Title
Operation and performance of VRF systems: Mining a large-scale dataset

Authors
Qian, M
Yan, D
Hong, T
et al.

Publication Date
2021

DOI
10.1016/j.enbuild.2020.110519

Peer reviewed

Operation and Performance of VRF Systems: Mining a Large-Scale Dataset

Mingyang Qian[1], Da Yan[1]*, Tianzhen Hong[2], Hua Liu[3]

[1] Building Energy Research Center, School of Architecture, Tsinghua University, Beijing, China

[2] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, USA

[3] State Key Laboratory of Air-conditioning Equipment and System Energy Conservation, Zhuhai, China

*Corresponding author email: D. Yan, yanda@tsinghua.edu.cn

Abstract

The energy consumption of air-conditioning systems has gained increasing attention as it contributes significantly to the global building energy use. The variable refrigerant flow (VRF) system is a common air-conditioning system applied widely in residential and office buildings in China. Understanding the actual operation and performance of VRF systems is fundamental for the energy-efficient design and operation of VRF systems. Previous research on VRF system operation used either limited field data covering certain building types and climate zones or used a questionnaire to obtain a larger dataset. However, they did not capture the wide applications of VRF systems quantitatively across all building types, climate zones, and operating conditions. To fill this gap, statistical and clustering analysis was conducted on the newly proposed key performance indicators of approximately 287,000 VRF systems for residential and commercial buildings in all five climate zones in China. The main findings are: (1) VRF systems are mainly used for cooling in all climate zones in China; (2) among all building types, the duration of use is lowest in residential buildings and highest in hotels and medical buildings; (3) the distribution of the ideal VRF cooling coefficient of performance (COP) is similar across all climate zones and building types; whereas the COPs of ideal VRF heating in the Severe Cold region and Cold regions are lower than those in other climate zones; and (4) partial load operations for VRF systems are common in residential buildings and office buildings due to the part-time-part-space operation mode. These findings can inform the actual application of VRF systems in China, supporting the design, operation, industry standard development, and performance optimization of VRF systems.

Keywords: **VRF system, data mining, big data, system performance, China, energy efficiency**

Highlights:

- New key performance indicators were proposed to benchmark actual application of 287,000 VRF systems in all five climate zones in China.
- Data mining on a large-scale dataset is a powerful tool to obtain statistical distributions of VRF systems for different types of buildings and climate zones.
- Cluster was applied to analyze typical operating load patterns of VRF systems for residential and office buildings.
- Partial load operations are common for VRF systems in residential buildings and office buildings.
- Findings inform the design, operation, industry standard development, and performance optimization of VRF systems.

1 Introduction

Cooling buildings is the fastest growing end use of energy and comprises 6% of overall building energy use because of global warming and increases in population and economic growth [1]. Therefore, there is a strong need to reduce cooling energy use by improving energy efficiency and energy conservation in buildings [2].

The variable refrigerant flow (VRF) system is widely used for space cooling, which is a kind of ductless multi-split systems providing cooling for multiple rooms using one or more outdoor units and multiple indoor evaporator units. In air conditioning market, mini- and multi-ductless splits accounted for 77% of total capacity in the world [1]. The VRF systems are widely applied in China with a 50.35% share of the central air conditioning market [3]. The VRF systems can easily operate indoor units to meet occupant's comfort demand and different rooms' loads [4].

Owing to the flexible adjustability and controls of VRF systems, their operation and performance are complicated. The actual operation and performance of VRF systems are fundamental to their design and operation. Due to the stochastic nature of occupant behavior in buildings with VRF systems, there are wide variations in their operation and performance, e.g., use duration, performance, and load ratio [5]. It is important to get the actual operation and performance of VRF systems. In order to capture the actual operation of VRF systems, various measured cases are necessary by reflecting stochastic and diverse occupant behavior. Previous research has already conducted field test for VRF systems operation in office buildings [6]. Large amount of questionnaires is a good tool to get various occupant behavior of using air conditioning. The design and sampling of questionnaire can affect the quality of reflecting the actual use of air conditioning [7]. However, there is still a significant gap between the questionnaire results and real operation of air conditioning. For VRF energy performance, laboratory tests and field measurement are traditional methods to get actual VRF systems energy performance in buildings. Laboratory test is measuring the performance of air conditioning equipment and system by controlling rating conditions [8]. However, the previous research pointed that rating conditions in laboratory are far different from the actual operation conditions in buildings. Therefore, several previous research conducted field measurements to get actual operation performance of air conditioning systems. Zhang et al. applied new method for measuring field performance of VRF systems to a VRF system in an office building over four weeks during the heating season [9]. The field measurements are always conducted on a limited number of buildings and certain period because of high cost of field measurement. The actual operation performance of air conditioning is also influenced by the uncertain occupant behavior and changing outdoor weather conditions [10]. Therefore, a large amount of actual field measurements is important to capture the actual operation and performance of VRF systems.

It has been previously observed that a large amount of measured data in real buildings is becoming available in recent years, which is benefited from the wide deployment of low-cost sensors, meters, as well as the internet of things [11]. Mathew et al. established United States' national building performance database based on energy data from more than one million commercial and residential buildings [12]. Environmental data is another common type of measured data in real buildings from indoor environmental monitoring system or smart thermostat. Stopps and Touchie collected data from 56 thermostats installed in two multi-family residential buildings [13]. Cetin obtained ON–OFF operation patterns and energy consumption of air conditioning system in 189 homes from energy-management systems [14]. Other systems in buildings such as internet of things (IoT) devices and Wi-Fi systems may also record occupant data, which can help the research of air conditioning operation. Rafsanjani et al. collected data of ten occupants in an office building during a six-week period in summer by IoT devices [15]. The increasing amount of actual field measurements makes it possible to get the actual operation and performance of air conditioning systems.

The existing body of research suggests that data mining methods can help capture new knowledge from big data sets [16] and have been applied to different fields of HVAC. For occupant behavior of HVAC, D'Oca and Hong derived occupancy patterns in buildings based on a data set of 16 offices using a data mining framework [17]. For energy performance of HVAC, Kontokosta and Tull used linear regression (OLS), random forest, and support vector regression (SVM) algorithms for city's energy benchmarking data and developed a predictive model to predict electricity and natural gas use for every property in the city [18]. For HVAC load pattern analysis, unsupervised method clustering is a common way to conduct load profiling [19]. Clustering is one of the most popular descriptive data mining techniques because it consumes less time and needs less supervision. Quintana et al using load shape clustering to detect uncharacteristic electricity use behavior [20]. Li et al. proposed clustering analysis and association rules mining to identify and interpret the power consumption patterns and associations for VRF systems [21]. Lu et al. conducted a new Gaussian Mixture Model (GMM) clustering to identify temperature related sub-pattern and people behavior related sub-pattern, and the clustering result is further utilized to improve the accuracy of prediction models [22].

Several attempts have been made to conduct data mining and machine learning method on VRF system modeling, control, and fault detection and diagnosis studies [23]. VRF system's energy consumption and modeling is a main researching topic for data mining. Qian et al. got 344 samples of operating data from real residential buildings to calculate the performance of a large-scale VRF [24]. Guo et al. proposed a virtual VRF power sensor to get energy consumption of VRF systems [25]. Liu et al. proposed a data-mining-based method to benchmark VRF system energy performance [26]. Li et al. used multiple linear regression (MLR) and non-linear support vector regression (SVR) to improve the prediction for refrigerant charge [27]. Learning VRF system's operation mode for control is another research point by data mining. Moon developed an artificial neural network (ANN) model to predict the heating energy cost during the next control cycle for VRF systems [28]. Liu attempted to get typical operation mode of VRF systems in residential buildings through clustering analysis [29]. Fault detection and diagnosis is also the research hit conducted by machine learning and data mining. Liu et al. proposed a data mining method for fault detection and diagnosis of VRF systems [30]. Sun et al. proposed a hybrid model combined support vector machine (SVM) with wavelet de-noising (WD) for diagnosing refrigerant charge faults [31].

Data from several studies suggest that most AC systems in China operate in a part-time-part-space mode. An et al. [32] and Qian et al. [33] found that the main AC operation mode is a part-time-part-space mode for fan-coil-unit (FCU) systems in residential communities. Yu et al. conducted field test in five office buildings to compare VAV and VRF systems and founded the part-time-part-space mode for office buildings with VRF systems [10]. Hu et al. used large scale of questionnaires to conclude that AC systems in China are operated in a part-time-part-space mode [7]. The data used by previous research is limited field test and questionnaires, which may need more real measured data to support the conclusion. A large VRF dataset may further support the reality of part-time-part-space mode for AC systems in China.

Although, some research has been carried out on operation and performance on VRF systems and AC systems operating modes, all the data used by previous research is usually small scale, covering only a single building type and one climate zone. The diverse loads and operation performance of wide-scale adoption of VRF systems have not been revealed.

The contributions of this paper is proposed new key performance indicators of 287,000 VRF systems in China by conducting statistical and clustering analysis, which cover various building types, climate zones, and operating conditions. This research contributed a large amount of VRF systems and high coverage for different climate zones in China. The remaining of the paper is organized as: Section 2 describes the dataset and the analysis method, Section 3 presents the key performance indicators, Section 4 illustrates potential applications of the outcomes, Section 5 discusses the policy implications and limitations of the current study, and Section draws the conclusion.

2 Dataset and Methods

2.1 Dataset introduction

The VRF dataset was established using the operation data of VRF systems monitored at the outdoor units. The VRF dataset includes two types of data: The installation information of the VRF systems, e.g., the building type, climate zone, and cooling and heating capacity, where users' information has been anonymized, and the operational data, e.g., operation duration of outdoor units.

2.1.1 Installation information of VRF systems

The dataset contained 287,992 VRF systems installed in China; 93% of the samples have location (climate zone) information. There are five climate zones in China: Severe Cold region, Cold region, Hot Summer and Cold Winter zone, Hot Summer and Warm Winter zone and Mild region [34]. The climate zone is defined by average temperature of the coldest month(ATCM) and of the hottest month(ATHM), which are listed in Table 1.

Table 1 China's climate zone classification

Climate zone	**Main indices**	**Typical cities**
Severe Cold	ATCM ≤ -10 °C	Harbin, Huhehaote, Changchun, Shengyang, Urumqi
Cold	-10 ≤ ATCM ≤ 0 °C	Beijing, Tianjin, Jinan, Qingdao, Shijiazhuang, Xi'an
Hot Summer and Cold Winter	0 ≤ ATCM ≤ 10 °C 25 ≤ ATHM ≤ 30 °C	Shanghai, Chongqing, Wuhan, Nanjing, Hangzhou
Hot Summer and Warm Winter	ATCM > 10 °C 25 ≤ ATHM ≤ 30 °C	Fuzhou, Guangzhou, Shantou, Xiamen
Mild	0 ≤ ATCM ≤ 13 °C 18 ≤ ATHM ≤ 25 °C	Guiyang, Kunming, Tengchong

In addition, 12% of the samples have information about the building type (one of the nine building types). All VRF systems in the dataset are heat pump type system without heat recovery, i.e., these VRF systems can only provide either cooling or heating to a group of zones at a time, no simultaneous cooling and heating. The distributions of VRF systems across different climate zones and building types are shown in Figure 1 and Figure 2. Most VRF systems are applied in the Cold climate region, the Hot Summer and Cold Winter climate zone, and the Hot Summer and Warm Winter climate zone. The Hot Summer and Cold Winter climate zone has the most VRF applications for residential buildings mainly due to the need of cooling in summer and heating in winter. The main building types in the dataset are residential buildings (more than 50%) and office buildings.

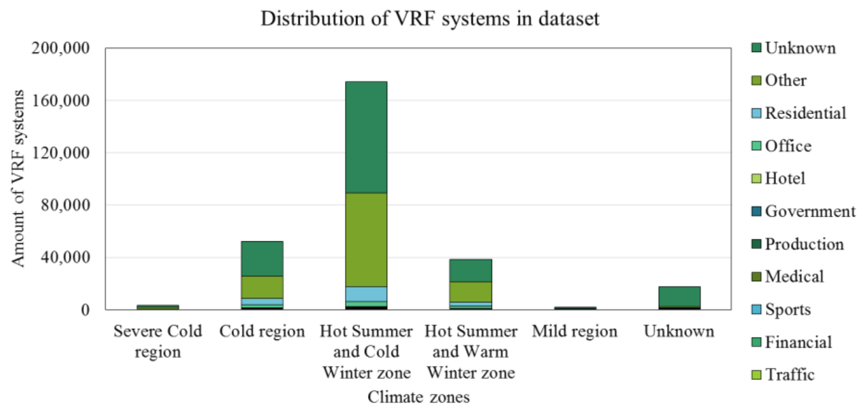


Figure 1 VRF application distribution in different climate zones

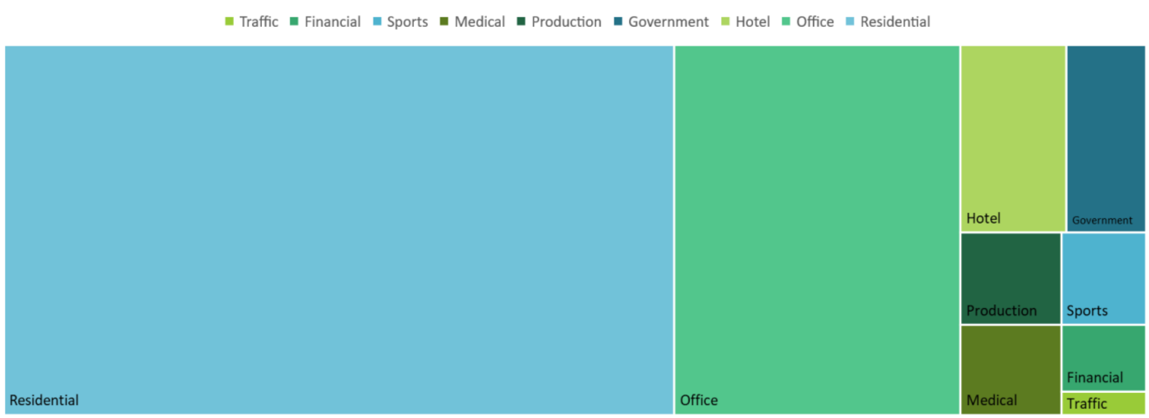


Figure 2 Distribution of VRF systems in different building types and climates

For the capacity of outdoor units in these VRF systems, there is a large difference between the VRF systems in residential buildings and those in commercial buildings. VRF capacity distribution across building type is shown in Figure 3. The capacity of VRF systems for most residential buildings is less than 20 kW (5.7 ton). The distributions of VRF capacity for commercial building types are similar, with most VRF system capacity greater than 60 kW (17 ton), except for those in government buildings.

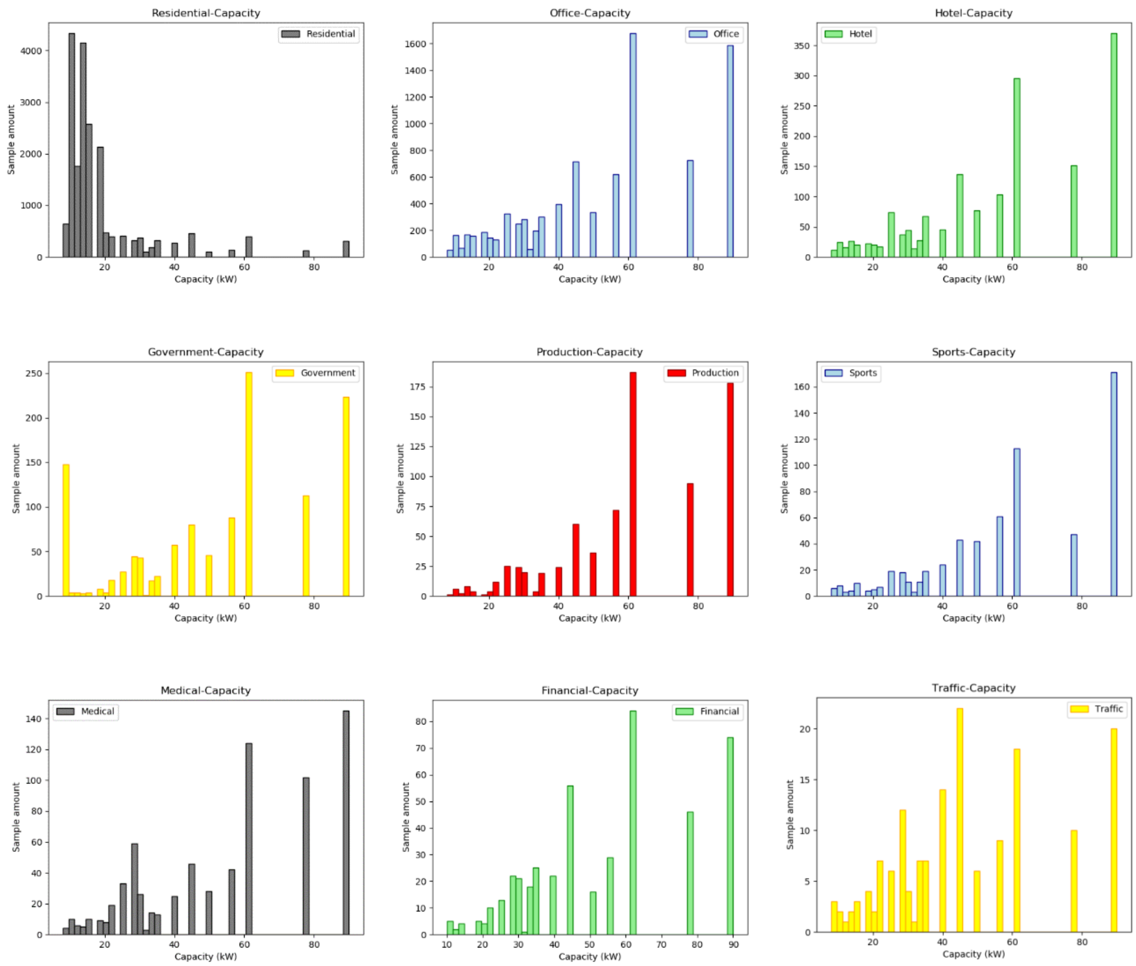


Figure 3 VRF capacity distribution for different building types

For residential and office buildings, the VRF system distributions in all climate zones are presented in Figure 4. The distributions of VRF within residential buildings are similar to each other. VRF systems are more likely to be less than 20 kW (5.7 ton) in the Severe Cold region for office buildings compared with other climate zones.

	Residential buildings	Office buildings

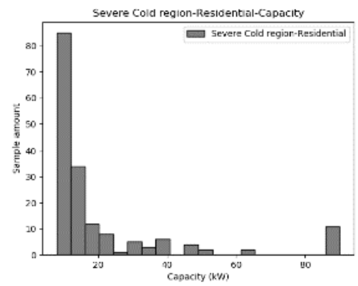
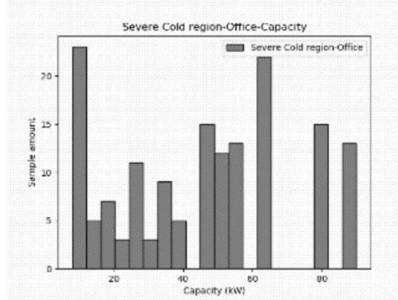
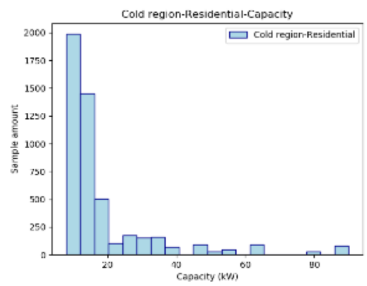
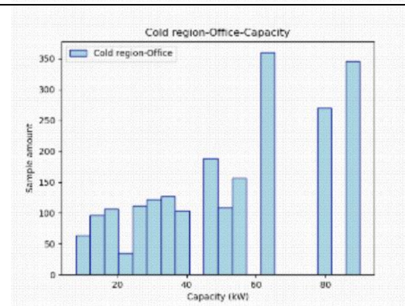
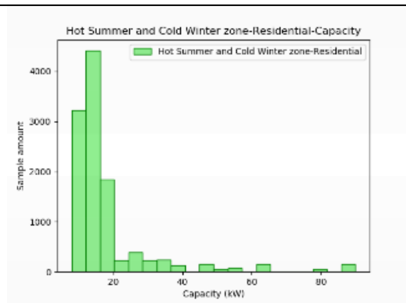
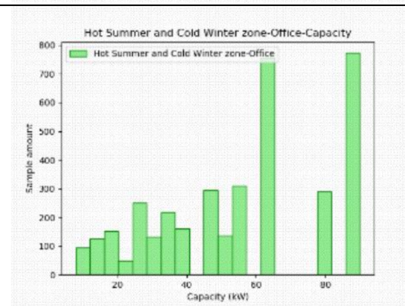
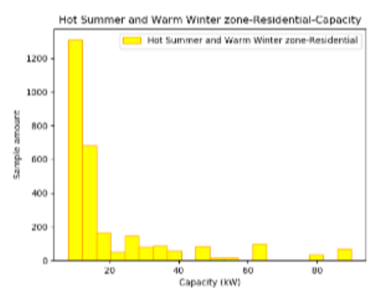
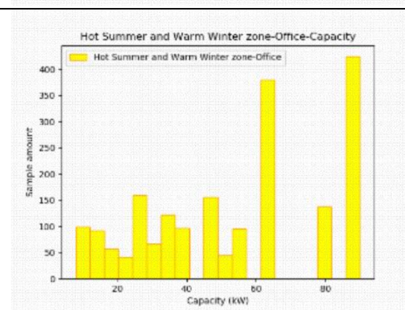
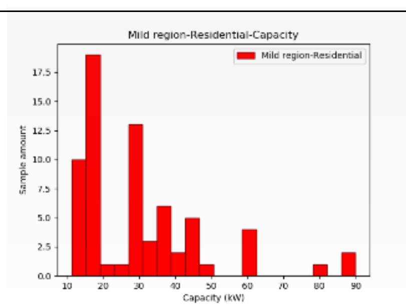
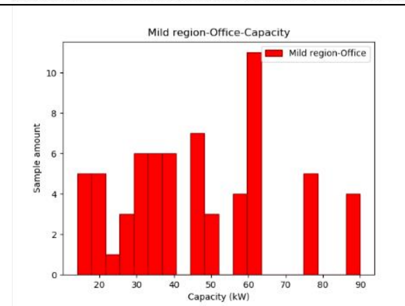
Severe Cold region		
Cold region		
Hot Summer and Cold Winter Zone		
Hot Summer and Warm Winter Zone		
Mild region		

Figure 4 VRF system capacity distribution of residential, office, hotel, and government buildings in different climate zones

2.1.2 Operational data of VRF systems

For the operational data, the dataset covered the period from December 2017 to December 2018. The operational data was measured by sub-meters installed in VRF systems, which can help insure the normal operation of VRF system. The direct measured data was stored in local outdoor unit hourly. The accumulated operating hour data under different conditions was calculated from the direct measured data, which intend to help protect user's privacy. The accumulated operating hour data was remotely logged to the cloud, which is the data source of this research(Table 2). Each temperature range and load ratio range were decided by the manufacturer. The main operation data used in this research include the evaporation temperature (T_e), condensing temperature (T_c), and load ratio (LR). T_e and T_c are directly measured at the outdoor units. Load ratio(LR) is calculated based on Indoor units on-off state and the name plate rated capacities. LR is defined as the ratio of the name plate rated capacities of the running indoor units to the name plate capacities of all installed indoor units. LR is calculated using equation (1).

$$LR = \frac{\sum capacity\ of\ running\ indoor\ units}{\sum capacity\ of\ all\ indoor\ units} \tag{1}$$

Table 2 Main operation data of VRF systems of this research

Direct measured data (Store in local VRF)	**Calculated data (Log to cloud)**
Evaporation temperature (K)	Accumulated hours under different evaporation temperature range (hour)
Condensing temperature (K)	Accumulated hours under different condensing temperature range (hour)
Indoor units on-off state	Accumulated hours under different Load ratio range (hour)

2.2 Methodology

The overall methodology is illustrated in Figure 5. First, the original accumulated use/operating hours were pre-processed into the use duration for different operating conditions of the VRF systems. Second, to obtain the real use status and evaluate the performance of the VRF system, three key performance indicators (KPIs) —use duration (representing operating time in a cycle), ideal coefficient of performance (ICOP, representing the theoretical efficiency), and load ratio pattern (representing part load operating conditions) — were proposed. Statistical and clustering analyses were conducted to determine the distribution of three KPIs in different building types and climate zones. Finally, a recommendation on VRF system design was proposed based on the data analysis.

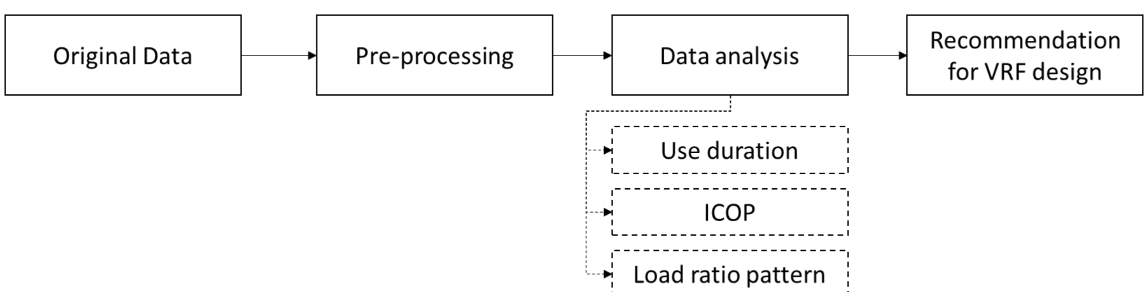


Figure 5 Methodology roadmap

2.2.1 Preprocessing the collected VRF dataset

The data preprocessing contained data formulation and data cleaning. The calculated data logged to the cloud of VRF dataset is the accumulated operating hours of VRF systems under different operation conditions for 3–4 months. The data was processed to determine the use duration at different operation conditions. Example data is presented in Table 3. Then, data cleaning was conducted, such as removing erroneous or missing data. For example, if the calculated use duration is negative or larger than the record time, the data sample will be removed from the dataset.

Table 3 Use duration sample data

	LR = (0,25%) use duration (h)	LR = [25,50%) use duration (h)	LR = [50,75%) use duration (h)	LR = [75,100%) use duration (h)	LR = 100% use duration (h)	Record time (h)
Dec. 2017–Feb. 2018	286	90	75	51	39	1728
Feb. 2018–Apr. 2018	60	1	0	0	0	1416
Apr. 2018–Jul. 2018	754	71	101	62	64	2496
Jul. 2018–Sept. 2018	328	86	67	63	40	888
Sept. 2018–Dec. 2018	0	0	0	0	0	1896

2.2.2 Key performance indicators of VRF systems

In order to evaluate the performance of VRF systems, three key performance indicators (KPIs), namely, the use duration, ICOP, and the load-ratio, were proposed.

Use duration is a key indicator of how occupants operate/use the VRF systems. Use duration is defined as the running/operating time of VRF systems, which is determined by the users of VRF systems. It was calculated by summing the running time at different load ratio conditions, as shown in Table 3. The statistical analysis of the use duration of VRF was conducted for different climate zones and building types. The accumulated proportion of VRF systems with various use durations was calculated.

ICOP was proposed to evaluate the ideal performance of VRF systems in different climate zones and building types. Different cities have totally different heating and cooling periods. In order to compare different cases in different periods, ICOP here was uniformly defined as the evaporation efficiency. $ICOP_{T_c,T_e}$ for different evaporation and condensation temperature was calculated as shown in equation (2).

$$ICOP_{T_c,T_e} = \frac{T_e}{T_c - T_e} \qquad (2)$$

where T_e and T_c are the evaporation and condensation temperatures (K), respectively.

The original data of evaporation and condensation temperatures determine the use duration for operating conditions within a temperature range. Each temperature range was decided by the manufacturer. We used the average temperature of each temperature range to calculate ICOP. The ICOP for each temperature range is shown in Table 4. Finally, the ICOP for each season was calculated using the proportion of time for each operating condition in equation (3):

$$ICOP = \frac{\sum ICOP_{T_c,T_e} \times Time_{T_c,T_e}}{\sum Time_{T_c,T_e}} \tag{3}$$

where $Time_{T_c,T_e}$ is the use duration for the operation conditions with T_c and T_e (h) and $\sum Time_{T_c,T_e}$ is the use duration of the VRF systems for each period.

Table 4 ICOP calculation results

$ICOP_{T_c,T_e}$	T_c < 283 K	T_c = [283,303) K	T_c = [303,328] K	T_c > 328 K
T_e < 243 K	6.1	4.9	3.4	2.9
T_e = [243,263) K	8.4	6.3	4.1	3.4
T_e = [263,273) K	17.9	10.7	5.6	4.5
T_e = [273,280] K	42.6	16.8	7.1	5.4
T_e > 280 K	93.4	21.6	7.9	5.8

To determine the operating load of VRF systems, the use duration data was analyzed for the load ratio, as shown in Table 3. This study assumed December 2017 to February 2018 as the heating season and July 2018 to September 2018 as the cooling season. Clustering analysis is common unsupervised method to conduct load profiling [19]. This research used accumulated hours under different load ratio for heating/cooling as the input data for clustering analysis(Figure 6).

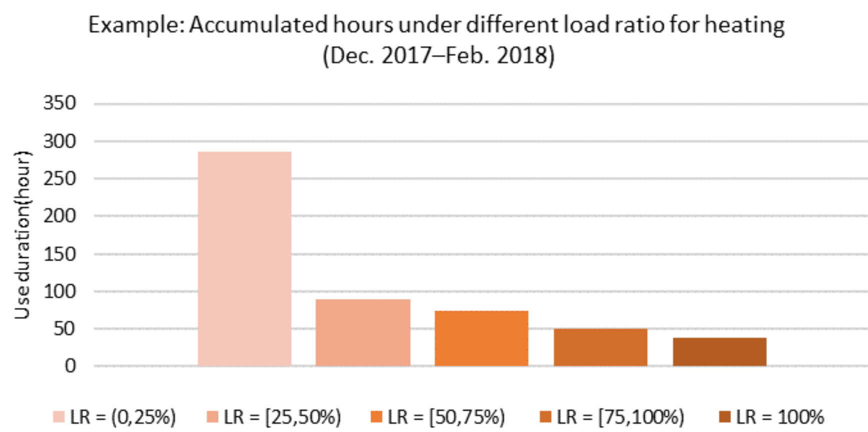


Figure 6 Clustering input data example: Accumulated hours under different load ratio for heating

Clustering analysis of the VRF load-ratio pattern for heating and cooling was conducted by the K-means algorithm, which is a common method for curve clustering [35]. Lavin and Klabjan [36, 37] and Green et al.[37] found that using the core K-means clustering algorithm for load profiling is efficient. Euclidean distances were used to conduct the clustering, as shown in equation (4).

$$d(x,y) = \left(\sum_{i=1}^{5} |x_i - y_i|^p\right)^{\frac{1}{p}}, \quad (4)$$

where $x = (x_1,x_2...,x_5)$, $y = (y_1,y_2...,y_5)$ are accumulated hours under different load ratio, two objects in a Euclidean n-space and p = 2.

In order to get typical load profile, on one hand, the research used the Davies–Bouldin index (DBI), a metric that is used for evaluating clustering algorithms [38]. A lower DBI represented better clustering, which is calculated by Equation (5), and can help to determine the number of clusters. Through the calculation, we found best cluster number for different climate zones are different, ranging from 4 to 6. On the other hand, this research intended to get typical load pattern, which may need accounted for at least 10% of samples and to compare different climate zones' results. So the cluster number was set as 5 for each climate zones.

$$DBI = \frac{1}{N}\sum_{i=1}^{N} \max_{j \neq i} \left(\frac{S_i + S_j}{M_{i,j}} \right) \tag{5}$$

where S_i is a measure of scattering within the cluster; $M_{i,j}$ is a measure of separation between cluster i and cluster j; and N is the number of clusters.

3 Analysis of key performance indicators of VRF systems

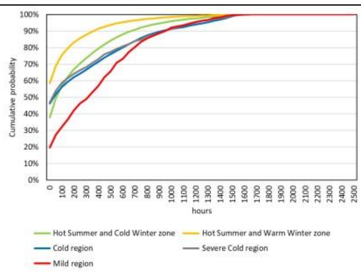
3.1 Use duration of VRF systems

3.1.1 Use duration of VRF systems in different climate zones

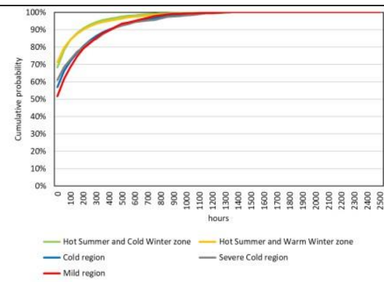
The cumulative probability of use duration in different climate zones are calculated based on the dataset. Figure 7 provides the cumulative probability curves of use duration in different climate zones for the same each time period in 2018. From December 2017 to February 2018, the use duration of 90% of the systems among all climate zones was less than 950 h (64% of the whole time, 1488 h). The use duration in the Hot Summer and Warm Winter zone was shorter than that of other climate zones, and more than 58% of systems were not used during these three months. The cumulative probability curves of the Severe Cold region are almost the same as those of the Cold region. From February 2018 to April 2018, the use duration of 90% of the systems in all climate zones was less than 450 h (32% of the whole time, 1416 h). The cumulative probability curves of the Hot Summer and Cold Winter zone and Hot Summer and Warm Winter zone were similar, with use durations shorter than those of the other three climate zones. In addition, more than 68% of the systems in these two zones did not operate during this period. The curves of the other three zones were similar in this period. From April to July 2018, the use duration of 90% of the systems in all climate zones was less than 1250 h (57% of the whole time, 2184 h). From July to September 2018, the use duration of 90% systems in all climate zones was less than 800 h (55%

of the whole time, 1464 h). Both from April to July and from July to September in 2018, the use duration of the Hot Summer and Warm Winter zone is longer than other climate zones', which may result from the higher outdoor temperature in this zone. From September 2018 to December 2018, the use duration of 90% of the systems in all climate zones was less than 800 h (36% of the whole time, 2208 h). The use duration of Hot Summer and Cold Winter zone was less than that of other zones, and more than 50% of the systems did not operate in this period.

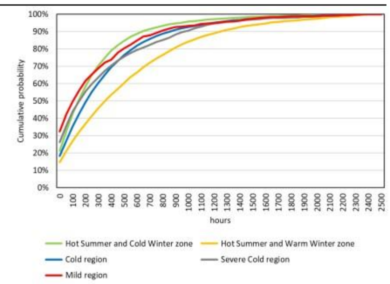
Dec. 2017–Feb. 2018(1488 hours)



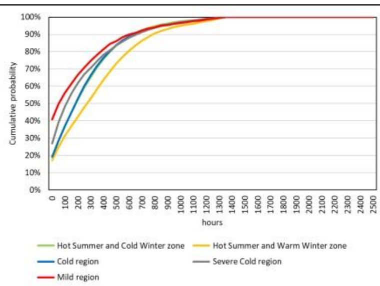
Feb. 2018–Apr. 2018(1416 hours)



Apr. 2018–Jul. 2018(2184 hours)



Jul. 2018–Sept. 2018(1464 hours)



Sept. 2018–Dec. 2018(2208 hours)

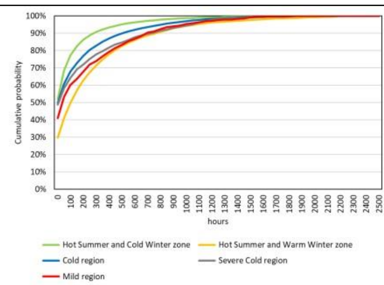
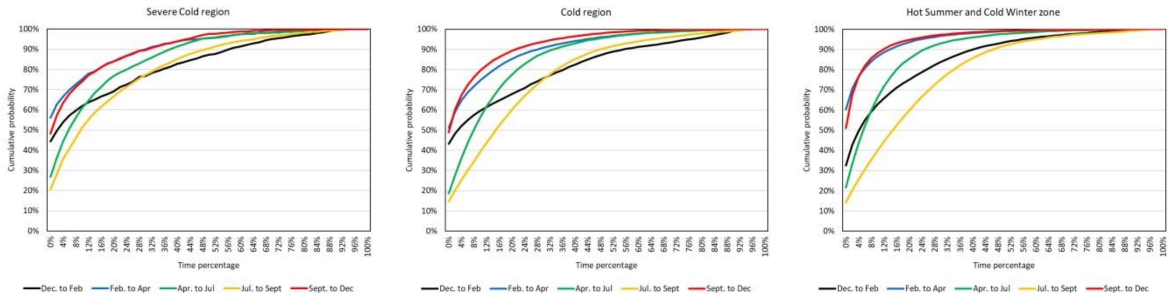


Figure 7 Cumulative probability curve of use duration of different climate zones in same time period

Figure 8 presents the cumulative probability curves of the use duration in the same climate zone for different time periods in 2018. The whole time for different time period is shown in Figure 7. Except for the Mild region, all the climate zones had longer use durations from July to Sept in 2018 than in other time periods, which may reveal that the VRF systems were used mainly for cooling in this time period for these climates. In the Mild region, the VRF use duration is longest from December to February, which reflected that heating was the main function there. The use durations of transition seasons (February–April and September–December) were shorter than those of other time periods in 2018, except for September–December in the Hot Summer and Warm Winter zone.

Severe Cold region	Cold region	Hot Summer and Cold Winter Zone



Hot Summer and Warm Winter Zone

Mild region

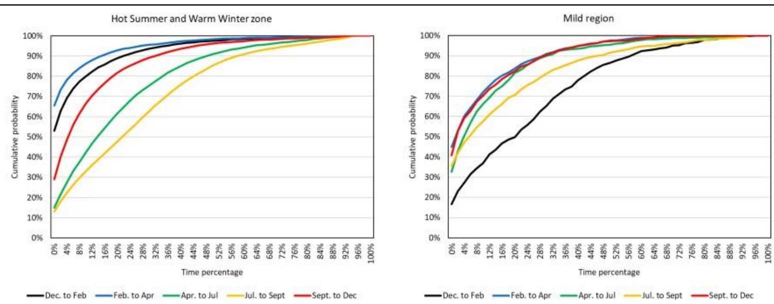


Figure 8 Cumulative probability curve of use duration in different time periods of the same climate zone

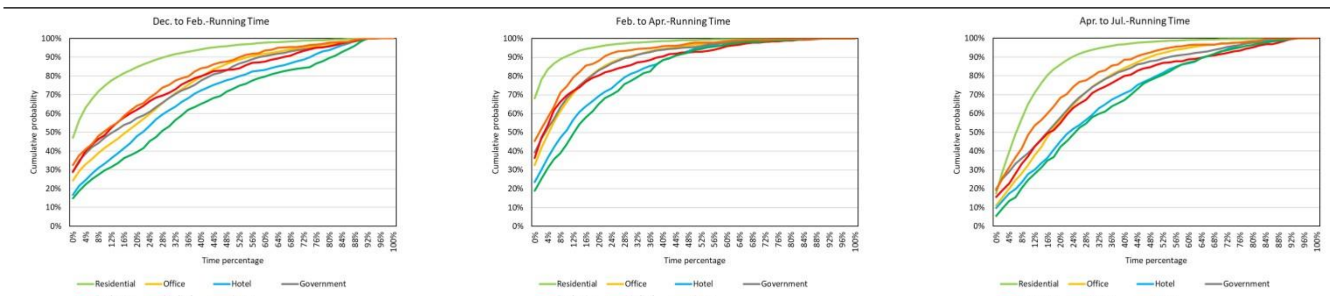
3.1.2 Use duration of VRF systems in different types of buildings

The cumulative probability of use duration in different types of buildings was generated based on the dataset. Figure 9 presents the cumulative probability curve for use duration in different types of buildings in the same time period. It illustrates that the use duration of residential buildings was shorter than that in other climate zones for the whole period. The use duration of 90% systems in residential buildings was less than 650 h in each period. The use duration of Hotel and medical were longer than most of others in each period. The use duration of 90% of the systems in hotel and medical buildings was less than 1700 h in each period.

Dec. 2017–Feb. 2018(1488 hours)

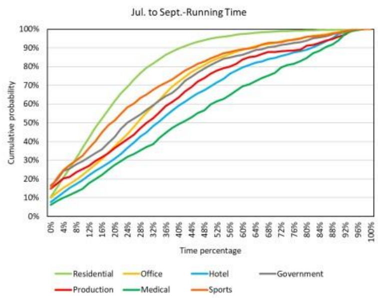
Feb. 2018–Apr. 2018(1416 hours)

Apr. 2018–Jul. 2018(2184 hours)



Jul. 2018–Sept. 2018(1464 hours)

Sept. 2018–Dec. 2018(2208 hours)



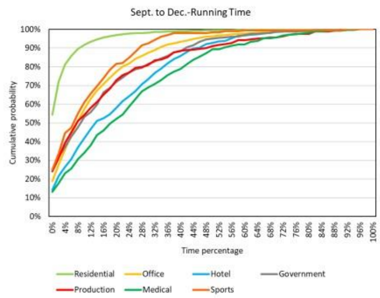
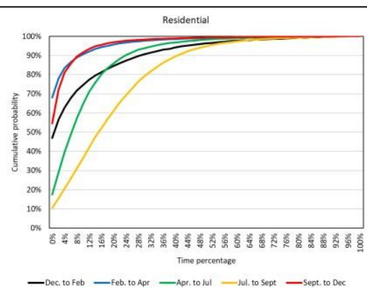
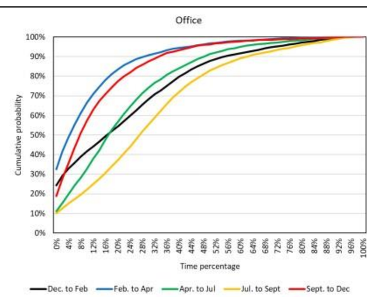
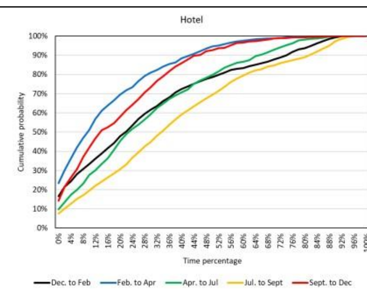
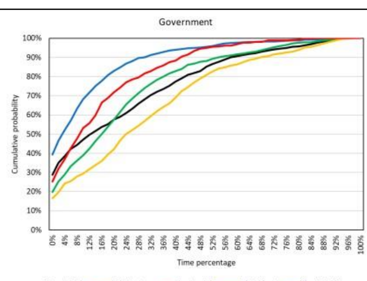
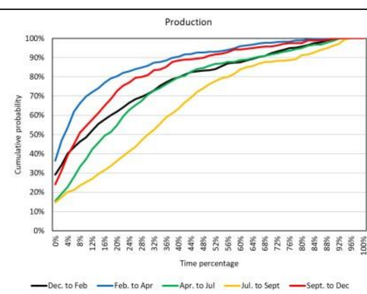
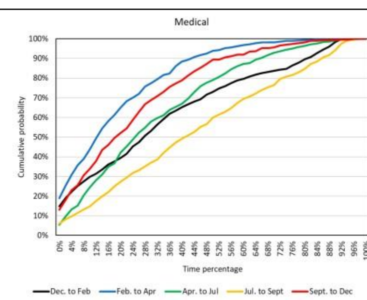


Figure 9 Cumulative probability Curve of use duration in different types buildings in same time period

Figure 10 provides the cumulative probability curve of use duration in different time periods for the same building type. The whole time for different time period is shown in Figure 9. For all types of buildings, the use duration from July to September is longer than that of other periods. This revealed that the VRF systems were mainly used for cooling for all types of buildings. The use durations of transition seasons (February–April and September–December) were shorter than those of other time periods in 2018 for all types of buildings. For residential buildings, 47% of the systems did not operate from December to February. For office buildings, 25% of the systems did were not operated from December to February. No more than 25% of the systems in hotels and 20% of the systems in medical buildings were not operated in each period, which may reveal that the usage in hotel and medical buildings is high in all seasons.

Residential	Office	Hotel
		
Government	Production	Medical
		
Sports		

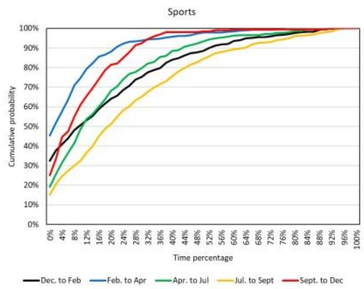


Figure 10 Cumulative probability curve of use duration in different time periods for the same building type

3.2 ICOP of VRF systems

3.2.1 ICOP of VRF systems in different climate zones

The sample probability distributions of ICOP in different climate zones were generated from the dataset. Figure 11 compares the ICOP of different climate zones in same time period. The distributions from April to July and from July to September are similar. For other time period, shown in Figure 12, ICOP with peak probability of Severe Cold region was lower than other climates, especially from September to February, which resulted from the lower outdoor air temperature. This may require more attention from designers.

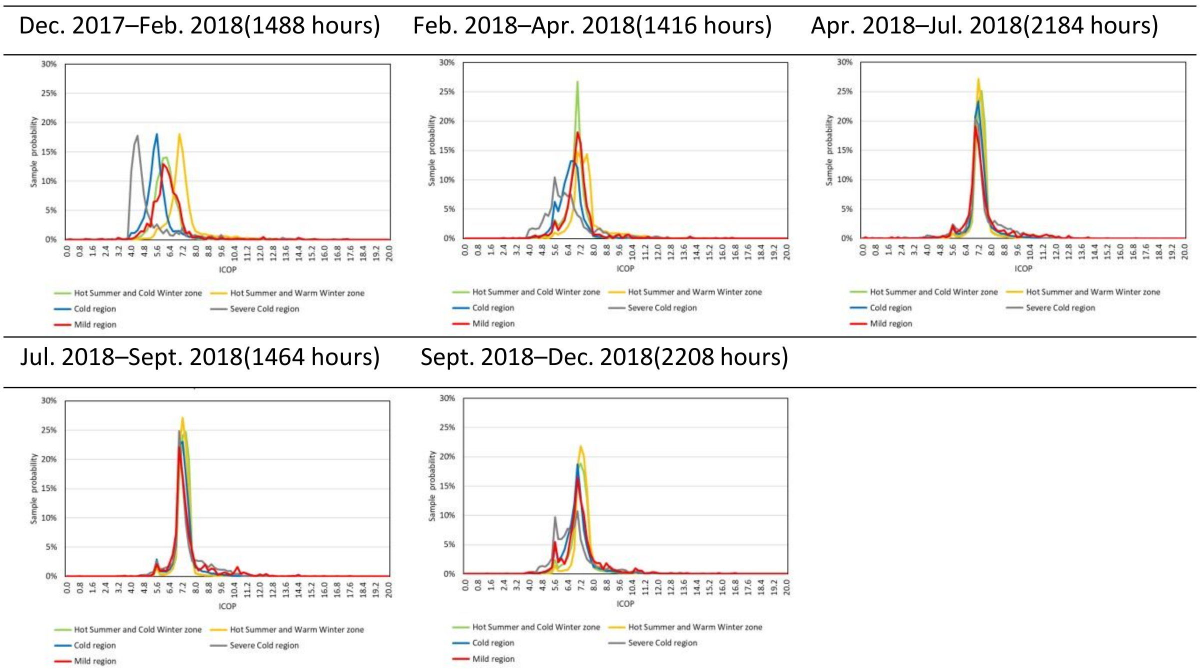


Figure 11 Sample probability distribution of ICOP of different climate zones in the same time period

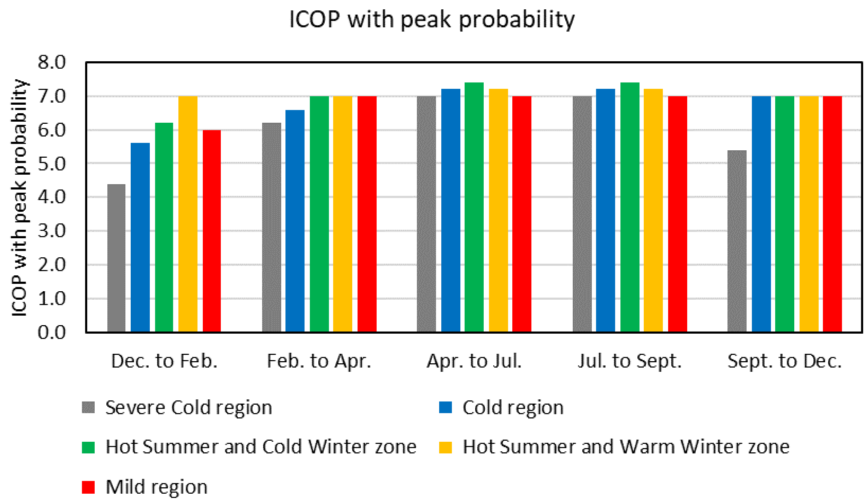
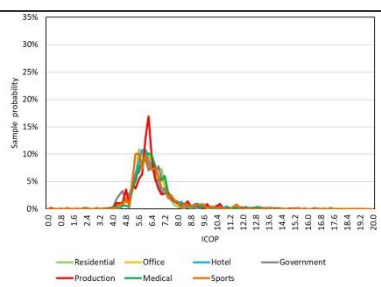
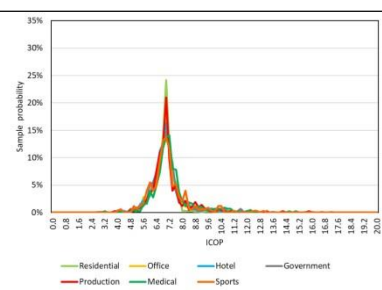
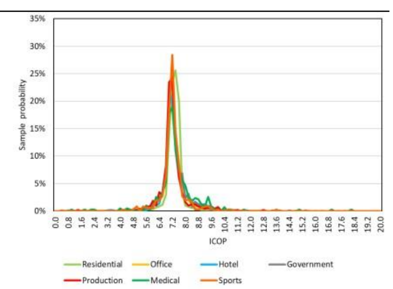


Figure 12 ICOP with peak probability for different climate zones

For the ICOP in different time periods of the same climate zone(see Appendix A for full distributions), except for the Hot summer and Warm Winter zone, ICOP with peak probability from December to February is lower than other time period shown in Figure 12, which resulted from the low outdoor temperature during this period.

3.2.2 ICOP of VRF systems in different types of buildings

The sample probability distributions of ICOP in different building types were calculated from the dataset. Figure 13 illustrates that the distributions in the same period for different types of buildings are similar. The distributions from April to July and from July to September are more concentrated than those of other periods. This reflected that the outdoor weather is more stable for using VRF from April to September. ICOP distributions for different building types in different period are similar (see Appendix A for full distributions). It reflected that outdoor weather is the main influencing factor rather than building types.

Dec. 2017–Feb. 2018(1488 hours)	Feb. 2018–Apr. 2018(1416 hours)	Apr. 2018–Jul. 2018(2184 hours)
		
Jul. 2018–Sept. 2018(1464 hours)	Sept. 2018–Dec. 2018(2208 hours)	

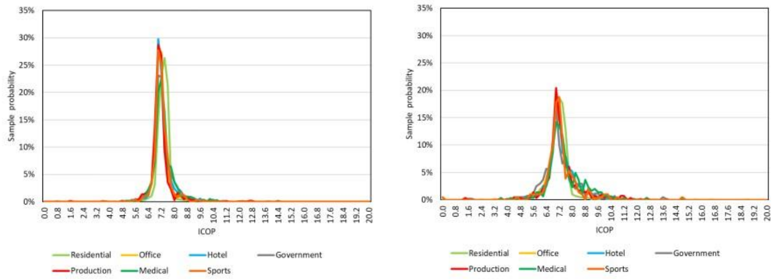
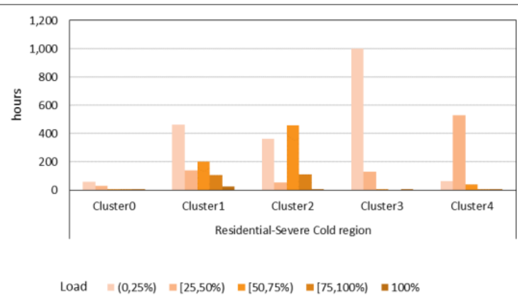
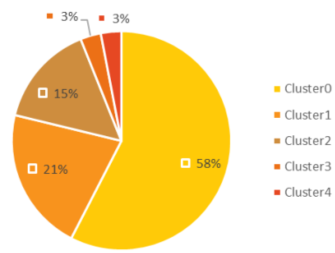
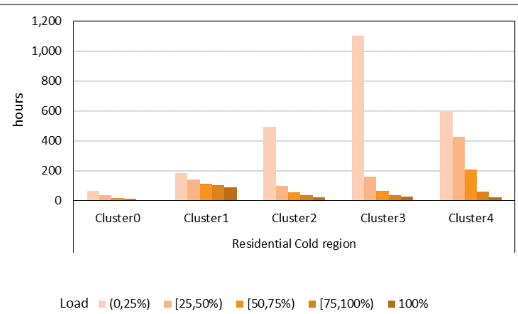
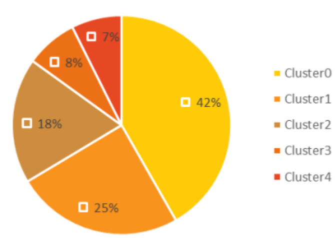


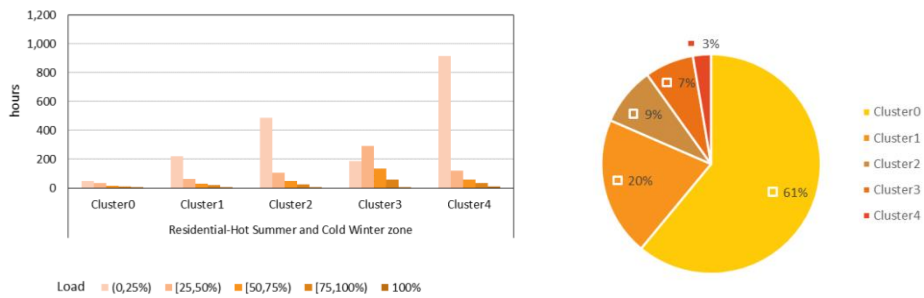
Figure 13 Sample probability distribution of ICOP of different building types in same period

3.3 Typical load-ratio pattern for residential and office buildings

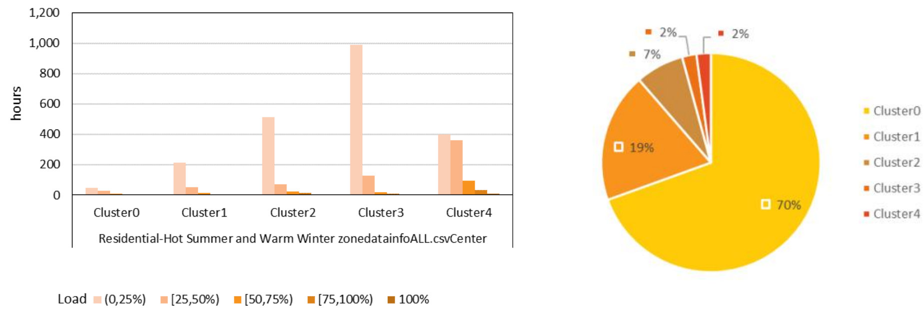
Through clustering, the typical load-ratio patterns in all five climate zones for residential buildings were generated from the large-scale dataset. Figure 14 presents the cluster centers of a typical load pattern of VRF systems of residential buildings for heating. For all the climate zones, one can conclude that the main operating condition of VRF systems was at less than 25% load ratio, which reflected that households used VRF for heating with only 1–2 indoor units most of the time. Among all climate zones, the VRF system average operating hours for heating of typical load-ratio patterns of Hot Summer and Warm Winter were shorter than those in other climates.

	Cluster centers	Numbers of each cluster
Severe Cold region		
Cold region		

Hot Summer and Cold Winter Zone



Hot Summer and Warm Winter Zone



Mild region

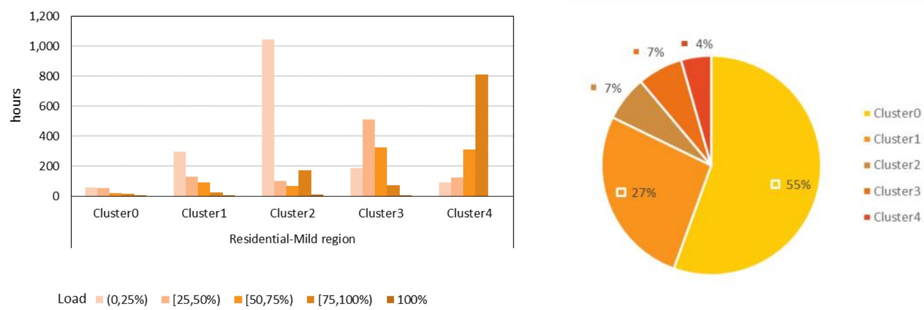
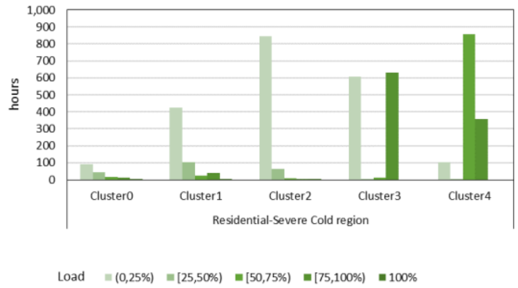


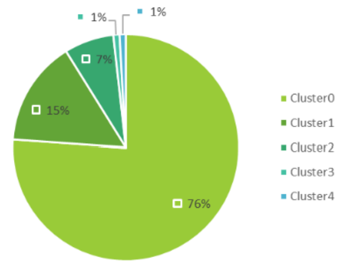
Figure 14 Clustering results of residential buildings for heating in different climate zones

Figure 15 shows the typical load pattern of VRF systems of residential buildings for cooling. For all the climate zones, similar to VRF for heating, it can be concluded that VRF systems for cooling mainly operate under lower than 25% load ratio, which reflected that households used VRF with only 1–2 indoor units most of the time. Compared with VRF systems for heating, there were more typical load-ratio patterns (accounted for more than 10% of all systems) in VRF cooling conditions, which may reveal more typical-operation cooling modes than heating modes for VRF systems in residential buildings.

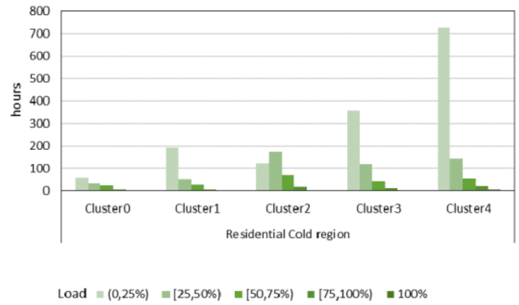
Cluster centers	Numbers of each cluster

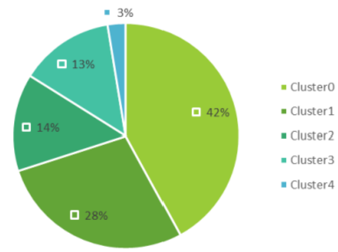
Severe Cold region



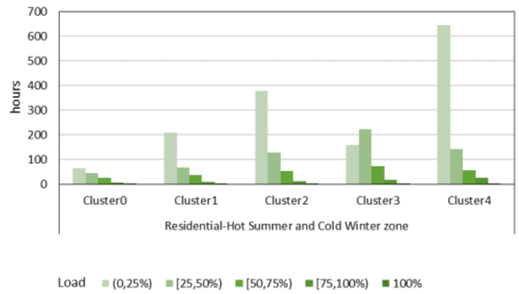


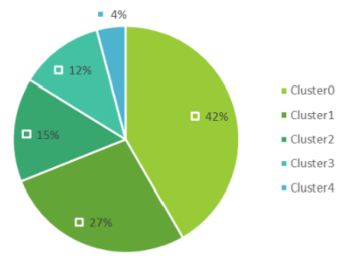
Cold region



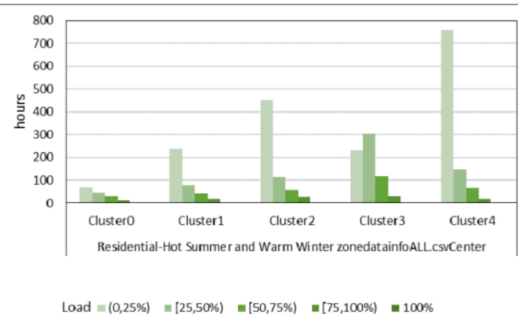


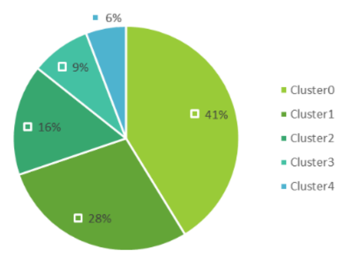
Hot Summer and Cold Winter Zone



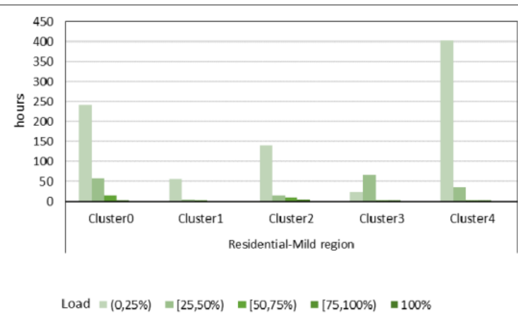


Hot Summer and Warm Winter Zone





Mild region



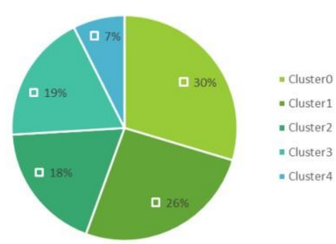


Figure 15 Clustering results of residential buildings for cooling in different climate zones

Through clustering, the typical load-ratio pattern in all five climate zones for office buildings were generated from large scale dataset. For office buildings in all the climate zones, it can be observed that VRF systems were mainly operating at a load ratio of less than 25% in heating mode, which may reflect that users in offices used VRF for heating with only one or two indoor units for most time (see Appendix B for full figure and description of typical load pattern of VRF systems of office buildings for heating).

Similar with VRF systems in residential buildings, among all climate zones, the office VRF system average operating hours for heating at typical load-ratio patterns in Hot Summer and Warm Winter were shorter than in other climates. For Cluster0 in the Mild region, the average operating hours were longer than those of Cluster0 in other climate zones were.

For VRF cooling in office buildings, it can be observed that, as with VRFs for heating, the main operating conditions of VRF systems for cooling were at a lower than 25% load ratio (see Appendix B for full figure and description of typical load pattern of VRF systems of office buildings for cooling).

For all the typical load patterns of residential and office buildings, a really lower load is common. This is because of the part-time-part-space operation mode of users/occupants who tended to activate only one or two indoor units most of the time. This further confirmed same finding from previous study on AC use measured in residential buildings [32] and office buildings [10].

4 Potential applications of outcomes

Based on the data-mining results of the large VRF dataset, a few potential applications and recommendations can be drawn for VRF systems to improve their design and operation in China.

4.1 Special design for VRF in specific climate zones and buildings

For traditional design, the cooling and heating demands for specific climate zones with different schedules could be calculated for residential buildings and commercial buildings for a typical day. Based on the cooling and heating demand of buildings, designers would select VRF systems according to the peak load for a typical day. However, according to the data-mining results of this research, some interesting phenomenon were observed for different seasons in one year for a specific climate zone and building type. From the use duration analysis, the use duration of cooling by VRF systems is longer in the whole year except for the Mild region. It is recommended to focus on cooling performance in the design stage. For the Mild region, data analysis reflected that heating was the main function. The results recommended to focus on heating performance in the design stage for the Mild region. For specific building types, the usage of VRF in hotel and medical buildings was high among all building types year-round, even during transition seasons (February–April and September–December). Thus, the design of VRF systems for hotel and medical buildings must consider not only typical cooling and heating days, but also those for transition seasons. From the ICOP analysis of VRF systems, it was revealed that the ICOP with peak probability in the Severe Cold region from December to February was lower than that of the other climate zones because of the low outdoor air temperature. This informs the optimization of VRF systems, which is to increase the heating performance of VRFs for cold climates with low outdoor air temperature. More statistical analysis of typical outdoor temperature in the Severe Cold region based on the dataset should be conducted for further research.

4.2 VRF performance evaluation standard improvement

Manufacturers conduct performance tests on sample VRF systems before selling them. The current evaluation standard in China for VRF performance is GB/T18837-2015 [39]. The evaluation method of the standard is to test the annual performance factor (APF) of the VRF system. The APF is calculated according to equation (6):

$$\mathrm{APF} = \frac{CSTL + HSTL}{CSTE + HSTE}, \quad (6)$$

where CSTL is the total load of the cooling season (Wh), HSTL is the total load of the heating season (Wh), CSTE is the total electricity consumption of the cooling season (Wh), and HSTE is the total electricity consumption of the heating season (Wh).

The total load and electricity of the cooling and heating seasons were determined by test data from three cooling conditions (nominal load, medium load, and minimum load) and four heating conditions (maximum load, nominal load, medium load, and minimum load) combined with use duration from another China standard GB/T 17758-2010 [40]. The standard provided the use duration for different cities and outdoor temperatures, rather than that for different load conditions. The typical load-ratio pattern analysis in this research can help provide use duration for different load conditions. In addition, the typical load-ratio pattern analysis in this study also provided the use duration for residential buildings and office buildings. The typical use duration with different load ratios, based on real monitoring data, can improve the evaluation of VRF performance to reflect the actual operating conditions.

4.3 Performance optimization for VRF systems

From the typical load-ratio pattern analysis, it revealed that long time low part load ratio operation is common for VRF systems in residential buildings and commercial buildings. This triggers two potential optimization recommendations for VRF systems. First, for some VRF systems, the frequency of full load is near zero. It is recommended to properly reduce VRF outdoor unit capacity, which can decrease initial investment of VRF systems and raise the operating load ratio to improve efficiency to save energy and increase the satisfaction level of users. Second, for most VRF systems, there is high frequency of both partial and full loads. In this case, a new optimization of VRF system design is to use multiple compressors: one highly efficient compressor meeting the base loads and other compressors with variable speed controls to improve the partial load efficiency. The typical load patterns proposed in this paper can guide the selection of different-capacity compressors in one VRF system for different types of buildings across each climate zone.

5 Discussion

5.1 Policy implications

VRF systems are applied widely in China, mainly for residential and office buildings. VRF systems have high control flexibility and can meet the comfort needs of individual spaces, which is supported by data in this research. Data analysis in this research revealed the high diversity of occupant behavior and the part-time-part-space use mode is the main operation mode of VRF systems in China. In addition, data-mining results indicated that partial load issue is common. All the phenomenon found through data mining may raise the attention of manufacturers and government and inform their policy making.

Manufacturers may need to use data-mining results to optimize the VRF design and operation controls. It is better for the government to use real installation and operation data of VRF systems for related industry VRF standard development, especially for performance evaluation and rating of VRF systems. Furthermore, government may value the VRF system performance in special building types such as hotel and medical buildings because of their high use duration. Government can also take more effort to promote energy conservation (e.g., not to oversize the VRF system) based on data mining outcome of VRF systems.

5.2 Limitations of the current work

This research is a preliminary attempt of big data application in the HVAC field in China. The dataset contained 287,000 VRF systems in China. However, data type in this research is use duration data and data range of each parameter is defined by the VRF manufacturer. There are no time series data, making it hard to get deeper knowledge from the operation data. Besides, there is only one data source for the research, which made it hard to conduct the validation work. If there is other data source with a large number of VRF samples in the future, the validation work can be done. In addition, the information of building type is missing for most installations, making it impossible to analyze data in detail for different building types. With the development of sub-metering and data science, time series dataset with more system information will be available in the premise of guaranteeing users' privacy. It will be valuable to inform HVAC design and operation, which can help achieve energy efficiency of VRF system thus reducing energy use, utility costs, and GHG emissions for the building sector in China.

6 Conclusion

This study performed statistical and clustering analysis of a large number of VRF systems for residential and commercial buildings in China across all five climate zones. Distributions of use duration and ideal performance of VRF systems across different climate zones and building types were derived from the actual VRF operation data. Represented load ratio patterns of VRF systems of residential and office buildings across all climate zones of China were developed, which can be used to inform VRF design, sizing and performance optimization. The main findings are as follows:

a. VRF systems are mainly used for cooling in all climate zones of China, except in the Mild climate region, where space heating is also a main function along with cooling.
b. There is a great diversity of VRF use duration. Among all the building types, the use duration of residential buildings is the lowest, whereas that of hotels and medical buildings is the highest.
c. The distribution of ideal VRF cooling COP is similar across all climate zones and building types, whereas the COPs of ideal VRF heating in the Severe Cold region and the Cold region are lower than those in other climate zones due to the low outdoor temperature conditions in the two cold climate zones.
d. Partial load operations for VRF systems are common in residential buildings and office buildings. This is because of the part-time–part-space operation mode with users/occupants usually activating only one or two indoor units most of the time. This agrees with findings from previous studies on AC use in residential buildings and office buildings in China.
e. Data mining on large-scale operation dataset can reveal the real application status of VRF systems in China. The data-mining results provide knowledge supporting the design, operation, industry standard development, and performance optimization of VRF systems.

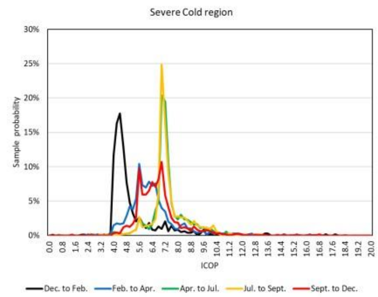
Based on the data-mining results of the large VRF dataset, recommendations for designers and manufacturers are as followed:

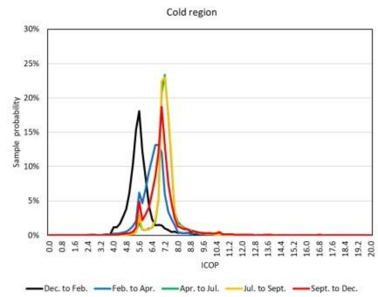
a. In design stage, as the main function of VRF is cooling except in Mild region, it is recommended to focus on cooling performance.
b. For special building types, such as hotel and medical buildings, the design of VRF systems must consider not only typical cooling and heating days, but also those for transition seasons.
c. For manufacturers, the optimization direction of VRF systems is to increase the heating performance of VRFs for cold climates with low outdoor air temperature.
d. As partial load operations for VRF systems are common in residential buildings and office buildings, it is recommended to properly reduce VRF outdoor unit capacity, which can decrease initial investment of VRF systems and raise the operating load ratio to improve efficiency to save energy and increase the satisfaction level of users.
e. Based on the typical load pattern, for most VRF systems, there is high frequency of both partial and full loads. A new optimization of VRF system design is to use multiple compressors to improve the partial load efficiency.

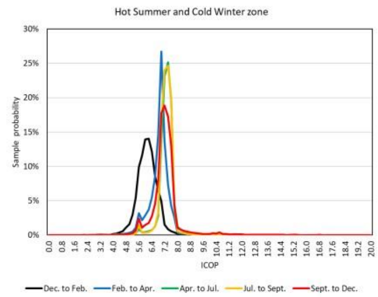
Appendix A

Figure 16**Error! Reference source not found.** presented the sample probability distribution of ICOP in different time periods of the same climate zone.

Severe Cold region	Cold region	Hot Summer and Cold Winter Zone

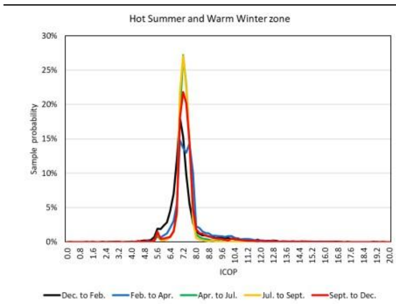






Hot Summer and Warm Winter Zone

Mild region



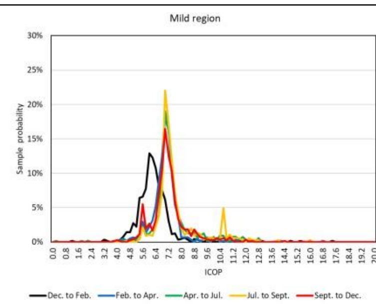


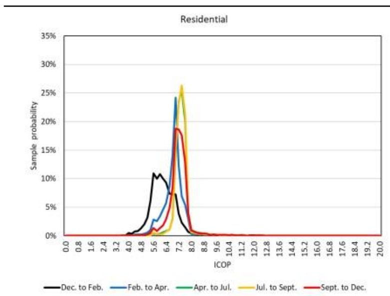
Figure 16 Sample probability distribution of ICOP in different periods of the same climate zone

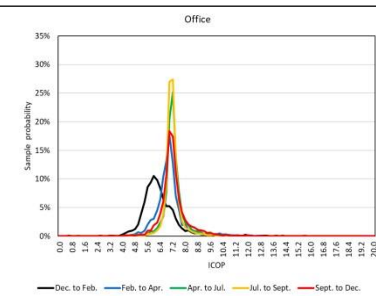
Figure 17**Error! Reference source not found.** shows that ICOP distributions for different building types in different period are similar. It reflected that outdoor weather is the main influencing factor rather than building types.

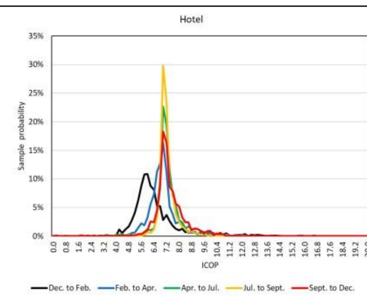
Residential

Office

Hotel



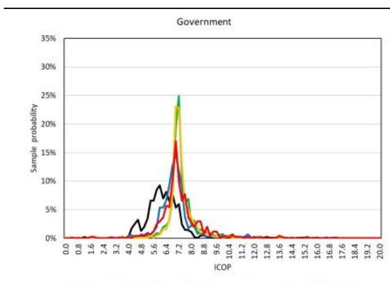


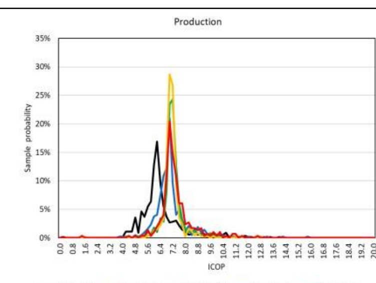


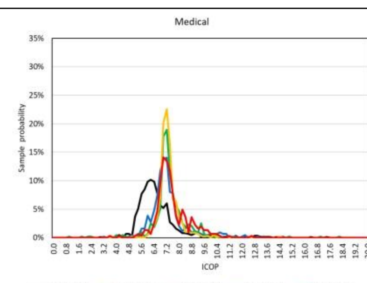
Government

Production

Medical







Sports

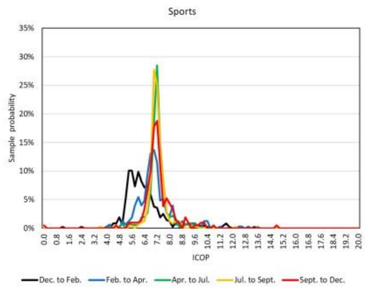
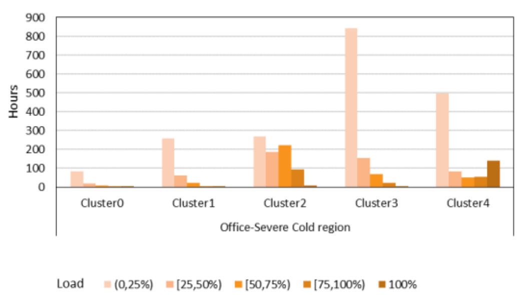
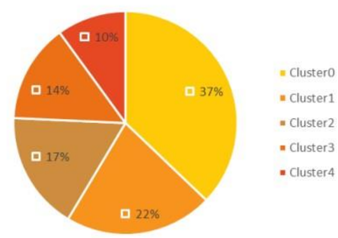
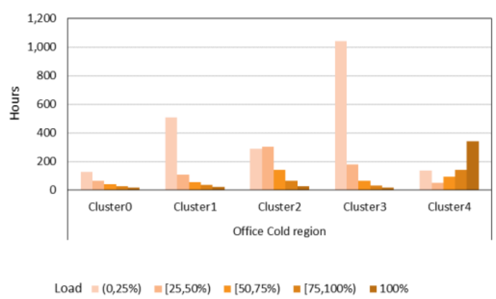
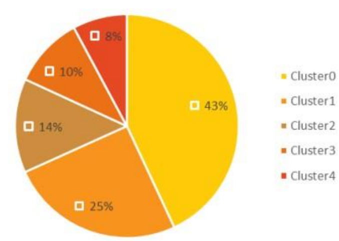
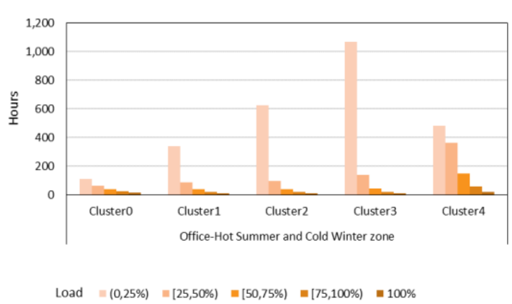
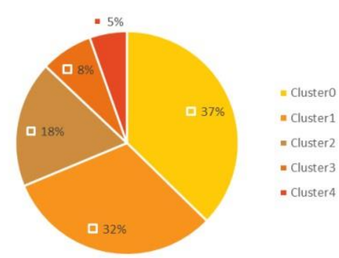


Figure 17 Sample probability distribution of ICOP in different periods of the same building types

Appendix B

Figure 18 provided typical load pattern of VRF systems of office buildings for heating.

	Cluster centers	Numbers of each cluster
Severe Cold region		
Cold region		
Hot Summer and Cold Winter Zone		

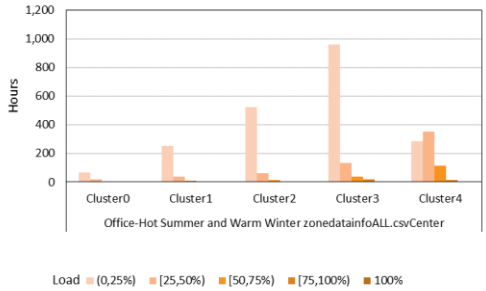
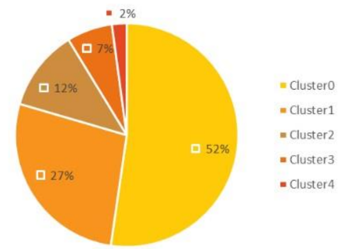
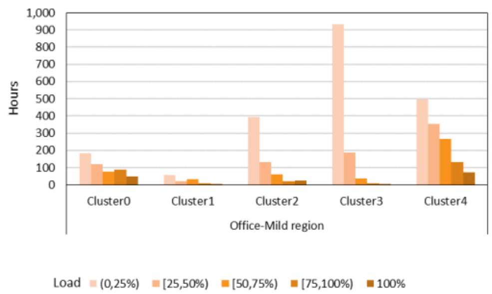
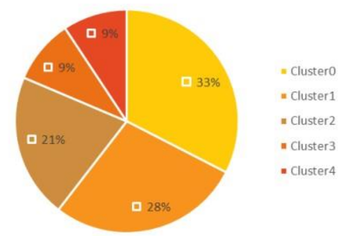
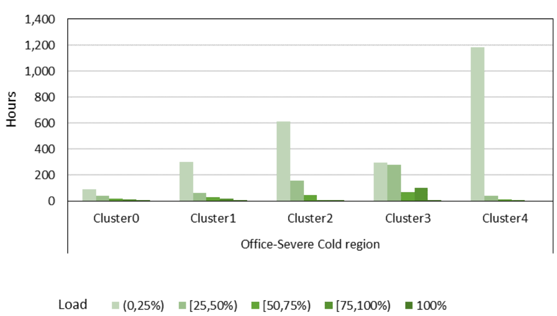
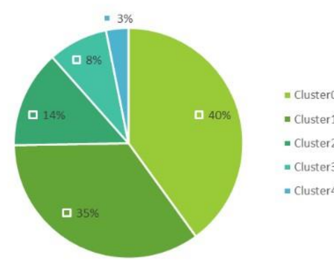
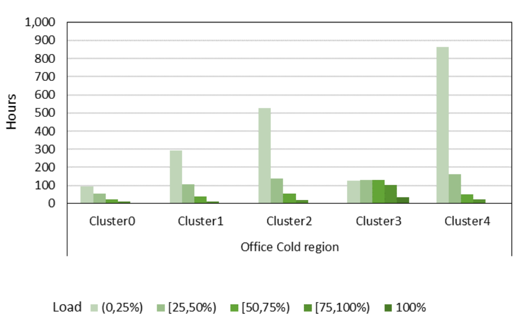
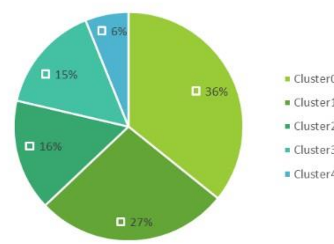
Hot Summer and Warm Winter Zone		
Mild region		

Figure 18 Clustering results of office buildings for heating in different climate zones

Figure 19**Error! Reference source not found.** shows the typical load pattern of VRF systems in office buildings for cooling.

	Cluster centers	Numbers of each cluster
Severe Cold region		
Cold region		

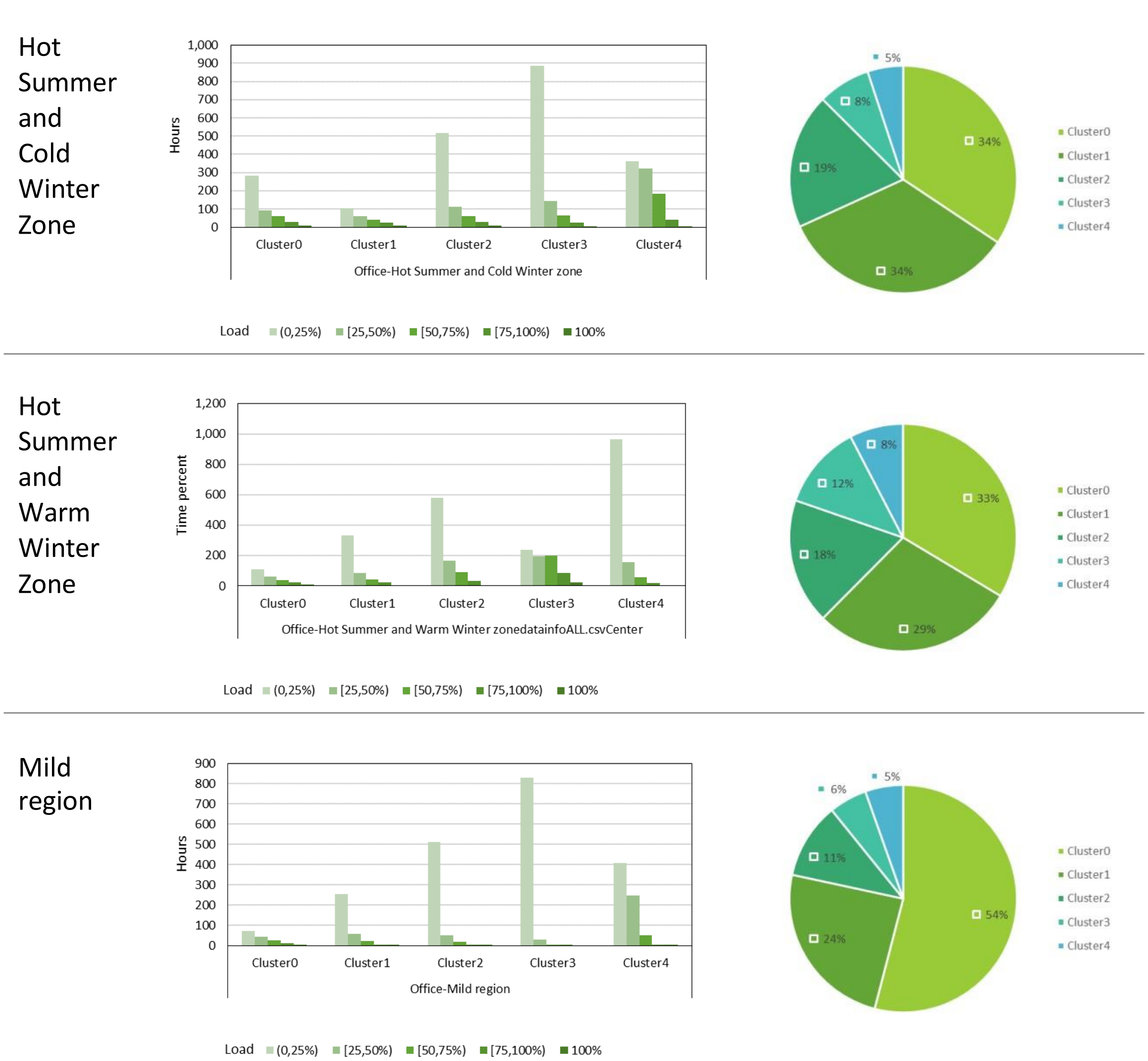


Figure 19 Clustering results of office buildings for cooling in different climate zones

Acknowledgments

The project was supported by the State Key Laboratory of Air-Conditioning Equipment and System Energy Conservation (ACSKL2018KT16).

References

1. Agency, I.E., *The Future of Cooling*. IEA, Paris, https://www.iea.org/reports/the-future-of-cooling, 2018.

2. Chua, K.J., et al., *Achieving better energy-efficient air conditioning – A review of technologies and strategies*. Applied Energy, 2013. **104**: p. 87-104.

3. *The 2018 China Central Air Conditioning Market Summary Report(in Chinese).* Mechanical and Electrical Information, 2019(04): p. 29-46.

4. Yun, G.Y., J.H. Lee, and H.J. Kim, *Development and application of the load responsive control of the evaporating temperature in a VRF system for cooling energy savings.* Energy and Buildings, 2016. **116**: p. 638-645.

5. Yan, D., et al., *IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings.* Energy and Buildings, 2017. **156**: p. 258-270.

6. Zhang, G., et al., *Influence of occupant behavior on the energy performance of variable refrigerant flow systems for office buildings: A case study.* Journal of Building Engineering, 2019. **22**: p. 327-334.

7. Hu, S., et al., *Investigation and analysis of Chinese residential building occupancy with large-scale questionnaire surveys.* Energy and Buildings, 2019. **193**: p. 289-304.

8. Chou, J.-S. and D.-K. Bui, *Modeling heating and cooling loads by artificial intelligence for energy-efficient building design.* Energy and Buildings, 2014. **82**: p. 437-446.

9. Zhang, G., et al., *New method for measuring field performance of variable refrigerant flow systems based on compressor set energy conservation.* Applied Thermal Engineering, 2019. **154**: p. 530-539.

10. Yu, X., et al., *Comparative study of the cooling energy performance of variable refrigerant flow systems and variable air volume systems in office buildings.* Applied energy, 2016. **183**: p. 725-736.

11. Fan, C., et al., *Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review.* Energy and Buildings, 2018. **159**: p. 296-308.

12. Mathew, P.A., et al., *Big-data for building energy performance: Lessons from assembling a very large national database of building energy use.* Applied Energy, 2015. **140**: p. 85-93.

13. Stopps, H. and M.F. Touchie, *Reduction of HVAC system runtime due to occupancy-controlled smart thermostats in contemporary multi-unit residential building suites.* IOP Conference Series: Materials Science and Engineering, 2019. **609**: p. 062013 (6 pp.)-062013 (6 pp.).

14. Cetin, K.S. and A. Novoselac, *Single and multi-family residential central all-air HVAC system operational characteristics in cooling-dominated climate.* Energy and Buildings, 2015. **96**: p. 210-220.

15. Rafsanjani, H.N. and A. Ghahramani, *Towards utilizing internet of things (IoT) devices for understanding individual occupants' energy usage of personal and shared appliances in office buildings.* Journal of Building Engineering, 2020. **27**: p. 12.

16. Han, J., J. Pei, and M. Kamber, *Data mining: concepts and techniques*. 2011: Elsevier.

17. D’Oca, S. and T. Hong, *Occupancy schedules learning process through a data mining framework.* Energy and Buildings, 2015. **88**: p. 395-408.

18. Kontokosta, C.E. and C. Tull, *A data-driven predictive model of city-scale energy use in buildings.* Applied Energy, 2017. **197**: p. 303-317.

19. Miller, C., Z. Nagy, and A. Schlueter, *A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings.* Renewable & Sustainable Energy Reviews, 2018. **81**: p. 1365-1377.

20. Quintana, M., P. Arjunan, and C. Miller, *Islands of misfit buildings: Detecting uncharacteristic electricity use behavior using load shape clustering.* Building Simulation, 2020.

21. Li, G., et al., *Data partitioning and association mining for identifying VRF energy consumption patterns under various part loads and refrigerant charge conditions.* Applied energy, 2017. **185**: p. 846-861.

22. Lu, Y., et al., *GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system.* Energy and Buildings, 2019. **190**: p. 49-60.

23. Wan, H., et al., *A review of recent advancements of variable refrigerant flow air-conditioning systems.* Applied Thermal Engineering, 2020. **169**: p. 114893.

24. Qian, M., et al., *Power consumption and energy efficiency of VRF system based on large scale monitoring virtual sensors.* Building Simulation, 2020.

25. Guo, Y., et al., *Development of a virtual variable-speed compressor power sensor for variable refrigerant flow air conditioning system.* International Journal of Refrigeration, 2017. **74**: p. 73-85.

26. Liu, J., et al., *Evaluation of the energy performance of variable refrigerant flow systems using dynamic energy benchmarks based on data mining techniques.* Applied energy, 2017. **208**: p. 522-539.

27. Li, G., et al., *Extending the virtual refrigerant charge sensor (VRC) for variable refrigerant flow (VRF) air conditioning system using data-based analysis methods.* Applied Thermal Engineering, 2016. **93**: p. 908-919.

28. Moon, J.W., et al., *Development of a control algorithm aiming at cost-effective operation of a VRF heating system.* Applied Thermal Engineering, 2019. **149**: p. 1522-1531.

29. Hua Liu, M.Q., Da Yan, *Analysis of Large Scale Air Conditioner User Behaviour in China Based on Data Mining Method.* Proceedings of Building Simulation 2019: 16th Conference of IBPSA, 2019: p. 2151-2156.

30. Liu, J., et al., *Abnormal energy identification of variable refrigerant flow air-conditioning systems based on data mining techniques.* Applied Thermal Engineering, 2019. **150**: p. 398-411.

31. Sun, K., et al., *A novel efficient SVM-based fault diagnosis method for multi-split air conditioning system's refrigerant charge fault amount.* Applied Thermal Engineering, 2016. **108**: p. 989-998.

32. An, J., D. Yan, and T. Hong, *Clustering and statistical analyses of air-conditioning intensity and use patterns in residential buildings.* Energy and Buildings, 2018. **174**: p. 214-227.

33. Qian, M., et al., *Evaluation of thermal imbalance of ground source heat pump systems in residential buildings in China.* Building Simulation, 2020.

34. MOHURD, *GB 50176-2016 Code for thermal design of civil building.* 2016.

35. Meesrikamolkul, W., V. Niennattrakul, and C.A. Ratanamahatana. *Shape-based clustering for time series data*. in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. 2012. Springer.

36. Lavin, A. and D. Klabjan, *Clustering time-series energy data from smart meters.* Energy Efficiency, 2015. **8**(4): p. 681-689.

37. Green, R., I. Staffell, and N. Vasilakos, *Divide and Conquer? ${k}$-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System.* IEEE Transactions on Engineering Management, 2014. **61**(2): p. 251-260.

38. Davies, D.L. and D.W. Bouldin, *A cluster separation measure.* IEEE transactions on pattern analysis and machine intelligence, 1979(2): p. 224-227.

39. Institute, H.G.M.R., et al., *GB/T 18837-2015 Multi-connected air-condition (heat pump) unit*. 2015, General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China; China National Standardization Administration. p. 56.

40. Institute, H.G.M.R., et al., *GB/T 17758-2010 Unitary air conditioners*. 2010, General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China; China National Standardization Administration. p. 72.

Declaration of interests

 The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

CRediT author statement

Mingyang Qian: Conceptualization, Methodology, Software, Formal analysis , Data Curation , Writing - Original Draft

Da Yan.: Supervision, Conceptualization, Methodology, Writing - Review & Editing.

Tianzhen Hong: Conceptualization, Methodology, Writing - Review & Editing

Hua Liu: Investigation, Resources