

Distribution Network Assessment using a Causal-Guided Diffusion Model of Synthetic Residential Load Profiles

Maria Cassidy

*Electronic and Electrical Engineering, University of Strathclyde, UK,
E-mail: maria.cassidy-carrasco@strath.ac.uk*

Junyi Lu

Electronic and Electrical Engineering, University of Strathclyde, UK, E-mail: junyi.lu@strath.ac.uk

Blair Brown

Electronic and Electrical Engineering, University of Strathclyde, UK, E-mail: blair.brown@strath.ac.uk

Stephen D. J. McArthur

Electronic and Electrical Engineering, University of Strathclyde, UK, E-mail: s.mcarthur@strath.ac.uk

Bruce Stephen

Electronic and Electrical Engineering, University of Strathclyde, UK, E-mail: bruce.stephen@strath.ac.uk

Motivated by the need for decarbonization, power systems are increasingly dependent on advanced models to build reliable power distribution networks. As these distribution networks become increasingly data-driven, customer load profiles represent valuable resources for operational characterization and decision-making. Nevertheless, these datasets are impacted by privacy, economic and quality barriers, challenging their accessibility and re-use. Against this backdrop, the new range of high performing generative models have shown increasing ability to produce realistically synthetic data. Despite their potential, these models are challenged by their inherent complexity and opacity, particularly when processing high-dimensional and stochastic data. Consequently, this absence of transparency offers limited operational value when models are developed in a standalone manner, as they cannot provide the actionable guidance necessary for operators to make informed decisions. To address these challenges, this paper introduces a causal-guided generative model coupled with a power flow simulation for disentangled guidance of realistic synthetic smart meter data to support distribution network analysis and planning. The proposed approach combines a causal inference module with a conditional diffusion model to generate load profiles that are statistically accurate and causally consistent in both social practice and electrically coherent terms. This framework offers a better understood means of reliably augmenting limited smart meter datasets, with potential to support demand-side flexibility assessment, reinforcement studies and the evaluation of policy interventions across the power distribution system.

Keywords: Causal inference, LV networks, interpretability, Generative AI, Decision-making, reliability.

1. Introduction

Achieving carbon neutrality necessitates intelligent, adaptive power systems capable of managing increasingly volatile Low Voltage (LV) networks. LV networks have traditionally lacked extensive monitoring and, despite the recent deployment of smart meters, coverage remains incomplete and inconsistent across customer bases. Even when complete data exist, Distribution System Operators (DSOs) face substantial regulatory

barriers that limit the sharing and utilization of high-granularity load data for system planning. Synthetic data generation has emerged as a solution to augment limited datasets while preserving privacy, with Generative AI (GenAI) models becoming increasingly popular. However, existing data-driven approaches to load synthesis face two critical limitations.

First, as power systems undergo digitalization, ensuring their trustworthiness requires models

that are robust, and interpretable. However, current AI/ML models rely predominantly on correlation rather than causation. Causal learning enables models to respect underlying causal mechanisms, such as the directionality and magnitude of influences within specific contexts Rawal et al. (2025). In addition, causal understanding reflects human reasoning and enhances decision-making based on knowledge of the underlying causes Nauta et al. (2019). While effective for predictive tasks, correlation-based models struggle to disentangle the true drivers of demand. This often leads to misleading conclusions, such as incorrectly attributing load increases to routine rather than temperature effects, given their status as covariates. Causal inference offers a mechanism to transcend simple correlation, mitigating bias, and providing actionable insights into the operational "physics" of the grid.

Second, regarding the synthesis of domestic loads, most models generate individual customer profiles in isolation, which is problematic when addressing aggregate behavior Cassidy et al. (2025). This is because random individual sampling ignores inter-customer correlations Haben et al. (2019); Shove and Walker (2014). Examples used in Distribution Network Assessment include the Elexon Unrestricted, representing typical profiles across customer classes, and the CREST thermal-electrical model, which has been identified to struggle to capture aggregate shapes, transitions, and peak durations Ramírez-Mendiola et al. (2017). This oversight can lead to synthetic datasets that underestimate coincident peaks or obscure critical edge cases, precisely the behaviors that risk triggering constraint violations and cascading failures in real-world networks Cassidy et al. (2025).

To address these gaps, we offer the following contributions:

- Novel application of Locally Partially Linear Double Machine Learning (LDML) to characterize causal sensitivity of a secondary substation to temperature across various temporal contexts.
- Coupling of LDML causal estimates with a Dif-

fusion Model to generate physically grounded and statistically robust domestic and substation demand scenarios, for assessing LV network reliability under stress conditions.

- Interpretability of quantified causal insights in the context of the physical network.

In the following section 2, we detail the causal inference framework, followed by the Diffusion Model architecture and the implementation of the guidance mechanism to ensure synthetic outputs follow causal insights 3. Section 4 presents a case study of a Slovenian DSO to train and validate the approach. We finalize with a conclusion and future steps in section 5. By merging three different environments (causal inference, diffusion models and power flow simulation as shown in Fig.1), we generate faithful synthetic data grounded in causal understanding of thermal drivers to provide network planners with actionable, physically-validated insights.

2. Causal Inference of Substation Temperature Sensitivity

In power networks, lower temperatures drive heating, while higher temperatures lead to increased cooling loads. However, this causal link is not stationary, with the temperature-load relationship being more significant depending on the hour of day, compared to the correlations when considering the complete temperature and load profiles Haben et al. (2019); Shove and Walker (2014).

Although the physical nature of consumer demand profiles are less deterministic than, for example, fault dynamics of power networks assets, associated domain knowledge allows us to establish the causal direction of exogenous variables. Network load is the ultimate outcome from demand, while temperature and calendar variables (hour, day of the week, month, holiday, etc.) cause load mediated by energy-based activities. However, a predictive model might assume that seasonality is the driver, e.g., because temperatures change with seasons. In causality, these variables are known as confounders. In fact, Haben et al. (2019) found seasonality to hold more predictive power than temperature. However, this does not imply temperature is not the cause of load, but that

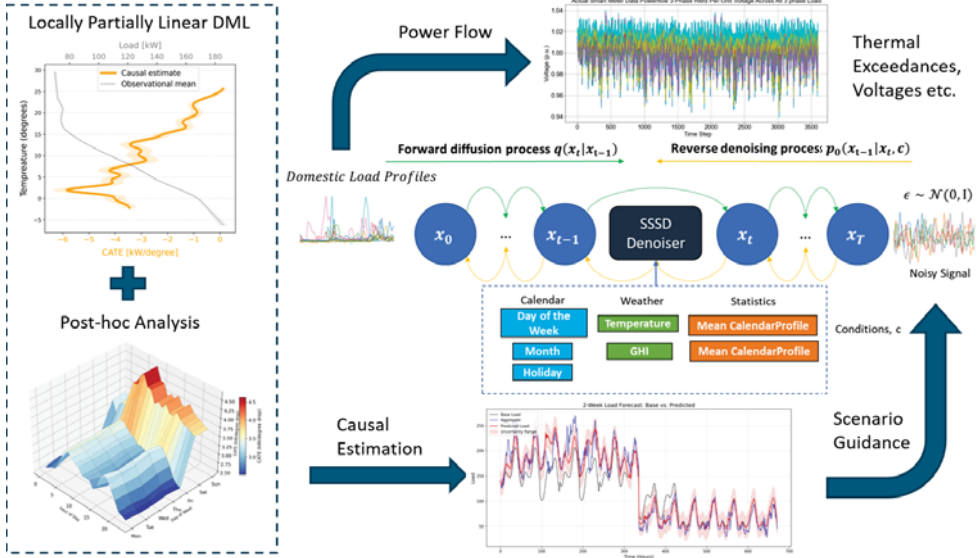


Fig. 1. Causally-guided Diffusion Model framework for Low-Voltage Network Assessment.

seasonality masks that causal effect. For example, if we have 0°C on a winter day and suddenly it shifted to 15°C, the load would drop regardless of the month. Furthermore, the impact of extreme temperatures transcends demand peaks. With both freezing temperatures and heatwaves directly affecting the physical integrity of grid equipment, increasing failure rates of transformers, switchgear, conductors and cables with high replacement costs Guddanti et al. (2025).

2.1. Double Machine Learning

To move beyond regression models which struggle to differentiate between causation and confounding features, we adopt a DML framework Chernozhukov et al. (2018); Cacciarelli et al. (2024). This allows us to disentangle the causal impact of temperature from complex, and high-dimensional data, mitigating the effect of confounding interactions. Estimating causal effects at the individual customer level is frequently unfeasible due to extreme stochastic noise. However, LV substations provide a more stable signal, capturing regional load dynamics, where the physics of load aggregation can be reliably modeled.

We assume a partially linear structure where the outcome p_i (substation load) and treatment t_i (forecast temperature) are influenced by a vector of confounders a_i (calendar effects, seasonal trends) and centered with finite variance error terms ϵ and η :

$$\begin{aligned} P_i &= \lambda t_i + f(a_i) + \epsilon_i \\ t_i &= g(a_i) + \eta_i \end{aligned} \quad (1)$$

To estimate the linear effect λ of t_i on p_i , we perform a two-stage residualization:

- (i) **Nuisance Estimation:** We approximate the functions $f(a_i)$ and $g(a_i)$ using machine learning models. To reduce overfitting and ensure consistent estimation of λ , K-fold cross-validation is used for $f(a_i)$ and $g(a_i)$ separately to prevent overfitting and regularization bias Chernozhukov et al. (2018).
- (ii) **Orthogonalization:** We compute the residuals $\tilde{p}_i = p_i - \hat{f}(a_i)$ and $\tilde{t}_i = t_i - \hat{g}(a_i)$. By regressing $\tilde{p}$ on $\tilde{t}$, we obtain λ , which represents the Average Treatment Effect (ATE). ATE reflects the unconditional and linear response of the outcome incurred by a unit increase on the treatment, obtained from the

entire population, Cacciarelli et al. (2024).

2.2. Local Extensions via Boxcar Kernels

The global ATE often fails to capture the nuances of non-linear data; to address this, the authors in Cacciarelli et al. (2024), implemented a LDML using a sliding window approach facilitated by a boxcar kernel. They formulate this as the Conditional Average Treatment Effect (CATE) by applying the previous approach to predefined subsets of data.

Aligned with the understanding of temperature-load relationship as a function of time, we alter the algorithm based on temporal features. Initially, dividing the dataset into weekday/weekend labels as we expect different routines, followed by the application of a rolling window to localize the temperature effect at different hours of the day. Rather than asses every hour individually, we asses the previous and following hour as well, as temperature effects will be similar around each hour. This hourly rolling window provides the model with additional data to stabilize the causal estimates. Within each label and hourly window, we apply a boxcar kernel to isolate data subsets near the point of interest, as illustrated in Algorithm 1.

Algorithm 1: Locally Partially Linear DML as a function of time

Input : Dataset $\mathcal{D} = \{(a_i, t_i, y_i)\}_{i=1}^n$, labels L , kernel size m , step size s , number of bootstrap replications B , number of folds K

Output: Nonlinear CATE estimates stratified by $l \in L$ and hour h

```

for each label  $l \in L$  do
     $\mathcal{D}_l \leftarrow$  Filter  $\mathcal{D}$  where label =  $l$ 
    for each hour  $h \in \{0, \dots, 23\}$  do
         $\mathcal{D}_{l,h} \leftarrow$  Filter  $\mathcal{D}_l$  for hour window  $[h - 1, h, h + 1]$ ;
        // For remaining process refer to Cacciarelli et al. (2024)
    end
end
Aggregate  $\hat{\lambda}$  across fitting point  $x_u$  for all labels and hours;

```

2.3. Causal Estimation

The target load $\hat{P}_t$ estimation is derived from the LDML framework, captured in Eq.(1), and recon-

figured in a counterfactual fashion. Note it is not the purpose of this work to evaluate counterfactual scenarios, but rather, past interventions since real data is available for validation. Given the current temperature T_t and the context-specific interpolators $\mathcal{F}$ for a label L , and hour h the load is estimated as:

$$\hat{P}_t = \mathcal{F}_{load,L,h}(T_t) + \Delta P_{causal} \quad (2)$$

where the base load is estimated by the load nuisance estimator, and the causal impact ΔP_{causal} represents the load shift induced by the temperature deviation from the baseline temperature, as modulated by the CATE:

$$\Delta P_{causal} = \mathcal{F}_{lambda,L,h}(T_t) \cdot (T_t - \mathcal{F}_{temp,L,h}(T_t)) \quad (3)$$

3. Domestic Load Profile Synthesis

The efficacy of domestic load modeling is a challenging task given the fundamental high volatility of residential load data Haben et al. (2019), which may lack the contextual granularity required for stable modeling due to privacy constraints. To tackle this, the substation causal insights are used downstream, as illustrated in Fig.1. In this scenario the LDML acts as a 'teacher' with knowledge of the aggregate behavior, to guide the Diffusion Model, operating as the 'student', towards generating faithful domestic profiles which adhere to the substation causal estimate.

3.1. Conditional Denoising Diffusion Model

We utilize a Denoising Diffusion Probabilistic Model (DDPM), which generates realistic load data by reversing a progressive Gaussian noise-injection process, governed by two processes.

The forward process $q(x_{1:T}|x_0)$ transforms real profiles x_0 into noise over multiple diffusion steps $t \in [1, T]$. This process is controlled by the changing variance $\beta_t \in (0, 1)$ of the noise added at each timestep and given by Eq.4, where $\alpha_t = 1 - \beta_t$ and $\tilde{\alpha}_t = \prod_{i=1}^t \alpha_i$

$$q(x_t|x_0) := \mathcal{N}(x_t; \sqrt{\tilde{\alpha}_t}x_0, (1 - \tilde{\alpha}_t)\mathbf{I}) \quad (4)$$

The reverse process is conditional on external drivers c to remove noise using a neural network.

The transition is defined as:

$$p_\theta(x_{t-1}|x_t, c) := \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t|c), \tilde{\beta}_t \mathbf{I}) \quad (5)$$

where μ_θ represents a neural network and $\tilde{\beta}_t = \frac{1-\tilde{\alpha}_t-1}{1-\tilde{\alpha}_t} \beta_t$.

To effectively estimate the additive noise in each diffusion step, we implemented the SSSDS4 network Alcaraz and Strodthoff (2023) $\epsilon_\theta(x_t, t|c)$, where x_t is the noisy input at timestep t , and c represents conditioning weather, calendar and average profile features, illustrated in Fig.1.

3.2. Guided-Diff

To ensure the generated samples adhere to the identified grid physics, we adopt the constraining approach proposed in Coletta et al. (2023), whereby the noise estimate ϵ_θ is adjusted by the gradient of an external constraint function f_c at inference, without the need to retrain the model. In our framework, f_c is formulated as a loss function targeting both the mean and maximum values of the generated load signal, as in Eq.(6).

$$f_c(x_0) = \|\text{mean}(x_0) - \hat{P}_{target}\|_2^2 + \|\max(x_0) - P_{peak}\|_2^2 \quad (6)$$

4. Case Study

4.1. Elektro Dataset and Model Implementations

A comprehensive observed dataset from a Slovenian DSO Elektro Ljubljana (2023) is used as a case study to demonstrate the production of CATE and guided diffused load profiles to inform distribution network planning. A model is developed to perform time-series power-flow simulations driven by the measured load profiles. The dataset includes smart-meter measurements of bidirectional active and reactive power, and transformer monitoring. The studied Slovenian radial distribution network consists of eight feeders supplied by a 20 kV substation, serving 158 customers (29 single-phase, 129 three-phase), including four with PV, and comprising 8,614 m of underground cable.

We train the LDML model using a 2-year period (May 2021-May 2023), with LightGBM as

the nuisance model learner, utilizing 5-fold cross-fitting. Furthermore, we controlled for the following confounders: solar irradiance, hour, month, holiday and day of the week. The sliding window approach was configured with a fixed window size (m) of 1000 observations and a step size (s) of 100 observations. To mitigate local variance of CATE curves, the resulting mean effect estimates were refined using a Gaussian smoothing filter as in Cacciarelli et al. (2024).

For the generative architecture and diffusion settings, we used the configuration used in Alcaraz and Strodthoff (2023), but with residual network and skip channels of dimension of 128. The generative model was trained on 10,000 randomized domestic load profiles, spanning the same period utilized for the substation's causal framework.

The substation causal effect estimate was benchmarked using a 6-month validation set (May-November 2023) to validate the overall performance of estimated causal priors that could be used in the downstream task, while two discrete weeks from February and July 2023 were reserved to test the overall framework's operational utility: a winter week featuring -12°C troughs and a summer week reaching 30°C peaks. By extracting the causal estimates for the two-week temperature stress case, we performed the causal-guided synthesis and the power flow validation. These scenarios represent periods of high network stress and therefore provide conservative evidence for the reliability of planning decisions.

4.2. Evaluation Strategy

To evaluate the estimated causal sensitivities between temperature and substation load, we employed stress tests to identify potential model configuration bias and evaluate robustness of the yielded CATE estimands for each hour and day type:

- (i) Sample split ratios: If the parameter estimates obtained for different k-folds are not stable, it indicates the dataset is too small or noisy, and the model has not learned the nuisance parameters properly.
- (ii) Replacement of ML algorithms: To test po-

tential model bias, we compare our LightGBM against a Random Forest.

To measure these discrepancies, we employ the coefficient of variation (CV) by measuring the mean and standard deviation of the CATE outputs, grouped by label (day and hour) and temperature bins of 5°C. A threshold of 0.2 is used to validate the robustness. To quantify the reliability of causal estimates within each regime, we computed the 90% confidence interval of the CATE via a 50-iteration bootstrap resampling process within sliding temperature windows.

We assessed the accuracy of the causal effect estimation approach by measuring the weighted Mean Average Percentage Error (MAPE) and the Root Mean Square Error (RMSE) on both the six-month validation period and the two-week temperature test period. Given the inherent uncertainties of the estimated features and the nature of the CATE as an average value, we utilized the 90th percentile of validation forecast errors to define the predictive uncertainty bounds for these extreme cases and measured the pinball loss. Finally, we compared the power flow voltage profiles of guided versus non-guided extreme temperature outcomes from the diffusion model.

4.3. Experimental Results

4.3.1. Causal Inference Results

The CATE estimates visualized in Fig.2 demonstrate a clear inverse relationship between ambient temperature and power demand, as evidenced by the consistently negative coefficients. These sensitivities directly reflect the Slovenian thermal load profile, where space heating (60%) dominates household demand in 2023, with electricity (35%) and heat pump (7%) loads surpassing wood fuels (33%) Statistical Office of the Republic of Slovenia (SURS) (2024). In addition, higher temperature trends seem to indicate an absence of temperature sensitivities, which is consistent with the current Slovenian context. Traditionally, cooling has been managed through ventilation, while space cooling, despite rising, still represents a small share of total energy usage ODYSSEEMURE (2024).

A distinct fan-shaped divergence exists across the diurnal cycle in which morning, midday, and early evening periods exhibit increased temperature effects and uncertainties, reflecting diverse residential heating demand and human activity cycles. Weekend profiles follow a similar diurnal trend with stronger temperature effects; however, due to being a reduced subset of data, these profiles cover a narrower temperature range compared to the more data-rich weekday subsets. Moreover, this analysis could provide grid operators with a clearer view of thermal electrification progress across substations. This understanding becomes imperative as lower sensitivity indicates a higher proportion of wood fuel, which in the coming years will be subject to grid risks as these substations electrify.

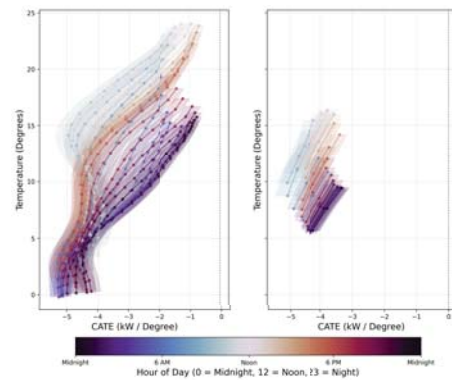


Fig. 2. CATE of temperature treatment of load, for weekdays (left) and weekends (right), across the day

We evaluated the reliability of each hourly CATE estimate using the CV-score heatmap shown in Fig.3. While the model maintains a high degree of stability across most operating states, instability emerges as temperatures rise around night time periods. These outliers coincide with regions of lower data density, suggesting that the model is overfitting to noise in these specific windows.

The 6-month validation period scored a MAPE of 15.81%, which is performant given the wider variety of load regimes. Notably, the 2-week inference dataset achieved a lower overall average

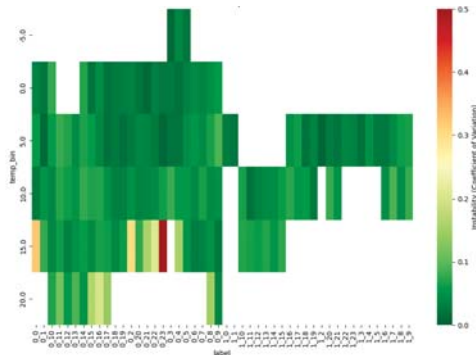


Fig. 3. CV heatmap of CATE estimates summarizing robustness tests, illustrating stability in green and instability in red. The labels are represented first by weekday (0) and weekend (1) signatures, followed by the hour (0-23)

error of 9.87%. However, this accuracy is relative to the average load, which is higher in the extreme scenario (127.94 kW against 78.8 kW). Furthermore, the RMSE revealed that the extreme case had slightly larger errors (14.74 kW against 13.79 kW). This is captured in Fig.4, where the predicted load - resulting from the addition of the disentangled temperature effect on load and the confounded base load - showed some limitations in tracking demand peaks during winter extremes when compared to the real aggregate signal. Although, this was compensated through the uncertainty range, where the pinball loss for both quantiles remains low, with the upper quantile (2.25) exhibiting slightly lesser performance than its counterpart (2.00). Warm temperature effects, shown in the second segment of Fig.4, were captured considerably well, indicating that this substation may not have a large AC load.

4.3.2. Power Flow Simulation

Fig.5 illustrates how integrating a simple generalized causal expectation improves the reliability of substation assessment compared to the baseline. In the summer scenario, the baseline model frequently yields under-voltage conditions, suggesting that the underlying signal fails to capture the necessary reactive power support or overestimates concurrent peak demand. Conversely, during win-

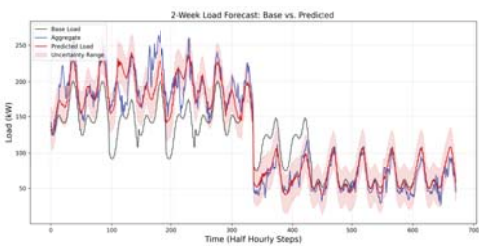


Fig. 4. Causal Forecast of winter week followed by the summer period.

ter, the baseline model tends to produce over-voltages, likely due to an underestimation of the thermal-driven load peaks. Although our causal approach is significantly more effective and maintains the voltage profile closer to real limits, it still encounters challenges in fully mitigating under-voltages during extreme winter situations.

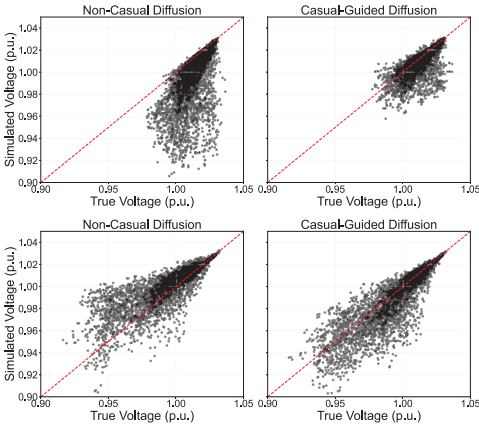


Fig. 5. Scatter Comparison of Simulated vs. True LV Network Voltages for Non-Casual and Casual-Guided Diffusion Models under Extreme Summer (top) and Winter (bottom) Scenarios.

Table.1 quantifies these performance gains, demonstrating that the Causal-Guided model achieves a reduction in error of approximately 70% for both RMSE and MAPE during the summer period. While improvements are more modest in the winter scenario, the causal framework still outperforms the baseline, strengthening substation load assessment across seasonal extremes.

Table 1. Model Performance Metrics

Error (p.u.)	Non-Casual	Casual-Guided
Summer RMSE	0.028	0.009
Summer MAPE	1.89%	0.56%
Winter RMSE	0.016	0.013
Winter MAPE	1.11%	0.96%

5. Conclusion and Future Work

In this work, we have demonstrated that a causal inference framework provides a quantifiable representation of the thermal drivers behind substitution load. By isolating temperature effects, we provide an interpretable representation of demand curves across temporal contexts to provide network operators with insights into space heating and cooling balance and risks of individual intra-day, weekly and seasonal periods. While the current model represents average expectations, it has proven highly effective for characterizing normal operations and current heatwave thermal events. Coupling of these statistical insights with a power flow simulator, improves grid engineering assessment, translating abstract causal estimates into actionable physical limits.

Future developments could explore Quantile Treatment Effects (QTE) as an alternative for a more granular assessment of edge effects, potentially providing a more accurate capture of the extreme winter peaks currently under-estimated by the localized mean effects.

Acknowledgement

This work is supported by the Engineering Physical Sciences Research Council (EPSRC) and the Technology and Innovation Centre's Scottish Low Carbon Power and Energy Partnership (TIC LCPE) in collaboration with the Data and Analytics team within ScottishPower UKIT Architecture group, under grant agreement EP/W524670/1.

References

- Alcaraz, J. M. L. and N. Strodthoff (2023). Diffusion-based time series imputation and forecasting with structured state space models.
- Cacciarelli, D., P. Pinson, F. Panagiotopoulos, D. Dixon, and L. Blaxland (2024). Do we actually understand the impact of renewables on electricity prices? a causal inference approach. Working Paper, Imperial College London and National Energy System Operator (NESO).
- Cassidy, M., B. Brown, S. D. J. McArthur, S. D. Vidhyasankar, D. Kirk, and B. Stephen (2025). Towards trustworthy residential load profile synthesis using a diversity driven diffusion model. In *2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, pp. 1–5.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21, C1–C68.
- Coletta, A., S. Gopalakrishnan, D. Borrajo, and S. Vyetenko (2023). On the constrained time-series generation problem. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Curran Associates Inc.
- Elektro Ljubljana (2023). Opsci - open sci: Platform for sharing low voltage distribution network technical data. Online Project Documentation, <https://www.elektro-ljubljana.si/projekti/ArtMID/1374/ArticleID/2156/OPSCI-Open-SCI>.
- Guddanti, K. P., A. K. Bharati, S. Nekkallapu, J. Mcwherter, and S. L. Morris (2025). A comprehensive review: Impacts of extreme temperatures due to climate change on power grid infrastructure and operation. *IEEE Access* 13, 49375–49415.
- Haben, S., G. Giasemidis, F. Ziel, and S. Arora (2019). Short term load forecasting and the effect of temperature at the low voltage level. *International Journal of Forecasting* 35, 1469–1484.
- Nauta, M., D. Bucur, and C. Seifert (2019). Causal discovery with attention-based convolutional neural networks. *Machine Learning and Knowledge Extraction* 1, 312–340.
- ODYSSEE-MURE (2024). Slovenia energy efficiency trends and policies. Technical Report, Enerdata and the European Commission.
- Ramírez-Mendiola, J. L., P. Grünwald, and N. Eyre (2017). The diversity of residential electricity demand – a comparative analysis of metered and simulated data. *Energy and Buildings* 151, 121–131.
- Rawal, A., A. Raglin, D. B. Rawat, B. M. Sadler, and J. McCoy (2025). Causality for trustworthy artificial intelligence: Status, challenges and perspectives. *ACM Computing Surveys* 57.
- Shove, E. and G. Walker (2014). What is energy for? social practice and energy demand. *Theory, Culture & Society* 31, 41 – 58.
- Statistical Office of the Republic of Slovenia (SURs) (2024). Household energy efficiency, 2023. Statistical Report, <https://www.stat.si/StatWeb/en/news/Index/12783>.