Applied Energy Symposium 2021: Low carbon cities and urban energy systems
September 4-8, 2021, Matsue, Japan
Paper ID: 2

Performance improvement control methods for solar-air source heat pump systems based on meteorological k-means clustering

Lu Chen[1], Dong Liu [2*]

1 School of Mechanical Engineering Tongji University

2 School of Mechanical Engineering Tongji University

ABSTRACT

To solve significant differences in the performance of solar-air source heat pump systems under different weather conditions for an office building of scientific research in Shanghai, it proposes to classify the meteorological data by cluster analysis in this paper. Mainly, it uses solar insolation and outdoor temperature as two primary indicators with ten secondary indicators for further analysis. By this means, this study classifies 90 days of winter meteorological data in Shanghai into eight categories. Data standardization, factor analysis, and k-means clustering are the critical methods, and Bayes discriminant verifies the correct rate of 98.9% in the paper. Furthermore, it selects the typical day of every selected class to analyze the heat pump operating time effect on the system COP. Finally, the maximum system COP was used to determine the heat pump operating time as the performance optimization target. Meaningfully, the intra-class daily data was verified to prove that the clustering result was highly accurate and reliable for further research and provides a solid related system control strategy.

Keywords: solar energy, air-source heat pump, k-means clustering, Bayes discriminant, factor analysis

NOMENCLATURE

Abbreviations	
SSE	Sum of the Squared Errors
Symbols	
T_1	Collector's outlet water temperature, °C
T_2	Bottom return water temperature of the tank, °C
T_3	The upper input mixing monitoring temperature (65°C), °C
T_4	The load-side water discharge temperature of the tank, °C
k1_airpump	The air source heat pump-side pump valve control signal
k2_s	The solar-side pump valve control signal
Z	Standardized indicators
X	Original data
$\bar{X}$	Average original data
s	The standard deviation
x_{ij}	The jth characteristic variable of the ith overall
COP_{sys}	System COP
$Q_c(t)$	Solar Collector Heat, kW
$Q_r(t)$	Heat pump heat production, kW
$W_1(t)$	Energy consumption of solar-side circulation pumps, kW
$W_2(t)$	Energy consumption of heat pump -side circulation pump, kW
$W_r(t)$	Air-source heat pump energy consumption, kW

1. INTRODUCTION

As one of the renewable energy sources, it widely uses solar energy because of its universality, colossal quantity, harmlessness, and longevity. However, it often encounters many dilemmas in the process of solar energy development. For example, it needs to determine solar energy extraction according to the weather and seasonal changes. The solar heat pump technology system, on the other hand, can make up for the shortcomings of the traditional system of unstable and scattered heat sources to improve the utilization of solar energy [1]. In recent years, there have been numerous studies on solar-air source heat pump systems. Mostly the authors choose specific meteorological days for experiments. The results do not have systematic characteristics, and the experiments are more dependent on the results meteorological data of the selected dates. Selecting random data may cause unreliable experiment results. Thus, some researchers have performed much preliminary analysis of meteorological data and applied their methods to experiments.

Jamer Jiménez Mares et al. [2] used hierarchical clustering to forecast energy demand to determine how weather conditions affect users' electricity consumption. Similarly, Liu [3] used the average solar irradiance and ambient temperature to select the closest typical day of four seasons and made the experimental analysis of solar-air source heat pump systems in hot-summer and cold-winter regions. Finally, Zeng et al. [4] used systematic clustering to classify winter meteorological data in Xichang and discriminate heat pump operating time with the goal of the lowest energy consumption. The clustering analysis has excellent application value in optimizing systems affected by the weather because of meteorological data's continuous, seasonal, and time-series characteristics.

However, only monitoring heat pump energy consumption does not reflect the system optimization characteristics, so the integrated analysis of solar-air source heat pump systems has more excellent research prospects. Therefore, this paper proposes a k-means clustering method for research, with solar insolation and outdoor temperature as the main influencing factors. Combined with statistical methods such as multi-index factor analysis and Bayes discrimination, it conducts winter weather clustering of Shanghai solar-air source heat pump systems. Furthermore, it regulates the heat pump operating time for different types of clustering according to experimental results. Finally, it uses system COP as the performance target for further validation and research.

2. MODEL DESCRIPTION

2.1 System Parameters

The solar-air source heat pump hot water system in this paper uses an R&D building in Shanghai for design. Meanwhile, the hot water system is based on solar heating and supplemented by an air source heat pump. The building has four floors and provides a domestic hot water supply for 50 researchers, and the supply time interval is set from 8:00 to 18:00. According to the provisions in the standard [5], the hot water consumption in this system takes the value of 10L/p·d. The mixed outlet water temperature of the system is 60°C, and the water is regulated to 45°C by the heat exchanger. The local average cold water temperature specifies the system's cold-water temperature, which is 15°C. The system's remaining key equipment parameter settings are designed according to the standards [6-7]in Tab.1. In this model's assumption, the water pump is used as the circulating power, with the flow rate fixed. Meanwhile, the water using process does not replenish cold water to the tank.

Tab 1 Key equipment parameters

Part Name	Parameter	Data
Solar Collector	Area /m^2	10
	Inclination /°	36
	Intercept efficiency	0.727
	Slope efficiency /$W\cdot(m^2\cdot K)^{-1}$	5.516
	Recommended workflow /$kg\cdot(m^2\cdot s)^{-1}$	0.02
Solar side circulation pump	Rated flow rate /$kg\cdot h^{-1}$	1079
	Rated Power/W	27.78
Tank	Volume /m^3	0.55
	Heat loss coefficient /$kJ\cdot(h\cdot m^3)^{-1}$	5
Air-source heat pump	Rated heating capacity/kW	18
	Rated heating power /kW	5.5
Heat pump side circulation pump	Rated flow rate /$kg\cdot h^{-1}$	540
	Rated Power/W	34.26

2.2 Model Setup

TRNSYS software sets up the computational model in this paper, and Fig.1 shows its specific flow chart.

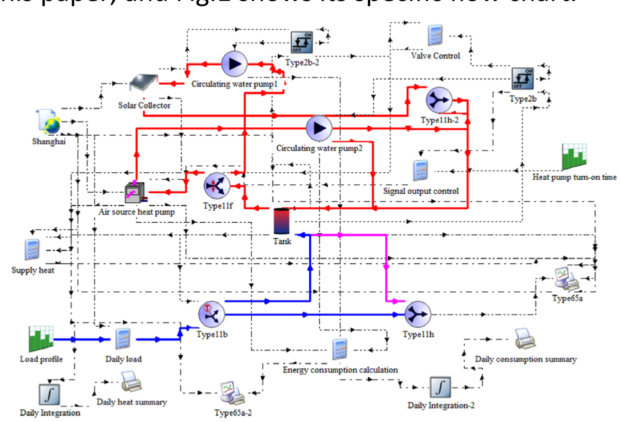

Fig 1 TRNSYS system model flow chart

2.3 System Control

According to the weather conditions, the system divides the operation mode settings into three types: solar system, air-source heat pump system, and dual-system combined operation. Mainly, the temperature difference controller, timing controller, and valve controller are the regulation components.

2.3.1 Heat Collection System Circulation Control

Heat collection system circulation contains two temperature difference controllers. The signal transmission of temperature difference controller 1 is mainly regulated based on T_1 and T_2. The upper and lower limit of the range of the temperature difference ΔT between the two is 8°C and 2°C. Likewise, the signal transmission of the heat pump-side temperature difference controller 2 is mainly to set T_3 and T_4 for regulation. The upper and lower limit of the temperature difference ΔT range of both is 5°C and 1°C.

2.3.2 Valve Flow Control

"Valve control" calculator controls the valve flow proportion between the two-circulation pump. The main logic is to set two control signals: k1_airpump and k2_s. When the valves are open, the sum of the output signal is 2, and the actual flow rate regulates the proportion of shunt. When only one side of the pump runs, the running side signal is 1, and the other side is 0. Therefore, the control logic formula of valve V_c is

$$V_c = gt(k1\_airpump + k2\_s, 1.5) * \left(\frac{540}{540+1079}\right) + k1\_airpump * (1 - k2\_s) \tag{1}$$

2.3.3 Start-stop Control for The Air-source Heat Pump

"Signal output control" calculator controls the operation of the air-source heat pump, which is mainly controlled jointly by the temperature difference controller 2 and the timing controller "Heat pump operating time." The output signal of temperature difference controller 2 is set to y_x. The heat pump turns on for the timing controller when the corresponding time is set to 1 and turns off when 0. The output signal of the joint control signal is set to y. The start-stop control logic formula is as follows：

$$y = \mathrm{int}(0.5 * (y_x + y_t)) \tag{2}$$

3. METEOROLOGICAL CLUSTERING METHODS AND OPERATIONAL REGULATION STRATEGIES

3.1 Meteorological Clustering Methods

Cluster analysis is the formal study of algorithms that group data using measured or perceived features. The main standard algorithms are such as k-means clustering, hierarchical clustering. This paper mainly adopts the k-means clustering method, an iterative solution clustering analysis algorithm, rapid and classical. The basic steps of this algorithm are as followed.

The k-means algorithm computes the process.

- Randomly generate k clustering centroids.
- Calculate the distance from the data points to each center.
- Assign the data points to the nearest center.
- Calculate the new clustering centers.
- Determine whether the data points are reassigned; if reassigned, return to the second step; if not, complete the clustering.

3.2 SPSS-based Meteorological Clustering process

3.2.1 Basic Evaluation Indicators

According to the characteristics of meteorological conditions, solar irradiation and outdoor temperature are the main factors affecting the system's performance. For different daily radiation patterns, the irradiation has a significant difference. While the outdoor temperature changes with the peak of solar radiation, there is an inevitable delay, so the solar radiation cannot be a simple indicator. In this paper, the maximum system COP is performance measurement, and solar irradiation and outdoor temperature are two primary indicators. According to the daily time interval, each primary indicator is divided into five secondary indicators.

3.2.2 Data Standardization

The data must be standardized before the clustering analysis due to the indicators' different units and dimensions. Therefore, this paper uses the Z-score standardization for studying. The calculation formula is as follows. After standardization, the generated data indicators were further indicators.

$$Z = \frac{X - \overline{X}}{s} \tag{3}$$

3.2.3 Multi-index Factor Analysis

In this study, the solar irradiance and outdoor temperature data from December, January, and February in

Shanghai are divided into ten sub-indicators by daily period. In the calculation, the processing speed is slow due to a large amount of data, while commonality exists among indicators. Therefore, by selecting the high-dimensional data through the commonality factor, the data can be reduced in dimensionality. Using SPSS software for factor analysis, the factors' number is judged mainly based on the total variance explained. There are two main methods to judge the effectiveness of factor analysis: Bartlett's test and the KMO test, in which the validity of that was further determined.

3.2.4 k-means Clustering Analysis

After dimensionality reduction, the initialized data are generated for clustering. Since the k-means algorithm is calculated based on the average value of objects in a cluster, distance is used as the similarity function, interpreting the meteorological data obtained as a spatial vector. Its distance is expressed as the Euclidean distance, defined as follows.

$$d(i,j)=\sqrt{\left|x_{i1}-x_{j1}\right|^2+\left|x_{i2}-x_{j2}\right|^2+...+\left|x_{ip}-x_{jp}\right|^2} \quad (4)$$

For the k-means clustering algorithm, the clustering effect is better when the resultant clusters are denser. Moreover, the difference between clusters is more evident. The range of the clusters' number was set according to the related paper [4]. Moreover, the optimal clusters' number is determined according to the method of SSE calculation, and when the curve of SSE value is abruptly changed to smooth, the data of the turning point is found.

3.2.5 Bayes Discriminant Analysis

In this study, Bayes discriminant analysis was chosen to test the k-means clustering data's rationality. The basic principle of Bayes discriminant is that there are k aggregates, the density function of each aggregate distribution is $f_i(X)$, the prior probability is given as q_i, i=1, 2..., k. The posterior probability to be judged is calculated from the Bayes formula as:

$$p_i=\frac{q_i f_1(x)}{\sum_{i=1}^{k} q_i f_i(x)} \quad (5)$$

Based on Bayes discrimination, the target is to find the maximum p_i and thus assign samples to the corresponding clusters.

3.2.6 System Energy Efficiency Monitoring Objectives

For testing the system's performance, this paper evaluated and monitored the system COP. Therefore, system operation optimization aims to obtain the maximum COP_{sys} for the whole day of the calculation. The COP_{sys} calculation formula for solar-air source heat pump systems specifies as the following.

$$COP_{sys}=\frac{\int_0^t\left[Q_c(t)+Q_r(t)\right]dt}{\int_0^t\left[W_1(t)+W_2(t)+W_r(t)\right]dt} \quad (6)$$

Therefore, this study uses the method of controlling the operating time of the heat pump unit under different operating conditions to further analyze its effect on the system COP. In the simulation, a time step of 1h was chosen as the optimal time to turn on the heat pump and the optimal operation mode of the system when the maximum COP_{sys} of the whole day was obtained.

4. RESULTS AND DISCUSSION

4.1 Evaluation Indicator Selection

Based on selecting clustering indicators for this study obtained from the analysis above, the specific indicators were classified in Tab. 2 below.

Tab 2 Meteorological clustering evaluation indicators

First Indicators	NO.	Secondary indicators	Time
Daily solar irradiation (kWh)	1	F_1	Sunrise~10:00
	2	F_2	10:00~12:00
	3	F_3	12:00~14:00
	4	F_4	14:00~16:00
	5	F_5	16:00~18:00
Outdoor temperature (°C)	6	T_1	Sunrise ~10:00
	7	T_2	10:00~12:00
	8	T_3	12:00~14:00
	9	T_4	14:00~16:00
	10	T_5	16:00~18:00

4.2 Results of factor analysis and cluster analysis

The meteorological data were standardized to obtain ZF1~ZF5 and ZT1~ZT5. The ten secondary index data were further factor analyzed by SPSS software: the KMO sampling fitness value is 0.652, which is greater than 0.5, so it meets the conditions of factor analysis; The Bartlett's sphericity test value was close to 0, which was much smaller than 0.05, so the factor analysis result was valid. Four principal components were extracted as factors by principal component analysis, and their cumulative percentage reached 95.963%, which was greater than 90%, so the extracted factors could be the representatives. The four index factors obtained were FAC1~FAC4 as new variables for the next step of cluster analysis.

Based on the related paper's research, the cluster numbers were initially set from 5 to 15. In the k-means clustering, *SSE* is one indicator to discern whether the number of clusters is reliable. The calculation is obtained as shown in Fig.2.

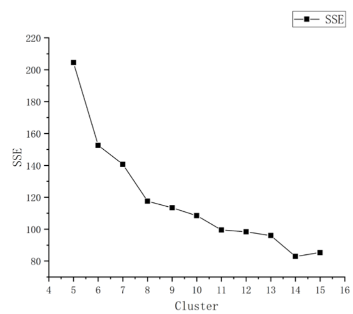

Fig.2 Different clusters' SSE value

According to Fig.2, When the simulated *k* value reaches the actual number of clusters, the decline in *SSE* decreases abruptly. However, it then levels off as the *k* value increases, corresponding to this curve's "elbow" as the correct number of data clusters. Therefore, the selected number of clusters according to Fig.2 is 8.

4.3 Bayes Discriminant Analysis results

By Bayes discriminant analysis, the data of the above indicators were subjected to obtain eight discriminant functions. The four indicators' data were brought into the discriminant functions to derive further the posterior probabilities, which led to the analysis results. Moreover, both the original data and the cross-processed data were validated. The correct classification rate of the original data count was 98.9%, and that of the cross-processed data was 93.3%. Thus, it is reasonable to cluster the winter meteorological data of Shanghai into eight categories in this study.

4.4 Typical Day and Heat Pump Operating Time

By k-means clustering analysis, the days of the meteorological data closest to the center, which have the smallest distance from the mass center of each cluster, are taken as typical days. Therefore, the typical day statistics table is obtained in Tab.4.

Tab 4 Typical Days Statistics

Class	Intra-class numbers	Typical Day Date
1	4	Dec.1
2	3	Jan.19
3	17	Dec.20
4	7	Feb.15
5	4	Dec.19
6	17	Jan.12
7	21	Feb.13
8	17	Jan.13

The typical daily meteorological data from the clusters were brought into the TRNSYS model simulation, and the data was obtained by using the system COP formula. These data only consider the heat pump to meet users' water demand, to get the effect of the heat pump operating time on the system COP, as shown in Fig.3 below. The total solar irradiation of clusters on the first and eighth are significant, and the temperature change trend is similar. Thus, the heat pump is turned on around noon when the heat pump's performance is well. Clusters 2, 3, 4, and 7 have medium solar irradiance and slightly different temperature changes; The second and seventh clusters have low temperatures in the morning, which determines the heat pump operating time based on the actual calculation of the system COP. The fifth and sixth clusters have such small temperature changes, and total solar irradiance that fully uses solar energy cannot meet the users' hot water demand. So, the heat pump needs to be turned on in the morning.

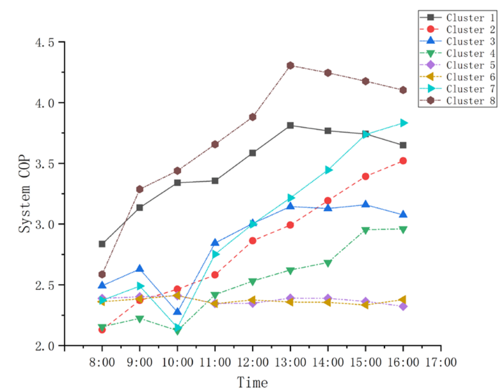

Fig 3 The relationship between heat pump operating time and system COP on a typical day

To ensure the system's highest energy efficiency and best performance, the maximum system COP was set as the target, and the data were summarized to obtain Tab.5.

Tab 5 Heat pump operating time and system COP

Class	Heat pump operating time	System COP
1	13:00	3.811
2	16:00	3.520
3	15:00	3.158
4	16:00	2.960
5	10:00	2.410
6	10:00	2.412
7	16:00	3.832
8	13:00	4.305

The above relationship between heat pump operating time and the system COP needs accuracy verification to discern whether the heat pump operating time obtained from the typical day calculation is satisfied for other days within the clusters. Taking the fourth cluster as an example, most data calculated when reaching the maximum system COP daily is consistent with the heat pump operating time set on a typical day. Furthermore, the individual days have the difference between their operating times within the one-hour interval as shown in Tab.6, with high accuracy and small error, which can further illustrate the reliability of the heat pump operating time on a typical day.

Tab 6 The fourth cluster validation statistics

Number	Date	Heat pump operating time
1	Feb.15	16:00
2	Feb.21	16:00
3	Dec.11	16:00
4	Feb.1	15:00
5	Jan.29	14:00
6	Jan.30	16:00
7	Dec.4	15:00

5. CONCLUSIONS

This paper uses k-means cluster analysis, factor analysis, and Bayes discrimination to cluster winter meteorological data in Shanghai. Furthermore, it selects typical days to analyze meteorological characteristics and determine the heat pump operating time to achieve the maximum system COP. As a result, the following conclusions were obtained.

1) The use of k-means clustering is reliable in analyzing meteorological data for solar-air source heat pump hot water systems.
2) The factors affecting solar-air source heat pump hot water systems were determined by dividing solar irradiation, outdoor temperature as two primary indicators and ten sub-indicators. As a result, the results for clustering are proved reasonable and valid, and the clustering accuracy reaches 98.9%.
3) The clustering data were simulated using TRNSYS software. Moreover, the system COP was used as the optimization target to regulate the operating time of the air-source heat pump. Furthermore, the obtained data were verified by intra-cluster days with high accuracy, which provides a reference basis for future research on energy efficiency improvement of solar and air-source heat pump systems.

REFERENCE

[1] Jiao Junxing. Discussion on the application and development of solar heat pump technology in HV-AC. Building Materials and Decoration, 2018;(47): 179-180.

[2] Jiménez Mares J, Navarro L, Quintero M. C, Pardo M. A Methodology for Energy Load Profile Forecasting Based on Intelligent Clustering and Smoothing Techniques. Energies (Basel). 2020;13(16):4040.

[3] Liu Yuxi. Study on Operating Simulation and Application of Air Source Heat Pump Assisted Solar Water Heating System in Hot Summer and Cold Winter Area. Chongqing University, 2011.

[4] Zeng Naihui, Yuan Yanping, Sun Liangliang, Yao Pan, Deng Zhihui. Meteorological classification of air source heat pump assisted solar water heating system based on clustering analysis. Acta Energies Solaris Sinica, 2017;38 (11): 3067-3076.

[5] GB 50364-2018, Technical standard for solar water heating system of civil buildings. China: Ministry of Housing and Urban-Rural Development, 2018.

[6]GB 50555-2010, Standard for water saving design in civil building. China: Ministry of Housing and Urban-Rural Development, 2010.

[7]GB/T 26973-2011, Specification of solar water heating systems assisted with air-source heat pump (for capacity of water tank more than 0.6m^3). China: Standardization Administration of China, 2011.