Improving short-term electricity forecast on smart meter data using graph neural networks

Sascha Petznick (6106749), Soham Prajapati (5712297),
Varun Pradhan (5655263), Henrik Brinch van Meenen (6070698)

Abstract

In response to the increasing demands for power and sustainability, smart meters have emerged as a crucial technology for capturing data that can be utilized for forecasting consumer demands. This forecasting aids utility companies in ensuring efficient and stable power delivery. This study utilizes anonymized timeseries data of 5567 household smart-meters in the London area, collected from November 2011 to February 2014, for short-term forecasting. An abstract network-based strategy is employed to capture temporal similarities between households, constructing a graph for analysis. Dynamic Graph-Vector Autoregressive Recursion (G-VAR) method is proposed to perform the forecasting and it compared with Seasonal Autoregressive Integrated Moving Average (SARIMA) method. Although the forecasting results are not optimal, spectral analysis of the data provides insights to support these findings.

1 Introduction

The adoption of renewable energy is transforming power grid operations, with even non-industrial consumers now acting as both consumers and producers via photovoltaic panels and small wind turbines. While environmentally beneficial, this shift complicates energy availability, frequency and voltage stability, and production efficiency for network operators. To manage these challenges, operators must predict consumer load within short-term time frames of up to six hours to effectively plan conventional power plant dispatch. In this scenario, smart meters can help by measuring household power consumption with high temporal resolution, alerting consumers to outages, and sending usage data to utilities multiple times an hour.

Forecasting can then be performed using classic timeseries techniques like SARIMA or nonlinear neural network techniques like Long Short-Term Memory (LSTM) networks [1], on the measured timeseries, which allows network operators to ensure stable dispatch. However, one drawback of these methods is their ability to incorporate and/or model dependencies between timeseries. For example, it is highly likely that households in close locality have similar consumption profiles. Such prior knowledge can be incorporated by using graph-based method making the forecasts more accurate.

There are already examples of this in literature. For example recent work by Lin et. al., [5] called Graph Wavenet. Which performs load forecasting with a self-adaptive adjacency matrix on anonymised data. This matrix is proposed to explore the hidden spatial dependency between nodes without any prior information. Other works like [8] improve this method by proposing dynamic adjacency matrix based on additive attention mechanism. Rather than using plain GNNs they use exponential moving average graph convolutional neural network (EMA-GCN) to avoid over-smoothing. These methods have the inherent problem of trying to learn a high number of parameters due to estimating the adjacency matrix. Additionally, they ignore the possibility for the graph topology to change over time. In a grid context this will happen as customers change providers, new housing is constructed, etc. and so time-varying topology is important to account for.

Preprint. Under review.

In this project, a more interpretable and less computationally expensive method is proposed, by extending the Graph-Vector autoregressive (G-VAR) model to support time-varying topology, to perform short-term forecasting of six hours. This approach enhances robustness to irregular data and eliminates the need for data imputation. Key findings include an investigation into low forecasting accuracy, utilizing techniques such as spectral analysis of the graph data and an examination of the learned graph filters.

2 Methods

We first define the mathematical notation to then introduce the dynamic G-VAR. Consider a weighted undirected graph: $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W}_{\mathcal{G}})$, where $\mathcal{V}$ is the set of N nodes, $\mathcal{E}$ is the edge set, and $\mathbf{W}_{\mathcal{G}}$ is the graph shift operator which is considered here as the adjacency matrix with $[\mathbf{W}_{\mathcal{G}}]_{i,j} = [\mathbf{W}_{\mathcal{G}}]_{j,i} = 1$ if $(i, j) \in \mathcal{E}$ and zero otherwise. For this work, the nodes of the graph will be the smart-meter of a given household. In order to create a graph based on the individual timeseries, from the analysis provided in section 3.1, the households with similar installation dates can be assumed to be located close to each other. This hints a geographical proximity of the households. This idea was used for constructing the adjacency matrix $\mathbf{W}_{\mathcal{G}}$. It employs a k-nearest neighbor graph, where nodes are sorted by their enabled date. An edge is formed between nodes if their start dates are among the k closest start dates, with an edge weight of 1. The graph's sparsity is controlled by k, a hyperparameter set to 50, resulting in a 94%.

Now, a snapshot-based temporal graph can be defined at any time t, as $\mathcal{G}_T^S = \{\mathcal{G}_t\}$ which accounts for only those nodes which are active at a particular time instant defined as $|V|_t$. For these nodes, using the similarities from $\mathbf{W}_{\mathcal{G}}$, makes a sub-adjacency matrix $\mathbf{W}_{\mathcal{G}_t}$. The spectral analysis was conducted on the graph using a snapshot where the maximum number of nodes were active, specifically at index 7752 on November 19, 2012, 00:00:00, with 5526 active nodes. The eigendecomposition of the Laplacian matrix $\mathbf{L}_{7752} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$. And the graph discrete Fourier transform (GFT) coefficient showing the signal variability spectrum of the corresponding signal $\hat{\mathbf{x}} = \mathbf{V}^H\mathbf{x}_{7752}$. The GFT coefficient $\hat{\mathbf{x}}_k$ is plotted with the eigenvalue λ_k in figure 1. A high zero frequency component indicates smoothness of the graph signal, implying minimal variation between adjacent nodes. This suggests that the signal values are not strongly influenced by the graph's topology, as the Laplacian acts as a differential operator capturing signal variations across the graph. This observation raises concerns about the suitability of graph-based methods for this dataset. However, as shown in 1 eigenvalues around 2000 and 4000 indicate higher frequency components in the signal, suggesting potential advantage of graph-based methods for further exploration.

Moving on, we modify the formulation introduced in [4] to suit dynamic topology and predicting at tth instant gives dynamic G-VAR (DG-VAR) as, $\hat{\mathbf{x}}_t = -\sum_{p=1}^{P}\sum_{k=0}^{K} h_{kp}\mathbf{S}_{t-p}^k\mathbf{x}_{t-p}$ where, $\mathbf{S}_{t-p}$ is the normalized aligned sub-adjacency matrix at $(t-p)$th time instant: $aligned(\mathbf{D}^{-1/2}\mathbf{W}_{\mathcal{G}_{t-p}}\mathbf{D}^{-1/2}$). Due to dynamic topology and changing sizes of $\mathbf{W}_{\mathcal{G}_t}$, sub-adjacency matrices and graph signal x_{t-p} are aligned based on the union of active nodes N_{act} from the last P timestamps, ensuring compatible sizes across snapshots. This involves creating larger adjacency matrices to accommodate all nodes, filling zeros where nodes are inactive. The is aligned similarly. equal total $(K+1)P$ parameters.

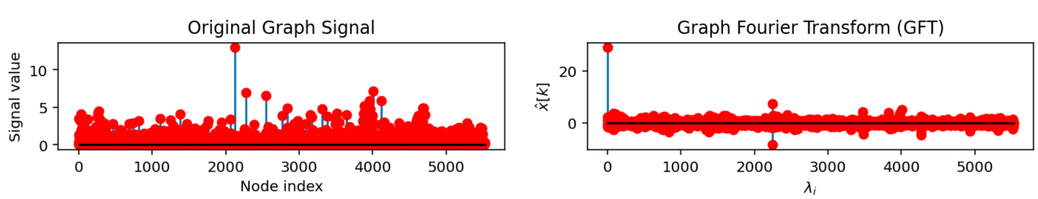


Figure 1: Spectral analysis of the graph data at index 7752

This formulation predicts only one-step ahead, to extend to a six-hour horizon, we can apply it iteratively: using the predicted value from each step as the input for the next step for a stable forecast. In this iterative approach, each prediction $\hat{\mathbf{x}}_{t+i}$ utilizes the latest adjacency matrix $\mathbf{S}_{t-1}$ from the given window, focusing solely on active nodes N_{act} across the last P instants. This approach assumes the 1-hour prediction is limited to nodes in N_{act}, disregarding link prediction tasks and potential fluctuations in node count within the targets. Finally, the loss function considered here is $\mathcal{L} = \frac{1}{H}\Sigma_{\tau=1}^{H}||\hat{\mathbf{x}}_{t+\tau} - \mathbf{x}_{t+\tau}||_2^2$ since this is a node regression problem.

For the SARIMA model, we look at individual timeseries of each household.The univariate seasonal ARIMA model is readily available and of the form as given in [3]. In order to identify the seasonal and non-seasonal coefficients of the model, we employ a grid search over the parameters until order 3 and minimise the Akaike information criterion [7] to find the most suitable model. This procedure works for every time-series individually. However, the analysis might indicate that the same model order could be used across all or at least most households.

Within the work of this report, we will address the issue of explainability. In this project's context, explainability will be mainly concerned with the importance of neighbours' information of a node for the six hour prediction of that node leading to a *local explanation*. This can be interpreted as looking for a sparse explanation. Of course, there might be the chance that the importance is spread across neighbours, and instead, the most recent timesteps of each neighbour are considered the most important ones.

In section 2, we introduced the assumption of spatial proximity based on the first day of communication. Now, it is reasonable to assume that nodes with the **same** installation date were installed in the same street or neighbourhood, however, nodes that are only close with our given metric do not necessarily need to be close because the installation of smart meters wouldn't require district-wise installation leading to no spatial proximity of the nodes. Thus, we will guide our investigation with the following hypothesis: We expect the most important features to be the ones of the node itself, followed by the nodes that share their installation date.

3 Numerical Experiments and Results

All the codes used to obtain the results in this work can be accessed at [6].

3.1 Dataset

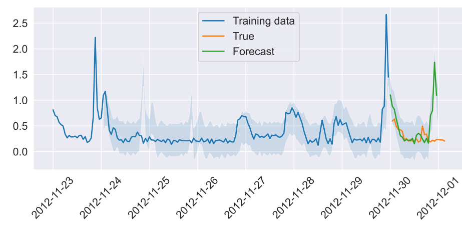
In this work, the dataset used from [2] provides anonymised power-consumption timeseries data sampled every 30 minutes for 5567 households in the London area. This data was collected between November 2011 and February 2014. This dataset is big (about 8GBs), so preprocessing of the data was performed based on the following observations, (1) The number of households active during the entire duration of data collection is varying and there were a few duplicates as well. (2) No data imputation was required since every timeseries is contiguous in the duration of the smartmeters transmit data. (3) About 95% of the data is contained in the year 2012-13. Therefore, the dataset was segmented for only this period. (4) It is a reasonable assumption that for six hour forecast, training the data sampled over an hour instead of 30 minutes will be adequate. Therefore, the data was resampled for 1 hr further reducing the size of the data. (5) Five smart-meters had zero duration of activity and were discarded from the dataset. Further analysis of the data included determining the start and end time when each smart meter enabled and disabled communication, respectively, to extract the installation time, discontinuation time, and total operational duration.

3.2 SARIMA

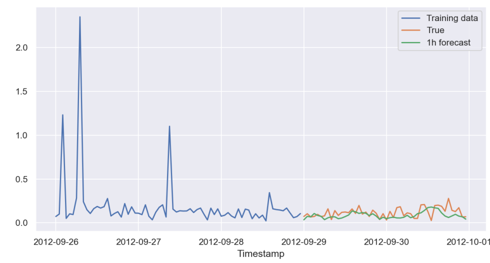
For the model selection of the SARIMA models, we decided to use daily seasonality based on the auto-correlation and partial auto-correlation. This insight will also support the use of a 24 hour observation window for the DG-VAR.

The models based on the AIC criterion varied quite a lot between households. As an example household, it was found to be $(p, d, q) = (1, 0, 1)$ and $(P, D, Q)_s = (1, 1, 0)_{24}$[1]. The resulting forecast for $24\,\mathrm{h}$ is displayed in Figure 2a. The scheme fails to predict the missing peak for the next day, leading to a higher prediction of power than in reality.

[1] See appendix for more insight.



(a) SARIMA model with $(p, d, q) = (1, 0, 1)$ and $(P, D, Q)_s = (1, 1, 0)_2 4$ for a representative household.



(b) DG-VAR model with K=2, P=24 for a representative household.

Figure 2: Both models were fit using the first 9 months of the year. Prediction is displayed for the next 24 hours with its 95% confidence interval for SARIMA and for the next 48 hours for DG-VAR. Note that we only display the last week of the training data for SARIMA and the last 3 days for DG-VAR.

The work presented here would be required for every single household for which we would like to make a prediction. Morevoer, the analysis showed that a single SARIMA model cannot even fit **most** households.

3.3 Dynamic G-VAR

Figure 2b shows the 1-hour forecast using DG-VAR with $K = 2$ and $P = 24$, using a 24-hour horizon to predict the output with 72 parameters. Training utilized 9 months of data (up to 2012-09-28) and testing 3 months. The results shown in the figure are not statistically significant. Additionally, with about 5000 active nodes at times, DG-VAR becomes computationally intensive due to matrix multiplications. Moreover, spectral analysis indicated that the graph topology may not offer significant advantages for this dataset which is proven here.

4 Discussion

The poor results of dynamic G-VAR may be due to the initial hypothesis, which assumed geographical similarity based on the time of installations, as supported by the spectral analysis. However, the results do not match our expectations; this could already be observed during training. The test loss did not decrease along the training loss.

When used to predict over a longer time interval of 6 hours, the predictions tend to be exaggerated. Since we're accumulating the outputs of the GCN layer over a 24-hour window for 1-hour predictions, any errors and deviations accumulate over time. This is a general downside of this approach as can be seen from the equations.

Moreover, there are some problems with the implementation of the approach. Computing the aligned and normalized sparse adjacency matrices leads to computationally intensive pre-computations. These computations also gave motivation not to proceed the initial idea of using a spatio-temporal graph structure with product graphs.

Addressing the challenge of explainability, the auto-correlation of the SARIMA model implied that lags of up to two are the most important for prediction. Based on that insight, we would expect the explanations generated for the DG-VAR to value the last two steps of the time-series the highest. We would expect a valid explanation to include these two input features as they carry significant statistical information.

Otherapproaches like EvolveGCN, product graphs, and Encoder-Processor-Decoder were considered. Product graphs were attempted but proved computationally expensive due to the Cartesian and Kronecker products over 5000-sized adjacency matrices. Additionally, sparse-tensor operations in Python do not support the necessary subscripting functionalities. The other two methods can be tried in future works.

5 CRediT author statement

Sascha Petznick challenge; data curation; method: S-ARIMA; results: S-ARIMA; discussion. **Soham Prajapati** data curation; conceptualization; method: G-VAR; results: G-VAR **Varun Pradhan** conceptualization; data curation; results: G-VAR **Henrik Brinch van Meenen**: conceptualization; optimization of code

References

[1] Riccardo Bonetto and Michele Rossi. Machine learning approaches to energy consumption forecasting in households. *arXiv preprint arXiv:1706.09648*, 2017.

[2] Greater London Authority. Smartmeter energy consumption data in london households, 2024. Accessed: 2024-05-19.

[3] R.J. Hyndman and G. Athanasopoulos. Forecasting: principles and practice, 2021. Accessed on June 30, 2024.

[4] Elvin Isufi, Andreas Loukas, Nathanaël Perraudin, and Geert Leus. Forecasting time series with varma recursions on graphs. *IEEE Transactions on Signal Processing*, 67(18):4870–4885, 2019.

[5] Weixuan Lin, Di Wu, and Benoit Boulet. Spatial-temporal residential short-term load forecasting via graph neural networks. *IEEE Transactions on Smart Grid*, 12(6):5373–5384, 2021.

[6] Sascha Petznick, Soham Prajapati, Varun Pradhan, and Henrik Brinch van Meenan. Dynamic short-term forecasting. `https://github.com/spetznick/ml4gd-project-group-3.git`, 2024.

[7] P. Stoica and Y. Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004.

[8] Chuyuan Wei, Dechang Pi, Mingtian Ping, and Haopeng Zhang. Short-term load forecasting using spatial-temporal embedding graph neural network. *Electric Power Systems Research*, 225:109873, 2023.

Appendix

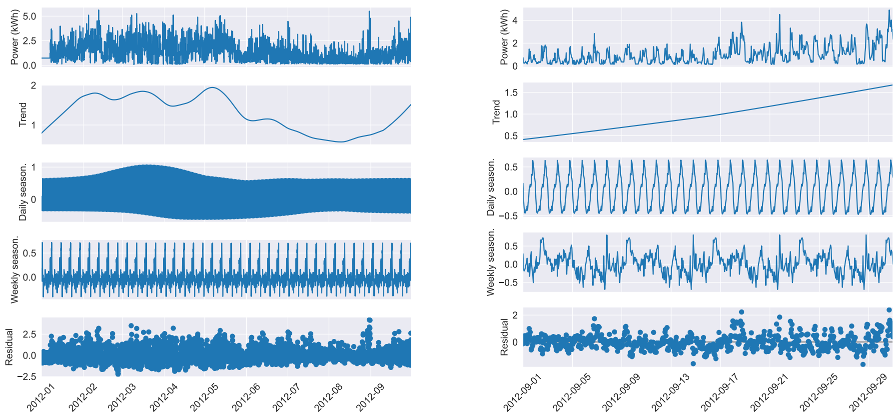


Figure 3: Multi-seasonal decomposition of a single household using daily and weekly seasons displayed for the whole year and a single week. Note that the leading data points were imputed with the first existing value of that timeseries.

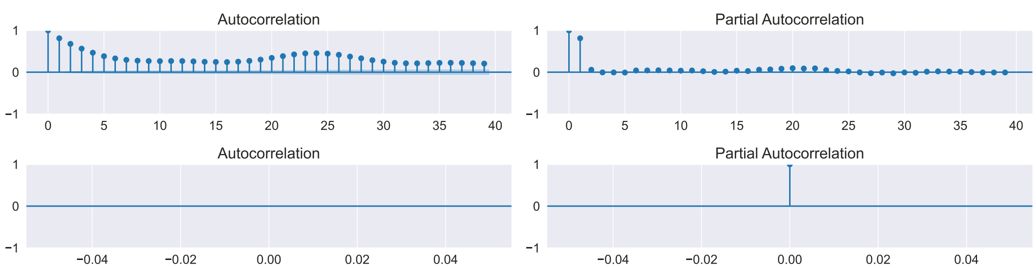


Figure 4: Autocorrelation and partial autocorrelation of a representative household (first row) and the according plot for the first-order difference of the data(second row).