HAL
open science

Global sensitivity analysis as a support for the generation of simplified building stock energy models

Alessio Mastrucci, Paula Perez-Lopez, Enrico Benetto, Ulrich Leopold, Isabelle Blanc

To cite this version:

Alessio Mastrucci, Paula Perez-Lopez, Enrico Benetto, Ulrich Leopold, Isabelle Blanc. Global sensitivity analysis as a support for the generation of simplified building stock energy models. Energy and Buildings, 2017, 149, pp.368 - 383. 10.1016/j.enbuild.2017.05.022 . hal-01564749

HAL Id: hal-01564749

https://minesparis-psl.hal.science/hal-01564749

Submitted on 19 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Global sensitivity analysis as a support for the generation of simplified building stock energy models

Alessio Mastrucci[a,1], Paula Perez-Lopez[b], Enrico Benetto[a], Ulrich Leopold[a], Isabelle Blanc[b]

[a] Luxembourg Institute of Science and Technology (LIST), Environmental Research and Innovation (ERIN) Department, 5, Avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg

[b] MINES ParisTech, PSL Research University, Centre Observation, Impacts, Energy (O.I.E.), CS 10207, F-06904 Sophia Antipolis, France

Abstract

Buildings are responsible for 40% of total final energy consumptions in Europe. Numerous bottom-up models were recently developed to support local authorities in assessing the energy consumption of large building stocks and reduction potentials. However, current models rarely consider uncertainty associated to building usage and characteristics within the stock, resulting in potentially biased results.

This study presents a generic model simplification approach using uncertainty propagation and stochastic sensitivity analysis to derive fast simplified (surrogate) models to estimate the current building stock energy use for improved urban planning. The methodology includes an engineering-based energy model as input to global sensitivity analysis (GSA) using the elementary effects (EE) screening and Sobol' method for key parameter identification and regression analysis to derive simplified models for entire building stocks.

The application to the housing stock of Esch-sur-Alzette (Luxembourg) showed that the parameters explaining most of the variability in final energy use for heating and domestic hot water are floor area, set-point temperature, external walls U-values, windows and heating system type. Results of the simplified models were validated against measured data and confirmed the validity of the approach for a simple yet robust assessment of the building stock energy use considering uncertainty and variability.

Keywords

Building stocks; Energy modelling; Urban housing; Uncertainty Analysis; Global Sensitivity Analysis; Parameter screening; Sobol' indexes; Surrogate models; Decision support; Sustainable urban planning.

---

[1] Present address: International Institute for Applied Systems Analysis (IIASA) - Energy Program (ENE), Schlossplatz, 1 A-2361 Laxenburg, Austria. Email: mastrucc@iiasa.ac.at

Abbreviations

A_f: Building footprint area (m^2);

A_n: Floor surface area (m^2);

CC: Construction class (light, medium, heavy);

DH: detached house;

DHW: Domestic hot water

e: Coefficient relative to the wind protection class (-);

EE: Elementary effects method

$F_{f,i}$: Shading correction factor for fins (-);

$F_{G,i}$: Glazed area fraction of windows (-);

F_g: Reduction factor for heating system control (-);

$F_{h,i}$: Shading correction factor for the horizon (-);

$F_{O,i}$: Shading correction factor for overhangs (-);

$F_{t,f}$: Coefficient of exposition of the ground floor (-);

$F_{t,r}$: Coefficient of exposition of the ceiling (-);

$F_{t,w}$: Coefficient of exposition of walls (-);

GIS: Geographical Information Systems;

Gl_{id}: Type of window;

GSA: Global Sensitivity Analysis;

HS_{id}: Heating system type;

MC: Monte-Carlo sampling;

Mean absolute percentage error (MAPE);

MFH: Multi-family house;

N: number of buildings;

n50: Tightness of the building envelope (h^{-1});

n_v: Ventilation air changes rate (h^{-1});

Orient: Orientation;

$Q_{E,V,T}$: Final energy use *(kWh/a);*

$q_{E,V}$: Final energy use intensity *(kWh/m²a);*

$Q_{h,M}$: Monthly heat demand for space heating (kWh);

$Q_{i,M}$: Monthly internal heat gains (kWh);

q_{iM}: Monthly internal heat gains intensity (W/m^2);

$Q_{s,M}$: Monthly solar heat gains (kWh);

$Q_{tl,M}$: Monthly heat losses (kWh);

RH: Row house;

RMSE: Root mean square error;

SFH: Single-family houses (detached);

S_i: First-order Sobol' indices;

S_{ij}: Higher-order Sobol' indices;

T_i: Indoor set-point temperature (°C);

U_f: U-value of the ground floor (W/m^2K);

U_r: U-value of the roof (W/m^2K);

U_w: U-value walls (W/m^2K);

V: Building volume (m^3);

V_{rL}: Share of the heated volume served by mechanical ventilation (-);

W: External walls area (m^2);

WWR: Windows-to-Wall ratio (-);

ΔU_{WB}: Correction factor for thermal bridges (W/m^2K);

η_L: Heat recovery efficiency for mechanical ventilation (-);

η_M: Heat gain utilization factor for monthly calculation (-).

1. Introduction

The building sector is responsible for 40% of the total final energy consumption in Europe [1] and has been identified as one of the most promising sectors in terms of carbon reduction potential [2]. In Europe, a wide legislative framework has been developed to promote the reduction of building energy consumption and carbon emissions, including the building energy performance directive [1] and the energy efficiency directive [3]. The latter entails a reduction of building consumption in a

stock perspective. Public authorities are therefore in the need to assess the energy demand of buildings at a large scale to support decision in energy planning strategies.

A number of bottom-up building stock energy models have been recently developed to cover this need (see e.g. [4–13]). The archetypes technique [14] is one of the most common approaches. In this approach, a series of typical buildings are identified to represent specific building cohorts, e.g. depending on the type of building, period of construction, etc. Then, the average energy performance of the individual typical buildings is calculated using an energy model and finally extrapolated to the regional or national level based on the representative weight of the modelled sample.

Bottom-up models represent a valuable tool to assess the energy performance of large building stocks. However, they are still affected by a number of shortcomings. Current models rarely account for the variability of building characteristics and usage within the same building cohort. This might limit the accuracy of the results due to an oversimplification of the building stock complexity. In addition, building stock models rely on a large number of input parameters that are commonly highly uncertain at a large scale due to limited data availability and accuracy. While uncertainty and sensitivity of input parameters have been widely investigated regarding the energy performance of individual buildings [15], only few studies focused on building stocks [16–20]. Uncertainty propagation and stochastic sensitivity analysis are therefore needed to assess the energy performance prediction and the identification of key parameters at the building stock level [18,21]. Simplified models, also known as surrogate models [22], based on uncertainty propagation and sensitivity analysis can provide policy makers with fast and reliable estimations to evaluate scenarios and energy efficiency strategies [18]. Simplified building stock models have therefore the potential of addressing uncertainty propagation by simplifying more complex engineering-based models.

This study aims at developing a methodology for the analysis of variability and uncertainty related to the building stock energy use and the subsequent generation of building cohort-specific simplified energy models, based on key parameters, for decision support in sustainable planning and policies. A generic model simplification approach using a complex engineering-based energy model in global sensitivity analysis (GSA) is developed to derive the simplified surrogate models.

The methodology consists of the following consecutive steps: 1) building stock characterization, 2) design of the engineering-based reference model, 3) GSA – including a) model input and parameter screening using the elementary effects (EE) method, b) characterization of input parameter probability distributions and Monte-Carlo (MC) sampling, c) calculation of the final energy use by propagation of uncertainties through the reference energy model and validation against measured or statistical energy consumption data, d) computation of Sobol' sensitivity indices at building cohort level for key parameter identification, 4) model simplification for each building cohort using multiple linear regression, 5) final aggregation at the stock level. Simplified models and building stock aggregation are finally presented as an easy-to-use tool to assess the overall building stock energy performance for decision support.

This generic simplification modelling approach contributes to improving the archetype technique as it accounts for the full variability and uncertainty linked to input parameters instead of considering one single representative building for every

cohort. The simplified models are capable of providing a simple, yet robust, estimation of the energy performance distribution for the different building cohorts and for the entire stock. The approach was tested and validated for the housing stock of Esch-sur-Alzette (Luxembourg).

2. State of the art

2.1. Building stock energy models

Building energy models are commonly classified into engineering, statistical and hybrid models [23,24]. Engineering (also called physical) models assess the energy use of buildings based on the solving of the equations that describe their physical behaviour and heat transfer. A variety of models exist, ranging from simplified steady-state models to detailed dynamic models [25]. Steady-state models use long time steps (monthly to seasonal) in the calculation of the energy balance and account for dynamic effects of the building mass in storing and releasing heat by empirical gain and loss utilization factors (quasi-steady-state methods). Dynamic models use short time steps to calculate the energy balance (e.g. hourly) and directly account for dynamic effects.

Statistical models estimate the energy behaviour of buildings by using a function deduced from samples of training data without using any explicit heat transfer modelling [24]. Regression analysis, conditional demand analysis and neural networks are among the techniques at the basis of statistical models. Hybrid (or "grey box") models aim at coupling engineering and statistical models by using several strategies. They are suitable to analyse the energy behaviour of buildings in case of incomplete or uncertain data [26]. Surrogate models are a particular case of hybrid models which allow a coherent data-driven model simplification assuring a replication of the more complex, often engineering-based, model accounting for uncertainties [22].

At the building stock level, bottom-up statistical and engineering methods have been widely applied [5,27]. They imply the estimation of the energy behaviour of individual (or groups of) buildings and end-uses and successive extrapolations at the building stock level. Conversely, hybrid building stock models have been rarely developed and have the potential to overcome some of the limits of current modelling approaches in properly taking into account uncertainty and variability related to large building stocks.

2.2. Uncertainty analysis

Uncertainty analysis focuses on quantifying uncertainty in model output [28] and is a major milestone to analyse complex models [21]. Uncertainty analysis has been widely applied to study the energy performances of individual buildings [29–31], highlighting the importance of including both physical and occupants' behaviour related parameters [32]. Whilst additional sources of uncertainty arise when the scale of the analysis expands from individual buildings to entire building stocks [16], only a few studies addressed uncertainty at this level [16–20]. The techniques used for uncertainty propagation at the building stock level include MC sampling [19,33] and factorial sampling analysis [34]. According to Booth et al. [16], four sources of

uncertainty can be discerned in building stock energy models: *chance variability* (also known as "first order" uncertainty), due to the random variability within a group of homogeneous buildings; *heterogeneity* due to the variation in characteristics of buildings belonging to different cohorts; *parameter uncertainty* (similar to "epistemic" uncertainty) due to parameters that could theoretically be measured if sufficient evidence was available ("second order" uncertainty) and assumptions; *ignorance* due to lack of knowledge. Our study addresses a method to handle *parameter uncertainty* and *heterogeneity* in the building stock analysis by applying the concept of building cohorts to isolate homogeneous groups of buildings and subsequently extrapolating results at the building stock level.

2.3. Sensitivity analysis

Sensitivity analysis studies how uncertainty associated to model outputs can be apportioned to different sources of uncertainty in the model input, parameters and structure [35]. Two main types of sensitivity analysis can be distinguished in building performance analysis [15]: local sensitivity methods and global sensitivity methods (GSA). In local sensitivity methods each input parameter is varied while holding the other constant to study the output variability. Global sensitivity methods assess the significance of input parameters by varying all other model inputs at the same time. Screening methods, such as the EE method [36], are qualified as global methods and can be used to reduce computational cost in high-dimensional or computationally expensive models [37]. EE typically does not characterise the uncertainty according to their complete probability distribution but rather randomly varies inputs by adding proportions and analyse the effects. This nevertheless, is useful when identifying the most sensitive model inputs for which a successive computationally expensive uncertainty propagation method will have to be performed.

At large scale (urban level or larger), most studies apply local sensitivity analysis because of its ease of use and in results interpretation whereas GSA techniques are rarely implemented due to heavier computation burdens. However, local sensitivity analysis is limited as it does not allow for the exploration of the whole input space and further inclusion of interaction between parameters [21].

We propose to use GSA taking advantage of previous studies which explored GSA for the Life Cycle Assessment (LCA) of energy pathways [38–41] and other fields [22,42,43]. Advantages of this method lie in its suitability for complex nonlinear and non-additive models and in its higher reliability of results even at the cost of increased computational time [15]. For these reasons, GSA techniques such as EE combined with Sobol' method were demonstrated to be appropriate and more robust than other techniques for building energy performance applications [44].

3. Methodology

This section describes the generic model simplification approach covering the following main steps (Fig.1): 1) building stock characterization based on building cohorts; 2) design of the reference energy model; 3) GSA for the identified building cohorts,

including input parameter screening (EE), uncertainty characterization and propagation (MC), identification of key parameters through Sobol' method; 4) simplification of the energy models; 5) aggregation of results at the stock level. This methodology is applicable to the analysis of existing building stocks in the current state. Possible adaptations to consider the future evolution of the building stock are addressed in the discussion (section 5).

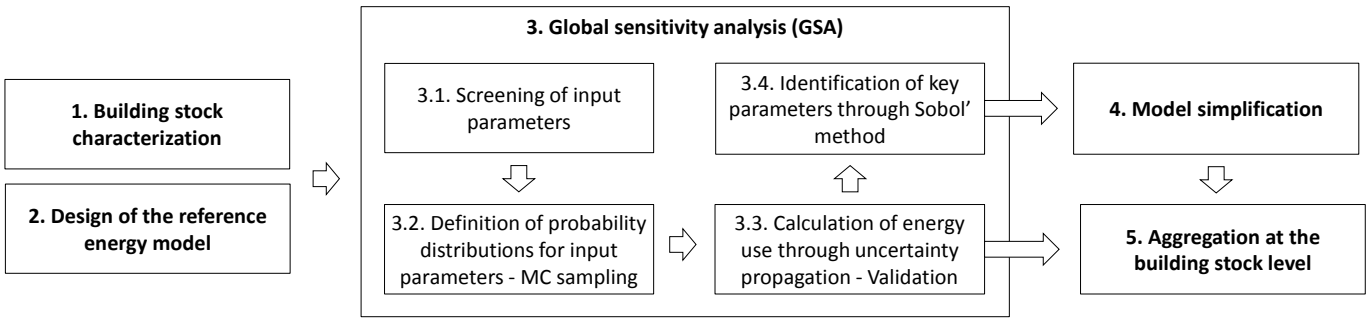

Figure 1 Steps of the generic model simplification approach.

3.1. Building stock characterization

The first step of the methodology consists in the identification of the targeted building stock and its characterization according to a series of building cohorts. The methodology can be directly applied for existing building stocks and buildings of different functions. Some adaptations of the methodology might be required for the application to new building stocks, e.g. to consider the uncertainty linked to the future evolution of the stock. The division of the building stock into cohorts is advised due to the great variability of characteristics among buildings and the resulting difficulty in having a unique model representing the entire building stock [14]. Grouping buildings by cohorts allows for a more accurate representation of the building stock thus maintaining the differences specific to each category.

Each cohort should be composed by buildings having relatively similar characteristics, for instant sharing the same building type and period of construction. A number of building libraries exist at the country or regional level, e.g. the project TABULA [45] provides a standard for a common definition of building typologies in European countries.

3.2. Design of the reference energy model

A reference engineering-based model for the calculation of energy performance of buildings should be selected. Steady-state models are easier to implement and computationally lighter, however they are limited in the calculation of peak loads, due to the simplification in accounting for dynamic effects of the building [23]. In contrast, dynamic models provide more accurate estimation of heating and cooling load, at the cost of higher inputs requirement and computation load [25,46]. Therefore, the choice of the modelling approach highly depends on the goal and scope of the study (e.g. estimation of heating/cooling load, building types, etc.), the possibility of automating simulation runs and the computational capacity.

In order to apply GSA using Sobol' method, a set of independent input parameters for the reference model is required. In the case of building energy models, dependent input parameters are not uncommon. If two parameters are not independent, a

possible solution is to keep the first one and introduce the second one by means of its residual variability [38,47]. Alternatively, other sensitivity methods may be applied instead of Sobol' method, for instance the Fourier Amplitude Sensitivity Testing – FAST [28]. However, the FAST method is not suitable for discrete distributions and it is not able to consider the interaction effects [15].

3.3. Global sensitivity analysis

The GSA step can be organized in four sub-steps. First, a screening of the input parameters is performed to select the most significant ones. Second, suitable probability distributions are identified for the selected parameters and a set of *n* scenarios are stochastically generated by a sampling method, for instance the Monte Carlo technique. Third, the model output (in this case, the energy use of buildings) is computed by running the reference energy model for the *n* scenarios. Finally, sensitivity indices are estimated for each parameters applying the Sobol' method, leading to the identification of key parameters.

3.3.1. Screening of input parameters

A screening step is suggested to identify the most important input parameters to be included in the GSA. This screening step is necessary as the number of input parameters is commonly high for energy models, regardless of the model type, and might result in unreasonable computation burdens for running the GSA. Another motivation to undertake this screening step is the fact that the boundary values must only be provided for each input parameter instead of the full values distribution required by the GSA which are potentially difficult and costly to determine.

The Elementary Effect (EE) method by Morris [36] is proposed as the screening method as it is simple but effective in identifying the important parameters among the many that can be contained in a model [28,48–50]. The computation cost for the method is $r \cdot (p+1)$ simulations, where p is the number of factors and r is the number of repetitions for the design. Two sensitivity measures are proposed by Morris: the mean μ estimating the overall influence of the factor on the output and the standard deviation σ assessing the effects due to interactions with other factors and nonlinear effects. The sensitivity analysis measures provided by this method are qualitative as they make it possible to rank input factors but not to quantify their effects on the output [15].

For every building cohort, the EE method is applied to isolate the most influencing parameters to be further characterized by identifying their probability distribution and subsequently introduced into the uncertainty analysis and GSA.

3.3.2. Characterization of the probability distribution of input parameters

This step consists in characterizing the probability distribution of the selected input parameters. Probability distribution can be obtained from different sources depending on the parameter, for instance from a statistical survey [19] or from previous studies and literature values [51]. Once the probability distribution of input parameters has been identified, the Monte-Carlo (MC) technique can be used to generate a set of random values following the given distributions.

3.3.3. Uncertainty propagation and validation of the energy use of buildings

In this step, the energy model is run for the sample of values stochastically generated in the previous step and a vector of energy use values is generated for each of the building cohorts. Results should be validated by comparison with a sample of measured values, statistical data for the region or country of interest, data from literature, etc. Validation can be performed at different levels, namely at the building level or cohort (aggregated) level, based on the energy consumption data availability.

3.3.4. Identification of key parameters through Sobol' method

The methodology used for the identification of key parameters is based on the Sobol' method. This technique is a variance-based method of GSA that consists in determining a set of indices called "Sobol' indices" or "variance-based sensitivity indices" [52,53]. Given the output $Y=f(x)$ of a deterministic model f (in this case the reference building stock energy model) under the conditions indicated by Sobol' [52], both the function itself and the corresponding variance can be decomposed in individual summands associated with each of the model inputs (or parameters) and their interactions. Hence, Sobol' indices express the share of the total variance of the output Y that is caused by a given parameter (first-order Sobol' indices) or a combination of parameters (n-order Sobol' indices) [53]. First-order Sobol' indices are calculated according to Eq. (1):

$$S_i = \frac{Var_i(Y)}{Var(Y)} \qquad (1)$$

Where S_i represents the first-order Sobol' index for parameter i, $Var_i(Y)$ is the variance due to the individual effect of parameter i, and $Var(Y)$ is the total variance of the output.

Higher-order indices can be calculated in an analogous way, as shown in Eq. (2) for second-order interactions:

$$S_{ij} = \frac{Var_{ij}(Y)}{Var(Y)} \qquad (2)$$

Where S_{ij} represents the Sobol' index for the effect of the interaction between parameter i and parameter j, $Var_{ij}(Y)$ is the variance due to the combined effect of parameters i and j and $Var(Y)$ is the total variance of the output. First-order and total-order indices were computed in this study.

3.4. Model simplification

According to the results of GSA, a set of key variables (generally the parameters with the highest Sobol' indices) are selected to be introduced into a simplified model. The number of selected parameters may vary depending on the objectives of the study. A possible criterion for their selection consists in defining a minimum threshold for the cumulative sum of Sobol' indices and picking up the parameters having the highest indices and contributing together to achieve the previously defined threshold.

A simplified model is generated for every building cohort based on the selected key parameters through multiple linear regression analysis and respective regression coefficients. A regression equation allows the assessment of the energy use for each building cohort based on key parameters only. In case of non-linear relationships or interactions among input parameters, higher-order or interaction terms may be added in the regression equation without increasing complexity of the model.

Nevertheless, we also have to warn for overfitting a regression model if interactions and higher orders do not explain substantially more of the overall variability. The fit provided by the simplified models should be checked against the original results by using suitable measures, e.g. root mean square error (RMSE), mean absolute percentage error (MAPE).

Two possible applications are envisaged for the simplified models (Fig. 2) *detailed* and *archetype-based*. In the *detailed* application, the model of each building cohort is fed with suitable probability distributions of the key input parameters and a distribution of the energy use is obtained as an output.

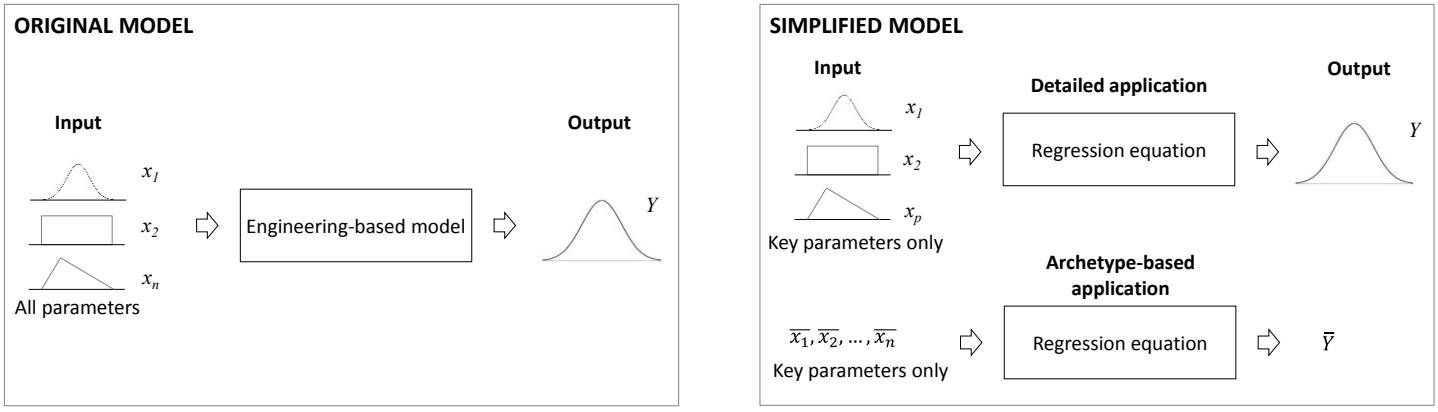

Figure 2 Overview of the original and simplified model applications for individual building cohorts.

In the case of the *archetype-based* application, a single set of average or typical values for the identified key parameters is assigned to each cohort in place of their probability distributions. The set of values describes an "archetype" building representative of the cohort. One single calculation is then run for each cohort, i.e. archetype building, by applying the regression equations. In contrast to the *detailed* application, only an average energy use value is obtained for each cohort instead of a full distribution.

3.5. Aggregation at the building stock level

Results are finally aggregated at the level of the building stock and compared with measured or statistical values for validation. The following equation can be applied to obtain the total energy use of the building stock $Q_{E,V,T}$:

$$Q_{E,V,T} = \sum_{i=1}^{n} N_i \cdot \overline{Q_{E,V,i}} \qquad (3)$$

where n is the number of building cohorts, N_i is the number of buildings belonging to the i-th cohort and $\overline{Q_{E,V,i}}$ the average energy use per building for the i-th cohort. For the original model and the detailed application of the simplified model, $\overline{Q_{E,V,i}}$ can be calculated as the median value of the output distribution. The median may be introduced instead of the mean value due to the recurrence of skewed distributions for energy use within the cohorts. Similarly, the first and third quartiles may be calculated as a measure of the deviation from the median energy use for every cohort and for the entire stock using proper error propagation formulas. For the archetype-based application of the simplified model, the values $\overline{Q_{E,V,i}}$ are directly obtained in output from the regression equations for each of the building cohorts.

The described methodology leads to the assessment of building stock energy use for the current state. For prospective studies focusing on the evolution of the building stock over time, adaptations might be needed to upscale results at the stock level accounting for the uncertainty on future size, composition of the stock and implementation of retrofitting measures.

4. Application to the case study and results

The generic methodology described above was applied and tested to the housing stock of Esch-sur-Alzette (Luxembourg), composed by more than 6000 residential buildings. The city was selected for its rich housing stock encompassing many building typologies and periods of construction and for the ready availability of quality geo-spatial data for buildings.

4.1. Building stock characterization

The classification of the housing stock was performed according to the national classification in the Second Energy Efficiency Action Plan [54]. Residential buildings are grouped in three types: single-family detached houses (DH), single-family row-houses (RH) and multi-family houses (MFH). Four periods of construction were considered instead of the three periods proposed by the National Plan. This choice is motivated by significant differences in the building geometry and distribution of heating systems highlighted by the analysis of the available dataset. Hence, a total of 12 building cohorts are obtained by combining the various housing types and periods of construction.

The geospatial dataset provided by the municipality of Esch-sur-Alzette was used to further characterise the housing stock. The dataset includes a vector polygon file for the year 2014 containing the building footprints and related attributes, for instance building functions and year of construction. In a previous study [55], we processed the building data and obtained information on the geometry and housing type of every building by means of Geographical Information Systems (GIS). Table 1 summarises the main characteristics of the housing stock in Esch-sur-Alzette obtained by GIS analysis. With the term floor surface area we refer here to the heated net floor area of the building (named *energy reference floor area* in [56]).

Table 1 Characteristics of the housing stock in Esch-sur-Alzette (Luxembourg) extracted from the database of the city.

Housing Type	Period of construction	Number of buildings	Median floor surface area *(m^2)*
DH	<1949	30	234
	1949-68	38	200
	1969-94	37	226
	>1994	11	270
RH	<1949	2418	150
	1949-68	1195	141
	1969-94	382	167
	>1994	232	151
MFH	<1949	910	332
	1949-68	333	413
	1969-94	303	575
	>1994	189	581

4.2. Design of the reference energy model

The reference energy model was designed based on the national regulation for the energy performance of buildings in Luxembourg [56]. This semi-steady model calculates the final energy use intensity per floor surface area for space heating and domestic hot water (DHW) using a monthly time step. Space cooling was not considered as it is significant only to a certain extend for residential buildings in Luxembourg and the National regulation recommends to avoid the installation of air conditioning systems in residential buildings.

The choice of a semi-steady state at a monthly resolution is motivated by the fact that it provides reasonable accuracy and computation load for urban energy scenarios for stocks containing thousands of buildings [57]. In addition, the exclusive focus on space heating and DHW further justify this choice as dynamic aspects are particularly important for the calculation of cooling loads.

4.2.1. Model description

The model follows a procedure similar to EN ISO 13790 [58] for the calculation of energy use intensity for space heating with a monthly time step. The monthly energy demand for space heating is computed by Eq. (4):

$$Q_{h,M} = Q_{tl,M} + \eta_M \cdot (Q_{s,M} + Q_{i,M}) \qquad (4)$$

where $Q_{h,M}$ *(kWh)* represents the heat demand for space heating, $Q_{tl,M}$ *(kWh)* the heat losses for ventilation and transmission in kWh, η_M *(-)* the heat gain utilisation factor depending on the thermal inertia on the building, $Q_{s,M}$ and $Q_{i,M}$ *(kWh)* the solar and internal heat gains. The final energy use intensity $q_{E,V}$ *(kWh/m²a)* is subsequently calculated by applying proper factors accounting for the efficiency of the heating and DHW systems and provided as a yearly value normalized on the reference floor surface area. Simplifications of the calculation for existing building were adopted in accordance with the regulation and regarded thermal bridges, shading coefficients, DHW demand, efficiency of the heating system and DHW systems. Describing the entire model is out of the scope of this paper and we refer to the cited regulation [56] for further information.

4.2.2. Model implementation

In order to proceed to the screening of input parameters and to the GSA, the reference model has to be designed with independent input parameters only. A scaling approach was chosen similarly to other studies [38], by expressing dependent parameters as a function of input parameters. Dependency was observed for several parameters linked to building geometry (floor surface area, volume, external walls area, etc.), building envelope (windows U-value and g-value) and technical systems (heating system type and efficiency).

Fig. 3 shows the relationship between all parameters considered in this study. Several authors have already investigated the relationship between building geometry-related parameters. Bon [59] and Steadman et al. [60] examined the relationship between building volume *(V)* and external wall area *(W)* and found that it follows the allometric rule $W \sim V^{0.77}$. This relationship can be explained as the building shape distorts to capture as much surface area as possible to get natural light as it increases in size. Analogous relationships were found by GIS analysis for the housing stock of Esch-sur-Alzette (Table 2), with

very similar results for RH and MFH building types. A slightly different relationship was found for DH, justified by the lower compactness of buildings belonging to this type. Relationships were also found between volume and footprint area, similarly to Steadman et al. [60], and between volume and reference floor area [55].

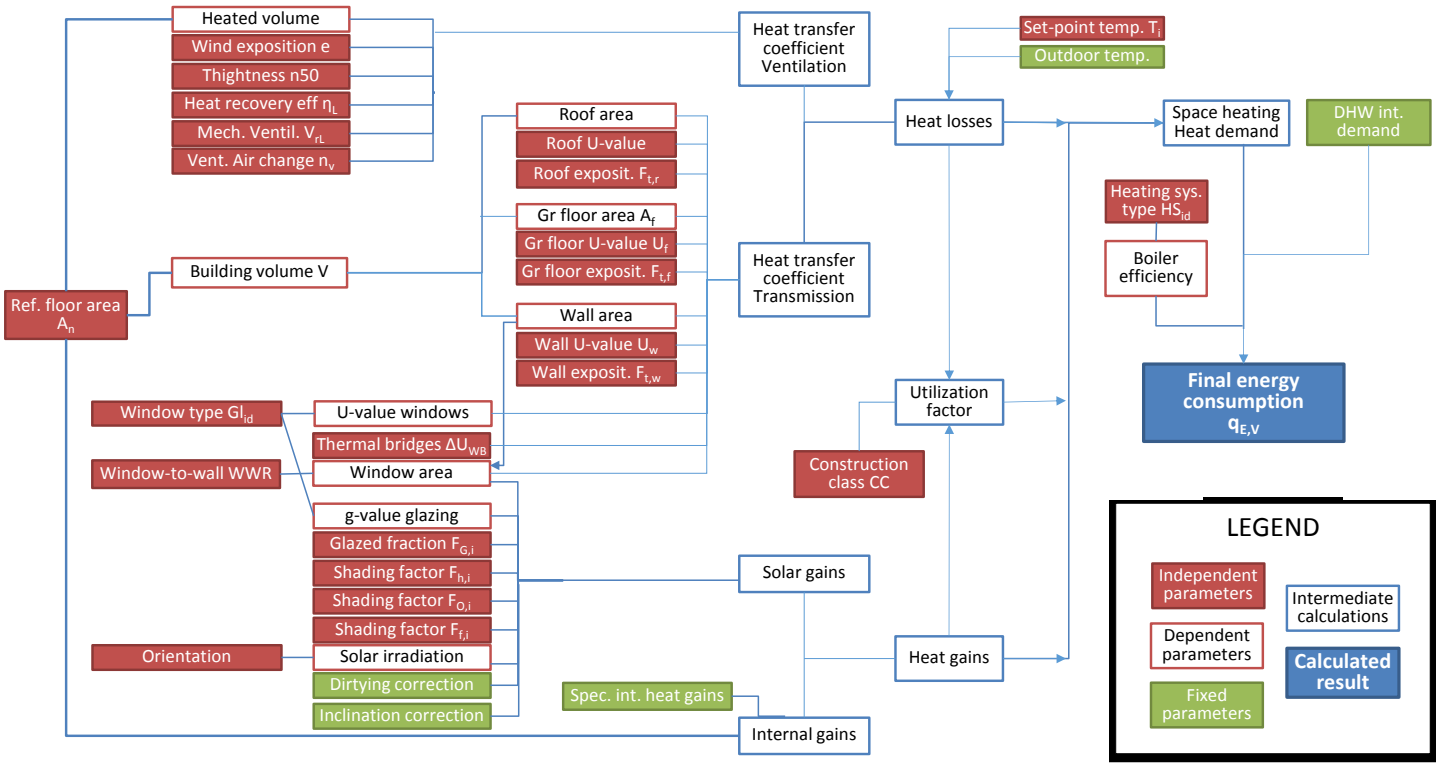

Figure 3 Overview of the reference engineering-based energy model.

Table 2 Relationship between geometrical parameters estimated for different types of residential buildings in Luxembourg.

Housing type	Relationship between geometrical parameters		
	External walls (W) and Volume (V)	Footprint area (A_f) and Volume (V)	Volume (V) and Floor area (A_n)
DH	$W \sim V^{0.84}$	$A_f \sim V^{0.66}$	$V \sim 4.4073\ A_n$
RH	$W \sim V^{0.77}$	$A_f \sim V^{0.66}$	$V \sim 4.4073\ A_n$
MFH	$W \sim V^{0.79}$	$A_f \sim V^{0.66}$	$V \sim 4.2213\ A_n$

Table 3 Input data for windows. Adapted from [61].

Window type	Glazing type	Frame type	Construction period	U-value	g-value
1	Single	Wood	Until 1968	5.00	0.87
2	Double	Aluminium (no thermal break)	1969-83	4.30	0.75
3	Double	Aluminium (with thermal break)	1984-94	3.20	0.75
4	Double	PVC	1969-94	3.00	0.75
5	Double	Wood	Until 1994	2.70	0.75
6	Double	PVC – Aluminium	From 1995	1.90	0.60
7	Double	Wood	From 1995	1.60	0.60

Table 4 Input data for heating systems depending on the heating system type, housing type and building heat demand. Adapted from [56].

Heating system type	Housing type	Building heat demand *($kWh/m^2\ a$)*				
		≤ 50	100	150	200	≥ 250
		Global heating system efficiency *(-)*				
1 - Conventional boiler	DH, RH	0.66	0.71	0.74	0.75	0.76

	MFH	0.75	[56,62]0.80	0.82	0.83	0.84
2 - Condensing boiler	DH, RH	0.78	0.85	0.87	0.89	0.90
	MFH	0.83	0.88	0.91	0.92	0.93

Windows and heating systems were implemented as discrete variables (Table 3-Table 4). A set of 7 window types were identified according to the national reference [61] and linked to specific U-values and g-values. Only natural gas and fuel oil heating systems were included in the analysis as other energy carriers represent a minor share in the housing stock. Heating system global efficiency values were assumed based on the national regulation [56] and depending on the boiler type (conventional, condensing), housing type and heating demand level. We finally implemented the reference energy model in a script written in R [63] and automated the calculation of the final energy use intensity of buildings.

4.3. Global sensitivity analysis

4.3.1. Screening of input parameters

A set of 25 parameters were selected for the initial screening of model inputs (Table 5). Parameters were further classified according to generally accepted groups: building positioning, geometry, envelope, occupants' behaviour and technological systems. Special care was dedicated in considering parameters where information is typically lacking at the building stock level, such as window-to-wall ratio (WWR) and exposition of envelope elements (e.g. presence of attic for the upper closure). Other studies have highlighted the importance of occupants' behaviour in addition to building configuration and physical parameters in regards to energy use [32,64]. We introduced three parameters related to occupants' behaviour being identified among the most influential for heating energy by other studies [51,65]: heating set point, ventilation rate and internal gains.

Boundary values were identified according to the results of the GIS analysis (floor surface area), normative limits and reference calculation values [56,61], other studies [62,66] and qualified assumptions.

Table 5 Input parameters selected for the screening step.

Parameter	Cohort	Min	Max	Data source
Building positioning				
Orient	-	North	South	QA
e (-)	-	0.04	0.07	Values corresponding to low and medium wind exposition [56]
Building geometry				
An (m^2)	DH <1949	82.0	657.9	GIS
	DH 1949-68	96.9	393.4	
	DH 1969-94	119.8	411.8	
	DH >1994	148.6	511.5	
	RH <1949	38.3	277.6	
	RH 1949-68	44.8	267.6	
	RH 1969-94	81.4	363.8	
	RH >1994	66.9	267.8	
	MFH <1949	75.3	753.1	
	MFH 1949-68	102.3	949.0	
	MFH 1969-94	145.2	1702.2	
	MFH >1994	126.0	2042.6	
WWR (-)	<1949; 1949-68	0.15	0.30	[66], QA
	1969-94; >1994	0.20	0.40	
Building envelope				

U_w (W/m^{2}K)	<1949; 1949-68	0.29	2.0	[56,61,62], QA
	1969-94	0.29	1.1	
	>1994	0.29	0.5	
U_r (W/m^{2}K)	<1949; 1949-68	0.23	2.5	[56,61,62], QA
	1969-94	0.23	1.4	
	>1994	0.23	0.8	
U_f (W/m^{2}K)	<1949; 1949-68	0.29	2.0	[56,61,62], QA
	1969-94	0.29	1.2	
	>1994	0.29	0.5	
CC	-	med.	heavy	Values corresponding to medium-weight and heavy construction [56]
ΔUWB (W/m2K)	<1949; 1949-68	0.05	0.15	[56]
	1969-94; >1994	0.05	0.10	[56]
Ft,w (-)	-	0.7	1.0	[56]
Ft,r (-)	-	0.8	1.0	[56]
Ft,f (-)	-	0.5	0.7	[56]
Glid	<1949; 1949-68	1	7	[61]
	1969-94	2	7	[61]
	>1994	6	7	[61]
FG,i (-)	-	0.6	0.9	[56]
Fh,i (-)	-	0.6	0.95	[56]
FO,i (-)	-	0.6	0.95	[56]
Ff,i (-)	-	0.75	0.95	[56]
n50 (h-1)	<1949; 1949-68	3.0	8.0	[56,62]
	1969-94; >1994	2.0	6.0	
Occupants' behaviour				
T_i (°C)	<1949; 1949-68	16.0	20.0	[56,62], QA
	1969-94; >1994	17.0	21.0	
n_v (h^{-1})	-	0.1	0.4	
q_{iM} (W/m^2)	DH; RH	2.24	3.36	[56]
	MFH	2.88	4.32	
Technological systems				
HS_{id}	-	1	2	[67], QA
F_g (-)		0.7	1.0	[56]
V_{rL} (-)	-	0.0	0.5	QA
η_L (-)	-	0.00	0.85	QA

Notes: QA = qualified assumption.

A screening test (EE method) was run for each of the 12 building cohorts described in Section 4.1 using the function "morris" in the R package "sensitivity" [63] and setting the number of repetitions *r to* 10. For each of the 12 cohorts, parameters were ranked based on the mean μ estimates. In order to select a set of sensitive input parameters for the GSA, we considered the number of times each parameter appeared in the first ten ranking out of the 12 runs and we finally selected the parameters scoring within the first ten ranking for at least 50% of the runs (Fig. 4). While many of the selected parameters are related to the building geometry (A_n, *WWR*) and envelope characteristics (U_r, $F_{t,r}$, U_w, U_f, *n50*), the heating set point temperature T_i and the type of heating system HS_{id} were also retained. Other parameters related to the occupants' behaviour resulted less influential and were excluded from the subsequent step. The complete results of the screening test are reported in the supplementary material.

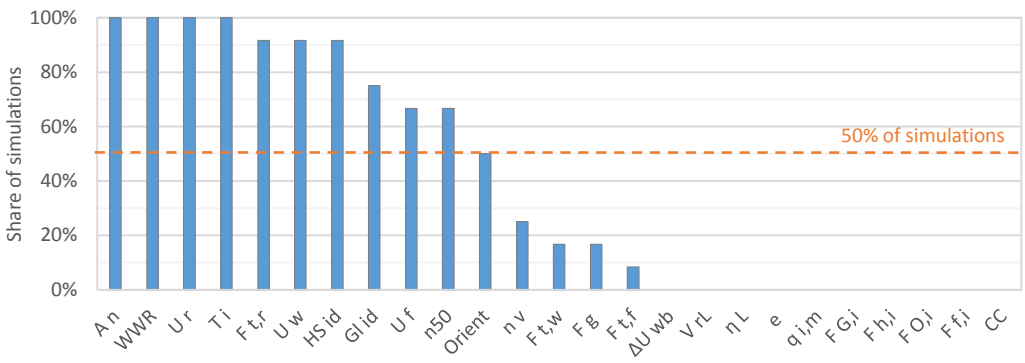

Figure 4 Results of the screening steps: times within the first ten ranking of input parameters out of the 12 runs. The parameters selected for the further GSA step are the ones above the dashed vertical line.

4.3.2. Characterization of the probability distribution of input parameters

Probability distributions of the 10 selected input parameters for GSA (Table 6) and were determined in several ways. Floor surface area *(A_n)* distributions were generated based on the building GIS data: lognormal distributions were fitted for each of the building cohorts and then tested against the original sample by means of the Kolmogorov-Smirnoff test [68]. The probability distributions for window-to-façade ratios, indoor set-point temperature and envelope tightness were identified according to relevant literature [19,51,62,66]. U-values were assumed to be normally distributed, similarly to Mauro et al. [19], and centred on typical values for Luxembourg depending on the period of construction [61]. Cut-offs were applied to U-value distributions in the proximity of normative limits according to the current construction practice. The discrete distributions for the type of heating system and windows were obtained based on statistical data [67], data on past renovations provided by the municipality and qualified assumptions. A series of 10 000 sample values for every input parameter and for each building cohort was finally generated by MC based sampling.

Table 6 Input values for the probability distribution of input parameters.

Parameter	Distribution	Housing Type	Construction period < 1949		1949-68		1969-94		>1994		Data sources
A_n *(m^2)*	Lognormal	DH	μ = 5.46	σ = 0.67	μ = 5.30	σ = 0.36	μ = 5.42	σ = 0.34	μ = 5.60	σ = 0.36	GIS
		RH	μ = 5.01	σ = 0.34	μ = 4.95	σ = 0.33	μ = 5.12	σ = 0.36	μ = 5.02	σ = 0.28	
		MFH	μ = 5.81	σ = 0.47	μ = 6.02	σ = 0.56	μ = 6.35	σ = 0.64	μ = 5.05	σ = 0.31	
WWR *(-)*	Lognormal	-	μ = -1.61	σ = 0.30	μ = -1.39	σ = 0.10	μ = -1.20	σ = 0.10	μ = -1.20	σ = 0.10	Values [66], QA. Distr. [51]
U_w *(W/m^2K)*	Trunc. Normal	-	μ = 1.10	σ = 0.33	μ = 1.10	σ = 0.33	μ = 0.80	σ = 0.24	μ = 0.50	σ = 0.15	Values [61,62]. Distr. [19].
U_r *(W/m^2K)*	Trunc. Normal	-	μ = 1.40	σ = 0.42	μ = 1.40	σ = 0.42	μ = 0.50	σ = 0.15	μ = 0.30	σ = 0.09	Values [61,62]. Distr. [19].
U_f *(W/m^2K)*	Trunc. Normal	-	μ = 1.20	σ = 0.36	μ = 1.20	σ = 0.36	μ = 0.80	σ = 0.24	μ = 0.60	σ = 0.18	Values [61,62]. Distr. [19].
Gl_{id} *(-)*	Discrete non-uniform	-	v = 1, 2, 3, 4, 5, 6, 7		v = 1, 2, 3, 4, 5, 6, 7		v = 2, 3, 4, 5, 6, 7		v = 6, 7		Values [61].
		DH	p = 10, 10, 10, 30, 30, 5, 5		p = 10, 5, 10, 30, 35, 5, 5		p = 5, 5, 35, 35, 10, 10		p = 50, 50		Distr. data past renovations,
		RH	p = 10, 10, 10, 30, 30, 5, 5		p = 10, 5, 10, 30, 35, 5, 5		p = 5, 5, 35, 35, 10, 10		p = 50, 50		QA.
		MFH	p = 10, 10, 10, 30, 30, 5, 5		p = 10, 5, 10, 30, 35, 5, 5		p = 15, 30, 25, 10, 10, 10		p = 70, 30		
$F_{t,r}$ *(-)*	Discrete uniform	-	v = 0.8, 0.9, 1.0		v = 0.8, 0.9, 1.0		v = 0.8, 0.9, 1.0		v = 0.8, 0.9, 1.0		[56], QA.
n50 *(h^{-1})*	Lognormal	-	μ = 1.39	σ = 0.20	μ = 1.39	σ = 0.20	μ = 1.10	σ = 0.10	μ = 1.10	σ = 0.10	Values [56,62]. Distr. [51]
T_i *(°C)*	Normal	-	μ = 17.0	σ = 1.0	μ = 17.0	σ = 1.0	μ = 19.0	σ = 1.0	μ = 20.0	σ = 1.0	Values [56,62], QA. Distr: [19,51]
HS *(-)*	Discrete non-uniform		v = 1, 2, 3, 4		v = 1, 2, 3, 4		v = 1, 2, 3, 4		v = 1, 2, 3, 4		[67], QA.
		DH	p = 27.3, 13.6, 50.5, 8.5		p = 34.9, 17.4, 40.8, 6.9		p = 31.1, 15.5, 45.7, 7.7		p = 40.7, 20.3, 33.4, 5.6		
		RH	p = 27.3, 13.6, 50.5, 8.5		p = 34.9, 17.4, 40.8, 6.9		p = 31.1, 15.5, 45.7, 7.7		p = 40.7, 20.3, 33.4, 5.6		
		MFH	p = 51.4, 25.7, 19.7, 3.3		p = 50.0, 24.9, 21.5, 3.6		p = 46.0, 23.0, 26.5, 4.5		p = 45.8, 22.8, 26.9, 4.5		

Notes: p = probability; v = values; μ = mean; σ = standard deviation; QA = qualified assumption.
Cut-off for truncated normal distributions (U_w, U_r, U_f):

4.3.3. Uncertainty propagation and validation of the energy use of buildings

Uncertainties were propagated throughout the engineering-based reference energy model using the generated sample of input values. The distributions of the final energy use intensity for space heating and DHW for the several building cohorts were then compared against reference and statistical values to check their reliability (Fig. 5). Results demonstrate a generally good agreement with reference values [54], being their difference between the calculated median lower than 10% in most cases. Additional comparison was performed against the distribution of measured final energy use intensity provided for a sample of buildings in Luxembourg by Merzkirch *et al* [62] showing a good agreement.

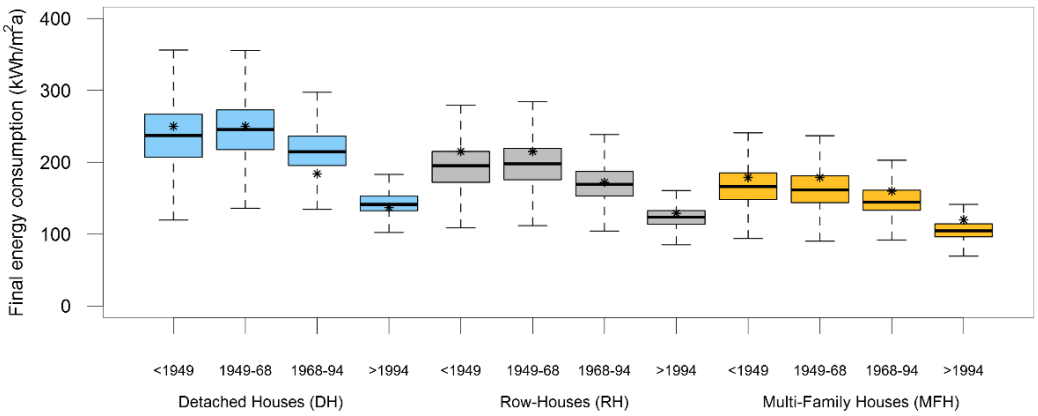

*Figure 5 Estimated distribution of final annual energy use intensity for heating and DHW in different housing types and periods of construction. Marked with *, reference values for Luxembourg* [54].

The results highlight that the cohorts composed of older buildings have higher energy use intensities and larger distributions due the higher degree of uncertainty and variability (e.g. uncertainty on past renovations, building characteristics and usage, wider range of construction years compared to the other periods of construction, etc.). DH have the higher energy use intensities, followed by RH and MFH, mainly because of the lower degree of compactness of their geometry (Table 2) resulting in higher heat losses per unit of volume.

4.3.4. Identification of key parameters through Sobol' method

After the validation of final energy use results, a GSA was performed for every building cohort and Sobol' sensitivity indices were determined out of 10000 simulation runs for every building cohort. Figs. 6-7 show the results of GSA for the housing stock of Esch-sur-Alzette depending on the housing type and period of construction.

Results demonstrate a significant difference in the first-order sensitivity indexes of input parameters among the different building categories. Overall, results are highly sensitive to geometric characteristics *(An)*, user behaviour *(Ti)* and heating systems type *(HS_{id})* for most of the building categories. The floor surface A_n is significant in particular for MFHs due to the high range of building sizes in the sample. The heating set-point temperature T_i has higher sensitivity indices for newer buildings because of the higher set-point values resulting in a larger gradient between indoor and outdoor temperature. Nevertheless, the high sensitivity indices obtained for the *(Ti)* confirm the importance of the occupants' behaviour even for

older buildings, where the building characteristics can commonly explain a large part of the variability [64]. The heating system type has a higher influence on the results for SFHs as the different energy carriers (gas, oil) and boiler types (traditional, conventional) are more evenly distributed whereas conventional gas boilers represent the majority of heating systems installed in MFHs.

Parameters related to the thermal characteristics of the building envelope, such as U-value of external walls and roof and the type of windows, demonstrated a higher influence for older buildings. This can be explained by both higher U-values and larger variability due to different building characteristics and renovation states of buildings. In contrast, more recent buildings are characterized by lower and more uniform values due to the thermal regulation requirements. Other parameters related to the building envelope characteristics have lower sensitivity indices in all simulations, namely $F_{t,r}$, *WWR*, U_f and *n50*. Results were similar for total-order indices and are reported in the supplementary material.

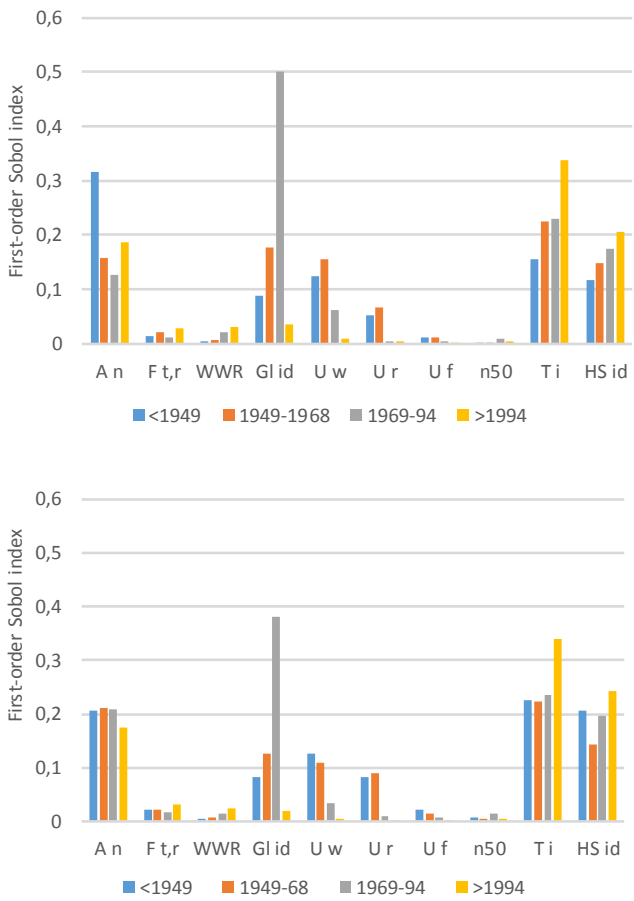

Figure 6 Results of the GSA for DH (left) and RH (right) and different construction periods.

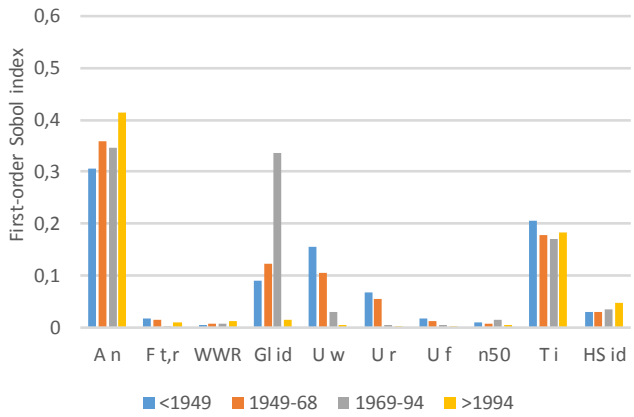

Figure 7 Results of the GSA for MFH of different construction periods.

4.4. Model simplification

A set of key parameters was selected for the generation of the simplified models based on the GSA results. For every building cohort, we set the threshold for the cumulative sum of Sobol' indices to 0.7 and selected the parameters having the highest indices and assuring together the threshold level. A multiple linear regression model was developed for every building cohort to get the final energy use intensity of buildings Y (kWh/m^2a) expressed as a function of the key parameters x_i specific of the cohort:

$$Y = \beta_0 + \sum_i^n \beta_i x_i + \varepsilon \quad (5)$$

where β_0 is the regression intercept and ε the random error term.

For every building cohort, the previously selected independent variables were introduced in the model. A logarithmic transformation was applied to A_n being the distribution of this variable lognormal and therefore manifestly skewed. This kind of transformation improves significantly the accuracy of the prediction in the case of variables with skewed distributions, as demonstrated by other authors [69,70]. The variables Gl_{id} and HS_{id} were treated as factorial. Higher order or interaction terms were not introduced in the equations, being the first-order Sobol' indices explicative of the largest part of the variability when compared with the total-order indices. The Ordinary Least Squares technique was used to fit the models and the model assumptions and accuracy of the results was carefully checked by means of proper diagnostics, including tests for heteroscedasticity, independence of errors, multi-collinearity and outliers.

The variables selected for every building cohort and the respective regression coefficients obtained by regression analysis are reported in Table 7. The results proved that an accurate prediction of the building final energy use intensity can be made by selecting only 3 up to 5 variables. Among the predictive variables, A_n and T_i are important for every building cohort as a consequence of their high sensitivity indices (Section 4.3.4). The variable U_w was selected only for buildings constructed until 1968 while Gl_{id} is significant mainly for buildings between 1949 and 1994. Finally, HS_{id} is used as a predictor for DH and RH, regardless of the period of construction, and for MFH built after 1994.

A simplified model was subsequently generated for every building cohort and expressed by the following generic equation:

$$Y = \beta_0 + \beta_1 \cdot \log(A_n) + \beta_2 \cdot Gl_{id} + \beta_3 \cdot U_w + \beta_4 \cdot T_i + \beta_5 \cdot HS_{id} \quad (6)$$

where $\beta_{0...}$ β_5 represent the regression coefficients specific to the considered building cohort (Table 7). Eq. (6) can be used to predict the final energy use intensity of buildings for each cohort accounting for the set of key parameters using the *detailed* or *archetype-based* applications of the simplified models.

Table 7 Results of the multiple linear regression for the prediction of the final energy use intensity for heating and DHW in different building cohorts.

Housing Type	Period of construction	Regression coefficients and respective input parameters						R^2	*Sum of Sobol' 1st ord. indices*
		β_0	β_1 $log(A_n)$	β_2 Gl_{id}	β_3 U_w	β_4 T_i	β_5 HS_{id}		
DH	<1949	122.65	-38.38	-	64.87	17.71	-37.19	0.775	0.714
	1949-68	173.53	-40.18	-11.38	62.14	18.46	-39.38	0.840	0.864
	1969-94	213.68	-29.81	-14.03	-	14.29	-33.86	0.802	0.978
	>1994	92.81	-16.47	-	-	8.46	-21.91	0.801	0.729
RH	<1949	133.27	-38.06	-	44.67	14.15	-30.46	0.708	0.765
	1949-68	168.38	-39.17	-7.75	42.82	14.54	-31.17	0.798	0.811
	1969-94	185.28	-28.87	-9.64	-	11.08	-26.32	0.813	0.979
	>1994	94.39	-16.39	-	-	6.70	-18.62	0.834	0.756
MFH	<1949	116.16	-28.04	-5.37	38.72	11.34	-	0.752	0.754
	1949-68	124.23	-27.05	-6.46	35.10	11.19	-	0.779	0.766
	1969-94	139.16	-20.09	-7.36	-	8.72	-	0.739	0.853
	>1994	78.84	-12.94	-	-	6.16	-10.84	0.888	0.700

Results of the simplified model (*detailed* application) were compared with the results of the original model for each cohort showing a good agreement (Fig. 8). RMSE and MAPE were additionally calculated and are reported in the supplementary material. Input parameters and results of the simplified models (*archetype-based* application) are reported in Table 8. Comparison of the final energy use intensities with the median values of the distributions obtained using the original model showed a low deviations, being the error lower than 5% for all building cohorts.

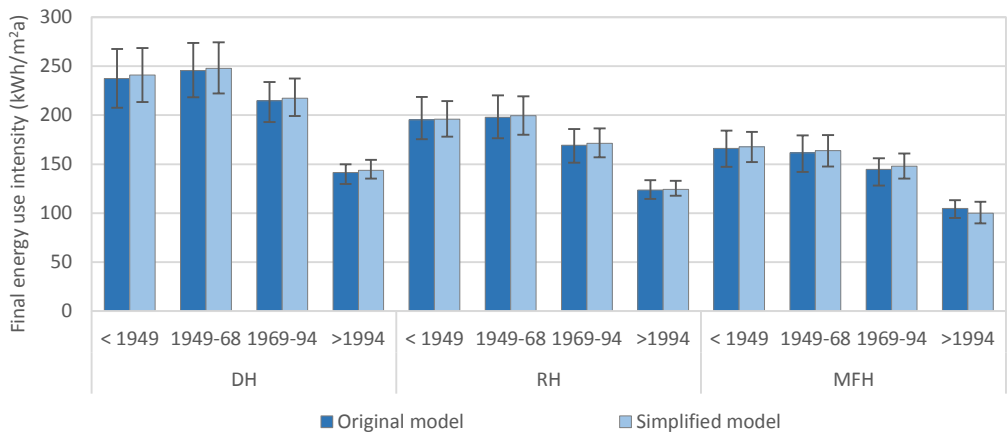

Figure 8 Final energy use intensity for heating and DHW of different building cohorts as a result of the original and detailed simplified models. The bars indicate the median values and the whiskers the first and third quartiles.

Table 8 Results of the archetype-based simplified models for each building cohort and comparison with the original model.

Housing type	Period of construction	Input parameters - simplified model	Final energy for heating and DHW *(kWh/m²a)*

		A_n (m^2)	Gl_{id} (-)	U_w (W/m^2K)	T_i (°C)	HS_{id} (-)	Simplified model	Original model (median)	Error (%)
DH	<1949	234	-	1.1	17	1	248.04	237.31	4.3
	1949-68	200	4	1.1	17	1	257.96	245.61	4.8
	1969-94	226	5	-	19	1	219.53	214.67	2.2
	>1994	270	-	-	20	1	148.10	141.35	4.6
RH	<1949	150	-	1.1	17	1	202.03	195.30	3.3
	1949-68	141	4	1.1	17	1	206.94	197.93	4.4
	1969-94	167	5	-	19	1	173.52	169.39	2.4
	>1994	151	-	-	20	1	127.44	123.73	2.9
MFH	<1949	332	4	1.1	17	-	167.39	166.23	0.7
	1949-68	413	4	1.1	17	-	164.36	161.62	1.7
	1969-94	575	4	-	19	-	147.77	144.64	2.1
	>1994	581	-	-	20	1	108.89	104.70	3.8

4.5. Aggregation at the building stock level

Results of the original and simplified models were aggregated at the stock level for the city of Esch-sur-Alzette (Fig. 9) using Eq. (3) and data on the number of buildings (Table 1). DH cohort has the lowest total energy use due to its limited number of buildings in the stock. The high energy use values for RH and MFH built before 1949 can be explained by the combination of the large number of buildings and a high energy use intensity. Uncertainty intervals were derived from the computation of first and third quartiles for the original and simplified models (detailed application). They are also larger for cohorts with older buildings as a result of the greater uncertainty and variability depending, for instance, on past renovations and less homogeneous building envelope and heating systems characteristics. Uncertainty intervals become smaller by moving to cohorts including more recent buildings due to lower uncertainty and variability, e.g. due to the alignment with new thermal regulations. The results demonstrated a small difference between the original, detailed and archetype-based applications of the simplified models.

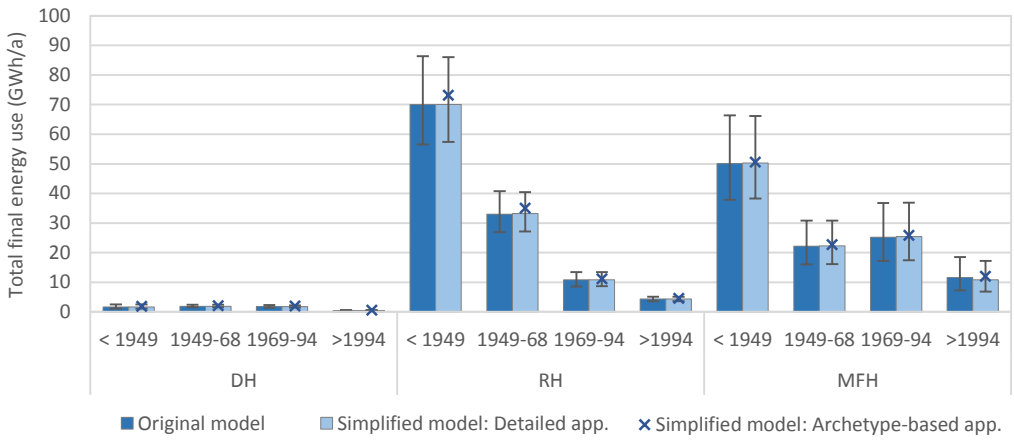

Figure 9 Total final energy use for heating and DHW and uncertainty intervals for different building cohorts using different models at the stock level.

Results were finally aggregated at the entire stock level as reported in Table 9. A total final energy use of *232.92 GWh/a* with an uncertainty interval ranging from *-9.6%* to *+12.5%* was obtained using the original model. Results of the detailed application of the simplified model are very similar, being the difference with the original prediction lower than *1%* and the

uncertainty intervals comparable with the original model. The prediction given by the archetype-based application of the simplified model is slightly higher than the original prediction being the total energy use *240.34 GWh/a* and the error around *3%.*

The calculated total final energy use values were finally compared with measured values provided by the municipality for validation purpose. The measured consumption falls within the uncertainty range provided by the original model and resulted 5% lower than the predicted value. The same conclusions apply for the detailed application of the simplified model, while the archetype-based application introduces a slightly higher overestimation of the total energy use. Overall, results are considered as satisfactory due to the high level of uncertainty and variability occurring at the level of the entire housing stock.

Table 9 Total final energy use for heating and DHW and uncertainty intervals predicted by different models at the housing stock level.

Model	Total final energy use (GWh/a)			Error with the original model prediction (%)
	Inferior limit	Prediction	Superior limit	
Original	210.64	**232.92**	262.14	0.00
Simplified: Detailed application	211.31	**232.85**	261.41	0.03
Simplified: Archetype-based application	-	**240.34**	-	3.18

5. Discussion

This study introduced a generic approach for the simplification of energy models for building stocks in the current state combining the archetype aggregation approach, GSA to identify the key parameters and regression analysis to get surrogate models per building cohort. The identification of key parameters is important in building stock energy modelling due to the high level of uncertainty and variability. Splitting energy modelling per building cohort was found to be necessary to handle the quite heterogeneous behaviour of each building cohort through tailored cohort model.

The generated simplified models provide a simple but robust tool to stakeholders and decision-makers for assessing both the energy use of each building cohort and the entire building stock energy use without running a full energy model. Compared to current engineering-based models, e.g. archetypes, this methodology makes it possible to fully account for the uncertainty and variability in the stock in contrast with deterministic approaches. Another advantage of this methodology lies in its generality and applicability to several contexts and scales, e.g. ranging from urban to national scale. The methodology can be adapted depending on the goal of the study and used regardless of the type of energy model.

The application to the test case city of Esch-sur-Alzette proved the validity of the approach in assessing the current final energy use of residential buildings for space heating and DHW under uncertainty and variability. Results of the GSA demonstrated that it is possible to cover most of the variability in energy use by considering a limited number of key parameters. The simplified models generated out of the key parameters are denoted by small deviation of the results from the original complete model (less than *1%* for the detailed application of the simplified model and around *3%* for the archetype-based application for total energy use) and from measured values. The developed simplified models and equations can be further used for similar study in other cities in Luxembourg and neighbouring countries. Nevertheless, the distribution of input

parameters, in particular floor surface area, building characteristics and technical systems, might be different from Esch-sur-Alzette and should be carefully verified prior to the application. In all other cases, the same generic methodology can be applied to develop simplified models from the urban to the national scale.

Limitations of this methodology lie in the need for detailed input data on the building stock necessary to generate probability distributions of input parameters. Lacking sufficiently detailed information and making imprecise assumptions might result in inaccurate results [71]. However, the increasing availability of GIS data on buildings will enhance and facilitate the generation of probability distributions, especially for parameters related to geometry and construction characteristics.

Occupants' behaviour is an important source of uncertainty for the energy use of buildings [32]. In this study we considered a limited number of parameters, such as heating set point temperature, ventilation rates and internal gains, based on the indications of previous studies [51,65] and we confirmed the importance of heating set point temperature as one of the most influential parameters on the energy use for existing housing. Additional parameters, e.g. occupation and operation schedules, should be included in future studies to address occupants' behaviour in a more comprehensive way, especially for new or retrofitted buildings where the occupants' behaviour is expected to have a higher influence on the energy use.

This assessment was limited to the present state of the building stock and serves as a basis for diagnosis of current energy use under uncertainty and for the identification of hotspots. However, uncertainty in the future evolution of the stock should be included to properly address prospective mitigation scenarios. Future developments are envisaged to improve the upscaling methodology so to include the modelling of stock dynamics and are expected in two directions: accounting of construction and demolition activities leading to a different size and composition of the stock; retrofitting of existing buildings to improve the energy performance of existing buildings.

Regarding the Sobol' method, independent input parameters are required. Input parameters in energy models are frequently interrelated, demanding for further pre-processing and adaptation to make the Sobol' method applicable. Nevertheless, the identified relationships of dependence between input parameters can be generalised and further support the development of analogous models for other contexts and building types. Another limitation is represented by the possible high computation burden associated with running multiple energy simulations and GSA, especially if a dynamic energy model is used. This can, nevertheless, be improved through stratification in the sampling to minimise the number of Monte Carlo runs, e.g. Latin Hypercube Sampling [72]. Also the parallelisation and usage of distributed computing approaches, such as Apache Spark can be a solution to solve computational burden by using more efficiently cloud computing infrastructures [73].

In spite of these limitations, the methodology has proven to be effective in identifying key parameters and generating simplified models for the assessment of building stock energy use for energy planning and policy support.

6. Conclusion

This paper presented an innovative approach to develop fast simplified models while accounting variability and uncertainty related to the energy modelling of a building stock in the current state. The model simplification is based on the identification of key input parameters using Sobol' indices in a GSA, representative of the stock. This simplified modelling approach helps to support decision making for sustainable energy planning.

The application of the methodology to the existing housing stock of Esch-sur-Alzette (Luxembourg) showed that the most significant input parameters explaining the variability of final energy use are the floor surface area, indoor set-point temperature, U-value of external walls, window type and heating system type. Results of the simplified models at the building stock scale were validated against aggregated measured data and confirmed its validity for a simple yet robust assessment of the final energy use of large building stocks.

The proposed approach and the developed simplified models represent an effective tool to support local authorities in assessing the current energy performance of large building stocks which represent an essential step for the identification of hotspots and consequent mitigation strategies. Its generic formulation allows further development and application to different context and scales, from the urban to the national scale. Future developments include the extension of the approach to the environmental impact assessment of large building stocks from a LCA perspective, the application to other contexts, the refinement of the upscaling methodology and the inclusion of building stock dynamics to assess uncertainty in the development of future mitigation scenarios.

Acknowledgements

This study is part of the postdoctoral project "DAEDALUS" supported by the National Research Fund, Luxembourg (FNR), Grant agreements AFR - 7579115. MINES ParisTech, PSL Research University is kindly acknowledged for having supported the scientific visit of Alessio Mastrucci which made this work possible at MINES ParisTech, Centre O.I.E. (Observation Impacts Energy), in particular with Prof. Philippe Blanc. We would like to thank the City Council of Esch-sur-Alzette for contributing with relevant discussion and data provision.

References

[1] European Parliament, Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast), Off. J. Eur. Union. (2010) 13–35. doi:doi:10.3000/17252555.L_2010.153.eng.

[2] Z. Ma, P. Cooper, D. Daly, L. Ledo, Existing building retrofits: Methodology and state-of-the-art, Energy Build. 55 (2012) 889–902. doi:10.1016/j.enbuild.2012.08.018.

[3] European Parliament, Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on

energy efficiency, Off. J. Eur. Union Dir. (2012) 1–56. doi:10.3000/19770677.L_2012.315.eng.

[4] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835. doi:10.1016/j.rser.2008.09.033.

[5] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfeld, Z. Stevanovic, M. Djurovic-Petrovic, A review of bottom-up building stock models for energy consumption in the residential sector, Build. Environ. 45 (2010) 1683–1697.

[6] É. Mata, A. Sasic Kalagasidis, F. Johnsson, Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK, Build. Environ. 81 (2014) 270–282. doi:10.1016/j.buildenv.2014.06.013.

[7] M. Österbring, É. Mata, L. Thuvander, M. Mangold, F. Johnsson, H. Wallbaum, A differentiated description of building-stocks for a georeferenced urban bottom-up building-stock model, Energy Build. 120 (2016) 78–84. doi:10.1016/j.enbuild.2016.03.060.

[8] R. Nouvel, M. Zirak, V. Coors, U. Eicker, The influence of data quality on urban heating demand modeling using 3D city models, Comput. Environ. Urban Syst. 64 (2017) 68–80. doi:10.1016/j.compenvurbsys.2016.12.005.

[9] G. Dall'O', A. Galante, M. Torri, A methodology for the energy performance classification of residential building stock on an urban scale, Energy Build. 48 (2012) 211–219. doi:10.1016/j.enbuild.2012.01.034.

[10] I. Theodoridou, M. Karteris, G. Mallinis, A.M. Papadopoulos, M. Hegger, Assessment of retrofitting measures and solar systems' potential in urban areas using Geographical Information Systems: Application to a Mediterranean city, Renew. Sustain. Energy Rev. 16 (2012) 6239–6261.

[11] A. Mastrucci, O. Baume, F. Stazi, U. Leopold, Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam, Energy Build. 75 (2014) 358–367.

[12] B. Howard, L. Parshall, C. Thompson, S. Hammer, J. Dickinson, V. Modi, Spatial distribution of urban building energy consumption by end use, Energy Build. 45 (2012) 141–151.

[13] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling - A review of a nascent field, Build. Environ. 97 (2016) 196–202. doi:10.1016/j.buildenv.2015.12.001.

[14] S. Moffatt, Stock Aggregation. Methods for Evaluating the Environmental Performance of Building Stocks, 2004. http://www.iisbe.org/annex31/index.html.

[15] W. Tian, A review of sensitivity analysis methods in building energy analysis, Renew. Sustain. Energy Rev. 20 (2013) 411–419. doi:10.1016/j.rser.2012.12.014.

[16] A.T. Booth, R. Choudhary, D.J. Spiegelhalter, Handling uncertainty in housing stock models, Build. Environ. 48 (2012) 35–47. doi:10.1016/j.buildenv.2011.08.016.

[17] A.T. Booth, R. Choudhary, Decision making under uncertainty in the retrofit analysis of the UK housing stock: Implications for the Green Deal, Energy Build. 64 (2013) 292–308. doi:10.1016/j.enbuild.2013.05.014.

[18] M. Kavgic, D. Mumovic, a. Summerfield, Z. Stevanovic, O. Ecim-Djuric, Uncertainty and modeling energy

consumption: Sensitivity analysis for a city-scale domestic energy model, Energy Build. 60 (2013) 1–11. doi:10.1016/j.enbuild.2013.01.005.

[19] G.M. Mauro, M. Hamdy, G.P. Vanoli, N. Bianco, J.L.M. Hensen, A new methodology for investigating the cost-optimality of energy retrofitting a building category, Energy Build. 107 (2015) 456–478. doi:10.1016/j.enbuild.2015.08.044.

[20] F. Branger, L.-G. Giraudet, C. Guivarch, P. Quirion, Global sensitivity analysis of an energy–economy model of the residential building sector, Environ. Model. Softw. 70 (2015) 45–54. doi:10.1016/j.envsoft.2015.03.021.

[21] J. Keirstead, M. Jennings, A. Sivakumar, A review of urban energy system models: Approaches, challenges and opportunities, Renew. Sustain. Energy Rev. 16 (2012) 3847–3866. doi:10.1016/j.rser.2012.02.047.

[22] A.N. Aijazi, L.R. Glicksman, Comparison of regression techniques for surrogate models of building energy performance, ASHRAE IBPSA-USA SimBuild 2016 - Build. Perform. Model. Conf. (2016) 327–334. https://www.ashrae.org/File Library/docLib/Events/Simbuild2016/Papers/C043.pdf.

[23] N. Fumo, A review on the basics of building energy estimation, Renew. Sustain. Energy Rev. 31 (2014) 53–60. doi:10.1016/j.rser.2013.11.040.

[24] A. Foucquier, S. Robert, F. Suard, L. Stéphan, A. Jay, State of the art in building modelling and energy performances prediction: A review, Renew. Sustain. Energy Rev. 23 (2013) 272–288. doi:10.1016/j.rser.2013.03.004.

[25] H. Wang, Z. (John) Zhai, Advances in building simulation and computational techniques: A review between 1987 and 2014, Energy Build. 128 (2016) 319–335. doi:10.1016/j.enbuild.2016.06.080.

[26] H. Zhao, F. Magoules, A review on the prediction of building energy consumption, Renew. Sustain. Energy Rev. 16 (2012) 3586–3592. doi:http://dx.doi.org/10.1016/j.rser.2012.02.049.

[27] L. Swan, V. Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835.

[28] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis, The primer, John Wiley & Sons, 2008. doi:10.1002/9780470725184.ch6.

[29] K.J. Lomas, H. Eppel, Sensitivity analysis techniques for building thermal simulation programs, Energy Build. 19 (1992) 21–44.

[30] S. De Wit, G. Augenbroe, Analysis of uncertainty in building design evaluations and its implications, Energy Build. 34 (2002) 951–958. doi:10.1016/S0378-7788(02)00070-1.

[31] I.A. Macdonald, J.A. Clarke, Applying uncertainty considerations to energy conservation equations, Energy Build. 39 (2007) 1019–1026. doi:10.1016/j.enbuild.2006.11.008.

[32] A.S. Silva, E. Ghisi, Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation, Energy Build. 76 (2014) 381–391. doi:10.1016/j.enbuild.2014.03.001.

[33] T. Pettersen, Variation of energy consumption in dwellings due to climate, building and inhabitants, Energy Build. 21 (1994) 209–218.

[34] Y. Yamaguchi, J.L.M. Hensen, Y. Shimoda, T. Asai, M. Mizuno, District level energy management using a bottom-up modeling approach, Ninth Int. IBPSA Conf. (2005) 1385–1392.

[35] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo, Sensitivity analysis practice: A guide to scientific models, 2004. doi:10.1016/j.ress.2005.11.014.

[36] M.D. Morris, Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics. 33 (1991) 161–174. doi:10.2307/1269043.

[37] F. Zhao, S.H. Lee, G. Augenbroe, Reconstructing building stock to replicate energy consumption data, Energy Build. 117 (2016) 301–312. doi:10.1016/j.enbuild.2015.10.001.

[38] P. Padey, R. Girard, D. Boulch, I. Blanc, From LCAs to Simplified Models: A Generic Methodology Applied to Wind Power Electricity, Environ. Sci. Technol. 47 (2012) 1231−1238. doi:dx.doi.org/10.1021/es303435e.

[39] C. Marini, I. Blanc, Towards prospective life cycle assessment: How to identify key parameters inducing most uncertainties in the future? Application to photovoltaic systems installed in Spain, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 8581 LNCS (2014) 691–706. doi:10.1007/978-3-319-09150-1_51.

[40] C. Marini, P. Padey, I. Blanc, D. Le Boulch, Assessing the prospective environmental impacts of photovoltaic systems based on a simplified LCA model, in: SETAC Eur. SETAC Eur. 24th Annu. Meet., Basel, 2014.

[41] M. Lacirignola, B. Meany, P. Padey, I. Blanc, A simplified model for the estimation of life-cycle greenhouse gas emissions of enhanced geothermal systems, Geotherm. Energy. 2 (2014) 8. doi:10.1186/s40517-014-0008-y.

[42] S. Cucurachi, R. Heijungs, Characterisation factors for life cycle impact assessment of sound emissions, Sci. Total Environ. 468–469 (2014) 280–291. doi:10.1016/j.scitotenv.2013.07.080.

[43] P. Azadi, G. Brownbridge, S. Mosbach, O. Inderwildib, M. Kraft, Simulation and life cycle assessment of algae gasification process in dual fluidized bed gasifiers, Green Chem. 17 (2015) 1793–1801. doi:10.1039/C4GC01698J.

[44] A.-T. Nguyen, S. Reiter, A performance comparison of sensitivity analysis methods for building energy models, Build. Simul. 8 (2015) 651–664. doi:10.1007/s12273-015-0245-4.

[45] TABULA National Building Typologies. Last consulted: February 2017., (n.d.). http://episcope.eu/building-typology/.

[46] B. Peuportier, I. Blanc, Simulation tool with its expert interface for the thermal design of multizone buildings, Int. J. Sol. Energy. 8 (1990) 109–120.

[47] I. Kioutsioukis, S. Tarantola, A. Saltelli, D. Gatelli, Uncertainty and global sensitivity analysis of road transport emission estimates, Atmos. Environ. 38 (2004) 6609–6620. doi:10.1016/j.atmosenv.2004.08.006.

[48] Y. Mery, L. Tiruta-Barna, E. Benetto, I. Baudin, An integrated “process modelling-life cycle assessment” tool for the

assessment and design of water treatment processes, Int. J. Life Cycle Assess. 18 (2013) 1062–1070. doi:10.1007/s11367-012-0541-5.

[49] Y. Mery, L. Tiruta-Barna, I. Baudin, E. Benetto, E. Igos, Formalization of a technical procedure for process ecodesign dedicated to drinking water treatment plants, J. Clean. Prod. 68 (2014) 16–24. doi:10.1016/j.jclepro.2014.01.011.

[50] Andrianandraina, A. Ventura, T. Senga Kiess??, B. Cazacliu, R. Idir, H.M.G. van der Werf, Sensitivity Analysis of Environmental Process Modeling in a Life Cycle Context: A Case Study of Hemp Crop Production, J. Ind. Ecol. 19 (2015) 978–993. doi:10.1111/jiec.12228.

[51] N. Heeren, C.L. Mutel, B. Steubing, Y. Ostermeyer, H. Wallbaum, S. Hellweg, Environmental Impact of Buildings - What Matters?, Environ. Sci. Technol. 49 (2015) 9832–9841. doi:10.1021/acs.est.5b01735.

[52] I.M. Sobol', Sensitivity analysis for nonlinear mathematical models, Math. Model. Comput. Exp. 1 (1993) 407–414.

[53] P. Iooss, B., Lemaître, A review on Global Sensitivity Analysis methods, in: Uncertain. Manag. Simulation-Optimization Complex Syst., Springer, New York, 2015: pp. 101–122.

[54] Le Gouvernement du Grand-Duché de Luxembourg, Second National Energy Efficiency Action Plan, (2011).

[55] A. Mastrucci, A. Marvuglia, E. Popovici, U. Leopold, E. Benetto, Geospatial characterization of building material stocks for the life cycle assessment of end-of-life scenarios at the urban scale, Resour. Conserv. Recycl. (2016). doi:10.1016/j.resconrec.2016.07.003.

[56] Règlement grand-ducal du 26 mai 2014. Performance énergétique des bâtiments d'habitation et fonctionnels, Meml. - J. Off. Du Gd. Du Luxemb. A – N° 99. 99 (2014). http://www.legilux.public.lu/leg/a/archives/2014/0099/a099.pdf.

[57] R. Nouvel, A. Mastrucci, U. Leopold, O. Baume, V. Coors, U. Eicker, Combining GIS-based statistical and engineering urban heat consumption models: Towards a new framework for multi-scale policy support, Energy Build. 107 (2015). doi:10.1016/j.enbuild.2015.08.021.

[58] CEN, EN ISO 13790:2008 Energy performance of buildings - Calculation of energy use for space heating and cooling, (2008).

[59] R. Bon, Allometry in the Topologic St ructure of Architectural Spatial Systems, Ekistics. 36 (1973) 270–276.

[60] P. Steadman, S. Evans, M. Batty, Wall area, volume and plan depth in the building stock, Build. Res. Inf. 37 (2009) 455–467. doi:10.1080/09613210903152531.

[61] L. Ministère de l'Economie et du commerce extérieur, Leitfaden zur Verwendung von U-Wert-Typologien im Rahmen der Energie-passerstellung für bestehende Gebäude, v. 1.1., (2007).

[62] A. Merzkirch, T. Hoos, S. Maas, F. Scholzen, D. Waldmann, Wie genau sind unsere Energiepässe?: Vergleich zwischen berechneter und gemessener Endenergie in 230 Wohngebäuden in Luxemburg Wie genau sind unsere Energiepässe?, Bauphysik. 36 (2014) 40–43. doi:10.1002/bapi.201410007.

[63] R Core Team, R: A language and environment for statistical computing., (2016). https://www.r-project.org/.

[64] O. Guerra Santin, L. Itard, H. Visscher, The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock, Energy Build. 41 (2009) 1223–1232. doi:10.1016/j.enbuild.2009.07.002.

[65] B. Peuportier, S. Thiers, A. Guiavarch, Eco-design of buildings using thermal simulation and life cycle assessment, J. Clean. Prod. 39 (2013) 73–78. doi:10.1016/j.jclepro.2012.08.041.

[66] R. Nouvel, C. Schulte, U. Eicker, D. Pietruschka, V. Coors, Citygml-Based 3D City Model for Energy Diagnostics and Urban Energy Policy Support, Proc. BS2013 13th Conf. Int. Build. Perform. Simul. Assoc. (2013) 218–225.

[67] STATEC, Statistic Portal Grand Duchy of Luxembourg, last consulted August 2016, (n.d.). http://www.statistiques.public.lu/ (accessed June 20, 2008).

[68] W.J. Conover, Practical Nonparametric Statistics, John Wiley & Sons, New York, 1971.

[69] J.Z. Kolter, J. Ferreira, A Large-scale Study on Predicting and Contextualizing Building Energy Usage, Proc. Conf. Artif. Intell. (AAAI), Spec. Track Comput. Sustain. AI, 2011. (2011) 8.

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

[71] M. Lacirignola, P. Blanc, R. Girard, P. Pérez-López, I. Blanc, LCA of emerging technologies: addressing high uncertainty on inputs' variability when performing global sensitivity analysis, Sci. Total Environ. 578 (2017) 268–280. doi:10.1016/j.scitotenv.2016.10.066.

[72] M. Stein, Large Sample Properties of Simulations Using Latin Hypercube Sampling. Technometrics. 29 (1987) 143–151.

[73] M. Zaharia, R.S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M.J. Franklin, A. Ghodsi, J. Gonzalez, S. Shenker, and I. Stoica. Apache Spark: a unified engine for big data processing. Commun. ACM 59 (2016) 56-65. doi:10.1145/2934664