

Comparison of Time-Series Foundation Models on power plant emission data

SangJoon Lee

Electronics and Telecommunications Research Institute (ETRI)

Gwangju, Republic of Korea

sjlee85@etri.re.kr

Seok Kap Ko

ETRI

Gwangju, Republic of Korea

softgear@etri.re.kr

Abstract—Time-series forecasting is an important issue in the power plant industry, where its application can vary from pre-emptive maintenance to demand prediction to emission control. However, due to the lack of quality data and environmental conditions of different facilities, such forecasting has generally been limited to elementary statistical analysis. The recent development of the time-series foundation model (TSFM) shows promising zero-shot forecasting quality, enabling a plug-and-play AI model that can be utilized by facilities with a minimal amount of recorded data. We evaluated some of the leading TSFMs in the AI industry to forecast power plant emission data. We compare their zero-shot forecasting accuracy as well as its applicability for anomaly detection and fine-tuning.

Index Terms—Time-series forecasting, Time-series foundation model, Large scale model, Probabilistic forecasting, Power plant emissions

I. INTRODUCTION

With the ongoing global warming issue and the development of artificial intelligence technology, the need for an intelligent energy system has become more important than ever. Many relevant technologies rely on the time-series forecasting of various components; for example, energy demand forecasting is essential for optimal energy distribution [1], renewable energy production forecasting is needed for energy storage system scheduling [2], and the power plant control system can be improved by predicting each sensor data the system collects [3].

However, there are several problems in developing a prediction model. First, many companies that seek to adopt such a system often lack sufficient data to train a relevant model. Moreover, each time-series we want to predict can be highly domain-specific, so a model that can predict one time-series domain can be an overfit to another domain. Traditional methods such as ARIMA and SARIMA can be implemented without an abundant amount of data, but they have limitations on prediction accuracy and struggle to capture the correlation between different streams of data.

Recently, there have been attempts to address this issue using large model-based approaches. Similarly to the now-popular large language models (LLMs), Nixtla introduced

TimeGPT [4], the first large model specifically designed to handle time-series data. Since then, many TSFMs have been introduced, including TimesFM [5] by Google, Chronos [6] by Amazon, and Moirai [7] by Salesforce.

Each TSFM are trained with massive amount of data, varying from nearly 100B to 300B. Because they are trained with such a wide variety of time-series data, they generally tend to have a good zero-shot forecasting capability. Like LLMs, TSFM can be fine-tuned to be specialized to the specific needs, while requiring far less amount of domain-specific data than to develop a model from scratch. Moreover, because its task is restricted to time-series data handling, the size of each model is much smaller than that of an LLM; some of the smallest TSFM consists of less than 10M parameters, and even the largest model has less than 1B parameters.

Time-series is similar to language processing in the sense that it handles a sequence of numerical tokens, so all TSFMs are based on the transformer structure. The major structural difference between different TSFMs is whether the structure is constructed with an encoder, decoder, or both. However, later models were all implemented with a decoder structure, following the current trend in the development of LLM models.

In addition, some later models are implemented with the Mixture-of-Experts(MoE) structure [8], which allows a larger and more accurate model while maintaining a similar inference time. Unfortunately, because of their structural complexity, MoE-based models generally do not support the software development kit for fine-tuning.

One of the key features of TSFMs is their ability to generate sampled forecasts by predicting the parameters of a mixture distribution, enabling probability analysis of the sampled output. Probabilistic forecasting is particularly important for analysis related to the energy system, as anomalies in the system can often lead to safety issues. For example, if the range between the prediction quantiles is large, it would indicate that the prediction model is not very confident about the prediction it provides. And if the monitored value has a safety boundary, checking whether the high/low quantiles of the prediction break the safety borderline can serve as a trigger alert, even if the mean value of the prediction is within the boundary. Moreover, if data recorded at a certain time t_n is outside the quantile range of prediction based on sequential data collected up to time t_{n-1} , such data could be treated as

This work was supported by Electronics and Telecommunications Research Institute(ETRI) grant funded by the Korean government [25ZK1100, Honam region regional industry-based ICT convergence technology advancement support project].

anomalies outside of the distribution (OOD).

II. POWER PLANT EMISSION DATA

To assess the applicability of TSFM on actual energy system data, we use Clean Air Markets Program Data (CAMPD) provided by the United States Environmental Protection Agency [9]. CAMPD provides emission data for the main pollutants for each of the enlisted power plants in the United States.

We collect the two-month-long hourly emission data of the Barry power plant in Alabama, recorded during November and December 2024. We predict the hourly emissions of CO_2 , SO_2 and NO_x for this plant.

It should be noted that TSFMs provide covariate support; for example, if the gross load of the power plant is scheduled beforehand, it can serve as a dynamic numerical covariate component to assist in prediction, and the weekday indicator can also serve as a dynamic categorical covariate.

III. TSFM COMPARISON

For each model listed in Table I, we predict one month of emission data from the given power plant by accumulating predictions of 6 hour data each, each prediction being based on the recorded data of the preceding 168 hours (1 week). Each prediction predicts the three categories of emission gases simultaneously. Specifically, the models being used are TimesFM 2.0 model (500M), Chronos base model (200M), Chronos-Bolt base model (205M), Moirai base model (91.4M) and Moirai-MoE base model (935M). We make the prediction for the 1-hour and 6-hour horizon. Each prediction will generate 100 sample scenarios, and we extract the mean, 0.1 quantile, 0.5 quantile (median) and 0.9 quantile. The prediction mean will be compared with the ground truth data to compute the mean absolute percentage error (MAPE).

Figures 1 to 5 present a comparison between the prediction of the 6 hour horizon of each TSFM and the ground-truth emission data for December 2024. Figure 6 displays the comparison of MAPE between models and emission gases. Each model exhibits similar MAPE for SO_2 and CO_2 emissions, as they follow similar patterns but on different scales. Due to its abrupt spike behavior, all models struggle to accurately predict NO_x emissions.

We also record the number of ground truth data that would be counted as anomalies for each TSFM. If a given ground-truth record is greater than the 0.9 quantile prediction value, it will be counted as a high anomaly, and if it is less than the 0.1 quantile prediction value, it will be counted as a low anomaly. Note that depending on the characteristic of the data we predict, the quantile criterion can be asymmetric; for example, emission data may have to be more sensitive to high anomalies and less sensitive to low anomalies.

Figure 7 compares the number of anomaly counts based on each TSFM. Compared to other models, the original Chronos seems to be over-confident about its prediction, as we get significantly more anomaly marks based on its prediction.

Finally, we measure the execution time of each prediction. This is particularly important if the domain of interest demands

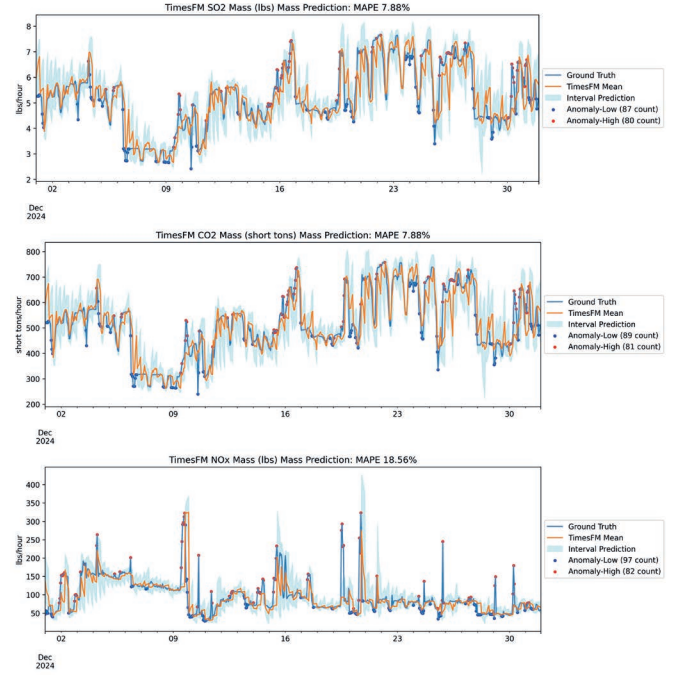


Fig. 1. Comparison of 6-hour horizon TimesFM forecast and ground truth emission data

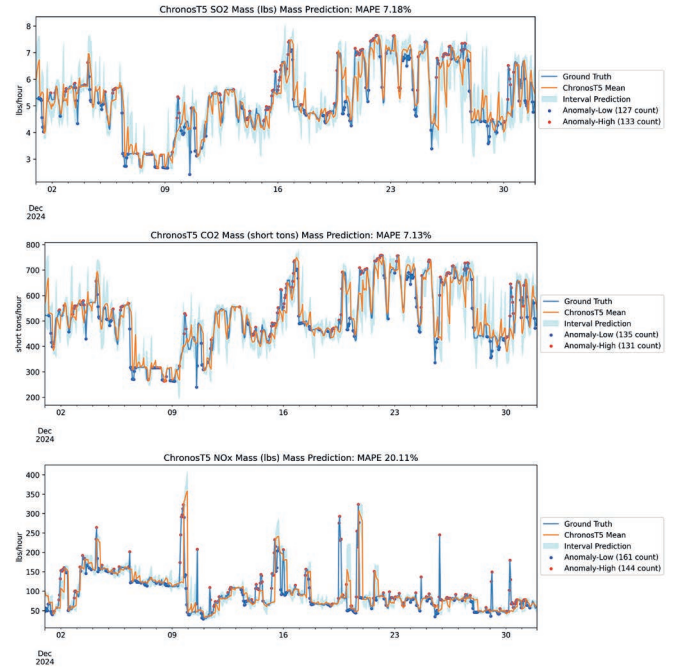


Fig. 2. Comparison of 6-hour horizon Chronos forecast and ground truth emission data

a stream of data frequently. Our machine specification is Intel i9-14900KS CPU, 128GB RAM, NVIDIA RTX 4090 24GB GPU with CUDA 12.6 and Windows 11.

Table II shows the overall execution time for predicting 1-month emission data, of which the prediction window is 6-hours each. For TimesFM and Chronos-Bolt models, we did

TABLE I
PROPERTY COMPARISONS OF TSFM

Model	Training Dataset Size	Model Size	Model Structure	Fine-tune	Open Source
TimeGPT (Nixtla)	100B	Unknown	Encoder-Decoder	Yes	No
TimesFM (Google)	100B+	200/500M	Decoder	Yes	Yes
Chronos (Amazon)	84B	8/20/46/200/710M	Encoder-Decoder	Yes	Yes
Chronos-Bolt (Amazon)	84B	9/21/48/205M	Encoder-Decoder	No	Yes
Moirai (Salesforce)	231B	13.8/91.4/311M	Encoder	Yes	Yes
Moirai-MoE (Salesforce)	231B	117/935M	Decoder	No	Yes

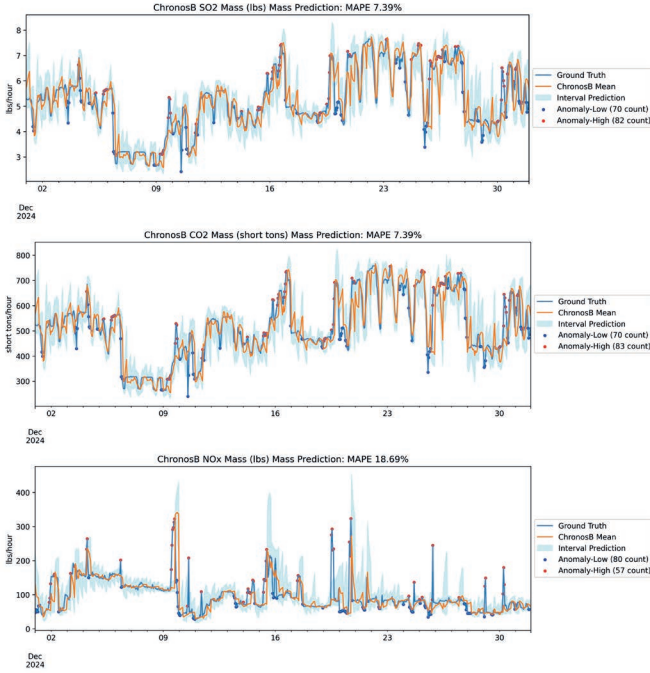


Fig. 3. Comparison of 6-hour horizon Chronos-Bolt forecast and ground truth emission data

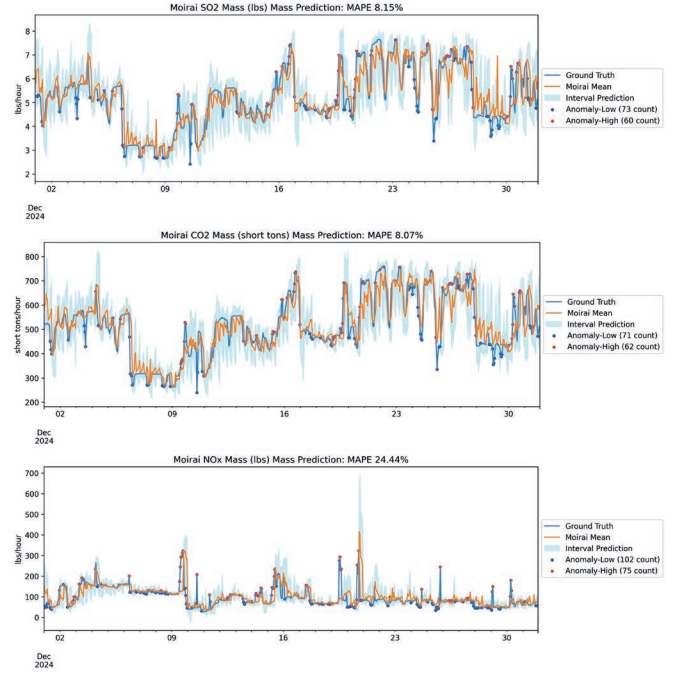


Fig. 4. Comparison of 6-hour horizon Moirai forecast and ground truth emission data

another experiment without using the GPU.

IV. ANALYSIS AND CONCLUSION

We compared the zero-shot prediction capability of hourly power plant emission data for public TSFMs. For more predictable SO_2 and CO_2 emissions, each model showed similar MAPE, and the models of the Moirai family showed worse precision for less predictable NO_x emission.

The Chronos-Bolt model shows the fastest prediction speed while maintaining accuracy and anomaly detection ability similar to other models. However, this particular model does not provide an official fine-tuning SDK, so its applicability may be limited to zero-shot predictions. Also, if the prediction model runs on a GPU-less system, then the TimesFM model could be a better option.

If we want to fine-tune a model to the domain of our interest, Moirai and TimesFM would be preferred; however, it is worth noting that Moirai has more smaller model variants that would require less computation load for fine-tuning.

In the future, we plan to evaluate the trade-offs between various prediction windows and context settings, and fine-tuning results to see the improvement over zero-shot forecasting. We also look forward to extending the application of TSFM to other power plant control data, such as turbine RPM and steam pressure.

ACKNOWLEDGEMENT

This work was supported by Electronics and Telecommunications Research Institute(ETRI) grant funded by the Korean government [25ZK1100, Honam region regional industry-based ICT convergence technology advancement support

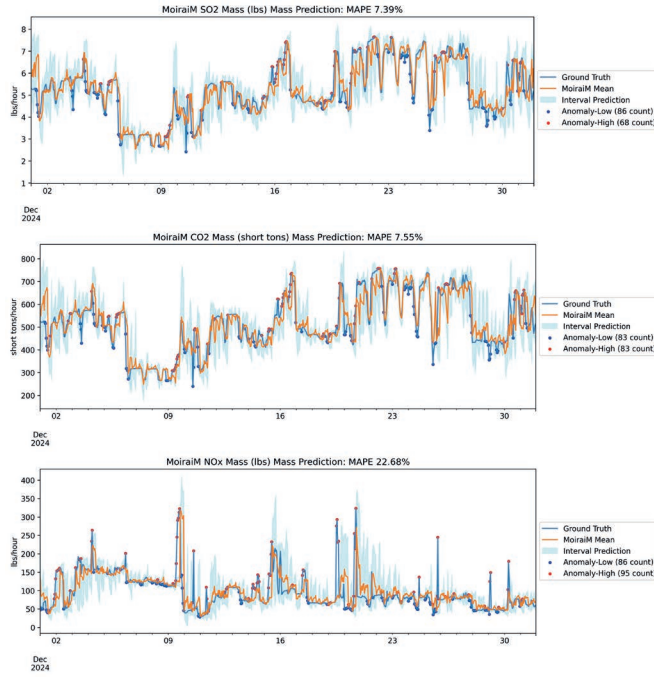


Fig. 5. Comparison of 6-hour horizon Moirai-MoE forecast and ground truth emission data

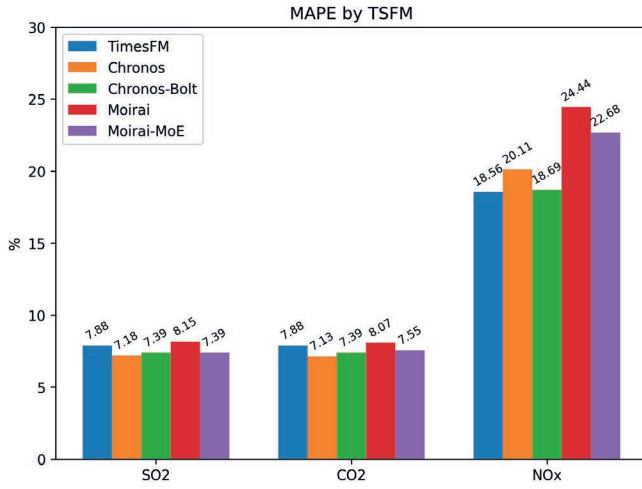


Fig. 6. MAPE comparison of each TSFM

project]. The authors express deep gratitude to the members of the research group for their constructive discussions and support.

REFERENCES

- [1] S. Lee, H. Lee, and S. K. Ko, "Stochastic programming method for volatile resource distribution problem," in *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2024, pp. 936–939.
- [2] S. Lee, Y. M. Hwang, H. Lee, and S. K. Ko, "Optimal ess scheduling method based on probabilistic forecasting," in *The 56th KIEE Summer Conference*. KIEE, 2025, pp. 2054–2055.

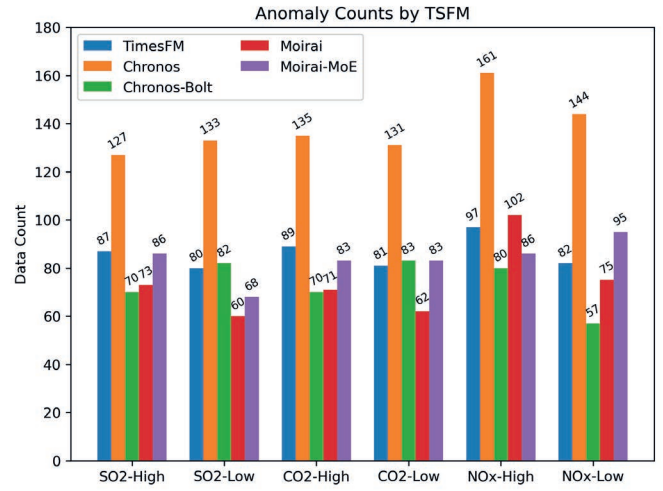


Fig. 7. Anomaly detection of each TSFM

TABLE II
OVERALL PREDICTION TIME OF DEC 2024 EMISSION DATA
(6-HOUR HORIZON FORECAST)

TSFM	Execution time (s)
TimesFM	9.23
TimesFM (CPU Only)	73.38
Chronos	22.98
Chronos-Bolt	5.78
Chronos-Bolt (CPU Only)	368.70
Moirai	8.14
Moirai-MoE	26.72

- [3] S. Lee, H. Lee, and S. K. Ko, "Cost-optimizing ess environmental control based on internal resistance and time-series forecasting of external conditions," *The Transactions of The Korean Institute of Electrical Engineers*, vol. 73, no. 8, pp. 1369–1375, 2024.
- [4] A. Garza and M. Mergenthaler-Canseco, "Timegpt-1," 2023.
- [5] A. Das, W. Kong, R. Sen, and Y. Zhou, "A decoder-only foundation model for time-series forecasting," in *Forty-first International Conference on Machine Learning*, 2024.
- [6] A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. Pineda Arango, S. Kapoor, J. Zschiegner, D. C. Maddix, M. W. Mahoney, K. Torkkola, A. Gordon Wilson, M. Bohlke-Schneider, and Y. Wang, "Chronos: Learning the language of time series," *Transactions on Machine Learning Research*, 2024. [Online]. Available: <https://openreview.net/forum?id=gerNCVqqtR>
- [7] G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo, "Unified training of universal time series forecasting transformers," in *Forty-first International Conference on Machine Learning*, 2024.
- [8] X. Liu, J. Liu, G. Woo, T. Aksu, Y. Liang, R. Zimmermann, C. Liu, S. Savarese, C. Xiong, and D. Sahoo, "Moirai-moe: Empowering time series foundation models with sparse mixture of experts," *arXiv preprint arXiv:2410.10469*, 2024.
- [9] United States Environmental Protection Agency (EPA), "Power sector data," <https://campd.epa.gov>, n.d., washington, DC: Office of Atmospheric Protection, Clean Air and Power Division. Available from EPA's Clean Air Markets Program Data website.