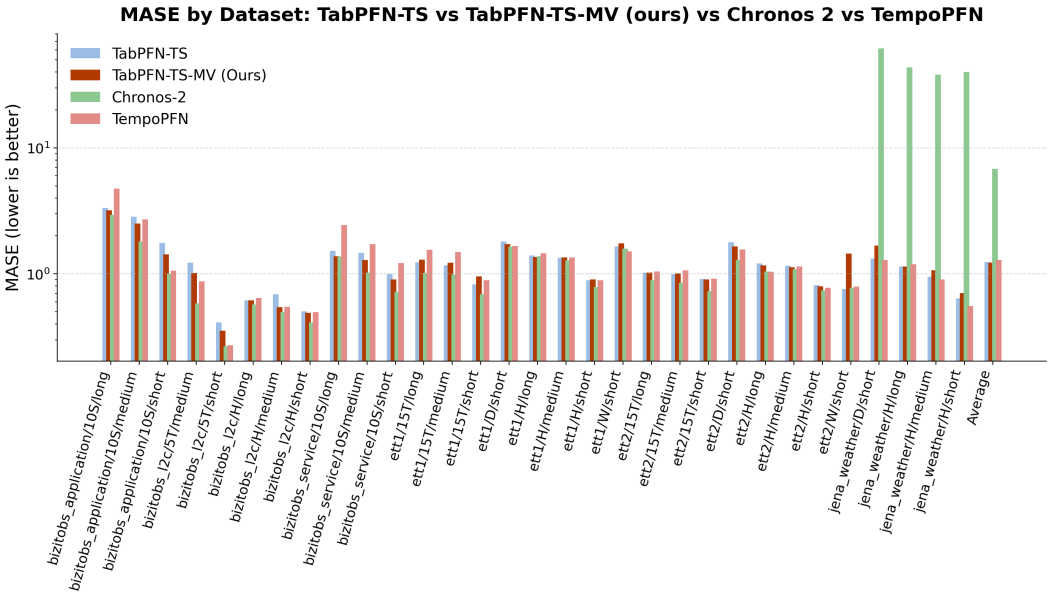

Figure 3: **Performance of TabPFN-TS-MV compared to other SOTA time series approaches:** The y-axis is log scale, Tabpfn-TS-MV (ours) and Chronos 2 supports channel dependent (CD) multivariate predictions while TabPFN-TS and TempoPFN only support univariate (channel independent) predictions.

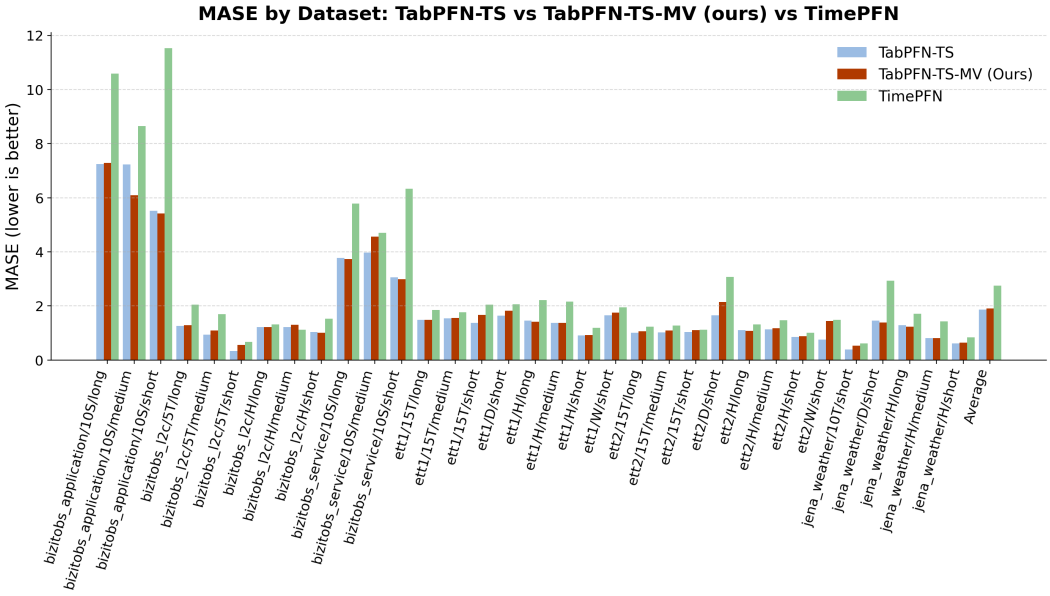

Figure 4: **Performance of TabPFN-TS-MV compared to TimePFN:** While TimePFN supports joint multivariate time series forecasting it only supports a maximum context window of 96 time steps. This poses a significant limitation as a majority of the datasets in Gift-Eval consist of much longer time series. All approaches are limited to a context window of 96 time steps (per variate) for fair comparison

E INVESTIGATING STRATEGIES FOR MITIGATING DISTRIBUTIONAL SHIFTS BETWEEN CHANNELS

As observed in Han et al. (2024), distributional differences between channels pose a significant challenge when performing joint multivariate forecasting. To address this, we evaluate two strategies for mitigating distributional discrepancies. First, we employ first-order differencing, where the model predicts the increment $\Delta y_t = y_t - y_{t-1}$ rather than the absolute value y_t. Second, we experiment with channel-wise standardization, training the model to predict normalized values $z_t = \frac{y_t - \mu}{\sigma}$. Crucially, these operations are performed independently for each channel, and the inverse transformations are applied to the predictions prior to calculating evaluation metrics. Our experiments reveal that while standardization improves accuracy, differencing actually degrades performance. The relative performance of these approaches is illustrated in Figure 5. Consequently, we adopt standardization for our final framework.

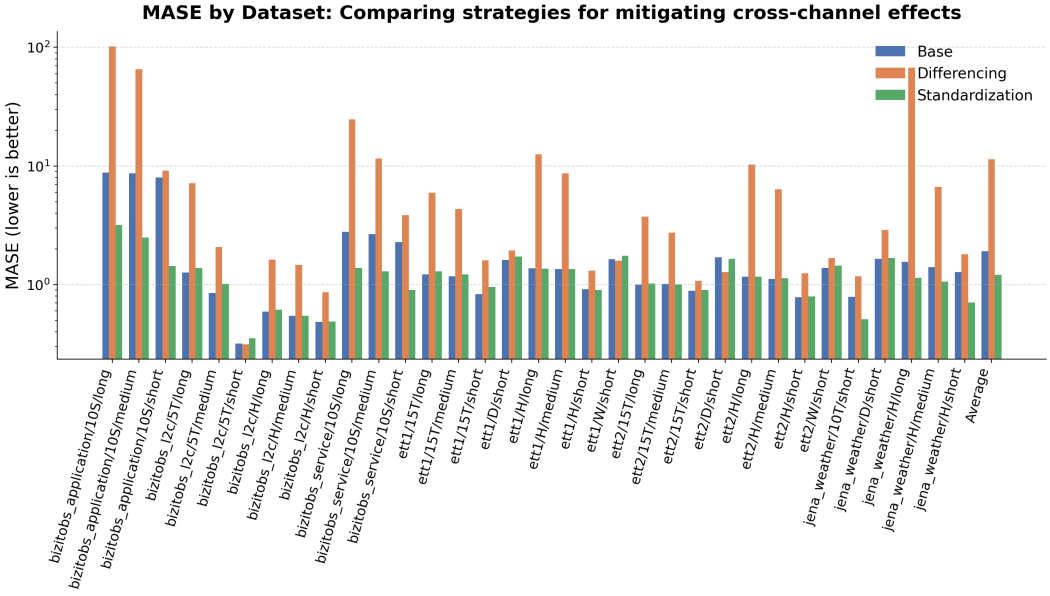

Figure 5: **Performance comparison between different strategies for mitigating cross-channel distributional shifts.** The y-axis is in log scale. We observe that while the standardizing approach greatly improves performance, first order differencing degrades performance. We incorporate the standardization approach into our framework.