| | |
|---|---|
| ***Title*** | A methodology for calibration of building energy models at district scale using clustering and surrogate techniques |
| ***Authors(s)*** | Tardioli, Giovanni, Narayan, Aditya, Kerrigan, Ruth, Oates, Mike, O'Donnell, James, Finn, Donal |
| ***Publication date*** | 2020-11-01 |
| ***Publication information*** | Tardioli, Giovanni, Aditya Narayan, Ruth Kerrigan, Mike Oates, James O'Donnell, and Donal Finn. "A Methodology for Calibration of Building Energy Models at District Scale Using Clustering and Surrogate Techniques." Elsevier, November 1, 2020. https://doi.org/10.1016/j.enbuild.2020.110309. |
| ***Publisher*** | Elsevier |
| ***Item record/more information*** | http://hdl.handle.net/10197/12268 |
| ***Publisher's statement*** | This is the author's version of a work that was accepted for publication in Energy and Buildings. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Energy and Buildings (226, (2020)) https://doi.org/10.1016/j.enbuild.2020.110309 |
| ***Publisher's version (DOI)*** | 10.1016/j.enbuild.2020.110309 |

# Non-deterministic calibration of district scale building energy models using clustering and surrogate techniques ☆

Giovanni Tardioli[a,b,*], Aditya Narayan[b], Ruth Kerrigan[a], Michael Oates[a], James O'Donnell[b], Donal P. Finn[b]

[a] *Integrated Environmental Solutions Ltd., Castleforbes House, Castleforbes Road, D01 A8N0, Ireland*
[b] *School of Mechanical and Materials Engineering, University College Dublin (UCD), Belfield, Dublin 4, Ireland*

---

**Abstract**

Prediction of building energy use, when performed at urban scale, is influenced by the choice of modelling approach, as well as the quality of available data. In the case of data scarcity, one of the main limitations of current urban scale building energy simulation models, is the use of deterministic approaches to specify modelling inputs for entire classes of buildings. This modelling approach is characterised by three major shortcomings: first, data uncertainties are not comprehensively considered; second, a rigorous method of identification of groups of buildings to be populated by similar modelling parameters is not specified; and third, strategies to calibrate the developed energy models are missing. Considering these challenges, the current paper presents a non-deterministic calibration method for groups of buildings. The methodology utilises four techniques: (i) the use of clustering to identify building groups and associated representative buildings within the urban context; (ii) the application of an automatic building energy modelling approach to simulate buildings within these groups; (iii) the application of data-driven models, used as emulators of the dynamic simulation engine, in a computationally efficient manner; and (iv) the exploitation of a non-deterministic Bayesian calibration framework to identify sets of representative parameters for the building clusters. Datasets, based on 2646 buildings from the city of Geneva, Switzerland, are used as a case study. Validation with measured data, shows that energy consumption and energy intensity predictions when considered at an aggregated scale, are within +/- 5%. On an individual building basis, the validation error is within +/- 20 % for approximately 70% of the buildings.

*Keywords:* `Data-driven models`, `district scale energy models`, `non-deterministic modelling`, `emulators`, `Bayesian calibration`, `clustering`

## Nomenclature

| | |
|---|---|
| $\delta()$ | Model discrepancies |
| $\epsilon_m()$ | Measurement error |
| $\epsilon_r()$ | Residuals |
| $\eta()$ | Simulation function |
| $\lambda$ | Length parameter |
| $\mu$ | Mean value |
| $\omega$ | Precision parameter |
| $\Phi()$ | Hyperparameters |
| $\sigma$ | Variance |
| $\Sigma()$ | Covariance |
| $\theta()$ | Modelling parameters |
| $\zeta()$ | Observed physical system |
| $ACH$ | Air Change per Hour [-] |
| $ANN$ | Artificial Neural Network |
| $BEM$ | Building Energy Model |
| $C$ | Cluster |
| $c()$ | Kernel function |
| $DDM$ | Data-Driven Model |
| $DHW$ | Domestic Hot Water $[m^3/s]$ |
| $f()$ | Emulator function |
| $GLR$ | Generalised linear regression |
| $GP$ | Gaussian Process |
| $HVAC$ | Heating Ventilation and Air Conditioning |
| $I$ | Identity matrix |
| $K$ | Modelling parameter |
| $L()$ | Likelihood function |
| $LHS$ | Latin Hypercube Sampling |
| $MAPE$ | Mean Absolute Percentage Error |
| $MCMC$ | Markov Chain Monte Carlo |
| $MLR$ | Multiple linear regression |
| $N$ | Number of predictions |
| $N()$ | Normal distribution |
| $N_b$ | Number of buildings |
| $n_k$ | Number of variations |
| $N_s$ | Number of simulations |
| $OEFEB$ | One Emulator For Each Building |
| $OEFEC$ | One Emulator For Each Cluster |
| $OEFRB$ | One Emulator For Each Representative Building |
| $p$ | Predicted value |
| $P()$ | Posterior distribution |
| $p()$ | Prior distribution |
| $PCA$ | Principal Component Analysis |
| $R^2$ | Coefficient of Determination |
| $RMSE$ | Root Mean Squared Error |
| $RNF$ | Random Forest |
| $SS_{RES}$ | Residual Sum of Squares |
| $SS_{TOT}$ | Total Sum of Squares |
| $SVM$ | Support Vector Machine |
| $x$ | Independent variable |
| $y$ | Metered value |
| $y()$ | Observed variable |

# 1. Introduction

Increasing urbanisation in the coming years is expected to place additional demands on city environments, including infrastructure, air quality, natural resources and living standards [1].

☆ * Corresponding author *Email address:* giovanni.tardioli@iesve.com (Giovanni Tardioli )

Worldwide, city authorities are addressing urbanisation burdens by activating transition programmes toward efficient organisation of mobility, resources and the built environment [2]. Central to these is energy, where several cities have set ambitious emissions and energy consumption abatement targets as horizons to shape future built environments [3]. In this context, buildings play an important role being one of the major contributors to energy consumption in cities [4, 5]. Short and long-term energy strategies for the building sector increasingly require detailed information on building energy use at a high spatial and temporal resolution [6–9] to facilitate scenario evaluation, policy assessment and future planning [10–19].

Data analysis ([20–23]) and large scale simulation [6, 24–28] are receiving attention as ways to assess building energy needs at an urban scale. In this context, building data may be affected by high heterogeneity in quality, accuracy and quantity. Furthermore, data scarcity resulting from privacy protection schemes, economic interests or technological barriers [29, 30] compounds these challenges. Data problems affect building energy simulation as well. On the one hand, further integration of data techniques in simulation work flows [31] is expected to increase in the coming years. On the other hand, missing data, stochasticity, uncertainties related to important modelling parameters, together with building heterogeneity, complex interconnected physical phenomena, computational overheads, and requirements of dedicated software architectures, challenge traditional building energy modelling approaches [32]. Large validation errors have been recorded for urban scale modelling especially at disaggregated level [33, 34].

Building energy models to reliably predict energy consumption require a process of calibration [35–37]. To date, however, the majority of building calibration studies focus at the individual building level [38]. Model calibration is a time-consuming task that involves expert knowledge and extensive metered datasets [39]. Automatic methods for parameter screening and adjustment have been proposed, but such approaches remain confined to single building applications [40–43]. There are few studies which propose solutions for urban scale model calibration.

For urban scale studies, when single building data are missing, input values derived from national standards, representative buildings, normative values, literature or expert knowledge [44–46] are often used for buildings falling into the same typology and period of construction [47].

One limitation associated with these classification methods is that they do not analyse multiple building features in parallel, nor do they deploy quality indices for the assessment of the best partitioning strategy among buildings. Accurate building grouping is still an open research question. Advanced data-mining techniques, such as clustering, can identify groups of buildings, screen multiple features in parallel and, therefore, enable a more comprehensive analysis of building characteristics [48–50].

Once building groups are identified, deterministic simulation methods based on the use of single values of modelling parameters, cannot correctly represent energy distributions and modelling uncertainties for entire groups of buildings [51]. Non-deterministic methods, have been used to better represent uncertainties within model calibration [52, 53] suggesting distributions of modelling parameters rather than fixed figures. Among these methods, Bayesian calibration plays an important role, being able to take into account expert knowledge, uncertainties of input parameters, model inaccuracies and measurement errors [54]. Bayesian calibration has been extensively considered at an individual building level [43, 55]. When used at an urban scale, Bayesian methods require adaptation to be applied for entire groups of buildings [56].

Calibration methods for groups of buildings should consider methods to reduce computational overheads and test different simulation strategies. Emulators, acting as a proxy of simulation engines, reduce computational time. Most of the literature to date considers linear regression or Gaussian Processes [57] as emulators of the simulation engine in Bayesian calibration studies. Linear models may not be the most accurate predictive approaches and Gaussian Processes are computationally expensive for large datasets and for high dataset dimensionality. The use of alternative machine learning techniques, such as neural networks (NN) and support vector machines (SVM), as part of urban scale non-deterministic calibration frameworks is missing from research to date. These methods offer an attractive trade-off between prediction accuracy, computational time and scaling capabilities. From a modelling point of view, building emulators could be developed using different surrogate strategies and at different levels of aggregation and simplification, such as: one emulator for each building or one emulator for the entire cluster of buildings. Studies which compare different surrogate strategies are also missing.

The overall aim of the current paper is to investigate a combination of clustering techniques and Bayesian calibration methods for the identification of the most representative ranges of modelling parameters for entire building clusters. In particular, the specific objectives of this research are as follows:

- to develop and test a Bayesian calibration framework to populate large clusters of building energy models with appropriate parameter distributions;
- to investigate the inclusion of machine learning techniques in an urban scale Bayesian calibration framework which act as a proxy of dynamic building simulation engines;
- to study and validate different emulator modelling structures (surrogate techniques) applied to clusters of buildings within the Bayesian framework.

The structure of the paper is as follows. In Section 2, the state of the art of large scale modelling techniques and Bayesian calibration is presented with focus on the role of data driven models (DDM) as emulators. In Section 3, an overview of the research methodology is presented with tools and methods used in the current work. Section 4 describes an urban district case study, where results of the proposed approaches are presented and discussed. Section 5 outlines conclusions and future work.

## 2. Literature Review

Most of urban scale modelling studies are based on deterministic approaches to evaluate energy use of a large number of buildings [6, 25, 33]. Deterministic methods, considering only a fixed number of values of modelling parameters, fail to accurately represent energy use distributions and modelling uncertainties associated with urban scale analysis [51]. In such applications, metered building consumption is mostly used to validate the models by results comparison, but not to inform and support the calibration procedure. In these cases, the simulation model is influenced by either expert knowledge assumptions, input data or modelling uncertainties. Urban scale building energy models require a process of calibration before being used as reliable decision support tool. Manual calibration in data-rich scenarios shows high levels of accuracy [58], however, it is a time demanding procedure [39, 59]. As such, it may be not feasible for a large number of buildings without investing substantial resources. In most of the cases, manual calibration is a result-driven or evidence-based procedure, where the energy modeller tunes selected parameters to achieve defined results [35]. Alternatives to manual calibration are non-deterministic approaches such as Bayesian methods or optimisation techniques, iterative approaches, evidence based techniques and signature methods [39]. Among these different modelling solutions, Bayesian methods offer interesting possibilities, such as: to take into account various sources of uncertainties, to include expert knowledge in the calibration process and to use information of metered data to lead the calibration process.

Bayesian calibration has been used to calibrate simulation models in a variety of different research fields, and to intrinsically take into account uncertainties related to the modelling approach ([60–64]). It relies on the Bayes theorem (Eq. 1). The theorem states that the most likely distribution of a variable can be inferred considering its prior distribution and its likelihood evaluated with comparison to metered data:

$$P(\theta,\phi|y) \simeq p(\theta,\phi) \cdot L(y|\theta,\phi) \tag{1}$$

where $P(\theta,\phi|y)$ is the posterior distribution of the modelling parameters, $p(\theta,\phi)$ is the prior distribution of the parameters and $L(y|\theta,\phi)$ is a likelihood function.

Table 1 summarises a literature review of applications of Bayesian methods in the building energy modelling sector. The table focuses only on large scale studies. It provides information regarding scale of application, building typology, type of emulator, building grouping technique and target parameters. Generally, Bayesian methods are mostly applied at single building level: for additional information the interested reader is redirected to [55, 57, 65–70].

With reference to Table 1, the majority of urban scale studies focus on the residential sector ([51, 56]. In these studies, Bayesian methods and group modelling approaches are applied to entire classes of buildings. Building grouping is performed considering simple filtering (classification) techniques based on categorical variables such as period of construction, building typology and building final use. Buildings in the same classes may have very different properties (i.e., energy consumption, energy intensities, geometric properties, etc.). More advanced grouping methods such as clustering allow to analyse many building features in parallel and to define building groups based on similarities.

Clustering methods have received particular attention because they enable for a finer segmentation of the building stock compared to traditional classification procedures. Clustering does not impose particular weights on some variables producing an un-biased grouping outcome. Once groups are identified, group modelling solutions, such as the identification of modelling parameters by Bayesian methods, can be applied. Nevertheless, to the extent of the authors knowledge, there are no studies which investigate the use of Bayesian approaches on building groups identified by clustering techniques.

Table 1: Bayesian calibration applied at large scale: literature review

| Ref. | Scale | Typology | Emulators | Resolution | Classification | Target variables [units] |
|---|---|---|---|---|---|---|
| [51] | City | Residential | Linear | Annual-monthly | none | Heating set point $[°C]$, cooling set point $[°C]$, infiltration $[ACH]$, occupant density $[people/m^2]$, equipment power density $[W/m^2]$, domestic hot water $[m^3/(s*m^2)]$ |
| [56] | District | Residential | GP | Yearly | none | Window U-value $[W/m2K]$, wall U-value $[W/m2K]$, infiltration $[ACH]$, windows-wall ratio (north)$[-]$, windows-wall ratio (south)$-$, temperature set point $[°C]$, appliances heat load $W/m^2$ |
| [71] | Building stock | Commercial | Linear | Annual | none | Appliances heat load $[W/m^2]$, lighting heat load $[W/m^2]$, gross floor area $[m^2]$, windows wall ratio $[m^2]$, aspect ratio, height $[m]$, infiltrations $[ACH]$, cooling COP[-], cooling set point $[°C]$, heating set point $[°C]$, wall U value $[W/(m^2K)]$, roof U value $[W/(m^2K)]$ |
| [72] | 35 flats | Residential | GP | Daily | Cat. var. | Heating set point $[°C]$, fraction of space heating $[-]$, windows wall ratio $[-]$, heating COP %, air leakage at 50 Pa $[ls \cdot m^2]$, wall U value $[W/(m^2K)]$, windows U value $[W/(m^2K)]$ |
| [73] | Schools | Educational | MLR | Annual | Cat. var. | Heating set point $[°C]$, fraction of space heating $[-]$, windows wall ratio $[-]$, heating COP %, air leakage at 50 Pa $[ls \cdot m^2]$, wall U value $[W/(m^2K)]$, windows U value $[W/(m^2K)]$ |
| [74] | Housing stock | Residentials | – | Annual | none | Heating set point $[°C]$, fraction of space heating $[-]$, windows wall ratio $[-]$, heating COP %, air leakage at 50 Pa $[ls \cdot m^2]$, wall U value $[W/(m^2K)]$, windows U value $[W/(m^2K)]$ |
| [75] | 35 super-markets | Commercial | – | Annual | none | lighting intensity $[W/m^2]$, appliance intensity $[W/m^2]$, occupants intensity $[W/m^2]$, air change ratio $[times/hour]$, air conditioning COP, length of cabinet $[m]$, COP refrigerators$[-]$, defrost heater capacity $[W/m]$, circulation fan capacity $[W/m]$, anti-sweat heater capacity $[W/m]$, cabinet lighting capacity $[W/m]$, air infiltration ratio, air velocity of air curtain $[v/s]$ |
| [75] | 2182 apartments | Residential | – | Monthly | none | Cooling COP$[-]$,Infiltration $[ACH]$,Internal Gain $[W/m^2]$, Cooling set point temperature $[°C]$, life style factor $[-]$ |

For large scale urban energy studies scalability and computational time is paramount as well as accuracy. Table 1 shows that the majority of the research studies related to Bayesian methods employ Gaussian processes (GP) or linear methods acting as proxy of the simulation engine [55–57, 67, 69, 76]. Gaussian processes, offer the advantage to fit the training data and to define confidence intervals for predictions. Time requirements for training of a GP emulator is often high [65]. For large datasets or for high dimensionality of the dataset, the use of GP emulators is computationally expensive and prohibitive. Data pre-processing is often considered to reduce the dimensionality of the dataset applying variable reduction methods such as PCA (Principal Component Analysis) [43], or limiting the investigation to subset of informative metered data [70].

The use of alternative 'lightweight' approaches (linear models), characterised by low training and prediction times, has been recently presented [68]. Although linear models have simple structures and for this reason short training and prediction times, they may not be the most accurate models.

Speed of calculation, flexibility in the exploration of high data dimensionality, data-volume scalability and prediction accuracy are fundamentals pre-requisite for urban scale modelling approaches. The use of non-linear machine learning techniques (such as neural networks, support vector machines or random forests) as emulators of dynamic simulation engines have not been comprehensively studied as part of urban scale non deterministic calibration methods.

In this research, the way emulators (i.e., GLM, SVM, NN, RNF) are deployed is indicated as *surrogate strategy*. This can be performed in various ways: training one emulator for each building [51] or one emulator for a group of buildings [56]. There are no studies which compare the use of different surrogate strategies on the same case study. This gap has already been identified by the research community in recent studies [56]. In addition, recently, it has been highlighted ([68]) that there is limited research which aims to assess the accuracy of different surrogate strategies and emulators.

The overall contribution of the present work is to develop a large scale Bayesian calibration framework to identify modelling parameters for clusters of similar buildings in an urban context testing different surrogate strategies and modelling techniques. Therefore, the objectives of the presented work is threefold. First, clustering algorithms are used to identify groups sharing similarities in terms of several building features to achieve a finer segmentation of the urban building stock. Second, different machine-learning techniques are employed as emulators of urban scale dynamic simulations and their performance are evaluated and compared. Third, different surrogate strategies are adopted in a large scale Bayesian calibration framework and compared.

## 3. Methodology

The methodology involves the use of a work-flow as shown schematically in figure 1, and in more detail in figure 2. The input of the method is a set of building stock data available from the city case study. The output are distributions of modelling parameters characterising entire groups of building energy models. A description of the main steps of the methodology are provided in the following sections.

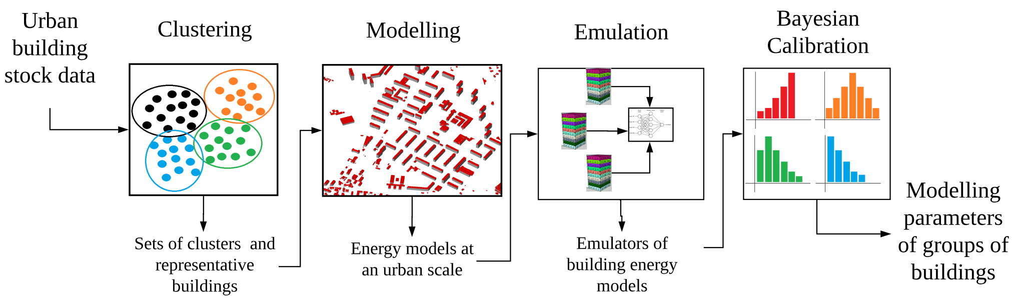



Figure 1: Overview of the research methodology from city data to modelling parameters

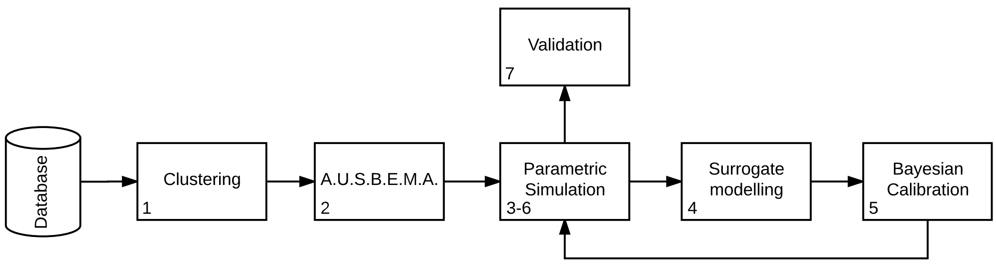



Figure 2: Overview of the research methodology in terms of work-flow

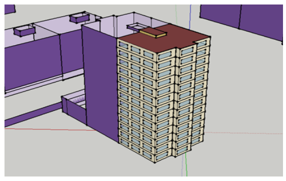
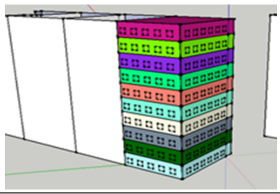
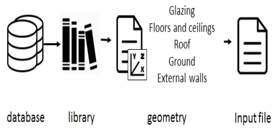
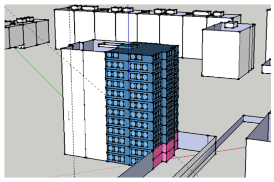
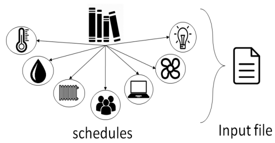
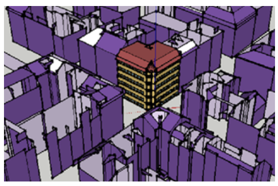
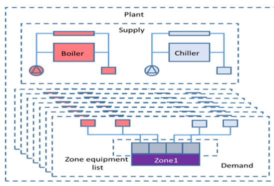
### 3.1. Clustering

A methodology, exploiting the combination of building classification, clustering analysis and predictive modelling techniques for the identification of representative buildings and building clusters, was developed by the authors as part of previous research work ([50]). A short summary of the methodology follows: (i) starting from multiple urban-scale datasets, classification techniques based on categorical variables (i.e., period of construction, final use and building typology) are used; (ii) four clustering algorithms (kmeans, hierarchical agglomerative, divisive and partitioning around medoids) are applied to identify building clusters and representative buildings. (iii) Normalisation techniques to process the input data are applied and multiple validation indices are evaluated to recommend the best segmentation of the building dataset in terms of numbers of clusters; (iv) post application of predictive modelling, using a random forest approach, facilitates the incorporation of a larger portion of building stock in the established clusters with high classification accuracy. The outcome of this steps are a number of building clusters and associated representative buildings (medoids).

### 3.2. Modelling

The clustering outcome (representative buildings and building within the same cluster) is forwarded to an automated urban scale building energy modelling approach, here after called *AUSBEMA*. The automated approach uses building data of the city case study to generate input files for dynamic simulation using EnergyPlus. A Python script maps and replicates the EnergyPlus simulation template (idf file). To achieve this, the model uses data including the geo-location of a building, 3D model of the city and information from different sources. Multilevel spatial analysis is achieved by defining spatial boundaries in terms of buildings included in the analysis, thereby facilitating analysis at group, district or city level. AUSBEMA is able to account for the urban environment in terms of building surroundings and building adjacencies and shading scenarios of the target building. Table 2 summarises the main properties of the automated modelling approach AUSBEMA. The outcomes of this step are a set of energy models of the buildings in the case study.

Table 2: Automated Urban Scale Building Energy Modelling Approach (AUSBEMA) main features

<table>
<tr><td>Geometry Parsed from CityGML using building geo-location such as building ID. Data are filtered in roofs, external walls and ground surfaces. Roofs and ground surfaces data are converted in BEM inputs; external walls data are further processed to define storeys, floors and ceilings. Glazing surfaces and shapes are generated defining initial geometry values and populating surface data using a window to wall ratio.</td><td></td></tr>
<tr><td>Thermal zones and boundary conditions Implemented with a multi-zone modelling approach. A thermal zone for each storey is defined using geometry information. Surfaces are linked to each thermal zone as well as the definition of internal gains, schedules, Heating Ventilation and Air Conditioning (HVAC) and set points.</td><td></td></tr>
<tr><td>Constructions and materials A database of materials and constructions properties are generated using information derived by different urban datasets. Building constructions and materials are selected automatically based on the period of construction of the building used as classification variable.</td><td>
</td></tr>
<tr><td>Building adjacencies Adjacencies are detected processing CityGML data and building relative positioning. This allows the definition of different boundary conditions. The script is able to handle complex boundary scenarios such as partial adiabatic adjacencies, internal walls, adiabatic ceiling and floors definition and to modify wind, sun and ground exposure accordingly.</td><td></td></tr>
<tr><td>Schedules and internal gains Representative schedules are created from literature data and from normative values. Schedules of equipment, lights, infiltration, occupancy, heating and cooling uses, water uses and set points are imported automatically into the input file using the building typology as classification variable. Internal gains are modelled based on library data and the definition of internal gains intensities.</td><td>
</td></tr>
<tr><td>Shading objects The definition of shading objects for a target building is based on a selective approach. The proximity of an object to one building is evaluated using geo-referenced information of each building and evaluating a distance measure between one target and the surrounding buildings. The model selects shadings evaluating a distance matrix of the target building and the surrounding constructions.</td><td></td></tr>
<tr><td>Domestic hot water and HVAC systems The structure of the DHW supply system and the connections with the demand side are defined automatically using tag-names for the definition of objects enabling the automatic interconnection of the parts of the systems. HVAC systems are defined starting from templates and enriching the input file with information of the demand side for each thermal zone of the building.</td><td>
</td></tr>
</table>

### 3.3. Parametric simulation for design of experiments

The experimental design, in term of number of different modelling scenarios, is performed using parametric simulation. The energy models of the buildings, classified as per results of clustering and generated by AUSBEMA, are used into a parametric simulation tool.

Initially, parametric variables are selected and defined. This step may involve sensitivity analysis or the identification of important variables considering previous studies from the literature. Model variations are generated using a parametric tool. This requires the definition of the initial distribution of modelling parameters. Upper and lower limits may be defined using normative inputs, expert knowledge, realistic values for the typology of buildings under analysis, or directly using measurements. The total number of simulations is given by equation 2:

$$Ns = N_b \cdot \prod_{k=1}^{K} n_k \tag{2}$$

where Ns is the number of simulations, K is the total number of variables, n is the number of values the variable k can have, $N_b$ is the number of buildings considered in the analysis. Equation 2 can be rewritten as Equation 3, if the number of parameter variations is the same for all the parameters.

$$Ns = N_b \cdot n^K \tag{3}$$

The number of simulations can grow very quickly, if a large number of calibration parameters and relative variations are selected or if the number of buildings considered in the analysis is high. To reduce the number of simulations a Latin Hypercube Sampling (LHS) [77] approach is used ([78]). The parametric simulation should allow to comprehensively explore the range of values of the modelling parameters, since the predictive models are able to provide the most accurate predictions in those ranges. The output of this passage will be a large set of simulation results which will be used for the training of the emulators.

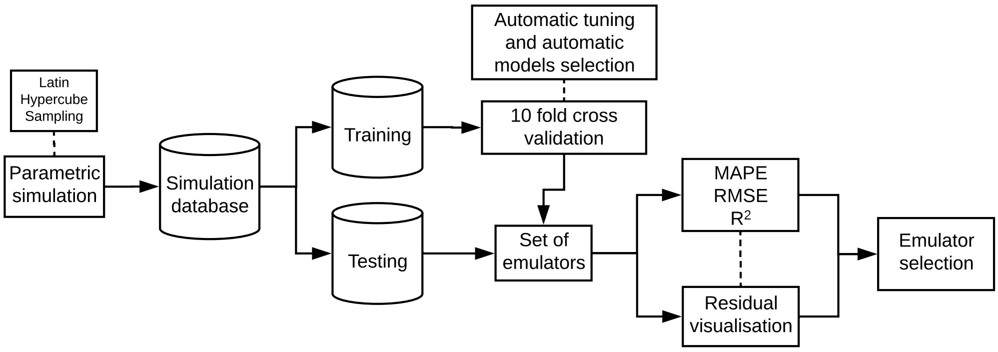



Figure 3: Emulator selection process. LHS: latin hypercube sampling.

### 3.4. Surrogate strategies

Emulators, also known as meta-models [39], are predictive approaches used to fit simulation outputs [79]. Their training differs from standard regression methods, as they are trained with simulated data rather than actual measurements. The main objective of the emulators is to map the dynamic behaviour of a simulation engine. They are employed when simulation is computationally expensive or when a large number of simulations are required (e.g., automatic calibration or optimisation studies).

With reference to Figure 3, a multi-step approach is performed to select emulators: firstly, simulation results, are gathered in a database; secondly, emulators are trained and selected; finally, the accuracy of the emulators is evaluated and best performing models are chosen. After the parametric simulation, the available results are split in two components of different sizes for training and testing: in this application, a larger portion is used for training (70%) and the remainder used for testing (30%). A 10-fold cross validation technique (Figure 3) is used to prevent over-fitting, efficiently exploring the information of the dataset and auto-tuning algorithmic parameters [80].

In this study, four widely used predictive models are employed as emulators of the dynamic simulation software: generalised linear models (GLM) [81–84], artificial neural networks (NN) [85–88], support vector machines (SVM) [89–92] and random forests (RNF) [87, 93–95]. With reference to Figure 3, the best models from each algorithm family are selected, the accuracy of the predictive models is manually evaluated on a totally unseen dataset (testing dataset: 30% of the overall dataset). Model accuracy is evaluated using three quality

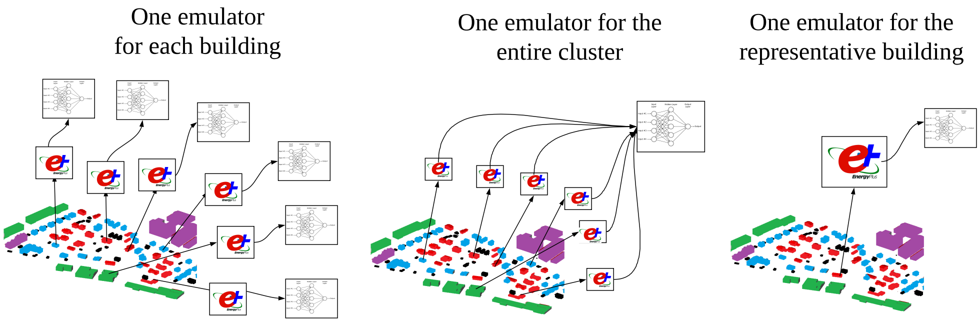



Figure 4: Surrogate strategies

measures: mean absolute percentage error (MAPE)[96–99], root mean square error (RMSE) [86, 97, 100, 101] and coefficient of determination ($R^2$) [102]. An analytical description of the quality measures follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - p_i}{y_i}\right| \cdot 100 \tag{4}$$

where MAPE is the mean absolute error, N is the total number of predictions, y is the actual value of measurement i and p is its predicted value;

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - p_i)^2} \tag{5}$$

where RMSE is the root mean squared error, N is the total number of predictions, y is the actual value of measurement i and p is its predicted value;

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{N}(y_i - p_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{6}$$

where $SS_{res}$ is the residual sum of squares and $SS_{tot}$ is the total sum of squares N is the total number of predictions, y is the actual value of measurement i and p is its predicted value and $\overline{y}$ is the mean value. Visualisation techniques are used to further judge the quality of the emulators, such as actual versus predicted value plots.

Three surrogate strategies are adopted. They are characterised by different deployment time and data requirements. Figure 4 shows the three strategies.

#### 3.4.1. One emulator for each building (OEFEB)

For this approach, one energy model for each building in the cluster and one emulator for each building is required. In terms of computational time and resources for development this is the most demanding approach. An emulator for each building is trained using the simulation outcomes. In this case, each emulator is able to perform predictions only on one building therefore each building emulator is required in the Bayesian framework.

#### 3.4.2. One emulator for each cluster (OEFEC)

In this approach, one energy model for each building in the cluster is required to conduct the calibration process. Only one emulator is developed for the entire cluster of energy models. As in the OEFRB approach, one emulator is trained in order to recognise differences across buildings in the same cluster using whole-building features (i.e., volumes, heating surfaces). Compared to the OEFRB method, the predictive model is trained with the results of all the building energy models and not only the representative building.

#### 3.4.3. One emulator for representative building and a scaling solution (OEFRB)

Representative buildings identified by clustering are used as a medium to identify modelling information for all the buildings in the same cluster. Only one emulator derived from the representative building of each cluster and its energy model are considered in the analysis. Whole-building variables (i.e., volume and heating surface) of all the buildings in the cluster are adopted to generate a scaling solution (regression function) and to predict the consumption distribution of the buildings in the clusters starting from the consumption of the representative building. Therefore, a regression method is trained to recognise differences among buildings in terms of consumption and to predict consumption values based on whole building features.

### 3.5. The Bayesian framework

For each surrogate strategy, selected predictive algorithms are used in the Bayesian framework as emulators of the dynamic simulation engine. The output of Bayesian methods are the most likely posterior distributions of modelling parameters. Measurements of a physical process can be analytically described as [103]:

$$y(x_i) = \zeta(x_i) + \epsilon_m(x_i) \qquad i = 1, ..., n \tag{7}$$

where $y(x_i)$ indicates the value of the observation, $x_i$ indicates the set of independent variables considered in the analysis, $\zeta(x_i)$ the actual physical system and $\epsilon_m(x_i)$ the measurement error. Equation 7 can be re-written as:

$$y(x_i) = \eta(x_i, \theta) + \delta(x_i) + \epsilon_m(x_i) \qquad i = 1, ..., n \tag{8}$$

where the physical system has been substituted by the outcome of a simulation method (dynamic or not) $\eta(x_i, \theta)$ for variables $x_i$ and simulation parameters $\theta$. $\delta(x_i)$ indicates the model inaccuracies or discrepancies between the actual physical process and the model. Model discrepancies are assumed to be dependent only to model inputs. The random observation error is considered to follow a Gaussian distribution $\epsilon_m(x_i) \simeq N(0, \sigma_m^2)$. If the simulation process is too time consuming and a large number of simulations are required for the calibration analysis, emulators can be included in the process to substitute the dynamic simulation method. Considering the use of emulators in the modelling approach, equation 8 is re-written as:

$$y(x_i) = f(x_i, \theta) + \delta(x_i) + \epsilon_m(x_i) \qquad i = 1, ..., n \tag{9}$$

Traditionally, model emulator and model discrepancies are modelled with a Gaussian process defined as follows:

$$f(x_i, \theta) = N(\mu, \Sigma), \qquad \delta(x_i) = N(\mu_\delta, \Sigma_\delta) \tag{10}$$

The Gaussian process emulator offers the advantage of a perfect fit on the training dataset. Nevertheless, GP emulators training is computationally demanding. When the size of the simulation is large, due to the number of buildings, the number of simulations to perform or the number of modelling calibration parameters, other 'lightweight' emulators may be introduced as suggested in [65]. Regression algorithms do not offer a perfect fit on the training dataset. Hence, a residual error is considered as follows:

$$y(x_i) = f(x_i,\theta) + \epsilon_r(x_i) + \delta(x_i) + \epsilon_m(x_i) \qquad i = 1,...,n \tag{11}$$

where $f(x_i,\theta)$ is a regression function of variables $x_i$ and modelling parameters $\theta$. $\epsilon_r(x_i)$ is the regression error in terms of residuals introduced to take into account the limit of the regression approach. The regression error is considered to follow a Gaussian distribution $\epsilon_r(x_i) \simeq N(0,\sigma_r^2)$, where $\sigma_r^2$ indicates the variance of the residuals of the predictive approach: such assumption can be verified analysing the residual distribution. The model discrepancy $\delta(x_i)$ is modelled as an unbiased Gaussian distribution $\mu_\delta = 0$ and covariance matrix $\Sigma_\delta = C(x,x)$ evaluated using a squared exponential kernel:

$$c_{ij} = exp\left(-\frac{1}{\lambda}\sum_s \omega_s(x_{is} - x_{js})^2\right) \tag{12}$$

The matrix and its elements are defined once the hyper-parameters ($\phi = (\lambda,\omega)$), $\lambda$ (length parameter) and $\omega$ (precision) are set for all the set of independent variables.

All the observations are assumed to form a joint Gaussian distribution. The model can be written as:

$$y \simeq N(\mu,\Sigma), \qquad \mu = f(x,\theta), \qquad \Sigma = \Sigma_\delta + (\sigma_m^2 + \sigma_l^2)I \tag{13}$$

where $\mu$ is the mean value of the regression function and $\Sigma$ is the total covariance matrix.

The Bayes's rule states that the posterior distribution for design parameters $\theta$ and hyper-parameters $\phi$, conditioned on the data $y$ is defined as:

$$P(\theta,\phi|y) \simeq p(\theta,\phi) \cdot L(y|\theta,\phi) \tag{14}$$

The prior probabilities $p$ are generated defining distribution of the modelling parameters and hyper-parameters. The likelihood function is defined as:

$$L(y|\theta,\phi) = |\Sigma|^{1/2}\, exp\left(-\frac{1}{2}\left[(y-\mu)^T\Sigma^{-1}(y-\mu)\right]\right) \tag{15}$$

### 3.5.1. Markov Chain Monte Carlo sampling

Non-linear phenomena govern the dynamic simulations and the subordinate emulation processes. It is impractical to derive an analytical expression of the posterior distributions of the modelling parameters and hyper-parameters especially in case of high data dimensionality [104]. Estimates of the distributions of the parameters can be generated as a chain of samples using a Markov Chain Monte Carlo (MCMC) approach whose convergent distribution is the posterior distributions of the parameters. In this study, the Metropolis-Hasting algorithm acts as a random walk through the parameters space [105]. At each iteration, the algorithm evaluates the probability ('probability of move') of a sample to be accepted as new stage in the random walk. This is done evaluating the ratio of a distribution, proportional to the target probability distribution, between the previous step and the suggested new point [106]. Considering the law of large numbers, the ratio is compared to a random number generated from a uniform distribution ($r_n \in (0,1)$) and accepted only if it is greater than the number [107]. The algorithm requires an initial phase of convergence to the actual posterior distributions, therefore a number of samples are initially discarded as they are not representative of the distribution ('burn in samples').

## 3.6. Parametric simulation of calibrated parameters

A set of posterior distributions describe the most likely values of the modelling parameters. As such, posterior distributions are sampled to generate parametric simulations for validation purposes. A scripting approach is required to sample from the posterior distributions and to generate a finite set of simulations. This is practically achieved by developing a code capable to modify the parametric simulation file changing its definition and calling the set of parameters from external text files.

## 3.7. Validation

Results of parametric simulations, generated from the identified posterior distributions of the modelling parameters, are used to estimate the accuracy of the results by comparison with metered data. Validation is performed for each of the surrogate strategy, comparing simulated and actual data in terms of total building consumption and energy index distribution for the urban case study. An analysis of the performance achieved at single building level is conducted to understand which modelling technique provide the best results.

# 4. Case Study and Results

## 4.1. Case study

A district (Meyrin) of the city of Geneva (CH) is considered in this research work to test and validate the presented methodology. Figure 5 shows the location of the district within the city boundaries. The district, as defined in the cadastral database of the city, is comprised of approximately 3370 buildings. An open data repository, SITG (Le système d'information du territoire à Genève)[108]), provides building information through a web portal. Several datasets store information of the building stock. Figure 6 describes the typologies of buildings within the district. The majority of the buildings are residential (mostly multifamily buildings) and other typologies which include a variety of small units (e.g., garages, verandas, spaces $< 20m^2$, etc). A smaller percentage is covered by commercial, services and mixed use buildings. In terms of period of construction, the majority of the buildings were constructed during the years 1960 to 1980. In terms of building energy use, the datasets provide building consumption as the sum of heating and domestic hot water requirements. Figure 7 shows a comparison between the energy consumption distribution normalised by the number of buildings for the district and from a larger data sample (of about 9000 buildings) from the overall city. The distribution of the total consumption is characterised by a right skewed distribution. When compared to the distribution of the larger sample of buildings from the city the two distributions follow similar trends in terms of total consumption and energy index suggesting that the energy analysis of the district is quite representative of the metered consumption of the city.

A 3D CityGML model of the city (LOD2) was used in parallel with the automated urban scale modelling approach AUSBEMA presented in Section 3.2 to translate the geometry of the buildings as inputs to the energy models. A bird-eye view of the district and the associated 3D model representation is provided in Figure 8. The modelling approach creates a multi-zone model for each building under analysis. Input parameters for the initial characterisation of the input files are gathered from the available datasets of the city. Values from the literature and from expert knowledge have been utilised in the case of missing data. Weather data time-

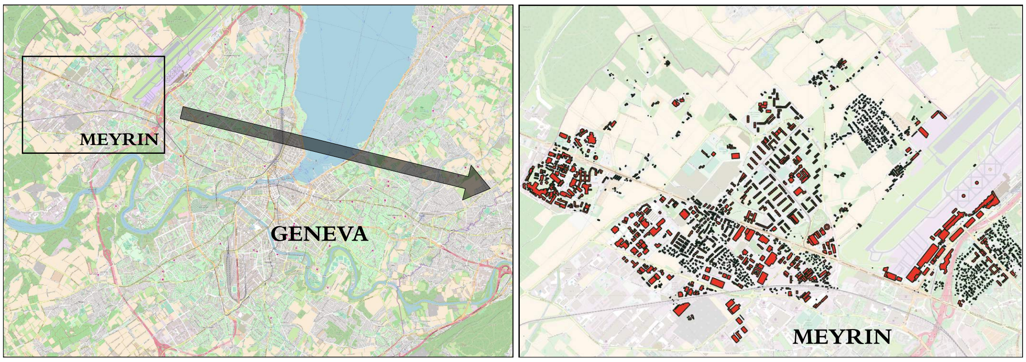



Figure 5: The district case study GIS overview. Map data: ©OpenStreetMap contributors. Building data: SITG.

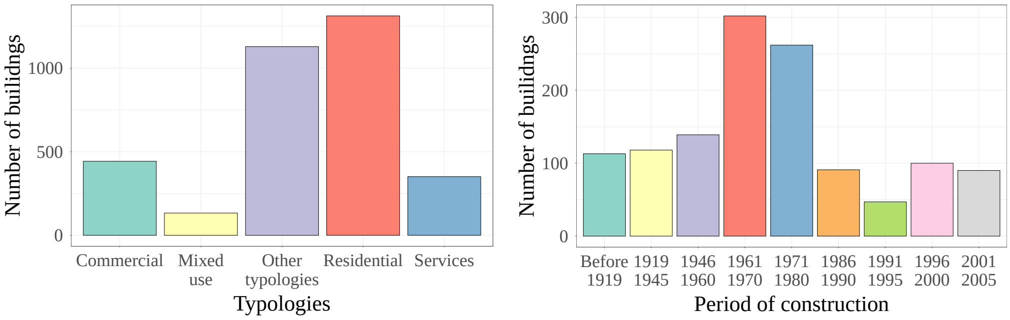



Figure 6: Left: building typologies in Meyrin. Right: buildings period of construction in Meyrin.

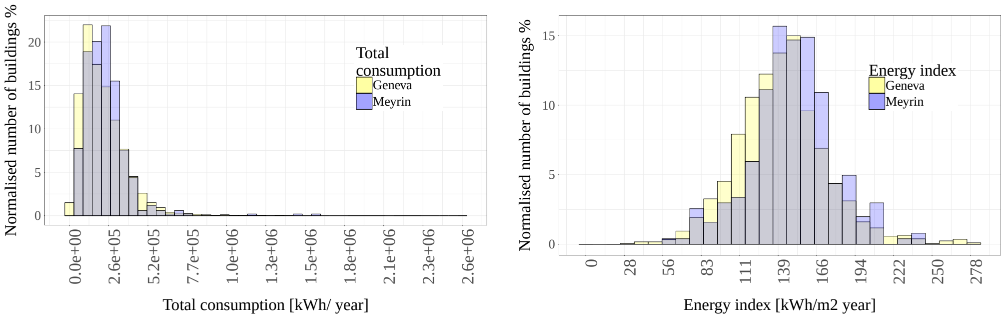



Figure 7: Total consumption (left) and energy index distributions (right) for the district case study and a larger sample of buildings (9000) within the city case study

series are collected from a meteorological station which is approximately 3km away from the Meyrin (Geneva Jonction [109]).

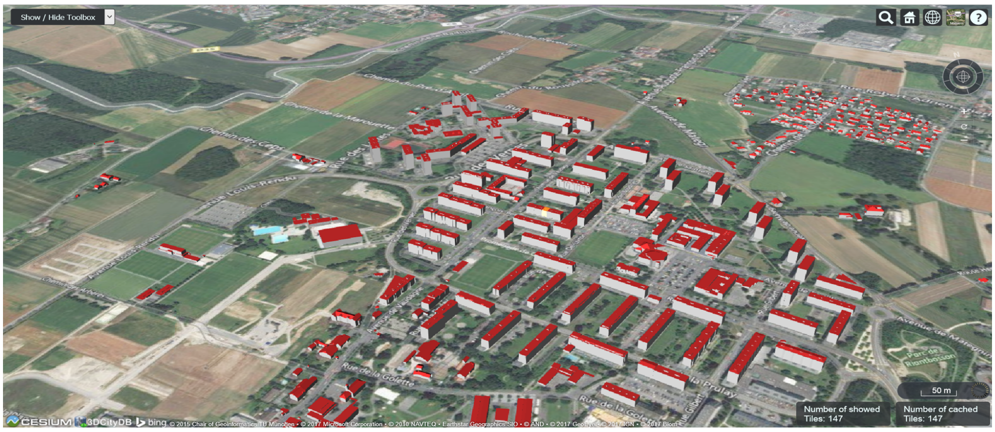



Figure 8: 3D building model (CityGML) of the Meyrin district case study. Map data: ©Cesium.

The residential multi-family buildings (326 buildings), the largest class and typology in the district, are considered in the following steps of the analysis. Modelling values have been gathered from normative guidelines relative to building energy modelling standards in Switzerland. In particular, normative SIA 380/4, SIA 382/1, SIA 385/2 ([110]) are considered. A direct engagement with experts of the Department of Planning, Housing and Energy (DALE), allowed the modelling parameters to be refined and integrated. In addition, the following research works [111, 112] were consulted for the initial characterisation of the models. A summary of the modelling parameters used to characterise the energy models are shown in Table 3.

The district case study is served by a district heating network facility which connects a large portion of the buildings under analysis and provides the necessary heating demand. The remaining buildings in the district case study generally employ centralised boiler systems for heating and DHW production.

## 4.2. Clustering results

The full set of results related to clustering has been presented elsewhere by the authors [50]. In this section, a summary of the results is presented. The use of the methodology, described

Table 3: Initial modelling parameters organised by period of construction

| Name | Units | Period of construction | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Before 1919 | 1919 1945 | 1946 1960 | 1961 1970 | 1971 1980 | 1981 1990 | 1991 2000 | 2001 2005 |
| U-values ext-wall | $W/Km^2$ | 1.40 | 1.8 | 1.35 | 1.20 | 0.63 | 0.48 | 0.29 | 0.21 |
| U-values roof | $W/Km^2$ | 1.30 | 1.20 | 1.18 | 0.85 | 0.54 | 0.46 | 0.29 | 0.21 |
| U-values ext-ground | $W/Km^2$ | 1.48 | 1.30 | 1.27 | 0.85 | 0.55 | 0.46 | 0.29 | 0.21 |
| U-values int part. | $W/Km^2$ | 1.86 | 1.44 | 1.40 | 1.40 | 0.65 | 0.56 | 0.29 | 0.21 |
| U-values windows | $W/Km^2$ | 3.50 | 3.50 | 3.30 | 3.30 | 2.50 | 2.50 | 1.40 | 1.40 |
| Infiltration | $ACH$ | 1.5 | 1.2 | 1 | 0.95 | 0.85 | 0.65 | 0.55 | 0.45 |
| Occupancy | $m^2/person$ | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| Lighting intensity | $W/m^2$ | 3.5 | 3.5 | 3.5 | 3.5 | 3.5 | 3.5 | 3.5 | 3.5 |
| Equipment intensity | $W/m^2$ | 15 | 15 | 15 | 15 | 15 | 15 | 15 | 15 |
| DHW efficiency | % | 0.75 | 0.75 | 0.8 | 0.8 | 0.85 | 0.85 | 0.9 | 0.9 |
| Water use | $m^3/s$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ | $1 \cdot 10^{-5}$ |
| Heating set point | °C | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |

in Section 3.1, allows the classification of a large portion of the building stock of the city case study (13500 buildings) into clusters. A total of 67 clusters were identified.

For the smaller number of buildings of the district case study, a total of 14 clusters and representative buildings can classify and represent the population of 326 buildings considered in this analysis as shown in Figure 9.

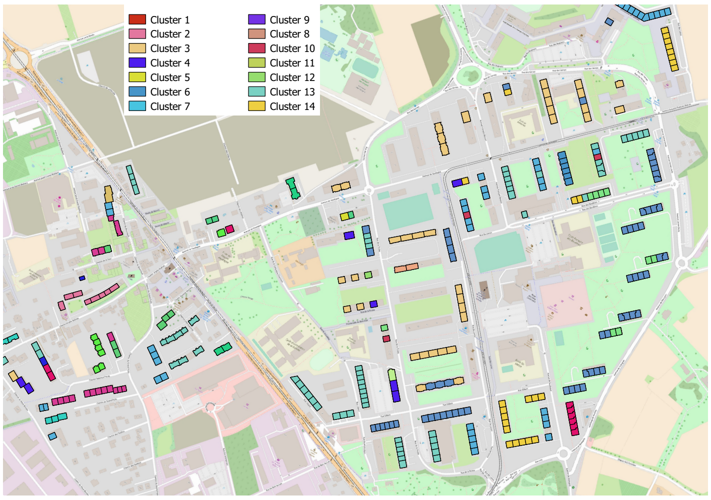



Figure 9: Clustering classification of the buildings within the district case study. Map Data ©OpenSreetMap contributors.

Clustering results and information of the building stock are forwarded to AUSBEMA to generate single building energy models classified in groups following the results of the clustering analysis.

## 4.3. Parametric Simulation

The parametric simulator used in this research work is JE+ [113]. Parameter screening and sensitivity analysis conducted in related literature work are used to define important building parameters affecting building energy consumption. Seven calibration parameters are selected as inputs with high impact on building energy consumption [51, 55, 71, 114–120]. Table 4 provides information of the selected parameters, their range of variation and the sampling technique adopted and the literature references where the parameters have been previously considered. The set up of the experiment follows an initial definition of the variation of the selected modelling parameters. Uniform distributions were selected to define the initial ranges of variation. The spectrum of variation is selected to ensure that the training of the emulators can examine the dynamic behaviour of the simulation engine at different energy consumption levels.

Table 4: Definition of experiment: parameters, distributions, variation ranges, increments, sampling strategy and literature references.

| Name | Units | Distribution | Range | Increment | Sampling | Reference |
|---|---|---|---|---|---|---|
| Lighting intensity | $W/m^2$ | uniform | $\in \mathbb{R}$: (1-10) | 1 | LHS | [71, 114–116] |
| Occupancy density | $m^2/person$ | uniform | $\in \mathbb{R}$: (1-50) | 10 | LHS | [71, 114, 116] |
| Infiltrations | $ACH$ | uniform | $\in \mathbb{R}$: (0.1-1.5) | 0.1 | LHS | [71, 114–118] |
| Heating set point temp. | $C$ | uniform | $\in \mathbb{R}$: (17-23) | 1 | LHS | [71, 114, 115, 118–120] |
| DHW use | $m^3/s$ | uniform | $\in \mathbb{R}$: ($0.5\cdot10^{-6}$-$1.5\cdot10^{-5}$) | 0.1 | LHS | [51, 55] |
| DHW sys. efficiency | % | uniform | $\in \mathbb{R}$: (0.7-0.95) | 0.05 | LHS | [51, 55, 115] |

## 4.4. Surrogate strategies: analysis and selection of the emulators

The results of the parametric simulations are collected in a database which is used for the training of the predictive algorithms (emulators: nnet, svm, glm, rnf) using different surrogate strategies as defined in Section 3.4 (OEFRB,OEFEB,OEFEC). The selection of the best emulators for the three surrogate strategies is achieved by evaluating the predictive performance metrics presented in Section 3.4.

To the authors knowledge, the OEFRB (one emulator for each representative building) has not previously been used in the context of meta-modelling therefore assessment of the predictive and modelling capabilities is required. The regression algorithm, trained only with representative building information and using whole building variables as a scaling approach, is characterised by large accuracy errors. The average MAPE is 26.89, the RMSE is 91150 kWh and the coefficient of determination is 0.93. Considering this, to improve the quality of the predictive approach, iterative increments of the number of buildings information sampled from the district case study and merged with representative buildings information is provided to the learning framework.

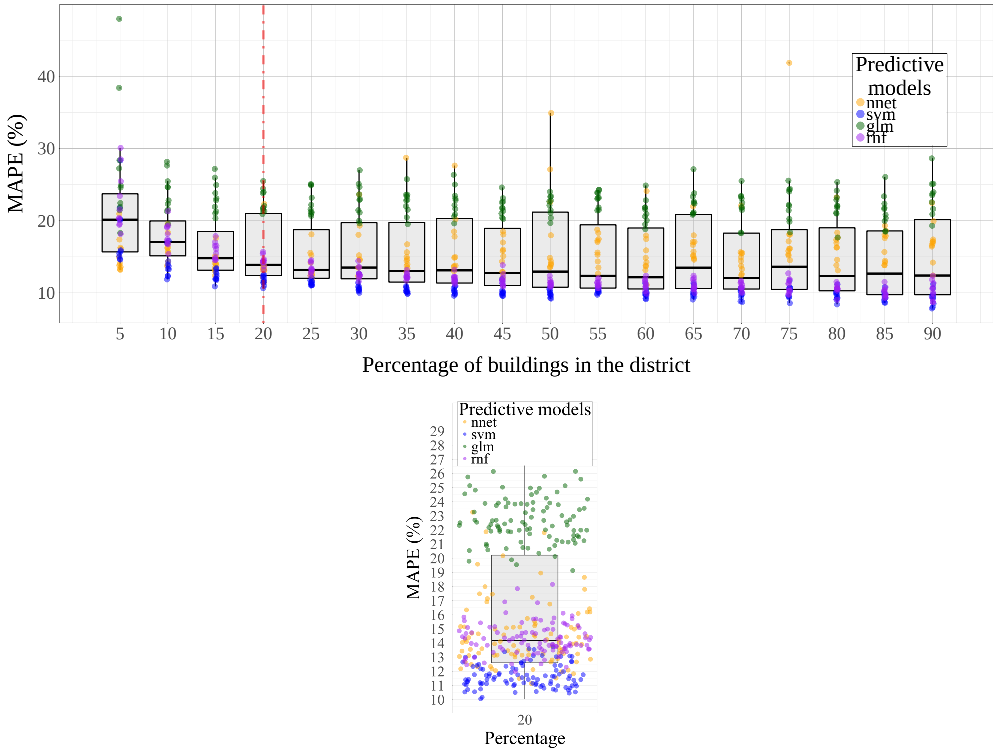



Figure 10: Top: OEFRB performance for increasing number of buildings information as percentage of total number of buildings in the district. Bottom: OEFEB random sampling for a selected number of informative buildings. (Nnet: neural networks; svm: support vector machines; glm: generalised linear models; rnf: random forests)

Figure 10 shows how increments of the number of buildings used as sources of information for the training of the predictive models generate lower MAPE, suggesting that it is possible to achieve more accurate predictions after the inclusion of a certain number of explanatory buildings from the district under analysis. An analysis of the results allows the percentage of total buildings, which provides a good trade-off between prediction accuracy and total number of energy models, to be identified. In this case study, a random selection of 20% of the total number of buildings of the district is able to improve substantially the accuracy of the predictive model. This requires the development of additional building energy models for the selected percentage of buildings.

For each percentage considered, a large number of random samples are selected to assess if some set of buildings are more explanatory than others. Figure 10 shows that the accuracy of the predictive method can have a large range of variation. This is related to two main factors: (i) the type of predictive algorithm and (ii) the subset of building sampled from the district to populate the training dataset.

For example, repeating random sampling of the same percentage (20%) as shown in Figure 10, it is possible to achieve in some cases an error equal to MAPE 10% (SVM). Predictive modelling families achieve different levels of error, with SVM and NNET being the most accurate models. Considering the results shown in Figure 10, for each family of emulators, it is possible to select the most accurate methods. Thus, these models are included in the Bayesian calibration framework.

For all the surrogate strategies, accuracy measures are assessed in the test dataset (out-of-sample testing) for each of the 14 clusters and for each emulator typology. Results are presented in Figure 11. The barcharts show the performance metrics in terms of RMSE MAPE and $R^2$ for each building cluster and for each emulator.

In terms of prediction accuracy and surrogate strategies, the OEFEB approach produces the most accurate emulators followed by the OEFEC and lastly the OEFRB method.

Figure 12 provides detailed information about the accuracy achieved in the three different surrogate strategies in terms of MAPE. Results of the predictive accuracy of each selected emulator are shown in Table 5 organised by cluster number.

For the OEFEB technique, the RMSE ranges between about 2480 kWh and 16800 kWh for the most accurate predictive algorithms. The MAPE ranges between 1.6 and 3.7 and coefficient of determination ranges between 0.998 and 0.983. NNET exhibits the best predictive performance for most of the clusters for this surrogate strategy.

For the OEFEC modelling techniques, the accuracy is lower with the MAPE ranging from 1.6 to 7.2. The RMSE ranges from 3368 kWh to 30701 kWh. NNET performs better and in some instances SVM outperforms the other approaches. RNF performs well in all the clusters analysed however it is less accurate than NNET and SVM. GLM had the worst performance across the testing datasets.

The OEFRB approach achieved the worst results with MAPE and RMSE higher than in the previous cases. The coefficient of determination is still high (94%-95%) suggesting that the results are equally spread in terms of underestimation and overestimation.

Visualisation of predicted versus simulated results facilitates the understanding of the behaviour of each algorithm in terms of over or under estimations at different levels of energy consumption. An example is provided in Figure 13. Linear models were found to have strong under or over estimation for low and high level of consumption compared to the other methods. This skewed behaviour is not easy to assess by only quality indices. NNET, SVM and RNF perform well over the entire spectrum of energy consumption.

Residuals of the emulators are assumed to follow a Gaussian distribution as outlined in Equation 11. This property is tested for each emulator. Generally emulators respect the assumption. This was somewhat expected given the high values of the coefficient of determination in the testing dataset. This is not always valid for linear models which tend to generate skewed distribution compromising the initial modelling assumptions.

In terms of computational requirements, GLM are the emulators with the lowest computational overhead. They require less time for model training and for the prediction phase when included in the MCMC framework. SVM and RNF have similar computational time cost both for training and in the prediction phase. NNET are the most accurate predictive methods across many building clusters and surrogate strategies, however, they require the highest computational time in the prediction phase.

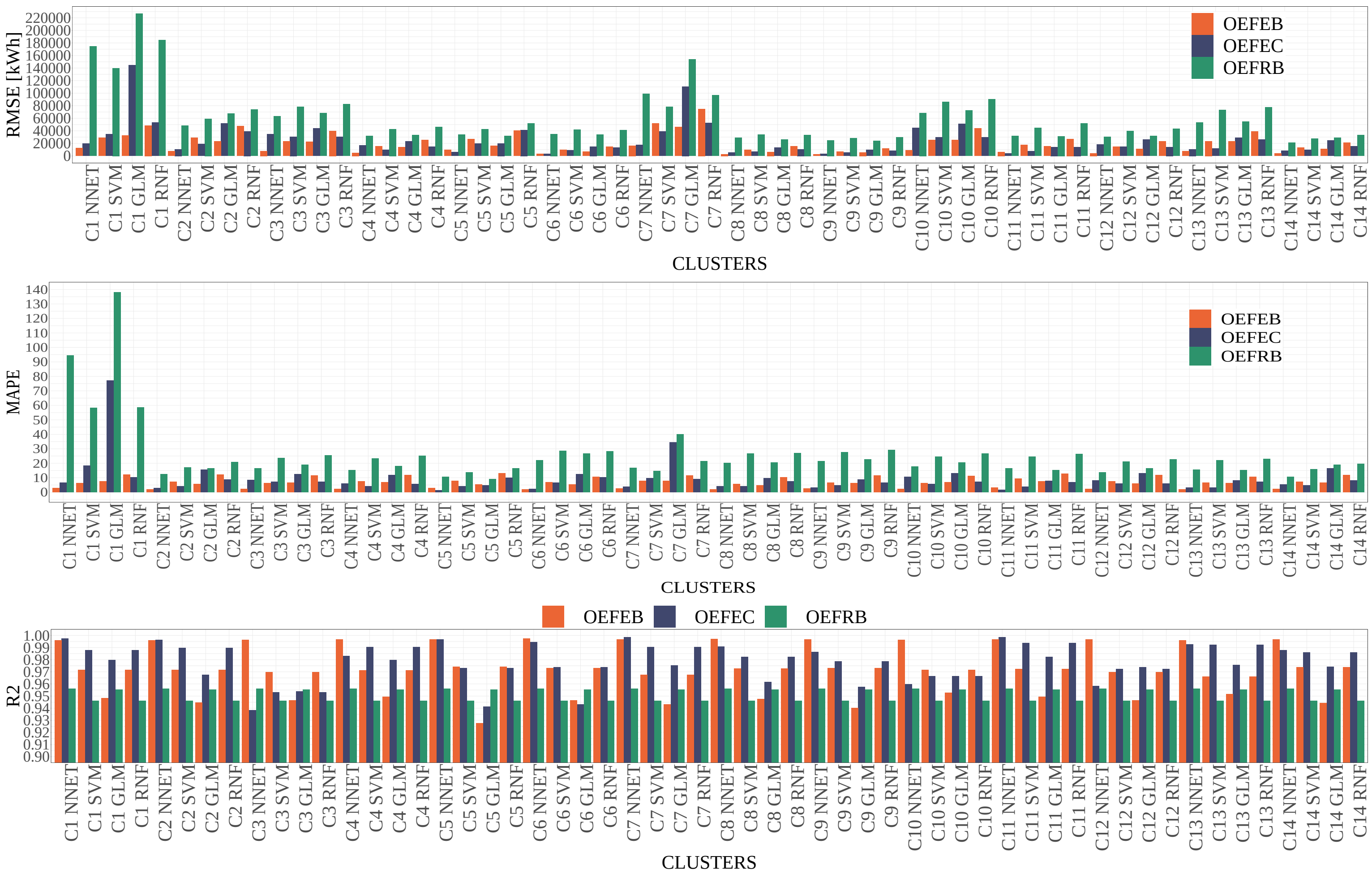



Figure 11: RMSE, MAPE and R2 of various emulators for each of the 14 building clusters. (NNET: neural networks; SVM: support vector machines; GLM: generalised linear models; RNF: random forests)

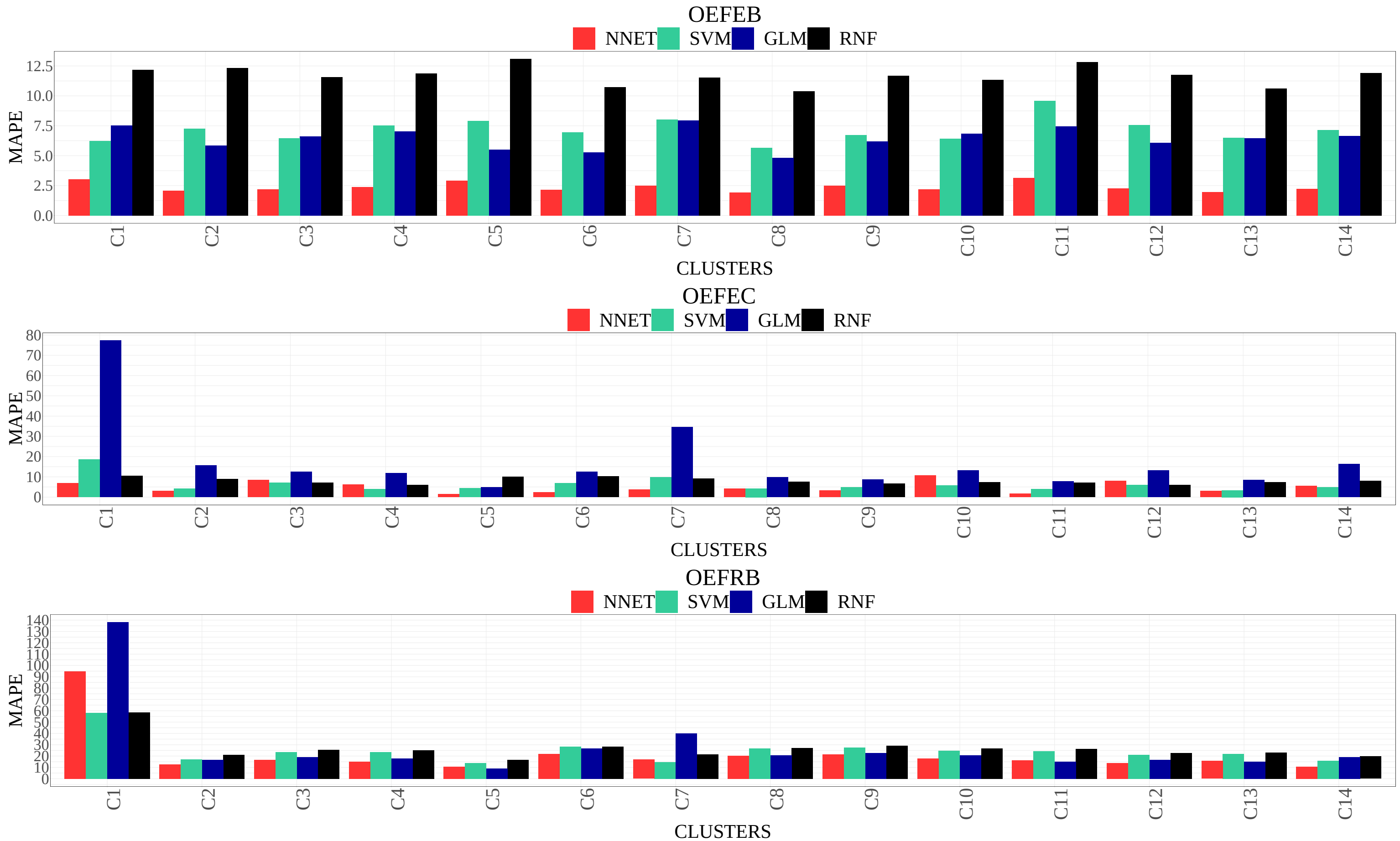



Figure 12: MAPE of the emulators for the three surrogate strategies for each building cluster. NNET: neural networks; SVM: support vector machines; GLM: generalised linear models; RNF: random forests

Table 5: Best emulator performance for each building cluster

| Cluster number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Emulator_OEFEB | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET | NNET |
| MAPE_OEFEB | 3 | 2.1 | 2.2 | 2.4 | 2.9 | 2.2 | 2.5 | 1.9 | 2.5 | 2.2 | 3.2 | 2.3 | 2 | 2.2 |
| RMSE_OEFEB | 12322 | 7720 | 7956 | 5091 | 9712 | 3274 | 16501 | 2879 | 2561 | 8940 | 6430 | 4419 | 7376 | 4245 |
| R2_OEFEB | 0.996 | 0.996 | 0.996 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.997 | 0.996 | 0.997 | 0.997 | 0.996 | 0.997 |
| Emulator_OEFEC | NNET | NNET | SVM | SVM | NNET | NNET | NNET | NNET | NNET | SVM | NNET | SVM | NNET | SVM |
| MAPE_OEFEC | 6.8 | 3 | 7.2 | 4 | 1.5 | 2.4 | 3.8 | 4.2 | 3.2 | 5.8 | 1.6 | 5.9 | 3.1 | 4.9 |
| RMSE_OEFEC | 19907 | 10854 | 30701 | 9570 | 5899 | 3368 | 17509 | 5662 | 3485 | 29704 | 3964 | 14975 | 10789 | 9978 |
| R2_OEFEC | 0.997 | 0.996 | 0.953 | 0.99 | 0.997 | 0.995 | 0.999 | 0.991 | 0.987 | 0.966 | 0.999 | 0.972 | 0.993 | 0.986 |
| Emulator_OEFRB | SVM | NNET | NNET | NNET | GLM | NNET | SVM | NNET | NNET | NNET | GLM | NNET | GLM | NNET |
| MAPE_OEFRB | 58.2 | 12.6 | 16.7 | 15.3 | 9.1 | 22.1 | 14.8 | 20.1 | 21.5 | 17.7 | 15.2 | 13.9 | 15.2 | 10.6 |
| RMSE_OEFRB | 139913 | 48091 | 63333 | 31865 | 31919 | 35024 | 78338 | 28927 | 25049 | 68579 | 31445 | 30496 | 54739 | 21470 |
| R2_OEFRB | 0.946 | 0.956 | 0.956 | 0.956 | 0.956 | 0.956 | 0.946 | 0.956 | 0.956 | 0.956 | 0.956 | 0.956 | 0.956 | 0.956 |

### 4.5. Posterior distributions and parametric simulation

The best emulators are included in the Bayesian calibration framework. The full Bayesian framework presented in Section 3.5 has been used for the OEFEC and OEFRB method, the framework has been modified to work with single emulators for the OEFEB surrogate strategy. Markov Chain Monte Carlo sampling is initiated with 15000 iterations and using a burn in sample of 25%. Once the prior distribution of the modelling parameters and the hyper-parameters are defined, the framework converges to the posterior distribution. An example of the results obtained from this process is reported in Figure 14 for the set of modelling parameters considered and for a cluster of building in the district.

The results are quite informative for energy modelling. Generally the calibration framework is able to generate skewed distribution refining the prior distribution of the modelling parameters.

### 4.6. Model Validation

Results of the parametric simulation were evaluated for each building and for each surrogate strategy in each cluster. Figure 15 shows the comparison of the metered consumption of the buildings and the outcome of the simulations.

Results for the OEFEB method show high accuracy in terms of validation. The calibration framework is capable of generating and characterise models which recreate the energy use distribution in terms of total consumption and energy index of the district. The simulated


Figure 13: Example of Emulators predictive accuracy: simulated versus predicted building total consumption.

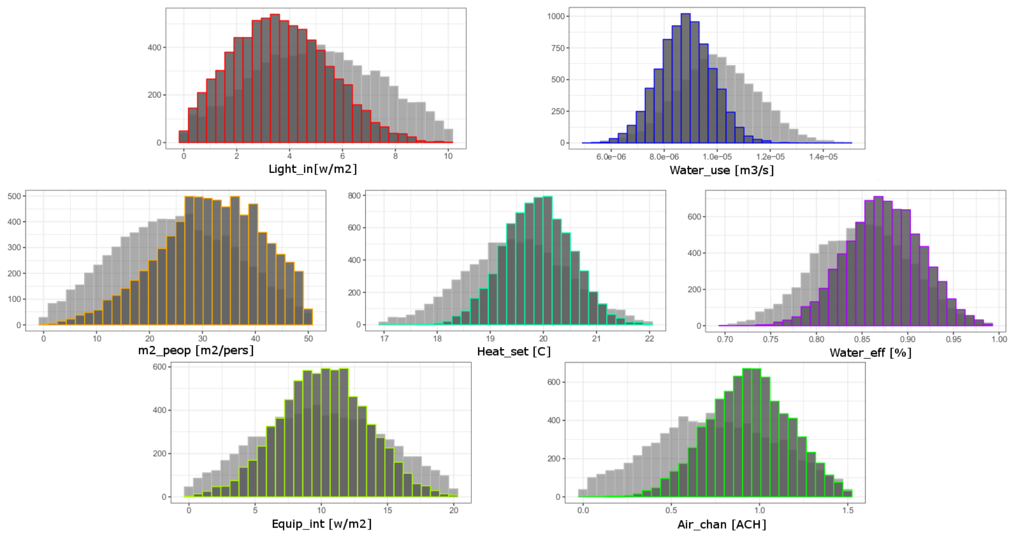



Figure 14: Example of prior and posterior distributions of modelling parameters. Prior: grey; posterior: coloured.

outcomes represent the ranges of building total consumption representing both very large energy consumers as well as smaller energy consumers. In terms of percentage error, the majority of the buildings generate results with error for the majority in the range of ±20% distributed over a nearly Gaussian shaped distribution. There are a few outliers which generate larger discrepancies for which investigation or additional building data is required to reduce errors. The aggregated sum of the consumption for the set of buildings under analysis is 73716 MWh. The sum of the simulation for each building provides an estimate of 74318 MWh. This means that at an aggregated level the model has an error equal to 2.3% (326 buildings).

The OEFEC method produces good results overall with distribution in terms of total building consumption and energy index which accurately match the metered values. In terms of single building results the validation error seems to be well centred on 0 and the error follows a quasi-Gaussian distribution. The model performs worse than the OEFEB approach with errors at single building level spread over a larger range of variation. At an aggregated level, the simulation outcome is equal to 67715 MWh which corresponds to a 8.2% of error.

The OEFRB method presents good results overall. The modelling approach is able to recreate the distribution of the total consumption of the buildings in the district case study as well as the distribution in terms of energy index. This approach presents the larger errors in terms of single building distribution and the results show that compared to the other two methods the building scale validation accuracy is the poorest. Nevertheless, errors are equally distributed following a Gaussian distribution. For this reason, at an aggregated level the model provides quite low simulation errors: the total simulation provides a value of 72246 MWh which is equal to approximately 2% of error.

Figure 16 shows the validation process for a larger sample of buildings not included in the previous analysis. Buildings outside the boundaries of the district and part of the city case study are selected to test the reliability of the calibration process. Specifically, a total number of 2646 multi-family buildings are first modelled using the automated modelling approach and thus characterised considering the identified parameters. Results of clustering techniques in terms of cluster membership are available for each building considered in this analysis. Therefore, it is possible to use the distributions of the parameters identified during the calibration process for buildings falling in the same clusters. Parametric simulation is repeated for each building. The final results are visualised for comparison in terms of total

Table 6: Percentages of buildings within error ranges for the three proposed approaches.

| Method | Range | -100 | -90 | -80 | -70 | -60 | -50 | -40 | -30 | -20 | -10 | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | -90 | -80 | -70 | -60 | -50 | -40 | -30 | -20 | -10 | 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| OEFEB | % | 0.0 | 0.0 | 0.1 | 0.2 | 0.6 | 2.7 | 5.8 | 8.9 | 13.0 | 19.1 | 13.6 | 12.0 | 9.8 | 5.9 | 4.7 | 3.2 | 1.5 | 1.0 | 0.7 | 0.3 |
| OEFEC | % | 0.0 | 0.0 | 0.1 | 0.3 | 0.7 | 2.9 | 6.3 | 10.3 | 14.5 | 17.5 | 14.8 | 11.2 | 8.6 | 4.9 | 3.3 | 2.0 | 1.4 | 0.6 | 0.4 | 0.2 |
| OEFRB | % | 0.0 | 0.0 | 0.2 | 0.2 | 0.7 | 2.7 | 6.1 | 9.4 | 13.2 | 17.8 | 14.4 | 11.4 | 8.7 | 5.3 | 4.3 | 2.3 | 1.6 | 0.6 | 0.6 | 0.5 |

consumption, energy index and error at single building level.

In terms of total distribution, all the three proposed techniques are able to generate simulated distributions which correctly match the metered data. The energy index distribution is well mapped by the three proposed approaches. All of them are able to correctly match the average value of the distribution and the overall shape in terms of minimum and maximum value. The metered energy index exhibits a distinctive peak distribution corresponding to its mean. It is not unusual to detect the exact same value of consumption in the metered consumption database for different buildings. This generates differences with the simulated values which are not rounded or approximated in fixed ranges.

In terms of single building errors, the three techniques produce good results with distribution centred on zero. The OEFEC method produces the closest Gaussian-shaped distribution compared to the other two techniques, which follow nearly-triangular distributions.

Table 6 summarises the percentage of buildings with a validation error within a specific range. The percentage of buildings with errors in the range of +/- 30% for the OEFEB method is 76.4, for the OEFEC approach 76.9, for the OEFRB is 74.9. Overall, across the range of values, all the methods present a small overestimation of the building energy consumption.

At an aggregated level, all the methods perform well with error 0.75% for the OEFEB method, 2.5% for the OEFEC and 1.34% for the OEFRB approach on a total consumption of 580.03 GWh.

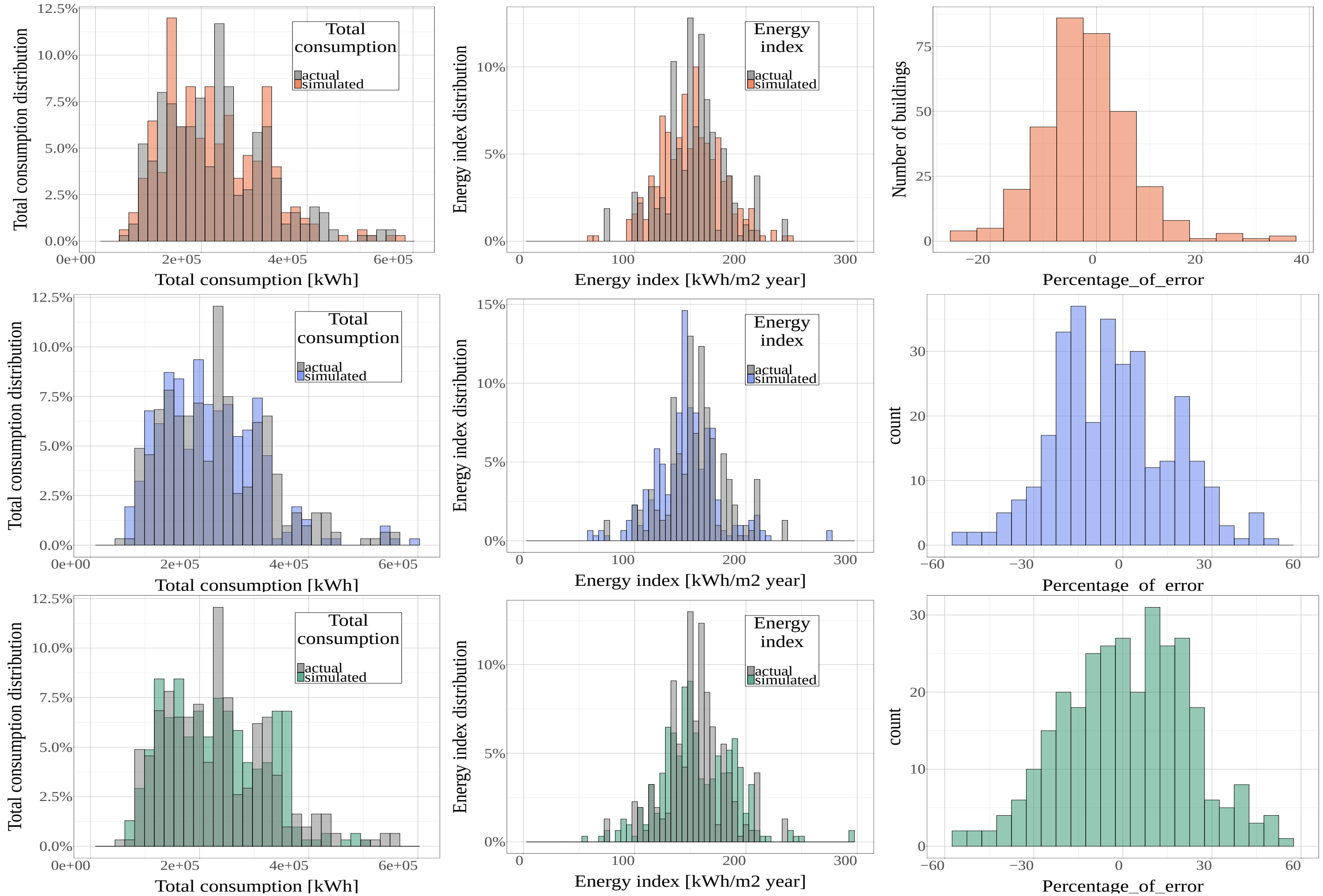



Figure 15: Validation of simulation results and comparison of the three surrogate strategies within the district case study (326 buildings): top OEFEB, center OEFEC, bottom OEFRB. First column total consumption (heating-DHW uses); second column energy index (heating-DHW uses); third column single building error

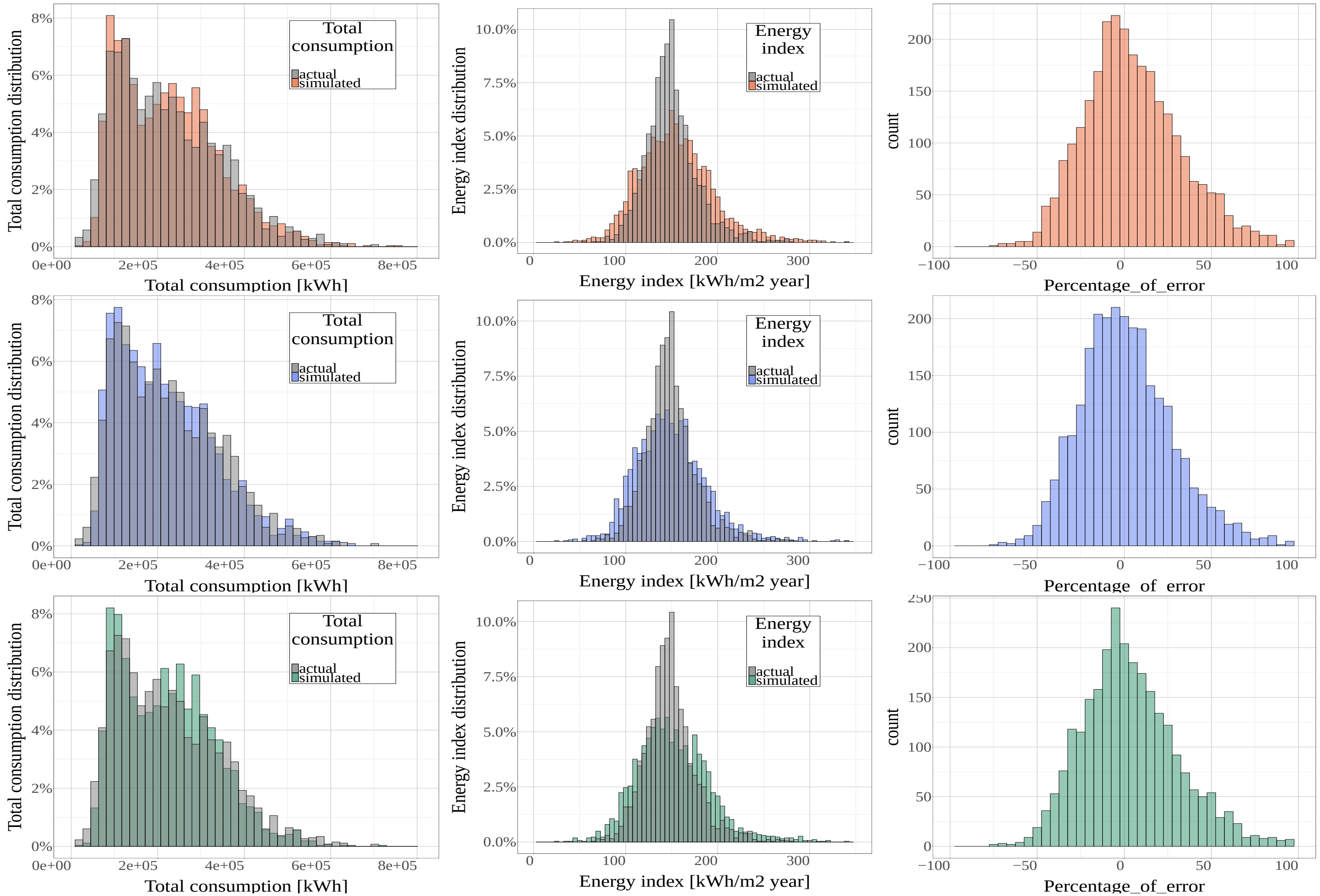



Figure 16: Validation of simulation results and comparison of the three surrogate strategies in a large sample of buildings within the city case study (2646 buildings): top OEFEB, centre OEFEC, bottom OEFRB. First column total consumption (heating-DHW uses); second column energy index (heating-DHW uses); third column single building error

# 5. Conclusions

An approach that utilises Bayesian theory for model calibration has been evaluated on building energy models generated by an automated urban energy modelling approach and grouped by the means of clustering algorithms. The research shows the possibility of conducting building energy model calibration at a district/city scale using such non-deterministic methods. In addition, the study underlines the possibility of exploiting limited available data from the building stock to infer distributions of building modelling parameters for building clusters and, therefore, to characterise groups of building energy models. In terms of urban scale modelling, the possibility of identifying modelling parameters from a limited number of buildings is important. This allows to partially overcome problems of data scarcity at an urban level and provides a scaling solution for large scale application.

The inclusion of machine learning techniques in the Bayesian calibration framework showed interesting results. Alternative predictive algorithms to the traditional Gaussian process were considered as emulators of a dynamic simulation engine. Different families of predictive models showed the possibility to reach high levels of prediction accuracy with errors related mostly to the family of emulators and surrogate strategy adopted. The use of accurate emulators is key to obtain reliable results. It is fundamental however to generate manageable emulators with reduced training and prediction times. This issue is important, especially, when the number of buildings to be included in the analysis is large or the dimensionality of the datasets is high. The analysis of the prediction capability of the emulators showed better results for non-linear approaches such as neural networks and support vector machines, when compared to more traditional algorithms such as linear models. Linear models showed poor quality and skewed behaviour. A recommendation arising from the results of the current research is that the coefficient of determination alone should not be considered as totally reliable quantifier for quality assessment of the emulators. Results show that it is not capable of providing information for specific under or over estimations. To overcome these issues, the adoption of multiple accuracy measures and visualisation of the results is recommended. The quality and shape of the residuals of the models should be always assessed (and ideally be Gaussian shaped) to validate the hypothesis related to the use of the Bayesian framework.

The adoption of various surrogate modelling techniques showed the flexibility of the Bayesian framework for different modelling approaches. Results are affected by the modelling solution adopted. Nevertheless, all the results obtained in this study are well aligned or surpass in accuracy the results showed in the current literature [34]. In terms of quality of the results, it is possible to achieve high level of accuracy utilising the single building emulator technique (OEFEB) as shown in this paper and in [51]. However, the limited amount of single building data may be not sufficient to create informative modelling parameters and therefore to extend the results to a larger group of buildings. For this reason, the OEFEB method, lacks of flexibility. Moreover at single building level, other calibration approaches may be preferred. Overall, the results showed better performance for the methods when applied at an aggregated scale, with small errors for the majority of the modelling solutions. The quality of the results at aggregated level and single building level do not seem to be related. While for small single building errors the aggregated error is expected to be small, this is not the case for large single building errors; when aggregated errors tend to eliminate each other and for this reason to provide accurate results overall if the model does not present particular bias. Therefore, error assessment at single building level is always important for urban scale studies.

Future work is required to test similar approaches for urban building energy model calibration at a higher temporal resolution. Emulators can be trained at a finer temporal granularity with high accuracy, as shown in [32], however, a series of additional research issues need to be considered. Additional data problems, reliability of the urban scale energy modelling approach, temporal dependency of calibration parameters, computational requirements for core calibration are only a few of the many modelling burdens to overcome before accurate automated calibration can be successfully performed at high time resolution and at an urban scale. In this context, the authors believe that integration of data-driven techniques in traditional building energy modelling workflows can be key to address these challenges.

## 6. Acknowledgement

The authors would like to acknowledge the SITG data provided by the City of Geneva and the European Commission for providing financial support during the research under the FP7-PEOPLE-2013 Marie Curie Initial Training Network Ci-NERGY project with Grant Agreement Number 606851.

# References


[1] S. Fox, T. Goodfellow, Cities and development, Routledge, 2016.

[2] Smart City Wien Framework Strategy, Vienna City Administration, 2017.
URL

[3] https://data.cdp.net: Cities emissions reduction targets, .

[4] B. Güneralp, Y. Zhou, D. Urge-Vorsatz, M. Gupta, S. Yu, P. L. Patel, M. Fragkias, X. Li, K. C. Seto, Global scenarios of urban density and its impacts on building energy use through 2050, Proceedings of the National Academy of Sciences,doi:10.1073/pnas.1606035114.
URL

[5] F. Creutzig, G. Baiocchi, R. Bierkandt, P.-P. Pichler, K. C. Seto, Global typology of urban energy use and potentials for an urbanization mitigation wedge, Proceedings of the National Academy of Sciences 112 (20) (2015) 6283–6288.

[6] J. A. Fonseca, A. Schlueter, Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts, Applied Energy 142 (2015) 247–265. doi:http://dx.doi.org/10.1016/j.apenergy.2014.12.068.
URL

[7] J. A. Fonseca, T.-A. Nguyen, A. Schlueter, F. Marechal, City energy analyst (cea): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts, Energy and Buildings 113 (2016) 202 – 226. doi:http://dx.doi.org/10.1016/j.enbuild.2015.11.055.
URL

[8] M. Manfren, P. Caputo, G. Costa, Paradigm shift in urban energy systems through distributed generation: Methods and models, Applied Energy 88 (4) (2011) 1032 – 1048. doi:http://dx.doi.org/10.1016/j.apenergy.2010.10.018.
URL

[9] G. Chicco, P. Mancarella, Distributed multi-generation: A comprehensive view, Renewable and Sustainable Energy Reviews 13 (3) (2009) 535 – 551. doi:http://dx.doi.org/10.1016/j.rser.2007.11.014.
URL

[10] P. D. Diamantoulakis, V. M. Kapinas, G. K. Karagiannidis, Big Data Analytics for Dynamic Energy Management in Smart Grids, Big Data Research 2 (3) (2015) 94–101. doi:10.1016/j.bdr.2015.03.003.
URL

[11] K. Zhou, S. Yang, Understanding household energy consumption behavior: The contribution of energy big data analytics, Renewable and Sustainable Energy Reviews 56 (2016) 810–819. doi:10.1016/j.rser.2015.12.001.

[12] K. Zhou, C. Fu, S. Yang, Big data driven smart energy management: From big data to big insights, Renewable and Sustainable Energy Reviews 56 (2016) 215–225. doi:10.1016/j.rser.2015.11.050.

[13] R. M. Benders, R. Kok, H. C. Moll, G. Wiersma, K. J. Noorman, New approaches for household energy conservation-In search of personal household energy budgets and energy reduction options, Energy Policy 34 (18) (2006) 3612–3622. doi:10.1016/j.enpol.2005.08.005.

[14] T. Erickson, M. Li, Y. Kim, A. Deshpande, S. Sahu, T. Chao, P. Sukaviriya, M. Naphade, The Dubuque Electricity Portal: Evaluation of a City-Scale Residential Electricity Consumption Feedback System, Chi 2013 (2013) 1203–1212doi:10.1145/2470654.2466155.
URL

[15] A. Arteconi, N. J. Hewitt, F. Polonara, Domestic demand-side management (DSM): Role of heat pumps and thermal energy storage (TES) systems, Applied Thermal Engineering 51 (1-2) (2013) 155–165. doi:10.1016/j.applthermaleng.2012.09.023.
URL

[16] H. Shahrokni, D. Lazarevic, N. Brandt, Smart urban metabolism: Towards a Real-Time understanding of the energy and material flows of a city and its citizens, Journal of Urban Technology 22 (1) (2015) 65–86. doi:10.1080/10630732.2014.954899.
URL

[17] K. Hara, M. Uwasu, Y. Kishita, H. Takeda, Determinant factors of residential consumption and perception of energy conservation: Time-series analysis by large-scale questionnaire in Suita, Japan, Energy Policy 87 (2015) 240–249. doi:10.1016/j.enpol.2015.09.016.
URL

[18] D. Müller, A. Monti, S. Stinner, T. Schlösser, T. Schütz, P. Matthes, H. Wolisz, C. Molitor, H. Harb, R. Streblow, Demand side management for city districts, Building and Environment 91 (2015) 283–293. doi:10.1016/j.buildenv.2015.03.026.

[19] P. M. Á. D. Uribarri, U. Eicker, D. Robinson, Energy performance of decentralized solar thermal feed-in to district heating networks, Energy Procedia 116 (2017) 285–296. doi:10.1016/j.egypro.2017.05.075.
URL

[20] M. Batty, Big data , smart cities and city planning, Dialogues in Human Geography 3 (3) (2013) 274–279. doi:10.1177/2043820613513390.

[21] R. Kitchin, The real-time city ? Big data and smart urbanism, GeoJournal 79 (1) (2014) 1–14. doi:10.1007/s10708-013-9516-8.

[22] M. M. Rathore, A. Ahmad, A. Paul, S. Rho, Urban planning and building smart cities based on the Internet of Things using Big Data analytics, Computer Networks 101 (2016) 63–80. doi:10.1016/j.comnet.2015.12.023.
URL

[23] I. A. T. Hashem, V. Chang, N. B. Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, H. Chiroma, The role of big data in smart city, International Journal of Information Management 36 (5) (2016) 748–758. doi:10.1016/j.ijinfomgt.2016.05.002.
URL

[24] D. Robinson, F. Haldi, J. Kämpf, P. Leroux, D. Perez, A. Rasheed, U. Wilke, Citysim: Comprehensive micro-simulation of resource flows for sustainable urban planning, in: Proc. Building Simulation, 2009, pp. 1614–1627.

[25] U. Eicker, J. Schumacher, M. Zirak, N. Bartke, R. Nouvel, V. Coors, An automated method for urban energy simulation based on 3d city models, in: Proceedings of BS2015:14th Conference of International Building Per-

formance Simulation Association, Hyderabad, India, Dec.7-9, 2015.
URL

[26] G. Sousa, B. M. Jones, P. A. Mirzaei, D. Robinson, A review and critique of uk housing stock energy models, modelling approaches and data sources, Energy and Buildings 151 (2017) 66 – 80. doi:https://doi.org/10.1016/j.enbuild.2017.06.043.
URL

[27] G. Kazas, E. Fabrizio, M. Perino, Energy demand profile generation with detailed time resolution at an urban district scale: A reference building approach and case study, Applied Energy 193 (Supplement C) (2017) 243 – 262. doi:https://doi.org/10.1016/j.apenergy.2017.01.095.
URL

[28] M. Wetter, C. van Treeck, IEA EBC Annex 60: New Generation Computing Tools for Building and Community Energy Systems, 2017.
URL

[29] M. Dohler, I. Vilajosana, X. Vilajosana, J. LLosa, Smart Cities : An Action Plan, Proceedings of Barcelona Smart Cities Congress (2011) 1–6.
URL

[30] A. Zanella, N. Bui, A. Castellani, L. Vangelista, M. Zorzi, Internet of Things for Smart Cities, IEEE Internet of Things Journal 1 (1) (2014) 22–32. doi:10.1109/JIOT.2014.2306328.
URL

[31] J. A. Clarke, J. L. M. Hensen, Integrated building performance simulation : Progress , prospects and requirements, Building and Environment 91 (2015) 294–306. doi:10.1016/j.buildenv.2015.04.002.
URL

[32] G. Tardioli, R. Kerrigan, M. Oates, J. O. Donnell, D. Finn, A Data-Driven Modelling Approach for Large Scale Demand Profiling of Residential Buildings, in: International Building Physics Simulation Association (IBPSA), San Francisco (CA), US, 2017.

[33] 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 and Buildings 107 (2015) –. doi:http://dx.doi.org/10.1016/j.enbuild.2015.08.021.
URL

[34] C. F. Reinhart, C. C. Davila, Urban Building Energy Modeling - A Review of a Nascent Field, Building and Environment 97 (2016) 196–202. doi:10.1016/j.buildenv.2015.12.001.
URL

[35] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renewable and Sustainable Energy Reviews 37 (2014) 123–141. arXiv:j.rser.2014.05.007, doi:10.1016/j.rser.2014.05.007.
URL

[36] G. Mustafaraj, D. Marini, A. Costa, M. Keane, Model calibration for building energy efficiency simulation, Applied Energy 130 (2014) 72–85.

[37] M. Royapoor, T. Roskilly, Building model calibration using energy and environmental data, Energy and Build-

ings 94 (2015) 109–120.

[38] Z. Yang, B. Becerik-Gerber, A model calibration framework for simultaneous multi-level building energy simulation, Applied Energy 149 (2015) 415–431. doi:10.1016/j.apenergy.2015.03.048.
URL

[39] E. Fabrizio, V. Monetti, Methodologies and advancements in the calibration of building energy models, Energies 8 (4) (2015) 2548–2574.

[40] J. Robertson, B. Polly, J. Collis, Evaluation of automated model calibration techniques for residential building energy simulation, National Renewable Energy Laboratory, 2013.

[41] T. Hong, J. Kim, J. Jeong, M. Lee, C. Ji, Automatic calibration model of a building energy simulation using optimization algorithm, Energy Procedia 105 (2017) 3698–3704. doi:10.1016/j.egypro.2017.03.855.
URL

[42] G. Chaudhary, J. New, J. Sanyal, P. Im, Z. O'Neill, V. Garg, Evaluation of autotune calibration against manual calibration of building energy models, Applied Energy 182 (2016) 115–134.

[43] F. Monari, A quasi-Bayesian framework for the development of computer models (2001). arXiv:1710.06671.
URL

[44] Y. Shimoda, T. Fujii, T. Morikawa, M. Mizuno, Residential end-use energy simulation at city scale, Building and environment 39 (8) (2004) 959–967.

[45] S. P. Corgnati, E. Fabrizio, M. Filippi, V. Monetti, Reference buildings for cost optimal analysis: Method of definition and application, Applied energy 102 (2013) 983–993.

[46] Y. Yamaguchi, Y. Shimoda, District-scale simulation for multi-purpose evaluation of urban energy systems, Journal of Building Performance Simulation 3 (4) (2010) 289–305.

[47] C. Cerezo Davila, C. F. Reinhart, J. L. Bemis, Modeling Boston: A workflow for the efficient generation and maintenance of urban building energy models from existing geospatial datasets, Energy 117 (2016) 237–250. doi:10.1016/j.energy.2016.10.057.
URL

[48] A. Schaefer, E. Ghisi, Method for obtaining reference buildings, Energy and Buildings 128 (2016) 660–672. doi:10.1016/j.enbuild.2016.07.001.
URL

[49] N. Ghiassi, A. Mahdavi, Reductive bottom-up urban energy computing supported by multivariate cluster analysis, Energy and Buildings 144 (2017) 372–386. doi:10.1016/j.enbuild.2017.03.004.
URL

[50] G. Tardioli, R. Kerrigan, M. Oates, J. O'Donnell, D. P. Finn, Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach, Building and Environment (2018) –doi:https://doi.org/10.1016/j.buildenv.2018.05.035.
URL

[51] J. Sokol, C. C. Davila, C. F. Reinhart, Validation of a bayesian-based method for defining residential archetypes in urban building energy models, Energy and Buildings 134 (2017) 11 – 24. doi:https://doi.org/10.1016/j.enbuild.2016.10.050.

URL
[52] J. Virote, R. Neves-Silva, Stochastic models for building energy prediction based on occupant behavior assessment, Energy and Buildings 53 (2012) 183–193. doi:10.1016/j.enbuild.2012.06.001.
URL
[53] P. A. Mirzaei, Recent challenges in modeling of urban heat island, Sustainable Cities and Society 19 (2015) 200–206. doi:10.1016/j.scs.2015.04.001.
URL
[54] M. C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464. doi:10.1111/1467-9868.00294.
URL
[55] Y. Heo, G. Augenbroe, R. Choudhary, Quantitative risk management for energy retrofit projects, Journal of Building Performance Simulation 6 (4) (2013) 257–268. doi:10.1080/19401493.2012.706388.
[56] M. H. Kristensen, R. Choudhary, R. H. Pedersen, S. Petersen, Bayesian Calibration Of Residential Building Clusters Using A Single Geometric Building Representation, The 15th International Conference of IBPSA.
[57] Y. Heo, D. J. Graziano, L. Guzowski, R. T. Muehleisen, Evaluation of calibration efficacy under different levels of uncertainty, Journal of Building Performance Simulation 8 (3) (2015) 135–144.
[58] G. Mustafaraj, D. Marini, A. Costa, M. Keane, Model calibration for building energy efficiency simulation, Applied Energy 130 (2014) 72–85. doi:http://dx.doi.org/10.1016/j.apenergy.2014.05.019.
URL
[59] T. A. Reddy, I. Maor, C. Panjapornpon, T. A. Reddy, Calibrating Detailed Building Energy Simulation Programs with Measured Data - Part I : General Methodology ( RP-1051 ) Simulation Programs with Measured Data - Part I : General Methodology ( RP-1051 ) 9669 (April). doi:10.1080/10789669.2007.10390952.
[60] N. C. Kumar, A. K. Subramaniyan, L. Wang, Improving high-dimensional physics models through bayesian calibration with uncertain data, in: Proceedings of ASME Turbo Expo, 2012.
[61] S. Yamamoto, M. Miyazaki, T. Iwano, S. Kitazawa, Bayesian calibration of simultaneity in audiovisual temporal order judgments, PloS one 7 (7) (2012) e40379.
[62] M. Van Oijen, C. Reyer, F. Bohn, D. Cameron, G. Deckmyn, M. Flechsig, S. Härkönen, F. Hartig, A. Huth, A. Kiviste, et al., Bayesian calibration, comparison and averaging of six forest models, using data from scots pine stands across europe, Forest Ecology and Management 289 (2013) 255–268.
[63] D. S. Mebane, K. S. Bhat, J. D. Kress, D. J. Fauth, M. L. Gray, A. Lee, D. C. Miller, Bayesian calibration of thermodynamic models for the uptake of co 2 in supported amine sorbents using ab initio priors, Physical Chemistry Chemical Physics 15 (12) (2013) 4355–4366.
[64] A. Hawkins-Daarud, S. Prudhomme, K. G. van der Zee, J. T. Oden, Bayesian calibration, validation, and uncertainty quantification of diffuse interface models of tumor growth, Journal of mathematical biology 67 (6-7) (2013) 1457–1485.
[65] Q. Li, G. Augenbroe, J. Brown, Assessment of linear emulators in lightweight Bayesian calibration of dynamic building energy models for parameter estimation and performance prediction, Energy and Buildings 124 (2016) 194–202. doi:10.1016/j.enbuild.2016.04.025.
[66] W. Tian, S. Yang, Z. Li, S. Wei, W. Pan, Y. Liu, Identifying informative energy data in Bayesian calibration of

building energy models, Energy and Buildings 119 (2016) 363–376. doi:10.1016/j.enbuild.2016.03.042.
URL

[67] Y. Kang, M. Krarti, Bayesian-Emulator based parameter identification for calibrating energy models for existing buildings, Building Simulation 9 (4) (2016) 411–428. doi:10.1007/s12273-016-0291-6.
URL

[68] H. Lim, Review on stochastic modeling methods for building stock energy prediction (2017) 607–624doi:10.1007/s12273-017-0383-y.

[69] J. Yuan, V. Nian, B. Su, Q. Meng, A simultaneous calibration and parameter ranking method for building energy models, Applied Energy 206 (2017) 657–666. doi:10.1016/j.apenergy.2017.08.220.
URL

[70] A. Chong, K. P. Lam, M. Pozzi, J. Yang, Bayesian calibration of building energy models with large datasets, Energy and Buildings 154 (2017) 343–355. doi:10.1016/j.enbuild.2017.08.069.
URL

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

[72] A. Booth, R. Choudhary, D. Spiegelhalter, Handling uncertainty in housing stock models, Building and Environment 48 (2012) 35 – 47. doi:https://doi.org/10.1016/j.buildenv.2011.08.016.
URL

[73] W. Tian, R. Choudhary, A probabilistic energy model for non-domestic building sectors applied to analysis of school buildings in greater London, Energy and Buildings 54 (2012) 1–11. doi:10.1016/j.enbuild.2012.06.031.
URL

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

[75] Y. Yamaguchi, Y. Suzuki, R. Choudhary, A. Booth, Y. Shimoda, Urban-scale energy modelling of food supermarket considering uncertainty, The proceedings of Building Simulation.

[76] Y. J. Kim, C. S. Park, Stepwise deterministic and stochastic calibration of an energy simulation model for an existing building, Energy and Buildings 133 (2016) 455–468. doi:10.1016/j.enbuild.2016.10.009.
URL

[77] R. L. Iman, Latin hypercube sampling, Encyclopedia of quantitative risk analysis and assessment.

[78] M. Stein, Large sample properties of simulations using latin hypercube sampling, Technometrics 29 (2) (1987) 143–151.

[79] L. Uusitalo, A. Lehikoinen, I. Helle, K. Myrberg, An overview of methods to evaluate uncertainty of deterministic models in decision support, Environmental Modelling and Software 63 (2015) 24–31. doi:10.1016/j.envsoft.2014.09.017.
URL

[80] T. J. Hastie, R. J. Tibshirani, J. H. Friedman, The elements of statistical learning : data mining, inference, and prediction, Springer series in statistics, Springer, New York, 2009, autres impressions : 2011 (corr.), 2013 (7e corr.).
URL

[81] H.-X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews 16 (2012) 3586–3592. doi:10.1016/j.rser.2012.02.049.

[82] B. Gunay, W. Shen, G. Newsham, Inverse blackbox modeling of the heating and cooling load in office buildings, Energy and Buildings 142 (2017) 200–210. doi:10.1016/j.enbuild.2017.02.064.
URL

[83] Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines, International Journal of Electrical Power and Energy Systems 67 (2015) 431–438. doi:10.1016/j.ijepes.2014.12.036.
URL

[84] K. Amasyali, N. M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews 81 (2018) 1192–1205. doi:10.1016/j.rser.2017.04.095.
URL

[85] J. G. Jetcheva, M. Majidpour, W. P. Chen, Neural network model ensembles for building-level electricity load forecasts, Energy and Buildings 84 (2014) 214–223. doi:10.1016/j.enbuild.2014.08.004.
URL

[86] E. M. Burger, S. J. Moura, Gated ensemble learning method for demand-side electricity load forecasting, Energy and Buildings 109 (2015) 23–34.

[87] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. Capretz, G. Bitsuamlak, An ensemble learning framework for anomaly detection in building energy consumption, Energy and Buildings 144 (2017) 191–206. doi:10.1016/j.enbuild.2017.02.058.
URL

[88] Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks, Energy and Buildings 121 (2016) 284–297. doi:10.1016/j.enbuild.2015.12.050.
URL

[89] A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussin, H. Abdullah, R. Saidur, A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renewable and Sustainable Energy Reviews 33 (2014) 102–109. doi:10.1016/j.rser.2014.01.069.
URL

[90] F. Magoulès, M. Piliougine, D. Elizondo, Support Vector Regression for Electricity Consumption Prediction in a Building in Japan, in: Proceedings - 19th IEEE International Conference on Computational Science and Engineering, 14th IEEE International Conference on Embedded and Ubiquitous Computing and 15th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, CSE-EUC-DCABES 2016, IEEE, 2017, pp. 189–196. doi:10.1109/CSE-EUC-DCABES.2016.184.
URL

[91] Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes, Energy

and Buildings 130 (2016) 519–531. doi:10.1016/j.enbuild.2016.08.081.
URL

[92] F. Zhang, C. Deb, S. E. Lee, J. Yang, K. W. Shah, Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique, Energy and Buildings 126 (2016) 94–103. doi:10.1016/j.enbuild.2016.05.028.
URL

[93] A. Tsanas, A. Xifara, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energy and Buildings 49 (2012) 560–567. doi:10.1016/j.enbuild.2012.03.003.

[94] A. Kusiak, M. Li, Z. Zhang, A data-driven approach for steam load prediction in buildings, Applied Energy 87 (2010) 925–933. doi:10.1016/j.apenergy.2009.09.004.

[95] J. Ma, J. C. Cheng, Identifying the influential features on the regional energy use intensity of residential buildings based on Random Forests, Applied Energy 183 (2016) 193–201. doi:10.1016/j.apenergy.2016.08.096.
URL

[96] M. a. Razi, K. Athappilly, A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models, Expert Systems with Applications 29 (1) (2005) 65–74. doi:10.1016/j.eswa.2005.01.006.

[97] M. Espinoza, C. Joye, R. Belmans, B. De Moor, Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series, IEEE Transactions on Power Systems 20 (3) (2005) 1622–1630. doi:10.1109/TPWRS.2005.852123.

[98] G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Blay, M. Alcázar-Ortega, New artificial neural network prediction method for electrical consumption forecasting based on building end-uses, Energy and Buildings 43 (2011) 3112–3119. doi:http://dx.doi.org/10.1016/j.enbuild.2011.08.008.
URL

[99] W. Ma, S. Fang, G. Liu, R. Zhou, Modeling of district load forecasting for distributed energy system, Applied Energy 204 (2017) 181–205. doi:10.1016/j.apenergy.2017.07.009.
URL

[100] X. Lü, T. Lu, C. J. Kibert, M. Viljanen, Modeling and forecasting energy consumption for heterogeneous buildings using a physicalâĂŞstatistical approach, Applied Energy 144 (2015) 261–275. doi:10.1016/j.apenergy.2014.12.019.
URL

[101] K. T. Williams, J. D. Gomez, Predicting future monthly residential energy consumption using building characteristics and climate data: A statistical learning approach, Energy and Buildings 128 (2016) 1–11. doi:10.1016/j.enbuild.2016.06.076.

[102] N. J. D. Nagelkerke, A note on a general definition of the coefficient of determination, Biometrika 78 (3) (1991) 691–692.
URL

[103] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, R. D. Ryne, Combining Field Data and Computer Sim-

ulations for Calibration and Prediction, SIAM Journal on Scientific Computing 26 (2) (2004) 448–466. doi:10.1137/S1064827503426693.
URL

[104] W. R. Gilks, S. Richardson, D. Spiegelhalter, Markov chain Monte Carlo in practice, CRC press, 1995.

[105] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics 21 (6) (1953) 1087–1092.

[106] T. A. Statistician, N. Nov, Understanding the Metropolis-Hastings Algorithm Siddhartha Chib ; Edward Greenberg 49 (4) (2007) 327–335.

[107] C. P. Robert, G. Casella, The metropolis—hastings algorithm, in: Monte Carlo Statistical Methods, Springer, 1999, pp. 231–283.

[108] Sitg: Le système d'information du territoire à genève, , accessed: 01/03/2017.

[109] Geneva weather data, , accessed: 2017-01-07.

[110] Swiss society of engineers and architects normative, , accessed: 2017-03-04.

[111] D. Perez, K. Henri, Urban Area Energy Flow Microsimulation for Planning Support : a Calibration and Verification Study, International Journal on Advances in Systems and Measurements 6 (3) (2013) 260–271.

[112] J. Khoury, Rénovation énergétique des bâtiments résidentiels collectifs: état des lieux, retours d'expérience et potentiels du parc genevois, Ph.D. thesis (12/16 2014).
URL

[113] Y. Zhang, I. Korolija, 9th SET Conference Proceedings, Shanghai, China, in: SET Conference, 2010.

[114] K. Menberg, Y. Heo, R. Choudhary, Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information, Energy and Buildings 133 (2016) 433–445. doi:10.1016/j.enbuild.2016.10.005.
URL

[115] W. Tian, P. De Wilde, Uncertainty and sensitivity analysis of building performance using probabilistic climate projections: A UK case study, Automation in Construction 20 (8) (2011) 1096–1109. doi:10.1016/j.autcon.2011.04.011.
URL

[116] C. J. Hopfe, J. L. Hensen, Uncertainty analysis in building performance simulation for design support, Energy and Buildings 43 (10) (2011) 2798–2805. doi:10.1016/j.enbuild.2011.06.034.
URL

[117] 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 and Buildings 60 (2013) 1–11. doi:10.1016/j.enbuild.2013.01.005.
URL

[118] M. H. Kristensen, S. Petersen, Choosing the appropriate sensitivity analysis method for building energy model-based investigations, Energy and Buildings 130 (2016) 166–176. doi:10.1016/j.enbuild.2016.08.038.
URL

[119] A. Mastrucci, P. Pérez-López, E. Benetto, U. Leopold, I. Blanc, Global sensitivity analysis as a support for the generation of simplified building stock energy models, Energy and Buildings 149 (2017) 368–383. doi:10.

1016/j.enbuild.2017.05.022.
URL

[120] D. Garcia Sanchez, B. Lacarrire, M. Musy, B. Bourges, Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods, Energy and Buildings 68 (PART C) (2014) 741–750. doi:10.1016/j.enbuild.2012.08.048.
URL