Applied Machine Learning for Short-Term Electric Load Forecasting in Cities - A Case Study of Eskilstuna, Sweden

Pontus Netzell[a,*] Hussain Kazmi[b] Konstantinos Kyprianidis[a]

[a]*Future Energy Center, Mälardalen University, Västerås, Sweden* [b]*Dept. of Electrical Engineering, KU Leuven, Leuven, Belgium*

[*]pontus.netzell@gmail.com

Abstract

With the growing demand, electrification, and renewable proliferation, the necessity of being able to forecast future demand in combination with flexible energy usage is tangible. Distribution network operators often have a power capacity limit agreed with the regional grid, and economic penalties await if crossed. This paper investigates how cities could deal with these issues using data-driven approaches. Hierarchical electric load data is analyzed and modeled using Multiple Linear Regression. Key calendar variables holidays, industry vacation, "Hour of day" and "Day of week" are identified alongside the meteorological heating-, and cooling degree hours, global irradiance, and wind speed. This inexpensive algorithm outperforms the benchmark "weekly Naïve" with a relative Root Mean Squared Error of 35% for the year-long rolling origin evaluation. Learnings from the data exploration and modeling are then used to evaluate the AI-based model Light Gradient Boosting Machine. Using similar explanatory variables for this expensive algorithm results in a relative error of 45%, although it outperforms the previous one during the summer. The models have varying strengths and weaknesses and could advantageously be combined into an ensemble model for improving accuracy. Incorporating detailed knowledge of local renewable electricity production in combination with hierarchical forecasting could further increase accuracy. With domain knowledge and statistical analysis, it is possible to create robust load forecasts with acceptable accuracy using easily available machine-learning libraries. Both models have good potential to be used as input to economic optimization and load shifting.

1 Introduction

A part of the solution to reach the global climate goals is to use renewable energy sources, which are volatile, intermittent, and non-dispatchable by nature (Huber et al., 2014). This poses several questions about continued grid stability and conventional power plants need to adapt to this reality by operating more flexibly, ramping up and down at a pace not traditionally seen (Beiron et al., 2020). Uncertainty and volatility in electricity production from variable renewable energy sources could be handled with demand response (Meliani et al., 2021) and the utilization of energy storage for load shifting (Cebulla et al., 2017).

In Sweden, the electrification of the transport and industry sector is crucial for carbon emission reduction, leading to significant growth in electricity demand. Two outstanding examples of industrial growth are the HYBRIT green steel project in the northern parts (Öhman et al., 2022), and the southern Mälardalen region due to its dense population and the addition of new electricity intense industries. Electricity has traditionally been transferred through the national grid from northern hydropower plants, and more recently offshore wind turbines, to the energy-intense southern half of the country. However, due to the rapid growth and end of life for several southern nuclear power plants there are short-term issues in the transfer capabilities, meaning the southern demand cannot be sustainably fulfilled with northern electricity. Intense reinforcement and expansion of the high-voltage grid may eventually make it possible to supply the additional demand. In the long-term however, the increase in electricity usage in the northern areas could lead to a shortage of energy to transfer to the south. Therefore, there is a need for increased local production in the energy-intensive southern cities and regions for a robust and resilient local energy system (Nik et al., 2021).

In Eskilstuna, a city located in the Mälardalen region, the dispatchable local electricity production currently makes up a small portion of the total demand, the rest is imported. The addition of several megawatt-size (MW-size) photovoltaic (PV) parks, and a wind turbine park will increase the yearly energy self-usage ratio. However, it does not resolve the issue on an hourly and seasonal basis, as there is no substantial electricity generation from PV in the evenings as well as during winter in Sweden. With the growing demand, electrification, and renewable proliferation, the necessity of being able to forecast future demand in combination with flexible energy usage is tangible. Reliable forecasts can enable system operators and utilities to better manage the demand and supply balance in real-time, and control energy storage units for shifting load from high to low production periods, i.e. from day to night, or summer to winter. Use cases for forecasts range from long-term world trends and national changes to medium- and short-term changes on a regional or city-scale level (Hong et al., 2020). Forecasting is essential for the energy and power sector and the area has gotten attention for many decades, but with increasing computational power and new advanced models, the area is regaining focus. Individual investigations are necessary as each dataset is unique and more complex models do not equal increased accuracy. Managing energy assets based on bad forecasts can lead to higher operating costs and, in a worst-case scenario, blackouts in the power grid.

Forecasting can be divided into three main parts using a systems engineering perspective; *Input*, *Model*, and *Output* (Hong & Fan, 2016). Size of historical data for training and the selection of both dependent and independent variables are examples of *Input* variables. If the data is disaggregated by geographical location, then hierarchical forecasting can be chosen as the *Model* technique (Hong et al., 2020). Other *Model* variants are the selection of e.g. non-linear or linear, black-box or non-black-box models, and their respective parameters. The predictions (*Output*) can be combined into ensembles, which is usually considered the best practice (Wang et al., 2018). The application of the forecasts matters, peak prediction generally demands an approach that is different from forecasts used for operational optimization of energy units (Gajowniczek & Ząbkowski, 2017). While numerous forecasting techniques have been proposed, there is no one-size-fits-all, a detailed analysis of the specific case is needed for maximizing the forecast accuracy.

This paper focuses on forecasts and their usage on the urban and sub-urban electricity demand levels in a city via a case study of the Eskilstuna Strängnäs Energi och Miljö (ESEM) electrical grid and energy system. Short-Term Load Forecasting (STLF) is applied to the geographically disaggregated hourly average electric load. The aim of this study is to create and explore a framework to analyze and evaluate forecasting models and determine which calendar and meteorological input variables are best suited for forecasting the electricity demand in cities similar to the studied city. Multiple Linear Regression (MLR) and Light Gradient Boosting Machine (LGBM) are compared to the benchmark ”weekly Naïve” to determine whether advanced AI-based methods provide additional value compared to simpler benchmarks. Implementing these forecasts for control of energy storage units and other flexible assets is discussed, and possible strengths and weaknesses of the two models are emphasized.

The rest of the paper is structured as follows: In section 2: Methodology, data acquisition, algorithm creation, and model selection are presented. In section 3: Results and Discussion, the choice of explanatory variables and model results is presented and criticized. The study is concluded in section 4: Summary and Conclusions, where the road ahead is elaborated.

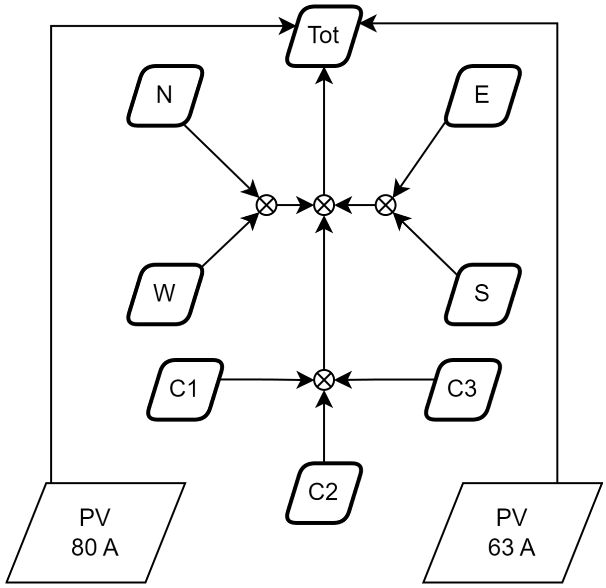

Figure 1. Three central and four outer transformer stations, together with PV-park of different sizes, make up the city’s total energy usage.

2 Methodology

2.1 Data acquisition and pre-processing

The dataset used herein comprises hourly average electrical load in MW from 2020-09-11 to 2022-10-31 (2 years, 1 month, 20 days) and is collected for all entry points (transformer stations) between the regional and the local grid. Local electricity production, e.g. small-scale hydropower generation, has been accounted for according to which transformer station they are connected to. The summation of the seven transformer station loads, together with the generation from all large (63 and 80 Ampere) PV installations, makes up the total energy usage of the city, denoted as ”total energy usage” in this paper (and ”Tot” in Fig. 1). The electricity generation from smaller local PV installations, such as private households, is not included in the total energy usage. The PV installations are not separated into individual time series according to their location; therefore, they must be excluded from the grouped forecasting part of this study. Instead, each individual transformer station and the sum are used for grouped forecasting.

2.2 Data exploration, correlation, and other statistics

To build an accurate forecast model, several meteorological and calendar explanatory variables are evaluated in terms of correlation with total energy usage and improvement in model accuracy. Some of the meteorological variables are reanalysis data of wet and dry temperature, wind speed, rain, and global irradiance from SMHI (2023). Measured in-situ temperature from the central power plant is also used, including smoothed variants, i.e. moving averages with different window sizes. Cross-effects can be calculated by multiplying meteorological and calendar variables (Hong et al., 2010). Degree days and -hours for heating and cooling, which is the temperature difference below or above a certain threshold multiplied by time (Chabouni et al., 2020), are examples of cross effects.
The correlation coefficient between each transformer station's load and the reanalysis dry temperature varies between -0.32 to -0.80 (-0.56 for the total load). Such a varying correlation with temperature is indicating the different patterns of usage for different parts of the city. A closer look reveals that the dry temperature gives higher accuracy more often than the wet.
By plotting the load versus different categories, e.g. in a box plot with the hour of the day on the x-axis, the daily load distribution is shown. The load is significantly lower during the night compared to the day. During autumn, winter, and spring a morning peak at 09:00 ±1h, and an afternoon peak at 17:00 ±1h, is identified. However, the load pattern during summer is different, with a single peak at 11:00 ±1h.
Public holidays are considered non-typical days (Eroshenko et al., 2017) where the load is significantly lower. Additive decomposition of the trend, seasonal and residual components (Hyndman & Athanasopoulos, 2018) is applied using the Python library Statsmodels (Seabold & Perktold, 2010). Similar to Işık et al. (2023) the MSTL (Multi Seasonal Trend Decomposition using LOESS (Locally Estimated Scatterplot Smoothing)) reveals daily and weekly seasonality.

2.3 Forecast models and benchmark

The benchmark model is selected as the well-known, in energy forecasting, "weekly Naïve" (copy-paste the previous week's values as the forecast for the next). It captures the weekly seasonality in the data and therefore outperforms the "daily Naïve" (Kolassa et al., 2023). A persistence-based benchmark, meaning finding and copying days that are more similar than simply the weekly pattern, is used in a recent forecasting competition (Farrokhabadi et al., 2022). It can lead to a more accurate Naïve benchmark but at a higher cost of implementation and reduced transferability to other cases, therefore not selected in this study.

The machine learning algorithm MLR is widely used for electric load forecasting and produces forecasts at low computational cost (Kuster et al., 2017). Eq. 1 shows MLR with two independent variables (Hong et al., 2010) as an example:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e \tag{1}$$

where Y is the dependent variable, X_1 and X_2 are independent variables, βs are parameters to estimate, and e is the error term. See Supapo et al. (2017) for a more detailed explanation of MLR. Even though it cannot capture nonlinear relationships by definition, MLR is used because of its scalability and interpretability, while also achieving state-of-the-art performance in many cases. The AI-model LGBM, on the other hand, was highly represented in a recent energy predictor competition (Miller et al., 2020). It is recognized as suitable for electric power modeling, and explained in more detail in the open literature (Tan et al., 2021). One obvious benefit of this model is that it can capture non-linear relationships while still remaining computationally feasible.
The proven track record and community support alongside their simplicity (no hyperparameter tuning), and compatibility (use of the same past and future covariates) conclude that MLR and LGBM are suitable for this comparative study. The models are available in the Python library Darts, which is used in this study (Herzen et al., 2022).

2.4 Algorithm creation

The algorithm (referred to as *Historical Forecasts* in this study) is depicted in Fig. 2. First, necessary inputs are given to the algorithm; forecast horizon, size of historical load for training, number of lagged (past) target values to use, how many hours to jump before making a new prediction, how many predictions to make before retraining, and when to stop. Future and past covariates including their lags can also be given to the model, e.g. temperature and day of the week. A prediction start date is given for splitting the data, otherwise, it will start as soon as possible given the size of the training and available data set. The model is trained, and predictions are made according to the inputs, and at some points retrained. *Historical Forecasts* uses a rolling window approach for the rolling origin evaluation of the forecast (Hewamalage et al., 2023). Each prediction, error, and error metric are saved for further analysis.

2.5 Forecast evaluation: the full-year run

A rolling origin evaluation of the model is applied via the *Historical Forecasts* algorithm using a forecast horizon of 168h, jumping 17h forward between

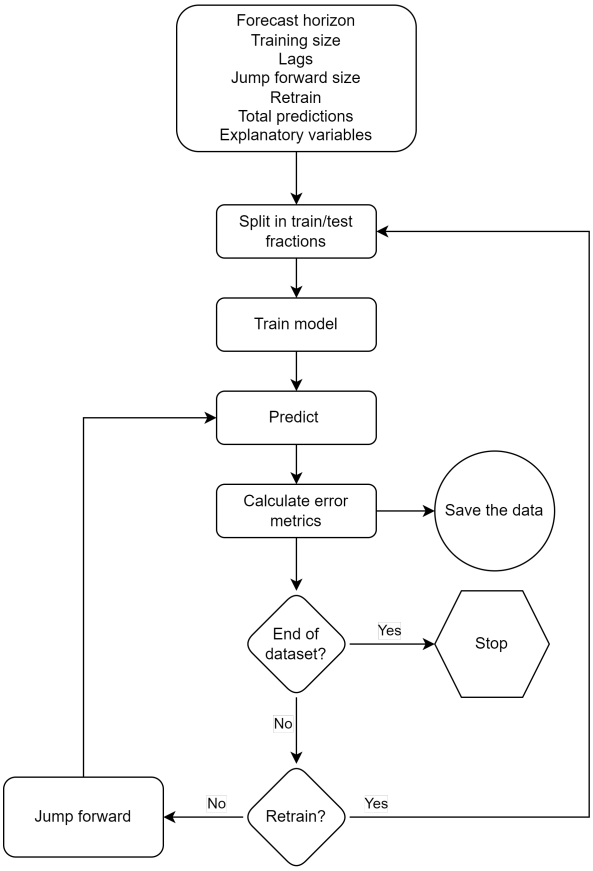

Figure 2. Flowchart of the algorithm *Historical Forecasts* created for this study.

each prediction, retraining every 100th prediction and hence doing a total of 515 predictions. The forecast horizon is selected to match the horizon of available weather forecasts, and the jump between predictions is chosen as a prime number to minimize the chance of resonance with any of the seasonal patterns. Past lags for the load and lags for the future- and past covariates are set to 168h. The evaluated period is approximately 1 year and 1 week, referred to as "the full-year run".

Given the size of available data, the maximum training size is approximately one year, one month, and two weeks for doing a full-year run. Varying the training size between lower than a year, one year, one year plus two weeks, and maximum size, the "one year plus two weeks" gave the best accuracy. Including a year of training data and predicting a week ahead means the model has seen the predicted week once, but adding at least one more week to the training data means the predicted week has been seen twice. Including maximum available data showed no significant accuracy improvement. This study is not aiming to prove the optimal training size, as there are many other possible approaches that have not been evaluated. The models are trained four times over the course of the full-year run, as the load profile and temperature dependency is known to be different for the four seasons. During the analysis where the number of retraining was varied, it was shown that re-training too often (every day or week) did not necessarily have a positive effect on accuracy, and certainly not on computational expense. Not re-training at all gave increasingly diverging errors, therefore the final re-training is set to four times.

For every full-year run, 86 520 errors (multiply forecast horizon by the total number of predictions) are analyzed, together with 515 average errors (one for each prediction made), and a single average error. Which error metric to be used for different datasets can be derived from Hewamalage et al. (2023). Root-Mean-Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are concluded as the two most common ones used for STLF of electrical load in Nti et al. (2020), the former used in this study and presented in MW. The relative RMSE (rRMSE), defined as the RMSE of MLR and LGBM respectively divided by RMSE for the "weekly Naïve", is used to quantify the performance against the benchmark.

An extensive analysis is done where the least computationally expensive model MLR is used for running hundreds of full-year runs, each generating errors that are compared. Periods with the largest errors, such as public holidays, are focused on separately, as well as the yearly peak, and the summer period. One explanatory variable is added after the other manually, including several combinations, the model parameters are varied, and the results are evaluated. Through combinations of visual inspection of the animations and plotting the model errors in different graphs, calculating and comparing the error metrics, the key explanatory variables are concluded. When no significant improvement is achieved with this semi-structured scrutiny of the MLR, the analysis is stopped. The same analysis is not done with the LGBM due to the computational expense, where only a few selected parameter changes are made to verify the model behavior, e.g. reducing training size reduces accuracy.

3 Results and Discussion

3.1 *Final set of explanatory variables*

By applying the methodology and analyzing the results, eight explanatory variables are selected, denoted as "the final set", shown in Table 1. The impact of adding each explanatory variable to the models is analyzed. The first row of the table shows the average RMSE for the full-year run, including only one explanatory variable; "Day of Week". In the second row, the "Hour of day" is added to the models, and the resulting full-year run RMSE is presented. Consequently, the models in the last row contain all the seven above explanatory variables, including the eighth, Wind speed.

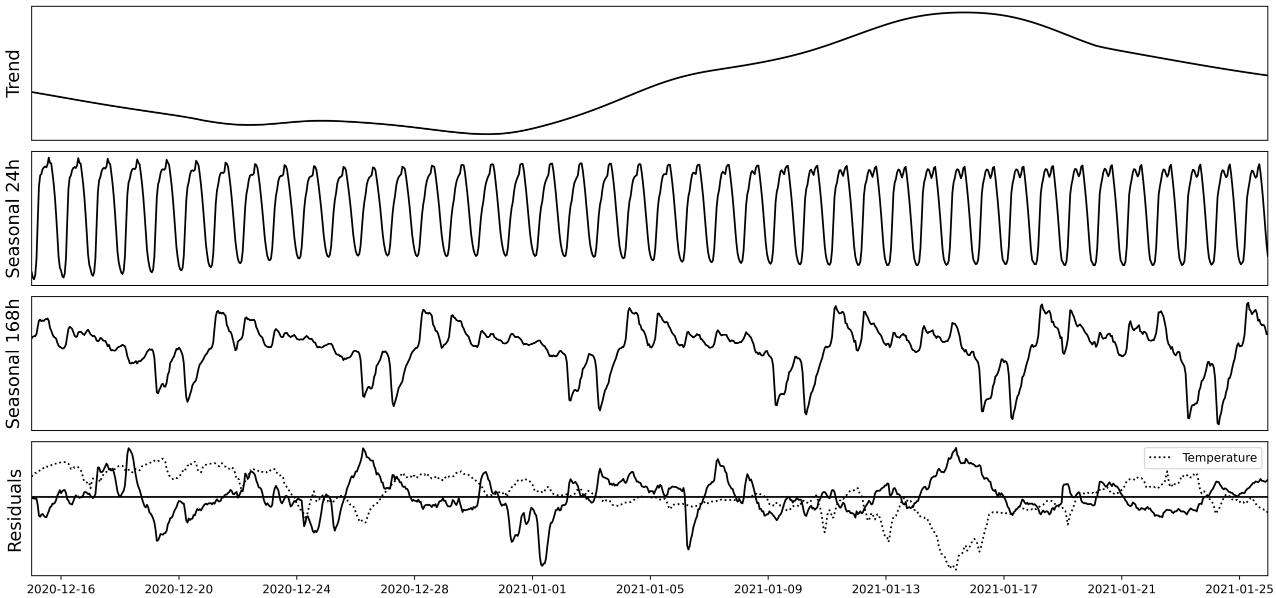

Figure 3. Additive load MSTL with seasonality periods of 24 and 168. Temperature is added to the bottom residuals graph.

Table 1. Final set of explanatory variables and the full-year run average RMSE (MW) for both MLR and LGBM, consecutively adding the explanatory variables in order of appearance

Explanatory variable	MLR	LGBM
Day of week [0-6]	4.54	4.74
Hour of day [0-23]	4.56	4.72
Holidays [0 OR 1]	4.44	4.53
Industry vacation [0 OR 1]	4.31	4.35
Heating hours [Kh, < 10°C]	2.20	2.89
Global irradiance [W/m^2]	2.12	2.68
Cooling hours [Kh, > 20°C]	2.06	2.65
Wind speed [m/s]	2.04	2.71

The choice of the calendar variables ”Day of week” and ”Hour of day” as explanatory variables are justified with the load decomposition, as a daily and weekly seasonal pattern is shown in Fig. 3. Analyzing the bottom residuals graph shows a negative correlation with temperature for this winter example. When not explained by temperature, large peaks in the residuals can be explained with knowledge of public holidays (Christmas and New Year). Further, there is a significant reduction in load due to the common industry practice of closing their operations during the summer vacation period. A binary variable which is set to zero for those four weeks is added, further improving the accuracy shown in Table 1. A variable for covering the thermal load is needed, as electricity is used for heating and cooling. Degree hours are part of the final set, as they give better results than degree days and temperature. Global irradiance and Wind speed improve the accuracy, apart from several of the other meteorological variables, and are therefore included. For justifying degree hours and the use of Holidays, MSTL is applied to the entire dataset, and the residuals are plotted against outdoor temperature in Fig. 4, with public holidays plotted separately. A portion of the residuals are significantly lower than the rest of the residuals during public holidays. Excluding public holidays and adding a LOESS line of the best fit gives a curve explaining how the residuals vary with temperature, depicted as ”Smoothed” in Fig. 4. Residuals are negatively correlated with temperatures below 10°C while positively correlated with temperatures above 20°C.

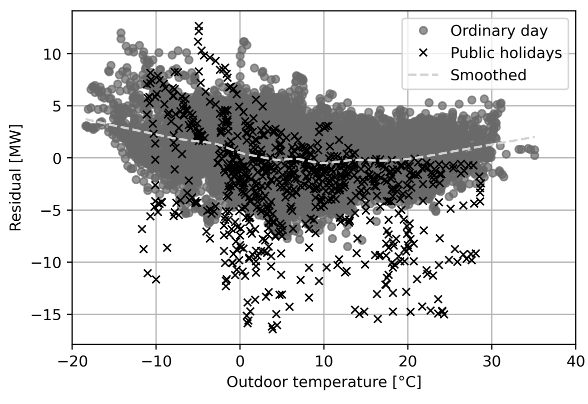

Figure 4. MSTL residuals vs temperature.

3.2 *Impact of explanatory variables*

Adding certain explanatory variables means only a slight improvement in the full-year run accuracy, and their existence in the final set needs to be questioned. Adding Wind speed in LGBM reduces large errors for some hours of the year at a cost of a higher average error for the full-year run. Quantifying the economic impact of reducing high errors for a few e.g. windy days, at the cost of a slightly worse overall performance, is a possible way to solidify the existence, and estimate the worth, of the explanatory variables.

The addition of the second explanatory variable ”Hour of day” means that the forecast is performing worse for the MLR as seen in Table 1. However, ”Hour of day” is making the LGBM forecast better and therefore kept in the final set, also because it shows a correlation with the electrical load in the data exploration. Discussions with the stakeholders about the future use of the models can also help in determining whether an explanatory variable should be included in the model.

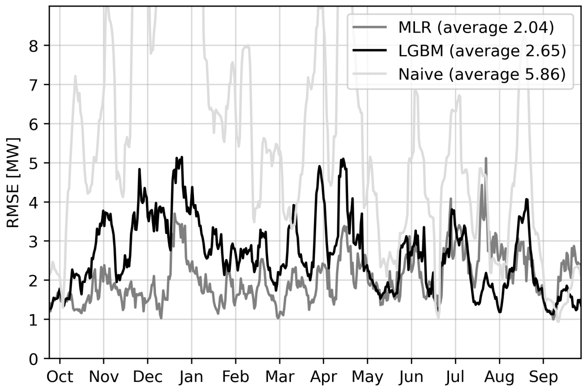

Figure 5. RMSE for MLR, LGBM and ”weekly Naïve”.

3.3 *Model results comparison*

The results in this paper show that there is no one-size that fits all. When comparing on a single error metric for the full-year runs, the MLR is concluded as superior in terms of accuracy over the LGBM. This despite the fact that a single metric is not giving any detailed insights into the performance of each model. Comparing the performance over the course of the year gives different winners for different periods.

The full-year run results for comparing MLR with the LGBM, including the Naïve benchmark, are shown in Fig. 5. The best-performing models according to Table 1 give an average RMSE of 2.04 (rRMSE of 35%) for MLR and 2.65 (rRMSE of 45%) for LGBM, compared to 5.86 for ”weekly Naïve” in the full-year run. MLR is performing better for three of the four seasons of the year, while LGBM is periodically more accurate during summer, as Fig. 5 shows. A single noon peak during summer, and morning and afternoon dual peaks for the rest of the year in the dataset could be a reason why the LGBM outperforms MLR during periods of the summer and vice versa.

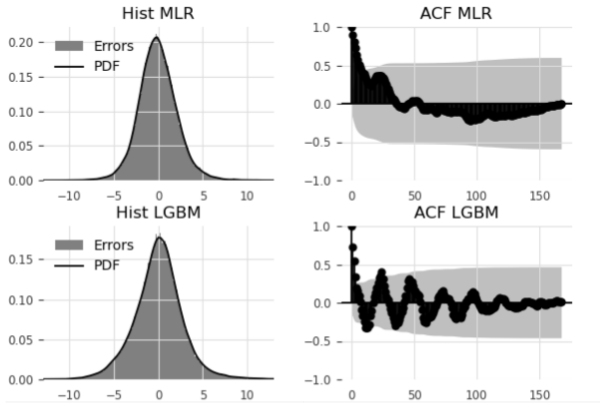

Figure 6. Left) Probability Density Function of the errors. Right) AutoCorrelation Function of the errors.

Another way of comparing the two best-performing models is by analyzing the shape of the histogram of the errors (all 86 520 errors for the full-year run), as seen to the left in Fig. 6. Both models produce errors close to a normal distribution centered close to zero for the full-year run. The centers of the distributions are slightly tilted towards a negative number for MLR, and a positive number for LGBM.

The autocorrelation plots, to the right in Fig. 6, are shown for the same prediction (out of the 515 predictions made, i.e. the 55th) for both models. First, they show that most of the past (lagged) errors are not significantly autocorrelated, except for the first 3–10 errors. This is concluded as serial autocorrelation, meaning if the model is wrong in one direction for the first time step, it will likely be wrong in the same direction in the next step. Second, seasonal autocorrelation is also observed, meaning if the prediction is too low one day, it is likely to be too low on the following day, in a seasonal pattern. Third, these plots highlight that both forecast models produce different errors from an autocorrelation perspective, and are therefore suitable for combining.

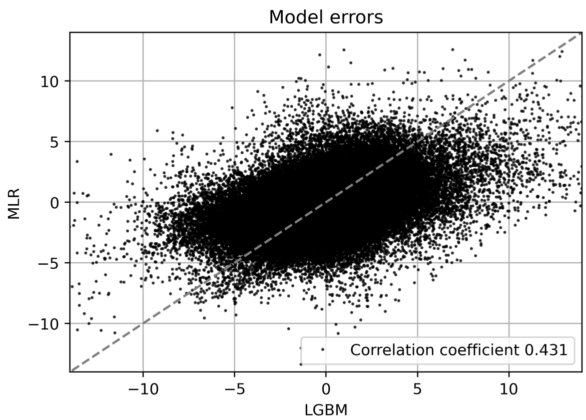

Figure 7. MLR vs LGBM full-year run model errors.

The errors from the two best-performing models are plotted against each other in Fig. 7, and they show a weak correlation. This, together with the distribution of the errors in Fig. 6, shows good potential for combining into an ensemble model. Accuracy improvements are expected when combining models according to the literature (Wang et al., 2018) and forecasting competitions (Miller et al., 2020) but a deeper analysis of this specific case study is needed. The MLR is the winner computationally-wise, it takes about 60 times more time for the LGBM to finish the full-year run.

3.4 Grouped forecasting

The available data are spatially separated and grouped forecasting is applied. The same explanatory variables as the best-performing MLR model and the same settings, e.g. forecast horizon, have been used for the modeling of each individual transformer station. A full-year run is made for all seven transformer stations and the predictions are added together, called Predict Then Sum (PTS). This is compared to the model trained on the sum of the individual transformer stations, called Sum Then Predict (STP), which is slightly different from the total energy usage used in this paper (see Methodology for explanation). The difference between PTS and STP is larger during the heating period, shown in Fig. 8. In general, the grouped PTS forecasting method performs worse, apart from a few exceptions for the full-year run.

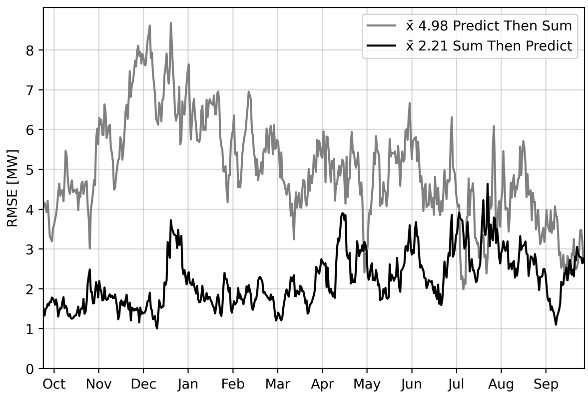

Figure 8. RMSE when forecasting each individual transformer station separately versus their sum.

Comparing PTS with STP show that forecasting on a more aggregated scale is preferred. The correlation between the electrical load and each transformer station, presented in the Methodology section, varies between -0.32 and -0.80. This suggests that the importance of temperature (or temperature variants such as degree hours) when describing the load can vary considerably. Knowledge of each PV park and customized models for each transformer station could improve accuracy further.

3.5 Model dynamics

The algorithm *Historical forecasts* can produce animations of the predictions plotted against the actual load. In Fig. 9, frame 6 out of 25 from an example prediction, which includes the yearly peak of 2021, is shown. The best-performing MLR model is used to produce this frame. The inputs used can be seen at the top left, and the error metrics calculated for this frame are at the bottom left. The bottom error graph shows the difference between the prediction and the actual load for the 168h forecast horizon framed between the two vertical dashed lines. Zooming in around the day of the yearly peak and analyzing the accuracy of the prediction three days before shows that the model underpredicts with approximately 3 MW and 31 MWh for the entire day. If the same analysis is done a few frames later, just 12h before, the numbers change to 1 MW and 10 MWh respectively.

During visual inspection of the predictions, dynamic behaviors are revealed. In some cases, when the model underpredicts the load on the next day, it also underpredicts the day after that, and the following, meaning the average prediction is too low. Dealing with seasonal data makes the presence of seasonal autocorrelation expected. Not until the first day of underprediction has passed does this level out and the model corrects the average level to fit the upcoming days better. This is a good simulation example of how the forecast would have reacted in such a case, it does not know it is underpredicting until the days pass.

Another important performance indicator, which is not straightforwardly easy to measure, is the trustworthiness and explainability of the models. The LGBM produces predictions that do not have a smooth pattern, meaning the first derivative of the predictions during midday alters between positive and negative values consecutively. MLR on the other hand produces predictions where the first derivative less often changes sign and can be seen in Fig. 9. Introducing a forecast model to decision-makers or operators, which don't like or trust it, could affect its usefulness, success, and arguably profitability (Kolassa et al., 2023).

4 Summary and Conclusions

In this paper, a framework to analyze and evaluate forecasting models is explored. The performance of two models, MLR and LGBM, are evaluated using a dataset from the local grid operator of Eskilstuna. Different sets of explanatory- and model variables are tested, concluding the calendar variables; "Day of week", "Hour of day", Holidays and Industry vacation period, and the meteorological variables; Global irradiance and Wind speed together with the cross-effect variables; Heating hours below 10°C and Cooling hours above 20°C as the final set. While

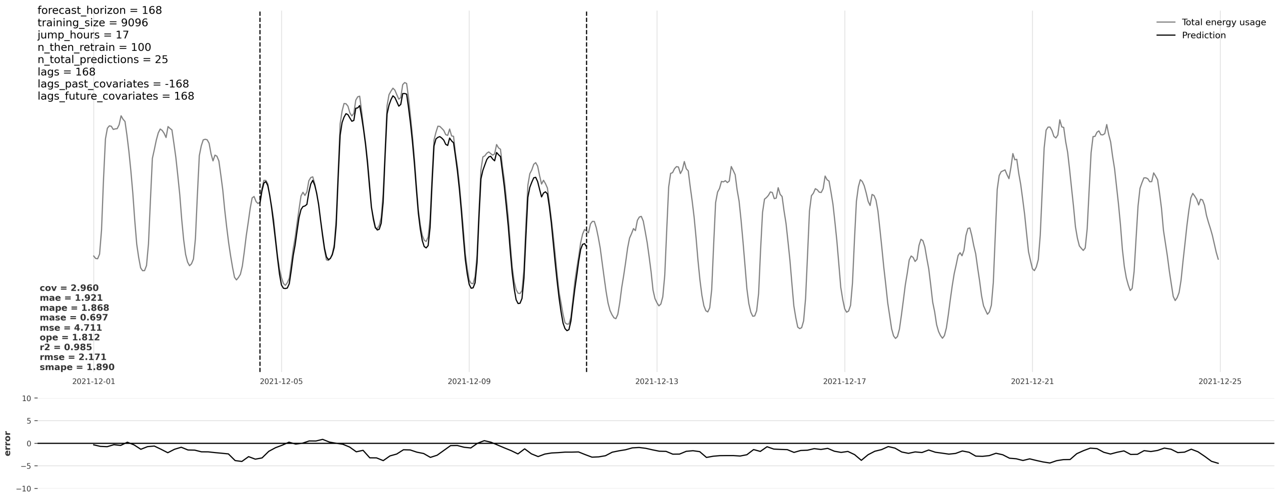

Figure 9. A frame of the animation produced by the algorithm Historical Forecast.

public holidays and non-typical periods still cause the largest errors, adding binary explanatory variables for these significantly improves accuracy. The best-performing MLR and LGBM models outperform the ”weekly Naïve” benchmark model with rRMSE of 35% and 45% respectively. MLR is producing lower errors compared to the more computationally expensive LGBM for the heating period, while it is difficult to unanimously declare a winner for the summer period. Adding an economical dimension can help in determining the acceptable level of accuracy, and suggested measures for enhancing the models is hierarchical and ensemble forecasting. Adding further models (e.g. Artificial Neural Networks based), has the potential to improve the accuracy. Techniques for selecting training data, and optimizing re-training intervals can be investigated further. Expanding the study to include more forecasting models and techniques, and additional explanatory variables (e.g. the national forecast of PV production), would be an interesting path to deepen the knowledge of this specific case.

4.1 The road ahead and future usage

The electrical grid environment is rapidly evolving. Changes in usage patterns, price volatility, and the installation of intermittent renewable energy are just some of the factors that affect the future. In that context, an important aspect when deciding the best model is the ability to adapt and change to better fit the load evolution, adaptability was a key concern during the Covid-19 period (Farrokhabadi et al., 2022). MW-size PV and wind parks are being commissioned from one day to the next, making the historical data less relevant for forecasting. Although important, this study is not evaluating the model's adaptability, or robustness.

Before implementing forecasts in real-life applications, such as planning and controlling electrical energy storage, an economic dimension should preferably be added to the analysis. Large economic penalties can be the consequence of underpredicting the annual peak on the grid. The best-performing MLR model produced an error in the order of 3 MW for the yearly peak and 31 MWh for that entire day, three days in advance. Depending on the size and availability of local storage, this may or may not be acceptable. This despite the fact that the model accuracy seems to be in line with, or even better than, the overall Swedish national load forecast produced by Svenska kraftnät (Kazmi & Tao, 2022). Forecasting national load is arguably an easier task due to its long history and considerably more in-house knowledge. Internal discussions with the local grid operator of Eskilstuna suggest that the errors are not acceptable, considering possible future energy storage investments. Consequently, point load forecasts, produced here, could be used for everyday short-term management and hourly spot-price optimization, while other methods should be used for peak prediction and peak-shaving as shown in the literature (Gajowniczek & Ząbkowski, 2017).

A future energy storage system will be connected to one of the seven transformer stations. Depending on the location, the forecasting method needs to be considered and customized. This customization, and detailed information about PV parks, can result in high-accuracy forecasts (Hong et al., 2020). Connecting these forecasts to an energy storage management- and sizing problem is the potential next step of this study.

Acknowledgement

The authors hereby thank ESEM for their support, Christer Wiik and Sara Jonsson from the Electrical grid department, and Per Örvind from the Energy department. Thanks to Dr. Moksadur Rahman and Prof. Erik Dahlquist for valuable input and discussions.

References

Beiron, J., Montañés, R. M., Normann, F., & Johnsson, F. (2020). Combined heat and power operational modes for increased product flexibility in a waste incineration plant. *Energy*, *202*, 117696. doi: https://doi.org/10.1016/j.energy.2020.117696

Cebulla, F., Naegler, T., & Pohl, M. (2017). Electrical energy storage in highly renewable european energy systems: Capacity requirements, spatial distribution, and storage dispatch. *Journal of Energy Storage*, *14*, 211–223. doi: https://doi.org/10.1016/j.est.2017.10.004

Chabouni, N., Belarbi, Y., & Benhassine, W. (2020). Electricity load dynamics, temperature and seasonality nexus in algeria. *Energy*, *200*, 117513. doi: https://doi.org/10.1016/j.energy.2020.117513

Eroshenko, S. A., Poroshin, V. I., Senyuk, M. D., & Chunarev, I. V. (2017). Expert models for electric load forecasting of power system. In *2017 ieee conference of russian young researchers in electrical and electronic engineering (eiconrus)* (pp. 1507–1513). St. Petersburg and Moscow, Russia. doi: 10.1109/EIConRus.2017.7910859

Farrokhabadi, M., Browell, J., Wang, Y., Makonin, S., Su, W., & Zareipour, H. (2022). Day-ahead electricity demand forecasting competition: Post-covid paradigm. *IEEE Open Access Journal of Power and Energy*, *9*, 185–191. doi: 10.1109/OAJPE.2022.3161101

Gajowniczek, K., & Ząbkowski, T. (2017). Two-stage electricity demand modeling using machine learning algorithms. *Energies*, *10*(10), 1547. doi: https://doi.org/10.3390/en10101547

Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G., … others (2022). Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research*, *23*(1), 5442–5447. doi: 10.48550/arXiv.2110.03224

Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: common pitfalls and best practices. *Data Mining and Knowledge Discovery*, *37*(2), 788–832. doi: https://doi.org/10.1007/s10618-022-00894-5

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, *32*(3), 914–938. doi: https://doi.org/10.1016/j.ijforecast.2015.11.011

Hong, T., Gui, M., Baran, M. E., & Willis, H. L. (2010). Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *Ieee pes general meeting* (pp. 1–8). Minneapolis, MN, USA. doi: 10.1109/PES.2010.5589959

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., & Zareipour, H. (2020). Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, *7*, 376–388. doi: 10.1109/OAJPE.2020.3029979

Huber, M., Dimkova, D., & Hamacher, T. (2014). Integration of wind and solar power in europe: Assessment of flexibility requirements. *Energy*, *69*, 236–246. doi: https://doi.org/10.1016/j.energy.2014.02.109

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts: Melbourne, Australia.

Işık, G., Öğüt, H., & Mutlu, M. (2023). Deep learning based electricity demand forecasting to minimize the cost of energy imbalance: A real case application with some fortune 500 companies in türkiye. *Engineering Applications of Artificial Intelligence*, *118*, 105664. doi: 10.1016/j.engappai.2022.105664

Kazmi, H., & Tao, Z. (2022). How good are tso load and renewable generation forecasts: Learning curves, challenges, and the road ahead. *Applied Energy*, *323*, 119565. doi: https://doi.org/10.1016/j.apenergy.2022.119565

Kolassa, S., Rostami-Tabar, B., & Siemsen, E. (2023). *Demand forecasting for executives and professionals* (1st ed.). Chapman and Hall/CRC: Boca Raton, United States.

Kuster, C., Rezgui, Y., & Mourshed, M. (2017). Electrical load forecasting models: A critical systematic review. *Sustainable cities and society*, *35*, 257–270. doi: https://doi.org/10.1016/j.scs.2017.08.009

Meliani, M., Barkany, A. E., Abbassi, I. E., Darcherif, A. M., & Mahmoudi, M. (2021). Energy management in the smart grid: State-of-the-art and future trends. *International Journal of Engineering Business Management*, *13*, 18479790211032920. doi: https://doi.org/10.1177/18479790211032920

Miller, C., Arjunan, P., Kathirgamanathan, A., Fu, C., Roth, J., Park, J. Y., … others (2020). The ashrae great energy predictor iii competition: Overview and results. *Science and Technology for the Built Environment*, *26*(10), 1427–1447. doi: https://doi.org/10.1080/23744731.2020.1795514

Nik, V. M., Perera, A., & Chen, D. (2021). Towards climate resilient urban energy systems: a review. *National Science Review*, *8*(3), nwaa134. doi: https://doi.org/10.1093/nsr/nwaa134

Nti, I. K., Teimeh, M., Nyarko-Boateng, O., & Adekoya, A. F. (2020). Electricity load forecasting: a systematic review. *Journal of Electrical Systems and Information Technology*, *7*(1), 1–19. doi: https://doi.org/10.1186/s43067-020-00021-8

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th python in science conference* (Vol. 57, pp. 10–25080). Austin, TX. doi: 10.25080/Majora-92bf1922-011

SMHI. (2023). *Open data api docs - meteorological forecasts.* Retrieved from `https://opendata.smhi.se/apidocs/metfcst/index.html`

Supapo, K., Santiago, R., & Pacis, M. (2017). Electric load demand forecasting for aborlan-narra-quezon distribution grid in palawan using multiple linear regression. In *2017ieee 9th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (hnicem)* (pp. 1–6). Manila, Philippines. doi: 10.1109/HNICEM.2017.8269480

Tan, Y., Teng, Z., Zhang, C., Zuo, G., Wang, Z., & Zhao, Z. (2021). Long-term load forecasting based on feature fusion and lightgbm. In *2021 ieee 4th international conference on power and energy applications (icpea)* (pp. 104–109). Busan, Republic of Korea. doi: 10.1109/ICPEA52760.2021.9639313

Wang, Y., Chen, Q., Sun, M., Kang, C., & Xia, Q. (2018). An ensemble forecasting method for the aggregated load with subprofiles. *IEEE Transactions on Smart Grid*, *9*(4), 3906–3908. doi: 10.1109/TSG.2018.2807985

Öhman, A., Karakaya, E., & Urban, F. (2022). Enabling the transition to a fossil-free steel sector: The conditions for technology transfer for hydrogen-based steelmaking in europe. *Energy Research & Social Science*, *84*, 102384. doi: https://doi.org/10.1016/j.erss.2021.102384