TempoPFN: Towards Synthetic Pre-training of Linear RNNs for Zero-shot Time Series Forecasting

Vladyslav Moroshan$^{*\diamondsuit}$**, Julien Siems**$^{*\diamondsuit}$**, Arber Zela**$^{\diamondsuit}$**, Timur Carstensen**$^{\diamondsuit\clubsuit}$**, Frank Hutter**$^{\heartsuit\clubsuit\diamondsuit}$
University of Freiburg$^{\diamondsuit}$, ELLIS Institute Tübingen$^{\clubsuit}$, Prior Labs$^{\heartsuit}$
*Equal contribution vlad.moroshan@gmail.com juliensiems@gmail.com

Abstract

Foundation models for zero-shot time series forecasting face challenges in efficient long-horizon prediction and reproducibility, with existing synthetic-only approaches underperforming on challenging benchmarks. This paper presents TempoPFN, a univariate time series foundation model based on linear Recurrent Neural Networks (RNNs) pre-trained exclusively on synthetic data. The model uses a GatedDeltaProduct architecture with state-weaving for fully parallelizable training across sequence lengths, eliminating the need for windowing or summarization techniques while maintaining robust temporal state-tracking. Our comprehensive synthetic data pipeline unifies diverse generators—including stochastic differential equations, Gaussian processes, and audio synthesis—with novel augmentations. In zero-shot evaluations on the Gift-Eval benchmark, TempoPFN achieves top-tier competitive performance, outperforming all existing synthetic-only approaches and surpassing the majority of models trained on real-world data, while being more efficient than existing baselines by leveraging fully parallelizable training and inference. We open-source our complete data generation pipeline and training code, providing a reproducible foundation for future research.

1 Introduction

Recent advances in large language models have inspired foundation models for time series forecasting that enable zero-shot predictions across diverse datasets without fine-tuning [3, 4, 11, 34]. By treating historical observations as input context, these models democratize forecasting for non-experts and excel in data-scarce domains.

However, current approaches face critical limitations. Transformer-based models struggle with long-horizon forecasting due to quadratic complexity and error accumulation [40]. While non-linear RNNs like those in TiReX [4] maintain temporal state, they require sequential processing that limits scalability. Although some recent models attempt synthetic-only pre-training including ForecastPFN [13], CauKer [35], and Mamba4Cast [7] none reported state-of-the-art performance on the Gift-Eval benchmark. TabPFN-TS [19], which adapts a tabular foundation model to time series, achieves strong Gift-Eval performance but does not release its synthetic pre-training data, limiting reproducibility and extensibility.

We introduce **TempoPFN** (see Table 1 and **??**), a time series forecasting foundation model using *linear RNNs with GatedDeltaProduct recurrence* [29] for parallelizable training and inference across the sequence length. Unlike TiRex [4] which argued that non-linear RNNs like sLSTM are necessary for time-series forecasting due to their state-tracking capabilities we find that linear RNNs based on the GatedDeltaProduct recurrence are sufficient, in line with recent research demonstrating how linear RNNs can perform state-tracking [15]. Our synthetic data pipeline unifies diverse generators with novel augmentations, ensuring exclusive synthetic pre-training to prevent benchmark leakage.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Tabular Data workshop at EurIPS 2025.

Unlike TabPFN-TS, we open-source our complete data generation pipeline and training code as a basis for future research (available at `https://github.com/automl/TempoPFN`).

In summary, our contributions are:

- The **TempoPFN** architecture, to our knowledge, the first univariate time series foundation model based on *linear RNNs with GatedDeltaProduct recurrence*. Our architecture and input representation allows the prediction of all future time-stamps in parallel, producing coherent quantile forecasts, without patching or windowing heuristics. We further propose a state-weaving mechanism for linear RNNs that facilitates bidirectional information flow across horizons without overhead.
- We design a **synthetic data pipeline** combining existing and novel synthetic generators with a cascade of augmentations, ensuring diverse temporal structures without relying on real-world data, thereby eliminating benchmark leakage and mitigating privacy concerns associated with training on real-world data. We release a fully open-source synthetic data generation pipeline for time series forecasting that achieves competitive competitive performance on Gift-Eval.
- Compared to nonlinear RNNs and transformer time series foundation models, TempoPFN achieves **top-tier competitive zero-shot performance on Gift-Eval**, surpassing all other synthetic-only approaches and the vast majority of models trained on real-world data. This result is achieved without any non-linearity in the recurrence, demonstrating that linear RNNs as a scalable and powerful alternative to non-linear RNNs and transformers for time series foundation models.

2 Background and Related Work

Time Series Forecasting. Time series forecasting aims to predict future values $y_{T+1:T+H}$ from historical observations $y_{1:T}$. Traditional methods such as ARIMA [8] and exponential smoothing [21] typically produce point estimates, while probabilistic forecasting captures uncertainty by modeling the predictive distribution $p(y_{T+1:T+h} \mid y_{1:T})$. The advent of deep learning has significantly expanded this toolkit, introducing transformers [33] and modern recurrent architectures [5, 16].

A major recent development is *zero-shot forecasting*, where models pre-trained on diverse corpora can directly predict unseen time series without requiring fine-tuning. This paradigm mirrors the transformative shift seen in natural language processing and computer vision, where foundation models enable efficient cross-domain adaptation without costly per-task training.

Most successful zero-shot approaches leverage transformer architectures. Chronos [3], TimesFM [11], and MOIRAI [34] use techniques such as patching, frequency-specific projections, and masked modeling to handle heterogeneous time series data. Building on this foundation, MOIRAI-MOE [23] incorporates a sparse mixture-of-experts architecture to achieve token-level specialization and improved robustness. Among true zero-shot models, MOIRAI currently demonstrates state-of-the-art performance on Gift-Eval while carefully avoiding overlap with evaluation benchmarks.

An alternative approach focuses on synthetic data pretraining. ForecastPFN [13] trains exclusively on synthetic distributions featuring multi-scale trends, seasonality, and Weibull noise, enabling Bayesian zero-shot inference through single forward passes. TimePFN [31] extends this framework to multivariate scenarios using Gaussian process kernels and linear coregionalization. Similarly, TabPFN-TS [19] represents time series in tabular format and leverages TabPFNv2 [18], achieving competitive performance despite the limited availability of its underlying synthetic data.

Recent work has also revisited recurrent architectures for long-horizon forecasting. TiRex [4], currently the top performer on Gift-Eval, uses xLSTM [5] pre-trained on synthetic Gaussian processes, Chronos datasets, and carefully selected Gift-Eval subsets, enhanced with data augmentation techniques including amplitude modulation, censoring, and spike injection. In contrast, TempoPFN takes a different approach by exploiting linear RNNs with GatedDeltaProduct mechanisms and negative eigenvalues, enabling fully parallelizable training without requiring patching or summarization while relying exclusively on synthetic pretraining data to eliminate real-world leakage concerns.

3 TempoPFN

3.1 Architecture

The TempoPFN architecture is designed to forecast univariate time series across a full prediction horizon in a single forward pass, as illustrated in Figure 1. It consists of four main stages: input representation, backbone, non-causality through state weaving, and prediction.

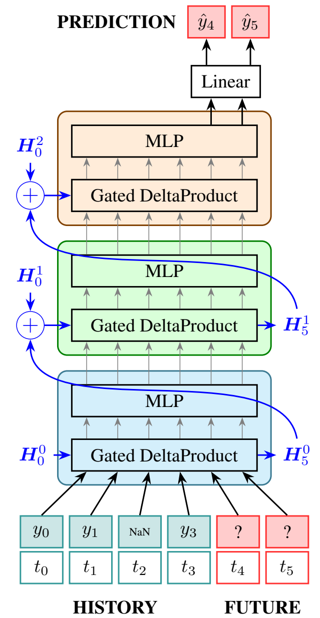


Figure 1: The **TempoPFN** architecture (3 blocks), using stacked GatedDeltaProduct blocks, learnable initial states $\boldsymbol{H}_0^i$ and state-weaving.

Input representation. TempoPFN uses an input representation in which history (timesteps + values) and future (timesteps) are concatenated into one token sequence enabling communication between future time-steps for coherent predictions. In contrast to TiReX, which presummarizes time-steps into windows of size 32, TempoPFN directly operates on the individual time-steps. Each time step t_i is encoded using `gluonts` [2] time features (e.g., seasonality indicators, day-of-week, or index-based encodings) that are linearly projected into the embedding dimension of the model. For historical steps, observed values y_i are projected via a linear layer, while missing values are handled by a learnable `NaN` embedding. The historical embedding is obtained by additively combining the value embedding and the time-feature embedding. For future time steps, only the time-feature embedding is used.

Backbone. The core of TempoPFN is a stack of 10–20 encoder layers, each based on the *Gated DeltaProduct block* from the `flash-linear-attention` library [36], originally derived from the LLaMA architecture [32]. Each block consists of three components: (1) *token mixing* through a Gated DeltaProduct recurrence with short one-dimensional convolutions (kernel size 8–16), (2) *pre-normalization* applied before the recurrent unit to stabilize training, and (3) a gated MLP for channel-wise feature transformation. This design combines the parallelization advantages of linear recurrences with the expressivity of lightweight convolutional and feedforward operations.

Non-causality via state weaving. Whereas DeltaProduct was originally developed for autoregressive language modeling, forecasting across a full horizon does not require causal masking. To exploit this property, we introduce *state weaving*. Specifically, the final hidden state of each layer $\boldsymbol{H}_t^i$ is added to the learnable initial state of the next layer $\boldsymbol{H}_0^{i+1}$. This mechanism enables bidirectional information flow across the entire sequence length without additional parameters or computational overhead through explicit bidirectionality [1, 20], effectively lifting the causal constraint while maintaining computational efficiency.

Prediction. At the output stage, embeddings corresponding to the forecast horizon are extracted from the final encoder block. These embeddings are passed through a linear projection head that outputs multiple *quantiles* of the predictive distribution, enabling probabilistic forecasting. Overall, this design allows to directly predict all future values given a history using a single forward pass.

3.2 Synthetic Data and Training

Synthetic Data Pipeline. Our training relies exclusively on synthetic data generated from 10 diverse generators (detailed in Appendix C). We combine established generators including ForecastPFN [**?**], KernelSynth [**?**], Gaussian Process [7], and CauKer [35] with novel generators: Sawtooth, Step Function, Anomaly, Spikes, Sine Wave, Audio-Inspired (modeling stochastic rhythms, financial volatility, network topology, and fractals), and our most sophisticated contribution—a stochastic differential equation (SDE) generator based on regime-switching Ornstein-Uhlenbeck processes with time-varying parameters and optional fractional Brownian motion.

The synthetic series undergo a comprehensive augmentation pipeline (detailed in Appendix D) that applies probabilistic transformations across six categories: Invariances (time reversal, sign inversion), Structure (regime changes, shock-recovery), Seasonality (amplitude modulation, calendar effects), Analytic (differential operators, integration), Discrete (censoring, quantization), and Artifacts (resampling). We implement time-varying TS-Mixup for novel combinations and apply random convolutional filtering. This ensures diverse temporal patterns while maintaining exclusive synthetic pre-training to prevent benchmark leakage.

Pre-training Setup. TempoPFN's pre-training is conducted exclusively on synthetic data. The training corpus consists of approximately 10 million time series from our generators, each with a maximum length of 2048. We train our main model (34.69M parameters) using a two-stage protocol: 3 million iterations at peak LR 2×10^{-4}, followed by 2 million iterations at 2×10^{-5}, using the AdamW optimizer [25] and quantile loss. Complete training details are in Appendix E.

4 Results

We evaluate TempoPFN on the Gift-Eval benchmark, a comprehensive zero-shot forecasting suite covering diverse real-world datasets across domains and horizons. TempoPFN surpasses TabPFN-TS, the strongest model trained exclusively on synthetic data. Remarkably, despite relying solely on synthetic training data, TempoPFN matches or exceeds several leading models trained on real-data, including Chronos Bolt B (0.574/0.808), TimesFM 2.0 (0.550/0.758), and YingLong 50m (0.567/0.822), ranking 6th overall in CRPS and 5th in MASE. Figure 2 summarizes these quantitative comparisons against state-of-the-art baselines.

Qualitative results (Appendix H) show that TempoPFN produces coherent predictive distributions without artifacts. Compared to TabPFN-TS, TempoPFN generates smoother uncertainty bounds while maintaining competitive point forecasts, likely because our architecture allows future time-steps to communicate rather than predicting them in isolation. In many longer predictions made by TiRex, we observe high-frequency artifacts in the quantile predictions, which we hypothesize result from TiRex's windowing approach that compresses time-series into chunks of size 32. Since TempoPFN requires no windowing, we did not notice similar artifacts.

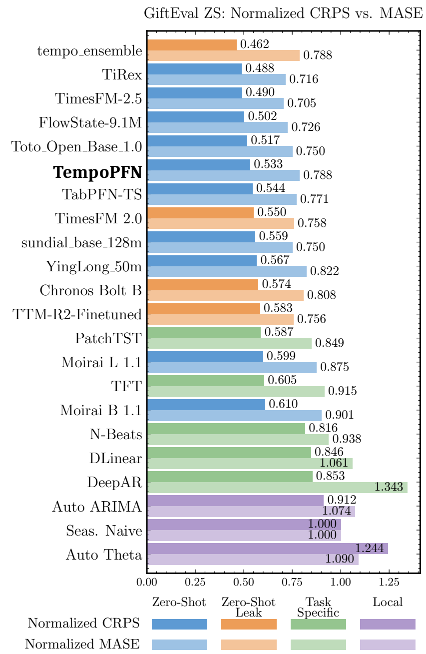


Figure 2: Comparison of TempoPFN performance against other models on GiftEval benchmark. We compute normalized scores for CRPS and MASE. Colors represent the class of time series model.

Our ablation studies (Appendix G) demonstrate that every synthetic generator contributes to performance, with the SDE generator being most critical—its removal increases CRPS by 26% from 0.578 to 0.729. Augmentations yield minor improvements in MAE and CRPS across horizons. Architectural ablations reveal complex interactions between state-weaving and negative eigenvalues in the state transition matrix.

5 Conclusion

We introduce *TempoPFN*, which combines a time-series architecture based on linear RNNs with an open-source synthetic data generation pipeline. Our method achieves performance comparable to models trained on real-world data when evaluated on the time-series benchmark Gift-Eval. Our architecture uses expressive linear RNNs that maintain state-tracking capabilities while preserving linearity, in contrast to approaches like TiRex, which use non-linear RNNs. This design enables efficient processing while retaining modeling capacity for complex time-series patterns. The synthetic data generation pipeline integrates both new and existing data generators within an augmentation framework. This pipeline combines mixing techniques, distortions, and differential transformations to produce diverse synthetic time-series data. Future work could focus on extending our current univariate synthetic time-series framework to multivariate scenarios. Additionally, incorporating pre-training on diverse real-world time series datasets could further enhance forecasting accuracy and generalization. Finally, investigating the performance of Linear RNN architectures against Transformer-based models for zero-shot forecasting represents a potential direction for further research.

Acknowledgments and Disclosure of Funding

We extend our heartfelt gratitude to the maintainers of the open-source libraries that were instrumental to this research, including Accelerate, xFormers, PyTorch, Flash-Linear-Attention, PyTorch Lightning, Weights & Biases, GluonTS, NeuralForecast, sktime, sktime-neuro, and mtenv. We also acknowledge the valuable contributions of Jonas Köhncke, Valentin Pratz, and the entire team at Cranberry-Lemon University for their insightful discussions and support. This work was supported in part by the NeurIPS Foundation and Google's TPU Research Cloud (TRC).

References

References

[1] Arshia Afzal, Elias Abad Rocamora, Leyla Naz Candogan, Pol Puigdemont, Francesco Tonin, Yongtao Wu, Mahsa Shoaran, and Volkan Cevher. Linear attention for efficient bidirectional sequence modeling. *arXiv preprint arXiv:2502.16249*, 2025.

[2] Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Danielle C. Maddix, Syama Rangapuram, David Salinas, Jasper Schulz, Lorenzo Stella, Ali Caner TÃ¼rkmen, and Yuyang Wang. Gluonts: Probabilistic and neural time series modeling in python. *Journal of Machine Learning Research*, 21(116):1–6, 2020.

[3] 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. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

[4] 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. *arXiv preprint arXiv:2505.23719*, 2025.

[5] Maximilian Beck, Korbinian Pöppel, Markus Spanring, Andreas Auer, Oleksandra Prudnikova, Michael Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. xlstm: Extended long short-term memory. In *Thirty-eighth Conference on Neural Information Processing Systems*, 2024.

[6] Ali Behrouz, Peilin Zhong, and Vahab Mirrokni. Titans: Learning to memorize at test time. *arXiv preprint arXiv:2501.00663*, 2024.

[7] Sathya Kamesh Bhethanabhotla and Omar Swelam. Mamba4cast: Efficient zero-shot time series forecasting with state space models. *arXiv preprint arXiv:2410.09385*, 2024.

[8] George E. P. Box and Gwilym M. Jenkins. Some recent advances in forecasting and control. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 17(2):91–109, 1968.

[9] Tri Dao and Albert Gu. Transformers are ssms: generalized models and efficient algorithms through structured state space duality. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24, 2024.

[10] Luke Nicholas Darlow, Artjom Joosen, Martin Asenov, Qiwen Deng, Jianfeng Wang, and Adam Barker. Tsmix: Time series data augmentation by mixing sources. In *Proceedings of the 3rd Workshop on Machine Learning and Systems*, EuroMLSys '23, page 109–114, New York, NY, USA, 2023. Association for Computing Machinery.

[11] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. *Proceedings of the 41st International Conference on Machine Learning*, 2024.

[12] Angus Dempster, François Petitjean, and Geoffrey I Webb. Rocket: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34: 1454–1495, 2020. doi: 10.1007/s10618-020-00701-z.

[13] 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.

[14] Xavier Gonzalez, Andrew Warrington, Jimmy T.H. Smith, and Scott W. Linderman. Towards scalable and stable parallelization of nonlinear rnns. In *Advances in Neural Information Processing Systems*, volume 37, pages 5817–5849. Curran Associates, Inc., 2024.

[15] Riccardo Grazzi, Julien Siems, Jörg K.H. Franke, Arber Zela, Frank Hutter, and Massimiliano Pontil. Unlocking State-Tracking in Linear RNNs Through Negative Eigenvalues. In *The Thirteenth International Conference on Learning Representations*, 2025.

[16] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

[17] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[18] Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin Hoo, Robin Tibor Schirrmeister, and Frank Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045):319–326, 2025.

[19] Shi Bin Hoo, Samuel Müller, David Salinas, and Frank Hutter. The Tabular Foundation Model TabPFN Outperforms Specialized Time Series Forecasting Models Based on Simple Features. In *NeurIPS 2024 Third Table Representation Learning Workshop*, 2024.

[20] Sukjun Hwang, Aakash Lahoti, Ratish Puduppully, Tri Dao, and Albert Gu. Hydra: Bidirectional state space models through generalized matrix mixers. In *Advances in Neural Information Processing Systems*, volume 37, pages 110876–110908. Curran Associates, Inc., 2024.

[21] Rob J. Hyndman, Anne B. Koehler, J. Keith Ord, and Ralph D. Snyder. *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, 2008.

[22] Kazuki Irie, Róbert Csordás, and Jürgen Schmidhuber. Practical computational power of linear transformers and their recurrent and self-referential extensions. In *The 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.

[23] Xu Liu, Juncheng Liu, Gerald Woo, Taha Aksu, Yuxuan Liang, Roger Zimmermann, Chenghao Liu, Junnan Li, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. Moirai-MoE: Empowering Time Series Foundation Models with Sparse Mixture of Experts. In *Forty-second International Conference on Machine Learning*, 2025.

[24] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts. In *International Conference on Learning Representations*, 2017.

[25] Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019.

[26] Eric Martin and Chris Cundy. Parallelizing linear recurrent neural nets over sequence length. In *International Conference on Learning Representations*, 2018.

[27] Bo Peng, Ruichong Zhang, Daniel Goldstein, Eric Alcaide, Xingjian Du, Haowen Hou, Jiaju Lin, Jiaxing Liu, Janna Lu, William Merrill, Guangyu Song, Kaifeng Tan, Saiteja Utpala, Nathan Wilce, Johan S. Wind, Tianyi Wu, Daniel Wuttke, and Christian Zhou-Zheng. RWKV-7 "Goose" with Expressive Dynamic State Evolution. In *Second Conference on Language Modeling*, 2025.

[28] Imanol Schlag, Kazuki Irie, and Jürgen Schmidhuber. Linear Transformers Are Secretly Fast Weight Programmers. In *International Conference on Machine Learning*, 2021.

[29] Julien Siems, Timur Carstensen, Arber Zela, Frank Hutter, Massimiliano Pontil, and Riccardo Grazzi. DeltaProduct: Increasing the Expressivity of DeltaNet Through Products of Householders. In *Advances in Neural Information Processing Systems*, volume 38. Curran Associates, Inc., 2025.

[30] Yu Sun, Xinhao Li, Karan Dalal, Jiarui Xu, Arjun Vikram, Genghan Zhang, Yann Dubois, Xinlei Chen, Xiaolong Wang, Sanmi Koyejo, Tatsunori Hashimoto, and Carlos Guestrin. Learning to (learn at test time): RNNs with expressive hidden states. *arXiv:2407.04620 [cs.LG]*, 2024.

[31] Ege Onur Taga, Muhammed Emrullah Ildiz, and Samet Oymak. TimePFN: effective multivariate time series forecasting with synthetic data. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025.

[32] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

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

[35] Shifeng Xie, Vasilii Feofanov, Marius Alonso, Ambroise Odonnat, Jianfeng Zhang, and Ievgen Redko. Cauker: Classification time series foundation models can be pretrained on synthetic data only. *arXiv preprint arXiv:2508.02879*, 2024.

[36] Songlin Yang and Yu Zhang. FLA: A Triton-Based Library for Hardware-Efficient Implementations of Linear Attention Mechanism, January 2024. URL `https://github.com/sustcsonglin/flash-linear-attention`.

[37] Songlin Yang, Bailin Wang, Yikang Shen, Rameswar Panda, and Yoon Kim. Gated Linear Attention Transformers with Hardware-Efficient Training. In *International Conference on Machine Learning*, 2024.

[38] Songlin Yang, Bailin Wang, Yu Zhang, Yikang Shen, and Yoon Kim. Parallelizing Linear Transformers with the Delta Rule over Sequence Length. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[39] Luca Zancato, Arjun Seshadri, Yonatan Dukler, Aditya Golatkar, Yantao Shen, Benjamin Bowman, Matthew Trager, Alessandro Achille, and Stefano Soatto. B'MOJO: Hybrid State Space Realizations of Foundation Models with Eidetic and Fading Memory. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[40] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, 2023.

A Our Contributions

Table 1: Contributions of TempoPFN: the first fully open-source time series forecasting foundation model with competitive performance; with fully synthetic pretraining and fast training and inference.

Criterion	Tirex	TabPFN-TS	Mamba4Cast	Chronos	TempoPFN
Fully open-source data pipeline	✗	✗	✓	✗	✓
Open-source training code	✗	✗	✓	✓	✓
Competitive with SOTA performance	✓	✓	✗	✓	✓
Fast training and inference	✓	✗	(✓)	(✓)	✓
Purely synthetic pretraining	✗	(✓)	✓	✗	✓

B Related Work

Linear RNNs. Recurrent neural networks have seen a resurgence in interest with the emergence of linear RNNs also known as state-space models. While non-linear RNNs are non-trivially parallelizable [14], linear RNNs can be parallelized using a chunk-wise parallel form [37] or an associative scan [16, 26]. Formally, linear RNN layers transform input sequences $\boldsymbol{x}_{1:t} \in \mathbb{R}^l$ into outputs $\hat{\boldsymbol{y}}_{1:t} \in \mathbb{R}^p$ through the recurrence

$$\boldsymbol{H}_i = \boldsymbol{A}(\boldsymbol{x}_i)\boldsymbol{H}_{i-1} + \boldsymbol{B}(\boldsymbol{x}_i) \text{ with output } \hat{\boldsymbol{y}}_i = \mathrm{dec}(\boldsymbol{H}_i, \boldsymbol{x}_i) \text{ for } i \in \{1, \dots, t\} \tag{1}$$

where $\boldsymbol{A} : \mathbb{R}^l \to \mathbb{R}^{n\times n}$ parameterizes the state-transition matrix, $\boldsymbol{B} : \mathbb{R}^l \to \mathbb{R}^{n\times d}$ governs state inputs, and $\mathrm{dec} : \mathbb{R}^{n\times d} \times \mathbb{R}^l \to \mathbb{R}^p$ produces the layer output. Variants such as Mamba [9], GLA [37], and mLSTM [**?**] adopt diagonal transitions, while others explore richer parameterizations. More expressive formulations relax diagonal state-transition constraints, as seen in DeltaNet [22, 28, 38], TTT-Linear [30], RWKV-7 [27], B'MOJO [39], and Titans [6].

Within this framework, we use **DeltaProduct** [29] as our token-mixing mechanism, which generalizes DeltaNet's non-diagonal transitions by expressing $\boldsymbol{A}(\boldsymbol{x}_i)$ as a product of n_h generalized Householder transformations, enabling a rank-n_h transformation of the matrix-valued hidden state $\boldsymbol{A}(\boldsymbol{x}_i) = \prod_{j=1}^{n_h}\big(\boldsymbol{I} - \beta_{i,j}\, \boldsymbol{k}_{i,j}\boldsymbol{k}_{i,j}^\top\big)$. For each token $\boldsymbol{x}_i$, the model generates n_h normalized keys $\boldsymbol{k}_{i,j} = \psi(\boldsymbol{W}_j\boldsymbol{x}_i)/\|\psi(\boldsymbol{W}_j\boldsymbol{x}_i)\|_2$, values $\boldsymbol{v}_{i,j} = \boldsymbol{V}_j\boldsymbol{x}_i$, and coefficients $\beta_{i,j} = \phi(\boldsymbol{U}_j\boldsymbol{x}_i)$ using learnable matrices $\boldsymbol{W}_j, \boldsymbol{V}_j, \boldsymbol{U}_j$, SiLU activation ψ [17], and a sigmoid-based gating function ϕ. Siems et al. [29] found that increasing n_h leads to significantly improved length-extrapolation, language modeling, and state-tracking on permutation tasks, all capabilties which are equally desirable for time-series forecasting.

C Detailed Synthetic Data Generation

This section provides comprehensive details on synthetic data generators used in TempoPFN.

CauKer. To increase the diversity and structural complexity of our training data, we used the CauKer generator from [35]. This method produces multivariate time series by sampling from a structural causal model (SCM) where each variable is a Gaussian process. A random Directed Acyclic Graph (DAG) defines the dependencies between nodes, with each node having a maximum number of parents. Root nodes in the DAG are sampled from a GP prior $y_i \sim \mathcal{GP}(m(t), \kappa(t, t'))$, using complex composite kernels κ (combined via + or *) and stochastic mean functions $m(t)$ (e.g., linear $at + b$, exponential $a\exp(bt)$, or functions with anomalous impulses). Child nodes then apply nonlinear activation functions (e.g., ReLU, sigmoid, sin) to affine combinations of their parents' values, introducing intricate, non-Gaussian dependencies. We generated multivariate series with 21 channels and treated each channel as an independent univariate time series.

KernelSynth. The KernelSynth generator, based on Chronos [**?**], samples independent univariate time series from Gaussian process priors $y \sim \mathcal{GP}(0, \kappa(t, t'))$. It constructs composite kernels κ by randomly combining base kernels (using addition or multiplication) from a large bank including periodic kernels (ExpSineSquared), stationary kernels (RBF, RationalQuadratic), and noise kernels (WhiteKernel).

Gaussian Process. The Gaussian Process generator, inspired by Mamba4Cast [7], extends the GP sampling approach with greater complexity. It constructs a composite kernel by combining up to six base kernels from a weighted bank that includes Matern, linear, periodic, and polynomial kernels. We inject periodic peak spikes aligned with the dominant periodicity of the sampled kernel, creating sharp, recurring events on top of smooth GP trajectories.

ForecastPFN. The ForecastPFN generator, adapted from **?**], creates time series with configurable trends, seasonality, and noise patterns. The trend component combines linear and exponential elements multiplicatively: $\tau(t) = [b + s_l(t + o_l)] \times s_e^{(t+o_e)}$. The seasonality component is multiplicative: $s(t) = \prod_f \left(1 + s_f \sum_h \left[c_{f,h} \sin\left(\frac{2\pi h(t+o_f)}{p_f}\right) + d_{f,h} \cos\left(\frac{2\pi h(t+o_f)}{p_f}\right)\right]\right)$. The final series values are $\tau(t) \cdot s(t) \cdot (1 + n(t))$, where $n(t)$ is Weibull-distributed noise.

Sawtooth. Creates univariate series with linear ramping patterns. The core waveform is $y_t = A \cdot \text{frac}((t/P) + \phi)$ for upward ramps, with minimal linear trends and low-amplitude seasonal components added to prevent overly idealised signals.

Step Function. Constructs complex piecewise constant series by concatenating multiple subseries. Each subseries is generated from patterns (stable, gradual trends, spikes, oscillations, random walks) with specific lengths, changepoints, step sizes, and drift, with optional Gaussian smoothing at transitions.

Anomaly. Produces constant baseline signals contaminated with periodic spike anomalies. Spike timing follows patterns (single, clustered, or mixed) with period variance and jitter, while magnitudes follow defined regimes (constant, trending, cyclical, or correlated random).

Spikes. Creates series where sharp spikes are the primary feature on flat baselines. Spikes have consistent direction and shape (V-shaped, inverted-V, or chopped variants with plateaus), generated in either "burst" or "spread" modes.

Sine Wave. Produces complex oscillatory patterns by summing 1 to 3 sinusoidal components with time-varying amplitude and frequency modulation: $y_t = \sum_{i=1}^{N} A_i(t) \sin(\phi_i(t)) + (at + b) + \epsilon_t$.

Audio-Inspired Generators. Four novel generators using procedural audio synthesis via the `pyo` library: (1) *Stochastic Rhythm* creates multi-layered, event-driven polyrhythmic patterns; (2) *Financial Volatility* mimics market dynamics with trends, volatility clustering, and sudden jumps; (3) *Network Topology* simulates network traffic with base flow, noise bursts, congestion dips, and DDoS-like spikes; (4) *Multi-Scale Fractal* produces self-similar patterns through parallel band-pass filters with logarithmically spaced frequencies.

Stochastic Differential Equations (SDEs). Our most sophisticated generator uses a regime-switching, time-inhomogeneous Ornstein–Uhlenbeck process: $dy_t = \theta(t, r_t)\big(\mu(t, r_t) - y_t\big)\,dt + \sigma(t, r_t)\,dW_t$ where $\theta(t, r_t)$ is mean reversion speed, $\mu(t, r_t)$ the time-varying mean, and $\sigma(t, r_t)$ the volatility. Each parameter depends on both time t and a latent regime $r_t \in \{0, 1\}$ that evolves as a Markov chain. Parameters shift abruptly across regimes while drifting smoothly over time through polynomial, sinusoidal, logistic, or piecewise-linear trends. When long memory is enabled, standard Brownian motion is replaced with fractional Brownian motion with Hurst exponent $H \in [0.3, 0.8]$.

D Data Augmentation Pipeline Details

Our augmentation pipeline applies transformations in a structured sequence. Base series undergo optional normalization (80% probability) using random scalers (Robust, MinMax, Median, or Mean). Early-stage TS-Mixup [10] creates convex combinations of multiple source series with probability $p = 0.5$. The core augmentation step samples 2-5 distinct transformation categories with weighted probabilities: Invariances (0.6), Structure (0.6), Seasonality (0.5), Signal Processing (0.4), Discrete Effects (0.6), and Measurement Artifacts (0.3). From each selected category, one specific transformation is randomly chosen and applied in fixed global order. Optional stochastic convolution filtering (probability $p = 0.3$) applies 1-3 random 1D convolutions with randomized parameters. Late-stage TS-Mixup provides additional combination opportunities, followed by finishing transformations including minor global scaling and low-magnitude Gaussian noise injection.

Transformation Categories. Invariance transformations promote robustness through temporal reversal ($\mathbf{x} \rightarrow \mathbf{x}_{T:1}$) and sign inversion ($\mathbf{x} \rightarrow -\mathbf{x}$), preserving temporal dependencies while testing

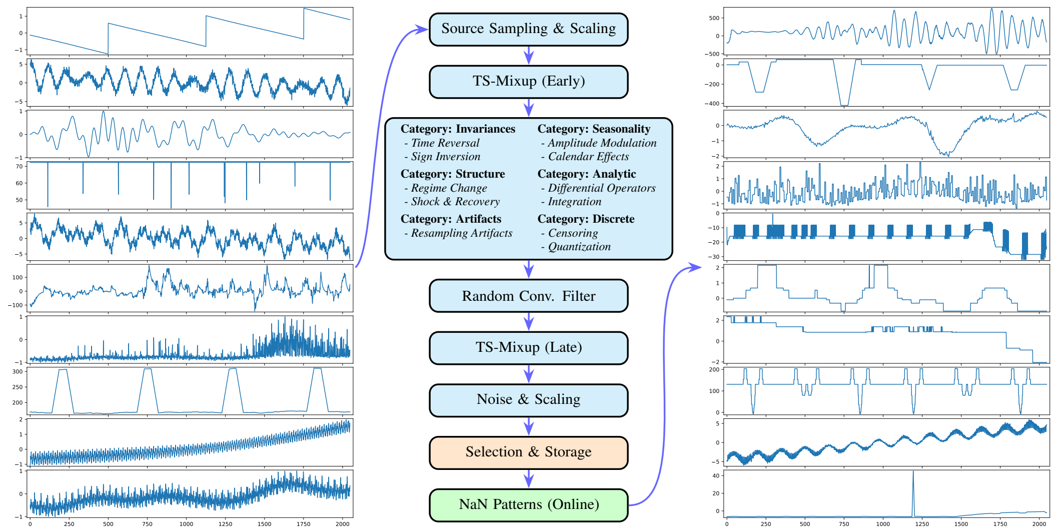


Figure 3: Synthetic data augmentation pipeline. Synthetic time-series undergo probabilistic transformations across six categorical groups with optional TS-Mixup combinations and stochastic convolution filtering.

directional conventions. **Structural modifications** inject non-stationarity via regime changes with piecewise affine transforms across random change-points, and shock-recovery dynamics using exponential decay impulses $I(t) = Ae^{-(t-t_0)/\tau}$ with randomized parameters.

Seasonal effects simulate real-world periodicities through calendar injections that apply multiplicative factors for weekend dips, month-end spikes, and holiday-like impulses using timestamp metadata. Amplitude modulation applies localized scaling to random segments, simulating time-varying volatility. **Signal processing** transformations include Gaussian smoothing followed by finite-difference operators (Sobel, Laplacian, higher-order derivatives up to 4th order) and numerical integration, with outputs rescaled to preserve original value ranges. Random convolution layers with highly randomized parameters [12] provide additional signal transformation capabilities.

Measurement artifacts introduce realistic data collection imperfections: censoring clips values at random quantiles (similarly used by TiRex [4]), non-uniform quantization maps values to discrete levels using quasi-random Sobol sequences, and resampling artifacts downsample and upsample series with various interpolation methods.

Combination strategies We implement TS-Mixup [3] to generate novel series through convex combinations of 2-10 source series, with mixing weights sampled from Dirichlet distributions and extend it with time-dependent mixing using smooth simplex path interpolation.

E Training Details and Hyperparameters

Data Composition and Sampling. The training corpus consists of approximately 10 million synthetic time series (500k–2M per generator), with batches composed of mixed samples from our generators. We apply higher weights to Cauker and augmented data to promote diversity in the training distribution.

Dynamic Structure Construction. Our training uses dynamic, per-sample construction of time series structures. For each training instance, we first randomly sample a total sequence length from a weighted distribution that favors longer contexts: {128: 0.05, 256: 0.10, 512: 0.10, 1024: 0.10, 1536: 0.15, 2048: 0.50}. When length shortening is applied, we use either cutting or subsampling with equal probability (50/50 split). Next, we perform a random history-future split, with forecast horizon lengths sampled from the range specified by the GIFT benchmark. This two-stage sampling creates highly variable training examples that simulate diverse forecasting tasks.

Data Augmentation. We apply several augmentation techniques during training: (1) Scaler augmentation with 0.5 probability, randomly selecting among minmax, median, or mean scalers (excluding the main robust scaler); (2) NaN augmentation that injects realistic missing data patterns into the history based on GIFT-Eval statistics.

Training Infrastructure. Pretraining uses PyTorch with distributed data parallelism (DDP) across 8–16 NVIDIA A100 or H100 GPUs and mixed precision (`bfloat16`), a requirement for the DeltaProduct implementation (in FLA: `https://github.com/fla-org/flash-linear-attention`).

Training Protocol. The training consists 4 million iterations with peak learning rate 2×10^{-4} and cosine annealing [24] schedulers (warmup ratio 0.003, minimum LR ratio 0.01).

We use the AdamW optimizer [25] with weight decay of 0.01, an effective batch size of approximately 200, and quantile loss over the [0.1–0.9] range. Other than weight decay, no explicit regularization such as dropout or early stopping is applied.

Architecture Selection. We ablated deeper models (8–16 layers) and found no consistent architectural winner. We selected the 10 layer model with embedding dimension of 512.

Table 2: Hyperparameters for main TempoPFN model.

Category	**Parameter**	**Value**
Model	**Total Parameters**	**41.98**
	Embedding size (`embed_size`)	512
	Encoder layers	10
	Number of heads (`num_heads`)	4
	Encoder attention mode	`chunk`
	Short convolution kernel size	32
	State weaving	`True`
	Quantiles for loss	[0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]
Training	Total training series	$\approx$ 10,000,000
	Max series length	2048
	Total training iterations	4,000,000
	Batch size (per GPU)	40
	Gradient accumulation steps	5
	Effective batch size	200
	Peak learning rate	2×10^{-4}
	LR scheduler	Cosine annealing
	Min learning rate ratio	0.01
	Warmup ratio	0.003
Optimization	Optimizer	AdamW
	β_1	0.9
	β_2	0.98
	Weight decay	0.01
	Adam ϵ	1×10^{-6}
	Gradient clipping	100.0
Augmentations	Length shortening	`True` (cut/subsample: 50/50)
	NaN augmentation	`True`
	Scaler augmentation prob.	0.5 (minmax/median/mean)
	Batch composition	Mixed (proportions favoring augmented/Cauker)
Hardware	GPUs	8–16 $\times$ A100/H100
	Precision	`bfloat16`

F Experimental Results

F.1 Ranking Results on Gift-Eval

Figure 4 shows the average ranks of all models across different forecasting horizons. TempoPFN consistently maintains strong performance across all horizon lengths, demonstrating robust generalization from synthetic pre-training data.

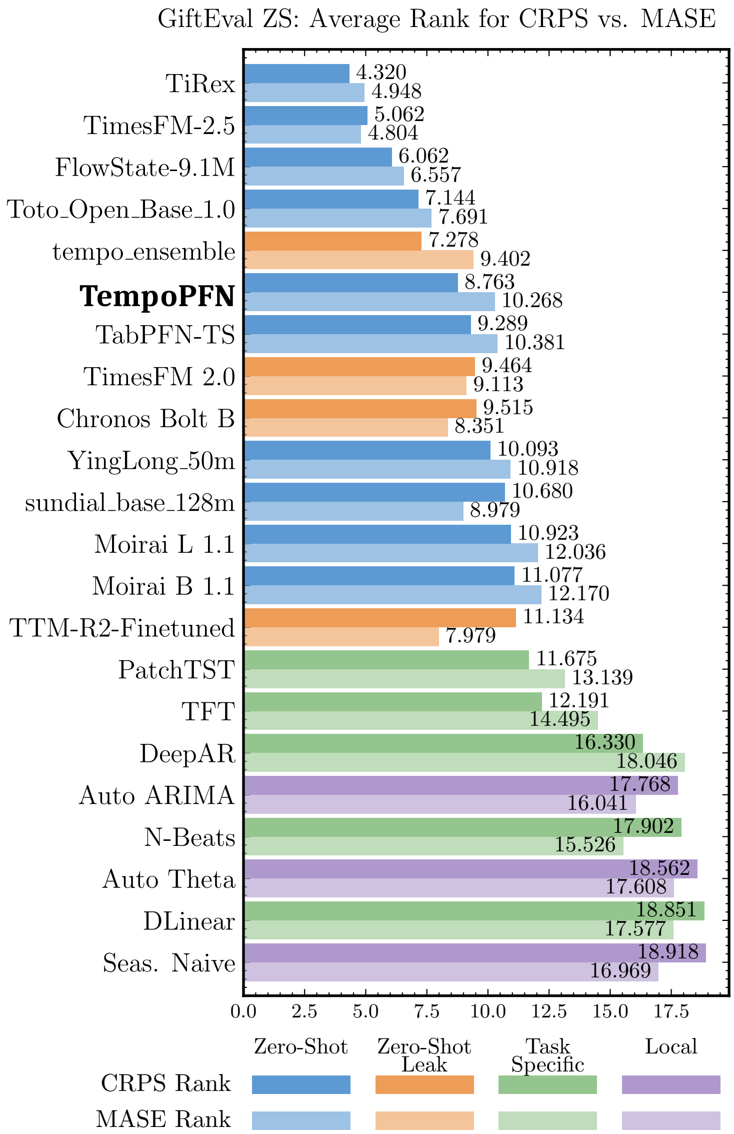


Figure 4: Average ranks of models on GiftEval benchmark across different horizons for CRPS and MASE metrics.

F.2 Robustness to Missing Values (NaNs)

We compare the robustness of TempoPFN and TiRex towards missing values (NaN) in the data. Figure 5 shows that both models exhibit performance degradation as the percentage of NaNs increases, however, TempoPFN is significantly more robust. While the normalized CRPS score (relative to TempoPFN's CRPS at 0% NaNs) for both models rises with more NaNs, TiRex's performance deteriorates more rapidly, with its median CRPS increasing by over 11% when 90% of the data is missing. In contrast, TempoPFN's median error increases by only about 4% under the same conditions, showcasing its superior stability and resilience when faced with incomplete time series data.

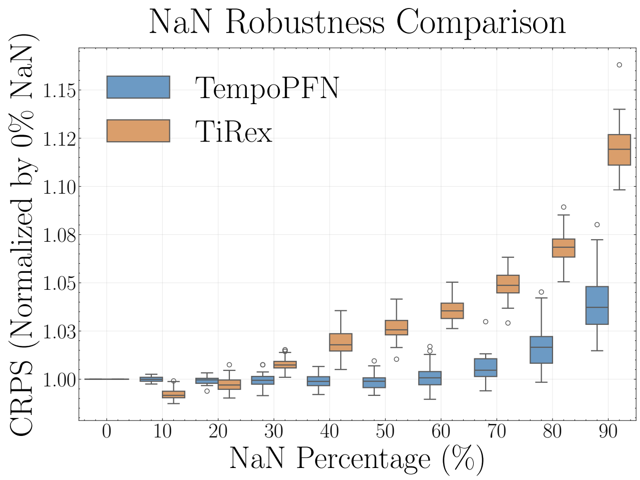


Figure 5: Normalized CRPS (relative to TempoPFN's CRPS at 0% NaNs) of TempoPFN and TiRex as a function of the percentage of missing values (NaN) in the data.

G Ablation Studies

Ablation Study Setup. To manage the significant computational cost of pre-training, not all ablation experiments were conducted for a full 4M-iteration training schedule. Therefore, these studies are designed to provide strong directional evidence on the relative importance of each component, rather than to measure their full, converged performance impact.

G.1 Ablating Synthetic Generators

To assess the individual contribution of each synthetic data source, we conducted an ablation study by retraining our model while excluding one synthetic time series generator at a time. As detailed in Table 3, the results reveal a clear hierarchy of importance, consistently observed across short, medium, and long-term forecasting horizons, with every generator proving beneficial for high performance. The highest impact data generator is our proposed SDE generator; its removal caused the most severe performance degradation, increasing the overall CRPS by 26% from 0.578 to 0.729. This highlights the importance of exposing the model to time series with mean-reverting and noisy, continuous-time dynamics. Significant, albeit smaller, performance losses were also observed upon removing generators responsible for complex seasonality (Cauker), abrupt changes (Step), and transient events (Spike), underscoring the necessity of a diverse pre-training corpus that captures a wide array of structural and stochastic patterns. In Table 5 in the appendix, we also compare our base model trained using all generators with models trained using a single generator at a time.

G.2 Ablating the Augmentation Pipeline

To quantify the impact of augmentations, we conducted controlled ablations, training models with identical hyperparameters on synthetic data with and without the full augmentation suite (see Table 4). Results on the GIFT-Eval benchmark reveal substantial gains: augmented training yields 15–25% relative improvements in MAE and CRPS across forecasting horizons, underscoring the role of our complex data augmentation pipeline for extrapolation to real-world data.

G.3 Architectural Ablations

Results on architectural ablations are provided in Table 6, where we notice that disabling either the 'weaving' mechanism or the allowance for negative eigenvalues in the state transition matrix (which enables state-tracking) individually results in a marginal performance improvement over the base model. However, this benefit is canceled out when both components are removed simultaneously, suggesting a complex interaction where the presence of one architecture feature compensates for the absence of the other.

Table 3: Ablation study of synthetic data priors using a leave-one-out methodology. The 'Ablation' column indicates the single prior excluded from the training mixture. Performance is measured by CRPS and MASE (lower is better). Rows are colored to indicate the performance impact on the overall CRPS when a prior is removed: High Impact ($> 25\%$ increase) and Medium Impact ($\geq 10\%$ increase). **N** = Novel prior (our contribution), **A** = Adapted from open-source.

		Gift-ZS Overall		Gift-ZS Short		Gift-ZS Medium		Gift-ZS Long	
Ablation	**Source**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**
Base Model	–	0.577	0.842	0.563	0.763	0.566	0.900	0.631	1.019
- GP	**A**	0.591	0.830	0.576	0.749	0.605	0.924	0.618	0.981
- Kernel	**A**	0.611	0.885	0.589	0.796	0.637	0.981	0.648	1.056
- ForecastPFN	**A**	0.617	0.885	0.588	0.791	0.643	0.981	0.674	1.075
- Sawtooth	**N**	0.628	0.900	0.597	0.800	0.661	1.012	0.684	1.091
- Sinewave	**N**	0.628	0.899	0.594	0.799	0.677	1.032	0.676	1.070
- Anomaly	**N**	0.630	0.897	0.592	0.794	0.684	1.024	0.683	1.079
- Step	**N**	0.640	0.927	0.605	0.819	0.686	1.063	0.689	1.120
- Stochastic Rhythm	**N**	0.642	0.911	0.601	0.802	0.701	1.043	0.699	1.111
- Spike	**N**	0.645	0.936	0.619	0.836	0.678	1.059	0.684	1.115
- Cauker	**A**	0.656	0.928	0.605	0.810	0.728	1.084	0.729	1.132
- SDE (OU Process)	**N**	**0.729**	**1.031**	**0.684**	**0.916**	**0.799**	**1.184**	**0.789**	**1.225**
seasonal_naive	–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: The impact of data augmentation on model performance, shown by normalized performance values. Lower values indicate better performance, demonstrating the significant improvement gained from using data augmentations.

	Gift-ZS Overall		Gift-ZS Short		Gift-ZS Medium		Gift-ZS Long	
Model	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**
w/ Aug	**0.577**	**0.842**	**0.563**	**0.763**	**0.566**	**0.900**	**0.631**	**1.019**
w/o Aug	0.610	0.875	0.582	0.783	0.617	0.963	0.643	1.059
seasonal_naive	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5: Ablation study of synthetic data priors where the model is trained on only a single prior at a time. To ensure tractability and manage computational costs, each ablation experiment was trained for a limited duration of **500,000 iterations**. This should be contrasted with the several million iterations used for our final, fully-trained model, as detailed in the main paper. The 'Base Model' was trained on all priors including augmentations. Lower values indicate better performance. Bold highlighting shows the best-performing individual priors, while underlined values indicate second-tier performance. Novel priors are our contributions, while Adapted priors are modified from open-source implementations.

		Gift-ZS Overall		Gift-ZS Short		Gift-ZS Medium		Gift-ZS Long	
Ablation	**Source**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**
Base Model	–	0.578	0.842	0.563	0.763	0.566	0.900	0.631	1.019
+ Cauker	Adapted	**0.600**	**0.875**	**0.583**	**0.789**	**0.615**	**0.964**	**0.631**	**1.043**
+ GP	Adapted	0.632	0.897	0.607	0.812	0.666	0.993	0.666	1.053
+ Kernel	Adapted	0.638	0.926	0.622	0.835	0.656	1.042	0.661	1.082
+ ForecastPFN	Adapted	0.715	1.027	0.695	0.918	0.760	1.172	0.726	1.206
+ SDE (OU Process)	Novel	**0.815**	**1.148**	**0.763**	**1.017**	**0.897**	**1.334**	**0.879**	**1.354**
+ Sinewave	Novel	0.868	1.223	0.854	1.113	0.901	1.375	0.872	1.397
+ Stochastic Rhythm	Novel	0.953	1.337	0.940	1.252	1.004	1.472	0.938	1.440
+ Sawtooth	Novel	1.187	1.534	1.162	1.362	1.294	1.802	1.152	1.781
+ Spike	Novel	1.215	1.318	1.019	1.250	1.565	1.411	1.498	1.416
+ Anomaly	Novel	1.310	1.522	1.487	1.610	1.145	1.430	1.075	1.399
+ Step	Novel	2.199	1.702	1.272	1.280	4.325	2.398	4.693	2.549
seasonal_naive	–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 6: Ablation study of core architectural components. All models were trained for **2M iterations** with d=512, L=10, convolution size 16, and the base configuration uses H=4 Householder matrices with weaving enabled and negative eigenvalues allowed. We systematically ablate key components to understand their individual contributions. Results are sorted by overall CRPS (best to worst). **Bold** indicates best performance, underline indicates second-best. Lower values for CRPS and MASE indicate better performance.

	Gift-ZS Overall		Gift-ZS Short		Gift-ZS Medium		Gift-ZS Long	
Configuration	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**
Ablating Positional Encoding:								
Base Model (Sin. Pos. Enc. Off)	**0.561**	**0.820**	**0.553**	**0.751**	**0.556**	**0.880**	**0.590**	**0.962**
Sinusoidal Positional Encoding	0.648	0.937	0.596	0.809	0.731	1.120	0.715	1.152
Ablating Number of Householder Matrices (H):								
H=6	**0.556**	0.823	**0.545**	**0.750**	**0.552**	0.886	**0.590**	0.972
Base Model (H=4)	0.561	**0.820**	0.553	**0.751**	0.556	**0.880**	**0.590**	**0.962**
H=2	0.562	0.822	0.549	0.750	0.560	0.892	0.598	0.964
H=1 (DeltaNet equivalent)	0.573	0.845	0.556	0.761	0.579	0.918	0.613	1.020
Ablating Negative Eigenvalues and Weaving:								
Neg. Eig. Off, Weaving On	**0.559**	0.821	**0.553**	0.753	**0.550**	0.881	**0.584**	0.957
Neg. Eig. Off, Weaving Off	0.560	**0.818**	0.554	**0.750**	0.548	**0.879**	0.590	**0.955**
Base Model (Neg. Eig. On, Weaving On)	0.561	0.820	0.553	0.751	0.556	0.880	0.590	0.962
Ablating Convolution Size:								
Conv. size 32	**0.559**	**0.816**	**0.543**	**0.737**	0.566	0.897	0.594	0.968
Base Model (Conv. size 16)	0.561	0.820	0.553	0.751	**0.556**	**0.880**	**0.590**	**0.962**
seasonal_naive	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 7: Ablation study of model scale and depth. All models were trained for **4M iterations** with H=4, convolution size 32, weaving enabled, and negative eigenvalues allowed. The **Base Model** uses d=512 embedding dimensions and L=10 layers. We explore the trade-off between model width (embedding dimension) and depth (number of layers) while keeping the parameter count roughly constant. Results are sorted by overall CRPS (best to worst). **Bold** indicates best performance, underline indicates second-best. Lower values for CRPS and MASE indicate better performance.

	Gift-ZS Overall		Gift-ZS Short		Gift-ZS Medium		Gift-ZS Long	
Configuration	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**
Base Model (d=512, L=10)	**0.533**	**0.788**	**0.532**	0.727	**0.523**	**0.840**	**0.544**	**0.912**
d=512, L=10, Weaving Off	0.537	0.790	**0.532**	**0.723**	0.533	0.862	0.553	0.914
d=384, L=16 (Narrower, Deeper)	0.539	0.792	0.532	0.727	0.533	0.850	0.563	0.921
d=576, L=8 (Wider, Shallower)	0.540	0.794	0.536	0.732	0.529	0.849	0.561	0.921
seasonal_naive	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

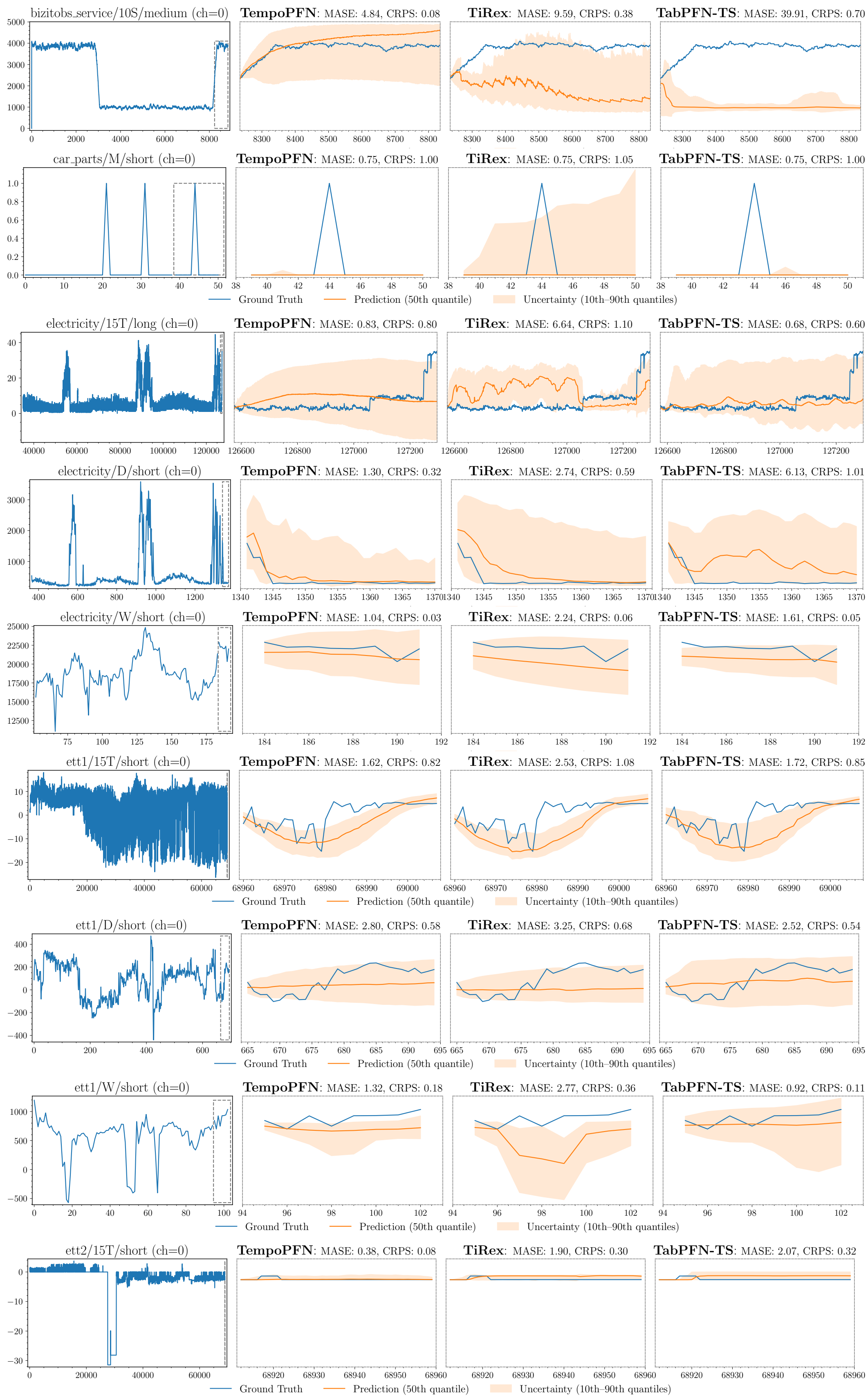
Table 8: Ablation study of learning rate schedulers. All models use the base architecture (d=512, L=10, H=4, convolution size 32, weaving enabled) and were trained for **2M iterations** with peak learning rate 2e-4. The **WarmupStableDecay** scheduler implements a three-phase approach with warmup (0.3% of steps), stable plateau (90% of steps), and cosine decay (9.7% of steps). **CosineWithRestarts** performs 4 periodic learning rate resets. **CosineWithWarmup** uses standard cosine annealing with warmup. **Cosine** applies basic cosine decay without warmup. Results are sorted by overall CRPS (best to worst). **Bold** indicates best performance, underline indicates second-best. Lower values for CRPS and MASE indicate better performance.

	Gift-ZS Overall		Gift-ZS Short		Gift-ZS Medium		Gift-ZS Long	
LR Scheduler	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**	**CRPS**	**MASE**
WarmupStableDecay	**0.554**	**0.812**	**0.544**	**0.740**	**0.550**	0.877	**0.584**	0.956
CosineWithWarmup	0.559	0.817	0.552	0.751	**0.550**	**0.873**	0.588	**0.955**
CosineWithRestarts	0.559	0.820	0.552	0.755	0.552	0.874	0.585	0.953
Cosine (no warmup)	0.561	0.820	0.553	0.751	0.556	0.880	0.590	0.962
seasonal_naive	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

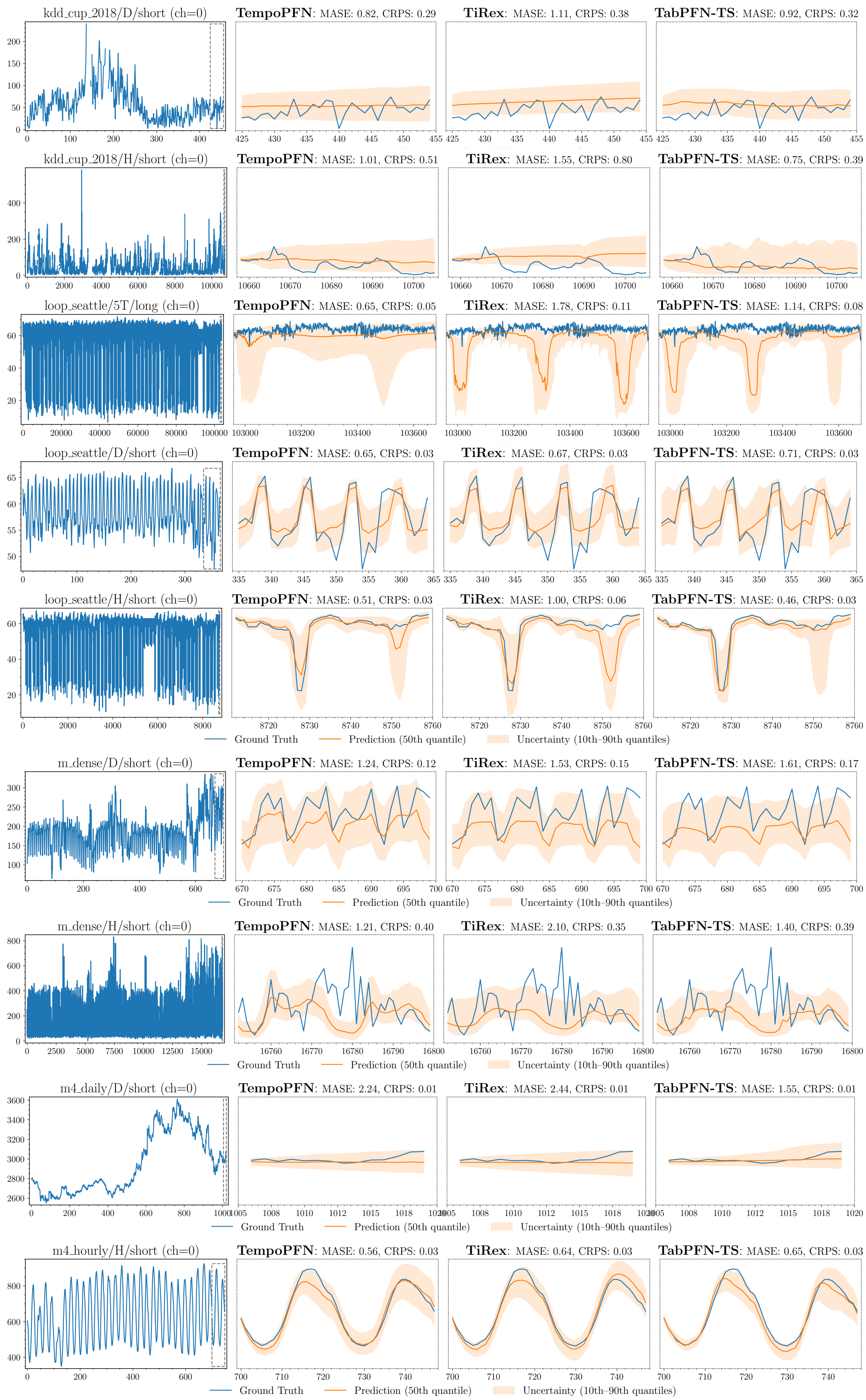
H Qualitative Results on the Gift-Eval Benchmark

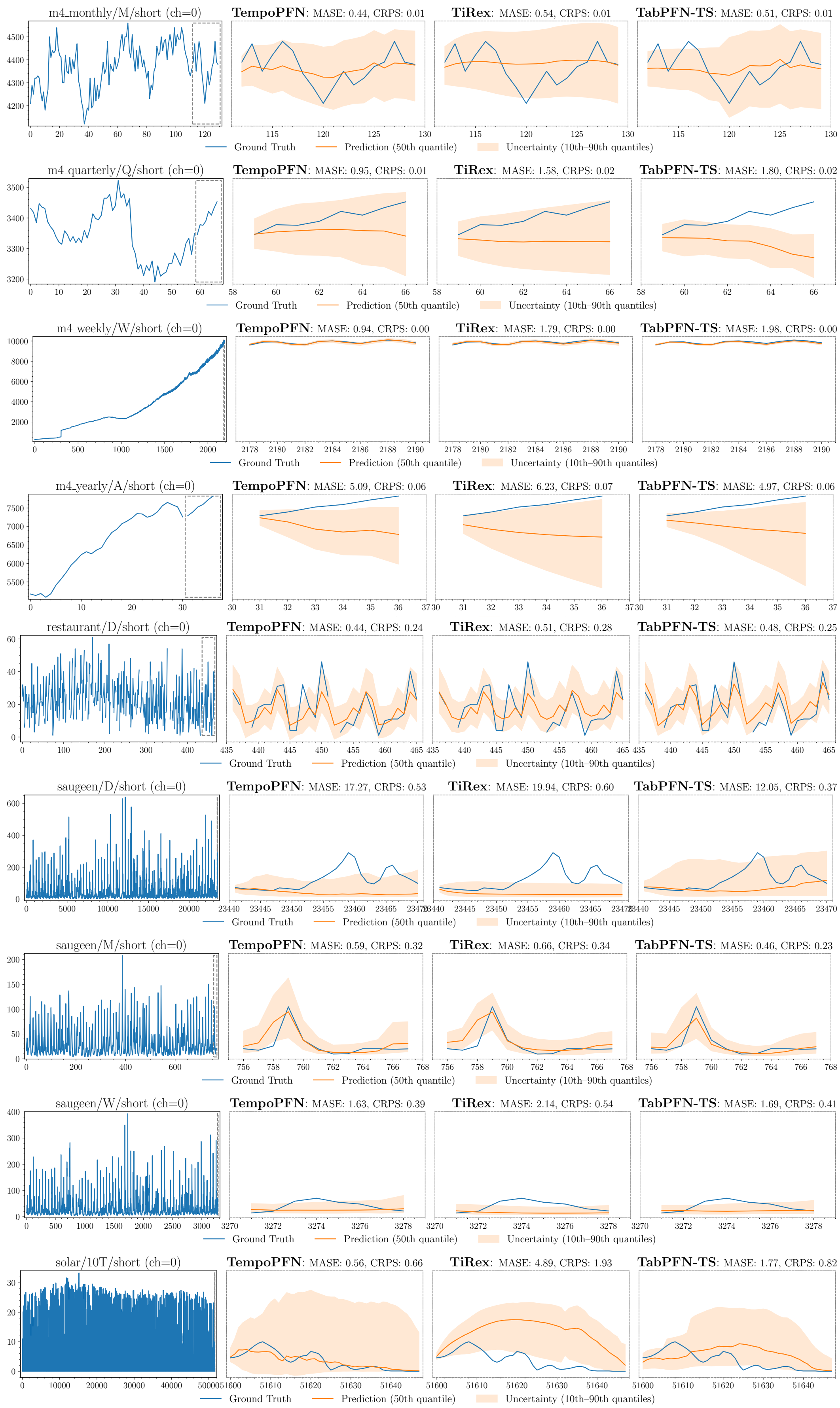
This section presents qualitative examples of forecasts from the Gift-Eval benchmark. Each example is illustrated through a horizontal arrangement of four plots. The first plot on the left displays the full historical context. The subsequent three plots provide a zoomed-in view of the forecast horizon, comparing the predictions from top-performing zero-shot models in the following order: TempoPFN, TiRex, and TabPFN-TS.

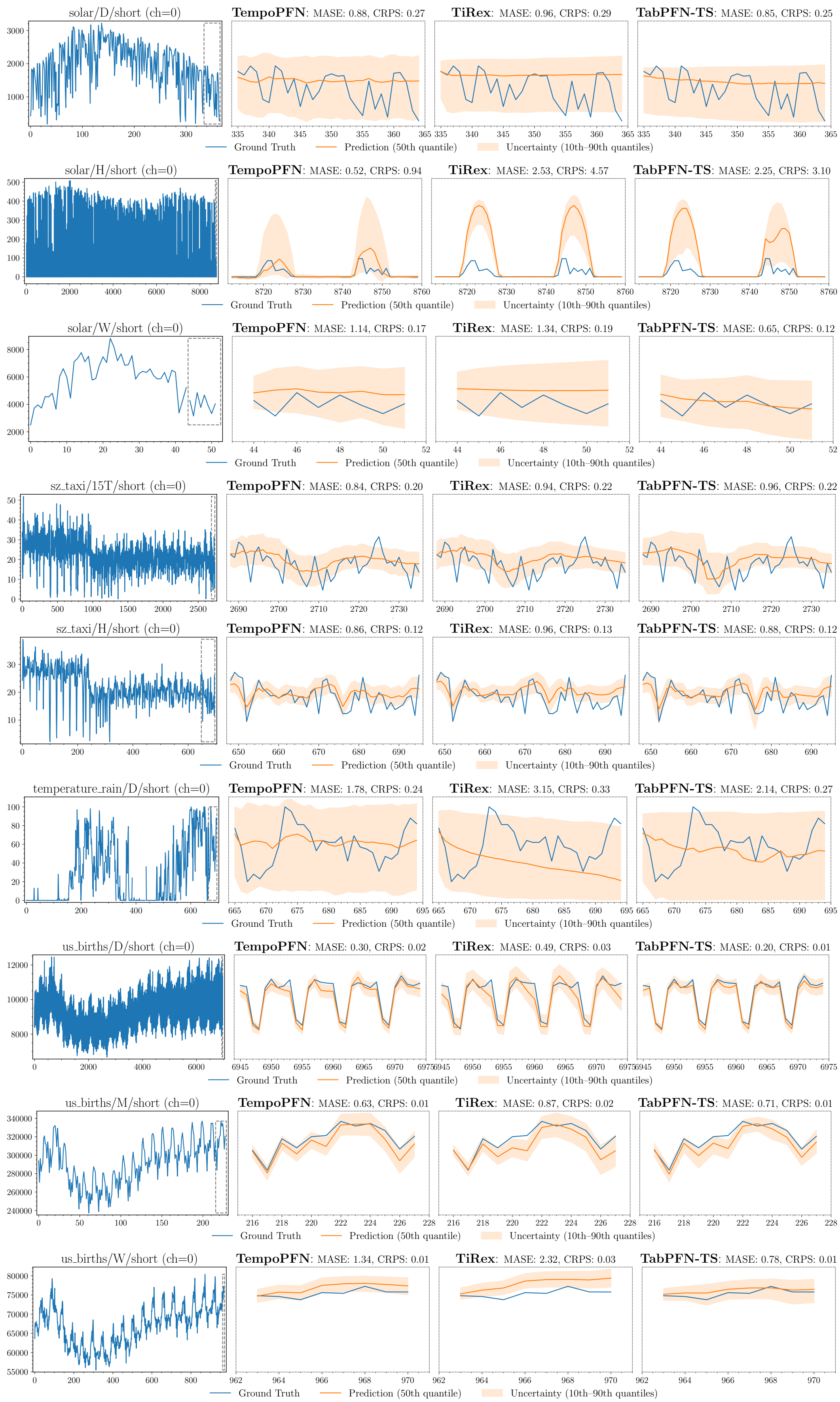
It is important to note the different context lengths used by each model during evaluation. The **TiRex** baseline was evaluated using the full available context for each series. In contrast, **TabPFN-TS** utilized a context of 4096 time steps, while our model, **TempoPFN**, operated with a maximum context length of **3072**.


bizitobs_service/10S/medium (ch=0)
TempoPFN: MASE: 4.84, CRPS: 0.08
TiRex: MASE: 9.59, CRPS: 0.38
TabPFN-TS: MASE: 39.91, CRPS: 0.70
car_parts/M/short (ch=0)
TempoPFN: MASE: 0.75, CRPS: 1.00
TiRex: MASE: 0.75, CRPS: 1.05
TabPFN-TS: MASE: 0.75, CRPS: 1.00
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
electricity/15T/long (ch=0)
TempoPFN: MASE: 0.83, CRPS: 0.80
TiRex: MASE: 6.64, CRPS: 1.10
TabPFN-TS: MASE: 0.68, CRPS: 0.60
electricity/D/short (ch=0)
TempoPFN: MASE: 1.30, CRPS: 0.32
TiRex: MASE: 2.74, CRPS: 0.59
TabPFN-TS: MASE: 6.13, CRPS: 1.01
electricity/W/short (ch=0)
TempoPFN: MASE: 1.04, CRPS: 0.03
TiRex: MASE: 2.24, CRPS: 0.06
TabPFN-TS: MASE: 1.61, CRPS: 0.05
ett1/15T/short (ch=0)
TempoPFN: MASE: 1.62, CRPS: 0.82
TiRex: MASE: 2.53, CRPS: 1.08
TabPFN-TS: MASE: 1.72, CRPS: 0.85
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
ett1/D/short (ch=0)
TempoPFN: MASE: 2.80, CRPS: 0.58
TiRex: MASE: 3.25, CRPS: 0.68
TabPFN-TS: MASE: 2.52, CRPS: 0.54
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
ett1/W/short (ch=0)
TempoPFN: MASE: 1.32, CRPS: 0.18
TiRex: MASE: 2.77, CRPS: 0.36
TabPFN-TS: MASE: 0.92, CRPS: 0.11
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
ett2/15T/short (ch=0)
TempoPFN: MASE: 0.38, CRPS: 0.08
TiRex: MASE: 1.90, CRPS: 0.30
TabPFN-TS: MASE: 2.07, CRPS: 0.32
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)


ett2/D/short (ch=0)
TempoPFN: MASE: 0.60, CRPS: 0.04
TiRex: MASE: 0.75, CRPS: 0.03
TabPFN-TS: MASE: 1.12, CRPS: 0.06
ett2/H/long (ch=0)
TempoPFN: MASE: 0.87, CRPS: 0.07
TiRex: MASE: 1.56, CRPS: 0.10
TabPFN-TS: MASE: 2.62, CRPS: 0.18
ett2/W/short (ch=0)
TempoPFN: MASE: 0.71, CRPS: 0.10
TiRex: MASE: 1.31, CRPS: 0.12
TabPFN-TS: MASE: 0.64, CRPS: 0.10
hierarchical_sales/D/short (ch=0)
TempoPFN: MASE: 2.79, CRPS: 0.75
TiRex: MASE: 2.99, CRPS: 0.76
TabPFN-TS: MASE: 3.34, CRPS: 0.91
hierarchical_sales/W/short (ch=0)
TempoPFN: MASE: 0.56, CRPS: 0.43
TiRex: MASE: 0.75, CRPS: 0.61
TabPFN-TS: MASE: 0.70, CRPS: 0.53
hospital/M/short (ch=0)
TempoPFN: MASE: 0.71, CRPS: 0.22
TiRex: MASE: 0.85, CRPS: 0.24
TabPFN-TS: MASE: 0.61, CRPS: 0.19
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
jena_weather/10T/short (ch=0)
TempoPFN: MASE: 3.78, CRPS: 1.00
TiRex: MASE: 6.00, CRPS: 0.79
TabPFN-TS: MASE: 4.61, CRPS: 0.96
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
jena_weather/D/short (ch=0)
TempoPFN: MASE: 2.77, CRPS: 0.01
TiRex: MASE: 3.03, CRPS: 0.01
TabPFN-TS: MASE: 3.83, CRPS: 0.01
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
jena_weather/H/short (ch=0)
TempoPFN: MASE: 1.31, CRPS: 0.00
TiRex: MASE: 2.48, CRPS: 0.01
TabPFN-TS: MASE: 3.42, CRPS: 0.01


kdd_cup_2018/D/short (ch=0)
TempoPFN: MASE: 0.82, CRPS: 0.29
TiRex: MASE: 1.11, CRPS: 0.38
TabPFN-TS: MASE: 0.92, CRPS: 0.32
kdd_cup_2018/H/short (ch=0)
TempoPFN: MASE: 1.01, CRPS: 0.51
TiRex: MASE: 1.55, CRPS: 0.80
TabPFN-TS: MASE: 0.75, CRPS: 0.39
loop_seattle/5T/long (ch=0)
TempoPFN: MASE: 0.65, CRPS: 0.05
TiRex: MASE: 1.78, CRPS: 0.11
TabPFN-TS: MASE: 1.14, CRPS: 0.08
loop_seattle/D/short (ch=0)
TempoPFN: MASE: 0.65, CRPS: 0.03
TiRex: MASE: 0.67, CRPS: 0.03
TabPFN-TS: MASE: 0.71, CRPS: 0.03
loop_seattle/H/short (ch=0)
TempoPFN: MASE: 0.51, CRPS: 0.03
TiRex: MASE: 1.00, CRPS: 0.06
TabPFN-TS: MASE: 0.46, CRPS: 0.03
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m_dense/D/short (ch=0)
TempoPFN: MASE: 1.24, CRPS: 0.12
TiRex: MASE: 1.53, CRPS: 0.15
TabPFN-TS: MASE: 1.61, CRPS: 0.17
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m_dense/H/short (ch=0)
TempoPFN: MASE: 1.21, CRPS: 0.40
TiRex: MASE: 2.10, CRPS: 0.35
TabPFN-TS: MASE: 1.40, CRPS: 0.39
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m4_daily/D/short (ch=0)
TempoPFN: MASE: 2.24, CRPS: 0.01
TiRex: MASE: 2.44, CRPS: 0.01
TabPFN-TS: MASE: 1.55, CRPS: 0.01
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m4_hourly/H/short (ch=0)
TempoPFN: MASE: 0.56, CRPS: 0.03
TiRex: MASE: 0.64, CRPS: 0.03
TabPFN-TS: MASE: 0.65, CRPS: 0.03
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)


m4_monthly/M/short (ch=0)
TempoPFN: MASE: 0.44, CRPS: 0.01
TiRex: MASE: 0.54, CRPS: 0.01
TabPFN-TS: MASE: 0.51, CRPS: 0.01
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m4_quarterly/Q/short (ch=0)
TempoPFN: MASE: 0.95, CRPS: 0.01
TiRex: MASE: 1.58, CRPS: 0.02
TabPFN-TS: MASE: 1.80, CRPS: 0.02
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m4_weekly/W/short (ch=0)
TempoPFN: MASE: 0.94, CRPS: 0.00
TiRex: MASE: 1.79, CRPS: 0.00
TabPFN-TS: MASE: 1.98, CRPS: 0.00
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
m4_yearly/A/short (ch=0)
TempoPFN: MASE: 5.09, CRPS: 0.06
TiRex: MASE: 6.23, CRPS: 0.07
TabPFN-TS: MASE: 4.97, CRPS: 0.06
restaurant/D/short (ch=0)
TempoPFN: MASE: 0.44, CRPS: 0.24
TiRex: MASE: 0.51, CRPS: 0.28
TabPFN-TS: MASE: 0.48, CRPS: 0.25
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
saugeen/D/short (ch=0)
TempoPFN: MASE: 17.27, CRPS: 0.53
TiRex: MASE: 19.94, CRPS: 0.60
TabPFN-TS: MASE: 12.05, CRPS: 0.37
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
saugeen/M/short (ch=0)
TempoPFN: MASE: 0.59, CRPS: 0.32
TiRex: MASE: 0.66, CRPS: 0.34
TabPFN-TS: MASE: 0.46, CRPS: 0.23
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
saugeen/W/short (ch=0)
TempoPFN: MASE: 1.63, CRPS: 0.39
TiRex: MASE: 2.14, CRPS: 0.54
TabPFN-TS: MASE: 1.69, CRPS: 0.41
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
solar/10T/short (ch=0)
TempoPFN: MASE: 0.56, CRPS: 0.66
TiRex: MASE: 4.89, CRPS: 1.93
TabPFN-TS: MASE: 1.77, CRPS: 0.82
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)


solar/D/short (ch=0)
TempoPFN: MASE: 0.88, CRPS: 0.27
TiRex: MASE: 0.96, CRPS: 0.29
TabPFN-TS: MASE: 0.85, CRPS: 0.25
Ground Truth
Prediction (50th quantile)
Uncertainty (10th–90th quantiles)
solar/H/short (ch=0)
TempoPFN: MASE: 0.52, CRPS: 0.94
TiRex: MASE: 2.53, CRPS: 4.57
TabPFN-TS: MASE: 2.25, CRPS: 3.10
solar/W/short (ch=0)
TempoPFN: MASE: 1.14, CRPS: 0.17
TiRex: MASE: 1.34, CRPS: 0.19
TabPFN-TS: MASE: 0.65, CRPS: 0.12
sz_taxi/15T/short (ch=0)
TempoPFN: MASE: 0.84, CRPS: 0.20
TiRex: MASE: 0.94, CRPS: 0.22
TabPFN-TS: MASE: 0.96, CRPS: 0.22
sz_taxi/H/short (ch=0)
TempoPFN: MASE: 0.86, CRPS: 0.12
TiRex: MASE: 0.96, CRPS: 0.13
TabPFN-TS: MASE: 0.88, CRPS: 0.12
temperature_rain/D/short (ch=0)
TempoPFN: MASE: 1.78, CRPS: 0.24
TiRex: MASE: 3.15, CRPS: 0.33
TabPFN-TS: MASE: 2.14, CRPS: 0.27
us_births/D/short (ch=0)
TempoPFN: MASE: 0.30, CRPS: 0.02
TiRex: MASE: 0.49, CRPS: 0.03
TabPFN-TS: MASE: 0.20, CRPS: 0.01
us_births/M/short (ch=0)
TempoPFN: MASE: 0.63, CRPS: 0.01
TiRex: MASE: 0.87, CRPS: 0.02
TabPFN-TS: MASE: 0.71, CRPS: 0.01
us_births/W/short (ch=0)
TempoPFN: MASE: 1.34, CRPS: 0.01
TiRex: MASE: 2.32, CRPS: 0.03
TabPFN-TS: MASE: 0.78, CRPS: 0.01

H.1 Quantitative Comparison on the Gift-Eval Benchmark

Table 9: Normalized CRPS (Continuous Ranked Probability Score) for various zero-shot models on the GiftEval benchmark. Scores are normalized against a seasonal naive baseline, with values less than 1.0 signifying superior probabilistic forecasts. The top two performing models are highlighted.

Dataset	TempoPFN	TiRex	FlowState-9.1M	Toto_Open_Base_1.0	TabPFN-TS	YingLong_50m	Chronos Bolt B	TTM-R2-Finetuned	Moirai L 1.1	Moirai B 1.1
bitbrains_fast_storage/5T/long	**0.555**	0.558	0.701	0.568	0.752	0.586	0.635	0.796	0.608	0.622
bitbrains_fast_storage/5T/medium	**0.516**	**0.516**	0.651	0.525	0.792	0.537	0.631	0.690	0.531	0.553
bitbrains_fast_storage/5T/short	0.334	0.334	0.422	**0.306**	0.547	0.343	0.375	0.465	0.340	0.341
bitbrains_fast_storage/H/short	0.632	0.683	0.684	0.609	0.655	0.610	0.757	1.007	0.632	**0.600**
bitbrains_rnd/5T/long	0.602	0.541	0.619	**0.501**	0.697	0.563	0.643	0.648	0.577	0.566
bitbrains_rnd/5T/medium	0.530	**0.500**	0.619	0.537	0.701	0.546	0.517	0.670	0.508	0.527
bitbrains_rnd/5T/short	0.477	0.367	0.420	**0.362**	0.552	0.390	0.398	0.424	0.379	0.405
bitbrains_rnd/H/short	0.523	0.522	0.535	0.477	0.597	0.546	0.502	0.718	**0.455**	0.467
bizitobs_application/10S/long	**1.041**	1.142	1.152	1.148	1.062	1.315	2.383	1.261	2.053	2.624
bizitobs_application/10S/medium	**0.560**	0.944	0.985	0.802	0.955	1.147	2.425	1.138	1.958	2.436
bizitobs_application/10S/short	**0.293**	0.377	0.389	0.333	0.426	0.506	1.549	0.638	1.099	0.941
bizitobs_l2c/5T/long	0.885	0.917	**0.355**	0.821	0.472	0.848	1.138	0.373	0.783	0.854
bizitobs_l2c/5T/medium	0.681	0.684	**0.411**	0.606	0.502	0.690	0.856	0.475	0.788	0.730
bizitobs_l2c/5T/short	0.276	0.303	0.323	0.265	0.321	0.298	0.284	**0.263**	0.303	0.297
bizitobs_l2c/H/long	0.333	0.297	**0.263**	0.392	0.311	0.401	0.295	0.736	0.638	0.526
bizitobs_l2c/H/medium	0.301	0.284	**0.245**	0.394	0.263	0.396	0.281	0.746	0.685	0.761
bizitobs_l2c/H/short	0.381	0.413	**0.362**	0.382	0.402	0.516	0.363	0.516	1.073	0.946
bizitobs_service/10S/long	1.043	1.001	1.001	**0.955**	0.965	1.220	2.111	1.095	1.946	2.151
bizitobs_service/10S/medium	**0.393**	0.639	0.562	0.574	0.860	0.981	2.022	0.896	1.455	1.892
bizitobs_service/10S/short	0.289	0.327	0.318	**0.286**	0.482	0.444	1.267	0.354	0.788	1.053
car_parts/M/short	0.590	0.576	0.585	**0.522**	0.563	0.817	0.578	0.641	0.685	0.580
covid_deaths/D/short	0.234	0.273	0.339	**0.215**	0.324	0.468	0.371	0.287	0.362	0.346
electricity/15T/long	1.010	**0.656**	0.666	0.761	0.716	0.744	0.748	0.690	0.878	1.022
electricity/15T/medium	0.907	**0.655**	0.671	0.762	0.734	0.754	0.734	0.683	0.913	0.940
electricity/15T/short	0.682	0.503	0.562	0.603	0.587	0.565	**0.496**	0.497	0.776	0.728
electricity/D/short	0.639	**0.525**	0.529	0.566	0.603	0.560	0.531	0.560	0.666	0.583
electricity/H/long	0.652	0.607	0.571	**0.544**	0.706	0.682	0.637	0.587	0.671	0.557
electricity/H/medium	0.685	0.617	0.603	**0.589**	0.688	0.677	0.636	0.615	0.685	0.645
electricity/H/short	0.730	**0.571**	0.632	0.652	0.677	0.744	0.602	0.643	0.732	0.711
electricity/W/short	0.611	0.459	**0.439**	0.641	0.550	0.636	0.477	0.605	0.625	0.775
ett1/15T/long	0.847	0.721	**0.676**	0.738	0.763	0.712	0.875	0.750	1.052	0.803
ett1/15T/medium	0.864	0.777	**0.713**	0.810	0.785	0.763	0.874	0.847	1.064	1.008
ett1/15T/short	0.753	0.662	0.670	0.670	0.691	0.704	**0.654**	0.724	0.936	0.800
ett1/D/short	**0.658**	0.678	0.685	0.695	0.729	0.664	0.703	0.838	0.700	0.737
ett1/H/long	0.598	0.551	0.573	0.566	0.626	**0.550**	0.660	0.597	0.628	0.609
ett1/H/medium	0.634	0.578	**0.569**	0.584	0.651	0.578	0.696	0.628	0.621	0.648
ett1/H/short	0.775	0.747	**0.733**	0.806	0.807	0.746	0.753	0.814	0.786	0.820
ett1/W/short	0.827	1.000	**0.786**	0.843	0.911	0.931	0.949	0.906	0.834	0.837
ett2/15T/long	0.768	0.721	0.680	**0.665**	0.761	0.702	0.839	0.718	0.866	1.031
ett2/15T/medium	0.776	0.744	**0.703**	0.748	0.808	0.741	0.889	0.776	0.846	0.878
ett2/15T/short	0.711	0.685	0.683	0.706	0.756	0.696	0.691	**0.662**	0.833	0.805
ett2/D/short	0.783	0.594	0.674	0.727	0.821	0.613	0.610	**0.553**	0.611	0.618
ett2/H/long	**0.498**	0.549	0.536	0.521	0.669	0.514	0.562	0.504	0.601	0.529
ett2/H/medium	0.567	0.568	0.560	0.547	0.648	0.548	0.616	0.560	0.634	**0.537**
ett2/H/short	0.730	0.744	**0.710**	0.725	0.821	0.730	0.713	0.742	0.775	0.807
ett2/W/short	0.687	**0.645**	0.779	0.792	0.738	0.817	0.660	0.702	0.815	0.652
hierarchical_sales/D/short	0.336	**0.327**	0.336	0.328	0.341	0.340	0.332	0.349	0.334	0.331
hierarchical_sales/W/short	0.412	0.416	**0.408**	0.428	0.415	0.467	0.424	0.435	0.431	0.429
hospital/M/short	0.840	0.830	**0.766**	0.838	0.860	0.926	0.913	0.841	0.821	0.810
jena_weather/10T/long	0.240	0.220	0.214	**0.210**	0.224	0.254	0.269	0.286	0.325	0.297
jena_weather/10T/medium	0.248	0.239	0.234	**0.231**	0.254	0.272	0.270	0.324	0.338	0.320
jena_weather/10T/short	0.204	0.200	0.185	**0.172**	0.219	0.238	0.210	0.285	0.331	0.345
jena_weather/D/short	0.239	0.218	0.226	0.240	0.224	0.233	**0.214**	0.343	0.243	0.236
jena_weather/H/long	0.147	**0.134**	0.159	0.136	0.245	0.145	0.147	0.449	0.145	0.156
jena_weather/H/medium	0.164	**0.153**	0.154	0.154	0.169	0.165	0.158	0.204	0.168	0.166
jena_weather/H/short	0.273	**0.266**	0.272	0.274	0.273	0.284	0.274	0.403	0.291	0.284
kdd_cup_2018/D/short	0.561	0.577	0.565	0.574	**0.537**	0.551	0.552	0.593	0.565	0.557
kdd_cup_2018/H/long	0.476	0.353	0.489	0.488	0.510	0.475	**0.320**	0.507	0.404	0.446
kdd_cup_2018/H/medium	0.563	0.437	0.576	0.582	0.593	0.553	**0.397**	0.597	0.510	0.581
kdd_cup_2018/H/short	0.701	0.498	0.697	0.736	0.763	0.694	**0.450**	0.758	0.661	0.710
loop_seattle/5T/long	0.867	0.680	0.582	0.602	0.710	0.790	1.015	0.659	**0.387**	0.407
loop_seattle/5T/medium	0.910	0.692	0.589	0.617	0.741	0.848	0.989	0.661	**0.328**	0.386
loop_seattle/5T/short	0.701	0.606	0.618	0.598	0.650	0.710	0.675	0.634	**0.512**	0.565
loop_seattle/D/short	0.417	0.409	**0.407**	0.430	0.418	0.425	0.424	0.441	0.438	0.428
loop_seattle/H/long	0.405	**0.328**	0.338	0.345	0.334	0.364	0.406	0.360	0.397	0.425
loop_seattle/H/medium	0.451	0.400	0.420	**0.398**	0.412	0.430	0.470	0.429	0.433	0.492
loop_seattle/H/short	0.669	**0.569**	0.583	0.609	0.608	0.609	0.624	0.642	0.634	0.709
m4_daily/D/short	1.048	**0.827**	0.998	0.898	0.928	0.975	0.865	0.970	1.244	1.646
m4_hourly/H/short	0.759	0.536	0.545	0.921	0.788	0.641	0.674	0.919	**0.527**	0.591
m4_monthly/M/short	**0.749**	0.756	0.759	0.795	0.768	0.860	0.769	0.821	0.780	0.769
m4_quarterly/Q/short	0.763	0.753	0.770	0.791	0.798	0.892	0.786	0.809	**0.749**	**0.749**
m4_weekly/W/short	0.689	**0.589**	0.602	0.806	0.608	0.681	0.627	0.722	0.764	0.792
m4_yearly/A/short	0.872	0.857	0.780	0.887	0.858	1.096	0.886	0.864	**0.758**	0.766
m_dense/D/short	0.291	0.292	0.308	0.328	**0.269**	0.354	0.304	0.311	0.420	0.458
m_dense/H/long	0.438	0.291	0.289	0.307	0.394	0.417	0.405	0.306	**0.272**	0.291
m_dense/H/medium	0.459	0.318	0.304	0.321	0.424	0.415	0.417	0.322	**0.297**	0.326
m_dense/H/short	0.627	0.472	0.478	0.540	0.564	0.573	**0.455**	0.514	0.466	0.509
restaurant/D/short	0.381	**0.377**	0.378	0.439	0.388	0.401	0.390	0.397	0.399	0.393
saugeen/D/short	0.642	0.669	**0.562**	0.604	0.638	0.627	0.578	0.694	0.694	0.605
saugeen/M/short	0.697	0.717	0.664	0.671	**0.621**	0.707	0.665	0.764	0.728	0.782
saugeen/W/short	0.546	**0.477**	0.486	0.531	0.539	0.512	0.494	0.606	0.586	0.576
solar/10T/long	0.748	**0.480**	0.501	0.523	0.491	0.830	0.657	0.722	1.144	1.340
solar/10T/medium	0.664	0.549	0.544	0.539	**0.498**	0.819	0.666	0.759	1.140	1.270
solar/10T/short	0.538	0.640	0.605	0.629	**0.533**	0.663	0.595	0.633	0.693	0.714
solar/D/short	0.501	0.506	0.497	0.518	**0.481**	0.516	0.514	0.541	0.522	0.528
solar/H/long	0.339	**0.237**	0.309	0.307	0.325	0.326	0.376	0.563	0.322	0.334
solar/H/medium	0.409	**0.281**	0.365	0.350	0.331	0.380	0.390	0.591	0.366	0.350
solar/H/short	0.573	**0.457**	0.563	0.555	0.567	0.576	0.504	0.691	0.563	0.571
solar/W/short	0.763	0.779	0.583	0.887	**0.572**	1.452	0.632	0.817	1.016	1.120
sz_taxi/15T/long	0.497	**0.460**	0.483	0.472	0.567	0.477	0.581	0.558	0.498	0.488
sz_taxi/15T/medium	0.558	**0.533**	0.548	0.542	0.608	0.549	0.644	0.567	0.567	0.556
sz_taxi/15T/short	0.669	**0.649**	0.656	0.657	0.676	0.658	0.654	0.682	0.696	0.690
sz_taxi/H/short	0.650	**0.634**	0.647	0.642	0.656	0.640	0.636	0.662	0.683	0.669
temperature_rain/D/short	0.451	0.434	0.434	0.441	0.449	0.462	0.424	0.518	**0.378**	0.422
us_births/D/short	0.169	0.179	0.198	0.217	**0.133**	0.258	0.217	0.171	0.224	0.223
us_births/M/short	0.814	0.894	0.887	0.754	0.909	0.766	1.156	**0.680**	0.939	0.886
us_births/W/short	0.553	0.654	0.616	0.750	**0.550**	0.743	0.668	0.739	0.922	0.897

Table 10: Normalized MASE scores of different zero-shot models on the GiftEval benchmark. Scores are relative to a seasonal naive baseline, where values below 1.0 indicate better performance. The models achieving the best and second-best scores are highlighted.)

Dataset	TempoPFN	TiRex	FlowState-9.1M	Toto_Open_Base_1.0	TTM-R2-Finetuned	TabPFN-TS	Chronos Bolt B	YingLong_50m	Moirai L 1.1	Moirai B 1.1
bitbrains_fast_storage/5T/long	0.846	0.808	0.913	**0.789**	0.827	1.014	0.834	0.886	0.840	0.851
bitbrains_fast_storage/5T/medium	0.896	0.815	1.033	**0.807**	0.876	1.072	0.871	0.889	0.836	0.860
bitbrains_fast_storage/5T/short	0.716	0.609	0.883	**0.591**	0.648	0.879	0.662	0.719	0.728	0.697
bitbrains_fast_storage/H/short	0.910	0.825	0.851	**0.728**	0.927	0.912	0.824	0.856	0.839	0.909
bitbrains_rnd/5T/long	1.005	0.954	1.007	**0.953**	1.001	1.107	0.970	1.002	0.977	0.985
bitbrains_rnd/5T/medium	1.001	**0.967**	1.009	0.972	0.997	1.064	0.979	1.004	0.982	0.997
bitbrains_rnd/5T/short	0.940	0.845	0.996	**0.837**	0.889	1.030	0.865	0.915	0.888	0.923
bitbrains_rnd/H/short	0.999	0.971	0.982	**0.934**	1.004	1.106	0.977	0.972	0.982	1.005
bizitobs_application/10S/long	1.086	1.150	1.042	1.021	1.187	**0.965**	3.270	1.518	2.445	4.210
bizitobs_application/10S/medium	**0.800**	1.035	0.962	0.856	1.134	0.925	3.612	1.516	2.746	4.756
bizitobs_application/10S/short	**0.466**	0.571	0.597	0.556	0.734	0.563	2.468	0.841	2.011	2.373
bizitobs_l2c/5T/long	0.840	0.819	0.356	0.809	**0.347**	0.457	0.853	0.816	0.770	0.770
bizitobs_l2c/5T/medium	0.629	0.668	**0.418**	0.606	0.441	0.513	0.706	0.677	0.794	0.686
bizitobs_l2c/5T/short	0.278	0.299	0.317	0.262	**0.250**	0.311	0.282	0.292	0.289	0.295
bizitobs_l2c/H/long	0.478	0.425	**0.384**	0.559	0.844	0.466	0.390	0.554	0.891	0.764
bizitobs_l2c/H/medium	0.403	0.358	**0.312**	0.501	0.753	0.324	0.328	0.512	0.828	0.874
bizitobs_l2c/H/short	0.381	0.420	0.371	0.387	0.448	0.400	**0.356**	0.532	0.947	0.823
bizitobs_service/10S/long	1.216	1.104	1.011	**0.953**	1.086	0.999	3.875	1.715	3.167	4.447
bizitobs_service/10S/medium	1.076	0.946	0.866	**0.820**	0.953	0.928	3.768	1.605	2.931	4.536
bizitobs_service/10S/short	0.947	0.677	0.658	**0.644**	0.653	0.721	2.706	0.953	1.885	2.799
car_parts/M/short	0.700	0.698	0.744	**0.675**	0.698	0.706	0.712	1.057	0.752	0.695
covid_deaths/D/short	0.785	0.830	0.738	0.695	**0.657**	0.837	0.828	0.929	0.778	0.738
electricity/15T/long	1.170	0.759	0.778	0.897	**0.747**	0.812	0.801	0.863	1.126	1.134
electricity/15T/medium	1.027	0.726	0.744	0.858	**0.714**	0.773	0.749	0.825	1.121	1.156
electricity/15T/short	0.751	0.557	0.630	0.667	**0.530**	0.670	0.545	0.628	0.996	0.897
electricity/D/short	0.787	**0.716**	0.722	0.747	0.730	0.751	0.729	0.744	0.760	0.755
electricity/H/long	0.896	0.802	0.795	0.811	**0.781**	0.861	0.812	0.869	0.892	0.827
electricity/H/medium	0.871	0.780	0.785	0.792	0.782	0.838	**0.774**	0.831	0.862	0.855
electricity/H/short	0.794	**0.641**	0.689	0.719	0.663	0.763	0.643	0.807	0.795	0.803
electricity/W/short	0.751	0.691	**0.653**	0.857	0.729	0.740	0.707	0.832	0.857	0.919
ett1/15T/long	1.050	0.871	**0.850**	0.898	0.870	0.939	0.954	0.869	1.176	0.940
ett1/15T/medium	0.995	0.868	**0.833**	0.909	0.902	0.919	0.893	0.865	1.094	1.044
ett1/15T/short	0.830	0.745	0.737	0.742	0.751	0.794	**0.728**	0.769	0.990	0.883
ett1/D/short	0.918	0.952	**0.916**	0.933	1.056	0.922	0.940	0.953	0.984	0.978
ett1/H/long	0.937	**0.886**	0.906	0.926	0.941	0.996	0.916	0.909	0.981	0.933
ett1/H/medium	0.860	0.785	**0.782**	0.811	0.819	0.896	0.877	0.808	0.855	0.861
ett1/H/short	0.875	0.849	**0.834**	0.885	0.869	0.908	0.847	0.850	0.875	0.906
ett1/W/short	0.851	1.003	**0.820**	0.893	0.902	0.938	0.959	0.980	0.854	0.871
ett2/15T/long	1.014	0.907	**0.855**	0.860	0.918	0.965	0.928	0.908	1.126	1.284
ett2/15T/medium	0.923	0.859	**0.806**	0.878	0.872	0.933	0.877	0.865	1.008	1.046
ett2/15T/short	0.731	0.695	0.707	0.736	**0.681**	0.788	0.718	0.724	0.937	0.899
ett2/D/short	1.648	0.917	1.053	1.161	**0.868**	1.029	0.951	0.943	1.036	0.942
ett2/H/long	**0.879**	1.008	0.972	0.955	0.919	1.280	0.918	0.940	1.134	0.993
ett2/H/medium	0.843	0.846	0.838	0.821	0.831	1.008	0.830	**0.816**	0.953	0.832
ett2/H/short	0.809	0.809	**0.782**	0.796	0.811	0.894	0.794	0.815	0.848	0.874
ett2/W/short	1.150	1.021	1.256	1.267	1.009	0.983	**0.949**	1.221	1.683	1.093
hierarchical_sales/D/short	0.662	0.653	0.660	**0.648**	0.666	0.669	0.655	0.677	0.657	0.657
hierarchical_sales/W/short	0.711	0.704	**0.695**	0.726	0.709	0.713	0.715	0.757	0.731	0.729
hospital/M/short	0.834	0.829	0.824	0.851	**0.815**	0.830	0.860	0.899	0.834	0.842
jena_weather/10T/long	0.851	0.836	0.868	**0.833**	1.008	0.876	0.862	0.843	1.040	1.001
jena_weather/10T/medium	0.867	0.842	0.873	**0.835**	1.001	0.874	0.852	0.888	0.969	0.994
jena_weather/10T/short	0.401	0.402	0.373	**0.358**	0.431	0.418	0.411	0.452	0.455	0.471
jena_weather/D/short	0.811	**0.648**	0.666	0.760	0.840	0.781	0.668	0.687	0.725	0.731
jena_weather/H/long	0.907	0.778	0.856	0.800	1.153	1.109	0.811	0.832	**0.695**	0.836
jena_weather/H/medium	0.934	0.949	0.965	0.847	0.969	1.225	**0.841**	0.991	1.003	0.919
jena_weather/H/short	0.739	**0.716**	0.728	0.752	0.802	0.759	0.741	0.740	0.809	0.766
kdd_cup_2018/D/short	0.803	0.820	0.817	0.810	0.800	0.784	0.799	**0.782**	0.802	0.802
kdd_cup_2018/H/long	0.776	0.554	0.789	0.779	0.752	0.819	**0.512**	0.756	0.649	0.719
kdd_cup_2018/H/medium	0.753	0.561	0.759	0.754	0.721	0.790	**0.490**	0.721	0.668	0.735
kdd_cup_2018/H/short	0.727	0.490	0.712	0.739	0.709	0.784	**0.448**	0.697	0.667	0.704
loop_seattle/5T/long	0.952	0.783	0.660	0.678	0.697	0.805	0.990	0.882	**0.445**	0.473
loop_seattle/5T/medium	1.004	0.796	0.671	0.697	0.701	0.836	0.985	0.949	**0.390**	0.453
loop_seattle/5T/short	0.856	0.745	0.756	0.737	0.732	0.784	0.823	0.864	**0.638**	0.703
loop_seattle/D/short	0.518	**0.505**	**0.505**	0.534	0.519	0.524	0.521	0.541	0.529	0.521
loop_seattle/H/long	0.727	**0.583**	0.595	0.610	0.585	0.599	0.644	0.636	0.679	0.744
loop_seattle/H/medium	0.721	0.634	0.662	0.628	**0.623**	0.661	0.688	0.684	0.675	0.770
loop_seattle/H/short	0.775	**0.657**	0.664	0.696	0.678	0.705	0.696	0.693	0.731	0.820
m4_daily/D/short	1.352	**0.942**	1.096	1.010	1.007	1.279	0.976	1.131	1.275	1.638
m4_hourly/H/short	0.695	**0.589**	0.613	0.721	0.862	0.619	0.701	0.808	0.743	0.814
m4_monthly/M/short	0.733	**0.732**	**0.732**	0.780	0.751	0.760	0.753	0.836	0.776	0.757
m4_quarterly/Q/short	0.733	0.725	0.720	0.766	0.730	0.764	0.764	0.875	**0.712**	**0.712**
m4_weekly/W/short	0.903	**0.679**	0.728	0.864	0.700	0.738	0.748	0.816	0.929	1.012
m4_yearly/A/short	0.863	0.864	**0.746**	0.857	0.819	0.834	0.884	1.098	0.749	0.759
m_dense/D/short	0.413	0.414	0.439	0.457	0.438	**0.406**	0.429	0.498	0.573	0.659
m_dense/H/long	0.751	0.491	0.504	0.528	0.497	0.692	0.635	0.666	**0.471**	0.497
m_dense/H/medium	0.660	0.464	0.441	0.464	0.452	0.633	0.561	0.567	**0.436**	0.468
m_dense/H/short	0.683	0.528	0.534	0.591	0.544	0.609	**0.521**	0.630	0.522	0.563
restaurant/D/short	0.684	**0.674**	0.678	0.779	0.695	0.694	0.696	0.713	0.711	0.700
saugeen/D/short	0.917	0.933	**0.802**	0.869	0.878	0.922	0.832	0.888	0.964	0.853
saugeen/M/short	0.809	0.790	0.766	0.775	0.773	**0.720**	0.757	0.763	0.774	0.854
saugeen/W/short	0.678	0.588	**0.586**	0.657	0.690	0.662	0.611	0.623	0.693	0.708
solar/10T/long	1.481	**0.928**	0.977	1.011	1.144	1.000	1.229	1.564	2.239	2.319
solar/10T/medium	1.196	0.969	0.970	0.951	1.075	**0.906**	1.108	1.324	1.963	2.039
solar/10T/short	0.823	0.982	0.928	0.934	**0.745**	0.854	0.896	1.015	1.004	0.995
solar/D/short	0.846	**0.841**	0.858	0.875	0.851	0.854	0.849	0.856	0.854	0.882
solar/H/long	0.929	**0.688**	0.879	0.894	1.311	1.011	0.964	0.911	0.952	0.999
solar/H/medium	1.061	**0.792**	0.937	0.926	1.323	0.924	0.996	0.993	0.981	0.954
solar/H/short	0.962	**0.759**	0.916	0.870	0.926	0.943	0.854	0.922	0.919	0.938
solar/W/short	0.804	0.830	0.570	0.970	0.678	**0.539**	0.666	1.415	1.041	1.129
sz_taxi/15T/long	0.787	**0.736**	0.766	0.749	0.851	0.810	0.789	0.754	0.802	0.777
sz_taxi/15T/medium	0.789	**0.753**	0.772	0.764	0.764	0.793	0.784	0.774	0.797	0.782
sz_taxi/15T/short	0.733	**0.712**	0.721	0.719	0.733	0.730	0.717	0.721	0.760	0.754
sz_taxi/H/short	0.781	0.763	0.780	0.769	0.780	0.776	**0.762**	0.768	0.814	0.797
temperature_rain/D/short	0.689	0.666	0.671	0.679	0.702	0.685	0.648	0.725	**0.597**	0.651
us_births/D/short	0.211	0.219	0.245	0.267	0.208	**0.170**	0.260	0.324	0.270	0.273
us_births/M/short	0.875	0.941	0.920	0.764	**0.673**	0.941	1.215	0.799	1.014	0.951
us_births/W/short	0.583	0.687	0.651	0.790	0.729	**0.572**	0.696	0.789	0.940	0.921

Table 11: CRPS performance summarized by average rank for zero-shot models on the GiftEval benchmark. A lower rank signifies superior probabilistic forecasting performance. The models achieving the first and second-best overall average ranks are highlighted.

Dataset	TempoPFN	TiRex	FlowState-9.1M	Toto_Open_Base_1.0	Chronos Bolt B	TabPFN-TS	YingLong_50m	TTM-R2-Finetuned	Moirai L 1.1	Moirai B 1.1
bitbrains_fast_storage/5T/long	**1.000**	2.000	8.000	3.000	7.000	9.000	4.000	10.000	5.000	6.000
bitbrains_fast_storage/5T/medium	**1.000**	2.000	8.000	3.000	7.000	10.000	5.000	9.000	4.000	6.000
bitbrains_fast_storage/5T/short	2.000	3.000	8.000	**1.000**	7.000	10.000	6.000	9.000	4.000	5.000
bitbrains_fast_storage/H/short	4.000	7.000	8.000	2.000	9.000	6.000	3.000	11.000	5.000	**1.000**
bitbrains_rnd/5T/long	6.000	2.000	7.000	**1.000**	8.000	10.000	3.000	9.000	5.000	4.000
bitbrains_rnd/5T/medium	5.000	**1.000**	8.000	6.000	3.000	10.000	7.000	9.000	2.000	4.000
bitbrains_rnd/5T/short	9.000	2.000	7.000	**1.000**	5.000	10.000	4.000	8.000	3.000	6.000
bitbrains_rnd/H/short	6.000	5.000	7.000	3.000	4.000	9.000	8.000	10.000	**1.000**	2.000
bizitobs_application/10S/long	**2.000**	4.000	6.000	5.000	10.000	3.000	8.000	7.000	9.000	11.000
bizitobs_application/10S/medium	**1.000**	3.000	5.000	2.000	10.000	4.000	8.000	7.000	9.000	11.000
bizitobs_application/10S/short	**1.000**	3.000	4.000	2.000	11.000	5.000	6.000	7.000	10.000	8.000
bizitobs_l2c/5T/long	8.000	9.000	**1.000**	5.000	11.000	3.000	6.000	2.000	4.000	7.000
bizitobs_l2c/5T/medium	5.000	6.000	**1.000**	4.000	10.000	3.000	7.000	2.000	9.000	8.000
bizitobs_l2c/5T/short	3.000	7.000	10.000	2.000	4.000	9.000	6.000	**1.000**	8.000	5.000
bizitobs_l2c/H/long	5.000	3.000	**1.000**	6.000	2.000	4.000	7.000	10.000	9.000	8.000
bizitobs_l2c/H/medium	5.000	4.000	**1.000**	6.000	3.000	2.000	7.000	9.000	8.000	10.000
bizitobs_l2c/H/short	3.000	6.000	**1.000**	4.000	2.000	5.000	8.000	7.000	11.000	9.000
bizitobs_service/10S/long	6.000	4.000	5.000	**1.000**	10.000	2.000	8.000	7.000	9.000	11.000
bizitobs_service/10S/medium	**1.000**	4.000	2.000	3.000	11.000	5.000	7.000	6.000	9.000	10.000
bizitobs_service/10S/short	2.000	4.000	3.000	**1.000**	11.000	7.000	6.000	5.000	8.000	10.000
car_parts/M/short	7.000	3.000	6.000	**1.000**	4.000	2.000	10.000	8.000	9.000	5.000
covid_deaths/D/short	2.000	3.000	6.000	**1.000**	9.000	5.000	10.000	4.000	8.000	7.000
electricity/15T/long	10.000	**1.000**	2.000	7.000	6.000	4.000	5.000	3.000	8.000	11.000
electricity/15T/medium	8.000	**1.000**	2.000	7.000	5.000	4.000	6.000	3.000	9.000	10.000
electricity/15T/short	8.000	3.000	4.000	7.000	**1.000**	6.000	5.000	2.000	10.000	9.000
electricity/D/short	9.000	**1.000**	2.000	6.000	3.000	8.000	5.000	4.000	10.000	7.000
electricity/H/long	7.000	5.000	3.000	**1.000**	6.000	10.000	9.000	4.000	8.000	2.000
electricity/H/medium	9.000	4.000	2.000	**1.000**	5.000	10.000	7.000	3.000	8.000	6.000
electricity/H/short	8.000	**1.000**	3.000	5.000	2.000	6.000	10.000	4.000	9.000	7.000
electricity/W/short	6.000	2.000	**1.000**	9.000	3.000	4.000	8.000	5.000	7.000	10.000
ett1/15T/long	8.000	3.000	**1.000**	4.000	9.000	6.000	2.000	5.000	11.000	7.000
ett1/15T/medium	7.000	3.000	**1.000**	5.000	8.000	4.000	2.000	6.000	11.000	10.000
ett1/15T/short	8.000	2.000	4.000	3.000	**1.000**	5.000	6.000	7.000	10.000	9.000
ett1/D/short	**1.000**	3.000	4.000	5.000	7.000	8.000	2.000	10.000	6.000	9.000
ett1/H/long	6.000	2.000	4.000	3.000	10.000	8.000	**1.000**	5.000	9.000	7.000
ett1/H/medium	7.000	3.000	**1.000**	4.000	10.000	9.000	2.000	6.000	5.000	8.000
ett1/H/short	5.000	3.000	**1.000**	7.000	4.000	8.000	2.000	9.000	6.000	10.000
ett1/W/short	2.000	11.000	**1.000**	5.000	9.000	7.000	8.000	6.000	3.000	4.000
ett2/15T/long	7.000	5.000	2.000	**1.000**	8.000	6.000	3.000	4.000	9.000	11.000
ett2/15T/medium	6.000	3.000	**1.000**	4.000	10.000	7.000	2.000	5.000	8.000	9.000
ett2/15T/short	7.000	3.000	2.000	6.000	4.000	8.000	5.000	**1.000**	10.000	9.000
ett2/D/short	9.000	2.000	7.000	8.000	3.000	10.000	5.000	**1.000**	4.000	6.000
ett2/H/long	**1.000**	7.000	6.000	4.000	8.000	10.000	3.000	2.000	9.000	5.000
ett2/H/medium	6.000	7.000	4.000	2.000	8.000	10.000	3.000	5.000	9.000	**1.000**
ett2/H/short	5.000	7.000	**1.000**	3.000	2.000	10.000	4.000	6.000	8.000	9.000
ett2/W/short	4.000	**1.000**	7.000	8.000	3.000	6.000	10.000	5.000	9.000	2.000
hierarchical_sales/D/short	7.000	**1.000**	6.000	2.000	4.000	9.000	8.000	10.000	5.000	3.000
hierarchical_sales/W/short	2.000	4.000	**1.000**	6.000	5.000	3.000	10.000	9.000	8.000	7.000
hospital/M/short	6.000	4.000	**1.000**	5.000	9.000	8.000	10.000	7.000	3.000	2.000
jena_weather/10T/long	5.000	3.000	2.000	**1.000**	7.000	4.000	6.000	8.000	10.000	9.000
jena_weather/10T/medium	4.000	3.000	2.000	**1.000**	6.000	5.000	7.000	9.000	10.000	8.000
jena_weather/10T/short	4.000	3.000	2.000	**1.000**	5.000	6.000	7.000	8.000	9.000	10.000
jena_weather/D/short	7.000	2.000	4.000	8.000	**1.000**	3.000	5.000	10.000	9.000	6.000
jena_weather/H/long	5.000	**1.000**	8.000	2.000	6.000	9.000	4.000	10.000	3.000	7.000
jena_weather/H/medium	5.000	**1.000**	3.000	2.000	4.000	9.000	6.000	10.000	8.000	7.000
jena_weather/H/short	3.000	**1.000**	2.000	6.000	5.000	4.000	7.000	10.000	9.000	8.000
kdd_cup_2018/D/short	5.000	9.000	7.000	8.000	3.000	**1.000**	2.000	10.000	6.000	4.000
kdd_cup_2018/H/long	6.000	2.000	8.000	7.000	**1.000**	10.000	5.000	9.000	3.000	4.000
kdd_cup_2018/H/medium	5.000	2.000	6.000	8.000	**1.000**	9.000	4.000	10.000	3.000	7.000
kdd_cup_2018/H/short	6.000	2.000	5.000	8.000	**1.000**	10.000	4.000	9.000	3.000	7.000
loop_seattle/5T/long	9.000	6.000	3.000	4.000	11.000	7.000	8.000	5.000	**1.000**	2.000
loop_seattle/5T/medium	9.000	6.000	3.000	4.000	10.000	7.000	8.000	5.000	**1.000**	2.000
loop_seattle/5T/short	9.000	4.000	5.000	3.000	8.000	7.000	10.000	6.000	**1.000**	2.000
loop_seattle/D/short	3.000	2.000	**1.000**	8.000	5.000	4.000	6.000	10.000	9.000	7.000
loop_seattle/H/long	8.000	**1.000**	3.000	4.000	9.000	2.000	6.000	5.000	7.000	10.000
loop_seattle/H/medium	8.000	2.000	4.000	**1.000**	9.000	3.000	6.000	5.000	7.000	10.000
loop_seattle/H/short	9.000	**1.000**	2.000	5.000	6.000	3.000	4.000	8.000	7.000	10.000
m4_daily/D/short	9.000	**1.000**	7.000	3.000	2.000	4.000	6.000	5.000	10.000	11.000
m4_hourly/H/short	7.000	2.000	3.000	10.000	6.000	8.000	5.000	9.000	**1.000**	4.000
m4_monthly/M/short	**1.000**	2.000	3.000	8.000	6.000	4.000	10.000	9.000	7.000	5.000
m4_quarterly/Q/short	4.000	3.000	5.000	7.000	6.000	8.000	10.000	9.000	**1.500**	**1.500**
m4_weekly/W/short	6.000	**1.000**	2.000	10.000	4.000	3.000	5.000	7.000	8.000	9.000
m4_yearly/A/short	7.000	4.000	3.000	9.000	8.000	5.000	11.000	6.000	**1.000**	2.000
m_dense/D/short	2.000	3.000	5.000	7.000	4.000	**1.000**	8.000	6.000	9.000	10.000
m_dense/H/long	10.000	3.000	2.000	6.000	8.000	7.000	9.000	5.000	**1.000**	4.000
m_dense/H/medium	10.000	3.000	2.000	4.000	8.000	9.000	7.000	5.000	**1.000**	6.000
m_dense/H/short	10.000	3.000	4.000	7.000	**1.000**	8.000	9.000	6.000	2.000	5.000
restaurant/D/short	3.000	**1.000**	2.000	10.000	5.000	4.000	9.000	7.000	8.000	6.000
saugeen/D/short	7.000	8.000	**1.000**	3.000	2.000	6.000	5.000	10.000	9.000	4.000
saugeen/M/short	5.000	7.000	2.000	4.000	3.000	**1.000**	6.000	9.000	8.000	10.000
saugeen/W/short	7.000	**1.000**	2.000	5.000	3.000	6.000	4.000	10.000	9.000	8.000
solar/10T/long	7.000	**1.000**	3.000	4.000	5.000	2.000	8.000	6.000	10.000	11.000
solar/10T/medium	5.000	4.000	3.000	2.000	6.000	**1.000**	8.000	7.000	10.000	11.000
solar/10T/short	2.000	7.000	4.000	5.000	3.000	**1.000**	8.000	6.000	9.000	10.000
solar/D/short	3.000	4.000	2.000	7.000	5.000	**1.000**	6.000	10.000	8.000	9.000
solar/H/long	8.000	**1.000**	3.000	2.000	9.000	5.000	6.000	10.000	4.000	7.000
solar/H/medium	9.000	**1.000**	5.000	4.000	8.000	2.000	7.000	10.000	6.000	3.000
solar/H/short	8.000	**1.000**	5.000	3.000	2.000	6.000	9.000	10.000	4.000	7.000
solar/W/short	4.000	5.000	2.000	7.000	3.000	**1.000**	11.000	6.000	9.000	10.000
sz_taxi/15T/long	6.000	**1.000**	4.000	2.000	10.000	9.000	3.000	8.000	7.000	5.000
sz_taxi/15T/medium	6.000	**1.000**	3.000	2.000	10.000	9.000	4.000	7.000	8.000	5.000
sz_taxi/15T/short	6.000	**1.000**	3.000	4.000	2.000	7.000	5.000	8.000	10.000	9.000
sz_taxi/H/short	6.000	**1.000**	5.000	4.000	2.000	7.000	3.000	8.000	10.000	9.000
temperature_rain/D/short	8.000	5.000	4.000	6.000	3.000	7.000	9.000	10.000	**1.000**	2.000
us_births/D/short	2.000	4.000	5.000	7.000	6.000	**1.000**	10.000	3.000	9.000	8.000
us_births/M/short	4.000	7.000	6.000	2.000	11.000	8.000	3.000	**1.000**	9.000	5.000
us_births/W/short	2.000	4.000	3.000	8.000	5.000	**1.000**	7.000	6.000	10.000	9.000

Table 12: MASE performance presented as average ranks for zero-shot models across the GiftEval benchmark. A lower average rank indicates consistently higher accuracy across datasets. The two models with the best overall average ranks are highlighted.

Dataset	TempoPFN	TiRex	FlowState-9.1M	Toto_Open_Base_1.0	Chronos Bolt B	TTM-R2-Finetuned	TabPFN-TS	YingLong_50m	Moirai L 1.1	Moirai B 1.1
bitbrains_fast_storage/5T/long	6.000	2.000	9.000	**1.000**	4.000	3.000	11.000	8.000	5.000	7.000
bitbrains_fast_storage/5T/medium	8.000	2.000	10.000	**1.000**	5.000	6.000	11.000	7.000	3.000	4.000
bitbrains_fast_storage/5T/short	6.000	2.000	10.000	**1.000**	4.000	3.000	9.000	7.000	8.000	5.000
bitbrains_fast_storage/H/short	8.000	3.000	5.000	**1.000**	2.000	10.000	9.000	6.000	4.000	7.000
bitbrains_rnd/5T/long	9.000	2.000	10.000	**1.000**	3.000	7.000	11.000	8.000	4.000	5.000
bitbrains_rnd/5T/medium	8.000	**1.000**	10.000	2.000	3.000	5.000	11.000	9.000	4.000	6.000
bitbrains_rnd/5T/short	8.000	2.000	9.000	**1.000**	3.000	5.000	11.000	6.000	4.000	7.000
bitbrains_rnd/H/short	7.000	2.000	5.000	**1.000**	4.000	9.000	11.000	3.000	6.000	10.000
bizitobs_application/10S/long	5.000	6.000	4.000	3.000	10.000	7.000	**1.000**	8.000	9.000	11.000
bizitobs_application/10S/medium	**1.000**	6.000	4.000	2.000	10.000	7.000	3.000	8.000	9.000	11.000
bizitobs_application/10S/short	**1.000**	4.000	5.000	2.000	11.000	6.000	3.000	7.000	9.000	10.000
bizitobs_l2c/5T/long	9.000	8.000	2.000	6.000	10.000	**1.000**	3.000	7.000	4.500	4.500
bizitobs_l2c/5T/medium	5.000	6.000	**1.000**	4.000	9.000	2.000	3.000	7.000	10.000	8.000
bizitobs_l2c/5T/short	3.000	8.000	10.000	2.000	4.000	**1.000**	9.000	6.000	5.000	7.000
bizitobs_l2c/H/long	5.000	3.000	**1.000**	7.000	2.000	9.000	4.000	6.000	10.000	8.000
bizitobs_l2c/H/medium	5.000	4.000	**1.000**	6.000	3.000	8.000	2.000	7.000	9.000	10.000
bizitobs_l2c/H/short	3.000	6.000	2.000	4.000	**1.000**	7.000	5.000	8.000	10.000	9.000
bizitobs_service/10S/long	7.000	6.000	4.000	**1.000**	10.000	5.000	2.000	8.000	9.000	11.000
bizitobs_service/10S/medium	7.000	4.000	2.000	**1.000**	10.000	5.000	3.000	8.000	9.000	11.000
bizitobs_service/10S/short	6.000	4.000	3.000	**1.000**	10.000	2.000	5.000	7.000	9.000	11.000
car_parts/M/short	5.000	4.000	8.000	**1.000**	7.000	3.000	6.000	11.000	9.000	2.000
covid_deaths/D/short	6.000	8.000	4.000	2.000	7.000	**1.000**	9.000	10.000	5.000	3.000
electricity/15T/long	11.000	2.000	3.000	7.000	4.000	**1.000**	5.000	6.000	9.000	10.000
electricity/15T/medium	9.000	2.000	3.000	7.000	4.000	**1.000**	5.000	6.000	10.000	11.000
electricity/15T/short	8.000	3.000	5.000	6.000	2.000	**1.000**	7.000	4.000	10.000	9.000
electricity/D/short	10.000	**1.000**	2.000	6.000	3.000	4.000	7.000	5.000	9.000	8.000
electricity/H/long	10.000	3.000	2.000	4.000	5.000	**1.000**	7.000	8.000	9.000	6.000
electricity/H/medium	10.000	2.000	4.000	5.000	**1.000**	3.000	7.000	6.000	9.000	8.000
electricity/H/short	7.000	**1.000**	4.000	5.000	2.000	3.000	6.000	10.000	8.000	9.000
electricity/W/short	6.000	2.000	**1.000**	8.000	3.000	4.000	5.000	7.000	9.000	10.000
ett1/15T/long	10.000	4.000	**1.000**	5.000	8.000	3.000	6.000	2.000	11.000	7.000
ett1/15T/medium	8.000	3.000	**1.000**	6.000	4.000	5.000	7.000	2.000	11.000	10.000
ett1/15T/short	8.000	4.000	2.000	3.000	**1.000**	5.000	7.000	6.000	10.000	9.000
ett1/D/short	2.000	6.000	**1.000**	4.000	5.000	11.000	3.000	7.000	9.000	8.000
ett1/H/long	7.000	**1.000**	2.000	5.000	4.000	8.000	10.000	3.000	9.000	6.000
ett1/H/medium	7.000	2.000	**1.000**	4.000	9.000	5.000	10.000	3.000	6.000	8.000
ett1/H/short	6.000	3.000	**1.000**	8.000	2.000	5.000	10.000	4.000	7.000	9.000
ett1/W/short	2.000	11.000	**1.000**	5.000	8.000	6.000	7.000	9.000	3.000	4.000
ett2/15T/long	9.000	3.000	**1.000**	2.000	6.000	5.000	7.000	4.000	10.000	11.000
ett2/15T/medium	7.000	2.000	**1.000**	6.000	5.000	4.000	8.000	3.000	10.000	11.000
ett2/15T/short	6.000	2.000	3.000	7.000	4.000	**1.000**	8.000	5.000	10.000	9.000
ett2/D/short	11.000	2.000	9.000	10.000	5.000	**1.000**	7.000	4.000	8.000	3.000
ett2/H/long	**1.000**	9.000	6.000	5.000	2.000	3.000	11.000	4.000	10.000	7.000
ett2/H/medium	7.000	8.000	6.000	2.000	3.000	4.000	11.000	**1.000**	9.000	5.000
ett2/H/short	5.000	4.000	**1.000**	3.000	2.000	6.000	10.000	7.000	8.000	9.000
ett2/W/short	7.000	5.000	9.000	10.000	**1.000**	4.000	2.000	8.000	11.000	6.000
hierarchical_sales/D/short	7.000	2.000	6.000	**1.000**	3.000	8.000	9.000	10.000	4.000	5.000
hierarchical_sales/W/short	4.000	2.000	**1.000**	7.000	6.000	3.000	5.000	10.000	9.000	8.000
hospital/M/short	6.000	3.000	2.000	8.000	9.000	**1.000**	4.000	10.000	5.000	7.000
jena_weather/10T/long	4.000	2.000	6.000	**1.000**	5.000	10.000	7.000	3.000	11.000	9.000
jena_weather/10T/medium	4.000	2.000	5.000	**1.000**	3.000	11.000	6.000	7.000	8.000	9.000
jena_weather/10T/short	3.000	4.000	2.000	**1.000**	5.000	7.000	6.000	8.000	9.000	10.000
jena_weather/D/short	9.000	**1.000**	2.000	7.000	3.000	10.000	8.000	4.000	5.000	6.000
jena_weather/H/long	8.000	2.000	7.000	3.000	4.000	11.000	10.000	5.000	**1.000**	6.000
jena_weather/H/medium	4.000	5.000	6.000	2.000	**1.000**	7.000	11.000	8.000	10.000	3.000
jena_weather/H/short	3.000	**1.000**	2.000	6.000	5.000	9.000	7.000	4.000	10.000	8.000
kdd_cup_2018/D/short	7.000	10.000	9.000	8.000	3.000	4.000	2.000	**1.000**	5.500	5.500
kdd_cup_2018/H/long	7.000	2.000	9.000	8.000	**1.000**	5.000	10.000	6.000	3.000	4.000
kdd_cup_2018/H/medium	7.000	2.000	9.000	8.000	**1.000**	4.000	10.000	5.000	3.000	6.000
kdd_cup_2018/H/short	8.000	2.000	7.000	9.000	**1.000**	6.000	10.000	4.000	3.000	5.000
loop_seattle/5T/long	9.000	6.000	3.000	4.000	10.000	5.000	7.000	8.000	**1.000**	2.000
loop_seattle/5T/medium	11.000	6.000	3.000	4.000	9.000	5.000	7.000	8.000	**1.000**	2.000
loop_seattle/5T/short	9.000	5.000	6.000	4.000	8.000	3.000	7.000	10.000	**1.000**	2.000
loop_seattle/D/short	3.000	**1.000**	2.000	9.000	5.000	4.000	7.000	10.000	8.000	6.000
loop_seattle/H/long	9.000	**1.000**	3.000	5.000	7.000	2.000	4.000	6.000	8.000	10.000
loop_seattle/H/medium	9.000	3.000	5.000	2.000	8.000	**1.000**	4.000	7.000	6.000	10.000
loop_seattle/H/short	9.000	**1.000**	2.000	5.000	6.000	3.000	7.000	4.000	8.000	10.000
m4_daily/D/short	10.000	**1.000**	6.000	5.000	2.000	4.000	9.000	7.000	8.000	11.000
m4_hourly/H/short	4.000	**1.000**	2.000	6.000	5.000	10.000	3.000	8.000	7.000	9.000
m4_monthly/M/short	3.000	2.000	**1.000**	9.000	5.000	4.000	7.000	10.000	8.000	6.000
m4_quarterly/Q/short	6.000	4.000	3.000	9.000	7.000	5.000	8.000	10.000	**1.500**	**1.500**
m4_weekly/W/short	8.000	**1.000**	3.000	7.000	5.000	2.000	4.000	6.000	9.000	11.000
m4_yearly/A/short	7.000	8.000	**1.000**	6.000	9.000	4.000	5.000	11.000	2.000	3.000
m_dense/D/short	2.000	3.000	6.000	7.000	4.000	5.000	**1.000**	8.000	9.000	10.000
m_dense/H/long	10.000	2.000	5.000	6.000	7.000	4.000	9.000	8.000	**1.000**	3.000
m_dense/H/medium	10.000	4.000	2.000	5.000	7.000	3.000	9.000	8.000	**1.000**	6.000
m_dense/H/short	10.000	3.000	4.000	7.000	**1.000**	5.000	8.000	9.000	2.000	6.000
restaurant/D/short	3.000	**1.000**	2.000	10.000	6.000	5.000	4.000	9.000	8.000	7.000
saugeen/D/short	7.000	9.000	**1.000**	4.000	2.000	5.000	8.000	6.000	10.000	3.000
saugeen/M/short	9.000	8.000	4.000	7.000	2.000	5.000	**1.000**	3.000	6.000	10.000
saugeen/W/short	7.000	2.000	**1.000**	5.000	3.000	8.000	6.000	4.000	9.000	10.000
solar/10T/long	8.000	**1.000**	2.000	5.000	7.000	6.000	3.000	9.000	10.000	11.000
solar/10T/medium	8.000	3.000	4.000	2.000	7.000	6.000	**1.000**	9.000	10.000	11.000
solar/10T/short	2.000	7.000	5.000	6.000	4.000	**1.000**	3.000	11.000	10.000	8.000
solar/D/short	2.000	**1.000**	8.000	9.000	3.000	4.000	6.000	7.000	5.000	10.000
solar/H/long	5.000	**1.000**	2.000	3.000	7.000	11.000	10.000	4.000	6.000	8.000
solar/H/medium	10.000	**1.000**	4.000	3.000	8.000	11.000	2.000	7.000	6.000	5.000
solar/H/short	10.000	**1.000**	4.000	3.000	2.000	7.000	9.000	6.000	5.000	8.000
solar/W/short	5.000	6.000	2.000	7.000	3.000	4.000	**1.000**	11.000	9.000	10.000
sz_taxi/15T/long	6.000	**1.000**	4.000	2.000	7.000	10.000	9.000	3.000	8.000	5.000
sz_taxi/15T/medium	8.000	**1.000**	4.000	3.000	7.000	2.000	9.000	5.000	10.000	6.000
sz_taxi/15T/short	7.000	**1.000**	5.000	3.000	2.000	8.000	6.000	4.000	10.000	9.000
sz_taxi/H/short	8.000	2.000	6.000	4.000	**1.000**	7.000	5.000	3.000	10.000	9.000
temperature_rain/D/short	8.000	4.000	5.000	6.000	2.000	9.000	7.000	10.000	**1.000**	3.000
us_births/D/short	3.000	4.000	5.000	7.000	6.000	2.000	**1.000**	10.000	8.000	9.000
us_births/M/short	4.000	7.000	5.000	2.000	11.000	**1.000**	6.000	3.000	10.000	8.000
us_births/W/short	2.000	4.000	3.000	8.000	5.000	6.000	**1.000**	7.000	10.000	9.000