https://dx.doi.org/10.1016/j.applthermaleng.2022.119931

Real-Time Energy Performance Benchmarking of Electric Vehicle Air Conditioning Systems Using Adaptive Neural Network and Gaussian Process Regression

Fangzhou Guo[1], Zhijie Chen[1], Fu Xiao[1, 2*], Ao Li[1] and Jian Shi[3]

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

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

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

Abstract

One major concern of the electric vehicle is the limited driving range per charge. Among its auxiliary systems, the heating, ventilation, and air conditioning (HVAC) system consumes the largest amount of electricity and can have a significant impact on the driving range when operational faults occur. This paper proposes a real-time benchmarking method to continuously evaluate the energy performance of a large number of electric vehicle air conditioning systems. Each system is benchmarked based on the energy consumption of its peer systems. Considering the energy consumption is influenced by several impact factors including the ambient environment and the cooling capacity, this study encodes the impact factors with an Autoencoder to measure the similarities between operating conditions. Also, considering the difference in operating conditions between a system and its peers, the uncertainties in energy performance benchmarks of the peer systems are quantified by the Gaussian process given its probabilistic nature. For each peer system, a Gaussian process regression model is developed as a benchmark, and the performance of the target system is assessed by comparing its measured energy consumption with the averaged benchmarks of all its comparable peers, accounting for the uncertainties. With the continual learning algorithm adopted, the Autoencoder can be updated periodically to adapt to real-time operational data with minimal computational cost. The real-time benchmarking method is applied to electric bus air conditioners in Haikou and Sanya, China, and can effectively identify malfunctioning systems. This method can be conveniently deployed on cloud for smart health management of public electric vehicles in smart cities.

Keywords

Energy performance benchmarking; air conditioning system; electric vehicle; data-driven model; Gaussian process regression; adaptive neural network

Highlights

- Energy performance of plenty of electric vehicle air conditioners is benchmarked.
- Every system is benchmarked based on the average performance of its peer systems.
- The Autoencoder enables measurement of similarity between operating conditions.
- Gaussian process regression is adopted to quantify the benchmarking uncertainty.
- The method is applied for real-time health management of public electric vehicles.

1 Introduction

Transportation relies heavily on fossil fuels, accounting for 37% of global carbon dioxide emissions from end-use sectors [1]. Among the sources of fossil fuels, gasoline and distillate fuels (mostly diesel fuel) are two petroleum products commonly used by on-road vehicles such as cars, motorcycles, trucks, and buses. Gasoline is the dominant transportation fuel in the U.S. accounting for 54% of U.S. transportation energy, followed by distillate fuels which account for 23% [2]. With the rapidly increasing transportation demand, energy consumption and emissions from transportation continue to rise [3].

With increasing concerns about global warming, governments and organizations are taking measures to reduce greenhouse gas emissions and accommodate the rising energy demand with other energy sources. In the transportation sector, the electric vehicle (EV) is an alternative to internal combustion engine vehicles since it generates fewer emissions. Electric vehicle deployment has witnessed fast growth over the past ten years, with 16.5 million on the roads worldwide by the end of 2021 [4].

Compared with internal combustion engine vehicles, one main disadvantage of the electric vehicle is the driving range. The major energy consumption of an electric vehicle comes from the engine. As to the auxiliary systems, the heating, ventilation, and air conditioning (HVAC) system accounts for the largest share of energy consumption and has dramatic impacts on the driving range [5,6]. Fiori et al. [5] showed that the driving range could be reduced by 24% when a heating system is operated, and by 10% when a cooling system is operated. Also, Wang et al. [7] and Liu et al. [8] modeled the effect of ambient temperature on electricity consumption, and concluded that an average of 9.66% of energy could be saved by eradicating unreasonable auxiliary loads. Therefore, researchers have been focusing on increasing the driving range of electric vehicles by improving the performance of the HVAC systems.

The energy performance of a vehicle HVAC system can be improved by (1) installing additional parts such as heat recovery systems [9], (2) using alternative refrigerants such as carbon dioxide [10–12], (3) adopting advanced control strategies [13,14], and (4) performing

predictive maintenance through benchmarking, fault detection and diagnosis (FDD) [15–18]. The first three approaches concentrate on improving the system operational efficiency in the design phase, while predictive maintenance aims at preventing possible equipment failures in advance during system operation. Performance benchmarking and FDD can both prompt and automate the predictive maintenance process. Generally, benchmarking methods can identify if a system deviates from its normal behavior by referring to the benchmarks, and FDD methods can be applied to the anomalous systems for further verification and diagnosis. Finally, service technicians can be dispatched to examine and repair the identified faulty systems.

Energy performance benchmarking of an individual system in general has two steps. Firstly, a benchmark of performance is computed considering key impact factors such as the ambient conditions. The benchmark of a system can be its historical performance, its ideal performance, and the performance of its peer systems. Secondly, the energy performance of the system is compared with the benchmark by an Energy Performance Index (EPI). If the real performance is significantly different from the benchmark, the system may malfunction and requires further diagnosis. Zhou et al. [19] benchmarked the energy performance of residential air conditioners based on their predicted power consumption under uniform weather conditions and air conditioner settings. Three key elements in energy performance benchmarking, namely the type of benchmark, EPI, and impact factors, were identified in their study. The type of benchmark determines the reference performance that a system should compare with. The EPI is a measurement of the energy performance of a system relative to its benchmark. The impact factors are physical quantities that do not indicate the energy performance of a system by themselves, but could strongly affect the EPI and therefore must be considered in the analysis. However, this benchmarking method is not applicable to vehicle HVAC systems as it is unrealistic to assume that they operate under uniform conditions. Research on energy performance benchmarking of vehicle air conditioners has not been found. This paper fills the gap by proposing an energy performance benchmarking method for electric vehicle air conditioning systems. Furthermore, different from traditional benchmarking methods which run a benchmark using static historical data and at long

intervals, the benchmarking method developed in this study can realize continual benchmarking using real-time system operation data.

Under the background of public transportation electrification in more and more cities, energy performance benchmarking of a large number of electric bus air conditions is urgently needed. Meanwhile, electric buses gradually adopt the state-of-the-art Internet of Things (IoT) technology in smart cities for city-level management [20]. The operating conditions of vehicle engines and auxiliary equipment can be recorded by IoT sensors and uploaded to a cloud server instantaneously, such that performance monitoring and benchmarking can be conducted on a remote platform. In view of this, this study develops a novel energy performance benchmarking method for a large number of electric buses in cities by adopting the peer-performance benchmarking approach. The benchmark energy performance of a vehicle air conditioner is established by aggregating the performance of its peer systems, which have the same configuration and operate in the same climate region. As electric buses are usually mass-produced and the types of electric buses running in a city are highly likely to be the same, it is feasible to consider the same type of electric bus air conditioners as peer systems.

This study specifically addresses three major challenges in peer-performance benchmarking of a large number of bus air conditioning systems in real-time. First, vehicle air conditioners operate under different conditions, so the similarity of operating conditions should be measured to determine if the energy performance of two systems is comparable. However, the operational data of each system is multivariate time series and similarity measurement is difficult. This study applies the Autoencoder to encode the multivariate time series into vectors, enabling similarity measurement through the calculation of Euclidean distances between vectors.

Second, to perform real-time energy assessment continually, the developed Autoencoder should be updated periodically to adapt to the changing operating conditions of systems. However, retraining the Autoencoder from scratch is time-consuming and computationally expensive. This paper incorporates the idea of lifelong learning and proposes to apply continual learning algorithms to reduce the computational cost of model updating. Continual

learning is a concept to train a model on multiple datasets sequentially, and during the training process of a new dataset, the loss of knowledge and patterns learned from preceding datasets is minimized. This concept has been recently employed in many different applications. For example, Zhou et al. [21] used a continual learning method termed elastic weight consolidation (EWC) for dynamic building energy load prediction and demonstrate its high accuracy, better stability, and low computational cost. Lee et al. [22] used the gradient episodic memory (GEM) method to improve and accelerate the performance of a multi-client power consumption model in smart grid systems.

Third, in peer-performance benchmarking, the energy performance of a target system is benchmarked against many of its peer systems. Because the target system and its peer systems do not operate under the exact same conditions, it is unreliable to compare their energy performance directly. To address this issue, this study proposes to establish regression models for each peer system which correlate the energy consumption with impact factors. Peer performance of the target system is the energy consumption predicted by the regression models. This paper applies the Gaussian process to establish the regression models to improve the reliability and robustness of the benchmarking method against diversities and uncertainties of operation conditions. Gaussian process regression assumes the posterior distribution of the model output follows the Gaussian distribution, and thus uncertainty of the predicted energy consumption could be quantified by the variance of the Gaussian distribution. When the operating conditions between the target and one of its peer systems are very different, the variance of the predictions will be very large due to the dissimilarity of impact factors. The predictions with large variance are considered inaccurate and are removed in benchmarking.

The remainder of this paper is organized as follows. Chapter 2 details the proposed real-time energy performance benchmarking method. Chapter 3 applies the method to real operational data of vehicle air conditioners. The performance of a continual learning algorithm applied to the Autoencoder is evaluated, and the effectiveness of the whole benchmarking method is assessed by verified faulty systems.

2 Proposed Real-Time Energy Performance Benchmarking Method

This chapter details the solutions developed to the challenges in real-time energy performance benchmarking. Section 2.1 provides an overview of the benchmarking method. Section 2.2 describes the configuration of the vehicle air conditioning systems and discusses the critical impact factors of system energy consumption. Section 2.3 introduces the Autoencoder, a type of artificial neutral network, and the continual learning algorithm which adapts the Autoencoder to real-time system operation data for continual benchmarking. Section 2.4 details the Gaussian process regression and explains the importance of kernels in the benchmarking process. Lastly, Section 2.5 presents the calculation of the Energy Performance Index (EPI) along with the following statistical analysis to identify abnormally high EPI values and changes of EPI.

2.1 Overview of the benchmarking method

Figure 1 shows the framework to evaluate the energy performance of the target system, i.e., *System A*, through peer-performance benchmarking. The benchmark of *System A* is the average performance of its peer systems, e.g., *System B, C*... The energy performance of *System A* is measured by its energy consumption, and EPI is defined as the difference of the measured energy consumption to the energy consumption predicted by its benchmarks. The energy consumption is affected by a few impact factors such as the ambient conditions and the actual cooling capacity, which are a combination of scalars and time series. Thus, an Autoencoder model is built to encode the crucial impact factors into a fixed-length vector. Then, regression models are established for each peer system to correlate energy consumption with the encoded impact factors. After that, the regression models produce predictions of the energy consumption at the impact factor values of *System A*, and the predictions are used to compute the EPI for *System A*.

However, since the operating cycles of a peer system cannot cover all operating conditions of *System A*, the trained regression models may produce poor predictions for some inputs due to extrapolation. Thus, this study adopts the Gaussian process to establish the regression model.

One advantage of the Gaussian process regression is that it can obtain the variance of each prediction to quantify the confidence level. The variance could be computed by kernels using the Euclidean distance between inputs of the training and test data. Note that in the framework shown in Figure 1, the training inputs of the Gaussian regress models (i.e. $y = f(x)$) are the encoded impact factors (i.e. x_i where i denotes the operating cycle) of a peer system, and the test inputs are the encoded impact factors of *System A*. If any test inputs are very different from the training inputs, then predictions of those test inputs will have high variance and are considered invalid (marked as red y_i^j in Figure 1, where j denotes the peer system). The invalid predictions are removed by a variance filter.

Finally, the EPI of *System A* is calculated, and statistical analysis is conducted to detect the outliers of EPI for evaluating the performance of *System A*. With the same framework, the performance of every system in the group of peer systems can be evaluated.

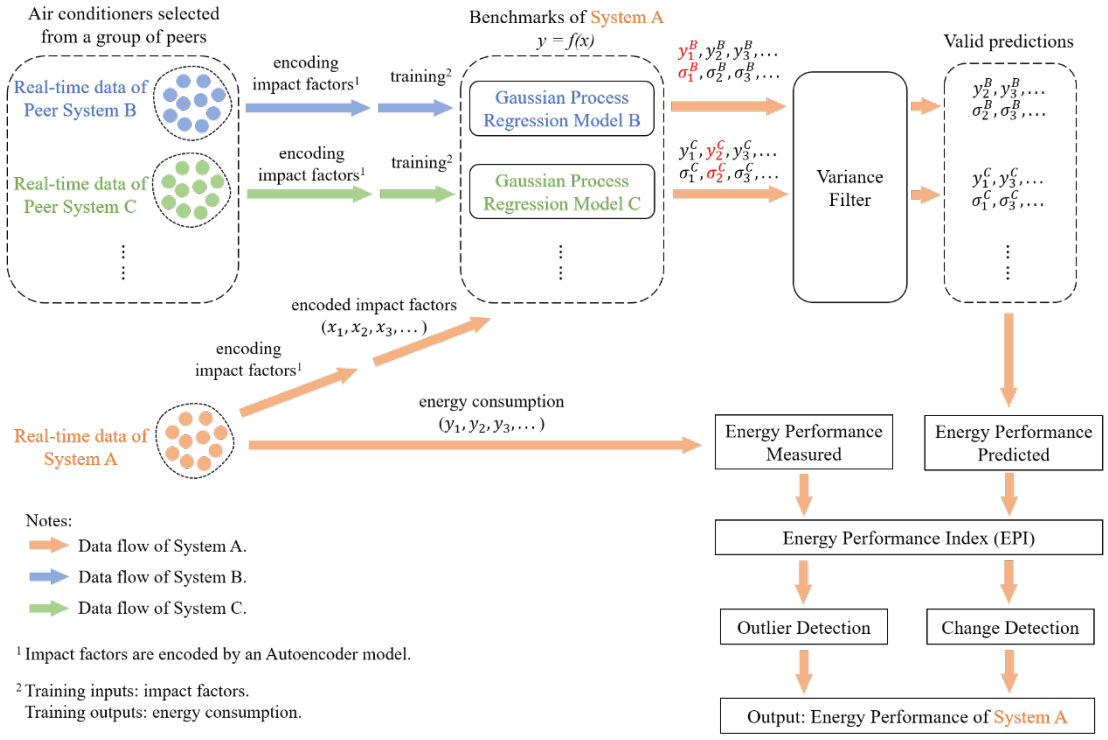


Figure 1: Framework of the proposed peer-performance benchmarking method for electric vehicle air conditioning systems.

Additionally, the flowchart in Figure 2 depicts the process to execute the real-time energy performance benchmarking. Each day, raw operating data from a group of vehicle air

conditioners is extracted from the cloud database. The raw data is cleaned and the operating cycles are extracted through data preprocessing. To efficiently compute the similarity of impact factors between operating cycles, an Autoencoder model is developed to encode the impact factors of each cycle to a vector. Then, Gaussian process regression models are established for each individual system to correlate the encoded impact factors with the energy consumption. After that, for each target system to be benchmarked, the expected energy consumptions of its operating cycles are predicted by regression models from its peer systems at the impact factor values of the target system. During this process, the similarity of the encoded impact factors between the target system and its peer systems are evaluated by kernels in the Gaussian process regression models, and predictions with large variance (computed by kernels) are eliminated. Based on the predicted energy consumption along with the measured energy consumption of the target system, the EPI of the target system can be calculated. Finally, detection thresholds of EPI are constructed by kernel density estimation using the previously calculated EPI values from all target systems, and a system is considered anomalous if its EPI is above the thresholds.

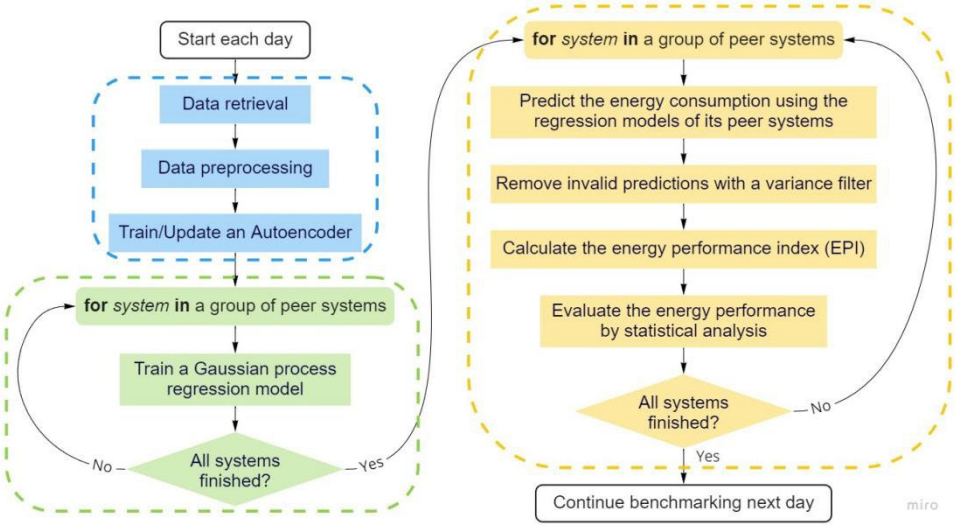


Figure 2: Flowchart of the real-time energy performance benchmarking method.

2.2 Description of the Vehicle Air Conditioning System and the Operational Data

The dataset used in this study is collected from a lot of electric buses in Haikou and Sanya, China from 2021 to 2022. The air conditioning systems of these electric buses have the same configuration, as shown in Figure 3, so the systems in the same city can be considered peer systems to each other. The refrigeration cycle of these systems consists of two modes: the indoor cooling mode and the battery cooling mode. The indoor cooling mode is usually switched on during the daytime when the buses carry passengers, and the battery cooling mode is usually switched on during the nighttime when the bus batteries are recharged in the bus station. Occasionally both modes are switched on, but this study does not consider this situation. In Figure 3, when the electronic expansion valve connecting to the evaporator (EEV1) is open, the system operates in the indoor cooling mode hunting for the indoor temperature setpoint; when the electronic expansion valve connecting to the heat exchanger (EEV2) is open, the system operates in the battery cooling mode and the cooling water is circulated to cool the vehicle battery. A variable-speed compressor is used in the system. The compressor frequency is controlled based on the indoor temperature setpoint error in the indoor cooling mode and based on the supply water temperature setpoint error in the battery cooling mode. The fans are also variable-speed. The indoor fan is controlled based on the indoor temperature setpoint error, and the outdoor fan is controlled based on the condenser pressure.

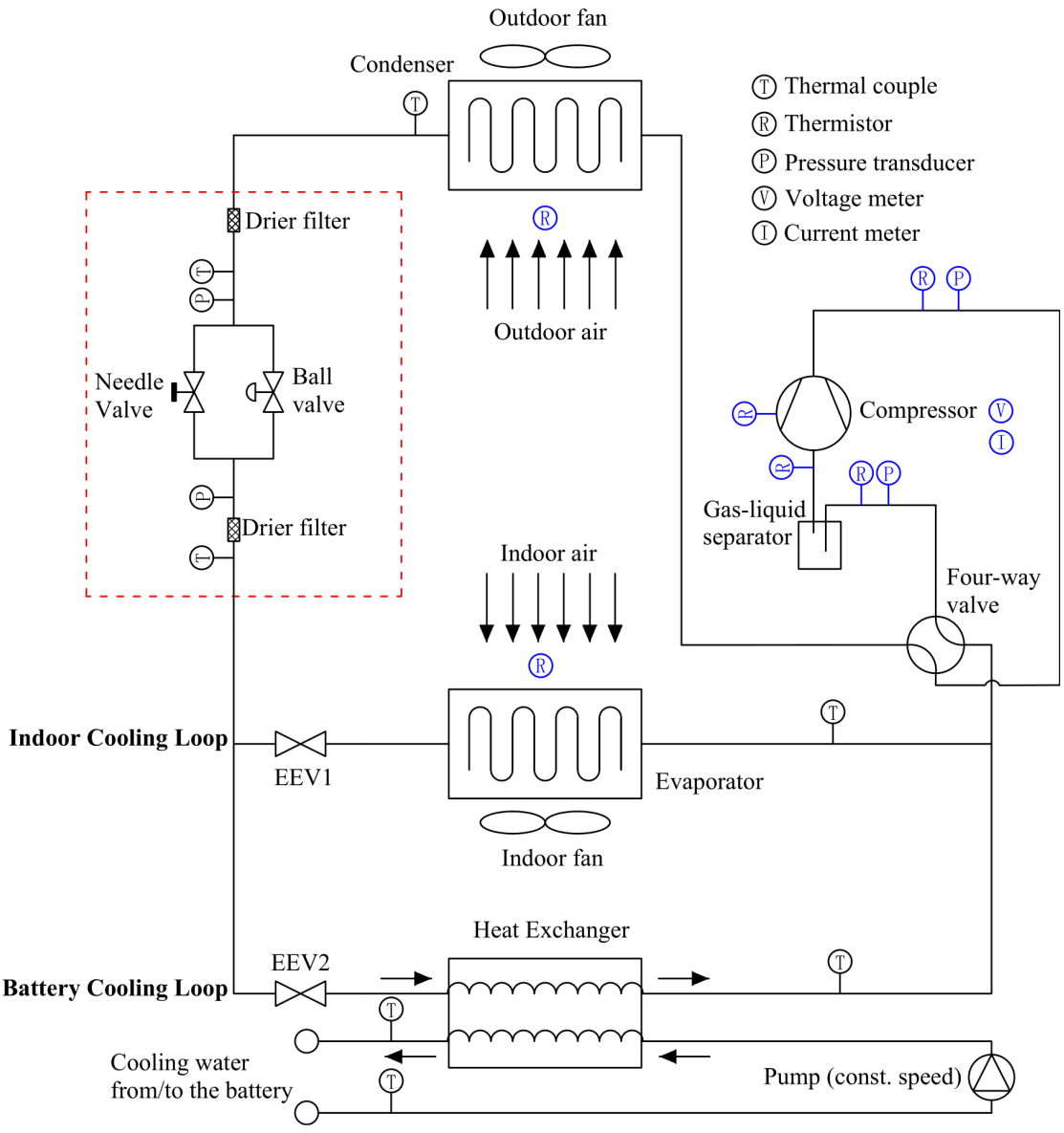


Figure 3: Configuration of the vehicle air conditioner.

Previous research has focused on the indoor cooling mode and benchmarked the efficiency of the vehicle air conditioners based on the relationships between the overall power consumption, the control signals, and the indoor and outdoor temperature [23]. However, it can still be improved in two aspects. Firstly, the developed neural networks require to be fully retrained every day, which is computationally expensive and inefficient. Secondly, because installing a flow meter on the vapor compression cycle is uneconomic for the mass-produced vehicle air conditioners, the previous research cannot calculate the cooling capacity of the operating systems from the sensor data. Without the cooling capacity, the coefficient of performance (COP) of a system is unavailable. Therefore, the previous research is mainly used to detect

inefficiencies in the compressor and fan motors rather than evaluating the performance of the whole refrigeration cycle.

This paper focuses on solving the above two issues. For the first issue, a continual learning method is adopted to update the Autoencoder (a neural network model). Besides, Gaussian process regression models, a machine learning method which requires much shorter time for training than neural networks, are used to create benchmarks of system power consumption. For the second issue, data in the battery cooling mode is also used in the analysis. In the battery cooling mode, the cooling water return and supply temperature are both available, and the cooling water is driven by a constant-speed water pump. Therefore, system cooling capacity can be obtained and taken into consideration in benchmarking. In this way, the overall performance of the refrigeration cycle is measurable. Also, compared to indoor cooling cycles, performance in battery cooling cycles is not affected by disturbances caused by the door opening and closing, passengers entering or leaving the bus, etc., and hence the operational data has less noise. This study focuses on the battery cooling mode as well as the indoor cooling mode, extracting additional impact factors from the heat exchanger and comparing the energy consumption between systems.

Figure 4 shows the durations of all operating cycles in the battery cooling mode and indoor cooling mode for 40 air conditioning systems from April 2021 to January 2022. Note that cycles shorter than 3 minutes are not analyzed in this study. The operating cycles in the battery cooling mode is fewer than the indoor cooling mode and generally has shorter cycle duration. Additionally, Table 1 shows the statistics of average power consumption in each cycle for these systems. In general, operating cycles in the battery cooling mode have powers between 3.4kW and 3.8kW, and cycles in the indoor cooling mode have powers between 3.6kW and 4.3kW.

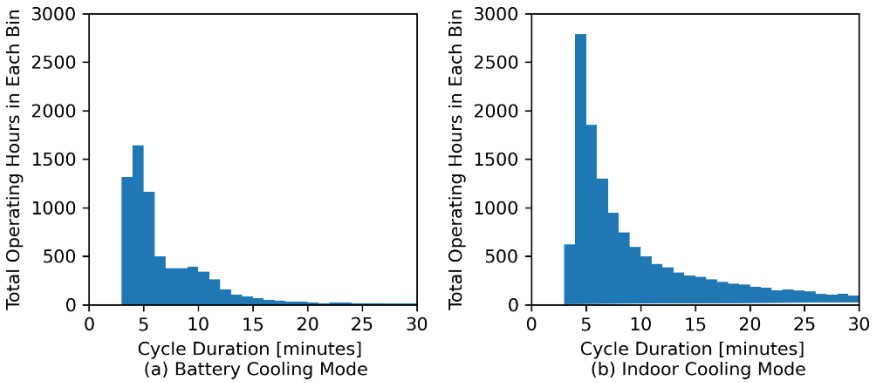


Figure 4: Cycle durations of the electric vehicle air conditioning systems (a) in the battery cooling mode and (b) in the indoor cooling mode.

Table 1: Statistics of average power consumption in each cycle for 40 bus air conditioners in Haikou from April 2021 to January 2022.

	Minimum	Lower quartile	Median	Upper quartile	Maximum
Battery cooling mode	1.5kW	3.4kW	3.6kW	3.8kW	5.2kW
Indoor cooling mode	1.5kW	3.6kW	3.9kW	4.3kW	10.7kW

Further, Figure 5 shows two typical operating cycles, one in the battery cooling mode and the other one in the indoor cooling mode. In the battery cooling cycle, after the start-up period, although the total power consumption is relatively constant, the cooling water return and supply temperatures gradually decrease and are not steady. In the indoor cooling cycle, the compressor frequency changes continuously to provide the required cooling load. Thus, instead of looking for steady-state periods in the system operation, this study extracts impact factors from each operating cycle and evaluates the system energy performance on the cycle level.

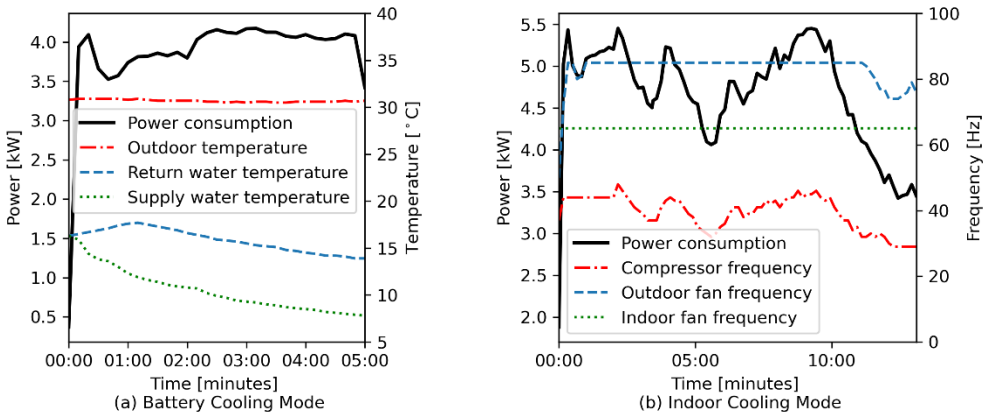


Figure 5: Change of physical quantities of typical operating cycles (a) in the battery cooling mode and (b) in the indoor cooling mode.

The energy consumption of an air conditioning system depends on several impact factors including the ambient environment and the actual cooling capacity. These impact factors are not directly related to operational faults and degradation, but can affect the benchmarks of the energy performance. Table 2 shows the selected impact factors in the battery cooling mode according to domain expertise. The ambient environment is represented by the outdoor temperature and the cooling water return temperature. Because a constant-speed pump is used to circulate the cooling water, the actual cooling capacity is represented by the cooling water temperature difference. In addition, the cycle duration may also affect the energy performance, so it is also included in the list of impact factors. Table 3 shows the selected impact factors in the indoor cooling mode. Since analysis in this mode mainly focuses on problems of fans and compressors, the fans and compressor frequency are included as impact factors.

Table 2: Selected impact factors in the battery cooling mode.

Physical parameter	Symbol	Unit	Type
Cooling water temperature difference	dT_w	[°C]	Time series
Cooling water return temperature	T_{wr}	[°C]	Time series
Outdoor temperature	T_o	[°C]	Time series
Cycle duration	t	[s]	Scalar

Table 3: Selected impact factors in the indoor cooling mode.

Physical parameter	Symbol	Unit	Type
Compressor frequency	f_{comp}	[Hz]	Time series
Indoor fan frequency	f_{fani}	[Hz]	Time series
Outdoor fan frequency	f_{fano}	[Hz]	Time series
Indoor Temperature	T_i	[℃]	Time series
Outdoor Temperature	T_o	[℃]	Time series

2.3 Encoding of Impact Factors with Adaptive Neural Network

To fairly benchmark the energy performance of a target system, its power consumption in each cycle should be compared with cycles from peer systems that have similar values of the impact factors. In this study, however, the impact factors of each cycle are variable-length time series (see Table 2 and Table 3), such that measurement of similarity is difficult. Thus, this study first applies an Autoencoder to compress the time series into a fixed-length vector, and then append the scalar (i.e. cycle duration in the battery cooling mode) to the vector. In this way, all impact factors in a cycle can be represented by a fixed-length vector.

Autoencoder is one type of artificial neural network which learns an efficient representation for a set of unlabeled data [24], typically used to reduce data dimensions. The inputs and outputs of an Autoencoder are both the original high-dimensional data. The Autoencoder first encodes the high-dimensional data into lower dimensions, and then decodes the lower-dimensional data back to the original data while minimizing the data reconstruction error. After the impact factors in each cycle are represented by a vector, similarities between vectors are measured by kernels in the Gaussian process.

The Autoencoder model needs to be updated periodically to adapt to changing operating conditions in real-time energy performance assessment. In this research, because the assessment is conducted every day with new streaming data and the encoded impact factors are required for similarity measurement in the Gaussian process regression model, the Autoencoder model is also updated every day to adapt to the new data. Besides, to ensure the

323 regression model has better generalizability, usually operational data in multiple days is
324 needed to develop a regression model for a system, so the Autoencoder is updated by data
325 from a fixed-length sliding window of multiple days.

326 Generally, there are three types of algorithms to update the neural network, namely
327 accumulative learning, incremental learning, and continual learning. In accumulative learning,
328 every day the Autoencoder is retrained from scratch using all data in the sliding window, so
329 this algorithm is also termed sliding window retraining. However, the computational cost of
330 accumulative learning is very high.

331 To reduce the computational cost, incremental learning utilizes the neural network models
332 that are already developed to update the new model. First of all, the model weights trained on
333 the first day are used as initial weights to train the model on the second day. In this way, the
334 new model could converge to the required precision with fewer epochs. Second, the number
335 of input data in the training process is reduced. Because the sliding windows between two
336 consecutive days have a large overlap, the model on the second day could just be fine-tuned
337 by new data collected on the same day. Thus, this algorithm is also termed sliding window
338 fine-tuning. Since neural network models usually partition input data into batches and are
339 trained on each batch, the number of batches can reduce by a few times in this algorithm and
340 the computational cost is significantly lower.

341 However, the drawback of incremental learning is catastrophic forgetting [25–28], which
342 refers to a situation where the model quickly loses knowledge of the previous tasks after
343 being trained with the current task. This is because when a model is trained sequentially on
344 multiple tasks, weights of the model that may be important for previous tasks will be
345 optimized to meet the objectives of new tasks. In this study, due to the gradual change of
346 climate conditions and corresponding control strategies, statistical properties of the data can
347 vary, termed concept drift, causing the Autoencoder to show high reconstruction error on data
348 that is not used in fine-tuning.

349 To overcome this problem, this paper adopts the Elastic Weight Consolidation (EWC)
350 algorithm [29]. EWC falls under the concept of continual learning, which refers to training a

model on multiple tasks sequentially without forgetting the knowledge obtained from the previous tasks. The fundamental idea of EWC is to give more constraints to the model weights that are important to previous tasks and fewer constraints to the unimportant weights. The loss function for the EWC algorithm is as follows:

$$L'(\theta) = L(\theta) + \frac{\lambda}{2}\sum_{k} b_k\left(\theta_k - \theta_k^b\right)^2 \quad (1)$$

where $L'(\theta)$ is the loss function for EWC, $L(\theta)$ is the loss function for the current task, θ_k are the weights to be optimized in the model, θ_k^b are the weights learned from the previous task, b_k measures the relative importance of each weight to the previous task, λ is a hyper-parameter which sets the importance of the previous task. The key in the EWC algorithm is to determine b_k for each weight in the model, and Kirkpatrick et al. [29] use the diagonal values of the Fisher information matrix to approximate b_k. In statistics, for a distribution with parameters θ_k that model an observable random variable *X*, the Fisher information matrix measures the amount of information that *X* carries about θ_k. In practice, a list of samples is drawn from the data trained in the previous task, and if the samples carry high Fisher information of some weights θ_k, then these weights are considered important for the previous task.

2.4 Power Prediction with Gaussian Process Regression

After the impact factors of each operating cycle are represented by a fixed-length vector, Gaussian process regression is applied to correlate the impact factors with the power consumption for each individual system. Note that the regression model of a system is used to benchmark the performance of its peer systems excluding itself.

Gaussian process regression is a supervised machine learning method that is able to generate probabilistic predictions in Gaussian distribution. Compared with other machine learning methods, the Gaussian process can compute the empirical confidence interval for each prediction using kernels. As an example, in Figure 6, the blue line shows the mean of the

probabilistic predictions, and the shaded area shows the confidence interval in one standard deviation. The model has lower uncertainty in the regions close to the training samples and higher uncertainty in the regions far away from the training samples, where the similarities (i.e. distances) between data can be measured by kernels. In performance benchmarking of vehicle air conditioning systems, because the air conditioners may not operate under uniform operating conditions, the regression model developed for one system sometimes cannot generalize to the operating conditions of another system, resulting in poor model predictions. However, the Gaussian process can quantify the uncertainty of the predictions so that the poor predictions can be rejected in the benchmarking process.

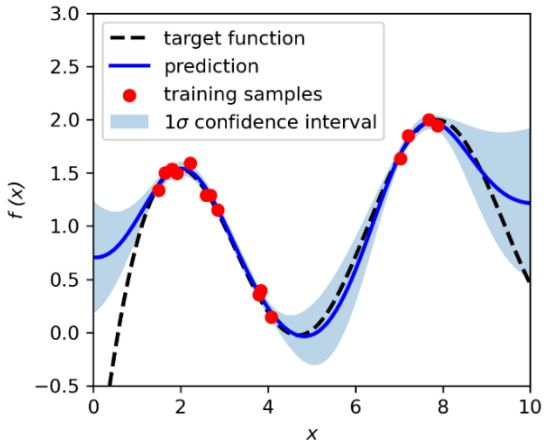


Figure 6: Illustration of Gaussian process regression. White noise is added to the training samples. The target function is $f(x) = 1 - e^{1-x} + \sin(x)$.

The methodology of Gaussian process regression is as follows. Consider a training set ($\boldsymbol{x}$, y) with M data points and the input $\boldsymbol{x}$ has N features. The output y can be represented by a function $f(\boldsymbol{x})$ and noises:

$$y = f(\boldsymbol{x}) + \varepsilon \tag{2}$$

where ε is the white noise with Gaussian distribution, i.e. $\varepsilon \sim N(0, \sigma_y^2)$. Then, given a test set $\boldsymbol{x}_*$ with M_* data points and its predictions denoted by $\boldsymbol{f}_*$, Gaussian process regression assumes $(\boldsymbol{y} \quad \boldsymbol{f}_*)^T$ follows a multivariate normal distribution:

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{x}) \\ \boldsymbol{\mu}(\boldsymbol{x}_*) \end{pmatrix}, \begin{pmatrix} \boldsymbol{K}_y & \boldsymbol{K}_* \\ \boldsymbol{K}_*^T & \boldsymbol{K}_{**} \end{pmatrix} \right) \tag{3}$$

and

$$\boldsymbol{K}_y = \boldsymbol{K} + \sigma_y^2 \boldsymbol{I}_M \tag{4}$$

where $\boldsymbol{y}$ has a length of M and $\boldsymbol{f}_*$ has a length of M_*, $\boldsymbol{\mu}(\boldsymbol{x})$ is the mean function of $\boldsymbol{x}$, $\boldsymbol{\mu}(\boldsymbol{x}_*)$ is the mean function of $\boldsymbol{x}_*$, $\boldsymbol{I}_M$ is an identity matrix of size M, $\boldsymbol{K}$ is an $M \times M$ matrix defined as the kernels between $\boldsymbol{x}$ and $\boldsymbol{x}$, $\boldsymbol{K}_*$ is an $M \times M_*$ matrix defined as the kernels between $\boldsymbol{x}$ and $\boldsymbol{x}_*$, and $\boldsymbol{K}_{**}$ is an $M_* \times M_*$ matrix defined as the kernels between $\boldsymbol{x}_*$ and $\boldsymbol{x}_*$.

It can be proved that the conditional probability density function of $\boldsymbol{f}_*$ is also in normal distribution:

$$P(\boldsymbol{f}_* \mid \boldsymbol{x}, \boldsymbol{x}_*, \boldsymbol{y}) = N(\boldsymbol{f}_* \mid \boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*) \tag{5}$$

The mean function ($\boldsymbol{\mu}_*$) and the covariance matrix ($\boldsymbol{\Sigma}_*$) in Equation (5) are calculated by:

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}(\boldsymbol{x}_*) + \boldsymbol{K}_*^T \boldsymbol{K}_y^{-1} (\boldsymbol{y} - \boldsymbol{\mu}(\boldsymbol{x})) \tag{6}$$

$$\boldsymbol{\Sigma}_* = \boldsymbol{K}_{**} - \boldsymbol{K}_*^T \boldsymbol{K}_y^{-1} \boldsymbol{K}_* \tag{7}$$

where the diagonal of $\boldsymbol{\Sigma}_*$ is the variance of each predicted $\boldsymbol{f}_*$. Note that $\boldsymbol{\Sigma}_*$ is fully determined by the kernel matrices.

The key in Gaussian process regression is the determination of kernels, because kernels specify how similarity is measured between $\boldsymbol{x}$ and $\boldsymbol{x}_*$ (i.e. the encoded impact factors). In this study, the similarity is measured by the Radial Basis Function (RBF) kernel and the Rational Quadratic (RQ) kernel with Euclidean distance adopted:

$$k_{RBF}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_1 \exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\theta_2^2}\right) \tag{8}$$

$$k_{RQ}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_3 \left(1 + \frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\theta_4 \theta_5^2}\right)^{-\theta_6} \tag{9}$$

where $\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|$ is the Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, and θ_1 to θ_6 are the parameters of the kernels.

Additionally, the white kernel is also added to explain the noise level of the data:

$$k_{white}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} \theta_7 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (10)$$

where θ_7 is a parameter that quantifies the variance of the data.

The final kernel adopted in this study is the summation of the above three kernels, and the parameters (θ_1, θ_2, …, θ_7) are optimized by the maximum likelihood estimation (MLE) given the training data ($\boldsymbol{x}$, $\boldsymbol{y}$):

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_{RBF}(\boldsymbol{x}_i, \boldsymbol{x}_j) + k_{RQ}(\boldsymbol{x}_i, \boldsymbol{x}_j) + k_{white}(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (11)$$

After Gaussian process regression models have been developed for each individual system, the energy performance of a target system can be benchmarked by the regression models of other systems. Consider a target system that has operational data ($\boldsymbol{x}_*$, P_{meas}) with M_* data points, where each $\boldsymbol{x}_*$ is a vector of the impact factors encoded by the Autoencoder and P_{meas} is the measured power consumption of the system. Then, predictions and prediction variances of the test data $\boldsymbol{x}_*$ can be evaluated by the regression models from all peer systems, denoted by P_{pred} and σ^2_{pred} respectively. After that, a variance filter is applied to eliminate the poor predictions due to the large dissimilarity between the test data $\boldsymbol{x}_*$ (i.e. data of the target system) and training data $\boldsymbol{x}$ (i.e. data of the peer systems). The variance filter calculates the upper quartile (i.e. the 75th percentile, also denoted by Q3) of the variance σ^2_{pred} for all predictions, and deletes the predictions with σ^2_{pred} greater than Q3. The remaining predictions are considered valid predictions, and are used to compute the energy performance index (EPI).

2.5 Calculation and Analysis of the Energy Performance Index (EPI)

With the above Gaussian process prediction models and the variance filter, valid power predictions are obtained and serve as the benchmarks for the target system. The energy performance index (EPI) of the target system is the difference of the measured power consumptions to the predicted power consumption. First of all, because the power consumption $P_{meas,i}$ for each cycle i of the target system usually has predictions $P_{pred,ij}$ from more than one peer system (j denotes the numbers of the peer systems), the median of predictions for each cycle is calculated:

$$P_{pred,i} = \text{median}\left(P_{pred,ij}\right) \tag{12}$$

where i is the cycle number of the target system and j is the peer system number. Note that this benchmarking method uses operational data without fault or fault-free labels, so a couple of faulty systems may exist in the list of peer systems. Although the majority of peer systems are assumed to be fault-free, a small fraction of faulty systems with abnormal predictions of power consumption can still bias the benchmark of the target system. Therefore, this research calculates the median value of the predicted power consumptions from each peer system. Since the median is more robust to outliers (i.e. abnormal predictions of power consumption) than the mean, the whole benchmarking method will be more robust to faulty peer systems.

Additionally, from the perspective of error propagation, the test error of $P_{pred,i}$ is significantly lower than the test error of $P_{pred,ij}$ predicted by each benchmarking model (i.e. Gaussian process regression model). That is to say, if the test error in each benchmarking model is estimated by the cross-validation error, then the error of $P_{pred,i}$ can be much smaller than those cross-validation errors.

Then, the difference of the measured power consumption to the predicted power consumption for each cycle i is calculated. EPI is defined as the median of the differences:

$$EPI = \text{median}\left(P_{meas,i} - P_{pred,i}\right) \tag{13}$$

EPI quantifies the energy performance of the target system relative to the average energy performance in the population. In the battery cooling mode, because the measured power and

predicted power are compared at the same cooling capacity, an EPI value greater than one indicates the system has lower coefficient of performance (COP) at the same ambient environment, and vice versa. In the indoor cooling mode, because the cooling capacity is unavailable, the change of EPI with time is more important because it indicates change of performance, usually performance degradation since spontaneous performance improvement is unlikely.

For the purpose of predictive maintenance, however, the method should also define a threshold to separate outliers from normal values of EPI or detect changes of EPI to identify performance changes. This study adopts the kernel density estimation (KDE) for outlier detection of EPI with battery cooling mode data and a change point detection algorithm proposed by Rogers et al. [30] to identify changes of EPI with indoor cooling mode data.

KDE is a statistical method which aims at estimating the probability density function (PDF) of a variable. Because the cooling capacity is measurable in the battery cooling mode, the EPI values can represent the COP of a system. A faulty system may have abnormally high EPI values, but because a system cannot perform much better than its peers, abnormally low EPI values should not appear. KDE is used to construct detection thresholds for EPI and identifies the EPI values that are above the thresholds as outliers. In this research, KDE treats each EPI value as a Gaussian kernel and then sums up all kernels to obtain the full PDF of EPI. From the PDF, thresholds for EPI can be determined by the percentiles of the distribution. This research uses the 99^{th} and 99.9^{th} percentiles as thresholds. If the EPI of a system is identified as an outlier for a few consecutive days, then the system is considered anomalous and technicians are required to be dispatched to examine the system.

The change point detection algorithm is based on the Student's *t* test. Given a time series of EPI, the detection algorithm performs *t* test at every possible bisections of the series. The method was proved to be effective and the authors used the Monte Carlo simulation to construct thresholds for changes at several confidence levels. In the indoor cooling mode, the change point detection algorithm is chosen to identify changes of energy performance. This research uses the 99.9% confidence level as a threshold to detect a change. If the change of

EPI for a system is greater than the 99.9% confidence level, then the energy performance of the system is considered having a change.

3 Implementation and Evaluation of the Benchmarking Method

The real-time energy performance benchmarking method is applied to electric bus air conditioning systems with the same configurations in Haikou and Sanya, China. The operational data is collected during the Spring of 2021 and the Summer of 2022. Section 3.1 details the data preprocessing procedures to remove sensor faults and data communication issues. Section 3.2 presents the architecture of the Autoencoder model to encode impact factors. Section 3.3 discusses the schemes to perform real-time benchmarking every day with lifelong learning algorithms. Then, the computational cost and performance of the lifelong learning algorithms are evaluated in Section 3.4. After that, in Section 3.5, the benchmarking method is applied to the battery cooling mode data of 40 systems which are essentially fault-free. Two detection thresholds of EPI are constructed with kernel density estimation. Finally, Section 3.6 applies the benchmarking method to the new operational data in both battery and indoor cooling modes in 2022, during which faults in the outdoor or indoor fan are imposed on three systems. The effectiveness of the benchmarking method is evaluated with the faulty systems.

3.1 Data Preprocessing

After raw data is retrieved from the database, data preprocessing is necessary before applying any benchmarking methods. This is because the sensors may malfunction sometimes, and also only the periods during which a system is in operation is useful for analysis. Data preprocessing is conducted with the following steps:

1. A temperature record is removed if its value is less than or equal to zero. Occasionally a thermistor may list the temperature as 0℃, but all of the monitored temperatures in these systems should be positive.

2. For each system, all of the on-cycles are extracted from the raw data and then sorted in chronological order. Note that because the fans, pumps, and the compressor do not start and stop simultaneously, the compressor frequency is used to determine the on-off status of a system.

3. Critical impact factors in Table 2 and Table 3 are extracted from the data.

4. An operating cycle is removed if the average power is lower than 1.5kW. Generally, the average power of a cycle is greater than 3kW, but occasionally a current meter may malfunction and record zero current.

5. An operating cycle is removed if the cycle duration is lower than 3 minutes or the number of updates in the cycle is fewer than 10. Note that the IoT sensors are set to update data every 10 seconds, but communication failure can occur.

6. Data in each operating cycle is resampled every 10 seconds for analysis, and the first half minute of each cycle is eliminated due to the transient behavior during system start-up.

7. Lastly, the impact factors are scaled between 0 and 1 using the global minimum and maximum values of each impact factor.

3.2 Architecture of the Autoencoder Model

An Autoencoder is developed after raw data is preprocessed. The architecture of the Autoencoder is shown in Figure 7. The number of features (i.e. variable-length impact factors) is three in the battery cooling mode and five in the indoor cooling mode. The multivariate time series is padded to a maximum length of 720 and the padded values are masked in the model. Note that each time series has been resampled every 10 seconds, and a maximum length of 720 allows an operating cycle as long as two hours to be included in the analysis, which is long enough for all operating cycles. The encoder and decoder are symmetrical, both with two long short-term memory (LSTM) layers. In the bottleneck layer, the input data is compressed into a vector with 16 elements.

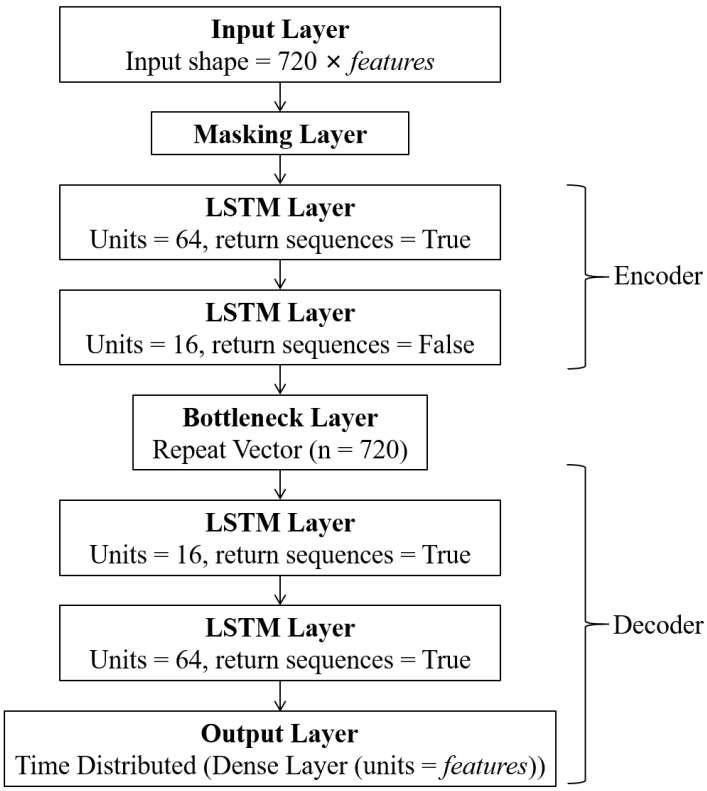
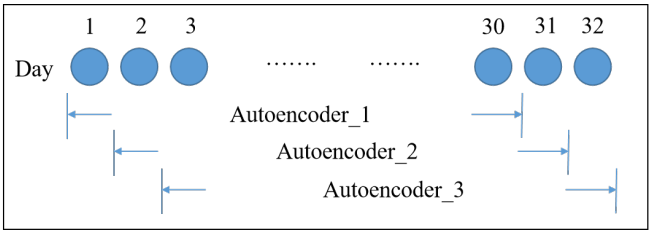


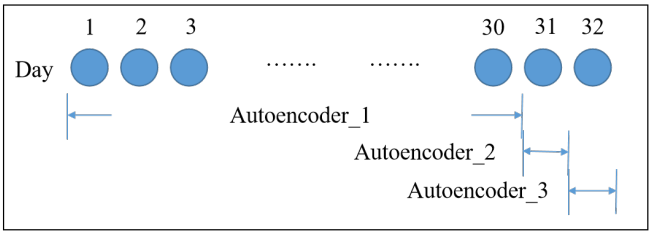
Figure 7: Architecture of the Autoencoder model in the battery cooling mode.

3.3 Schemes of Real-Time Energy Performance Benchmarking

To perform real-time benchmarking, the Autoencoder model is updated every day to adapt to new streaming data. Figure 8 shows the scheme of model update using accumulative learning, incremental learning, and continual learning respectively. In accumulative learning (Figure 8a), the Autoencoder is retrained from scratch each day by data in a 30-day sliding window. In incremental learning and continual learning (Figure 8b), the first model is still trained by data from the last 30 days. Starting from the second model, the Autoencoder is only fine-tuned by the new data collected on the current day. The next section will compare these three algorithms and show the advantage of using the continual learning algorithm.



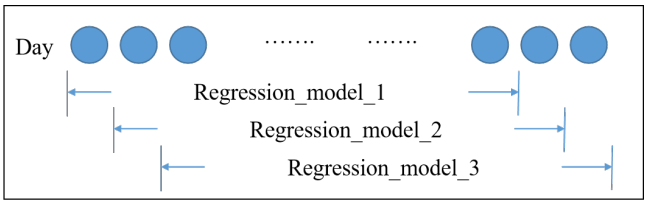
(a) Using the accumulative learning algorithm, the Autoencoder is retrained from scratch by data in a 30-day sliding window.



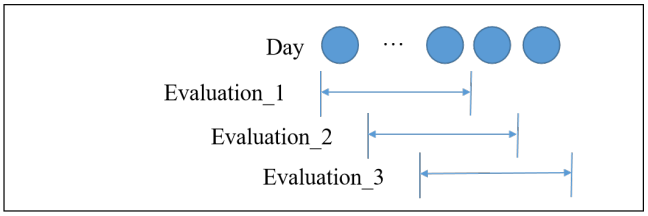
(b) Using the incremental learning or continual learning algorithm, starting from the second model, the Autoencoder is fine-tuned only by data of the current day.

Figure 8: Scheme of the Autoencoder model update using accumulative learning, incremental learning, and continual learning.

Figure 9 shows the scheme to develop Gaussian process regression models and energy performance assessment in real-time benchmarking. The Gaussian process power prediction models are established from the operational data of each system within the last 30 days in the battery cooling mode and the last 14 days in the indoor cooling mode. The energy performance of the 40 bus air conditioning systems is assessed every day based on the operating cycles within the last 3 days in the battery cooling mode and the last 1 day in the indoor cooling mode.



(a) For each peer system, the Gaussian process regression model is trained by data in the last 30 days in the battery cooling mode and by data in the last 14 days in the indoor cooling mode.



(b) For each target system, the energy performance is evaluated by its operating cycles in the last 3 days in the battery cooling mode and in the last 1 day in the indoor cooling mode.

Figure 9: Scheme of the real-time energy performance benchmarking method.

3.4 Evaluation of the Adaptive Neural Networks

This section compares the computational cost and performance between accumulative learning, incremental learning, and continual learning algorithms. In continual learning, the EWC algorithm is adopted. The data is collected from the battery cooling mode of 40 electric bus air conditioners in Haikou starting from April 1st, 2021. Because the Autoencoder model needs operational data in the last 30 days, the first Autoencoder model is developed on April 30th using the full set of training data from April 1st to April 30th (i.e. accumulative learning). Then beginning from May 1st, the accumulative learning algorithm still uses the full set of data in the last 30 days to retrain the model. But in incremental learning and continual learning algorithms, the Autoencoder is updated only by data collected on the same day.

3.4.1 Computational Cost Evaluation

First, the computational cost between the three algorithms is compared. The Autoencoder models are updated from May 1st to May 10th, and the computational time of the 10 models is recorded. The tasks are executed on a supercomputer with GPU cores. The average number of training samples for the Autoencoder models are 8,701. As discussed in Section 3.2, each

589 sample is a multivariate time series with a dimension of 720×3. Table 4 shows the average
590 training time of the 10 models using the three algorithms. Continual learning with the EWC
591 algorithm on average spends 7.6 minutes to update a model, which is only 4.8% of the time
592 compared with the accumulative learning algorithm and 158% of the time compared with the
593 incremental learning algorithm. Accumulative learning has a very high computational cost
594 and in practice is not considered. Incremental learning and continual learning both have much
595 shorter training time and their performance is compared next.

596 Table 4: Training time comparison between three learning algorithms

	Accumulative learning	Incremental learning	Continual learning
Average training time (min)	158.6	4.8	7.6
Standard deviation (min)	32.4	0.9	1.1

597

598 **3.4.2 Performance Evaluation Between Incremental and Continual Learning Algorithms**

599 After evaluation of the computational cost, the performance of the Autoencoder model using
600 incremental learning and continual learning algorithms is compared. Figure 10(a-c) shows the
601 reconstruction error of the three impact factors after encoding and decoding without EWC
602 applied (i.e. incremental learning). The first model on April 30th is trained by accumulative
603 learning, and then the models are updated until May 30th. The reconstruction error is
604 evaluated by the mean absolute percentage error (MAPE) defined as:

605 $$MAPE = \frac{1}{n}\sum_{0}^{n}\frac{\sum_{0}^{T_i}\left|F_{meas} - F_{pred}\right|}{\sum_{0}^{T_i} F_{meas}} \times 100\% \quad (14)$$

606 where F_{meas} is the measured value of an impact factor, F_{pred} is the predicted value of the
607 impact factor, T_i is the length of time series of cycle *i*, and n is the number of training
608 cycles. In each subplot of Figure 10, the lower right axis shows the date when the model is
609 updated, and the height of each bar shows the MAPE of the Autoencoder model for data
610 trained m days ago ($0 \leq m \leq 29$). As an example, the data collected on April 30th is only

611 trained on the same day and is considered being trained 0 days ago for the model on April 30th,
612 1 day ago for the model updated on May 1st, and 2 days ago for the model updated on May 2nd,
613 etc. When $m \geq 30$, the data is no longer used in benchmarking. As m increases, the
614 phenomenon of catastrophic forgetting is observed in the outdoor temperature and cooling
615 water temperature difference.

616 In comparison, Figure 10(d-f) shows the reconstruction error with EWC applied (i.e.
617 continual learning). The hyper-parameter λ in Equation (1) is set to 0.8 after grid search.
618 MAPE of the model almost remains the same as m increases. Thus, continual learning with
619 EWC is useful for lifelong training tasks and is applied in real-time peer-performance
620 benchmarking.

621

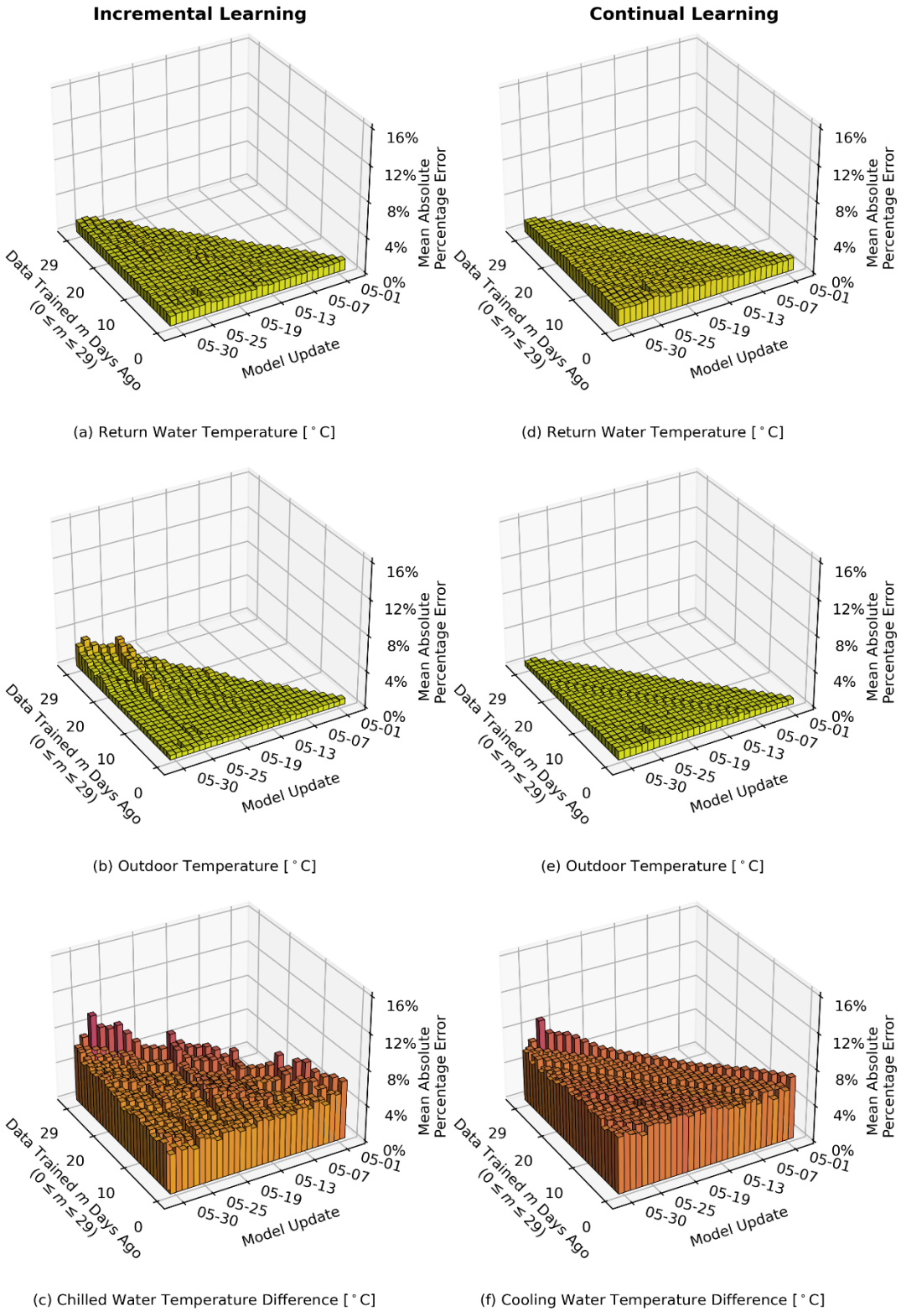
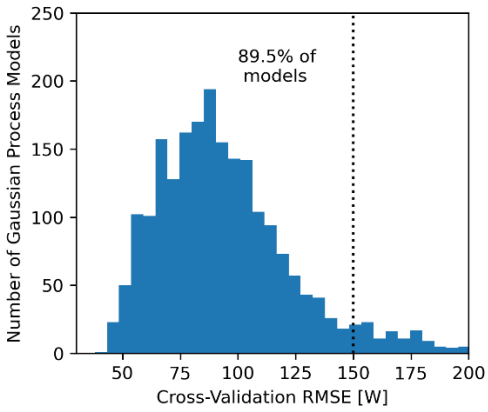


Figure 10: Mean absolute percentage error (MAPE) of the three selected impact factors in the Autoencoder model. (a-c) Incremental learning (EWC not applied, $\lambda = 0$); (d-f) Continual learning (EWC applied, $\lambda = 0.8$).

627 **3.5 Construction of Detection Thresholds for EPI in the Battery Cooling Mode**

628 In this section, the proposed benchmarking method is applied to the battery cooling mode
629 operational data of 40 air conditioning systems in May and June 2021 (two months). The
630 systems have only operated for about one year during that time, and are not expected to suffer
631 from severe operational faults. The Gaussian process regression models are developed using
632 the Scikit-learn package [31]. In the benchmarking method, five-fold cross validation is
633 performed for each Gaussian process regression model. Figure 11 shows the distribution of
634 the root mean square error (RMSE) for the models. To prevent overfitting, only regression
635 models with a cross-validation RMSE of below 150W are used for benchmarking, which
636 accounts for about 90% of models. Note that when the models are used for prediction, only 75%
637 predictions with lower prediction error will be used to calculate EPI (see Section 2.4).



638

639 Figure 11: Cross-validation root mean square error (RMSE) of all Gaussian process
640 regression models using battery cooling mode data in Haikou. 89.5% of models have an
641 RMSE of below 150W.

642 Figure 12 displays the distribution of the everyday EPI for the 40 systems. The majority of
643 EPI values are close to zero, and a few EPI values are above 150W. To separate the outliers
644 from normal EPI values, this study applies kernel density estimation (KDE) to generate a
645 distribution of EPI, and then find the 99^{th} and 99.9^{th} percentiles of EPI in the distribution. The
646 99^{th} percentile is the lower alert level, systems with EPI above this level may be in the early
647 stage of some fault and deserve attention. The 99.9^{th} percentile is the higher alert level,
648 systems with EPI above this level for a few consecutive days need inspection and

maintenance. Note that the probability of getting a false alarm under the 99th percentile is 1%, and under the 99.9th percentile is 0.1%.

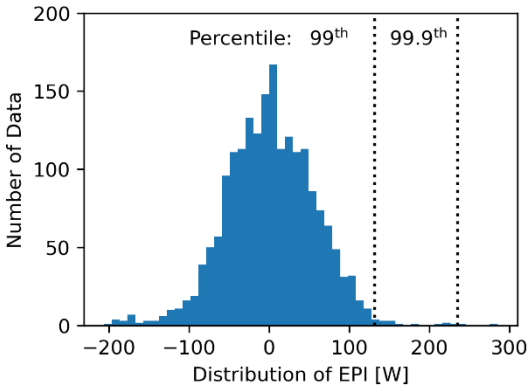


Figure 12: Distribution of the everyday EPI for all systems in Haikou from May to June 2021. The 99th percentile of EPI is 132W and the 99.9th percentile of EPI is 235W.

Figure 13 shows the change of EPI for the 40 electric vehicle air conditioning systems in May and June. Generally, the EPI values are within the 99th percentile, and no system has abnormally high EPI values lasting for a long period. Thus, these systems are considered to operate under healthy conditions. As these 40 electric vehicles are relatively new and no maintenance was called during the period according to the maintenance record, the result is credible.

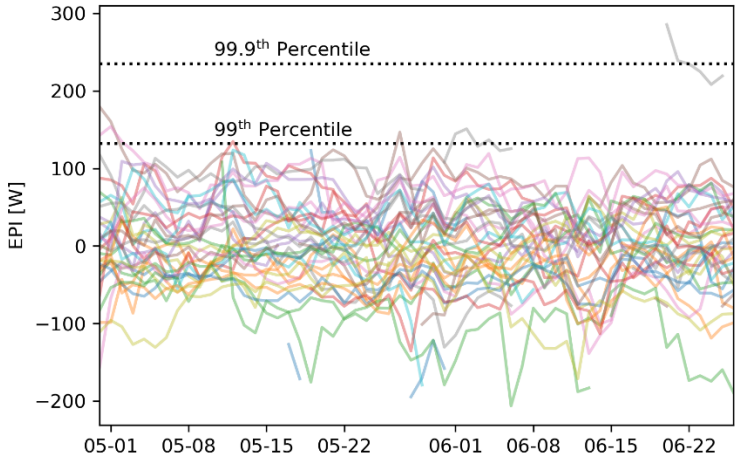


Figure 13: Variations of EPI for the 40 electric vehicle air conditioning systems from May to June 2021. Each curve represents the energy performance of a single system. The 99th and 99.9th percentiles of EPI are shown by the dotted line.

3.6 Validation of the Benchmarking Method with Faulty Systems

This section applies the benchmarking method to electric bus air conditioners installed in city bus fleets. Section 3.6.1 evaluates the performance of the benchmarking method on battery cooling mode data, and Section 3.6.2 evaluates the performance on indoor cooling mode data. In total three faulty systems with indoor or outdoor fan problems are shown, and results show that the method can successfully detect these types of faults.

3.6.1 Validation with Data in the Battery Cooling Mode

The previous section constructs two detection thresholds of EPI for battery cooling mode data, then this section applies the benchmarking method to the same 40 systems operating in July 2022. On July 19th, 2022, a fault is imposed on one of the systems by lowering the outdoor fan speed by 30%. Referring to Figure 14, after the outdoor fan speed is decreased from $f_{o,1}$ to $f_{o,3}$, the condenser temperature will rise, and the variable-speed compressor will accelerate and work harder (frequency rises from $f_{comp,1}$ to $f_{comp,3}$). Although the outdoor fan power is reduced, because the compressor power accounts for a larger share of the system power consumption, the whole system is expected to consume more energy (i.e. power increases

from P_1 to P_3) to provide the same amount of cooling capacity. However, because the faulty system is evaluated at the same cooling capacity ($Q_2 = Q_1$), the power consumption of its benchmarks (P_2) remains unchanged. As a result, EPI of the faulty system, defined as the difference of power consumption between a system and its benchmarks, will increase from $P_1 - P_2$ to $P_3 - P_2$.

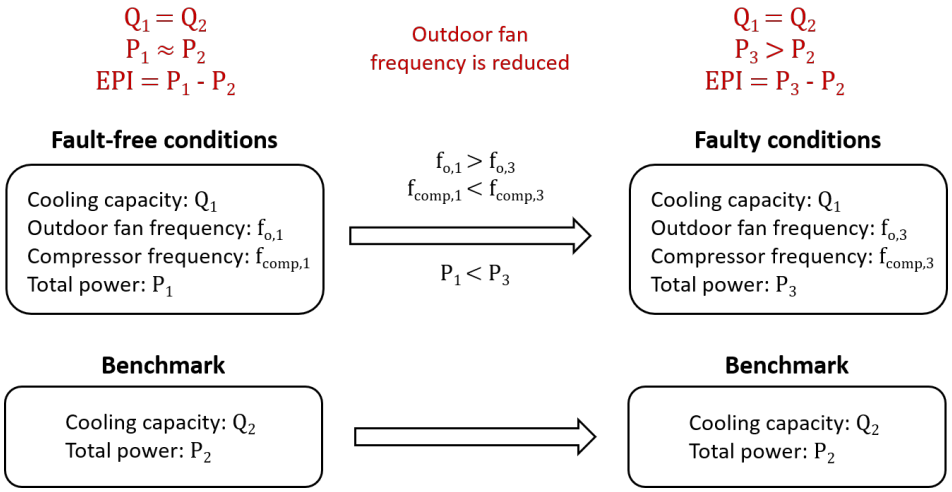
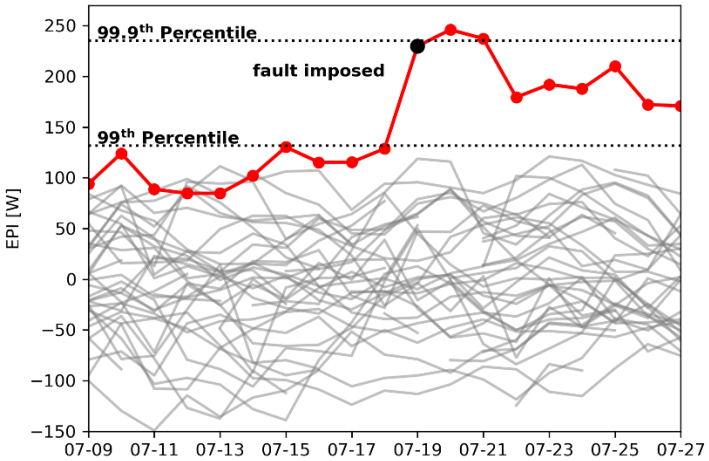


Figure 14: Impact of the outdoor fan fault to benchmarking in the battery cooling mode.

Figure 15 shows the performance of the 40 systems in July 2022. The performance of the faulty system is marked in red while the other systems are shown in grey. The thresholds are adopted from Section 3.5. After the fault is imposed, the EPI of the system rises from 129W to 246W, approaching the higher alert level (i.e. 99.9th percentile). Note that in the battery cooling mode, the compressor usually operates in partial load and has relatively high efficiency, so the approximately 117W increase in energy consumption is reasonable.



693

694 Figure 15: Variations of EPI in the battery cooling mode for the 40 electric vehicle air
695 conditioning systems in July 2022. The red curve shows the faulty system with declined
696 outdoor fan speed, and the grey curves show the performance of the other systems. The 99th
697 and 99.9th percentiles of EPI obtained from the data in May and June 2021 (see Section 3.5)
698 are shown by the dotted line.

699

700 **3.6.2 Validation with Data in the Indoor Cooling Mode**

701 This study also applies the benchmarking method to the indoor cooling mode data with two
702 case studies. In the first case study, data of 10 vehicle air conditioners in Sanya is processed to
703 prove the benchmarking method is also effective with only 10 peer systems. Same as above,
704 the outdoor fan fault is also imposed on one system. However, because the impact factors to
705 benchmark indoor cooling mode data also include fan and compressor frequencies, the impact
706 of the outdoor fan fault on EPI is different. Referring to Figure 16, the original outdoor fan
707 frequency ($f_{o,1}$) decreases by 30% in faulty conditions assuming the fan is unable to operate
708 at the required frequency (i.e. $f_{o,3\_real} = 0.7f_{o,1}$). But usually the real frequency is not
709 measured, and the frequency data comes from the control signal ($f_{o,3\_control}$) which still
710 requires the fan to operate at $f_{o,1}$. Thus, after the fault occurs, power consumption of the
711 faulty system will decrease from P_1 to P_3 due to a decline of real outdoor fan frequency, but
712 the power consumption is still evaluated at $f_{o,3\_control}$. As a result, EPI of the faulty system
713 will decrease from $P_1 - P_2$ to $P_3 - P_2$. Note that the decrease of power consumption does

not indicate the system has better performance. Instead, because the real outdoor fan frequency is lower, the cooling capacity of the faulty system also decreases from Q_1 to Q_3, just unmeasurable in the indoor cooling mode.

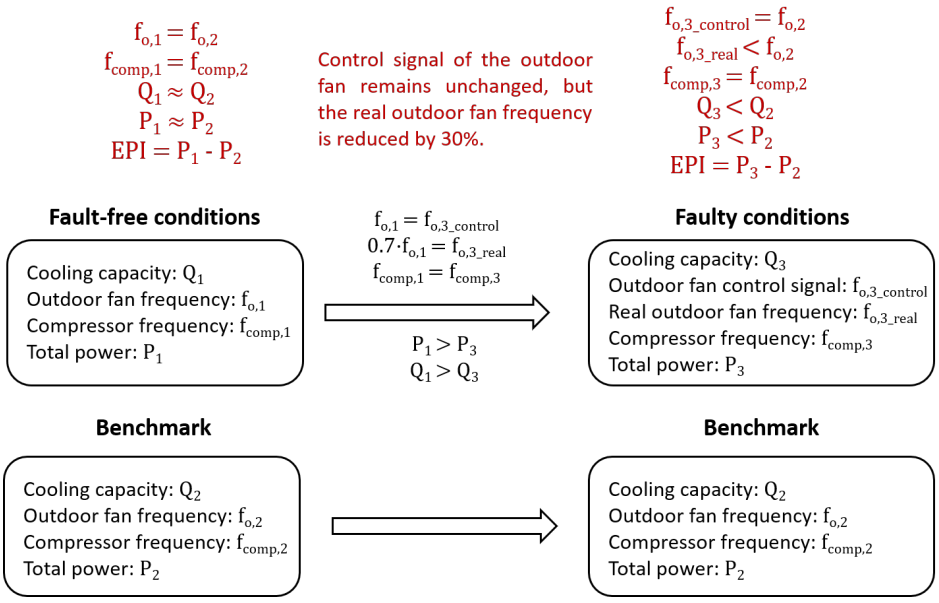


Figure 16: Impact of the outdoor fan fault to benchmarking in the indoor cooling mode.

Figure 17 shows the cross-validation RMSE of Gaussian process regression models using the indoor cooling mode data of 10 Sanya vehicle air conditioning systems. Only models with RMSE of below 75W are used in Benchmarking, and almost all models satisfy this criterion.

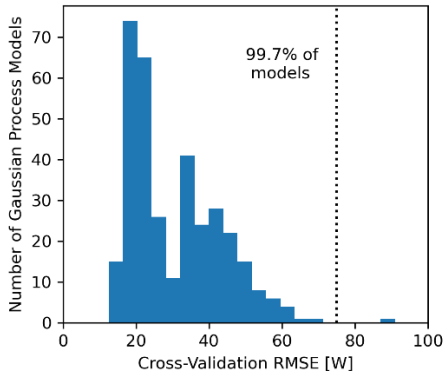


Figure 17: Cross-validation root mean square error (RMSE) of all Gaussian process regression models using indoor cooling mode data in Sanya. Almost all models have an RMSE of below 75W.

Figure 18 shows the benchmarking results of the 10 systems with the faulty one marked in red. From the upper subplot, after fault is imposed, the EPI drops significantly from -100W to -300W. Note that the EPI values between the indoor cooling mode and battery cooling mode are incomparable since they are evaluated at different impact factors. The lower plot performs change detection for all 10 systems in a 14-day sliding window. If a change is detected in any window with over 99.9% confidence level, the magnitude of change will be plotted. The plotted changes are *statistically* significant changes. However, the magnitude of most statistically significant changes is relatively small probably caused by disturbance and modeling error. The study only considers changes that greater than 150W as *practically* significant changes. The result shows only the faulty system has practically significant changes since fault occurs. After 14 days passed, the data before fault occurs is no longer included in the sliding window so no practically significant change is detected any more.

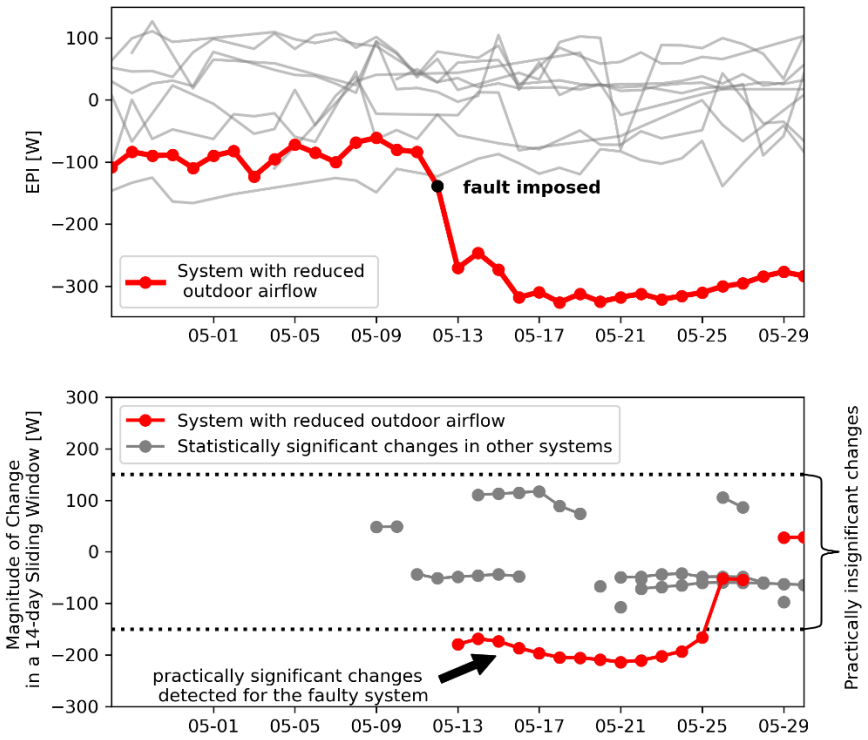
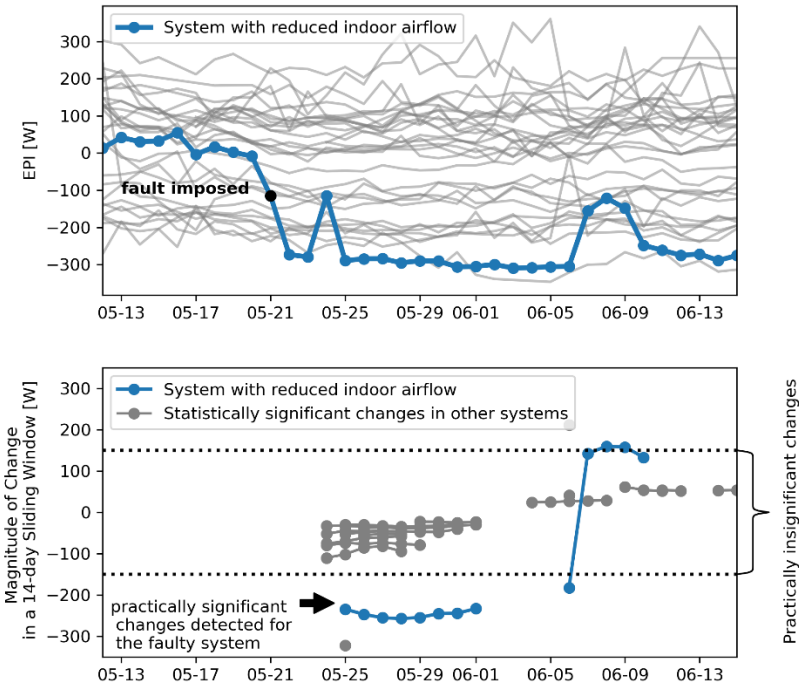


Figure 18: EPI values and change detection of EPI in the indoor cooling mode for the 10 electric vehicle air conditioning systems in May 2022. The red curve shows the faulty system with declined outdoor fan speed, and the grey curves show the performance of the other systems.

743 In the second case study, the benchmarking method is applied to the indoor cooling mode data
744 of another 40 vehicle air conditioners in Haikou in May and June 2022. The aim is to prove
745 the method is also capable of detecting indoor fan issues. The detection results are shown in
746 Figure 19. Based on the similar thermodynamic principles discussed in Figure 16, a reduced
747 indoor airflow will also cause the EPI to decrease. In Figure 19, EPI of the faulty system
748 decreases from about 0W to -300W, and the decline can be detected by the change detection
749 algorithm with a 14-day moving window until June 1st.



750

751 Figure 19: EPI values and change detection of EPI in the indoor cooling mode for the 40
752 electric vehicle air conditioning systems in May and June 2022. The blue curve shows the
753 faulty system with declined indoor fan speed, and the grey curves show the performance of
754 the other systems.

755

756 **4 Conclusion**

757 This paper presents a real-time benchmarking method to assess the energy performance of
758 electric vehicle air conditioning systems. Because electric vehicles are mass-produced and
759 becoming popular, the energy performance of each air conditioning system is evaluated by the

performance of its peer systems. The principal benefit of this approach is that streaming data from IoT sensor records can be directly processed by the benchmarking method without faulty or fault-free labels.

The proposed method focuses on the battery cooling refrigeration cycle of electric vehicles, and evaluates the energy consumption of each target system using the aggregated performance of its peer systems. To ensure energy performance between systems is compared under the same operating conditions, critical impact factors of the system energy consumption are identified. The Autoencoder model compresses the impact factors into fixed-length vectors with minimal loss of information, enabling similarity measurement between impact factors. With the impact factors encoded, Gaussian process regression models are developed for each individual system to benchmark its peer systems excluding itself. Similarities of the encoded impact factors between systems are computed through the RBF and RQ kernels in the Gaussian process. When the operating conditions of a target system are very different from some peer system, power consumptions predicted from the Gaussian process regression model of the peer system will have high variance. Predictions with high variance are rejected later with a variance filter. Lastly, the EPI is computed by comparing the real energy consumption of a system to its benchmarks predicted by the peer systems.

The proposed benchmarking models are updated continuously to learn knowledge from the new streaming data. To reduce the computational cost of model retraining while avoiding the catastrophic forgetting issue, a continual learning algorithm termed EWC is applied to update the Autoencoder model. The EWC algorithm is proved to be effective in learning knowledge in new training tasks while remembering the knowledge in old training tasks.

The benchmarking method is applied to monitor the EPI of a lot of electric bus air conditioning systems in Haikou and Sanya, China. The operational data in 2021 is basically fault-free, and thus is used to construct two detection thresholds to identify the outliers of EPI in the battery cooling mode. Then, this study applies the benchmarking method to identify faulty systems with reduced outdoor and indoor fan speed. The faulty systems have a moderate increase of EPI in the battery cooling mode and significant decrease of EPI in the indoor cooling mode, which are consistent with thermodynamic principles. The results proved

the benchmarking method is effective in evaluating the energy performance of vehicle air conditioning systems.

Acknowledgement

The authors gratefully acknowledge the support of this research by the Postdoctoral Fellowship Scheme at the Hong Kong Polytechnic University and the Research Grants Council of the Hong Kong SAR (C5018-20GF).

References

[1] IEA Transport, (2021). https://www.iea.org/topics/transport (accessed October 22, 2022).
[2] Use of energy for transportation - U.S. Energy Information Administration (EIA). https://www.eia.gov/energyexplained/use-of-energy/transportation.php (accessed October 23, 2022).
[3] Transport – Topics, IEA. (n.d.). https://www.iea.org/topics/transport (accessed October 22, 2022).
[4] Global EV Outlook 2021 – Analysis, IEA. https://www.iea.org/reports/global-ev-outlook-2021 (accessed October 22, 2022).
[5] C. Fiori, K. Ahn, H.A. Rakha, Power-based electric vehicle energy consumption model: Model development and validation, Applied Energy. 168 (2016) 257–268. https://doi.org/10.1016/j.apenergy.2016.01.097.
[6] Z. Zhang, C. Liu, X. Chen, C. Zhang, J. Chen, Annual energy consumption of electric vehicle air conditioning in China, Applied Thermal Engineering. 125 (2017) 567–574. https://doi.org/10.1016/j.applthermaleng.2017.07.032.
[7] J. Wang, K. Liu, T. Yamamoto, T. Morikawa, Improving Estimation Accuracy for Electric Vehicle Energy Consumption Considering the Effects of Ambient Temperature, Energy Procedia. 105 (2017) 2904–2909. https://doi.org/10.1016/j.egypro.2017.03.655.
[8] K. Liu, J. Wang, T. Yamamoto, T. Morikawa, Exploring the interactive effects of ambient temperature and vehicle auxiliary loads on electric vehicle energy consumption, Applied Energy. 227 (2018) 324–331. https://doi.org/10.1016/j.apenergy.2017.08.074.
[9] E. Afrasiabian, R. Douglas, M. Geron, G. Cunningham, A numerical evaluation of a novel recovery fresh air heat pump concept for a generic electric bus, Applied Thermal Engineering. 209 (2022) 118181. https://doi.org/10.1016/j.applthermaleng.2022.118181.
[10] H.-S. Lee, M.-Y. Lee, Cooling Performance Characteristics on Mobile Air-Conditioning System for Hybrid Electric Vehicles, Advances in Mechanical Engineering. 5 (2013) 282313. https://doi.org/10.1155/2013/282313.
[11] M.-Y. Lee, H.-S. Lee, H.-P. Won, Characteristic Evaluation on the Cooling Performance of an Electrical Air Conditioning System Using R744 for a Fuel Cell Electric Vehicle,

826 Energies. 5 (2012) 1371–1383. https://doi.org/10.3390/en5051371.
827 [12] Y. Song, H. Wang, Y. Ma, X. Yin, F. Cao, Energetic, economic, environmental
828 investigation of carbon dioxide as the refrigeration alternative in new energy bus/railway
829 vehicles' air conditioning systems, Applied Energy. 305 (2022) 117830.
830 https://doi.org/10.1016/j.apenergy.2021.117830.
831 [13] X. Huang, K. Li, Y. Xie, B. Liu, J. Liu, Z. Liu, L. Mou, A novel multistage constant
832 compressor speed control strategy of electric vehicle air conditioning system based on
833 genetic algorithm, Energy. 241 (2022) 122903.
834 https://doi.org/10.1016/j.energy.2021.122903.
835 [14] M. Yao, Y. Gan, J. Liang, D. Dong, L. Ma, J. Liu, Q. Luo, Y. Li, Performance simulation
836 of a heat pipe and refrigerant-based lithium-ion battery thermal management system
837 coupled with electric vehicle air-conditioning, Applied Thermal Engineering. 191 (2021)
838 116878. https://doi.org/10.1016/j.applthermaleng.2021.116878.
839 [15] Q. Lei, C. Zhang, J. Shi, J. Chen, Machine learning based refrigerant leak diagnosis for a
840 vehicle heat pump system, Applied Thermal Engineering. (2022) 118524.
841 https://doi.org/10.1016/j.applthermaleng.2022.118524.
842 [16] A. Rafati, H.R. Shaker, S. Ghahghahzadeh, Fault Detection and Efficiency Assessment
843 for HVAC Systems Using Non-Intrusive Load Monitoring: A Review, Energies. 15 (2022)
844 341. https://doi.org/10.3390/en15010341.
845 [17] Z. Tian, Ch. Qian, B. Gu, L. Yang, F. Liu, Electric vehicle air conditioning system
846 performance prediction based on artificial neural network, Applied Thermal Engineering.
847 89 (2015) 101–114. https://doi.org/10.1016/j.applthermaleng.2015.06.002.
848 [18] J.-D. Wu, J.-Y. Ke, F.-Y. Shih, W.-J. Shyr, Fault diagnosis for vehicle air conditioning
849 blower using deep learning neural network, Journal of Low Frequency Noise, Vibration
850 and Active Control. (2022) 146134842210858.
851 https://doi.org/10.1177/14613484221085891.
852 [19] Y. Zhou, C. Lork, W.-T. Li, C. Yuen, Y.M. Keow, Benchmarking air-conditioning energy
853 performance of residential rooms based on regression and clustering techniques, Applied
854 Energy. 253 (2019) 113548. https://doi.org/10.1016/j.apenergy.2019.113548.
855 [20] Y. Zhao, P. Liu, Z. Wang, L. Zhang, J. Hong, Fault and defect diagnosis of battery for
856 electric vehicles based on big data analysis methods, Applied Energy. 207 (2017) 354–
857 362. https://doi.org/10.1016/j.apenergy.2017.05.139.
858 [21] Y. Zhou, X. Tian, C. Zhang, Y. Zhao, T. Li, Elastic weight consolidation-based adaptive
859 neural networks for dynamic building energy load prediction modeling, Energy and
860 Buildings. 265 (2022) 112098. https://doi.org/10.1016/j.enbuild.2022.112098.
861 [22] C. Lee, S.-H. Kim, C.-H. Youn, An Accelerated Continual Learning with Demand
862 Prediction based Scheduling in Edge-Cloud Computing, in: 2020 International
863 Conference on Data Mining Workshops (ICDMW), IEEE, Sorrento, Italy, 2020: pp. 717–
864 722. https://doi.org/10.1109/ICDMW51313.2020.00103.
865 [23] Z. Chen, F. Guo, F. Xiao, X. Jin, J. Shi, W. He, Development of Data-Driven
866 Performance Benchmarking Method for A Large Number of Bus Air Conditioners,
867 International Journal of Refrigeration. (in review).
868 [24] M.A. Kramer, Nonlinear principal component analysis using autoassociative neural
869 networks, AIChE J. 37 (1991) 233–243. https://doi.org/10.1002/aic.690370209.

[25] R.M. French, Catastrophic forgetting in connectionist networks, Trends in Cognitive Sciences. 3 (1999) 8.
[26] J.L. McClelland, B.L. McNaughton, R.C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory., Psychological Review. 102 (1995) 419.
[27] M. McCloskey, N.J. Cohen, Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem, in: Psychology of Learning and Motivation, Elsevier, 1989: pp. 109–165. https://doi.org/10.1016/S0079-7421(08)60536-8.
[28] R. Ratcliff, Connectionist models of recognition memory: constraints imposed by learning and forgetting functions., Psychological Review. 97 (1990) 285.
[29] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A.A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, R. Hadsell, Overcoming catastrophic forgetting in neural networks, Proc. Natl. Acad. Sci. U.S.A. 114 (2017) 3521–3526. https://doi.org/10.1073/pnas.1611835114.
[30] A.P. Rogers, F. Guo, B.P. Rasmussen, A Change Point Detection Algorithm with Application to Smart Thermostat Data, 126 14.
[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) 2825–2830.