Energy demand uncertainty analysis based on a probabilistic building characteristics approach

Rui Guo¹,², Simon Meunier³,⁴, Dirk Saelens¹,²
¹KU Leuven, Department of Civil Engineering, Kasteelpark Arenberg 40, Heverlee, 3001, Belgium
²EnergyVille, Thor Park 8310, Genk, 3600, Belgium
³Université Paris-Saclay, CentraleSupélec, CNRS, Gif-sur-Yvette 91192, France
⁴Sorbonne Université, CNRS, GeePs, Paris 75252, France

Abstract
Understanding district energy demand is crucial to address the growing demand for energy-efficient buildings and renewable energy integration at regional levels. This study proposes a novel probabilistic building characterisation method for district energy simulations, combining various quantile regression methods, several copula methods, and a building simulation framework to quantify the uncertainty of heat demand in Belgian dwellings and districts. The linear quantile regression method was selected to build the marginal (uncorrelated) distribution of U-values of building elements using the explanatory variables combination of building type, protected volume, ground floor area, and roof area. The C-vine copula method is then chosen to build the multivariate distributions of U-values and generate random correlated samples. Meanwhile, the heat demand functions based on the average U-values for dwellings are obtained using the building energy simulation framework. A number of detached and terraced dwellings were randomly selected to form Belgian rural and urban districts. At last, the multivariate distributions of the U-values of building elements and thus the distribution of the average U-values of selected dwellings are generated. The average U-values are used as inputs to the heat demand function to quantify the uncertainty in the heat demand of Belgian dwellings and districts. The findings reveal that the heat demand of detached dwellings has a larger range of values than terraced dwellings, while the range of heat demand values is larger for Belgian urban districts compared to rural districts.

Highlights
- Most appropriate combination of explanatory variables and quantile regression method for obtaining the marginal distributions of U-values of building elements are found.
- Most appropriate copula method for building multivariate distributions is found.
- Heat demand as a function of the average U-value is well-fitted linearly for the considered cases.
- Uncertainties in the heat demands of Belgian dwellings and rural and urban districts are quantified.

Introduction
The need to tackle climate change and reduce the environmental impact of existing buildings has led to increased interest in improving energy efficiency and incorporating renewable energy sources at the district level (Capone et al., 2021). Urban building energy models (UBEMs) have become essential for quantifying operational building energy use at various scales. Quantifying the district energy demand can contribute to studies such as the low-voltage grid stability analysis (Meunier, Protopapadaki, Baetens, et al., 2021) and the district heating networks analysis (Morvaj et al., 2016). Geographic Information System (GIS)-based UBEMs often use geospatial data to generate building models and estimate energy consumption. However, this approach heavily depends on the quality and availability of input data, such as building geometries, construction materials, and occupancy patterns (Nouvel et al., 2015). Numerous studies rely on public data sources, which can vary in quality and resolution, leading to uncertainties in the simulation results (Mastrucci et al., 2014). Furthermore, UBEMs often utilise prototype buildings, oversimplifying the actual variability in the existing building stock due to insufficient building-level input data. This approach can oversimplify the actual variability in the existing building stock, leading to a significant error of up to 21%, particularly for smaller-scale examples (Reinhart & Cerezo Davila, 2016). The problem of underestimating variability can only be addressed by removing the prototype buildings and describing the characteristics of each building individually. However, obtaining the characteristics of each individual building can be extremely challenging. One possible approach is to characterise the full variability of existing neighbourhoods without using the energy performance certificate (EPC) database or prototype buildings. A probabilistic building characterisation method was proposed to model the variability of the Belgian existing residential building stock using the EPC database (De Jaeger et al., 2021). The method estimates a realistic distribution of building U-values and window-to-wall ratios based on known data (location, geometry and construction year). However, only the linear quantile regression (QR) method and one set of explanatory variables (e.g., building construction year, building geometry) for getting the multivariate distributions of response
variables such as U-values, were used. Moreover, the data from the multivariate distribution were not fed as building-level parameters into UBEMs to perform uncertainty analysis on the energy demand of residential neighbourhoods. Therefore, there is a need to develop probabilistic approaches that can better capture the uncertainties and dependencies among building parameters for quantifying the uncertainties in energy demand at building and district levels.

In response to this gap, this study proposes a novel probabilistic building characterisation method for district energy simulations, which combines different QR methods, various copula methods, and a building simulation framework to quantify the uncertainty of heat demand in dwellings and districts. This analysis offers valuable insights into the correlation between U-values of building elements and uncertainty analysis of heat demand at building and district levels, which can inform policy decisions and energy efficiency measures tailored to the specific characteristics of dwellings and districts.

**Methodology**

Figure 1 illustrates the proposed research framework of the probabilistic building characterisation method for the district energy simulations. To obtain the marginal distributions for the U-values of building elements (i.e., response variables), linear QR, neural network QR, and gradient boosting QR are used. These methods are based on known inputs (i.e., explanatory variables). The most appropriate QR method and combinations of explanatory variables that produce the smallest prediction errors are selected. Then, the multivariate distributions of the U-values are determined based on the marginal distributions. The Gaussian, C, D, and R-vine copula methods are compared to build the multivariate distributions and draw correlated samples. The most suitable copula method, which provides the smallest dissimilarity between the actual multivariate distribution of U-values, is identified.

Meanwhile, the heat demand (space heating + domestic hot water (DHW)) of detached and terraced dwellings under different insulation levels and stochastic residential occupancy profiles are simulated using OpenIDEAS (R. Baetens et al., 2015), an open framework for integrated building and district energy simulations. The heat demand per floor area as a function of the average U-value for detached and terraced dwellings is then linearly fitted separately. Besides, the detached and terraced dwellings are randomly selected based on the shares and numbers of these two building types in Belgian rural and urban districts.

By adopting the most appropriate copula method, the multivariate distributions of the U-values of building elements, and thus the distribution of the average U-value of selected dwellings in urban and rural districts are drawn based on the corresponding explanatory variables. In the next step, 1000 random samples are generated from these multivariate distributions, and their average U-values are used as inputs to the heat demand function to perform uncertainty analysis for the heat demand of dwellings and the rural and urban districts.

**Marginal distributions for U-values**

Three QR methods are used to obtain the marginal distributions of U-values of the window, roof, ground floor, and external walls based on the available data like building geometry and construction year as explanatory variables. The explanatory variables considered in this study are postal code, building type, construction year, total floor area, protected volume, ground floor area, façade area, and roof area, as they are available for all dwellings. Thus, a total of 768 parametric simulations are conducted considering the number of QR methods and the number of explanatory variable combinations to find the most appropriate QR method and combination. The available data used in this study involves 243 dwellings from the Stebo vzw company in Belgium, including the explanatory variables and U-values.
The marginal distribution represents all possible probabilities of the response variable without regard to the values of other relevant response variables and without any correlation. To estimate model quantiles conditional distribution of the response variable’s quantiles as a function of observed covariates, linear QR (Koenker & Bassett, 1978), gradient boosting QR (Scikit, 2019), and neural network QR (Cannon, 2011) are adopted and compared. Linear QR uses an ordinary least squares method, gradient boosting QR uses an ensemble of decision trees, and neural network QR uses an artificial network of artificial neurons. Table 1 lists the relevant parameters for gradient boosting and neural network QR methods.

Table 1: Parameter setup for gradient boosting and neural network QR methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient boosting QR</td>
<td>Estimators</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>Max depth</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Learning rate</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Subsample</td>
<td>1.0</td>
</tr>
<tr>
<td>Neural network QR</td>
<td>Hidden layers</td>
<td>2 (32, 16)</td>
</tr>
<tr>
<td></td>
<td>Activation</td>
<td>tanh</td>
</tr>
<tr>
<td></td>
<td>Output activation</td>
<td>linear</td>
</tr>
<tr>
<td></td>
<td>Optimizer</td>
<td>Adam</td>
</tr>
<tr>
<td></td>
<td>Learning rate</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Epochs</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Batch size</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Validation split</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Early stopping</td>
<td>patience: 15</td>
</tr>
</tbody>
</table>

QR models can approximate the full cumulative distribution function of a random variable by characterising a complete range of quantiles. The StatsModels, scikit-learn, and keras Python packages are used to set up the QR models. Invalid data points are removed before fitting QR models by visually observing scatter plots and applying filters. For instance, only values between the 1st and 99th percentiles are kept for the construction year. Only values between the 2.5th and 97.5th percentiles are retained for all other geometrical parameters. The training and test datasets are split using a 90/10 ratio. QR models are fitted for each response variable (i.e., U-values) and each qth quantile ranging from 0.01 to 0.99, resulting in 396 models. Subsequently, the explanatory variables for each dwelling are fed into each QR model to generate the marginal distributions for the test dataset. In this way, the cumulative distribution functions for the four response variables are characterised by aggregating the QR models for each qth quantile.

Correlated samples for U-values
Generating samples using marginal distributions of response variables can result in unrealistic samples in the multivariate case, as the response variables may not be completely independent. If the random variables are correlated, and samples are generated using their marginal distributions solely, the generated values would not represent realistic samples. It is important to infer the multivariate relations between different response variables, such as their multivariate distribution or correlations, to generate realistic samples. For example, a dwelling with a well-insulated façade will likely have a well-insulated roof. Thus, obtaining the marginal distributions of the different variables is only the first step, and inferring their multivariate relationships is crucial for generating realistic samples.

The Gaussian, C, D, and R-vine copulas methods are proposed and compared to build multivariate distributions from marginal distributions and to draw correlated samples. All four methods are implemented through the Copulas Python package and offer distinct ways to incorporate correlations between U-values. The Gaussian copula is the simplest and most widely known copula model, which captures the dependence structure using a linear correlation matrix (Nelsen, 2007). In contrast, C-vine and D-vine copulas are based on regular vine structures, which are more flexible in modelling complex dependence patterns among multiple variables (Aas et al., 2009). The C-vine copula is characterised by its hierarchical structure, while the D-vine copula adopts a sequential structure, making it more suitable for high-dimensional data (Kurowicka & Cooke, 2006). The R-vine copula is the most flexible and general among these methods. It allows for arbitrary pair-copula constructions, providing a robust framework to model various dependence structures in multivariate distributions (Dißmann et al., 2013).

Performance indicators
To enhance available energy performance-related data within UBEMs, this study proposes a probabilistic method for characterising buildings, which considers four parameters: the U-values of the floor, external walls, windows, and roof. Initially, QR is utilised to generate marginal distributions for these four variables. Next, various copula methods are presented for constructing multivariate distributions from marginal distributions and generating random correlated samples. An extensive comparison of different implementations of probabilistic building characterisation methods has been carried out based on the following two performance indicators.

Average absolute difference (ABD) between empirical coverages and prediction intervals. To verify the accuracy of the marginal distributions, their empirical coverage is evaluated at different prediction intervals. Specifically, this study considers the 50%, 80%, 90%, and 98% prediction intervals and calculates their respective empirical coverage. For instance, the empirical coverage of the 98% prediction interval is discussed in detail. In this case, the empirical coverage represents the percentage of buildings in the test dataset for which the actual value falls within the predicted 1st and 99th quantiles and should ideally be close to the theoretical value of 98%. ABD is the mean value of all the discrepancies between the empirical coverages and prediction intervals.

Achieved Results to Date
The proposed probabilistic methods are implemented using Python packages such as scikit-learn, StatsModels, and keras. The datasets are split into training and test sets, and the models are trained and validated accordingly. The performance of the models is evaluated using various metrics, including the average absolute difference (ABD) between empirical coverages and prediction intervals. The results show that the proposed methods are effective in characterising buildings with varying energy performance-related data. Further, the models are compared with existing methods to demonstrate their efficacy and robustness.

To conclude, the development of probabilistic methods for characterising buildings is essential for enhancing available energy performance-related data in UBEMs. The proposed methods are effective in generating realistic samples and enhancing the accuracy of the predictive models. However, further research is needed to improve the performance of the models and to extend their applicability to a wider range of building types and geographical locations.
Normalized maximum mean discrepancy (NMD). To verify the accuracy of the multivariate distributions, NMD is utilised to assess the dissimilarity between the real multivariate distribution of all dwellings and the multivariate distributions generated from the copula methods. In this study, 16 random samples are generated for all the dwellings of the dataset. NMD is a variant of the maximum mean discrepancy, defined as the maximum difference in expectations over functions in the unit ball of a reproducing kernel Hilbert space. The NMD is particularly useful for assessing the similarity between two distributions, with smaller values indicating a higher degree of similarity. Independent samples are drawn from each distribution to compute the NMD between two probability distributions, and their empirical means (the average values of the data points in the sample) are calculated. A kernel function, which measures the similarity between data points in the feature space, is then chosen to calculate the NMD.

Building energy simulation framework

An open framework for integrated building and district energy simulations, OpenIDEAS (R. Baetens et al., 2015), is used for building energy simulations. IDEAS Modelica library, a tool within the OpenIDEAS, is used to build the model and conduct the building energy simulations. The typical methodological year weather data in Uccle, Belgium, generated from Meteonorm (Meteotest, 2009), are selected for the simulations.

A Belgian generic dwelling structure (Protopapadaki & Saelens, 2017) shown in Figure 2 is considered. Each dwelling is treated as a separate building, consisting of two thermal zones: the day-zone, which includes the ground floor living area and kitchen, and the night-zone, encompassing bedrooms and corridors on the first floor. The building features a pitched roof, and the two zones share a common floor without air circulation. Internal heat gains are divided, with 70% allocated to the day-zone and 30% to the night-zone, based on the assumption that more appliances are present in these areas and occupants spend more active time in the kitchen and living room (Protopapadaki & Saelens, 2017). The dwellings are assumed to have a North/South orientation.

Building parameters for EU countries in 2020 were collected in Work Package 1 (WP1) of the sEEnergies reports (SEEnergies Reports, 2022). Accordingly, the window-wall ratio for Belgian detached dwellings is set at 0.3 and 0.17 for terraced dwellings. Furthermore, the mean floor areas for the Belgian detached and terraced dwellings in 2020 are set as 80 m² and 76 m², respectively.

To explore the impact of building element U-values on heat demand, three levels of U-values (i.e., High-U, Medium-U, and Low-U) are selected. The High-U scenario for Belgian dwellings is derived from the U-values in 2020 collected in WP1 (SEEnergies Reports, 2022), while the Low-U scenario for each building element is 0.15 W/m²·K for the exterior wall, 1.00 W/m²·K for the window, 0.15 W/m²·K for the roof, and 0.25 W/m²·K for the basement. These Low-U scenario values are more ambitious than those projected for 2050 in WP1. For the Medium-U scenario, U-values for corresponding building elements are calculated as the average between the High-U and Low-U scenarios. The average dwelling U-value is calculated by Equation (1).

$$U_{\text{average}} = \frac{\sum U_{\text{building element}} \cdot A_{\text{building element}}}{\sum A_{\text{building element}}}$$

(1)

where $A_{\text{building element}}$ represents the area of different building elements and $U_{\text{building element}}$ is the U-value of the corresponding building element. Table 2 shows the U-values under different insulation levels and the corresponding average U-values. Notably, with the increase of insulation, the infiltration air change rate per hour decreases. Thus, the infiltration air change rates per hour for High-U, Medium-U, and Low-U scenarios are 0.55 h⁻¹, 0.35 h⁻¹, and 0.15 h⁻¹, respectively.

Table 2: U-values of building elements under different insulation levels.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>High U-values of building elements (W/m²·K)</td>
<td>Wall: 0.70, Window: 1.93, Roof: 0.74, Basement: 0.66 Average (Detached): 0.87, Average (Terraced): 0.76</td>
</tr>
<tr>
<td>Medium U-values of building elements (W/m²·K)</td>
<td>Wall: 0.42, Window: 1.46, Roof: 0.45, Basement: 0.45 Average (Detached): 0.56, Average (Terraced): 0.47</td>
</tr>
<tr>
<td>Low U-values of building elements (W/m²·K)</td>
<td>Wall: 0.15, Window: 1.00, Roof: 0.15, Basement: 0.25 Average (Detached): 0.28, Average (Terraced): 0.22</td>
</tr>
</tbody>
</table>

The stochastic residential occupancy model utilises the Python StROBe Package, as described in (Ruben Baetens & Saelens, 2016), which is a part of the OpenIDEAS framework. StROBe provides missing boundary conditions related to human behaviour in integrated district energy assessment simulations, such
as appliance and lighting usage, space heating settings, and domestic hot water redraws. To capture the uncertainty from the occupant behaviour, a pool of 300 stochastic Belgian occupancy profiles derived from StROBe is generated. Consequently, two pools of 300 dwelling heat demand (space heating + DHW) profiles are generated for each insulation level: one pool for detached dwellings and another one for terraced ones. Accordingly, 1800 heat demand profiles are obtained. After obtaining the heat demand per floor area for detached and terraced dwellings under three insulation levels, heat demand functions are fitted as a function of the dwelling average U-value.

The number of dwellings in Belgian rural areas is $1.86 \times 10^5$, while the number of dwellings in Belgian urban areas is $4.95 \times 10^6$ (Eurostat, 2011). In rural areas, all the dwellings are assumed to be detached while, in urban areas, 67% of dwellings are assumed to be detached, and 33% are assumed to be terraced (Meunier, Protopapadaki, Persson, et al., 2021). As explained in (Meunier, Protopapadaki, Persson, et al., 2021), these assumptions lead to a conservative estimation (i.e. overestimation) of the energy demand. (Guo et al., 2023) states that the average number of Belgian dwellings per low-voltage feeder is 17 for rural areas and 20 for urban ones. Therefore, the considered rural and urban districts comprise 17 and 20 dwellings in this study.

**Results and discussion**

**Accuracy of the marginal distributions**

After conducting 768 parametric simulations considering 256 combinations of explanatory variables and three QR methods, the most appropriate combination of explanatory variables is building type, protected volume, ground floor area, and roof area, and the most appropriate QR is linear QR. This yields the smallest ABD between empirical coverages and prediction intervals at 2.4%. The superior performance of linear QR over the other two QR methods may be due to the linear nature of the underlying relationship between the explanatory variables, the marginal distributions of U-values, and the potential suboptimal tuning of the more complex gradient boosting and neural network models.

Table 3 presents the empirical coverages for different prediction intervals of all U-values under the most appropriate combination of explanatory variables and QR. The empirical coverages are relatively close to their respective prediction intervals. This indicates that the linear QR models provide reasonably accurate estimates of the U-value marginal distributions.

**Table 3: Empirical coverages (%) on the different prediction intervals for all U-values.**

<table>
<thead>
<tr>
<th>Output variable</th>
<th>Prediction interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50%</td>
</tr>
<tr>
<td>$U_{\text{window}}$</td>
<td>47.6%</td>
</tr>
<tr>
<td>$U_{\text{roof}}$</td>
<td>47.6%</td>
</tr>
<tr>
<td>$U_{\text{ground floor}}$</td>
<td>47.6%</td>
</tr>
<tr>
<td>$U_{\text{external walls}}$</td>
<td>47.6%</td>
</tr>
</tbody>
</table>

Figure 3 presents the probability distributions predicted by the QR models for the four U-values for a selected dwelling. The predicted distributions demonstrate that the QR models accurately capture the U-value marginal distributions for this dwelling. The real values fall within the predicted ranges, which tends to indicate that the selected combination of explanatory variables and the linear QR method provide a reliable estimation of the U-value distributions. This further reinforces the findings from Table 3, which shows that the empirical coverages are relatively close to their respective prediction intervals.

![Figure 3: Probability distributions predicted by the QR models for U-values of one selected semi-detached dwelling from the test dataset. The red dot line indicates the actual U-values for the selected dwelling in Zoenhoven and built in 1982. The dwelling’s features include a floor area of 251 $m^2$, a protected volume of 720 $m^3$, a ground floor area of 126 $m^2$, a façade area of 213 $m^2$, and a roof area of 146 $m^2$. Histograms are generated using 100,000 randomly selected samples from the cumulative distribution function for improved readability.](image_url)

**Accuracy of the correlated samples**

Table 4 presents the NMD between the real multivariate distribution and the multivariate distributions generated from the various copula methods, which quantifies the
performance of these methods in capturing the dependence structure of the U-values. A lower NMD value indicates better performance. Among the four copula methods analysed, the C-vine method exhibits the lowest NMD at 0.018, suggesting that it is the most accurate in modelling the dependence structure among the considered methods. Conversely, the R-vine method demonstrates the highest NMD at 0.028, indicating a relatively weaker performance than the other methods. The Gaussian and D-vine methods show similar NMD values, at 0.023 and 0.022. Therefore, the C-vine copula method is selected to build the multivariate distributions based on the marginal distributions of U-values.

Table 4: NMD of different copula methods.

<table>
<thead>
<tr>
<th>Copula method</th>
<th>NMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>0.023</td>
</tr>
<tr>
<td>C-vine</td>
<td>0.018</td>
</tr>
<tr>
<td>D-vine</td>
<td>0.022</td>
</tr>
<tr>
<td>R-vine</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Dwelling heat demand as a function of U-value

Figure 4 shows the annual space heating and DHW demands per floor area under varying average U-values for terraced and detached dwellings and their respective fitted heat demand functions. Each data point corresponds to the average value of 300 heat demand results based on a pool of 300 stochastic Belgian occupancy profiles. The linear regression method is used to fit the functions. The resulting coefficients of determination ($R^2$) are 0.99, signifying a strong linear relationship between the heat demand and the building average U-value. A close observation of the fitted functions reveals a stronger influence of the average U-value on the heat demand for detached dwellings than terraced dwellings. This is evident from the steeper slope of the heat demand function for detached dwellings, indicating a more significant increase in heat demand as the average U-value increases. For the considered cases, the fitted function can thus be used directly to estimate heat demand without running detailed building energy simulations.

Figure 4: Annual space heating and domestic hot water (DHW) demands per floor area of dwellings under different insulation levels and fitted heat demand functions.

Uncertainty analysis for heat demand

Figure 5 shows the distributions of heat demand per floor area of a randomly selected detached dwelling in the rural district and a terraced dwelling in the urban district. The detached dwelling exhibits a wider distribution of heat demand, with values ranging from 47.9 kWh/m² to 553.4 kWh/m² and a median of 216.0 kWh/m². In contrast, the terraced dwelling shows a more condensed distribution, with values ranging from 16.2 kWh/m² to 340.8 kWh/m² and a median of 122.0 kWh/m².

The observed disparity between the heat demand distributions of detached and terraced dwellings may be mainly attributed to the exposed surface areas of the two dwellings. Detached dwellings generally have a larger exposed surface area, resulting in greater heat loss and higher heat demand than terraced dwellings. In addition, terraced dwellings benefit from shared walls with adjacent properties, which can reduce heat loss and lead to lower heat demand. This comparative result also aligns with the fitted heat demand functions for these two types of dwellings illustrated in Figure 4, where the detached dwelling demonstrates higher heat demand than the terraced one under the same insulation level.

Figure 5: Distributions of heat demand per floor area of selected detached and terraced dwellings.

Figure 6 shows the distributions of heat demand in rural and urban districts. The heat demand distribution (369.3 MWh to 738.1 MWh) is observed to be broader for the urban district (432.7 MWh to 915.4 MWh), implying a higher variability in heat demand compared to the rural one. With a median annual heat demand of 609.4 MWh, the urban district exhibits a higher demand than the rural one, with a median value of 532.2 MWh. The first possible reason for these differences is the higher number of dwellings in the urban district. Even though the urban district comprises six terraced dwellings,
which typically have lower heat demand per floor area than detached dwellings, the overall number of dwellings (i.e., 20) is greater than that of the rural district (i.e., 17). The higher number of dwellings in the urban district could notably contribute to the observed variability and the higher median heat demand. In addition, factors such as building age, insulation quality, and architectural design could also influence the heat demand distributions and median values for rural and urban districts.

We also observe a significant overlap in heat demands between the two districts, which ranges from 432.7 MWh to 738.1 MWh. This suggests that certain dwelling types (detached or terraced) in both districts might share similar heat demand characteristics. The distribution of district heat demands also indicates the need to conduct an uncertainty analysis to better understand the factors influencing these similarities and differences.

![Figure 6: Total heat demand distributions of rural and urban districts.](image)

**Conclusion**

This study proposes a novel probabilistic building characterisation method for uncertainty analysis in heat demand of detached and terraced dwellings in considered Belgian rural and urban districts using various quantile regression and copula methods and a building energy simulation framework. The results highlight the superior performance of linear quantile regression for estimating marginal distributions of U-values based on the most appropriate combination of explanatory variables, including building type, protected volume, ground floor area, and roof area. The C-vine copula method is the most appropriate approach to model the dependence structure among U-values. The strong linear relationship between heat demand and average U-values for detached and terraced dwellings is demonstrated, and the corresponding heat demand functions are linearly fitted. The fitted functions can be used directly to estimate heat demand without running detailed building energy simulations. The uncertainty analysis indicates that the detached dwelling exhibits a higher median heat demand per floor area than the terraced dwelling due to its larger exposed surface area. Additionally, the urban district shows a higher median heat demand than the rural district, mainly attributable to the differences in dwelling characteristics (e.g., building age) and the greater number of dwellings in the urban district. In future research, the uncertainty in heat demand can be employed to assess the electricity consumption of heating devices (particularly heat pumps) and estimate the impact on the stability of low-voltage electrical grids. Such insights can be valuable for energy planners and policymakers, enabling them to make well-informed decisions regarding energy infrastructure investments.

**Acknowledgement**

This work is part of the EU H2020 “sEEnergies” project (https://www.seeenergies.eu/) and the Belgian FLUX 50 ICON “MUPEDD” project. “sEEnergies” project has received funding from the European Union’s Horizon 2020 Research and Innovation Action under Grant Agreement No 846463. “MUPEDD” project has received funding from Flanders Innovation & Entrepreneurship.

**Reference**


