Tilburg University

PREDICTING ELECTRICITY DEMAND WITH DEEP LEARNING TECHNIQUES

MARIA-ANTIGONE RUMPF

THESIS PROPOSAL
DATA SCIENCE & SOCIETY

STUDENT NUMBER

563586

COMMITTEE

dr. Giuseppe Cascavilla

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

February 10th, 2025

WORD COUNT

1768

PREDICTING ELECTRICITY DEMAND WITH DEEP LEARNING TECHNIQUES

MARIA-ANTIGONE RUMPF

1 PROJECT DEFINITION, MOTIVATION & RELEVANCE

Problem Definition: Forecasting electricity demand is a central issue for government-level institutions and grid suppliers. Stakeholders rely on approximations of future electricity demand, and are at risk of financial losses and grid instability when the market is over- or under-supplied. Optimizing electricity demand forecasting has therefore been an important task for policymakers and operators.

Machine Learning techniques in electricity demand forecasting have been on the rise, as they capture non-linearities and dependencies that standard models cannot. A lot of research is available on *short-term* demand forecasting, predicting energy demand based on historical data, time data, and weather data. *Short-term* forecasting (hours and days ahead of time) is helpful to operators to optimize their supply, and to ensure grid stability. However, these models have ignored *long-term* forecasting (month ahead of time), which has additional value.

This research employs Temporal Fusion Transformers (TFT), a Deep Learning model, in order to predict *long-term* forecasting of grid-level electricity demand in Germany. New exogenous features are taken into account, namely the availability of renewable energy, which influence longer-term trends. Relative importance of features is extracted post-hoc, to quantify the impact of renewable energy production on electricity demand. The model's efficiency is compared against seasonal ARIMA (SARIMA), a statistical model widely used as a benchmark for long-term forecasting.

Societal Relevance: The societal relevance of *long-term* energy demand forecasting on the national grid-level is twofold: both for operators and policymakers. First, operators can benefit from *long-term* forecasting for strategic planning, such as reserve generation to meet demand spikes and avoid blackouts, or scheduling maintenance without disrupting grid transmission. Second, for policymakers, understanding the effect of renewable energy sources on electricity demand will help inform the best implementation policy for the clean energy transition, without disrupting energy supply for the population.

Scientific Relevance: The scientific relevance is to evaluate the use of TFT for *long-term* time-series predictions. The method is successful in identifying *shorter-term (daily)* dependencies, however is not widely used to reap the benefits of *longer-term (monthly)* forecasting. Identifying whether it manages to beat robust benchmark models such as seasonal SARIMA propels the state-of-the-art in *long-term* forecasting.

2 LITERATURE REVIEW

Transformer models are primarily used in Natural Language Processing, using attention mechanisms. Lim et al. (2021) introduced TFT to time-series forecasting, and it has since become a State-of-the-Art method, trumping traditional methods like LSTM or XGBoost. Since, TFT's are widespread in financial forecasting, to predict stock volatility and price trends in the cryptocurrency market.

#	Research Paper	Field	Forecasting Goal	Method	Results
1	Zheng et al.	Energy Forecasting	Short-term, building-level energy consumption	TFT and clustering, for feature	Improved accuracy and interpretability
2	Nazir et al. (2023)	Energy Forecasting	Short-term, individual-level energy consumption	TFT and multi-head attention	High accuracy in short-term, limited long-term prediction ability
3	Jenko & Costa (2024)	Energy Forecasting	Short-term, multi-target forecasting: demand, price, and carbon emissions	TFT with multi-task learning	Improved accuracy over baselines
4	Giacomazzi et al. (2023)	Energy Forecasting	Short- and medium-term, grid-level and substation-level electricity demand	TFT compared to LSTM and ARIMA	TFT outperforms LSTM at substation level, but not at grid-level
5	Laborda et al. (2023)	(Macro-)economic Forecasting	Medium- and long-term, multi-country GDP	Multi-horizon TFT predictions	TFT outperforms ARIMA, especially during economic turbulence
6	Ho & Hung (2024)	Financial Forecasting	Long-term stock price	CEEMD-based TFT model	Improved accuracy over baselines

Figure 1: Overview of forecasting studies using TFT models

There are two key aspects in the literature that this project combines. Firstly, recent studies have extended TFT to energy forecasting, however only in the *shorter-term*, meaning hour-ahead, day-ahead and week-ahead predictions. Papers 1-4, listed in Figure 1, indicate the forecasting targets and methods. Zheng et al. (2023) conducted building-level energy forecasting using TFT, using clustering for feature extraction instead of granular predictors, strongly emphasizing interpretability for shorter-term forecasting, in favor of energy management and energy-saving on the building level. Giacomazzi et al. (2023) similarly forecasted short-term electricity load forecasting at the national grid-level, however on a short-term basis, and therefore used LSTM as a benchmark model, whereas ARIMA is a more appropriate long-term horizon. This research found that TFT did not outperform LSTM models in short-term forecasting in general, however was able to improve the prediction of more granular targets, at the substantion level. Similarly, Nazir et al. (2023) also applied TFT for short-term grid forecasting for residental consumption, where TFT did indeed outperform LSTM. Jenko and Costa (2024) used TFT for multi-target forecasting, to instantly predict electricity demand, energy price and carbon emissions against traditional methods, where TFT also performed better. In sum, the results point to interpretability as a major asset in TFT's forecasting ability, and that TFT as a trend results in improved accuracy over statistical baselines. This research, however, does not extend to longer-term predictions.

Secondly, other papers in the fields of economics and financial predictions have indeed leveraged TFT methods to *longer-term* forecasting, such as yearly or monthly predictions, as seen in Figure 1, where Papers 5 and 6 have also shown accuracy improvements over the baselines. In this example, TFT has been used for long-term forecasting, and benchmarked against ARIMA. Han et al. (2023) forecasted macroeconomic Chinese industry data, finding that ARIMA performed poorly due to linear assumptions leveraged against volatile economic conditions. Volatile conditions also apply to energy demand, as certain socioeconomic factors have defined energy use in the last decade. Laborda et al. (2023) predicted GDP values, where over extended horizons, TFT outperformed ARIMA.

Contribution: The added value of this research to the field of energy forecasting is two-fold. Firstly, forecasting grid-level electricity demand in a *long-term* time horizon (30 days ahead), to add to the well researched of the well researched short-term (next-day) approach, using TFTs. Secondly, the use of renewable energy availability as an exogenous feature, which has not been accounted for in the literature. To achieve this, the study takes the long-term horizon forecasting approach, described by macroeconomic studies, which benchmarks a TFT model against the standard statistical model of ARIMA. To use a more robust benchmark specifically for energy prediction, the study uses the adapted seasonal ARIMA (SARIMA) to account for cyclical weather trends. This approach is applied to an under-explored field of *long-term* electricity demand forecasting.

3 RESEARCH STRATEGY & RESEARCH QUESTIONS

This research forecasts electricity demand on several time horizons, including the previously unexplored long-term horizon (30 days ahead) in Germany, for the last decade. A Temporal Fusion Transformer model employs multi-horizon forecasting, to analyze whether deep learning significantly improves forecasting accuracy, as compared to a robust statistical baseline, SARIMA. The study also extracts feature importance, to identify for policymakers and operators the key predictors which affect how Germany's grid demand fluctuates.

Research Questions:

- RQ 1: To what extent is the Temporal Fusion Transformer (TFT) model able to outperform seasonal ARIMA for 30-day electricity demand forecasting?

  RQ1 applies the model to forecast electricity demand in Germany. The literature indicates that TFT has been successful in short-term time horizon forecasting (one day-ahead) of electricity demand. This question addresses the gap between using the Deep Learning method to forecast longer-term time horizons. 30-days ahead is chosen as the long-term time horizon of interest, because it relates to monthly-forecasting, which is unexplored in this domain. Seasonal ARIMA (SARIMA) is selected as a robust benchmark, used in long-term forecasting of macroeconomic indicators in the literature.

  – SRQ 1.1: How do short-term (day-ahead), medium-term (7 days ahead) and long-term (30 days ahead) time horizons compare in TFT's ability to accurately forecast electricity demand?

  SRQ 1.1 builds on RQ1 to compare how long-term time horizon captures predictions differently from the short- and medium-term, by leveraging TFT's ability to apply multi-time horizons instantaneously. This targets how well long-term forecasting can be applied to electricity demand data using TFT, by directly comparing how the model is able to capture different trends.

- RQ 2: How do the predictor features contribute to most accurate long-term electricity demand forecasting, by comparing SHAP values between TFT and SARIMA?

RQ 2 derives interpretability and feature importance, in order to satisfy the real-world value of predicting energy prices, allowing relevant stakeholders to conduct strategic planning. SHAP values, an explainable AI (XAI) technique, identify the relative weight of each feature. TFT and SARIMA each also offer internal interpretable opportunities: via attention weights for TFT and autoregressive patterns for SARIMA. These are also identified in the analysis, however SHAP values allow for direct comparison between the models.

- SRQ 2.1: To what extent do different time horizons change what features most accurately predict electricity demand?

SRQ 2.1 builds on feature importance extraction, by analyzing how the TFT model finds different features relevant for analysis in the short-, medium-, and long-term. This informs the gap in the literature about long-term forecasting of energy prices, and encourages future research by identifying which predictors make short-term or long-term effects on energy demand.

Feature Engineering includes creating past demand value via using lags at key intervals for each time horizon. These lags are treated as hyperparameters for SARIMA, which are tuned during validation via grid search, and are automatically chosen by TFT from a predefined set. Temporal encoding will also be used on the timestamps, to extract, hours, days and months.

4 METHODOLOGY AND EVALUATION

4.1 *Dataset Description*

The open-access dataset, derived from ENTSO-E Transparency Platform, shows electricity demand at 15 minute intervals, from the beginning of 2015 to the end of 2024. The decade-long span allows for long-term trend identification. Figure 2 shows the selected features. The inclusion of exogenous features allow the TFT to identify correlations from unseen inputs.

The dataset accounts for over 200,000 rows of data, appropriate for a Deep Learning algorithm such as TFT. It is formated in a CSV file, in an appropriate time-series format.

Target variable	Actual electricity demand (MW) in Germany
Time-based features	Hour
	Day of the week
	Month
	Season
Past demand (lag features):	Lags of the target variables. Lags on 1h, 6h, 24h, and 7 days.
Exogenous features: weather conditions and availability of renewable energy sources.	Temperature
	Direct solar radiation
	Diffuse solar radiation
	Solar energy generation
	Onshore wind energy generation
	Offshore wind energy generation

Figure 2: Overview of forecasting studies using TFT models

4.2 *Algorithms and Software*

Benchmark Model: Seasonal ARIMA (SARIMA) is a robust benchmark for long-term time-series forecasting. ARIMA is a widely used statistical model, and a common benchmark against Deep Learning methods, as it can also capture temporal dependencies. The literature review identifies long-term horizon forecasting approaches, outside of the domain of energy prediction, which use ARIMA as a baseline. SARIMA is a more robust choice, due to its ability to handle cyclical trends, such a weather.

Deep Learning Model: The Temporal Fusion Transformer model is optimized for multi-horizon time-series forecasting and to handle exogenous features. This research extends TFT's use within energy forecasting literature, to test it against the benchmark for *longer-term* time horizons. TFT's ability to compare several time horizons within one round of model training allows for key comparisons between short-, medium-, and long-term forecasting performance.

4.3 *Evaluation Method*

Performance Metrics: This study uses two performance metrics for the regression task performed by the TFT model: Mean Aboslute Error (MAE) and Root Mean Squared Error (RMSE). These are widely used accuracy in both TFT and SARIMA. MAE measures the average absolute difference between predicted and actual values, meanwhile RMSE places a higher penalty on larger errors. The combination ensures a balance, in case small and large errors differ in real-world importance.

Baseline: The comparative approach is carried out by measuring the performance of SARIMA against the Deep Learning architecture of Temporal Fusion Transformers. The TFT model's results are compared against the three time horizons, namely short-term (day ahead predictions), medium-term (week ahead predictions) and long-term (month ahead predictions).

Validation and Testing: The split for time-series data, using a time-based train-test split, without shuffling. The split is 80:10:10, reserving the validation set for hyperparameter tuning. Random search is used for model optimization of both the ARIMA and TFT models, meanwhile grid search is used for the selection of the optimal lag value for ARIMA.

Interpretability: SHAP values are used for post-hoc feature importance extraction for both the baseline and TFT model. For further analysis, both models have inherent extractable values, namely attention weights for TFT and autoregressive patterns for ARIMA.

5 MILESTONES AND PLAN

Deadline	**Description**
February 28th	Submit thesis proposal
April 1st	Finish preprocessing, EDA & feature engineering
April 24th	Finish training algorithms
May 5th	Finish error analysis & feature importance analysis
May 15th	Finish Results & Discussion
May 15th-19th	Final review of thesis or extra time in event of delays
May 19th	Submission deadline

Table 1: Thesis Timeline

REFERENCES

Giacomazzi, E., Haag, F., & Hopf, K. (2023). Short-term electricity load forecasting using the temporal fusion transformer: Effect of grid hierarchies and data sources. *Proceedings of the 14th ACM International Conference on Future Energy Systems*, 353–360. https://doi.org/10.1145/3575813.3597345

Han, Y., Tian, Y., Yu, L., & Gao, Y. (2023). Economic system forecasting based on temporal fusion transformers: Multi-dimensional evaluation and cross-model comparative analysis. *Neurocomputing*, *552*, 126500. https://doi.org/https://doi.org/10.1016/j.neucom.2023.126500

Jenko, J., & Costa, J. P. (2024). Using temporal fusion transformer predictions to maximise use of renewable energy sources. *2024 International Workshop on Artificial Intelligence and Machine Learning for Energy Transformation (AIE)*, 1–10. https://doi.org/10.1109/AIE61866.2024.10561261

Laborda, J., Ruano, S., & Zamanillo, I. (2023). Multi-country and multi-horizon gdp forecasting using temporal fusion transformers. *Mathematics*, *11*(12). https://doi.org/10.3390/math11122625

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764. https://doi.org/https://doi.org/10.1016/j.ijforecast.2021.03.012

Nazir, A., Shaikh, A. K., Shah, A. S., & Khalil, A. (2023). Forecasting energy consumption demand of customers in smart grid using temporal fusion transformer (tft). *Results in Engineering*, *17*, 100888. https://doi.org/https://doi.org/10.1016/j.rineng.2023.100888

Zheng, P., Zhou, H., Liu, J., & Nakanishi, Y. (2023). Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture. *Applied Energy*, *349*(1), 121607. https://doi.org/https://doi.org/10.1016/j.apenergy.2023.121607