Title

Incorporating machine learning with building network analysis to predict multi-building energy use

Authors

Xu, Xiaodong
Wang, Wei
Hong, Tianzhen
et al.

Publication Date

2019-03-01

DOI

10.1016/j.enbuild.2019.01.002

Peer reviewed

Incorporating machine learning with social network analysis to predict multi-building energy use

Xiaodong Xu [a,*], Wei Wang [b, c], Tianzhen Hong [c, *], Jiayu Chen [b]

[a] School of Architecture, Southeast University, 2 Sipailou, Nanjing, Jiangsu Province，China
[b] Department of Architecture and Civil Engineering, City University of Hong Kong, Y6621, AC1, Tat Chee Ave, Kowloon, Hong Kong

[c] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

Abstract

Predicting energy use of campuses or city district buildings has recently gained more attention due to dynamic large-scale building energy demands. This data enlightens public's awareness of energy use and informs building energy policy. Understanding the correlation of energy use patterns between buildings is a key issue to analyzing multi-building energy use. Moreover, how to apply this inter-building relationship to multi-building energy prediction, using significantly less amount of building energy data, is still an open question. To solve such problems, this study proposed an interdisciplinary research method to predict multi-building energy use by integrating a social network (SN) analysis with an Artificial Neural Network (ANN) technique. The SN method was used to identify reference buildings and determine correlations between reference buildings and non-reference buildings. The ANN technique was applied to learn correlations and historical building energy use, and then used to predict multi-building energy use. To validate the SN-ANN method, 17 buildings in the Southeast

* Corresponding author. tel.: +86 (25) 8379-5689; xxdseu@126.com (Xiaodong Xu)

* Corresponding author. tel.: +1 (510) 486-7082; thong@lbl.gov (Tianzhen Hong)

University campus, located in Nanjing, China, were studied. These buildings have three years of actual monthly electricity use data, and were grouped into four types: office, educational, laboratory, and residential. The results showed the integrated SN-ANN method achieved an accuracy of 90.28% for the predicted energy use for all building groups. Finally, this study provides insights into advancing the interdisciplinary research on multi-building energy use prediction.

Keywords:

Multi-building; energy use prediction; social network analysis; artificial neural networks; machine learning

1. Introduction

Buildings are a main energy consumer, demanding more than 40% of primary energy usage [1]; while in cities, buildings can consume up to 75% of total primary energy usage [2]. In particular, electricity use is a main driver. The latest Electric Power Monthly data reported in January 2018 by United States Department of Energy (DOE) indicated that electricity consumption from both commercial and residential buildings represented 77.5% of all the electricity produced in the U.S. [3]. The International Energy Agency (IEA)'s Energy in Buildings and Communities (EBC) Programme annexes discussed methods to analyze total energy use in buildings to reduce energy use and associated emissions [4,5]. The use of building energy modeling has significantly improved building energy efficiency and reduced environmental impact [6,7]. A considerable number of studies have been conducted to develop efficient energy models for single buildings [8–10]. In recent years, some researchers have recognized the importance of energy use studies in large-scale areas with distributed building groups to analyze distributed building energy use patterns and optimize net-zero building or distribution energy systems [11,12], also, for city-scale buildings through benchmarking building energy use and reducing city building emissions [13,14]. Focus on analyzing and modeling urban building energy use at the large scale can potentially provide insights into large-scale building energy use patterns and opportunities to save energy [15,16]. Li et al. analyzed 51 high-performance office buildings in the U.S., Europe, and Asia using portfolio analysis and individual detailed case studies based on actual energy use data of buildings [17]. Pang et al. brought together real-time data sharing, a database for assessing past and present weather data, a network for communicating energy-saving strategies between building owners, and a set of modeling tools for real-time building energy simulation, all in an effort to promote large-scale energy efficiency in neighboring buildings [18]. Fonseca and Schlueter proposed one integrated model for the characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. The model calculated the power and temperature requirements for residential, commercial, and industrial sectors using spatial (building location using geographic information system, GIS) and temporal (hourly) dimensions of analysis [19]. To predict energy use of a large group of buildings, Panao and Brito presented a bottom-up building stock energy model [20]. They predicted hourly electricity consumption of residential buildings and validated the model by using smart meter data of roughly 250 dwellings [20]. Kalogirou et al. utilized the electricity data of 225 buildings and applied back propagation of neural networks to predict the required heating load of buildings [21]. Constantine used data-driven prediction models, including linear regression, random forest, and support vector regression, to predict city-scale electricity and natural gas usage in New York City buildings [22]. The project encompassed 23,000 buildings, with model validation at the building and ZIP code levels [22]. Similarly, Hsu studied multi-family buildings in New York City and used clusterwise regression and cluster validation methods to determine building energy use [23]. Jain et al. applied a sensor-based forecasting approach coupled with support vector regression modeling and examined the impact of temporal and spatial granularity on to energy consumption of multi-family buildings [24].

In modeling large-scale building energy use, more researchers have started to realize the impact and interrelationship between building groups. The concept of the Inter-Building Effect (IBE) was introduced to understand the complex mutual impacts within spatially proximal buildings [25–27]. Han et al. explored mutual shading and reflection for IBE on building energy performance with two realistic urban contexts in Perugia, Italy [28]. Han and Taylor further simulated the IBE on energy consumption by embedding phase change materials into the building envelopes [29].

Innovative software or web-based applications have been developed to analyze and predict energy use of multiple buildings in distributed or urban areas. For example, the City Building Energy Saver (CityBES), an Energyplus-based web application, provides a visualization platform, focusing on energy modeling and analysis of a city's building stock to support district or city-scale building energy efficiency programs [30–32], as well as to predict energy use for informing building retrofits. Based on CityBES, Chen et al. analyzed the impacts of building geometry modeling on urban building energy models to understand how a group of buildings perform together [33]. City Energy Analyst (CEA) provides a computational framework for the analysis and optimization of energy systems in neighborhoods and city districts. CEA has a unique interface to facilitate the spatiotemporal analysis of energy patterns for energy savings [34]. Usually, with these software or web-based applications every building modeled is explicitly done in EnergyPlus-based detail. While it can be accurate, it is time consuming and requires absorbent amounts of data.

To reduce the complexity of urban building energy models, some studies advocate reduce-order building models or building prototype models. Felsmann used reduced order building energy system modeling, e.g., district heating or cooling systems, to create large-scale urban energy simulations [35]. Heidarinejad et al. developed a framework to rapidly create urban scale reduced-order building energy models relying on the contributions of different influential variables to the internal, external, and system thermal loads [36]. Then the framework was validated by applying typical building geometries for simulations [36]. Zhao et al. developed a reduced order building energy model to estimate single building energy performance; then applied regression and Markov chain Monte Carlo techniques to integrate physics-based energy modeling to replicate the single building model [37]. The resulting model was an efficient energy model development at the city scale [37].

One method to reduce data demands includes the development and replication of prototype building models. The U.S. DOE has developed a suite of prototype building models covering 80% of the commercial building stock in the U.S. to support the analysis of urban energy use. This database include 16 commercial reference building types across different climate zones [38, 39]. Similarly, Mastrucci et al. analyzed six types of dwellings by using a GIS-based statistical downscaling approach and adopted a multiple linear regression model for estimating energy savings at the city scale [40]. Caputo et al. used four archetypes to characterize the energy performance of the built environment in a city or neighborhood, and to evaluate the effects of different energy strategies [41]. Such prototype buildings or archetypes extend the knowledge beyond individual buildings for efficient energy models of neighborhoods or cities. Furthermore, city-scale building energy benchmarking policy provides a holistic dataset foundation and enables comparison of energy performance between similar buildings [14, 42, 43]. Holistic building energy consumption data can be used for defining reference buildings by investigating the closeness of building groups, for which, cluster analysis is one of the most efficient methods. Deb and Lee [44] studied the determining key variables influencing energy consumption in 56 office buildings through cluster analysis. The clustering approach focused on a small number of representative, reference buildings from a large building dataset [45, 46]. Gaitani et al. [47] applied several variables, including the heated floor area, building age, insulation of the building envelope, number of classrooms and students, operation hours, and age of heating system, using principal component and cluster analysis methods to establish the reference buildings. Tardioli et al. developed a novel framework utilizing a combination of building classification, clustering, and predictive modelling to identify a total of 67 representative buildings out of a dataset of 13,614 mixed-use buildings in the city of Geneva [48].

One challenge in defining reference buildings, from a stock of existing ones, is how to efficiently create groups. One technique, that has received limited attention, is to use machine learning to integrate multi-buildings into an energy use prediction model. With this complexity arise, such as how to (1) capture the impact and interrelationship between multi-buildings best and (2) use reference building energy datasets to learn and predict multi-building energy use.

To address these prediction gaps, this study presents a novel data-driven interdisciplinary method, integrating social network analysis and artificial neural network (SN-ANN) techniques, to predict multi-building energy use. Energy use patterns between buildings are leveraged to identify reference buildings and create a building network. Built on the network, the SN-ANN model aims to apply energy use and building features from a small reference group (e.g., n buildings) to accurately and efficiently predict energy use of a larger group (e.g., n + m buildings). This is done by investigating the network between the reference n-buildings group and the non-reference m-buildings group. To validate the technique, the proposed approach was evaluated using campus buildings at Southeast University, China. Three-years of monthly energy use data from 2015 to 2017, was used. Seventeen buildings were selected, covering four use types namely: office buildings, educational buildings, laboratory buildings, and residential buildings.

The main contribution of this work is in the unique interdisciplinary method of combining SN and ANN to create the building relationships and network with building energy use patterns. This technique efficiently learns the building feature and network for multi-building energy use, and provides a framework for analyzing building energy use patterns in large-scaled areas. Moreover, the proposed algorithm is validated using building groups with actual data to demonstrate significant accuracy in the results. While energy prediction is critical, the data-driven energy modeling also opens many other applications, such as performance monitoring, control and optimization of building groups, distributed energy systems and micro-grids implementation, which need co-operations between buildings.

The remaining sections are outlined as follows. In Section 2, provides an introduction to the methodology, including the data-processing, network extraction, and network-based artificial neural network algorithm. The case study and model configuration are given in Section 3. Section 4 presents the results on energy use patterns, network between buildings, energy use prediction results, and prediction accuracy assessments. In Section 5, discussions of the findings, implications and the future work beyond the existing results are deliberated. Finally, Section 6 concludes this study.

2. Methodology

To establish the SN-ANN relationship, three main components were first conducted: (1) the feature selection for building energy use prediction, (2) the extraction of the reference buildings and the building of the network between buildings, and (3) the integration of the social network based artificial neural network algorithm.

2.1 Feature selection

Before implementing the SN-ANN method, pre-processing of raw data is necessary to eliminate erroneous or missing measurements in the energy use data. In this study, the Lagrange polynomials for interpolation filter is applied due to its computational efficiency and causality which are important in time-series applications. If we have time-series data as $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we can formulate the interpolation for the default measurement as shown in Eq. 1.

$$y = \sum_{i=0}^{n} y_i \prod_{j=0, j \neq i}^{n} \frac{x - x_j}{x_i - x_j} \quad (1)$$

Where, y is the interpolation value and n is the size of the data used for interpolation.

After the dataset is filtered, the time series data of energy use data for one building can be described in Eq. 2.

$$X = (x_1, x_2, \ldots, x_t, \ldots, x_n)^T \quad (2)$$

Where, 1, 2, …, n is the discrete time step.

For feature-based prediction, the feature is a variable which contains the information relevant for object recognition. In forecasting energy use, it should include the use, trend, and the determined factors of building energy. Therefore, we considered two kinds of features, the value and the change of energy use parameter. The change of a parameter was equated using Eq. 3.

$$\Delta x = x_t - x_{t-1} \tag{3}$$

To better predict building energy use, some determined factors are also considered in this study, including the year-built, construction type, number of stories, building area, and roof type.

2.2 Network extraction

Predicting the total energy use or the demand in a distributed building group is difficult and complicated, especially at the city scale. Moreover, the task of collecting historical energy use datasets for large-scale building groups is a big issue. One way to overcome these complexities is to exact typical building geometries and simulate or estimate the total energy use using typical buildings. However, such methods ignore actual energy use patterns which are influenced not only by building geometries, but also occupancy, operation mode, and so on…

This study uses the social network (SN) analysis method to extract and enhance connections between buildings through their historical energy use patterns. SN analysis is the process of investigating connections between networked nodes (individual actors, people, or things) through the use of networks and graph theory. In an effort to reduce the number of buildings used, we established connections and relationships between buildings in a building group. From this the total energy use of distributed buildings with part of the total building set is inferred. To build the connections of individual buildings using SN analysis, two main approaches are used: (1) distance (e.g., Euclidean distance), which is usually used to calculate the difference between two objects, and (2) correlation (e.g., Pearson correlation coefficient), which is usually used to find the similarity between the two objects. In this study, we use the Pearson correlation coefficient method to calculate the connections between buildings shown in Eq. 4.

$$c_{EU_i,EU_j}=\frac{E\left(EU_i EU_j\right)-E\left(EU_i\right)E\left(EU_j\right)}{\sqrt{E\left(EU_i^2\right)-\left(E\left(EU_i\right)\right)^2}-\sqrt{E\left(EU_j^2\right)-\left(E\left(EU_j\right)\right)^2}} \tag{4}$$

Where, c_{b_i,b_j} is the correlation coefficient between building *i* and *j*. EU_i and EU_j are the energy use dataset of building *i* and *j*.

Two steps are taken to extract the networks. The first step is to identify the reference buildings by using Eq. 4. The reference buildings are used to predict building energy use. The second step is to build networks between reference building and non-reference buildings.

2.3 Social Network based Artificial Neural Network model (SN-ANN model)

To predict the energy use of multi-buildings using a building subset, this study proposed the Social Network based Artificial Neural Network (SN-ANN) model shown in Fig.1. The ANN algorithm is used to solve problems similar to the human brain. ANN consists of a network of simple neuron elements connecting the output to the input with the directed and weighted graph. The capabilities of the ANN algorithm fall within the realm of regression analysis including time series prediction and modeling, classification, including pattern recognition and sequential decision making, and so on. Meanwhile, since this study considers multi-features in the prediction model, the ANN model is also well fit to tackle the different scale of feature datasets.

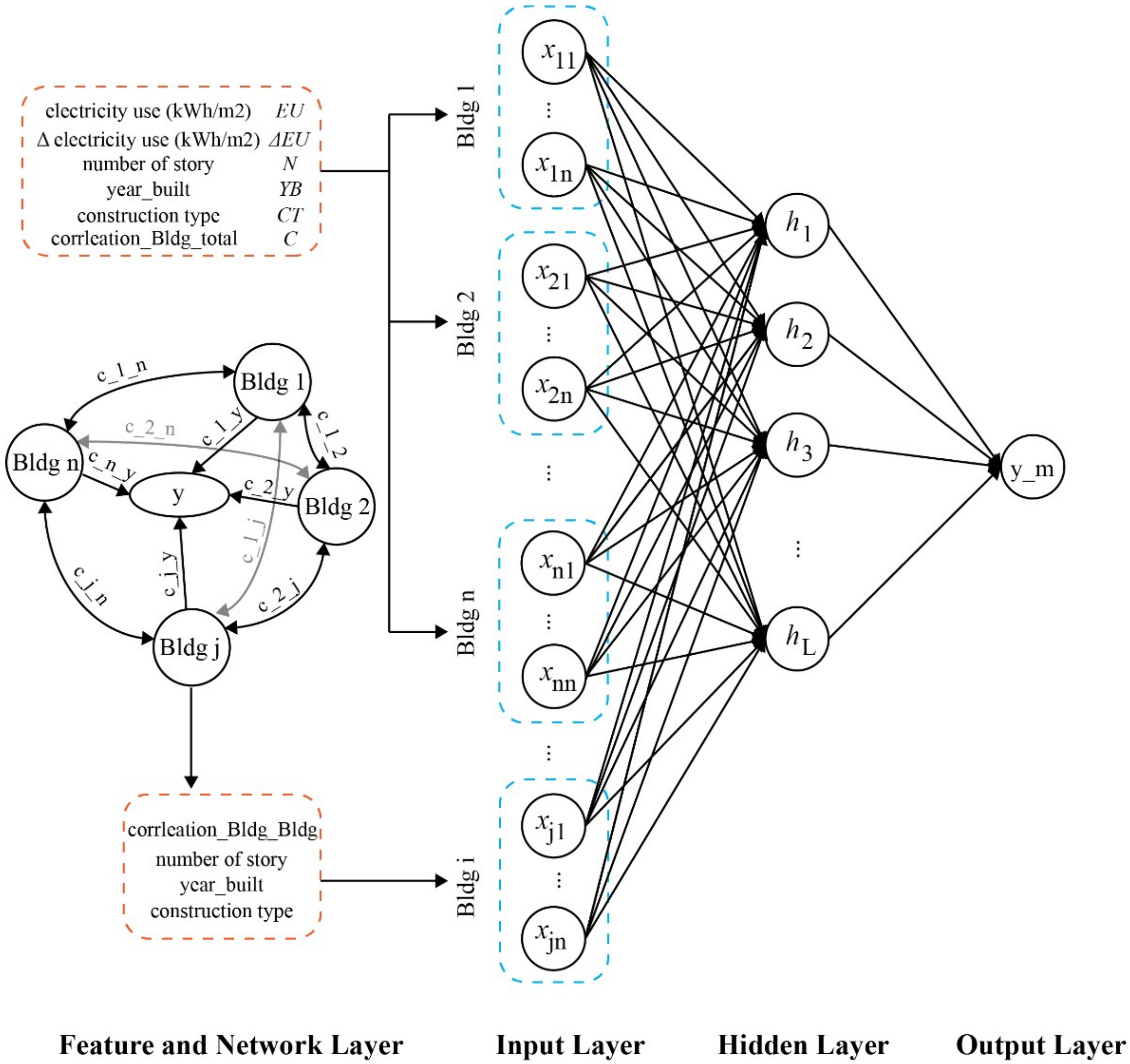

Fig. 1 The construction of the SN-ANN model.

The SN- ANN algorithm has four layers: the feature layer, input layer, hidden layer, and the output layer. In the feature layer, we selected the building information property dataset and the SN dataset to represent the compiled dataset composing the input vector for the input layer of SN-ANN model.

Suppose that *EU* represents the energy use vector, Δ*EU* represents the energy use difference vector, *NS* represents the number of building story vector, *YB* represents the year-built vector for buildings, *CT* represents the construction type vector, and *C* represents the correlation network vector between buildings, then, the input vector for the input layer can be illustrated in Eq 5.

$$X_{input}=(x_1, x_2, \ldots, x_i, \ldots, x_N)^T=(EU, \Delta EU, NS, YB, CT, C)^T \quad (5)$$

Where, N is the size of input layers and n is the size of buildings.

$$EU = (e_1, e_2, \ldots, e_i, \ldots, e_n) \quad (6)$$

$$\Delta EU = (\Delta e_1, \Delta e_2, \ldots, \Delta e_i, \ldots, \Delta e_n) \quad (7)$$

$$NS = (ns_1, ns_2, \ldots, ns_i, \ldots, ns_n) \quad (8)$$

$$YB = (yb_1, yb_2, \ldots, yb_i, \ldots, yb_n) \quad (9)$$

$$CT = (ct_1, ct_2, \ldots, ct_i, \ldots, ct_n) \quad (10)$$

$$C = (c_{\text{¿}}, c_{\text{¿}}, \ldots, c_{\text{¿}} \ldots, c_{\text{¿}}, \ldots, c_{\text{¿}}) \quad (11)$$

To eliminate the impact caused by different scales of the feature dataset, we need to normalize the different feature vectors. The output of the hidden layer, the output of the output layer, weights from the hidden layer, and weights from the hidden layer to the output layer are defined in Eq. 13, Eq. 14, Eq. 15, and Eq. 16, respectively.

$$X_{input}=(x_1, x_2, \ldots, x_i, \ldots, x_n)^T \quad (12)$$

$$H_{hidden}=(h_1, h_2, \ldots, h_j, \ldots, h_m)^T \quad (13)$$

$$Y_{output}=(y_1, y_2, \ldots, y_k, \ldots, y_l)^T \quad (14)$$

$$V=(v\text{¿¿}1, v_2, \ldots, v_j, \ldots, v_m)\text{¿} \quad (15)$$

$$W=(w\text{¿¿}1, w_2, \ldots, w_k, \ldots, w_l)\text{¿} \quad (16)$$

Where, m and l are the length of the hidden layer and the output layer, respectively. The v_j donates the weight vector of the j^{th} neural cell of hidden layer and w_k donates weight vector of the k^{th} neural cell of output the layer. The length of input t layer is determined by the number of elements of the input data while the length of the hidden layer (m) is randomly selected. Four occupancy prediction objects are used, which are maximum, average, and minimum number of occupants and vacant. However, the ‘vacant’ status of occupancy is defined when the output of the minimum of occupancy is zero. Therefore, the length of output layer (*l*) which equals the number of expected output elements is selected as 3. The mathematic information transfer between each layer can be expressed in Eq. 17 and Eq. 18. The activation function is shown in Eq. 19.

$$h_j=f \text{ ¿} \tag{17}$$

$$y_k=f \text{ ¿} \tag{18}$$

$$f(x)=\frac{1}{1+e^{-x}} \tag{19}$$

The ground truth of energy, Eq. 20, assumes one output neuron, with the squared error function shown in Eq. 21.

$$d=(d_1, d_2, \ldots, d_k, \ldots, d_l)^T \tag{20}$$

$$E=\frac{1}{2}\sum_{k=1}^{l}(d_k-y_k)^2 \tag{21}$$

The gradient descent approach computes the derivative of the squared error function and iterates through different weights to minimize the error shown in Eq. 22 and Eq. 23. This is discussed more in Section 3.2.

$$\Delta v_{jk}=\eta\left(\sum_{k=1}^{l}(d_k-y_k)y_k(1-y_k)w_{jk}\right)h_i(1-h_i)x_i \tag{22}$$

$$\Delta w_{ij}=\eta(d_k-y_k)y_k(1-y_k)h_i \tag{23}$$

3. Case study

3.1 Description of the case study

A distributed building group from Southeast University (SEU) was selected to validate the proposed multi-building energy consumption prediction algorithm. The SEU is located in the center of Nanjing City, Jiangsu Province, China. The SEU has a total area of 3.9 km^2 and consists of 53 buildings on the main campus, including office buildings, laboratories, educational buildings, multiple-use buildings, residential buildings, and other building types. A group of other buildings includes some auxiliary service buildings, such as a kindergarten, an elementary school, retail stores, canteen, and so on. The multi-use buildings usually consist of classrooms, research rooms, office rooms, lab areas, etc. With only two multi-use buildings on the campus and the pattern of their energy use hard to identify, the multi-use buildings and a few other buildings were not considered. The four types of buildings analyzed include: office, educational, laboratory, and residential. The dataset of the four building groups, provided by the General Affairs Department of SEU, includes energy use, year built, construction type, wall material, total building floor area, use type, and the number of stories. Energy use data was collected monthly from 2015 to 2017. The buildings without complete three-year data were excluded, resulting in: six office buildings (O1-O6), four educational buildings (E1-E4), four laboratories (L1-L4), and three residential buildings (R1-R3) being used for validation purposes (Fig. 2). Details of each office, educational, laboratory and residential building groups are presented in Tables 1, 2, 3, 4, respectively.

Fig. 2 Diagram of the footprint and location of the building groups.

Table 1. Office building group

Building Type	Year Built	Construction Type	No. of story	Buildin g Area	Roof Type
Office Building 1 (O1)	1980	Reinforced concrete structure	16	16910	Flat roof
Office Building 2 (O2)	1927	Reinforced concrete structure	3	5072	Sloping roof
Office Building 3 (O3)	1957	Brick-concrete structure	4	3938	Sloping roof
Office Building 4 (O4)	1922	Brick-concrete structure	2	4500	Sloping roof
Office Building 5 (O5	1990	Brick-concrete structure	8	11748	Flat roof
Office Building 6 (O6)	1991	Reinforced concrete structure	4	7106	Flat roof

Table 2. Educational building group

Building Type	Year Built	Construction Type	No. of story	Buildin g Area	Roof Type
Education building 1 (E1)	1980	Brick-concrete structure	3	3630	Sloping roof
Education building 2 (E2)	1987	Brick-concrete structure	6	5595	Flat roof

Education building 3 (E3)	1982	Brick-concrete structure	6	7482	Flat roof
Education building 4 (E4)	1982	Brick-concrete structure	3	2859	Flat roof

Table 3. Laboratory building group

Building Type	Year Built	Construction Type	No. of story	Building Area	Roof Type
Laboratory 1 (L1)	1994	Brick-concrete structure	4	2993	Flat roof
Laboratory 2 (L2)	1955	Reinforced concrete structure	6	10902	Flat roof
Laboratory 3 (L3)	1957	Reinforced concrete structure	1	949	Sloping roof
Laboratory 4 (L4)	1957	Reinforced concrete structure	1	1421	Sloping roof

Table 4. Residential building group

Building Type	Year Built	Construction Type	No. of story	Building Area	Roof Type
Residence building 1 (R1)	1980	Reinforced concrete structure	4	1313	Flat roof
Residence building 2 (R2)	1990	Reinforced concrete structure	16	12906	Flat roof
Residence building 3 (R3)	1980	Reinforced concrete structure	15	9980	Flat roof

3.2 Model configuration and assessment

To eliminate the impact of different scales embedded in the different building features, it was necessary to configure the inputs of the SN-ANN model before model training. For the year-built feature, the age varied from 1922 to 1994 with a clear separation around 1960 to 1970. The model sets a binary value of 0 for buildings built before year 1965, otherwise 1. Construction type also required two values. The model sets the value of “reinforced concrete structure” as 0 and the value of “brick-concrete structure” as 1. Similarly, the value of “flat roof” and “sloping roof” was set as 0 and 1, respectively. The value of the number of stories was normalized as in [0, 1], while the building floor area was used to calculate the building energy use intensity. Table 5 shows the details of different scales of feature data for model input. Fig. 3 shows the learning process of the SN-ANN model. With the EUI dataset, Eq. 4 was used to identify reference buildings.

To dynamically update the network between buildings, one time window with length ΔT was applied in the model to calculate the correlations. During model training, the initial parameters for the neural networks were defined randomly, including the weights and bias of each neural between each layer. The gradient descent rule was applied to learn and update the weights and bias shown in Eq. 22 and 23 until the errors between predicted values and actual values were minimized. In validation, this study compared the predicted energy use proposed by the SN-ANN and ANN models with actual measured energy use data. To assess the model performance, three indices were used to compare the results for accuracy, the mean absolute percentage error (Eq. 24), the root mean squared error (Eq. 25) and the Q-Q plot curve.

(1) Mean Absolute Percentage Error (MAPE) shows the mean percentage error between the predicted occupant count and the actual number of occupants.

$$MAPE\left(EUI^{p}\right)=\frac{1}{N}\sum_{i=1}^{N}\left|\left(EUI_{i}^{a}-EUI_{i}^{p}\right)/EUI_{i}^{a}\right| \tag{24}$$

(2) Root Mean Squared Error (RMSE) shows the magnitude of the estimation error.

$$CV\,RMSE\left(EUI^{p}\right):=\frac{\sqrt{\sum_{i=1}^{N}\left(EUI_{i}^{a}-EUI_{i}^{p}\right)^{2}/N}}{\sum_{i=1}^{N}EUI_{i}^{a}/N} \tag{25}$$

Where, the EUI^a and EUI^p are the actual and predicted building EUI, respectively. N is the sample size.

(3) Q-Q plot curve is a graphical plot used to compare the true positive rate and the false positive rate as the criterion changes.

Table 5. Details of the SN-ANN model input

Feature	Source/Scope	Description
Year Built		Discrete, Binary
Construction Type	Reinforced concrete structure, Brick-concrete structure	Binary
No. of story	[1, 16]	Discrete, Normalization
Roof Type	Flat roof, Sloping roof	Binary
Building Area	[1313, 16910]	To calculate Energy Use Intensity
Energy Use Intensity	N.A.	Normalization
Correlation	Eq. 4	Normalization

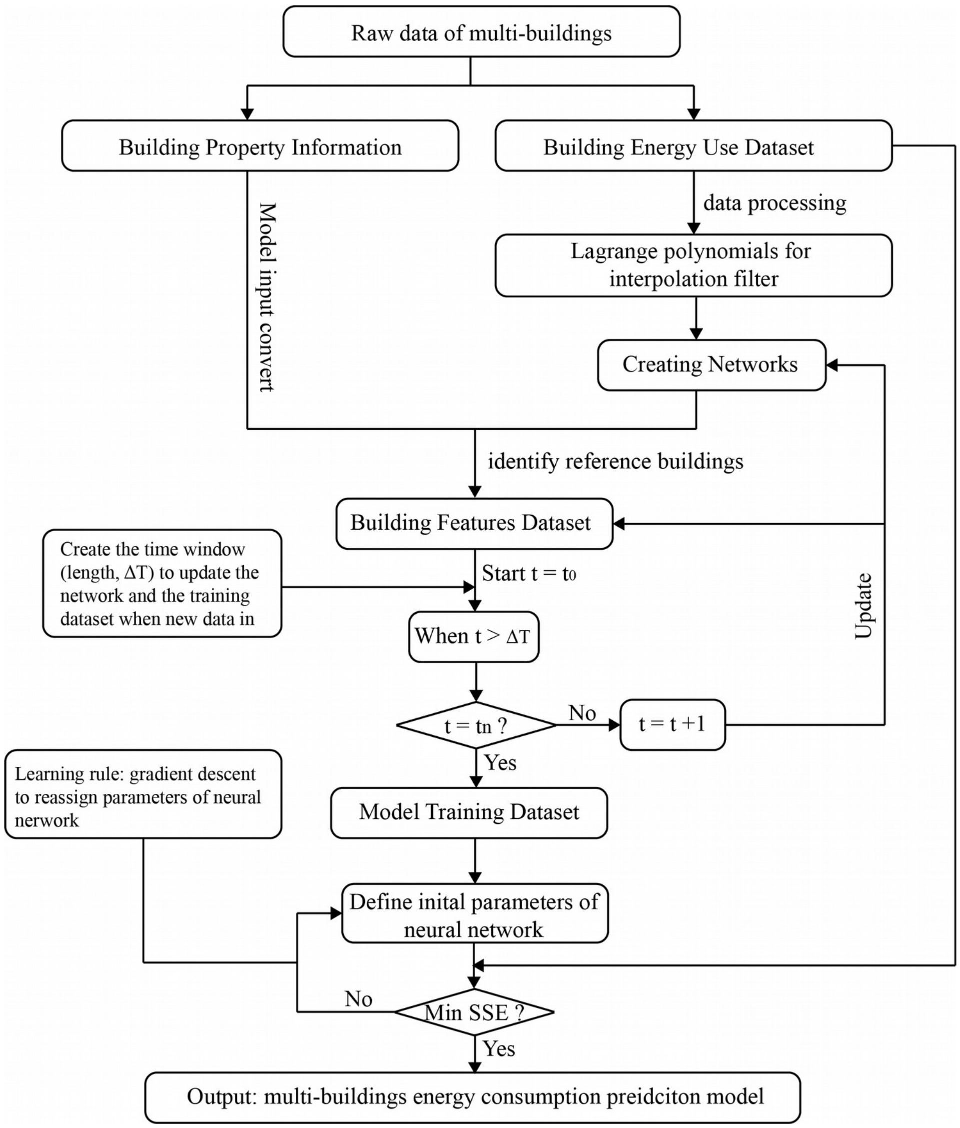

Fig. 3 Overview of the learning process of the SN-ANN model

4. Results and assessment

4.1 Building Energy Use Intensity Results

The energy use intensity measurement results of the office, educational, laboratory, and residential buildings are shown in Figs. 4(a-d), 5(a-d), 6(a-d), 7(a-d), respectively. For each building group, the energy use intensity distribution in year 2015, 2016, 2017, and the total building energy use intensity box plot are presented.

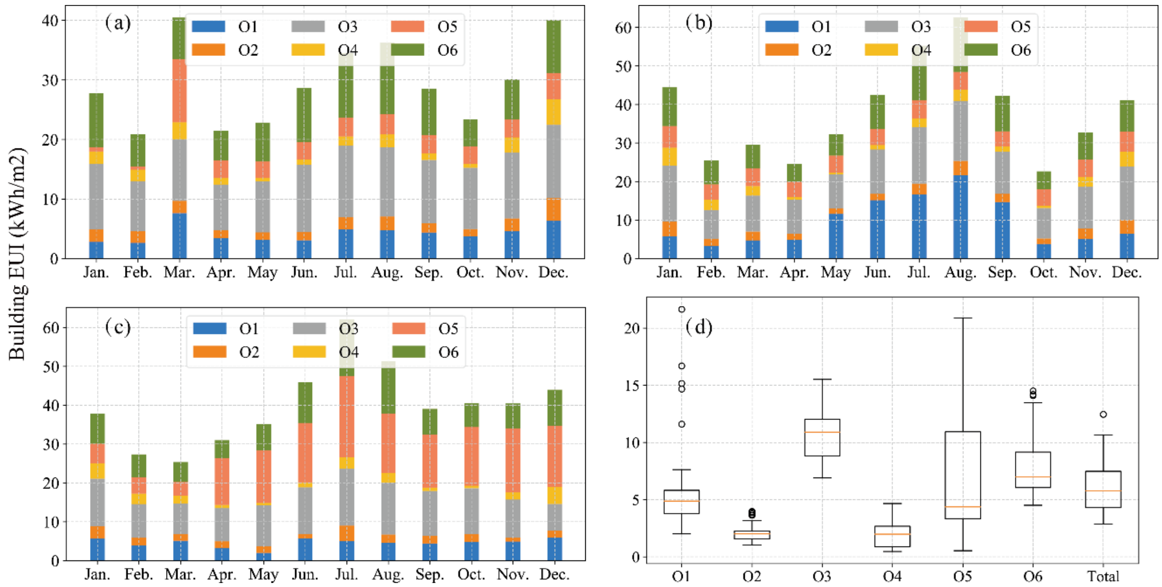

Fig. 4 (a-d): EUIs of the office building group (a, b, c) and total building energy use (d).

The results show in Figure 4(a-c) that the three biggest energy consumers are O1, O3, and O6 in 2015 and 2016, and O3, O5, and O6 during 2017. Figure 4(d) shows the combined total of the EUI of the office building group. The EUI trends show that the office buildings generally consumed more energy in the winter (December and January), and summer (July and August) months, due to typical seasonal patterns. The results show the EUI of O1 varies from 2 to 22 kWh/m^2, with an average of 6 kWh/m^2 and a standard deviation of 4 kWh/m^2. The O2 and O4 buildings consumed less energy and have minimum EUIs of about 1 and 0.5 kWh/m^2, respectively. The average EUIs for O2 and O4 buildings is about 2 kWh/m^2 (for both), with maximums of 4 and 4.7 kWh/m^2 and a standard deviation of 0.8 and 1.2 kWh/m^2, respectively. Observed from Table 6, O2 and O4 have smaller floor areas and are also older, than O1 and O3. O3 has the biggest average EUI of 11 kWh/m^2 and a minimum EUI of 7 kWh/m^2. For the entire office building group, the EUI varies from 3 to 12.5 kWh/m^2 with an average of 6 kWh/m^2.

Table 6. EUIs of the office building group (kWh/m²)

	Min.	Mean	Std.	Max.
O1	2.01	6.14	4.35	21.64
O2	1.05	2.13	0.83	3.97
O3	6.89	10.76	2.25	15.51
O4	0.46	2.00	1.21	4.66
O5	0.54	6.81	5.35	20.89
O6	4.51	8.01	2.88	14.50
O_total	2.86	6.15	2.29	12.46

Fig. 5(a-d) and Table 7 show the EUIs and the combined total building energy use intensity box plot of the educational buildings. The educational buildings are relatively new and smaller (Table 2) compared with the office buildings and the EUI results are reflective of this fact. Additionally, the educational buildings empty during summer (July and August) and winter (February) breaks. For E1, E2 and E4, the EUI varies (minimum to maximum) from 1 to 3 kWh/m², from 1 to 5 kWh/m², from 1 to 3 kWh/m², respectively. While E2, the biggest energy consumer, varied from 4 to 14 kWh/m².

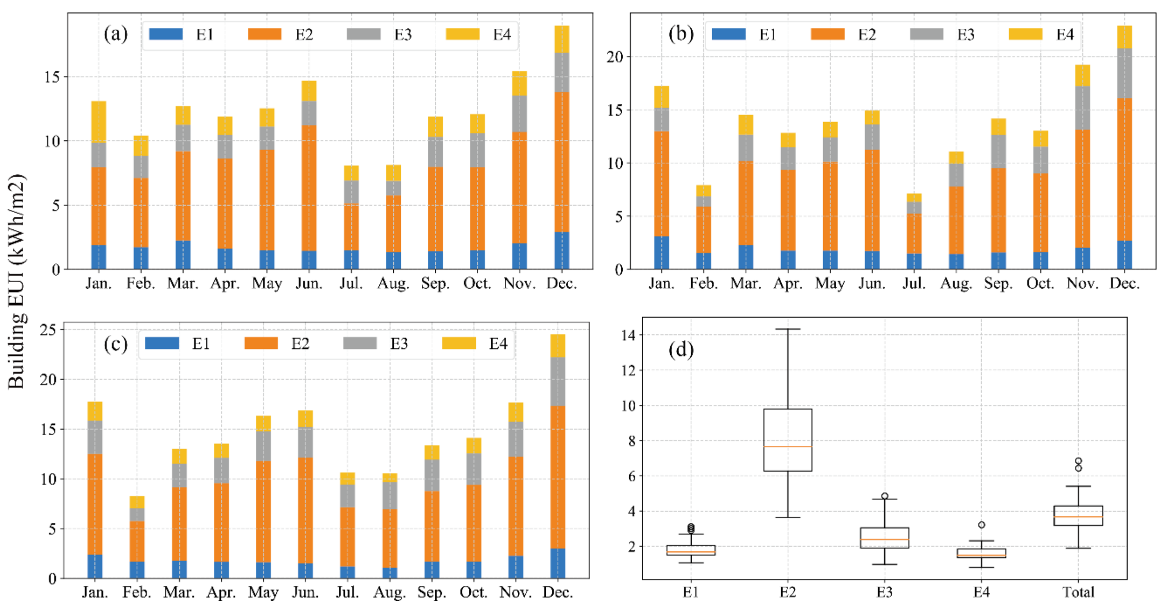

Fig. 5 (a-d): The EUIs of the educational building group (a, b, c) and total building energy use (d).

Table 7. EUIs of the educational building group (kWh/m^2)

	Min.	Mean	Std.	Max.
E1	1.06	1.83	0.49	3.11
E2	3.63	7.84	2.54	14.34
E3	0.97	2.51	0.90	4.86
E4	0.79	1.58	0.45	3.22
E_total	1.89	3.77	1.13	6.85

Fig. 6(a-d) and Table 8 show the EUIs of the laboratory buildings with no unclear trend. The average EUI for L4 is substantially less than the EUIs of the other laboratory buildings. L4 varied from nearly zero (0.04 kWh/m^2) to 6 kWh/m^2 and with an average of 1 kWh/m^2. Although L3 has a similar EUI maximum to minimum range as L4 (varying from 0.2 to 7.3 kWh/m^2) the average EUI of L3 was 4.8, far higher than that of L4. L2 consumes the largest amount of energy and has the largest building area. Its EUI varies from 6.4 to 11.4 kWh/m^2, with an average of 8.5 kWh/m^2. While for L1, its EUI ranged from 0.7 to 11 kWh/m^2, with an average of 3.6 kWh/m^2.

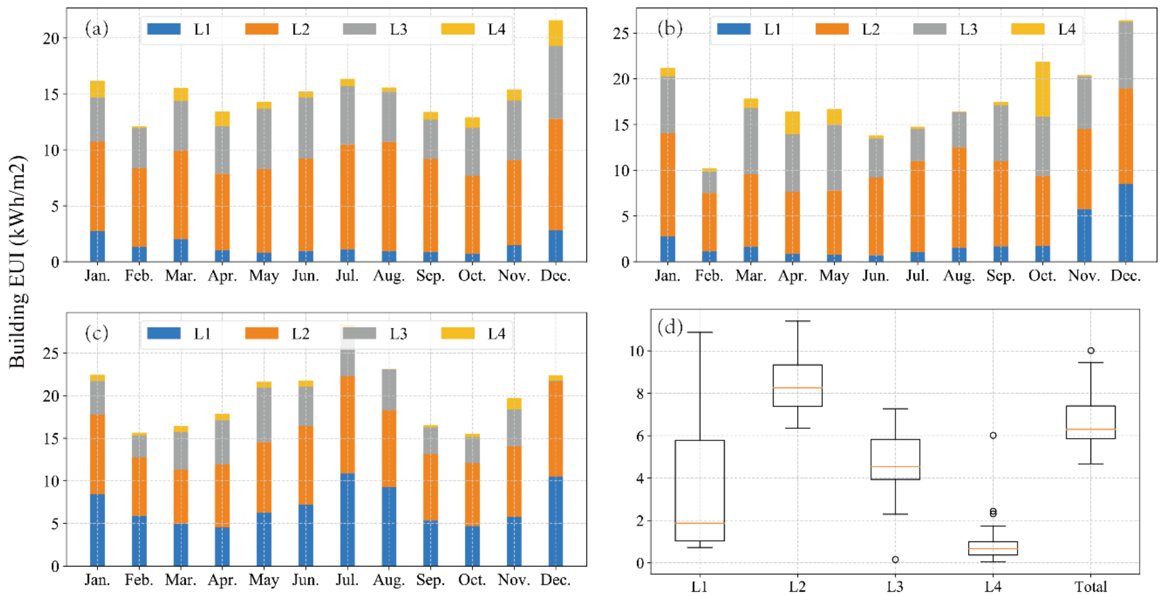

Fig. 6 (a-d): The EUIs of the laboratory building group (a, b, c) and total building energy use (d).

Table 8. EUIs of the laboratory building group (kWh/m^2)

	Min.	Mean	Std.	Max.
L1	0.72	3.58	3.11	10.87
L2	6.36	8.48	1.44	11.40
L3	0.16	4.75	1.53	7.27
L4	0.04	0.89	1.05	6.01
L_total	4.67	6.70	1.29	10.01

Fig 7(a-d) and Table 9 shows the EUIs of the residential buildings. The EUI trend for the student residential buildings is very similar to the educational buildings, as they are utilizing the same educational schedule. During the summer and winter breaks, the occupancy and operation of the residential buildings decreases, thus the building energy use decreases accordingly. This study selected three adjacent residential buildings which were built in the same year, with the same construction wall and roof type. The R2 and R3 are high-rise buildings with 16 and 15 stories, respectively. R2 has a floor area of 12,906 m^2, larger than R3. The EUIs of R2 vary from 0.2 to 17 kWh/m^2 with an average of 4 kWh/m^2; while EUIs of R3 vary from 0.2 to 6 kWh/m^2 with an average of 2 kWh/m^2. While for R1, its EUI is from 0.7 to 10 kWh/m^2 with an average of 3.5 kWh/m^2.

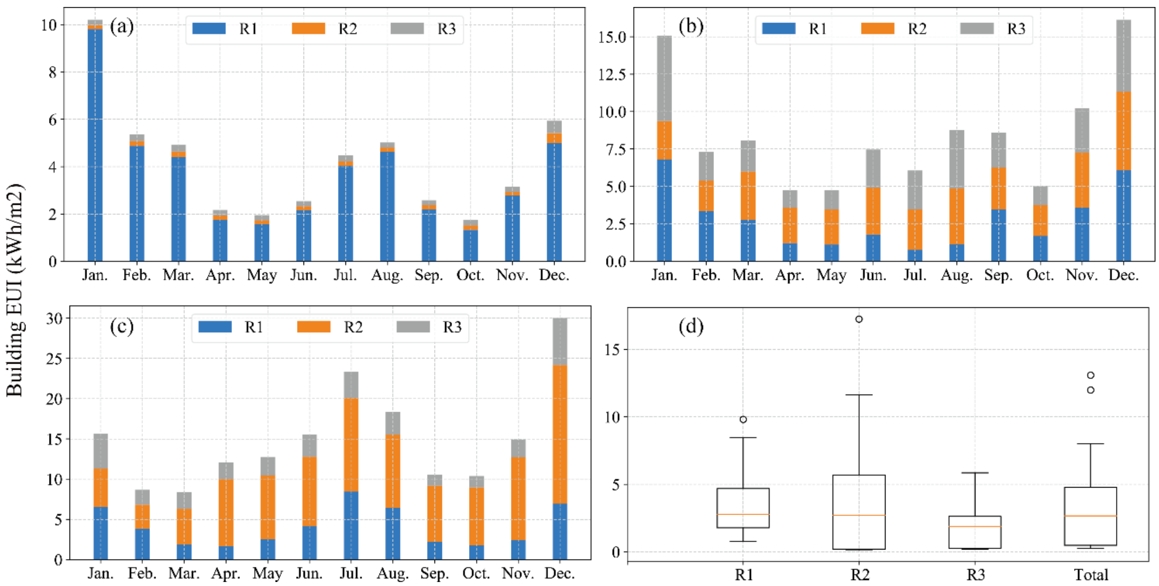

Fig. 7(a-d): The EUIs of the residential building group (a, b, c) and total building energy use (d).

Table 9. EUIs of the residential building group (kWh/m^2)

	Min.	Mean	Std.	Max.
R1	0.77	3.54	2.25	9.81
R2	0.16	3.82	4.03	17.24
R3	0.21	1.89	1.59	5.85
R_total	0.27	3.35	3.12	13.08

4.2 Network analysis and prediction accuracy

This section discusses the applications of the social network analysis and the social network based machine learning technique, as well as presents the multi-building prediction accuracy. Fig. 8 (a-d) shows the networks between the buildings in each group, by calculating the correlations of individual building EUI with the total building group EUI, in year 2015 to 2017.

In the office building group, buildings O1, O3, and O6 were identified as the reference buildings (Fig. 8a). The correlations between O1, O3, O6 and the total building's EUI are 0.7, 0.6, and 0.7, respectively. Building O1 shows the most relevant trend to the total building EUI. Observed from the networks between the non-reference buildings and the reference buildings, it is found that most non-reference buildings do not have much relevancy to the reference buildings and their EUI trend correlations are generally less than 0.6 (except for O2 and O3 with a correlation of 0.7). This is especially true for O1 and O5 as they shared a negative network correlation.

For the educational building group building E4 is the only non-reference building (Fig. 8b). The building with the most relevant trend to the total building's EUI trend is building E2 with a high correlation of 0.98. The building E3 is also highly correlated to the total building's EUI trend with a correlation of 0.91.

While considering the networks in the laboratory buildings, buildings L1 and L2 are the reference buildings with correlations of 0.72 and 0.92, respectively (Fig. 8c). For the non-reference buildings, building L4 is negatively relevant to both buildings L1 and L2, showing opposite EUI trends between L4 and L1, L2, respectively. Meanwhile building L3 shows a much lower relation of EUI trend to the reference buildings L1 and L2.

In the residential buildings group, buildings R2 and R3 are identified as the reference buildings with high correlations of 0.82 and 0.96 (Fig. 8d). Building R1 shows a positive relationship between the two reference buildings, but with low correlations.

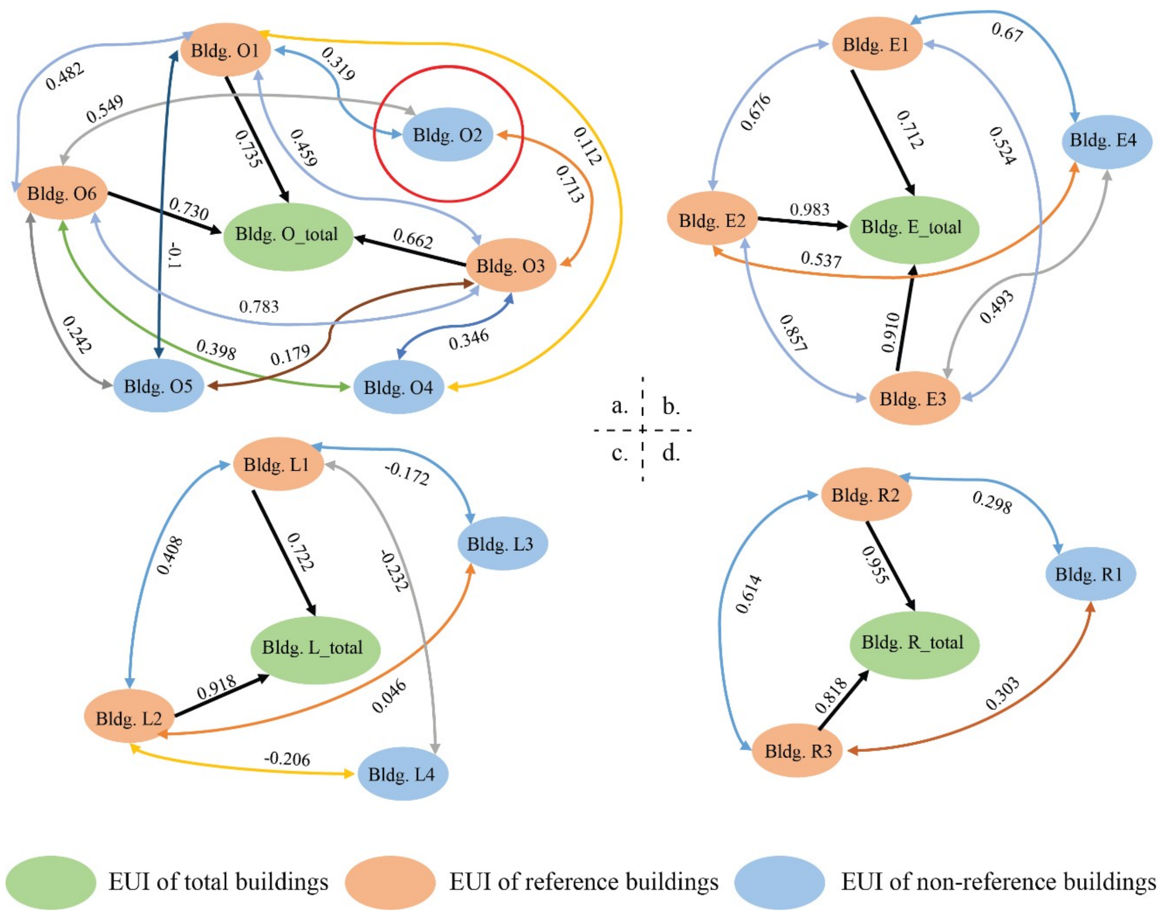

Fig. 8 (a-d): The social network analysis results of (a) office, (b) educational, (c) laboratory and (d) residential buildings.

To compare the proposed SN-ANN model, two baseline results were applied (Fig. 9). The first one compares the actual building EUI to the predicated EUI using the SN-ANN model for a six-month period from July to December. The second compares the actual EUI with the predicted building EUI generated by applying the reference buildings in the ANN model while ignoring networks between the buildings. In most cases the actual and predicted values are reasonable. To further understand the prediction performance Q-Q plots were generated.

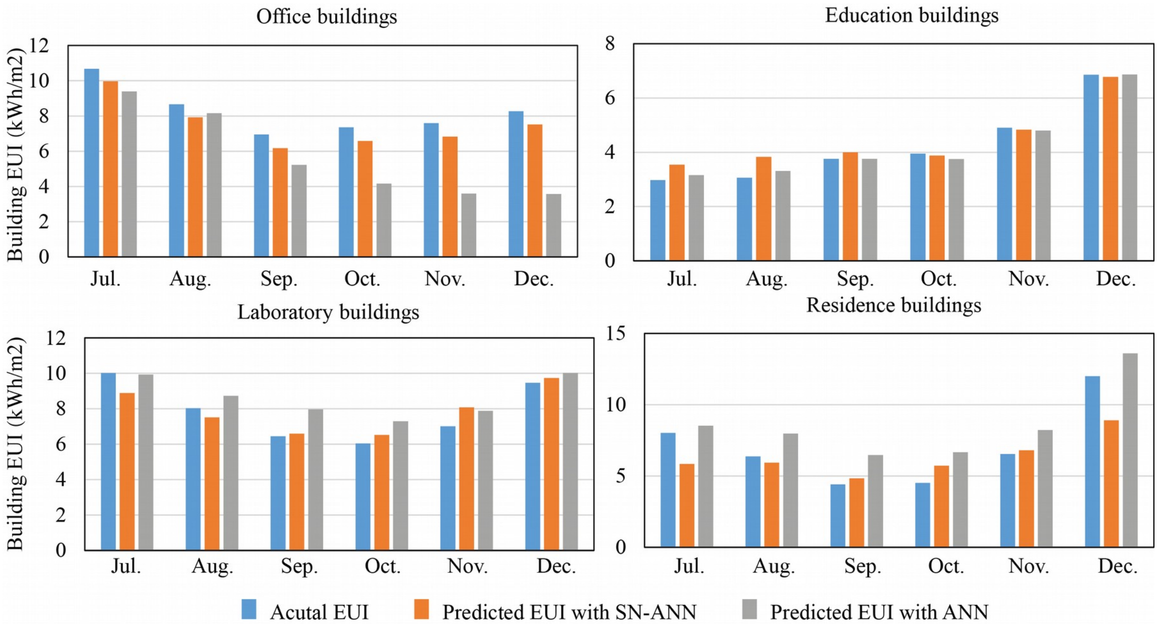

Fig. 9 The building EUI prediction results for the office, educational, laboratory and residential buildings.

The Q-Q plot represents the quantiles of the actual data set compared against the quantiles of the predicted data set. Fig. 10 presents an assessment of the results for the office and educational buildings. The results from the office buildings Q-Q plot validate the predicted SN-ANN model results with R^2 of 1. Moreover, the SN-ANN model performed better than using just the ANN model (R^2 of 0.6309). Additionally, the SN-ANN predicted results gravitate closer to the line Y = X, which indicates the SN-ANN model can predict more accurately the actual EUI. However, for the educational buildings, although the two models achieved good prediction performance (R^2 of 0.9578 for the SN-ANN model and R^2 of 0.988 for the ANN model), the ANN model predicted results far better when assessing the line Y = X.

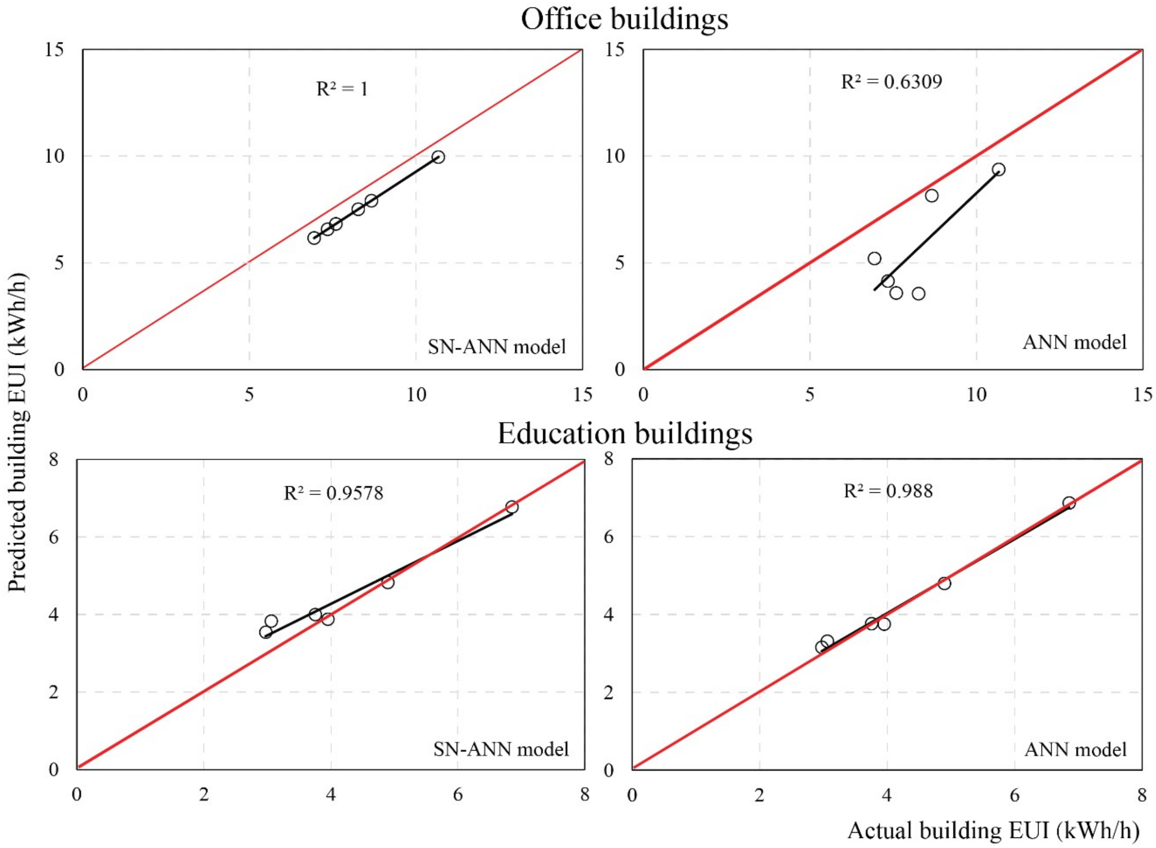

Fig. 10 The Q-Q plot for the office and educational buildings.

Fig. 11 presents the Q-Q plot results for the laboratory and residential buildings. Observed from the results, the ANN model achieved better predicted results in both building groups than the SN-ANN model as indicated by the R^2 values. However, compared with the ANN model, the SN-ANN model has more predicted building energy results falling with the line Y = X. This means the SN-ANN model achieved more accurate prediction results. A numerical comparison of the ANN and SN-ANN results are presented in Table 10 with MAPE and RMSE calculations. Compared with the ground truth, the SN-ANN model for the office building group showed greater than a 23% improvement in prediction accuracy over the ANN model for both the MAPE (9.3% SN-ANN, 33% ANN) and the RMSE (9.1% SN-ANN, 36% ANN). For the educational buildings, it shows that although the SN-ANN model achieved an acceptable accuracy, the ANN model can lead to a better total building EUI without considering the networks between the non-reference building E4 and the reference buildings E1, E2, and E3. This might be because there are already two reference buildings (E1, E2) that are highly related to the total building energy EUI. It is adequate to predict the total building EUI with the reference buildings. While for the laboratory buildings, we can find the SN-ANN model can have a better performance with the MAPE and RMSE of 8% and 9%, respectively, while the two assessment indices are both 12% for the ANN model. In the residence buildings, the RMSE results show the SN-ANN model doesn’t improve a lot on the robustness of the total building EUI prediction compared with the ANN model. However, the SN-ANN can improve the accuracy of EUI prediction results from 27% to 17% compared with the ANN model. Finally, the overall assessment shows the SN-ANN model based prediction accuracy with MAPE is 11% while the ANN model based accuracy with MAPE is 19%. Also, the robustness can be improved from 26% to 14% when comparing the SN-ANN with the ANN model. It can be concluded that the social network analysis can greatly improve the prediction performance of the ANN model by integrating the networks between the reference buildings with the total building energy use and the non-reference buildings.

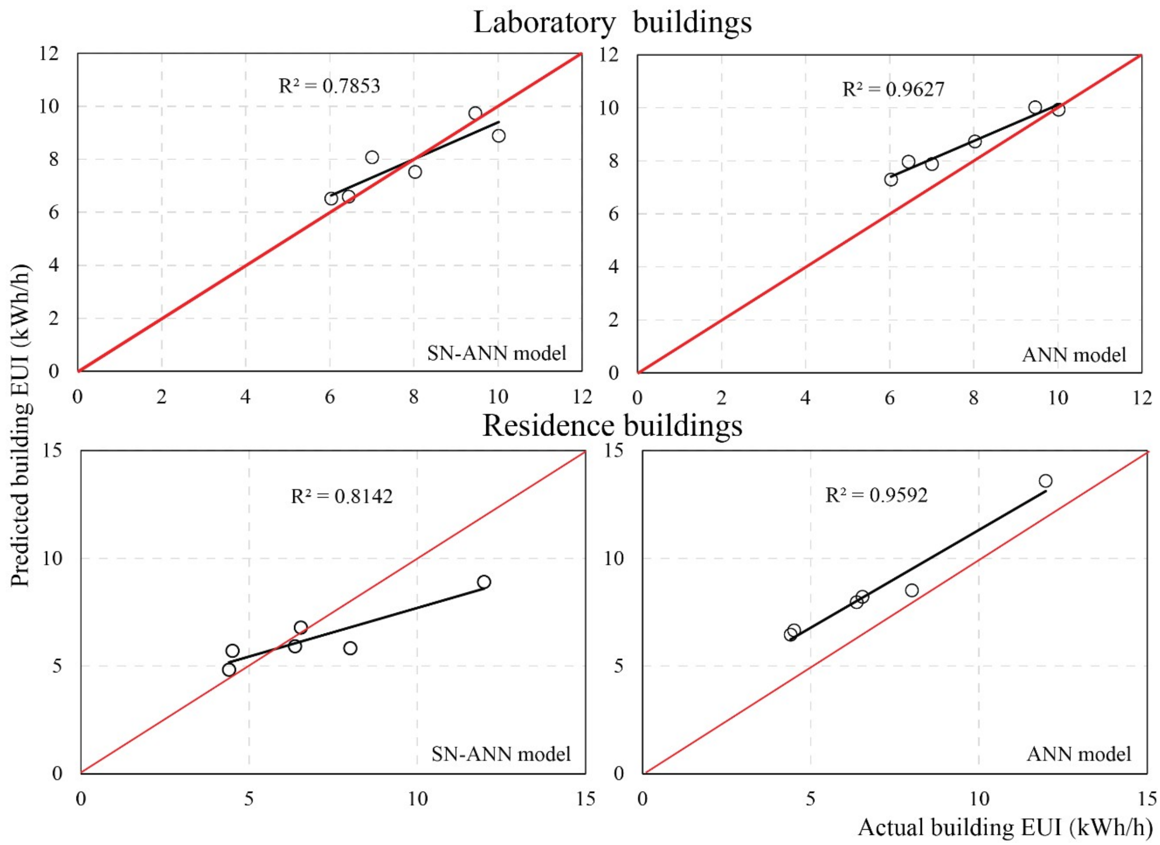

Fig. 11 The Q-Q plot for laboratory and residential buildings.

Table 10 Comparison of the predicted results from the SN-ANN and ANN models

	SN-ANN model		ANN model	
	MAPE	RMSE	MAPE	RMSE
Office buildings	9.33%	9.12%	32.6%	36.1%
Education buildings	9.21%	9.56%	3.6%	3.66%
Laboratory buildings	7.66%	9.04%	12.1%	12.2%
Residence buildings	16.68%	23.55%	27.46%	24.19%
Total	10.72%	14.52%	18.94%	26.06%

In the proposed SN-ANN model, three networks are vital for the building group EUI prediction. They are: (1) the network between the reference buildings' EUI and the total buildings' EUI, (2) the network between each other of the individual reference buildings' EUI, and (3) the network between the reference buildings' EUI and the non-reference buildings' EUI. For office buildings, it shows in Fig. 9 that the predicted building EUI results based on the SN-ANN and ANN models are less than the actual building EUI. This may be attributed to two important reasons, that the three networks are weak in the office buildings group and the sum of reference buildings' EUI is far less than the total building' EUI. However, the accuracy using the SN-ANN model is improved substantially when compared with the ANN model, which only used the reference buildings' EUI to predict the total buildings' EUI. While for the educational buildings, the findings are on the contrary; the networks between each reference building's EUI and the total buildings' EUI and the networks between reference buildings, are both strong so that the accuracies based on the SN-ANN and the ANN models are both significant with the latter being a little better. This indicates that the ANN model, using only the reference buildings' EUI, is good enough to predict the total building energy use.

For laboratory buildings, the accuracy using the SN-ANN model is also improved. Although the network between the reference buildings' EUI and the total buildings' EUI is strong, it shows that most of the networks between the reference buildings' EUI and the non-reference buildings' EUI are negative. This phenomenon causes the predicted results based on the SN-ANN model to be less than those based on the ANN model. In the residential building group, the accuracy is improved and networks between each other of reference buildings' EUI, and the reference buildings' EUI and the total buildings' EUI are both strong. Considering the standard deviation of each building group's EUI pattern, we found that the standard deviation of the office building group and the residential building group are relatively high. Applying the SN-ANN model in those two groups can improve accuracy when predicting the total building EUI.

In conclusion, the findings indicate that the SN-ANN model is more suitable and accurate for those buildings, in which the networks (or correlations) between reference buildings' EUI and the total buildings' EUI are weak and standard deviation of building groups' EUI is relatively high. While for other building groups, e.g., the educational buildings, in which the three networks are strong, it might infer that the ANN model will be accurate enough to predict the total building EUI.

5. Discussion

This study presented interdisciplinary research integrating social network analysis and an artificial neural network algorithm. First, we applied the social network analysis method to identify: (1) the reference buildings, the building's energy use that closely matched the total buildings energy use (correlation coefficient >= 60%), (2) the networks between the reference buildings, (3) the total building energy use trend, and (4) the non-reference buildings. In the second step, we integrated machine learning techniques with the social networks to predict the total building energy use. The final validation step used buildings from Southeast University that were divided into four building types, including office, educational, laboratory and residential groups. Noteably, each group had unique operational hours resulting in different peak energy use. The results demonstrated the proposed SN-ANN model predicted the multi-building EUI with satisfactory accuracy.

For city planners and energy policymakers, understanding energy use dynamics is critical to (1) knowing where and how energy is being consumed across the morphologic and socioeconomic contours of the city, (2) providing situational awareness of energy use to better allocate resources and target policy interventions, and (3) identifying cost-efficient savings opportunities across the city. Campuses consisting of a big group of buildings are an important component of a city, and optimization of distribution energy systems in a campus can help reduce energy use and GHG emissions in the city. In this light, the SN-ANN model can provide insights into which buildings are the most significant users and to predict the dynamic building EUI. The SN-ANN model provides a method for better allocating the distribution of energy resources. This study provides interdisciplinary framework outlining how to define reference buildings and apply them to predict multi-building energy. Finally, the multi-building energy prediction model can estimate the energy use, which is one key feature of the grid-interactive efficient buildings [49].

This study has some limitations. Firstly, it only selected the Southeast University as a case study with a limited number of buildings, and did not test the method in other types of building groups, such as the multi-use buildings. This study also does not validate the performance of the proposed SN-ANN model to predict non-campus building energy use at the city scale. Secondly, when applying the social network analysis method, three networks were created and inputted into the SN-ANN model. However, this study did not investigate if all the networks are needed or which network is better for multi-building energy prediction. Also, the findings in this study indicated that the SN-ANN model is more suitable for those building groups with higher standard deviation of EUI patterns. Therefore, more effort and richer building datasets are needed to support further research. Thirdly, further research is needed to determine whether the proposed SN-ANN model can be adopted for energy prediction of larger building groups, e.g., city-scale buildings. For this, the modeling to identify all the reference buildings at the city scale can be intensive and thus requiring cloud computing or high-performance computing to handle large-scale problems.

6. Conclusions

This study proposed a multi-building energy use prediction model by integrating social network analysis (SN) and machine learning techniques. SN methods were used to identify the reference buildings and establish correlations between the reference buildings and (1) the total building energy use and (2) non-reference buildings. The next step was to integrate the network into the artificial neural network (ANN) method. Important building property information, like the building height, number of stories, year built, roof type, and construction type were considered in the model. To validate the proposed AN-SNN method, this study selected the Southeast University as a case study, with four building groups tested including: office, educational, laboratory, and residential groups. To test the performance of the proposed SN-ANN model, we selected the ground truth energy use data and the ANN-based predicted energy use as two baselines. The results show the proposed SN-ANN and ANN models can achieve the prediction MAPE accuracies of 9.3% and 32.6%, respectively for office buildings; 9.2% and 3.6%, respectively for the educational buildings; 7.7% and 12.1%, respectively for the laboratory buildings; and 16.7% and 27.5%, respectively for the residential buildings. Considering the robustness of the RMSE results, the proposed SN-ANN and ANN models can achieve the prediction accuracies of 9.1% and 36.1%, respectively for the office buildings; 9.6% and 3.6%, respectively for the educational buildings; 9.0% and 12.2%, respectively for the laboratory buildings; and 23.5% and 24.2%, respectively for the residential buildings. Also, observed in the overall results, the SN-ANN model can predict the multi-building energy use with an accuracy of MAPE and RMSE of about 10.72% and 14.52%, respectively, demonstrating that the proposed model can efficiently and accurately predict the multi-building energy use. Moreover, the findings indicate that the SN-ANN model is more suitable for buildings, in which the networks between reference buildings' EUI and the total buildings' EUI are weak and the standard deviation of building groups' EUI is relatively high. While for other building groups, in which the three networks are strong, the ANN model proved to be accurate enough to predict the total building EUI. The proposed interdisciplinary SN-ANN model, presented in this study, provides a new, attractive empirical approach to urban building energy use prediction. In future work, we will apply and enhance the SN-ANN model to a larger group of buildings in city districts or an entire city depending upon availability of monthly energy use data.

Acknowledgment

This work was also supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy and Buildings 2008;40:394–8. doi:10.1016/j.enbuild.2007.03.007.

[2] City Energy Project | A Joint Project of NRDC + IMT n.d. http://www.cityenergyproject.org/ (accessed May 3, 2018).

[3] U.S. Energy Information Administration (EIA). Electric Power Monthly with data for January 2018. Washington, DC: 2018.

[4] Hong T. IEA EBC annexes advance technologies and strategies to reduce energy use and GHG emissions in buildings and

communities. Energy and Buildings 2018;158:147–9. doi:10.1016/J.ENBUILD.2017.10.028.

[5] Yoshino H, Hong T, Nord N. IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. Energy and Buildings 2017;152:124–36. doi:10.1016/J.ENBUILD.2017.07.038.

[6] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews 2012;16:3586–92. doi:10.1016/j.rser.2012.02.049.

[7] Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, et al. A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews 2018;82:1027–47. doi:10.1016/j.rser.2017.09.108.

[8] Zhao HX, Magoulès F. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews 2012;16:3586–92. doi:10.1016/j.rser.2012.02.049.

[9] Zheng Z, Zhuang Z, Lian Z, Yu Y. Study on Building Energy Load Prediction Based on Monitoring Data. Procedia Engineering 2017;205:716–23. doi:10.1016/j.proeng.2017.09.894.

[10] Kadir Amasyali, Nora M. El-Gohary. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews 2018;81:1192–205. doi:10.1016/J.RSER.2017.04.095.

[11] Kneifel J, Webb D. Predicting energy performance of a net-zero energy building: A statistical approach. Applied Energy 2016;178:468–83. doi:10.1016/J.APENERGY.2016.06.013.

[12] Sun Y, Huang G, Xu X, Lai AC-K. Building-group-level performance evaluations of net zero energy buildings with non-collaborative

controls. Applied Energy 2018;212:565–76. doi:10.1016/J.APENERGY.2017.11.076.

[13] Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings. Applied Energy 2017;197:303–17. doi:10.1016/j.apenergy.2017.04.005.

[14] Hsu D. Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Applied Energy 2015;160:153–63. doi:10.1016/j.apenergy.2015.08.126.

[15] Li W, Zhou Y, Cetin K, Eom J, Wang Y, Chen G, et al. Modeling urban building energy use: A review of modeling approaches and procedures. Energy 2017;141:2445–57. doi:10.1016/j.energy.2017.11.071.

[16] Reinhart CF, Cerezo Davila C. Urban building energy modeling - A review of a nascent field. Building and Environment 2016;97:196–202. doi:10.1016/j.buildenv.2015.12.001.

[17] Li C, Hong T, Yan D. An insight into actual energy use and its drivers in high-performance buildings. Applied Energy 2014;131:394–410. doi:10.1016/j.apenergy.2014.06.032.

[18] Xiufeng P, Tianzhen H, Mary ANN Piette. Improving building performance at urban scale with a framework for real-time data sharing. Proceedings of the Symposium on Simulation for Architecture & Urban Design 2013:221.

[19] Fonseca JA, Schlueter A. Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts. Applied Energy 2015;142:247–65. doi:10.1016/j.apenergy.2014.12.068.

[20] Oliveira MJN, Ao P, Brito MC, Oliveira Panão MJN. Modelling aggregate hourly electricity consumption based on bottom-up building stock. Energy & Buildings 2018;170:170–82. doi:10.1016/j.enbuild.2018.04.010.

[21] Kalogirou S, Neocleous C, Schizas C. Building Heating Load Estimation Using Artificial Neural Networks. Proceedings of the 17th International Conference on Parallel Architectures and Compilation Techniques 1997:1–8.

[22] Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings. Applied Energy 2017;197:303–17. doi:10.1016/j.apenergy.2017.04.005.

[23] Hsu D. Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data. Applied Energy 2015;160:153–63. doi:10.1016/j.apenergy.2015.08.126.

[24] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy 2014;123:168–78. doi:10.1016/j.apenergy.2014.02.057.

[25] Han Y, Taylor JE, Pisello AL. Toward mitigating urban heat island effects: Investigating the thermal-energy impact of bio-inspired retro-reflective building envelopes in dense urban settings. Energy and Buildings 2015;102:380–9. doi:10.1016/J.ENBUILD.2015.05.040.

[26] Pisello AL, Castaldo VL, Taylor JE, Cotana F. Expanding Inter-Building Effect modeling to examine primary energy for lighting. Energy and Buildings 2014;76:513–23. doi:10.1016/J.ENBUILD.2014.02.081.

[27] Pisello AL, Taylor JE, Xu X, Cotana F. Inter-building effect: Simulating the impact of a network of buildings on the accuracy of building energy performance predictions. Building and Environment 2012;58:37–45. doi:10.1016/J.BUILDENV.2012.06.017.

[28] Han Y, Taylor JE, Pisello AL. Exploring mutual shading and mutual reflection inter-building effects on building energy performance. Applied Energy 2017;185:1556–64. doi:10.1016/J.APENERGY.2015.10.170.

[29] Han Y, Taylor JE. Simulating the Inter-Building Effect on energy consumption from embedding phase change materials in building envelopes. Sustainable Cities and Society 2016;27:287–95. doi:10.1016/J.SCS.2016.03.001.

[30] Building Technology and Urban Systems Division at Lawrence Berkeley National Laboratory; City Building Energy Saver n.d. https://citybes.lbl.gov/.

[31] Chen Y, Hong T. Creating Building Datasets for CityBES 2017.

[32] Chen Y, Hong T, Piette MA. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. Applied Energy 2017;205:323–35. doi:10.1016/J.APENERGY.2017.07.128.

[33] Chen Y, Hong T. Impacts of building geometry modeling methods on the simulation results of urban building energy models. Applied Energy 2018;215:717–35. doi:10.1016/j.apenergy.2018.02.073.

[34] Fonseca JA, Nguyen TA, Schlueter A, Marechal F. City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts.

Energy and Buildings 2016;113:202–26. doi:10.1016/j.enbuild.2015.11.055.

[35] Felsmann C, Robbi S, Eckstädt E. Reduced Order Building Energy System Modeling in Large-Scale. 13Th Conference of International Building Performance Simulation Association 2013:1216–23.

[36] Heidarinejad M, Mattise N, Dahlhausen M, Sharma K, Benne K, Macumber D, et al. Demonstration of reduced-order urban scale building energy models. Energy and Buildings 2017;156:17–28. doi:10.1016/j.enbuild.2017.08.086.

[37] Zhao F, Lee SH, Augenbroe G. Reconstructing building stock to replicate energy consumption data. Energy and Buildings 2016;117:301–12. doi:10.1016/J.ENBUILD.2015.10.001.

[38] U.S. Department of Energy. Commercial Prototype Building Models. Building Energy Codes Program 2016:4–7.

[39] Deru M, Field K, Studer D, Benne K, Griffith B, Torcellini P, et al. U.S. Department of Energy Commercial Reference Building Models of the National Building Stock. Golden, CO (United States): 2011. doi:10.2172/1009264.

[40] Mastrucci A, Baume O, Stazi F, Leopold U. Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam. Energy and Buildings 2014;75:358–67. doi:10.1016/j.enbuild.2014.02.032.

[41] Caputo P, Costa G, Ferrari S. A supporting method for defining energy strategies in the building sector at urban scale. Energy Policy 2013;55:261–70. doi:10.1016/j.enpol.2012.12.006.

[42] Yang Z, Roth J, Jain RK. DUE-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and

stochastic frontier analysis. Energy and Buildings 2018;163:58–69. doi:10.1016/J.ENBUILD.2017.12.040.

[43] Gaitani N, Lehmann C, Santamouris M, Mihalakakou G, Patargias P. Using principal component and cluster analysis in the heating evaluation of the school building sector. Applied Energy 2010;87:2079–86. doi:10.1016/J.APENERGY.2009.12.007.

[44] Deb C, Lee SE. Determining key variables influencing energy consumption in office buildings through cluster analysis of pre- and post-retrofit building data. Energy and Buildings 2018;159:228–45. doi:10.1016/j.enbuild.2017.11.007.

[45] European Commission. Commission Delegated Regulation (EU) No 244/2012 of 16 January 2012 supplementing Directive 2010/31/EU of the European Parliament and of the Council on the energy performance of buildings by establishing a comparative methodology framework for calculating. 2012.

[46] Arambula Lara R, Pernigotto G, Cappelletti F, Gasparella A. Energy audit of schools by means of cluster analysis. Energy and Buildings 2015;95:160–71. doi:10.1016/J.ENBUILD.2015.03.036.

[47] Gaitani N, Lehmann C, Santamouris M, Mihalakakou G, Patargias P. Using principal component and cluster analysis in the heating evaluation of the school building sector. Applied Energy 2010;87:2079–86. doi:10.1016/J.APENERGY.2009.12.007.

[48] Tardioli G, Kerrigan R, Oates M, O'Donnell J, Finn DP. 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;140:90–106. doi:10.1016/J.BUILDENV.2018.05.035.

[49] ENERGY.GOV. Buildings and the Grid 101: Why Does it Matter for Energy Efficiency 2017. https://energy.gov/eere/buildings/articles/buildings-and-grid-101-why-does-it-matter-energy-efficiency.