Article

Not peer-reviewed version

Zero-Shot Weekly Load Forecasting for the Early-Stage Operation of Energy Communities

Philipp Wohlgenannt , Lukas Moosbrugger , Fiona Feuerstein , Vinzent Vetter , Elias Eder , Peter Kepplinger *

Posted Date: 1 June 2026

doi: 10.20944/preprints202606.0013.v1

Keywords: energy community; load forecasting; zero-shot; foundation model; early-stage operation

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 June 2026 doi:10.20944/preprints202606.0013.v1

Article

Zero-Shot Weekly Load Forecasting for the Early-Stage Operation of Energy Communities

Philipp Wohlgenannt, Lukas Moosbrugger, Fiona Feurstein, Vinzent Vetter, Elias Eder and Peter Kepplinger *

illwerke vkw Endowed Professorship for Energy Efficiency, Energy Research Centre, Vorarlberg University of Applied Sciences, Hochschulstrasse 1, 6850 Dornbirn, Austria
* Correspondence: peter.kepplinger@fhv.at

Abstract

Accurate load forecasting is essential for the optimal operation of local energy communities. However, newly established communities usually lack sufficient historical data, limiting the use of supervised forecasting and optimization methods. This paper investigates the zero-shot time-series foundation model TiRex for weekly load forecasting in emerging energy communities. After an initial one-week observation period, the model is deployed without task-specific training, feature engineering, or exogenous input data. The approach is evaluated using ten months of real measurements from an energy community consisting of six residential households, one school, and one industrial building, including local photovoltaic generation. Using only historical load data, TiRex generates weekly forecasts in a walk-forward setup that reflects realistic operation. In addition to point forecasts, predictive quantiles are provided to represent forecast uncertainty. Performance is compared with a weekly persistence baseline, where each forecast week equals the load profile of the previous week. Despite requiring no training or external data, TiRex reduces forecasting error by about 20%, achieving a mean RMSE of 1.50 kW versus 1.89 kW for persistence. As an additional benchmark, the transformer-based foundation model Chronos is tested in three model sizes, reaching RMSE values between 1.59 kW and 1.65 kW under identical conditions. The results show that zero-shot forecasting can provide accurate and stable performance under realistic deployment conditions, even with distributed photovoltaic generation and volatile residual profiles. This enables forecasting-based operation from the earliest deployment stage and supports a later transition to supervised deep learning models as more data become available.

Keywords: energy community; load forecasting; zero-shot; foundation model; early-stage operation

---

1. Introduction

Energy communities are becoming an increasingly important element of the decentralized energy transition. By enabling citizens, municipalities, and small enterprises to jointly generate, consume, store, and share electricity, they support renewable energy integration, local value creation, and greater flexibility of distribution grids. Their growing relevance has been reinforced by recent European regulatory frameworks that formally recognize citizen and renewable energy communities (Lowitzsch et al., 2020). To operate such communities efficiently, accurate load forecasting is essential, since reliable demand estimates support energy scheduling, storage management, peer-to-peer coordination, and cost-optimal market participation (Sousa et al., 2019; Seiler et al., 2024).

However, newly established energy communities typically face a cold-start problem. During the first weeks or months of operation, only limited historical smart-meter data are available, while load profiles are often highly variable due to heterogeneous participants, changing occupancy behavior, and local photovoltaic generation. Conventional machine learning and deep learning forecasting approaches usually require sufficiently large task-specific datasets for model training and hyperparameter tuning. As a result, their applicability in early-stage communities is limited, precisely when forecasting would already be valuable for operational planning and decision support (Deb et al., 2017; Moosbrugger et al., 2025).

A promising alternative is provided by pretrained time-series foundation models, which can generate forecasts for unseen datasets without task-specific retraining. Through zero-shot inference, these models can be deployed immediately in data-scarce environments and may therefore overcome the cold-start challenge of emerging energy communities. Recent foundation models such as Chronos (Ansari et al., 2024) and TiRex (Auer et al., 2025a; Auer et al., 2025b) have demonstrated strong generalization capabilities across heterogeneous time-series domainsmaking foundation models a relevant research direction for practical energy applications.

This paper investigates the zero-shot foundation model TiRex for weekly load forecasting in a virtual energy community based on real-world data. The model is evaluated using measured residual load data and compared against a weekly persistence baseline as well as the pretrained foundation model Chronos. In addition to point forecasts, probabilistic predictions are analyzed through predictive quantiles. The results show that zero-shot forecasting can provide a practical and accurate solution for supporting energy community operation from the earliest deployment stage.

The main contributions of this work are:

- A zero-shot forecasting framework for early-stage energy communities using TiRex
- A comparative evaluation against persistence forecasting and Chronos
- The demonstration that pretrained foundation models can enable forecasting-based operation despite limited historical data.

The remainder of this paper is structured as follows. Section 2 presents the dataset, the methodology and the forecasting framework. Section 3 discusses the simulation results. Section 4 outlines limitations and future research directions. Section 5 concludes the paper.

2. Methods

This section describes the dataset, forecasting framework, baseline methods, and evaluation procedure used in this study.

1.1. Dataset

The evaluation is based on ten months of openly available real measurement data (1 h resolution, see Figure 1) used to construct an emerging virtual energy community. The community consists of six residential households, one school, and one industrial building, representing heterogeneous consumption behavior. Local photovoltaic generation is included, and the forecasting target is the aggregated community residual load. The limited historical data reflects an early-stage deployment scenario, making the dataset well suited for evaluating zero-shot forecasting methods. (Open Power System Data, 2020)

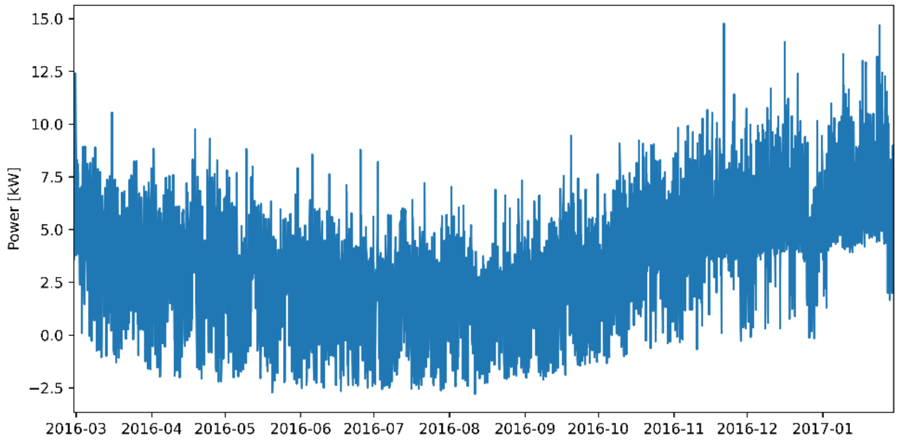


Figure 1. Available data of the virtual energy community over 10 months.

2.2. Forecasting Framework

The overall forecasting framework used in this study is illustrated in Figure 2. Historical residual load measurements from an emerging energy community are used as input to the time-series foundation model TiRex, which generates probabilistic forecasts for the following seven days. The forecasting target is the hourly aggregated community residual load. The forecasting system relies exclusively on historical load measurements and does not require any additional input variables such as weather data or calendar features.

2.3. Forecasting Problem

Let y_t denote the aggregated residual load of the energy community at time step t. The objective is to forecast the hourly future load for a fixed forecast horizon H based on previously observed measurements. In this study, the forecast horizon is set to seven days ($H = 168$ hours). At each forecasting step, the model receives all previously observed load values and generates probabilistic forecasts for the next seven days. The forecasting task can therefore be formulated as

$$\hat{y}_{t+1:t+H} = f(y_{1:t}) \tag{1}$$

where $f(\cdot)$ represents the forecasting model and H denotes the forecast horizon.

2.4. Zero-Shot Forecasting with TiRex

The time-series foundation model TiRex (34 million parameters) is applied in a zero-shot forecasting setting. No exogenous variables such as weather data or calendar features are provided. In contrast to conventional supervised forecasting approaches, the model is used without task-specific training, parameter tuning, or feature engineering. For each forecast step, the model receives the sequence of previously observed load measurements and directly generates probabilistic forecasts for the subsequent seven days. As additional observations become available over time, the context window increases continuously, allowing the model to incorporate progressively more historical information. The model outputs multiple predictive quantiles, enabling the estimation of forecast uncertainty.

2.5. Baseline Methods

To assess forecasting performance, the TiRex model is compared with two baseline ap-proaches. As a simple benchmark, a weekly persistence model is used. In this approach, the load profile of the previous week is directly used as the forecast for the following week. In addition, the transformer-based time-series foundation model Chronos-T5 is evaluated as an alternative zero-shot forecasting approach. To analyze the impact of model capacity, three different model sizes are considered: Chronos with approximately 8 million, 20 million, and 46 million parameters.

2.6. Evaluation Procedure and Metrics

The forecasting models are evaluated using a walk-forward validation scheme to simulate realistic operational deployment conditions. After an initial observation period of one week, the first forecast is generated for the following seven days. Subsequently, the da-taset is extended with newly observed measurements and the next forecast is produced. As a result, the amount of historical context available to the model increases over time, while the forecast horizon remains constant. This procedure is repeated throughout the dataset, resulting in a sequence of weekly forecasts that are evaluated against the corre-sponding ground truth measurements.

Forecasting performance is evaluated using several commonly used error metrics. Let y_t denote the observed load and $\hat{y}_t$ the forecasted load at time step t, with N representing the total number of observations. The forecast bias measures the systematic deviation between forecasts and observations and is defined as

$$Bias = \frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t) \tag{2}$$

The mean absolute error (MAE) evaluates the average magnitude of forecast errors

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|\hat{y}_t - y_t| \tag{3}$$

while the root mean squared error (RMSE) assigns higher weight to larger deviations

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2} \tag{4}$$

The symmetric mean absolute percentage error (sMAPE) is defined as

$$sMAPE = \frac{100}{N}\sum_{t=1}^{N}\frac{|\hat{y}_t - y_t|}{(|y_t| + |\hat{y}_t|)/2} \tag{5}$$

To enable scale-independent comparison, the normalized mean absolute error (nMAE) is calculated relative to the mean observed load

$$nMAE = \frac{MAE}{\bar{y}} \cdot 100 \tag{6}$$

where $\bar{y}$ denotes the mean observed load. The coefficient of determination R^2 evaluates the proportion of variance explained by the model

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2} \tag{7}$$

To analyze systematic prediction behavior, the proportions of overpredictions (OP) and underpredictions (UP) are also reported. Overpredictions correspond to time steps where the predicted value exceeds the observation $(\hat{y}_t > y_t)$, while underpredictions occur when the predicted value is lower than the observation $(\hat{y}_t < y_t)$. The corresponding proportions are calculated as

$$OP = \frac{1}{N}\sum_{t=1}^{N}\mathbb{1}(\hat{y}_t > y_t) \tag{8}$$

$$UP = \frac{1}{N}\sum_{t=1}^{N}\mathbb{1}(\hat{y}_t < y_t) \tag{9}$$

where $\mathbb{1}(\cdot)$ denotes the indicator function.

In addition to point forecast accuracy, the probabilistic quality of the forecasts is evaluated using reliability diagrams and probability integral transform (PIT) histograms. Reliability diagrams assess the calibration of predicted quantiles by comparing nominal and empirical coverage probabilities. The PIT histogram evaluates whether the predictive distributions are statistically consistent with the observed data, where a uniform distribution indicates well-calibrated probabilistic forecasts. Finally, the evolution of forecasting errors across the forecasting horizon is analyzed using the mean absolute error as a function of the forecast horizon. This analysis provides insight into how forecast accuracy changes over the seven-day forecast window and allows the identification of systematic error increases for longer lead times.

3. Results and Discussion

Figure 2 illustrates an example forecast where the model forecasts the residual load of the third week based on the preceding two weeks of historical data. The red line represents the median forecast, while the shaded region indicates the probabilistic forecast interval between the 10th and 90th quantiles. The results show that the model successfully captures the dominant daily patterns of the load profile and provides uncertainty estimates that cover most of the observed values during the forecast horizon.

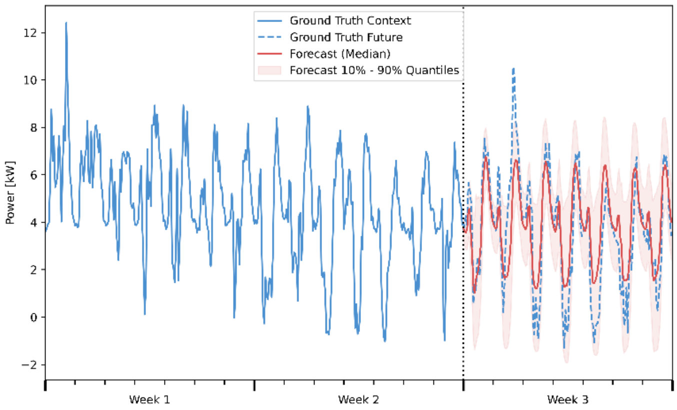
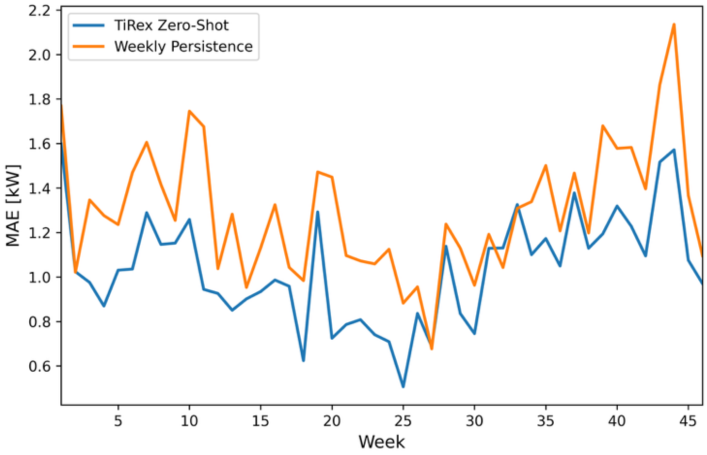


Figure 2. Example of a weekly residual load forecast generated by TiRex using two weeks of historical observations as input context.

Figure 3 shows the weekly MAE of the TiRex zero-shot model compared to a weekly persistence baseline. TiRex consistently achieves lower forecast errors across most weeks, demonstrating its ability to capture temporal patterns beyond simple weekly repetition.



doi:10.20944/preprints202606.0013.v1

Figure 3. Weekly MAE comparison between the TiRex zero-shot model and the weekly persistence baseline.

Table 1 summarizes the overall forecasting performance of the evaluated models.

Table 1. Results of different forecasting methods.

Model	TiRex	Chronos	Chronos	Chronos	Persistence
Parameters	34M	8M	20M	46M	-
Bias (kW)	0.32	0.27	0.21	0.36	0.05
MAE (kW)	1.04	1.13	1.09	1.11	1.30
RMSE (kW)	1.50	1.65	1.59	1.64	1.89
sMAPE (%)	50.20	52.47	52.95	53.13	55.87
nMAE (%)	37.55	33.93	32.92	33.43	48.60
R2	0.47	0.64	0.67	0.65	0.17
OP (%)	54.54	51.66	51.00	56.13	51.26
UP (%)	45.46	48.34	49.00	43.87	48.74

The results show that the TiRex zero-shot model achieves the lowest MAE and RMSE among the considered approaches, outperforming all Chronos variants as well as the persistence baseline. In contrast, the Chronos models achieve slightly better performance in terms of nMAE and significantly better values in terms of R2. However, percentage-based metrics such as sMAPE and nMAE should be interpreted with caution, as the residual load frequently approaches or crosses zero, which can inflate percentage errors. Similarly, the R2 metric may become unstable when the signal frequently remains close to zero. Overall, the results indicate that the evaluated foundation models show comparable performance, with no single model consistently dominating across all evaluation metrics. Instead, each model exhibits specific strengths depending on the chosen performance criterion. An interesting observation is that no clear relationship between model size and forecasting performance can be identified for the Chronos variants. Nevertheless, all foundation models outperform the persistence baseline across all considered metrics except for the forecast bias.

The probabilistic calibration of the TiRex forecasts is evaluated using a reliability diagram and the PIT. As shown in Figure 4, the empirical coverage closely follows the perfectly calibrated diagonal but remains slightly below it across all quantiles, indicating a mild undercoverage.

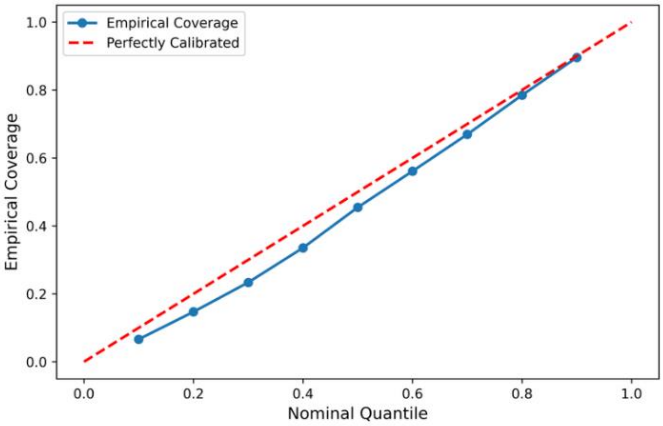


Figure 4. Reliability diagram comparing nominal and empirical quantile coverage of the probabilistic TiRex forecasts.

The PIT histogram (Figure 5) is approximately uniform, suggesting that the predictive distributions are reasonably calibrated without strong systematic bias. Slight deviations from perfect calibration can be observed for lower quantiles, where empirical coverage is marginally below the nominal level. Detailed coverage statistics are provided in Appendix A.

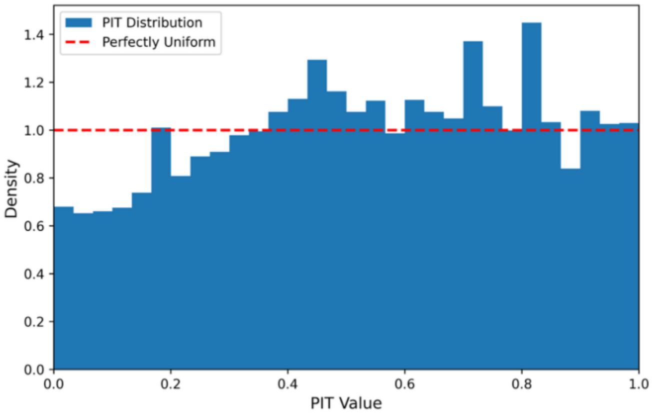


Figure 5. PIT histogram used to assess the calibration quality of the probabilistic forecasts.

Figure 6 illustrates the MAE across the forecast horizon. The error exhibits a pronounced daily pattern corresponding to the daily structure of the load. Importantly, no systematic increase in error can be observed over the seven-day horizon, indicating stable forecasting performance across longer lead times.

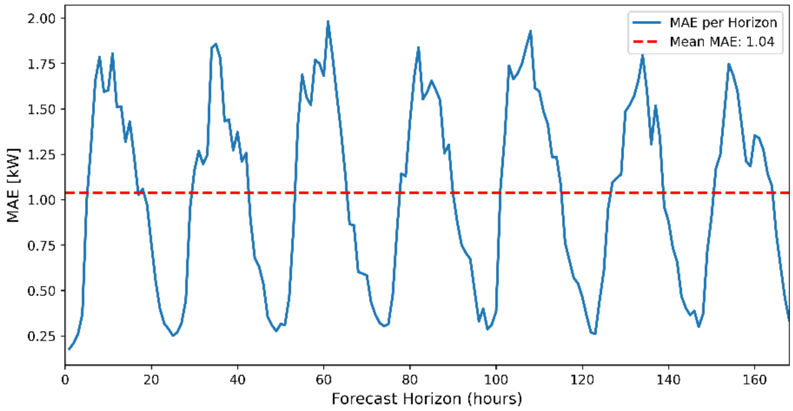


Figure 6. MAE as a function of the forecast horizon over the seven-day forecast window.

4. Limitations and Future Research Directions

While the presented results demonstrate the potential of zero-shot foundation models for load forecasting in early-stage energy communities, several limitations remain. The evaluation is based on data from a single virtual energy community, and further studies are required to assess the transferability of the results to different community structures and operating conditions. In addition, the local photovoltaic generation remains relatively small compared to the overall community demand, limiting the variability introduced by renewable feed-in. Furthermore, the study focuses exclusively on a zero-shot forecasting setup without analyzing when a transition to locally trained models becomes beneficial. Future research could investigate the amount of historical data required for community-specific models to outperform foundation models. Another limitation arises from the deliberately minimal preprocessing required by the proposed approach. The forecasting models rely exclusively on historical load measurements and do not incorporate exogenous variables such as weather data or calendar information. While this enables simple deployment in practical applications, incorporating additional features could potentially improve forecasting performance. However, such feature-based approaches typically require task-specific model training and therefore move away from the zero-shot foundation model paradigm.

Beyond forecasting accuracy, the generated probabilistic forecasts could directly support operational energy management strategies in emerging energy communities. In particular, forecast-based control approaches such as MPC could utilize the predicted residual load and associated uncertainty estimates for battery scheduling, energy trading, and demand-side flexibility coordination. Since the proposed forecasting framework operates without task-specific training or exogenous input data, it can provide operational support already during the earliest deployment stages, where only limited historical measurements are available. Future work could therefore investigate the integration of zero-shot forecasting models into MPC-based energy management systems for real-time community operation.

5. Conclusion

This paper investigated the application of the zero-shot time-series foundation model TiRex for weekly load forecasting in early-stage energy communities with limited histori-cal data. The approach was evaluated using ten months of real measurement data from an emerging energy community comprising six residential households, one school, and one industrial building with local photovoltaic generation. Forecasts were generated in a realistic walk-forward setup with a forecast horizon of seven days. The results demonstrate that accurate load forecasts can be obtained even without task-specific training, feature engineering, or exogenous input data. In particular, TiRex achieves the lowest RMSE among the evaluated models, yielding a mean RMSE of 1.50 kW compared to 1.89 kW for the weekly persistence baseline, corresponding to an error reduction of approximately 20%. As additional benchmarks, three variants of the transformer-based foundation model Chronos (8M, 20M, and 46M parameters) were evaluated, achieving RMSE values between 1.59 kW and 1.65 kW under identical conditions. While the evaluated foundation models show comparable performance overall, they consistently outperform the persistence baseline across most evaluation metrics. These findings highlight the strong potential of zero-shot foundation models for forecasting applications in emerging energy communities, where only limited historical data is available during the early stages of operation. The results indicate that such models can provide reliable load forecasts from the very beginning of community operation, thereby supporting forecasting-based coordination and energy management in early-stage energy communities.

Acknowledgments: Philipp Wohlgenannt was funded by the European Union under the programme “Investment in Growth and Jobs (IBW)/European Regional Development Fund (ERDF) & Just Transition Fund (JTF) Austria 2021–2027” and the State of Vorarlberg. Peter Kepplinger and Lukas Moosbrugger gratefully acknowledge the financial support from the Austrian Research Promotion Agency FFG for the Hub4FlECs project (COIN FFG 898053). AI-based assistance tools (including OpenAI ChatGPT and Google Gemini) were used to support language editing, programming assistance, and manuscript preparation. The authors take full responsibility for the scientific content, methodology, results, and conclusions.

Appendix A

Table 2. TiRex Quantile Coverage.

Quantile	Empirical	Difference
0.10	0.0663	-0.0337
0.20	0.1470	-0.0530
0.30	0.2336	-0.0664
0.40	0.3354	-0.0646
0.50	0.4546	-0.0454
0.60	0.5611	-0.0389
0.70	0.6694	-0.0306
0.80	0.7849	-0.0151

0.90	0.8957	-0.0043

Table 3. TiRex Interval Coverage.

Interval	Empirical	Difference
0.10-0.90	0.8295	+0.0295
0.20-0.80	0.6379	+0.0379
0.30-0.70	0.4358	+0.0358
0.40-0.60	0.2257	+0.0257

References

1. Ansari, A., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S.S., Arango, S.P., Kapoor, S., Zschiegner, J., Maddix, D.C., Mahoney, M.W., Torkkola, K., Gordon Wilson, A., Bohlke-Schneider, M., & Wang, Y. (2024). Chronos: Learning the language of time series. arXiv preprint, arXiv:2403.07815. https://doi.org/10.48550/arXiv.2403.07815
2. Auer, A., Klotz, D., Böck, S., & Hochreiter, S. (2025a). Pre-trained forecasting models: Strong zero-shot feature extractors for time series classification. Proc. of the NeurIPS 2025 Workshop on Recent Advances in Time Series Foundation Models (BERT2S). https://arxiv.org/abs/2510.26777
3. Auer, A., Podest, P., Klotz, D., Böck, S., Klambauer, G., & Hochreiter, S. (2025b). TiRex: Zero-shot forecasting across long and short horizons with enhanced in-context learning. Proc. of the Thirty-Ninth Annual Conference on Neural Information Processing Systems. https://arxiv.org/abs/2505.23719
4. Deb, C., Zhang, F., Yang, J., Lee, S.E., & Shah, K.W. (2017). A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev., 74, 902–924. https://doi.org/10.1016/j.rser.2017.02.085
5. Lowitzsch, J., Hoicka, C.E., & van Tulder, F.J. (2020). Renewable energy communities under the 2019 European Clean Energy Package – Governance model for the energy clusters of the future? Renew. Sustain. Energy Rev., 122, 109489. https://doi.org/10.1016/j.rser.2019.109489
6. Moosbrugger, L., Seiler, V., Wohlgenannt, P., Hegenbart, S., Ristov, S., Eder, E., & Kepplinger, P. (2025). Load forecasting for households and energy communities: Are deep learning models worth the effort? arXiv preprint, arXiv:2501.05000. https://doi.org/10.48550/arXiv.2501.05000
7. Open Power System Data. (2020). Household data. Version 2020-04-15. Detailed household load and solar generation in minutely to hourly resolution. Open Power System Data Platform. https://data.open-power-system-data.org/household_data/2020-04-15
8. Seiler, V., Moosbrugger, L., Huber, G., & Kepplinger, P. (2024). Assessing model predictive control for energy communities' flexibilities. In Intelligente Energie-und Klimastrategien: Energie-Gebäude-Umwelt. Holzhausen. https://doi.org/10.57739/978-3-903207-89-9
9. Sousa, T., Soares, T., Pinson, P., Moret, F., Baroche, T., & Sorin, E. (2019). Peer-to-peer and community-based markets: A comprehensive review. Renew. Sustain. Energy Rev., 104, 367–378. https://doi.org/10.1016/j.rser.2019.01.036