ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

A comparison of open time-series foundation models for industrial manufacturing applications

Can Calisir and Simon Leszek *

Technische Universität Berlin - Machine Learning Group
Straße des 17. Juni 135, 10623 Berlin - Germany

Abstract. Large-scale, pre-trained foundation models have recently been introduced for time-series modeling. While typically evaluated on broad forecasting benchmarks, we assess their suitability for industrial manufacturing. We benchmark three open-source time series foundation models (TSFMs) on two representative datasets: steel-plant energy consumption and computer numerical control (CNC) milling spindle current. In the structured, pattern-driven steel setting, TSFMs consistently outperform classical baselines, even without task-specific training. In contrast, the highly dynamic CNC process reveals limited TSFM gains without domain-specific signals, with simple models excelling once control covariates are provided. These results highlight both the promise and current limitations of TSFMs for real-world industrial applications.

1 Introduction

Foundation models – large-scale models trained on diverse datasets using self-supervised learning – have become a central paradigm in machine learning [1]. Initially driven by advances in large language models (LLMs), they have been extended to vision and multi-modal domains, achieving strong performance across downstream tasks by learning generalizable representations providing pre-trained backbones for various domains. Time-series modeling is now benefiting from this shift, evolving from traditional statistical methods (e.g., ARIMA) and deep learning architectures (e.g., LSTMs, CNNs) towards large-scale time-series foundation models (TSFMs) [2, 3, 4, 5]. Pre-trained on extensive, diverse corpora using self-supervised objectives [6], these models aim to capture universal temporal patterns, facilitating robust transfer learning across forecasting and anomaly detection tasks. The recent emergence of open-source variants is particularly significant, lowering adoption barriers for both academic and industrial applications.

This development holds particular promise for industrial applications, where data from sensors, controllers and monitoring systems is abundant yet often lacks labels. Manufacturing relies heavily on such time-series data for critical tasks like predictive maintenance, fault detection and forecasting [7]. Foundation models offer a path to streamline these operations by minimizing manual feature

*This work was funded by the German Federal Ministry for Research, Technology and Space through the project RIESIQ (under ref 16IS24087A).

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

engineering and enabling robust transfer learning across diverse equipment and operating conditions, a crucial capability when a single machine produces varying product lines. In this work, we evaluate three selected open-source TSFMs on time-series forecasting using two manufacturing datasets to assess whether they can overcome the costs of developing custom models for heterogeneous systems [8] and meet specific industrial requirements regarding accuracy and efficiency.[1]

2 Experimental Setting

In the following we present models and datasets for our TSFM evaluation. Across all settings we compare zero-shot performance and fine-tuning strategies, including full fine-tuning and linear probing, where only the final prediction head is trained. Furthermore, we distinguish between *univariate* forecasting, where only past values of the target variable are available, *past covariates*, where additional covariates up to the time of prediction are included, and *known covariates*, which are available at prediction time for the prediction horizon. In the *known covariates* setting, regression baselines treat data points independently, utilizing only the covariates at time t as input (without the target variable), whereas foundation models retain their sequential nature. We evaluate the forecasting performance by computing the mean absolute error (MAE) on unseen test data.

2.1 Models

Chronos 2 [3]: Chronos casts time-series forecasting as a language modeling task. We utilize the 120M-parameter Chronos-2 (Base) , which employs a group attention mechanism to aggregate information across related series. This architecture is able to model multivariate dependencies and supports both past and known future covariates.

MOMENT [4]: MOMENT is an open-source family of foundation models built on a transformer encoder architecture and pre-trained on a large, multi-domain corpus. It learns general-purpose temporal representations via a self-supervised masked prediction task, reconstructing randomly masked patches to capture underlying patterns. We employ the 385M-parameter MOMENT-Large version, adapting it via fine-tuning since its randomly initialized linear head rules out zero-shot use. Our experiments utilize full fine-tuning or linear probing (freezing the encoder) for domain adaptation.

GTT [5]: GTT employs a transformer-based encoder architecture applying multi-scale attention blocks, temporal embeddings and partitioned computations to efficiently capture dependencies across varying time resolutions while maintaining stability through residual connections and normalization. Notably, the model was pre-trained on in-house manufacturing-specific data, among others.

[1] Supplementary data, including detailed results and model hyperparameters, is available at: `https://github.com/CanCalisir/tsfm-industrial-benchmark`

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

We use the 57M-parameter GTT-Large model, evaluating it under zero-shot, linear probing and fine-tuning settings.

Baselines: We compare the TSFMs against a set of baselines including a mean-predictor (*Average*) as the simplest statistical benchmark and *AutoARIMA* [9] as a more elaborate classical baseline. Furthermore, we use a basic Multi-Layer Perceptron (*MLP*) and *DLinear* [10], which decomposes time series for efficient linear processing. We also employ CatBoost (*CAT*) [11], a gradient boosted decision tree algorithm, as an additional classical machine learning model.

2.2 Datasets

Steel Industry Dataset [12]: The steel industry dataset includes multivariate time series data gathered from a steel manufacturing plant. It spans a full year of operations with 15-minute resolution. We adopt a sliding-window approach with a window length of 512 and a stride of 96 to predict energy usage for the following day (96 time steps), as its accurate forecasting is vital for operational cost optimization and demand-side management. The dataset is split (66%/17%/17%) chronologically. As known covariates, we use cyclic sine/cosine encodings of hour-of-day and day-of-week to capture predictable temporal consumption patterns.

CNC Mill Tool Wear [13]: This dataset comprises multivariate time series from 18 experiments where a CNC machine mills S-shaped contours into wax blocks, each lasting between 46.2 s and 231.4 s (10 Hz sampling rate). We predict spindle current because it directly correlates with tool load and wear, serving as a critical indicator for process stability and predictive maintenance [14]. We adopt a sliding-window approach with a window length of 512, a prediction horizon of 50 and a stride of 50 to construct supervised forecasting samples. Before modeling, machine idle times were removed and the data was split into training (58%), validation (23%) and test (19%) sets, ensuring even representation of all process configurations and sufficient data volume for robust evaluation. In multivariate scenarios, we incorporate past covariates (actual velocities and accelerations) and known covariates (process parameters from the machine control, such as planned velocities, accelerations, along with feed rate and clamp pressure).

3 Results

Fig.1 summarizes the forecasting performance of the selected models across both datasets and input-feature configurations and also reports their parameter counts, allowing us to relate accuracy to model size.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

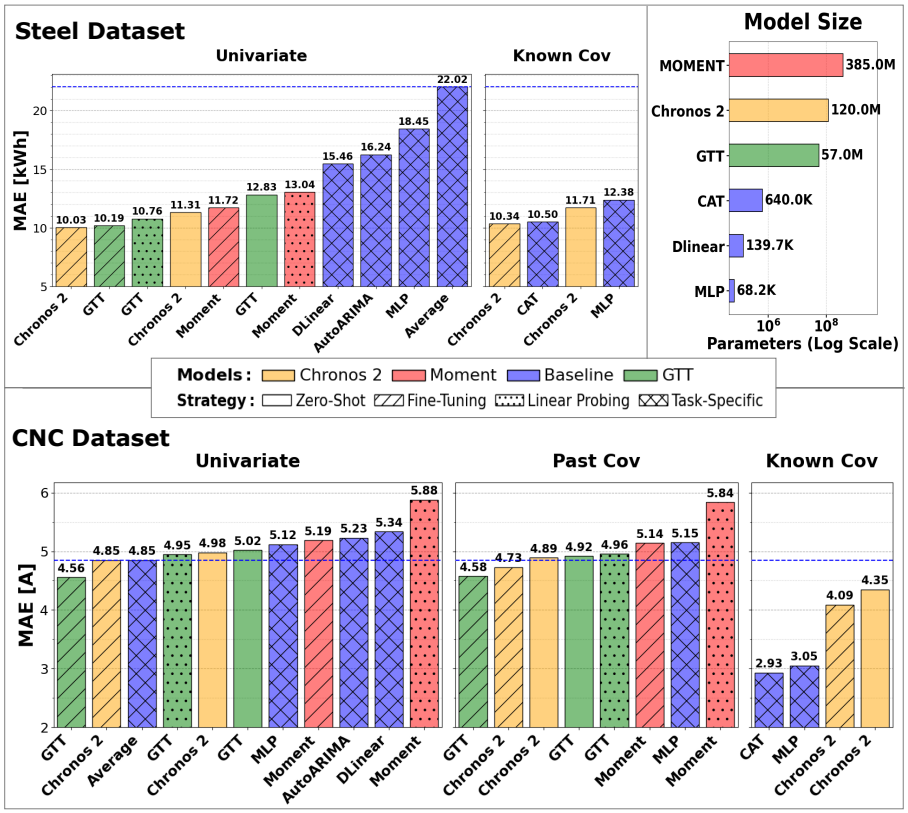

Fig. 1: Results for the steel (top left) and CNC (bottom) dataset by input features. Comparison by MAE, lower is better. Color encodes model family, patterns training strategies and overall sorting is based on performance. Dashed blue line is mean-predictor performance for reference. The top right panel reports the number of parameters (log scale) for the models

3.1 Steel Industry Data

Fig.1 (top) displays the results (MAE) for the steel industry dataset. In the univariate setting, TSFMs consistently outperform baselines, led by fine-tuned Chronos 2 (10.03) and GTT (10.19). Even zero-shot Chronos 2 (11.31) and GTT (12.83) surpass the strongest baselines (DLinear, AutoARIMA). MOMENT improves via fine-tuning (11.72) over linear probing (13.04), justifying backbone updates. With known covariates, CAT (10.50) and MLP (12.38) are competitive, though Chronos 2 (10.34) still leads this configuration. Notably, the best univariate TSFM (10.03) outperforms even covariate-informed models, demonstrating that TSFMs effectively capture temporal patterns from history alone.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

3.2 CNC Mill Tool Wear Data

Fig.1 (bottom) presents the CNC mill results (MAE). In the univariate setting, fine-tuned GTT is the only model to outperform the mean-predictor (4.56 vs. 4.85), while fine-tuned Chronos 2 matches it, improving over its zero-shot baseline (4.98). Adding past covariates provides slight gains for fine-tuned Chronos 2 (4.73) but not for GTT. In the known covariates setting, domain adaptation remains beneficial: fine-tuned Chronos 2 achieves 4.09 compared to 4.35 in zero-shot mode. However, despite these improvements, it trails behind classical baselines like CAT (2.93).

4 Discussion & Outlook

We compared three open-source TSFMs on two manufacturing datasets to evaluate their suitability for industrial use. The datasets represent distinct regimes. The steel case study involves long-horizon forecasting dominated by regular temporal patterns—essentially a classic univariate time-series problem well-suited to TSFMs. Here, all TSFMs outperformed classical baselines on the univariate prediction task, notably also in the zero-shot regime. This highlights the methods' potential to represent highly versatile tools in such settings. Whether the additional computational cost justifies the improvements over classical methods using temporal (known) features, which also performed well, depends on the individual application case. As shown by the parameter counts in Fig.1, the TSFMs are several orders of magnitude larger than the classical baselines, and even within the TSFM family there is substantial variation (57M-385M parameters), so any accuracy gains must be weighed against their higher memory footprint and potential inference latency. The CNC task, by contrast, is far more dynamic and control-based (known) covariates are essential for accurate modeling of the process. In the univariate setting, only GTT—in part trained on in-house industrial time-series—beats the mean-predictor, suggesting that domain-focused training data may be key to unlocking the full potential of TSFMs in industrial applications. While Chronos 2 effectively leverages control signals to reduce error, it still falls short of simple regressors like CAT and MLP. This suggests that for high-frequency control loops, lightweight baselines could offer a superior trade-off between accuracy and computational cost compared to foundation models. In such latency-critical settings, the computational complexity of foundation models may preclude their deployment despite their architectural advancements.

In the future, TSFMs should be evaluated on more diverse datasets and tasks, similar to [15], to assess whether our findings generalize to other industrial processes. Moreover, additional industrial-specific requirements, such as robustness to noise and missing data as well as the ability to handle irregular sampling should be evaluated. Finally, future work should investigate the transparency of time-series-specific foundation models, where model complexity, the regression nature of the task [16, 17], and the modality itself [18] pose challenges.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

References

[1] R. Bommasani, D. A. Hudson *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2180.07258*, 2022.

[2] Q. Wen, T. Zhou *et al.*, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.

[3] A. F. Ansari, O. Shchur *et al.*, "Chronos-2: From univariate to universal forecasting," 2025. [Online]. Available: https://arxiv.org/abs/2510.15821

[4] M. Goswami, K. Szafer *et al.*, "Moment: A family of open time-series foundation models," in *International Conference on Machine Learning*, 2024.

[5] C. Feng, L. Huang *et al.*, "General time transformer: an encoder-only foundation model for zero-shot multivariate time series forecasting," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 3757–3761.

[6] Y. Liang, H. Wen *et al.*, "Foundation models for time series analysis: A tutorial and survey," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, ser. KDD '24. ACM, Aug. 2024, pp. 6555–6565.

[7] L. Ren, Z. Jia *et al.*, "Deep learning for time-series prediction in iiot: progress, challenges, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 35, no. 11, pp. 15 072–15 091, 2023.

[8] J. Wang, Y. Ma *et al.*, "Deep learning for smart manufacturing: Methods and applications," *Journal of Manufacturing Systems*, vol. 48, pp. 144–156, 2018.

[9] R. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for r," *Journal of Statistical Software*, vol. 26, 07 2008.

[10] A. Zeng, M. Chen *et al.*, "Are transformers effective for time series forecasting?" *arXiv preprint arXiv:2205.13504*, 2022.

[11] L. Prokhorenkova, G. Gusev *et al.*, "Catboost: unbiased boosting with categorical features," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017, pp. 6638–6648.

[12] S. V. E, C. Shin *et al.*, "Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city," *Building Research & Information*, vol. 49, pp. 127 – 143, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:224916577

[13] S. Suresh, "Tool wear detection in CNC mill," https://www.kaggle.com/datasets/shasun/tool-wear-detection-in-cnc-mill, 2020, accessed: 2025-10-06.

[14] R. Teti, K. Jemielniak *et al.*, "Advanced monitoring of machining operations," *CIRP Annals*, vol. 59, no. 2, pp. 717–739, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0007850610001976

[15] M. A. Farahani, F. E. Kalach *et al.*, "Time-series forecasting in smart manufacturing systems: An experimental evaluation of the state-of-the-art algorithms," *Robotics and Computer-Integrated Manufacturing*, vol. 95, p. 103010, 2025.

[16] S. Letzgus, P. Wagner *et al.*, "Toward explainable artificial intelligence for regression models: A methodological perspective," *IEEE Signal Processing Magazine*, vol. 39, no. 4, pp. 40–58, 2022.

[17] S. Letzgus, K.-R. Müller *et al.*, "Xpertai: Uncovering regression model strategies for sub-manifolds," in *Explainable Artificial Intelligence. xAI 2025*, ser. Communications in Computer and Information Science, R. Guidotti, U. Schmid *et al.*, Eds. Springer, Cham, 2025, vol. 2578.

[18] J. Vielhaben, S. Lapuschkin *et al.*, "Explainable AI for time series via virtual inspection layers," *Pattern Recognition*, p. 110309, 2024.