*energies*

Article

# Systematic Method for the Energy-Saving Potential Calculation of Air-Conditioning Systems via Data Mining. Part I: Methodology

**Rongjiang Ma [1,2], Shen Yang [1,3,*], Xianlin Wang [1,4,5], Xi-Cheng Wang [4,5], Ming Shan [1], Nanyang Yu [2] and Xudong Yang [1]**

1 Department of Building Science, Tsinghua University, Beijing 100084, China; rongjiangma@139.com (R.M.); wangxianlingree@126.com (X.W.); mshan@mail.tsinghua.edu.cn (M.S.); xyang@tsinghua.edu.cn (X.Y.)
2 School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China; rhinos@126.com
3 Human-Oriented Built Environment Lab, School of Architecture, Civil and Environmental Engineering, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
4 State Key Laboratory of Air-Conditioning Equipment and System Energy Conservation, Zhuhai 519070, China; xichengwanggree@163.com
5 Gree Electric Appliances, Inc., Zhuhai 519070, China
* Correspondence: yangshen13@tsinghua.org.cn

**Abstract:** Air-conditioning systems contribute the most to energy consumption among building equipment. Hence, energy saving for air-conditioning systems would be the essence of reducing building energy consumption. The conventional energy-saving diagnosis method through observation, test, and identification (OTI) has several drawbacks such as time consumption and narrow focus. To overcome these problems, this study proposed a systematic method for energy-saving diagnosis in air-conditioning systems based on data mining. The method mainly includes seven steps: (1) data collection, (2) data preprocessing, (3) recognition of variable-speed equipment, (4) recognition of system operation mode, (5) regression analysis of energy consumption data, (6) constraints analysis of system running, and (7) energy-saving potential analysis. A case study with a complicated air-conditioning system coupled with an ice storage system demonstrated the effectiveness of the proposed method. Compared with the traditional OTI method, the data-mining-based method can provide a more comprehensive analysis of energy-saving potential with less time cost, although it strongly relies on data quality in all steps and lacks flexibility for diagnosing specific equipment for energy-saving potential analysis. The results can deepen the understanding of the operating data characteristics of air-conditioning systems.

**Keywords:** energy saving potential; data mining; recognition; optimization; operational data



**Citation:** Ma, R.; Yang, S.; Wang, X.; Wang, X.-C.; Shan, M.; Yu, N.; Yang, X. Systematic Method for the Energy-Saving Potential Calculation of Air-Conditioning Systems via Data Mining. Part I: Methodology. *Energies* **2021**, *14*, 81. https://dx.doi.org/10.3390/en14010081

Received: 23 November 2020
Accepted: 23 December 2020
Published: 25 December 2020



## 1. Introduction

Air-conditioning systems account for 50% to 60% of the total energy consumption of buildings [1,2]. Therefore, energy-saving for air-conditioning systems would be the essence of reducing building energy consumption. Energy saving for air conditioning systems can be implemented in the system design process or during system operation, called energy-saving diagnosis. Generally, an eligible energy-saving diagnosis for air-conditioning systems can save energy, reduce system maintenance costs, extend equipment service life, and improve system control and occupant comfort.

Typically, a conventional energy-saving diagnosis includes three phases: (1) by referring to as-built drawings and the energy consumption records of the air-conditioning system, researchers obtain the basic information and status of energy consumption of the system as well as by communicating about the existing problems or troubles of the system from the operators. (2) More detailed tests, analysis, and calculation regarding existing problems should be conducted, including the typical condition and various operating conditions of different seasons according to operation records. (3) Corresponding solutions and energy-saving potential analysis are presented in summary reports. This diagnosis process was summarized as the OTI method [3], namely, observation/question, test/calculation, and identification/resolution. Building envelopes, fresh air supplies, cooling and heating sources, transmission and distribution systems, and so on are considered in this method. The method can provide the overall energy-saving potential of the air-conditioning systems and detailed diagnosis reports on each focused component of the systems. Therefore, the OTI method has been widely applied in engineering for air-conditioning system diagnosis. However, several problems exist in this method, which may adversely influence the method's condition and effect. First, the difficulties in field measurements would lead to a cost-intensive diagnosis. For instance, in the diagnosis of chillers in air-conditioning systems, the measurement and calculation of various operation parameters and indicators under typical operating modes are necessary to evaluate the energy efficiency of the chillers. For a complex system, this step would take several weeks or even months. Second, the quality of the final diagnosis results depends on the proficiency of the involved technicians and investigators to a large extent. If researchers lack experience in the diagnosis of air-conditioning systems, the decisions and adjustments will be less reliable. Therefore, it is of demand and interest to develop a more cost-efficient and dependable diagnosis method for air-conditioning systems.

Data mining is the application of specific algorithms for extracting patterns from data and has been applied to several fields with large datasets [4]. Nowadays, an increasing number of air-conditioning systems have acquired energy consumption data in the system operation process. The robust accumulation of system running data enables data mining to reveal the quality of system operation. Recent studies have applied data mining to deal with diagnosis problems in air conditioning systems and building energy consumption research, as the results were difficult to obtain by conventional methods [5]. For instance, energy modeling has been conventionally conducted in the building design process due to time-consuming data entry. However, Kim et al. [6] demonstrated that data-mining-based energy modeling could improve the energy efficiency of building design during the design phase. Ahmed et al. [7] investigated the impact of connecting building characteristics and designs with their performance using data mining techniques. The results show the high accuracy and reliability of these techniques in predicting low-energy, comfortable rooms. Data mining with association analysis was conducted for neural network modeling for an air-conditioning system via a new modeling method based on artificial intelligence algorithms proposed by Wang et al. [8]. Some studies [9–11] developed new strategies based on data mining to detect and diagnose the faults of heating, ventilation, and air-conditioning (HVAC) systems. Meanwhile, the detection and classification of abnormal energy consumption in buildings were investigated using the data mining approach in several studies [12,13]. Although abnormal utility consumption could be identified by the proposed method in conjunction with building management systems, outlier detection was only the first step for diagnosing energy efficiency and energy conservation of the system. The proposed method cannot illustrate the causes of generated outliers and optimization solutions for system operation. An improved method based on data mining proposed by Seem et al. [14] applied outlier detection to determine whether the energy consumption for a particular day is significantly different from the previous energy consumption. This method could help save diagnosis time by avoiding manual detection and reducing operation costs by detecting problems that previously would have gone unnoticed. Nevertheless, this approach reported the results by comparing the actual energy consumption and normal energy consumption (baseline). Thus, it cannot determine the potential capacity of energy conservation of the system or ensure that the system operates in the most optimized condition. Recent developments in data mining applied to building energy systems have covered load prediction [15–19], pattern identification [20–22], and fault detection and diagnosis [23–26]. However, these studies mainly focused on specific data mining technologies and algorithms rather than a systematic approach in diagnosing the energy saving potential in an air-conditioning system. Some researchers proposed a more general energy saving advisory approach based on data mining for building energy systems [27–29]. Although they did not focus on air-conditioning systems, their frameworks inspired us to have a systematic data-mining-based method for energy saving diagnosis for air-conditioning systems.

This paper presents the first of two publications proposing a systematic methodology to calculate the energy-saving potential of an air-conditioning system based on data mining. In this paper, we provide a comprehensive overview of the framework of the proposed systematic method, details of each step, and comparison with traditional OTI method from a systematic perspective. Readers are expected to gain the general logic flow, the advantages and disadvantages of the proposed method, and guidelines to apply the method based on data completeness. The companion paper [30] presents a detailed application case with specific models and algorithms applied in each step, where readers are expected to understand technical details and available data-mining technologies for applying the method in a complex air-conditioning system.

## 2. Introduction to the Systematic Method

### 2.1. Framework of the Method

The flow chart in Figure 1 illustrates the operating process of the method for energy-saving potential analysis. The method included seven steps, as shown. (1) Data collection: energy consumption data and system operation data; for example, the energy consumption of chillers and pumps, temperature of supply water, and flowrate of chilled water, should be collected comprehensively. (2) Data preprocessing: the acquired data should be preprocessed before analysis to ensure data quality and consequent reasonable analysis results. The preprocessing may include several steps, such as cleaning duplicate data and obtaining necessary evaluation parameters calculated from existing data. (3) Recognition of variable-speed equipment: for air-conditioning systems consisting of both variable- and constant-speed equipment, the identification of each equipment type should be conducted as the energy-saving analysis of the two types of equipment differs. The recognition of variable-speed equipment could be implemented based on the analysis of energy consumption data distribution of each piece of equipment. (4) Recognition of system operation mode: for systems with various operation modes, for example, air-conditioning systems coupled with ice storage systems, to recognize the operation mode corresponding to detailed energy consumption data is vital in analyzing the energy-saving potential of the systems. The recognition can be conducted via decision tree modeling on operation data. (5) Regression analysis of energy consumption data: the relationship between energy consumption data and corresponding operation parameters should be analyzed for each component to obtain basic inputs for energy-saving potential analysis. For example, regression analysis between energy consumption and water flow rate should be conducted for water pumps in air-conditioning systems. (6) Constraints analysis of system running: in the actual running of an air conditioning system, various constraints would restrict the system. These constraints should be defined before optimizing the system running to ensure that the optimized results can be realized in actual operation. (7) Energy-saving potential analysis: by setting the target as minimizing energy consumption or running cost of the air-conditioning system, the optimization of system running, corresponding energy, and cost-efficiency can be achieved via several specific optimization algorithms, such as the particle swarm optimization method. The main advantages of the proposed method include a more comprehensive analysis of energy-saving diagnosis, less involvement of professional researchers, and less time. A detailed explanation of each step is presented in Sections 2.2–2.8. It is worth mentioning that following the common practice of introducing a framework of new systematic methods [27–29], the descriptions of each step are given overall in a general way so that future research can flexibly apply the method to various cases based on data completeness of cases (Section 4.2), instead of being constrained to specific models or algorithms. An example for a case study is presented in Section 3 and more details can be found in the companion paper [30].

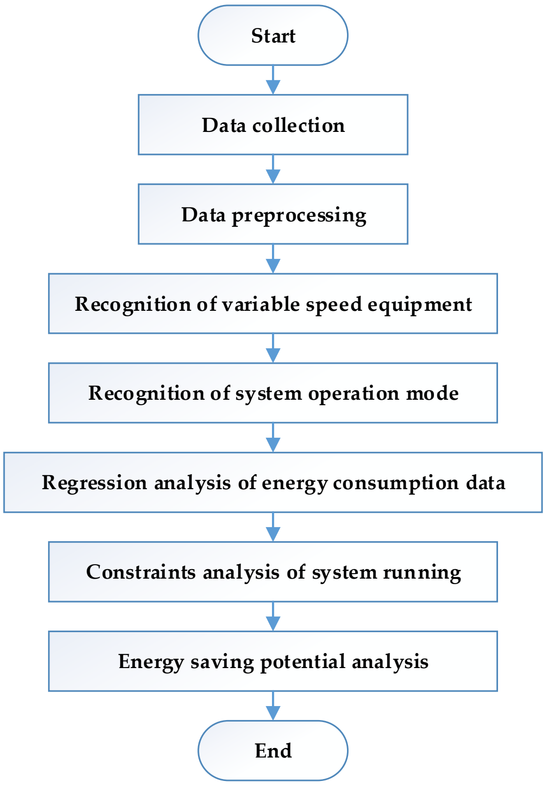


**Figure 1.** Flow chart of the proposed energy-saving potential analysis method based on data mining.

### 2.2. Data Collection

Consider a typical all-air air-conditioning system in Figure 2 as an example. Table 1 summarizes the data required for the data-mining-based energy-saving potential analysis method. The required data can be classified into two types: outdoor weather data and the running data of the air conditioning system. The outdoor weather data include atmospheric temperature and relative humidity, easily obtained through a weather station near the building. Alternatively, for a system equipped with outdoor air sensors, outdoor data can be acquired directly via the system. The running data of the system can be further categorized into two aspects, namely, energy consumption data and running status data. The energy consumption data can be obtained from the energy consumption information platform of the target building. The running status data mainly include the water flow rate and temperature of the chilled and condensing water, respectively. In addition, the opening status of valves in the system is necessary to determine the system's operation mode. Ideally, for the sake of comprehensive analysis and optimization, the energy consumption and running status data of each device in the air-conditioning system needs to be recorded and collected. However, it is recognized that in many buildings, such a refined and abundant data package is not available. The influence of the data incompleteness on the data-mining-based method will be discussed hereinafter (Section 4.2).

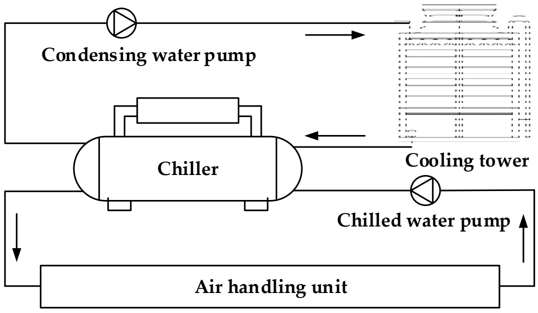


**Figure 2.** Schematic of a typical all-air air-conditioning system.

**Table 1.** Data requirements for the data-mining based energy-saving potential analysis method.

| Objects | Parameters |
|---|---|
| Outdoor air | Temperature<br>Humidity (or relative humidity) |
| Chiller | Temperature of (supply/return) chilled water<br>Temperature of (supply/return) condensing water<br>Electrical energy consumption<br>Chilled water flow<br>Condensing water flow |
| Chilled water pump | Water flow<br>Electrical energy consumption |
| Condensing water pump | Water flow<br>Electrical energy consumption |
| Plant | Temperature of (supply/return) chilled water in main pipe<br>Temperature of (supply/return) condensing water in main pipe<br>Total plant electrical energy consumption<br>Opening of major valve (For change of operating modes) |
| Cooling tower | Temperature of (supply/return) working fluid<br>Water flow<br>Electrical energy consumption |
| Air handling unit | Temperature of (supply/return) air<br>Air flow<br>Electrical energy consumption |

### *2.3. Data Preprocessing*

As the collected data are usually large and possibly influenced by noise and missing and inconsistent information, data preprocessing is vital to ensure data quality and the robustness of the analysis results. Based on the operating data features of air-conditioning systems, we can preprocess the data using the following three approaches.

- *Missing data preprocessing.* Accidents occurring during data monitoring, transfer, and storage can cause missing data, leading to deficient information on the system running status, thereby impacting the accuracy of the analysis results. There are several ways to deal with missing data, depending on their importance and quantity. At some time-points, the data of some key parameters, such as energy consumption and/or output temperature of the chiller, are missing during the collection, especially for a long period. Subsequently, we need to perform the list-wise deletion to preprocess the dataset, which means that all the data at the time points have to be removed from the following analysis. Alternatively, to ensure data continuity and isometry, we can set the weight of variables at the time points to zero. For analysis that does not require time-series conceptualization, for example, regression analysis of energy consumption and flow rate for pumps, only the data that pair with the missing one need to be excluded, which can be named pairwise deletion.

- *Data cleaning*. In terms of the data from an air-conditioning system, the data cleaning step mainly covers three aspects: (1) Duplicate data cleaning. Duplicate data may exist when data are gathered from different platforms. Duplicated data should be removed from the dataset to reduce computing costs and avoid data confusion. (2) Halt data cleaning. When an air-conditioning system halts, the collected energy consumption should be zero. Nevertheless, other parameters, such as chilled water temperature, are recorded in the dataset, which can cause confusion when we look into the trends of these parameters. In addition, for the statistics of the energy consumption data (see Section 2.4), zero data strongly influences the data distribution, and thus cannot represent the running conditions of the system. Therefore, halt data should be excluded. (3) Conflict of data cleaning. Sometimes, owing to the influences of noise and sensor failure, several related parameters could exhibit conflicts with each other. For instance, a conflict occurs when the water temperature supply is higher than the return water for a running chiller. Additional underlying conflicts against the normal principles of a running air-conditioning system may also exist. For example, there is a large disparity between the measured cooling load and the theoretical one (calculated based on the chilled water flow rate and the temperature difference between the return and supply water). In such cases, all the data at the corresponding time points with conflicts should be deleted for the following analysis.
- *Data extension*. When preprocessing the raw data, we need to add several columns to the dataset based on the calculation and summary of raw data variables. For instance, by calculating the difference between the chilled water supply and return, we can add a new variable to the dataset. Apart from adding variables by calculation, labeling of the data can also be appended. For example, after the recognition of the system operation mode (see Section 2.5), the labeling of the operation mode can be added to the dataset.

### *2.4. Recognition of Variable-Speed Equipment*

Variable-speed equipment is now widely used in air-conditioning systems. However, some systems, particularly renovated ones, would have both variable- and constant-speed equipment. The two types of equipment differ in their energy consumption features. Considering centrifugal pumps commonly equipped in air-conditioning systems as an example, the energy consumption of constant-speed pumps exhibits an approximately linear relationship with the water flow rate. For variable-speed pumps, the two parameters show a cubic function. In some scenarios, the speed characteristics of the equipment can be obtained directly from the information platform. Otherwise, we need to distinguish the speed type of all equipment based on the energy consumption data distribution. Owing to the positive relationship between running speed and energy consumption, the energy consumption of variable-speed equipment was evenly distributed within its range. By contrast, constant-speed equipment has a more concentrated energy distribution, near the maximum value. Based on the different features of energy consumption distribution, we can recognize the two types of equipment via the statistics of their energy consumption data. Our previous study [31] demonstrated that the coefficient of the median (defined as (max-median)/range) could successfully distinguish the speed type of the equipment in an air-conditioning system.

### *2.5. Recognition of System Operation Mode*

Systems equipped with more than one heat/cooling source can have several operation modes. For example, to take the advantage of the peak-valley electricity price, an air-conditioning system can be coupled with an ice storage system. In such a complex integrated system, there are six distinct operation modes: (a) shutdown, (b) ice build, (c) cooling by ice only, (d) cooling by chillers only, and (e) cooling by ice with chillers.

Recognizing the operation mode is important for analyzing energy-saving potential across the various operation modes. The groups of equipment work differently according to the incompatible system operation mode. Therefore, we can distinguish the system operation modes by analyzing the working status of equipment groups. Generally, recognition includes two steps, namely, qualitative and quantitative analysis.

The qualitative analysis aims to determine the number of operation modes in the system. As mentioned above, equipment works in groups in each operation mode. Thus, the energy consumption characteristics of the equipment group in operation mode are concentrated and differ from those in other modes. Based on this feature, we can qualitatively distinguish the operation modes by clustering analysis of energy consumption data of representative equipment, such as the chillers and pumps of different groups. The number of clusters can determine the number of operation modes in the running system. The energy consumption data characteristics of each cluster can demonstrate the operation mode the cluster. In some scenarios, researchers may have prior knowledge about the quantity and specification of the system operation modes. Therefore, the clustering analysis can be used to verify the precedent knowledge and provide an overview of energy consumption in each operation mode.

After obtaining the specification of the system operation modes, we need to determine when the system switches from one operation mode to another through quantitative analysis. The classification method would be an appropriate tool for recognition. By summarizing and refining the regulations within the dataset using various classified models, the classification method can identify the criteria to switch the system operation mode. The summarized criteria can be used to detect abnormal launches of equipment. Meanwhile, it provides the basis for calculating the energy-saving potential concerning the switching of operation mode.

### 2.6. Regression Analysis of Energy-Consumption Data

The regression analysis aims to quantify the relationship between energy consumption and running parameters (water flow rate and temperature) for each equipment, including chillers and pumps. The fitting models can be selected from the recommendations of EnergyPlus (9.3.0, National Renewable Energy Laboratory, Golden, CO, USA) and TRNSYS (v. 17, Thermal Energy System Specialists, LLC, Madison, WI, USA). The obtained model can then be used to determine predictive energy consumption in the subsequent energy saving potential calculation.

### 2.7. Constraints Analysis of System Operation

In reality, an air-conditioning system is restricted by different types of constraints during operation. These constraints also set margins to the following optimization calculation of the system operation.

The fundamental constraint of the system operation is that the cooling supply should meet the requirement of the cooling load and not exceed the maximum cooling capacity of the system. Another basic constraint is that the operation parameters, including water flow rate and temperature, should not exceed the indicated threshold of any equipment. Preferably, in order to ensure the accuracy of the optimization results, the values of the operation parameters should not exceed the range used in the precedent regression analysis (Section 2.6). For complicated systems coupled with an additional cooling source, similar constraints should be applied for each subsystem. The basic principle of the constraints analysis is the energy balance and rational range of running parameters.

### 2.8. Energy-Saving Potential Analysis

The energy-saving potential is defined as the difference between the actual energy cost and the benchmark cost (Equation (1)). Alternatively, we can add the factor of electricity price to have the cost-saving potential (Equation (2)), which may be more attractive to building owners.

$$\Delta W = W_{actual} - W_{benchmark} \tag{1}$$

$$\Delta J = \sum e \Delta W \tag{2}$$

where $\Delta W$ represents the energy-saving potential, $W_{actual}$ is the actual energy cost, $W_{benchmark}$ is the benchmark energy cost, $\Delta J$ represents the cost-saving potential, and $e$ is the electricity price.

The benchmark energy cost is the optimization result of the system operation. To calculate the benchmark value, we need to apply optimization algorithms. The basic principle is to set the total energy/cost as the objective function, which is the sum of all the equipment that has gone through the regression analysis. Afterward, by adjusting the operation parameters in the regression models obtained, we can determine the system's optimized running status and the corresponding lowest energy/cost. Naturally, the optimization process should run with the constraints summarized in Section 2.7. After obtaining the optimized (benchmark) energy/cost of the system, we can obtain the energy/cost-saving potential of the target air-conditioning system by comparing it with the actual cost.

## 3. Case Study

To validate the effectiveness of the proposed method, we performed an energy-saving potential analysis in an air-conditioning system equipped in a five-floor commercial building. The building was located in Shenzhen, China, with 30,000 m$^2$ air conditioning area. The system was coupled with an ice storage system (circulation medium: glycol-water solution) to take advantage of the lower electricity price at night, as shown in Figure 3. The system contained two chillers, nine pumps (three chilled-water pumps, three condensing-water pumps and three glycol-water pumps), two plate heat exchanges, and five air-handling units. Following the steps shown in Figure 1, the energy/cost-saving potential analysis was implemented and described in brief as follows, whereas details were presented in the companion paper [30].

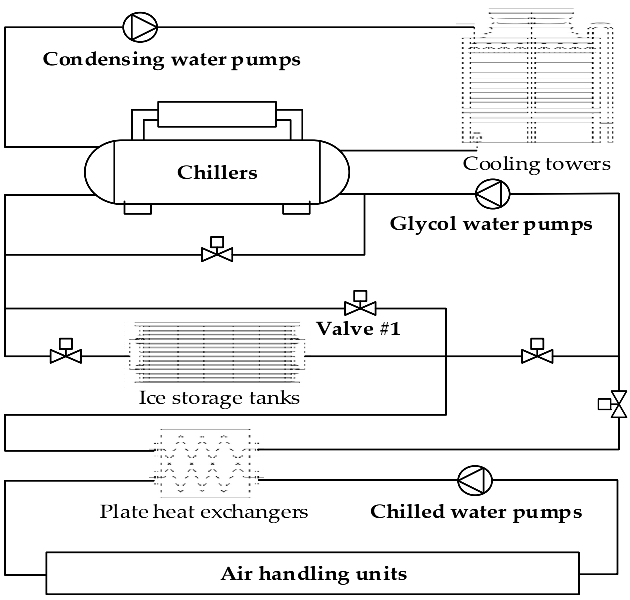


**Figure 3.** Schematic of the air-conditioning system involved in the case study.

### 3.1. Data Collection and Preprocessing

The air-conditioning system was equipped with a comprehensive data-monitoring platform, where the electricity consumption of each component and the operating data of chillers and pumps (water temperature and flowrate) were recorded. We collected such data from 22 July 2011 to 20 August 2013 at 1 h interval for energy-potential analysis during the selected period, with a total number of 1,345,975 data points.

Following data preprocessing procedure described in Section 2.3, missing data were deleted in list-wise and pair-wise ways to ensure the completeness of data pairs for analysis. Duplicated data and system operating data during shutdown status were removed. In addition to data removal, data of temperature differences between supply and return chilled-water, between the water supply and return chilled-water of the chiller, and between the water supply and return condensing-water of the chiller were calculated and added into the dataset. Finally, the units of the included parameters were transformed to the SI (Système International) scheme. After date preprocessing, we obtained 1416 pairs of continuous time-series data for following analysis.

### 3.2. Recognition of Variable-Speed Chillers and Pumps

With the coefficient of median ($C_m$) proposed in our previous study [31], the variable-speed chillers and pumps can be recognized by their energy consumption data only. It is found that four pumps had $C_m$ higher than 0.30, while the $C_m$ values of the rest five pumps and two chillers were all below 0.15 [31]. Based on such observations, we successfully distinguished that two chilled-water pumps and two glycol-water pumps operated at variable-speed mode, while the rest three condensing-water pumps, one chilled-water pump, one glycol-water pump and two chillers were constant-speed equipment.

### 3.3. Recognition of System-Operation Mode

As previously mentioned in Section 2.5, such a complex system was expected to have five distinct modes, namely "shutdown" (operation-mode code: M0), "ice build" (M1), "cooling by chiller(s) only" (M2), "cooling by ice only" (M3), and "cooling by both chiller(s) and ice" (M4). We applied the classification and regression trees (CART) algorithm [32] with good interactivity to identify the operation modes based on energy consumption and operation data. By evaluating the heterogeneity with the Gini Index [33], CART selected the splitting variables that maximized the Gini Index reduction, which were considered as the splitting criterion. Table 2 lists the results of the classification of system-operation modes and the recognizing rules. Details can be found in the companion paper [30].

**Table 2.** Recognizing rules of system operating mode.

| Operating Modes | Codes | Recognizing Rules [1] |
|---|---|---|
| Shutdown | M0 | Rule 2 & Rule 4 & Rule 6 & Rule 8 |
| Ice build | M1 | Rule 1 |
| Cooling by chiller(s) only | M2 | Rule 2 & Rule 3 |
| Cooling by ice only | M3 | Rule 2 & Rule 4 & Rule 6 & Rule 7 |
| Cooling by ice with chiller(s) | M4 | Rule 2 & Rule 4 & Rule 5 |

[1] Rule 1: time period 23:00–8:00; Rule 2: time period 8:00–23:00; Rule 3: opening of valve #1 >97%; Rule 4: opening of valve #1 ≤97%; Rule 5: energy-consumption of chillers >0; Rule 6: energy-consumption of chillers = 0; Rule 7: energy-consumption of chilled water pumps >0; Rule 8: energy-consumption of chilled water pumps = 0.

### 3.4. Regression and Constraint Analysis

Based on the recognition results in Section 3.2, different types of pumps had varying regression models. The regression model for constant-speed centrifugal pumps followed linear correlation between energy consumption and flowrate, whereas for variable-speed pumps the cubed relationship was selected. As for the chillers, common regression models were adopted from previous studies [34,35].

The system operation should be subjected to the following constraints that are necessary to consider during optimization:

(1) The cooling capacity supplied by the chiller plant room must be sufficient to meet the cooling load of the case system;

(2) The cooling capacity supplied by chillers should not exceed the maximum cooling capacity of chillers;
(3) The sum of the accumulation of the cooling capacity in the ice storage tanks and the current remaining cooling capacity should not exceed their maximum accumulation of cooling capacity;
(4) The accumulation of cooling capacity in the ice storage tanks should be equal to the cooling capacity supplied by the chillers during ice build, and cannot exceed the maximum cooling storage speed;
(5) The cooling release of the ice storage tanks should not exceed the remaining cooling capacity as well as the maximum cooling release speed.

### *3.5. Energy/Cost Saving Potential Analysis with Optimization Algorithms*

Combing the results obtained from steps above, we applied several optimization algorithms to calculate the energy/cost saving potential of the case system during the period investigated. Equation (2) was selected as the aim function. Three algorithms were examined for optimal results, namely particle swarm optimization (PSO), genetic algorithm (GA), and ant colony optimization (ACO). The optimization results are shown in Table 3, which indicate that PSO exhibited the optimized costs of the system operation relative to GA and ACO, and the cost-saving potential reached as high as 24.03% (307,213.5 vs. 400,467.6 CNY). Daily optimization results and details of the PSO algorithm can be found in the companion paper [30].

**Table 3.** Comparison of optimal costs of the case system calculated by three algorithms (particle swarm optimization (PSO), genetic algorithm (GA) and ant colony optimization (ACO)) for 59 d from 22 June to 19 August 2013.

| Day | Actual Costs (CNY) | Optimal Costs (CNY) | | |
|---|---|---|---|---|
| | | PSO | GA | ACO |
| 1–10 | 72,839.5 | 59,024.7 | 63,326.8 | 60,748.6 |
| 11–20 | 65,922.0 | 53,591.7 | 63,711.4 | 57,737.5 |
| 21–30 | 66,967.1 | 50,673.4 | 61,151.6 | 53,671.9 |
| 31–40 | 58,514.0 | 46,051.8 | 49,331.2 | 47,603.5 |
| 41–50 | 74,628.9 | 53,489.5 | 58,802.9 | 56,950.5 |
| 51–59 | 70,631.7 | 50,734.6 | 59,945.4 | 54,075.9 |
| Total | 400,467.6 | 307,213.5 | 350,321.7 | 324,513.9 |

## 4. Discussion

### *4.1. Comparison with the Conventional Method*

As summarized in the Introduction, precedent data-mining-based studies in air-conditioning systems mainly focused on respective data-mining technologies that pertain to a specific step or optimization algorithm included in our systematic method. To the best of our knowledge, the method proposed in this paper is the first systematic one targeting energy-saving potential analysis of air-conditioning systems. Therefore, it is not feasible to compare the method and results with those from previous studies. Instead, the advantages and disadvantages of the proposed method are illustrated by comparison with the conventional OTI method from the following four aspects.

#### 4.1.1. Consumed Resources

The traditional OTI method initially needs to understand the basic information of the system. Then, the researchers need to communicate about the current issues existing in the system with the system management. Based on the collected information, they have to decipher a diagnostic program to output the energy-saving potential in the end. For complicated systems, such as those involved in this study, the whole process can take up to several weeks or even months. However, the proposed method can only spend several days on a thorough analysis of the system. Similarly, the human resources involved would be much less for the proposed method than the conventional method. In addition, the requirement of the air-conditioning knowledge level for the researchers is less for the proposed method. However, as computers do almost all the methodological work, and due to the large amount of data, the demand for computational resources is generally higher for the proposed method.

#### 4.1.2. Data in Use

The proposed method relies more on the data-monitoring platform, whereas the OTI method mainly uses data from the field test. Therefore, without a complete data monitoring system, the proposed method cannot work properly for air-conditioning systems. Hence, the proposed method has higher demands of data monitoring and completeness relative to the OTI method. An advantage of the proposed method using monitoring data is the synchronization of the data. Typically, in a data platform, we can obtain a package of energy consumption and running status parameter data at each time point, ensuring data synchronization for analysis. In contrast, the conventional OTI method normally needs to perform field tests on the main equipment one by one. Thus, the data obtained from different equipment may have time lags and will probably influence the analysis results if the system runs in an unsteady state. However, as the method is based on data mining, the range and amount of data are generally larger than the conventional method. If such a vast amount of data is not processed decently, the proposed method will be easier to face "data disaster" and output non-realistic results.

#### 4.1.3. Technical Details

The conventional OTI method mainly focuses on the running status and energy consumption of specific equipment and, thus, it cannot provide a comprehensive energy saving potential analysis as the proposed new method. The proposed method can provide a broader picture of the system running constraints. Therefore, it can provide more reasonable solutions compared to the OTI method in terms of system constraints.

#### 4.1.4. Main Characteristics

Concerning the flexibility of the method, it is acknowledged that the proposed method is less flexible than the conventional OTI method, as the latter can pay attention to specific components of the systems. Nevertheless, in terms of universality and scalability, the proposed method performs better than the traditional method, providing a comprehensive picture of the energy-saving potential analysis, and is achieved through remote control and online diagnosis.

#### 4.1.5. Summary

In summary (Table 4), the conventional OTI method is problem-oriented, while the proposed systematic method is data-oriented. Hence, the methodologies applied in each step of the proposed method are evolving from those existing in data science, whereas the conventional OTI method mostly relies on field examinations. The main advantages of the proposed method include a more comprehensive analysis of energy saving diagnosis, less involvement of professional researchers, and less time. However, it strongly relies on the data quality of the monitoring platform and lacks flexibility for diagnosing specific equipment.

It should be noted that the proposed systematic method is not in competition with the conventional OTI method. They possess complementary aims and pathways and should thus be considered as supplementary to each other. Therefore, in engineering practice, it is recommended to combine the data-mining based method and the conventional OTI method to deal with various problems faced in air-conditioning systems.

**Table 4.** Comparison between the proposed systematic method and conventional observation, test, and identification (OTI) method.

| Item | Proposed Systematic Method | Conventional OTI Method |
|---|---|---|
| Orientation | Data-oriented | Problem-oriented |
| Basic principle | Energy saving potential optimization calculation based on daily operation data | Energy saving diagnosis by field investigation and test |
| Main pathway | Optimizing operation strategies based on existing equipment and systems | Troubleshooting existing equipment or retrofitting equipment |
| Consumed resources | Low consuming of time and human resources, but high demand of computational resources | Time- and researcher-consuming but low level of computational demands |
| Data in use | Strong dependence on data monitoring platform; high-level demand of data scope and quantity; better data synchronization. | Strong dependence on field measurement; potential issues in data synchronization |
| Technical details | Reasonable solutions owing to a broader picture of system running constraints | Potential issues in realizing proposed solutions |
| Main characteristics | Less flexibility, especially with respect to individual component; remote control and online diagnosis | Wider applicability in both equipment and systems; inevitable field presence |

It is also worth mentioning that this paper aims at introducing the framework and logics of the newly proposed systematic method. Hence, we did not put emphasis on the comparisons of specific data-mining methods and optimization algorithms. Future research can follow the steps of the proposed systematic method, when various data-mining technologies can be flexibly applied according to data completeness and specific cases. For example, Section 3.4 in the companion paper [30] provided detailed comparisons for the selection of models.

### *4.2. Influence of Data Completeness*

As mentioned above, the proposed method based on data mining strongly depends on data completeness. In many scenarios, due to the lack or deficiency of the system monitoring platform, we cannot collect all the data listed in Table 1 and thus cannot implement energy-saving diagnosis using the new method. However, according to different data completeness levels, we can still analyze the energy consumption of the system and acquire energy-saving information to some extent, as summarized in Figure 4.

With only the annual total energy consumption of the air-conditioning system, we can perform longitudinal and cross-sectional comparisons to determine whether the energy consumption for the year is normal. Specifically, the longitudinal comparison means comparing the annual total energy cost with that of previous years to check if the value exceeds the precedent range. The cross-sectional comparison compares the energy consumption with other systems in similar buildings to rate the target's performance. When we can add the dimension of time to have time-series total energy consumption data, more data mining methods can be applied to analyze the system's energy consumption, for example, energy consumption prediction. Furthermore, if the data of the subsystems, such as pumps, are available, we can determine which subsystems have energy-saving potentials. When the time-series energy consumption data can be further refined for each piece of equipment, we can have a clear picture of which equipment has operational issues and how to improve the energy performance by analyzing running time, sequence, and other mining methods. Preferably, if the complete energy consumption and system operational data are accessible, a more comprehensive optimization analysis can be implemented, as in the method proposed in this study.

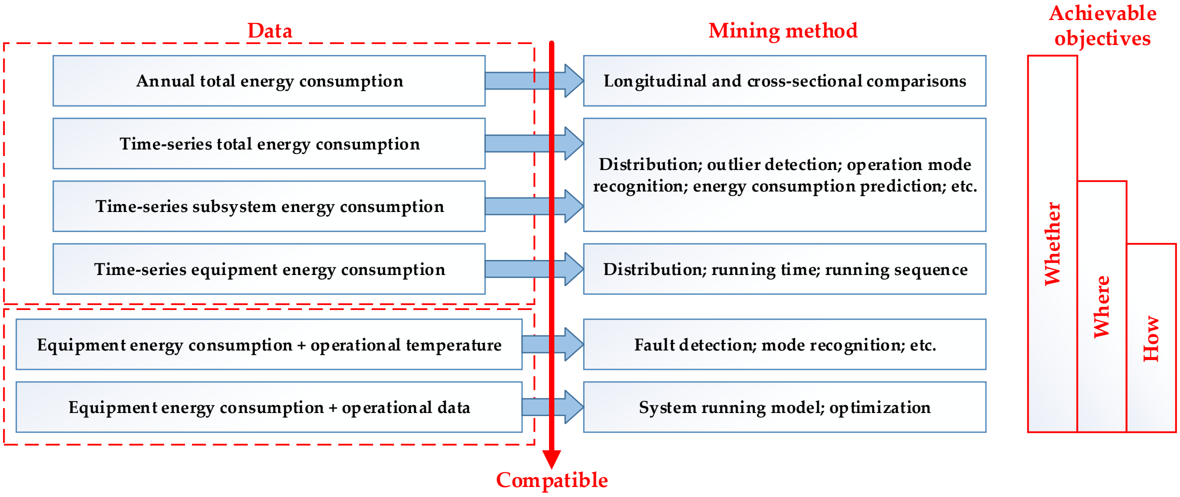


**Figure 4.** Influence of data completeness on the mining method and achievable objectives.

### *4.3. Limitations and Future Outlook*

This study has the following limitations:

- As presented in Section 4.1, It has been a common practice to illustrate the advantages and disadvantages of a newly proposed framework in a qualitative way, especially for new systematic methods [27–29]. Nevertheless, in order to make a more straightforward comparison, there are three future directions to work on: (1) to apply the systematic method to a large number of case studies to have an overall range of the performance evaluation and then to compare them with the performance range of the conventional OTI method; (2) to apply both methods in one case study simultaneously and to directly compare the performance of two methods; and (3) given the complementary features of the two methods, to apply both methods together in one case study and to obtain more comprehensive diagnosis results. We encourage more research to target the aforementioned three directions in order to demonstrate in a straightforward way the advantages and disadvantages of the proposed method, and to investigate optimal approaches to coordinate it to traditional methods.
- The main objective of this study has been to calculate the energy saving potential by optimizing the system operating parameters, when the primary cooling load demand can be met. Therefore, indoor environmental quality (IEQ, including but not limited to air change rate and thermal comfort) is not considered yet. We believe that other than focusing on energy consumption, further research can also attempt to apply the proposed method to consider the optimization of IEQ.

## 5. Conclusions

This study proposes a systematic method for energy-saving diagnosis in air-conditioning systems using data mining. The method mainly includes seven steps: (1) data collection, (2) data preprocessing, (3) recognition of variable-speed equipment, (4) recognition of system operation mode, (5) regression analysis of energy consumption data, (6) constraints analysis of system running, and (7) energy-saving potential analysis. A case study with a complicated air-conditioning system coupled with an ice storage system demonstrated the effectiveness of the proposed method, and the details were reported in the companion paper. Compared with the traditional OTI method, the data-mining-based method can provide a more comprehensive analysis of energy-saving potential in less time, although it strongly relies on data quality and lacks flexibility for diagnosing specific equipment. The newly proposed method can also deepen the understanding of the operational data characteristics of air-conditioning systems. Further research is warranted to examine the applicability of various data-mining technologies to realize a systematic diagnosis method.

**Author Contributions:** Conceptualization and project administration, R.M.; methodology and data analysis, R.M. and S.Y.; academic guidance, resources and supervision, X.W., X.-C.W., N.Y., X.Y., and M.S.; writing—original draft, R.M., and S.Y.; writing—review and editing, X.W., X.-C.W., M.S., X.Y., N.Y., and S.Y.; funding acquisition, X.Y., M.S., and R.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the State Key Laboratory of Air-conditioning Equipment and System Energy Conservation (Grant No. ACSKL2018KT1201), National Science and Technology Pillar Program during the Thirteenth Five-year Plan Period (Grant No. 2018YFD1100702), and Beijing Science and Technology Program (Grant No. Z181100005418005).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data was obtained from Energy Solution Development (Shenzhen) Company Limited, which is not allowed to share data other than that already disclosed herein.

**Acknowledgments:** The authors would like to thank Yuyang Feng of Energy Solution Development (Shenzhen) Company Limited for providing the case data for this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lam, T.N.T.; Wan, K.K.W.; Wong, S.L.; Lam, J.C. Impact of climate change on commercial sector air conditioning energy consumption in subtropical Hong Kong. *Appl. Energy* **2010**, *87*, 2321–2327. [CrossRef]
2. Li, C.; Zhou, J.; Cao, Y.; Zhong, J.; Liu, Y.; Kang, C.; Tan, Y. Interaction between urban microclimate and electric air-conditioning energy consumption during high temperature season. *Appl. Energy* **2014**, *117*, 149–156. [CrossRef]
3. Ginestet, S.; Marchio, D.; Morisot, O. Improvement of buildings energy efficiency: Comparison, operability and results of commissioning tools. *Energy Convers. Manag.* **2013**, *76*, 368–376. [CrossRef]
4. Fayyad, U.M.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. *Advances in Knowledge Discovery and Data Mining*; AAAI Press: Cambridge, MA, USA, 1996.
5. Zhao, Y.; Zhang, C.; Zhang, Y.; Wang, Z.; Li, J. A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis. *Energy Built Environ.* **2020**, *1*, 149–164. [CrossRef]
6. Kim, H.; Stumpf, A.; Kim, W. Analysis of an energy efficient building design through data mining approach. *Autom. Constr.* **2011**, *20*, 37–43. [CrossRef]
7. Ahmed, A.; Korres, N.E.; Ploennigs, J.; Elhadi, H.; Menzel, K. Mining building performance data for energy-efficient operation. *Adv. Eng. Inform.* **2011**, *25*, 341–354. [CrossRef]
8. Wang, Z.; Han, N.; Wang, Y. Studies on neural network modeling for air conditioning system by using data mining with association analysis. In Proceedings of the 2011 International Conference on Internet Computing and Information Services, ICICIS 2011, Hong Kong, China, 17–18 September 2011; pp. 423–427.
9. Hou, Z.; Lian, Z.; Yao, Y.; Yuan, X. Data mining based sensor fault diagnosis and validation for building air conditioning system. *Energy Convers. Manag.* **2006**, *47*, 2479–2490. [CrossRef]
10. Rakhshani, E.; Sariri, I.; Rouzbehi, K. Application of data mining on fault detection and prediction in boiler of power plant using artificial neural network. In Proceedings of the POWERENG 2009—2nd International Conference on Power Engineering, Energy and Electrical Drives Proceedings, Lisbon, Portugal, 18–20 March 2009; pp. 473–478.
11. Du, Z.; Fan, B.; Jin, X.; Chi, J. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Build. Environ.* **2014**, *73*, 1–11. [CrossRef]
12. Liu, D.; Chen, Q.; Mori, K.; Kida, Y. A method for detecting abnormal electricity energy consumption in buildings. *J. Comput. Inf. Syst.* **2010**, *6*, 4887–4896.
13. Li, X.; Bowers, C.P.; Schnier, T. Classification of energy consumption in buildings with outlier detection. *IEEE Trans. Ind. Electron.* **2010**, *57*, 3639–3644. [CrossRef]
14. Seem, J.E. Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. *Energy Build.* **2005**, *37*, 127–139. [CrossRef]
15. Bagnasco, A.; Fresi, F.; Saviozzi, M.; Silvestro, F.; Vinci, A. Electrical consumption forecasting in hospital facilities: An application case. *Energy Build.* **2015**, *103*, 261–270. [CrossRef]
16. Zhao, D.; Zhong, M.; Zhang, X.; Su, X. Energy consumption predicting model of VRV (Variable refrigerant volume) system in office buildings based on data mining. *Energy* **2016**, *102*, 660–668. [CrossRef]

17. Fan, C.; Xiao, F.; Zhao, Y. A short-term building cooling load prediction method using deep learning algorithms. *Appl. Energy* **2017**, *195*, 222–233. [CrossRef]
18. Naji, S.; Keivani, A.; Shamshirband, S.; Alengaram, U.J.; Jumaat, M.Z.; Mansor, Z.; Lee, M. Estimating building energy consumption using extreme learning machine method. *Energy* **2016**, *97*, 506–516. [CrossRef]
19. Amber, K.P.; Ahmad, R.; Aslam, M.W.; Kousar, A.; Usman, M.; Khan, M.S. Intelligent techniques for forecasting electricity consumption of buildings. *Energy* **2018**, *157*, 886–893. [CrossRef]
20. Li, G.; Hu, Y.; Chen, H.; Li, H.; Hu, M.; Guo, Y.; Liu, J.; Sun, S.; Sun, M. Data partitioning and association mining for identifying VRF energy consumption patterns under various part loads and refrigerant charge conditions. *Appl. Energy* **2017**, *185*, 846–861. [CrossRef]
21. Qiu, S.; Feng, F.; Li, Z.; Yang, G.; Xu, P.; Li, Z. Data mining based framework to identify rule based operation strategies for buildings with power metering system. *Build. Simul.* **2019**, *12*, 195–205. [CrossRef]
22. Fan, C.; Xiao, F.; Song, M.; Wang, J. A graph mining-based methodology for discovering and visualizing high-level knowledge for building energy management. *Appl. Energy* **2019**, *251*, 113395. [CrossRef]
23. Zhao, Y.; Li, T.; Zhang, X.; Zhang, C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renew. Sustain. Energy Rev.* **2019**, *109*, 85–101. [CrossRef]
24. Karami, M.; Wang, L. Fault detection and diagnosis for nonlinear systems: A new adaptive Gaussian mixture modeling approach. *Energy Build.* **2018**, *166*, 477–488. [CrossRef]
25. Yan, K.; Ma, L.; Dai, Y.; Shen, W.; Ji, Z.; Xie, D. Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis. *Int. J. Refrig.* **2018**, *86*, 401–409. [CrossRef]
26. Guo, Y.; Tan, Z.; Chen, H.; Li, G.; Wang, J.; Huang, R.; Liu, J.; Ahmad, T. Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. *Appl. Energy* **2018**, *225*, 732–745. [CrossRef]
27. Ashouri, M.; Haghighat, F.; Fung, B.C.M.; Lazrak, A.; Yoshino, H. Development of building energy saving advisory: A data mining approach. *Energy Build.* **2018**, *172*, 139–151. [CrossRef]
28. Fung, B.C.M.; Haghighat, F. Extracting knowledge from building-related data—A data mining framework. *Build. Simul.* **2013**, *6*, 207–222.
29. Xiao, F.; Fan, C. Data mining in building automation system for improving building operational performance. *Energy Build.* **2014**, *75*, 109–118. [CrossRef]
30. Ma, R.; Yang, S.; Wang, X.; Wang, X.; Shan, M.; Yu, N.; Yang, X. Systematic method for the energy-saving potential calculation of air conditioning systems via data mining. Part II: A detailed case study. *Energies* **2021**, *14*, 86. [CrossRef]
31. Ma, R.; Wang, X.; Shan, M.; Yu, N.; Yang, S. Recognition of variable-speed equipment in an air-conditioning system using numerical analysis of energy-consumption data. *Energies* **2020**, *13*, 4975. [CrossRef]
32. Yan, R.; Ma, Z.; Zhao, Y.; Kokogiannakis, G. A decision tree based data-driven diagnostic strategy for air handling units. *Energy Build.* **2016**, *133*, 37–45. [CrossRef]
33. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Elsevier Inc.: Waltham, MA, USA, 2012; ISBN 9780123814791.
34. Zhang, G.; Chen, L. Opmization of energy consumption in chilled plant. *Fluid Mach.* **2012**, *40*, 75–80.
35. Ma, R. The Study of Energy Efficiency Diagnosis Methodology for Air-Conditioning Systems with the Actual Operating Data. Ph.D. Thesis, Southwest Jiaotong University, Chengdu, China, 2014.