https://dx.doi.org/10.1016/j.ijrefrig.2022.12.027

Development of Data-Driven Performance Benchmarking Methodology for A Large Number of Bus Air Conditioners

Zhijie Chen[1], Fangzhou Guo[1], Fu Xiao[1,2]*, Xiaoyu Jin[1], Jian Shi[3], Wanji He[3]

[1] Department of Building Environment and Energy Engineering, Hong Kong Polytechnic University

[2] Research Institute for Smart Energy, The Hong Kong Polytechnic University

[3] Shenzhen Envicool Technology Co., Ltd., Shenzhen, China

Abstract

Bus air conditioners (ACs) are responsible for providing a comfortable cabin environment for passengers. Identifying the bus ACs with degraded performance from a large number of city buses is a critical and challenging task in the development of smart cities. This study developed a data-driven benchmarking methodology to detect anomalous operations with degraded energy performance from a large number of bus ACs. For each target AC to be benchmarked, its similar operation data in other ACs, termed comparable peer samples, are first identified by a Long-Short-Term-Memory (LSTM) autoencoder-based similarity measurement method. The comparable peer samples are then used to develop a LSTM network-based reference model for predicting the power consumption of the target AC. A key energy performance indicator termed power consumption ratio (*PCR*) is defined for the target AC as the ratio of its measured power to the predicted power. Statistical analysis-based trend and change detection algorithms are designed to identify a trend or change of *PCR* over a few days for anomalous detection. To validate the benchmarking methodology, two fault experiments were conducted in field-operating bus ACs, and the results show

encouraging potentials of the proposed methodology for health monitoring of a large number of ACs serving the city bus fleet.

Keywords: Bus air conditioner, Benchmarking, Multivariate time series analysis, Deep learning.

Nomenclature

Symbol

b	Bias vector of the neural network
d	Distance
h	Context vector of the autoencoder
fan	Fan speed ($\mathrm{r}\cdot\ \mathrm{min}^{-1}$)
fre	Compressor frequency ($\mathrm{r}\cdot\ \mathrm{min}^{-1}$)
power	Power consumption (W)
S	Degree of EEV opening
W	Weight matrix of the neural network

Abbreviation

AC	Air Conditioner
DTW	dynamic time warping
EUI	Energy use intensity
GHG	Greenhouse gas
IEA	International energy agency
IoT	Internet of things
KPI	Key Performance Index
LSTM	Long-short term neural network
PCA	Principal component analysis
PCR	Power consumption ratio
SAX	Symbolic Aggregate Approximation
VRF	Variable refrigerant flow
XGBoost	eXtreme Gradient Boosting

Subscripts

cond (c)	Condenser
comp	Compressor
dec	Decoder
evap (e)	Evaporator
EEV	Electronic Expansion Valve
enc	Encoder
indoor	Indoor (cabin)
max	Maximum value
measured	Measured value
outdoor	Outdoor
ref	Refrigerant
predicted(pre)	Model Predicted Output

1. Introduction

According to the estimation of International Energy Agency (IEA), worldwide energy consumption of vehicle air conditioners (ACs) reached 2 million barrels of oil equivalent per day (Mboe/d) in 2019 (IEA, 2019). This number would triple to 5.8 Mboe/d in 2050, if no measures were taken in the future to improve the energy efficiency and reduce the energy waste caused by various faults. Meanwhile, the greenhouse gas (GHG) emission from vehicle ACs accounts for 420 million tons of CO_2 equivalent ($MtCO_2$-eq) per year. 70% of the GHG emission is caused by energy consumption of the vehicle ACs, while the rest is the consequence of refrigerant leakage, a common fault in vehicle ACs (Zhang et al., 2017). Besides environmental concerns, the large energy consumption of vehicle ACs also intensifies the range anxiety of electrical vehicles (Wang et al., 2015). IEA (2019) shows that AC operation can reduce the driving range of electric vehicles by up to 50% on hot and humid days.

Continuously monitoring the energy performance of ACs serving city buses, identifying the ones with low efficiency and arranging maintenance in a timely manner can reduce the energy waste caused by AC faults and enhance passengers' experience riding on public transportation. The 4^{th} industry revolution boosts the fast development of internet of things (IoT), smart sensor and big data analysis technologies, which enable remotely monitoring the operation of a large number of public transport vehicles (Zhao et al., 2017). However, due to a lack of data analytical methods, anomalies are still identified by experts manually. Advanced data analytical methods are required to analyze the large number of data collected from public transport ACs.

Benchmarking is a process of comparing performance against practices of the same nature, under the same circumstances and with similar measures (Schaad and Hofer, 2020). The most common benchmarking we experience in our daily life is the energy efficiency labels on household appliances, for example, the EnergyStar label in America (Brown et al., 2002). Before AC manufacturers launch their products into the market, their ACs should be tested in psychrometric laboratories according to certain standards such as GB/T 7725 for residential AC in China (Yuanxia et al., 2011) and AHRI 210/240 in America (Cheng et al., 2021). In order to fairly benchmark the products, these standards include detailed regulations for sensor accuracy, indoor/outdoor temperature settings, humidity settings etc. However, laboratory experiments-based benchmarking has its own disadvantages. First, high cost of these experiments means it can only test a few selected samples. Second, after the products are launched, faults and aging during operation may lead to the deviation of performance from benchmarking results in labs (Wang et al., 2022). As a result, an increasing number of researchers are focusing on benchmarking the performance of ACs using real operation data.

Attention should be paid to three aspects during the development of benchmarking framework (Zhou et al., 2019). First, key performance index (*KPI*) should be defined considering the operation performance can be evaluated from a few aspects including energy consumption, CO_2 emission and passenger comfort level. Second, the type of benchmarking should be determined from previous-performance benchmarking, intended-performance benchmarking, and peer-performance benchmarking. Intended-performance benchmarking compares the performance of an entity against some standard models. These models are usually white-box models built with specific physical parameters (Fumo et al.,

2010). Intended-performance benchmarking may yield meaningful results but cannot be applied to massive entities considering the cost of model development (Li et al., 2014). Instead, peer-performance benchmarking compares the performance of an entity against its peers, which has the advantage of both quality and scalability (Zhou et al., 2019). Third, the impact factors which may influence *KPI* values should be identified, and their impacts should be eliminated during the benchmarking process. Model-based prediction and clustering are two common methods to eliminate the influence of impact factors.

Model-based prediction is the core idea behind most benchmarking methods. In this approach, the relationship between the impact factors and the *KPI* is described by a regression model using data from all benchmark entities, regarded as the reference model. Then, the measured *KPI* of each benchmark entity is compared with the prediction of the reference model. Usually, the regression model is developed by data-driven methods such as polynomial regression, stochastic frontier analysis (Ding and Liu, 2020), quartile regression (Roth and Rajagopal, 2018) and random forest (Jin et al., 2022).

Apart from establishing regression models for impact factors and *KPI*, clustering is also used to eliminate or minimize the influence of some impact factors on *KPI*. The purpose is to cluster the benchmark entities that share similar values based on the impact factors, such that performance evaluation and comparison can be conducted within each cluster. Note that *KPI* is not compared between entities in different clusters. As an example, Zhan et al. (2020) considered that building energy consumption (i.e. the *KPI*) is affected by various operational patterns caused by different occupant behaviors. Thus, the authors first categorized buildings into different clusters according to their daily energy usage pattern,

and then compared the energy use intensity (*EUI*) of different buildings within each cluster. As a result, the impact of operational patterns on *EUI* is minimized by clustering.

Although relatively adequate research effort has been made on building energy performance benchmarking, such as commercial buildings (Chung, 2012) and multi-family complexes (Jeong et al., 2016), literature regarding AC energy performance benchmarking is limited. Qian et al. (2021) conducted a comprehensive analysis on the operation data collected from a large number of variable refrigerant flow (VRF) systems installed in residential and office buildings in China. The authors used the ideal coefficient of performance (ICOP), cycle duration and load ratio as KPIs to evaluate the VRF system performance in different climate zones, building types and operating conditions. Their research outputs are valuable for building HVAC design purpose. Liu et al. (2017) adopted decision tree to benchmark a VRF system based on its historical performance. The historical data was clustered according to the indoor/outdoor temperature and load ratio. Then, benchmarks were established based on the median value of power consumption in each cluster. Zhou et al. (2019) established a peer-performance benchmark framework for multiple residential ACs. The impact factors included room area, outdoor temperature and humidity, indoor temperature, indoor setting point and duration of an operation cycle. Wang et al. (2022) used eXtreme Gradient Boosting (XGBoost) to establish the relationship between power consumption and outdoor temperature, indoor occupant number for 129 unitary ACs installed in university dormitories. The Shapley value between outdoor temperature and power consumption was adopted as *KPI* to distinguish the low-efficiency ACs from normal ACs.

Research on energy performance benchmarking of vehicle ACs has not been found. Compared with residential ACs, public transport vehicle ACs have highly dynamic operations and frequent "on-off" operation cycles due to frequent door open/close, varying number of passengers and intermittent shading effect on road (Chen et al., 2022). The majority of the previous research on residential ACs only adopts steady-state operation data, which is inappropriate to evaluate the dynamic interaction between vehicle ACs and environment (Dhillon et al., 2021).

A data-driven reference model is developed based on the comparable peer samples to predict the power consumption at the operating conditions of the target AC. After that, a key performance indicator termed power consumption ratio (*PCR*) is defined for the target AC as the ratio of its measured power to the predicted power in each day. Finally, through trend and change detection, the target AC is considered becoming anomalous when a trend or change of *PCR* over a few days is detected.

This paper proposes a benchmarking methodology for a large number of bus ACs. The benchmarking methodology first integrates deep learning-based multivariate time series analysis to identify comparable peer samples with similar dynamic operating conditions. Then, for a target AC to be benchmarked, its comparable peer samples identified from other systems are used to develop a reference model. After that, a key performance indicator termed power consumption ratio (*PCR*) is computed based on the measured performance of the target AC and the predicted performance from the reference model. Finally, trend and change point detection algorithms are adopted to identify any change or trend of *PCR*, which is possibly caused by emerging faults or equipment aging. When a change or trend is recognized, alerts will be sent to maintenance crew in time.

The rest of this paper is organized as follows. Chapter 2 discusses the general design of the proposed benchmarking methodology and critical challenges that need to be tackled. Chapter 3-5 present the detailed techniques used in the benchmarking methodology. In Chapter 6, the benchmarking methodology is validated by data collected from field-operating bus fleet ACs. Finally, a few conclusions are drawn in Chapter 7.

2. Overview of the benchmarking methodology

Before introducing the proposed benchmarking methodology, this chapter first addresses four critical challenges, namely the definition of *KPI*, processing of multivariate time series, identification of comparable peer samples, and analysis of benchmarking results. Then, an overview of the benchmarking methodology is presented.

2.1 Critical challenges in performance benchmarking of a large number of bus ACs

This section will detail four critical challenges in performance benchmarking of a large number of bus ACs. The first one is the determination of *KPI* index for bus ACs. *KPI* should reflect the operation performance of the benchmark entities. Although a few metrics have been proposed in literature to evaluate the energy performance of AC such as the Seasonal Energy Efficiency Ratio (Jain et al., 2021), these metrics are difficult to be applied to field operation data of bus AC, mainly because of the lack of measurements of cooling capacity.

The second challenge is the data structure. As mentioned in the introduction, due to the highly dynamic operation of bus ACs, the basic element of data analysis in this research is multivariate time series representing each AC “on-off” cycle. Instead of analyzing each data point separately, the full dataset is segmented and data in each “on-off” cycle is

processed as a whole since the dynamics of system operation are embedded in the interrelationship among time series. These multivariate time series with different length will be benchmarked in this research. When one “on-off” cycle is benchmarked, it will be referred as benchmark sample while the others will be referred as peer samples.

The third challenge is that, to ensure fair benchmarking, the performance comparison should be carried out under comparable or similar operating conditions (i.e. operation cycles with similar dynamics of working environment and control input). Thus, similarity measurement among these multivariate time series with different length is necessary. However, it poses great challenge to the development of benchmarking methodology.

Extensive efforts have been made in previous research to measure the distance between data points, including the Euclidean distance, the Chebyshev distance, the Manhattan distance etc. (Li et al., 2022). The Euclidean distance can also be used as a metric to evaluate the similarity between univariate time series with the same length (Zhan et al., 2020). For univariate time series with different length, researchers propose a few methods to measure their similarity including dynamic time warping (DTW) (Iglesias and Kastner, 2013) and Symbolic Aggregate Approximation (SAX) (Habib et al., 2016). However, very few research focuses on the similarity measurement between multivariate time series, especially when the time series are in different length.

Li (2019) proposed a method combining principal component analysis (PCA) and K-means to cluster a number of multivariate time series with different length. Although this method can identify whether two time series are in the same cluster, the similarity between

multivariate time series still cannot be evaluated through this method. Besides, the number of clusters should be pre-defined, which is a drawback inherited from K-means.

The fourth challenge is to analyze the calculated KPI values. The majority of benchmarking methods proposed in previous researches only compared the value of *KPI* among different benchmark entities. However, the change of *KPI* with time also contains critical information of system performance. Therefore, to actively monitor the performance of bus ACs, statistical methods including trend and change detection algorithms are adopted to analyze how *KPI* value evolves with time and alerts are generated to notify the maintenance crew if anomalies are detected during ACs' operation.

2.2 Major steps of the benchmarking methodology

The purpose of establishing the benchmarking methodology for a large number of bus ACs is to evaluate their operation performance in order to identify the ones with relatively low efficiency. The typical system structure and IoT sensor locations of bus ACs are shown in Fig.1. In view of the challenge of defining *KPI* for field operating bus ACs, power consumption is used as *KPI* to evaluate the overall energy performance of bus ACs. Note that overall power consumption includes the power of the compressor, the indoor fan, and the outdoor fan.

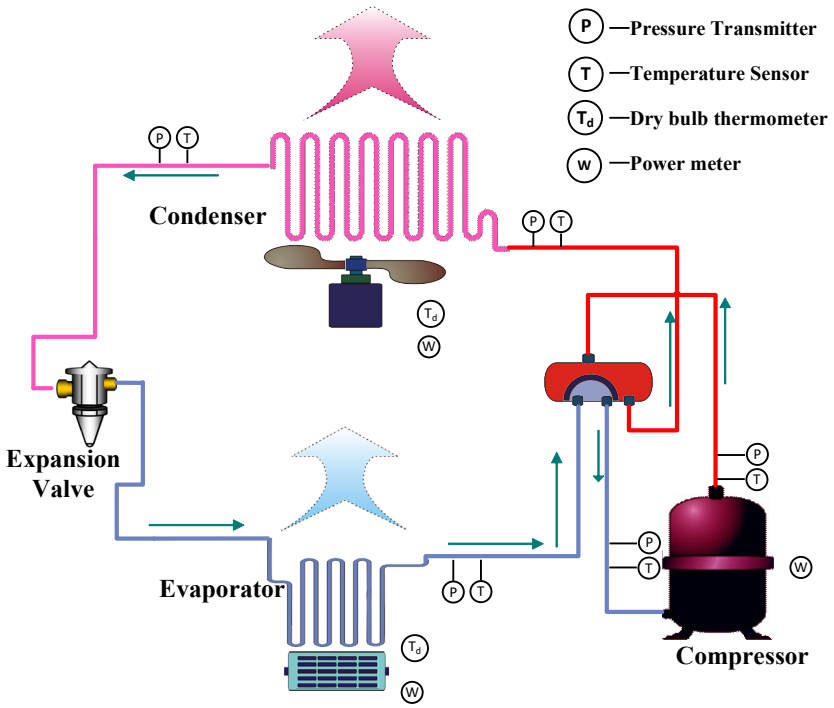

Fig. 1 System structure and IoT sensor locations of the studied vehicle AC

During operation, a few impact factors may influence the power consumption of vapor compression cycle, including the indoor/outdoor environment and control inputs. For instance, the power consumption is generally higher as the outdoor temperature increases. Different control input, like variable or constant compressor speed may also influence the energy efficiency, but these situations are not caused by faults or deficiencies. Therefore, the influence of those impact factors should be eliminated or minimized in the benchmarking process. In this research, two parameters representing environmental conditions (i.e. indoor and outdoor air temperature (T_{indoor} , $T_{outdoor}$)) and four parameters indicating control inputs (i.e. compressor frequency f_{comp}, EEV opening S_{EEV}, indoor fan speed fan_{indoor} and outdoor fan speed $fan_{outdoor}$) are identified as impact factors. The hypothesis is that power consumption of bus ACs under the same operation condition (represented by the abovementioned six impact factors) should be very much similar.

The basic idea behind the proposed benchmarking methodology is depicted in Fig.2. For each bus AC, operation data is generated during its daily operation. The first step to benchmark a bus AC is to find its comparable peer operation samples by analyzing their dynamic working conditions and control inputs. Through a comparable peer sample selection mechanism, comparable operation samples from peer ACs are selected and a reference model can be developed. To actively monitor the performance of the bus ACs, one reference model will be built for each bus AC on each day.

The inputs of reference model include the six impact factors identified in the above discussion while the output of the reference model is overall power consumption of AC. Since reference models are developed based on the selected comparable peer operation samples, data-driven techniques are more suitable.

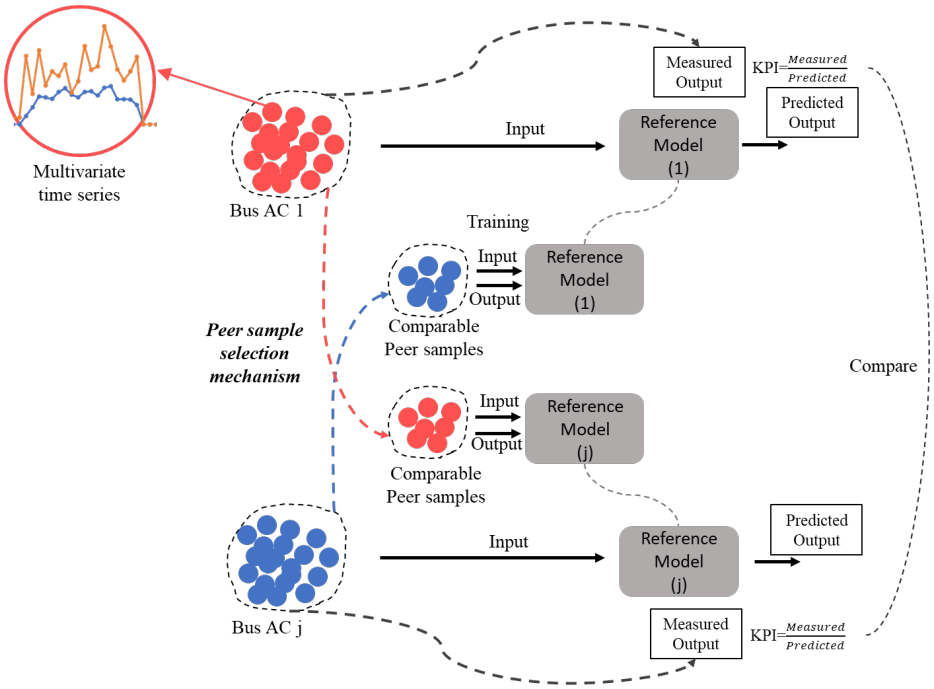

Fig. 2 Basic idea of the proposed benchmarking methodology

Finally, power consumption ratio (*PCR*) is computed to evaluate the energy efficiency of AC system on each day. *PCR* is defined as the ratio of the measured energy consumption of an AC system to the predicted energy consumption using its reference model, as expressed in Equation 1:

$$PCR_{i,j} = \frac{\sum_{t=1}^{t=T} power_{measured,t}}{\sum_{t=1}^{t=T} power_{predicted,t}} \quad (1)$$

where the subscript i is the index for the day, j is the ID of bus ACs and t represents the t^{th} operation cycle of No. j bus AC on the i^{th} day, and T refers to the total number of operation cycles on the i^{th} day.

The major steps of the proposed benchmarking methodology are depicted in Fig. 3, namely data preprocessing, multivariate time series similarity comparison based on Long short-term neural network (LSTM)-autoencoder, reference model development based on comparable peer samples, and statistical analysis of *PCR* for anomaly detection.

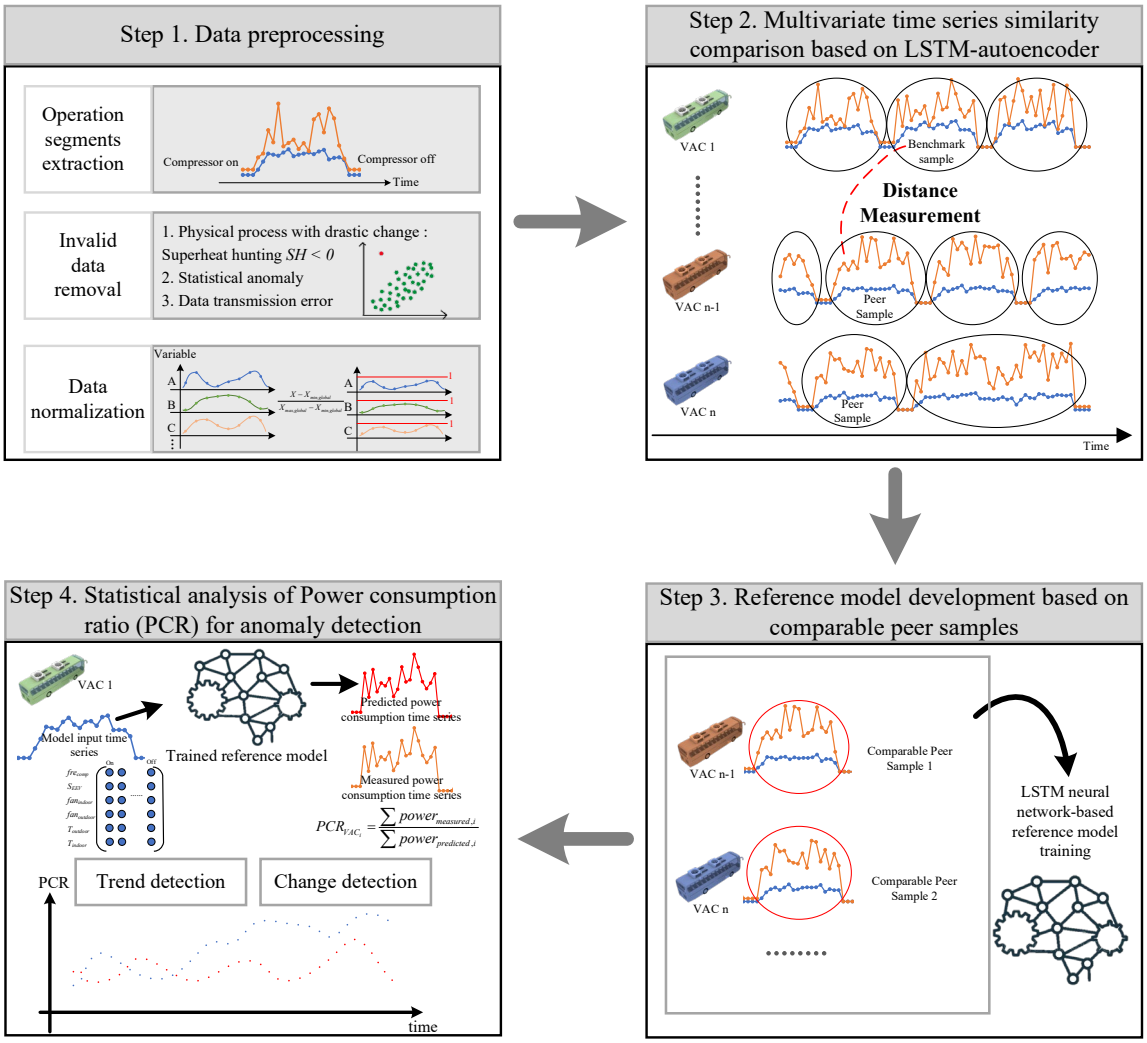

Fig. 3 Detailed steps of the proposed benchmarking methodology

During the data preprocessing stage, multivariate time series representing AC "on-off" operation cycles are extracted from the raw measurement data. Then, data cleaning and normalization are conducted to prepare the data for further analysis.

Multivariate time series similarity comparison based on LSTM-autoencoder is designed to analyze all these "on-off" operation cycles and identify comparable peer samples for each benchmarking sample. After that, reference models based on comparable peer samples are

established to predict the power consumption of ACs, and the predicted values are compared to the measured ones.

Lastly, statistical analysis is conducted to identify the change of *PCR* for each bus AC, which could indicate operational faults or degradation. In this study, the change of *PCR* is identified by an abrupt change detector and a trend detector.

3. Similarity measurement of multivariate time series based on LSTM-autoencoder

After data preprocessing, the multivariate time series data segments of impact factors from all the benchmark entities are obtained in the form of Equation 2, which is treated as the basic sample element in the following research. Similarity between these multivariate time series is compared in order to locate peer samples that share similar dynamic working conditions and control inputs for benchmark.

$$\text{Operation segment} = \begin{array}{c|cccc} & \textit{Variable I} & \textit{Variable II} & .... & \textit{Variable N} \\ \textit{Time index 1} & a_{11} & a_{12} & .... & a_{1n} \\ \textit{Time index 2} & a_{21} & a_{22} & .... & a_{2n} \\ .... & .... & .... & .... & .... \\ \textit{Time index n} & a_{n1} & a_{n2} & .... & a_{nn} \end{array} \tag{2}$$

3.1 LSTM-autoencoder

As a promising deep learning structure for handling multivariate time series, LSTM-autoencoder structure is adopted in this research to locate comparable operation samples with similar dynamics of impact factors from peer ACs. As shown in Fig.4, autoencoder is

a typical unsupervised machine learning method. Its architecture contains two parts: the encoder and the decoder. The encoder part attempts to transform the input of the neural network into a hidden state called context vector. The process happening in encoder can be expressed in Equation 3, where w_{enc}, b_{enc} and σ_{enc} refer to the weights, bias and activation functions respectively for the encoder.

$$h_x = Encoder(x) = \sigma_{enc}(W_{enc}x + b_{enc}) \tag{3}$$

The objective of decoder is to reconstruct the original input from the context vector generated by the encoder. Computation in the decoder can be expressed in Equation 4, where w_{dec}, b_{dec} and σ_{dec} refer to the weights, bias and activation functions respectively for the decoder.

$$\hat{X} = Decoder(h_x) = \sigma_{dec}(W_{dec}h_x + b_{dec}) \tag{4}$$

The reconstruction error between the input X and output $\hat{X}$ of the autoencoder neural network is minimized in the training process. The loss function is defined in Equation 5, where mean square error is adopted to evaluate the reconstruction error:

$$Minimizeloss = MSE(X, \hat{X}) \tag{5}$$

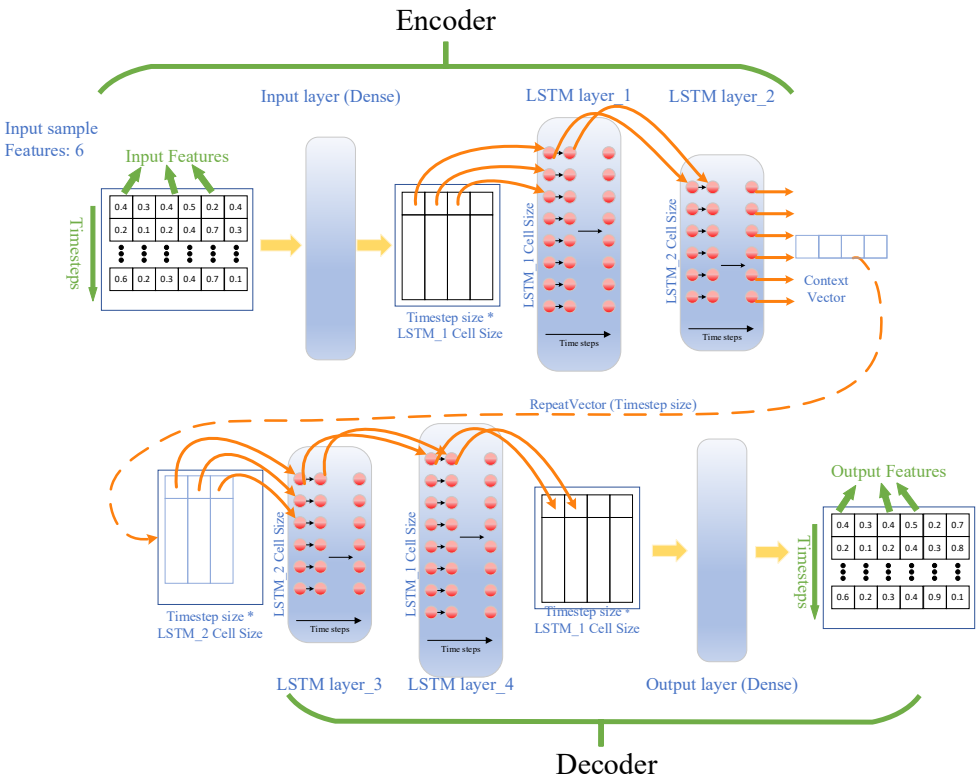

Fig. 4 LSTM-autoencoder based context vectors extraction for multivariate time series

3.2 Distance calculation based on context vector

The context vector extracted from the autoencoder contains important information of the input data and can be used as a good representation of the input data. LSTM neural network has been widely adopted for handling time series (Li et al., 2021). When LSTM is combined with autoencoder as shown in Fig.4, the LSTM-autoencoder can transform multivariate time series with different length into context vectors of the same size. Then, the Euclidean distance between context vectors can be used to measure the distance between the original multivariate time series. A smaller distance between two context vectors indicates the dynamics of their corresponding multivariate time series share more similarity. Calculation of the distance between two multivariate time series can be

expressed in Equation 6, where X and Y are two multivariate time series and h_x and h_y are their corresponding context vectors obtained by the LSTM-autoencoder.

$$dist(X,Y) \approx dist(h_x,h_y) = \sqrt{\sum_{i=1}^{n}(h_{x,i}-h_{y,i})^2} \tag{6}$$

In the proposed benchmark methodology, peer samples that have the similar dynamics of working environment and control inputs can be identified for each benchmark sample, which is accomplished by transforming the original multivariate time series into corresponding context vectors through LSTM-autoencoder and measuring the Euclidean distance of context vectors. The whole LSTM-autoencoder adopts the symmetric structure in this research. As shown in Fig.4, LSTM layer 1-4 have 64,32,32 and 64 neurons respectively. All LSTM layers use ReLU as the activation function (Schmidt-Hieber, 2020), while the output layer uses the sigmoid function to ensure the output ranges from 0 to 1.

4. Data-driven reference model development based on comparable peer samples

Chapter 3 details the similarity measurement between multivariate time series using the context vectors of LSTM-autoencoder. After that, data-driven reference models are developed based on the selected comparable peer samples to predict how much power would peer ACs consume under the same dynamic environmental conditions and control inputs. In this way, the daily power consumption of each bus AC can be evaluated.

4.1 Selection of comparable peer samples

The first step of peer sample selection is to calculate the distance matrix describing the similarity of the impact factors between daily operation cycles of bus ACs. As discussed in Section 2.2, the impact factors include 6 variables, namely indoor and outdoor temperature, compressor frequency, EEV opening, indoor fan speed, and outdoor fan speed. The distance matrix can be expressed in Equation 7. Each sample in this matrix refers to a multivariate time series and the distances among these samples are calculated based on the method proposed in Chapter 3. The distance matrix is symmetric, i.e. $d_{i,j}=d_{j,i}$.

		Bus AC 1			*Bus AC 2*				*Bus AC n*		
		Sample 1	*Sample 2*	*Sample 3*	*Sample 4*	*Sample 5*	*Sample 6*	....	*Sample n*	*Sample n + 1*	*Sample n + 2*
Bus AC 1	*Sample 1*	/	/	/	$d_{1,4}$	$d_{1,5}$	$d_{1,6}$		$d_{1,n}$	$d_{1,n+1}$	$d_{1,n+2}$
	Sample 2	/	/	/	$d_{2,4}$	$d_{2,5}$	$d_{2,6}$		$d_{2,n}$	$d_{2,n+1}$	$d_{2,n+2}$
	Sample 3	/	/	/	$d_{3,4}$	$d_{3,5}$	$d_{3,6}$		$d_{3,n}$	$d_{3,n+1}$	$d_{3,n+2}$
Bus AC 2	*Sample 4*	$d_{4,1}$	$d_{4,2}$	$d_{4,3}$	/	/	/		$d_{4,n}$	$d_{4,n+1}$	$d_{4,n+2}$
	Sample 5	$d_{5,1}$	$d_{5,2}$	$d_{5,3}$	/	/	/		$d_{5,n}$	$d_{5,n+1}$	$d_{5,n+2}$
	Sample 6	$d_{6,1}$	$d_{6,2}$	$d_{6,3}$	/	/	/		$d_{6,n}$	$d_{6,n+1}$	$d_{6,n+2}$
....	....							/			
Bus AC n	*Sample n*	$d_{n,1}$	$d_{n,2}$	$d_{n,3}$	$d_{n,4}$	$d_{n,5}$	$d_{n,6}$	$d_{n,7}$	/	/	/
	Sample n + 1	$d_{n+1,1}$	$d_{n+1,2}$	$d_{n+1,3}$	$d_{n+1,4}$	$d_{n+1,5}$	$d_{n+1,6}$	$d_{n+1,7}$	/	/	/
	Sample n + 2	$d_{n+2,1}$	$d_{n+2,2}$	$d_{n+2,3}$	$d_{n+2,4}$	$d_{n+2,5}$	$d_{n+2,6}$	$d_{n+2,7}$	/	/	/

(7)

After the calculation of distance matrix, two peer samples with the shortest distances from the operation data of every bus AC in the last 14 days are selected as comparable peer samples for each benchmark sample.

4.2 Development of data-driven reference models

The energy performance of a bus AC is benchmarked based on the peer samples that share the most similar dynamics of working environment and control inputs. After the process of comparable peer sample selection, a reference model is established for each bus AC on each day using the selected peer samples. The model inputs are the 6 impact factors, and

the model output is the overall power consumption of bus AC. Because the inputs and outputs are both time series, multiple layers of LSTM cells are stacked to build the reference model, as shown in Fig. 5. After training based on comparable peer samples, the developed reference models are used to predict the benchmark of overall power consumption for a bus AC. Finally, *PCR* is calculated according to Equation 1 in Chapter 2.

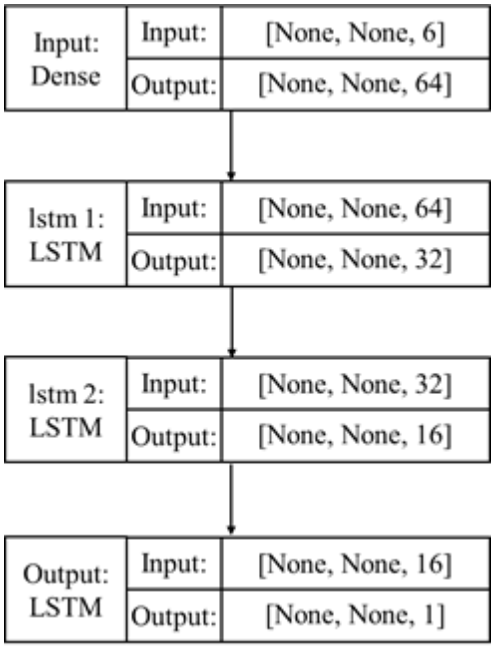

Fig. 5 Structure of the data-driven reference model based on the LSTM neural network

5. Statistical analysis of PCR for anomaly detection

In order to identify ineffective behavior which may be caused by faults, further analysis of the benchmarking results is necessary. A typical way of identifying anomalies is to set thresholds for benchmarking results. If the benchmarking results exceed the threshold, then it is deemed as abnormal operation and requires further analysis before scheduling maintenance. The thresholds should be determined by balancing the sensitivity of the algorithm and the cost of maintenance. Anomaly detection algorithms with high sensitivity

may result in frequent false alarms while low sensitivity may result in more energy waste and lower passenger comfort level.

Actually, it is very hard to determine a certain threshold of *PCR* for benchmarking a large number of field operating bus ACs. Many factors may influence the energy performance of bus ACs. Though bus ACs are generally mass-produced with the same design parameters and similar manufacturing quality, the subsequent installation and commissioning process can still result in differences in performance between systems. Evaluation of the manufacturing process or the installation process, however, is beyond the scope of this study. This research focuses on identifying gradual trends or abrupt changes of *PCR* for each system, because a change or trend of *PCR* usually indicates a deviation of operation performance in AC, which require further diagnosis and analysis.

5.1 Trend detection of PCR based on Mann-Kendall test

The non-parametric Mann-Kendall trending test is adopted in this research to assess the monotonic trend of each bus AC's *PCR* over time. The null hypothesis of Mann-Kendall test is that no monotonic trend exists in the time series, and the alternative hypothesis is that the data follows a trend. The Mann-Kendall test has been widely used in meteorology (Gocic and Trajkovic, 2013) and hydrology (da Silva et al., 2015) research for time series trend analysis. It can deal with missing data and is a non-parametric method (i.e. Any assumption of the data distribution is not required). The Mann-Kendall test uses statistic S to compare each data point and its subsequent data points in a time series $x_1, x_2, \ldots x_n$. If subsequent data point x_j is larger than x_i, then S is increased by 1. If subsequent data point

x_j is smaller than x_i, then S is decreased by 1. Otherwise, S does not change. The calculation of statistic S can be expressed by the Equation 8 and 9.

$$S = \sum_{i=1}^{n-1} \sum_{j=k+1}^{n} sgn(x_j - x_i) \tag{8}$$

$$sgn(x_j - x_i) = \begin{cases} +1 & if\,(x_j - x_i) > 0 \\ 0 & f\,(x_j - x_i) = 0 \\ -1 & f\,(x_j - x_i) < 0 \end{cases} \tag{9}$$

where n is length of the time series. Researchers have proved that statistic S is approximately normally distributed when n>8. The mean value of S is 0 and the variance can be expressed as:

$$var(S) = \frac{n(n-1)(2n+5)}{18} \tag{10}$$

Then statistic Z is obtained by, which follows the standard normal distribution:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{var(S)}} & if\ S > 0 \\ 0 & if\ S = 0 \\ \dfrac{S+1}{\sqrt{var(S)}} & if\ S < 0 \end{cases} \tag{11}$$

Given a significance level α, if $|Z| > Z_{(1-\alpha/2)}$, where $Z_{(1-\alpha/2)}$ is the $100(1-\alpha/2)^{th}$ percentile in the standard normal distribution, then the trend of time series is deemed as statistically significant and the null hypothesis is rejected. If $Z>0$, there is a possible monotonic increasing trend in the time series, and vice versa.

5.2 Change point detection of PCR based on *t*-statistic

The Mann-Kendall test can detect if there is an increasing or decreasing trend in the *PCR* time series. If a trend exists, then a change point detection method can be applied to locate the point where *PCR* starts to change. This study adopts a t-test based change point detection method which is proposed to analyze the operation data of residential air conditioning systems (Rogers et al., 2020). The original *t*-statistic is designed to describe the difference of two data groups in the mean value, as expressed in Equation 12:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \tag{12}$$

where $\bar{X}$ and $\bar{Y}$ are the mean value of these two data groups, s_x and s_y are the corresponding standard deviation.

Rogers et al. (2020) proposes a modified t-statistic considering X and Y may have drastically different size, as shown in Equation 13 and 14. In order to determine the location of the change point in a time series, the whole series is bisected between each of the two consecutive data points, creating two sub-groups of time series each time. Then, for all possible bisections of the time series, the *t*-statistics is calculated for the two sub-groups. The point of bisection that has the largest *t*-statistics is considered the change point in the time series.

$$t = \left| \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \right| \tag{13}$$

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2} \tag{14}$$

Chapter 3-6 introduce the technical details of the proposed benchmarking methodology. The major concern is how to identify comparable peer operation segments with similar dynamics of working environment and control input. Besides, trend and change detection algorithms of *PCR* are designed to identify any anomalous deviation of operation and locate the point when it starts to deviate.

6. Deployment of the benchmarking methodology and results

In order to test and validate the effectiveness of the proposed benchmarking methodology, the methodology is applied to benchmark the performance of ACs in two city buses fleets. The first fleet has 7 buses in Guangzhou, China, and the second fleet has 10 buses in Sanya, China. In the experiment, faults were manually added to one bus AC in each fleet. The bus ACs are equipped with advanced IoT sensors to monitor the operating conditions and system performance. The sensors continuously upload data to a central database. A cloud-based server is also developed to visualize the collected data and is ready for deploying advanced artificial intelligence algorithms. Due to the humid subtropical monsoon climate in Guangzhou and tropical climate in Sanya, the cooling demand of bus ACs is very large all year round.

6.1 Introduction of the studied bus ACs

All these 17 bus ACs are of the same model. The structure of these ACs is depicted in Fig. 1. The systems have a nominal cooling capacity of 30,000 kcal/h (34.9 kW) and use R407C refrigerant. The control strategies contain the following aspects:

- Compressor speed and indoor fan speed are controlled to maintain the cabin temperature around its setpoint.
- EEV opening is controlled to ensure the degree of superheat is around a certain setpoint (usually 10K).
- Outdoor fan speed is controlled based on condensing pressure.

All these control strategies are accomplished by proportional controllers. The compressor and fans are driven by inverters, which may work under automatic or manual mode chosen by the bus driver. In the automatic mode, the compressor and indoor fan speed are regulated according to the difference of cabin temperature and its setpoint. In the manual mode, the speed of indoor fan is fixed. They may work under the “high”, “medium”, or “low” speed set by the bus driver. Even under the same working environment, the choices of control strategy may still lead to difference in the overall energy consumption. Thus, the control inputs should be considered as critical impact factors in benchmarking.

6.2 Description of faults implemented in selected bus ACs

In order to validate the effectiveness of the proposed benchmarking methodology for anomaly detection of a large number of bus ACs, experiments were conducted by manually imposing two types of faults during operation of the selected bus ACs.

In the first fault experiment, a positive offset of 5 K was manually added to the evaporator outlet temperature sensor of one bus AC (named GZ-AC #1) in the bus fleet running in Guangzhou from the afternoon of 12th May to 31st May, while the other ACs operated under normal conditions. Since the degree of superheat is designed to be maintained at around 10K by the EEV, the fault caused a 5K increase of superheat received by the control system, as shown in Equation 15. As a result, the actual superheat was expected to decrease by 5K.

$$SH_{control,receive} = SH_{actual} + \Delta SH_{offset} = SH_{actual} + 5K \tag{15}$$

In the second fault experiment, a 30% decrease in the outdoor fan frequency was imposed to one of the bus ACs in Sanya (named SY-AC #1) to validate the proposed benchmarking methodology with condenser fan fault. The fault was simulated by decreasing the control signal of the outdoor fan frequency from the afternoon of 12th May to 31st May. After modification of the condenser fan control logic, the control system still outputs the normal condenser fan frequency according to the original normal control settings, but condenser fan drivers receive the faulty signals, which is 30% smaller.

6.3 Data preprocessing

The operation data of these 17 bus ACs from 16th April 2022 to 31st May 2022 is collected for analysis. The sampling interval of the sensors is 15 seconds. During the data preprocessing stage, "on-off" operation data segments are extracted from the raw measurement data. After extracting operation cycles, data with sensor or communication error is removed.

Then, data normalization is conducted to scale the original data between 0 and 1. This is an essential step in data preprocessing for the majority of machine learning methods (Singh and Singh, 2020). Since multiple bus ACs are considered in this research, the original data is normalized according to Equation 16:

$$X^* = \frac{X - X_{global,min}}{X_{global,max} - X_{global,min}} \tag{16}$$

where X refers to a certain variable and the subscript "global" refers to the global minimum or maximum value of that variable in the operational data of all bus ACs.

After data preprocessing, original recordings are converted into multivariate time series samples with different length representing “on-off” operation cycles. Fig.6 depicts the number of “on-off” operation cycles of 7 the bus ACs in Guangzhou from 16th April 2022 to 31st May 2022. Due to the intermittent solar irradiance, random numbers of passengers, and frequent opening and closing of doors, bus ACs generally work under dynamic environment and frequent “on-off” cycles are very common. Compared with the operation in May, the operation in April is less frequent. The number of data segments of different bus ACs also varies since some buses work on densely populated routes during peak hours.

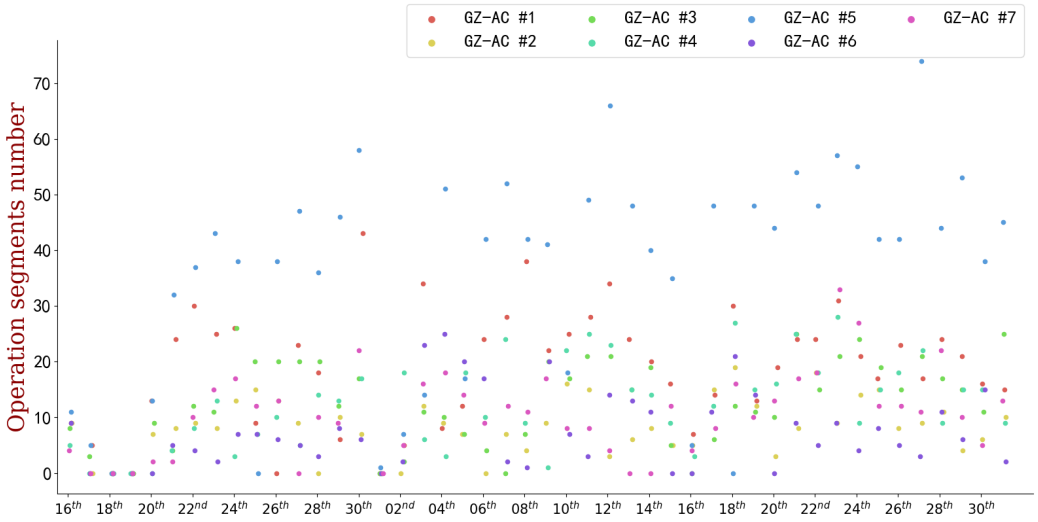

Fig. 6 Operation segments count of 7 bus ACs in Guangzhou from 16th April to 31st May

6.4 Results of multivariate time series distance calculation

As mentioned in Section 3.2, LSTM-autoencoder is established in the proposed benchmark methodology to transform multivariate time series with different length into context vectors with the same size, distance measurement is conducted among context vectors so that data segments with similar dynamics of working environment and control input can be recognized as comparable peer samples. The developed LSTM-autoencoder is trained

with the Adam optimizer. The learning rate is set at 6×10^{-6}, and the model is trained with 100 epochs. An epoch means training the neural network with all the training data for one cycle. After 100 epochs, the training loss reaches an acceptable level, which means the developed LSTM-autoencoder has obtained the knowledge of the dynamic interrelationship among different variables. The context vectors extracted from the LSTM-autoencoder can be used as good representations of the dynamic operation condition time series.

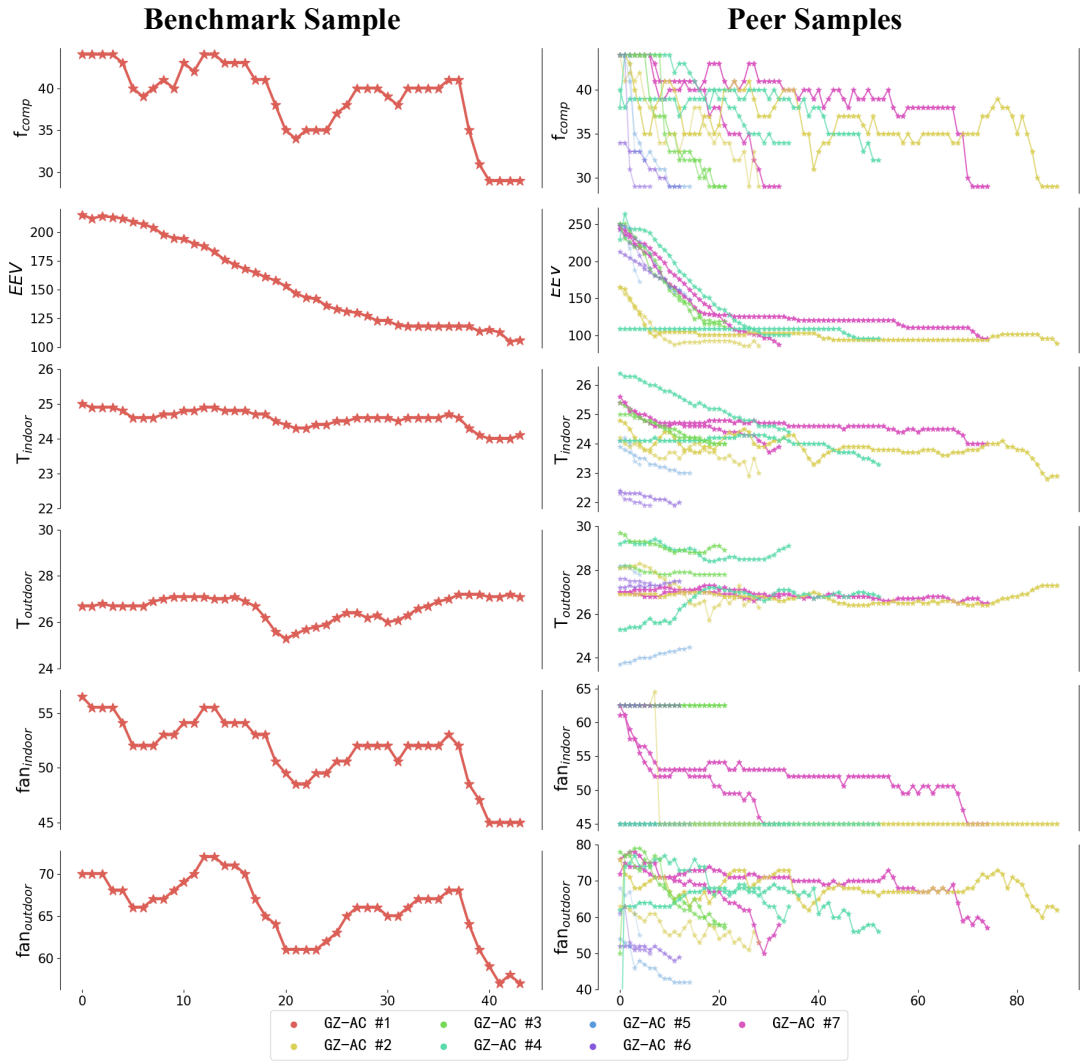

Fig.7 An example results of peer sample identification (benchmark sample belongs to the operation of GZ-AC #1 on May 5th)

Fig.7 presents example results of comparable peer sample identification. The benchmark sample is an "on-off" operation cycle of GZ-AC #1 on May 5th. The peer samples are the data segments of the other 6 bus ACs in the past 14 days. Each color in the figure represents a bus AC. As shown in the example, although these multivariate time series are of different length, the dynamics of each identified peer sample share high similarity with the benchmark sample. The developed LSTM-autoencoder is capable of identifying peer samples with similar dynamics for benchmarking.

6.5 Benchmarking and validation results

6.5.1 Test results of bus fleet ACs running in Guangzhou

The identified comparable peer samples for each benchmark sample are used to train reference models as the benchmark of system operation performance. Then, *PCR* is calculated with the daily operation data for each bus AC. Fig.8 depicts the *PCR* value of all bus ACs in Guangzhou operating from 1st May to 31st May. Apart from GZ-AC #1, the *PCR* value of the other bus ACs fluctuates around 1.0, which indicates their energy performance is more or less the same as their peers.

After introducing the evaporator outlet sensor offset fault on 12th May, the *PCR* of GZ-AC #1 suddenly increases from 1.02 to 1.06 and then fluctuates around this new level. The maximum *PCR* of GZ-AC #1 can reach as high as 1.10, which means 10% more power consumption is caused by this fault.

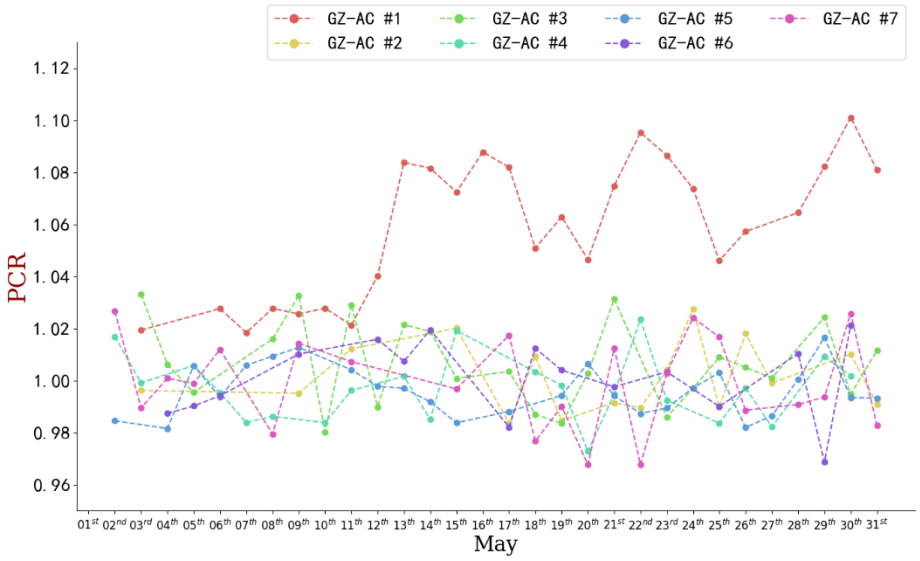

Fig. 8 *PCR* of all bus ACs in Guangzhou from 1st May to 31st May

The violin plots in Fig.9 show the trend of actual degree of superheat for GZ-AC #1 before and after the implementation of evaporator outlet temperature sensor fault. The black dots represent the median of all superheat values in each day. As shown in Fig.9, although the EEV control logic is designed to maintain the superheat of evaporator outlet at 10K before 12th May, the actual superheat varies largely due to the highly dynamic operation. As explained in the introduction, the highly dynamic operation conditions of bus AC results in frequent "on-off" operation cycles. Sometimes, these cycles only last for a few minutes and the EEV controllers installed in the bus ACs are incapable of regulating the refrigeration system to the designed superheat degree in such short time.

After 12th May, the superheat degree has a clear drop due to the implementation of evaporator outlet temperature sensor offset fault. A lower superheat indicates higher probability of compressor slugging (Laughman et al., 2006). Sometimes, the actual superheat even drops below 0K. Superheat oscillation (Xia et al., 2019) and possible

compressor slugging will increase the power consumption of refrigeration cycle. Thus, an increase of *PCR* is observed in GZ-AC #1 after 12th May.

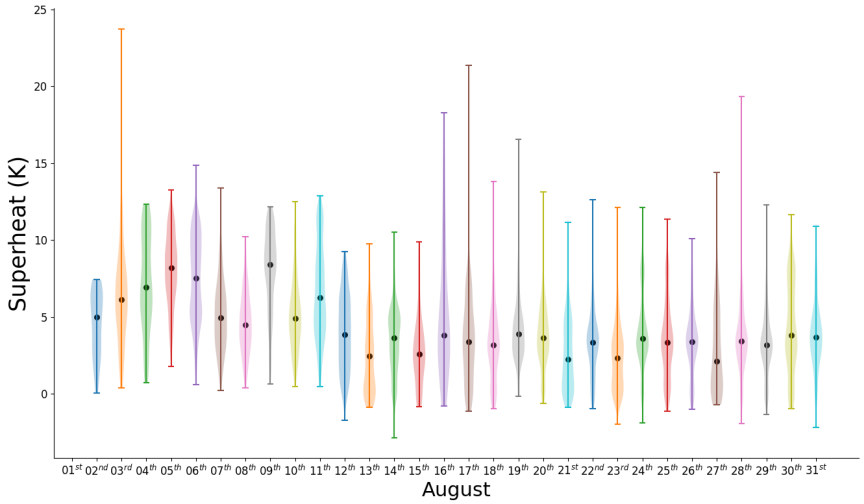

Fig. 9 Actual degree of superheat of GZ-AC #1 from 1st May to 31st May

The Mann-Kendall trend test is applied to determine whether any trend of *PCR* exists. As shown in Fig.10, a significant increasing trend is identified for the *PCR* of GZ-AC #1 with over 99% confidence level. The *PCR* of other bus ACs does not show any trend.

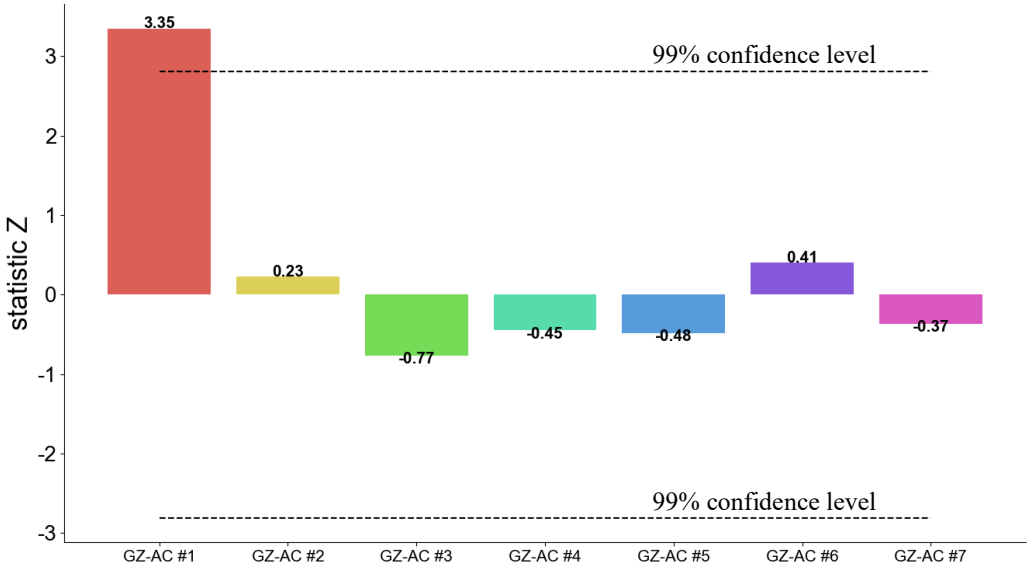

Fig. 10 Mann-Kendall trend test result for *PCR* of different bus ACs in Guangzhou

The *t*-statistic based change point detection is then used to determine the point where *PCR* of GZ-AC #1 starts to change. As shown in Fig.11, the *t*-statistic of *PCR* values for GZ-AC #2-#7 maintain at a very low level from 1st May to 31st May, which is another proof that little changes occur in their operation performance during the period. On the other hand, the *t*-statistic of *PCR* values of GZ-AC #1 has a quick rise around 12th May, which indicates an abrupt change happens to the performance of GZ-AC #1. The change point detection result is consistent with the fault experiments conducted on GZ-AC #1.

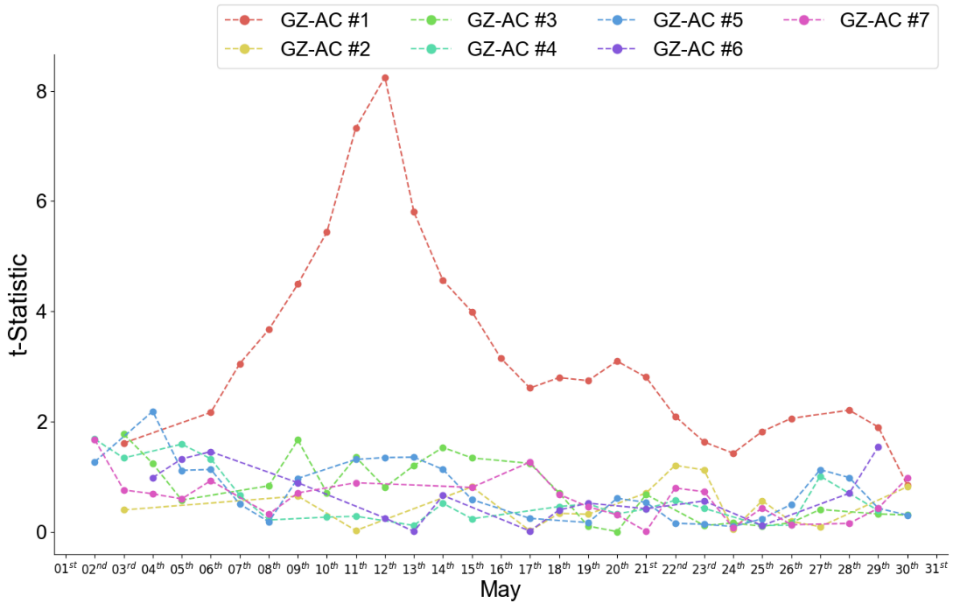

Fig. 11 *PCR* change point detection result of different bus ACs in Guangzhou

6.5.2 Test results of bus fleet ACs running in Sanya

The benchmarking results of bus ACs in Sanya are depicted in Fig.12. The energy performance of most bus ACs doesn't change very much during the test as their *PCR* values maintain at a fixed level. However, *PCR* of SY-AC #1 drops from 1.0 to 0.95 after 12th May and maintains at this new level. The change of *PCR* indicates some change happened in the operation of SY-AC #1 in Sanya.

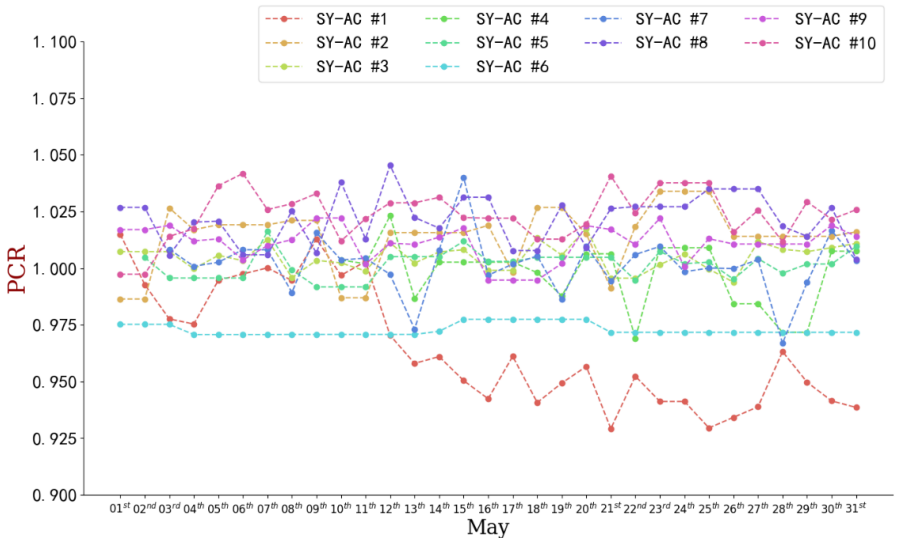

Fig. 12 *PCR* of bus ACs in Sanya from 1st May to 31st May

After implementation of condenser fan fault, compared with other bus ACs which receive the same condenser fan control signals, the power consumption of SY-AC #1 is lower due to a lower actual condenser fan speed.

Mann-Kendall trend test also detects the decreasing trend in the *PCR* of SY-AC #1, as shown in Fig. 13. Compared with SY-AC #1, *PCR* values of other ACs in Sanya do not show any significant trend. The change point detection algorithm is then adopted to identify the point when *PCR* of SY-AC #1 starts to change. As shown in Fig. 14, the *t*-statistic of *PCR* for SY-AC #1 reaches the peak during 11th May and 12th May. Therefore, possibly a performance change occurred during these days, which is consistent with the time when the fault was introduced to SY-AC #1.

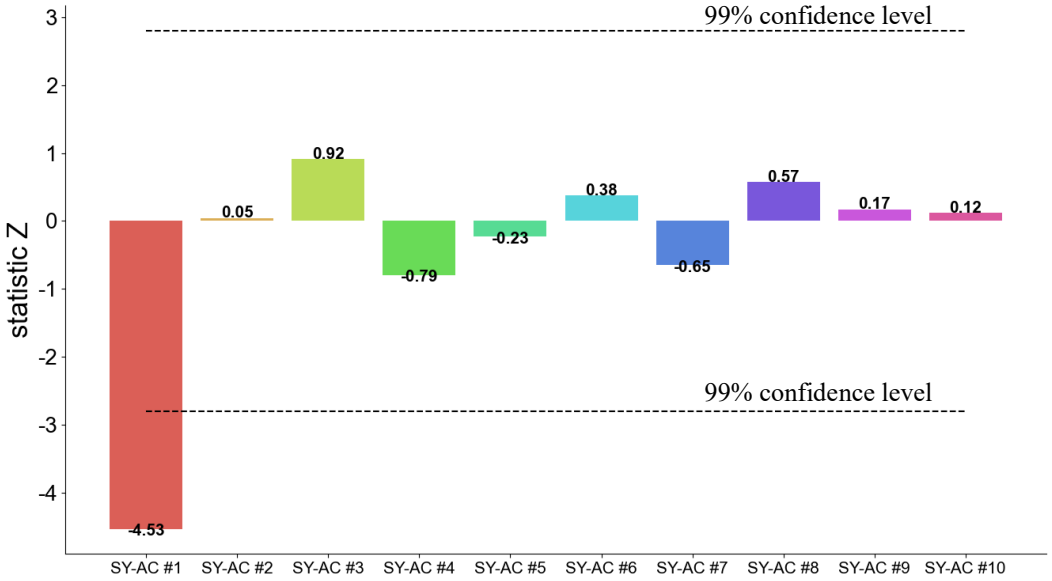

Fig.13 Mann-Kendall trend test result for *PCR* of different bus ACs in Sanya

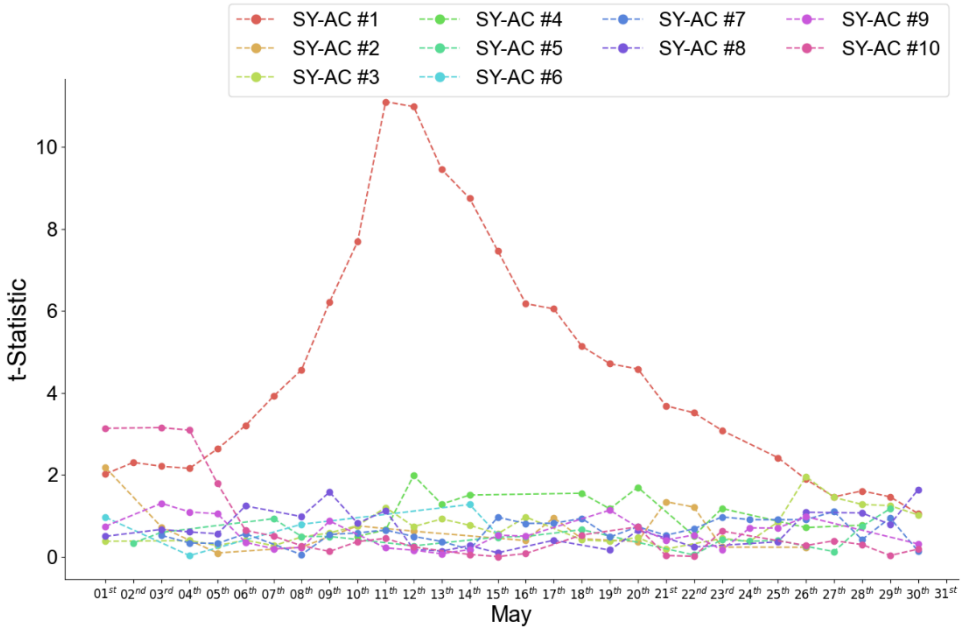

Fig.14 *PCR* change point detection result of different bus ACs in Sanya

Through manual implementation of sensor offset fault on a bus AC in Guangzhou and condenser fan fault on another bus AC in Sanya, faulty operation data is obtained to validate the effectiveness of proposed benchmarking methodology in detecting the change of energy performance for bus ACs. The validation results agree with the expected outcomes and the benchmarking methodology shows the capability of detecting anomalies.

7. Conclusions

Health monitoring and performance evaluation of massive public transport ACs are vital to the development of future smart cities. This paper presents a newly developed benchmarking methodology to evaluate the operation performance of a large number of bus ACs considering their unique operation dynamics.

In order to identify the anomalous operation among a large number of bus ACs, this study aims at benchmarking their energy performance. A LSTM-autoencoder based distance measurement method is proposed to identify AC operation cycles that share similar dynamics of working conditions as comparable peer samples for benchmarking. After selection of comparable peer samples, data-driven reference models specialized in time series prediction are trained. Finally, the ratio of the measured energy consumption to the energy consumption predicted by the reference model is defined as *PCR*, and *PCR* is used to evaluate the energy performance of each bus AC. Additionally, trend and change detection algorithms are applied to analyze the benchmarking results and detect anomalies in time.

The developed benchmarking methodology has been deployed on 7 bus ACs operating in Guangzhou and 10 bus ACs in Sanya, China for validation. Experiments are conducted to manually implement faults in ACs, including evaporator outlet temperature offset fault and condenser fan fault. The proposed benchmarking methodology exhibits the capability of detecting anomalous power consumption pattern caused by these faults among a large number of bus ACs.

In future research, more field tests on various faults will be conducted on bus ACs to validate the proposed benchmarking methodology. Different fault types and multiple fault scenarios will be considered in these experiments to test the robustness and effectiveness of the proposed benchmarking methodology.

Acknowledgement

The authors gratefully acknowledge the support of this research by the Research Grants Council of the Hong Kong SAR (C5018-20GF).

Reference

Brown, R., Webber, C., Koomey, J.G., 2002. Status and future directions of the Energy Star program. Energy 27, 505-520.

Chen, Z., Xiao, F., Shi, J., Li, A., 2022. Dynamic model development for vehicle air conditioners based on physics-guided deep learning. International Journal of Refrigeration 134, 126-138.

Cheng, L., Dhillon, P., Horton, W.T., Braun, J.E., 2021. Automated laboratory load-based testing and performance rating of residential cooling equipment. International Journal of Refrigeration 123, 124-137.

Chung, W., 2012. Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings. Applied Energy 95, 45-49.

da Silva, R.M., Santos, C.A.G., Moreira, M., Corte-Real, J., Silva, V.C.L., Medeiros, I.C., 2015. Rainfall and river flow trends using Mann–Kendall and Sen's slope estimator statistical tests in the Cobres River basin. Natural Hazards 77, 1205-1221.

Dhillon, P., Horton, W.T., Braun, J.E., 2021. Comparison of Steady-State and Dynamic Load-Based Performance Evaluation Methodologies for a Residential Air Conditioner.

Ding, Y., Liu, X., 2020. A comparative analysis of data-driven methods in building energy benchmarking. Energy and Buildings 209, 109711.

Fumo, N., Mago, P., Luck, R., 2010. Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. Energy and Buildings 42, 2331-2337.

Gocic, M., Trajkovic, S., 2013. Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. Global and Planetary Change 100, 172-182.

Habib, U., Hayat, K., Zucker, G., 2016. Complex building's energy system operation patterns analysis using bag of words representation with hierarchical clustering. Complex Adaptive Systems Modeling 4, 8.

IEA, 2019. Cooling on the Move. IEA, Paris.
Iglesias, F., Kastner, W., 2013. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. Energies 6.
Jain, N.R., Rawal, R., Vardhan, V., Dey, S., 2021. Performance metrics for room air-conditioners: energy, comfort and environmental impacts. Buildings and Cities 2.
Jeong, J., Hong, T., Ji, C., Kim, J., Lee, M., Jeong, K., 2016. Development of an integrated energy benchmark for a multi-family housing complex using district heating. Applied Energy 179, 1048-1061.
Jin, X., Xiao, F., Zhang, C., Li, A., 2022. GEIN: An interpretable benchmarking framework towards all building types based on machine learning. Energy and Buildings 260, 111909.
Laughman, C.R., Armstrong, P.R., Norford, L.K., Leeb, S.B., 2006. The detection of liquid slugging phenomena in reciprocating compressors via power measurements.
Li, A., Fan, C., Xiao, F., Chen, Z., 2022. Distance measures in building informatics: An in-depth assessment through typical tasks in building energy management. Energy and Buildings 258, 111817.
Li, Z., Han, Y., Xu, P., 2014. Methods for benchmarking building energy consumption against its past or intended performance: An overview. Applied Energy 124, 325-334.
Rogers, A.P., Fangzhou, G., Rasmussen, B.P., 2020. A Change Point Detection Algorithm with Application to Smart Thermostat Data. ASHRAE Transactions 126, 567-579.
Roth, J., Rajagopal, R., 2018. Benchmarking building energy efficiency using quantile regression. Energy 152, 866-876.
Schaad, K., Hofer, S., 2020. Benchmarking, in: Redlein, A. (Ed.), Modern Facility and Workplace Management: Processes, Implementation and Digitalisation. Springer International Publishing, Cham, pp. 115-138.
Schmidt-Hieber, J., 2020. Nonparametric regression using deep neural networks with ReLU activation function. The Annals of Statistics 48, 1875-1897, 1823.
Singh, D., Singh, B., 2020. Investigating the impact of data normalization on classification performance. Applied Soft Computing 97, 105524.
Wang, J., Besselink, I., Nijmeijer, H., 2015. Electric Vehicle Energy Consumption Modelling and Prediction Based on Road Information. World Electric Vehicle Journal 7, 447-458.
Wang, W., Zhou, Z., Lu, Z., 2022. Data-driven assessment of room air conditioner efficiency for saving energy. Journal of Cleaner Production 338, 130615.
Xia, Y., Ding, Q., Jiang, Z., Deng, S., Song, M., 2019. Development of a superheat controller for mitigating hunting in a direct expansion air conditioning system. Energy Procedia 158, 2085-2091.
Yuanxia, W., Jianghong, W., Fang, X., Chaopeng, L., Hongqi, L., Jianhong, C., 2011. Survey of residential air-conditioning-unit usage behavior under south China climatic conditions, 2011 International Conference on Electric Information and Control Engineering, pp. 2711-2714.
Zhan, S., Liu, Z., Chong, A., Yan, D., 2020. Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking. Applied Energy 269, 114920.
Zhang, Y., Yang, W., Huang, Z., Liu, D., Simpson, I., Blake, D.R., George, C., Wang, X., 2017. Leakage Rates of Refrigerants CFC-12, HCFC-22, and HFC-134a from Operating Mobile Air

Conditioning Systems in Guangzhou, China: Tests inside a Busy Urban Tunnel under Hot and Humid Weather Conditions. Environmental Science & Technology Letters 4, 481-486.
Zhao, Y., Liu, P., Wang, Z., Zhang, L., Hong, J., 2017. Fault and defect diagnosis of battery for electric vehicles based on big data analysis methods. Applied Energy 207, 354-362.
Zhou, Y., Lork, C., Li, W.-T., Yuen, C., Keow, Y.M., 2019. Benchmarking air-conditioning energy performance of residential rooms based on regression and clustering techniques. Applied Energy 253, 113548.