

An evolving learning method –growing Gaussian mixture regression—for modeling passive chilled beam systems in buildings

Liping Wang^{a,*}, James Braun^{b,c}, Sujit Dahal^a

^a Civil and Architectural Engineering, University of Wyoming, 1000 E University Ave, Laramie, WY 82071, United States

^b School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN 47907, United States

^c Center for High Performance Buildings, Ray W. Herrick Laboratories, Purdue University, 177 S Russell St, West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:

Received 3 January 2022

Revised 25 May 2022

Accepted 27 May 2022

Available online 30 May 2022

Keywords:

Data-driven model

Evolving learning

Growing Gaussian mixture regression

Chilled beam

ABSTRACT

Although chilled beam systems are considered as one of the best performing HVAC systems, few studies have looked into data-driven-based modeling for chilled beam systems. Furthermore, no learning-based data-driven methods have been applied to chilled beam systems in buildings. In this study, we used an evolving learning method, growing Gaussian mixture regression (GGMR), to predict cooling rates for passive chilled beam (PCB) systems where the training, evolution, and validation were carried out using data from real system measurements and from building energy simulation. GGMR updates key parameters such as weight coefficients, means, and covariance matrices of Gaussian components to adapt to changes in system operation beyond training data. This case study demonstrated that GGMR is an effective evolving learning-based data-driven method for accurately predicting cooling rates of PCB systems. The selection of key performance parameters of GGMR models including the number of components, training data size, and the learning rate was discussed in this paper. It is recommended that GGMR models could be further explored for predicting the performance of other complex HVAC systems such as radiant slab or mixed-mode ventilation systems.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Chilled beams can be categorized into passive chilled beams and active chilled beams [1]. A passive chilled beam exchanges cooling energy with spaces through natural convection. In addition, displacement ventilation [2] or variable air volume systems are commonly employed to meet space heating and ventilation requirements. Active chilled beams exchange energy with spaces through forced convection. For active chilled beams, room air is induced across finned coils and then mixed with primary air. Active chilled beams can be used for space heating or cooling.

A chilled beam system is considered one of the high-performance HVAC systems. Convection is the primary heat transfer for chilled beam systems [3]. Compared to conventional cooling systems, supply chilled water temperatures can be significantly higher, and thus the number of hours for running a free cooling chilled water system can be extended and the efficiency of the chiller operation can also be improved [4]. In addition, fan energy consumption can be reduced due to reduced airflow rate although the savings may be compromised by increased pump energy consumption for the chilled water distribution system [5]. Kim et al. [4] sim-

ulated four passive chilled beam system configurations in different U.S. climatic zones and concluded that a passive chilled beam system could result in up to 20% energy savings depending on climate zones in comparison with a variable air volume system. Rumsey and Weale [6] predicted a 57% reduction in energy use for a 3,716 m² laboratory in California, U.S. conditioned by an active chilled beam system and dedicated outdoor air system with additional features that included a heat recovery ventilator, cogeneration, waterside economizer, and a photovoltaic system that met 10% of electricity demands. The system was compared with a baseline having a variable air volume system with none of the additional features. In addition, computational fluid dynamic (CFD) simulations have been employed to evaluate indoor air thermal stratification and cooling effectiveness of chilled beam systems [7–9].

Existing modeling efforts for predicting chilled beam energy use can be categorized into first-principle-based [10], data-driven-based [11], or hybrid [12] methods. Kim et al. [13] reviewed modeling methods for passive ceiling cooling including chilled beams, chilled ceiling panels, and radiant slabs, and categorized the passive ceiling modeling approaches into component models, indoor environmental models, and integrated models between the two. Betz and McNeill [10] summarized the capability of building

* Corresponding author.

energy programs for modeling chilled beams and their limitations. Kim et al. [11] created regression models for predicting cooling rates and chilled surface temperature of passive chilled beams based on experimental results. The input parameters to the regression model include water flow rate, water supply temperature, the air temperature above the chilled beam, and area-weighted uncooled surface temperature in the space. Chen et al. [12] developed a dynamic model for predicting the performance of active chilled beams. Air mixing and cooling coil heat transfer were modeled based on first principles and experimental data were used to fit unknown model parameters. Filipsson et al. [14] modeled an active chilled beam based on the number of transfer units (NTU) method and empirical equations for air induction ratio. With the complexity of system operation, calibration of first-principle-based simulation models for chilled beam systems is a challenging task, and therefore, this has limited the application of first-principle-based models of chilled beam systems for operational use such as *retro*-commissioning, advanced controls, and fault detection and diagnostics. The intent of existing energy simulation efforts for chilled beam systems is to facilitate chilled beam system design rather than operation and maintenance. Data-driven-based or hybrid modeling for chilled beam systems are promising approaches that could inform system operation and increase energy savings for high-performance systems, but have received limited study [15].

With the advancement of sensing and monitoring technologies, large amounts of building system performance data have increasingly become available. Various data-driven-based algorithms [16,17] have been investigated for predicting energy consumption and indoor environmental conditions in buildings. This is mainly due to the advantages of data-driven models, such as easier implementation for model predictive controls [18] and less model development effort in comparison with first-principle-based models. The main challenge of data-driven-based modeling is that the developed data-driven model is sensitive to training data size, the accuracy of the training data, and the operating conditions under which the data were collected [19,20]. The training data available do not include all possible operating conditions that the target systems and components experience throughout their life cycles. It is expected and important that data-driven models for advanced controls and fault diagnostics can adapt and learn with the evolution of building systems. Evolving learning refers to the feature that the model evolves with the changes in the dynamic system of interest, specifically HVAC operating environments, building systems, and components in our study. The evolving learning method overcomes this limitation through updating key parameters over time.

Currently, no learning-based data-driven methods have been used to model high-performance HVAC systems, such as passive chilled beam systems. The novelty of the study is the development, application and evaluation of an evolving learning method, growing Gaussian mixture regression (GGMR), as a modeling approach for predicting cooling rates of passive chilled beam systems. This study represents the first application of an evolving learning method for predicting the performance of complex building systems.

2. Methodology

Before discussing the GGMR algorithm, we introduce the Gaussian mixture model (GMM) and Gaussian mixture regression (GMR). Both GMM and GMR have been successfully applied as effective data-driven methods for building energy use prediction [19,21,22] and fault detection and diagnosis [23,24] for the built environment. The methods have also been actively used in other

disciplines such as material design [25], energy storage [26], and robotics [27].

A Gaussian mixture model (GMM) is a parametric joint probability density function represented as a weighted sum of Gaussian component densities (Eq. (1)).

$$\text{prob}(\mathbf{u}, \mathbf{y}) = \sum_{m=1}^M \alpha_m N(\mathbf{u}, \mathbf{y} | \boldsymbol{\mu}_m, \boldsymbol{\sigma}_m) \quad (1)$$

$$\boldsymbol{\mu}_m = \begin{bmatrix} \mu_{m,X} \\ \mu_{m,Y} \end{bmatrix} \quad (2)$$

$$\boldsymbol{\sigma}_m = \begin{bmatrix} \sigma_{m,XX} & \sigma_{m,XY} \\ \sigma_{m,YX} & \sigma_{m,YY} \end{bmatrix} \quad (3)$$

where $\mathbf{u}$ is a zero-mean row vector of measurement for independent variables $\mathbf{X}$; $\mathbf{y}$ is the observation for dependent variable Y ; $\text{prob}(\mathbf{u}, \mathbf{y})$ is a multivariate probability density function of a Gaussian Mixture Model, α_m is non-negative mixing coefficient of the m -th component, $\boldsymbol{\mu}_m$ (Eq. (2)) and $\boldsymbol{\sigma}_m$ (Eq. (3)) are the mean vector and covariance matrix of the m -th Gaussian component. N represents a probability density function of a Gaussian component. M is the number of GMM components. Training data are required to develop the mixture model for key parameters. The expectation-maximization (EM) algorithm is used for finding maximum likelihood estimates of GMM parameters such as α_m , $\boldsymbol{\mu}_m$ and $\boldsymbol{\sigma}_m$ based on a training dataset.

Gaussian mixture regression (GMR) [28] is then used to predict the value of the dependent variable $\hat{y}$ (Eq. (4)) given the observations $\mathbf{u}$ of independent variables and GMM parameters. $\hat{y}$ represents the prediction of the dependent variable Y . $\alpha_m(\mathbf{u})$ (Eq. (5)) is the mixing coefficient vector given observed independent variables. Mean value $\mu_{m,Y|X}$ of dependent variable Y given observations is updated based on Eq. (6).

$$\hat{y} = E(Y | \mathbf{X} = \mathbf{u}) = \sum_{m=1}^M \alpha_m(\mathbf{u}) \mu_{m,Y|X} \quad (4)$$

$$\alpha_m(\mathbf{u}) = \frac{\alpha_m \text{prob}(\mathbf{u}; \mu_{m,X}, \sigma_{m,XX})}{\sum_{m=1}^M \alpha_m \text{prob}(\mathbf{u}; \mu_{m,X}, \sigma_{m,XX})} \quad (5)$$

$$\mu_{m,Y|X} = \mu_{m,Y} + \sigma_{m,Y|X} (\sigma_{m,YY})^{-1} (\mathbf{u} - \mu_{m,X}) \quad (6)$$

For every new data, growing Gaussian mixture regression (GGMR) continuously updates $\boldsymbol{\mu}_m$ and $\boldsymbol{\sigma}_m$ in GMM and α_m in GMR, which were static and were calculated based on training data. The parameters of the component with the highest probability density (non-zero) including the weight coefficients α_m , means $\boldsymbol{\mu}_m$, and covariances $\boldsymbol{\sigma}_m$ are updated using a recursive filtering approach, as shown in Eqs. (7) to (9) [29].

$$\alpha_m(t) = (1 - \beta) \alpha_m(t-1) + \beta q_m \quad (7)$$

$$\boldsymbol{\mu}_m(t) = (1 - \rho_m) \boldsymbol{\mu}_m(t-1) + \rho_m \mathbf{d}(t) \quad (8)$$

$$\boldsymbol{\sigma}_m(t) = (1 - \rho_m) \boldsymbol{\sigma}_m(t-1) + \rho_m (\mathbf{d}(t) - \boldsymbol{\mu}_m(t)) (\mathbf{d}(t) - \boldsymbol{\mu}_m(t))^T \quad (9)$$

where $\mathbf{d}(t) = [\mathbf{u}(t), \mathbf{y}(t)]$, q_m is the expected posterior of the m th component, β is the learning rate, and ρ_m is the learning factor corresponding to updating $\boldsymbol{\mu}_m$ and $\boldsymbol{\sigma}_m$. GGMR can adapt to changes in operation beyond training data or the replacement of HVAC equipment or systems.

3. Case study buildings

In this section, we described the two passive chilled beam (PCB) systems used as case studies: the PCB system in a modified DOE large office reference building prototype implemented as an EnergyPlus model and a real PCB system in the Living Lab 1, Herrick Laboratories at Purdue University. We used GGMR models to predict cooling rates for both PCB systems. The cooling rates $\dot{Q}$ (kW) of PCB systems are defined by Eq. (10).

$$\dot{Q} = \dot{m}C_p(t_r - t_s) \quad (10)$$

where $\dot{m}$ is supply water flow rates of PCB systems (kg/s); C_p is the specific heat of water (kJ/(kg·K)); t_r is the return water temperature of PCB systems (°C); t_s is the supply water temperature of PCB systems (°C).

3.1. The DOE large office reference building prototype

The EnergyPlus model for the DOE large office reference building prototype [30] was used and modified in this study. The large office building prototype has 12 stories and a basement. The geometry model of this prototype is shown in Fig. 1. Thermal zones of the basement, the first floor, the middle floor (6th floor), and the top floor were explicitly added in the model. To account for floor area, zone loads, and internal gains, and system and equipment sizing of typical floors for the 2nd through 11th stories, a “multiplier” of 10 was applied to the middle floor’s thermal zones. The typical floor plan of the building is shown in Fig. 2. There are five conditioned thermal zones per floor including four perimeter zones and one core zone. The average window to wall ratio is 40%.

In the EnergyPlus models for the large office building prototype, there are four single duct variable air volume (VAV) systems serving the basement, first floor, middle floors, and top floor, respectively. There are two water-cooled chillers and one natural gas boiler providing chilled water and heating hot water to water coils. The EnergyPlus models across different climate zones comply with ANSI/ASHRAE/IESNA Standard 90.1-2004 [31] for new construction.

In this study, the HVAC system models were modified. Models for a passive chilled beam system and a baseboard water heating system were employed for cooling and heating. The air systems were modified to only meet latent loads and ventilation requirements. The passive chilled beam system accounts for a majority of the building’s sensible cooling loads. The system diagram of

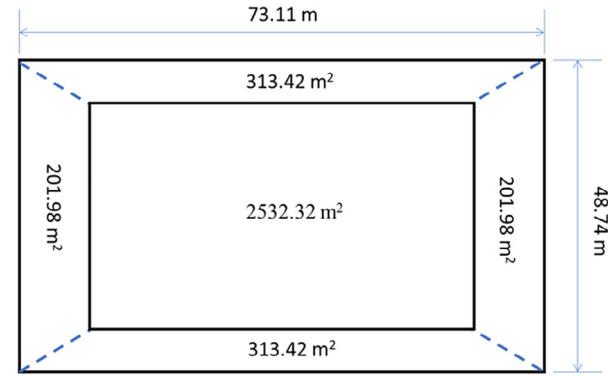


Fig. 2. Typical floor plan of the large office building prototype.

the modified mechanical system is shown in Fig. 3. A water-to-water heat exchanger was added to the secondary loop. The chilled water supply temperature setpoint was 15 °C to the passive chilled beams in the thermal zones.

3.2. Living Lab 1, Herrick Laboratories at Purdue University

Living Lab 1—one of the four living labs in the Herrick Laboratories at Purdue University—is equipped with high-performance HVAC systems. Living Lab 1 (W: 9.5 m × L: 10.7 m) is an open office, which has a capacity of 20 people and features a passive chilled beam system (PCB), a hydronic radiant slab system (RS), displacement ventilation from a variable air volume system (VAV), and a south-facing double-skin façade with motorized openings. The indoor environment and the HVAC systems are intensively monitored and controlled via a Niagara/AX software framework through Tridium JACE controllers which can connect devices (sensors and actuators) and systems utilizing diverse communication protocols.

For the purpose of this study on modeling PCB cooling rates, both PCB and VAV systems control the space temperature simultaneously and serve space cooling while the RS system and façade ventilation were off. As shown in Fig. 4, there are 30 passive chilled beam panels installed in the suspended ceiling of Living Lab 1. The dimension of each chilled beam panel is 1.2 m (L) × 0.52 m (W) × 0.14 m (H). The 30 panels are divided into three banks (North, Middle, and South). A control valve for each bank can modulate water flow rates. For this study, the supply water temperature to the PCBs was controlled based on a PI feedback loop to maintain the space temperature thermostat at a setpoint (22.8 °C). The supply water temperature was adjusted by mixing hot water that was heated using the campus district heating with cold water cooled by campus chilled water using heat exchangers that are depicted in Fig. 4. This is not a typical arrangement and control approach for a PCB system because it was designed to be used for experimental purposes and not for energy efficiency. A typical PCB system would be connected directly to cooling equipment controlled to maintain a constant supply water temperature to the PCBs that is above the dewpoint of the room air with feedback control of the PCB water flows to maintain the zone temperature setpoint. This was not possible for this experimental system because the cooling is provided with a campus-wide chilled water system that provides water at a temperature that is well below room air dewpoint. Even though this system is not typical, it is still useful for studying the application of GGMR method to a complex HVAC system. During operation in cooling mode for this study, the supply water temperature setpoints to the PCBs varied between 16.7 and 22.8 °C for the living lab. The system control sequence was set to ensure that the supply water temperature setpoint was always above the dew point of the space air to avoid any risk

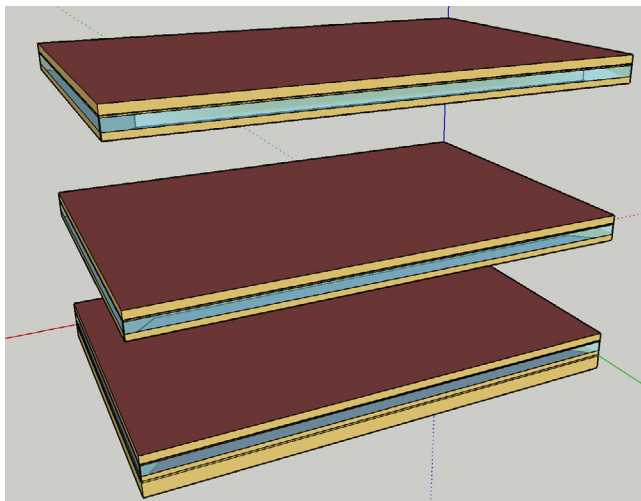


Fig. 1. Geometry model of the large office building prototype.

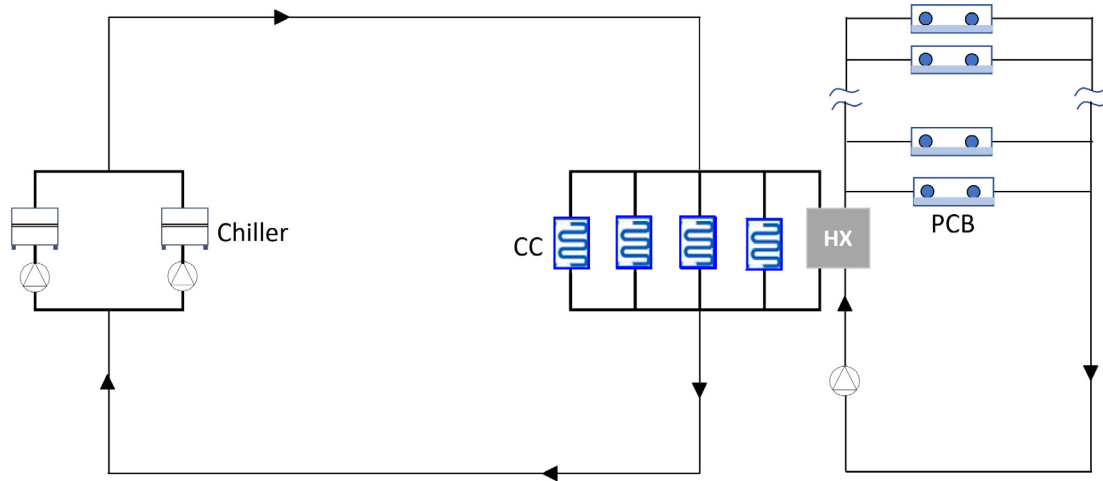


Fig. 3. System diagram of the chilled water loop for the large office building model (HX: heat exchanger, CC: cooling coil, PCB: passive chilled beam).

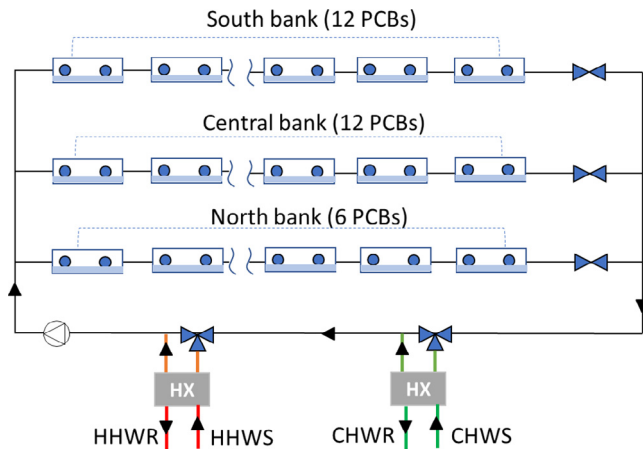


Fig. 4. System diagram of passive chilled beam system for Living Lab 1 (HX: heat exchanger, PCB: passive chilled beam).

of condensation. When measured supply water temperatures are above supply water temperature setpoint, cooling exchange between campus chilled water and the PCB system was enabled. Otherwise, heating exchange between campus hot water and the PCB system was enabled. The supply water temperature setpoint was used to set the control valve positions based on PI feedback loops for heating/cooling exchange between water in the PCB system and the campus hot and chilled water supply.

4. Selection of input parameters for GGMR

Both simulation data from EnergyPlus and measurement data from Living Lab 1 were first normalized using Eq. (11) before being used for GGMR model creation and testing. Each parameter z was normalized based on its mean $\bar{z}$ and standard deviation s . The normalized parameter was represented by z' in Eq. (11). The purpose of normalization is to ensure that the parameters with large values

don't have too much influence on model prediction in data-driven models [19].

$$z' = \frac{z - \bar{z}}{s} \quad (11)$$

4.1. EnergyPlus simulation data

The simulation data from the EnergyPlus model for the large office building located in Miami, Florida was used to demonstrate the process of selecting input parameters for GGMR models that predict cooling rates of the passive chilled beam system.

Correlation coefficients were first used to evaluate the correlation between dependent variables and potential input parameters. The correlation coefficients between hourly cooling rates and input parameters such as environmental parameters and operation schedules are summarized in Table 1. We selected the parameters having relatively high correlations with cooling rates. Based on the correlation coefficient results from Table 1, outdoor air temperature (OAT), outdoor air relative humidity (RH), zone air temperature (ZAT), HVAC system operation status (e.g., on/off), and solar radiation have a relatively strong correlation with PCB cooling rates. The HVAC system operation status determines when HVAC systems are available for running. Therefore, those parameters are selected as initial input parameters for performance evaluation.

A few scenarios with different combinations of input parameters were tested. During the selection of input parameters, the number of components M was set as 10 and the learning rate β was 0.001 for GGMR models. Performance data for January (744 hourly data points) were used for model training and the data from February to December (8016 hourly data points) for testing. For evolving learning method, the required training data volume is relatively small because key parameters of the means, covariance matrices, and weight factors of Gaussian components are updated with new testing data.

Table 2 summarized the list of input parameters for each scenario. In scenario 1, OAT, RH, ZAT, operation status, direct radia-

Table 1

Correlation coefficients between hour cooling rate of the passive chilled beam system and potential input parameters for EnergyPlus models. (OAT: outdoor air temperature; RH: relative humidity; ZAT: zone air temperature).

	OAT	RH	ZAT	Operation status	Direct radiation	Diffuse horizontal radiation
PCB cooling rate (kW)	0.41	-0.35	-0.77	0.89	0.53	0.6

Table 2

List of input parameters for each scenario using EnergyPlus simulation data.

List of Input Parameters	
S1	OAT, RH, ZAT, operation status, direct radiation, and diffuse radiation
S2	OAT, ZAT, operation status, and direct radiation
S3	OAT, ZAT, operation status, direct radiation, and historical cooling rates for the previous three hours
S4	OAT, ZAT, operation status, direct radiation, historical cooling rates for the previous three hours, and the occupancy ratio

Table 3

Statistical analysis results for performance prediction of scenarios 1–4 with EnergyPlus simulation data.

PCB cooling rate	S1	S2	S3	S4
CV-RMSE	20.5%	25.2%	8.7%	5.8%
MBE	−1.8%	−4.6%	−0.8%	−0.3%
R ²	0.936	0.888	0.990	0.995

tion, and diffuse radiation were selected for PCB cooling rates. In scenario 2, RH and diffuse radiation were removed and the input parameters were OAT, ZAT, operation status, and direct radiation. In scenario 3, PCB cooling rates for the previous three hours were added as inputs to scenario 1 to consider the high thermal capacitance of both the PCB system and the large office building. In scenario 4, the 3-hour time history for PCB cooling rates was added to scenario 2 and an additional input was included that is the ratio of the number of occupants to full occupancy. The occupancy input is important in accounting for differences that occur between weekdays and weekends.

Statistical testing results for the coefficient of variation of root mean square error (CV-RMSE), normalized mean bias error (MBE), and R squared (R²) of scenarios 1–4 are summarized in Table 3. Eqs. (12) to (13) are for calculating CV-RMSE, MBE, and R².

$$CV - RMSE = \frac{1}{\bar{Y}} \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{N}}$$

$$MBE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)}{N \times \bar{Y}} \quad (13)$$

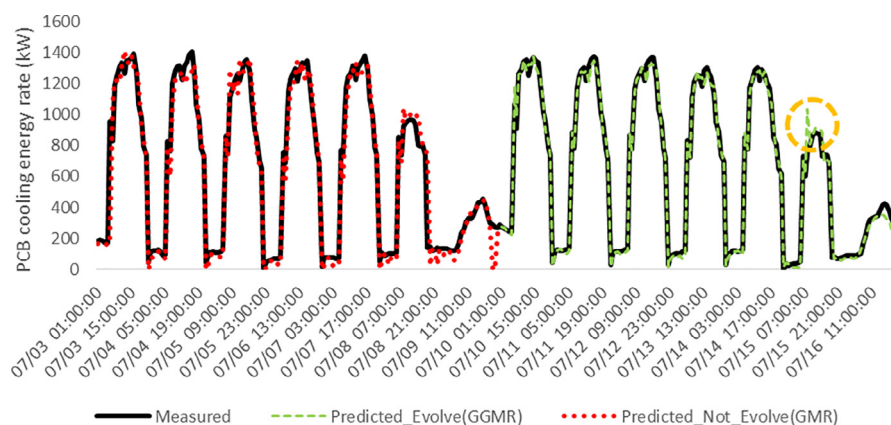
$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (14)$$

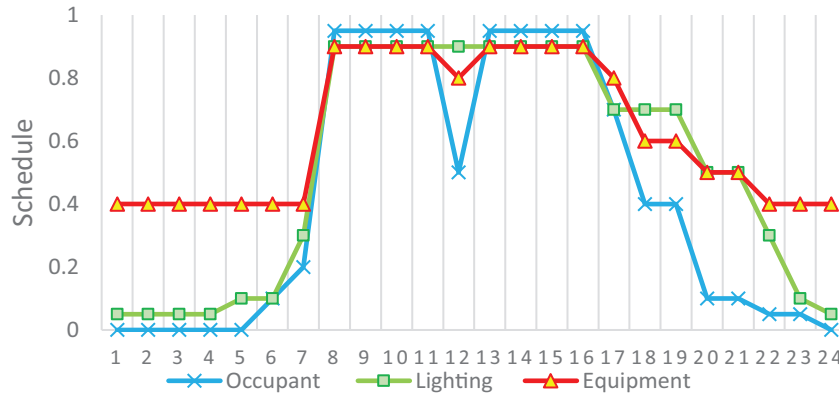
where N is the number of samples, Y represents the measurements, $\hat{Y}$ represents the predictions from GGMR.

For scenario 1, the coefficient of variance of the root mean square error (CV-RMSE) was 20.5% with Gaussian component evolution in the GGMR. For scenario 2 without RH and diffuse radiation as inputs, the CV-RMSE increased to 25.2%. However, the model for this input scenario still meets the modeling requirements of ASHRAE Guideline 14 [32]. For scenario 3 that uses recent historical data for PCB cooling rates, the prediction performance improved significantly with a CV-RMSE of 8.7% compared to 20.5% without the time history. Scenario 4 that also employed the 3-hour time history, but with an occupancy input and without RH and diffuse solar as inputs, had the best prediction accuracy with a CV-RMSE of 5.8%.

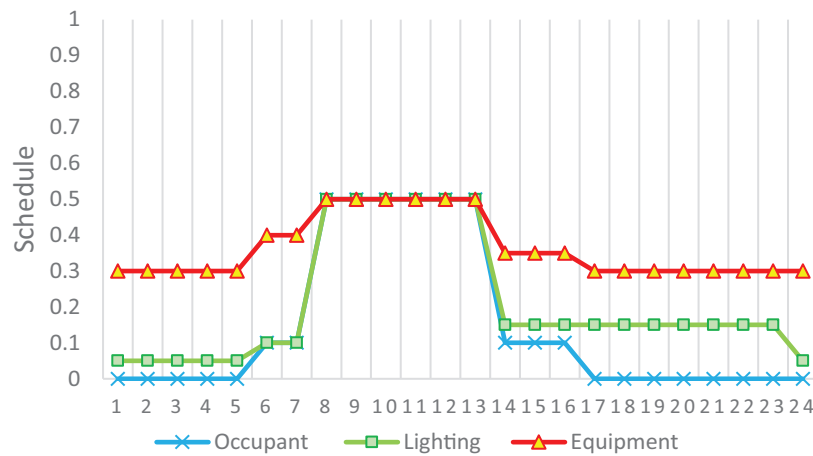
All of the results presented in Table 3 employed Gaussian component evolution associated with the GGMR method. The CV-RMSE was reduced from 14.8% with GMR (without the evolution) to 8.7% with the involvement of Gaussian components in GGMR. Fig. 5 contrasts the prediction difference before and after evolving the GMR model and shows predicted PCB cooling rates before evolving (GMR) from July 3 to July 9 and after evolving (GGMR) from July 10 to July 16 in comparison with EnergyPlus model simulation results when employing the scenario 3 inputs. The “Measured” PCB cooling rates were from EnergyPlus model simulation results. Both the GMR model (before evolving) and the GGMR model (after evolving) are compared to the “Measured” results in this figure. Before evolving, the predicted PCB cooling rates from the GMR model tend to underestimate both peak cooling demands and the cooling rates when the cooling rates were low. The hourly averaged offset between the predicted cooling rates from the GMR model and “Measured” cooling rates for 07/03–07/09 is −2.9% (−19.2 kW) of the hourly averaged “Measured” cooling rates. After evolving by updating the means, covariance matrices, and weight coefficients of Gaussian components, the predicted PCB cooling rates from the GGMR model matched the peak cooling demands relatively well. The hourly averaged offset between the predicted cooling rates from the GMR model and “Measured” cooling rates for 07/10–07/16 is 1.0% (6.38 kW) of the hourly averaged “Measured” cooling rates. Although predicted cooling rates for the PCB system from GGMR for scenario 3 and the simulation data from EnergyPlus model agreed well in general, the GGMR model substantially overestimated PCB cooling rates on Saturday morning as shown (circled) in Fig. 5.

The differences in internal heat gains between weekdays and weekends most likely account for inaccurate predictions of PCB cooling rates for GGMR during those weekend hours. The schedules for occupants, lighting, and equipment for weekdays and

**Fig. 5.** Comparison of predicted PCB cooling rates before evolving (GMR) and after evolving (GGMR) for scenario 3.



(a)



(b)

Fig. 6. Schedules for occupant, lighting, and equipment for the large-office building prototype EnergyPlus model (a) weekdays (b) Saturdays.(b)

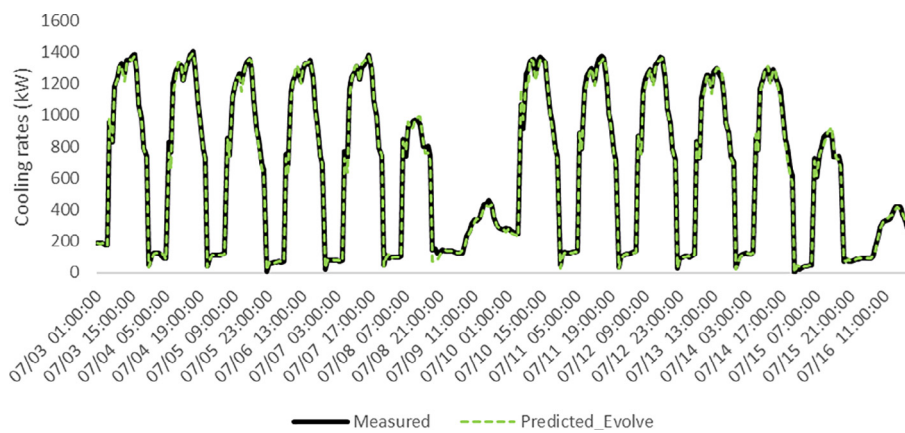


Fig. 7. Predicted PCB cooling rates after evolving (GGMR) for scenario 4.

Saturdays are shown in Fig. 6. Schedules for internal heat gains on Sundays were constant values (no occupants, lighting: 0.05, equipment: 0.3). From Fig. 6, it can be observed that schedule patterns for weekdays and weekends are consistent between occupants,

lighting, and equipment. Therefore, in scenario 4, the input parameters for the GGMR model included only the occupant schedule category. Fig. 7 shows comparisons of time series data between GGMR predictions in scenario 4 and EnergyPlus simulation results

for the two weeks. In comparison with the GGMR prediction in scenario 3, the GGMR model in scenario 4 agrees extremely well with the simulation data and can accurately predict PCB cooling rates for both weekdays and weekends.

Based on prediction results, we selected the input parameters from scenario 4 for the GGMR model. The input parameters in scenario 4 are OAT, ZAT, HVAC operation status, occupant schedule, direct solar radiation, and PCB cooling rates for the previous three hours.

4.2. Living Lab 1 measurement data

A similar procedure was used to select input parameters and performance parameters for the Living Lab PCB system. During the selection process for input parameters, we used a learning rate of $\beta = 0.001$ and the number of components of $M = 10$.

Trend data between April 1, 2021 and July 26, 2021 were collected via the Niagara system. Compared with the EnergyPlus simulation results, the prediction of PCB cooling rates for this real system is challenging due to the dynamics of the PCB system operating with actual control sequences including PI control loops in real-time. 5-minute time interval trend data for Living Lab 1 were obtained from Niagara and then preprocessed to remove any missing data and/or outliers. Measurement data obtained during equipment maintenance or when sensor communication issues occurred were also removed for baseline model development. Based on the current control sequence, the PCB system can provide both heating and cooling to the space. However, only cooling rates from PCB were considered when the supply water temperature was lower than the space temperature.

As shown in Fig. 8, significant and frequent fluctuations of cooling rates were observed for data with the five-minute interval. Instead of directly using the five-minute data, we averaged the five-minute cooling rates over each hour as a means of filtering the significant fluctuations as shown in Fig. 8. The hourly average cooling rates for the PCB system were used for GGMR model creation.

A similar approach was taken for the selection of input parameters as was used for the EnergyPlus simulation data. Based on the correlation coefficients summarized in Table 3 for use of hourly

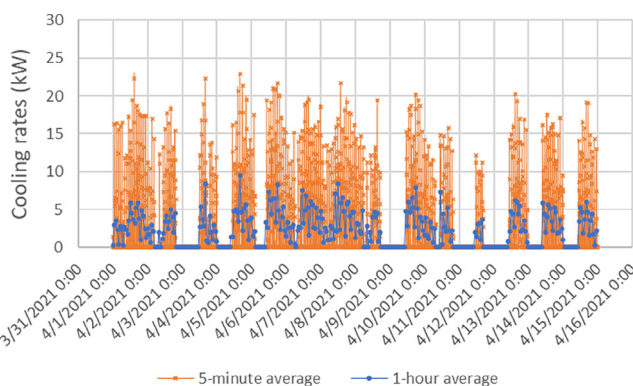


Fig. 8. Comparison of PCB cooling rates with 5-minute average and 1-hour average.

Table 4

Correlation coefficients between cooling rates of the passive chilled beam system, and potential input parameters for Living Lab 1. (OAT: outdoor air temperature; ZAT: zone air temperature).

	OAT	RH	ZAT	PCB valve	Chilled water valve	Double façade temperature
Cooling rates (kW)	0.47	-0.42	0.37	0.12	0.55	0.61

average data, the input parameters considered for cooling rate predictions are the outdoor air temperature, outdoor relative humidity, zone air temperature, average PCB control valve position, chilled water valve position, and double façade temperature. The training data for the PCB system were from April 01, 2021 to May 31, 2021 and the testing data were from June 1, 2021 to July 26, 2021.

Four scenarios with different input parameters were summarized in Table 4. The overall test performance results for each scenario were shown in Table 5. In scenario 1, OAT, outdoor relative humidity, ZAT, average PCB control valve position, chilled water valve position, and double façade temperature were selected as the input parameters for GGMR. For scenario 1, the CV-RMSE was 32.4% after the evolvement of Gaussian components with an R-square of 0.707 for testing data. The statistical results do not meet the requirement for CV-RMSE by ASHRAE Guideline 14 [32].

In scenario 2, we took a nested GGMR modeling approach for predicting cooling rates using predicted PCB water flow rates. The PCB cooling rates have a strong correlation coefficient with water flow rates (0.44) because the cooling rates are determined with an energy balance using water flow rates and the temperature difference between supply and return water. The supply water flow rate for the passive chilled beam system was first predicted using a Gaussian mixture model (GMR). With other input parameters in scenario 1 including OAT, outdoor relative humidity, ZAT, average PCB control valve position, chilled water valve position, and double façade temperature, the predicted chilled water flow was fed into a GGMR for the prediction of the PCB cooling rates. The flow rates in GMR were predicted using the PCB control valve position and differential pressure as inputs. The CV-RMSE for flow prediction was 4.5% with an R-square of 0.980. The CV-RMSE for cooling rate prediction in scenario 2 was reduced to 28.0%.

Similar to scenario 3 for predicting cooling rates of the PCB system in EnergyPlus simulation, historical data for PCB cooling rates of the previous three hours were also added to the GGMR model based on scenario 2. The prediction was improved with slightly reduced CV-RMSE and increased R-square. In scenario 4, OAT and outdoor relative humidity were removed from the input parameters. Statistical results for GGMR in scenario 4 were similar to the results in scenario 3. Input parameters in scenario 4 were selected for further testing. These parameters include ZAT, average PCB control valve position, chilled water valve position, double façade temperature, predicted flow rate, and historical cooling rates for the previous three hours.

5. Evaluation of performance parameters

In this section, we consider the impacts of performance parameters in the GGMR model training on performance prediction with the selected list of input parameters of scenario 4 for PCB systems in both simulation and living lab 1. These performance parameters include the number of components M , train data size N , and the learning rate β .

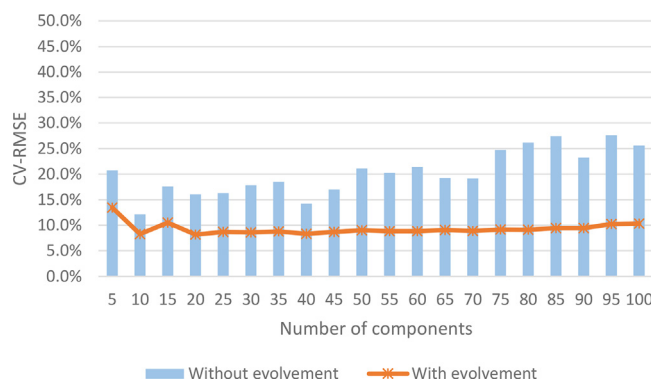
5.1. Number of components m and training data size N

The number of components affects the accuracy and complexity of Gaussian mixture models. Because of the evolvement of key

Table 5

List of input parameters for each scenario using Living Lab 1 measurement data.

List of Input Parameters	
S1	OAT, outdoor relative humidity, ZAT, average PCB control valve position, chilled water valve position, and double façade temperature
S2	OAT, outdoor relative humidity, ZAT, average PCB control valve position, chilled water valve position, double façade temperature, and predicted flow rate
S3	OAT, outdoor relative humidity, ZAT, average PCB control valve position, chilled water valve position, double façade temperature, predicted flow rate, and historical cooling rates for the previous three hours
S4	ZAT, average PCB control valve position, chilled water valve position, double façade temperature, predicted flow rate, and historical cooling rates for the previous three hours.

**Fig. 9.** The variation of CV-RMSE with the number of components in the Gaussian mixture model for EnergyPlus simulation with one-month training data.**Table 6**

Statistical analysis results for performance prediction of each scenario with Living Lab 1 measurement data.

PCB cooling rate	S1	S2	S3	S4
CV-RMSE	32.4%	28.0%	27.3%	27.4%
MBE	4.5%	4.5%	−0.8%	3.2%
R ²	0.707	0.79	0.809	0.805

parameters in growing Gaussian mixture regression models with measurement data, traditional model discrimination methods such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) methods [33] are not applicable for identifying the optimal number of components. Instead of using AIC or BIC, a range of numbers of components was tested for data from both EnergyPlus simulation and Living Lab 1. The number of components was selected based on the best statistical results.

The impacts of the number of components in growing Gaussian mixture regression models on testing accuracy for EnergyPlus simulation data are illustrated in Fig. 9 for one month of training data. Without the evolutionment of Gaussian components, the optimum number of components is 15. With the evolutionment in GGMR, the optimum number of components was 20.

Table 7

Statistical analysis results for performance prediction with increased training data size.

PCB cooling rate	EnergyPlus (N = 1-mon.)	EnergyPlus (N = 6-mon.)	LL1 (N = 2-mon.)	LL1 (N = 3-mon.)
Optimum number of components	20	60	15	20
CV-RMSE	4.6%	3.8%	25.8%	19.6%
MBE	−0.3%	−0.2%	−0.2%	−2.2%
R ²	0.997	0.998	0.833	0.921

Table 6 shows the impact of training data size on model prediction accuracy for both the simulation and measurement data. We observed that the optimum number of components varies with the number of training data points. The optimum number of components for GGMR was 60 with six months of training data in comparison with 20 components with one month of training data for EnergyPlus simulation. When measurement data from Living Lab 1 was used, the optimum number of components for GGMR was 15 with two-month training data while the optimum number of components was 20 with three-month training data. For both simulation and measurement data, the model prediction accuracy increased with the amount of training data. For the simulation data, increasing the training data size from one month to six months led to a decrease of CV-RMSE for cooling rate prediction from 4.6% to 3.8% and an increase in the R-square from 0.997 to 0.998. For the measurement data, the CV-RMSE for the prediction of PCB cooling rates decreased from 25.8% to 19.6% and the R-square increased from 0.833 to 0.921 by adding one more month of training data. Access to good-quality training data would improve the accuracy of the GGMR prediction.

5.2. Learning rate β

Learning rate is another key performance parameter in GGMR. We tested five different learning rates (0.0001, 0.001, 0.01, 0.1, and 0.2) with one-month training data and 20 Gaussian components using EnergyPlus simulation data. Those learning rates were tested with two-month training data and 15 Gaussian components using measurement data from Living Lab 1. Results for the impact of learning rate on testing accuracy are presented in Table 7.

The accuracy of PCB cooling rate prediction when the learning rate β was set as 0.001 and 0.01 is relatively higher than that with other tested learning rates. The selection of the learning rate in GGMR represents a trade-off between the updating speed of the Gaussian component and model stability. When the learning rate was increased to 0.1 and 0.2, the prediction accuracy of the GGMR model developed using EnergyPlus simulation data was compromised. The prediction with high learning rates became unstable and occasionally did not converge, requiring much more simulation time than when using smaller learning rates. The prediction accuracy was also significantly reduced with relatively large learning rates for measurement data from Living Lab 1. For example, the CV-RMSE was increased from 25.8% to 42.9% for PCB cooling rate prediction when the learning rate was increased from 0.1 to 0.2. A learning rate of 0.001 was selected for PCB cooling rate prediction for the testing data from EnergyPlus simulation and a learning rate of 0.01 was used for the testing data of Living Lab 1 measurement.

6. Discussion

6.1. Measurement data vs simulation data

The input parameters of the GGMR model for the data from EnergyPlus simulation include OAT, ZAT, operation status, occupant schedule, direct solar radiation, and PCB cooling rates for

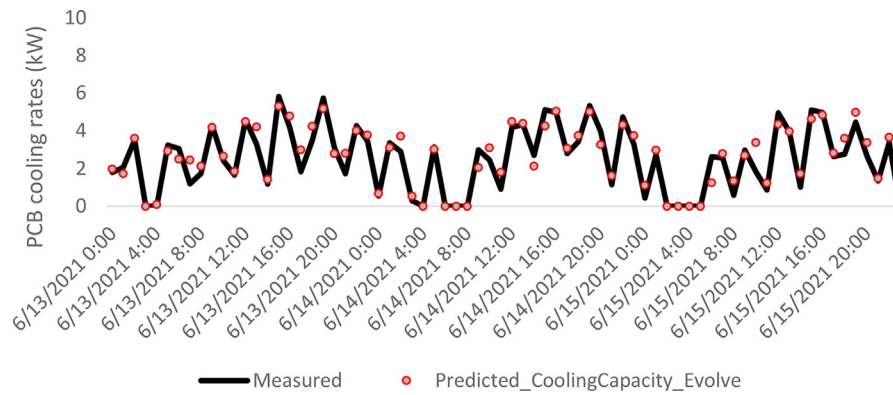
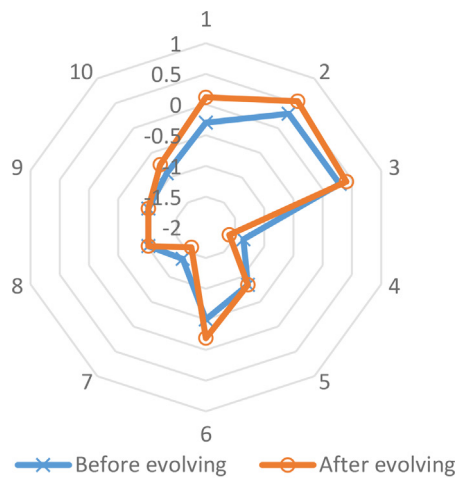


Fig. 10. The hourly PCB cooling rate prediction for testing data from Living Lab 1 measurements.

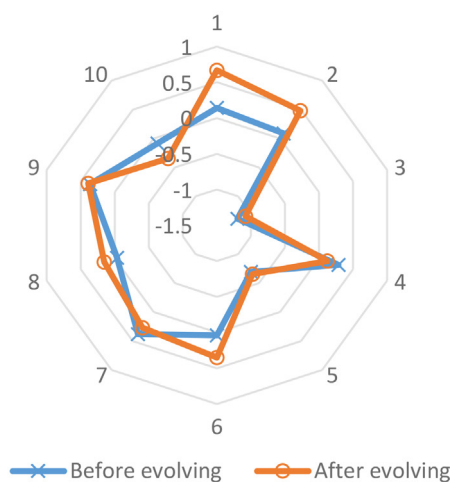
the previous three hours. The input parameters of the GGMR model for the data from Living Lab 1 measurements include ZAT, average PCB control valve position, chilled water valve position, double façade temperature, predicted flow rate, and historical cooling

rates for previous three hours. For both GGMR models, historical cooling rates in previous hours improved PCB cooling rate predictions. Since they have a direct effect on space cooling loads, both OAT and direct solar radiation were selected as input parameters for the model built from EnergyPlus simulation. Without measurement data for solar radiation, the double façade space temperature is a surrogate that indirectly reflects the effect of solar radiation heat gains on Living Lab 1. For the model developed from EnergyPlus simulation data, HVAC operation status was also used to describe the availability of the PCB system. The PCB system operates 24/7 in Living Lab 1, and therefore the operation status was not included as one of the input parameters.

Fig. 10 presents hourly PCB cooling rate predictions and measurements associated with testing data for Living Lab 1. Two months of training data, 15 Gaussian components, and a learning rate of 0.01 were used for developing the GGMR model using the Living Lab 1 measurement data. We observed that hourly cooling rates ramp up from the early morning and usually peak at around 15:00 or 16:00 in the late afternoon although there are some fluctuations of hourly cooling rates throughout the day. These fluctuations due to imperfect control of the system make hourly cooling rates for the PCB system difficult to predict. This is also the reason that the GGMR model developed from simulation data performed much better (CV-RMSE = 4.6% and $R^2 = 0.997$) than the GGMR model developed from measurement data (CV-RMSE = 23.7% and $R^2 = 0.867$). Furthermore, the GGMR model developed from measurement data using the nested GGMR approach is more complicated than that developed from building simulation data. The flow rates were first predicted by a GMR model and then the predicted flow rates from the GMR model were used as one of the



(a) component 3



(b) component 5

Fig. 11. Comparisons of mean vectors for two components (components 3 and 5) of the model before and after evolving for Living Lab 1.

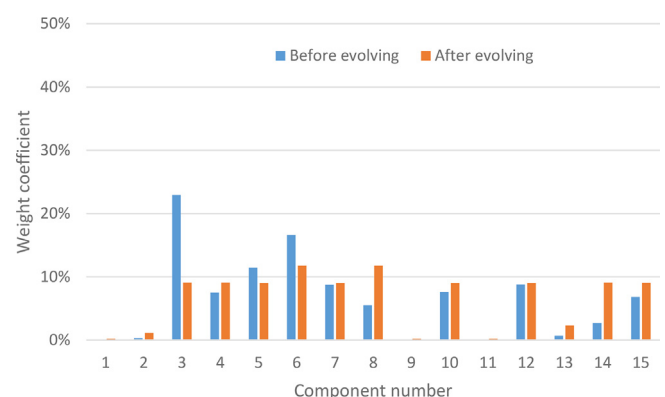


Fig. 12. The changes of weight coefficients before and after evolving for Living Lab 1.

Table 8

Statistical analysis results for performance prediction with different learning rates.

Learning rate		0.0001	0.001	0.01	0.1	0.2
EnergyPlus	CV-RMSE	4.6%	4.6%	4.7%	6.4%	11.6%
	MBE	−0.9%	−0.3%	−0.9%	0.2%	−0.2%
	R ²	0.997	0.997	0.997	0.994	0.981
Living Lab 1	CV-RMSE	26.3%	25.8%	23.7%	25.8%	42.9%
	MBE	0.08%	0.2%	0.5%	0.8%	−0.8%
	R ²	0.825	0.833	0.867	0.858	0.563

Table 9

Statistical analysis results for performance prediction with various climate zones.

	Miami (1A)	Houston (2A)	Phoenix (2B)	Atlanta (3A)	Los Angeles (3B)	Baltimore (4A)	Albuquerque (4B)
Number of components	20	25	25	20	40	40	25
CV-RMSE	4.6%	8.4%	5.2%	9.5%	4.8%	11.0%	7.7%
MBE	−0.3%	−1.1%	−0.4%	−0.9%	−0.4%	−1.3%	−1.1%
R ²	0.997	0.993	0.997	0.991	0.998	0.990	0.994

input parameters for predicting PCB cooling rates in the GGMR model.

Key parameters of Gaussian components including weight coefficients α_m , means μ_m , and covariance matrices σ_m of Gaussian components were continuously updated for every new data point during testing. Fig. 11 illustrates the changes of mean vectors for two components (components 3 and 5) of the model before and after evolving using the training data. Fig. 12 shows the changes of weight coefficients before and after evolving with testing data. The changes of weight coefficient ranged from 0.16% to 13.8%. Among the 15 components, components 3, 8 and 14 experienced relatively large changes of 13.8%, 6.2%, and 6.4%, respectively.

6.2. Simulation data from different climates

In addition to GGMR model development for Miami, FL (ASHRAE Climate 1A) based on EnergyPlus simulation data, GGMR models were developed for six other climate zones with relatively high cooling demands in the U.S. including Houston (2A), Phoenix (2B), Atlanta (3A), Los Angeles (3B), Baltimore (4A), and Albuquerque (4B). The GGMR models were created based on one-month training data and tested for the rest of 11 months using EnergyPlus simulation data. The learning rate was fixed at 0.001. In all the simulation models, the PCB cooling system accounted for the majority of the sensible cooling demands for the large office building prototype and VAV systems were only sized to account for latent loads and ventilation requirements.

As summarized in Table 8, the CV-RMSEs of GGMR models for predicting PCB cooling rates ranged from 4.6% to 11.0% for the seven climate zones with high R-squares all above 0.99. The input parameters for all the GGMR models were OAT, ZAT, operation status, occupant schedule, direct solar radiation, and PCB cooling rates for the previous three hours. It should also be noted that the optimum number of components varies with the climate zones. Table 9.

7. Conclusion and recommendation

In this study, the GGMR method was applied to predict cooling rates of passive chilled beam systems using data from real system measurements and building energy simulations. The key parameters in growing Gaussian mixture regression models, such as mean vector and covariance matrix for each Gaussian component, evolve as system performance over time due to seasonal variation or system degradation. Both CV-RMSE and MBE of the GGMR models for

PCB cooling rate prediction meet the requirements of ASHRAE Guideline 14 for hourly data prediction.

This case study demonstrated that GGMR is an effective evolving learning-based data-driven method for accurately predicting PCB system performance. The selected input parameters for the models are commonly monitored and can be trended. Baseline models representing the normal operation of a PCB system can be used for advanced applications such as fault detection and diagnosis and model predictive control. This GGMR method can potentially be applied to other high-performance complex systems in the built environment, such as radiant slabs and hybrid ventilation.

Although reasonable CV-RMSE and R-square for testing data were obtained using both simulation and measurement data, advanced features [29] of GGMR models could be further explored for complex HVAC systems. These features include training and creating new Gaussian components, splitting oversized Gaussian components, or merging Gaussian components with similarities. The core of a creating, splitting, and merging approach is to set up the mean vector and covariance matrix for newly generated components. Importantly, thresholds switching between updating, creating new components, splitting, and merging need to be selected and tested for each model. In addition, a hybrid modeling approach between evolving learning-based data-driven models (e.g. GGMR), physical models, and grey-box models (e.g. resistance-capacitance thermal network models) could be tested for complex systems that are used in the built environment. The hybrid approach could take advantage of the benefits of different methods for modeling building performance.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Liping Wang reports financial support was provided by National Science Foundation. Liping Wang reports a relationship with National Science Foundation that includes: funding grants.

Acknowledgments

This study was supported by the National Science Foundation EPSCoR Research Infrastructure program under Grant No. 1929209. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- [1] Kosonen, R. *Chapter 8 – Chilled Beams and Radiant Ceiling Systems*. In *Air Conditioning System Design*, R. Legg, Editor 2017, Butterworth-Heinemann. 151–166.
- [2] Z. Shi, D. Lai, Q. Chen, Performance evaluation and design guide for a coupled displacement-ventilation and passive-chilled-beam system, *Energy and Buildings* 208 (2020) 109654.
- [3] K. Roth et al., Chilled beam cooling, *ASHRAE Journal* 49 (9) (2007) 84.
- [4] J. Kim, A. Tzempelikos, J.E. Braun, Energy savings potential of passive chilled beams vs air systems in various US climatic zones with different system configurations, *Energy and Buildings* 186 (2019) 244–260.
- [5] S. Weidner, J. Doerger, M. Walsh, Cooling with less air, *ASHRAE Journal* 51 (12) (2009) 34–40.
- [6] P. Rumsey, J. Weale, Chilled beams in labs: eliminating reheat & saving energy on a budget, *ASHRAE Journal* 49 (1) (2007) 18.
- [7] Z. Shi, Z. Lu, Q. Chen, Indoor airflow and contaminant transport in a room with coupled displacement ventilation and passive-chilled-beam systems, *Building and Environment* 161 (2019) 106244.
- [8] W. Shan, D. Rim, Thermal and ventilation performance of combined passive chilled beam and displacement ventilation systems, *Energy and Buildings* 158 (2018) 466–475.
- [9] H. Koskela, H. Häggblom, R. Kosonen, M. Ruponen, Air distribution in office environment with asymmetric workstation layout using chilled beams, *Building and Environment* 45 (9) (2010) 1923–1931.
- [10] Betz, F. et al. Issues arising from the use of chilled beams in energy models. In *Proceedings of 5th National Conference of IBPSA-USA*. 2012.
- [11] J. Kim, A. Tzempelikos, W.T. Horton, J.E. Braun, Experimental investigation and data-driven regression models for performance characterization of single and multiple passive chilled beam systems, *Energy and Buildings* 158 (2018) 1736–1750.
- [12] C. Chen, W. Cai, K. Giridharan, Y. Wang, A hybrid dynamic modeling of active chilled beam terminal unit, *Applied Energy* 128 (2014) 133–143.
- [13] J. Kim, A. Tzempelikos, J.E. Braun, Review of modelling approaches for passive ceiling cooling systems, *Journal of Building Performance Simulation* 8 (3) (2015) 145–172.
- [14] P. Filipsson, A. Trüschel, J. Gräslund, J.-O. Dalenbäck, A thermal model of an active chilled beam, *Energy and Buildings* 149 (2017) 83–90.
- [15] B. Wu, W. Cai, H. Chen, A model-based multi-objective optimization of energy consumption and thermal comfort for active chilled beam systems, *Applied Energy* 287 (2021) 116531.
- [16] Z. Tian et al., A review of data-driven building performance analysis and design on big on-site building performance data, *Journal of Building Engineering* 41 (2021) 102706.
- [17] Y. Sun, F. Haghighat, B.C.M. Fung, Trade-off between accuracy and fairness of data-driven building and indoor environment models: A comparative study of pre-processing methods, *Energy* 239 (2022) 122273.
- [18] J. Wang et al., Data-driven model predictive control for building climate control: Three case studies on different buildings, *Building and Environment* 160 (2019) 106204.
- [19] L. Wang, R. Kubichek, X. Zhou, Adaptive learning based data-driven models for predicting hourly building energy use, *Energy and Buildings* 159 (2018) 454–461.
- [20] L. Zhang, Data-driven building energy modeling with feature selection and active learning for data predictive control, *Energy and Buildings* 252 (2021) 111436.
- [21] A. Srivastav, A. Tewari, B. Dong, Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models, *Energy and Buildings* 65 (2013) 438–447.
- [22] K. Li, Z. Ma, D. Robinson, J. Ma, Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering, *Applied Energy* 231 (2018) 331–342.
- [23] Y. Guo, H. Chen, Fault diagnosis of VRF air-conditioning system based on improved Gaussian mixture model with PCA approach, *International Journal of Refrigeration* 118 (2020) 1–11.
- [24] M. Karami, L. Wang, Fault detection and diagnosis for nonlinear systems: A new adaptive Gaussian mixture modeling approach, *Energy and Buildings* 166 (2018) 477–488.
- [25] N. Shimizu, H. Kaneko, Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design, *Materials & Design* 196 (2020) 109168.
- [26] M. Wei et al., Remaining useful life prediction of lithium-ion batteries based on stacked autoencoder and gaussian mixture regression, *Journal of Energy Storage* 47 (2022) 103558.
- [27] S. El Zaatari, W. Li, Z. Usman, Ring Gaussian Mixture Modelling and Regression for collaborative robots, *Robotics and Autonomous Systems* 145 (2021) 103864.
- [28] Sung, H.G., *Gaussian mixture regression and classification*, in *Statistics 2004*, Rice University: Houston, Texas.
- [29] Bouchachia, H. and C. Vanaret, *Incremental Learning Based on Growing Gaussian Mixture Models*. Vol. 2. 2011.
- [30] DOE. *Commercial Reference Buildings*. Available from: <http://energy.gov/eere/buildings/commercial-reference-buildings>.
- [31] ASHRAE, *ANSI/ASHRAE/IES Standard 90.1-2004 Energy Standard for Buildings Except Low-Rise Residential Buildings*, 2004, American Society of Heating, Refrigerating and Air Conditioning Engineers: Atlanta, Georgia.
- [32] ASHRAE, *ASHRAE Guideline 14: Measurement of Energy, Demand and Water Savings*, 2014, American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA.
- [33] P. Stoica, Y. Selen, Model-order selection: a review of information criterion rules, *IEEE Signal Processing Magazine* 21 (4) (2004) 36–47.