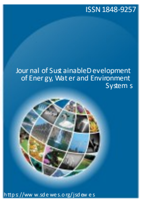
Journal of Sustainable Development of Energy, Water and Environment Systems

http://www.sdewes.org/jsdewes



Year 2026, Volume 14, Issue 4, 1140756

Original Research Article

Comparison of Data Aggregation Utilising a State-Provided Urban Building Energy Model and Actual Consumption Data for the City of Krefeld, Germany

Laura Schriefers*, Thomas Leidenbach, Arne Graßmann, Frank Alsmeyer
SWK E² - Institute for Energy Technology and Energy Management
Hochschule Niederrhein University of Applied Sciences, Reinarzstraße 49, 47805 Krefeld
e-mail: laura.schriefers@hs-niederrhein.de, thomas.leidenbach@hs-niederrhein.de

Cite as: Schriefers, L., Leidenbach, T., Graßmann, A., Alsmeyer, F., Comparison of Data Aggregation Utilising a State-Provided Urban Building Energy Model and Actual Consumption Data for the City of Krefeld, Germany, J.sustain. dev. energy water environ. syst., 14(4), 1140756, 2026, DOI: https://doi.org/10.13044/j.sdewes.d14.0756

ABSTRACT

Meeting climate targets requires sustainable energy systems and proactive municipal heat planning. This paper aims to investigate how different spatial aggregation levels impact the accuracy of heat demand prediction. For this purpose, the energy demand calculated by a state-provided and publicly available urban building energy model is compared with a comprehensive high-resolution consumption dataset from the city of Krefeld in Germany. The latter includes detailed consumption data for district heating, natural gas, heat pumps, and night storage heaters. Box plot diagrams and statistical performance indicators are applied to evaluate the precision of various spatial aggregation levels. The results demonstrate that aggregations at the level of postcode areas, statistical districts and cadastral sectors can provide a reliable foundation for planning purposes. Aggregation at the level of building blocks and heat lines provides an improvement compared to individual parcels, but it should be applied in planning practice with due consideration of the remaining uncertainty. This paper also examines the model deviation between simulated and measured data based on building age class, building type and heat carrier. Furthermore, this paper identifies an underestimation by the state-provided Urban Building Energy Model of around 25 %.

KEYWORDS

Heat demand, Urban planning, Municipal heat planning, Urban building energy modelling, UBEM, Data acquisition, Data Aggregation, UBEM Validation.

INTRODUCTION

The building sector is a major contributor to global energy use and related greenhouse gas emissions. It is responsible for around 30 % of global final energy consumption and 26 % of global energy-related emissions [1]. In the EU, the majority of energy demand in the building sector is used for space heating [2]. Space heating is still heavily reliant on fossil fuels [3]. With a transition to sustainable energy being essential for reaching the European climate targets [4], this also means a transition away from the current heating supply sources. Currently, individual fossil-fuel boilers are the most common technology in the building sector [5]. Although gas boilers were still the most common option sold in 2022 [5], this development is subject to change, with the EU already proposing a ban on new sales from 2029

---

* Corresponding author

onwards [6]. A prominent solution is to instead switch to electrically powered heating systems, such as heat pumps, based on the premise that the electricity supply will predominantly be renewable in the future [7]. However, a significant increase in electricity demand could potentially redefine the electricity sector. Therefore, alternative sources are also considered in order to minimise the strain on the electricity grid and its transformation [7]. One option is the supply of heat through district heating networks, which are considered an essential technology for the cost-effective decarbonisation of the EU energy system [8].

To take advantage of these benefits, thorough planning of the district heating system and its future expansions is essential [9]. Geographic Information Systems (GIS) are commonly used for this purpose to analyse the demand based on spatial positioning [8], [9]. To estimate the demand, Urban Building Energy Modelling (UBEM) is becoming an increasingly useful tool. UBEM uses computational modelling, in combination with GIS, to link urban data and predict energy use at a city scale [10]. In recent years, it has grown in popularity for calculating energy use for many buildings with limited resources [11]. It has helped draft new energy policies [11] and allows planners to understand current energy expenditures and their background [12]. A key challenge in UBEM is the limited validation of many models against measured data [13]. As a result, most studies do not perform systematic calibration through input adjustment to achieve convergence [14]. This potentially leads to inaccuracies in current works due to the inherent uncertainty in UBEM models [15], which propagate to the results [16]. This is in part due to a lack of availability of or access to the data, particularly at the building level [17].

The development of UBEM methodologies, particularly with respect to predictive accuracy, has been driven by increasing computational capabilities and the continued growth of research in this domain. For example, Dilsiz et al. [18] demonstrate in a case study of 70 university buildings that calibrating UBEMs with annual building-level data improves accuracy compared to using aggregated data. When detailed data is not available, aggregation by primary use type offers the best alternative, though it may mask important building-level errors.

Johari et al. [17] developed, calibrated, and validated a UBEM for parts of the Swedish cities of Borlänge (~2050 buildings) and Uppsala (~3500 buildings) using national open data, including Geographic Information System (GIS) and Energy Performance Certificate (EPC) records. The consumption data covers 10-11 % of all buildings in both cities. They showed that prediction accuracy improves particularly with spatial aggregation, as the Mean Absolute Percentage Error (MAPE) dropped from around 22-26 % at the building level to 10-13 % at the city level. This study provides detailed insights into the validation and calibration of UBEM. Particularly noteworthy is the revalidation of the calibrated UBEM to a new area with another consumption dataset. The results of the study are presented in detail, in particular with regard to the deviations between simulated and measured thermal energy demand for residential building types. However, the study does not further differentiate the spatial aggregation levels between the parcel (property) and the district scale. As a result, it remains unclear which intermediate aggregation level would provide an optimal balance between reduced uncertainty and spatial resolution for planning purposes.

Nouvel et al. [19] studied the Bospolder district in Rotterdam, comprising around 1000 buildings. Two models were validated using gas consumption data at the postcode level. Both showed strong agreement at the neighbourhood scale (deviations of 5-25 %), but accuracy decreased at finer spatial resolutions. The study examines the validation of UBEMs based on detailed consumption data for a centrally located district within the city. The analysis is limited to the Bospolder district without investigation of whether the results are transferable to other districts in Rotterdam.

Yang et al. [20] studied 29,030 residential buildings in Leiden, located in the Netherlands, and validated their four model variants using gas consumption data from 14,321 buildings. The models differed in the inclusion of refurbishment measures, occupant heating schedules, and weather data. The best model incorporated all three factors, which substantially improved accuracy. This model overestimated total gas use by only 6 %, although achieving high accuracy at the individual building level remained challenging due to the lack of detailed building and occupant data. The other three model variants showed deviations ranging from 8 % to 20 %, depending on the variations of the applied factors. In this study, only measurement data aggregated at the building block level were available for validation, preventing a detailed assessment at the individual building level.

Grundahl et al. [21] analysed a heat atlas for the whole country of Denmark. Denmark has an extensive database containing measurement data on heat consumption for over 1.2 million buildings from 2010 to 2015, which were analysed in this study. The results show a deviation of 3.6 % at the national aggregation level, while most districts deviate by only 5-10 %. The study shows how randomly selected samples affect accuracy depending on the number of buildings aggregated and how the accuracy of an urban zone depends on the number of buildings examined. The study does not address how to manage the potentially high uncertainty of simulated data in planning contexts where only a subset of a heat atlas is used.

In summary, the accuracy and applicability of UBEMs strongly depend on the availability and resolution of measurement data. While national datasets can provide sufficient information for calibration and validation, most studies rely on aggregated data at the building block [20] or city level [17], which limits detailed assessment at the individual building level. Spatial aggregation generally improves prediction accuracy, as shown by Johari et al. [17] and Dilsiz et al. [18]. Studies on single districts, such as Bospolder in Rotterdam [19], or larger datasets like the one in Leiden [20] indicate that finer-scale validation is often constrained by data aggregation. Even with extensive national databases covering over 1.2 million buildings in Denmark [21], the results show that accuracy improves with increased aggregation and larger sample sizes. Analyses should consider fine-grained spatial aggregations as well as thematic aggregations, such as building age classes and residential building types, to obtain a comprehensive understanding of the simulated heat demand in UBEMs. While previous studies have highlighted the effects of aggregation on UBEM accuracy, detailed citywide analyses covering multiple districts remain scarce. Grundahl [21] is based on a substantially larger dataset than most comparable investigations and, therefore, enables a more comprehensive city-wide assessment. In particular, in contrast to this work, which examined large-scale regional structures, this study explicitly focuses on the effects of different aggregation levels within a single city, thereby providing differentiated insights at the municipal planning scale. To address this gap, this paper analyses heat demand data for 34,504 buildings in the city of Krefeld, Germany, using a large-scale UBEM provided by the state of North Rhine-Westphalia (NRW) and including real usage data for gas and district heating. The aim is to investigate how different spatial aggregation levels, from individual buildings up to larger districts, as well as thematic aggregations, such as building age classes and residential building types, influence the accuracy of demand predictions and improve understanding of simulated heat demand in UBEMs.

To examine the deviation between the simulated and measured consumption data, this study performs different analyses, summarising the results using box plots. In addition, metrics such as the Mean Absolute Percentage Error (MAPE), the Coefficient of Variation of the Root Mean Square Error (CVRMSE), and the Normalized Mean Bias Error (NMBE) are used to evaluate accuracy. The analysis includes various geographical and thematic aggregations. Although accuracy improves noticeably at the building block and heat line level compared to individual parcels, it is not consistently reliable across all heat lines and building blocks to support fine-grained planning. The findings of this study highlight that aggregations up to cadastral sectors have a very low distribution. Furthermore, smaller aggregations such as building blocks and heat lines also significantly improve the accuracy in direct comparison to individual parcels. Finally, this study shows that the state-provided UBEM consistently underestimates the measured consumption of residential buildings.

METHOD

This section describes the methodological approach applied to investigate the influence of spatial and thematic aggregation on the predictive accuracy of the state-provided UBEM. The methods section is structured in four consecutive parts. First, the two underlying datasets are introduced. Second, the case study context is described in detail. Third, the data preprocessing and harmonisation steps are outlined. Finally, the comparative evaluation and the descriptive indicators used to quantify model accuracy are presented.

The state-provided UBEM

To investigate how spatial aggregation affects the accuracy of UBEM predictions, this paper analyses a large-scale UBEM using real consumption data from district heating, gas, and electricity in the entire city of Krefeld in Germany. This paper compares two datasets. The first dataset, "Data for Municipal Heat Planning NRW", has been published by the North Rhine-Westphalia Office of Nature, Environment, and Climate (LANUK) to support municipal heat planning in the federal state of NRW, which became mandatory in December 2024 [22]. It contains collected and estimated data on the building stock, based on a bottom-up archetype model [23] using IWU's TABULA typology of buildings in Europe [24]. Since the model is based on the results of the TABULA project, the heat demand is calculated using U-values for the various components of the building type. The model also uses climate data from the German Weather Service (DWD) for each region, known as the climate reference year. The UBEM of LANUK, from here on referred to as the LANUK dataset, includes construction age, building types, thermal properties, and heat gains, providing specific space heating and hot water demand values at the individual building level [23]. Data are assigned to cadastral sectors (groups of parcels), building blocks (variable number of buildings enclosed by streets), and district heating routes (potential or existing supply routes) [23].

The LANUK dataset includes the calculated heat demand for individual buildings. The dataset enables a holistic analysis of all buildings and can thus improve the efficiency and accuracy of heat planning.

Case Study

The city of Krefeld is investigated as an exemplary case study, providing the empirical basis for this validation. Krefeld covers an area of approximately 137 km² and has a population of around 230,000 inhabitants [25]. The local energy supplier, NGN Netzgesellschaft Niederrhein mbH (NGN), provided consumption data for the years 2018-2022 [26]. This dataset includes annual consumption data for district heating (DH), natural gas (Gas), as well as contracted heat pumps (HP) and heating devices using night storage heating tariffs (NSH). Heat pumps are sometimes billed at reduced electricity tariffs and are registered with the local energy supplier only in such cases. It is a plausible assumption that there are heat pumps in Krefeld which are not billed under their own tariff and are, therefore, not included in the dataset. Since it was outside the scope of this paper to identify HPs and their electrical draw from general meter data, only those identified in special heat pump tariffs were included. The data for each consumption point are listed for individual metering points, which are localized via addresses. While this real consumption data provides insights into the actual energy demand at defined measuring points, it does not provide any information about the supplied buildings. This means that information like the status of renovation, the number of buildings supplied by a single metering point, the number of occupants and their behaviour, building extensions, or the use of additional fuels not supplied via networks, was not directly analysed.

Data processing

In this study, the NGN consumption data are linked with the current energy demand estimated from the LANUK dataset. This enables the combination of metering points with building information and thus also a validation of the UBEM. Figure 1 presents the flowchart of the preprocessing and analysis method. In multiple steps, these datasets were merged into a total of two databases.

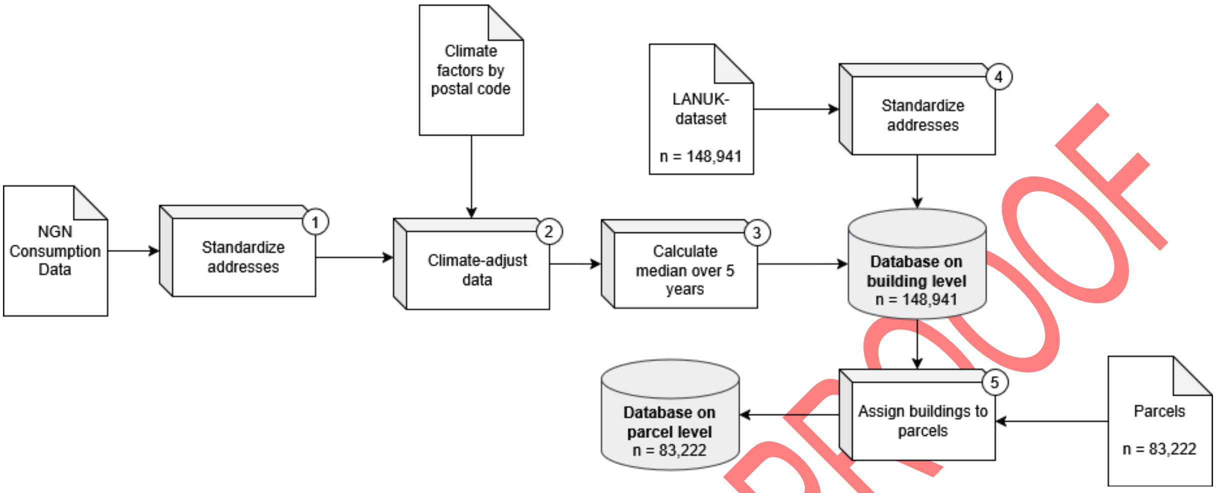


Figure 1. Flowchart of data processing steps to incorporate the NGN consumption data and the LANUK dataset into comparable databases

The LANUK dataset contains geographic coordinates for each building, as well as street names, house numbers, and address suffixes. The NGN consumption data includes addresses in the format of postal code, street name, house number, and address suffix. In Step 1, the address formatting was standardized and consolidated into a "standard address format" key, which was later used to merge the datasets.

Step 2 involves climate adjustment of all NGN consumption data. The LANUK dataset is based on the so-called "Test Reference Years" of the DWD [27]. These provide climate factors at the postal code level to convert past consumption data to an equivalent of the test reference year allowing comparison across years despite weather fluctuations [27]. Using these climate factors, the NGN consumption data for all five years and all heat carriers distributed via networks were climate-adjusted to compensate for the influence of varying weather conditions. Climate adjustment of the consumption values C_{Year} was performed according to Equation 1 using the climate factors CF_{Year} of the corresponding year, resulting in the climate-adjusted consumption $C_{CA,Year}$. In Step 3, the median of the 5 climate-adjusted consumption values from 2018 to 2022 was calculated. This method reduces the impact of outliers compared to an arithmetic mean.

$$C_{CA,Year} = C_{Year} \cdot CF_{Year} \tag{1}$$

The climate-adjusted consumption data were assigned to the database at the building level using the address keys and set to always correspond to a main building on a parcel in the database at the parcel level. Figure 2 schematically illustrates this process. The LANUK dataset distinguishes between main and auxiliary buildings on a parcel. For an auxiliary building, there can be either a space heating demand (e.g., building extensions) or no space heating demand (e.g., garages). Since these auxiliary buildings do not have their own metering points, the consumption values and estimated space heating demand of these buildings were aggregated for each parcel in Step 4. An example is shown in Figure 2. This last step is essential to enable a meaningful comparison of the datasets, thus serving as the foundation for all further aggregations and analyses.

Figure 2. Illustration of the distribution of consumption for different buildings and auxiliary buildings according to type to form individual parcels

Krefeld is home to industrial facilities, such as steel processing and chemical industry, which results in a very high energy demand. To exclude these industrial consumers, this paper focuses exclusively on parcels where the main buildings are classified as residential. Mixed-use parcels with partial residential use, such as residential apartments above a restaurant, were also included. As illustrated in Figure 2, the following analyses include only parcels for which both the actual consumption data (yellow dot) and a demand estimate (dark green) are available, and where the main building is classified as a residential building. For heat pump tariffs, a seasonal coefficient of performance of 3.5 is assumed to calculate the heat demand from electricity consumption [28]. The underlying UBEM database comprises approximately 148,941 buildings in total. After excluding all buildings not covered by the NGN consumption dataset and removing parcels used exclusively for industrial purposes, 62,843 buildings remain included in the analyses.

Comparative Result Analysis

The ratio p between two corresponding data points is determined as shown in Equation 2 using the base value E_{NGN} (NGN consumption data) and the estimated demand E_{LANUK} for space heating and hot water (LANUK dataset). It is assumed that in most cases, hot water production is already included in the measured consumption data [29], meaning it is provided directly by the heating system.

$$p = \frac{E_{LANUK}}{E_{NGN}} \tag{2}$$

Box plots (BPs) are used in this study to present the data comparison. They are suitable for the structured representation of numerical values, even if they are not normally distributed, and provide an overview of their location and spread. All BPs display the median as a light red central line within the box, which spans from the first quartile (Q1, 25th percentile) to the third quartile (Q3, 75th percentile). This range, known as the interquartile range, represents the central 50 % of the data. The whiskers extend to data points that lie within 1.5 times the interquartile range from the lower and upper quartiles, respectively. Data points beyond this range are classified as outliers and are shown separately, using individual markers.

The ratio p_i for each aggregation i is calculated using Equation 3. For this, the LANUK data for space heating and hot water are summed for all i parcels in an aggregation that meets the criteria, described as $\sum_i E_{LANUK,i}$. Additionally, the sum of all NGN consumption data points for the parcels is calculated, described as $\sum_i E_{NGN,i}$. From this, the corresponding ratio p_i between the datasets is derived.

$$p_i = \frac{\sum_i E_{LANUK,i}}{\sum_i E_{NGN,i}} \tag{3}$$

In addition to the ratio p_i, which provides insights into model accuracy across different aggregation levels through box plots, two further metrics are considered essential for evaluating the UBEM: model deviation and the uncertainty of the results, expressed as the dispersion around the mean. These metrics are quantified using the MAPE (Equation (4)), the NMBE (Equation (5)), and the CVRMSE (Equation (6)).

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{E_{LANUK,i} - E_{NGN,i}}{E_{NGN,i}} \right| \tag{4}$$

$$NMBE = \frac{\sum_{i=1}^{n} \left(E_{LANUK,i} - E_{NGN,i} \right)}{(n-1) \cdot \sum_{i=1}^{n} E_{NGN,i}} \tag{5}$$

$$CVRMSE = n \cdot \frac{\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(E_{LANUK,i} - E_{NGN,i} \right)^2}}{\sum_{i=1}^{n} E_{NGN,i}} \tag{6}$$

MAPE provides a measure of the average magnitude of errors relative to observed values, independent of their direction, and thus captures the overall deviation of the model's predictions. NMBE reflects systematic bias in the model by indicating whether predictions tend to over- or underestimate actual consumption. CVRMSE expresses the variability of prediction errors relative to the mean observed value, capturing the spread or uncertainty of the results across all analysed data points. Together, these metrics complement the ratio p_i and box plots by providing both central tendency and dispersion measures, enabling a comprehensive assessment of the UBEM's predictive performance.

RESULTS

The following section presents the results of the analysis. First, the general characteristics of the dataset are described, including the distribution of buildings by building age classes and residential building type. Subsequently, the impact of spatial aggregation on model accuracy is examined. The analysis then proceeds to thematic aggregations, considering building age classes, residential building types, and a combination of residential building types with heating systems.

Sample Characteristics: Age and Building Type Distribution

The database at the building level, based on the LANUK dataset, includes a total of 148,941 buildings distributed across 55,036 parcels in Krefeld. Out of these parcels, 32,155 were included in the analysis. This number deviates from the number of buildings because, in some cases, several main buildings are located on one parcel. Each parcel contains a space heating demand estimate from the LANUK dataset and consumption data from the NGN dataset. In addition, at least one building on the parcel is classified as a residential building. Figure 3 shows the distribution of buildings categorised by age group across the entire city of Krefeld, compared to the buildings that were analysed. Most of the analysed buildings were constructed before 1970. Interestingly, the period between 2000 and 2010 saw the fewest new buildings, with only around 580 buildings.

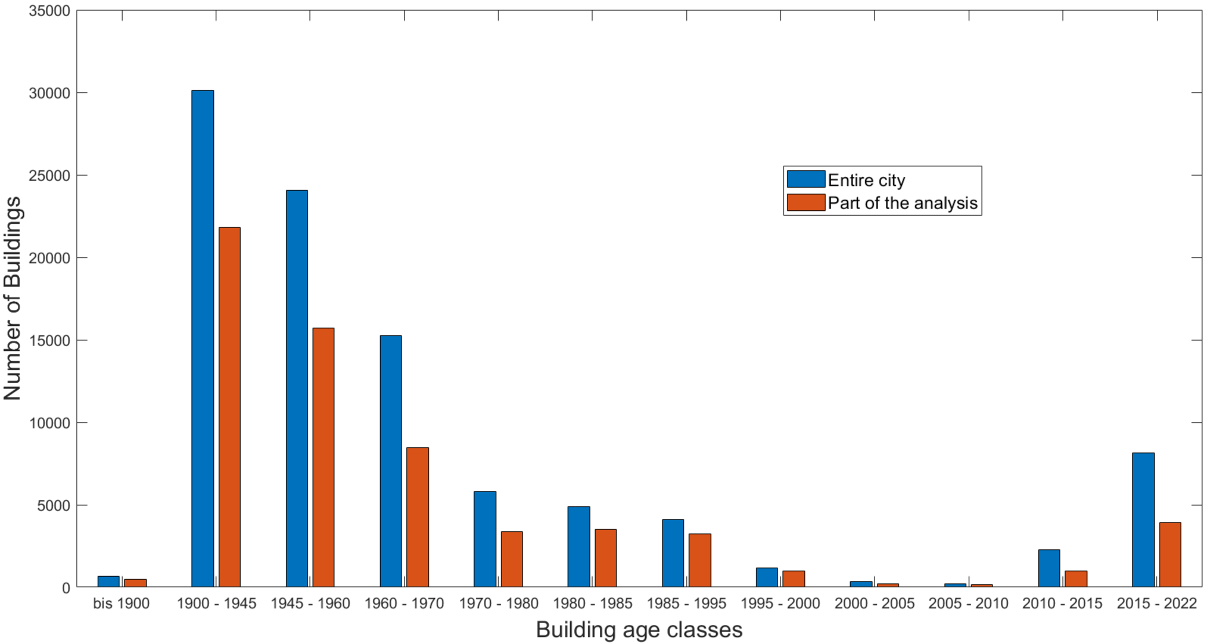


Figure 3. Distribution of the building age classes in the city of Krefeld in relation to the ones analysed in this study

Figure 4 shows the same distribution for the residential building types. As can be seen, the figure shows that most single-family houses, terraced houses, and apartment buildings were included in the analysis, while large apartment buildings are represented only to a lesser limited extent.

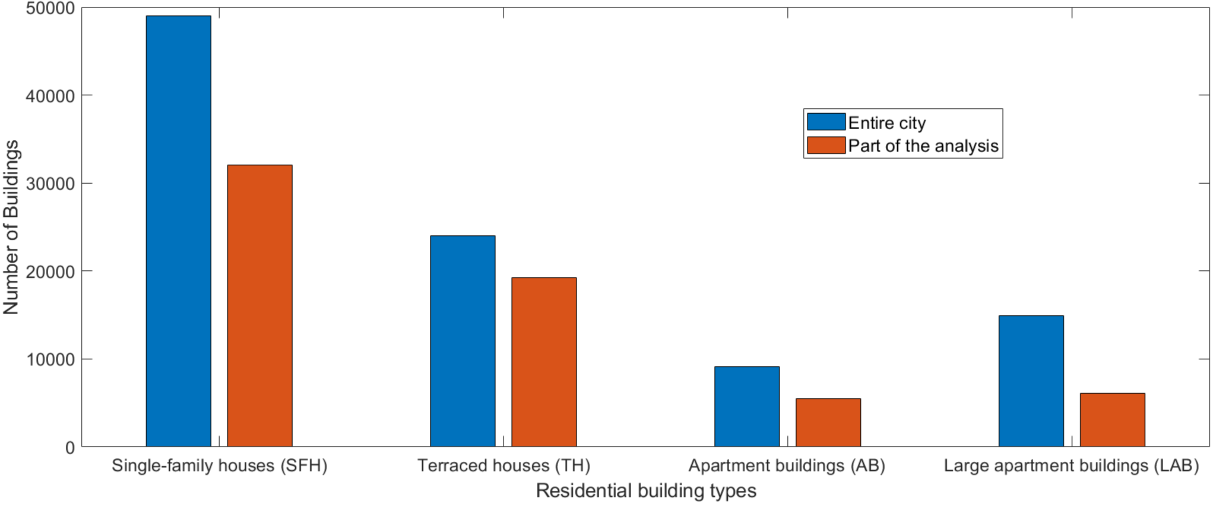


Figure 4. Distribution of the building types in the city of Krefeld in relation to the ones analysed in this study

Spatial Aggregation Analysis

The diagram in Figure 5 shows, from left to right, a single value and six BPs. The first individual value represents the largest aggregation, which is the comparison across the entire city area. The ratio for the city of Krefeld, $p_{Krefeld}$, is 76 %, meaning the LANUK UBEM underestimates consumption. The BPs for the remaining aggregations contain increasingly more values due to their finer resolution. The leftmost BP shows the comparison of the 11 postal code areas in Krefeld. This is followed by the 45 statistical districts, 223 cadastral sectors, 2,225 building blocks, and 4,143 heat lines. The last BP shows the results for the individual parcels.

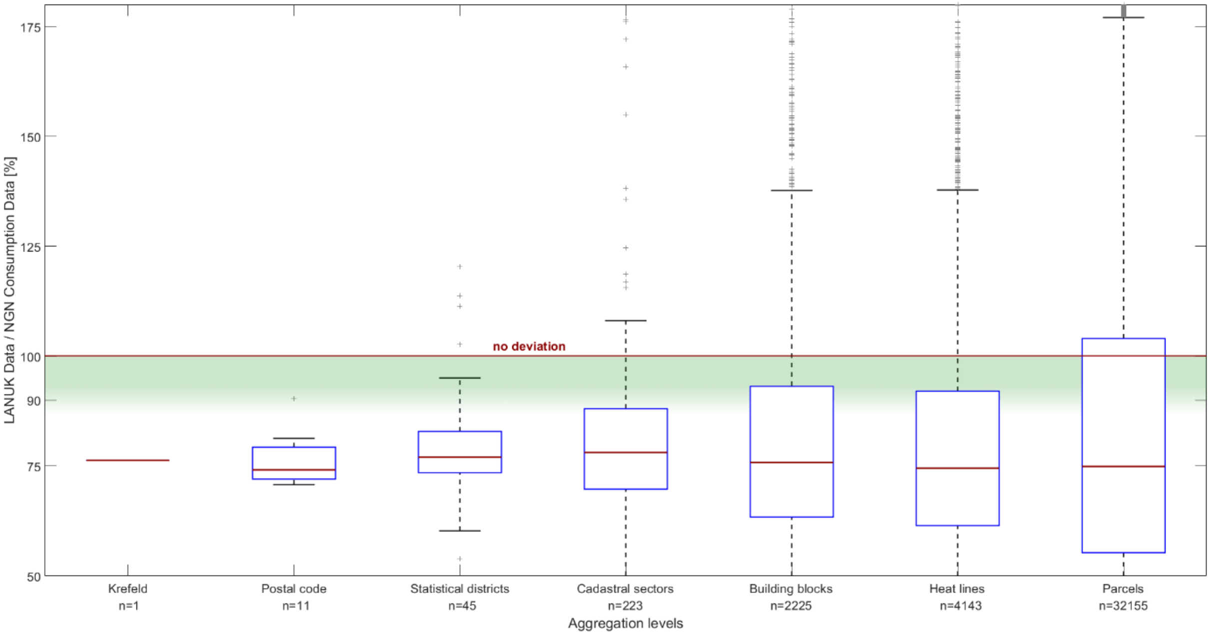


Figure 5. BP diagram of p sorted by different aggregation levels

The diagram incorporates an additional horizontal dark red line to provide a visual reference. This value indicates the point at which the LANUK dataset and the NGN consumption data are entirely aligned, suggesting no deviation exists between the two. The green area below 100 % is intended to indicate that there is an expected difference between the two datasets, with its gradient visually representing that this is not a hard limit but rather a range of expected deviations. Out of the analysed heat sources, gas is the least efficient energy source included, with an expected efficiency in conversion of at least 90 %, as evidenced by the findings reported in [30]. The actual distribution of the construction years of the gas boilers is unknown and cannot be estimated more precisely. Other factors, such as user behaviour and ventilation patterns, may also contribute to these deviations. Therefore, the expected deviation between the UBEM and the measured data should be within this green-marked range.

Figure 5 illustrates how increasing spatial aggregation affects the distribution of the results. With higher aggregation levels, the spread of deviations narrows and the number of outliers decreases. Initial improvements are visible at the level of building blocks and heat lines compared to the parcel level, while cadastral sectors further reduce the spread. Across all aggregation levels, the median values in Figure 5 indicate that the LANUK model consistently underestimates space heating demand compared to measured NGN consumption. The box plots in the figure provide a visual overview of the results, while the corresponding values in Table 1 complement this impression with detailed statistical performance indicators.

Table 1 summarises the key descriptive metrics and error indicators for the UBEM across different levels of spatial aggregation, complementing the insights provided by the box plots in Figure 5. At coarse aggregation levels, such as postal codes or statistical districts, the model exhibits the smallest deviations, with MAPE values around 22-24 % and CVRMSE values of 25-30 %, while the negative NMBE indicates an underestimation of actual consumption. As aggregation becomes finer, moving from cadastral sectors to individual parcels, MAPE and CVRMSE increase markedly, reflecting both higher variability and larger deviations at building-level resolution. Notably, at the parcel level, extreme outliers contribute to the very high MAPE and CVRMSE values, highlighting substantially reduced model accuracy at this aggregation level. Although median ratios remain relatively stable across aggregation levels, the overall magnitude of model deviation exceeds the anticipated range. Overall, Table 1 demonstrates how spatial aggregation affects both systematic and random components of model error and reinforces the observations derived from the box plots.

Table 1. Model accuracy metrics for different spatial aggregation levels

Aggregation level	**MAPE [%]**	**NMBE [%]**	**CVRMSE [%]**	**Median [%]**	**Outliers [%]**	n
Krefeld	24	-	-	76	0	1
Postal code	24	−26	25	74	9	11
Statistical districts	23	−24	30	77	11	45
Cadastral sectors	35	−24	38	78	11	223
Building blocks	31	−24	61	76	6	2225
Heat lines	42	−21	54	75	5	4143
Parcels	250	−24	201	75	7	32155

Thematic Aggregation: Building Age Classes

Figure 6 shows a BP diagram at the parcel level, grouped by the twelve building age classes based on the definition of the LANUK dataset. The greatest spread is observed for pre 1900 buildings. Buildings from 2010-2022 show the smallest spread, but also the highest deviation. Overall, the actual energy demand from the NGN consumption data is underestimated by the LANUK dataset for all building age classes.

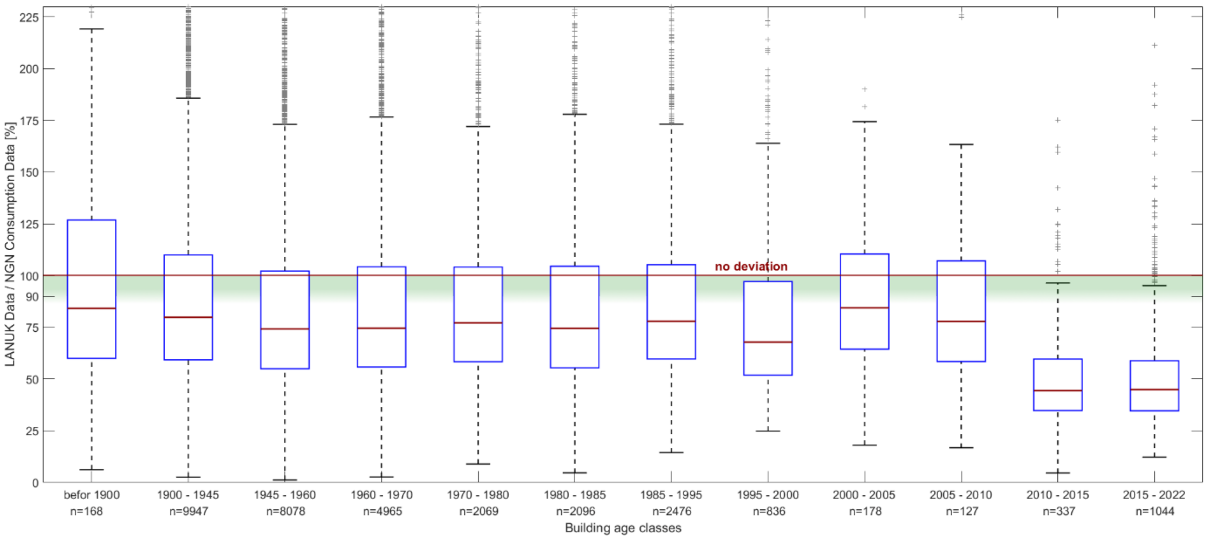


Figure 6. BP diagram of p arranged by building age classes

Table 2 illustrates that approximately 70 % of the buildings in Krefeld were constructed between 1900 and 1970, accounting for roughly 80 % of the total residential heat demand. This group also exhibits some of the highest MAPE values. For modern buildings (2010-2022), the table shows that NMBE ranges between -48 % and -53 %, indicating that simulating these building age classes is particularly challenging for the analysed UBEM. These energy-efficient buildings collectively account for only 1.7 % of Krefeld's total heat demand and represent around 4 % of the building stock.

Table 2. Model accuracy metrics for different building age classes

Aggregation level	**MAPE [%]**	**NMBE [%]**	**CVRMSE [%]**	**Median [%]**	$\sum E_{LANUK,i}$ [%]	$\boldsymbol{n}$
Before 1900	161	−23	136	84	0.71	168
1900 – 1945	246	−19	102	80	34.86	9947
1945 – 1960	348	−23	186	74	27.26	8078
1960 - 1970	325	−25	287	75	18.61	4965
1970 – 1980	265	−20	91	77	6.41	2069
1980 – 1985	61	−32	289	74	5.24	2096
1985 – 1995	117	−27	227	78	3.93	2476
1995 – 2000	46	−31	58	68	0.96	836
2000 – 2005	42	−27	96	84	0.17	178
2005 – 2010	45	−26	65	78	0.16	127
2010 – 2015	55	−48	116	44	0.47	337
2015 – 2022	120	−53	160	45	1.23	1044

Thematic Aggregation: Residential Building Types

Figure 7 shows a BP diagram aggregating the parcel data by four residential building types classified by the UBEM. The values for single-family houses (SFH) have the smallest spread, but also the highest deviation between consumption and space heating demand estimates. In contrast, the consumption of terraced houses (TH) and large apartment buildings (LAB) is less underestimated.

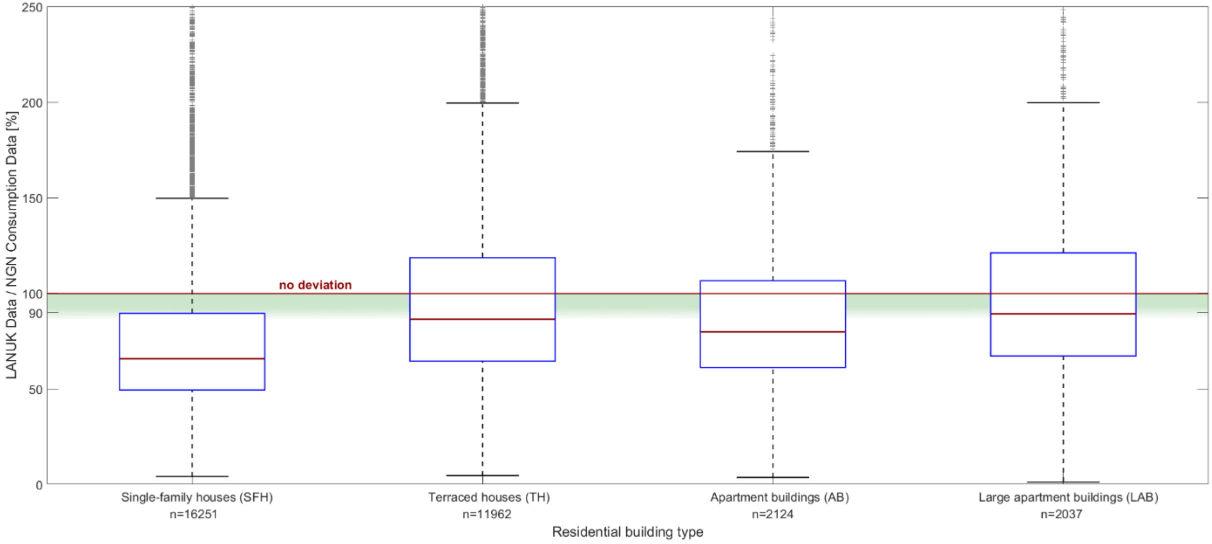


Figure 7. BP diagram of p sorted by residential building type

Table 3 presents the model accuracy metrics for the different residential building types. As already evident from Figure 5 and Table 1, the overall high MAPE values primarily result from the comparison at the parcel level and are, therefore, consistent with expectations. Of particular note is the exceptionally high MAPE for LAB, indicating that uncertainty in this building category is especially pronounced. Furthermore, the NMBE for SFH is considerably higher than for the other residential building types, showing that the model underestimates this category to a greater extent. SFH also represent the largest group, accounting for approximately 50 % of the buildings in the dataset.

Table 3. Model accuracy metrics for different building types

Aggregation level	**MAPE [%]**	**NMBE [%]**	**CVRMSE [%]**	**Median [%]**	$\sum E_{LANUK,}$ **[%]**	$\boldsymbol{n}$
Single-family houses (SFH)	199	−31	143	66	35.40	16251
Terraced houses (TH)	274	−16	74	86	26.28	11962
Apartment buildings (AB)	189	−19	95	80	12.92	2124
Lange apartment buildings (LAB)	566	−17	182	89	25.40	2037

Thematic Aggregation: Residential Building Types and Heating Systems

In Figure 8, the data are again grouped by residential building types, and each of the four BPs shown in Figure 7 is further subdivided by four types of heating systems. It becomes evident that night storage heating (NSH) exhibits a very large spread for all residential building types. Only heat pumps (HP) for apartment buildings (AB) show a comparable spread.

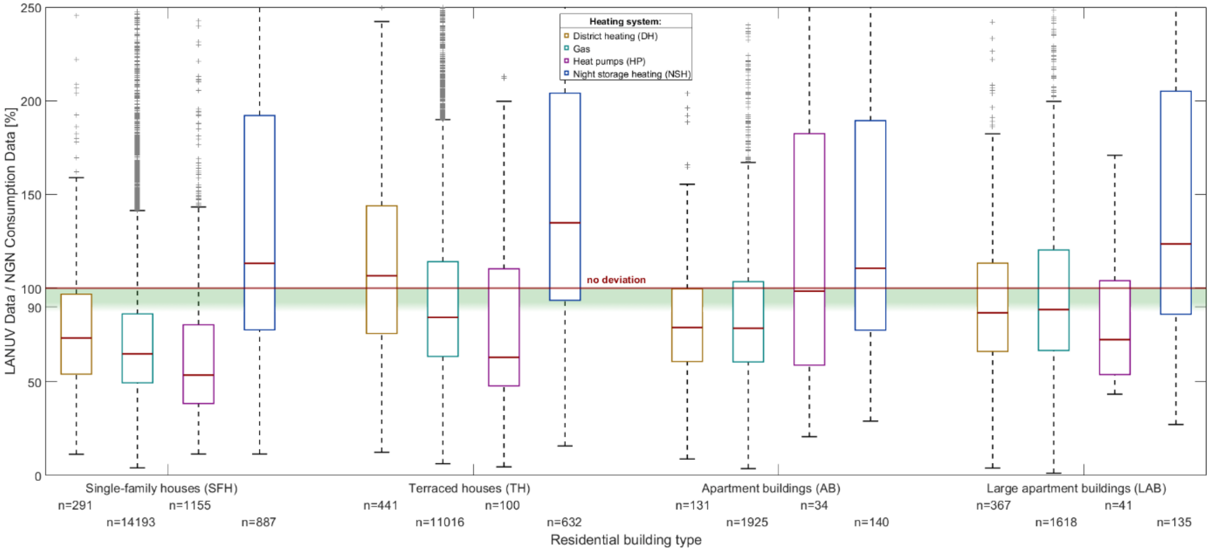


Figure 8. BP diagram of p sorted by residential building type and heat carrier

Table 4 presents the model accuracy metrics for the different residential building types in combination with the respective heat carriers. Gas dominates as a carrier for heat supply systems across all building types. The largest deviations occur for HP and especially NSH, while DH generally exhibits the lowest MAPE values. Notably, the high MAPE for gas-supplied LABs reflects the high uncertainty already observed in this building type from Table 3 and is not necessarily attributable to the combination of gas and LAB. Overall, the table demonstrates that the combination of building type and heat carrier can influence model validation.

Table 4. Model accuracy metrics for different building types and heat carrier

Aggregation level		**MAPE [%]**	**NMBE [%]**	**CVRMSE [%]**	**Median**	$\sum E_{LANUK,i}$ **[%]**	$\boldsymbol{n}$
Single-family houses (SFH)	DH	47	−22	113	73	2.05	291
	Gas	150	−33	144	65	29.50	14193
	HP	642	−38	81	54	1.43	1155
	NSH	417	13	136	113	2.09	887
Terraced houses (TH)	DH	50	−9	89	107	0.77	441
	Gas	261	−18	73	84	23.30	11016
	HP	62	−31	79	63	0.12	100
	NSH	602	24	106	135	1.67	632
Apartment buildings (AB)	DH	45	−25	87	79	1.02	131
	Gas	168	−20	93	79	10.91	1925
	HP	112	10	116	98	0.19	34
	NSH	540	17	92	111	0.86	140
Lange apartment buildings (LAB)	DH	76	−23	106	87	5.93	367
	Gas	647	−17	205	89	18.11	1618
	HP	49	−19	60	73	0.20	41
	NSH	617	46	140	124	1.85	135

DISCUSSION

In this section, two key aspects of the analysis are discussed separately. The first aspect is analysing the effects of different aggregations on the distribution of data. During the course of the analysis, a second aspect was observed: the discrepancy between the datasets. Therefore, this study also examines this aspect in order to identify possible causes for the deviation between calculation and reality and to provide insights into the model's deviation for different thematic aggregations.

The results in Figure 5 for parcels agree with previous studies [17], [31], [19], [20], and [21]. These studies show that an analysis at the parcel level is plagued by greater data variability due to individual occupant behaviour or the status of renovation and, therefore, has limited accuracy for urban heat energy planning. In the present study, the MAPE at the parcel level is around 250 %, considerably higher than in other studies, for example Johari et al. [17], where it was 78 % for the uncalibrated model. Figure 5 also confirms the hypothesis that geographical aggregation reduces the uncertainty in calculated demand. The analysis demonstrates a noticeable improvement in MAPE already at the building block and heat line levels. However, the spread of the data, expressed by the CVRMSE, remains relatively high. At the cadastral sector level, this variability is substantially reduced, indicating a more stable model prediction. While MAPE values remain high even at higher aggregation levels, their consideration together with the NMBE shows that the simulated data from the cadastral sector level onward are promising. For comparison, Johari et al. [17] achieved an improvement of approximately 10 % between the district level and the city level. Based on an estimate of building to population ratio, the district level roughly corresponds to the building block level of this study. Grundahl [21] demonstrated a reduction in data distribution for sample sizes of around 100 buildings. This roughly corresponds to the size of cadastral sectors and, in practical applications, represents too large a spatial unit to support decisions such as where to construct a new district heating pipeline.

In Figure 5 to Figure 8, outliers can be observed at the parcel level, frequently caused by unusually low consumption data points in the NGN consumption dataset. In some examined cases, the measured consumption was lower than a realistic space heating demand. This indicates that these buildings are also using fuels that are not supplied via networks, in addition to the recorded consumption. However, when analysing the data aggregated at the cadastral sector level, these outliers are already substantially reduced shown in Figure 5.

The results in Figure 5 for various aggregation levels show a deviation of approximately 25 % between LANUK demand estimates and real consumption data from NGN. This deviation is higher than expected based on the efficiency of the heating systems prior to the analysis. Occupant behaviour plays an important role in the estimation and inevitably causes a high spread of data at the parcel level, which is in line with the underlying assumptions. However, the systematic underestimation cannot be explained in this way. For the comparison, consumption data from 2018-2022 are used. Unusually frugal heating by occupants can, therefore, be ruled out, as the recent war-induced price increases only started in 2022 and thus do not affect the vast majority of the data [32]. In addition, particularly economical heating below the actual heat demand would reduce the deviation between the UBEM and measured data.

Figure 6 and Table 2 show the results for all 12 building age classes. Their NMBEs correspond to the medians calculated for the different aggregation levels (Figure 5), with slightly larger deviations between -19 % and -32 %. This is consistent with the wider distribution of data at the parcel level. However, a particularly striking observation is that buildings constructed from 2010 onwards exhibit a larger deviation between the datasets of -48 % and -53 %. A comparable effect could also be observed in the investigation of Yang et al. [20]. They show in a BP diagram that even the best UBEM variants for the building age class 2006 to 2014 is about 50 % higher on a median level than the measured value. This indicates a possible systematic error in the way the heat demand is estimated in modern buildings. Further study is needed for a better understanding of the underlying causes. As demonstrated by [33] and [34], a possible cause could be the occupant as the actual consumption can deviate considerably from calculated demand due to behavioural factors. Linked to this are studies that have shown that modern, thermally insulated buildings often exhibit higher room temperatures than older structures [28][35]. For example, [35] showed that warm water production in old buildings constructed before 1977 accounts for an average of 16.6 % of heating consumption. In buildings with EnEV 2002 insulation standards, this figure rises to 27.9 %. In individual cases, the proportion of warm water heating in new buildings can rise to up to 50 % [35]. In addition, the room temperature rises from ~18°C in old buildings to ~20°C or higher for buildings with EnEV 2002 insulation standards [35]. Residents tend to be less cautious about energy consumption in better-insulated buildings, counteracting the expected savings. This so-called "rebound effect" highlights the importance of occupant behaviour.

Figure 7 and Table 3 show that the NMBEs for the calculated heat demands of the four residential building types range from -16 % to -31 %. The heat demands of TH and LAB are calculated most accurately, with an underestimation of only 16 % (TH) and 17 % (LAB). With an underestimation of 31 %, the calculation of SFH heat demands represent the greatest challenge for the UBEM. The variance of the SFH heat demands is greater than that of the other residential building types, and they may, therefore, be particularly challenging to simulate.

Figure 8 and Table 4 provide insights into which consumption data are most suitable to validate a model. The data distribution for district heating and gas is narrower for most building types compared to NSH. The assigned NSH data shows high variability, which may be due to some of the data being a partial measure and not the full consumption. This greater variation could be due to the specific use of NSH for heating individual rooms within a housing unit. For these reasons, it seems ineffective to continue using night storage heating data for model validation. The figure also shows a noticeably wider distribution of data for HP in AB compared to other building types. The underlying cause for this should be investigated in more detail in future studies. In comparison, the results of Grundahl [21] indicate that detached SFH can be represented most accurately due to the extensive dataset available for model development. A comparable data collection was not available to LANUK during the development of their model. These results highlight the importance of real measured data at the building level to optimally support the development of UBEM for heat planning.

The analysis attempts to identify a systematic cause using box plots and statistical performance indicators for different building age classes, residential building types, and heat sources. Except for the more recently constructed buildings, the difference appears to be consistent across all building age classes. This suggests that calibration could be achieved using correction factors depending on the building type. It should be noted that heat source also influences the heat consumption, but this factor varies by building type.

In comparison to Johari et al. [17], Nouvel et al. [19], Yang et al. [20] and Grundahl [21], the deviations in this paper tend to lie at the upper end. Johari et al. [17] quantify model deviations with a MAPE of 22-26 % at the building level, decreasing to around 20 % at the neighbourhood level, and to about 10 % at the city level. These results are based on 2,044 buildings in Borlänge and 3,526 buildings in Uppsala, using EPCs for calibration and validation. In this paper, data for 32,155 out of a total of 55,036 parcels were used, enabling a more detailed and city-wide analysis. The study by Nouvel et al. [19] examines 1,000 buildings in one district. They refer to this area as block or neighbourhood level. This aggregation level corresponds to the size of a cadastral sector in this paper. The deviation between the two models is between 5 and 25 %. In Yang et al. [20], the total of 29,000 buildings in Leiden (Netherlands) are distributed across a total of 2,950 postal code areas. Yang et al. [20] evaluate 1,292 postal code areas. This corresponds approximately to the aggregation at the corridor or building block level in this study. Four variants of a UBEM were examined; the deviations ranged from 6 to 20 % depending on the variant. In a direct comparison, the UBEM of the LANUK examined in this study most closely corresponds to model variant 2, which is architecture-based, uses average weather data, and takes the state of renovation into account. This model 2 has a deviation of 12 % in Yang et al. [21]. Furthermore, Grundahl [21] investigates a heat atlas for the country of Denmark. Most regions achieved an agreement of 5-10 % compared to measured data, while the most challenging region still reached 89 % agreement. These regions are presumably comparable to the scale of an entire city in this study.

A possibility to improve the robustness of the results is a more detailed investigation of outliers and the application of an appropriate filtering strategy prior to evaluation. For instance, parcels with measured consumption values far below the simulated heat demand may indicate partially heated or temporarily vacant buildings, or buildings that rely on additional non-network energy sources. Such cases could distort the comparison between calculated demand and measured consumption. The present results do not strongly suggest that such filtering would reduce the observed underestimation, since excluding unusually low consumption values would rather increase the distance from the 100 % line. Nevertheless, implementing a systematic filtering procedure could be beneficial in future studies to further enhance the reliability of model validation.

In the present study, for all examined building blocks and heat lines, more than one measurement value was available. However, the number of buildings within a heat line or a building block can vary considerably. This may affect the statistical robustness of the respective aggregation level. A possible follow-up study could, therefore, place greater emphasis on the sample size within each aggregation unit and systematically investigate the minimum number of buildings required to obtain reliable results at the level of building blocks or heat lines.

This study clearly demonstrates the importance of validating a UBEM with real measured data prior to its application in order to ensure reliability. Based on the presented results, it is recommended that any municipality conduct a preliminary assessment before applying the UBEM locally. Only a single region of the UBEM was examined in this paper, for which data were available, and it remains uncertain whether the same systematic deviations would be observed in other cities or municipalities. For Krefeld, however, it is likely that a calibration of the UBEM based on the findings of this paper could be successfully implemented.

CONCLUSION

For accurate energy planning, it is important to understand how different spatial aggregation levels affect the accuracy of space heating demand predicted by UBEM. This study focuses on a single city, Krefeld, within the NRW region, as access to measured consumption data was only available for this area. While the analysis is therefore, limited to one municipality, the findings provide insights that can be relevant for other cities and municipalities across NRW. The results suggest that deviations between the UBEM predictions and actual measured consumption may be considerably larger than expected and, in some cases, differ in direction from initial assumptions. Such deviations highlight potential uncertainties not only in the present heat demand but also in modelled projections for future consumption, which could have implications for municipal heat network planning, particularly regarding the sizing, design, and expansion of distribution infrastructure.

Further research should focus on how cities and municipalities can strategically use measured data to adapt UBEMs to local conditions. Based on the present study conducted in Krefeld, the development of a standardized calibration procedure using existing consumption data from municipal utilities could help reduce model uncertainties while ensuring compliance with data protection requirements. Such an approach would enable more accurate predictions of heating energy demand at all geographical aggregation levels and provide a reliable basis for municipal heat planning.

Although only one municipality has been examined in detail so far, the findings suggest that the deviations observed in Krefeld may also occur in other cities in the region. The similarity in the historical development of the building stock, particularly for buildings constructed after World War II, as well as the comparable regulatory and climatic conditions in many North Rhine-Westphalian cities, support the potential transferability of these results. These parallels underscore the importance of local validation and calibration before applying UBEMs in other municipalities.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to SWK AG and SWK ENERGIE GmbH for their collaboration and funding of projects related to municipal heat planning in Krefeld, which are related to this study. Special thanks also go to NGN Netzgesellschaft Niederrhein mbH for their valuable support in providing the datasets and implementing the project. We also thank the city of Krefeld for their support and guidance throughout the project.

NOMENCLATURE

Symbols

C	Consumption in kWh	[m]
CF	Climate factor	[-]
E	Energy	[kWh]

n	Number of elements in one boxplot	[-]
p	Ratio	[%]
Subscripts and superscripts		
CA	Climate-adjusted	
$LANUK$	Value from LANUK dataset	
NGN	Value from NGN consumption data	
Abbreviations		
AB	Gross Domestic Product	
BP	Boxplot	
CVRMSE	Coefficient of Variation of the Root Mean Square Error	
DH	District heating	
DWD	Deutscher Wetterdienst (German Weather Service)	
EPC	Energy Performance Certificate	
GIS	Geographic Information System	
HP	Heat pumps	
IWU	Institut Wohnen und Umwelt GmbH	
LAB	Large apartment building	
LANUK	North Rhine-Westphalia Office of Nature, Environment and Climate	
MAPE	Mean Absolute Percentage Error	
NGN	NGN Netzgesellschaft Niederrhein mbH	
NMBE	Normalized Mean Bias Error	
NRW	North Rhine-Westphalia	
NSH	Night storage heating	
SFH	Single-family house	
TH	Terraced house	

REFERENCES

[1] Cozzi, L. and Gould, T., World Energy Outlook 2023, International Energy Agency (IEA), https://www.iea.org/reports/world-energy-outlook-2023, [Accessed: 12-December-2025].

[2] Statistical Office of the European Union (Eurostat), Disaggregated Final Energy Consumption in Households - quantities, 2023, https://doi.org/10.2908/NRG_D_HHQ.

[3] Statistical Office of the European Union (Eurostat), Complete Energy Balances, 2022, https://doi.org/10.2908/NRG_BAL_C

[4] European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, The European Green Deal, Brussels, Belgium, 2019.

[5] Braungardt, S., Tezak, B., Rosenow, J. and Bürger, V., Banning boilers: An analysis of existing regulations to phase out fossil fuel heating in the EU, *Renewable and Sustainable Energy Reviews,* Vol. 183, pp. 113442, 2023, https://doi.org/10.1016/j.rser.2023.113442

[6] European Union, Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Committee and the Committee of the Regions, EU 'Save Energy', Brussels, Belgium, 2025.

[7] Möller, B., Wiechers, E., Persson, U., Grundahl, L. and Connolly, D., Heat Roadmap Europe: Identifying local heat demand and supply areas with a European thermal atlas, *Energy,* Vol. 158, pp. 281-292, 2018, https://doi.org/10.1016/j.energy.2018.06.025

[8] Connolly, D., Lund, H., Mathiesen, B.V., Werner, S., Möller, B., Persson, U., Boermans, T., Trier, D., Østergaard, P.A. and Nielsen, S., Heat Roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system, *Energy Policy,* Vol. 65, pp. 475-489, 2014, https://doi.org/10.1016/j.enpol.2013.10.035

[9] Eslami, S., Noorollahi, Y., Marzband, M. and Anvari-Moghaddam, A., Integrating heat pumps into district heating systems: A multi-criteria decision analysis framework incorporating heat density and renewable energy mapping, *Sustainable Cities and Society,* Vol. 98, pp. 104785, 2023, https://doi.org/10.1016/j.scs.2023.104785

[10] Dabirian, S., Panchabikesan, K. and Eicker, U., Occupant-centric urban building energy modeling: Approaches, inputs, and data sources - A review, *Energy and Buildings,* Vol. 257, pp. 111809, 2022, https://doi.org/10.1016/j.enbuild.2021.111809

[11] Kong, D., Cheshmehzangi, A., Zhang, Z., Ardakani, S.P. and Gu, T., Urban building energy modeling (UBEM): a systematic review of challenges and opportunities, *Energy Efficiency,* Vol. 16, No. 6, 2023, https://doi.org/10.1007/s12053-023-10147-z

[12] Ang, Y.Q., Berzolla, Z.M. and Reinhart, C.F., From concept to application: A review of use cases in urban building energy modeling, *Applied Energy,* Vol. 279, pp. 115738, 2020, https://doi.org/10.1016/j.apenergy.2020.115738

[13] Malhotra, A., Bischof, J., Nichersu, A., Häfele, K.-H., Exenberger, J., Sood, D., Allan, J., Frisch, J., van Treeck, C., O'Donnell, J. and Schweiger, G., Information modelling for urban building energy simulation—A taxonomic review, *Building and Environment,* Vol. 208, pp. 108552, 2022, https://doi.org/10.1016/j.buildenv.2021.108552

[14] Kamel, E., A Systematic Literature Review of Physics-Based Urban Building Energy Modeling (UBEM) Tools, Data Sources, and Challenges for Energy Conservation, *Energies,* Vol. 15, No. 22, pp. 8649, 2022, https://doi.org/10.3390/en15228649

[15] Shen, P. and Wang, H., Archetype building energy modeling approaches and applications: A review, *Renewable and Sustainable Energy Reviews,* Vol. 199, pp. 114478, 2024, https://doi.org/10.1016/j.rser.2024.114478

[16] Johari, F., Peronato, G., Sadeghian, P., Zhao, X. and Widén, J., Urban building energy modeling: State of the art and future prospects, *Renewable and Sustainable Energy Reviews,* Vol. 128, pp. 109902, 2020, https://doi.org/10.1016/j.rser.2020.109902

[17] Johari, F., Shadram, F. and Widén, J., Urban building energy modeling from geo-referenced energy performance certificate data: Development, calibration, and validation, *Sustainable Cities and Society,* Vol. 96, pp. 104664, 2023, https://doi.org/10.1016/j.scs.2023.104664

[18] Dilsiz, A.D., Nweye, K.E., Wu, A.J., Kämpf, J.H., Biljecki, F. and Nagy, Z., How spatio-temporal resolution impacts urban energy calibration, *Energy and Buildings,* Vol. 292, pp. 113175, 2023, https://doi.org/10.1016/j.enbuild.2023.113175

[19] Nouvel, R., Mastrucci, A., Leopold, U., Baume, O., Coors, V. and Eicker, U., Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support, *Energy and Buildings,* Vol. 107, pp. 204-212, 2015, https://doi.org/10.1016/j.enbuild.2015.08.021

[20] Yang, X., Hu, M., Heeren, N., Zhang, C., Verhagen, T., Tukker, A. and Steubing, B., A combined GIS-archetype approach to model residential space heating energy: A case study for the Netherlands including validation, *Applied Energy,* Vol. 280, pp. 115953, 2020, https://doi.org/10.1016/j.apenergy.2020.115953

[21] Grundahl, L. and Nielsen, S., Heat atlas accuracy compared to metered data, *International Journal of Sustainable Energy Planning and Management,* Vol. 23, pp. 3-13, 2019, https://doi.org/10.5278/ijsepm.3174

[22] Landesamt für Natur, Umwelt und Klima Nordrhein-Westfalen (LANUK), Municipal Heat Planning in North Rhine-Westphalia (in German), https://www.lanuk.nrw.de/themen/klima/waerme-und-kommunale-waermeplanung/kommunale-waermeplanung, [Accessed: 12-December-2025].

[23] Landesamt für Natur, Umwelt und Klima Nordrhein-Westfalen (LANUK), Data on Municipal Heat Planning in North Rhine-Westphalia (in German), https://open.nrw/dataset/4dcb3c40-548c-4d32-b09b-decaf838c885, [Accessed: 21-July-2025].

[24] Institut Wohnen und Umwelt GmbH (IWU), TABULA, Development of building typologies for the energy assessment of residential buildings in 13 European countries (in German), https://www.iwu.de/forschung/gebaeudebestand/tabula/, [Accessed: 12-December-2025].

[25] The City of Krefeld, Krefelder Kommunalstatistik, https://www.krefeld.de/de/krefelder-kommunalstatistik, [Accessed: 28-02-2026].

[26] NGN Netzgesellschaft Niederrhein mbH, 2023, Consumption Dataset 2018-2022, NGN Netzgesellschaft Niederrhein mbH.

[27] Deutscher Wetterdienst (DWD), Climate factors (KF) for energy consumption certificates (in German), https://www.dwd.de/DE/leistungen/klimafaktoren/klimafaktoren.html, [Accessed: 21-July-2025].

[28] Schröder, F.P., Papert, O., Boegelein, T., Navarro, H. and Mundry, B., Real trends in specific energy consumption and representative residential temperature control with increasing modernisation levels in the housing stock (in German), *Bauphysik,* Vol. 36, No. 6, pp. 309-324, 2014, https://doi.org/10.1002/bapi.201410045

[29] Statistisches Bundesamt, Wohnen in Deutschland - Zusatzprogramm Wohnen des Mikrozensus 2022, https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Wohnen/Publikationen/Downloads-Wohnen/wohnen-in-deutschland-5122125229005.html?utm_source=chatgpt.com, [Accessed: 04-March-2026].

[30] Martinopoulos, G., Papakostas, K.T. and Papadopoulos, A.M., A comparative review of heating systems in EU countries, based on efficiency and fuel cost, *Renewable and Sustainable Energy Reviews,* Vol. 90, pp. 687-699, 2018, https://doi.org/10.1016/j.rser.2018.03.060

[31] Nutkiewicz, A., Yang, Z. and Jain, R.K., Data-driven Urban Energy Simulation (DUE-S): Integrating machine learning into an urban building energy simulation workflow, *Energy Procedia,* Vol. 142, pp. 2114-2119, 2017, https://doi.org/10.1016/j.egypro.2017.12.614

[32] Statista, Gaspreisentwicklung in Deutschland bis 2025 - Verbraucherpreisindex| Statista, https://de.statista.com/statistik/daten/studie/914807/umfrage/entwicklung-der-gaspreise-in-deutschland-verivox-verbraucherpreisindex/, [Accessed: 04-02-2026].

[33] Ali, U., Shamsi, M.H., Hoare, C., Mangina, E. and O'Donnell, J., Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, *Energy and Buildings,* Vol. 246, pp. 111073, 2021, https://doi.org/10.1016/j.enbuild.2021.111073

[34] Nejadshamsi, S., Eicker, U., Wang, C. and Bentahar, J., Data sources and approaches for building occupancy profiles at the urban scale – A review, *Building and Environment,* Vol. 238, pp. 110375, 2023, https://doi.org/10.1016/j.buildenv.2023.110375

[35] Felsmann, C. and Schmidt, J., Impacts of Consumption-Based Billing Depending on the Energy Performance of Buildings (in German), Final report, https://bved.info/wp-content/uploads/2013/10/Studie_UniDresden.pdf, [Accessed: 21-July-2025].