Analysis of Load Profile Generation Methods and Their Effect on the Results of Energy Optimization Models

Mohamad Wehbe
mohamad.wehbe@student.innoenergy.com
Instituto Superior Técnico, Universidade de Lisboa, Portugal
December 2023

***Abstract*—This paper presents an analysis of the effects of different load profile generation methods on the results of energy optimization models. Load profile generation methods are used to synthesize demand load data, which represents the energy consumption patterns of end-users over time. The paper breaks down the analysis into two parts: First an analysis of load profile generators is done through various methods that aim to characterize and benchmark their performance. The second part analyses energy model optimization under different load profile generation methods as inputs. The paper uses an open-source energy system modelling framework (OEMOF) to optimize a multi-energy system under various scenarios. The paper evaluates the performance and robustness of each method, uncovers their underlying assumptions, and biases, and explains how they could affect the results of energy model optimization. The paper ends with some best practices for working with Load Profile Generators.**

Keywords- Energy Optimization, Energy Modelling, Demand Profiles, Energy System Modelling Frameworks, Load Profile Generators.

I. INTRODUCTION

Assumptions play a pivotal role in energy system modelling and optimization. The accuracy and reliability of a model's output are intrinsically tied to the quality and validity of its input assumptions [1]. Among these inputs, demand load data stands out as particularly critical [2]. Demand load data represents the energy consumption patterns of end-users over time. For instance, underestimating peak demand can result in inadequate infrastructure planning, while overestimating can lead to oversizing and unnecessary costs and inefficiencies. Furthermore, the source and methodology used to derive this demand load data can introduce biases and uncertainties.

There are several methods to obtain this data. Direct measurements using smart meters or sensors provide real-time insights but might be limited in scope or raise privacy concerns. Deploying smart meters is also costly. Historical data, while valuable, might not always reflect future scenarios, especially in rapidly changing environments or in the face of significant policy or technological shifts. Another issue is the lack of availability of data and tools necessary to analyse the collected historical data[3]. A final method, surveys and statistical extrapolations, while useful, have notable limitations. Firstly, conducting and analysing surveys is time-intensive, often not aligning with the swift decision-making needs of energy planning. Secondly, biases can be introduced through survey design, question phrasing, and the administration process. Such biases can skew load data, impacting the accuracy of energy system models.

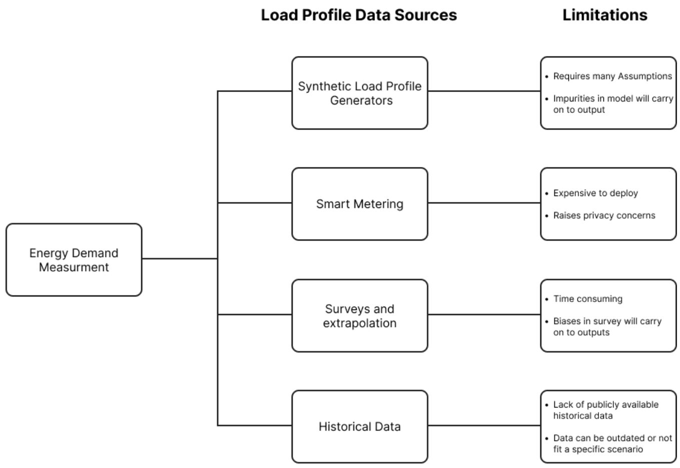


Figure I.1 Possible source of demand profiles and their respective limitations

Given these complexities in obtaining and validating load data, there's a growing interest in synthetic methods of data generation, leading us to the realm of Load Profile Generators (LPGs).

LPGs offer a more dynamic approach to generating demand load data. Instead of relying solely on historical records or direct measurements, LPGs synthesize load profiles based on a combination of factors, including user behaviour, appliance efficiency, weather patterns, and more. This synthetic generation allows for the creation of diverse and adaptable load profiles that can cater to a wide range of scenarios, from the introduction of new energy-efficient technologies to shifts in consumer behaviour due to policy changes.

a significant gap persists in understanding how the assumptions behind these LPGs, and the choice of one LPG over another, influence the outcomes of energy models and their optimization results. One paper however, does explore the effects of data sampling frequency on the resultant load profile features [4]. Concentrating on a specific type of energy model, such as residential systems, allows for a more precise and in-depth analysis. Yet, the overarching challenge remains: Without a comprehensive grasp of these effects, there's a risk that energy system models might yield results that, while mathematically sound, may not align with real-world scenarios, potentially leading to suboptimal decisions in system design and operation.

Given the identified problem, this research aims to:

1. How do the different LPG tools compare in terms of their required input data, methodologies, outputs, and applicability?

2. How do input assumptions impact the synthetic demand profiles produced by LPGs, and how do these profiles compare to historical data?

3. What is the extent of the influence of LPG input assumptions on the results of energy system models, particularly in their optimization outcomes?

4. What are the best practices for utilizing LPGs in energy system modelling to ensure that the generated profiles closely mirror real-world scenarios?

II. State of the Art

Four load profile generators are used in this study to generate load profiles. The energy modelling and optimization will subsequently be carried out using the Open Energy Modelling Framework (OEMOF)

A. Load Profile Generators (LPGs)

Demandlib [5] is a Python package that can create load profiles for electricity and heat demand based on standardized methods from BDEW (German Association of Energy and Water Industries) [6]. It can be used to model different sectors, such as households, businesses, industries, and agriculture, and different energy types, such as electricity, heat, and water. It can also generate customized profiles by adjusting parameters such as annual demand, building class, temperature, and holidays. Demandlib is based on python and can be installed through GitHub.

Districtgenerator is another open-source python-based tool on GitHub that is under current development [7]. Its primary purpose is to generate building-specific thermal, electrical, and occupancy profiles for residential districts. It integrates two open-source tools, TEASER [8], and Richardson.py [9]. It classifies buildings using the TABULA archetype approach which is a methodology for classifying residential buildings across Europe based on their energy-related characteristics [10].

The LoadProfileGenerator [11] is a specialized tool designed for the creation of load profiles detailing energy and water consumption in households. The tool's foundation is built upon a behaviour model which posits that individuals are driven by intrinsic desires, influencing their daily activities and, consequently, their consumption patterns [12].

synPRO [13] is a tool designed to generate electric load profiles of a single household, a commercial zone, or a community with multiple commercial and residential zones [14]. The loads have a minimum time resolution of 1 minute, specifically tailored for German households. The tool uses a bottom-up stochastic model to model electric, heating, domestic hot water, and electric car load profiles. To simulate a load profile users can use the tool online. Unlike the other tools that were covered, the tool’s source code is not published online.

B. The OEMOF Model

Open Energy Modelling Framework (OEMOF)[15] is an optimization model based on linear programming (LP) and mixed integer linear programming (MILP). It can be used to build an energy system model. The tool is open-source [16] and based on the python coding language. When building OEMOF, the goal was to create a flexible open-source tool that is capable of mathematically mapping the multidimensional interdependencies within the energy industry [17]. One thing that stands out in OEMOF is the great variety of different models that can be built and research questions that can be answered.

An important additional package that was used in this study to enhance the modelling capabilities of this model is ESyOpT. It is a package that introduces economic and technological context around basic OEMOF components [18]. It introduces more input parameters to each component and contributes more elements such as investment costs of a storage system or maximum power of a solar panel to the objective function and constraints.

III. Methodology

A. Research Philosophy

The problem being solved is the uncertainty that energy modelling experts face concerning the effects of their assumptions on the overall results of their study. To shed some light on this issue this study aims to understand how the underlying assumptions of energy system modelling from the load profile perspective affects the result of model optimization. As stated in the objective, it is important to understand any biases introduced by the LPG tools that show up in the load profiles and how these biases propagate to the optimization results. Once understood, some clarifications and recommendations can be made to help researchers better understand the impact of their load profile modelling assumptions on the overall optimization results.

Two testing methods are available. A black box method, and a white box method. In the context of this study a white box method would involve analysing the internal workings, algorithms, and processes of the LPG tools and the modelling framework. This would entail a deep dive into the mathematical formulations, algorithms, and data structures used by these tools. On the other hand, the black box method involves treating the LPG tools and optimization models as a closed system, where the focus is primarily on the inputs and outputs without considering the internal processes. By providing a set of standardized inputs and analysing the resulting load profiles and optimization results, we can infer the biases and assumptions made by the tool. For the purpose of this study the black box approach will be pursued since this method is less intensive and instead, it relies on empirical testing and observational analysis. By concentrating on observable outcomes, the black box method emphasizes empirical results. This can lead to more tangible and actionable insights, especially when the goal is to understand the impact of assumptions on optimization results.

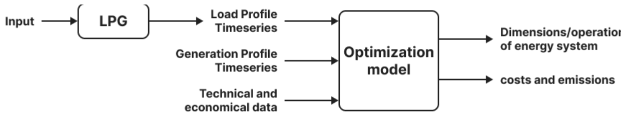


Figure III.1 Process flow of energy system optimization model with focus on load profile timeseries simulation using LPG

For this study, which aims to discern the effects of energy system modelling assumptions on optimization results, an inductive reasoning approach is most fitting. This approach, rooted in specific observations, aligns with the study's exploratory nature and its empirical emphasis, especially given the use of the black box method. Inductive reasoning allows the research to detect patterns and draw broader generalizations based on these observations, offering flexibility and the potential to develop new theories grounded in the data. In contrast to deductive reasoning, which tests a predefined hypothesis against observations, inductive reasoning starts from specific cases and moves towards broader insights, making it more suitable for this study's objectives. After we clarified the philosophy, we will move into describing the research design and justifying the decisions that were taken.

B. Research Design

The initial section centres on the second objective. This involves an analysis of simulated load profiles. Each tool is employed to craft a simulated load profile that closely matches historical data from various sources. Once established, these profiles are compared with the historical datasets to compare accuracy. However, validating the accuracy of a household load profile is complex. A direct comparison of generated load profiles with real-time household data for each timestep might not provide significant insights. Even with similar modelled activities, slight deviations in the timing of appliance use can lead to discrepancies between the profile and the real data. Merely contrasting the annual or monthly totals also doesn't provide a comprehensive understanding of the simulated load profile's precision [12]. Thus, various visual and statistical techniques are employed to assess load profile accuracy, with some techniques being adaptations from diverse studies. The analysis is segmented into two subsections each with its own methods:

- Analysis of load profiles
  - Visual/graphic analysis
    - Load profile characterisation method
    - Relative heat maps method
  - Statistical analysis
    - Annual energy demand
    - Monthly energy demand

In the initial subsection load profiles undergo two different methods of visual comparison. The first method of visual analysis, the load profile characterisation method, focuses on clustering and comparing load profiles according to some studied parameters. This visual analysis approach predominantly stems from a study on electric load profile scrutiny from the Lawrence Berkley National Laboratory [19] and another exploring patterns and variability of electric load profiles from the same institution [20].

The second visual analysis method, the relative heatmaps method, is used as a way to visualize the load profile difference between the actual and simulated dataset. The heatmap is created by subtracting the timesteps of the actual and simulated load profiles. As mentioned earlier, a direct statistical comparison between each actual and simulated timestep will not lead to useful results, nevertheless, visualising the difference can show in which general time of the year or day the simulated load profile overestimated or underestimated energy demands.

In the second subsection to answer objective 2 a statistical analysis was carried out. The statistical analysis also focused on indicators that were used in study papers with the aim of validating synthetic load profiles. [12], [14], [21]

Section 4.2 is dedicated to solving objective 3 which aims to understand the effects that the simulated load profiles had on the optimization results. In accordance with the black box and inductive approach, a common energy model was built and fed both the real and simulated load profiles the real and simulated load profiles. More details about the energy model can be found in section 3.2.3. The model created contains two different optimization problems. A design optimization problem and an operation optimization problem. By building a model that solves these problems, a comprehensive approach to understanding the influence of LPG input assumptions on various facets of energy system modelling was achieved.

The energy model built using OEMOF was based on a hypothetical scenario of a residential household in Europe. The model that was built had to have both the design and operation optimization problems that OEMOF can solve. With this in mind, the following model was created:

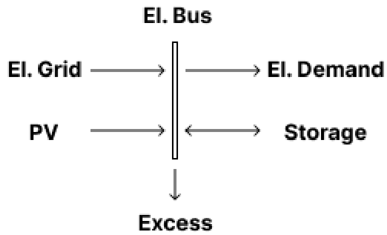


Figure III.2 OEMOF Energy Model used in this analysis

The energy system that was modeled is a grid connected residential household. The system can potentially have PV modules and energy storage installed. Only the household's electricity demand was modeled. The heat and water demands were neglected in this model. To create this energy model six components shown in figure 3.2 were included in the system. The PV and Storage components are ESyOpT components. The optimization model created has the following simplified objective function:

$$\min: \sum_{pv} x_{pv}^{\text{capacity}} \cdot c_{pv}^{\text{capacity cost}} + \sum_{s} x_{s}^{\text{capacity}} \cdot c_{s}^{\text{capacity cost}} + \sum_{s,t} x_{s}^{\text{flow}}(t) \cdot c_{s}^{\text{marginal cost}} + \sum_{g,t} x_{g}^{\text{imports}}(t) \cdot c_{g}^{\text{marginal cost}} \tag{1}$$

This objective function is composed of four segments. The first segment represents the investment cost of the PV system (design optimization problem). The second segment represents the investment cost of the storage system (design optimization problem). The third segment represents the operational cost of the storage system (operation optimization problem). Finally, the fourth segment represents the cost of the grid imports (operation optimization problem). The goal of the optimization was to minimize these costs that are associated with the energy system.

IV. RESULTS

The section is structured into two main subsections: visual analysis, which includes the load profile characterisation and relative heatmaps methods, and statistical analysis, focusing on annual and monthly energy demands. This framework ensures a detailed assessment of the accuracy of the simulated load profiles.

A. *Visual Analysis*

The UKdale dataset, which served as the benchmark in this analysis, presented a consistent base load during nighttime, particularly noticeable between 0.2 kW to 0.3 kW. This consistency contrasted with the simulations in various ways. For instance, the Demandlib simulation delivered a smoother base load throughout the year, but it was lower in magnitude, hovering between 0.14 kW to 0.17 kW. The Districtgenerator's base load profile was markedly different, characterized by its short duration and instability. LoadProfileGenerator maintained a steady base load of around 0.11 kW. SynPRO's profile emerged as distinct among the simulations, particularly during the winter when it reached a base load as high as 0.3 kW, even surpassing UKdale's own winter base load which sustained base load of over 0.25 rarely and for a short period of time.

In terms of peak consumption patterns, UKdale generally witnessed its morning peaks around 8, with slight variations across seasons. Evening peaks were predominantly observed around 18 or 19. In contrast, Demandlib, while maintaining consistency, did not mirror UKdale's pronounced morning peak, instead showing an early semi-peak leading up to a significant midday rise. The Districtgenerator's approach was even more distinct, with its load profile punctuated by small midnight peaks and inconsistent early morning peaks across seasons. LoadProfileGenerator's sharp midday peak, especially evident during the summer months, starkly contrasted with UKdale's moderated consumption pattern. SynPRO offered a diverse pattern, with multiple peaks throughout the day, and a particularly pronounced evening peak in winter that reflected some characteristics of the UKdale winter behavior.

The seasonal variations in the load profiles further highlighted the differences between UKdale and the simulations. UKdale's winter consumption was characteristically higher, especially during its peak hours. Demandlib's profile, on the other hand, exhibited a trend of reduced power during the winter, while maintaining more uniformity during the summer. This differed from UKdale's clear winter elevation in consumption. Districtgenerator's unique approach resulted in two distinct winter peaks during the afternoon and evening. LoadProfileGenerator's summer profile stood out due to its pronounced midday peak, deviating considerably from UKdale's moderated summer consumption. SynPRO's winter profile, although exhibiting higher overall consumption, also stood out due to its elevated base load.

Among the simulations themselves, distinctions were apparent. Demandlib's consistent load profile across seasons set it apart from the more varied profiles of the other tools. Districtgenerator, with its erratic load variations, starkly contrasted with the other simulations. LoadProfileGenerator, while maintaining shape consistency across seasons, was characterized by intense midday peaks, especially in summer, making it divergent from the others. SynPRO offered a mix, with its winter profile indicating higher consumption, providing a unique perspective among the simulations but being the only one to match with UKdale's seasonalities.

While UKdale offered a grounded and consistent benchmark, each simulation reflected varied patterns and behaviors. Some simulations, like Demandlib, exhibited overall consistency, while others, such as Districtgenerator, presented more fluctuating patterns.

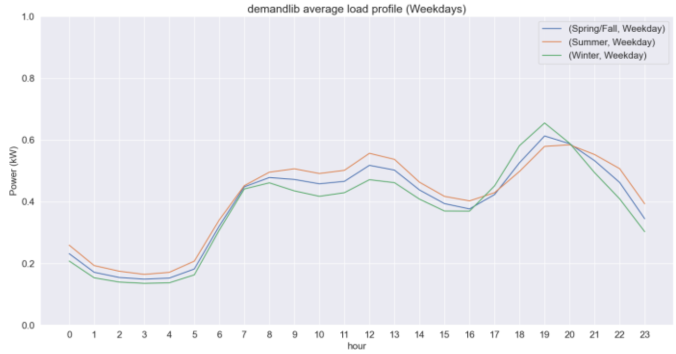


Figure IV.1 An example of the produced clustered load profiles.

The relative heatmaps method visually represented differences between the actual and simulated load profiles. By subtracting corresponding timesteps, the heatmap was generated. Although a direct statistical comparison of these timesteps wasn't advantageous, the visualization highlighted periods when the simulated profile either over- or underestimated demands.

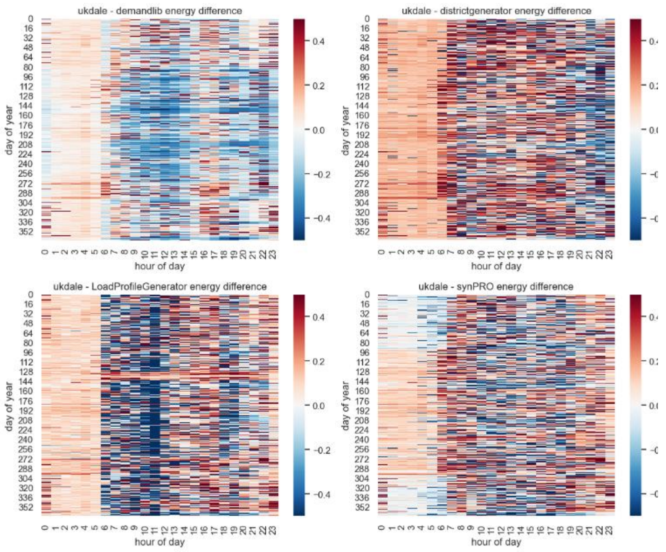


Figure IV.2 Relative heatmaps comparing UKdale and the simulations

Comparing the baseload in the early hours between around 0 to 6 shows that all the simulations underestimated the baseload except for synPRO in the winter. This is consistent with the findings from the load profile characterisation.

The early peaks between 6 and 12 are slightly overestimated by demandlib in the summer and heavily overestimated by LoadProfileGenerator all year round. Again, this was evident from the sharp peaks in the load profiles. A red patch however can be seen after day 130 during midday. This corresponds to a drop in demand in the 5th month which will be seen in the analysis in the next section. The very messy midday map of synPRO and Districtgenerator indicates irregularities in the peaks during these hours throughout the year.

B. *Statistical Analysis*

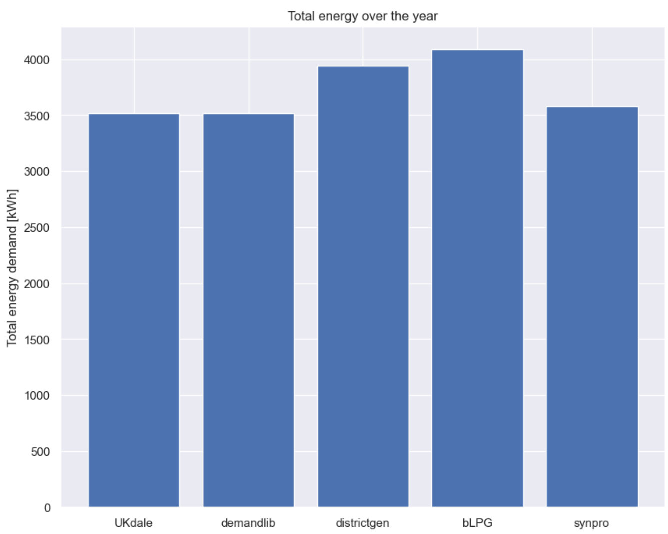


Figure IV.3 Total energy expenditure of real and simulatd load profiles over a year

Figure 4.3 depicts the total annual energy expenditure of the real and simulated load profiles over a year. The UKdale dataset serves as the benchmark, representing the actual energy demand, with a value of 3517.8 kWh. The remaining bars represent the output of various tools employed to simulate load profiles that aim to resemble the UKdale dataset.

Both the "demandlib" tool and the UKdale dataset exhibit identical total energy demand values at 3517.8 kWh. This is due to the input requirements of demandlib. When simulated the total annual energy expenditure was inputted into the simulation.

The synPRO tool's simulation results in an energy demand of 3581.5 kWh, which is slightly higher than the UKdale. The precent error is 8%.

On the other hand, the districtgenerator and LoadProfileGenerator tools yield significantly higher energy demand values. Specifically, districtgenerator results in a value of 3938.6 kWh, an error of 12%, and LoadProfileGenerator surpasses all with a value of 4087.2 kWh and an error of 16%.

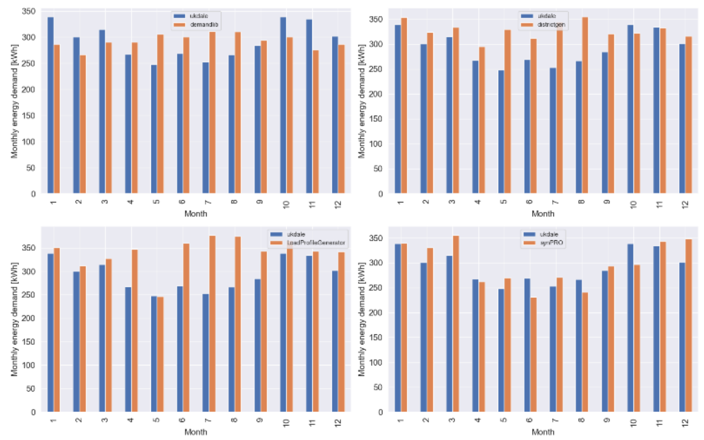


Figure IV.4 Monthly energy consumption of the UKdale dataset and the simulations

The series of graphs present the monthly electricity demand (in kWh) for a year, comparing the actual dataset, UKdale, with simulated results from different load profile generation tools. The UKdale monthly electricity demand displays a seasonal trend with lower demand in the summer months.

The energy demand from the demandlib simulation exhibited a weak seasonal trend. Contrary to UKdale, demandlib displayed higher electricity demand in the summer months. Hence, the summer demand is overestimated, and the winter demands are underestimated. The calculated MAPE of the demandlib simulation is 12.7%

Districtgenerator's simulation does not display a seasonal trend. The simulation provided similar values during the winter months but overestimated them in the summer. The calculated MAPE of the Districtgenerator simulation is 13.0%

LoadProfileGenerator also shows weak seasonal trend that increases in the summer except for a big dip in May. The simulation closely follows the UKdale dataset in the winter but shows a large margin of error in the summer months. The calculated MAPE is 16.3%

The synPRO simulation exhibits the lowest margin of error among all simulations. The calculated MAPE is 8.2%. This tool is also the only tool that matched the seasonal trend and showed a dip electricity usage during the summer months.

The same analysis was also done using a different dataset and simulations but with the same LPGs. This time the the analysis was benchmarked against the 21 selected houses in the WPuQ dataset. And because this dataset is an average of 21 houses, the seasonalities were slightly averaged out. Nevertheless winter in WPuQ datasets also showed higher energy usage when compared to summer.

Similar outcomes were observed in this anaylsis. The similarities are as follows. Demandlib underestimated enery usage in the winter and overestimated in the summer. It also exhibited weak seasonalities albeit slightly more energy usage in the summer. bLPG again underestimated energy usage in month 5. This is due to simulating vacation and time away from the house. It also overestimated in the summer months. In general the seasonalities among the tools did not change from one simulation to the other, but the intensities did change.

Some contrasts include districtgenerator underestimating instead of overestimating in energy usage. One notable difference is the very low forecasted monthly usage by synPRO. The main parameter that lead to this drop in occupants from four in the UKdale simulation to two in the WPuQ simulation.

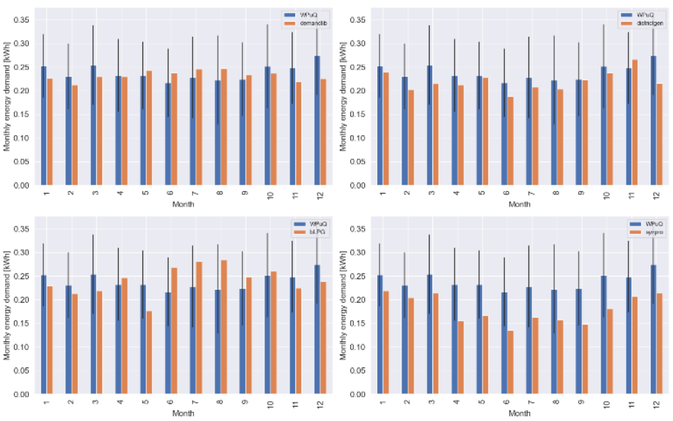


Figure IV.5 Monthly Energy Consumption of the WPuQ datasets and the simulations

The frequency analysis undertaken delved into the occurrence rates of specific power levels over the year's hours. While average values can sometimes be deceptive, examining the frequency of power levels paints a clearer picture of the alignment between synthetic and historical datasets.

The load duration curve representing the UKdale shows that power throughout the year rarely crosses the 1kW mark and that only 500 total hours (6%) of electricity consumption happens while the power is above 2kW. On the contrary, progressive load profile generators like LoadProfileGenerator and district generator produce values above 1kW 14% and 11% of the time respectively.

On the other hand, demandlib is too conservative and never produces any values above 1kW. Its max power value is 0.74kW. In contrast, the max value of UKdale dataset is 2.9 kW. Despite these low values, demandlib's power values over the hours does decay slower than the other simulations.

Looking at the load duration curve of the WPuQ dataset simulation a similar conclusion can be done. In this case however, since the WPuQ curve is an average of many households, its more conservative. Nevertheless, standard deviation lines show that it is unlikely for households to be producing some of the high-power values that the load profile generators are producing.

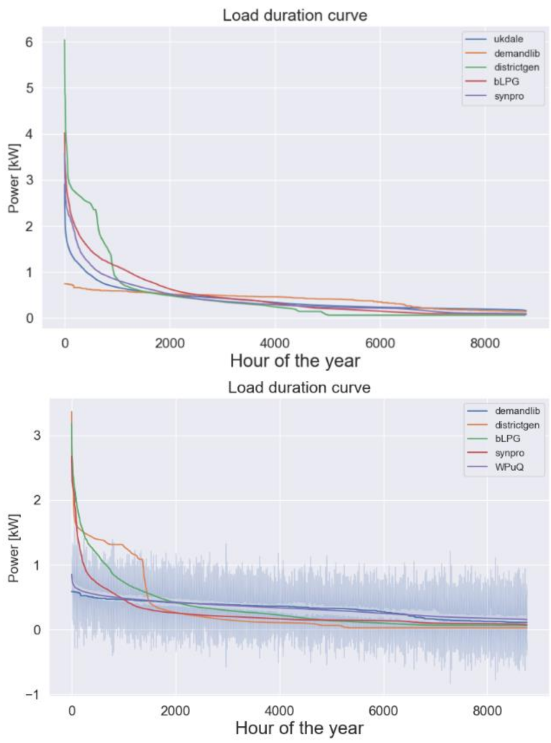


Figure IV.6 Load duration curbe of UKdale simulation on top and WPuQ simulation on the bottom.

C. *Analysis of Optimization Results*

In the design optimization results, a tenuous relationship was observed between demand and storage investment. Notably, even with comparable energy demands, synPRO and demandlib displayed distinct optimized energy storage capacities. Demandlib necessitated just 60% of the storage investment of synPRO and a marginally smaller PV system. This implied that the demandlib model sourced less overall energy from PV production and battery discharge. However, this was offset by a reduced energy excess or curtailment.

The correlation between grid imports and demand was also subtle, as highlighted by the results from districtgenerator and LoadProfileGenerator. While LoadProfileGenerator's demand was approximately 77% of districtgenerator's, its grid imports slightly exceeded those of districtgenerator. Moreover, LoadProfileGenerator called for more significant PV investments, but this was accompanied by a considerably higher curtailment rate.

In comparison to UKdale, synPRO's design optimization results were closely aligned across most parameters. Interestingly, even with identical demand, Demandlib necessitated a more modest investment than UKdale and executed fewer energy curtailments. The most pronounced curtailment was observed with LoadProfileGenerator. Yet, it also possessed the largest PV system. In terms of curtailment as a fraction of PV production, LoadProfileGenerator ranked second with 23%, slightly behind synPRO's 23.5%. UKdale exhibited a median curtailment ratio of 21%, with demandlib presenting the lowest curtailment.

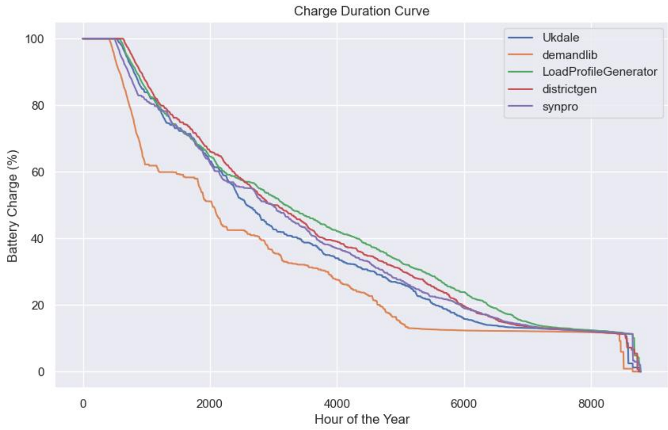
Regarding storage capacity investments, districtgenerator's model demanded the most significant commitment. In stark contrast, the investment needed by demandlib's model was markedly less than its counterparts, standing at merely 41% of what districtgenerator's model required.

Parameter`	*UKdale*	*Districtgenerator*	*SynPro*	*Demandlib*	*LI*
Grid Imports (kWh)	*1929.1*	*2225.0*	*1910.0*	*1824.5*	*171*
PV Production (kWh)	*2151.7*	*2432.7*	*2342.4*	*2143.8*	*334*
Storage Charge (kWh)	*499.1*	*954.5*	*526.6*	*281.1*	*894*
Storage Discharge (kWh)	*382.6*	*736.5*	*406.5*	*217.9*	*692*
Demand (kWh)	*3517.8*	*3938.6*	*3581.3*	*3517.8*	*408*
Excess (kWh)	*446.5*	*501.1*	*551.0*	*387.4*	*772*
PV capacity (kWp)	*2.3*	*2.6*	*2.5*	*2.3*	*3.5*
Storage capacity (kWh)	*1.9*	*2.9*	*1.8*	*1.2*	*2.6*
Storage power (kWp)	*0.9*	*1.4*	*0.9*	*0.6*	*1.3*

Figure IV.7Result of the optimization of the energy models.

The battery's behavior in different models was closely studied. To better understand how the battery operated over time, a charge duration curve was used. This curve showed how long the battery stayed at each charge level.

The result of almost all the charge duration curves is very similar. A notable outlier is the charge duration curve of demandlib's model. In this model the battery spent relatively more time at lower charge levels when compared to the other models.



V. CONCLUSION

A. LPG Biases

In the analysis, demandlib exhibited very conservative simulation. Compared to the actual data, as well as the other simulations, the clustered load profiles had little variation across weekend/weekdays and across the seasons. These characteristics are mainly due to the fact that demandlib is a tool that scales the H0 load profile which is a standard set of load profiles that is scaled to fit the model requirements. In the context of EU households with weather patterns similar to German weather patterns, and behaviours similar to German household behaviours. Demandlib provides relatively accurate results as can be seen from the optimization results.This however could not hold true for other regions were behaviour is different as the H0 load profile might not be an accurate representative of other residential household load profiles.

Ukdale and Demandlib also both had a recommended PV system size of 2.3 kW. Furthermore, the grid import error of demandlib was only 5%. This could be attributed to the accuracy in annual energy expenditure as well as the accurate shape characteristics. Nevertheless, since demandlib's load profile is relatively flat and even, it highly underestimates the required energy storage sizing since there are no peaks to require the flexibility provided by energy storage. Moreover, the stability of the load profile across the seasons also means that the batteries can be used more efficiently as demonstrated by the charge duration results.

Districtgen on the other hand exhibited inconsistencies in the shape and peaks. This can be mainly attributed to its stochastic nature. This also means that every new simulation will yield different results, albeit with the same inconsistencies and wild peaks. This means that using the tool will lead to very overestimates in investments and costs. In the design optimization analysis, district generator required the biggest storage capacity. This is required to cover the big evening peaks. Furthermore, Districtgenerator had the highest requirement of grid imports to cover the excess peaks that solar could not cover especially the ones that happened at midnight as can be seen from the following figure. This also means that for the same energy demand, Districtgenerator will underestimate the required PV system size as it overestimates the demand in the summer when sun is plenty, underestimates the winter requirements, and creates peaks and demand in hours of the day when there is no sun hence also leading to an overestimate in storage investments.

In both instances of the simulation, LoadProfileGenerator's overestimated the midday peaks which led to a huge PV capacity requirement in the optimization results since a lot of PV is needed to cover the midday peak. The high midday peak and large PV system also led to the highest energy curtailment. This assumption is true for any load profile with big morning or midday peaks. It is important to note however that LoadProfileGenerator's simulation model has many parameters, and the load profile can be engineered by introducing/removing appliances, usage limits, and occupant behaviour. Nevertheless, in a real situation where the researcher usually does not have all the details of a household, or if modelling multiple households is required, then creating the simulation model is either impossible or at the very least time consuming. It is also important to note that any mistakes in modelling will introduce different biases in the load profile and despite the time investment required, results can be very inaccurate as was the case in the models simulated for UKdale and WPuQ where random garden appliances that were included in the simulation and a stay at home occupant usage of these appliances led to huge midday peaks and huge PV investments in the optimization results.

SynPRO's results depended very heavily on the number of occupants. The tool itself has limited parameters for electric profile measurements. So, each of these parameters can have a huge effect on the simulated load profile. In the results, modelling a house with two occupants produced a total annual energy demand of 2165.5 kWh. However, modelling a house with 4 occupants produced an energy of 3581 kWh. In the first case, synPRO was very inaccurate and underestimated the energy demand. In the second case, it was accurate and even in the monthly energy demand analysis, it had the lowest MAPE of 8.2%. However, it is important to mention that synPRO was the only tool to assume higher electricity expenditure in the winter compared to the summer. And although in the UKdale case that was study, this was the correct assumption, some of the houses in the WPuQ dataset exhibited the opposite seasonality. Nevertheless, synPRO's design optimization results were very accurate and close to UKdale's results. The load duration curve produced was also accurate and closely related to the actual data.

Figure V.1 Districtgenerator's model operation in the month of june

B. Best Practices

The following section elucidates best practices derived from a comprehensive analysis of various load profile generation tools. These guidelines aim to aid researchers in achieving unbiased energy model optimization results and ensuring that simulated load profiles closely mirror real-world scenarios.

Parameter Selection: Choose tools that offer flexibility in parameter selection, enabling accurate representation of the specific scenario under study. The time

Regional Compatibility: Ensure that the foundational dataset of the tool (such as H0 for Demandlib) is in alignment with the behavioural and weather patterns of the region in question. Other tools, like RAMP, assert their capability to simulate load profiles in diverse contexts.

Stochasticity Awareness: For tools like Districtgen, be cognizant of the inherent unpredictability and its implications on results. Some tool rely on a stochastic methodology to produce load profiles. While this might provide non-representative load profiles, they might be more appropriate to use for aggregate households and communities as they will avoid being too specific.

Avoid Overfitting: While it's essential to align simulations closely with real-world data, be wary of overfitting. A model that is too tightly fit to a specific dataset might not generalize well to other similar scenarios.

Seasonal Analysis: Account for seasonal variations in energy consumption, as they can significantly influence PV and storage investment estimations.

Sensitivity Analysis: When using a new tool or creating a new model, it might be useful to perform a quick sensitivity analysis to understand how each parameter could affect the resultant load profile or the overall result of an energy model optimization.

REFERENCES

[1] E. A. Martínez Ceseña *et al.*, “Quantifying the Impacts of Modelling Assumptions on Accuracy and Computational Efficiency for Integrated Water-Energy System Simulations Under Uncertain Climate,” *IEEE Trans. Sustain. Energy*, vol. 13, no. 3, pp. 1370–1382, Jul. 2022, doi: 10.1109/TSTE.2022.3155073.

[2] A. Almuhtady, A. Alshwawra, M. Alfaouri, W. Al-Kouz, and I. Al-Hinti, “Investigation of the trends of electricity demands in Jordan and its susceptibility to the ambient air temperature towards sustainable electricity generation,” *Energy Sustain. Soc.*, vol. 9, no. 1, p. 39, Dec. 2019, doi: 10.1186/s13705-019-0224-1.

[3] H. Kazmi, Í. Munné-Collado, F. Mehmood, T. A. Syed, and J. Driesen, “Towards data-driven energy communities: A review of open-source datasets, models and tools,” *Renew. Sustain. Energy Rev.*, vol. 148, p. 111290, Sep. 2021, doi: 10.1016/j.rser.2021.111290.

[4] J. C. Hernandez, F. Sanchez-Sutil, A. Cano-Ortega, and C. R. Baier, “Influence of Data Sampling Frequency on Household Consumption Load Profile Features: A Case Study in Spain,” *Sensors*, vol. 20, no. 21, p. 6034, Oct. 2020, doi: 10.3390/s20216034.

[5] “Overview — demandlib 0.1.10 documentation.” Accessed: Oct. 24, 2023. [Online]. Available: https://demandlib.readthedocs.io/en/latest/readme.html

[6] M. Hellwig, “Enwicklung und Anwendung parametrisierter Standard-Lastprofile,” Jan. 2003.

[7] “RWTH-EBC/districtgenerator: Tool for demand profile generation in districts.” Accessed: Oct. 24, 2023. [Online]. Available: https://github.com/RWTH-EBC/districtgenerator

[8] “TEASER - Tool for Energy Analysis and Simulation for Efficient Retrofit.” RWTH Aachen University - E.ON Energy Research Center - Institute for Energy Efficient Buildings and Indoor Climate, Oct. 12, 2023. Accessed: Oct. 24, 2023. [Online]. Available: https://github.com/RWTH-EBC/TEASER

[9] “richardsonpy.” RWTH Aachen University - E.ON Energy Research Center - Institute for Energy Efficient Buildings and Indoor Climate, Jul. 07, 2023. Accessed: Oct. 24, 2023. [Online]. Available: https://github.com/RWTH-EBC/richardsonpy

[10] T. Loga, B. Stein, and N. Diefenbach, “TABULA Building Typologies in 20 European countries − making energy-related features of residential building stocks comparable,” *Energy Build.*, vol. 132, Jul. 2016, doi: 10.1016/j.enbuild.2016.06.094.

[11] “LoadProfileGenerator.” Accessed: Oct. 24, 2023. [Online]. Available: https://www.loadprofilegenerator.de//

[12] N. Pflugradt, “Modellierung von Wasser und Energieverbräuchen in Haushalten,” 2016.

[13] F. ISE, “SynPRO Home.” Accessed: Oct. 25, 2023. [Online]. Available: https://synpro-lastprofile.de/

[14] D. Fischer, A. Härtl, and B. Wille-Haussmann, “Model for electric load profiles with high time resolution for German households,” *Energy Build.*, vol. 92, pp. 170–179, Apr. 2015, doi: 10.1016/j.enbuild.2015.01.058.

[15] “A modular open source framework to model energy supply systems.” Accessed: Oct. 24, 2023. [Online]. Available: https://oemof.org/

[16] “oemof community.” Accessed: Oct. 24, 2023. [Online]. Available: https://github.com/oemof

[17] J. Nagel, *Optimization of Energy Supply Systems: Modelling, Programming and Analysis*, vol. 69. in Lecture Notes in Energy, vol. 69. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-319-96355-6.

[18] “design optimization — ESyOpT 30.01.2020 documentation.” Accessed: Oct. 24, 2023. [Online]. Available: https://ob220.pages.fraunhofer.de/esyopt/Modell/theoretical_background/design_opt.html

[19] P. Price, “Methods for analyzing electric load shape and its variability,” Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2010. Accessed: Oct. 27, 2023. [Online]. Available: https://www.osti.gov/biblio/985909

[20] H. Li, Z. Wang, T. Hong, A. Parker, and M. Neukomm, “Characterizing patterns and variability of building electric load profiles in time and frequency domains,” *Appl. Energy*, vol. 291, p. 116721, 2021.

[21] L. Chuan and A. Ukil, “Modeling and Validation of Electrical Load Profiling in Residential Buildings in Singapore”.