Generation of high spatial and temporal resolution heat demand profiles for Germany

Submitted by:

Aadit Malla

553395

Thesis submitted as a partial fulfilment of the requirements for the degree
Of
"Master of Engineering (MEng) in Energy and Environmental Management"

Supervisors:

1. Prof. Dr. Bernd Möller
2. Dr. Maria Vanegas-Cantarero

Energy and Environmental Management
Interdisciplinary Institute of
Environmental, Social, and Human Sciences
Europa-Universität Flensburg
Germany

August, 2021

Table of Contents

List of Acronyms

ACF	Auto-Correlation Function
ADF	Augmented Dicky Fuller test
BBR	Federal Office for Building and Regional Planning
BDEW	Bundesverband der Energie- und Wasserwirtschaft
CDS	Climate Data Store
CHP	Combined Heat and Power
COP	Coefficient of Performance
CTS	Commercial Trade and Service
DH	District Heating
DHN	District Heating Network
DHW	Drinking Hot Water
DIW	German Institute of Economic Research
DWD	Deutsche Wetterdienst
ECMFW	European Centre for Medium-Range Weather Forecasts
eGo	Open Electricity grid optimization
eTraGo	Electricity Transmission Grid Optimization
EUF	Europa Universität Flensburg
GJ	Giga Joule
HDD	Heating Degree Days
IDP	Intra Day Profiles
IEE	Institute for Energy Economics
IRENA	International Renewable Energy Agency
IWES	Institute of Energy and Wind system technology
IWU	Institut Wohnen and Umwelt
LPG	Load Profile Generator
MFH	Multi-Family Household
MRAE	Mean Relative Absolute Error
MV	Medium Voltage
MW	Mega Watts
MWh	MegaWatt Hours
NUTS	Nomenclature des unités territoriales statistiques
OEMOF	Open Network Modelling Framework

OPSD	Open Power System Data
RLI	Reiner Lemoine Institute
RMSE	Root Mean Square Error
SFH	Single-Family Household
SH	Space Heating
SLP	Standard Load Profiles
Ta	Allocation temperature
TRY	Test Reference Years
TUM	Technical University Munich
TUS	Time of Use Survey
TWh	Tera Watt Hour
VRE	Variable Renewable Energy

List of Equations

List of Figures

List of Tables

Acknowledgement

I express my sincerest gratitude to Europa Universität Flensburg, Department of Energy and Environment management for the opportunity and support throughout the master's thesis and the master's program. I am indebted to my thesis supervisors Prof. Dr. Bernd Möller and Dr. Maria Vanegas-Cantarero for their valuable guidance and encouragement throughout the study.

I also express extreme gratefulness to eGon research project and acknowledge the entire eGon team at EUF for providing the platform without which this study would not have been possible. I am wholeheartedly thankful to Mr. Ulf Philip Müller for his continuous supervision, support, and motivation. I am also grateful to Ms. Clara Büttner for her active support with quick responses to queries along with vital suggestions and encouragement for accomplishing the study. Also, a special thanks to Ms. Eva Wiechers for her continuous help with valuable inputs on the German heating sector.

Finally, a mention to my family and friends for their continuous support and motivation.

Executive Summary

The transition in the energy sector is seen as pivotal for the decarbonization of the German energy sector towards an envisioned carbon-neutral economy by 2050. But this transition seems to be primarily focusing on the electricity sector, with other high energy-consuming non-electric sectors still largely dependent on fossil fuel. Hence a detailed look into sector coupling becomes important. However, a primary challenge associated with it would be a significant rise in electrical demand. Thus, this requires a lot of policy-level changes, market alterations, and the development of efficient technologies. Furthermore, the process requires robust planning and a well-analyzed pathway incorporating both demand and supply which can be accomplished with energy modelling studies. However, the results of any energy modelling study are largely dependent on the availability of input data which in the non-electric sector is very limited as opposed to the electric sector. Electrical grid balancing requires intricate high spatial and temporal resolution data on both supply and demand. However, such detailed data are not available in the non-electric sector, primarily because the currently practiced independently operating systems do not necessitate it. The heating sector, one of the most energy-intensive sectors in Germany, has data only on the level of total annual demands with high-resolution data nonexistent, at least on open-source platforms. The lack of data makes detailed analysis difficult as most energy modeling studies rely on open-source data.

Thus, to overcome this gap in data availability on a high granularity, this study was undertaken in collaboration with eGon research project. The study aimed at the development of an open-source database of heat demand profiles. Though the final database and its usability are aimed at meeting the requirements of the eGon research project, its availability and use for any other work are not restricted. Under eGon, the data has planned use for grid optimization analysis. This made the need for correct representation of the profile peaks and variability important. Such variability and peak representation could not be achieved from the existing state of the art. In addition, the existing state of the art was also seen to have shortcomings when dealing with the expected level of granularity. Other methodology and databases identified over the study were observed to generate demand only on a microeconomic level (typically a single building), thus not meeting the macroeconomic requirements of the study. Hence the development of a new methodology.

The dependence of heat demand on numerous parameters including temperature, building properties, occupant behavior, and source of heating fuel makes statistical estimation difficult. Intra Day Profile (IDP) methodology was developed to incorporate all the heat demand-dependent parameters as best possible. The methodology was successfully implemented to develop heat demand profiles on a high spatial resolution of 100X100m^2 (per census cell) and an hourly temporal resolution. The correct representation of the temperature-demand relationship was given priority. In addition, the consideration of the building typology resulted in a better representation of thermal building mass. For replicating the occupant behavior, the methodology undertakes a stochastic approach based on consumption patterns observed from the Time of Use Survey. The induced randomness gives the profiles their uniqueness and higher variability.

The IDP methodology developed a pool of 24hrs. profiles each represented by a combination of household characteristics and temperature based on the controlled runs of the load profile generator, a micro-level demand estimation model. The random assignment of the pool profiles to each census cell and each day results in the high variability in the IDP profiles. For ease in data storage, the generated profiles were aggregated to potential district heating networks or induvial supply grids. Nevertheless, the database provides the capability to generate census cell level outputs through the python-based eGon-data library.

The absence of real-time measured data restricted comparison to existing studies for validation of the output. Comparative tests regarding nature, structure, and patterns in the profiles were conducted. Compared to the reference profiles, consistent similar results were observed on higher aggregation levels (regional or national) with replication of stationary nature of the profiles. On a census cell level, the similarity was much lower. However, considering the lack of variability in other methodologies this dissimilarity is desirable. Pearson's correlation results showed higher similarity on a daily resolution over hourly. Again, the desired temporal variability deems the results to be welcoming. The structural similarity tests further validated Pearson's coefficient results. The generated profiles replicated structures of the OPSD database which is based on the existing state of art. No structural similarity could be obtained with the DIW database. The demand profiles also showed patterns mimicking the temperature profiles. The demand of any given hour was seen to be dependent, with statistical significance, on the prior 3 consecutive 24th-hour demands. Furthermore, the best representative of the temperature-demand relationship was identified to be a sigmoid curve, with correlation coefficients ranging between values agreed by other secondary literature.

All in all, the IDP methodology was able to develop profiles that meet the desired spatial and temporal resolution, obtaining of which from existing methodologies was not possible. The open-source availability of this database provides the possibility of its use for a wide range of energy modeling studies. The IDP methodology also showed possibility of replacing the state of the art on a high spatial resolution. But further detailed application and usability tests would be suggested to confirm it.

1 Introduction

Energy transition (Energiewende) in Germany has targets to reach a climate-neutral, nuclear-free economy by 2050. The transition towards cleaner and renewable energy has been identified as a major requisite for the transformation of the energy sector from fossil dependence to carbon neutrality. Decarbonization of the energy sector requires immediate implementation of climate mitigation approaches, 90% of which is estimated to be achieved through Renewable energy integration and energy efficiency measures This transition necessitates a cohesion between numerous aspects including information/smart technology, policy frameworks, and market instruments (IRENA, 2020). To ensure a successful transition, Germany has emphasized the integration of RE technologies with a significant reduction in fossil-based energy generation. But the transition has been observed to primarily focus on the electricity sector, with other high energy-consuming sectors (non-electric) such as gas, heating, and mobility still heavily dependent on fossil-based sources. This is verified by the 36% share of RE in the electricity sector compared to just 13% in the latter (Umwelt Bundesamt, 2018). To guarantee a complete transition for a carbon-neutral energy system, the consideration of sectoral coupling is crucial to ensure the inclusion of all energy-consuming systems and their integration into the renewable supply chain (Kannan, 2018). Energy system modeling provides a possibility to analyze the transition and provide a nexus between energy systems and economic models which integrates all these energy-consuming sectors on the demand side and with the supply side.

An increasing trend in the global heating demands because of population growth and improvement in the quality of life has been imminent. But only 10% of this increasing demand is met from renewable energy sources globally. The heating demand in Germany comprises most of the final energy consumption of all buildings. For the residential sector as of 2020, the space heating demand was observed to be 464 TWh and DHW of 103 TWh whereas for the non-residential buildings the total SH was 245 TWh and SH of 23 TWh (Erigggksen Freja, 2020). Of the total final energy consumption in the German residential sector, 13% account for DHW, and 70% account for space heating (Fischer et al., 2016). The majority of this demand is met by natural gas with the heating sector consuming 53% of the total gas sold in Germany (Hellwig, 2003). Therefore, the electrification of this large share of demand would make a great stride towards the country's target of a carbon-neutral economy.

A crucial step in energy system data modeling is data processing, thus ensuring the formulation of the data as per the requirement of the modeling tool (Fleischer, 2021). This is in turn determined by the availability of required input data and its granularity. However, there always exists a trade-off while identifying a working level of data granularity with regards to computational complexity and the desired detail in the output. For instance, there could be cases where even the availability of high granular data would require a level of aggregation for better analysis and reduced computational loads. On the contrary, there could be instances where the higher degree of granularity provides a better result with a higher degree of confidence such that the flatness of the optimization is reduced. Though open-source data are available especially in the European context, the availability of the same on a high temporal and spatial resolution (high degree of granularity) is very limited (Fleischer, 2021). Therefore, before model formulation and interpretation of the results, the pre-processing of the data becomes equally important.

Temporal and spatial high-resolution models of both generation and demand are requirements for the system analytical evaluation of energy systems comprising of a high share of renewable energy supply. Though high spatial and temporal resolution data on the generation side are available at high transparency, demand-side data are rare and require further assessment (Gotzens et al., 2020). For the generation side, primarily electricity, the availability of high degree granular data can be assured. This is bona fide in the case of variable renewable energy (VRE) sources where high-resolution data is a prerequisite for system optimization. For wind and solar, understanding the capability and complementarity of these resources to meet the demand on a very high resolution is crucial to increase the share of these generation sources in the power system (Couto & Estanqueiro, 2020). Even if in case the generation data is not openly available close estimates can be made based on the installed capacity and the location of the system.

However, the same cannot be said on the demand side where accurate estimation becomes difficult in the absence of measured data. More so on a higher resolution due to its dependence on consumer behavior. This is even more true in the case of thermal demands over electricity in terms of data granularity (Fleischer, 2020), considering that the heat demands are largely affected by the ambient temperature (Chramcov, 1982) and region-specific consumer behavior are also defined by the fuel source used for heat supply (Kozarcanin et al., 2019). Nevertheless, understanding all demand patterns and their behavior is important for ensuring an optimal coupled energy system. However, such data is largely absent in the required granularity level. Thus, the thesis focuses on the development of a methodology and a final database that can imitate the heat demand observed in Germany in a high spatial and temporal resolution such that it may be used in the future, for sector coupled energy modelling studies.

1.1 Background of the Study

This master's thesis study (heron referred to as **the thesis**) was conducted in collaboration with eGon research project (referred hereon **as the project**) team at EUF to develop a methodology and analyze the heat demand profile in Germany developed using the same. With this, a high spatial and temporal resolution (100*100-meter sq. spatial resolution/census cell and hourly temporal resolution) demand profile was generated which could meet the requirements of the project optimization tool. The inputs required for the study were acquired from project partners who have been working independently on numerous parameters which have an impact on the head demand. The result of the study is an important component of the demand database used by energy modelling tool eTraGo, developed by the project.

The project has been working on the development of a transparent, cross-network planning instrument for the electricity system to determine economically favorable network expansion scenarios. The project aims to investigate the necessary grid expansion in the German electricity grid caused by the integration of renewable energy and other energy sectors (Reiner Lemoine Institut, 2015). With their extensive work in the electricity sector, the project has been expanding its model to integrate areas of gas, e-mobility, and heating. This permits the examination of cross-sectional synergies for the future energy system (EUF, 2020). Considering the fluctuation in the supply of renewable energy and the changing demand pattern of sector coupling, the electrical grid is subjected to new challenges. However, the integration of the above-mentioned non-electric sector provides the potential to increase the stability of the grid. Hence the project has been focused on investigating the benefits of the obtained flexibility through the coupling. The project at EUF has been focusing on the integration of the heating system for the system modelling. The thesis is thus aimed at achieving it.

1.2 Motivation

The heating sector is a vital component of a coupled energy system considering its high share in the overall energy demand. Numerous studies have been conducted in the past covering the heating demand sector in Germany aimed at best replication of the demand profile. These studies have developed national-level standard load profiles based on historical measurements and statistical analysis which have been implemented in numerous energy modelling and sector coupling studies. However, the limitations of these methodologies are quite prominent (Fallahnejad & Eberl, 2016). The most popular used methodology the BDEW methodology, explained in section 2.2.2, makes use of gas demand profiles as a proxy for heat demand. Also, a major shortcoming of these methodologies would be the lower variability offered in a high spatial resolution. The method also overlooks the peaks in the profiles, whose correct estimation is crucial considering its impact on the system costs in the coupled system (Zeyen et al., 2021). Also, discussions in the scientific community were encountered where issues regarding the lack of availability of such granular data were debated with no conclusive solution (Open Energy Modelling Initiative, 2021). The restrictions are mainly associated with legal and financial constraints associated with real-time measurement. Also, related complication with the handling of high-resolution data is an issue. In addition, most top-down methodology overlooks the consumption on an appliance level thus restricting analysis on individual technologies (Kleinertz et al., 2017) and also misinterpret peaks with reduced profile variability (Fischer et al., 2016). Hence this study is aimed to fill this gap on the availability of data, by developing a methodology making use of openly accessible data and create a database that is publicly accessible. This methodology will be hereon referred to as the Intra Day Profile (IDP) methodology.

1.3 Research questions

The thesis aims to develop and make use of this IDP methodology and cover the gap regarding the unavailability of a high-resolution realistic heat demand profile. The study will be carried out in collaboration with the project and will focus on developing a final output that meets the resolution requirements of the project. The objectives of the proposed study are as stated below:

Objective 1: Develop and implement a methodology to generate high spatial and temporal resolution demand profiles for Germany

- Identify technical and economic parameters affecting heat demand and generate generalized base profiles with minimal variations to these parameters.
- Identification of open-source data sources applicable for the generation of high-resolution demand profiles.

Objective 2: Validate the generated results and ensure their applicability for energy modeling studies

- Identify and evaluate valid state-of-the-art reference sources
- Identify statistical measures for the validation of the heat demand profiles and implement them for the testing of the results

The study aims to answer the following research questions:

- How can the shortcomings of the existing state-of-the-art be curtailed to ensure better replication of peaks and variability in the heat demand profiles?
- What optimal set of assumptions can be undertaken to best generate these heat demand profiles ensuring minimal data input and computational load and maximum accuracy?
- Can the existing SLPs be substituted by the heat demand profiles developed from the proposed IDP methodology?

1.4 Limitations of the study

- Newer building classes after 2009 have not been integrated into the LPG model and hence the actual latest IWU definitions have not been considered. Alterations could be made to the LPG to improve the results.

- The closed source LPG model limits the alteration in the IDP pool, hence limiting it to the pre-generated base profiles. This on one hand would mean consistent replicability, but limitation regarding the possibility of inclusion of higher variability.
- Weekday and weekend factors have not been considered for the study. Its consideration would mean additional input for the generation of the base profiles and further division in the intraday pool. This could be further implemented in future work, but its significance on the census profiles is unknown.
- The profiles for comparison and validation are of different years. Hence the comparison in the most case has been done with the normalization of profiles. Such an approach limits the direct comparison of the demand magnitudes.
- The unavailability of data on desired census level for the CTS sector resulted in lesser variability in the profiles compared to the residential profile. Changes could be made if better census level CTS demand can be acquired in the future.
- Though statistical tests were conducted for the validation of the profiles, the actual real statistical behavior is unknown. Considering the reference profiles are statistical estimates themselves, similarity to them does not necessarily prove their similarity to the real profiles. Therefore, if any close source real measured values are acquired in the future, the comparison results would be able to truly validate the profiles.
- Since the statistical comparison of single-year time series is uncommon, finding statistical methods that ideally fit the requirements of the study was difficult. Hence some common time-series comparative methods are implemented with adjustments as per need. Nevertheless, the made adjustments are verified to be feasible based on secondary literature.

2 Literature Review

2.1 Methods on Estimating Heating Demand Profiles

Fischer et al., 2016 and Ruhnau et al., 2019 define three possible measures for the estimation and best replication of energy demand: standard load profile, reference load profile, or through statistical data-driven approaches. In the case of heat demand, the correct estimation requires the consideration of weather, building properties, and consumer behavior. With these inputs, the demand profile can be evaluated by either one of the above-mentioned methods. Each of these approaches is briefly discussed below.

Standard Load Profiles

Estimation of profiles can be done with a representative load profile that can be generalized on a national level and referred to as Standard Load Profiles or Synthetic Load Profiles (SLP). SLPs are primarily used for the calculation of non-metered energy consumption patterns (Fallahnejad & Eberl, 2016). In the energy sector dealing with modeling and forecast, the practice of utilizing the SLPs is common mostly amongst grid operators and grid modelers (D. Peters, R. Völker, 2020). It becomes a convenient measure to best replicate a demand curve, especially when real-time measured data are scarce. SLPs are derived from historical measured values (Fallahnejad & Eberl, 2016) and hence considered a close representation of the consumption curve.

For heating demand, the commonly used hourly resolution SLP was identified to be based on the methodology developed by TU Munich, 2016 which has been elaborated in section 2.2.1 SLPs are developed based on an average historical value derived from measurements, linked to the ambient temperature, and scaled up to the annual heat demand (Fischer et al., 2016). Generally, the historical data is acquired from publicly available data sources, in most cases from local utility companies. Since the profiles are generated with an average of a large variety of data from numerous locations the profiles can be considered representative of all national heat demand behavior. However, the average values would also mean the reduction of accuracy concerning individual profiles. Nevertheless, this approach is also able to partly represent the individualities in the buildings as the demand depends on the building properties and the solar heat gains due to the building properties (Fischer et al., 2016). For German applications, the average load profile for heating gas is used as a heat load profile for households, as explained in (BDEW et al., 2020) section 2.2.2.

Fallahnejad & Eberl, 2016 further discuss in detail the common practice of the use of SLP to calculate the natural gas demand based on the temperature forecast. The most recent version of these load profiles described with the SigLinDE function as described in section 2.2.2 is used as a representative of the natural gas profiles. Considering the similarity in the heat demand and the natural gas profiles, the use of these profiles as a proxy for the largely unavailable heat demand time series has been a common practice.

A major issue associated with SLPs is the induced limitations with regards to their ability to forecast future long-term scenario demands since the profiles are based on historical values. Stegner et al., 2019 mention of the increasing reliance on intermittent energy resources make existing electrical SLP unrealistic for future prognosis. Also, Fallahnejad & Eberl, 2016 indicate that the lack of consideration of improved building typology, use of efficient heating system, changes in the family size, and floor area per person causes unreliability issues with the thermal load profiles. Smart metering systems make the possibility for the improvement in the electrical SLP possible (Stegner et al., 2019), but financial and technical constraints in the heating sector make the further improvement of the SLP difficult.

Fallahnejad & Eberl, 2016 made alterations to the SigLinDE profiles (the existing SLPs) to better represent the extreme temperature for the future forecast. The alterations were done such that no additional effect on the computational time of the profiles is observed. The profiles are generated with the assumption that all demands over 24°C are comprised of DHW and unaffected by the h-factor. A similar assumption has also been made in this study. The observance of extreme temperatures over 28°C and below -20°C is rare in Germany and the h-values are constant beyond this with the justification that the heating system is turned off or operating at full load respectively. The results from Fallahnejad & Eberl, 2016 showed an anticipated shift in the space heating requirements towards colder temperature which would otherwise have been overlooked if the demand profiles were based on traditionally used SLPs. Thus, indicating an improvement and higher realistic nature of the profile. Nevertheless, applying the methodology to every census cell would make the entire process cumbersome. This provided another strong reason for the development of a new methodology and moving away from the use of SLP.

Reference Profiles

In this method, the entire year is categorized into a limited number of representative days. These representative days are then assigned to every individual day based on certain pre-defined day characteristics. One such common model is the VDI-4655 where the year is divided into 10 representative days. The LPG model is validated against the VDI 4655 model which is further described in section 2.3 (Drauz, 2016). The VDI 4655 has a higher variability because of higher simultaneity as multiple events occur at the same time (Ruf et al., 2016). However, since the model is a closed source, there was no direct interaction during the thesis and all the information is based on other secondary literature that has implemented this model. Considering the priority that was given to open-source modelling by this study, the VID 4655 was not looked into in detail further.

Data-Driven Statistical Regression

These models can accurately predict the demands of known buildings but give poor results for new or unknown buildings. Fischer et al., 2016 also mention the use of physical models based on lumped energy balance equations and physical building properties for the replication of the profiles. However, these are more suitable for individual building energy simulations and not specifically applicable for large-scale (national level) demand-based studies as conducted in the thesis.

Summary

Though all the above-mentioned approaches replicate the demand profile to a certain degree, each approach has its limitations. Fischer et al., 2016 suggest the combination of any two of these models as a possible intervention for improvement on the results. Also, the possibility of considering different socio-economic factors which have an impact on the heat demand as a possible alternative for improvement. However, with regards to this study identifying these factors and acquiring information of the same on the undertaken high spatial resolution was challenging. Fischer et al., 2016 also mention the approach of randomization through clustering of building types. A similar approach of randomization has been incorporated in the methodology followed in the thesis. Also, a stochastic bottom-up approach and aggregation are seen as ideal, as these would permit for higher variability in the individual profiles making the results more realistic.

Though improvements are continuously made to the standard load profiles (Fischer et al., 2016), (Fallahnejad & Eberl, 2016), concerning the scope of this study the variability in the profiles is still absent (further explained section 2.2.2). Another major issue associated with the thermal SLPs is the smooth nature of the curves over a 24hrs. period as seen in Figure 2-1. Though such nature is acceptable on an aggregation level, when dealing with profiles on a high spatial-temporal resolution, the results are not accurate. To overcome these issues the need for improvement was identified. However, limiting the input data requirement and the computational time was crucial. Hence further research was conducted in this study for better representable and highly variable demand profiles with the introduction of a new methodology possible with a potential to replace the existing SLPs on a high spatial resolution.

2.2 The State of the Art

Though data on annual total heat demands are available in the desired spatial resolution (SEEnergies, 2020), its corresponding high temporal resolution is absent. Thus, to accommodate for this into energy system analysis, methodologies have been developed over the years to best replicate the heat demand patterns following either of the 3 approaches mentioned in section 2.1

The pioneer in the sector and basis of the existing state of the art was developed by the Bundesverband der Energie- und Wasserwirtschaft (BDEW) which makes use of the natural gas demand as the proxy for the heat demand in Germany. Other improved or simplified versions have emerged over the years which are based on the original methodology developed by Hellwig, 2003. Though the methodology was specifically developed to accommodate the demand patterns of the German heating sector.

2.2.1 TUM methodology

The methodology developed by Hellwig, 2003, hereon referred to as the TUM methodology (developed at the Technical University Munich) primarily focused on the analysis of statistical load profiles. The study extensively defines load profiles for various heat energy-consuming sectors, replicating the gas demand profiles. Due to the flexibility of the methodology and the closeness of the generated profiles to realistic values, this methodology has been undertaken as the basis of the SLP. The TUM methodology was primarily developed for the natural gas network to ensure a fair opportunity for new network operators and gas traders. This methodology allowed the mapping of all levels of natural gas customers throughout Germany providing an opportunity for new players to participate in the liberalized European energy market.

The SLP is popular and commonly used in energy modeling studies due to its ability to generate high temporal resolution profiles based on statistics thus not requiring any measured data. The SLP in its early days provided means of tracking gas volumes purchase on an hourly resolution. Also, the high temporal resolution provides a basis to ensure that consumers are charged fairly for their energy usage (Hellwig, 2003).

Measuring real-time values makes the process quite expensive, and unlike for electricity, real-time data for the balance of the grid is not mandatory for gas/heat grids. Hence the SLP provides a suitable alternative for dividing the annual total consumption into hourly resolution values. Also, if real-time data were to be measured, the equipment cost would have to be borne by the consumer thus making the whole process highly impractical. The statistical dependence of the methodology manages to provide a realistic load profile based on assumptions, and hence the methodology provides an optimum tradeoff for the cost savings compared to real-time data measurement and the accuracy of the generated profiles in terms of the temporal resolution. The data collected as inputs for the methodology is kept in limits in terms of costs, ensuring the non-discriminative representation of all consumers. Hence the TUM methodology can be stated as an optimum trade-off between the requirements for temporal detail and spatial accuracy.

Considering the high correlation between the natural gas demand and heat demand (Ruhnau et al., 2019), the gas profiles generated from TUM have also been applied for the generation of heat demand profiles. Also, the equal high dependence of both natural gas and heating demand on ambient temperature (Hellwig, 2003, p. 25), (Chramcov, 1982) thus gives a higher degree of conformity regarding the use of gas profiles as a proxy for the heat. Also, a direct link between the two can be drawn considering that natural gas was primarily used for heating, hence the similarity in the demand is anticipated. Ruhnau et al., 2019, p. 7 also confirm the possibility of interpreting the gas consumption as a proxy for the temporal profile of heat demand which has been validated by the study in the UK where both gas and heat demand data are available in daily resolution. The comparison of the real-time data and the heat model developed based on the gas demand profile showed a high correlation between the two ($R^2 = 0.95$). Hence the TUM methodology gives a basis for the estimation of the heat demand profiles based on natural gas supply. In the present case of the absence of data, this is seen as the best method for demand replication.

The TUM methodology considered the actual thermal insulation of buildings to generate the demand profiles which has a major impact on the heat demand. To account for the thermal mass (inertia)/ building storage capacity, the profiles developed by Hellwig, 2003 make use of geometric series of temperature allocation which takes into account factors of previous day temperatures to determine the allocated temperature with the equation below (Ruhnau et al., 2019). Such an approach gives a realistic representation of the temperature and in turn the overall demand (BDEW et al., 2020). The geometric series for daily temperature allocation is given by Equation 2-1.

$$T_a = \frac{T_n + 0.5T_{n-1} + 0.25T_{n-2} + 0.125T_{n-3}}{1 + 0.5 + 0.25 + 0.125}$$ *Equation 2-1*

Source: (Ruhnau et al., 2019)

Where T_a is the allocation temperature, T_n is the measured temperature at that interval, and T_{n-1}, $T_{n-2,}$ and T_{n-3} are the corresponding temperatures at that hour in the previous three days.

However, Hellwig, 2003 also mentions that further inclusion of the heat supply technology, resident use pattern, and the building geometry could further help improve the results. For the methodology developed in this study, the resident use pattern has been replicated by the occupancy model, the building geometry and age has been considered by the building classes considered while the generation of the base profiles which define the age of the building and in turn the level of insulation and the building/apartment floor area as per IWU, 2015. The heat supply technology has been fully covered as the methodology is primarily focused on demand-side estimation, however, the aggregation on a district heating system level also provides a possibility to study the effect of the supply system in further studies. Also as indicated by Kozarcanin et al., 2019 the heating energy source with regards to its economic value greatly affects the heating requirements irrespective of the ambient temperature hence the inclusion of such aspects is also expected to bring alterations to the standard profiles acquired from the TUM methodology.

Hellwig, 2003 identified a sigmoid function *Equation 2-2* as a best fitting measure for the regression between demand and temperature represented in terms of the hourly scaling factor, which was further verified by Louvet et al., 2019. TUM methodology also suggests the inadequacy associated with the use of an oversimplified linear or a polynomial function, which is seen to be an approach undertaken in a handful of studies including DIW Berlin, 2017 also referred in this study. For instance, the linear function though providing a simple means of calculation shows unrealistic demand behavior when the temperature is too high or too low. Similarly, the polygon function though develops a perfectly fitting regression model for a certain temperature range shows unrealistic behaviors during the two extremes. Overcoming these issues, the sigmoid curve makes use of the combination of best characteristics of both the linear and polynomial curves to give a better fitting line for the relationship.

However, a major drawback of the sigmoid curve would be the existence of a small amount of heating demand even in higher summer temperatures only for SH demand. This does not make a difference when profiles are required on a national or regional level aggregations but needs a level of correction when demand is estimated for individual buildings as done by (Louvet et al., 2019). Nevertheless, for aggregated demand times series, the demand upon exceeding certain limits is expected to remain constant considering certain industrial demand and demand from residential DHW. Also, the curve flattens as the temperature decreases.

$$h(v) = \frac{A}{1 + \left(\frac{B}{v - V_o}\right)^C} + D \qquad \textit{Equation 2-2}$$

Source: (Hellwig, 2003)

where A, B, C, and D are sigmoid parameters which are defined by the sector and building class considered, v is the allocated temperature and v_0 is the reference temperature. h(v) or the h factor gives the daily demand factor.

Hellwig, 2003 points out the importance of consideration of three different parameters which are the cause of the heat consumption. These include transmission losses, infiltration losses, and supply boiler efficiency. Heat demand/consumption is primarily made of the losses through transmission over the entire building shell. Transmission could be positive or negative. The gains include those from internal gains (from the occupants and the appliances used) and from solar irradiance which is stored in the building mass. The first two are considered in the methodology developed in the thesis considering that these are primarily dominated by building parameters and occupant behavior (Kaminska, 2019). The final effect is caused by the boiler efficiency and the losses associated with it. The boiler efficiency has been overlooked considering its analysis exceeding the scope of the study which is intended for estimation of the demand and not much focus has been shed on the supply side which results in the absence of detailed information on the availability of information on boiler use on a high spatial and temporal level as planned by this study. Also, the effect of the boiler is mainly constant throughout the year with a slight increase in losses observed during the summer months due to lower use. Hence an assumption has been made that overlooking this constant minor loss would have an insignificant difference on the result over the considered aggregation

Limitations of TUM

As with any other statistically developed model, there also exists some limitations and drawbacks with the TUM methodology. The generated profiles can represent the structurally determinable components of the curve, which include the proportions based on the consumption structure and the external influencing factors like the ambient temperature. But since the entire model is based on statistical assumptions, the load profile developed is only able to represent consumer demand on a certain aggregation level. For example, considering the residential sector the methodology is unable to represent the individual household curve fluctuations and only on an aggregated level of e.g. 100 households. Also, Hellwig, 2003 mentions that the profile on an aggregated level should be different than on an individual level and in case of observance of any similarity is entirely coincidental. But this difference in the TUM method is absent or not significant enough. This means that if hourly profiles were developed for two households at a given single location and with identical building properties, the profiles from TUM for the two would be identical as well with disregard to the occupant behavior. Also, there would exist a certain summer demand in both these profiles. Hinterstocker et al., 2015 also identified and verified the deviation between allocated temperature and residual load pointing out the drawbacks of the TUM methodology in terms of low allocation of demand during cold temperatures and low allocation of baseload during the warm temperatures. Nevertheless, the methodology provides a concrete means for estimating both heat and natural gas demand and has been continued to be used as the national standard and state of the art.

2.2.2 BDEW methodology

The TUM methodology has been further developed by the Bundesverband der Energie- und Wasserwirtschaft (BDEW) to better accommodate the model for a continuously densifying German gas network. BDEW is an interest group of the German energy industry working specifically in the sector of power production, grid operation, natural gas, electricity, and district heating. The TUM methodology was altered to further accommodate other impacting parameters and provide an improved representation of the demands to provide a basis for the policymakers to build a framework for the network operators. This methodology hereon will be termed the BDEW methodology. Many versions of the BDEW methodology have been developed primarily to overcome the drawbacks of the TUM methodology.

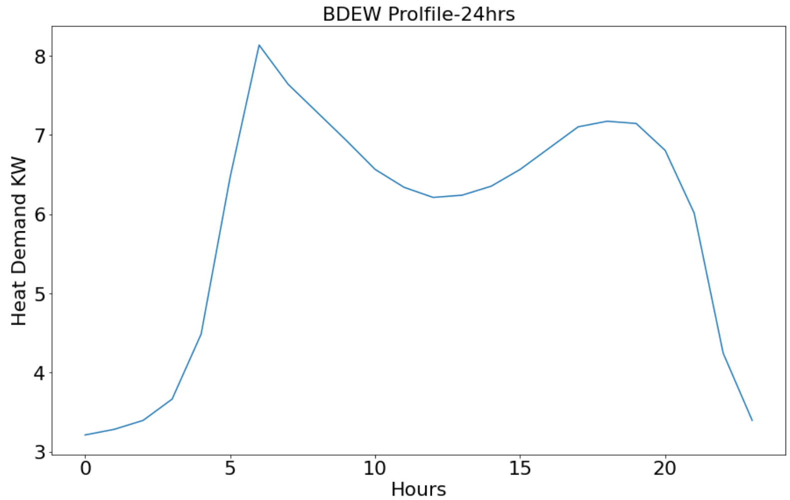
For this study Hinterstocker et al., 2015, Ruhnau et al., 2019 and BDEW et al., 2020 have been referred for a better understanding of the methodology and its improvement over the TUM methodology. The first noticeable difference between the two methodologies is the alteration of the demand-temperature regression curve from the sigmoid curve as used by TUM. The BDEW methodology proposes the use of a SigLinDe function which is a combination of the sigmoid and a linear function. The linear component of the function is used in the extreme temperature zones where the results of the sigmoid or polynomial function are not realistic. Thus, providing a better representation of the heat demand, the output of the methodology has been used as a standard and is referred to as the standard load profile. The SigLinDe function with its linear component is represented by *Equation 2-3* extracted from (Ruhnau et al., 2019).

$$h(v) = \frac{A}{1+\left(\frac{B}{v-V_o}\right)^{C}} + D + max\begin{Bmatrix} m_{space} \cdot T_a + b_{space} \\ m_{water} \cdot T_a + b_{water} \end{Bmatrix} + max \qquad \textit{Equation 2-3}$$

Source: (Ruhnau et al., 2019)

Other than the modified relationship between the demand and temperature BDEW is also based on the statistically determined standardized consumption developed based on a bottom-up model without the consideration of any existing network and exclusively based on process-specific parameters. The developed profiles are based on extensive investigation of individual measurements. However, as mentioned by Hellwig, 2003 the high statistical dependence of these standard load profiles does not truly represent the forecast procedures of the demand curves. In addition, BDEW et al., 2020 mention systematic deviation between allocation and residual load observed in the TUM methodology. Nevertheless, the overlooked seasonal factors associated with the demand curve are improved in BDEW compared to TUM.

The shape of a sample 24 hr. heat demand profile generated from the BDEW methodology would give a profile as seen in Figure 2-1 with the magnitude defined by the hourly demand factor and time of year.



Source: (Author, generated with demandlib)

Figure 2-1: BDEW SLP

Though the curve pattern replicates a generalized 24hrs. heat demand profile, the curves tend to be smoother and underestimate the peaks of DHW. Thus, the BDEW methodology still does not meet the requirements for variability proposed in this thesis. Therefore the study methodology aims to overcome this drawback, creating peakier profiles as seen in Figure 4-2 ((c),(d)).

BDEW software implementation

In line with the scope of the study, Open Network Modelling Framework (OEMOF) developed a python-based library called Demandlib was identified as a software-based implementation of the BDEW methodology. The demandlib library permits the generation of power and heat profiles for numerous sectors and scales them up as per the desired demand. The hourly heat demand values are calculated based on *Equation 2-4*:

$$Q(v) = KW.h(v).F.SF \qquad \textit{Equation 2-4}$$

Source: (Oemof Developing Group, 2016)

Where "*KW*" is the assumed daily consumption at a temperature of 8°C, h(v) is the daily demand factor from *Equation 2-2*. "*F*" is the weekday and "*SF*" is the hour factor. The entire workflow is based on the BDEW methodology with the required input being the annual hourly temperature profile and the total annual demand. The sigmoid function and associated parameters are evaluated based on the considered sector type input. The different sector types in the demandlib module along with the sigmoid factors for each are available in Annex A. The sigmoid factors for the latest building class were also used in the IDP methodology. In addition, other minor aspects of the demandlib have also been undertaken for this study.

Limitations of BDEW

Though a lot of issues associated with the TUM methodology are improved upon in the BDEW methodology, there still exist a few concerns. The variability concerning the spatial distribution and the inability of the BDEW methodology to forecast future demand forecasting considering the improvement in building insulation and use of efficient heating systems (Fallahnejad, 2017). Also Clegg & Mancarella, 2019 argue the lack of consideration of the newer building classes in the SLPs makes its use for estimating future forecasts irrelevant. Hence the IDP methodology was implemented to overcome these issues. Furthermore, Fleischer, 2020 also recommends the use of high granularity data for optimization of systems with high VRE share as reduced spatial resolution has an impact on the system least-cost solutions.

2.3 Load Profile Generator

Bottom-up engineering approaches for modelling and estimating energy demand are seen as strategic decision-making tools especially in the management of transition towards a low-carbon energy system (Fleiter Tobias; Rehfeldt Matthias, 2018). Drauz, 2016 in collaboration with Fraunhofer IEE developed an energetic bottom-up model which can estimate and reproduce the energy demand of the German residential sector per household stock. The model from here on will be referred to as the Load Profile Generator (LPG). The LPG model merges the electricity, heat, and water models to form a comprehensive energy demand generator. The aim of developing the model was to obtain a realistic replication of demand such that a reliable, economically viable, environmentally sound, and self-sustaining decentralized energy supply approach could be developed to meet it. The LPG model estimates the demand for two basic residential household stocks; Single-Family Households (SFH) and Multi-Family Households (MFH). Fraunhofer Institute of energy and wind system technology (IWES) primarily target the use of the LPG model for the optimal inclusion of renewable energy into the system ensuring power security with interest in CHP, a battery, a photovoltaic and a heat storage tank, and a peak voltage boiler with regards to their performance in Single and Multi-Family households (Kneiske & Drauz, 2017).

The thesis can be vaguely considered a continuation of Drauz, 2016. The major difference being that the LPG model on its own generates profiles on a micro level, for individual households, whereas the IDP methodology is more focused on optimizing the LPG output to generate profiles on a macro scale representing the demand profiles per census cell or any other aggregation level for the entire country. Alterations to the LPG wherever possible are made in this study such that data handling complexity is reduced but with minimal effect on the final output.

The LPG model makes use of the White Box modelling approach for estimating the demand patterns of the households. As a result, the output is independent of the historical energy consumption and can largely simulate the behavior of occupants in a building and base the energy consumption on it (Swan & Ugursal, 2009). To better simulate the occupant behavior, LPG considers the appliance distribution amongst the households. Also, the building characteristics have a crucial role especially in terms of space heating demand, and hence have been taken into consideration in the LPG model. Most importantly compared to other reference models analyzed by Drauz, 2016, the ability of the LPG to produce the individuality of each day gives it superiority over the reference profiles and is one of the main reasons for use in this study. This degree of individuality is brought about by the occupancy model. The LPG bases its stochastic occupancy model on Aragon et al., 2019 and Richardson et al., 2008 with the Time of Use probability derived from time use survey by the German Research Data Center (Statistisches Bundesamt, 2016).

The load profile generator is an extensive model comprising the ability to develop electricity, space heating (SH), and drinking hot water (DHW) profiles for a household with a given set of characteristics. These characteristics are provided as an input for each run of the model. Aligning with the scope of this study emphasis was given towards the understanding and implementation of the components of the model responsible for the generation of the SH and DHW demand. The appliances use in the household which is in turn determined by the household and occupant characteristics greatly affects the electrical and DHW demands (Muhammad, 2017),(Kadian et al., 2007). Hence correct allocation of such appliance per household stock and occupant number has been prioritized in the LPG model. For residential SH demand Berger & Worlitschek, 2018, indicate its high dependence on the ambient temperature and the building typology representing the thermal mass of the building. Though the final required temperature is dependent on the occupant's behavior, its effect on the SH demand is not seen to be significant in comparison to the impact of the ambient temperature.

The following section consists of a description of each of the individual LPG component models. Before analyzing the energy models, themselves, it was important to get an insight into the operation of the occupancy model since it is a crucial component for replicating the consumer behavior and central to the high degree of daily individuality observed in the model output.

2.3.1 Occupancy Model

A connection between the three energy models generated by the Load Profile Generator is made through the occupancy model. The bottom-up energy consumption is based on the occupant's state in a household which in turn determines the time and duration of the energy-consuming appliances used. The LPG makes use of the Richardson et al., 2008 model which provides a comprehensive and validated approach for the determination and estimation of the end-use probability. In the most simplistic term, the occupancy model distinguishes the status of occupants, if they are home or not, and if they are active or not.

A simplistic workflow of the LPG occupancy model can be described as follows. For the first period of 00:00h-00:10 hrs. the start state is determined based on the occupant number. The probability of occupancy is defined by the raw data collected from the Time of Use Survey (TUS) for 2012/2013. The survey's probability distribution is based on the daily activity of 300 participants in a 10 min resolution. It classifies the occupants into either active/inactive and present/absent state based on the probability of that state at that time resolution. Once the starting state is determined the following period is determined by the transition matrix. This provides the probability of an occupant in a certain state in the previous period switching to another state in the current period. The probability distribution for each day is obtained from Statistisches Bundesamt, 2016. And the process continues for all periods.

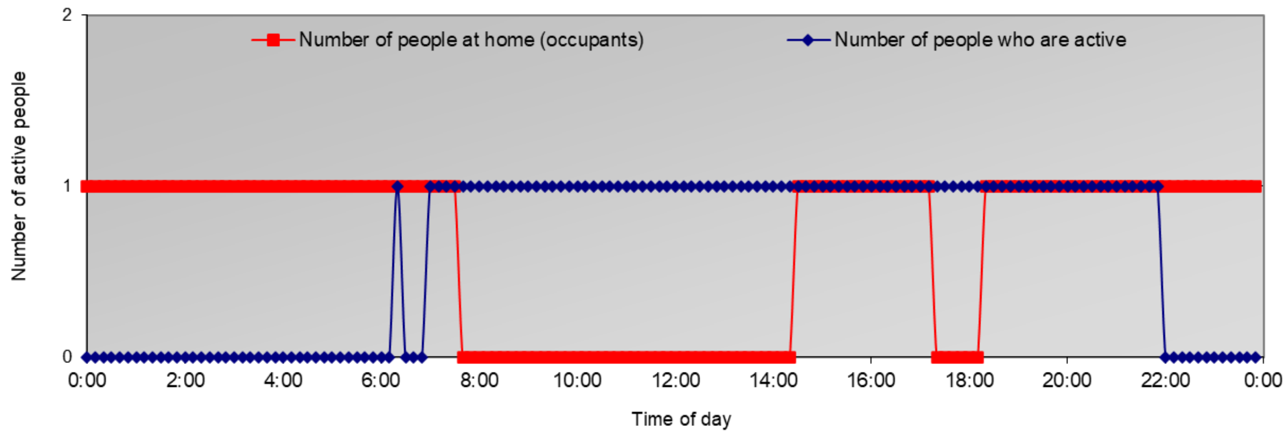
For example, a household with a single occupant has a probability of existing in either of the four-occupancy states (home-active (11), home-inactive (10), not home-active (01), not-home-inactive (00)). However, the probability of each of these states is not equally distributed. A sample occupancy status is shown in the figure below directly extracted from the Richardson model. At 00:00h the probability of the occupant being home and inactive is 80.2% and being home and active is 11.4%. Hence in Figure 2-2 the status state 00:00 is the home-inactive which has a significantly higher probability.

Table 2-1: Probability distribution of each stage at 00:00h

Occupancy State	**00**	**01**	**10**	**11**
Probability	0.0311	0.051	0.802	0.114

Source: (Richardson et al., 2008)

Now with the starting state of the occupancy determined, the transition matrix can be assigned. Again, the probability of occupant at home and inactive, continuing existing in that state at 00:10hrs is 99.8%. With this, the status of the occupant in the second interval is known and the process continues for an entire day.



Source: (Richardson et al., 2008)

Figure 2-2: Sample Occupancy Plot

The final output is a probabilistic daily occupancy status. For a day profile, the same method is implemented over the required period. Thus, the entire day is made up of 144 transition matrices in the Richardson model with each state depending on the probability observed for the corresponding period (Richardson et al., 2008). For the annual profile, the entire process is repeated by checking the corresponding day type. With the status of the occupant determined for a given time, the LPG then estimates the use pattern of appliances in the household/building. For simplicity and reduced computational time LPG only makes use of the occupancy output on an hourly resolution. Thus, the occupancy model is crucial to provide the variability offered by both the LPG and IDP results.

2.3.2 LPG SH Model

The following sections briefly cover the thermodynamic aspects for the working of the model. Extensive details are avoided considering the scope of work. Also, since the model is a closed source, analysis was only possible from secondary literature.

The SH demand based on the DIN V4180-6 in its simplest form is defined as represented in *Equation 2-5* (Drauz, 2016).

$$SH\ demand\ consumption\ = \sum heat\ losses - Utilization\ factor * \sum heat\ gains \qquad \text{Equation 2-5}$$

Source : (Drauz, 2016)

Numerous aspects associated with the respective heat losses and gains have been incorporated in the LPG to best replicate the heat demand. The LPG makes use of the concept of the monthly balance system method for the calculation of the space heating demand. The details on the mathematical calculations and equations used for the calculation of each of these steps are available in Drauz, 2016, pp. 11–17. In general, the LPG model considers three different sets of input parameters for the generation of the load profile. Firstly, the heat losses and gains are determined through the provided input on the temperature and the radiation data. Also considering the internal heat gains the presence of the occupant and the use of appliances are as per the occupancy model. Using the determined gains and losses the heating periods are determined. Also considered are the outside surface which is in direct contact with the ambient air. Finally, the SH profile of one year is determined.

For the implementation of the DIN-V 4180-6 model, LPG has a predefined set utilization factor of 0.95. For the ambient temperature, geometric progression of temperature interval of the last four days is used, which is also recommended by the BDEW methodology with equation Equation 2-1. Such an approach changes the dependency on the temperature in case sudden large fluctuations in the values are observed. For the indoor temperature, the LPG considers 22°C in case of active occupancy.

Drauz, 2016 validated the SH model against the VDI 4655 model, a reference profile demand estimation method. This is a closed source model designed specifically focused on CHP manufacturers to determine the norm degree of efficiency for CHPs is essential for the cost-effectiveness of the CHP. Drauz, 2016 mentions the capability of LPG to overcome the issue of oversimplicity associated with the profiles generated from the VDI 4655 model. The general operation of the model is based on the division of an entire year into 10-day types for sample SFH and MFH. The 10 typical days are classified as Summer and Winter (Weekday, weekend, sunny, and cloudy) and two transition days. Also, the model limits the number of people in an SFH to 12 and the number of apartments in MFH to 40. For every typical day, the VDI model generates an energy demand profile of one day with a time resolution of one minute for SFH and 15 minutes for MFH. The model makes use of the Deutsche Wetterdienst (DWD's) 15 climate zone classification for the extraction of temperature from different sources within the country. The same approach has been used in this study for the weather data needed for the estimation of the heat demand, as explained in section 3.1.1((Federal Office for Building and Regional Planning (BBR), 2014). Nevertheless improved results are developed from LPG than generated by this reference model as claimed by Drauz, 2016.

2.3.3 LPG DHW Demand

The generation of the DHW demand profiles is largely dependent on the occupancy status of any household irrespective of the ambient temperature and building dynamics. LPG's DHW model is based on the VDI 2067 which was in turn developed based on the Jordan Vejan model. The Jordan Vejan model assigns the use of DHW into four categories (short load, medium load, bath, and shower). The model requires a pre-defined input on the total water requirement of the household (daily water demand), which is independent of the occupant number. The model then generates arbitrary tapping incidents (the action of use of a tap) based on a probability function that determines the probability of DHW demand in each time step. Details on the model can be found in Drauz, 2016 and Jordan & Vajen, 2001. Based on the flow rate and the tapping start time the model determines the power demand for the DHW. The final DHW demand is based on the aggregation of each of the tapping incidents. The LPG makes use of a very similar approach the only difference being a reduced complexity with replacement of the four end-use categories by three. Also, the flow rate, taping length, and the frequency of use of every category are pre-defined in the model with a normal distribution of the flow rate and tapping length. The probability of each of the categories being utilized is entirely dependent on the occupancy model.

Interlinkage between the energy models

The occupancy model acts as an interlink between the three LPG energy models. Specific to this study the occupancy model determines the occupant's state which in turn quantifies the heating water demand. Also, the occupant's presence in the home considers the indoor temperature and thus the effect on the SH demand. The occupancy model also affects the SH demand considering the heat gains from the occupants. There exists an interlinkage between the 3 different components of the LPG. The SH demand is influenced by the electricity demand concerning the heat gains obtained from the appliance used. The electrical demand is in turn dependent on the DHW demand considering that the hot water is obtained with the use of electrical appliances. Since the LPG is used for the development of base profiles for the IDP methodology, all the above-mentioned aspects have been taken into consideration during implementation.

2.4 Existing and Identified Database for the validation of the generated profiles

As mentioned previously the absence of measured data resulted in difficulty in the validation of the profiles. As a result, other studies which were also based on estimations and statistical analysis had to be accounted as a reference to validate the generated outputs. Though not the ideal methodology for validation, the results of these studies have been self-validated by their respective authors with some also being used as a national standard. So quantitative similarity along with strong qualitative reasoning to support it would to a great extent provide validation to the IDP generated profiles. A similar validation approach was also undertaken by Drauz, 2016 to validate LPG results comparing the SH output with the VDI 4655 and the DHW demand with the Jordan/Vajen (Jordan & Vajen, 2001) models. Hence the validation approach undertaken in this study can be attributed as scientific.

For the validation, two different datasets, the BDEW based OPSD dataset and the linear and simplified DIW dataset were identified and used for comparison. The OPSD dataset provided a possibility of comparison with the standard load profiles which are aggregated on a national scale. Similarly, the DIW provided a unique perspective with comparison on a district heating (DH) level. In addition, a study based on the stochastic bottom-up heat demand model developed by Fischer et al., 2016 has also been discussed, as suggestions from this study were considered while developing the methods for the thesis. A brief description of each of these reference profiles is included below.

2.4.1 OPSD When2Heat Profiles

For the comparative analysis of the generated profile, the Open Power system data platform was identified as a reliable source. OPSD is a project that develops an open-source free of charge platform primarily dedicated to electricity research. OPSD works with the collection of different scales of energy data, validate and process it to ensure ease of use specifically focused on energy modelling applications. The project has put together a database for the heating demand profiles referred to as the When2Heat Heating profiles (Open Power System Data (OPSD), 2020) which is a simulated hourly country-aggregated heat demand and COP time series for 16 European countries spanning from 2008 to 2018. The When2Heat Heating profile makes use of the SLP approach for the generation of the national level heat demand estimate. Concerning the limitations in the availability of real-time measured data (Ruhnau et al., 2019), the when2heat (OPSD), though with its limitations, is seen as the best option for the validation of the implemented IDP methodology. Also, the availability of sectoral datasets provides the possibility for a much wider range of comparisons.

The OPSD demand profiles for space and water heating are computed by combining gas standard load profiles with spatial temperature and wind speed reanalysis data along with the population geodata (Ruhnau et al., 2019). The heat demand time series are based on the German gas standard load profile approach and defined by BGW and BDEW (BDEW et al., 2020). The inputs used for the generation of these profiles are the 2 m temperature, soil temperature level 4, and the 10 m wind speed acquired from (CDS Climate Data, 2018). The methodology assumes the heat demand to be proportional to the population, gas boilers capability to follow the original heat demand, and gas heated building as representative of the entire German building stock. The spatial time series are then weighted with the population geodata from Eurostat and then aggregated to generate a national-level time series. The national-level output is then normalized to a 1 TWh national average yearly demand.

Three major components for the generation of the OPSD demand profiles are the household-specific demand profiles, the daily demand factor, and the hourly demand factor as assigned by BDEW methodology in section 2.2.2. The temperature sensitivity of these profiles for varying weather conditions is related to the local wind speed which is also considered by this methodology. Hourly demand time series are derived for a location using the hourly demand factors. These hourly factors are dependent on the building types, and 10 different temperature ranges, and for the case of commercial buildings also dependent on the day of the week. To give the profiles a seasonal pattern, these are scaled up as per the daily demand factors. The OPSD methodology considers the use of SigLinDe function section 2.2.2 over a sigmoid function for determining the daily demand factors.

For the generation of the DHW profiles, considering the lack of need of SH on high ambient temperature days (temperature greater than 25°C), the water heating factors are multiplied with the high-temperature hourly demand factor to generate the heating water demand profiles. The SH demand is then calculated as the difference between the calculated total demand and the DHW demand (Ruhnau et al., 2019). Finally, the generated spatial demand time series are weighted against the population and scaled up to the actual demand. The study assumes the residential household share in the ratio of 70:30. The data sets from 2008 to 2013 are scaled to the annual final energy consumption using the EU building database and then corrected for final heat-energy conversion losses. Considering other data inputs in the IDP methodology, only 2011 values have been used in this study.

2.4.2 Heating demand for district heating networks

DIW Berlin, 2017 provides an overview of modelling of time series data based on the CHP maximum and minimum generation defined by the heating markets and the heating networks. The database provided by the DIW Berlin gives a basis for the evaluation of the generated profiles in comparison to a DH level aggregation. The need for a high temporal resolution heat demand time series for individual heating networks is a requisite to determine the commitment and dispatch of the powerplants within the heating network. This could in turn be used to model the operation of CHP units in DH networks.

DIW Berlin, 2017 also mentions the lack of public availability of high-resolution network data which has also been an issue for the validation of results in this study. In contrast the data on the annual energy consumption and subsequent data on the individual network associated CHP units are available to a great extent. Hence based on these available annual energy consumption data, DIW estimates the demand profiles on DH level of aggregation. The methodology developed and followed by DIW to generate the profiles simplifies the approach of the BDEW methodology replacing the sigmoid function relationship between the temperature and heat demand with a piecewise linearization approach as this reduces the number of parameters and improves calculation performance with limited loss of accuracy.

The methodology followed by the DIW study defines a sectoral sub-division of the heating demand into the residential and the industrial sector. A large portion of the residential demand is from space heating which is largely temperature dependent contrary to the industrial demand. The methodology follows the approach Felten et al., 2017 which focuses on understanding the interdependencies between the electricity and the heating market and respective market drivers. The methodology is developed primarily to optimally accommodate CHP into the power systems since a combined generation of heat and electricity from CHP plants have been proven to be one of the most efficient way of power generation and supply in the current market.

Regarding the inputs used by this methodology, the following have been highlighted as crucial :

- The daily mean temperature of the location
- CHP plants of a district heating network and its specifications
- Full load hours of heating networks calculated based on the annual heating demand and the installed capacities of the CHP

Felten et al., 2017 uses *Equation 2-6* for determining the demand at any instance ***t***:

$$Q(t) = Qo + \frac{Qmax - Qo}{TR - Tmin} \; max\ (0, TR - T(t)) \qquad \textit{Equation 2-6}$$

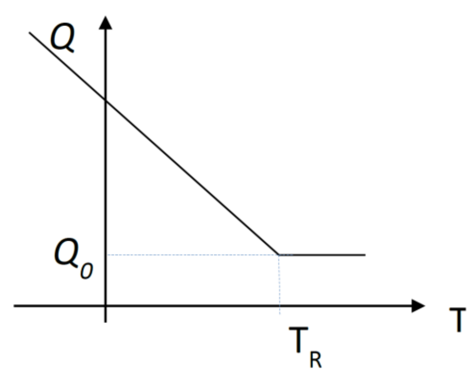
where,

Q_o is the base demand beyond a certain base temperature

Q_{max} is the heat demand corresponding to the minimum temperature

T_R is the base or threshold temperature

T_{min} is the minimum ambient temperature



Source: DIW Berlin, 2017

Figure 2-3: Linearized Demand-Temperature Relationship

The assumptions made in this approach indicate constant demand on exceeding certain predefined temperature (T_R), which holds considering the heating demand would be equal to a base demand/a constant value on exceeding a certain threshold base temperature. But similar to the explanation in BDEW methodology (Oemof Developing Group, 2016), the profiles tend to have certain baseload at all temperature intervals. But considering all DIW profiles are based on DH aggregation, the presence of these baseload is acceptable. Felten et al., 2017 determined base heat demand and the demand associated with the minimum ambient temperature for each heating network. For this, the annual demand and the annual peak demand of the network were used with the integral of the heating degree days.

One major assumption made DIW Berlin, 2017 study is the synonymous use of heat production as the heat demand disregarding the network losses. The results describe the primary heat energy consumption of the aggregation level and not necessarily the final heat energy consumption, which is the actual energy consumption of the residents. This would indicate that the demand seen in the results to be higher than the expected demand. Çomakli et al., 2004 calculated heat transmission losses of about 16% of the total heating system exergy. But the losses are also affected by the ambient temperature, which in turn affects the supply temperature and is largely region and system-specific (Terehovics et al., 2017). However, since acquiring and analyzing transmission loss data of all the networks is unrealistic and out of the scope of this study, it has been overlooked during analysis. Also, an educated guess is made that the increased demand in the primary energy consumption is directly proportional to the final energy consumption and would only affect the magnitude of the curve and not the shape itself, which is the actual comparison parameter of this study.

The DIW database was developed in close collaboration with the German association of district heating operators which account for a total of 260 members. DIW Berlin, 2017 mentions the already large and increasing number of networks to be a reason for the lack of detailed data availability per network. Also, the varying sizes of the network make generalizing the available data impossible. The spatial distribution and capacities of each of these networks are not available openly, however, due to the collaboration DIW Berlin, 2017 was able to access closed source data, and were able to identify the ten largest heating network in Germany based on the maximum heat production of installed CHP units in the network, maximum electricity production by the CHP in the network and annual heat production of the heating network. All data are based on 2015, so adjustments had to be made for comparison for this study. For the comparison in the thesis, amongst all identified databases, this database was closest to the real-time measured values.

The result obtained from the DIW study is the time series of 10 of the largest district heating networks in Germany. However, DIW Berlin, 2017 does not provide a spatial distribution of the networks openly. Therefore to identify the location and corresponding census cells under each of the DH networks, the database developed by Fleiter et al., 2020 was used.

Fleiter et al., 2020 work on the identification of industrial areas nearby high heat demand densities, which meet the criteria for development potential as a district heating network making use of the flue gases or excess heat released from these industries. Though the categorization of the district heating areas in the study is based on assumptions promoting the use of industrial flue gases, the outputs of district heating networks provide a proven estimation of the potential district heating areas in Germany. All neighboring census cells with heating demand over 500 GJ are clustered together as a single heating district. The study ideally defines such high-demand areas primarily to be urbanized locations where district heating is expected to be feasible. Without going into further details of the methodology available in Fleiter et al., 2020, these district heating areas were used as a source for spatial linkage between the DIW profiles and the census cells. The spatial distribution database of the potential network from Fleiter et al., 2020 was acquired from sEEnergies Open Data, 2020.

2.4.3 Estimation based on the heating degree days

A common and simplistic approach for estimating a heat demand profile is through the application of the concept of Heating Degree Days (HDD). Heating Degree days is a technical index taking based on the consideration of the ambient temperature and the average room temperature used to describe the heating requirements of a building (Kuru & Calis, 2019). HDD gives a measure of the severity of winter in terms of the outdoor dry bulb temperature and a measure of sensible heating required for a given location (Giannakopoulos & Psiloglou, 2006). Heating Degree Days are the primary basis of estimation of heating demands, especially for space heating. In the case of DHW, the effect of HDD is minimal as the demands are observed to be consistent throughout the year. As mentioned by BizEE Energy Lens, 2021, the most crucial and also the most difficult aspect of determining the heating degree days is the base temperature also referred to in some literature as the threshold temperature. Base temperature (threshold temperature) is the temperature level considered below which the HDD is calculated. It is the difference between the typical building temperature and the average internal heat gain (BizEE Energy Lens, 2021). However, this value is dependent on many factors and may vary from building to building. Therefore, considering a set value for a large region, for an entire climate zone as done in this study is already a drawback. However, since generating specific HDD profiles is not the primary scope of the study, this approach is taken to observe and analyze the behavior of the generated IDP curves with the HDD of specific zones. For the analysis, the base temperature has been assumed following the findings of Kozarcanin et al., 2019. The respective HDD profiles of the census cells were generated to observe the behavior of the profiles against it, which is also seen as a reliable measure of validation.

2.5 Other relevant aspects for heat demand estimation

2.5.1 Building classification and Characteristics

The housing stock for Germany has been divided into 12 use classes and 6 age classes. Though advanced models regarding the availability and estimation of the occupancy type and precise determination of the occupancy status of households are available, the LPG model is based on the German Household stock categorization, which limits the input data requirement compared to other models and methodology followed in Aragon et al., 2019. The construction type of a house is seen to have a great impact on the heating demand which is truer in the case of SH demand. The thermal mass of the building defines the additional temperature needed to measure the desired temperature (Hellwig, 2003). Hence correct allocation of the building classes becomes important. The load profile generator makes use of the classification based on the categorization done by the Institut Wohnen and Umwelt (IWU, 2015). Parameters such as the building geometry, overall heat transfer coefficient, adjacent buildings, heating system, modernization actions, and residential behavior are considered for this categorization. Based on the building house type and the commonly used CHP type for such households, the IWU provides an approximate average annual heating requirement of such building categories. Details of the household class and their characteristic properties are presented in Annex B.

2.5.2 Data Aggregation

The IDP methodology aims at the development of heat demand profiles on a high spatial and temporal resolution. The final database comprises 8760 hrs. profile for over 35 million census cells. This would result in an enormous data volume, limiting computation and problems associated with storage. Hence to overcome this issue eGon proposes the storage on aggregation levels defined as potential district heating areas for future forecasted scenarios. The project has determined Prospective Supply Districts which are accumulated census population cells that constitute an aggregation of the nearby cells with heating demand exceeding 100GJ. To restrict the size of the resulting district heating areas the cells, all areas exceeding a predefined threshold are split as per the borders of the nuts3 level division. The creation of the district heating areas is not a direct output of this study and details on its creation can be found in (openego, 2021a). All in all, the entire census population cells across Germany are categorized into 3785 district heating areas. However, the district heating aggregation is only observed to account for 12-15% of the total demand. Hence for the remaining census population cells, the aggregation was done based on the medium voltage (MV) grid level, which is also used in the eGon project for electrical grids extracted from open street map data. Details on the categorization of the high and medium voltage grid assignment are available in Openego, 2021. The aggregation assumes that these non-district heating demands will be supplied with individual CHP. Nevertheless, though the aggregation provides a measure for reduced data volume, Fischer et al., 2016 suggest the possibility of resulting unwanted effects such as peak summation and reduction in variability. Hence these aspects are kept in mind when evaluating the final resultant profiles.

2.6 Data sources used in the study

A large variety of data was needed for the implementation of the IDP methodology. This section provides an overview of the data used along with its sources

Table 2-2: Data Sources

Data	Description	Source
TRY Climate Zones	Spatial distribution of the country into 15 Climate Zones based on their geographic and altitude similarity	(Federal Office for Building and Regional Planning (BBR), 2014)
Temperature Profiles	Hourly annual temperature profiles for each TRY climate station representative of the zone	(ECMFW, 2020)
Daily Demand Factor (h-factor)	Scaling factor defined by the relationship between demand and temperature. (Sigmoid curve)	Generated by the author based on (Oemof Developing Group, 2016)
Annual Heat demand	Residential and CTS heat demand per census cell adjusted for forecast scenarios. Based on 2015 values.	(SEEnergies, 2020), (openego, 2021b)
Household stock	Number of SFH and MFH per census cell.	(DESTATIS Statistisches Bundesamt, 2014) rearranged by author
House Type	House Classification based on age of the building linking it to the materials used and in turn the thermal mass of the building	(IWU, 2015)
Base Profiles	Output of the LPG for the development of the IDP pool. 4 cities representing the demand profile in all of Germany	LPG outputs as per predefined input conditions by the author
Temperature Interval	Categorization of temperature classes per average daily temperature values	(Oemof Developing Group, 2016)
CTS Profiles	NUTS3 CTS profile disaggregated to individual census cells	(Gotzens et al., 2020)
District heating network	Potential District heating network clustering defined for future forecast scenarios	(openego, 2021a)

3 Methodology

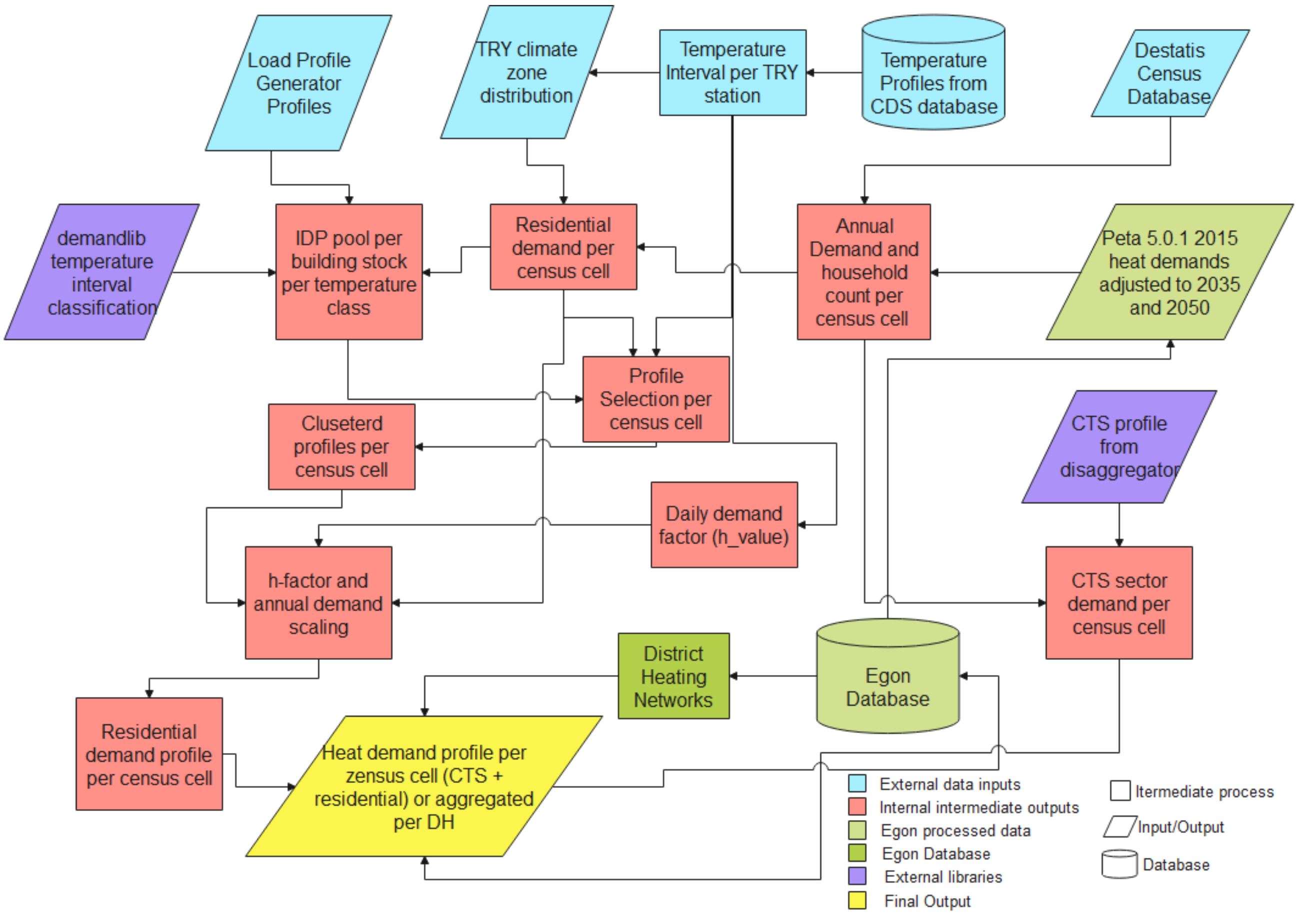
3.1 Developing the IDP methodology

The following methodology was undertaken for the development of highly variable high spatial and temporal resolution heat demand profiles. The methodology can generate profiles for the eGon's forecasted scenarios of 2035 and 2050 described in openego, 2021b. However, profiles generated are based on temperature data of 2011. The 2011 values were used to maintain consistency with other data sources used in the study including the household and CTS demand categorization which are based on the European Census 2011. Also, 2011 temperature showed a good representation and a close correlation to the average historical values and was observed to be appropriate for future projections. In addition, the 2011 data gives average wind speed values and higher solar irradiation which closely represents the future scenarios of 2035 and 2050 (Maruf, 2021). Hence, with this assumption, all further temperature-based calculations are based on these values. With the 2011 data and the census prognosis heat, energy demand values for the two scenario years have been forecasted.

Figure 3-1 shows a descriptive overview of the developed methodology with its detailed description provided in this chapter.

3.1.1 Temperature Profile Allocation

Estimation of both heat and gas profiles has a heavy dependence on weather data (Kozarcanin et al., 2019). This is more true in the case of space heating which is a primary component of heating demand in the residential sector, as changes in the ambient temperature largely fluctuate the heat demands (DIW Berlin, 2017). The selection of correct temperature profiles which give an actual representation of the temperature pattern of any area/region and a correct base temperature becomes important while estimating the heat demand (BizEE Energy Lens, 2021) In other studies (Oemof Developing Group, 2016),(DIW Berlin, 2017) where demand profiles have been generated on different aggregation scales, the use of average national or regional temperature data have been observed. However, in the case of this study, taking a single national-level data would overlook the primary aim and a critical aspect of this study, of obtaining a high spatial resolution, as the same temperature profile would mean a repetitive curve shape for all census cells resulting in low variability amongst the different curves. In an ideal case, census cell-specific temperature data would generate discrete profiles unique to this cell. But this would be impractical firstly considering the large volume of data to be handled and also lack of measurement stations per census cell.



Source: Author

Figure 3-1: IDP Methodology flow diagram

Though profiles with an hourly resolution per km^2 are available from (Krähenmann et al., 2018), these are statistically constructed and real-time historical values are preferred for this study. Also, if in case such measured values were available, the difference in temperature between corresponding or even a cluster of cells in a region would be negligible. Hence as an optimal tradeoff between identifying data on temperature profiles per census cell, the expected increase in the computation time and the loss of accuracy in the final output, the methodology of Test Reference Year (TRY) Climate zones (Federal Office for Building and Regional Planning (BBR), 2014) was identified and used in this study.

Try climate zones divides Germany into 15 clusters based on climatic conditions and geographic similarity. This division was principally done to generate TRY data sets which comprise of details on selected meteorological parameters on an hourly resolution for each of these zones from the year 1961 to 1990 put together in real weather segments primarily focusing on technical climatology but also designed for a wider range of users (Krähenmann et al., 2018). The new updated version of the TRY datasets which consists of future projections from 2021-2050 was developed by the German Weather Service (DWD) and is mainly applicable in the field of heating, cooling, and air conditioning. The categorization is based on the measurement data from 114 climate stations with homogenous data series. Federal Office for Building and Regional Planning (BBR), 2014 mentions variations in temperature values for every 100 m vertical deviation from the height of the station. Hence elevation was an important aspect considered in the paper for the clustering of the zones. But, overlooking other meteorological details provided in the TRY datasets, only the regional division provided by this methodology was used. As TRY climate division validates the temperature similarity in each of the regions, the use of these for allocation of temperature profile was justified.

For this study, each of these TRY climate zones is represented by a TRY temperature station. The hourly resolution temperature data for each station is acquired from ECMFW, 2020. The measured temperature values of each of these temperature stations are assumed to be representative of the temperature profiles of the entire zone. The regional division as per the TRY climate zone is presented in Figure 3-2. As digital zonal division maps could not be acquired, this study was initiated with the digitalization of the climate zones. This would allow accurate distribution of population census cells into each of these climate zones and thus would be crucial for the development of the profiles in the latter stages of the study.

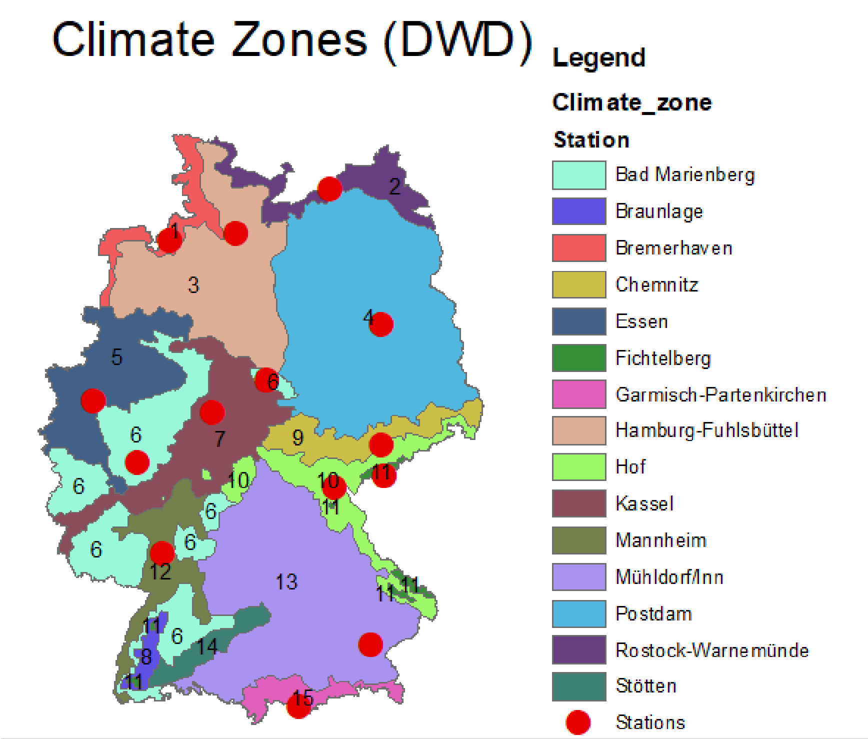


Figure 3-2: Division of TRY Climate Zones

Source: Defined by (Federal Office for Building and Regional Planning (BBR), 2014), digitalized by the author

ECMFW, 2020 provides 2m temperature data for grid resolution of 0.25 deg on an hourly resolution from 1979 to 5 days before the date of access. Profiles for each of the TRY climate zone stations were extracted for the year 2011 to co-align with other data sources used in eGon project.

3.1.2 Generation of the Base profiles

The first step towards the implementation of the IDP methodology was the generation and study of profiles generated from controlled LPG runs, referred to as the base profiles. The base profiles are a critical aspect of the IDP methodology as they provide the 24hrs. profiles with the desired degree of variability. Drauz, 2016 implemented the model with the generation of profiles for specific types of households in Kassel for validating the output of the LPG. The same approach under controlled input parameters was conducted for 4 selected cities: Lubeck, Kassel, Wurzburg, and Schleswig. The selection of the cities was done with an educated assumption that the temperature profiles of these cities would be able to capture the general temperature trends throughout the country.

The input provided for these LPG runs was controlled on two aspects. The house type and the household type. LPG is capable of taking inputs on all household classes up to class J, classified by (IWU, 2015) detailed in section 2.5. But only classes H, I, and J were considered for this study under the assumption that all older households before the considered three classes would be non-existent in the future forecast scenario of 2035 and 2050. The three classes are the most recent that can be used by the LPG and comprises houses constructed post-1984. Zeyen et al., 2021 mention that 75-90% of these building stocks are going to exist till 2050 thus justifying the assumptions. Though building classes K and L (the most recent building classes) have also been recently defined by IWU, these are recent additions to the building typology and have not been included in the LPG. Making alterations to the load profile generator is firstly out of the scope of this study and secondly not possible considering the tool is closed sourced. Hence continuing further analysis with the above mentioned three. With regards to the household type, 5 different categories have been chosen, namely P1a: parents with one child (family size: 3), P2a: parents with two children (4), PRa: retired couple (2), SRa: single retiree (1) and SOa: single (1) which are the most common family sizes in Germany (eurostat, 2021).

As a step to confirm the usability of these assumed controlled inputs, unaggregated sample individual SH and DHW profiles were generated for each of the 4 cities. The result was analyzed to verify if the individual SH and DHW profiles return agreeable results and if their aggregation provides desirable heating demand behavior. The analysis results in section 4.1 confirm this. Hence the combination of these 3-house types and 5 household types, were used to generate the base profiles. In addition, temperature and radiation data were provided for each run of the LPG based on the city under consideration.

Finally, the base profiles were generated for each city with a random combination of the 3X5 parameters. A total of 308 SFH and 930 MFH high-resolution base profiles (8760hrs.) were generated with the LPG. These were later used for the generation of the IDP pool. The number of profiles created is random and could be increased, which would in turn increase the size of the IDP pool, thus inducing further higher variability in the final profiles.

The generation of the final census level demand profiles entirely from LPG model runs was also considered during development. However, some issues were encountered which made the approach inefficient. These included the need for large input data on census level and the immensely long run time of the model. Acquiring and processing such large data at the census level was problematic (Fischer et al., 2016) and the run time could not be reduced considering the model was primarily designed for micro-level profile generation. Therefore, the development of the IDP methodology was further continued.

3.1.3 Temperature Interval allocations

The definition and the categorization of the numerous temperature classes used for this methodology are based on the approach undertaken by the demannlib library. Demanlib is a library generated as a part of OEMOF group but also with the capability of working as a standalone library. The library can generate hourly resolution curves for both the heating and electricity sectors. The library can generate heat demand profiles for both the residential sector based on house type SFH and MFH and also the CTS sector categorized into 14 different sub-types. However, considering the low anticipated variability of the BDEW profiles (as explained in section 2.2.2), the direct use of demandlib is not applicable for the generation of results desired by this study. Nevertheless, it provides a valuable means of reference. A major takeaway from the demandlib library for application in the IDP methodology was the division of the temperature classes as defined in Table 3-1.

Table 3-1: Classification of Temperature Intervals

Temperature Class	**Temperature Range**	**Temperature Class**	**Temperature Range**
Class-1	-20°C to -14°C	Class-6	6°C to 10°C
Class-2	-13°C to -10°C	Class-7	11°C to 15°C
Class-3	-9°C to -5°C	Class-8	16°C to 21°C
Class-4	-4°C to 0°C	Class-9	22°C to 25°C
Class-5	1°C to 5°C	Class-10	26°C and above

Source: (Oemof Developing Group, 2016)

For the temperature profiles over the 4 base profile cities, each day was categorized into their respective temperature class based on the ambient temperature T_a. Also, for every temperature station, each day was assigned a temperature class based on T_a. These classes were then used to select the correct combination of the profiles from the IDP pool, thus assigning each day with a 24hr. profile. This permitted a degree of control over the randomization as the temperature profile for a summer day would be selected only from the pool of profiles corresponding to that temperature interval. For the assignment of the daily ambient temperature geometric mean of the temperature of the previous four days for any particular hour (as defined by *Equation 2-1*) was used as this better represents the thermal mass of the building (Ruhnau et al., 2019).

3.1.4 Intra Day Profile Pool

Generation of the IDP Pool

Intra Day Profile pool is a collection of 24-hour profiles with each representing their respective temperature classes and household stock type (SFH and MFH). With the assignment of the daily temperature classes to the 4 base profile cities based on respective temperature profiles of 2011, every 24hrs. profiles for each base profile were assigned to their respective temperature class. Now based on these classes, each class of profiles was clustered into their respective class database, also keeping in mind the household stock division. However, once the profile is split into the so-called pools, the 24 hr. profile cannot be tracked back to the source city.

For example, considering Luebeck, the first day of the year for Luebeck as per 2011 profile was observed to be Class 5. Combining both SFH and MFH building stock types, 629 base profiles had been previously generated for Lubeck. Hence 629 different 24-hour profiles or IDPs were assigned to class 5 of the IDP pool from Luebeck day one. In total, Lubeck has 76 days assigned to Class 5. Hence, Lubeck contributes 47,804 profiles to the class 5 pool. The accumulated individual 24-hour profile is then normalized to its sum. With a similar approach, splitting all the hourly resolution profiles into IDPs, a total of 451,870 24-hours profiles were generated categorized as per their respective temperature class and house stock. To increase further variability, additional profiles from the load profile generator could be generated, however, the impact on the final profiles would not be significant. Hence the total profile output was limited to 1238 profiles. The final pool consisted of IDPs distributed amongst 9 different classes and 2 household stocks. The pool lacks IDPs for temperature classes 1 since none of the 5 cities had these temperature classes. But since temperatures below -14°C are seldom expected in any grid cell, the generation of IDP pool for this class was concluded to be unnecessary.

Random selection and incorporation of the IDP pools

The primary attribute of the developed IDP methodology was the possibility of generating varying heat demand profiles for a high spatial resolution of a census cell. This variability and uniqueness to each profile were implemented through the randomness introduced during the aggregation of the profiles. For every census cell, the total number of households was acquired from DESTATIS Statistisches Bundesamt, 2014, and re-categorized into the two household stock types. Then for each census cell, based on the day's temperature class, random profiles were selected. The number of selected profiles per day was dependent on the count of each household stock in the census cell. These daily profiles were then aggregated to generate a total 24-hour profile. A similar approach is continued for each day of the year, thus generating a randomized 8760 hrs. profile. For example, for a census cell in Bremerhaven TRY climate zone with 3 SFHs and 5 MFHs, for the first day when the temperature class is 5, a respective number of profiles were selected from each household stock from the class 5 IDPs of the pool. These 8 profiles are then aggregated to obtain a single 24-hour profile for the day. The same procedure was followed for all 365 days, and IDPs were assigned to form a series of 8760-hour profiles. The final 8760 values representing every hour were then normalized to their sum.

3.1.5 Daily Demand Factor

Following the selection of random profiles, the next step was to integrate the effect of temperature into the profiles. For this, the concept of the daily demand factor acquired from the demandlib library was used. The factor defines the relationship between the ambient temperature and the demand. This was necessary to give the selected curves a seasonal trend. The daily demand values are represented by the sigmoid curve as defined in *Equation 2-2*. Each parameter of the curve A, B, C, and D are mathematically calculated based on the house type under consideration. The demandlib library defines a specific parameter for every combination of household stock and house type. But only the latest combination has been used in the methodology keeping in mind the need for future forecast scenarios and limiting the computational time and working data volume. Nevertheless, the difference in the demand factor values due to these varying parameter combinations was not found to be significantly different, hence justifying the use of only the latest combination. Hence, for all household stock and house types considered, the latest of the parameters defined by demandlib (A=3.046, B=-37.183, C=5.672, D=0.116) were used. The sigmoid factors for other building typologies are available in Annex A. With this, the daily demand factor for all TRY climate stations (h-value) was calculated. Furthermore, the demandlib library also makes use of an hourly scaling factor which scales up each hour of the day based on the household stock and the temperature class of that particular day. But this has not been implemented here, as unlike the BDEW there already exists a large amount of variability due to the randomization. Thus, every hour of the day is unique and does not need to be further scaled. The final 8760 demand-factor was then normalized to its total to give individual census cell its respective demand pattern.

Finally, the demand patterns of each cell are scaled up to their respective annual demand values thus providing a final residential heating profile per census cell.

3.2 Generation of CTS Profiles

The commercial sector is another significant contributor to the heating demand. A slightly different approach was undertaken for the generation of the CTS profiles due to the unavailability of detailed data for the sector (unlike in the case of residential). The CTS profiles were derived with the combination of two different inputs. For the profile shape, the inputs were used from the disaggregator tool (Gotzens et al., 2020), which is a python-based library tailor-made for developing energy demands for the German heating sector. The tool, an output for the DemandRegio research project, is aimed towards the development of high spatial and temporal resolution electricity and demand profile generation. Considering the common practice of using gas profiles as a proxy for the heat demand (Hellwig, 2003, p. 25), the disaggregator-generated gas profiles were used as the curve pattern for heat profiles. Though the tool permits the selection of any desired year as the input for the generation of the profile, to maintain consistency with other data sources 2011 profiles were used. However, the tool is only able to give an output in NUTS-3 district level spatial resolution making use of the profiles generated with the standard load profiles.

To meet the census cell resolution requirement of the project, each census cells within the respective district were assigned the profile for the district. Hence all cells within a district have identical curves. Though this approach prevented a higher degree of variability in the curves between census cells within the same district as compared to the residential sector, this was seen as the most convenient way to develop realistic high spatial demand for the CTS sector. Once respective profiles were assigned for the cells, these were then scaled up based on the total CTS demands of the cell. The CTS demand was acquired from the census demand per cell aggregating all service sector demands of that cell.

3.3 Aggregation of the demand profiles

The methodology discussed in the previous sections was implemented to develop hourly resolution heat demand profiles on a high spatial resolution per census cell. But storage and processing of this large volume of data are difficult. Also, validation is an important phase of the study. And as mentioned, high spatial resolution demand forecasts are not available and therefore the aggregation of the data to the level on which a proven study is based becomes important. Hence for this, the aggregation of the census profiles was done on two levels. Firstly on a district heating level as defined under the egon project which clusters neighboring cells to form district heating regions. For those census cells which are not categorized under DH network, the aggregation is done under the medium voltage substations defined by OSM as detailed in section 2.5.2.

3.4 Validation of IDP methodology

An important aspect following the generation of the heat demand profiles was the validation of the generated database. The absence of real-time measured data slightly complicated the process. However, two different validated databases were identified, described in section 2.4. These were used as reference profiles for comparison. These include:

- OPSD when2heat database
- DIW Berlin district heating network

Both OPSD and DIW databases are both developed with a top-down approach. The OPSD provided the possibility of comparison on a national aggregation level and the DIW on a district heating/regional level. However, considering both these profiles are on a certain aggregation level, for individual census cell comparison, direct BDEW methodology was implemented by generating BDEW profiles for a given cell using demandlib. Irrespective of the aggregation level the comparison was done on either of the two scales: Daily 24 hours scale and Annual 8760 hours scale. The daily scale provides a comparison of hourly variability and aspects such as the behavior of peaks in the profiles. The annual scale analyzes possible trends, seasonality, and cyclicity. In summary, the following quantitative aspects were evaluated for testing the comparative similarity:

- Nature of the profiles and the effect of randomness on it
- The pattern in terms of seasonality and cyclicity shown by the profiles
- Structural similarity to the reference profiles
- Behavior in terms of the demand and temperature

Ruhnau et al., 2019 suggest demand time-series validation in terms of accuracy, comprehensiveness, and applicability. Similarity tests regarding the nature, pattern, and structure provide a means of validating the accuracy aspects. Also, the data generated are expected to give comprehensive results considering that the profiles are generated for high spatial resolution covering a detailed heat demand on the geographical locations. Also, the high variability in each of the profiles gives a more realistic coverage of the overall heating demand sector in Germany. In terms of applicability, though the initial proposed plan was the use of the database for grid optimization with eTrago modelling tool, this could not be implemented. Hence this leaves an opportunity for future studies to test the applicability of the database to further verify the generated outputs.

In addition to the comparison with the reference profiles, inter profile comparison is also crucial to understand the behavior of the curves. This mostly helps in validating the effect of randomization implemented in the study. The following approach was taken to ensure its validation:

- *The behavior of the individual curves:* This category covered the quantitative evaluation of the generated profiles which was able to provide details on the behavior of the curve. Referring to similar trends observed in other studies and making an educated assumption on the expected curve patterns, the obtained patterns, and the impact of randomness on them was validated. For this, the behavior of the curve in terms of the HDD was analyzed. Furthermore, the test for autocorrelative properties of each of the curves was conducted to observe the effect of randomness on the cyclicity and seasonality of the curves.
- *The effect of randomness on the curve*: Possible measures to limit the spread of the randomness in the demand-temperature curve and the impact of this on the final output was investigated. Also, the distribution of the generated profiles was analyzed and justification on the observed patterns was developed. Also, a test of stationarity provided a means to test the effect of randomness on the nature of the curve.

Statistical Tests for Validation

The analysis of time series data can be done based on its two components, systematic component (level, trend, and seasonality) and non-systematic component (randomness or noise). Athiyarath et al., 2020 and Kuru & Calis, 2019 mention RMSE and Pearson’s correlation being the best measure for identifying relation and likeliness between two time-series profiles. Also, the use of Pearson’s correlation coefficient, RMSE, and MAPE for time-series comparison and validation was applied by Schüler et al., 2015, Chen et al., 2017 and Idowu et al., 2015 amongst others. Furthermore, Chramcov, 1982 focused on studying the stationarity and autocorrelative features of the heat time series. Thus, based on the approaches undertaken by published literature, the following statistical methodologies were identified and used as the most suitable measure for the validation of the generated profiles with the reference profiles.

Correlation: For the comparison of the difference between the observed series and reference series, Pearson’s correlation was used to identify the linear relationship between the sample generated curves and the reference curves. More similarity between the curves is indicated by the higher value of the correlation coefficient. Also, to evaluate the effect of the randomness, the cross-correlation between the two was evaluated. Cross-correlation is the measure of similarity between two time-series calculated as a function of displacement of one of the sample series to the reference series. Cross-correlation was identified as a measure of the degree of similarity between the two. Even though a high Pearson’s correlation would show low cross-correlation (lower percentage of the difference between two profiles) between the two time series, this was only expected on a daily resolution and not on hourly resolution. The difference in the cross-correlation in two resolutions was used to understand and quantify the impact of randomness on the final model. Also, the summation of these errors (the difference in the two-time series) is a means of quantification to the error and used in the comparison of multiple census cells. To validate the Pearson's correlation coefficient's statistical significance, a statistical T-test was conducted.

Structural Similarity: As a further means of similarity test on the two demand profiles and validation of the correlation result, a Chow Break Test was conducted to test the structural similarity of the two profiles. This provides additional statistical significance to the obtained results.

RMSE: Another important measure for the comparison as determined to be the root mean square error between the two-time series and gives an accurate measure of the profiles. The RMSE provides a more concurrent comparison of the errors or differences in the two series and prevents the canceling out of the positive and negative errors, which could lead to a false representation of the outputs and analysis results. The RMSE of multiple random generated profiles were compared to the available reference profiles. The distribution of the RMSE values was used to evaluate the closeness of the curves with the reference. Also, the distribution of the RMSE values provided a possibility for the evaluation of the effect of randomness on the profiles. In addition, Mean Relative Absolute Error (MRAE) was also evaluated to obtain a better interpretation of the performance of the generated profile in comparison to the benchmark as suggested by Chen et al., 2017 due to RMSE's sensitiveness to outliers.

Nature and seasonality test of the demand profiles: The autocorrelation of the demand curve was conducted to evaluate the relationship or the causation of the demand at any instance as a result of the demand in the preceding instances. With these results, the impact of randomness on instantaneous demand could be identified. It also provided a more descriptive analysis along with the possibility of obtaining a valid justification for the observed patterns. Observation of autocorrelation was also important mainly in high linearly correlated profiles as it provides the possibility of extracting other useful features of the curve.

Test for Similarity: The elastic measure of dynamic time warping is another approach that was undertaken, which has been proven to provide a basis for the comparison of time series. The methodology provides the possibility to determine time-series similarity and the behavior of corresponding points in two-time series. The dissimilarity between two-time series is measured in terms of the overall cost for aligning the two-time series, meaning the greater the misalignment, the higher the cost. The DTW measure provides a simple yet superior measure to provide a quantification to the dissimilarity between the two time-series (Serrà &

Arcos, 2014). The approach was undertaken to compare the IDP profiles with the BDEW and also the generated profiles with the reference profiles on different aggregation levels.

To ensure consistency on all validation approaches in the study the following general aspects were followed:

- With an anticipated loss of accuracy with the use of average temperatures, any comparative analysis involving temperature aggregation levels was not used and only done on a census scale.
- For all census level analyses, the comparison was done on randomly selected sample cells. Only the results which are representative of all random sample cells have been presented in the thesis.

Additional steps for DIW comparison:
DIW Berlin, 2017 provided a database on heat time series for 10 of the largest DH networks in Germany. However, the database lacked the spatial distribution of these networks. Hence for spatially locating these district heating networks, DH distribution was acquired from sEEnergies Open Data, 2020. For instance, for Gelsenkirchen the nearest and the largest identified DH was the district heating network with a total annual demand of 61.287 PJ. Following this, the corresponding census cell under the coverage of this district heating network was identified and their corresponding heat time series were aggregated. The final aggregation was then compared to the DH_DE_Gelsenkirchenfrom DIW. The comparison of the annual total demand was important to confirm that the correct DH network had been identified. However, considering the methodology undertaken by DIW Berlin, 2017 and Fleiter et al., 2020 are different, the identification of the exact DH network would be difficult, hence a 25% of margin of error in total demand while identifying the network was accepted. A similar approach was undertaken for the other profiles. Nevertheless, the Berlin and Hamburg profiles were expected to carry a lot more weightage considering the same level of NUTS 3 and NUTS2 levels. Also, the largest DH networks in these locations were much easier to identify. This was finally followed by the above-mentioned statistical comparison.

4 Findings, Discussions, and Analysis

4.1 Identification of input parameters for the base profile

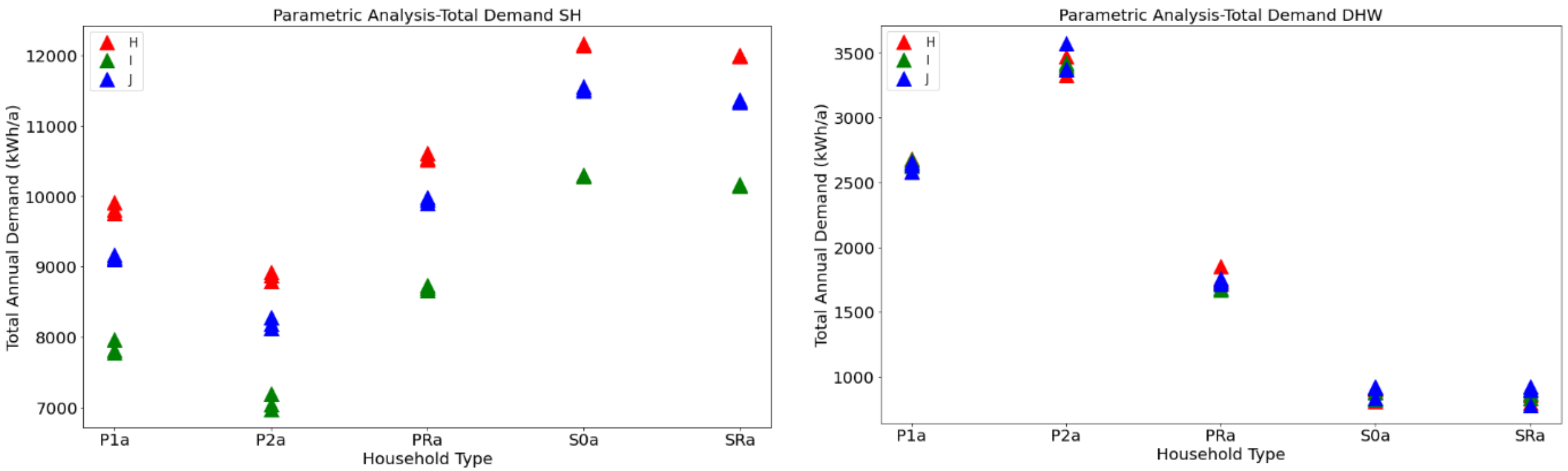
This section covers the understanding of the software implementation of the Load Profile Generator and validates the sufficiency of identified input parameters to correctly generate desired base profiles and overall final demand profiles.

Prior to the development of the final base profiles, individual separate sample SH and DHW profiles were generated with the LPG with the introduction of the two major limitations to the input. Only the latest 3 household types and 5 most common house types were considered as explained in section 3.1.2.

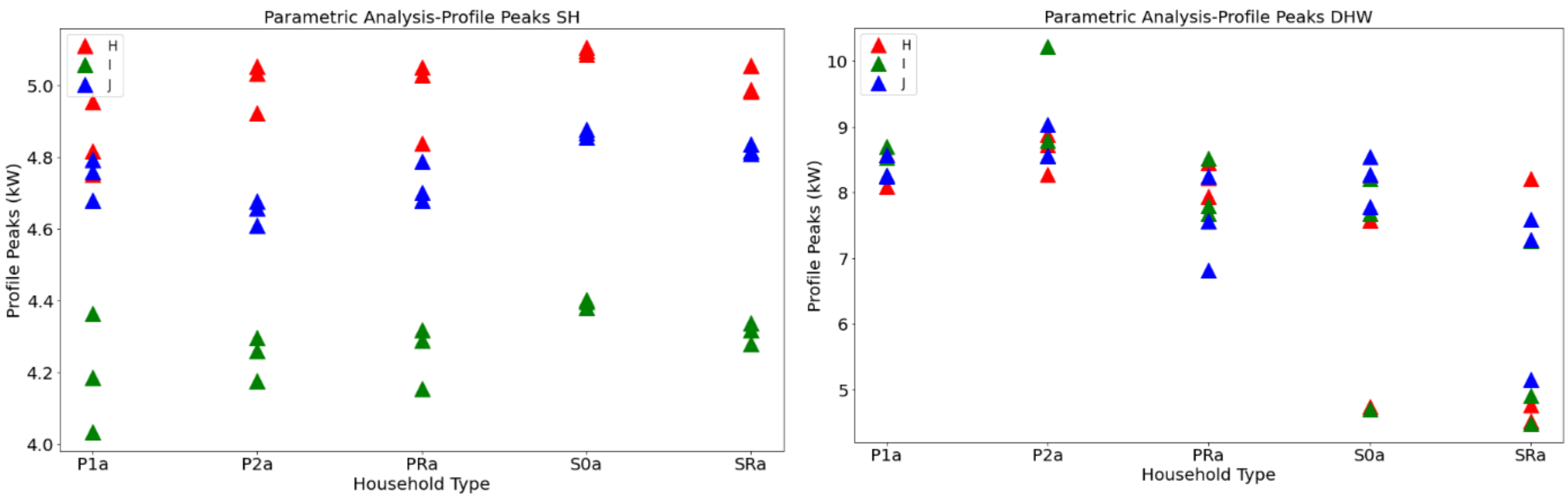
The combination of the 3 Household and 5 House type parameters iterated over 3 runs of the model generated 45 profiles each for SH and DHW. The 10 min temporal resolution of the LPG was adjusted to an hourly resolution. With regards to the household stock (SFH and MFH), this analysis was performed only for SFH, since excluding the randomity the overall curve patterns are expected to be the same for either type with variations only in the demand magnitude. Thus, a desirable result seen in the SFH results could be used to justify similar results for MFH.

The observed outputs have been visualized in Figure 4-1 for SH on the left and DHW on the right. The comparison showed the effect of different household types on the house type and vice versa. Though the limited number of samples were tested with regards to reproducibility, considering the randomness of the bottom-up model the obtained outputs can be generalized for all runs (iterations) with respective identical inputs. The colors of the dot plot represent each of the house types (H, I, J) compared against the household type. For each combination of H and HH types, the 3 points are plotted each representing one specific iteration of the LPG run.

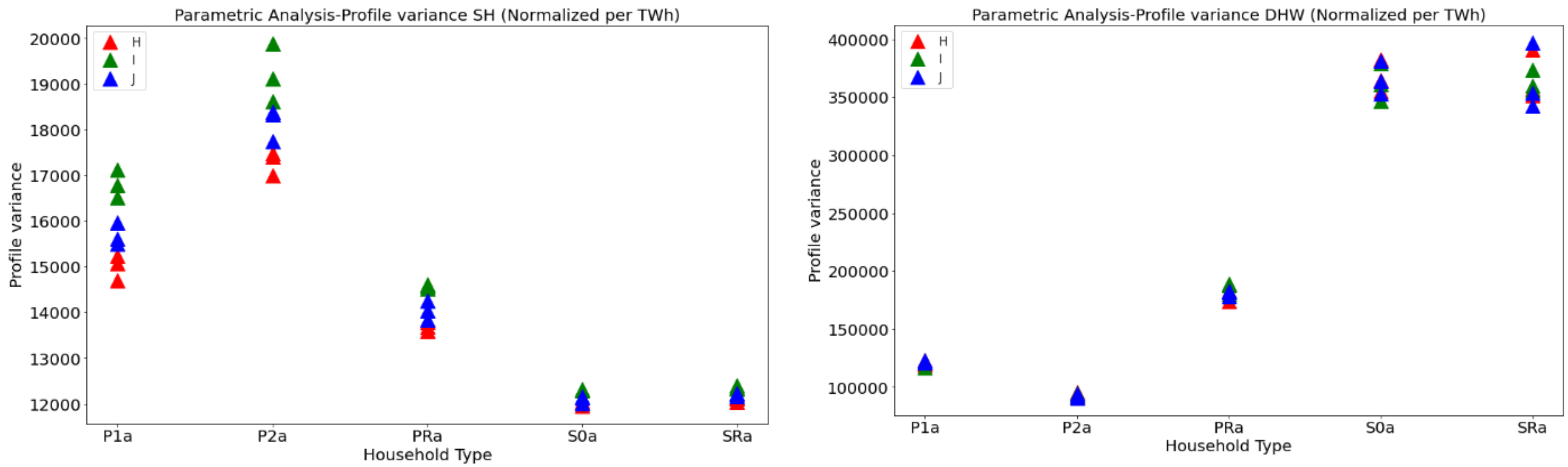
Space Heating Drinking Hot Water



(a) Total Demand



(b) Peaks



(c) Variance

Source: Author

Figure 4-1: Base Profile-Effect of Input
(Left: SH; Right: DHW)

SH base profiles:

At first glance, the variability in the SH profiles was observed to be much less compared to the DHW, which was anticipated. The variability and lower correlation between the sample profiles would mean a lower reproducibility/higher variability of the curve patterns. Considering the wide range of profiles needed, this lower reproducibility can be considered good, however, this made the evaluation of the effect of the parameters quite cumbersome. Considering the primary application of the generated results for grid optimization, aspects such as peaks, variance, and total annual demand which affect the design of grid systems were analyzed.

For SH limited variations were observed in the total annual demand generated from the bottom-up model, for each of the iterations, thus indicating a degree of control over the variability. The total annual demand is seen to be higher for H building types for all cases. The same is the observance in the peaks. Since this is the oldest building class, the thermal inertia in the buildings is the lowest hence these higher values in both the cases, thus providing a logical justification to the observed behavior. However as can be seen in all cases the newer building class, J has higher peaks and total annual demand compared to class I, even though J is a newer building class. As per the justification above on the thermal inertia compared to the building class, this should not have been the case. Buta further detailed look into the IWU, 2015 building class showed a much larger average floor area in class J, thus the higher values.

For the annual load graph, demand is seen to increase as the occupant number goes down. Though at first glance it may look unrealistic, this behavior is quite common for SH demand since its dependence on the occupant status is significantly lower than the DHW demand. This behavior of the profiles has also been explicitly mentioned by (Drauz, 2016). The dependence of SH on ambient temperature and irradiance also plays a part. But the major contributor to the observed result is the internal heat gains of the building. Higher occupant number would on one hand mean capturing of the heat released by the occupant themselves and on the other hand from the larger number of electrical appliances used from the higher number of occupants.

In terms of the SH peak, the variations in each of the iteration are observed to be highly variable than the annual demand. Thus, this demonstrates the ability of the LPG to generate varying peaks even for very similar total demand values, thus the large variability. Such high variability is a desired output of the IDP methodology. Also, for instance, comparing H and J, it was observed that the reduction in peaks from H to J is higher than the annual demand. This is in line with the expected behavior of improvement in building thermal characteristics which have a more significant impact on the demand peaks than on the total demand. For any house type, the effect of the household type on the annual demand is seen to be significant whereas the change in the peaks is almost linear thus indicating the minimal impact of the family size on the peaks.

For the comparison of the variances within the individual profiles, each of the SH profiles was normalized to demand per TWh of annual demand. This was important when comparing the individual profile variance between other profiles. The variances were observed to be directly proportional to the household occupant size. The higher occupant would mean a higher frequency of use of appliances and could be one explanation for the higher fluctuation. In all cases house type I was seen to have the highest variance. The exact cause of this variance could not be verified. Nevertheless, the consistency in this behavior can validate that this is not a random error.

DHW Base Profiles:

For DHW the graphs show that the effect of the house type on the curves is negligible. Irrespective of the parameter type, all 3 characteristics showed the same behavior for every iteration for each household type. Thus, indicating that DHW is significantly affected only by the occupant number and status. The total demands and the peaks are seen to be proportional to the occupant number whereas the variances observed in the profiles are inversely proportional to the occupant number. The relationship with the variance is explainable since the continuous tapping rates decrease with reduced occupants. Also comparing the variances in the DHW values are seen to vary 10 folds in comparison to the SH profiles. This thus validates the higher variability in the DHW profiles compared to the SH, also confirming the major cause of variability seen in both the base and final profiles is dew to DHW.

Overall, it can be concluded that the effect of the parameters on the SH demand is affected by both the house and household types but has a much lesser variability compared to DHW which is only affected by the household type. Combining these individual base profiles gave a final output that incorporates the true nature of heat demand profiles with controlled variability following a seasonal pattern. This was further confirmed with a Pearson's correlation of 0.96 and a MAPE of less than 12% MAPE with the BDEW generated demand profiles on a daily resolution. The visualization and analysis performed in this section help confirm the reliability of the pre-assumed input parameters. Also, in terms of the final profiles, no significant distinguishable differences were observed in the normalized shape of the curves for any of the three house types. Nevertheless, all three were still used for the generation of the actual base profiles.

A tabularized result summary is available in Annex C.

4.2 Intra Day Profiles

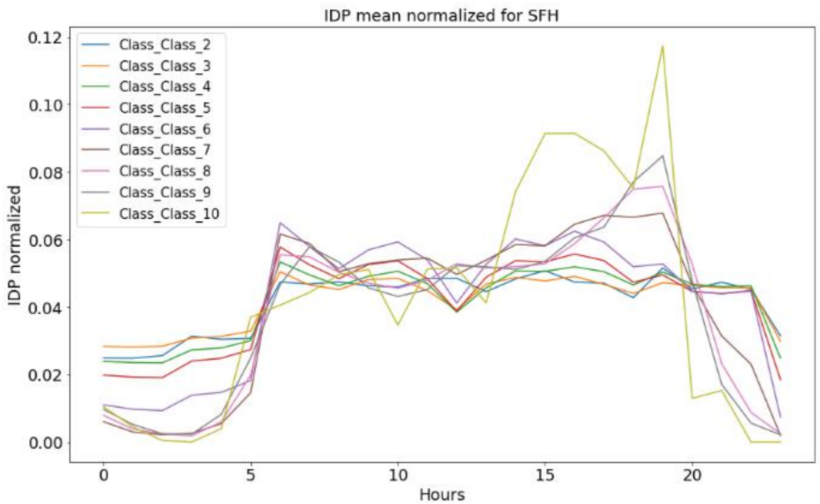
The main feature and an important component of the IDP methodology is the IDP pool. The pool is a collection of the normalized 24hrs. profile categorized based on the household stock and the temperature class. This pool is generated from the base profiles and is the basis for the generation of the final demand profiles.

The section focuses on the evaluation of the behavior of the individual IDPs. These 24hrs. profiles can provide a higher degree of variability than offered by other methodologies. Thus, this is a unique feature offered by the methodology. Figure 4-2 ((a),(b)) shows the statistical medians of individual IDP pool per class per household stock. It gives a visualization of the IDP curve patterns and their behavior with regard to the ambient temperature.

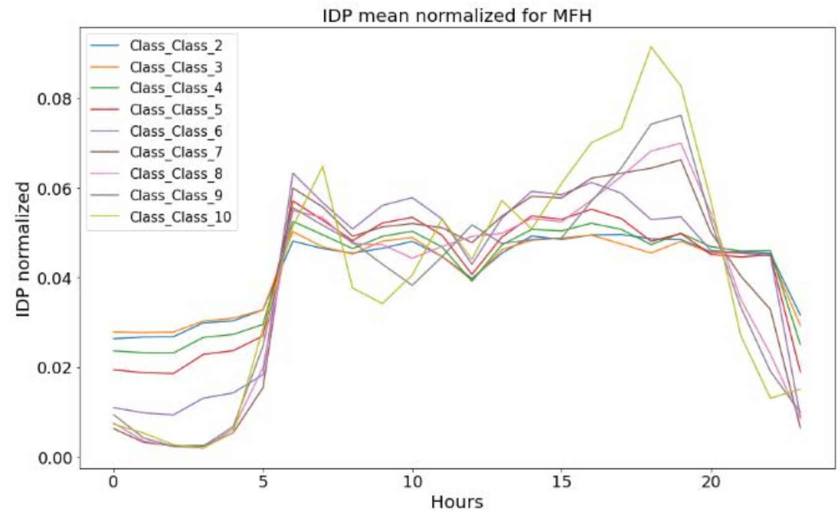
A common tendency of two distinct peaks is seen in all curves irrespective of the temperature class. A low morning peak ranging between 5:00 hrs. to 7:00 hrs., and a high peak in the evening ranging between 17:00 hrs. to 19:00 hrs. Considering these times to be the most likely when the occupants are home and active, the behavior observed is logically realistic. Similar intra-day patterns were also observed in Clegg & Mancarella, 2019.

The high-temperature class curves (representing summer days) were seen to show abrupt peaks which can be acquainted with the DHW demand. Also, the presence of these peaks in hours of the day with the probability of high active occupancy further backs the claim for these profiles being a DHW demand. The reason for the presence of the continuous peaks is due to the representation of the graphs in terms of medians which may not have necessarily captured the essence of the individual curve behavior. For this individual curve must be visualized. For lower temperature classes (representing winter days) a consistent baseload can be observed in the profile with no zero demand hours. This can be acquainted with the SH load and the peaks with DHW. As a result of the baseload, the lower classes curves are much flatter than the higher classes as this reduces the drastic difference in the magnitude of the peak as seen in a DHW profile.

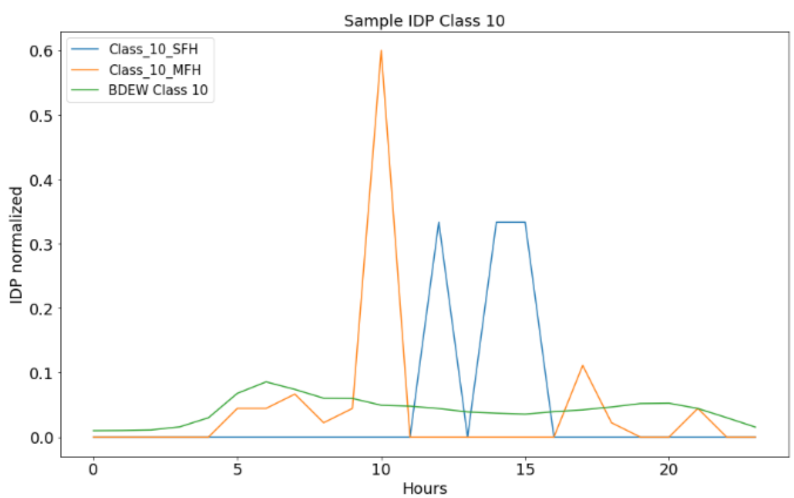
Comparing the household stock (SFH and MFH) the building behavior is quite identical in terms of the shape of the curve. The magnitude is however dependent on the random profiles and the medians do not represent it correctly. For better visualization of the actual curves, 2 random samples of class 2 and class 10 each are presented in Figure 4-2 ((c),(d)). In general, the peaks are seen to be more common in the SFH profiles compared to the MFH profile. This behavior was seen in all random samples and can be generalized to all profiles. In SFH the non-overlapping peaks are higher than those in MFH. Thus, SFH profiles are slightly peakier. However, since all profiles are aggregated, the difference is unnoticeable in the final aggregated profiles. In the figure, the green plot indicates the BDEW based SLP which shows the absence



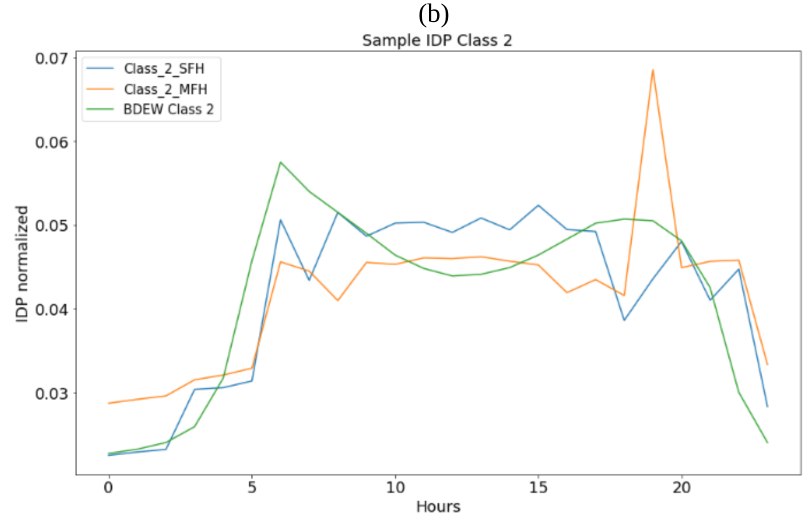
(a)



(b)



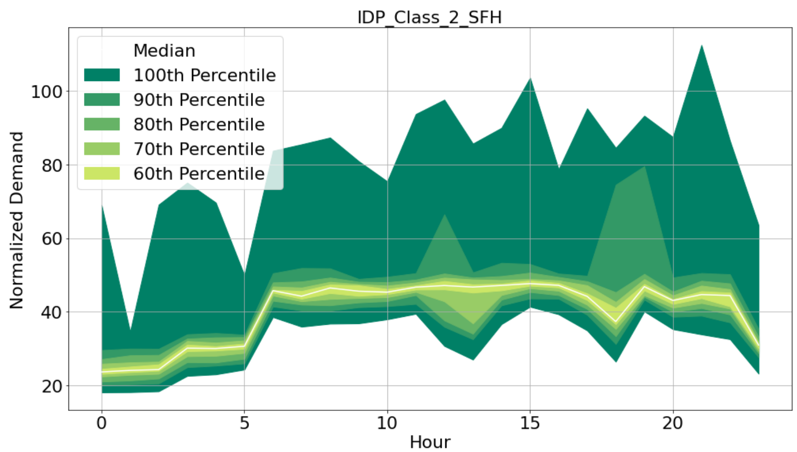
(c)



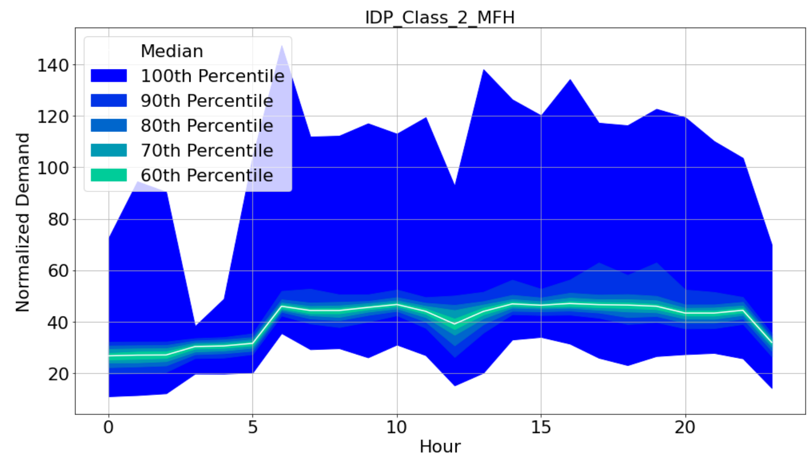
(d)

Source: Author

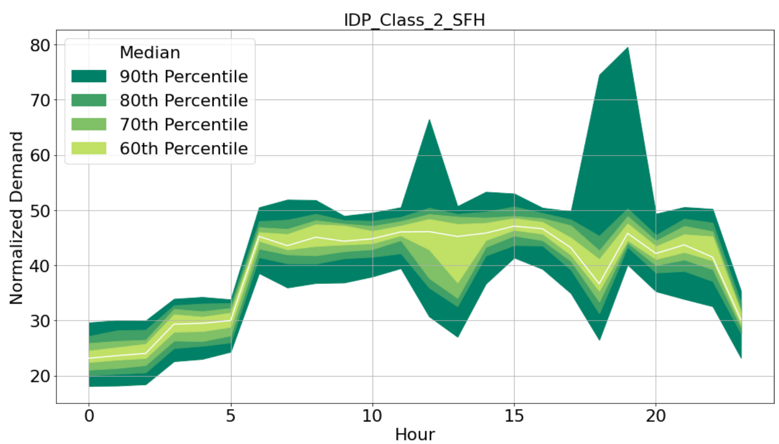
Figure 4-2: IDP Pool
((a)Median-SFH;(b) Median-MFH;(c)Class-10 random sample; (d)Class-2 random sample)



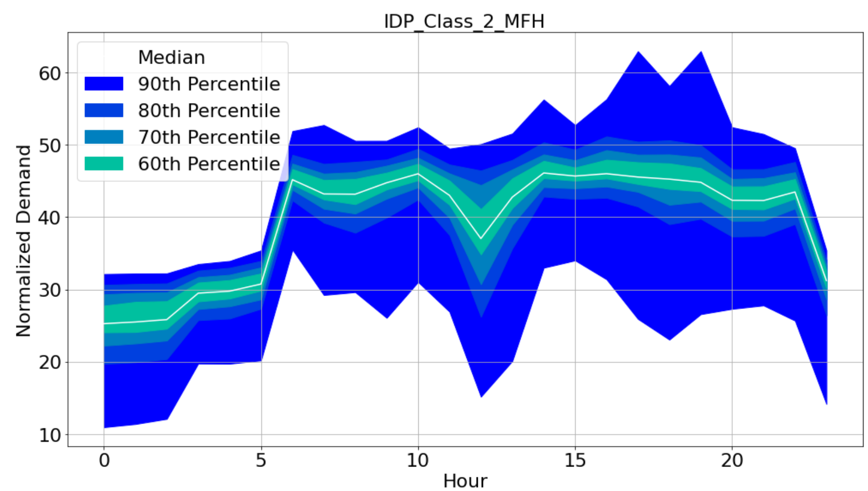
(a)



(b)



(c)



(d)

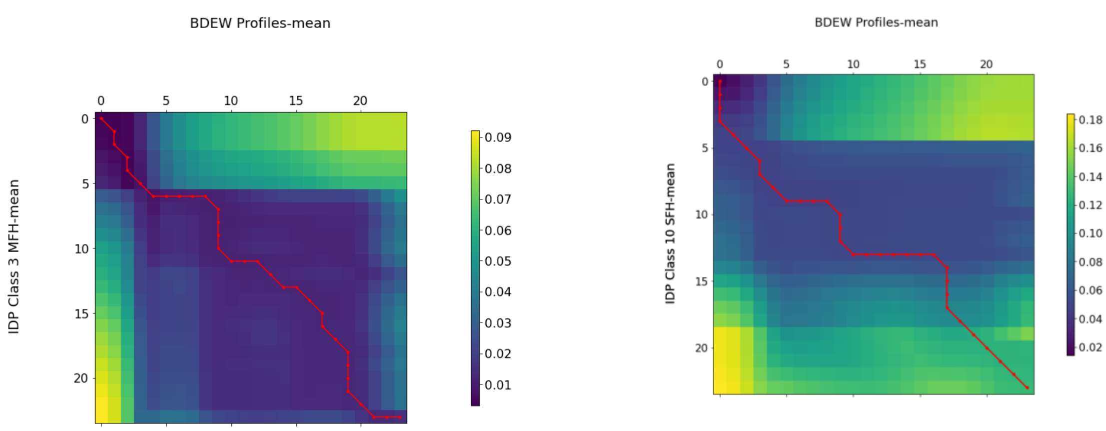
Source: Author

Figure 4-3:Variability in IDP
((a) Class 2-SFH; (b)Class 2-MFH; (c) Class 2-SFH (extremes removed); (d)Class 2-MFH(extremes removed))

of peaks. Correct estimation of heat demand peaks is important especially in an electrically supplied system as it associates with over 30% of the total system costs (Zeyen et al., 2021).

Furthermore, better visualization of individual IDPs can be seen from Figure 4-3 ((a),(b))presented for class2. Every range of color in the graph represents the 10th percentile distribution of the profile hourly values. The profiles here are normalized to MW per TWh of total annual demand. As seen in the figure, even with a degree of randomness induced, 80% of the class 2 IDP follow a close controlled behavior with the next 10% (between 80th to 90th) show slight variations and final 10% of the profiles with extreme variations. On the second row of Figure 4-3 ((c),(d))a more consistent behavior of the profiles can be observed when the upper extreme 10th percentile of the fluctuations are removed. Random assignment of the energy-consuming activities is the cause of these abrupt variations in the profiles. Similar behavior is observed for both SFH and MFH profiles for all temperature classes. The figures confirm the high amount of daily variability induced in the profiles. The profiles for other temperature classes are available in Annex D.

Finally, the approach of DTW was undertaken to compare the dissimilarity of each of the individual IDP with a typical BDEW profile to illustrate the variability provided by each profile. For the analysis, both BDEW and IDP profiles were normalized to their 24-hr. sum. The results were compared in terms of the least overall cost for aligning the two-time series. Figure 4-4 shows the DTW plot of the two extreme profiles with regards to their similarity with the BDEW profiles.



(a) (b)

Source: Author

Figure 4-4: Profile Comparision with DTW
((a) Class 2 vs BDEW;(b) Class 10 vs BDEW)

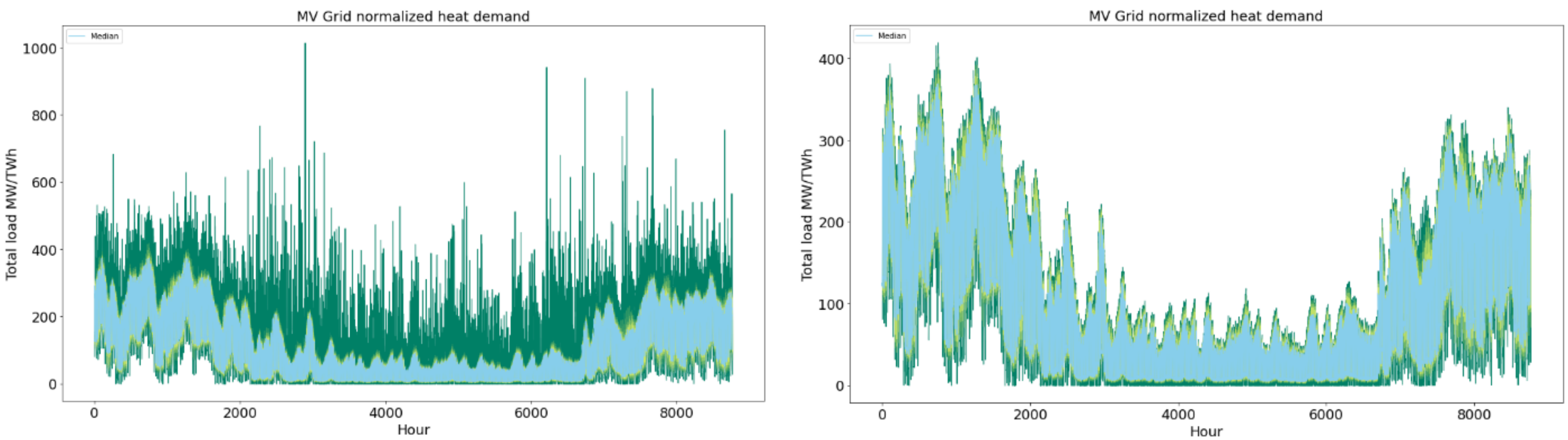
The red path in the diagram indicates the least cost path for the alignment of the two time-series. The inclined paths represent the match of the corresponding points in the two profiles. The horizontal lines represent the deletion of the corresponding point and the vertical insertion to accommodate matching. In Figure 4-4 (a) majority of the connection are with an inclined line thus indicating a higher match of the two profiles. On the contrary, Figure 4-4 (b) is dominated by horizontal lines thus indicating dissimilarity. In terms of the total alignment cost, Class 3 MFH had the lowest of 0.028 compared to the highest of 0.12 in Class 10 SFH. A higher value of the cost indicates dissimilarity. All other comparative plots are available in Annex E.

Summarizing the results of this subchapter, lower-class profiles, typically winter days have a much higher similarity with the BDEW SLP compared to higher classes. The result can be justified considering the feature of BDEW which has an absence of no baseload hours and smooth curve patterns. In contrast, summer days (higher temperature classes) have considerably peaky profiles due to the DHW load and absence of baseload due to no SH demand. Hence the profiles generated from IDP are much closer to the BDEW profiles on winter days compared to the summer days. From the above analysis, the introduction of measured variability with realistic results on a 24-hour scale is confirmed from the implementation of the IDP methodology. Hence supporting the further use of the IDP pool to generate the annual demand profiles.

4.3 The output (High Spatial and Temporal Resolution Heat Demand Profiles)

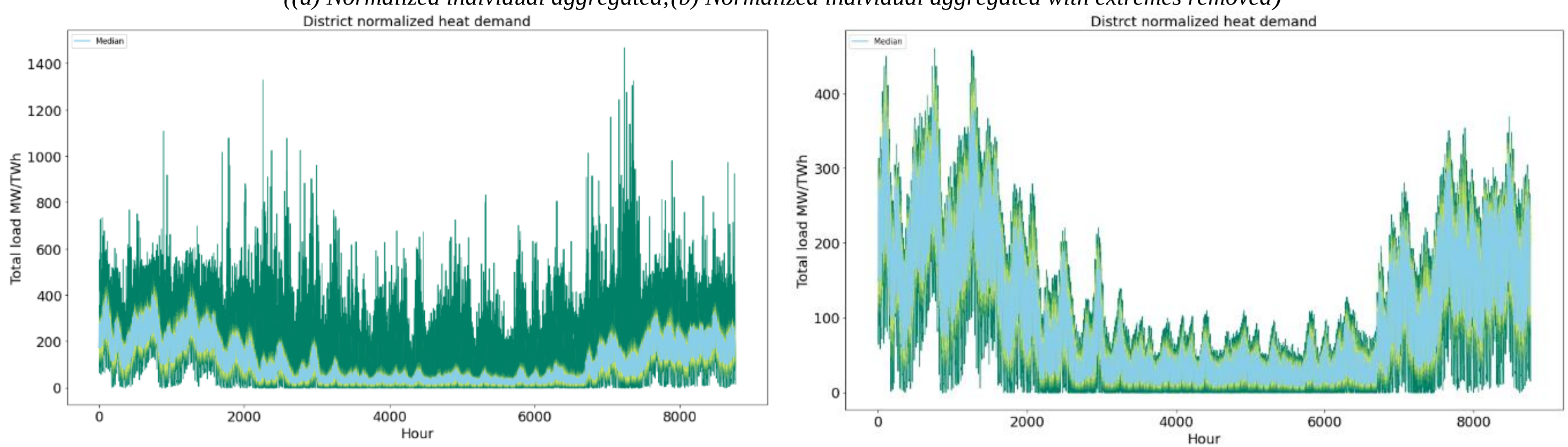
With the use of the defined methodology and implementation of the IDP methodology, the final heat demand profiles for individual census cells were generated. However, for simplicity in evaluation, reduction in data storage volume, and easier visualization, the profiles were aggregated to a pre-defined aggregation level. The aggregation was done under two levels: potential district heating networks and individual heating which have been pre-defined under the eGon project which has been detailed in section 3.3. The DH network covers about 12-15% of the total demand, whereas the remaining is aggregated under the individual heating aggregated under the medium voltage grid distribution acquired from Open Street Map (OSM).

Considering the strict criteria for aggregation of census cells into DH network as defined in section 2.5.2, over 75% of the census cells were seen to fall under the mv grid category with the assumption that these demands are most likely to be met using individual heat pumps. These demands are thus clustered into over 3000 sub-station ids. A wide range of fluctuations and statistical variations were seen in the individual profiles, largely determined by the number of cells aggregated under each station id. Also, profiles with substantially no variability were also observed. Nevertheless, the aggregation brings a reduction in the variability and the respective profiles tend to show a stationary behavior with the increase in the aggregation



(a) (b)

Figure 4-5: Individual Grid Final Profiles
((a) Normalized individual aggregated;(b) Normalized individual aggregated with extremes removed)



(a) (b)

Source: Author

Figure 4-6: District Heating Network Final Profiles
((a) Normalized individual aggregated;(b) Normalized individual aggregated with extremes removed)

(explained in section 4.4.1). The higher variability also indicates the difficulty in maintaining the grid balance due to unexpected peaks but no consistent demand.

Figure 4-5 visualizes the variability of the generated profiles based on the hourly median value represented by the blue line with shades of green representing every 10th percentile in the hourly values. The profiles were normalized since the large fluctuation resulted in difficulty in interpreting the graph. The extreme values represent the randomness associated with the individual station_id of the mv grid. However, in terms of general tendency, most of the curves seem to show a consistent controlled variable pattern with patterns similar to the OPSD profiles. In terms of energy, the variation between the median hourly values and the 90th percentile is not significantly different. 90% of the profiles have a peak of around 400 MW/TWh at around hour 1000. On the higher side, 10% of the profiles show a greater variance with the demand peaking as high as 1000 MW/TWh. These peaks are for those mv grids with a large number of census cells with overlapping demand. The median curve is seen to have a standard deviation of about 79% of the mean value. This level of variance is acceptable since this mainly results due to the scaling from the daily demand factor and not the individual IDP itself. Overall, the patterns observed in the curve are consistent for all generated curves and as mentioned above the large effect of randomness seen on the individual census cell profiles is observed to be minimal once aggregated.

Similarly, Figure 4-6 represents the results for the aggregation as per the potential district heating areas. For the analysis potential areas for 2035 scenarios were considered. The results obtained here were very similar to the ones discussed above. 90% of the generated profiles are similar and significantly smaller than the upper 10% of the hourly values. An observation from the plot was the slight increase in the normalized hourly peak value in the case of DH aggregation. Nevertheless, this is not an actual increase in consumption but is seen only due to the smaller total annual demand in the grid networks compared to the mv grid. All in all, the impact of aggregation on the reduction of variability was observed here as well. A correlation of 0.98 was calculated between the two aggregated curves (grid and district), which shows that the aggregation can omit the randomness.

Overall, the two aggregation levels provided a means for ease in the storage of the data and a measure for better interpretation of the results. The final database is accessible either in these aggregated formats or per census cell resolution.

4.4 Curve Characteristics Comparison

4.4.1 Nature and Pattern of the generated profiles

Evaluation of nature and pattern characteristics of time series provides a broadened perspective to understanding its behavior and there, in turn, giving insight on its suitability for energy modelling application. For the evaluation of the nature of the time-series the stationarity test was conducted. Though the exact desired nature of the results is unknown considering the absence of measured data; this test was done to compare the nature with the reference profiles and dig deeper into the cause of these results. Based on the literature Tang et al., 2013 augmented Dicky Fuller (ADF) test was conducted to test the profiles for stationarity.

ADF Null Hypothesis: "If failed to reject, the time series has a unit root, meaning it is non-stationary. It has some time-dependent structure" (Brownlee, 2016).

ADF test was conducted on each of the profiles: reference data, IDP output on all aggregation levels, and for randomly selected sample census cell profiles. The results are presented in Table 4-1.

Table 4-1: Stationarity Test Results

Dataset	Stationarity Status	Confidence Interval	Source
OPSD National Level	Stationary	95%	(Open Power System Data (OPSD), 2020)
IDP National Level	Stationary	95%	Author
IDP District Heating Network	Stationary	99%	Author
IDP MV Grid	Stationary	95%	Author
IDP Census Cell	Non- Stationary	-	Author
DIW Berlin	Non-Stationary	-	(DIW Berlin, 2017)
DIW Hamburg	Non-Stationary	-	(DIW Berlin, 2017)

The stochastic approach taken for the assignment of the daily profiles in the IDP methodology results in high variability and extensive fluctuation in the time-series standard deviation. Thus, the non-stationary behavior is a consequence of implementing a high variability to the individual census profiles. The non-stationarity nature of the heat demand profiles makes the development and timely update of the profiles important as abrupt changes in the demand pattern are hard to predict. As such analysis may lead to misinterpretation in data understanding and forecasting (Iordanova, 2020) the use of census level data for further forecasting is not recommended. However, once the aggregation is done, fluctuation in the variances and the means of the curves tend to decrease the resulting in the curve showing behavior of stationarity which has higher similarity to those generated from existing SLP.

As can be seen from the table the aggregation of the generated profiles tend to follow the behavior comparative to the reference profiles. This result aligns with the covariance comparison result obtained in section 4.5.2 where the correlation between the reference and the generated profiles were observed to increase with aggregation. However, considering the larger variability offered in the census level resolution, the IDP profiles are arguably able to provide higher accuracy to the heat demand estimation. A stationarity measure of the profile also indicates the absence of abrupt changes in the hourly values, an aspect whose correct representation would be critical for gird design and optimization for future implementation of the generated results.

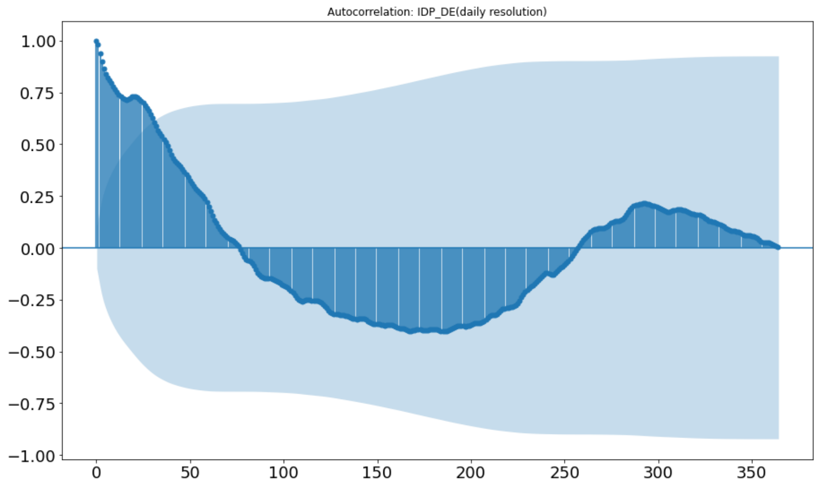
Nevertheless, the DIW profiles did not show a stationary result as would have been expected from an aggregation. A much larger seasonal variability is observed in the DIW profiles, meaning that the overall winter demands are much higher than in summer. The exact cause of this behavior could not be deduced, but this is the primary cause of the non-stationary behavior in DIW. The aggregation of census cells within both Berlin and Hamburg networks showed stationarity, thus keeping consistent with previous results of reduced variability on aggregation in the IDP methodology.

The nature assessment was followed by a comparative assessment of the profile patterns This was done to better understand and identify the cyclicity[1], seasonality, and other relevant patterns and their similarity to the reference profiles. For this autocorrelation of the generated profiles was determined as suggested by Tang et al., 2013. For a random census sample, the autocorrelative plot showed demand patterns close to the temperature cycles, thus indicating the strong effect of temperature on the demand profile mostly associated with residential SH.

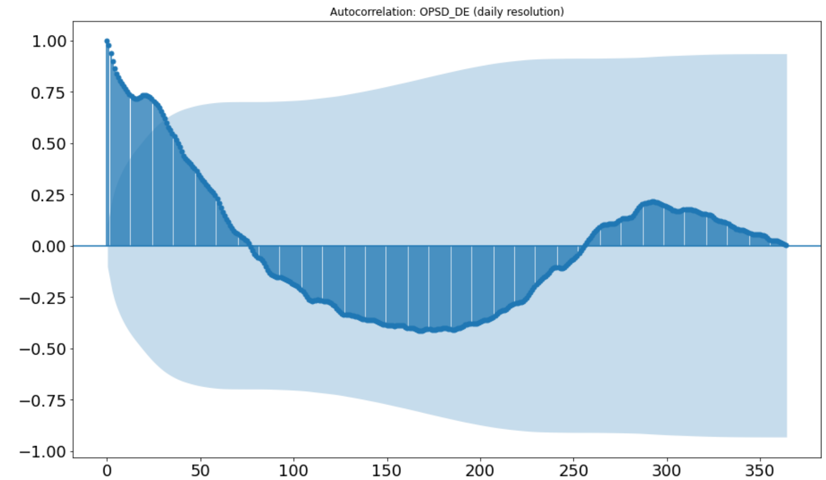
Overall a similar trend was observed with regards to the autocorrelative behavior of the profiles for both reference and generated. An obvious pattern of seasonality can be observed in the profile considering the decaying, but fluctuating demand correlation observed in the lags. The ACF plots are presented in Figure 4-7. For better visualization, only daily resolution plots are provided. In general, a significant correlation was observed up to a lag of 30 days. However, statistically significant lag was observed only on the first 3-4 days with a Pearson's coefficient of over 0.8. On an hourly scale, the first 3 days' lag peaked exactly on the time interval of the 24^{th} hour. This outcome is in line with the behavior observed in temperature-based time series. Intuitively, demand today is reasonably closer to the demand at the same time the next day. But further away from the starting point the patterns tend to change. However, as the lag tends closer to the 24th hour the correlation tends to increase, and the cycle continues.

---

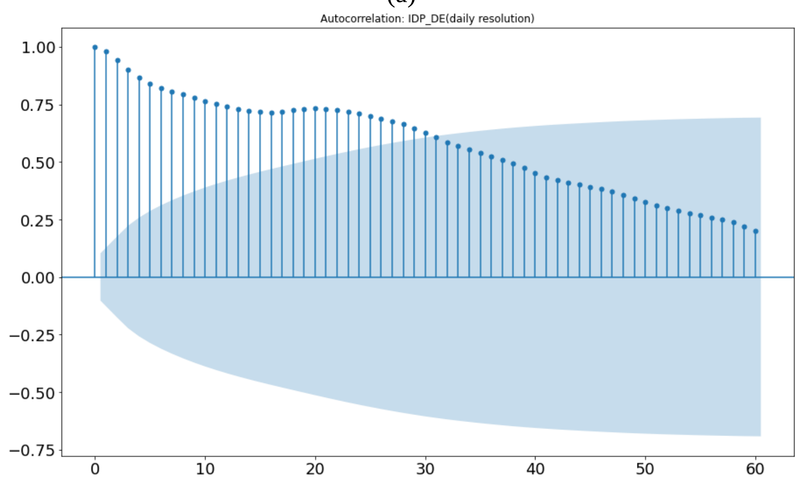
[1] Time series fluctuations that are not of fixed period is referred to as cyclicity and those associated with some aspect of the calendar is referred to as seasonality (Hzndman, 2011)



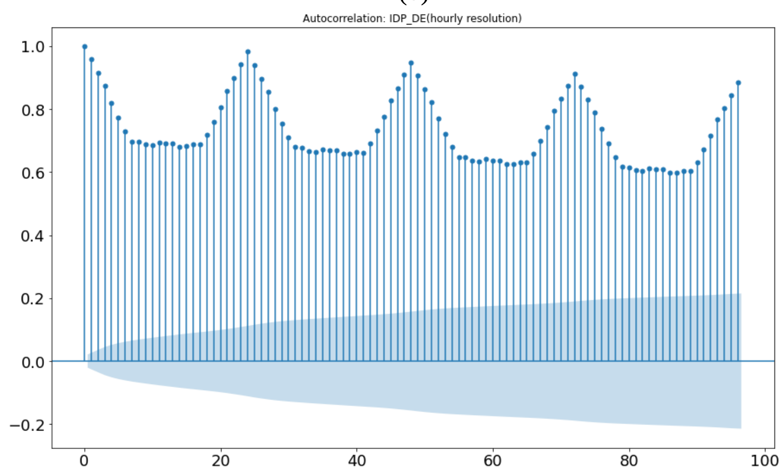
(a)



(b)



(c)



(d)

Source: Author

Figure 4-7: Autocorrelation Analysis
(a) Autocorrelation IPD-national profiles (daily resolution);(b) Autocorrelation OPSD-national profiles (daily resolution);(c) Autocorrelation IPD-national profiles (daily resolution-60 days);(d) Autocorrelation IPD-national profiles (Hourly resolution 100hrs.)

A cyclic behavior is observed in the patterns where the correlation from the 20th to the 24th-day lag tends to slightly increase. Though the exact cause of this pattern could not be verified, a similar trend was also observed in the reference profiles. Hence supporting its presence.

Though analysis of autocorrelation function of non-stationary time series is not ideally recommended, it was nevertheless conducted for the census level data. The ACF plot showed a similar pattern to the aggregated data but a correlation value with much lower statistical significance. Thus, this further shows the effect of the randomness and the higher degree of variability in the high-resolution profiles. As a result, as stated above any further statistical process based on these profiles may not be able to give a realistic model thus further reasoning for data processing on a certain aggregation level.

The test for autocorrelation for random sample census cells and all aggregation levels showed a seasonality factor that largely coincides with the behavior observed in the temperature patterns thus indicating an anticipated presence of a weather seasonality though the annual profile lacks any statistical seasonality. The curve shows two winter contiguous blocks of data and one summer contiguous block. Nevertheless, a similar level of cyclicity patterns was observed in both the generated and reference profiles Thus, qualitatively it can be validated that the profile closely relates to the reference profiles with the desired degree of similarity with the ambient temperature behavior.

4.4.2 Chow Break Test

A Chow Break Test was conducted to observe the structural similarity of the generated profiles with the reference profiles. The test is based on the null hypothesis that the two-time series can be represented by a single linear regression. The test assumes that the time series data are stationary with a stable change in the variations. The above-discussed stationarity and autoregressive analysis thus help identify the nature and pattern of the generated and reference time series and perform the Chow test. A chow break test is ideally used for observing a breakage in different subsets of data within a multivariate time series. For the application in the thesis, the generated series and the reference series were considered as two subsets and their linear regression concerning the ambient temperature was calculated. For the nationally aggregated data, the average temperature of all the TRY climate zones was used. Though the use of average temperature data was avoided in previous sections, no other alternative could be identified here.

The results obtained from the test showed a high degree of structural similarity of the generated aggregated profiles with the reference profiles on a daily resolution. Thus, validating the IDP methodology can produce results similar to the OPSD database. However, on the hourly resolution, the structure of the database is seen to be extremely different. Though the profiles do not tally on this resolution, the variability was a desired output of the study. In the case of the DIW comparison, the rejection of the null hypothesis was observed in all aggregation levels. Though the autocorrelation function for DIW profiles showed results very similar to the IDP and OPSD profiles, the structure of these profiles is quite different. A primary cause of this is the linear function used to establish the relationship between the demand and temperature in this methodology as mentioned in section 2.4.2. Due to this, a highly sloped winter-to-summer curve can be seen in the DIW profiles. On a census level, the structural similarity between the IDP and the existing state of the art could not be obtained. Nevertheless, this was an expected result considering the large variability obtained in the IDP methodology. The results of the Chow Break test are summarized in Table 4-2.

Table 4-2: Results Chow Break Test

	Resolution	F-statistics	P value	H_0 Status
OPSD national vs IDP National Total Demand	Hourly	31.75	≈ 0	Reject
OPSD national vs IDP National Total Demand	Daily	0.55	0.57	**Retain**
DIW Berlin vs IDP Berlin	Hourly	194.14	≈ 0	Reject
DIW Berlin vs IDP Berlin	Daily	51.04	≈ 0	Reject
DIW Hamburg vs IDP Hamburg	Daily	20	≈ 0	Reject
DIW Hamburg vs IDP Hamburg	Hourly	83.14	≈ 0	Reject
IDP Census vs BDEW census	Hourly	70.48	≈ 0	Reject
IDP Census vs BDEW census	Daily	5.22	0.005	Reject

The Chow Break test also can be used as a measure of statistical validation of the Pearson's correlation. The method has been implemented to measure the covariance of the generated and the reference profiles in the following sections.

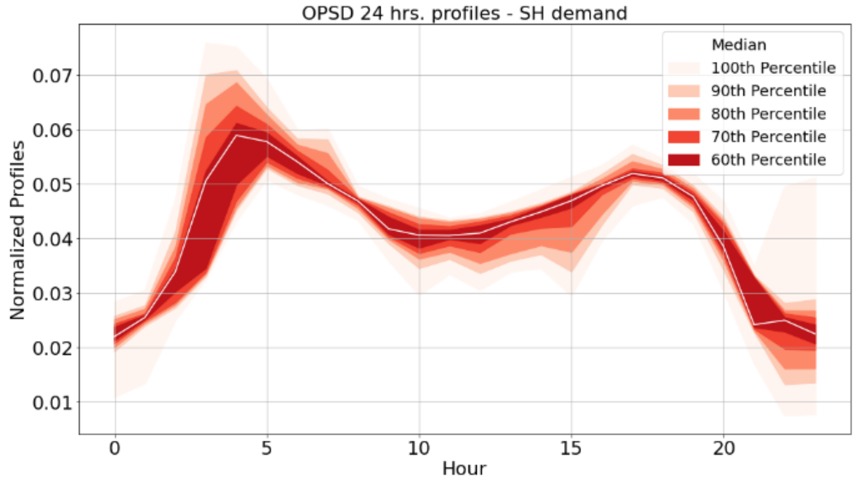
4.5 Comparative Validation

4.5.1 Comparison with the BDEW -24hrs. scale

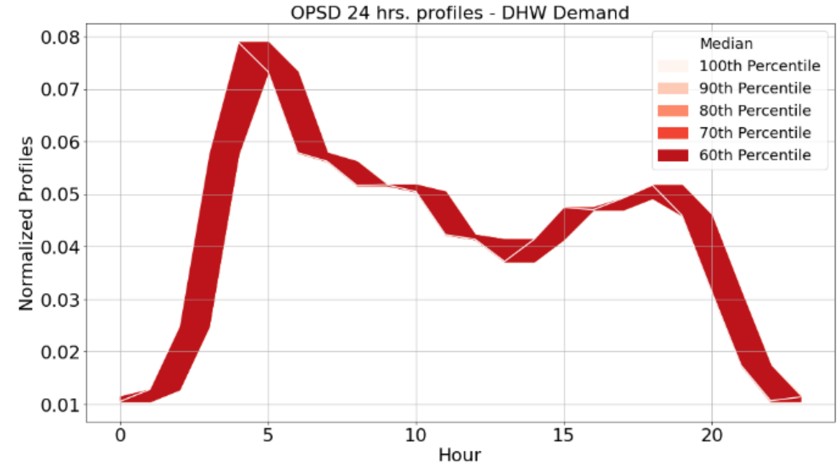
The identification of the OPSD database avoided the need for the creation of individual BDEW based census profiles and their aggregation into the national level, as was initially proposed. The OPSD database provided a much simpler and convenient alternative. The creation of the BDEW profiles for individual cells would have induced enormous computational time and data storage volume. Therefore, to avoid redundancy the OPSD was used for analysis for all national-level comparisons. Also, through literature, OPSD could be confirmed to be a well-established and reliable source for demand data on a national scale and has been used in other studies (Maruf, 2021) for providing the closest estimation to the heat demand profiles. Hence the database can be deemed feasible for use as a reference profile for the comparison and evaluation of the generated demand profiles.

As the OPSD profiles follow the BDEW methodology the variability in terms of the 24 hr. profiles is absent or very limited. Figure 4-8 represents the variability observed in the profiles at different levels for both reference and generated profiles. For better visualization, all 24 hr. profiles have been normalized to their sum (MW per MWh). Each shade of red represents the 10th percentile of the data for every hour in a 24hours period. Thus for every hour of the day, the spread of the demand values can be seen. The time of peak demand is consistent with a specific time of the day for all profiles. For OPSD, though SH demand Figure 4-8(a) patterns are partially justifiable, the behavior of the DHW profile Figure 4-8(b) is highly unrealistic. Regarding daily variability, limited variability can be observed in the SH profiles, but the DHW profiles are repetitive and identical for almost every day of the year and indicated by all values equal to the median. A maximum standard deviation of 0.0125 was observed in the normalized total heat demand profile Figure 4-8(c). Rather than an explanation of the numbers the variation in the profiles can be observed clearly when observed side by side with a sample output of the IDP methodology. Figure 4-8 (d)SH, (e) DHW and (f) Total demand represent the IDP output of one random cell. In general, a higher spread of the hourly curves can be observed in the IDP profiles compared to the OPSD profiles. Similar results were obtained for other sample profiles where the peaks were seen to be over 15 times the hourly mean. A clear pattern of the majority of days can be seen on the removal of the upper 10% extremes Figure 4-8 (g). However, once all census cells are aggregated to form the national-level data, the large variability is seen to disappear as seen in Figure 4-8 (h). Hence, the aggregation generates results very similar to the OPSD results in terms of variability. This thus indicates the capability of the method to generate similarly variable profiles on a high aggregation level.

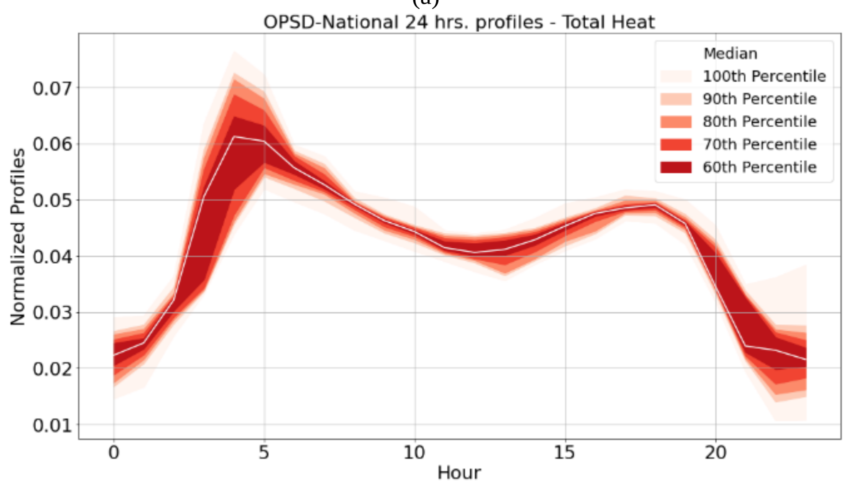
In terms of the national level peaks, the IDP is seen to be slightly skewed towards the evening peak compared to a morning peak on OPSD. This could be due to the higher occupant active probability during these hours but the exact cause could not be confirmed.


OPSD 24 hrs. profiles - SH demand
Normalized Profiles
Hour
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile

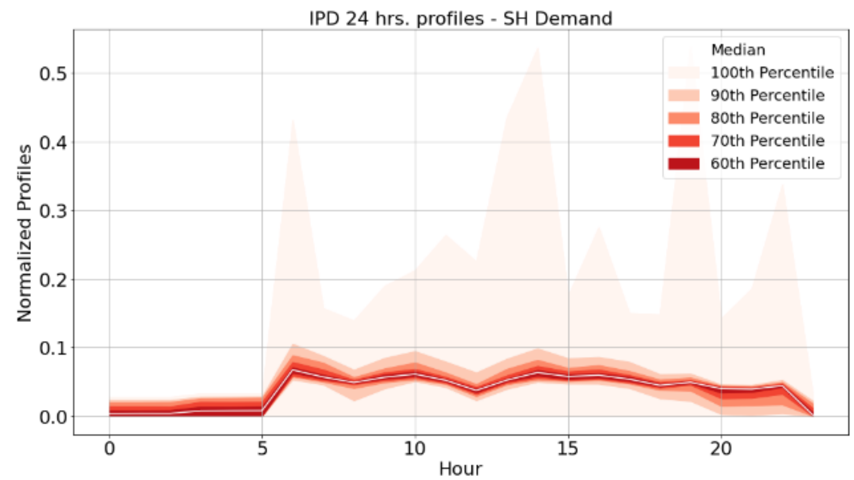
(a)


OPSD 24 hrs. profiles - DHW Demand
Normalized Profiles
Hour
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile

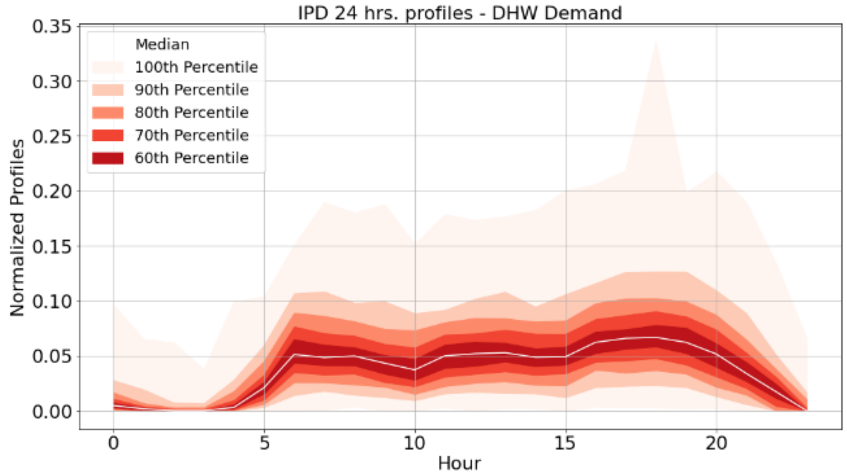
(b)


OPSD-National 24 hrs. profiles - Total Heat
Normalized Profiles
Hour
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile

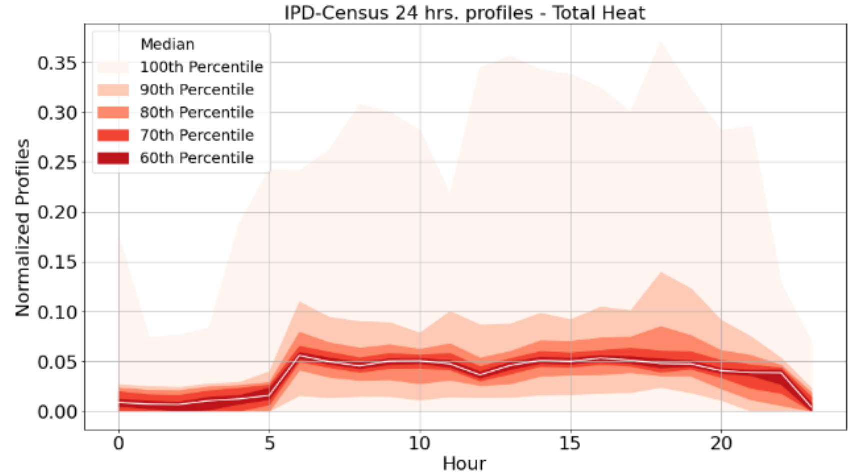
(c)


IPD 24 hrs. profiles - SH Demand
Normalized Profiles
Hour
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile

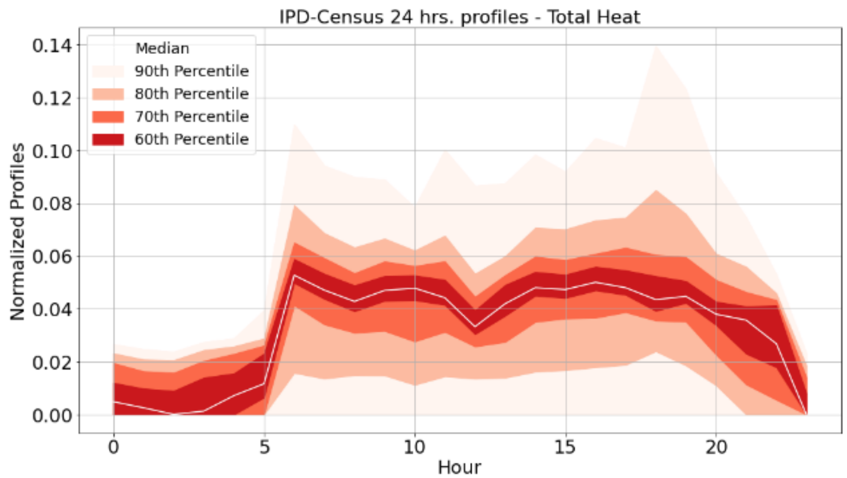
(d)



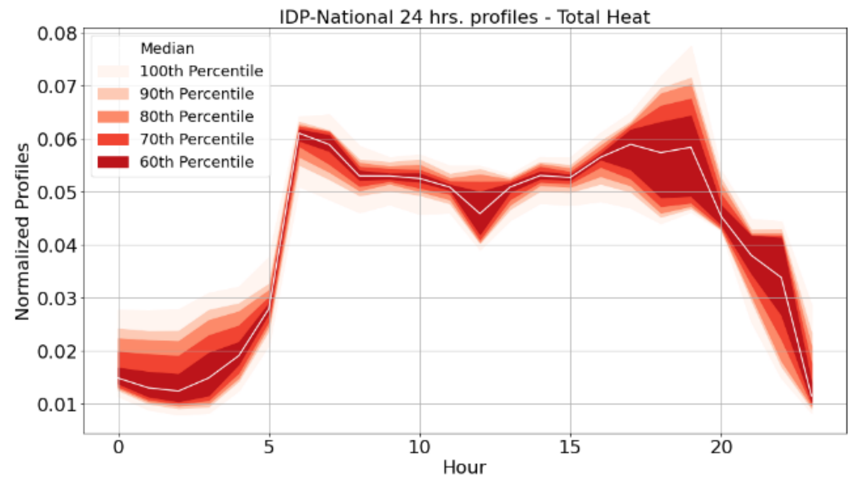
(e)



(f)



(g)



(h)

Source: Author

Figure 4-8: Profile Variability Comparison
(OPSD profiles[(a)SH,(b)DHW,(c)Total heat]; Sample IDP[(d)SH,(e)DHW,(f)Total heat];(g) Sample IDP tota heat extremes removed (h) IDP Total heat National)

For the IDP SH profile, constant normalized profiles are observed with the majority of the variations within the 10^{th} and 90^{th} percentile with a few outliers. This would be true as the load is almost entirely dependent on the ambient temperature. Also, the effect on the time of the day with regards to the occupancy status can be observed in the patterns with peaks in mornings and evenings when the occupants are highly likely to be active. This behavior was observable when the graphs were viewed on removing the outliers. Among all tested samples a small number of outlier profiles were identified which is significantly higher than the SH mean. These profiles comprised of a sudden surge and instantaneous drop in demand which are ideally a behavior seen in DHW profiles and not expected in SH demand. For all random samples, such profiles are observed on summer days where ideally SH would be absent. Due to limitations to the access to the LPG directly, based on the literature this unexpected behavior of the SH profile could only be explained by the stochastic nature of the profile assignment. Firstly, the existence of such 24hrs. profile in base profile outputs is due to a limitation of the load profile generator. As explained by Drauz, 2016, the model increases the indoor temperature to 22 °C when the occupant is active and the set temperature is below the ambient temperature. In winter, the heating is consistently on and hence such peaks are not observable as the energy needed to maintain a consistent temperature is less than to meet sudden surges (peaks). However, in summer mostly in the early morning, when the temperature might drop for few hours and in case the occupant is active then such surges are expected. Hence the presence of these profiles can be attributed to the occupancy model, where there are times in the early morning, even though low in probability, there could exist instances when the occupant is active. Of the total SH-IDP generated from the base profiles, only 0.2% of the profiles were seen to show such behavior, however, the probability of selection for LPG accumulation still exists. However, since the final IDP profiles are only concerned with the aggregated DHW and SH profiles, sources of such profiles are indistinguishable from the aggregated IDP pool. Therefore, the existence of this error in the generation of solely SH profiles from the LPG is self-corrected by the model when dealing with aggregated profiles SH and DHW profiles.

For IDP DHW, as expected a much larger degree of variance is observed. Here the large degree of fluctuations can be attributed to the occupancy model which largely defines the DHW demand. Compared to the OPSD, this output can be confirmed to give a more realistic profile or at least successfully provide the desired level of variability. The final total demand IDP profiles are generated from the base profiles, where most of the peaks are attributed to the DHW demand, and significantly larger variation is observed amongst profiles compared to the corresponding OPSD.

4.5.2 Comparison with BDEW – Annual Scale

The comparison of the annual profiles was done on two levels. Firstly, the comparison of the OPSD was done with the direct output of the LPG (base profiles) and then again, with the IDP profile. This approach was chosen as it would not just compare the generated profiles to the reference but also give an overview of the effect of the randomness associated with both LPG and IDP. Also considering the aggregation level and spatial resolution of all three profiles are different, to ensure consistency in the compared data, profiles have been normalized to hourly power consumption per TWh.

A comparison between the LPG and OPSD profiles was done to observe the similarity between the two available datasets. Though LPG is verified by Drauz, 2016, its real-world application has not been observed. Also, the input parameters for base profile generation runs needed validation with the existing state-of-the-art. The cause of the difference in the IPD output, if any, can be connected to the assumptions undertaken for the development of the base profiles. Hence the comparison of the LPG with the OPSD was also performed.

In comparison to the reference profiles, some interesting, as well as some expected results, were obtained as predicted before the conduction of the analysis. A high degree of fluctuation was observed between the OPSD and LPG DHW profiles. A very low correlational coefficient of only 0.089 was observed between these two profiles. This was expected as section 4.5.1 indicates daily OPSD-DHW profiles are identical with the daily mean representable by a horizontal line. In contrast, the LPG shows significant unpredictability and variability on the daily curves because of the randomness brought by the occupancy model. As summarized in Table 4-3 and Figure 4-9 the highest correlation between the two profiles was observed in the SH profiles, true considering the effect of randomness in this model is minimum. This also helps in rectifying the above-identified error in the SH output of the LPG. Thus, the presence of the unusual DHW like profiles can be overlooked in the annual profile as its impact is insignificant. The LPG total heat demand profiles are the aggregation of the SH and DHW profiles and show an average correlation between the two. An effect of randomness can be better visualized from the graph where the randomness decreases moving from DHW to total heat demand. In the case of the total heat demand, the correlation tends to increase on comparing the profiles from an hourly resolution to a daily resolution. On a daily resolution, the two curves seem to almost overlap each other with the correlational coefficient increasing from 0.76 to 0.96. This also clearly indicates that the profiles are primarily different because of the implemented randomness in the LPG model. Hence the OPSD methodology as a result of its limited variability limits in has limited application in high spatial resolution studies.

Comparison of the IDP profiles, for the census cells profiles, the similarity trends were seen to be much closer for daily demand resolution than compared to the hourly resolution. This was an expected result considering the high hourly variability. However, the comparison with the national aggregated IDP, the generated IDP was seen to be almost identical to the OPSD profiles. This was also verified by the low overall cost of alignment between the two profiles. The result is also in line with that of the previous section.

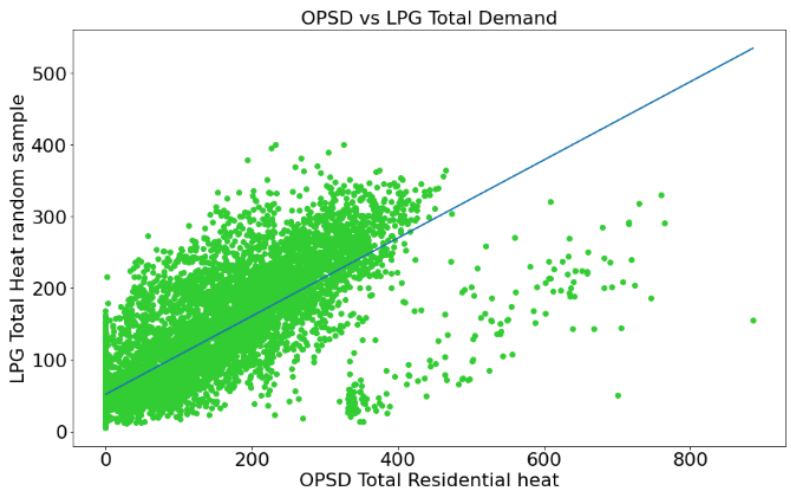
Table 4-3: Statistical Comparison-Summary (OPSD vs.Others)

	RMSE		**Correlation Coefficient**		**MAPE**	
	LPG	**IDP**	**LPG**	**IDP**	**LPG**	**IDP**
Total Residential heat demand (hourly resolution)	91.64	49.12	0.76	0.852	75%	48%
Total Residential heat demand (daily resolution)	36.4	15.45	0.96	0.979	39.81%	11.61%
Residential SH demand (hourly resolution)	55.98	-	0.88	-	-	-
Residential DHW demand (hourly resolution)	300.28	-	0.089	-	-	-
IDP national total heat (hourly resolution)	-	3.6	-	0.92	-	2.06%

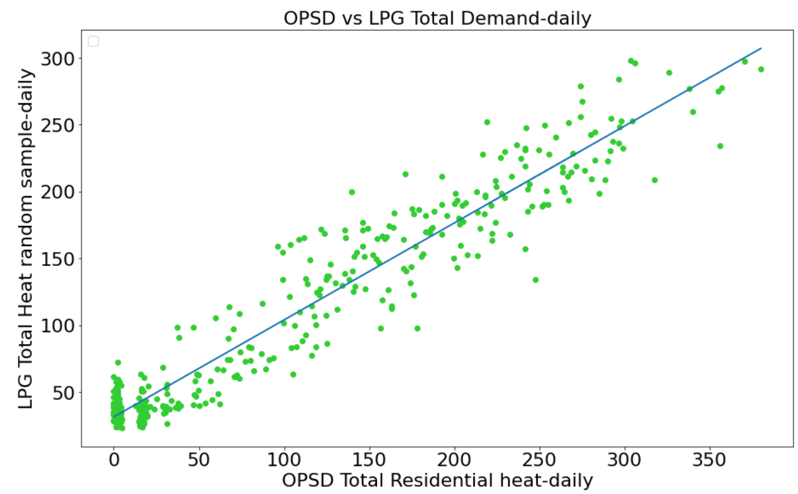
Table 4-4 summarizes the statistical results obtained from the comparison of the IDP and the DIW profiles. Comparisons were made for the Hamburg and Berlin largest district heating networks. Unlike with the OPSD profiles the difference in the parameter values on daily and the hourly resolution was not seen to be as significant. Nevertheless, in general, all profiles show less similarity with the DIW profiles compared to OPSD. This is primarily because of the piecewise linearization approach undertaken by DIW which results in a much steeper decline in demand in comparison to the temperature, as explained in section 2.4.2. The low correlation in the DIW Profiles can be observed in Figure 4-9 (g),(h).

Table 4-4: Statistical Comparison-Summary (IDP vs. DIW)

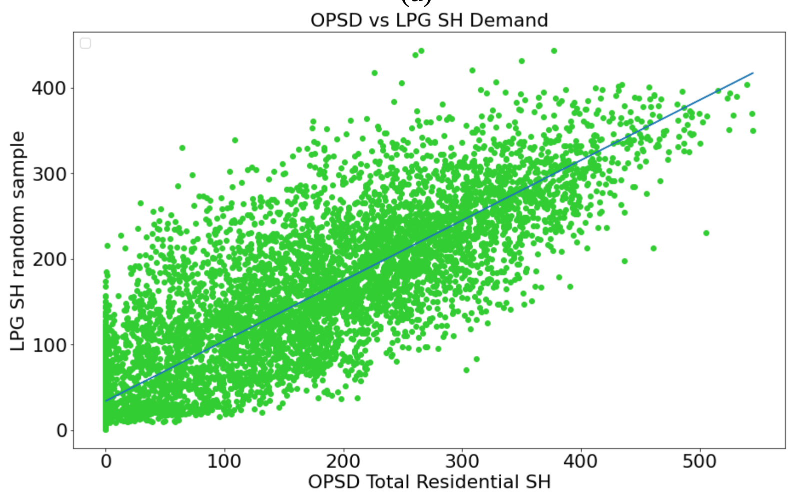
	RMSE	**Correlation Coefficient**	**MAPE**
Hamburg (daily)	52.39	0.76	54.23%
Hamburg (hourly)	69.39	0.65	69%
Berlin (daily)	55.44	0.75	47.9%
Berlin (hourly)	71.74	0.66	63.19%


OPSD vs LPG Total Demand
LPG Total Heat random sample
OPSD Total Residential heat
0
100
200
300
400
500
0
200
400
600
800

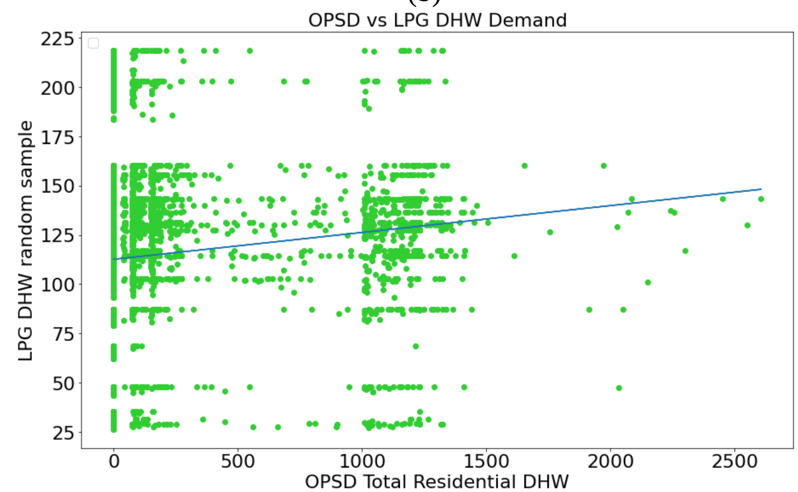
(a)


OPSD vs LPG Total Demand-daily
LPG Total Heat random sample-daily
OPSD Total Residential heat-daily
50
100
150
200
250
300
0
50
100
150
200
250
300
350

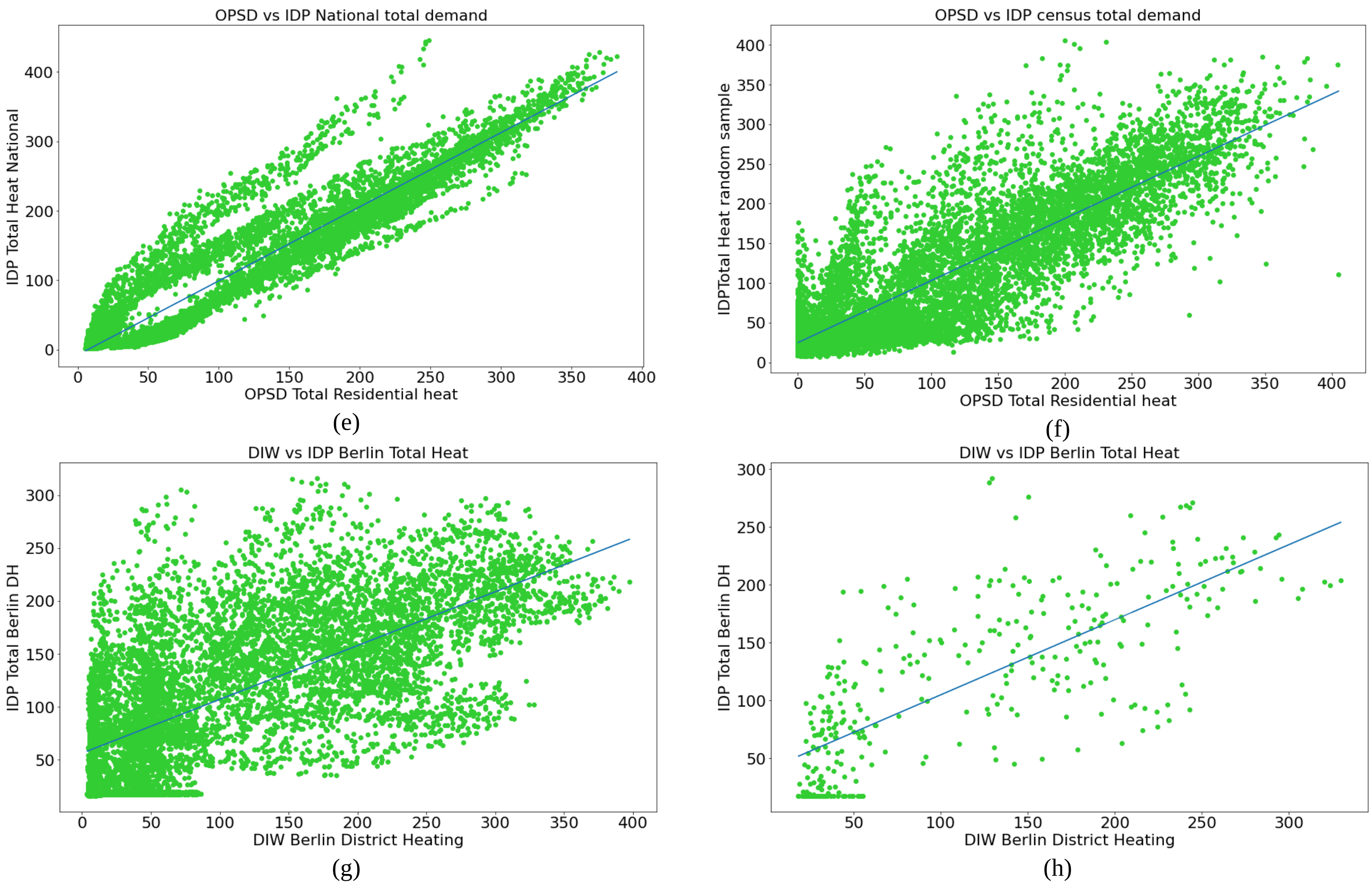
(b)


OPSD vs LPG SH Demand
LPG SH random sample
OPSD Total Residential SH
0
100
200
300
400
0
100
200
300
400
500

(c)


OPSD vs LPG DHW Demand
LPG DHW random sample
OPSD Total Residential DHW
25
50
75
100
125
150
175
200
225
0
500
1000
1500
2000
2500

(d)



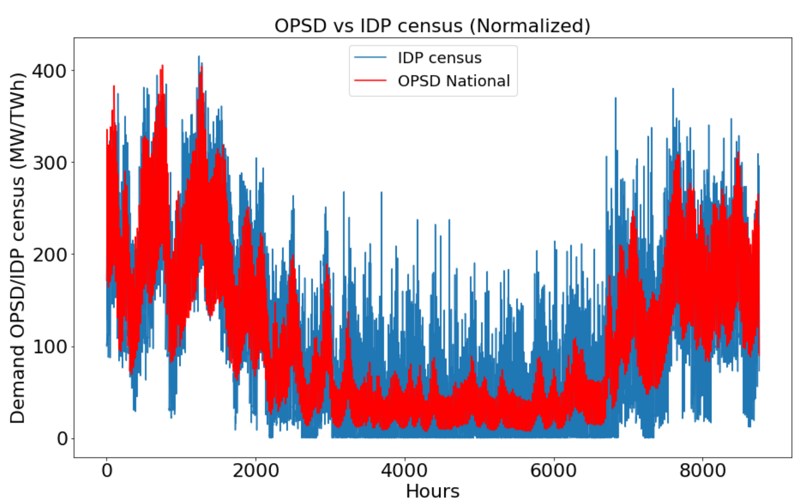
Source: Author

Figure 4-9: Correlational Plots

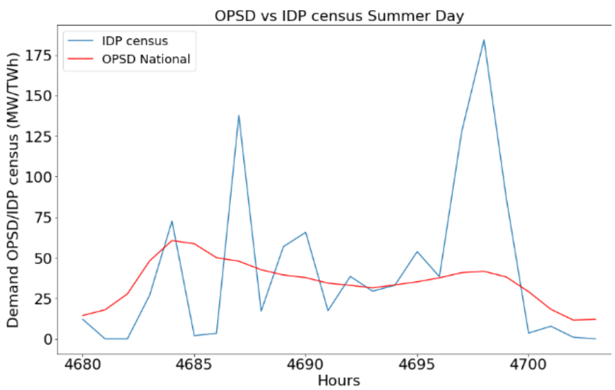
(OPSD vs LPG[(a)Total Demand (hourly resolution);(b)Total Demand(daily);(c) SH (hourly);(d) DHW (hourly)];OPSD vs IDP[(e)Total demand national (hourly);(f) Total demand census sample(hourly)];IDP vs DIW[(g) Total Demand Berlin (hourly);(h) Total Demand Berlin(daily)])

On comparison of the OPSD to a random IDP census cell, the correlational coefficient in an hourly resolution was in the range of 0.75 to 0.86. This is slightly higher than what was observed during the LPG-OPSD comparison. The main reason for this would be due to a slight decrease of variability in IDP results compared to the LPG profile on a census cell level as a result of the daily demand factor scaling. The restrictions induced while defining the daily demand factor and the restrictions on random IDP selection from the pool with temperature intervals are possible reasons for this behavior. Nevertheless, aggregated national level IDP profile showed a much higher correlational similarity with the OPSD even on hourly resolution, thus further validating the similarity between these two profiles. The plots can be seen in Figure 4-9 (e),(f). In the hourly plots, the deviated values from the best fit line are observed to be generated due to the DHW demand. Nevertheless, the correlational behavior is in line with the results of section 4.5.1. Further, the structural similarity of the two observed in section 4.4.2 provides a concrete justification on the ability of IDP methodology to replicate reference profiles on a national level with included higher variability on high spatial resolution, which is absent in other methodologies. This provides an immense possibility for replacing the existing SLPs.

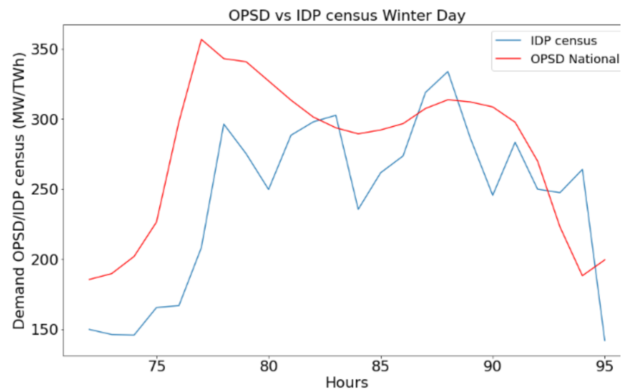
In terms of the coefficient of determination, 72% of the variance between the IDP and the OPSD is explainable, and 28% unexplainable. Thus, the IDP model provides 28% more variability in the profile. This in turn avoids the repetition of the daily profiles and gives unique patterns thus further confirming higher variability on the census cell level. Figure 4-10 provides a visualization of the IDP census cell against the OPSD on the annual profile. A clear variability and higher peaks can be seen in the IDP profiles as discussed in the previous sections. The much closer profile patterns in winter than in summer is in line with the results obtained in section 4.2. Similarly, the profiles for random summer and winter days show OPSD vs IDP behavior as discussed in. the same section.



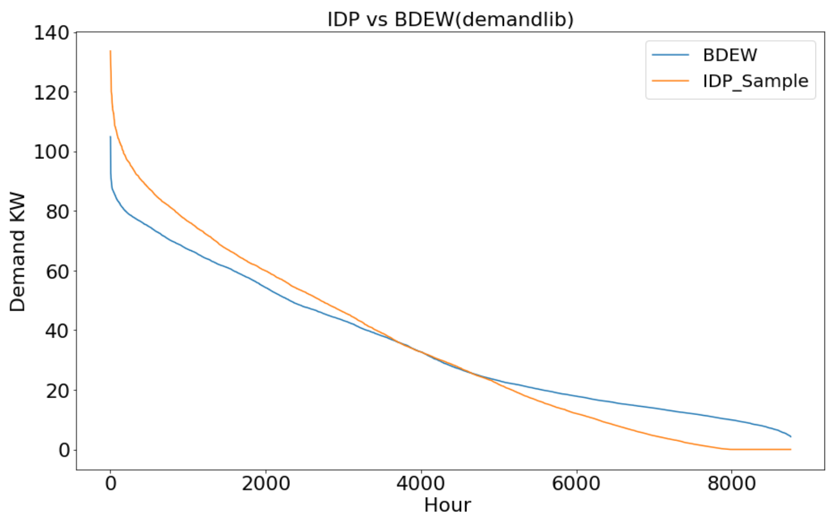
(a)



(b)



(c)



(d)

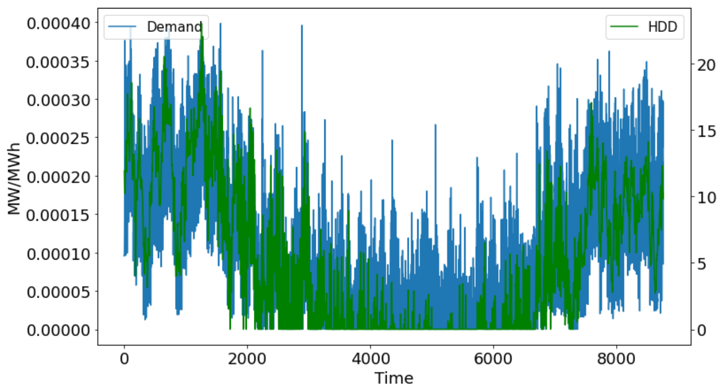
Source: Author

Figure 4-10: IDP census cell Vs OPSD normalized
Annual profile comparison normalized (b) Sample Summer Day Comparison (c) Sample Winter Day Comparison (d) Load Duration Curve

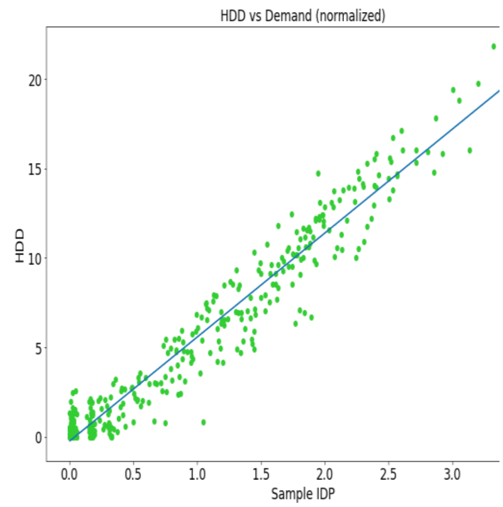
Figure 4-10 (d) shows the load duration curve comparison of the IDP profiles with the BDEW. For this comparison, BDEW based profiles were generated for a randomly selected census cell using demndlib. The comparison shows a much higher peak in the IDP profile compared to the BDEW, as discussed above. The IDP curves are much steeper compared thus indicating unstable inconsistent demands which can be explained by the variability. Using BDEW would have overlooked this characteristic and resulted in incorrect model results. Further, following the comparison, the IDP shows hours with zero demand value which is absent in BDEW. Consistent results were obtained from all sample profiles.

4.5.3 Demand-HDD relationship

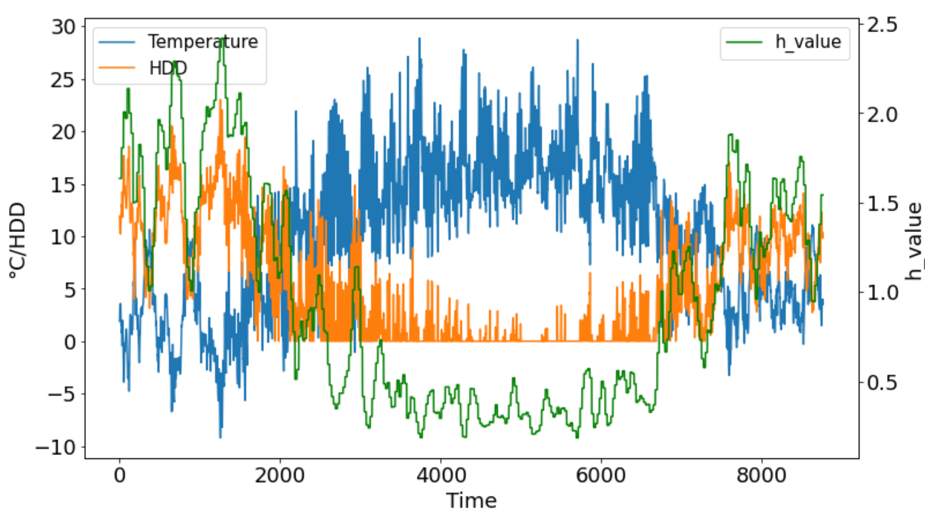
As mentioned in section 2.4.3, the use of heating degree days is the simplest method to estimate a realistic heat demand profile. In the absence of available measured data, it has provided a means of close estimation of heating profiles, especially for the residential sector. Hence a comparative analysis of the heating degree days with the generated profile was performed. The comparison was done with the hypothesis that a statistically significant correlation will be observed between the two curves. Also, to ensure better accuracy in the output, the comparison was done only on a census level (non-aggregated) level. This is because the aggregation comparison would require the use of average temperature values. Generalized national level HDD would have less accuracy to a census level demand.



(a)



(b)



(c)

Source: Author

Figure 4-11: HDD vs IDP profiles
((a) HDD-IDP census sample (plot);(b) HDD vs IDP;(c) Temperature-HDD-Daily demand factor)

A comparative analysis of the demand with the HDD showed the following results. As can be visually observed in Figure 4-11 (a) a wide range of fluctuation can be observed in the demand profile compared to the HDD profile with much lower variability. Though consistent summer and winter seasonal patterns were observed in both the curves, two significant differences can also be seen. During the winter, though the peaks coincide with the HDD plot a clear much lower demand can be observed. On the contrary in the summer months, even in the absence of HDD, significant demand is observed. This winter's lower demand is presumably due to the consideration of building typology and the summer demand due to the inclusion of DHW. This has been further investigated in detail later in the chapter.

As mentioned in section 2.4.3, estimating the base (threshold temperature) becomes important in determining the HDD. To obtain a generalized value, secondary literature was investigated to identify the most suitable for use for the entire country. Disregarding the effects of the altitude and the energy source of the heat, which greatly affects the threshold temperature, Kozarcanin et al., 2019 estimated a value of 13.8°C as the national average thermal temperature which is slightly higher than the one estimated by VDI 2067, which is 12°C (Kozarcanin et al., 2019, p. 12). For the analysis of this study, both values were used to generate HDD with a pre-assumption that the demand profiles must be closer to the 12°C HDD profiles considering the LPG model is based on the VDI models (Drauz, 2016). However, this was not observed to be the case. Nevertheless, during the analysis, it was identified that the difference in the HDD profiles generated from either of the base temperature values was not significantly different and did not affect the HDD profiles as expected. Hence further analysis was carried out making use of 13.8°C as the base temperature since the source of the VDI data could not be directly accessed. The above mentioned vaues are lower than base temperature values mentioned in other older literature, collection of which is provided in Giannakopoulos & Psiloglou, 2006. Nevertheless, considering that Kozarcanin et al., 2019 is a very recent publication, these values were used.

For this analysis, the generated heat demand profile was compared with the HDD profiles generated for the different stations from the temperature profile. Figure 4-11 (c) presents one of the results of a randomly selected census cell, where the IDP demand profile is compared to the corresponding HDD curve for the respective TRY climate station. As seen in Figure 4-11 (b) the generated profile against the HDD showed a high Pearson’s correlation coefficient of an R^2value of 0.98 against the (Kozarcanin et al., 2019) and an R^2 of 0.96 against the VDI values. This is in line with the correlational coefficient of over 0.9 suggested by Quayle Robert G., 1979. Irrespective of the considered threshold temperature the demand profile is highly correlational to the HDD. The high correlation was an expected result considering the dependence of the profiles on ambient temperature and its implementation on both IDP and LPG models. Since

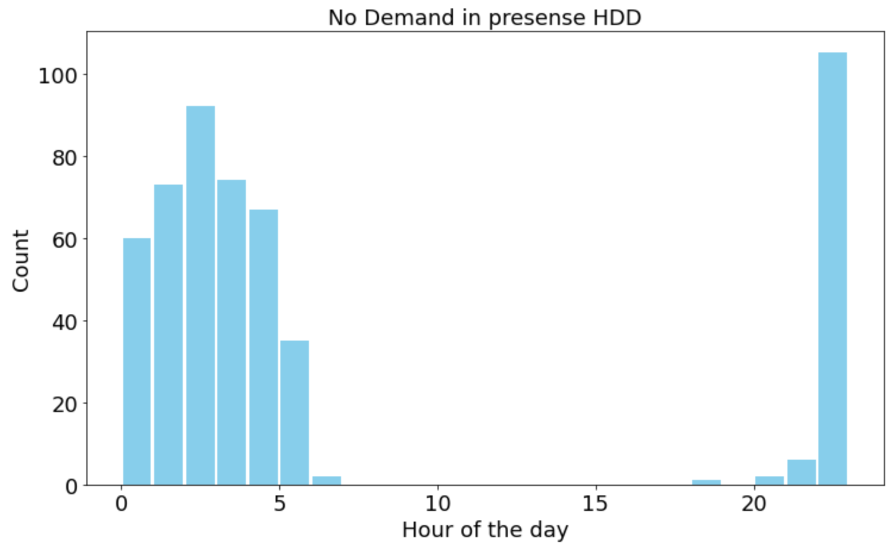
Kozarcanin et al., 2019 have verified the assigned temperature, the close correlation of the IDP profiles with the HDD indicates the realistic nature of the profiles to a certain degree.

Summer and winter deviation in HDD vs Demand

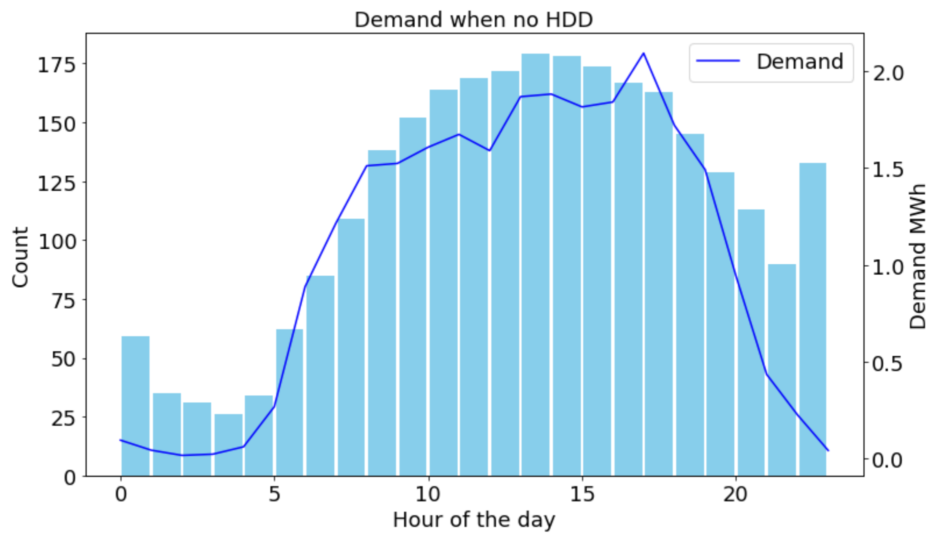
Though a high correlation was observed between the sample profile and the respective station HDD in all random results, there were instances in the lower quadrants of the demand values where a null value of one did not correspond to the null value of the other. These time resolutions can easily be pointed out when comparing the HDD profile with any randomly selected sample. However, it is to be noted that such patterns are only distinguishable in the individual census cell residential demand profiles. Identifying the source of any certain peak is not possible once aggregation with CTS is done. However, though the HDD and demand values contradict, realistic and interesting patterns were observed in these instances. In general, two different scenarios could be identified.

- Absence of demand in hours where the ambient temperature is below-set temperature.
- Presence of demand in hours where the ambient temperature is above the set temperature.

When the ambient temperature is less than the required set temperature (presence of HDD), instances were observed where the demands are absent. The majority of it was observed in the winter while some hours in the summer also observed this behavior. In both cases, such demand curve behavior was observed during late night and early mornings (in between 23:00-3:00) as seen in Figure 4-12 (a) where each bar in the graph indicates the hour count where no demand was observed even in the presence of HDD, aggregated over 24 hours for a whole year. The highest probabilistic occupancy status at this time interval is expected to be inactive, thus partially justifying the absence of the demand in terms of DHW, since the occupant is expected to be active for the DHW consumption to occur. Though the SH is not as much affected by the activity of the occupant, the justification for both the summer and winter months can be made considering the thermal heat storage in the buildings. The energy stored in the buildings maintains the desired comfort temperature, thus making the need for additional SH non-essential (Hellwig, 2003). Also, this indicates that HDD is a method designed for the estimation of heating demand and not vice versa as demand also depends on numerous other factors.



(a)



(b)

Source: Author

Figure 4-12: HDD vs Demand (Winter and Summer Abnormalities)
((a) No demand in presence of HDD;(b) Demand in absence of HDD)

In contrast, the opposing scenario is when there is the absence of HDD (temperature above set the temperature), but the presence of demand as seen in Figure 4-12 (b). As mentioned by As can be observed the presence of demand is centered in the hours where the occupants are expected to be home and active as per the occupancy model. Also, the majority of these demands are observed in the summer months where the baseload from SH is expected to be absent. Considering the observed distribution in both hourly and monthly resolution, the majority of this demand can be attributed to DHW. In addition, the aggregated demand in MWh for these time scenarios further justified the peak energy consumption to be during the occupant's high probabilistic active state. In early mornings and late nights, though the hour counts with demand are observed to be higher, the corresponding aggregated demand is significantly lower, thus further validating the above justification of the demand source being primarily DHW. Though Sarak & Satman, 2003 mention the absence of demand in the absence of HDD, this is not the case here as other non-temperature dependent components of the demand are also taken into consideration.

Since the load profile generator is a closed source model, detailed firsthand interaction and analysis were not possible during this study. Nevertheless, the above-given justification provides a realistic evaluation for the behavior of the curves and the randomness in the applied approach provides a better inclusion of behavioral aspects, whose mathematical replication would have been difficult. Though Kozarcanin et al., 2019 argue the possibility of determining the culture-specific heat demand behavior by studying the primary energy consumption, obtaining sufficient data for doing so on a high spatial resolution as planned by the methodology of the study becomes difficult. Also, the mathematical representation and inclusion of all these factors become difficult, thus a stochastic approach was implemented to ensure its maximum inclusion of these parameters. Hence the randomness seems to provide an adequate and desired consumer behavioral aspect of the heat demand profile.

Overall, this analysis could help qualitatively justify the aspects of the curve concerning realistic and expected heat demand patterns. However, for further improvement and detailed validation additional tests and comparisons were conducted.

HDD to h-factor

For the implemented IDP methodology, the interpretation of the effect of ambient temperature is slightly changed then the HDD, as building properties are also taken into consideration as described in section 3.1.5. A Pearson's correlation of 0.98 on a daily resolution and 0.93 on an hourly resolution was observed between the HDD and the h-factor curve. This slight difference is caused by the consideration of all these additional factors in the IDP methodology, and some mathematical changes implemented for better results. For instance, the IDP methodology makes use of a geometrical progression to calculate the temperature values as suggested by Hellwig, 2003 which is another reason for the deviation. Nevertheless, the h-factor can be stated as an improved version of HDD as a better and realistic relationship with the temperature can be replicated. The linear relationship in HDD is replaced by the sigmoid function. Even on consideration of older building classes of the correlation with HDD was not significantly different. Hence the approach of considering only the newer building classes to reduce complexity is justifiable.

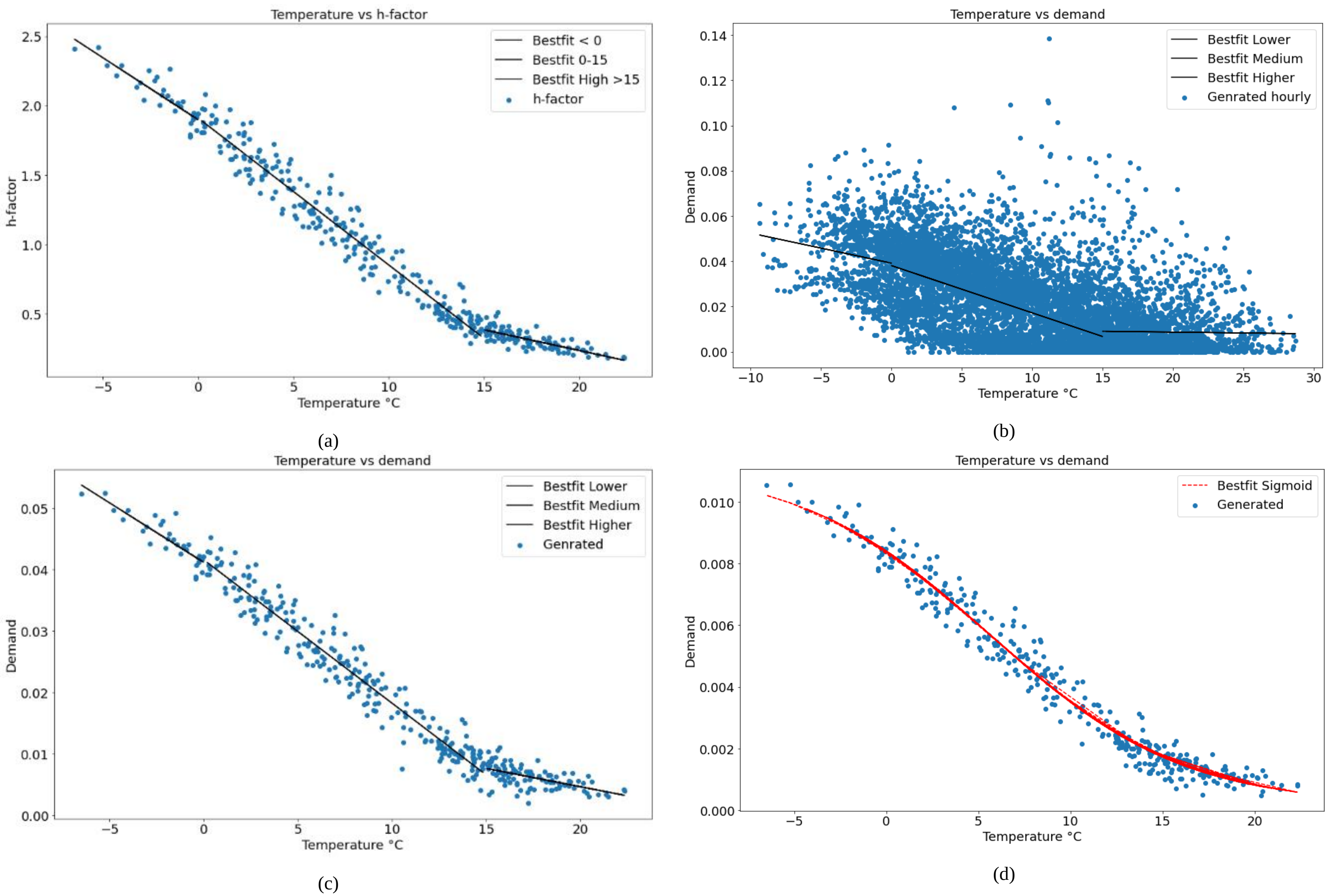
4.5.4 Demand-Temperature relationship

The best representation of the relationship between heating demand and temperature is a sigmoid curve (Hellwig, 2003) also implemented in (BDEW et al., 2020) with alterations to the equations. Hence for the generated demand profiles, the relationship with the temperature was plotted to compare the covariance and determine a best-fit equation.

Again, as with all other previous analyses, the use of average temperature data was avoided thus making the analysis only for census cell profiles. Firstly, the h-factor vs the temperature showed an anticipated result as seen in Figure 4-13(a) where the plot is divided into 3 sections based on the temperature represented by the 3 best-fit lines. The h-factor analysis was conducted only on a daily resolution considering the h-factor values are calculated based on daily average temperature values. A linear best fit line showed a high Pearson’s correlation in the middle-temperature range in between 0 to 15°C. In the extreme temperatures, the correlation was seen to be slightly lower. A better representation of the curves in these two endpoints would be with a sigmoid curve. Literature indicates a constant demand value on temperature extremes, an exceedance of 24°C in the summer, and temperatures below -5°C in the winter (Fallahnejad & Eberl, 2016). But such behavior of the curves could not be visualized as such extreme temperature values were not available in the sample data and are also seldom anticipated in the German temperature profiles (Fallahnejad & Eberl, 2016).

Comparing the demand with the temperature showed a lesser correlation especially on an hourly resolution. This is mainly due to the output obtained from the LPG generator which due to the presence of DHW provides random peaks which are unaffected by the temperature. Thus a low correlation with the best fit line is observed as seen in Figure 4-13(b).On the exclusion of the DHW demand, a much better best fit relation could be obtained thus showing the temperature and demand dependence for SH. Thus, it can be stated that implementing the desired variability in the demand profiles results in the decrease of the covariance between the demand and temperature in comparison to the desired results. Nevertheless, this degree of decreased covariance is acceptable, considering non-temperature dependent components of demand of DHW and CTS are also considered in the final profiles.

Table 4-5 below shows the summary of the calculated values of Pearson’s correlation between the h-factor/demand and the temperature. For a linear best fit line, also shown in Figure 4-13(b), the overall relationship is divided into 3 components based on the temperature value. As anticipated a higher correlation was obtained for the h-factor over demand since it is a direct output of the sigmoid function and demand is inclusive of DHW, CTS, and a further degree of randomness. This also validates the behavior



Source: Author

Figure 4-13: Demand vs Temperature
((a) Temperature vs. h-factor;(b) Temperature vs. Demand(hourly);(c) Temperature vs. Demand(daily); (d) Temperature vs. Demand(sigmoid bestfit)

seen in section 4.5.3 where demand is seen even in the absence of HDD. Since SH and temperature as seen to be very closely related, this abnormal behavior is only explainable by DHW.

Table 4-5: Temperature-Demand Covariance

Parameter	Resolution	Correlation
h-factor	Daily	Lower: - 0.88 Medium: - 0.97 Higher: - 0.77
Demand	Hourly	Lower: - 0.16 Medium: - 0.5 Higher: - 0.018
Demand	Daily	Lower: - -0.88 Medium: - 0.97 Higher: - 0.58

The obtained Pearson's correlation values closely match values presented in Eriksson, 2012 ranging between 0.88 to 0.92. Though Eriksson presents generation profiles, the behavior of the curves is expected to closely match the demand curves. Also, both the demand and generation of heat are closely linked to the temperature. To improve the results Eriksson, 2012, makes corrections to the profiles to better fit the sigmoid representation. But no further alterations have been done in this study considering the profiles have the presence of non-temperature dependent components as well. Thus, the underlying demand-temperature relationship for SH provides a means of validation to the generated output.

Finally, comparing paring the sigmoid best fit line to the of the OPSD and the IDP census level normalized profiles showed a high correlation of 0.97 on a daily resolution level. A high best fit sigmoid correlation of 0.96 was also seen in the hourly resolution of the best fit line. The hourly resolution also showed two distinctive profiles clusters which were further verified to be winter and summer demand. This hourly similarity was seen only in OPSD comparison.

All in all, the dependence of heat demand on temperature is significant. The results obtained from the above analysis illustrate the existence of this relationship in IDP profiles. The existing relationship is verified to be true considering similarities observed in reference profiles. Thus, this provides a strong validation with regards to the implemented methodology generating realistic demand profiles or at least one with the same level of accuracy as seen in existing studies.

5 Conclusion and Outlook

The vision of the German energy sector to attain carbon neutrality by 2050 requires robust planning. Electrification of all energy demand through sector coupling is seen as a major step towards achieving this vision. But to ensure a smooth transition, the planning phases require a detailed energy modelling study with regards to the analysis of the demand, generation, and capability of the supply grid. The nature of detail in energy modelling results is largely dependent on the granularity of the working data.

However, data availability on the required or desired granularity may be rare and difficult to acquire even in case of its availability. This holds truer especially in the case of heat demand data where the practice of user measurement is uncommon. Thus, data availability is only limited to the annual total demand, which is not sufficient if the analysis is regarding the electrification of these demands. Though the heating utilities are obliged to keep track of their supply, publishing these data is not mandatory. This leaves a large gap in the availability of high granularity data in the heating sector. In the absence of the measured data, a methodology with the capability to correctly estimate the demand profiles becomes important. Hence this study focuses on filling this gap in data availability to ensure open-source access to the high-resolution heating demand profiles for Germany.

For the thesis, the study in collaboration with eGon research project aimed at a granularity of 100 X 100 m^2 high spatial resolution and an hourly temporal resolution. Though the data available on the required granularity level is absent, methodologies and studies regarding the estimation of the heat demand profile are not new. On the contrary, studies conducted in the German heating sector are seen as a pioneer. A common practice of the application of the German methodologies for the estimation of heating demand in other European countries was also seen. In addition, existing open-source databases were also identified over the course of the study which provides heat demand profiles on some spatial aggregation level, in an hourly temporal resolution. These methodologies make use of the currently existing state-of-the-art Standard Load Profiles (SLP) or some form minor alterations to it. However, the SLP for heating demand is developed based on the natural gas supply, and though it gives validated realistic results in aggregated spatial resolution, the output is further away from reality in a high spatial resolution, primarily due to their inability to represent DHW which is a direct cause of peaks in the profile. This may not have a significant impact in the case of non-electric supply, but with the planned electrification of the sector, the correct acknowledgment of these peaks becomes especially important in terms of sufficiency in supply and capacity of the grid. Hence this deems the use of the existing SLPs unfeasible for such energy modelling analysis and thus making high variability heat demand profiles database as proposed by this study a necessity.

The thesis was broadly classified to cover two major objectives. The primary being the generation of the said profiles. For this a new methodology referred to as the Intra Day Profile (IDP) was developed and implemented. The study conducted under the project was principally designed to create a database that fits the requirements of the project's energy modelling tool. However, the database is also accessible for other applications. Secondly, the validation of these results in terms of accuracy and its correct representation of a realistic heat demand behavior was necessary. This was conducted through the implementation of numerous statistical measures primarily comparing it to other reference profiles.

Some of the salient features of the generated profiles are as follows:

- The final heat demand database is available in two future scenarios for 2035 and 2050. For ease in storage, the profiles have been aggregated into two categories of potential district heating and individual supply.
- A large variability with regards to the hourly resolution of each indivudal profile is noticeable. This is due to the underlying stochastic bottom-up approach for the generation of the profiles further randomized by the selection process of individual day profiles. This enables the generation of unique profiles per census cell.
- The overall trend of peakier profiles is observed compared to much smoother ones developed from other established methodologies. These peaks in the individual census profiles provide a better representation of the DHW demand. Thus, the profiles are arguably more realistic.
- The relationship between the demand and ambient temperature for the residential sector is represented by a sigmoid curve, truer on a daily resolution than on a highly variable hourly resolution.
- The profiles both on individual and aggregation levels show autocorrelative demand behavior. The demand of any given hour is depedendent on the corresponding hour of preceding 3-4 days with statistical significance, depending on the temperature patterns of the location.
- The level of aggregation is inversely proportional to the variability offered by the profiles. As the aggregation increases the variability tends to be closer to that offered by the existing SLPs.
- A justifiable behavior of summer demand dominated by DHW peaks and winter demand explainable in terms of the implementation of building thermal energy storage provides logical explanations to the generated profile patterns.
- The winter demand in the generated profiles is much closer to that in the reference profiles considering the presence of a continuous baseload which is common for all BDEW based profiles.

The statistical validation and comparison of the generated profiles showed the following results:

- The similarity in the nature of the curve is statistically verified on all aggregation levels. However, cell-level generated profiles showed a tendency of non-stationarity behavior induced due to the fluctuating variations in the individual profiles introduced by randomness.
- A consistent weather seasonality pattern was observed in profiles that closely follow the ambient temperature patterns. This on one hand provides validation considering the temperature dependence of the demand profiles and further shows the dominance of SH demand on the overall census demand.
- A Pearson's correlation test showed a high correlation between the generated profiles on a daily resolution and slightly lower on an hourly resolution. A Chow break test showed structural similarity between the generated profiles and OPSD reference profiles on an aggregated level.
- The least overall cost of profile alignment was observed with OPSD on the national level, with cost inversely proportional to the degree of aggregation. The DIW profiles were concluded to be considerably different from the generated results.

Even with a higher degree of variability on a high spatial resolution, a consistent and controlled variability was observed with increased aggregation. Hence it can be concluded that the generated profiles show potential for replacing the existing SLPs as detailed demand fluctuations are better interpreted in this methodology compared to the existing state of the art. Nevertheless, considering the similarity in variability between IDP and reference profiles once profiles are aggregated, the use of either, on a national level or a regional level aggregation for energy modelling is expected to generate similar results. However, for studies dealing with high spatial resolution, as proposed by eGon for making use of census level profiles for grid optimization, the analysis with IDP is expected to provide higher details and accurate results. Hence when dealing with high-resolution energy modelling analysis the replacement of SLPs with the IDP generated profiles would be recommended.

Finally, potential measures for improvements were identified with regards to the obtained results. However, these could not be implemented in the study. Prior to the initiation of the study a third objective regarding the application of the generated profiles in the eTrago energy modelling tool had also been proposed. However due to the time constraints and unprecedented issues with the modelling tool this objective could not be fulfilled. Hence as a continuation of the work initiated in this thesis, the application of the generated profiles could provide better insight into the usability of the profiles and the impact this additional load has on grid optimization in terms of grid extension and associated costs. Successful usability would provide means of further validation and confirm the statistical results. As a final measure possible alterations to the LPG model can be performed to include newer building classes which are also expected to provide improvements to the results. Also identification and implementation of improved data sources representing the CTS sector providing a higher degree of variability can be done to ensure the same level of granularity as in the residential sector

All in all, the study was able to generate demand profiles on a high spatial resolution of 100 X 100m^2 (census cell) for an hourly temporal resolution. The stochastic bottom-up approach for estimation of the hourly demand values enabled the generation of non-identical profiles for individual cells with a large degree of hourly variability. This high inter and intra profile variability is expected to provide a higher degree of detail and accuracy to results during the energy modelling application of these demand time-series. The results were verified and are available as open-source data for future applications.

6 References

Aragon, V., Gauthier, S., Warren, P., James, P. A. B., & Anderson, B. (2019). Developing English domestic occupancy profiles. *Building Research and Information*, *47*(4), 375–393. https://doi.org/10.1080/09613218.2017.1399719

Athiyarath, S., Paul, M., & Krishnaswamy, S. (2020). A Comparative Study and Analysis of Time Series Forecasting Techniques. *SN Computer Science*, *1*(3), 1–7. https://doi.org/10.1007/s42979-020-00180-5

BDEW, VKU, & GEODE. (2020). *BDEW / VKU / GEODE- Leitfaden Abwicklung von Standardlastprofilen Gas*. https://www.bdew.de/media/documents/20200331_KoV_XI_LF_SLP_Gas_clean_final.pdf. abgerufen am 10.12.2020

Berger, M., & Worlitschek, J. (2018). A novel approach for estimating residential space heating demand. *Energy*, *159*, 294–301. https://doi.org/10.1016/j.energy.2018.06.138

BizEE Energy Lens. (2021). *Degree Days - Handle with Care*. https://www.energylens.com/articles/degree-days

Brownlee, J. (2016). *How to Check if Time Series Data is Stationary with Python*. Machine Learning Mastery. https://machinelearningmastery.com/time-series-data-stationary-python/

CDS Climate Data. (2018). *ERA5 hourly data*. https://doi.org/https://doi.org/10.24381/cds.adbb2d47

Chen, C., Twycross, J., & Garibaldi, J. M. (2017). *A new accuracy measure based on bounded relative error for time series forecasting*. 1–23.

Chramcov, B. (1982). Forecast model of heat demand. *Forecast*, 1–9.

Clegg, S., & Mancarella, P. (2019). Integrated electricity-heat-gas modelling and assessment, with applications to the Great Britain system. Part I: High-resolution spatial and temporal heat demand modelling. *Energy*, *184*, 180–190. https://doi.org/10.1016/j.energy.2018.02.079

Çomakli, K., Yüksel, B., & Çomakli, Ö. (2004). Evaluation of energy and exergy losses in district heating network. *Applied Thermal Engineering*, *24*(7), 1009–1017. https://doi.org/10.1016/j.applthermaleng.2003.11.014

Couto, A., & Estanqueiro, A. (2020). Exploring wind and solar PV generation complementarity to meet electricity demand. *Energies*, *13*(6). https://doi.org/10.3390/en13164132

D. Peters, R. Völker, F. S. and K. von M. (2020). Are standard load profiles suitable for modern electricity grid models? *17th International Conference on the European Energy Market (EEM), 2020, Pp. 1-6*. https://doi.org/10.1109/EEM49802.2020.9221967

DESTATIS Statistisches Bundesamt. (2014). *Zensus 2011*. https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html

DIW Berlin. (2017). *Electricty, Heat and Gas Sector Data for Modeling the German System*.

Drauz, S. R. (2016). *Master Thesis Synthesis of a heat and electrical load profile for single and multi-family houses used for subsequent performance tests of a multi-component energy system*.

ECMFW. (2020). *ERA5 Datasets*. https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5

Eriksen Freja. (2020). *The heat is on to make German buildings "nearly" climate-neutral*.

https://www.cleanenergywire.org/news/heat-make-german-buildings-nearly-climate-neutral

Eriksson, N. (2012). Predicting demand in district heating systems. *Uppsala University*, 43. http://www.diva-portal.org/smash/get/diva2:530099/FULLTEXT01.pdf

EUF. (2020). *eGon. Open cross-network level and cross-sector planning tool to determine the optional use and expansion of flexibility options in Germany*. https://www.uni-flensburg.de/eum/forschung/laufende-projekte/egon/

eurostat. (2021). *Household composition statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Household_composition_statistics

Fallahnejad, M. (2017). *Long - Term Forecast of Residential & Commercial Gas Demand in Germany*. *March*.

Fallahnejad, M., & Eberl, B. (2016). *Creating Standard Load Profiles in Residential and Commercial Sectors in Germany for 2016, 2025 and 2040*. 8.

Federal Office for Building and Regional Planning (BBR). (2014). *Manual Test reference years of Germany for medium , extreme and future weather conditions*. *September*.

Felten, B., Baginski, J. P., & Weber, C. (2017). *KWK-Mindest- und Maximaleinspeisung - Die Erzeugung von Zeitreihen für die Energiesystemmodellierung*. *10*.

Fischer, D., Wolf, T., Scherer, J., & Wille-Haussmann, B. (2016). A stochastic bottom-up model for space heating and domestic hot water load profiles for German households. *Energy and Buildings*, *124*, 120–128. https://doi.org/10.1016/j.enbuild.2016.04.069

Fleischer, C. E. (2020). Minimising the effects of spatial scale reduction on power system models. *Energy Strategy Reviews*, *32*(September), 100563. https://doi.org/10.1016/j.esr.2020.100563

Fleischer, C. E. (2021). A data processing approach with built-in spatial resolution reduction methods to construct energy system models. *Open Research Europe*, *1*, 36. https://doi.org/10.12688/openreseurope.13420.1

Fleiter, T., Manz, P., Neuwirth, M., Mildner, F., Persson, U., Kermeli, K., Crijns-Graus, W., & Rutten, C. (2020). *Quantification of synergies between energy efficiency first. Excess heat potentials of industrial sites in Europe*.

Fleiter Tobias; Rehfeldt Matthias. (2018). *A methodology for bottom-up modelling of energy transitions in the industry sector: The FORECAST model*. https://doi.org/https://doi.org/10.1016/j.esr.2018.09.005

Giannakopoulos, C., & Psiloglou, B. E. (2006). Trends in energy load demand for Athens, Greece: Weather and non-weather related factors. *Climate Research*, *31*(1), 97–108. https://doi.org/10.3354/cr031097

Gotzens, F., Gillessen, B., Burges, S., & Hennings, W. (2020). *DemandRegio Harmonisierung und Entwicklung von Verfahren zur regionalen und zeitlichen Auflösung von Energienachfragen Abschlussbericht Vorwort*.

Hellwig, M. (2003). *Entwicklung und Anwendung parametrisierter*.

Hinterstocker, M., Eberl, B., & von Roon, S. (2015). *Weiterentwicklung des Standardlastprofilverfahrens Gas*.

Hzndman, R. (2011). *Cyclic and Seasonal Time Series*. https://robjhyndman.com/hyndsight/cyclicts/

Idowu, S., Saguna, S., Åhlund, C., & Schelén, O. (2015). Forecasting heat load for smart district heating systems: A machine learning approach. *2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014*, 554–559. https://doi.org/10.1109/SmartGridComm.2014.7007705

Iordanova, T. (2020). *An Introduction to Stationary and Non-Stationary Processes*. Investopedia. https://www.investopedia.com/articles/trading/07/stationary.asp

IRENA. (2020). *Energy Transition*. https://www.irena.org/energytransition

IWU. (2015). *Deutsche Wohngebäudetypologie zur Verbesserung der Energieeffizienz*.

Jordan, U., & Vajen, K. (2001). Realistic domestic hot-water profiles in different time scales. *Report for Solar Heating and Cooling Program of the International Energy Agency (IEA-SHC)Task*, *26*, 1–18. http://sel.me.wisc.edu/trnsys/trnlib/iea-shc-task26/iea-shc-task26-load-profiles-description-jordan.pdf

Kadian, R., Dahiya, R. P., & Garg, H. P. (2007). Energy-related emissions and mitigation opportunities from the household sector in Delhi. *Energy Policy*, *35*(12), 6195–6211. https://doi.org/10.1016/j.enpol.2007.07.014

Kaminska, A. (2019). *Impact of Heating COntrol Strategy and Occupant Behavior on the Energy Consumption in a Building with Natural Ventilation in Poland*. https://doi.org/https://doi.org/10.3390/en12224304

Kannan, R. (2018). Dynamics of long-term electricity demand profile: Insights from the analysis of Swiss energy systems. *Energy Strategy Reviews*, *22*(September 2017), 410–425. https://doi.org/10.1016/j.esr.2018.10.010

Kleinertz, B., Samweber, F., Mueller, M., Schifflechner, C., & Hinterstocker, M. (2017). *Development of realistic energy demand profiles for estimation of load flexibility of households for the integration of PV generation The Research Center for Energy Economics (FfE)*.

Kneiske, T. M., & Drauz, S. R. (2017). *Öffentlicher Abschlussbericht Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums*. *June 2019*. https://doi.org/10.13140/RG.2.2.15341.03045

Kozarcanin, S., Andresen, G. B., & Staffell, I. (2019). Estimating country-specific space heating threshold temperatures from national gas and electricity consumption data. *Energy and Buildings*, *199*(April), 368–380. https://doi.org/10.1016/j.enbuild.2019.07.013

Krähenmann, S., Walter, A., Brienen, S., Imbery, F., & Matzarakis, A. (2018). High-resolution grids of hourly meteorological variables for Germany. *Theoretical and Applied Climatology*, *131*(3–4), 899–926. https://doi.org/10.1007/s00704-016-2003-7

Kuru, M., & Calis, G. (2019). *Forecasting Heating Degree Days for Energy Demand Modeling*. *January*, 713–718. https://doi.org/10.3311/ccc2019-097

Louvet, Y., Orozaliev, J., & Vajen, K. (2019). *Measurement Evaluation and Simulation of an Innovative Drainback Solar Combi-System*. 1–12. https://doi.org/10.18086/eurosun2018.01.05

Maruf, M. N. I. (2021). A novel method for analyzing highly renewable and sector-coupled subnational energy systems-case study of schleswig-holstein. *Sustainability (Switzerland)*, *13*(7). https://doi.org/10.3390/su13073852

Muhammad, H. D. (2017). The energy audit process for universities accommodation in Malaysia: A preliminary study. *IOP Conference Series: Earth and Environmental Science*, *67*(1). https://doi.org/10.1088/1755-1315/67/1/012027

Oemof Developing Group. (2016). *Demandlib*. https://demandlib.readthedocs.io/en/latest/description.html

Open Energy Modelling Initiative. (2021). *Thermal demand profiles*. https://forum.openmod.org/t/thermal-demand-profiles/749

Open Power System Data (OPSD). (2020). *When2Heat Heating Profiles*. https://doi.org/https://doi.org/10.25832/when2heat/2019-08-06

openego. (2021a). *egon-district heating areas*. https://github.com/openego/eGon-data/blob/dev/src/egon/data/processing/district_heating_areas/__init__.py

openego. (2021b). *egon/data/importing/heat_demand_data*. https://github.com/openego/eGon-data/blob/dev/src/egon/data/importing/heat_demand_data/__init__.py

Openego. (2021). *High and Medium voltage substation*. https://github.com/openego/eGon-data/blob/dev/src/egon/data/processing/substation/hvmv_substation.sql

Quayle Robert G. (1979). Heating Degree Day Data Applied to Residential Heating Energy Conumption. *National Climatic Centre*.

Reiner Lemoine Institut. (2015). *eGon*. https://reiner-lemoine-institut.de/en/egon/

Richardson, I., Thomson, M., & Infield, D. (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings*, *40*(8), 1560–1566. https://doi.org/10.1016/j.enbuild.2008.02.006

Ruf, H., Kober, P., Meier, F., Irlbeck, M., & Heilscher, G. (2016). *Comparison of the synthetic reference load profile according to VDI 4655 with high-resolution measurement data from a modern house in Ulm*. *March*.

Ruhnau, O., Hirth, L., & Praktiknjo, A. (2019). Time series of heat demand and heat pump efficiency for energy system modeling. *Scientific Data*, *6*(1), 1–11. https://doi.org/10.1038/s41597-019-0199-y

Sarak, H., & Satman, A. (2003). The degree-day method to estimate the residential heating natural gas consumption in Turkey: A case study. *Energy*, *28*(9), 929–939. https://doi.org/10.1016/S0360-5442(03)00035-5

Schüler, N., Mastrucci, A., Bertrand, A., Page, J., & Maréchal, F. (2015). Heat demand estimation for different building types at regional scale considering building parameters and urban topography. *Energy Procedia*, *78*, 3403–3409. https://doi.org/10.1016/j.egypro.2015.11.758

SEEnergies. (2020). *QUANTIFICATION OF SYNERGIES BETWEEN ENERGY EFFICIENCY FIRST*. *846463*.

sEEnergies Open Data. (2020). *D5 1 DIstrict Heating Areas*. https://s-eenergies-open-data-euf.hub.arcgis.com/datasets/b62b8ad79f0e4ae38f032ad6aadb91a0_0/explore

Serrà, J., & Arcos, J. L. (2014). An empirical evaluation of similarity measures for time series classification. *Knowledge-Based Systems*, *67*, 305–314. https://doi.org/10.1016/j.knosys.2014.04.035

Statistisches Bundesamt. (2016). *Qualitätsbericht - Zeitverwendungserhebung ZVE 2012/2013*. *49*(0), 88. https://www.destatis.de/DE/Methoden/Qualitaet/Qualitaetsberichte/Einkommen-Konsum-

Lebensbedingungen/zeitverwendungserhebung-2012-2013.html

Stegner, C., Glaß, O., & Beikircher, T. (2019). Comparing smart metered, residential power demand with standard load profiles. *Sustainable Energy, Grids and Networks*, *20*, 1–10. https://doi.org/10.1016/j.segan.2019.100248

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, *13*(8), 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033

Tang, L., Wang, C., & Wang, S. (2013). Energy time series data analysis based on a novel integrated data characteristic testing approach. *Procedia Computer Science*, *17*, 759–769. https://doi.org/10.1016/j.procs.2013.05.098

Terehovics, E., Veidenbergs, I., & Blumberga, D. (2017). Exergy Analysis for District Heating Network. *Energy Procedia*, *113*, 189–193. https://doi.org/10.1016/j.egypro.2017.04.053

Umwelt Bundesamt. (2018). *Primary Energy Consumption*. https://www.umweltbundesamt.de/daten/energie/primaerenergieverbrauch#definition-und-einflussfaktoren

Zeyen, E., Hagenmeyer, V., & Brown, T. (2021). Mitigating heat demand peaks in buildings in a highly renewable European energy system. *Energy*, *231*(December). https://doi.org/10.1016/j.energy.2021.120784

ANNEX

Annex A: Demanlib building typology classification

	parameter_a	parameter_b	parameter_c	parameter_d	building_class	shlp_type
0	3.2279446	-37.42148	6.2222288	0.0828441	1	SFH
1	3.2107659	-37.41788	6.2024	0.0865217	2	SFH
2	3.1935978	-37.414248	6.1824021	0.0901957	3	SFH
3	3.1764404	-37.410583	6.1622336	0.0938662	4	SFH
4	3.159294	-37.406886	6.1418926	0.0975329	5	SFH
5	3.1421588	-37.403156	6.1213773	0.1011959	6	SFH
6	3.1250349	-37.399394	6.1006859	0.104855	7	SFH
7	3.1079226	-37.395599	6.0798166	0.1085103	8	SFH
8	3.0908222	-37.391772	6.0587677	0.1121617	9	SFH
9	3.0737337	-37.387913	6.0375374	0.1158091	10	SFH
10	3.1850191	-37.412416	6.1723179	0.0920309	11	SFH
11	2.5736652	-35.016944	6.131814	0.0996851	1	MFH
12	2.5516882	-35.023422	6.1680699	0.108708	2	MFH
13	2.529738	-35.030015	6.2051109	0.1177216	3	MFH
14	2.507817	-35.036736	6.2430159	0.1267238	4	MFH
15	2.4859161	-35.043598	6.2818214	0.1357193	5	MFH
16	2.4640414	-35.050587	6.321514	0.1447056	6	MFH
17	2.4421941	-35.057708	6.3621285	0.153682	7	MFH
18	2.4203748	-35.064962	6.4036973	0.1626484	8	MFH
19	2.398584	-35.072352	6.446253	0.1716044	9	MFH
20	2.3768224	-35.079882	6.4898289	0.1805499	10	MFH
21	2.5187775	-35.033375	6.2240634	0.1222227	11	MFH

Source: (Oemof Developing Group, 2016)

Annex B: IWU Building Typology

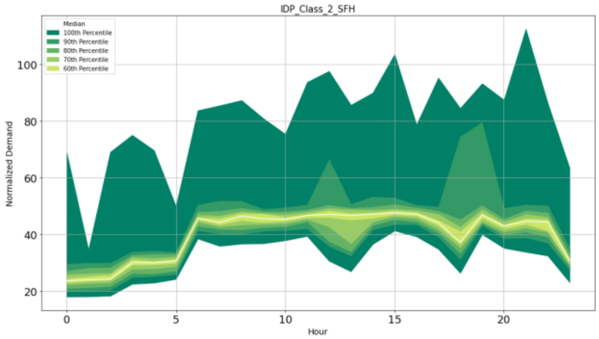
analysis of the housing stock 2011		housing age										sum	mean
		till 1860	1861-1918	1919-1948	1949-1957	1958-1968	1969-1978	1979-1983	1984-1994	1995-2001	2002-2009		
		A	B	C	D	E	F	G	H	I	J		
SFH	number of houses in thousand	330	966	1131	859	1509	1507	704	1160	1035	775	9976	997.6
SFH	number of apartments in thousand	399	1213	1389	1060	1948	1915	881	1397	1204	858	12263	1226.4
SFH	area in million m^2	46	135	150	116	218	233	110	178	158	119	1463	146.3
SFH	ratio area/apartment	115	111	108	109	112	122	125	127	131	139	-	120
MFH	number of houses in thousand	54	442	388	356	586	412	146	309	244	85	3023	302.3
MFH	number of apartments in thousand	214	2177	1911	2003	3348	2313	852	1826	1390	461	16495	1649.5
MFH	area in million. m^2	16	163	129	125	225	169	64	133	104	39	1168	116.7
MFH	ratio apartments/houses	4	5	5	6	6	6	6	6	6	5	-	5.5
number of houses in thousand		384	1438	1519	1215	2095	1919	850	1469	1279	860	18239	-
number of apartments in thousand		613	3390	3300	3063	5296	4228	1733	3223	2594	1319	39228	-
area in million m^2		62	298	279	241	443	392	174	311	262	158	2631	-

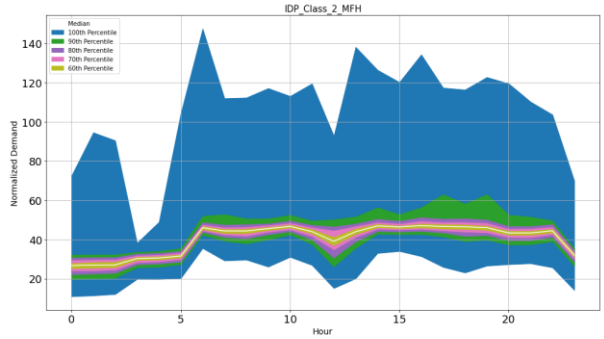
Source: (Drauz, 2016)

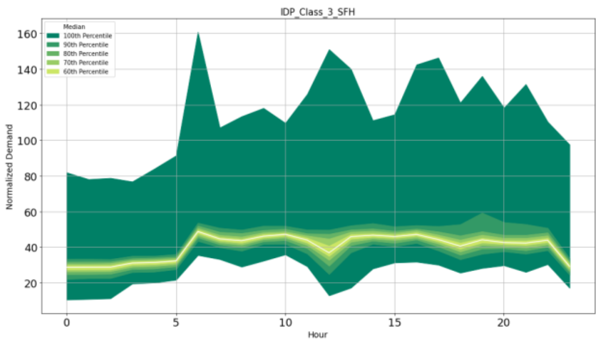
Annex C: LPG profile behavior-Summary

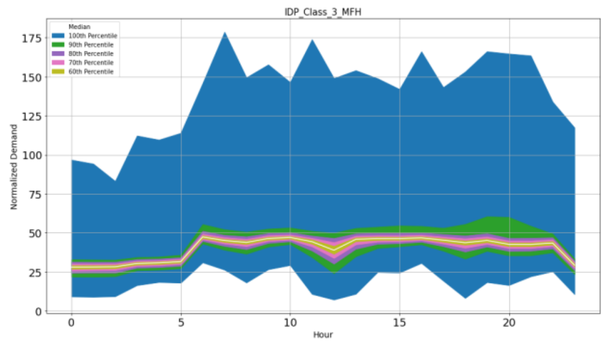
	SH	DHW
Total Demand	• Higher Demand in Lower building classes • Inverse effect of occupant number	• Effect of the household typology is almost completely absent • Greatly affected by the occupant number
Peak	• Higher Demand in Lower building classes • The effect of the occupant number is negligible	• Effect of the household typology is almost completely absent • Greatly affected by the occupant number
Variance	• Effect of the household typology is insignificant • Effect is proportional to the occupant number	• Effect of the household typology is almost completely absent • Inversely proportional to the occupant number

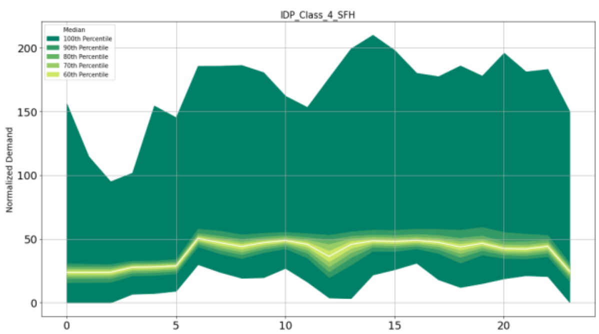
Annex D: Variability observed in IDP Pool

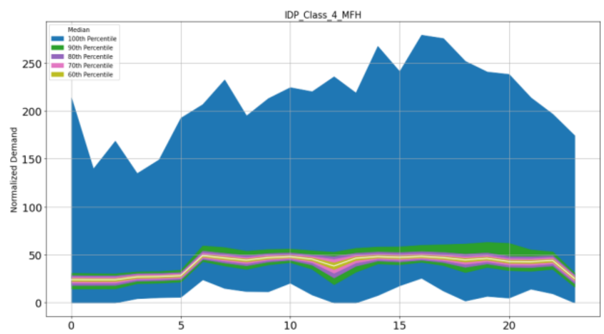


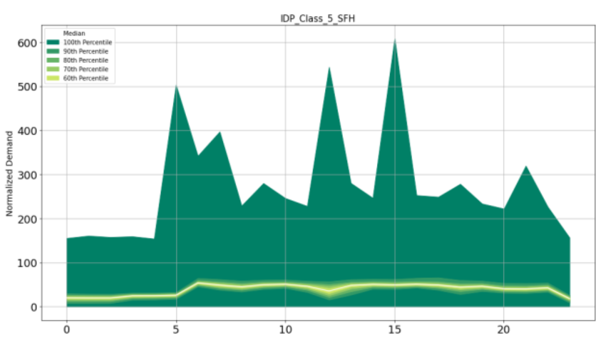


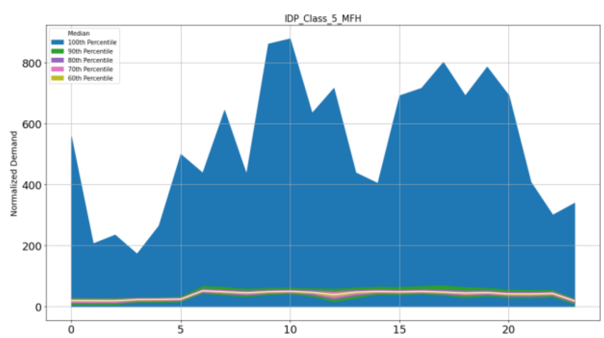


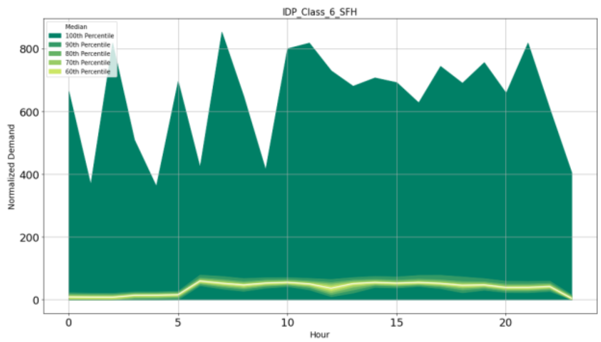


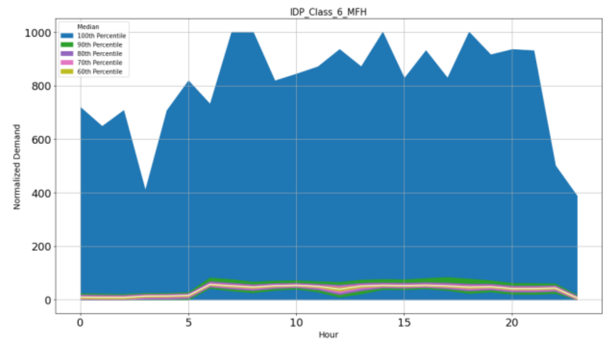


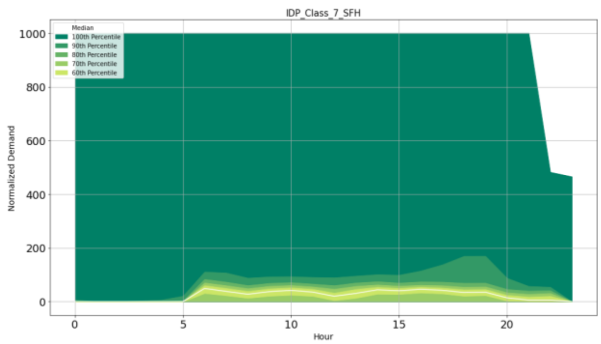


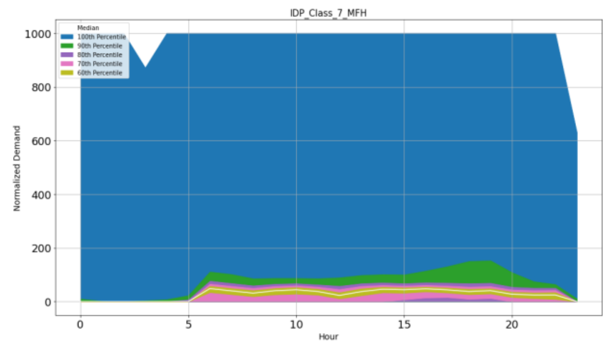


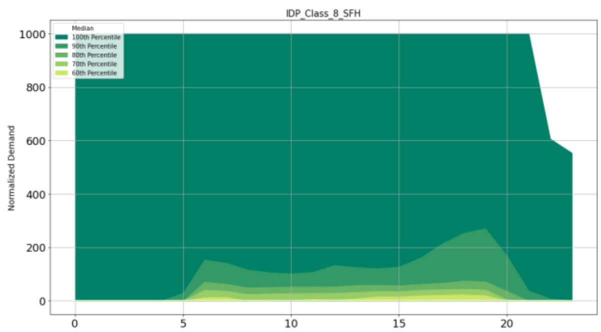


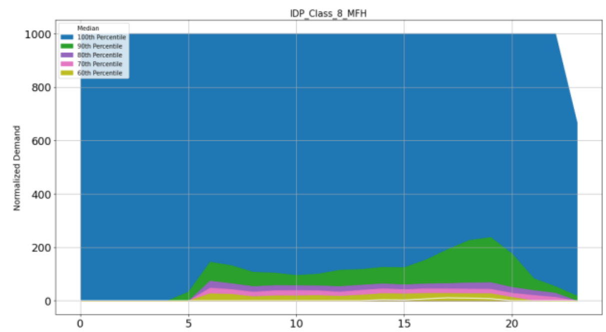

IDP_Class_6_SFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
0
5
10
15
20

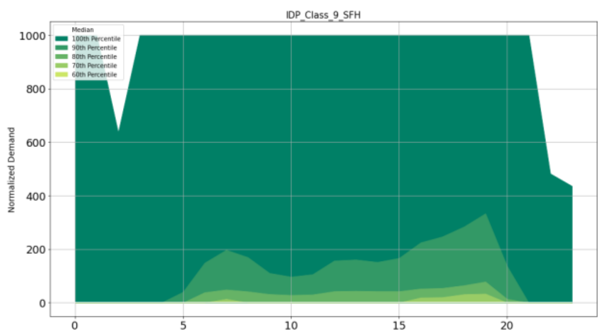

IDP_Class_6_MFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20

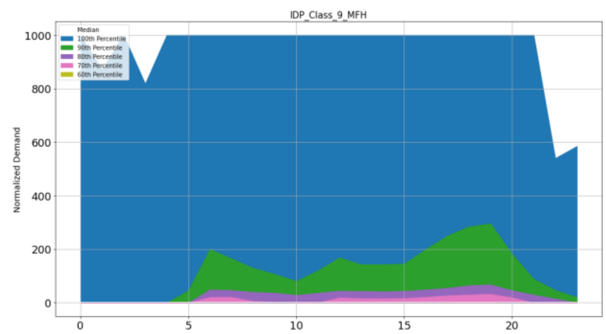

IDP_Class_7_SFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20

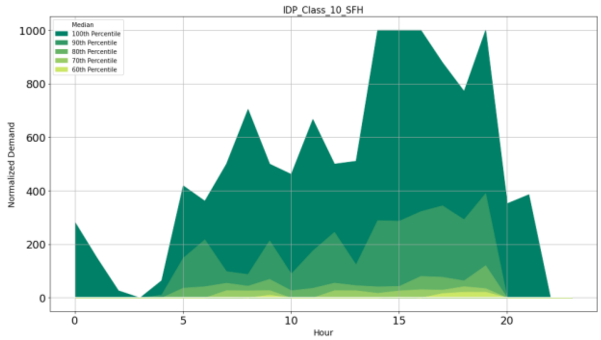

IDP_Class_7_MFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20

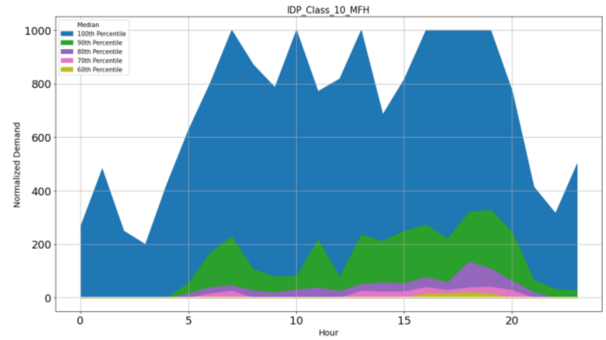

IDP_Class_8_SFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20


IDP_Class_8_MFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20


IDP_Class_9_SFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20


IDP_Class_9_MFH
Median
100th Percentile
90th Percentile
80th Percentile
70th Percentile
60th Percentile
Normalized Demand
Hour
0
200
400
600
800
1000
0
5
10
15
20

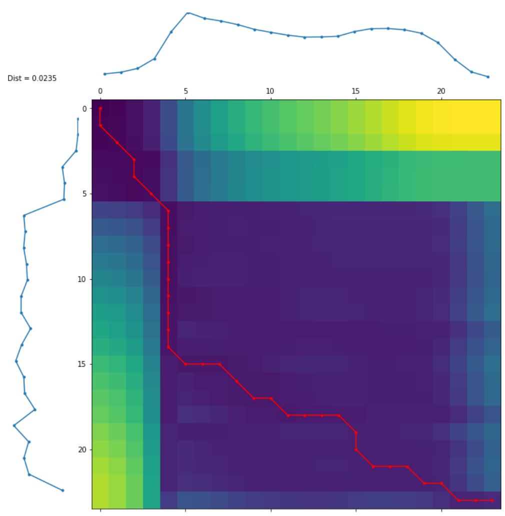


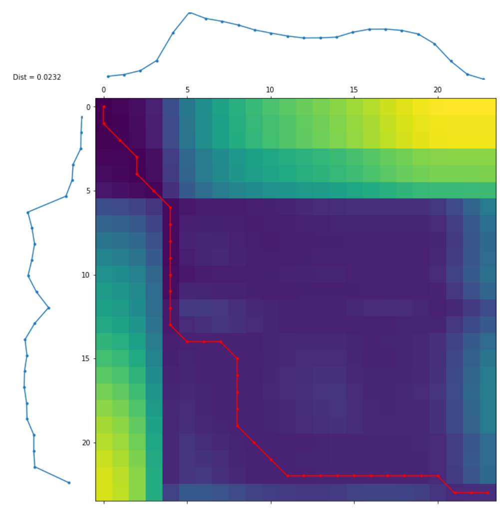


Annex E: IDP Pool comparison with BDEW profiles

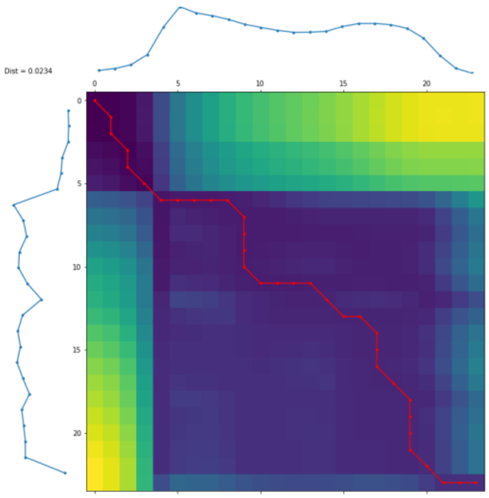
Class SFH MFH

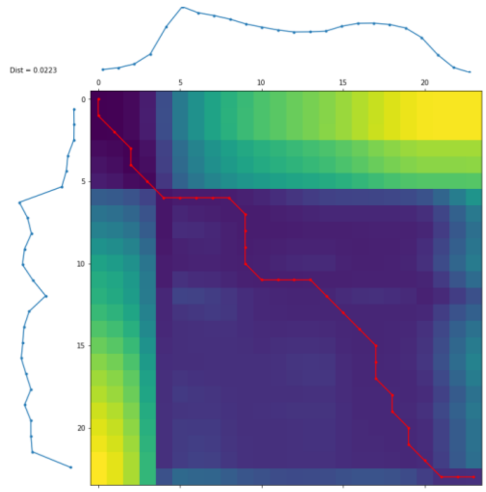
Class2



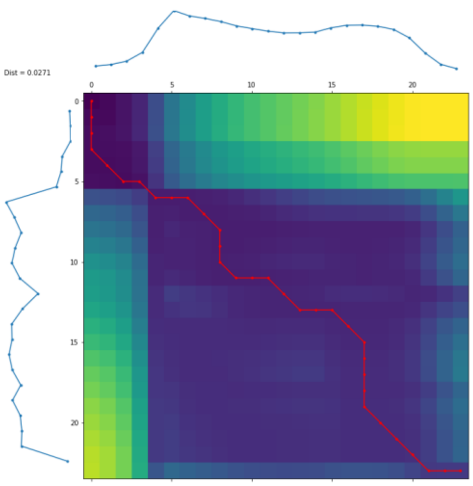


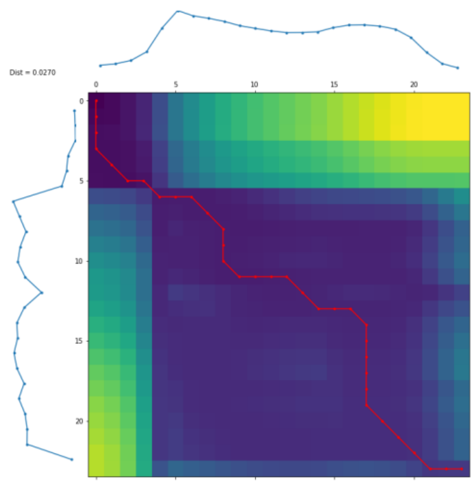
Class 3



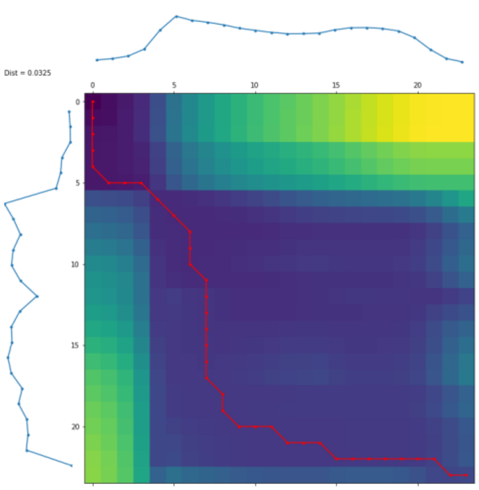


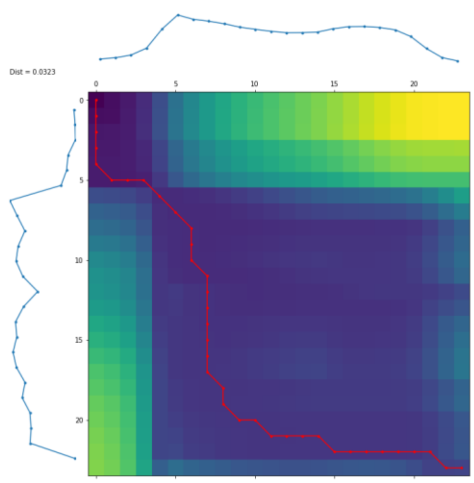
Class 4



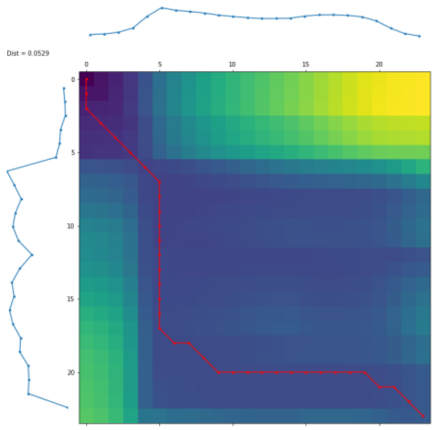


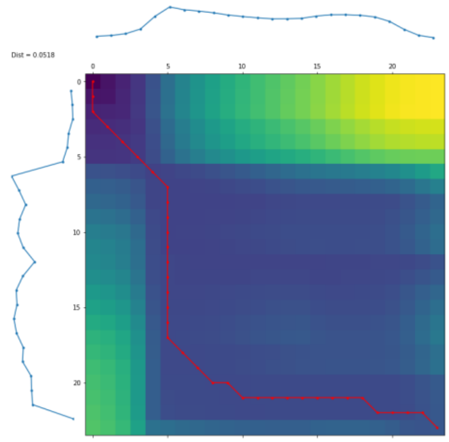
Class 5



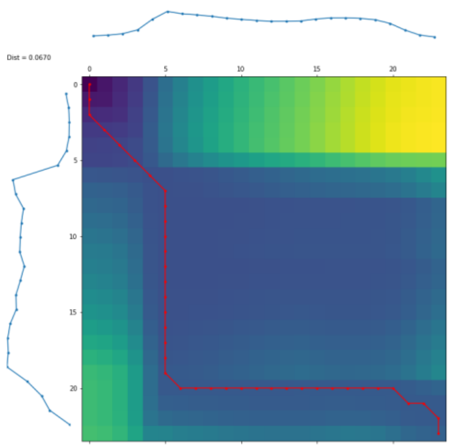


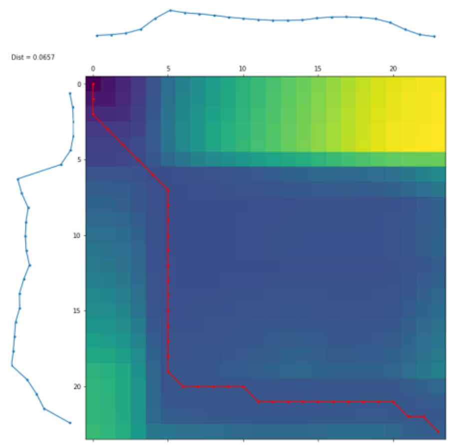
Class 6



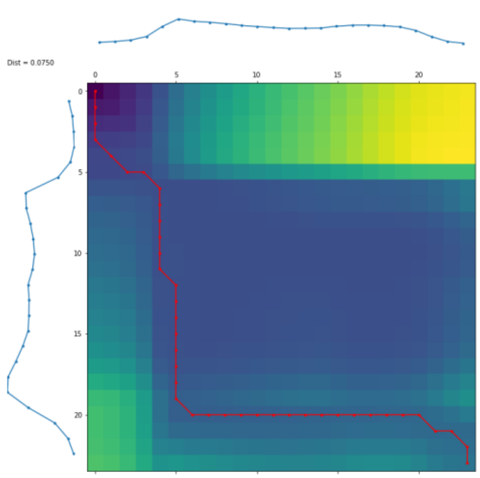


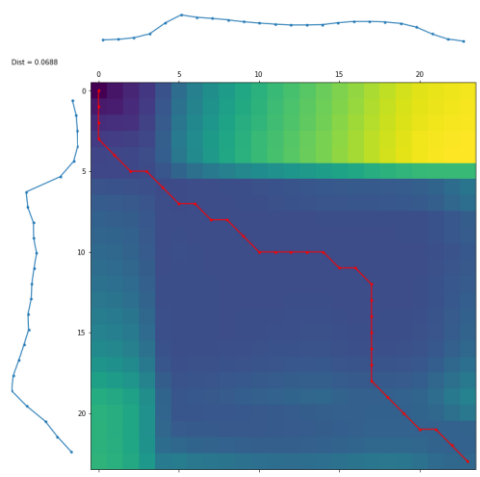
Class 7



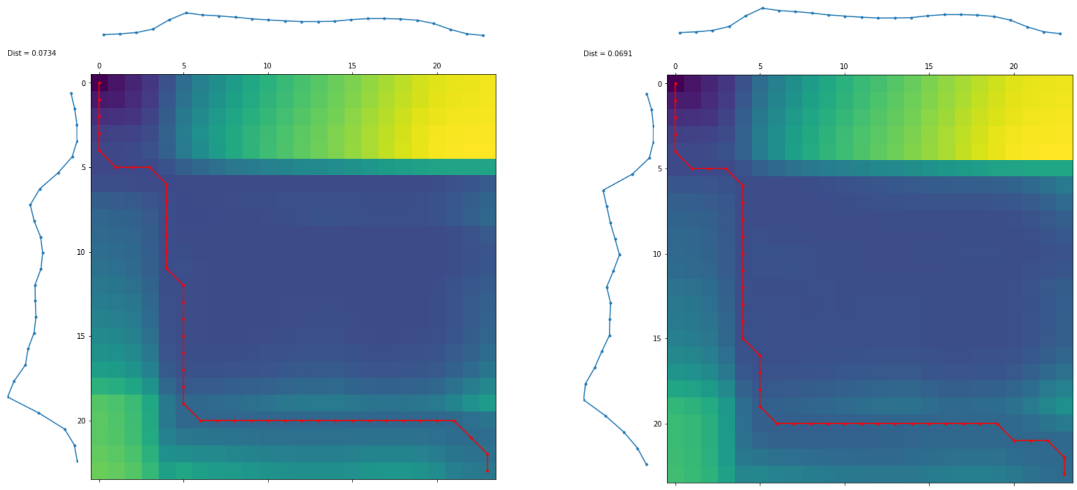


Class 8

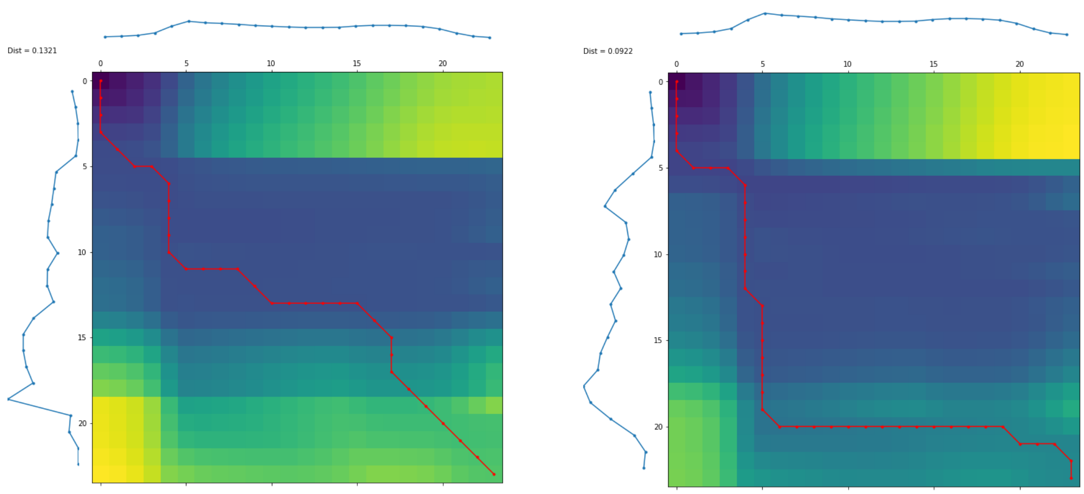




Class 9



Class 10



Declaration

I hereby expressly declare that I have prepared this work on my own using no sources, aids or resources other than those cited in it. In particular, I expressly affirm that I have not used any services or received support of any kind, paid or unpaid, from ghost-writer agencies, comparable service providers, or other third parties. All text passages cited or borrowed (either verbatim or in spirit) from printed, electronic or other sources have been duly acknowledged by me. I am aware that violations of this policy may result in a grade of "Insufficient/Fail" (5.0) with respect to my submitted work, and in more serious cases could lead to further measures by Europa-Universität Flensburg including my possible ex-matriculation from the university. I am aware of and agree to the fact that this text can be digitally stored and checked or scanned using anti-plagiarism software.

Place, date: *Signature*