An occupancy prediction model for campus buildings based on the diversity of occupancy patterns

Abstract: The complexity and randomness of occupancy often lead to the deviation between simulated energy and measured energy. It is of great significance to select appropriate approaches for occupancy prediction. This study applied Gaussian distribution model to fit occupancy of three functional buildings within a campus. Abandoned the average occupancy level widely adopted in exiting researches, a novel classified approach considering the diversity of occupancy patterns was proposed. The resulting Gaussian curves presented a better fitting performance for stable changes of occupancy. Although the sudden increase and decrease of occupancy greatly affects the prediction accuracy, the occupancy prediction error based on Gaussian distribution can still be controlled within ±15%. The energy of the case building was obtained by superposing simulated energy of each occupancy pattern. Rather than traditional method of averaging, the classified methodology eliminated the simulation errors produced from the inhomogeneous and stochastic occupancy. The exploration of changing regularity of energy affected by occupancy and time periods with energy saving potential were achieved by integrating energy data with occupancy data. And the detection of the degree and occurrence times of peak occupancy of various occupancy patterns of rooms provided a support for rational load distribution.

Keywords: Gaussian distribution, occupancy pattern, building simulation, energy saving.

1. Introduction

In current society, buildings are major energy consumers and their energy demand has significant impact on sustainable development (Han, Zhou, & Luo, 2015). To deliver energy efficient solutions without affecting occupant comfort, dynamic building performance simulation is widely used during the design stage of buildings to assist designers make decisions under various design scenarios (Y. Zhang, Bai, Mills, & Pezzey, 2018). When using simulation packages, however, a significant difference between the design expectation and actual operation has been realized once the building is in use, and such difference has been referred as performance gap (Schakib-Ekbatan, Çakıcı, Schweiker, & Wagner, 2015). Climate, building envelope, building energy service system, indoor design criteria, building operation and maintenance and occupant behavior were both considered as possible factors affecting performance gap(Yoshino, Hong, & Nord, 2017). The parameters related to the physical characteristics of the building are easy to determine, such as building size, orientation, construction material, Heating, Ventilation, and Air Conditioning (HVAC) system size and type, etc(Shen, Newsham, & Gunay, 2017). However, some time-varying parameters, such as occupant behavior input, are relatively difficult to predict. Occupant behavior has been regarded as one of the most significant considerations for building and system design, and occupancy has been suggested as a critical contributor in energy prediction model(Yan et al., 2015). It is practical to discover the current state of occupancy in many applications such as lighting control, however, for some applications such as ideal thermal environment control may require longer prediction horizons.(Huchuk, Sanner, & O'Brien, 2019). Therefore, providing a comprehensively and reliably long-term occupancy prediction model is still under development. Occupancy is a periodic time series, and many time-independent and non-linear models have been developed to analyze and predict the sensing occupancy(Candanedo & Feldheim, 2016). These models can be generally categorized as deterministic model, stochastic model and machine learning. Deterministic model could achieve the establishment of predictable and repeatable environment by averaging the diversity of individuals, space and time, including linear regression model, logical regression model, time series and Bayesian distribution(Gaetani, Hoes, & Hensen, 2016). Markov chain and Monte Carlo method were pointed to be the most commonly used stochastic models(Stazi, Naspi, & D'Orazio, 2017). And machine learning method included Artificial neural network (ANN), Support vector machine (SVM), Classification and Regression Trees (CART), etc.(Wang, Chen, & Hong, 2018).

During the last two decades, various occupancy model have been developed to mimic the randomness and diversity of occupants and generate stochastic occupancy profiles for building performance simulation(J. Li, Yu, Haghighat, & Zhang, 2019). So far, Markov chain and its various derivation were the most widely used in the process of occupant behavior simulation. Dong(Dong et al., 2010) extracted relevant occupancy information from CO_2 acoustic sensors and passive infrared sensors (PIR) to estimate occupant number in an open-plan office, fused with three algorithms Support Vector Machines(SVM), Artificial Neural Network(ANN) and Hidden Markov model(HMM). The estimation performance of HMM is similar to that of ANN but outperformed ANN in the description of occupancy presence profile due to its ability to discount sudden brief changes of occupancy level and maintain constant during static occupancy periods. In 2014, Dong(Dong & Lam) further developed a unique hidden Markov model based on Gaussian mixture model and applied this improved model into HVAC system control. Compared with the conventional schedule, 30.1% energy reduction in the heating season and 17.8% energy reduction in the cooling season have been realized. Given the ability of Autoregressive Hidden Markov Model(ARHMM) to establish correlations among the observed variables, Han(Z. Han, Gao, & Fan, 2012) developed a model to estimate the occupancy pattern based on the features extracted from a wireless sensor network equipped with inexpensive PIR, CO2 sensors, RH sensors, air velocity sensors and globe thermometer in a building. The results were compared with that obtained from Hidden Markov Model(HMM) and Support Vector Machines(SVM), which indicated that the ARHMM performed better than the other two methods with an average estimation accuracy of 80.78%. Wang(Wang, Chen, & Song stressed the time-series and stochastic characteristics of detected signals and proposed a novel Dynamic Markov Tim-Window Inference(DMTWI) model to predict reliable occupancy. When comparing the proposed approach with another two conventional Auto-Regressive Moving Average(ARMA) model and Support Vector Regression(SVR) model, similar x-accuracy (when the error of a prediction less than x occupants, the prediction regards as correct) with higher tolerance presented, however, for x-accuracy with lower tolerance, DMTWI performed bests for the weekday and weekend day and SVR performed bests for holiday. Salimi (Salimi, Liu, & Hammad, 2019) developed a time-dependent inhomogeneous Markov chain to predict occupancy of an open-plan office based on real occupancy patterns data. High accuracy results of occupancy patterns prediction as 86% and 68% on average for the lighting and HVAC systems control respectively indicated an acceptable performance of proposed Markov chain in distinguishing the temporal behavior of different occupants. However, this progress may be just limited to open-plan offices with occupant number no more than four, otherwise the prediction accuracy of occupancy will decrease significantly.

Unlike machine action, occupants behave differently to given set of circumstances.

In Candanedo's study regarding to the occupancy schedule definition based on temperature, humidity, light, and occupancy measurement, they noted that the occupancy pattern were highly varied day-to-day(Candanedo, Feldheim, & Deramaix, 2017). Exploration of the diversity of occupancy patterns from big data streams will allow for a better understanding of energy usage in buildings(Nguyen & Aiello, 2013). In recent decades, Machine learning method and data mining techniques had been regarded as a powerful tool to identify unsuspected relationships and to summarize the data in novel ways based on large observation datasets. Although data mining techniques were largely applied to research fields such as marketing, medicine, biology, engineering, medicine, and social science, their applications for building occupant behavior prediction and energy consumption simulation were still in elementary phases. However, highly effective technique of data mining will facilitate the exploration of occupant-building interaction(Wymelenberg, 2012). In this context, Yu tested several data mining technique from 2010 to 2012 as 'decision tree method'(Z. Yu, Haghighat, Fung, & Yoshino, 2010), 'cluster analysis'(Z. Yu, Fung, Haghighat, Yoshino, & Morofsky, 2011), and 'association rule mining'(Z. Yu, Haghighat, Fung, & Zhou, 2012), successively, for examining associations and correlations between building operation and occupant behavior data. The results obtained could help to prioritize efforts at modification of occupant behavior for energy conservation and to provide guidance for modeling of occupant behavior in numerical simulation. Simona(D' Oca & Hong, 2015) applied a three-step data mining framework as decision tree model-rule induction algorithm-cluster analysis along with open source data mining program RapidMiner to excavate occupancy patterns of 16 office spaces located in Germany. The application of this proposed framework on different data sets facilitated to enhance the robustness of the patterns description and prediction of individual or group energy-related behavior in office buildings. Meyn(Meyn, Surana, Lin, Oggianu, & Frewen, 2009) introduced a sensor-utility-network(SUN) algorithms to analyze and estimate occupancy in buildings using sensor measurements from diverse sources, models of facility use, and historical data. Compared with the naive approach that relies solely on flow measurements, the average estimation error based on SUN estimator at the building level reduced from 70% to 11%. Yu(T. Yu, 2010) applied genetic programming algorithm(GPA) to learn the behavior of an occupant in single-office room based on sensor data. The learned rules could be used into a building simulation tool to estimate the energy demand of a building in many areas as L-HVAC system, office appliances and the window opening activities. Hailemariam (Hailemariam, Goldstein, Attar, & Khan, 2011) equipped an office workspace with a tandem of heterogeneous sensor array to produce a real-time occupancy data, and applied Decision Trees classification to distinct the relative accuracy of various occupancy detection options. Having improved occupant detection accuracy from 97.9% to 98.4%, it was concluded that Decision Trees may well outperform simple thresholds when applied to multiple features derived from a single motion sensor.

The comparison of prediction accuracy of Markov chain model(including its derivation) and various machine learning algorithms in existing researches has been concluded and listed in Table 1.

Table 1. Accuracy comparison of occupancy prediction in existing researches

Reference	Modeling method (Markov chain)	Type of testing room	Accuracy (%)
Dong et al.(2010)	Hidden Markov chain	Multiple-office	73
Z. Han et al.(2012)	Autoregressive Markov chain	Multiple-office	80.78
Wang, Chen, &Song(2017)	Dynamic Markov chain	Multiple-office	80
Salimi et al. (2019)	Inhomogeneous Markov chain	Multiple-office	84
Reference	**Modeling method (Machine learning)**	**Type of testing room**	**Accuracy (%)**
Meyn et al. (2009)	Sensor utility network	Multiple-office	89
T. Yu(2010)	Genetic programming algorithm	Single-office	80~83
Hailemariam, Goldstein, Attar, &Khan(2011)	Decision tree algorithm	Multiple-office	98.4

However, researchers generally tend to take the 'average level' of occupancy data or a 'typical occupant' for modelling and building simulation without considering the diversity between occupants regarding to their behavior and presence, which will be bound to cause excessively peak-load prediction and irrational load distribution(Nägele, Kasper, & Girod, 2017). Based on the above consideration, Buttitta(Buttitta, Turner, Neu, & Finn, 2019) developed a new methodology to define occupancy-integrated archetypes and adopted data mining clustering techniques to embed occupant behavior profiles in these archetypes. When applying the occupancy-integrated archetypes rather than uniform occupancy archetypes into energy simulation, a thirty percentage of difference regarding to the heating demand had been discovered. Li (Z. Li & Dong, 2018) provided a unique data set containing the occupancy with different pattern varieties and explored the predictive power of Markov chains by different temporal scenarios as 15-min ahead, 30-min ahead, 1-h ahead and 24-h ahead. The excellent prediction performance indicated that it would be an applicable solution to implement such adaptive occupancy models for integrated predictive controls that handle multiple building optimization. Capozzoli(Capozzoli, Piscitelli, Gorrino, Ballarini, & Corrado, 2017) optimized the HVAC system according to the actual arrival and exit time points of occupants in an office building, by displacing the employees with the most similar occupancy patterns to the same thermal zone and performing the characteristic of the typical occupancy profiles of each sub-zone. The proposed methodology has proved to be able to draw low-cost real-life management solutions especially for buildings without intermittently occupancy and the energy saving of the HVAC system accounted for 14% in comparison to an occupancy-independent operation schedule.

However, this methodology of occupancy patterns classification was mostly applied to the office rooms of commercial buildings. It is still not clear how this methodology can be applied to the campus buildings, where occupancy dynamics are relatively regular but differ in various functional buildings and even in rooms with various occupancy patterns. This will cause differences in the time duration with energy saving potential between various buildings or among various rooms within the campus.

In Davide's study, the energy saving could reach to 6%~9% for the whole residential building, and 3%~13% for the individual apartment in case of vacancy(Bionda & Domingo-Irigoyen, 2017). In Oldewurtel's study, the simulation results of their proposed homogeneous occupancy patterns showed a 34% saving potential for the case of 5-day and 10-day intervals of average vacancy and occupancy for common office buildings in Switzerland equipped with integrated room automation facilities(Oldewurtel, Sturzenegger, & Morari, 2013). Compared with residential buildings and common office buildings, etc., the occupation of campus buildings are mostly limited by class time or office time, or else the great vacancy of rooms and a huge waste of energy will be resulted. Furthermore, during which time period would campus buildings with different functions exist great energy saving potential? This will also be worth deeply pondering and exploring.

Though the occupancy state can be observed directly in the time series of Markov chain, the future occupancy state only depends on the current state but not on the past state(Yang, Santamouris, & Lee, 2016). Instead, the modelling of occupancy model based on historical data profiles may significantly improve the prediction accuracy(Roselyn et al., 2019). In addition, the occupation and movement strongly depend on the time point, which caused a difficulty to obtain a detailed occupancy schedule while investigating and testing. Chen(Chen & Soh) proposed and compared different occupancy models (including inhomogeneous Markov chain and multivariate Gaussian) and data mining approaches (autoregressive integrated moving average, artificial neural network and support vector regression) to predict regular occupancy level in multi-occupant commercial buildings. In comparison, the occupancy prediction performance of stochastic occupancy models were more limited than that of data mining approaches. Data mining approaches could effectively determine the correlation between the previous and subsequent occupancy state. Hailemariam(Hailemariam, Goldstein, Attar, & Khan, 2011) measured several attributes of environment as light, sound, CO_2 level, power use, and motion by embedding a number of low-cost sensors of different types into the cubicle furniture, and performed classification to explore the relationships between various sensors and deduce the occupancy of the workspace at any given time through Decision Trees method. This individual feature achieved 97.9% accuracy of occupancy detection considering a simple threshold and further 98.4% accuracy combining multiple motion sensor features with a Decision Tree. Though the prediction and detection accuracy of occupancy could be significantly improved, machine learning method requires stricter sample data and the data should be collected by the combination of various sensors in the process of occupancy prediction, causing the increased cost in the meantime of improving the prediction accuracy(Ahmad & Chen, 2020).

Recently, Gaussian distribution have been frequently proposed to be used in the field of atmospheric diffusion assessment due to its special advantages (i.e. simple structure, strong operability, and flexible nonparametric inference ability)(Cao, Cui, Chen, & Chen, 2020). Gaussian function provides a flexible and accurate model for a wide range of probability distribution(J. Zhang, Yan, Infield, Liu, & Lien, 2019). A similarly multiple distribution of occupancy just appears throughout the whole cycle of a day. When applying Gaussian distribution into occupancy modelling, the universality of the occupancy profiles could be improved and the complexity of occupancy data could be eliminated. In addition, the differences of occupancy between various time series of a day could be judged by multiple Gaussian distribution.

Based, a Gaussian distribution was proposed to fit the occupancy for three functional building types, namely dormitory building, lecture building and office building within a campus unit in this study. In order to promote the general application of occupancy profiles of different campus buildings, the diversity modeling of occupancy patterns applied to each case building was mainly focused. Prior to this focus, the occupancy and vacancy time nodes of rooms for case buildings were collected and then clustered to determine categories of occupancy patterns. The fitting performance of Gaussian distribution for occupancy prediction was evaluated by multiplying the occupancy data of clustered patterns with their corresponding proportion. The two simulated energy adopting the methodology of classification at the room-scale and average level at the building-scale were together compared with the measured energy. The contribution of comparison lied in the verification of the feasibility of two methodologies. Combined the occupancy data and energy data, the explorations of time duration with great energy saving potential and theoretical basis for load distribution of rooms with diverse occupied pattern were realized.

The rest of this study was organized as follows: Section 2 introduced the methodology regarding to occupancy data collection, occupancy pattern clustering, and model development. Section 3 provided detailed research results and corresponding discussion, including the clustering of occupancy patterns, the fitting performance of Gaussian distribution, and verification of predicted occupancy level. The simulation and validation of energy consumption was presented in Section 4. Finally, the conclusion and limitation of the current research study were discussed in Section 5.

2. Methodology

In this study, the research framework was mainly established by the following four steps: 1) Classifying of occupancy patterns of rooms by clustering the working schedule for dormitory building and the occupancy and vacancy time nodes for lecture/office building; 2) Developing Gaussian distribution for occupancy prediction at the classified room-scale and unclassified building-scale, respectively; 3) Inputting two types of occupancy profiles obtained from the accumulation of clustered patterns and the average level of case buildings to simulate energy; and 4) Comparing the simulated energy with the actual energy to verify the application of the classified methodology and detect the time duration with great energy saving potential. Detailed framework of methodology adopted in this study have graphically depicted as Figure 1.

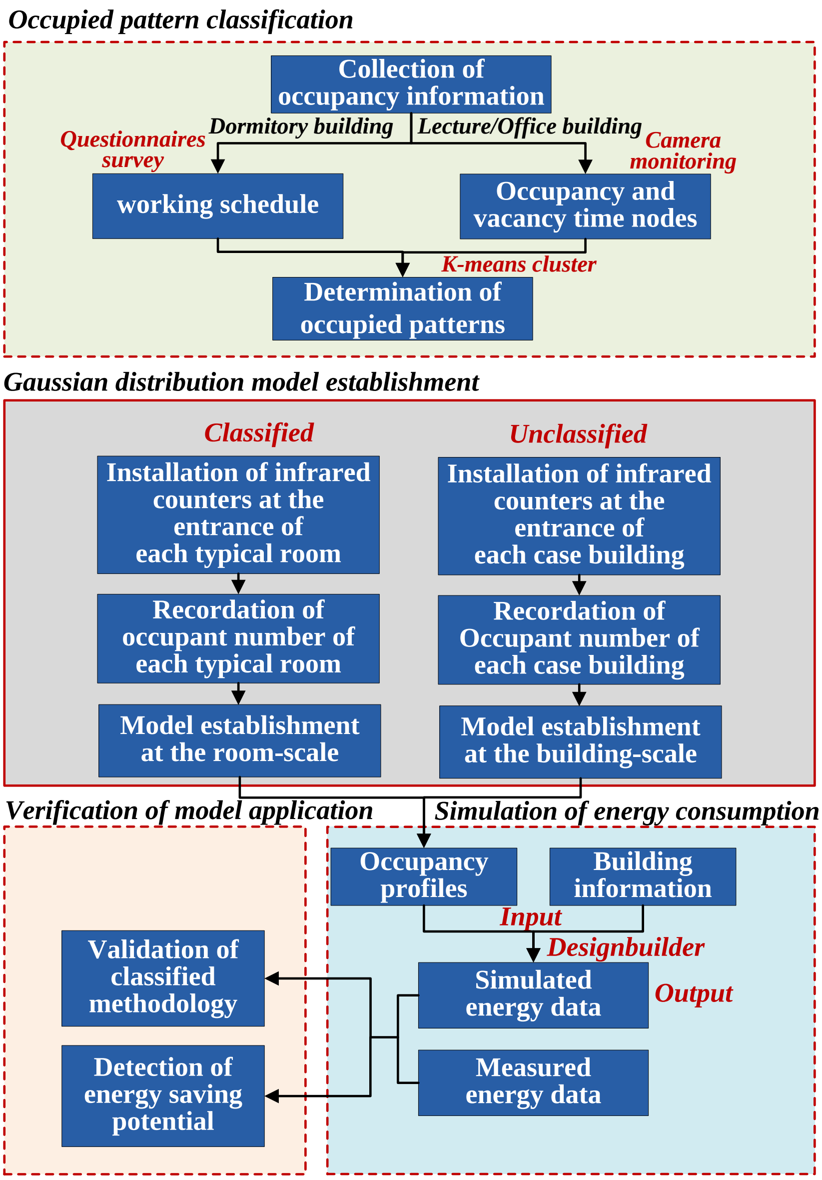


Figure 1. Framework for model development

2.1 Data collection

2.1.1 Case study buildings information

Three functional buildings namely as dormitory building, lecture building and office building within a campus located in Tianjin, China, were selected in this study.

The lecture building and the office building were serviced by a centralized cooling and heating system. But for the dormitory building, only split air conditioners were installed. The three case buildings could respectively well represent living, working and learning environment within a campus. And they all have the north-south orientation. The basic building information, in terms of functions, area, operation time and utilization, have been listed as Table 2.

Table 2 Basic information about the case study buildings

Building type	Total floor area (m^2)	Number of rooms (-)	Maximum capacity of occupant for one room (-)	Opening time (-)	Room functions (-)
Dormitory building	3400	200	4	00:00 - 24:00	Accommodation 80% Laundry/Activity room 20%
Lecture building	12670	50	65	06:30 - 22:30	Professional classroom 80% Machine/Activity room 20%
Office building	3690	40	8(Teachers' office) 25(Graduates' office)	06:30 - 22:30	Teachers' office 80% Graduates' office 20%

The dormitory building was in operation all the time, providing a living environment mainly for undergraduate and graduate students. The students could get in and out of their dormitory building by swiping their ID cards at any time. However, the lecture building and office building were in operation from 6:30 am to 22:30 pm, strictly conformed to management mechanism. The lecture building provided an environment for teaching and self-studying. And the office building provided an environment for teachers doing office work and graduates conducting researches. The class time and office time were from 8:30 to 12:00 am and from 13:30~17:00 pm.

This study was conducted on the weekdays during the period from 4 June to 15 July, 2018, when all buildings within campus unit were in normal operation. Based on the variety of utilization, the complete database were divided into three parts as shown in Figure 2.

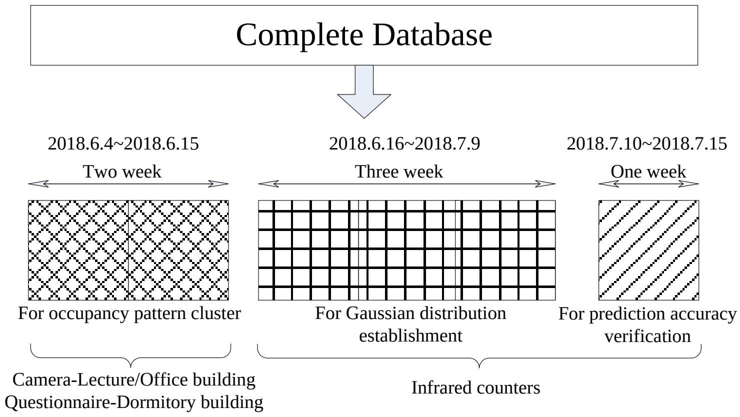


Figure 2. The details of complete database collection

The former two-week occupancy data of all rooms obtained by cameras and questionnaire were used to determine the occupancy patterns. The next three-week data recorded by infrared counters were used as training dataset for Gaussian distribution establishment, and the last one-week data were used for prediction accuracy verification. The concrete methods of data collection and data processing are described in section 2.1.2 and 2.1.3, respectively.

2.1.2 Determination of occupancy pattern

The data used for occupancy pattern clustering were collected from 4 June to 15 June, 2018. The collected occupancy and vacancy time nodes of all rooms were used to cluster and iterate until the clustering centers converge, and finally to determine the occupancy pattern. However, limited by experimental equipment, it may be not realistic to install related sensors for each room or real-time positioning system for each occupant. In order to obtain the occupancy and vacancy time nodes of each room, it is best to utilize the monitoring system equipped and controlled uniformly by administration. Secured the agreement of university institute and participants, the time nodes of all rooms within lecture building and office building were monitored and recorded by cameras and the scene under camera are shown as Figure 3. However, with privacy considerations, the camera monitoring system is unfeasible for the record of occupancy and vacancy information in the dormitory building, instead, the occupied regularity of dormitory rooms were obtained through questionnaires, focusing on the time nodes when occupants usually leave or return to the rooms. Eventually, a total of 150 valid questionnaires were surveyed, basically representing the working schedule of most dormitory rooms.

Although two-week data were collected, it was observed through image that the occupancy and vacancy time nodes of the same room were similar. In addition, the type of occupants were simple and mainly as teachers and students for such typical buildings as dormitory building, lecture building and office building within campus unit. And there existed no significant differences in the number of occupants for the rooms with similar using function restrained by the class and work arrangement, which could be ignored. The two-week data was enough to reflect the occupancy regularity of different rooms of these typical buildings, based on which the clustered results of occupancy pattern would be highly reliable.

Figure 3. The monitoring scenes under the cameras

After the occupancy and vacancy time nodes of each rooms for different building types have been obtained, K-means clustering method can be used to classify rooms with different occupancy patterns. It is a typical clustering algorithm based on distance. A closer the distance between any two objects represents a higher similarity. A cluster corresponding to a unique clustering center will be composed of these similar objects(Feng, Niu, Zhang, Wang, & Cheng, 2019). The key of K-means method is the determination of the clustering centers, k objects will be taken out randomly from the samples as the initial clustering centers, and they will be updated and iterated until the cluster criterion function converges based on Euclidean Distance, just as the clustering centers no longer changed. The mathematical expression of Euclidean Distance is shown as Equation (1~3).

$$\mathrm{d}(i, j) = \sqrt{(x_{i1} - x_{j2})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{ip} - x_{jp})^2} \quad (1)$$

$$i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \quad (2)$$

$$j = (x_{j1}, x_{j2}, \ldots, x_{jp}) \quad (3)$$

Where, i and j represent the objects described by p numerical attributes, d(i,j)

represents the Euclidean distance between object i and object j. In this study, object i and object j refer to any two occupancy patterns corresponding to different occupancy and vacancy time nodes.

The number of clusters was determined based on the actual situation, and the reasonability of the clusters was judged by evaluation indicators. Calinski-Harabaz index, as the most commonly applied indicator, represents the covariance between different clusters. The mathematical expression is shown as Equation (4).

$$CH(k) = \frac{tr(B_k)}{tr(w_k)} \bullet \frac{m-k}{k-1} \quad (4)$$

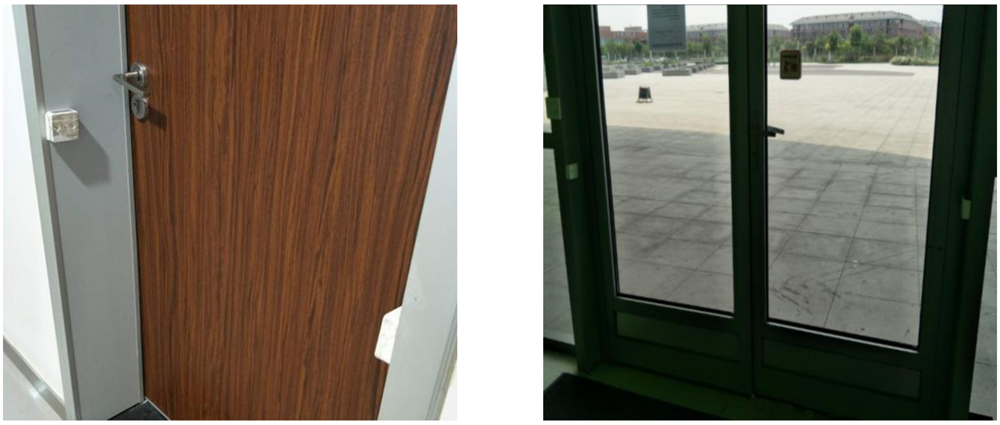
Where, m represents training samples, k represents the number of clusters, B_k represents the covariance matrix between clusters, w_k represents the covariance matrix of data, tr represents the trace of matrix. A larger $CH(k)$ value indicates a greater gap between clusters and a more effective clustering result.

2.1.3 Measurement of occupancy

The data used for model establishment and verification were collected from 16 June to 15 July, 2018. The occupant number was measured using infrared counters as Figure 4. The counter could effectively resist the illumination interference to reduce the counting errors. Its maximum records can reach to 99,999 times. However, one limitation of the counters lied in that occupants could be identified accurately only when they enter or leave the room or building in turn. Only one occupant could be sensed and recorded even if two or more occupants enter or leave room simultaneously. The entrance and exit width of tested rooms and case buildings are just limited to one occupant entering or leaving the room simultaneously, eliminating the error caused by the limitation.

Considering the variety of different occupancy patterns at the room-scale, infrared counters were installed at the entrance and exit of typical rooms representing clustered patterns to record the hourly occupant number as Figure 5 (a). Considering the average level of occupancy at the building-scale, infrared counters were installed at the entrance and exit of each case building to record hourly occupant number as Figure 5 (b). The infrared counters will generate an ‘in data’ or an ‘out data’ and update every time someone enters or leaves the room or building. Data will be summarized every hour such that the hourly occupant number could be obtained.

Figure 4. Monitoring device for occupant number



(a) At the room-scale (b) At the building-scale

Figure 5. The installation of infrared counters

2.2 Model development for occupancy prediction

This study adopted Gaussian distribution to fit the occupancy profiles. The function of Gaussian distribution, expressed as a density function, was defined as Equation (5).

$$N(x;\mu;\sigma)=\frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{5}$$

Taking the diversity of individual rooms into account, the cumulative occupancy of the case building could be expressed as Equation (6).

$$occ(t_m)=\sum_{i=1}^{n}a_i\varphi_i N(x;\mu_i;\sigma_i) \tag{6}$$

Where, $occ(t_m)$ represents the occupancy of the targeted object m(case building); n represents the category of occupancy pattern for one case building; a_i represents the weight coefficient of i^{th} occupancy pattern for one case building, the total proportion of which is 1, $\sum_{i=1}^{N} a_i = 1$; $N(x; \mu_i; \sigma_i)$ represents the function of Gaussian distribution of the i^{th} occupancy pattern; φ_i represents correction coefficient of Gaussian function $N(x; \mu_i; \sigma_i)$; μ_i and σ_i represents the corresponding mean value and variance, respectively. The width of each occupancy peak could be reflected by σ_i. Each set of μ_i and σ_i were respectively sorted in order of time series of a day.

The Gaussian distribution model for diverse occupancy patterns will be developed by averaging the fitting results of the former twenty-four days, conducted as the training dataset. And the variety of occupancy patterns shall bring out the divergence of the Gaussian distribution. It is of great significance to pre-classify occupancy patterns for discrimination of different functional rooms and accuracy improvement of occupancy prediction.

2.3 Statistical indicator

After the Gaussian model was fitted, the last one-week data shown as Figure 1 were used to evaluate the prediction accuracy. Here, RMSE (Root Mean Square Error) is used as the statistical performance indicator. It refers to the arithmetic square root of the expected value of the square of the difference between the predicted and measured occupancy. The value of RMSE is calculated as Equation (7~8).

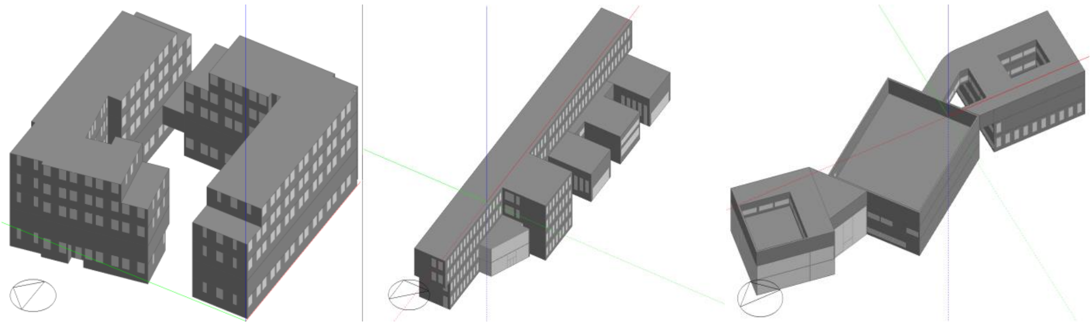
$$RMSE_{ave} = \frac{\sum_{d=1}^{n} RMSE_d}{n} \tag{7}$$

$$RMSE_d = \sqrt{\frac{\sum_{t=1}^{T} (occ_o(t) - occ_p(t))^2}{T}} \tag{8}$$

Where, $RMSE_d$ represents the root mean square error of one of verification days;n represents the number of verification days; $RMSE_{ave}$ represents the average $RMSE_d$ of verification datasets n; $occ_m(t)$ and $occ_p(t)$ represent the observed occupancy and predicted occupancy at time t, respectively; T represents the number of observed times for one verification day.

2.4 Simulation of energy consumption

The building energy consumption generally consists of two parts. One part refers to the fixed power unchanged by occupants such as the fire fighting and emergency lighting, maintaining open all the year round. Another part refers to the power affected directly by occupants, floating up and down as the changing occupancy(Ding, Wang, Wang, Han, & Zhu, 2019). In order to compare the influence of the occupancy profiles respectively obtained by classified and unclassified methodology on the energy simulation, only occupant-related part rather than the total energy was simulated by DesignBuilder. It is a comprehensive graphical interface simulation software based on the dynamic simulation program Energyplus. The physical models of the dormitory building, lecture building and office building were established respectively as Figure 6.



(a) Dormitory building (b) Lecture building (c)Office building

Figure 6. Physical models of three typical buildings

3. Results and discussion

3.1 Clustering of occupancy patterns

The occupancy patterns of different rooms varied even for the same building type. The detected occupancy and vacancy time nodes of rooms for different buildings were clustered by K-means method, respectively, judging the optimum number of clusters by Calinski-Harabaz index. The Calinski-Harabaz value corresponding to various cluster number were shown as Figure 7.

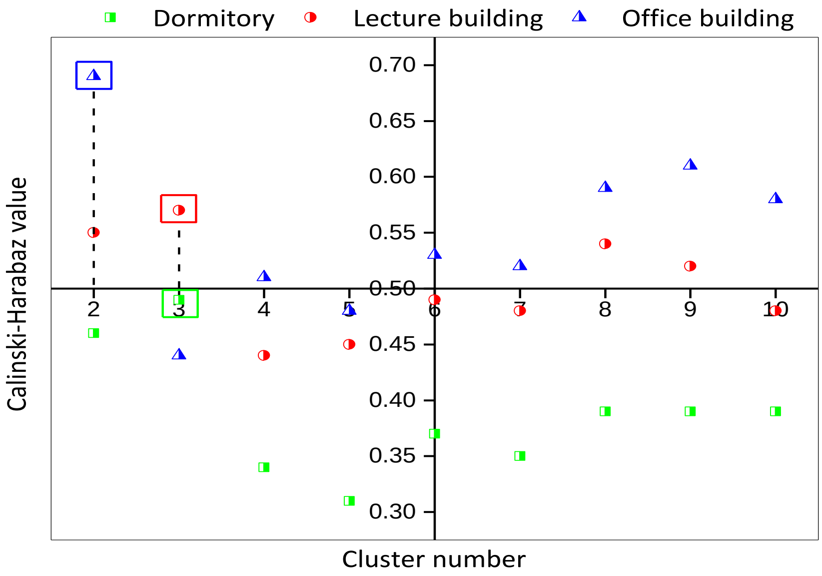
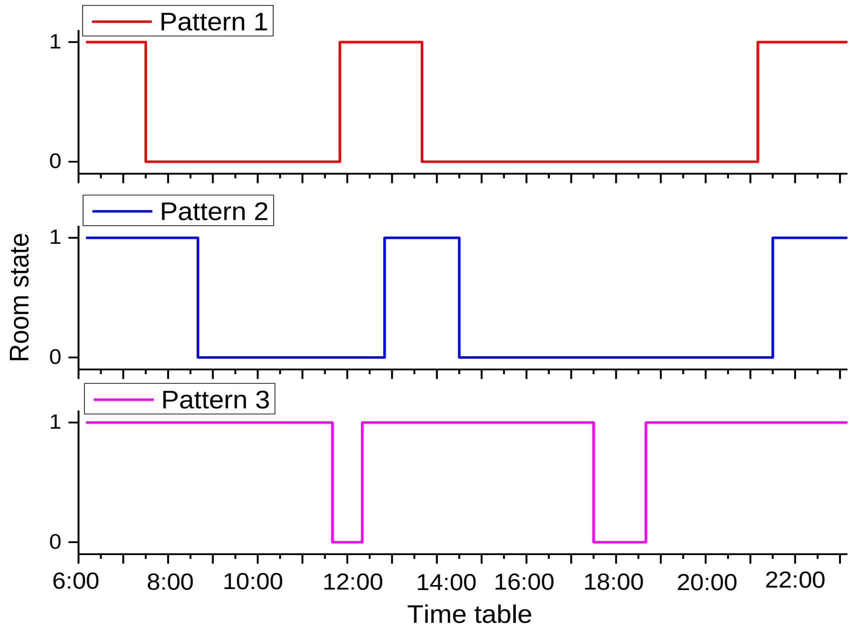
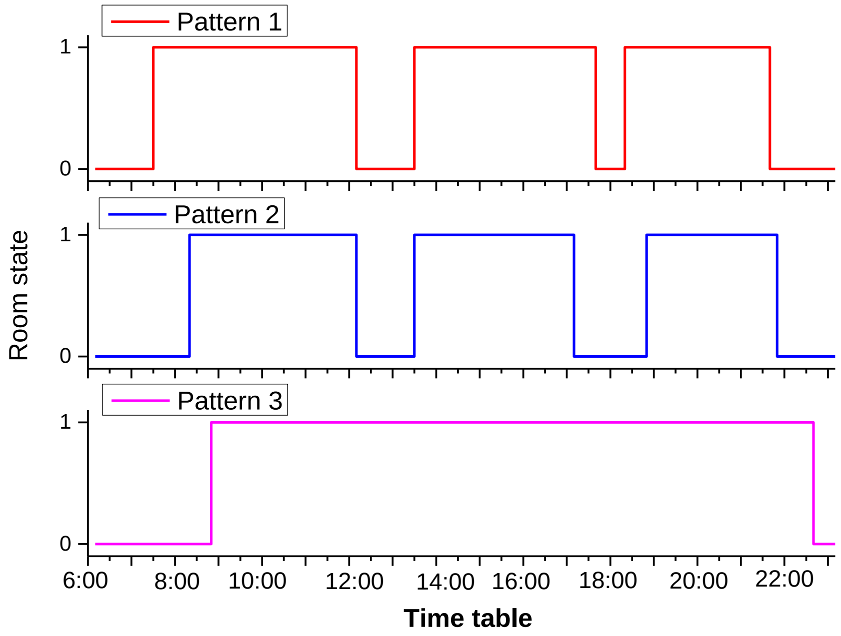


Figure 7. Calinski-Harbaz value corresponding to Cluster number

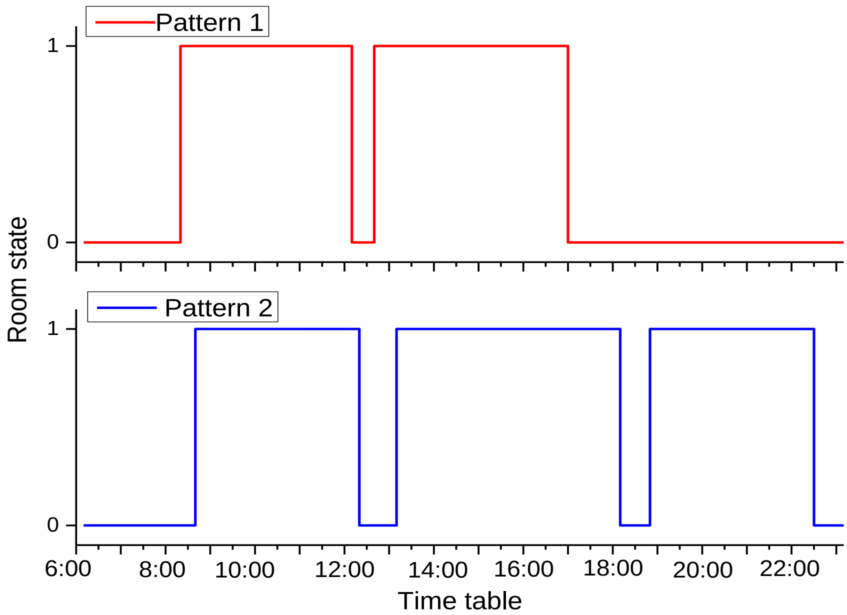
As a result, the Calinski-Harabaz value is the highest when the clustering room number of dormitory building and lecture building are both three, and the clustering room number of office building is two. Therefore, it is the optimum case to divide rooms of dormitory building and lecture building with different occupancy patterns into three categories, and to divide rooms of office building with different occupancy patterns into two categories. Only considering whether the rooms were occupied rather than the concrete occupancy, the room state was divided into two categories as 'occupied' and 'unoccupied'. In this study, the occupied state was defined as "1", and the unoccupied state was defined as "0". It can be seen from the using functions and operating mechanism of three campus buildings that the initial room state was '1' for dormitory building, and was '0' for lecture building and office building at the beginning of a day. Based on the diversity of room states corresponding to different time nodes, the occupancy patterns were classified and clustered by K-means method, and the corresponding occupancy and vacancy time nodes throughout a day of each clustering center obtained by multiple iterations for different building type were depicted as Figure 8, respectively. The proportions of each occupancy pattern were shown in Figure 9.



(a) Dormitory building



(b) Lecture building



(c) Office building

Figure 8. The occupancy and vacancy schedule of various occupancy patterns

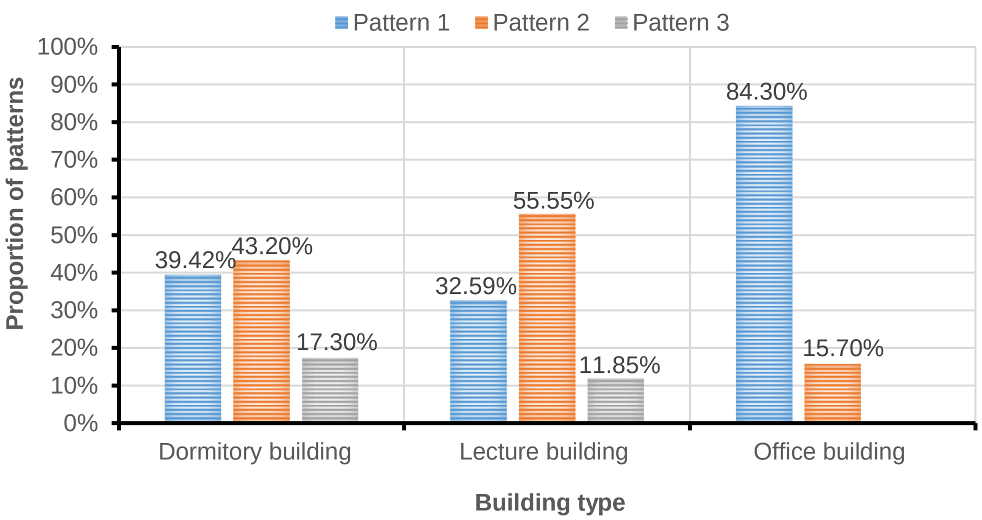


Figure 9. The proportion of each occupancy pattern for different building type

From Figure 8(a), the occupancy pattern 1 and pattern 2 of dormitory building slightly varied in the occupancy and vacancy time nodes. The room state of pattern 1 transformed from occupied to unoccupied at 7:30 am in the early morning and 13:40pm after noon break and transformed from unoccupied to occupied at 11:50 am back for noon break, and 21:10 pm back for sleep, symbolizing a dormitory room type occupied by early-hour keeper. In contrast, the occupancy and vacancy time nodes of pattern 2 integrally lagged behind that of pattern 1, which were mainly reflected in the time of leaving the room in the early morning, and returning to the room at nighttime, representing a dormitory room type occupied by a relatively late-hourly keeper. The two types of dormitory room were similarly occupied during two time periods for night sleep and noon break, respectively. From Figure 9, the dormitory room of occupancy pattern 2 accounted for the most, slightly higher than pattern 1, indicating a smaller part of occupants with early schedule than late schedule. However, the proportion of dormitory room of occupancy pattern 3 was the least just as 17.3%. This category of room had been occupied almost all day, representing a type lived in occupants used to house in their dorms.

From Figure 8(b), three similar time duration with occupied state appeared in occupancy pattern 1 and pattern 2 basically followed by the fixed time for class, representing the same classroom type mainly used for course teaching. But the three time duration still diversified in the time nodes of state transformation from occupied to unoccupied and from unoccupied to occupied. The occupancy time node in the early morning of pattern 1 was nearly 1*h* ahead of that of pattern 2. And the unoccupied state between the second and third occupied time duration of pattern 2 lasted 1h longer than pattern 1. Unlike the former two patterns, the third pattern of classroom had been occupied from early morning until nighttime, possibly representing an occupancy pattern type used for self-study unlimited by class time. The huge mobility and randomness of occupants in such pattern of classrooms caused the continuously occupied state even during lunch and dinner time periods. However, the proportion of occupancy pattern 3 was the least as 11.85%, indicating the classrooms of case lecture building were mostly intended for course teaching but less for self-studying.

From Figure 8(c), the office room of pattern 1 had been occupied for two time duration from 8:00 am to 12:00 am and from 12:40 pm to 17:00 pm, followed by office time. And the unoccupied time duration happened during lunch time and off-duty, respectively. This category of occupancy pattern was in line with the working schedule of teachers, representing an office environment provided for teachers handling office work. In comparison, the office room of occupancy pattern 2 had been occupied since 8:30 am until 22:30 pm except for lunch time and dinner time lasting for around one hour. The sustained occupied state of pattern 2 generally conformed to the working schedule of graduate students. Pattern 2 represented an office environment provided for graduate students studying or conducting researches. From Figure 9, the office rooms of occupancy pattern 1 accounted for a higher proportion than pattern 2, indicating the case office building was supplied mainly for teachers and secondarily for graduate students.

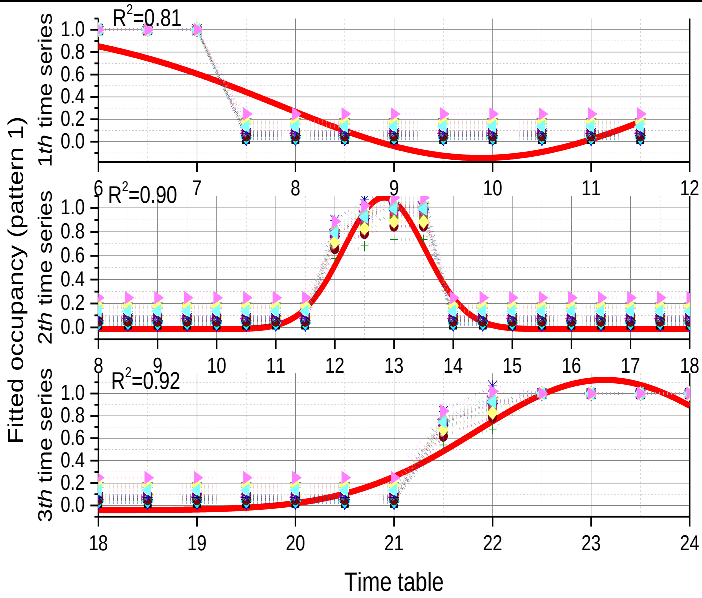
3.2 Fitting performance of Gaussian distribution

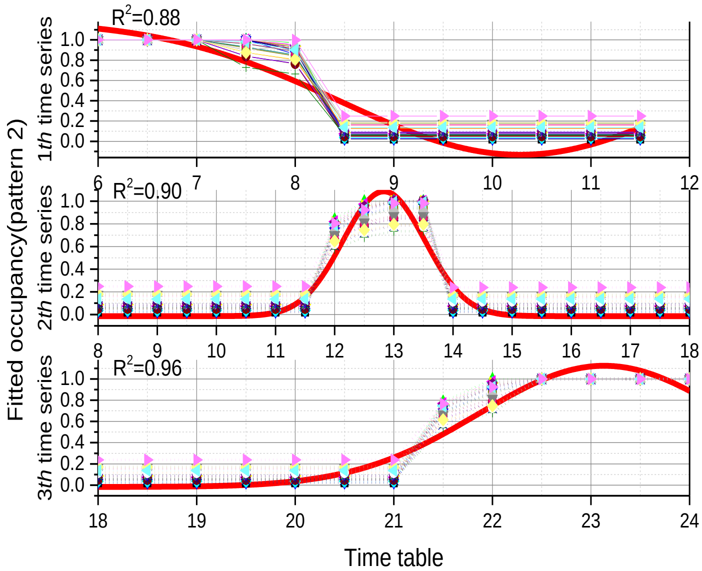
The training occupancy dataset introduced in section 2.2 were used to fit Gaussian curves of various occupancy patterns obtained above. C omplex model based on training dataset may cause results over-fitted in the meantime of improving prediction accuracy. A unique set of peak occupancy would be presented in a Gaussian curve according to the function expressed as Equation (5). Based on measured occupancy data, multiple occupancy peaks were discovered for occupancy patterns of case buildings, thus corresponding amount of time series were separated to fit the occupancy using Gaussian distribution. Table 2 showed the value and standard error of coefficients (φ_i, μ_i and σ_i) of each Gaussian distribution. And the corresponding Gaussian curves of dormitory building, lecture building, and office building were fitted as Figure 10 (a), 10 (b), and 10 (c), respectively.

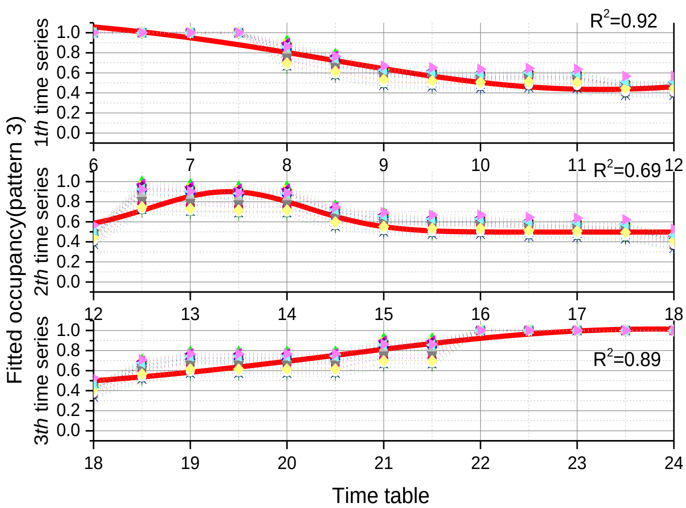
Table 2. The value and standard error of coefficients of Gaussian distribution

Building	Pattern	Time series	Coefficients [Value (Standard error)]		
			μ_i	σ_i	φ_i
Dormitory building	1	1	9.88(0.029)	2.04(0.37)	-0.76(0.09)
		2	12.83(0.064)	0.69(0.069)	1.87(0.029)
		3	23.13(0.17)	1.29(0.21)	3.60(0.097)
	2	1	10.28(0.028)	1.81(0.54)	-0.58(0.25)
		2	12.83(0.066)	0.69(0.071)	1.87(0.093)
		3	23.13(0.12)	1.27(0.14)	3.56(0.067)
	3	1	11.24(0.05)	2.79(0.099)	2.99(0.20)
		2	13.40(0.13)	0.80(0.15)	1.80(0.061)
		3	23.79(0.13)	3.23(0.16)	8.16(0.22)
Lecture building	1	1	10.18(0.18)	1.89(0.054)	5.01(0.30)
		2	15.28(0.14)	2.40(0.14)	5.16(1.1)
		3	19.91(0.09)	1.19(0.15)	1.04(0.03)
	2	1	10.28(0.14)	2.16(0.50)	4.64(0.24)
		2	15.85(1.12E-11)	0.028(1.90E-10)	0.48(1.75E-14)
		3	20.08(0.011)	0.99(0.15)	0.61(0.028)
	3	1	10.47(0.11)	1.62(0.19)	2.51(0.049)
		2	16.56(0.17)	1.15(0.24)	1.64(0.048)
		3	19.29(0.17)	1.67(0.27)	1.72(0.042)
Office building	1	1	10.33(0.17)	1.54(0.31)	4.50(0.2)
		2	15.37(0.13)	1.40(0.15)	4.08(0.1)
	2	1	10.35(0.094)	1.12(0.13)	2.74(0.088)

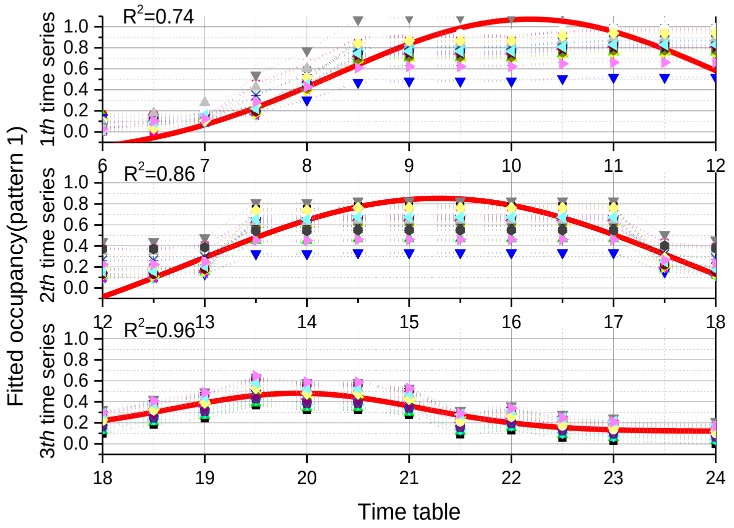
	2	15.73(0.093)	1.75(0.40)	4.29(0.3)
	3	20.70(0.072)	1.08(0.12)	2.59(0.077)

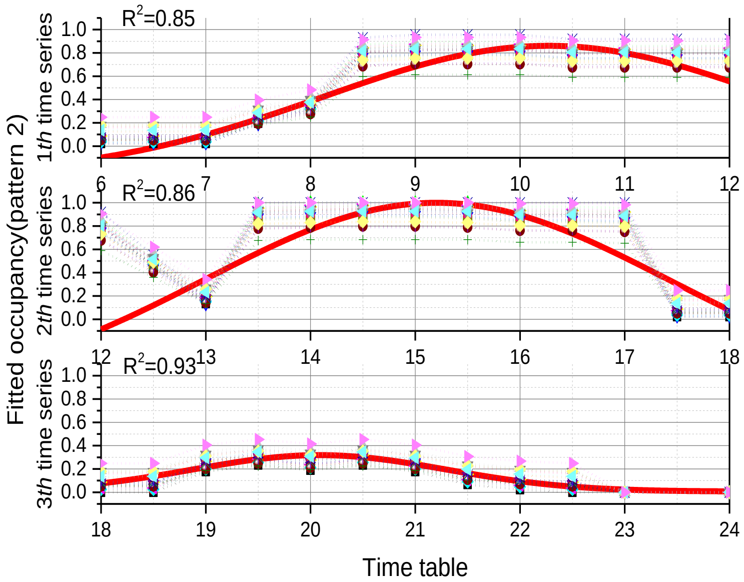


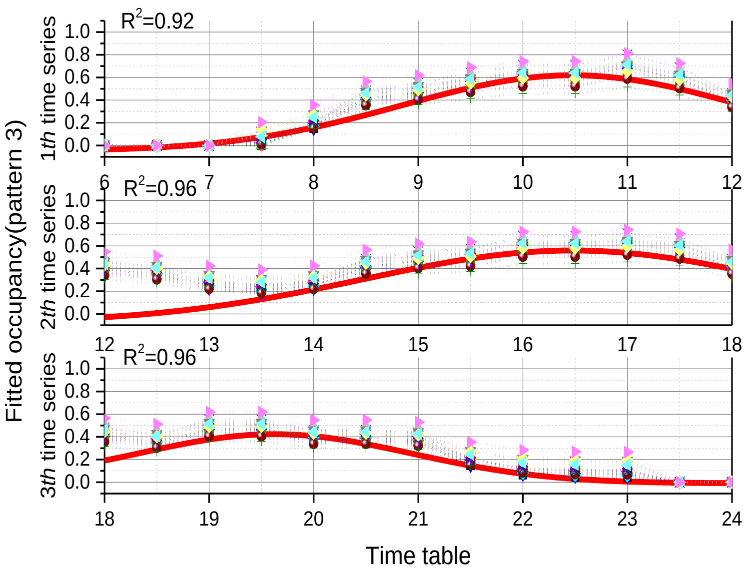




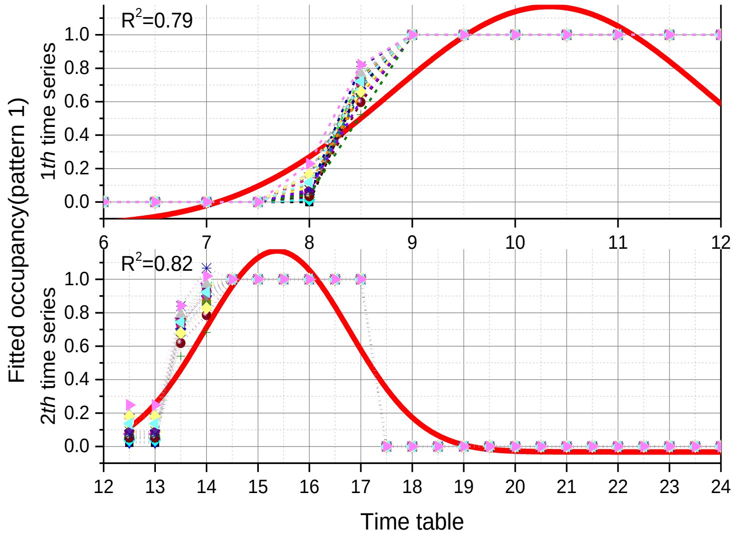
(a) Dormitory building

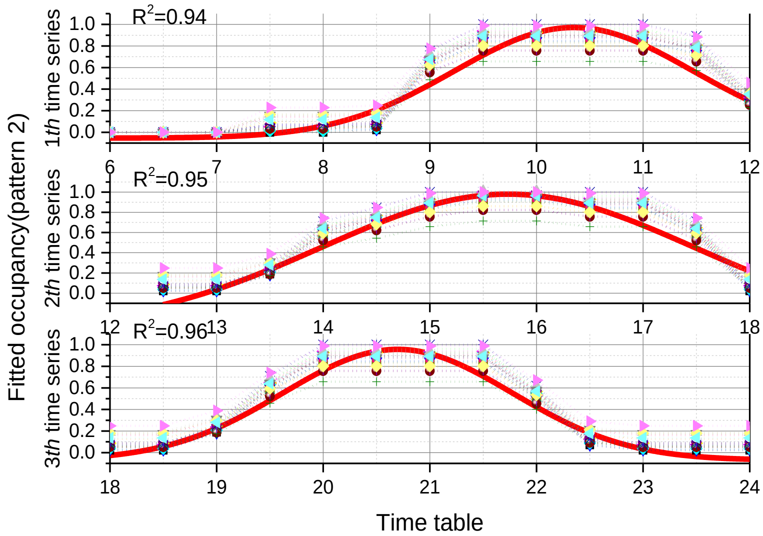






(b) Lecture building





(c) Office building

Figure 10. The fitting performance of Gaussian distribution for occupancy prediction

The occurrence time points and width of peaks of Gaussian curves could be clearly observed from the value of coefficients μ_i and σ_i as shown in Table 2. The standard errors of μ_i, σ_i and φ_i were both lower than 0.5, indicating better fitting performance of coefficients. Combined with the measured occupancy data of former 24 days shown as Figure 10, it can be seen that the occurrence time points and duration of occupancy peak for the same occupancy pattern corresponding to different days were almost exactly identical, and even more data would not make distinct influence on the fitting results of μ_i and σ_i. The amount of training dataset was enough to accurately fit the changing curve of occupancy of different time series.

It can be seen from Figure 10, three peaks of occupancy existed throughout the day in any other patterns of case buildings, except for office room of pattern 1 of office building with only two peaks. The three time periods as early morning, noon, and nighttime for break of dormitory building generally appeared peaks of occupancy. While for another two buildings, a day can be divided into three time periods taking lunch time and dinner time as two split points. A universal phenomenon is that the movement and activities of students are limited within campus area, causing a diversity of occupancy degree presented in classrooms and graduates' office rooms during above three time periods. In contrast, the movement and activities of teachers are more random and unlimited in the scope of campus, most of them tend to leave the campus for home or other places after office time, resulting in only two occupancy peaks presented in teachers' office room as pattern 1. The fitted Gaussian curves for any patterns exhibited smoothly ascending and descending changing process. The width of the Gaussian curves depended on the exiting time duration of measured peak and the changing intensity between the valley data and the peak data. Separated Gaussian model regularized the complexity and the noises of occupancy profiles and achieved better fitting performance with fewer coefficients. The fitted occupancy corresponding to each time point would exactly coincide with the measured data for the time duration with stable change tendency, just as emerged in occupancy pattern 3 of dormitory building, occupancy pattern 3 of lecture building, and occupancy pattern 2 of office building. The fitting goodness R^2 of above three patterns for any time series were both more than 0.90, representing an excellent fitting performance of Gaussian distribution for occupancy prediction. However, the sudden change of occupancy may cause fitted Gaussian curves a little delayed or advanced by measured data. Even though, such deviation had not affected the overall fitting performance of curves. The average prediction deviation of training days were shown as Figure 11. The phenomenon that errors exceeding zero represented lower fitted occupancy than measured data, on the contrary, the fitted occupancy was be higher than measured data when the error was below zero. The columns of error closer to zero could be identified as normal deviations caused by calculation no matter positive or negative.

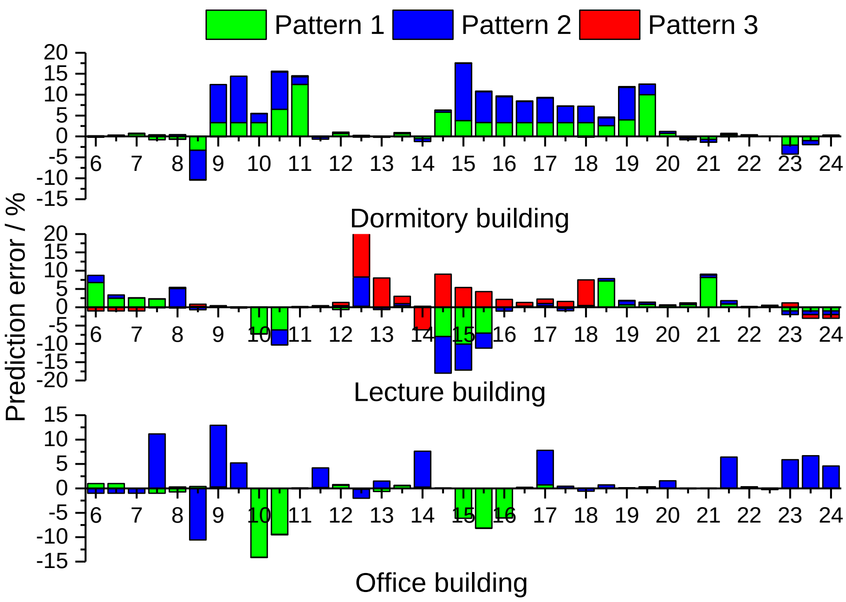
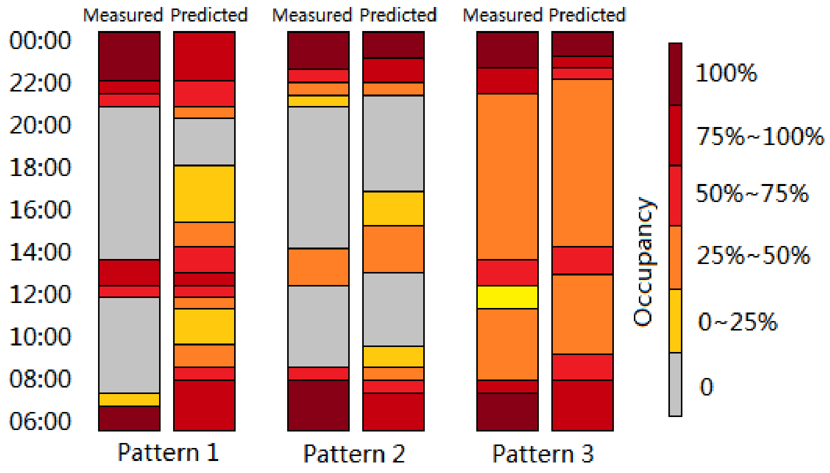


Figure 11. The deviation of occupancy prediction from the measurement

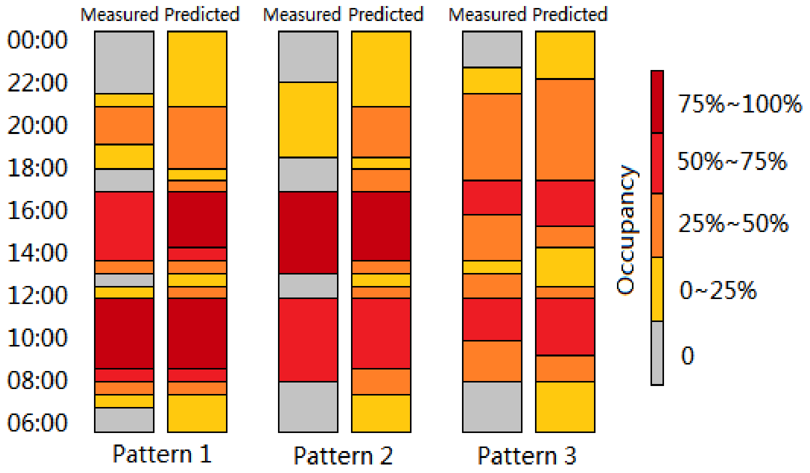
It can be seen from Figure 11, distinctly positive columns of error usually occurred when occupancy located at the valley, and gradually offset until occupancy turned to increase steadily. Conversely, distinctly negative error generally appeared at the peak of fitted Gaussian curves. Although prediction errors for sudden change between valley and peak occupancy were relatively higher, the overall errors were still controlled within ±15%. As is seen from Table 1, the prediction errors of most existing proposed occupancy models focused on multiple-office rooms were generally in the range of 10%~20%, thus 15% prediction error presented in this study was situated within an acceptable range.

3.3 Verification of predicted occupancy level

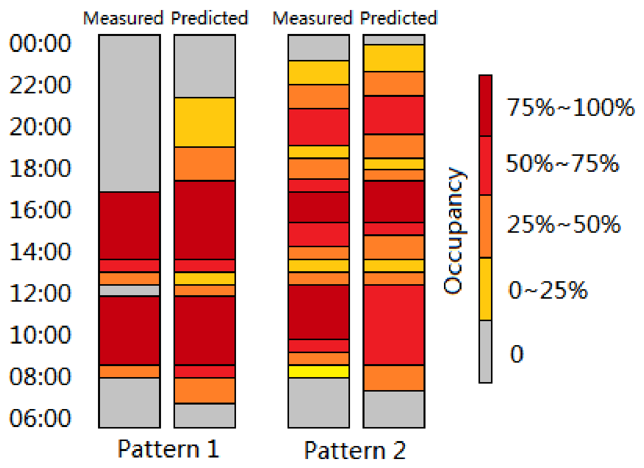
The later six days conducted as validation dataset were used to verify the application of fitted Gaussian distribution model. The average of measured and predicted data of the validation dataset expressed in the way of occupancy level, including 0%, 0~25%, 25~50%, 50~75%, 75~100%, and 100% were depicted as Figure 12 (a), 12 (b), and 12 (c), respectively. The occupancy level was ranked based on the depth of the color.



(a) Dormitory building



(b) Lecture building



(c) Office building

Figure 12. Comparison of measured and predicted occupancy level

It can be seen from Figure 12 (a), the measured and predicted occupancy level for any patterns of dormitory room had been highly consistent at two time points as the beginning and ending of a day with the highest occupancy, respectively. It can be observed directly that the primary difference between pattern 1 and pattern 2 lied in the occupancy level during the noon break. And the peak occupancy level of pattern 1 was obviously higher than that of pattern 2, supposing early-hour keepers more likely return to dormitory rooms for noon break. In comparison, the occupancy of patterns 3 changed steadily throughout the day, positioning the measured and predicted results generally at the level of 25~50%. This indicated that the dormitory room of pattern 3 lived less than two occupants on average used to house in their dorms.

Similarly smooth changing tendency of occupancy level were applied to classrooms of pattern 3 as Figure 11 (b) and office rooms of pattern 2 as Figure 11 (c). Observed from the color diversity of measured and predicted results of the two patterns, the occupancy corresponding to adjacent time points had been fluctuated at the same or contiguous scale of occupancy level. From the measured data shown in Figure 12 (b), the primary difference between another two patterns of lecture building lied in the start time of the state transformed from unoccupied to occupied. And the color diversity of the measured peaks between two patterns may be resulted from the class size. However, the differences have not resulted in significant distinction of predicted occupancy level between the two patterns with the same functionality. As is seen from Figure 12 (c), the measured occupancy of office room of pattern 1 plunged to zero when it reached to the time for off-duty limited by fixed office time, instead, the predicted occupancy level postponed to decline to zero, conforming to the smoothly downward trend of Gaussian curved presented in section 3.2. In order to further weigh and compare the level of prediction error of Gaussian distribution for occupancy of different rooms, $RMSE_{ave}$ index was used to evaluate in the order of time series and the results were shown as Figure 13.

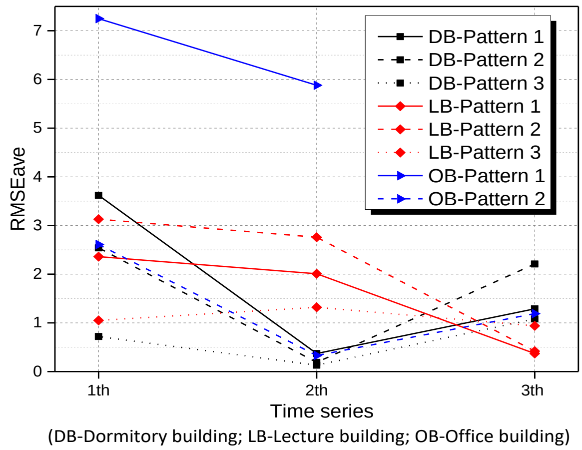


Figure 13. The average value of statistic indicator $RMSE_{ave}$ of verification datasets

Combined with Figure 10, it can be concluded that lower $REMSE_{ave}$ value always appeared in the time series of gentle occupancy change. Particularly, for dormitory room with the capacity of 4 and teachers' office room with the capacity of 8, the mobility of each occupant will bring about 25% and 12.5% occupancy fluctuate, respectively, which caused relatively drastic occupancy change. However, as was seen in Figure 10, Gaussian distribution was more suitable for the fit of stable occupancy, and thus the prediction performance for the time series with less mobility of occupants or the rooms with larger capacity were best with lower $REMSE_{ave}$ value.

4. Simulation and validation of energy consumption

Based on the total number of rooms and the proportions of rooms with various occupancy patterns introduced in section 2.1.1 and section 3.1, the number of various patterns of rooms for one case building could be obtained. The occupancy profiles obtained by Gaussian distribution coupled with relevant parameters including the environmental condition and building information, etc., will be input to DesignBuilder, accumulating each sub-item of simulated energy data as the second part of energy introduced in section 2.3. In light of the resulted energy divergence from the diversity of occupancy pattern, the energy of each pattern was simulated separately, multiplied by the number of this pattern of rooms, and then superposed to acquire the energy data of the whole case building. The comparison of measured and simulated energy data in the building level were shown as Figure 14. The 'classified' results represented the simulated energy data considering the diversity of occupancy pattern at the room-scale, while the 'unclassified' results represented the simulated energy data considering the average level of occupancy at the building-scale.

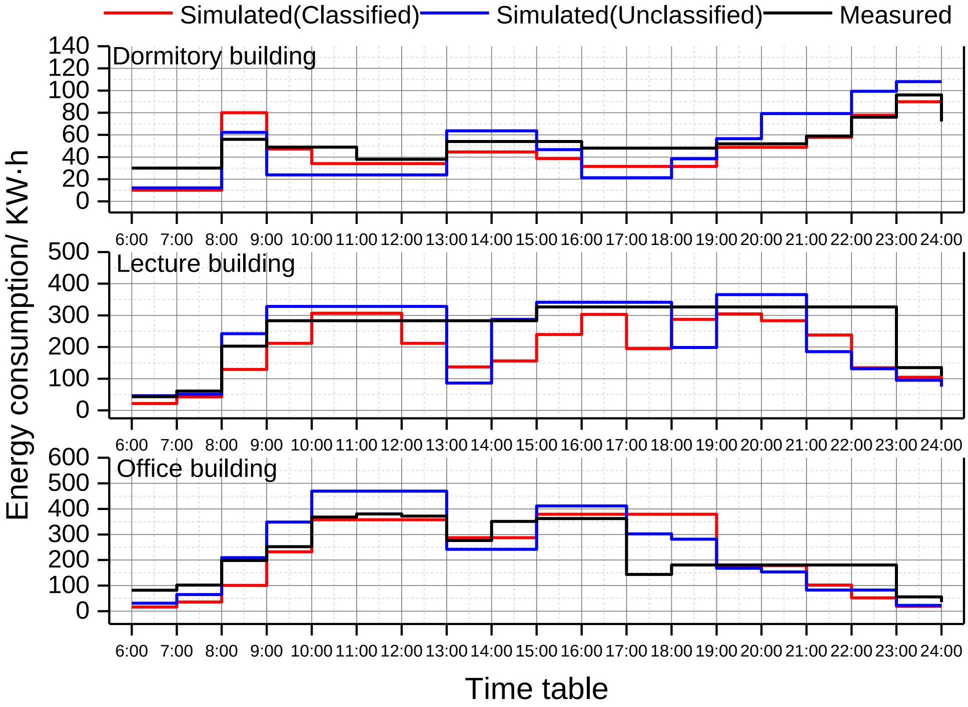


Figure 14. Comparison of measured and simulated energy consumption

It can be seen from Figure 14, the simulated energy data and changing tendency with classified accorded with the measured data more closely for any building type. The situation that several occupants lived or stayed in one room shared part of energy failed to be discriminated when applying the occupancy profiles obtained by modelling without classified, causing the simulated energy data to exceed the measured data in the increasing stage of occupancy, and conversely to fall behind the measured data in the decreasing stage of occupancy. Adopting unclassified occupancy patterns for modelling in the domain of energy simulation and load distribution would result in inaccuracy of energy prediction. While, the methodology of classification at the room-scale considered the room state and corresponding occupancy of each clustered pattern when fitting occupancy, avoiding the errors caused by inhomogeneous distribution of occupants.

Integrated with the fitted Gaussian curve for three time series shown as Figure 10 (b), both simulated energy data of lecture building presented a similar tendency with three peaks. However, the measured energy data maintained peak value during the daytime except for the morning and nighttime. The raise of occupancy brought about an increased building energy consumption, however, the drop of occupancy with their departure during lunch time and dinner time did not lead to the decrease of energy use. The leaving of occupants will not generally trigger the closing behavior of some public equipment such as lights, air conditioners, etc., instead, these equipment will keep original state once opened, inevitably causing a waste of energy during the valley time periods but appearing energy saving potential. Different from the lecture building, a majority of occupants of office building would choose to close the energy equipment when they leave their office rooms, making the occurrence time of the peak and valley of simulated and measured energy generally consistent. However, the energy data unlike the occupancy data of office building would not reduce to zero. Some high-powered equipment just like the laptop/computer were always in standby state of operation based on questionnaires. It will be advocated to close these energy equipment with departure. As is seen from Figure 12 (c), the occupancy level of office rooms of pattern 2 always positioned at a high level but the occupancy level of pattern 1 dropped to zero after work, causing the third peak energy lower than the first two peaks of energy. Therefore, it is not reasonable to adopt the same load capacity for different peak periods, instead, the distributed load for the third peak period shall be less than that for another two-peak periods.

5. Conclusion

The case study provided an exploration for occupancy modeling and energy simulation based on Gaussian distribution considering the diversity of occupancy patterns for different building types within a campus. It was observed that Gaussian distribution emerged a better fitting performance with lower $REMSE_{ave}$ value for the time series with less mobility of occupants or the rooms with larger capacity. Although distinct prediction deviations still existed for the mutation between valley and peak occupancy, the prediction accuracy was yet controlled within an acceptable range of $\pm$15%. Positive errors, representing lower fitted curves than measurement, usually appeared at the valley occupancy and be offset with smooth increase of occupancy. Conversely, negative errors, representing excessive fitting results compared with measurement, usually appeared at the peak and be eliminated by steady descend of curves.

Generally, the occupancy of some spaces within campus for course teaching or self-studying generally presented three peaks separated by lunch and dinner time, respectively. However, the occupancy of spaces within campus provided for teachers handling office presented only two peaks during work hours. The exploration of low energy or near zero energy output resulted from the unoccupied state of this pattern of teachers' office rooms after work provided basis for the decrease of this part of load distribution. Further combining the energy data with occupancy data, sometimes, the arrival of occupants at some public spaces brought about the increase of energy use but their departure did not result in the projected decrease of energy use. It deserved to be noted that the valley time duration certainly cause a waste of energy but demonstrate an energy-saving potential caused by irrational load distribution.

The detection of occupancy patterns conducted in this study could arouse in-depth pondering and excavation for both occupancy modelling and building performance simulation of a large variety of researches regarding to such functional buildings. And the combination of occupancy and energy data could provide firm foundation for the domains of load distribution and energy-saving implementation. However, although Gaussian distribution considering the diversity of occupancy pattern was validated as a qualified method to predict occupancy and energy simulation, it should be pointed out that the surveyed and clustered occupancy pattern was not comprehensive. For other exceptional cases, such as part of rooms in occupied/unoccupied state during the unoccupied/occupied time periods observed in this study should be noted and included in our future research. Additionally, whether the regularities of occupant distribution indoors detected in this case campus are applicable to most of both domestic and foreign campus buildings with various occupant behavior remains to be verified.

Acknowledgement

This work is supported by the Nature Science Foundation of Tianjin (19JCQNJC07000) and has been benefited from broader discussion of occupant behavior in the International Energy Agency Energy in Buildings and Communities Program (IEA EBC) Annex 79: Occupant behavior-centric building design and operation.

Reference

Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society, 54*, 102010. doi:https://doi.org/10.1016/j.scs.2019.102010

Bionda, D., & Domingo-Irigoyen, S. (2017). Energy saving potential of occupancy-based heating control in residential buildings. *Energy Procedia, 122*, 27-31. doi:https://doi.org/10.1016/j.egypro.2017.07.298

Buttitta, G., Turner, W. J. N., Neu, O., & Finn, D. P. (2019). Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. *Energy and Buildings, 198*, 84-99. doi:https://doi.org/10.1016/j.enbuild.2019.05.056

Candanedo, L. M., & Feldheim, V. (2016). Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. *Energy and Buildings, 112*, 28-39. doi:https://doi.org/10.1016/j.enbuild.2015.11.071

Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building. *Energy and Buildings, 148*, 327-341. doi:https://doi.org/10.1016/j.enbuild.2017.05.031

Cao, B., Cui, W., Chen, C., & Chen, Y. (2020). Development and uncertainty analysis of radionuclide atmospheric dispersion modeling codes based on Gaussian plume model. *Energy, 194*, 116925. doi:https://doi.org/10.1016/j.energy.2020.116925

Capozzoli, A., Piscitelli, M. S., Gorrino, A., Ballarini, I., & Corrado, V. (2017). Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings. *Sustainable Cities and Society, 35*, 191-208. doi:https://doi.org/10.1016/j.scs.2017.07.016

Chen, Z., & Soh, Y. C. Comparing occupancy models and data mining approaches for regular occupancy prediction in commercial buildings. *Journal of Building Performance Simulation*, 1-9.

D'Oca, S., & Hong, T. (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings, 88*, 395-408. doi:https://doi.org/10.1016/j.enbuild.2014.11.065

Ding, Y., Wang, Q., Wang, Z., Han, S., & Zhu, N. (2019). An occupancy-based model for building electricity consumption prediction: A case study of three campus buildings in Tianjin. *Energy and Buildings, 202*, 109412. doi:https://doi.org/10.1016/j.enbuild.2019.109412

Dong, B., Andrews, B., Lam, K. P., Höynck, M., Zhang, R., Chiou, Y.-S., & Benitez, D. (2010). An information technology enabled sustainability test-bed (ITEST) for occupancy detection through an environmental sensing network. *Energy and Buildings, 42*(7), 1038-1046. doi:https://doi.org/10.1016/j.enbuild.2010.01.016

Dong, B., & Lam, K. P. A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation, 7*(1), 89-106.

Feng, Z.-k., Niu, W.-j., Zhang, R., Wang, S., & Cheng, C.-t. (2019). Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization. *Journal of Hydrology, 576*, 229-238. doi:https://doi.org/10.1016/j.jhydrol.2019.06.045

Gaetani, I., Hoes, P.-J., & Hensen, J. L. M. (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings, 121*, 188-204. doi:https://doi.org/10.1016/j.enbuild.2016.03.038

Hailemariam, E., Goldstein, R., Attar, R., & Khan, A. (2011). *Real-time occupancy detection using decision trees with multiple sensor types*.

Han, Y., Zhou, X., & Luo, R. (2015). Analysis on Campus Energy Consumption and Energy Saving Measures in Cold Region of China. *Procedia Engineering, 121*, 801-808. doi:https://doi.org/10.1016/j.proeng.2015.09.033

Huchuk, B., Sanner, S., & O'Brien, W. (2019). Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. *Building and Environment, 160*, 106177. doi:https://doi.org/10.1016/j.buildenv.2019.106177

Li, J., Yu, Z., Haghighat, F., & Zhang, G. (2019). Development and improvement of occupant behavior models towards realistic building performance simulation: A review. *Sustainable Cities and Society, 50*, 101685. doi:https://doi.org/10.1016/j.scs.2019.101685

Li, Z., & Dong, B. (2018). Short term predictions of occupancy in commercial buildings—Performance analysis for stochastic models and machine learning approaches. *Energy and Buildings, 158*, 268-281. doi:https://doi.org/10.1016/j.enbuild.2017.09.052

Meyn, S. P., Surana, A., Lin, Y., Oggianu, S. M., & Frewen, T. A. (2009). *A Sensor-Utility-Network Method for Estimation of Occupancy in Buildings.* Paper presented at the Proceedings of the 48th IEEE Conference on Decision and Control, CDC 2009, combined withe the 28th Chinese Control Conference, December 16-18, 2009, Shanghai, China.

Nägele, F., Kasper, T., & Girod, B. (2017). Turning up the heat on obsolete thermostats: A simulation-based comparison of intelligent control approaches for residential heating systems. *Renewable and Sustainable Energy Reviews, 75*, 1254-1268. doi:https://doi.org/10.1016/j.rser.2016.11.112

Nguyen, T. A., & Aiello, M. (2013). Energy intelligent buildings based on user activity: A survey. *Energy and Buildings, 56*, 244-257. doi:https://doi.org/10.1016/j.enbuild.2012.09.005

Oldewurtel, F., Sturzenegger, D., & Morari, M. (2013). Importance of occupancy information for building climate control. *Applied Energy, 101*, 521-532. doi:https://doi.org/10.1016/j.apenergy.2012.06.014

Roselyn, J. P., Uthra, R. A., Raj, A., Devaraj, D., Bharadwaj, P., & Krishna Kaki, S. V. D. (2019). Development and implementation of novel sensor fusion algorithm for occupancy detection and automation in energy efficient buildings. *Sustainable Cities and Society, 44*, 85-98. doi:https://doi.org/10.1016/j.scs.2018.09.031

Salimi, S., Liu, Z., & Hammad, A. (2019). Occupancy prediction model for open-plan offices using real-time location system and inhomogeneous Markov chain. *Building and Environment, 152*, 1-16. doi:https://doi.org/10.1016/j.buildenv.2019.01.052

Schakib-Ekbatan, K., Çakıcı, F. Z., Schweiker, M., & Wagner, A. (2015). Does the occupant behavior match the energy concept of the building? – Analysis of a German naturally ventilated office building. *Building and Environment, 84*, 142-150. doi:https://doi.org/10.1016/j.buildenv.2014.10.018

Shen, W., Newsham, G., & Gunay, B. (2017). Leveraging existing occupancy-related data for optimal control of commercial office buildings: A review. *Advanced Engineering Informatics, 33*, 230-242. doi:https://doi.org/10.1016/j.aei.2016.12.008

Stazi, F., Naspi, F., & D'Orazio, M. (2017). A literature review on driving factors and contextual events influencing occupants' behaviours in buildings. *Building and Environment, 118*, 40-66. doi:https://doi.org/10.1016/j.buildenv.2017.03.021

Wang, W., Chen, J., & Hong, T. (2018). Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings. *Automation in Construction, 94*, 233-243. doi:https://doi.org/10.1016/j.autcon.2018.07.007

Wymelenberg, K. V. D. (2012). Patterns of occupant interaction with window blinds: A literature review. *Energy & Buildings, 51*(AUG.), p.165-176.

Yan, D., O' Brien, W., Hong, T., Feng, X., Burak Gunay, H., Tahmasebi, F., & Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings, 107*, 264-278. doi:https://doi.org/10.1016/j.enbuild.2015.08.032

Yang, J., Santamouris, M., & Lee, S. E. (2016). Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. *Energy and Buildings, 121*, 344-349. doi:https://doi.org/10.1016/j.enbuild.2015.12.019

Yoshino, H., Hong, T., & Nord, N. (2017). IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods. *Energy and Buildings, 152*, 124-136. doi:https://doi.org/10.1016/j.enbuild.2017.07.038

Yu, T. (2010). *Modeling Occupancy Behavior for Energy Efficiency and Occupants Comfort Management in Intelligent Buildings*.

Yu, Z., Fung, B. C. M., Haghighat, F., Yoshino, H., & Morofsky, E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings, 43*(6), 1409-1417. doi:https://doi.org/10.1016/j.enbuild.2011.02.002

Yu, Z., Haghighat, F., Fung, B. C. M., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings, 42*(10), 1637-1646. doi:https://doi.org/10.1016/j.enbuild.2010.04.006

Yu, Z., Haghighat, F., Fung, B. C. M., & Zhou, L. (2012). A novel methodology for knowledge discovery through mining associations between building operational data. *Energy and Buildings, 47*, 430-440. doi:https://doi.org/10.1016/j.enbuild.2011.12.018

Zhang, J., Yan, J., Infield, D., Liu, Y., & Lien, F.-s. (2019). Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. *Applied Energy, 241*, 229-244. doi:https://doi.org/10.1016/j.apenergy.2019.03.044

Zhang, Y., Bai, X., Mills, F. P., & Pezzey, J. C. V. (2018). Rethinking the role of occupant behavior in building energy performance: A review. *Energy and Buildings, 172*, 279-294. doi:https://doi.org/10.1016/j.enbuild.2018.05.017