Feasibility of Forecasting Highly Resolved Power Grid Frequency Utilizing Temporal Fusion Transformers

Sebastian Pütz*
Karlsruhe Institute of Technology
Karlsruhe, Germany
sebastian.puetz@kit.edu

Hadeer El Ashhab*
Karlsruhe Institute of Technology
Karlsruhe, Germany
hadeer.elashhab@kit.edu

Matthias Hertel
Karlsruhe Institute of Technology
Karlsruhe, Germany
matthias.hertel@kit.edu

Ralf Mikut
Karlsruhe Institute of Technology
Karlsruhe, Germany
ralf.mikut@kit.edu

Markus Götz
Karlsruhe Institute of Technology
Karlsruhe, Germany
markus.goetz@kit.edu

Veit Hagenmeyer
Karlsruhe Institute of Technology
Karlsruhe, Germany
veit.hagenmeyer@kit.edu

Benjamin Schäfer
Karlsruhe Institute of Technology
Karlsruhe, Germany
benjamin.schaefer@kit.edu

ABSTRACT

As our society moves toward a decarbonized energy system, we need to improve our ability to model, predict, and understand power system behavior and dynamics. The balance between generation and demand on short time scales is reflected by the power grid frequency, making it central to the control of power grids. Hence, an accurate understanding and forecasting of power grid frequency could ease the planning of control actions and thus improve system stability and help save costs. Whether deep learning approaches can provide forecasts of the highly resolved and noisy time series, as they are present in the case of power grid frequency, remains an open question. In this paper, we find that the Temporal Fusion Transformer (TFT) is able to outperform baseline models, while a comparably simple multilayer perceptron is not. By reducing the time resolution of the frequency time series, we investigate and quantify the trade-off between the energy consumption and prediction performance of the TFT. Furthermore, the inclusion of additional exogenous variables (e.g. calendar features, load, or generation) further improves the performance of the TFT. Utilizing the TFT's inherent interpretability, we identify the forecasted load ramp, the current hour, and the current month as the most relevant features.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Applied computing** → *Engineering*; • **Hardware** → **Energy distribution**.

*Both authors contributed equally to this research.

E-Energy '24, June 04–07, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0480-2/24/06
https://doi.org/10.1145/3632775.3661963

KEYWORDS

power grid frequency, time series forecasting, interpretable artificial intelligence, temporal fusion transformer, deep learning

ACM Reference Format:
Sebastian Pütz, Hadeer El Ashhab, Matthias Hertel, Ralf Mikut, Markus Götz, Veit Hagenmeyer, and Benjamin Schäfer. 2024. Feasibility of Forecasting Highly Resolved Power Grid Frequency Utilizing Temporal Fusion Transformers. In *The 15th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '24), June 04–07, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3632775.3661963

1 INTRODUCTION

The mitigation of climate change necessitates a fundamental transformation of our energy system [29]. Power grids play a pivotal role in this transformation and face several challenges. On the one hand, additional consumers are introduced to the system, such as electrical vehicles or heat pumps. These additional actors increase consumer complexity and volatility [3, 4]. On the other hand, generation behavior also becomes more complex and volatile. The massive integration of renewable wind and solar power generation is essential for a future sustainable energy system. However, their variability on long (seasonal) and short (up to subsecond) timescales [12, 25], their location dependence, and their missing inherent inertia are core challenges that need to be tackled for a stable and sustainable power grid [36]. To ensure a stable supply of electrical power, power generation and demand need to be balanced at all times [24]. The power grid frequency reflects this balance and is the central observable to maintain it on short time scales. A frequency value below the reference of 50 or 60 Hz indicates a shortage of generation, while a value above the reference indicates an abundance of generation. Large deviations from the reference frequency necessitate costly control actions to prevent a collapse of the power system. A precise forecast of power grid frequency could help in the planning and the optimized usage of control resources. Hence, building precise forecasts and a solid understanding of the diverse factors influencing power grid frequency is important for

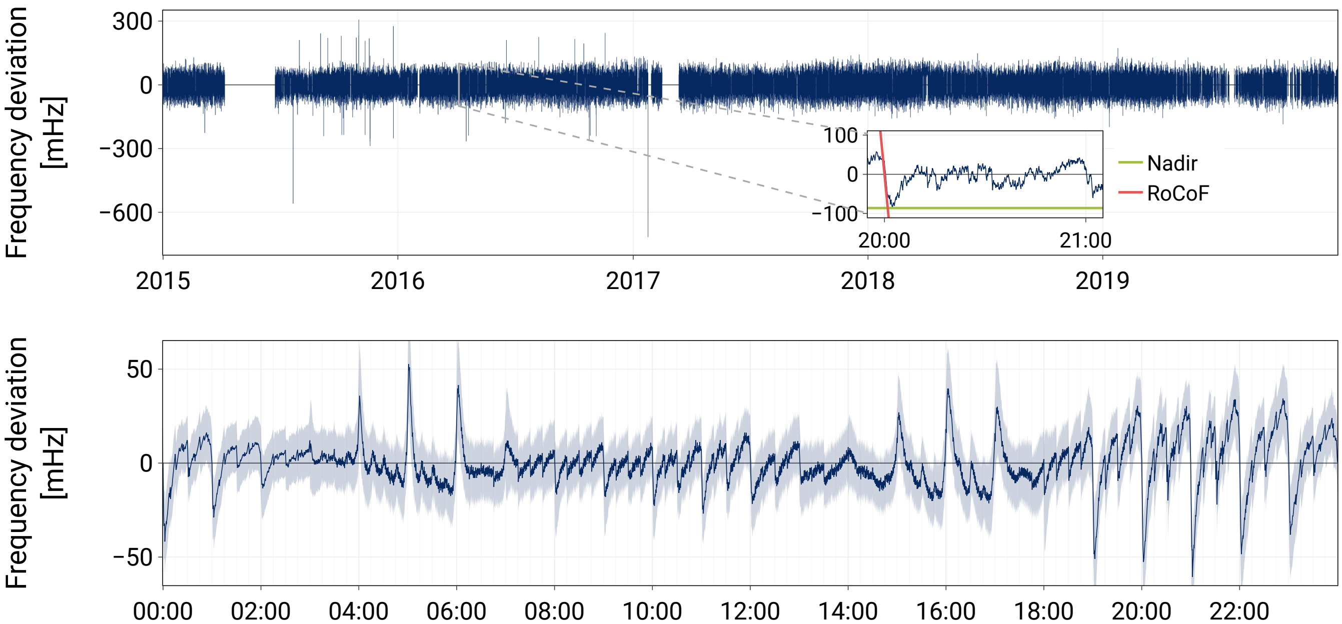

Figure 1: Continental European power grid frequency data. (top) Power grid frequency deviation from the reference frequency (50 Hz) over five years. The inset shows a typical one-hour excerpt. (bottom) The daily average profile of the Continental European power grid frequency which also acts as a benchmark model. The colored area marks the standard deviation.

navigating the challenges posed by the integration of renewable energy sources. Notably, power grid frequency data is more readily available compared to high-resolution data on (renewable) generation and demand for a synchronous area [6, 31, 34]. Large-scale unforeseen outages are rare events that we usually cannot predict. However, even during normal operational conditions, the power grid frequency shows deviations subject to various external factors, contributing to its non-constant and non-linear variability. Specifically, the deterministic frequency deviations can be observed at regular market intervals due to a change in generation scheduling [18, 35] (see Appendix B).

Forecasting power grid frequency presents unique challenges as models must accommodate long time series. High resolution is important as the short-term values and ramps are relevant for balancing actions, while a long forecasting horizon aids in better planning. Furthermore, as millions of consumers and thousands of generators are connected to large-scale grids, they display large and unpredictable fluctuations, which will likely only increase with additional renewable generation and new consumers. Hence, these grids pose a special opportunity to investigate the feasibility of forecasting highly resolved and noisy data with deep learning approaches. Ideally, these approaches should be interpretable, as the black-box nature of many algorithms can pose risks in critical infrastructures [1]. Moreover, the interpretability of the algorithms also helps in gaining scientific insights and informed decision-making [30].

2 RELATED WORK

A large part of the energy forecasting literature is focused on load forecasting, which usually employs up to hourly or quarter-hourly resolution [13]. There are comparably fewer studies on power grid frequency forecasting: There are studies on very short-term forecasts that applied Recurrent Neural Network architectures to forecast frequency with one-minute granularity for one-step-ahead predictions [38] or applied Machine learning methods to forecast grid frequency with horizons ranging from 183 ms up to 2 min (in 1 min increments) [5]. Other studies utilized time series related to power grid frequency, focusing on the forecasting of minutely power grid frequency-corrected demand in Great Britain [33] or forecasting of the intra-hour imbalance in Norway with 5 min granularity [32]. Recently, a physics-informed machine learning approach was introduced to construct an inductive biased probabilistic model of power grid frequency dynamics [16, 26]. The closest research in terms of time resolution, forecasting horizon, and dataset size is the work by [17], who used a weighted-nearest-neighbor (WNN) predictor to match similar frequency patterns. We use this model as a baseline. In the present paper, we apply the Temporal Fusion Transformer (TFT) to forecast power grid frequency. Examples of utilizing TFT in energy forecasting include prediction of hourly day-ahead photovoltaic power generation [23] and short-term electricity load forecasting [8, 14].

3 DATA

The power grid frequency data employed in this study originates from the publicly available measurements of TransnetBW [34], a German transmission system operator (TSO), in the Continental European (CE) power grid. Specifically, we use the pre-processed data by [19], which is shown in Figure 1 (top). The power grid frequency data spans a period from 2015 to 2019 with a 1 s resolution. We split our data as follows: The first three years (2015-2017) were used for training, while 2018 and 2019 were used for validation and testing respectively. Each target sample consists of one hour of data, while the preceding hour serves as input. At 1 s resolution, this yields a vector of length 3600 for both input and output. For experiments at lower resolutions, we downsample the data by averaging the data points. For example for a 15 s resolution, we have 240 timesteps for both input and output.

We use various techno-economic time series of the CE power grid from [20, 21] as external features. Our total dataset consists of 15 features with hourly resolution that are either originally based on data from the publicly available ENTSO-E Transparency Platform [6] or encode the prediction time. Six features are day-ahead forecasts of load and renewable generation provided by the TSOs, the scheduled generation, and the day-ahead electricity price for the target hour. Moreover, we include "ramp" features, i.e. we compute the difference of a feature value from its value in the previous hour. We complete the feature set with calendar features, specifically the current hour, weekday, and month.

4 METHODOLOGY

To compare the performance of the employed deep learning models, we use two baseline models as reference. The first baseline model we employ is the daily profile of the power grid frequency. It is defined as the average of all frequency recordings across all available days for a specific second of the day. This means that the daily profile is a time series consisting of $3600 \cdot 24 = 86400$ values. Due to its dependence on the daily load patterns, the daily profile — as shown in Figure 1 (bottom) — is the most pronounced recurring pattern in power grid frequency [17]. As a second baseline model, we use the weighted-nearest-neighbor (WNN) predictor introduced in [17].

In the present study, we utilize two deep-learning algorithms, namely a Multilayer Perceptron (MLP) and a Temporal Fusion Transformer (TFT) [9, 22]. The MLP is a fully connected feed-forward artificial neural network. The TFT is an attention-based architecture designed specifically for time series forecasting [22]. The architecture employs a gated variable selection mechanism, i.e. a feature selection tool designed to exclude irrelevant inputs. The gated variable selection networks and the multi-head attention offer inherent interpretability that proves particularly valuable in three scenarios: identifying significant events, capturing persistent temporal patterns, and identifying globally relevant variables. Due to the vastly different time resolutions of our external features and our target time series (hourly and secondly), we treat all external features as "static" in this study, i.e., we include a single value per feature for each sample. For specifics of the model architectures and experimental setup, refer to Appendix A.

To assess the performance of the forecasting models we use the Root Mean Square Error (RMSE) as our main evaluation metric. Furthermore, we include application-specific evaluation metrics, namely the Rate of Change of Frequency (RoCoF) and the Nadir, see also Figure 1 (top) and Appendix C [10]. The RoCoF describes the steepest slope of the frequency trajectory. The Nadir is the largest deviation from the reference frequency. Large values are problematic as they indicate a need for large amounts of control power (Nadir) and/or power ramps (RoCoF). We averaged results over five runs with varying random seeds for each deep learning model.

Table 1: Performance comparison of the models. We run each of the deep learning models five times with varying random seeds and report the mean and standard deviation of each metric.

Model	RMSE	RoCoF error	Nadir error
Daily Profile	16.3	0.768	42.1
WNN	15.5	0.766	40.1
MLP	15.795 ± 0.012	0.783 ± 0.007	38.1 ± 0.3
TFT	15.24 ± 0.04	0.735 ± 0.016	36.8 ± 0.3
Time resolution			
TFT *2s*	15.25 ± 0.05	0.75 ± 0.03	36.6 ± 0.5
TFT *5s*	15.25 ± 0.05	0.751 ± 0.017	36.7 ± 0.3
TFT *10s*	15.24 ± 0.04	0.79 ± 0.02	37.1 ± 0.3
TFT *15s*	15.26 ± 0.05	0.83 ± 0.02	37.2 ± 0.4
TFT *30s*	15.302 ± 0.019	0.939 ± 0.011	37.3 ± 0.2
TFT *60s*	15.40 ± 0.04	1.201 ± 0.013	38.2 ± 0.8
Addition of external features			
TFT	15.03 ± 0.06	0.744 ± 0.015	36.30 ± 0.14

5 EXPERIMENTS

We compare the MLP and TFT models to the two baseline models for the 1 s resolution data. We include the hour of the day as an additional static feature but no other external information. This information is inherently captured in both baseline models by design. As in [17], the WNN outperforms the daily profile. The results show that the MLP outperforms the daily profile, but its performance is worse than the WNN's. The TFT outperforms all other models. In terms of RMSE, it improves on the daily profile by 6% and on the WNN by 2% (see Table 1). This performance increase of the TFT compared to the WNN primarily stems from the first minutes of the forecast horizon, see Figure 2 (left). In the first five minutes, the performance increase exceeds the average, showing a 11% improvement over the WNN and a 19% improvement over the daily profile. The TFT seems to be more capable of modeling the market-driven deterministic frequency deviation at the start of the hour. Conclusively, the TFT also shows better performance for both, the RoCoF and the Nadir error (see Appendix 1) We note that at the beginning of each quarter-hourly trading interval, the models have a visibly larger error (see Figure 2 (left)).

The full utilization of the 1 s resolution data for hourly forecasting results in comparably long forecasting horizons (3600 timesteps). This can be harder for the model to predict and could be an unnecessary computational burden. Therefore, we vary the time resolution of the training data and quantify the changes in model performance and energy consumption. The downsampled resolution is computed by averaging data points. The final predictions of the model for the test set are then linearly interpolated and evaluated against the 1 s resolution data. We estimate the GPU energy consumption by integrating over half-minute power measurements from Weights

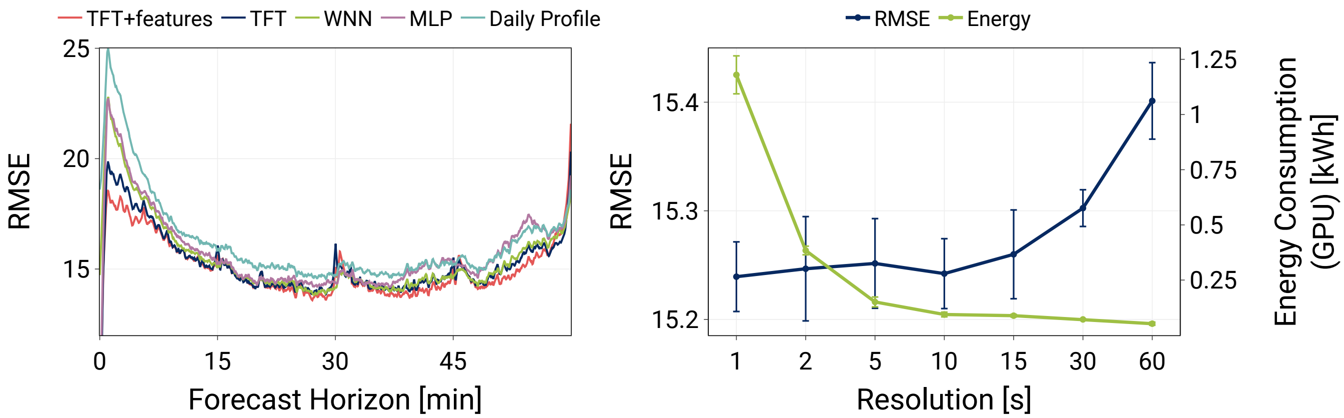

Figure 2: Performance and energy consumption. (left) Model performance of the included models (average of five seeds) compared to baselines for one-second resolution data over the forecasting horizon. TFT improvements are most noticeable in the first ten minutes. (right) TFT performance and energy consumption for different time resolutions (average and standard deviation of five seeds). Each consecutive model roughly doubles the resolution and halves the target length.

& Biases [2]. As visible in Figure 2 (right), lowering the resolution negatively impacts the RMSE, i.e. high-resolution data still contains relevant information for the TFT. Meanwhile, the training consumes less energy for coarser resolutions, with the most notable changes around target lengths of over 1000 timesteps (i.e. 1 s & 2 s resolutions). Notably all models utilizing the different resolutions are still able to beat the two baseline models. We notice that the RMSE up to about 10 s stays roughly the same and degrades more for higher resolutions. The RoCoF performance is steadily degrading and underperforms the baseline models already for resolutions lower than 5 s. The Nadir performance only decreases slightly for resolutions lower than 5 s and surpasses the baselines for all resolutions. Our results suggest that time resolutions up to 10 s seem to achieve viable performance in terms of RMSE, with the most notable degradation in RoCoF, while achieving about a tenfold reduction in power consumption. Further downsampling degrades performance without substantially reducing power consumption.

The inclusion of exogenous variables enhances the performance of the TFT by 1.5% on average and 4% in the first five minutes compared to the TFT without exogenous variables. This model further outperforms the baseline models (15% over WNN and 23% over daily profile in the first five minutes). To understand which features contributed to the improvement of the model's performance, we quantify the importance of features by evaluating the variable selection weights of the TFT's static variable selection network as proposed in [22]. We identify the load ramp forecast, the current hour, and the current month as the three most important features as shown in Table 2. The importance of the hour and month points to the effects of daily and yearly patterns. The load ramp forecast is probably a useful indicator for the deterministic frequency deviation at the start of the hour. To further understand the importance of certain time stamps, we evaluate the model's attention weights. The quarter-hourly patterns in the attention weights once again underline the impact of the market intervals on power grid frequency (see Appendix D for a more detailed discussion).

Table 2: Variable importance of static external features sorted by importance. We show the average and standard deviation of five runs with varying seeds. The three variables with the highest importance scores are bold.

Variable	Importance
hour	**0.106 ± 0.007**
load ramp forecast	**0.090 ± 0.010**
month	**0.089 ± 0.010**
scheduled generation ramp	0.085 ± 0.005
weekday	0.071 ± 0.004
day-ahead price ramp	0.067 ± 0.003
solar ramp forecast	0.063 ± 0.004
day-ahead price	0.059 ± 0.010
solar forecast	0.058 ± 0.007
load forecast	0.057 ± 0.006
onshore wind forecast	0.056 ± 0.019
offshore wind forecast	0.051 ± 0.007
scheduled generation	0.051 ± 0.013
onshore wind ramp forecast	0.051 ± 0.007
offshore wind ramp forecast	0.046 ± 0.013

6 CONCLUSION

In conclusion, our study emphasizes the importance of accurate power grid frequency forecasting in managing modern power systems, particularly as renewable energy integration increases. Our findings show that Temporal Fusion Transformers (TFT) outperform Multilayer Perceptrons (MLP) and offer better interpretability which is useful for understanding the reasoning behind the forecast and identifying which factors impact power grid frequency. Furthermore, this demonstrates that applying deep learning approaches to highly resolved time series data is feasible. The high resolution is useful if higher accuracy is needed or if application-specific metrics

(RoCoF or Nadir) have to be determined (see Table 1). We note that there exist algorithms, either attention-based or not, that are specifically built for long sequence time-series forecasting [11, 37, 39, 40]. Although these algorithms seem promising to be explored in a future study, we consider the TFT for this feasibility study due to its straightforward implementation of external (static) features and its easy access to inherent interpretations. Future research should explore more models on highly resolved data and consider different data, either from other power grids [31] or generation and imbalance data. With this contribution we hope to inspire more research in the area of highly resolved energy time series, advance forecasting of such time series, including power grid forecasting, and thereby support the stability and control of future renewable energy systems.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Helmholtz Association under grant no. VH-NG-1727 and the Networking Fund through Helmholtz AI. This work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

REFERENCES

[1] Tanveer Ahmad, Dongdong Zhang, Chao Huang, Hongcai Zhang, Ningyi Dai, Yonghua Song, and Huanxin Chen. 2021. Artificial Intelligence in Sustainable Energy Industry: Status Quo, Challenges and Opportunities. *Journal of Cleaner Production* 289 (2021), 125834. https://doi.org/10.1016/j.jclepro.2021.125834

[2] Lukas Biewald. 2020. Experiment Tracking with Weights and Biases. https://www.wandb.com/ Software available from wandb.com.

[3] Andreas Bloess, Wolf-Peter Schill, and Alexander Zerrahn. 2018. Power-to-Heat for Renewable Energy Integration: A Review of Technologies, Modeling Approaches, and Flexibility Potentials. *Applied Energy* 212 (2018), 1611–1626. https://doi.org/10.1016/j.apenergy.2017.12.073

[4] Rui Carvalho, Lubos Buzna, Richard Gibbens, and Frank Kelly. 2015. Critical Behaviour in Charging of Electric Vehicles. *New Journal of Physics* 17, 9 (2015), 095001. https://doi.org/10.1088/1367-2630/17/9/095001

[5] N. Chettibi, A. Massi Pavan, A. Mellit, A. J. Forsyth, and R. Todd. 2021. Real-Time Prediction of Grid Voltage and Frequency Using Artificial Neural Networks: An Experimental Validation. *Sustainable Energy, Grids and Networks* 27 (2021), 100502. https://doi.org/10.1016/j.segan.2021.100502

[6] ENTSO-E. 2024. ENTSO-E Transparency Platform. https://transparency.entsoe.eu/

[7] William Falcon and The PyTorch Lightning team. 2019. PyTorch Lightning. https://doi.org/10.5281/zenodo.3828935

[8] Elena Giacomazzi, Felix Haag, and Konstantin Hopf. 2023. Short-Term Electricity Load Forecasting Using the Temporal Fusion Transformer: Effect of Grid Hierarchies and Data Sources. In *Proceedings of the 14th ACM International Conference on Future Energy Systems* (New York, NY, USA) *(E-Energy ’23)*. Association for Computing Machinery, 353–360. https://doi.org/10.1145/3575813.3597345

[9] Ian Goodfellow, Courville, Aaron, and Bengio, Yoshua. 2016. *Deep Learning*. MIT press.

[10] Dominic Groß, Saverio Bolognani, Bala K. Poolla, and Florian Dörfler. 2017. Increasing the Resilience of Low-Inertia Power Systems by Virtual Inertia and Damping. In *Proceedings of IREP’2017 Symposium*. International Institute of Research and Education in Power System Dynamics …, 64.

[11] Albert Gu, Karan Goel, and Christopher Re. 2022. Efficiently Modeling Long Sequences with Structured State Spaces. In *International Conference on Learning Representations*. https://openreview.net/forum?id=uYLFoz1vlAC

[12] Dominik Heide, Lueder von Bremen, Martin Greiner, Clemens Hoffmann, Markus Speckmann, and Stefan Bofinger. 2010. Seasonal Optimal Mix of Wind and Solar Power in a Future, Highly Renewable Europe. *Renewable Energy* 35, 11 (2010), 2483–2489. https://doi.org/10.1016/j.renene.2010.03.012

[13] Tao Hong, Pierre Pinson, Yi Wang, Rafał Weron, Dazhi Yang, and Hamidreza Zareipour. 2020. Energy Forecasting: A Review and Outlook. *IEEE Open Access Journal of Power and Energy* 7 (2020), 376–388. https://doi.org/10.1109/OAJPE.2020.3029979

[14] Pham Canh Huy, Nguyen Quoc Minh, Nguyen Dang Tien, and Tao Thi Quynh Anh. 2022. Short-Term Electricity Load Forecasting Based on Temporal Fusion Transformer Model. *IEEE Access* 10 (2022), 106296–106304. https://doi.org/10.1109/ACCESS.2022.3211941

[15] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference for Learning Representations*.

[16] Johannes Kruse, Eike Cramer, Benjamin Schäfer, and Dirk Witthaut. 2023. Physics-Informed Machine Learning for Power Grid Frequency Modeling. *PRX Energy* 2, 4 (2023), 043003. https://doi.org/10.1103/PRXEnergy.2.043003

[17] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2020. Predictability of Power Grid Frequency. *IEEE Access* 8 (2020), 149435–149446. https://doi.org/10.1109/ACCESS.2020.3016477

[18] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2021. Exploring Deterministic Frequency Deviations with Explainable AI. In *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 133–139. https://doi.org/10.1109/SmartGridComm51999.2021.9632335

[19] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2021. Pre-Processed Power Grid Frequency Time Series. https://doi.org/10.5281/zenodo.5105820

[20] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2021. Revealing Drivers and Risks for Power Grid Frequency Stability with Explainable AI. *Patterns* 2, 11 (2021), 100365. https://doi.org/10.1016/j.patter.2021.100365

[21] Johannes Kruse, Dirk Witthaut, and Benjamin Schäfer. 2021. Supplementary Data: "Revealing Drivers and Risks for Power Grid Frequency Stability with Explainable AI". https://doi.org/10.5281/zenodo.5118352

[22] Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. 2021. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *International Journal of Forecasting* 37, 4 (2021), 1748–1764. https://doi.org/10.1016/j.ijforecast.2021.03.012

[23] Miguel López Santos, Xela García-Santiago, Fernando Echevarría Camarero, Gonzalo Blázquez Gil, and Pablo Carrasco Ortega. 2022. Application of Temporal Fusion Transformer for Day-Ahead PV Power Forecasting. *Energies* 15, 14 (2022), 5232. Issue 14. https://doi.org/10.3390/en15145232

[24] Jan Machowski, Zbigniew Lubosny, Janusz W Bialek, and James R Bumby. 2020. *Power System Dynamics: Stability and Control*. John Wiley & Sons.

[25] Patrick Milan, Matthias Wächter, and Joachim Peinke. 2013. Turbulent Character of Wind Energy. *Physical Review Letters* 110, 13 (2013), 138701. https://doi.org/10.1103/PhysRevLett.110.138701

[26] George S. Misyris, Andreas Venzke, and Spyros Chatzivasileiadis. 2020. Physics-Informed Neural Networks for Power Systems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*. 1–5. https://doi.org/10.1109/PESGM41954.2020.9282004

[27] Nixtla. 2024. Nixtla/Neuralforecast: Scalable and User Friendly Neural Forecasting Algorithms. https://github.com/Nixtla/neuralforecast/tree/main

[28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[29] Joeri Rogelj, Gunnar Luderer, Robert C. Pietzcker, Elmar Kriegler, Michiel Schaeffer, Volker Krey, and Keywan Riahi. 2015. Energy System Transformations for Limiting End-of-Century Warming to below 1.5 °C. *Nature Climate Change* 5, 6 (2015), 519–527. Issue 6. https://doi.org/10.1038/nclimate2572

[30] Ribana Roscher, Bastian Bohn, Marco F. Duarte, and Jochen Garcke. 2020. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* 8 (2020), 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

[31] Leonardo Rydin Gorjão, Richard Jumar, Heiko Maass, Veit Hagenmeyer, G. Cigdem Yalcin, Johannes Kruse, Marc Timme, Christian Beck, Dirk Witthaut, and Benjamin Schäfer. 2020. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. *Nature Communications* 11, 1 (2020), 6362. https://doi.org/10.1038/s41467-020-19732-7

[32] Tárik S. Salem, Karan Kathuria, Heri Ramampiaro, and Helge Langseth. 2019. Forecasting Intra-Hour Imbalances in Electric Power Systems. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (2019), 9595–9600. Issue 01. https://doi.org/10.1609/aaai.v33i01.33019595

[33] James W. Taylor and Matthew B. Roberts. 2016. Forecasting Frequency-Corrected Electricity Demand to Support Frequency Control. *IEEE Transactions on Power Systems* 31, 3 (2016), 1925–1932. https://doi.org/10.1109/TPWRS.2015.2444665

[34] TransnetBW GmbH. 2020. Regelenergie Bedarf + Abruf. https://www.transnetbw.de/de/strommarkt/systemdienstleistungen/regelenergie-bedarf-und-abruf

[35] Tobias Weißbach, Simon Remppis, and Hendrik Lens. 2018. Impact of Current Market Developments in Europe on Deterministic Grid Frequency Deviations and Frequency Restauration Reserve Demand. In *2018 15th International Conference on the European Energy Market (EEM)*. 1–6. https://doi.org/10.1109/EEM.2018.

8469210
[36] Dirk Witthaut, Frank Hellmann, Jürgen Kurths, Stefan Kettemann, Hildegard Meyer-Ortmanns, and Marc Timme. 2022. Collective Nonlinear Dynamics and Self-Organization in Decentralized Power Grids. *Reviews of Modern Physics* 94, 1 (2022), 015005. https://doi.org/10.1103/RevModPhys.94.015005
[37] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *Advances in Neural Information Processing Systems*, Vol. 34. Curran Associates, Inc., 22419–22430. https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html
[38] Ogun Yurdakul, Fatih Eser, Fikret Sivrikaya, and Sahin Albayrak. 2020. Very Short-Term Power System Frequency Forecasting. *IEEE Access* 8 (2020), 141234–141245. https://doi.org/10.1109/ACCESS.2020.3013165
[39] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 12 (2021), 11106–11115. Issue 12. https://doi.org/10.1609/aaai.v35i12.17325
[40] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In *Proceedings of the 39th International Conference on Machine Learning*. PMLR, 27268–27286. https://proceedings.mlr.press/v162/zhou22g.html

A EXPERIMENTAL SETUP

To determine optimal hyperparameters, we conducted a Bayesian search within the parameter space shown in Table 3, employing Weights & Biases [2]. We conducted one hyperparameter search at 1 s resolution for the TFT and used the best hyperparameters throughout this paper. Both models were trained using mean squared error (MSE) as the optimization criterion, with stochastic gradient descent (SGD) employed as the optimizer for the MLP and ADAM [15] for the TFT. During the final runs, we implemented early stopping with a patience of 20 epochs. The algorithmic implementations were adapted from [27] and executed using PyTorch and PyTorch Lightning [7, 28]. Notably, the hour feature was encoded using sine and cosine transformations for the MLP. We averaged results over five runs with varying random seeds for each deep learning model.

Table 3: Value ranges for hyperparameter tuning

MLP

Parameter	Range	Best
batch size	$\{64, 128, 256, 512, 1024\}$	256
hidden size	$\{256, 512, 1024, 2048, 4096\}$	4096
learning rate	$[0.0001, 10]$ (log uniform)	0.187
number of layers	$\{1, 2, 3, 4, 5, 6\}$	3

TFT

Parameter	Range	Best
attention heads	$\{1, 4\}$	1
hidden size	$\{64, 128, 256\}$	128
dropout rate	$[0.05, 0.5]$ (uniform)	0.27
learning rate	$[0.0001, 0.1]$ (log uniform)	8.5×10^{-4}

B DETERMINISTIC FREQUENCY DEVIATIONS

Deterministic frequency deviations (DFDs) are comparably large deviations from the reference frequency that occur at the beginning of market intervals [18, 35]. On the day-ahead markets in the Continental European power grid electricity is procured in hourly or quarter-hourly intervals. The rapid adjustment of power generation and demand following the intervals causes an almost step-like behavior — constant during the market intervals and jumping in between. Combined with the continuously changing demand that is led by consumer behavior, not necessarily directly bound to the market intervals, there arise power mismatches at the beginning of the market intervals. DFDs can affect the frequency quality and stability of the power system, as they deplete the frequency control reserves and make the grid vulnerable to additional disturbances.

C ROCOF & NADIR

The Rate of Change of Frequency (RoCoF) and the Nadir both describe relevant metrics in power system operation, linked to the stability of the system [10]. A large value in one or both quantities signifies a substantial imbalance and a potential stability problem of the system.

We calculate the RoCoF from the grid frequency deviation time series $f(t)$ similar to the procedure described in [18, 20]. To estimate the derivative df/dt, we compute the increments $f(t + 1\text{s}) - f(t)$. Next, using a rectangular rolling window with a length of 30 s, we smooth the increment time series. We then set the RoCoF to the absolute maximum of this smoothed time series. To evaluate our forecasts we apply the same procedure to the prediction and the ground truth and calculate the absolute difference between the two values.

The Nadir describes the largest deviation from the reference frequency in an Interval I and is therefore defined as

$$\text{Nadir}(I) = f(\arg\max_{t \in I} |f(t)|).$$

To calculate the error, we compute this number for the prediction and the ground truth and again calculate the absolute difference between the two values.

D INTERPRETABLE ATTENTION

To identify which parts of the past time series are important for the prediction of the TFT, we visualize the average attention weights of one of the TFT models with exogenous features, as proposed in [22]. We find that, apart from a very sharp peak at the last timestep before the prediction interval, the model attends to the past timesteps more or less uniformly for all time horizons. At each quarter hour, there are slight bumps visible in the attention which coincide with the market intervals. The impact of the market intervals is even more visible for the attention to the timesteps that the model predicted itself. We observe that both, for the t_0+20 min horizon as well as the $t_0 + 30$ min horizon, the model has a decreased attention between t_0 and $t_0 + 15$ min. Similarly, we notice a decreased attention between t_0 and $t_0 + 30$ min for the $t_0 + 50$ min and $t_0 + 60$ min horizons. This loss of attention probably links to the market-based memory loss observed in [17].

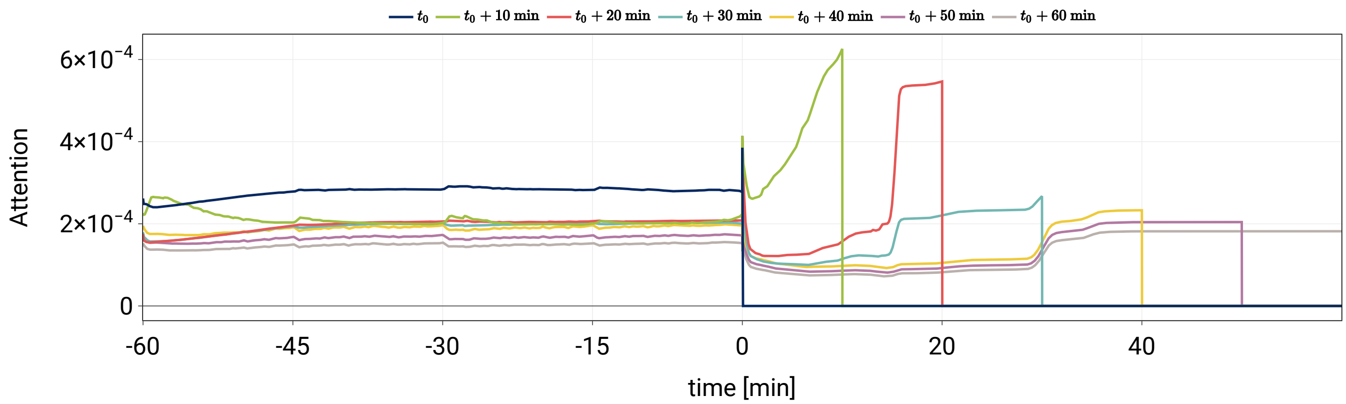

Figure 3: Average attention weights for forecasts at different horizons.