Probabilistic Forecasting for Building Energy Systems: Are Time-Series Foundation Models the Answer?

Young-Jin Park[1*], **François Germain**[2], **Jing Liu**[2], **Ye Wang**[2], **Toshiaki Koike-Akino**[2], **Gordon Wichern**[2], **Christopher Laughman**[2], **Navid Azizan**[1], and **Ankush Chakrabarty**[2†]
[1]Massachusetts Institute of Technology
[2]Mitsubishi Electric Research Labs

Abstract

Accurate time-series forecasting is essential for real-world applications such as predictive maintenance and feedback control. While deep neural networks have shown promise in recognizing complex patterns and predicting trends, their generalization capabilities are open to debate, and they typically do not perform well with limited data. In this paper, we examine the potential of time-series foundation models (TSFM) as a practical solution for addressing real-world (probabilistic) forecasting challenges. Our experiments using real building data demonstrate that, through fine-tuning TSFMs, we can achieve excellent predictions, even with limited data, and improve generalization in zero-shot prediction on unseen tasks.

1 Motivation

Probabilistic time-series forecasting plays a crucial role in a range of real-world applications such as energy systems, especially as predictive models for anomaly detection based on confidence intervals or stochastic model-based predictive control. Deep forecasting models are especially useful when accurate and tractable first-principles models (e.g., a physics-based model) are difficult to obtain. Consequently, recent developments have focused on deep learning methods, which can identify patterns from historical data and provide predictions; c.f. DEEPAR [18], N-BEATS [15], and temporal fusion transformers (TFT) [13]. While deep learning approaches can yield accurate time-series forecasts [16], they also often produce unreliable forecasts, sometimes even under-performing compared to traditional statistical models like seasonal ARIMA or classical MLPs [10]. Moreover, for small datasets, these approaches are prone to issues like overfitting or mode collapse [7, 14].

Foundation models (FMs) [3] have recently emerged as a powerful tool, demonstrating excellent performance across numerous machine learning domains. The core idea of FMs involves leveraging a massive dataset to pre-train a large model, which can then be tuned to various downstream use cases. This approach contrasts with classical deep learning, which typically learns from scratch using task-specific data. Two primary advantages of FMs include: (i) their capacity to learn various function landscapes, allowing them to be expressive enough to contribute to different problem domains; and (ii) their strong generalization capability which can be attributed to exploiting patterns learned from a wide range of pre-training datasets. This generalizability is typically exploited for specific use cases or downstream tasks through a transfer learning process referred to as fine-tuning.

Despite their impressive track record in language and vision [4, 11, 17], their applicability and effectiveness in real-world time-series forecasting problems remain largely unexplored. This paper examines the effectiveness of *time-series foundation models* (TSFMs), *with emphasis on real multi-month data obtained from building energy systems*. In particular, we present four pressing research

[*]Work completed during internship at MERL.
[†]Corresponding author. achakrabarty@ieee.org

NeurIPS 2024 Workshop on Time Series in the Age of Large Models.

questions and provide answers based on empirical studies using time-series signals collected from real-world building energy systems: (**RQ1**) *Zero-shot prediction quality*: How effective are baseline TSFMs? (**RQ2**) *Effectiveness of fine-tuning*: Can fine-tuning improve baseline TSFM forecasts? (**RQ3**) *Performance with limited data*: How do TSFMs perform on problems with limited data? (**RQ4**) *Generalization to unseen tasks*: Can TSFMs predict accurately on unseen systems?

2 Experimental Setup: Benchmark Models, Datasets, and Evaluation

TSFMs: Salesforce's MOIRAI [22], Google's TIMESFM [5], and Amazon's CHRONOS [2]. **Deep Forecasting Models:** LSTM[8]-based DEEPAR, N-BEATS , and TFT. See Appendix A for details.

For our experiments, we test the aforementioned forecasting methods on four different types of time-series signals, collected at 15-minute intervals from SUSTIE, a Mitsubishi Electric's net-zero energy building, during office workdays. The signals include room occupancy (`Occ`), carbon emissions (`CO2`), power consumption for illumination and appliances (`Light`), and energy consumption for heating, ventilation, and air conditioning (HVAC) equipment (`HVAC`).

To reduce bias in the experimental results, we collected each time series from 8 zones spanning 3 floors (i.e., 3+3+2=8). Moreover, we conducted the experiments using 4 different cross-validation splits, each representing a different season through a blocked-split approach; each signal spans 3 months, resulting in 8 zones $\times$ 4 seasons = 32 different time series. For each time series, the last 2 weeks of signals are excluded and used as test periods. Based on the building's operational specifications, we used 24-hour signals as the context length (i.e., look-back period) to predict the following 4 hours. We conduct a rolling-window analysis [23] and report the mean $\pm$ standard deviation error across these 32 time series.

3 Results

To evaluate the efficacy of different TSFMs compared to classical deep forecasting models such as TFT, we assess the performance based on: mean absolute scaled error (MASE), root-mean-squared scaled error (RMSSE), mean scaled interval score (MSIS), and weighted quantile loss (wQL); c.f. Appendix B for further details. Note that MASE and RMSSE evaluate the accuracy of the models' point predictions, while MSIS and wQL assess the quality of their probabilistic predictions. We compare against the popular N-BEATS which is deterministic: probabilistic metrics do not apply.

3.1 Zero-Shot Prediction Quality (RQ1)

Table 1: Comparison of forecasting performance. The winner/runner-up is highlighted in light blue/yellow. Fine-tuned TSFMs outperform baseline TSFMs and s.o.t.a. deep forecasting models.

		Deep Forecasting Models			TSFMs (Zero-Shot)			Fine-tuned TSFMs	
Dataset	**Metric**	N-BEATS	DEEPAR	TFT	MOIRAI	TIMESFM	CHRONOS	CHRONOS +FullFT	CHRONOS +LoRA
`Occ`	MASE	0.32 $\pm$ 0.06	0.27 $\pm$ 0.05	0.27 $\pm$ 0.05	0.70 $\pm$ 0.12	0.51 $\pm$ 0.07	0.60 $\pm$ 0.08	0.24 $\pm$ 0.04	0.23 $\pm$ 0.04
	RMSSE	0.33 $\pm$ 0.05	0.27 $\pm$ 0.04	0.28 $\pm$ 0.05	0.68 $\pm$ 0.11	0.50 $\pm$ 0.07	0.59 $\pm$ 0.07	0.25 $\pm$ 0.04	0.24 $\pm$ 0.04
	MSIS	-	4.24 $\pm$ 0.92	3.93 $\pm$ 0.89	9.11 $\pm$ 1.50	10.84 $\pm$ 1.55	8.50 $\pm$ 1.26	3.80 $\pm$ 0.72	3.58 $\pm$ 0.82
	wQL	-	0.44 $\pm$ 0.08	0.43 $\pm$ 0.08	1.09 $\pm$ 0.18	1.12 $\pm$ 0.17	1.11 $\pm$ 0.15	0.43 $\pm$ 0.07	0.40 $\pm$ 0.07
`CO2`	MASE	0.34 $\pm$ 0.09	0.48 $\pm$ 0.16	0.38 $\pm$ 0.10	0.70 $\pm$ 0.14	0.55 $\pm$ 0.12	0.50 $\pm$ 0.11	0.32 $\pm$ 0.08	0.31 $\pm$ 0.07
	RMSSE	0.31 $\pm$ 0.08	0.43 $\pm$ 0.14	0.34 $\pm$ 0.09	0.66 $\pm$ 0.13	0.50 $\pm$ 0.11	0.47 $\pm$ 0.10	0.29 $\pm$ 0.07	0.29 $\pm$ 0.07
	MSIS	-	7.75 $\pm$ 2.86	5.27 $\pm$ 1.73	8.54 $\pm$ 1.79	12.08 $\pm$ 2.59	6.96 $\pm$ 1.51	5.43 $\pm$ 1.38	4.54 $\pm$ 1.16
	wQL	-	0.15 $\pm$ 0.06	0.10 $\pm$ 0.03	0.18 $\pm$ 0.05	0.20 $\pm$ 0.06	0.14 $\pm$ 0.03	0.10 $\pm$ 0.03	0.09 $\pm$ 0.02
`Light`	MASE	0.63 $\pm$ 0.06	0.28 $\pm$ 0.05	0.27 $\pm$ 0.08	0.83 $\pm$ 0.09	0.50 $\pm$ 0.08	0.52 $\pm$ 0.07	0.22 $\pm$ 0.04	0.22 $\pm$ 0.04
	RMSSE	0.64 $\pm$ 0.06	0.30 $\pm$ 0.05	0.29 $\pm$ 0.07	0.80 $\pm$ 0.10	0.49 $\pm$ 0.08	0.51 $\pm$ 0.07	0.26 $\pm$ 0.04	0.25 $\pm$ 0.04
	MSIS	-	4.47 $\pm$ 1.17	3.79 $\pm$ 0.99	11.30 $\pm$ 1.10	10.60 $\pm$ 1.72	7.74 $\pm$ 1.13	3.38 $\pm$ 0.70	3.10 $\pm$ 0.66
	wQL	-	0.33 $\pm$ 0.07	0.31 $\pm$ 0.10	0.93 $\pm$ 0.10	0.77 $\pm$ 0.15	0.66 $\pm$ 0.10	0.28 $\pm$ 0.05	0.25 $\pm$ 0.05
`HVAC`	MASE	0.51 $\pm$ 0.46	0.41 $\pm$ 0.35	0.40 $\pm$ 0.30	0.79 $\pm$ 0.66	0.58 $\pm$ 0.51	0.42 $\pm$ 0.49	0.32 $\pm$ 0.24	0.32 $\pm$ 0.24
	RMSSE	0.48 $\pm$ 0.40	0.38 $\pm$ 0.31	0.35 $\pm$ 0.25	0.69 $\pm$ 0.54	0.48 $\pm$ 0.40	0.37 $\pm$ 0.41	0.30 $\pm$ 0.22	0.30 $\pm$ 0.22
	MSIS	-	7.59 $\pm$ 6.63	5.38 $\pm$ 3.91	9.38 $\pm$ 8.55	12.56 $\pm$ 11.10	7.08 $\pm$ 8.59	5.01 $\pm$ 3.79	5.00 $\pm$ 4.03
	wQL	-	0.94 $\pm$ 0.81	0.80 $\pm$ 0.58	1.42 $\pm$ 1.25	1.59 $\pm$ 1.42	0.98 $\pm$ 1.16	0.72 $\pm$ 0.57	0.70 $\pm$ 0.56

From Table 1, we observe the following. *Among the TSFMs,* TIMESFM *and* CHRONOS *perform similarly, while* MOIRAI *generally exhibits the highest errors on our datasets.* From the perspective

of distribution forecasting, CHRONOS consistently achieves the best performance amongst these three models. We attribute this to differences in the architectures and training methodologies, as this is in line with current conventional wisdom regarding generative models for continuous data (e.g., images), i.e., that models based on tokenized (i.e., discrete) representation often outperform models with continuous variables [19, 12]. In our case, we see CHRONOS, which operates on a tokenized representation of time series data and is trained via a classification cross-entropy loss, performing better than TIMESFM, i.e., a fully-connected network trained via a regression loss.

It is worth noting that our signals are not completely distinct but may share similarities with some of the pre-training datasets, such as `PedestrianCounts`, `SpanishEnergyAndWeather`, `KDDCup2018`, and `AustralianElectricity`. Nevertheless, the zero-shot performance of TSFMs falls short compared to models specifically trained on the downstream datasets, contradicting the original paper's claim of having comparable or occasionally superior zero-shot performance on new datasets. This discrepancy highlights the need for further examination, as discussed in Section 4.

Given that CHRONOS demonstrates the most promising preliminary performance, and its implementation is compatible with HuggingFace's Transformers library [21], we conduct the subsequent experiments using the CHRONOS models.

3.2 Effectiveness of (Parameter-Efficient) Fine-Tuning (RQ2)

Next, we investigate whether we can potentially improve upon the (underwhelming) zero-shot performance of TSFMs with fine-tuning. Concretely, we test the effects of full fine-tuning (Full-FT) along with parameter-efficient fine-tuning (PEFT) via low-rank adaptation (LoRA) [9], which has shown promise in fine-tuning large language models (LLMs) but has been under-explored for TSFMs.

From Table 1, we deduce that *fine-tuning proves to be effective when applying TSFMs to our real-world data*. Fine-tuned CHRONOS clearly and consistently outperforms the benchmark deep forecasting models with both FullFT and PEFT with 1K fine-tuning gradient steps, sometimes reducing the error metric by more than 50%. We also check the impact of LoRA rank by varying $r = 4$, 16, and 64 for 1K training iterations; see Table 3 in Appendix C. Interestingly, Table 3 implies that LoRA is not only comparable, but actually outperforms (albeit marginally) Full-FT. This is consistent with the observations of LoRA in language tasks [9], and mainly because LoRA can prevent overfitting. Since the PEFT results are all similarly good, we recommend a lower rank (e.g., $r = 4$) to reduce the computational expense, for limited data and few FT iterations. In fact, using LoRA $r = 4$ reduces our total training floating-point operations (FLOPS) by 33% and accelerates training by 2.3×, compared to Full-FT, on a NVIDIA RTX2080Ti GPU with 6 CPU cores.

3.3 Transferability with limited data (RQ3)

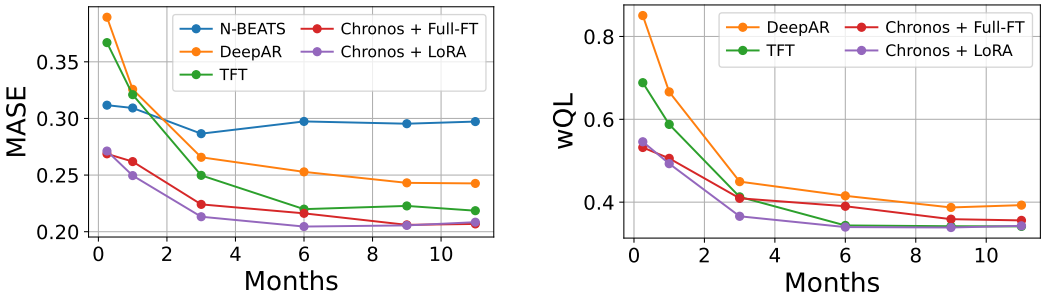

Figure 1: Forecasting accuracy by training dataset size on the `Occ` dataset.

So far, we have demonstrated that TSFMs can successfully forecast real-world building time-series data with fine-tuning. An immediate question is how much time-series data is needed to fine-tune such large models and whether a fine-tuned model can achieve high accuracy with less data. To explore this, we vary the training dataset size across periods of 1 week, and $\{1, 3, 6, 9, 11\}$ months. For a fair comparison, we keep the evaluation period fixed while extending the training duration accordingly. The plot of forecasting accuracy across different models for the `Occ` signal is shown in Figure 1, with the tabulated results provided in Appendix C Tables 4–9.

When sufficient training data is available, TFT demonstrates performance comparable to that of TSFMs. However, the forecasting accuracy of both deep learning baselines and TSFMs declines as the training dataset size decreases. Notably, the performance gap between tiny datasets (e.g., 1 week) and sufficient datasets (e.g., $\geq$ 6 months) is $\approx 0.5\times$ for CHRONOS compared to DEEPAR and TFT. This suggests that *TSFMs exhibit greater robustness to the small data problem*, presumably due to pre-training; pre-training enables models to capture general patterns from large datasets, which helps mitigate overfitting and enhances generalization [6].

3.4 Generalizability to unseen tasks (RQ4)

In real applications, one is often faced with predicting time-series for a new client/user: i.e., with extremely little prior data available for training. In such circumstances, we evaluate the generalizability of TSFMs to unseen (but known to be similar) thermal zones. We fine-tuned CHRONOS using data for 3 months from a zone on the 2nd floor of our commercial building, and then tested it on the one from the 3rd (Table 2) and the other from the 4th floor (Table 10) without any additional tuning. As in the previous sections, we conducted tests on the four splits.

We observe that TSFM with fine-tuning outperforms TFT predictions comprehensively in terms of errors between point estimates. As in previous sections, LoRA $r = 4$ often does slightly better than Full-FT. The improvement of fine-tuned TSFMs over TFT is especially apparent from the `Light` and `HVAC` categories, where the TSFM competitors generalize significantly better due to an informative prior learned from a pre-training dataset containing energy data (albeit from a completely different source). For `Occ` and `CO2` (which are correlated), since usage patterns across zones are not very dissimilar, TFT exhibits slight improvement in the probabilistic metrics, but is still outclassed by fine-tuned TSFMs. This demonstrates clearly how the pre-training prior of CHRONOS helps regularize the training and build predictions more robust to unseen patterns. With probabilistic forecasting deemed more complex, we surmise it is more susceptible to overfitting, occasionally competing against the benefits from the pre-trained prior and potentially explaining the few rare cases when TFT outperforms Full-FT on probabilistic metrics.

Table 2: Comparison of forecasting accuracy between TFT and TSFMs on *unseen* zone. The model is trained on a zone on the 2nd floor, then tested on another zone on the 3rd floor.

			Fine-tuned TSFMs					Fine-tuned TSFMs	
Dataset	**Metric**	TFT	CHRONOS +FullFT	CHRONOS +LoRA	**Dataset**	**Metric**	TFT	CHRONOS +Full-FT	CHRONOS +LoRA
`Occ`	MASE	0.271 ± 0.071	0.231 ± 0.061	0.229 ± 0.064	`Light`	MASE	0.262 ± 0.061	0.216 ± 0.056	0.223 ± 0.049
	RMSSE	0.278 ± 0.070	0.248 ± 0.062	0.243 ± 0.063		RMSSE	0.292 ± 0.056	0.255 ± 0.057	0.258 ± 0.050
	MSIS	4.040 ± 1.030	3.693 ± 1.179	3.629 ± 1.173		MSIS	3.818 ± 0.936	3.368 ± 1.045	3.239 ± 0.744
	wQL	0.432 ± 0.110	0.435 ± 0.114	0.416 ± 0.117		wQL	0.305 ± 0.062	0.281 ± 0.076	0.259 ± 0.059
`CO2`	MASE	0.345 ± 0.087	0.274 ± 0.074	0.255 ± 0.070	`HVAC`	MASE	0.436 ± 0.310	0.410 ± 0.235	0.383 ± 0.270
	RMSSE	0.309 ± 0.074	0.255 ± 0.068	0.239 ± 0.064		RMSSE	0.380 ± 0.292	0.363 ± 0.263	0.347 ± 0.276
	MSIS	4.875 ± 1.384	4.742 ± 1.230	3.624 ± 1.009		MSIS	6.895 ± 3.724	6.462 ± 4.056	6.285 ± 4.443
	wQL	0.100 ± 0.024	0.101 ± 0.023	0.079 ± 0.018		wQL	0.984 ± 0.494	0.988 ± 0.427	0.924 ± 0.438

4 Conclusions and Open Opportunities

While the underwhelming performance of zero-shot TSFMs is understandable given that the models have not been exposed to real building usage data or a wide range of occupancy signals during pretraining, this highlights that TSFMs are still in the early stages of development and are not yet ready for plug-and-play with real-world applications. Instead of relying on in-context inference alone, our preliminary study indicates that to reach their full potential, current TSFMs require fine-tuning until they are scaled significantly to comprise billions of parameters, similar to LLMs, and capable of learning from a massive data corpus.

One open challenge is multivariate forecasting, which is especially critical to building energy systems whose time series are correlated, as they are produced by dynamics that are strongly connected between zones. Current TSFMs predominantly focus on univariate time series, while real-world data is often multivariate. Furthermore, the most effective method to integrate additional static and/or dynamic covariates into TSFMs remains unclear. Finally, an interesting open problem is to enable domain adaptation e.g. from commercial building data, how to transfer knowledge to improve residential building forecasting problems, or from the building level to the city level.

References

[1] Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Danielle C. Maddix, Syama Rangapuram, David Salinas, Jasper Schulz, et al. GluonTS: Probabilistic and Neural Time Series Modeling in Python. *Journal of Machine Learning Research*, 21(116):1–6, 2020.

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

[3] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[5] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688*, 2023.

[6] Jacob Devlin. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[7] Cheng Fan, Fu Xiao, and Yang Zhao. A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy*, 195:222–233, 2017.

[8] Alex Graves and Alex Graves. Long short-term memory. In *Supervised Sequence Labelling With Recurrent Neural Networks*, pages 37–45. Springer, 2012.

[9] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[10] Seungjae Jung, Kyung-Min Kim, Hanock Kwak, and Young-Jin Park. A worrying analysis of probabilistic time-series models for sales forecasting. In *NeurIPS Workshops*. PMLR, 2020.

[11] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026. IEEE, 2023.

[12] Tianhong Li, Yonglong Tian, He Li, Mingyang Deng, and Kaiming He. Autoregressive image generation without vector quantization. *arXiv preprint arXiv:2406.11838*, 2024.

[13] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4): 1748–1764, 2021.

[14] Roberto Morcillo-Jimenez, Jesús Mesa, Juan Gómez-Romero, M Amparo Vila, and Maria J Martin-Bautista. Deep learning for prediction of energy consumption: an applied use case in an office building. *Applied Intelligence*, 54(7):5813–5825, 2024.

[15] Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*, 2019.

[16] Young-Jin Park, Donghyun Kim, Frédéric Odermatt, Juho Lee, and Kyung-Min Kim. A large-scale ensemble learning framework for demand forecasting. In *2022 IEEE International Conference on Data Mining (ICDM)*, pages 378–387. IEEE, 2022.

[17] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

[18] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3): 1181–1191, 2020.

[19] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[20] 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.

[21] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45. Association for Computational Linguistics, 2020.

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

[23] Eric Zivot and Jiahui Wang. *Modeling financial time series with S-PLUS*, volume 2. Springer, 2006.

Appendices

A Benchmark Models

Foundation Models We use our base model on official implementations of three recently developed TSFMs: Salesforce's MOIRAI [22], Google's TIMESFM [5], and Amazon's CHRONOS [2]. Despite some minor differences, these models utilize transformers [20] as their underlying architecture, with 91M, 200M, and 200M parameters, respectively. They are all trained on publicly available datasets encompassing various domains such as energy, weather, stock, and synthetic time-series datasets.

Deep Forecasting Models We compare these TSFMs to three prominent deep learning baselines: LSTM[8]-based DEEPAR, N-BEATS , and TFT; all implementations use GluonTS [1].

B Description of Performance Metrics

Let $Y := \{y_1, \ldots, y_t, \ldots\}$ denote a time-series. The main objective of probabilistic forecasting is to predict the conditional probability distribution

$$\pi := \pi(\overrightarrow{Y}_{t,P} \mid \overleftarrow{Y}_{t,C})$$

of the future target sequence

$$\overrightarrow{Y}_{t,P} := \{y_{t+1}, y_{t+2}, \ldots, y_{t+P}\}$$

based on a past context sequence

$$\overleftarrow{Y}_{t,C} := \{y_{t-C+1}, y_{t-C+2}, \ldots, y_t\}$$

for a given predictive window length $P \in \mathbb{N}$ and context window length $C \in \mathbb{N}$. Errors are computed between model predictions $\overrightarrow{Y}_{t,P}$ and ground truth $\overrightarrow{Y}^{\mathsf{true}}_{t,P} = \{y^{\mathsf{true}}_{t+1}, y^{\mathsf{true}}_{t+2}, \ldots, y^{\mathsf{true}}_{t+P}\}$. Let the absolute mean error of a naive forecasting:

$$\zeta_{MAE} = \frac{1}{T-C} \sum_{\tau=C+1}^{T} |y^{\mathsf{true}}_{\tau} - y^{\mathsf{true}}_{\tau-C}|.$$

where $T = |Y|$ is the length of a time-series. Similarly, let the root-mean-squared error of a navie forecasting:

$$\zeta_{RMSE} = \sqrt{\frac{1}{T-C} \sum_{\tau=C+1}^{T} |y^{\mathsf{true}}_{\tau} - y^{\mathsf{true}}_{\tau-C}|^2}.$$

The performance metrics chosen[‡] in the paper to evaluate model performance are as follows. Note that we do not want the metric to assign higher weights to easy-to-forecast sub time-series or (near) zero-series, since they, in fact, hold less practical importance in practice. Therefore, instead of using varying scales, we apply a constant scale across the series.

For MSIS, we use the 10% and 90% quantiles for lower and upper bounds, respectively. For wQL, we evaluate the 10%, 50%, and 90% quantiles..

B.1 Mean Absolute Scaled Error (MASE)

$$\mathsf{MASE}(\overrightarrow{Y}_{t,P}, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) = \frac{1}{\zeta_{MAE}} \cdot \frac{1}{P} \sum_{\tau=t+1}^{t+P} |y_\tau - y^{\mathsf{true}}_{\tau}|$$

[‡]Note that the evaluation metrics are modified arithmetically to ensure a constant scaling factor.

B.2 Root-Mean-Squared Scaled Error (RMSSE)

$$\mathsf{RMSSE}(\overrightarrow{Y}_{t,P}, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) = \frac{1}{\zeta_{RMSE}} \cdot \sqrt{\frac{1}{P}\sum_{\tau=t+1}^{t+P} |y_\tau - y_\tau^{\mathsf{true}}|^2}$$

For the probabilistic metrics some additional notation is required. Suppose $u_t = y_t^{(0.9)}$ and $l_t = y_t^{(0.1)}$ respectively denote the upper (90th) and lower (10th) quantiles considered during evaluation, $y_t^{(0.5)}$ denote the median, and $y_t^{(\beta)}$ denote the β-th quantile for $\beta \in (0, 1)$, all at time t. Let $\mathcal{I}(\cdot)$ denote an indicator function that is 1 when the parenthetical argument (usually a conditional statement) is satisfied, and 0 otherwise. Since we choose the 10-90 confidence interval, $\alpha = 0.1$ denotes the 10% coverage parameter.

B.3 Mean Scaled Interval Score (MSIS)

$$\mathsf{MSIS}(\pi, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) = \frac{1}{\zeta_{MAE}} \cdot \frac{1}{P}\sum_{\tau=t+1}^{t+P} \left[(u_t - l_t) + \frac{2}{\alpha}(l_t - y_t^{\mathsf{true}}) \cdot \mathcal{I}(y_t^{\mathsf{true}} < l_t) + \frac{2}{\alpha}(y_t^{\mathsf{true}} - u_t) \cdot \mathcal{I}(y_t^{\mathsf{true}} > u_t)\right]$$

B.4 Weighted Quantile Loss (wQL)

$$\mathsf{wQL}(\pi, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) := \frac{1}{3}\left[\mathsf{QL}^{(0.1)}(\pi, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) + \mathsf{QL}^{(0.5)}(\pi, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) + \mathsf{QL}^{(0.9)}(\pi, \overrightarrow{Y}^{\mathsf{true}}_{t,P})\right]$$

where,

$$\mathsf{QL}^{(\beta)}(\pi, \overrightarrow{Y}^{\mathsf{true}}_{t,P}) = \frac{\frac{2}{P}\sum_{\tau=t+1}^{t+P}\left[\beta \cdot \max(0,\ y_t^{\mathsf{true}} - y_t^{(\beta)}) + (1-\beta) \cdot \max(0,\ y_t^{(\beta)} - y_t^{\mathsf{true}})^{(j)}\right]}{\frac{1}{T-C}\sum_{\tau=C+1}^{T} |y_\tau^{\mathsf{true}}|}$$

C Supplementary Tables

Table 3: Comparison of forecasting accuracy between PEFT and FullFT.

Dataset	Metric	LoRA (rank=4)	LoRA (rank=16)	LoRA (rank=64)	FullFT
Occ	MASE	0.229 ± 0.042	0.230 ± 0.042	0.229 ± 0.042	0.235 ± 0.037
	RMSSE	0.242 ± 0.041	0.243 ± 0.040	0.242 ± 0.041	0.250 ± 0.037
	MSIS	3.577 ± 0.815	3.599 ± 0.814	3.588 ± 0.827	3.801 ± 0.724
	wQL	0.402 ± 0.074	0.403 ± 0.074	0.402 ± 0.075	0.434 ± 0.069
CO2	MASE	0.306 ± 0.074	0.305 ± 0.077	0.306 ± 0.077	0.316 ± 0.081
	RMSSE	0.287 ± 0.069	0.287 ± 0.071	0.287 ± 0.072	0.293 ± 0.074
	MSIS	4.543 ± 1.161	4.535 ± 1.132	4.538 ± 1.142	5.427 ± 1.377
	wQL	0.087 ± 0.021	0.086 ± 0.022	0.086 ± 0.021	0.104 ± 0.025
Light	MASE	0.215 ± 0.035	0.216 ± 0.035	0.216 ± 0.035	0.215 ± 0.036
	RMSSE	0.250 ± 0.037	0.251 ± 0.036	0.251 ± 0.037	0.256 ± 0.038
	MSIS	3.102 ± 0.659	3.134 ± 0.657	3.134 ± 0.668	3.376 ± 0.698
	wQL	0.254 ± 0.047	0.256 ± 0.046	0.256 ± 0.046	0.282 ± 0.052
HVAC	MASE	0.320 ± 0.239	0.319 ± 0.238	0.319 ± 0.238	0.315 ± 0.244
	RMSSE	0.299 ± 0.217	0.300 ± 0.217	0.300 ± 0.216	0.301 ± 0.216
	MSIS	4.995 ± 4.028	5.024 ± 4.091	4.979 ± 4.053	5.012 ± 3.790
	wQL	0.698 ± 0.564	0.700 ± 0.567	0.700 ± 0.566	0.716 ± 0.570

Table 4: Comparison of forecasting accuracy (**1 week**).

		Deep Forecasting Models			Fine-tuned TSFMs	
Dataset	**Metric**	N-BEATS	DEEPAR	TFT	CHRONOS +FullFT	CHRONOS +LoRA
Occ	MASE	0.340 ± 0.048	0.461 ± 0.096	0.559 ± 0.296	0.293 ± 0.052	0.293 ± 0.059
	RMSSE	0.339 ± 0.044	0.453 ± 0.088	0.530 ± 0.253	0.307 ± 0.048	0.308 ± 0.054
	MSIS	-	8.067 ± 1.999	7.850 ± 3.622	4.863 ± 0.869	4.983 ± 0.982
	wQL	-	0.967 ± 0.230	1.084 ± 0.544	0.571 ± 0.095	0.582 ± 0.110
CO2	MASE	0.467 ± 0.134	2.017 ± 1.367	0.705 ± 0.223	0.392 ± 0.108	0.381 ± 0.121
	RMSSE	0.412 ± 0.118	1.652 ± 1.106	0.612 ± 0.188	0.359 ± 0.100	0.351 ± 0.111
	MSIS	-	38.715 ± 26.897	11.621 ± 3.817	6.972 ± 1.867	6.644 ± 2.002
	wQL	-	0.730 ± 0.444	0.242 ± 0.074	0.137 ± 0.041	0.131 ± 0.047
Light	MASE	0.544 ± 0.103	0.693 ± 0.309	0.467 ± 0.159	0.275 ± 0.051	0.281 ± 0.056
	RMSSE	0.568 ± 0.088	0.659 ± 0.269	0.468 ± 0.136	0.315 ± 0.050	0.323 ± 0.054
	MSIS	-	11.757 ± 6.195	7.615 ± 3.464	4.555 ± 0.901	4.632 ± 0.950
	wQL	-	1.052 ± 0.499	0.716 ± 0.364	0.390 ± 0.075	0.396 ± 0.082
HVAC	MASE	0.462 ± 0.387	0.581 ± 0.633	0.553 ± 0.375	0.360 ± 0.288	0.341 ± 0.265
	RMSSE	0.442 ± 0.358	0.488 ± 0.485	0.461 ± 0.296	0.337 ± 0.242	0.318 ± 0.223
	MSIS	-	10.804 ± 12.535	8.857 ± 6.034	5.632 ± 4.229	5.319 ± 3.983
	wQL	-	1.450 ± 1.614	1.328 ± 0.939	0.853 ± 0.700	0.791 ± 0.636

Table 5: Comparison of forecasting accuracy (**1 month**).

Dataset	Metric	Deep Forecasting Models			Fine-tuned TSFMs	
		N-BEATS	DeepAR	TFT	Chronos +FullFT	Chronos +LoRA
Occ	MASE	0.329 ± 0.059	0.319 ± 0.069	0.358 ± 0.077	0.272 ± 0.052	0.267 ± 0.049
	RMSSE	0.333 ± 0.059	0.326 ± 0.063	0.356 ± 0.069	0.288 ± 0.051	0.283 ± 0.048
	MSIS	-	5.331 ± 1.310	5.396 ± 1.389	4.488 ± 0.930	4.544 ± 0.898
	wQL	-	0.609 ± 0.142	0.627 ± 0.135	0.518 ± 0.100	0.520 ± 0.092
CO2	MASE	0.368 ± 0.098	0.540 ± 0.227	0.524 ± 0.169	0.364 ± 0.095	0.351 ± 0.105
	RMSSE	0.333 ± 0.087	0.473 ± 0.191	0.458 ± 0.144	0.335 ± 0.085	0.326 ± 0.095
	MSIS	-	8.990 ± 3.892	8.458 ± 2.895	6.517 ± 1.688	6.009 ± 1.820
	wQL	-	0.173 ± 0.068	0.166 ± 0.055	0.125 ± 0.029	0.115 ± 0.032
Light	MASE	0.581 ± 0.068	0.315 ± 0.077	0.351 ± 0.146	0.242 ± 0.043	0.243 ± 0.040
	RMSSE	0.602 ± 0.058	0.336 ± 0.067	0.369 ± 0.123	0.284 ± 0.045	0.287 ± 0.041
	MSIS	-	5.057 ± 1.309	5.555 ± 3.271	3.961 ± 0.736	3.980 ± 0.717
	wQL	-	0.418 ± 0.112	0.465 ± 0.235	0.336 ± 0.068	0.331 ± 0.057
HVAC	MASE	0.466 ± 0.395	0.376 ± 0.283	0.431 ± 0.345	0.342 ± 0.267	0.327 ± 0.257
	RMSSE	0.448 ± 0.367	0.342 ± 0.240	0.371 ± 0.274	0.325 ± 0.233	0.307 ± 0.221
	MSIS	-	6.289 ± 4.537	6.224 ± 4.773	5.542 ± 4.146	5.342 ± 4.056
	wQL	-	0.867 ± 0.656	0.922 ± 0.708	0.811 ± 0.647	0.762 ± 0.613

Table 6: Comparison of forecasting accuracy (**3 months**).

Dataset	Metric	Deep Forecasting Models			Fine-tuned TSFMs	
		N-BEATS	DeepAR	TFT	Chronos +FullFT	Chronos +LoRA
Occ	MASE	0.322 ± 0.056	0.265 ± 0.046	0.274 ± 0.048	0.235 ± 0.037	0.229 ± 0.042
	RMSSE	0.325 ± 0.053	0.273 ± 0.042	0.280 ± 0.045	0.250 ± 0.037	0.242 ± 0.041
	MSIS	-	4.236 ± 0.920	3.925 ± 0.886	3.801 ± 0.724	3.577 ± 0.815
	wQL	-	0.439 ± 0.077	0.431 ± 0.079	0.434 ± 0.069	0.402 ± 0.074
CO2	MASE	0.337 ± 0.088	0.478 ± 0.158	0.378 ± 0.104	0.316 ± 0.081	0.306 ± 0.074
	RMSSE	0.306 ± 0.079	0.427 ± 0.141	0.340 ± 0.093	0.293 ± 0.074	0.287 ± 0.069
	MSIS	-	7.750 ± 2.859	5.265 ± 1.728	5.427 ± 1.377	4.543 ± 1.161
	wQL	-	0.146 ± 0.060	0.103 ± 0.033	0.104 ± 0.025	0.087 ± 0.021
Light	MASE	0.625 ± 0.063	0.283 ± 0.053	0.269 ± 0.080	0.215 ± 0.036	0.215 ± 0.035
	RMSSE	0.639 ± 0.058	0.304 ± 0.050	0.293 ± 0.070	0.256 ± 0.038	0.250 ± 0.037
	MSIS	-	4.467 ± 1.174	3.788 ± 0.986	3.376 ± 0.698	3.102 ± 0.659
	wQL	-	0.326 ± 0.072	0.309 ± 0.104	0.282 ± 0.052	0.254 ± 0.047
HVAC	MASE	0.512 ± 0.458	0.414 ± 0.346	0.400 ± 0.297	0.315 ± 0.244	0.320 ± 0.239
	RMSSE	0.479 ± 0.404	0.384 ± 0.307	0.352 ± 0.249	0.301 ± 0.216	0.299 ± 0.217
	MSIS	-	7.594 ± 6.631	5.383 ± 3.911	5.012 ± 3.790	4.995 ± 4.028
	wQL	-	0.935 ± 0.810	0.802 ± 0.584	0.716 ± 0.570	0.698 ± 0.564

Table 7: Comparison of forecasting accuracy (**6 months**).

Dataset	Metric	Deep Forecasting Models			Fine-tuned TSFMs	
		N-BEATS	DEEPAR	TFT	CHRONOS +FullFT	CHRONOS +LoRA
Occ	MASE	0.329 ± 0.060	0.271 ± 0.054	0.246 ± 0.045	0.230 ± 0.042	0.227 ± 0.044
	RMSSE	0.332 ± 0.058	0.278 ± 0.049	0.255 ± 0.043	0.245 ± 0.041	0.240 ± 0.043
	MSIS	-	4.485 ± 1.137	3.433 ± 0.874	3.675 ± 0.823	3.527 ± 0.877
	wQL	-	0.439 ± 0.081	0.385 ± 0.073	0.419 ± 0.078	0.387 ± 0.075
CO2	MASE	0.337 ± 0.086	0.482 ± 0.149	0.332 ± 0.090	0.319 ± 0.081	0.299 ± 0.074
	RMSSE	0.307 ± 0.077	0.431 ± 0.129	0.303 ± 0.081	0.296 ± 0.073	0.281 ± 0.068
	MSIS	-	7.830 ± 2.841	4.207 ± 1.201	5.300 ± 1.403	4.258 ± 1.075
	wQL	-	0.145 ± 0.053	0.087 ± 0.023	0.101 ± 0.026	0.080 ± 0.020
Light	MASE	0.638 ± 0.068	0.271 ± 0.042	0.243 ± 0.048	0.209 ± 0.035	0.220 ± 0.043
	RMSSE	0.649 ± 0.064	0.292 ± 0.038	0.269 ± 0.043	0.249 ± 0.038	0.250 ± 0.042
	MSIS	-	4.080 ± 0.736	3.255 ± 0.873	3.178 ± 0.700	3.125 ± 0.843
	wQL	-	0.309 ± 0.054	0.270 ± 0.054	0.263 ± 0.050	0.249 ± 0.054
HVAC	MASE	0.525 ± 0.462	0.459 ± 0.409	0.355 ± 0.256	0.323 ± 0.243	0.361 ± 0.280
	RMSSE	0.487 ± 0.406	0.419 ± 0.353	0.327 ± 0.229	0.305 ± 0.216	0.329 ± 0.248
	MSIS	-	8.947 ± 8.572	5.095 ± 3.728	5.133 ± 3.846	5.743 ± 5.047
	wQL	-	1.012 ± 0.884	0.706 ± 0.492	0.717 ± 0.558	0.790 ± 0.646

Table 8: Comparison of forecasting accuracy (**9 months**).

Dataset	Metric	Deep Forecasting Models			Fine-tuned TSFMs	
		N-BEATS	DEEPAR	TFT	CHRONOS +FullFT	CHRONOS +LoRA
Occ	MASE	0.327 ± 0.059	0.265 ± 0.044	0.242 ± 0.041	0.227 ± 0.041	0.228 ± 0.044
	RMSSE	0.331 ± 0.056	0.273 ± 0.040	0.252 ± 0.039	0.242 ± 0.041	0.240 ± 0.044
	MSIS	-	4.318 ± 0.786	3.456 ± 0.718	3.602 ± 0.801	3.515 ± 0.834
	wQL	-	0.432 ± 0.068	0.379 ± 0.065	0.406 ± 0.076	0.386 ± 0.076
CO2	MASE	0.342 ± 0.092	0.444 ± 0.189	0.327 ± 0.086	0.312 ± 0.082	0.299 ± 0.079
	RMSSE	0.312 ± 0.083	0.398 ± 0.164	0.299 ± 0.079	0.289 ± 0.075	0.282 ± 0.073
	MSIS	-	6.270 ± 3.144	4.333 ± 1.221	4.967 ± 1.339	4.193 ± 1.114
	wQL	-	0.130 ± 0.058	0.086 ± 0.021	0.095 ± 0.024	0.079 ± 0.021
Light	MASE	0.644 ± 0.061	0.272 ± 0.040	0.237 ± 0.046	0.210 ± 0.039	0.222 ± 0.047
	RMSSE	0.654 ± 0.057	0.293 ± 0.040	0.262 ± 0.042	0.246 ± 0.040	0.252 ± 0.044
	MSIS	-	4.127 ± 0.940	3.276 ± 0.706	3.104 ± 0.712	3.136 ± 0.819
	wQL	-	0.311 ± 0.052	0.265 ± 0.054	0.255 ± 0.055	0.249 ± 0.057
HVAC	MASE	0.526 ± 0.461	0.451 ± 0.396	0.354 ± 0.268	0.316 ± 0.239	0.368 ± 0.290
	RMSSE	0.487 ± 0.406	0.412 ± 0.343	0.322 ± 0.231	0.298 ± 0.213	0.330 ± 0.255
	MSIS	-	8.626 ± 7.904	4.946 ± 3.984	4.913 ± 3.926	5.746 ± 5.097
	wQL	-	0.997 ± 0.876	0.681 ± 0.505	0.683 ± 0.554	0.789 ± 0.669

Table 9: Comparison of forecasting accuracy (**11 months**).

Dataset	Metric	Deep Forecasting Models			Fine-tuned TSFMs	
		N-BEATS	DEEPAR	TFT	CHRONOS +FullFT	CHRONOS +LoRA
Occ	MASE	0.331 ± 0.060	0.265 ± 0.053	0.245 ± 0.049	0.226 ± 0.040	0.228 ± 0.045
	RMSSE	0.333 ± 0.057	0.273 ± 0.049	0.253 ± 0.046	0.240 ± 0.040	0.240 ± 0.043
	MSIS	-	4.252 ± 1.127	3.617 ± 0.967	3.561 ± 0.802	3.529 ± 0.854
	wQL	-	0.428 ± 0.081	0.380 ± 0.074	0.398 ± 0.074	0.385 ± 0.075
CO2	MASE	0.341 ± 0.089	0.403 ± 0.116	0.328 ± 0.096	0.310 ± 0.079	0.301 ± 0.079
	RMSSE	0.311 ± 0.080	0.363 ± 0.101	0.298 ± 0.086	0.289 ± 0.073	0.284 ± 0.072
	MSIS	-	6.415 ± 2.494	4.535 ± 1.196	4.795 ± 1.310	4.196 ± 1.118
	wQL	-	0.117 ± 0.036	0.086 ± 0.023	0.092 ± 0.023	0.080 ± 0.021
Light	MASE	0.649 ± 0.071	0.275 ± 0.042	0.241 ± 0.057	0.207 ± 0.036	0.224 ± 0.049
	RMSSE	0.658 ± 0.066	0.295 ± 0.040	0.267 ± 0.052	0.243 ± 0.038	0.254 ± 0.046
	MSIS	-	3.934 ± 0.940	3.155 ± 0.829	3.029 ± 0.677	3.182 ± 0.859
	wQL	-	0.311 ± 0.053	0.270 ± 0.065	0.246 ± 0.048	0.251 ± 0.059
HVAC	MASE	0.529 ± 0.464	0.453 ± 0.383	0.355 ± 0.266	0.318 ± 0.239	0.372 ± 0.293
	RMSSE	0.489 ± 0.409	0.413 ± 0.334	0.324 ± 0.233	0.297 ± 0.212	0.333 ± 0.257
	MSIS	-	8.325 ± 7.131	5.320 ± 4.065	4.914 ± 4.002	5.737 ± 5.142
	wQL	-	0.991 ± 0.848	0.682 ± 0.500	0.684 ± 0.557	0.796 ± 0.676

Table 10: Comparison of forecasting accuracy between TFT and TSFMs on *unseen* zone. The model is trained on a zone on the 2nd floor, then tested on another zone on the 4th floor.

Dataset	Metric	TFT	Fine-tuned TSFMs		Dataset	Metric	TFT	Fine-tuned TSFMs	
			CHRONOS +FullFT	CHRONOS +LoRA				CHRONOS +Full-FT	CHRONOS +LoRA
Occ	MASE	0.250 ± 0.077	0.224 ± 0.061	0.213 ± 0.064	Light	MASE	0.225 ± 0.054	0.223 ± 0.054	0.218 ± 0.048
	RMSSE	0.260 ± 0.073	0.242 ± 0.062	0.229 ± 0.063		RMSSE	0.256 ± 0.050	0.261 ± 0.055	0.249 ± 0.049
	MSIS	3.799 ± 1.144	3.417 ± 1.241	3.117 ± 1.230		MSIS	3.495 ± 0.804	3.423 ± 0.981	3.063 ± 0.707
	wQL	0.413 ± 0.112	0.409 ± 0.114	0.366 ± 0.117		wQL	0.256 ± 0.055	0.299 ± 0.074	0.261 ± 0.059
CO2	MASE	0.348 ± 0.081	0.294 ± 0.063	0.285 ± 0.057	HVAC	MASE	0.262 ± 0.288	0.216 ± 0.211	0.231 ± 0.254
	RMSSE	0.309 ± 0.070	0.264 ± 0.059	0.259 ± 0.054		RMSSE	0.256 ± 0.273	0.227 ± 0.242	0.225 ± 0.262
	MSIS	5.314 ± 1.177	5.129 ± 1.056	4.169 ± 0.679		MSIS	3.553 ± 3.418	3.398 ± 3.711	2.928 ± 4.215
	wQL	0.079 ± 0.027	0.081 ± 0.025	0.065 ± 0.018		wQL	0.486 ± 0.450	0.453 ± 0.373	0.421 ± 0.402