



# Research Repository UCD

<b>Title</b>	A Novel Hybrid Technique For Building Demand Forecasting Based On Data-driven And Urban Scale Simulation Approaches
<b>Authors(s)</b>	Tardioli, Giovanni, Kerrigan, Ruth, Oates, Mike, O'Donnell, James, Finn, Donal
<b>Publication date</b>	2019-09-04
<b>Publication information</b>	Tardioli, Giovanni, Ruth Kerrigan, Mike Oates, James O'Donnell, and Donal Finn. "A Novel Hybrid Technique For Building Demand Forecasting Based On Data-Driven And Urban Scale Simulation Approaches." IBPSA, September 4, 2019. <a href="https://doi.org/10.26868/25222708.2019.211169">https://doi.org/10.26868/25222708.2019.211169</a> .
<b>Conference details</b>	The 16th Conference of International Building Simulation Association (BS 2019), Rome, Italy, 2-4 September 2019
<b>Publisher</b>	IBPSA
<b>Item record/more information</b>	<a href="http://hdl.handle.net/10197/12261">http://hdl.handle.net/10197/12261</a>
<b>Publisher's version (DOI)</b>	<a href="https://doi.org/10.26868/25222708.2019.211169">10.26868/25222708.2019.211169</a>

Downloaded 2025-12-04 22:56:18

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd\_oa)



© Some rights reserved. For more information

---

# A Novel Hybrid Technique For Building Demand Forecasting Based On Data-driven And Urban Scale Simulation Approaches

Giovanni Tardioli<sup>1,2</sup>, Ruth Kerrigan<sup>1</sup>, Mike Oates<sup>1</sup>, James O'Donnell<sup>2</sup>, Donal P. Finn<sup>2</sup>

1: Integrated Environmental Solutions (IES) R&D, Dublin, Ireland

2: School of Mechanical & Materials Engineering, University College Dublin, Dublin, Ireland

## Abstract

This paper presents a novel feature engineering procedure to generate case study specific input variables for the training of data-driven models used to predict the heating demand of blocks of buildings. Traditionally, predictive model training is performed using sets of data from sensors (e.g. weather stations, metering systems). Feature engineering procedures such as the inclusion of innovative predictive variables in the forecasting framework are generally not considered. The method presented in this paper exploits results of calibrated physics-based building energy models that are included as an additional independent variable in combination with the traditional sets of predictors in an innovative forecasting framework. The method is tested on a district case study of the city of Geneva (CH) served by a district heating network. Results show that the presented approach improves the quality of the forecasting outcomes of state-of-the-art predictive algorithms. In this context, the accuracy of the simulation outputs affects the predictive capability of the presented forecasting procedure. In addition, normalised information derived from substation of the heating network of the district are informative for the predictive model.

## Introduction

Buildings have been identified as one of the major energy consumers in urban environments (Kylili and Fokaides, 2015). Heating and domestic hot water usage are amongst the major power-intensive processes in buildings (Sharif et al., 2015). Among possible strategies that could foster urban scale energy efficiency, district heating networks cover a primary role (Lund et al., 2014). An optimised design and operation of district heating systems can be performed by evaluating demand profiles of groups of buildings (blocks or clusters) which will be served by network substations (Marquant et al., 2018). Accurate forecasting is crucial to facilitate the correct use of the network, and to design energy saving techniques such as storage systems and optimised unit commitment of generation plants (Short et al., 2017).

Nevertheless, an evaluation of building demand at different time horizons and granularities is a difficult task due to unpredictable phenomena (such as uncertainties in weather variables and building occupancy) or operations

and control system constraints of the network (set-back temperatures, partial-load periods, inactivity due to legislation requirements, etc.) (Noussan et al., 2017). Thus, methods to enhance the predictive capability of forecasting techniques are important.

From a building perspective, physics-based simulations and data-driven approaches (Zhao and Magoulès, 2012) are used to characterise the heating demand at a single building level. In recent years, many urban scale modelling approaches based on building physics modelling have been presented (Eicker et al., 2015; Kazas et al., 2017). On the other hand, data-driven models (DDM), offering high levels of accuracy when historical data are available, are used to take into account occupant behaviour and control operations, and therefore, are mostly employed at an operational stage of the building (Wei et al., 2018). Nevertheless, they rely on the available building data, thus, dedicated metering systems or sensors are required (Molina-solana et al., 2017).

One of the major limitation of DDM techniques is the requirement in terms of large datasets for model training. Moreover, contrary to building physics modelling, DDM models are neither suitable for scenario evaluation nor for investigation of innovative design options. The quality of the predictions of DDM is directly related to the quality of the training datasets, as well as the selection and inclusion of training variables. Therefore, the use of feature engineering procedures and input selection methods to enhance the predictive capability of the forecasting method is a fundamental step to enhance the accuracy of the predictions. In this paper, an innovative use of building physics simulation, used as value-added data service, is presented. The hypothesis of the current work is that building physics simulation can process the available data by creating informative variables at different time and space granularities, that can be included in a predictive framework to enhance its accuracy. Therefore, three objectives are addressed in this paper:

1. to record if an improvement in the quality of the prediction of data-driven techniques occurs when independent variables are augmented with simulation results from building energy models;
2. to test if an improvement of the outcomes from building energy modelling, assessed by using comparisons

## Nomenclature

<i>ACH</i>	Air Change Per Hour
<i>ANN</i>	Artificial Neural Network
<i>BEM</i>	Building Energy Model
<i>CVRMSE</i>	Coefficient of Variation of the Root Mean Squared Error
<i>DDM</i>	Data Driven Model
<i>DHN</i>	District Heating Network
<i>DHW</i>	Domestic Hot Water
<i>GLR</i>	Generalised linear regression
<i>HVAC</i>	Heating Ventilation and Air Conditioning

<i>MAPE</i>	Mean Absolute Percentage Error
<i>MBE</i>	Mean Bias Error
<i>MLR</i>	Multiple linear regression
<i>N</i>	Number of predictions
<i><math>\hat{p}</math></i>	predicted value
<i><math>R^2</math></i>	Coefficient of Determination
<i>RMSE</i>	Root Mean Squared Error
<i>RNF</i>	Random Forest
<i>SVM</i>	Support Vector Machine
<i>y</i>	Metered value

with metered data, can generate an increment of the importance, as a predictor, of the simulated demand in the predictive framework;

3. to test if metered data collected from similar buildings can constitute useful information to enhance the prediction accuracy of the predictive modelling workflow.

The research methodology and the main objectives of the current paper are tested using a district of the city of Geneva (CH) as case study.

## Literature Review

Table 1 provides an overview of the main applications of DDM in the building energy sector for prediction of building heating requirements. The table is divided into algorithms, main inputs, outputs, resolution of the target variables, scale of the analysis and level of aggregation of the case studies.

There are many examples of the use of DDM in the literature for the forecasting of building heating demand (Ahmad and Chen, 2018). Although, there has been a growing interest among researchers in studying DDM for design purposes (Chou and Bui, 2014), the majority of the studies focus on the operational phase of the building (post-occupancy) where predictive models can take into account past occupant behaviour (Li and Wen, 2014).

Residential buildings, together with commercial and university buildings, are the most investigated building typologies (Powell et al., 2014). Aggregated case study are investigated gathering measurements of heating demand at generation points or sampled at substation level with an hourly time resolution (primary-secondary network connections or at customer point connection) due to sensors measurements availability (flow meters and thermocouples) (Noussan et al., 2017).

In terms of predictive resolution, the majority of the studies focus on prediction at a hourly time step, followed by daily and monthly forecasts. Generally the resolution of the target variable is at a hourly time step. Intra-hour prediction studies for heating demand forecasting are very rare and missing (Kapetanakis et al., 2017).

In terms of time horizons, studies focus on short time predictions such as hourly and next day load forecasting on next week analysis (Rahman and Smith, 2018). Common

parts of these studies is that for an extension of the prediction horizon the predictive error tends to increase due to a lack of supportive/informative measurements. Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), Support Vector Machines(SVM) are amongst the most used predictive algorithms. Emergent, innovative predictive techniques are deep-learning methods and ensemble learning approaches (Suryanarayana et al., 2018).

In terms of input selection (features), the majority of the studies focuses on 'filters' and 'wrappers' approaches. Table 1 shows that heating demand forecasting is performed using a small number of predictors. In particular, variables related to weather conditions are largely employed. Particular control methods, such as partial load operation during summer time or particular set-back temperatures at night-time, can partially de-correlate heating demand from external temperature. Time information, such as day-type (working day, week-end, holidays) and time of the day (hour), are recurrent categorical variables for prediction. Past consumption is used in different cases as an additional predictive variable of building demand. This is included as previous hour demand, previous day demand, etc. A large part of the studies includes ad-hoc predictors which are in most of the cases occupant related information.

Enhancements in the quality of the predicted demand at a urban scale can generate energy and cost savings. So far, new independent variables have been investigated or proposed as additional predictors in machine learning frameworks. In this study, two innovative predictive variables are considered to support prediction of groups of building heating demand: aggregated simulation outputs of blocks of buildings and a normalised heating demand of all the substations of the district.

## Methodology

The proposed methodology consists of three main steps as shown in Figure 1.

### 1 Physics based modelling.

Building data from different sources of the case study are collected and organised in a database. Physics-based modelling is used for the development of energy models of single buildings, as part of a building block.

An automated urban scale building energy modelling approach, hereafter called *AUSBEMA* has been developed. Accurate details of the modelling capabilities of the large-

Table 1: Literature review Table on DDM applied in the building energy sector

Case studies and models						Inputs										Output	
						Weather variables						Time variables		Other		Demand	
Ref.	Case Study	Algorithms				Resol.	Out air T.	Int air T.	Glob. rad	W. speed	W. dir.	R.H.	Day	Hour	Cons.	Other	
		MLR	ANN	SVM	Other												
Naji et al. (2016)	Res		X		X	Annual										X	H,C
Macas et al. (2016)	Offi		X			12 hours	X	X	X	X	X	X				X	H
Sholahudin and Han (2016)	Res		X			Hourly	X	X	X	X						X	H
Al-shammari et al. (2016)	Mix			X		Hourly	X								X	X	H
Idowu et al. (2016)	Mix	X	X	X	X	Hourly	X								X	X	H
Song et al. (2017)	Camp		X		X	Hourly	X		X	X	X	X	X			X	E,H
Gunay et al. (2017)	Offi	X	X			Hourly	X		X	X	X	X	X	X	X	X	H
Rahman and Smith (2017)	Comm	X	X		X	Hourly	X		X	X	X	X	X	X		X	H
Dahl et al. (2017)	Mix				X	Hourly	X		X	X	X				X	X	H
Paudel et al. (2017)	Res			X	X	Hourly	X		X				X	X		X	H
Fan et al. (2017)	Edu	X	X	X	X	Hourly							X	X	X	X	H
Yuan et al. (2018)	Mix	X	X			Daily	X									X	H
Suryanarayana et al. (2018)	Mix	X	X		X	Hourly	X						X	X	X	X	H
Geysen et al. (2018)	Mix	X	X	X	X	Hourly	X						X	X	X	X	H

scale simulation method are provided elsewhere (Prieto et al., 2019). The automated approach uses building data of the city case study to generate input files for dynamic simulation in EnergyPlus. A Python script maps and replicates the simulation template (idf file), enriching it with information from the case study. To achieve this, the model uses data including the geo-location of a building, the 3D model of the city (CityGML, LOD2) and information from different datasets of the urban building stock. Starting from the 3D geometry, the script can include in the input file of the simulation, a multi-zone representation (one zone for each floor) of the building, a simplified description of the heating and domestic hot water systems and can take into account the urban environment in terms of surroundings, adjacencies and shading scenarios of the target building. The outcomes of this step are a set of energy models of the buildings in the case study that can be used for simulation.

Simulations are aggregated to generate heating load profiles of the groups of buildings. An evidence based approach for manual calibration is used to tune building energy modelling parameters and match as much as possible the profile of the metered data. Metrics based on the ASHRAE guidelines 14-2014, i.e. CVRMSE and NMBE, are used to define the quality of the simulation results (ASHRAE, 2014). The metrics are described as follows:

$$CV(RMSE) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2}}{\bar{y}} \quad (1)$$

where N is the total number of predictions, y is the actual value of measurement i and p is its predicted value and  $\bar{y}$  is the average value and,

$$NMBE = \frac{1}{N} \frac{\sum_{i=1}^N (y_i - p_i)}{\bar{y}} \quad (2)$$

where N is the total number of predictions, y is the actual

value of measurement i and p is its predicted value and  $\bar{y}$  is the average value.

## 2 Data-driven modelling.

Metered data are collected from the building case studies (district heating substations) and stored in a database, then pre-processed and structured. Based on the control system of the district heating network, it may be useful to structure building demand profiles on the basis of three periods of operation: winter, summer and a transition period, spring-autumn.

A normalised district average profile is evaluated by aggregation and normalisation of the available metering information. This is used to derive an average hourly profile of the entire network. Various predictive algorithms are tested (ANN, MLR, SVN, RNF). These models are considered as the reference approaches (baseline methods) to estimate the improvement derived by including the two additional variables in the training phase. To estimate the prediction capability of each model, five metrics are used: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), Coefficient of Determination (R2), Coefficient of Variation of the Root Mean Squared Error (CVRMSE) and the Normalised Mean Biased Error (NMBE). Considering each of these, they can be defined as:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - p_i}{y_i} \right| \cdot 100 \quad (3)$$

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

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2} \quad (4)$$

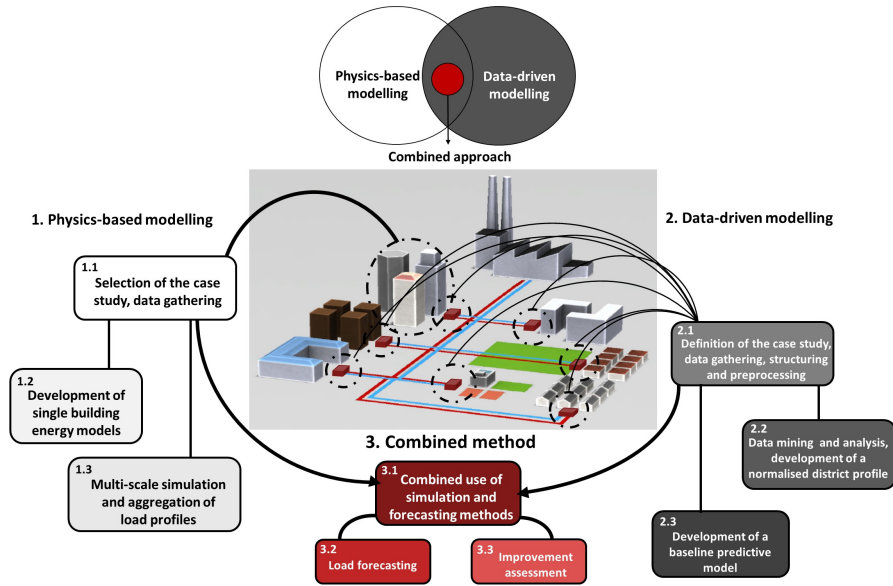


Figure 1: Methodology overview: main steps of the proposed combined modelling approach

where  $N$  is the total number of predictions,  $y$  is the actual value of measurement  $i$  and  $p$  is its predicted value;

$$R^2 \equiv 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^N (y_i - p_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (5)$$

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  $\bar{y}$  is the mean value.

Visualisation techniques are used to further judge the quality of the emulators, such as actual versus predicted plots. A procedure for the training of the predictive model is described in Figure 2. Data are initially split in a training and testing dataset (respectively 70 and 30% of the total). The predictive models are trained on the training sample. The training procedure is performed by using a 10-fold cross validation technique to automatically tune algorithmic parameters and select the best performing method. The out-of-sample dataset (30% of the initial dataset) is used to estimate the effective prediction capacity of each algorithm and to avoid overfitting. The final accuracy measures evaluated in the testing set rank the best algorithms.

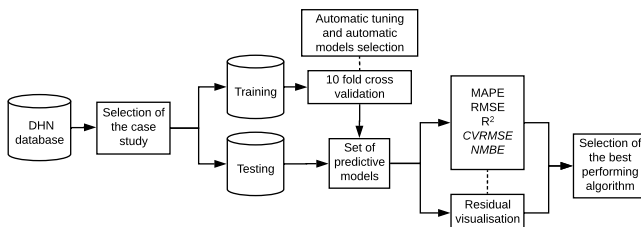


Figure 2: Modelling steps for the development of predictive models using data-driven techniques

In a second step, the importance of each predictor is considered and evaluated with a wrapper method: LASSO (Fonti and Belitser, 2017).

### 3 Combined method.

Building simulation results and data collected from similar buildings are included as additional predictors in the predictive approach. Model creation follows the same steps described in 2. The quality of the prediction is assessed training the predictive model on the training dataset augmented by simulation results. Importance of the predictors is assessed using the LASSO method. Several tests are performed at different time horizons to assess the quality of the predictive technique. Predictor importance is assessed for different groups of buildings in the district case study to identify relationships between quality of the simulation results and forecasting quality.

#### Case study

The district of Meyrin of the city of Geneva is considered as the case study for testing the research methodology. A large part of the buildings in the case study is supplied by a district heating network. The operating temperature at substation level is on average between 115°C in winter and 87°C in summer, when the network is operated at partial load. The majority of the buildings of the district are residential multi-family and single-family houses. Although mixed-use and commercial buildings are also present in the district, in this study, three blocks of residential buildings were considered. Each building block consists of: 4, 14 and 5 buildings, respectively. Each block of buildings is served by a different district heating substation. A summary of the information regarding the three different case studies is provided in Table 2.

A physics-based simulation method is used to model the three different building blocks after gathering data from the case study. An evidence based approach based on recursive parameter adjustment is used to calibrate the energy models. A summary of the modelling information used for the three case studies is provided in Table 3: An example of load profile for one of the group of buildings

Table 2: Building block

Case study (buildings)	Block 1	Block 2	Block 3
Buildings	4	14	5
Heating surface [ $m^2$ ]	7317	18187	7565
Volume [ $m^3$ ]	25368	63741	23561
Peak Load (2014)[kW]	245	890	398

within the network is provided in Figure 3.

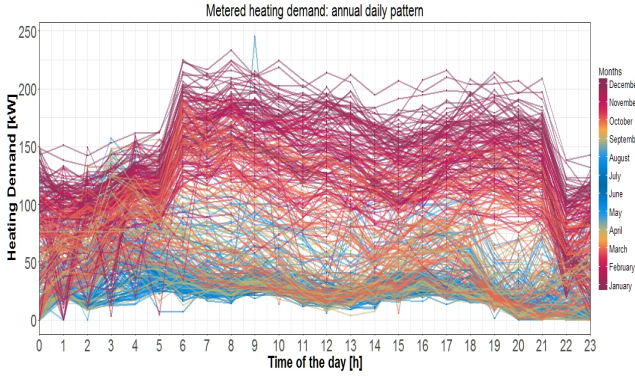


Figure 3: Example of heating demand profile of substation 1 supplying a residential block of buildings in the district case study.

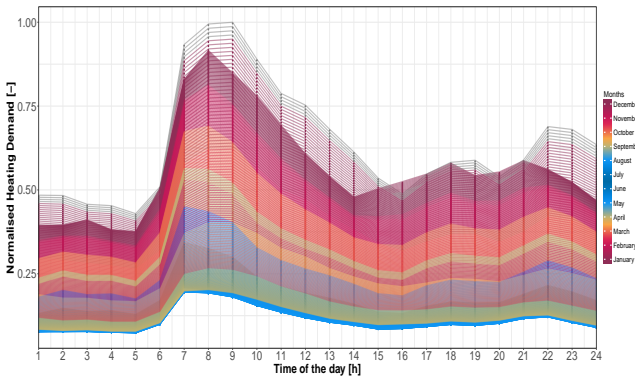


Figure 4: Normalised average daily heating profile  $P_{norm}$  for the year 2014 of all the substations in the district case study

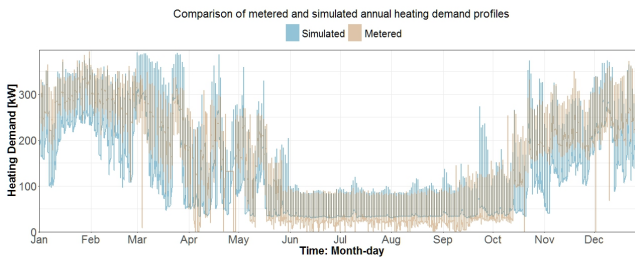


Figure 5: Example of comparison of simulated versus metered heating profiles for a substations case study

An average profile of the heating demand for the network is evaluated. The average profile is calculated using the following equation:

$$\overline{P_{norm}} = \frac{1}{N} \sum_{i=1}^N \frac{P_{s,i,j}}{P_{max,j}} \quad \forall i \in (1, 8760) \quad (6)$$

where,  $N$  is the number of substations in the district case study (45 substations),  $P_{s,i,j}$  is the metered profile at time step  $i$  of substation  $j$ ,  $P_{max,j}$  is the max profile of substation  $j$ . The heating profile of the network presents an irregular trend, when averaged across all the substations, due to different operating conditions in terms of final use and customers behaviour of the single substations. Therefore, a smoothing approach was used to derive a final trend. The profile is evaluated using an univariate interpolation method. A cubic equation ( $y = ax^3 + bx^2 + cx + d$ ) was used to fit the annual normalised profile for each hour of the day for the entire year. The result of the smoothing process is shown in Figure 4.

## Results

### Physics based modelling

An example of the comparison between metered data and the aggregated profile of a group of buildings is shown in Figure 5. Table 4 summarises the error metrics associated with the calibration process in terms of CVRMSE and NMBE. Results show different levels of accuracy achieved by the simulation models for the different case studies.

With reference to Table 4, and considering the quality guideline indicated for hourly calibration (ASHRAE Guideline 14.2014:  $CVRMSE < 30\%$ ,  $NMBE < 0.1$ ), in some instances and seasons, the models are able to achieve measures in the range of acceptability of calibrated simulations. The summer period is the most difficult to simulate given the low and variable heating demand of the building.

### Data-Driven modelling

As explained in part 2 (Data Driven Modelling) of the research methodology, a full data-driven model is trained on the available data. Figure 6 shows the results for one substation and for one week in winter. It is evident that all the predictive models are able to provide quite accurate forecasting when employed for predictions at short time horizons and supported by weather information and historical demand of the buildings connected to the substations. However, when the operation of the network is decoupled from the external temperature trend, the predictive error can be high (summer periods).

In terms of residuals (Figure 7), for the case study considered, the predictive model generates distributions centred on 0, attesting that the models are well trained and a low or absent bias is generated during the training phase with small noise learnt by the predictive algorithms. An example of the hierarchy of the strength of the predictors evaluated with the LASSO method is shown in Figure 8. Results attest the importance of the past demand of the network as well as the outside temperature to predict future heating demand. The average demand of the network is a useful predictor and it ranks quite high on the hierarchy list. An analysis of additional results (not reported here) attests that in several occasions the average profile of the network covers a higher place than the external dry-bulb temperature. It is time dependent and it seems to have a



Table 3: Modelling parameters utilised in Meyrin case study

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 partitions	$W/Km^2$	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	$m^3/s$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$	$10^{-5}$
Heating set point	$^{\circ}C$	20	20	20	20	20	20	20	20

Table 4: Quality of the simulation outcomes achieved during the training period for different seasons

Case study	Average CVRMSE%	Average NMBE [-]	Season	CVRMSE%	NMBE [-]
Sub1	32.93	-0.02	Win	13.54	0.027
			Spr/Aut	37.15	0.077
			Sum	48.12	-0.044
Sub2	19.93	0.0073	Win	10.09	0.022
			Spr/Aut	27.4	0.039
			Sum	22.31	-0.039
Sub3	22.23	0.023	Win	16.1	0.033
			Spr/Aut	34.9	0.028
			Sum	15.7	0.01

Table 5: Quality measure of a full Data-Driven approach for a substation case study.

	Seasons	MAPE [-]	RMSE [kW]	CVRMSE %	$R^2$ [-]	NMBE [-]
nnet	Winter	21.28	20.66	15.03	0.70	0.030
svm		<b>19.70</b>	<b>18.69</b>	<b>13.59</b>	<b>0.76</b>	<b>0.010</b>
glm		21.24	20.36	14.81	0.71	0.020
rnf		20.05	19.87	14.45	0.72	0.040
nnet1	Spr_Aut	29.90	19.48	24.86	0.75	0.120
svm1		29.87	16.52	21.09	0.82	0.010
glm1		41.81	20.02	25.55	0.73	0.060
rnf1		<b>27.19</b>	<b>17.17</b>	<b>21.92</b>	<b>0.80</b>	<b>0.090</b>
nnet2	Summer	102.63	8.76	31.29	0.67	0.030
svm2		<b>84.83</b>	<b>8.74</b>	<b>31.22</b>	<b>0.67</b>	<b>0.050</b>
glm2		255.33	9.42	33.65	0.62	0.010
rnf2		238.66	8.96	32.00	0.65	0.070

stronger influence during winter and spring-autumn seasons. Five accuracy measures have been evaluated for each model: an example is reported in Table 5.

The predictive models perform better during winter time and for short time horizons (not included in the table). Non linear models are the ones with the best predictive capabilities. The CVRMSE and NMBE are, for the majority of the time, below the ASHRAE guidelines ( $CVRMSE < 30\%$  and  $NMBE < 0.1$ ).

These results will be used as baseline for examine the effect of introducing simulated demand as predictive variable to the forecasting framework.

### Combined approach

The analysis of predictor importance using the wrapper method is repeated a second time to estimate the hierarchy of the predictors, when simulated demand is added to the

set of independent variables. Results are visualised for the next hour time frame horizon, as shown in Figure 8.

As indicated in Section 2, the importance of the variables for the prediction of building demand was evaluated using the LASSO method. The outcomes of the algorithm were further processed for practical visualisation purposes using a max-min normalisation method. This result is indicated with the variable ‘‘importance’’ as per Figures 8 and 9 and it enables the visualisation of the predictive hierarchy of the variables in a 0-1 range.

The simulated demand in the case presented here (but also in most of the cases investigated) ranks very high in the list of predictors. In most of the cases, it covers a role higher than the external temperature underlining the capacity of the simulation method to act as informative value added data service for the data-driven approach.

Percentages of improvement with respect to the baseline predictive model are shown in Figure 10 and 11. With reference to the figures, the results show that in the majority of the three case studies and in most of the time horizons considered in the analysis there is an improvement of the predictive capability of the models when the simulated demand data is added to the forecast model compared to the baseline predictive algorithm. These results are not always valid for all the predictive methods. Generally, the predictive models which achieve best performance in the baseline approach are positively affected by the inclusion of the simulated demand. Models which do not perform well in the predictive phase may not be affected by the inclusion of simulated demand. Interestingly, overall, the final prediction capability achieved selecting the best performing methods is in many cases improved. Percentage of improvement can reach up to +10% in terms of comparison with CVRMSE and  $R^2$  of the baseline model.

### Conclusion

The main objective of the current paper was to show how building simulation can have an important role in fostering the prediction capabilities of a forecasting method. This was tested by predicting building blocks heating demand in a district heating network. In this study, simulation methods act as a value added data service for predictive modelling. Results show that the average district heating

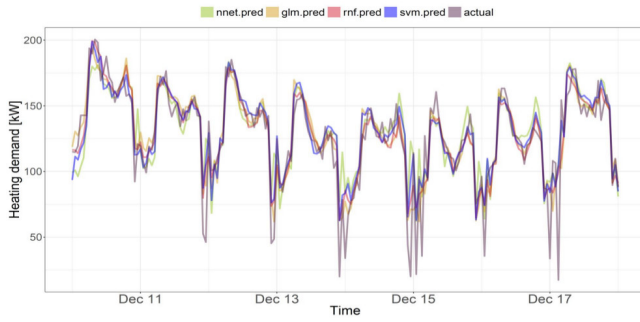


Figure 6: Forecasted winter heating demand for one block of buildings over one week.

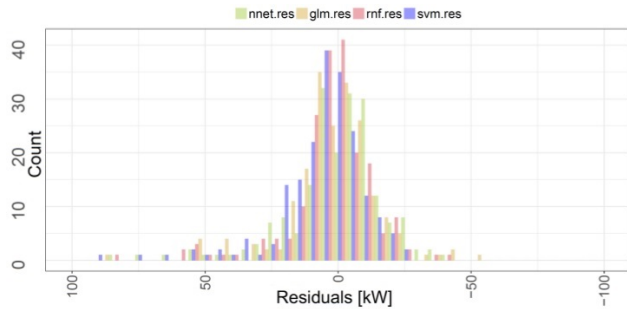


Figure 7: Residuals of the forecasting models.

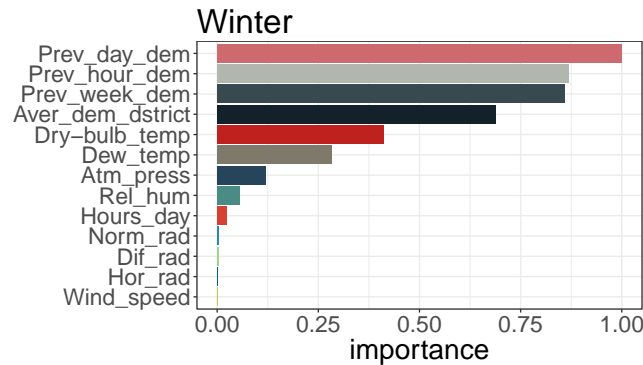


Figure 8: Full data-driven model: predictors importance output of the LASSO method

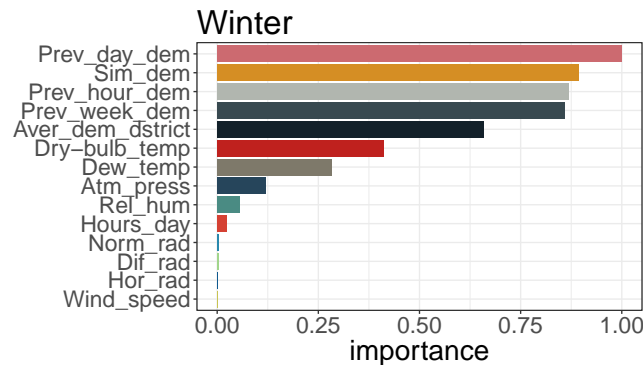
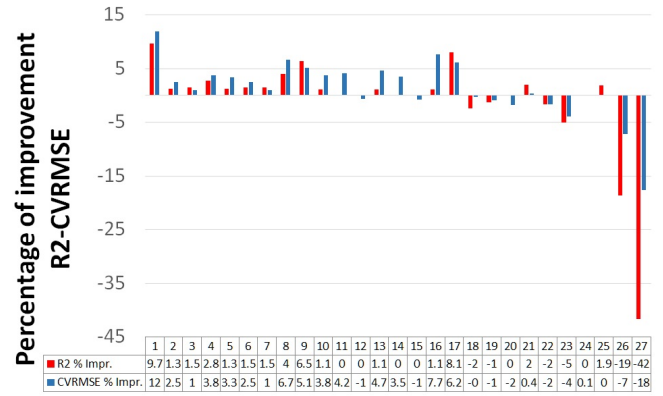
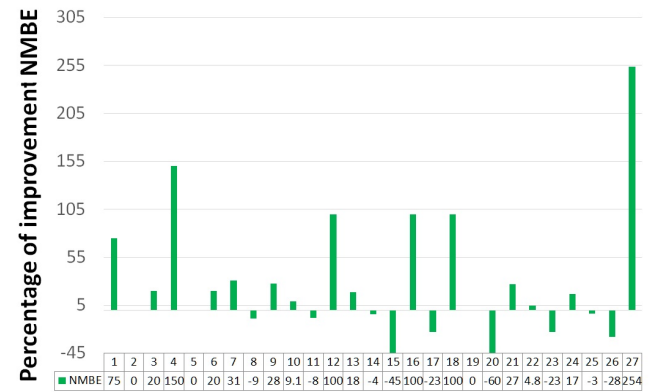


Figure 9: Simulated demand added to the predictive framework: predictor importance output of the LASSO method



### Case study

Figure 10: Percentage of improvement for the CVRMSE[%] and R2 compared to the baseline model



### Case study

Figure 11: Percentage of improvement for the NMBE compared to the baseline model

profile covers an important role as predictor when included in a traditional forecasting framework.

The predictive methods augmented by the inclusion of simulated demand achieve substantial improvements in some cases. Results show that the quality of the simulated demand is an important aspect to foster the quality of the predictions. Nevertheless, although the results of the proposed approach are promising, there are a series of limitations which should be further investigated in future studies. A more robust calibration approach should be carried out to test how the importance of the simulated demand changes at different stages of model calibration. Higher quality of the simulated demand could further improve the prediction capability of the data-driven model. A mapping procedure could be considered to test how building information, elaborated by the energy model, influences the results of simulations and subsequently the predictive capability of the simulated demand. This may advice the energy modellers on the best trade-off between predictive accuracy and time required for the creation of calibrated building energy models for heating demand forecasting.



## Acknowledgement

## References

- Ahmad, T. and H. Chen (2018). Energy & Buildings Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy & Buildings* 166, 460–476.
- Al-shammari, E. T., A. Keivani, and S. Shamshirband (2016). Prediction of heat load in district heating systems by Support Vector Machine with Fire fly searching algorithm. *Energy* 95, 266–273.
- ASHRAE, A. G. (2014). Guideline 14-2014, measurement of energy, demand and water savings, ashrae. *Atlanta*.
- Chou, J.-S. and D.-K. Bui (2014, oct). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings* 82(August), 437–446.
- Dahl, M., A. Brun, and G. B. Andresen (2017). Using ensemble weather predictions in district heating operation and load forecasting. *Applied Energy* 193, 455–465.
- Eicker, U., J. Schumacher, M. Zirak, N. Bartke, R. Nouvel, and V. Coors (Hyderabad, India, Dec.7-9, 2015). An automated method for urban energy simulation based on 3d city models. In *Proceedings of BS2015:14th Conference of International Building Performance Simulation Association*.
- Fan, C., F. Xiao, and Y. Zhao (2017). A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy* 195, 222–233.
- Fonti, V. and E. Belitser (2017). Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*.
- Geysen, D., O. D. Somer, C. Johansson, J. Brage, and D. Vanhoudt (2018). Operational thermal load forecasting in district heating networks using machine learning and expert advice. *Energy & Buildings* 162, 144–153.
- Gunay, B., W. Shen, and G. Newsham (2017, may). Inverse blackbox modeling of the heating and cooling load in office buildings. *Energy and Buildings* 142, 200–210.
- Idowu, S., S. Saguna, and C. Ahlund (2016). Applied machine learning : Forecasting heat load in district heating system. *Energy and Buildings* 133, 478–488.
- Kapetanakis, D. S., E. Mangina, and D. P. Finn (2017). Input variable selection for thermal load predictive models of commercial buildings. *Energy and Buildings* 137, 13–26.
- Kazas, G., E. Fabrizio, and M. Perino (2017). 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), 243 – 262.
- Kylili, A. and P. A. Fokaides (2015). European smart cities : The role of zero energy buildings. *Sustainable Cities and Society* 15, 86–95.
- Li, X. and J. Wen (2014). Review of building energy modeling for control and operation. *Renewable and Sustainable Energy Reviews* 37, 517–537.
- Lund, H., S. Werner, R. Wiltshire, S. Svendsen, J. Eric, F. Hvelplund, and B. Vad (2014). 4th Generation District Heating ( 4GDH ) Integrating smart thermal grids into future sustainable energy systems. *Energy* 68, 1–11.
- Macas, M., F. Moretti, A. Fonti, A. Giantomassi, G. Comodi, M. Annunziato, S. Pizzuti, and A. Capra (2016, jan). The role of data sample size and dimensionality in neural network based forecasting of building heating related variables. *Energy and Buildings* 111, 299–310.
- Marquant, J. F., L. A. Bollinger, R. Evins, and J. Carmeliet (2018). A new combined clustering method to Analyse the potential of district heating networks at large-scale. *Energy* 156, 73–83.
- Molina-solana, M., M. Ros, M. D. Ruiz, and J. Gómez-romero (2017). Data science for building energy management : A review. *Renewable and Sustainable Energy Reviews* 70(August 2015), 598–609.
- Naji, S., A. Keivani, S. Shamshirband, U. J. Alengaram, M. Z. Jumaat, Z. Mansor, and M. Lee (2016, feb). Estimating building energy consumption using extreme learning machine method. *Energy* 97, 506–516.
- Noussan, M., M. Jarre, and A. Poggio (2017). Real operation data analysis on district heating load patterns. *Energy* 129, 70–78.
- Paudel, S., M. Elmitri, S. Couturier, P. H. Nguyen, R. Kamphuis, B. Lacarrière, and O. Le Corre (2017). A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. *Energy and Buildings* 138, 240–256.
- Powell, K. M., A. Sriprasad, W. J. Cole, and T. F. Edgar (2014). Heating, cooling, and electrical load forecasting for a large-scale district energy system. *Energy* 74, 877–885.
- Prieto, M. P., P. M. A. de Uribarri, and G. Tardioli (2019). Applying modeling and optimization tools to existing city quarters. In U. Eicker (Ed), *Urban Energy Systems for Low-Carbon Cities*, pp. 333 – 414. Academic Press.
- Rahman, A. and A. D. Smith (2017). Predicting fuel consumption for commercial buildings with machine learning algorithms. *Energy & Buildings* 152, 341–358.
- Rahman, A. and A. D. Smith (2018). Predicting heating demand and sizing a stratified thermal storage tank using deep learning algorithms. *Applied Energy* 228(March), 108–121.
- Sharif, M. K. A., A. A. Al-abidi, S. Mat, K. Sopian, and M. H. Ruslan (2015). Review of the application of phase change material for heating and domestic hot water systems. *Renewable and Sustainable Energy Reviews* 42, 557–568.
- Sholahudin, S. and H. Han (2016). Simplified dynamic neural network model to predict heating load of a building using Taguchi method. *Energy* 115, 1672–1678.
- Short, M., T. Crosbie, M. Dawood, and N. Dawood (2017). Load forecasting and dispatch optimisation for decentralised co-generation plant with dual energy storage. *Applied Energy* 186, 304–320.
- Song, K., N. Kwon, K. Anderson, M. Park, H. S. Lee, and S. H. Lee (2017, dec). Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups. *Energy and Buildings* 156, 121–133.
- Suryanarayana, G., J. Lago, D. Geysen, P. Aleksiejuk, and C. Johansson (2018). Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods. *Energy* 157, 141–149.
- Wei, Y., X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, and X. Zhao (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews* 82(October), 1027–1047.
- Yuan, T., N. Zhu, Y. Shi, C. Chang, K. Yang, and Y. Ding (2018). Sample data selection method for improving the prediction accuracy of the heating energy consumption. *Energy and Buildings* 158.
- Zhao, H.-X. and F. Magoulès (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews* 16, 3586–3592.