A PROBABILISTIC APPROACH TO ALLOCATE BUILDING PARAMETERS WITHIN DISTRICT ENERGY SIMULATIONS

I. De Jaeger[1,2,3], J. Lago[1,3,4], and D. Saelens[1,2]

[1]EnergyVille, Thor Park 8310, BE-3600 Genk

[2]KU Leuven, Department of Civil Engineering, Building Physics Section, Kasteelpark Arenberg 40 – box 2447, BE-3001 Heverlee

[3]Flemish Institute for Technological Research (VITO), Boeretang 200, BE-2400 Mol

[4]Delft University of Technology, Delft Center for Systems and Control, Mekelweg 2, NL-2628CD Delft

ABSTRACT

Urban building energy models (UBEMs) are expected to play a key role in the integrated assessment of sustainability measures on both district and city level. However, due to limited availability of data sources, those models are often created through an archetype approach, which is a deterministic method to allocate building envelope characteristics to building groups. Unfortunately, this deterministic approach may underestimate the variability of the existing building stock, which is important when designing district energy systems to optimise the location of production and storage units within the system. In contrast to the deterministic approach, this work presents a new probabilistic approach to allocate building envelope characteristics within UBEMs that in combination with stochastic occupants enables to include the variability of existing districts. A thorough comparison of the deterministic and the probabilistic method is established for 820 buildings of the Boxbergheide district in Genk by performing dynamic energy simulations in the IDEAS Modelica library. For the studied district, a probabilistic building envelope characterisation with standard occupants increases the coefficient of variation (CV) on the energy demand for space heating, compared to a deterministic approach with standard occupants, from 17.8% to 46.4%. Including a probabilistic building envelope characterisation increases the variability on the energy demand for space heating to a larger extent than including stochastic occupants, which increases the CV to only 29.6%.

INTRODUCTION

To reduce the climate impact of the existing building stock, strategies include both increasing the energy efficiency and integrating renewable energy sources. However, these measures should be studied on a district or city level to include the synergy effects that result from the heterogeneity of the existing building stock. To do so, urban building energy models (UBEMs) are emerging, as they can be used to quantify the operational building energy use on district or city level through building-by-building simulation (Reinhart & Cerezo Davila, 2016). UBEMs are typically bottom-up building physics models (Kavgic et al., 2010), which enable to analyse the current status of the building stock and to assess possible future scenarios that combine energy efficiency measures with renewable energy integration. An additional strength of UBEMs is that they allow for studies on multiple levels: from street to district to city level.

In order to quantify the energy use of all buildings within a UBEM, a considerable number of input parameters is required for each building. These include geometry and location, occupant behaviour, building envelope characteristics, heating, ventilation and air conditioning (HVAC) systems, renewable energy systems, and building appliances. Building geometry and location can often be acquired from geographic information systems (GIS), since GIS are becoming more commonly available. Significant research efforts have been contributed to the influence of occupant behaviour, resulting in a number of tools to allocate occupant behaviour in a stochastic way (e.g. Baetens & Saelens, 2015). Nevertheless, the characteristics of the building envelope, the present HVAC and renewable energy systems, as well as the present building appliances still remain unavailable. Although these characteristics could be acquired per building, the data acquisition effort becomes infeasible on district or city level.

Due to the lack of building envelope and system characteristics on building level, UBEMs often make use of archetypes, which are either average buildings that represent a group of similar buildings or sample buildings (i.e. an actual building that is closest to the average of a group of similar buildings). The whole building stock is thus represented by a limited number of representative buildings or archetypes, e.g. using the resulting typologies of the TABULA project (Cuypers et al., 2014) or other studies (Swan & Ugursal, 2009). As an example, the TABULA typology for Belgium characterises the whole Belgian building stock by 30 archetypical dwellings, considering five different building types and six different construction periods (Cuypers et al. , 2014). These approaches decrease not only the data collection effort but also the required simulation time, which is particularly important when using dynamic simulation models in district simulations. Although UBEM simulations have been reported to correspond reasonably well to measured energy use data on higher aggregation levels, with errors of the considered studies ranging from 4 to 21% (Reinhart & Cerezo Davila, 2016), the errors increase significantly when focusing on the scale of individual buildings. In other words, the archetype approaches fail to include the non-negligible variability that is characteristic for the existing building stock (De Jaeger et al. , 2017), which is important for the optimal design of district energy systems. Particularly on a smaller scale (~100 dwellings) the use of archetypes may no longer be justifiable.

To model the actual variability in building energy use and to assess the feasibility as well as the optimal design of district energy systems, all buildings should be fully characterised; thus, additional data, i.e. both measured building energy use and building envelope and system characteristics, are desired. Measurement data can be obtained from distribution system operators and can be used both to calibrate the model (Sokol et al., 2017) and to estimate the simulation error. Data on the building envelope and system characteristics can be obtained from governmental databases such as the Energy Performance Certificates databases in Europe (Österbring et al., 2016). Energy Performance Certificates are labels that inform consumers of the energy efficiency of buildings they plan to purchase or rent. The Flemish Energy Performance Certificates database is therefore a valuable resource for energy performance-related data of buildings (i.e. building type, construction year, building geometry, thermal performance of the building envelope, information on the HVAC systems, …). However, privacy issues are often the key argument for not sharing the data.

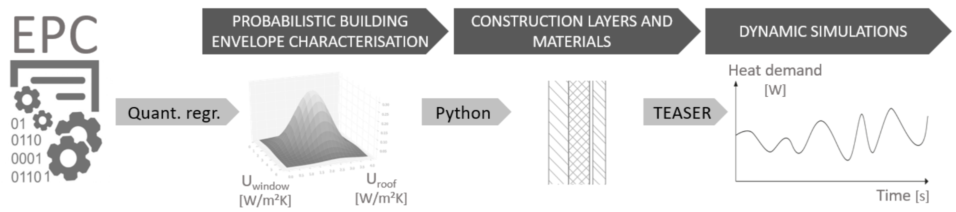
Within this context, this work presents a new probabilistic method to allocate the building energy related data to individual building models. The Flemish Energy Performance Certificate (EPC) database is employed to set up probability density functions for the building energy related data and to relate these to parameters that are known on individual building level through GIS and cadastral data – i.e. building geometry and construction year. Although this approach can be extended to the building energy systems, a first assessment of the usability of the EPC database focuses solely on the building envelope characteristics of Flemish single-family dwellings. To illustrate its added value, the presented probabilistic method is applied to the Boxbergheide district in the city of Genk. Then, it is compared to a deterministic approach that makes use of the Belgian TABULA typologies through an analysis of the building UA-value and the building energy demand for space heating, by including both standard occupants and stochastic occupants.

In the next Section, the probabilistic methodology to allocate building energy related data in UBEMs in is introduced. Also, the Boxbergheide district that was used to compare the probabilistic method to the deterministic method is presented. Subsequently, the results of this comparison are described. Then, the presented method as well as the comparison results are discussed. Finally, the conclusions are drawn.

METHODOLOGY

In this Section, the new method to allocate building energy related data to individual building models is presented. The method consists of three main parts, which are all implemented in Python and are illustrated in Figure 1. Firstly, the building envelope properties of Flemish single-family dwellings are characterised in a probabilistic manner based on the EPC database. To this end, quantile regression models are set up. Secondly, as a detailed building energy simulation model is used for this study, all construction element layers and materials should be deduced from the U-values that are generated by the quantile regression model. Thirdly, this information is automatically translated into detailed Integrated District Energy Assessment Simulations (IDEAS) building models (Jorissen et al., 2018). The IDEAS library is implemented in the Modelica language and allows simultaneous transient simulation of thermal, control, and electric systems at both building and district level. Finally, to show the differences between a deterministic and a probabilistic approach, the Boxbergheide district, on which both methods are applied, is introduced.

Figure 1: Graphical overview of the proposed probabilistic method to characterise districts.



From the Energy Performance Certificates database to a probabilistic building characterisation

The building envelope properties (i.e. U-values of the roof, outer walls, windows and ground floor) are characterised in a probabilistic way per building, based on available data. For Flanders, the available data for all buildings consists of building geometry and construction year. Building geometry data can be obtained from the Flemish GIS, but is for this study obtained from a CityGML model of the city of Genk with level of detail (LOD) 2 (Biljecki et al., 2016). The available building geometry and location data includes postal code, building type (terraced, semi-detached or detached dwelling), building volume, building height, ground floor area, façade area and roof area. In this work, the heated floor area is deduced from an assumed number of storeys based on the building height, which is described in (De Jaeger et al., 2017). The construction year can be obtained from the Flemish cadastral database, which is a property register that contains, among others, information on the ownership, land use, building dimensions, and building construction year for taxation purposes.

To obtain a probability distribution function for all U-values of the buildings, quantile regression models were built based on the Energy Performance Certificates database. Quantile regression (QR), introduced by Koenker and Bassett (1978), expresses all quantiles of the conditional distribution of the response variable as functions of observed covariates. This can be compared with Ordinary Least Squares (OLS), where only the conditional mean is estimated by minimizing the squared residuals. In particular, instead of estimating the mean, QR models estimate the complete conditional distribution of the response variable. In more detail, in OLS the sample mean μ of a variable y, which is an estimate of the unconditional population mean E(Y), is found by solving the following problem:

$$\min_{\mu \in \Re} \sum_{i=1}^{n} (y_i - \mu)^2$$

Likewise, an estimate of the conditional expectation function E(Y|x) can be equally found by OLS by replacing μ by a parametric function μ(x, β):

$$\min_{\beta \in \Re^p} \sum_{i=1}^{n} (y_i - \mu(x_i, \beta))^2$$

Similarly, in QR, the unconditional τ^{th} quantile of y, i.e. qτ, can be found by solving the following problem:

$$\min_{q\tau \in \Re} \sum_{i=1}^{n} \rho_\tau (y_i - q\tau)$$

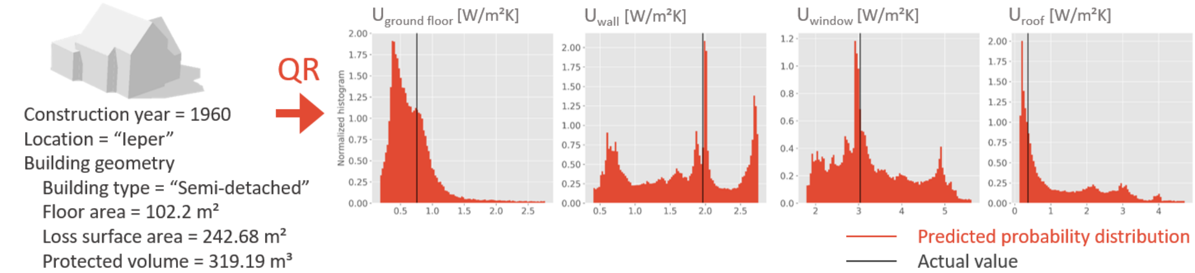
where $\rho_\tau(u) = \tau * u$ for $u > 0$ and $\rho_\tau(u) = (\tau - 1) * u$ for $u < 0$. Similarly, an estimate of the conditional τ^{th} quantile of y can be found by replacing $q\tau$ by a parametric function $q\tau(x_i, \beta)$:

$$\min_{\beta \in \Re^p} \sum_{i=1}^{n} \rho_\tau (y_i - q\tau(x_i, \beta))$$

A very interesting property of quantiles is that they are in fact a representation of the Cumulative Distribution Functions (CDF) of any variable. Particularly, given the CDF F(y) of a random variable y, $q\tau$, i.e. the τ^{th} quantile, relates to F(y) as follows: $F(q\tau) = \tau$. Therefore, QR models are able to characterize a complete range of quantiles and thus approximate the full CDF of y. The optimisation problem to estimate the quantiles is solved through linear programming (Koenker & Bassett, 1978). For this study, postal code, building type, construction year, total floor area, protected volume, ground floor area, façade area (opaque plus transparent), and roof area were considered as explanatory variables, since they are available for all buildings and a preliminary analysis showed their relevance. Subsequently, the CDFs for the various output variables are created by aggregating the QR models for each τ^{th} quantile, with τ ranging from 0.01 to 0.99. As an example, the resulting distributions are shown for one building in Figure 2.

The QR models were built based on a training set of 330000 EPC dwellings, leaving 82000 dwellings as a test set. For this study, an anonymised version of the EPC database – excluding address-related information –, the StatsModels, and the scikit-learn Python packages were used.

Figure 2: Probability distributions predicted by the QR model for all the U-values for one specific dwelling of the test set. To enhance readability, histograms are created based on 100000 random samples from the CDF.



To check the accuracy of the QR models, the empirical coverage of the predictions at different prediction intervals is evaluated. The 50%, 80%, 90% and 98% prediction intervals are considered and the empirical coverage on these intervals is computed (Table 1). As an example, the 90% prediction interval is discussed. For the 90% prediction interval, the empirical coverage is equal to the percentage of all buildings in the test set of which the real value falls within the predicted 5th and 95th quantiles and should ideally be close to the theoretical range of 90%. Table 1 shows that the empirical coverage is close to the theoretical range for all output variables and all considered prediction intervals.

Table 1: Empirical coverages on the 50%, 80%, 90% and 98% prediction interval for all output variables

Prediction interval → Output variable ↓	50%	80%	90%	98%
Roof U-value	0.4981	0.8009	0.9012	0.9807
Outer wall U-value	0.5019	0.7997	0.8982	0.9790
Ground floor U-value	0.4992	0.7992	0.8994	0.9806
Window U-value	0.4963	0.7980	0.8997	0.9793

From sampled U-values to construction layers and materials

Through the QR models, all U-values of all dwellings are characterised by CDFs. After sampling particular U-values for the roof, walls, ground floor and windows from these CDFs, the U-values are translated to a corresponding set of construction layers and materials, which is described now.

Table 2: Initial construction layers of the roofs, outer wall and ground floors, along with the predefined order of upgrade and the maximal thicknesses of the layers.

Construction element	Initial U-value of construction element [W/m²K]	Initial materials (from inside to outside)	Initial thickness [m]	Order of upgrade	Maximal thickness [m]
Roof	3.37	Gypsum plaster	0.01	/	/
		Timber	0.01	/	/
		Mineral wool (415 mm) between wooden rafters (35 mm)	0.00	1	Unlimited
Outer wall	2.52	Gypsum plaster	0.02	/	/
		Heavy masonry	0.10	1	0.14
		Mineral wool	0.00	3	Unlimited
		Non-ventilated air cavity	0.00	2	0.025
		Heavy masonry	0.09	/	/
Ground floor	2.75	Ceramic tile for finishing	0.02	/	/
		Screed	0.06	/	/
		Expanded polystyrene	0.00	1	Unlimited
		Dense cast concrete	0.14	/	/

As the majority of the Flemish buildings are heavyweight constructions, lightweight constructions are not considered and all buildings initially share the same constructions. If the initial U-value is lower than the sampled U-value, then initial construction remains unchanged. If the initial U-value is higher than the sampled U-value, the initial constructions are adjusted with respect to a predefined order of upgrades (Table 2). As an example, the method is described for the outer wall, but a similar approach is followed for the ground floor and roof. In order to reach the sampled U-value of the outer wall, three possible upgrades are predefined, each concerning a particular layer and material. The first upgrade concerns the inner heavy masonry layer. The required thickness of this layer to reach the sampled U-value is calculated and compared to the predefined maximal thickness of this layer, determining the final thickness. If the new U-value does not yet satisfy the sampled U-value, a non-ventilated air cavity with a maximal thickness of 2.5 cm is added. The resistance is calculated following EN ISO 6946. If the new

U-value still does not satisfy the sampled U-value, mineral wool is added until the sampled U-value is reached. This method allows a continuous distribution of U-values. However, this approach could not be extended to the windows, since the window glazing is only available for a limited number of U-values, which also determines the optical properties. Therefore, only 6 different glazing options were considered (glazing U-values of 0.7, 1.0, 1.1, 1.4, 2.9 and 5.8 W/m²K). Subsequently, for each glazing option, the required window frame U-value is calculated. The window frame U-value is restricted from 1.3 and 5.9 W/m²K. Finally, the glazing option that results in the U-value closest to the sampled U-value is allocated.

From a fully characterised neighbourhood to a dynamic simulation model

In this study, TEASER, developed by RWTH Aachen (Remmen et al., 2018), is used to translate geometrical CityGML models into Modelica models. However, in order to satisfy the needs of this study, TEASER was slightly adapted as described above as well as in (De Jaeger et al., 2017). Firstly, a CityGML model – either LOD1 or LOD2 – containing building geometry, building function, construction year, number of storeys and storey height is imported. Then, the CDFs of the U-values of the roof, the walls, the windows, and the ground floor are estimated. Subsequently, particular U-values are sampled randomly from these CDFs and are translated in construction layers and elements. Finally, all building descriptions are exported to detailed IDEAS building models.

The IDEAS library supports detailed building energy simulations modelling transient thermal phenomena within the building using a zonal modelling approach, assuming perfect mixture of the air inside the zone. A detailed description of the IDEAS library is given in (Jorissen et al., 2018). The adapted TEASER version is used to generate two-zone IDEAS building models, assuming that the ground floor represents the day zone while all the upper floors belong to the night zone. Each building is implemented with an ideal radiator heating system and no ventilation system. To calculate the ventilation losses, air infiltration is included, but window opening is not. The simulations are conducted for the heating dominated climate of Uccle (Belgium) for a period of 1 year. A 1-month initialization period is used. Simulations are performed in Dymola, using the Dassl solver with an output interval of 10 min.

An introduction to the Boxbergheide district

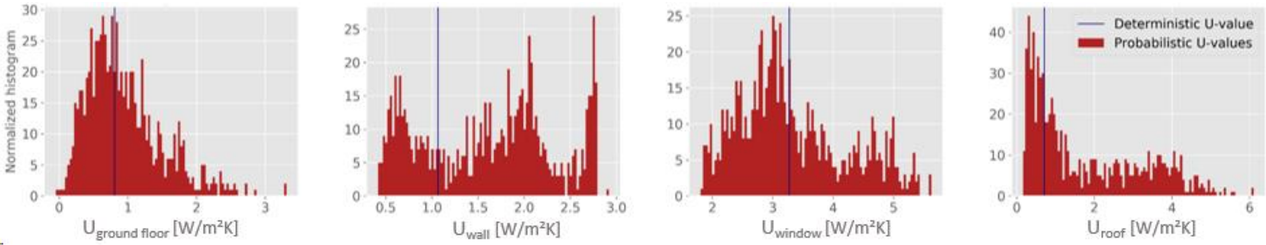
To illustrate its added value, the presented probabilistic method is compared to a deterministic approach, which makes use of the TABULA typologies. Both approaches are used to simulate 820 single-family dwellings of the Boxbergheide district of Genk. In both approaches, the geometry originates from an LOD2 CityGML model of the city of Genk. For the sake of simplicity, the construction year is randomly allocated, ranging from 1980 to 1989, and therefore the buildings are described by one of the Belgian TABULA typologies. The exact assumptions with a view to the construction layers and materials are given in (Protopapadaki et al., 2014). Due to the nature of the deterministic approach, all buildings share the same U-values for the roof, the walls, the windows, and the ground floor and the variability in energy simulation results for this approach is solely due to the variability in geometry. By contrast, in the probabilistic approach, each individual building is characterised by both their own geometry and their own U-values.

As the occupant behaviour appears to be decisive for the building energy demand (Baetens & Saelens, 2015), the comparison is performed for two types of occupants, resulting in four different modelling approaches. The first approach for modelling the occupants is to assume standard occupants, identical for all buildings, and modelled following the ISO 13 790 standard with an indoor air temperature set point for day zone and night zone respectively of 21°C/18°C in the occupied period, 18°C/20°C at night and 16°C/16°C in unoccupied periods. The internal gains are also set according to the standard. The second approach for modelling the occupants is to assume stochastic occupants. For this study, 100 occupant profiles were generated following the method of Baetens and Saelens (2015) and randomly allocated to the buildings. Obviously, the stochastic approach for the occupants adds more variability to both the deterministic and the probabilistic approach.

RESULTS

In this Section, the deterministic and the probabilistic approach are compared extensively for the Boxbergheide district. First, the focus is on the differences in allocated U-values. Then, the resulting distributions of the UA-value and the energy demand for space heating for the studied district are studied, while distinguishing between standard occupants and stochastic occupants.

Figure 3: The distributions of all U-values over the 820 studied buildings in the probabilistic approach, in contrast with the fixed U-value for all buildings within the deterministic approach.



For every building of the studied district, the CDFs for the U-values of the roof, the walls, the ground floor and the windows are predicted by the QR models. Next, particular U-values are sampled from these CDFs randomly, resulting in the U-value distributions on district level that are shown in Figure 3. In addition to the resulting U-value distributions of the probabilistic approach, the deterministic U-value is shown as well in Figure 3. On average, the deterministic U-value from TABULA is lower than the mean of the probabilistic U-values for this district. Multiple aspects cause the variability that is characteristic for existing districts. Differences in geometry, in building envelope fabrics and in occupant behaviour are considered within this study. Figure 4 compares both building UA-value and building energy demand for space heating for all four modelling approaches. The variations are also quantified in Table 3. The variability in UA-value and in energy demand for space heating due to the natural variability of the building geometry is shown by the deterministic approach with standard occupants (shown in grey in Figure 4). Including stochastic occupants rather than standard occupants almost doubles the variation on the energy demand for space heating (shown in blue in Figure 4), since the coefficient of variation increases from 0.178 to 0.296 (Table 3). As a result, buildings can have either a larger or a smaller energy demand for space heating, but on average their energy demand decreases from 25.4 MWh to 24.4 MWh (Figure 4).

Figure 4: Building UA-value and total energy demand for space heating as a function of the total loss area for all four modelling approaches.

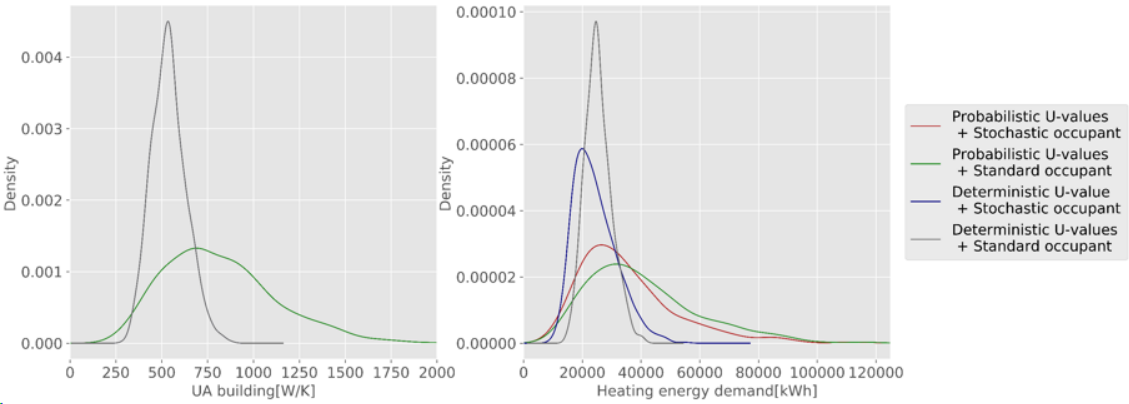


Table 3: Descriptive statistics of building UA-value and total energy demand for space heating for all four modelling approaches.

KPI →	UA-value of building [W/K]		Energy demand for space heating [kWh]			
Approach →	Deterministic	Probabilistic	Deterministic		Probabilistic	
Occupant → Statistic ↓	Standard = Stochastic	Standard = Stochastic	Standard	Stochastic	Standard	Stochastic
Mean	538.5	817.5	25372.1	24423.7	40564.1	35492.8
Standard deviation	94.1	304.9	4516.8	7227.7	18818.4	17373.5
Coeff. of variation	0.175	0.373	0.178	0.296	0.464	0.489
Minimum	298.8	242.3	14565.9	9894.5	9872.7	8222.4
First quartile (25%)	471.2	589.9	22198.3	18832.6	26602.5	23466.3
Median (50%)	535.0	777.3	24948.9	23176.5	36836.7	31385.4
Third quartile (75%)	598.4	995.1	28127.0	28791.3	50402.7	42786.2
Maximum	873.7	2253.0	41079.9	54731.1	118043.9	143531.9

By considering the natural variability in U-values, as achieved by the probabilistic approach, the variability in UA-value and in energy demand for space heating increases significantly compared to the deterministic approach for the standard occupants (shown in green in Figure 4), as the coefficients of variation increase respectively from 0.175 to 0.373 and from 0.178 to 0.464 (Table 3Table 3). By including stochastic occupants in the probabilistic approach rather than standard occupants (shown in red in Figure 4), the coefficient of variations on the energy demand for space heating increases even more from 0.464 to 0.489 (Table 3), but the impact of the stochastic occupants is not as large as within the deterministic approach. Based on the comparison of the deterministic approach with stochastic occupants and the probabilistic approach with standard occupants, including a natural variability on the building envelope fabric appears to enlarge the spread on the energy demand for space heating (coefficient of variation (CV) of 0.464) more than including stochastic occupants (CV of 0.296).

On district level, the energy demand for space heating differs among the deterministic and the probabilistic approach as a result of the differences in allocated building envelope fabrics. The percentage error of the probabilistic approach compared to the deterministic approach on the energy demand for space heating (calculated as (probabilistic – deterministic) / deterministic) is 59.9 % and 45.3 % respectively for the standard occupants and the stochastic occupants, while the percentage error on the building UA-value is 51.8 %. The smaller error for the stochastic occupants is explained by the lower temperature set points and thus energy use.

DISCUSSION

This paper presents a probabilistic method to introduce variability on the building envelope fabrics. To this end, quantile regression is deployed, as it enables to include the natural variability on the building envelope fabrics on two levels. First, as applied in this work, it enables to characterise districts in a probabilistic way and allocate a different U-value to every building as opposed to assuming a fixed U-value for all buildings of the 1980s. Second, it allows to characterise buildings in a probabilistic way and thus to perform uncertainty analyses of the energy demand on district level, which will be the scope of future work. This approach can also be extended to include building systems and thus quantify the district energy use in a probabilistic way.

Although this work highlights the added value of a probabilistic approach, the QR models should be enhanced. A disadvantage of the current models is that, since each output variable is estimated based on its particular QR model, the CDFs for the different output variables are not correlated. The dependence of the roof U-value on the wall U-value, as an example, was visually observed within the EPC dataset and this issue will therefore be corrected with novel methods in future work. Moreover, the use of the EPC dataset to determine distributions on the U-values should be considered with care because of two aspects. First, Energy Performance Certificates are only established before purchasing or renting a building and buildings are very likely to be renovated right after they have been purchased. Second, only Energy Performance Certificates of existing buildings are considered and their exact construction layers and materials are often unknown. If the exact composition of a construction element is unknown, conservative values are assumed. As a result, the EPC dataset gives a very conservative view of the existing building stock. For this reason, the assumed initial construction layers are not adjusted according to the EPC U-values and worse U-values are automatically replaced by a predefined maximal U-value. These drawbacks of using EPC data highlights the need for more accurate data on the current state of our existing building stock.

As a new probabilistic method to allocate building envelope fabrics was compared to the TABULA method, the resulting district models differ significantly in thermal performance, which encumber drawing conclusions for the influence on the energy demand on district level. However, previous literature (Reinhart & Cerezo Davila, 2016) showed that the error, using archetype approaches, was reasonably small compared to measured data. Nevertheless, the variation between the buildings is of significant importance for designing district energy systems (Happle et al., 2018) and defining the optimal location of production and storage units. Also, local implementation of renewable energy as well as energy flexibility assessment are benefitting from this work. Therefore, the conclusions of this work should be viewed from a district perspective, rather than from a city perspective. On city level, with a view to the occupancy model, deterministic models suffice for quantifying the annual energy demand and single building peak loads due to the averaging effect of stochastic models (Happle et al., 2018).

CONCLUSION

In this work, the existing building stock is characterised in a probabilistic manner by means of data-based CDFs of the U-values of the buildings and combined with stochastic occupancy data, allowing to capture the diversity in the existing building stock more accurately. This probabilistic approach was compared to a deterministic approach with a view to the building UA-values and the energy demand for space heating based on a district of 820 buildings while considering either standard or stochastic occupants. Including a probabilistic characterization of the building envelope fabrics increased the coefficient of variation for the studied district for the building UA-value from 17.5 % to 37.3 %, and for the energy demand for space heating from 17.8 % to 46.4 % and from 29.6 % to 48.9 % respectively for the standard and the stochastic occupant. Moreover, introducing variability on the building envelope fabric enlarges the spread on the energy demand for space heating more than including stochastic occupants. Including the spatial variation of the energy demand is beneficial for optimising the location of production and storage units, local implementation of renewable energy as well as energy flexibility assessment within district energy systems. The probabilistic definition of building archetypes, based on EPC data, is novel and is particularly interesting to perform uncertainty analyses and assess the impact of input data uncertainty on the district energy demand within future work.

ACKNOWLEDGEMENTS

This project receives the support of the European Union, the European Regional Development Fund ERDF, Flanders Innovation & Entrepreneurship and the Province of Limburg. The authors also gratefully acknowledge the Research Foundation Flanders (FWO) and the Flemish Institute for Technology (VITO) for funding this research. This research has also received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 675318 (INCITE). Additionally, the authors are thankful to the Flemish Energy Agency (VEA) for providing the EPC data of Flanders.

REFERENCES

Baetens, R., & Saelens, D. (2015). Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour. *Journal of Building Performance Simulation*, *9*(4), 431–447.

Biljecki, F., Ledoux, H., & Stoter, J. (2016). An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, *59*, 25–37.

Cuypers, D., Vandevelde, B., Van Holm, M., & Verbeke, S. (2014). *Belgische woningtypologie: Nationale brochure over de TABULA woningtypologie*.

De Jaeger, I., Reynders, G., & Saelens, D. (2017). Impact of spatial accuracy on district energy simulations. *Energy Procedia*, *132*, 561–566.

Happle, G., Fonseca, J. A., & Schlueter, A. (2018). A review on occupant behavior in urban building energy models. *Energy and Buildings*.

Jorissen, F., Reynders, G., Baetens, R., Picard, D., Saelens, D., & Helsen, L. (2018). Implementation and verification of the IDEAS building energy simulation library. *Journal of Building Performance Simulation*.

Kavgic, M., Mavrogianni, A., Mumovic, D., Summerfield, A., Stevanovic, Z., & Djurovic-Petrovic, M. (2010). A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, *45*(7), 1683–1697.

Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33.

Österbring, M., Mata, É., Thuvander, L., Mangold, M., Johnsson, F., & Wallbaum, H. (2016). A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model. *Energy and Buildings*, *120*, 78–84.

Protopapadaki, C., Reynders, G., & Saelens, D. (2014). Bottom-up modelling of the Belgian residential building stock: impact of building stock descriptions. In *Proceedings of 9th International Conference on System Simulation in Buildings* (Vol. 2, pp. 1–21). Liège, Belgium.

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling – A review of a nascent field. *Building and Environment*, *97*, 196–202.

Remmen, P., Lauster, M., Mans, M., Fuchs, M., Osterhage, T., & Müller, D. (2018). TEASER: an open tool for urban energy modelling of building stocks. *Journal of Building Performance Simulation*, *11*(1), 84–98.

Sokol, J., Cerezo Davila, C., & Reinhart, C. F. (2017). Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings*, *134*, 11–24.

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