Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 79–85
www.elsevier.com/locate/egyr

7th International Conference on Advances on Clean Energy Research, ICACER 2022 April 20–22, 2022, Barcelona, Spain

Simplified model for the short-term forecasting of heat loads in buildings

Markel Eguizabal[a,b], Roberto Garay-Martinez[c,*], Iván Flores-Abascal[b]

[a] Geuria Ingenieria & PRL, Nafarroa Kalea 19, Trapagaran, 48510, Spain
[b] ENEDI Research Group, Department of Energy Engineering, Faculty of Engineering of Bilbao, University of the Basque Country UPV/EHU, Pza, Ingeniero Torres Quevedo 1, Bilbao, 48013, Spain
[c] Institute of Technology, Faculty of Engineering, University of Deusto, Av. Universidades, 24, 48007 Bilbao, Spain

Received 4 October 2022; accepted 9 October 2022
Available online xxxx

Abstract

A data-driven model is used to predict one-hour ahead heat loads based on present and recent history of weather and heat loads. A computationally inexpensive method is built to deliver load forecasting based on existing data quality and resolution from smart meters. Optimal model formulation is discussed and optimized at 4-hour historical values. The model is trained and tested against synthetic data from a building energy simulation, resulting in absolute error <4% and R^2 values in the range of 0.92 to 0.94.

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances on Clean Energy Research, ICACER, 2022.

Keywords: Regression model; Heating demand; Building; Lagged value

1. Introduction

There is an increased research area dealing with building load assessment and forecasting based on smart meter data. This data is used for many applications such as load profile identification and clustering (Johra et al. [1], Gianniou et al. [2]), energy signature assessment [3,4], submetering & HVAC system identification [5,6] and control applications (Bergsteinsson et al. [7]).

With increased reliability, and interconnection, datasets from smart meters, together with meteorological data can boost our understanding of building load dynamics. One key application is the delivery of short-term data-driven heat load-prediction methods to Building Management Systems (BMS) to improve efficiency of building operation. These systems require accurate but computationally inexpensive heat load forecasts of buildings.

Improvements to BMSs are known to produce very relevant energy savings in the range of 7% to 52% (Fong et al. [8] and Peng et al. [9]). This is highly dependent on the type of building, occupancy patterns, new and benchmark control models.

* Corresponding author.
E-mail address: roberto.garay@deusto.es (R. Garay-Martinez).

https://doi.org/10.1016/j.egyr.2022.10.224

Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances on Clean Energy Research, ICACER, 2022.

There have been a number of research applications to integrate forecasting into BMSs. Some approaches have pursued the integration of engineering simulation tools (also known as White-box models) such as EnergyPlus. These deliver detailed simulations of buildings, but require of extensive modelling work and are still computationally expensive (Wei et al. [10]). Other alternatives such as semi empirical (also known as Grey-box models) are possible, where a reduced number of physical equations are calibrated against data. Although there are many examples of these systems such as (Colombo Paola et al. [11]), there is still a relevant way ahead to standardize model definition and calibration procedures. Lastly, purely mathematical modelling is also possible, where no physical relationship is sought, and there is no need of previous knowledge from the building is needed.

This last group is typically labelled as "black box", and comprise a large set of potential models, comprising alternatives such as statistical modelling (Caldera et al. [12]), regression and/or Support Vector Machine (SVM), and artificial neural network (Melo et al. [13]).

Although very powerful, these models need to be applied with care. Considering the suitability of the selected formulation to each particular application and ensuring the accuracy of the delivered forecasts. Considering the opaqueness of the methods, an unbiased short-and long-term calibration and operation must be ensured. Once this is verified, these models are potentially applicable over the vast majority of buildings and implemented in relatively simple building control systems.

There is however a very relevant risk of overfitting and uncertainty associated with the use of complex, black box models. Grey-box models such as Bacher and Madsen [14] have successfully calibrated models with 2 dominant time constants. Considering this, a relatively simple model should be able to perform short-term predictions of the heat load of buildings, given relatively stable/predictable occupancy patterns. But this concept is yet to be tested, by selecting the model shape, parameter identification criteria, and range of validity of the model.

In this paper, a simple data-driven model is proposed. It is formulated as an Autoregressive with Exogenous input (ARX) model. The model provides a one-hour ahead prediction of the heating load based on external information about the building (ambient temperature, solar irradiation and heat load). These variables are known to be key explanatory variables for similar models (Dong et al. [15]). Present and lagged values of variables are used. It is proven that a relatively small number of historical values of the variables substantially improves the quality of the model. The model is trained and tested against synthetic data from a building energy simulation.

In Section 2, the methodology that have been followed to obtain estimated heating demand is explained. In Section 3, a brief description of the building is done, followed by obtained results in Section 4. Finally, in Section 5 conclusions are shown, as well as the importance that this work could have for future research works.

2. Methodology

A five-step methodology is defined. Synthetic datasets are created (1), model structure is defined (2) and calibrated (3). Then the model is used for load estimation (4) and the quality of estimations are evaluated (5). This process is illustrated in Fig. 1.

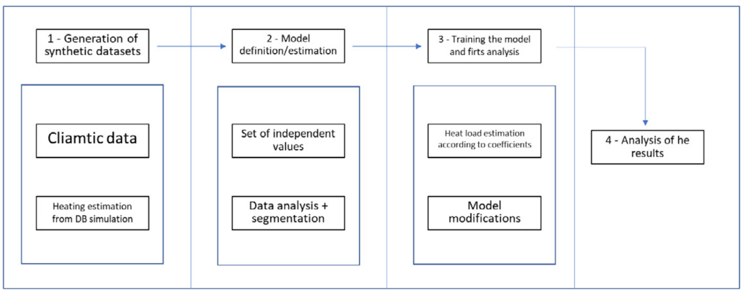

Fig. 1. Stepwise definition of the methodology.

2.1. Generation of synthetic datasets

The dataset is composed by weather (outdoor temperature, solar irradiance) and heat load data with hourly resolution. ASHRAE IWEC weather data for the cities of Bilbao and Madrid has been used. Heat loads have been calculated in EnergyPlus v8.9. Specific information on the simulated building can be found in Section 3.

2.2. Model definition

The model is expected to predict the actual heating load of the building based on climate data and previous heat loads. To do so, present and lagged values are used as input values. The optimal number of lagged values suitable for the model is not known. Up to 12 such lagged values are considered, but this figure is optimized in the process.

Lagged values are identified by numerals from 0 (actual value) to 12 (value referred to 12th previous hour). For heat loads values from 1 to 12 are used as the actual value is meant to be forecasted by the model. A common naming for the model is defined as ARX_TX_IY_QZ, where X, Y and Z are replaced by the number of lagged values considered in the model for Outdoor Temperature (X), Solar Irradiation (Y) and Heat Load (Z). The subscript "n" is used if a specific variable is not used. As an example, (1) presents the formulation associated with model ARX_T2_I2_Q1.

$$Q = C_0 + C_{T_o} * T_0 + C_{T_1} * T_1 + C_{I_0} * I_0 + C_{I_1} * I_1 * C_{Q_1} * Q_1 \tag{1}$$

2.3. Model calibration

The calibration is performed through the regression function in R. The model is focused on heating loads, so data from periods with existing heating load (October to March) are used. This is done so to avoid excessive relevance of periods with low or inexistent heat load in the regression process. The model is complemented with a low-level filter, where negative values are corrected to 0.

2.4. Results assessment & validation process

For each model, visual inspection has been complemented with R^2, p-value and absolute error have been calculated. The mean absolute error is used as optimization metric to define the optimum number of lagged values.

Two climates are used so that the climate for Bilbao is used for model calibration while the climate for Madrid is used for validation.

3. Building description

The modelled building corresponds with a real multi-rise building with 9 floors (8 + commercial ground floor), and 4 dwellings per floor. It exceeds the actual building regulation in Spain, by 38% (heating) and 50% (cooling). Table 1 provides a brief description of the building, and Table 2 delivers a summary of insulation levels.

Table 1. Description of building's main characteristics.

Number of dwellings	32
Built Surface	2898 m^2
Number of floors	8 + commercial ground floor
Volume to envelope ratio	4.1 m^3/m^2
Heating	Air source heat pumps
Façade openings (%)	30–35

Table 2. Summary of U-values.

Element	U-value [W/m^2 K]
Walls and floor in contact with outside air	0.263
Roofs in contact with outside air (inclined)	0.596
Roofs in contact with outside air (plane)	0.260
Windows	1.91
Common zone partition	0.522

Building usage is defined through daily profiles in line with the provisions of the building code.

The building has been simulated for the climates of Madrid and Bilbao. These are representative of different climate zones Csa/D and Cfb/C according to the Köpen-Geiger climate classification and Spanish technical code respectively.

4. Results and discussion

Figs. 2 and 3 present the outcomes of the optimal model against the datasets for Bilbao and Madrid. In both cases, it corresponds with the ARX_T0_In_Q1 model. As expected, maximum heating demand happens in Madrid (Fig. 2, right), because of its colder weather conditions. The accuracy of the model can be considered as acceptable for both locations, and to quantify it, absolute error and R^2 have been calculated.

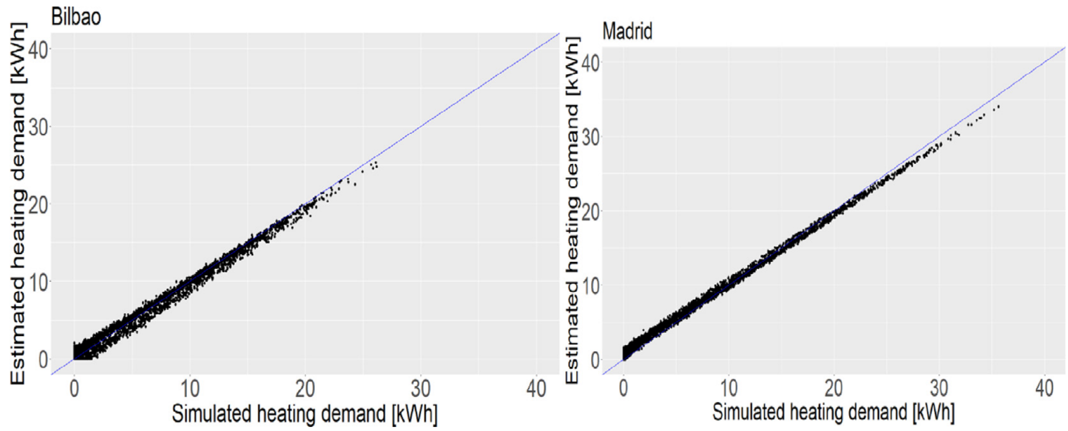

Fig. 2. Comparison between estimated and simulated heating demands for Bilbao (left) and Madrid (right).

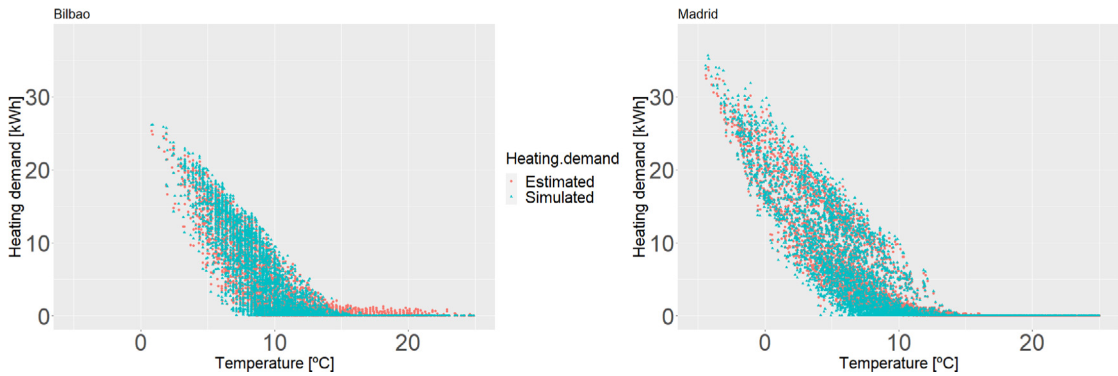

Fig. 3. Estimated and simulated heating demands vs. outdoor temperature in Bilbao (left) and Madrid (right).

Considering that there are too many data points in Fig. 3, subsets of some months (January and October) are presented in Fig. 4.

As identified from visual inspection, the model can estimate the heat loads with high accuracy. Some deviations in the range of 1–2 kW (<5% of full load) are observed for extreme cases. With high loads (i.e. >80% of full load), estimations are lower than the actual load, while the opposite happens with low loads (i.e. <30% of full load).

Fig. 4 (right) shows the warmest month in which heating system is working (October) for Madrid. As expected, maximum heat load is much lower than January's maximum value, and there is a remarkable amount of null heating loads. The error made month by month will be analysed further.

Regarding the determination of the optimal number of coefficients, only one lagged value is taken for the heat load of the building. While the inclusion of this coefficient showed a great improve in the performance of the models, further inclusion of lagged values did not show any relevant improvement.

The determination of the number of lagged values to be considered for Temperature and Irradiance signals has been performed based on the assessment of the resulting model absolute error. For the determination of absolute error, all the models of the same inclusion depth (i.e. all the models with 5 lagged Temperature signals) are considered, their absolute error is calculated, and then an average of these is performed.

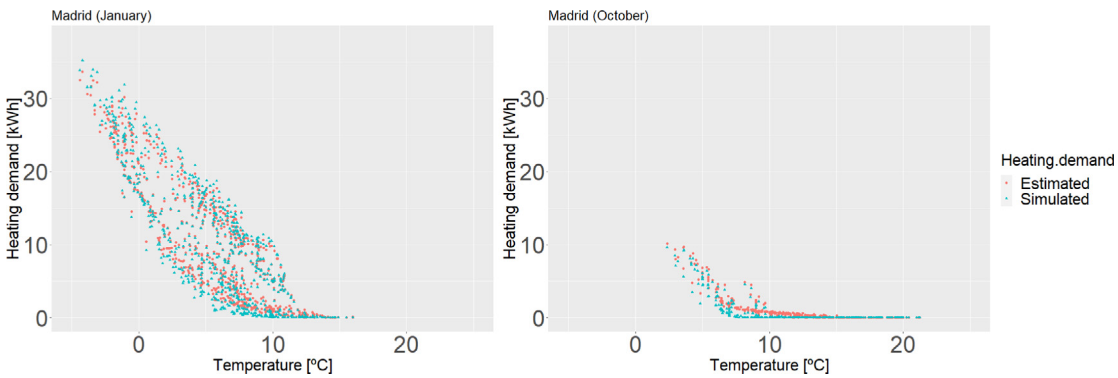

Fig. 4. Estimated and simulated heating demands vs outdoor temperature in January and October (Madrid).

Table 3. Average absolute error (kWh) depending on the inclusion depth on lagged values of Temperature and Irradiance.

Lag	0	1	2	3	4	5	6	7	8	9	10	11	12
Temperature	0,45	0,57	0,57	0,57	0,58	0,59	0,62	0,64	0,65	0,65	0,65	0,65	0,65
Irradiation	0,56	0,58	0,58	0,58	0,58	0,59	0,60	0,60	0,60	0,61	0,61	0,61	0,61

Table 3 presents average absolute error of models with varied inclusion depths lagged values of temperature and solar irradiation signals. For both signals, the error increases slightly with every addition of lagged values. Although using only present information (i.e. T_0 or I_0) minimizes the error, this configuration is potentially too exposed to the impact of weather forecasting errors. Thus, a regression model which consider the short-term history of the local weather conditions is preferred, with an ARX_TX_IY_Q1 configuration. As per the data in Table 3, lagged values in the 0–4 ranges are proposed both for Temperature and Solar Irradiation. In Fig. 5, monthly model outputs are summarized against the original simulation data. Estimation made by models with minimum (T_0, I_0) and maximum (T_4, I_4) lagged values are shown. It can be seen that both models present similar output quality.

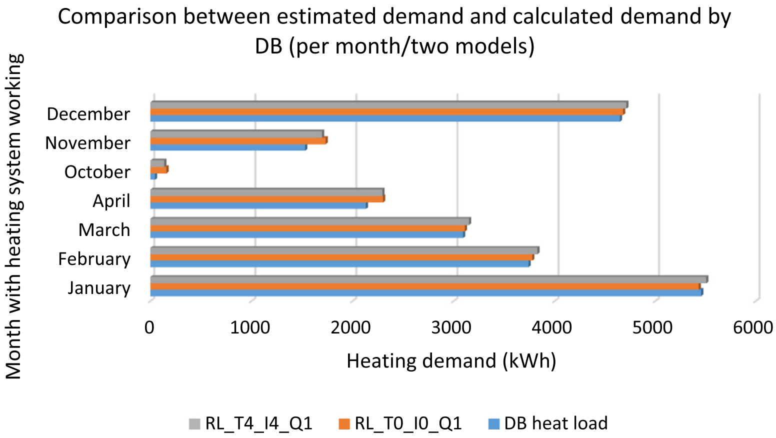

Fig. 5. Comparison between different models per month.

The model quality is assessed with R^2 and p values, as well as model error. p-values show that the model is in good agreement with the data, with values well below 0.05 in all cases. R^2 values are in the range of 0.92 to 0.94. Model error is in the range of 2.5% to 4%. In Figs. 6 and 7, the evolution of R^2 and model error is shown.

As a result, it has been mathematically verified that the proposed models can estimate properly the heat loads of the building. Moreover, as shown in Fig. 5, the model's accuracy improves for high heat loads.

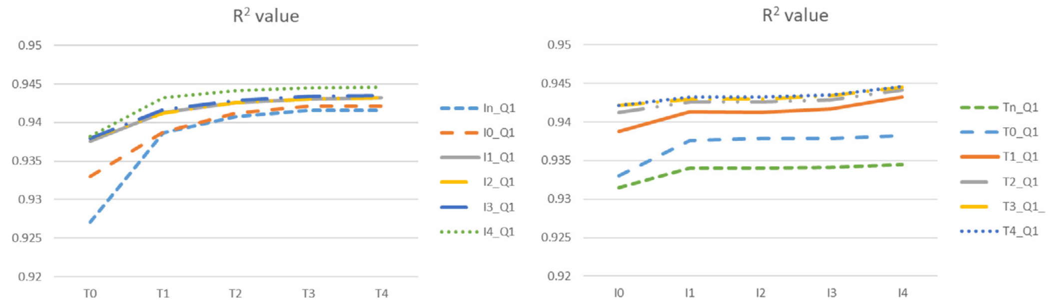

Fig. 6. R2 value depending on T (left) and I (right) lagged values.

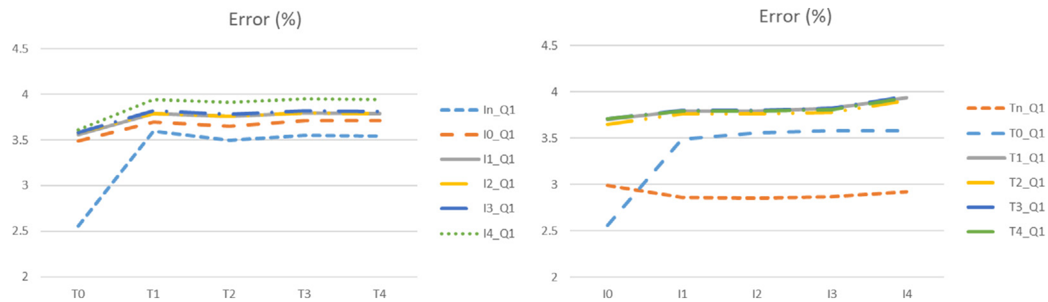

Fig. 7. Absolute error depending on T (left) and I (right) lagged values.

Considering the large number of considered model variants (35), detailed model information is omitted from this section. The coefficients of the most relevant configurations are presented in Table 4.

Table 4. Coefficients of ARX model.

Model	C_0	C_{T0}	C_{T1}	C_{T2}	C_{T3}	C_{T4}	C_{I0}	C_{I1}	C_{I2}	C_{I3}	C_{I4}	C_{Q1}
ARX_T1_I0_Q1	2,3241	−0,1641	–	–	–	–	–	–	–	–	–	0,8730
ARX_T4_I0_Q1	2,1470	−1,2899	1,3501	0,3227	−0,6801	0,1377	–	–	–	–	–	0,9067
ARX_T1_I1_Q1	1,6335	−0,0860	–	–	–	–	−2,2504	–	–	–	–	0,8996
ARX_T4_I1_Q1	2,1292	−1,1979	1,2935	0,3168	−0,6804	0,1199	−0,8630	–	–	–	–	0,9063
ARX_T1_I4_Q1	2,2252	−0,1466	–	–	–	–	−5,9987	3,4415	0,3067	2,3868	−1,5378	0,8946
ARX_T4_I4_Q1	1,9586	−1,1356	1,1388	0,2987	−0,5537	0,1181	−3,2692	2,7900	0,1906	2,3589	−3,0852	0,9137

5. Conclusions and future work

This paper presents an ARX model which performs a short-term prediction of heat loads in buildings. A model accuracy better than 4% is achieved over heating loads during winter periods.

The presented analysis assessed the accuracy of various model formulations with present and recent past observations of heat load, temperature, and solar irradiation. It is identified that the use of recent history (up to 4-h historical data) results in an accurate and stable forecasting tool. The model has been calibrated and validated with simulation data from a multi-rise building for two cities.

The model construction does not require any a-priori knowledge on the building (i.e. size, geometry…), so it can be used to model any heat load independently such as heat load data coming from smart meters. At the same time, the presented method does not require large computational efforts, allowing for its deployment over large sets of buildings at district and urban levels.

The present method has been developed using simulation data, with stable indoor climate and user patterns. Accordingly, to achieve the aforementioned applications, further research is needed to integrate the variability present in real-life data. This implies that further refinement will be required in to integrate thermostatic daily and weekly cycles, variable occupant behaviour, and testing against data from real buildings. At the same time, the method is focused in winter performance. This implies that methods for the segregation of winter and summer seasons are needed, along with suitable models for the forecasting of loads in summer periods.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] Johra H, Leiria D, Marszal-Pomianowska A, Tvedebrink T. Treatment and analysis of smart energy meter data from a cluster of buildings connected to district heating: A danish case. E3S Web Conf 2020;172(12004).
[2] Gianniou P, Liu X, Heller A, Nielsen PS, Rode C. Clustering-based analysis for residential district heating data. Energy Convers Manage 2018;165:840–50.
[3] Eriksson M, Akander J, Moshfegh B. Development and validation of energy signature method – case study on a multi-family building in sweden before and after deep renovation. Energy Build 2020;210:109756.
[4] Lumbreras M, Garay-Martinez R, Arregi B, Martin-Escudero K, Diarce G, Raud M, Hagu I. Data driven model for heat load prediction in buildings connected to district heating by using smart heat meters. Energy (Oxford) 2022;239:122318.
[5] Bacher P, de Saint-Aubain PA, Christiansen LE, Madsen H. Non-parametric method for separating domestic hot water heating spikes and space heating. Energy Build 2016;130:107–12.
[6] Westermann P, Deb C, Schlueter A, Evins R. Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. Appl Energy 2020;264:114715.
[7] Bergsteinsson HG, Nielsen TS, Møller JK, Amer SB, Dominković DF, Madsen H. Use of smart meters as feedback for district heating temperature control. Energy Rep 2021;7:213–21.
[8] Fong KF, Hanby VI, Chow TT. HVAC system optimization for energy management by evolutionary programming. Energy Build 2006;38(3):220–31.
[9] Peng Y, Rysanek A, Nagy Z, Schlüter A. Using machine learning techniques for occupancy-prediction-based cooling control in office buildings. Appl Energy 2018;(211):1343–58.
[10] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, Zhao X. A review of data-driven approaches for prediction and classification of building energy consumption. Renew Sustain Energy Rev 2018;82(1):1027–47.
[11] Paola Colombo, Rossano Scoccia, Marcello Aprile, Mario Motta, Livio Mazzarella. Minimalist RC network for building energy simulations: A case study based on openbps. E3S Web Conf 2020;197(02005).
[12] Caldera M, Corgnati SP, Filippi M. Energy demand for space heating through a statistical approach: Application to residential buildings. Energy Build 2008;40(10):1972–83.
[13] Melo AP, Versage RS, Sawaya G, Lamberts R. A novel surrogate model to support building energy labelling system: A new approach to assess cooling energy demand in commercial buildings. Energy Build 2016;131:233–47.
[14] Bacher P, Madsen H. Identifying suitable models for the heat dynamics of buildings. Energy Build 2011;43(7):1511–22.
[15] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build 2005;37(5):545–53.