J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 3675--3696

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on uncertainty modeling and decision-making system for intelligent energy management of building complexes based on Bayesian networks

Jing Xu[1,✉]

[1] *Department of Architectural Engineering, Bozhou Vocational and Technical College, Bozhou, Anhui, 236800, China*

ABSTRACT

The energy consumption problem of building complexes has become increasingly prominent along with the acceleration of urbanization. In order to achieve efficient energy saving in building complexes, this study proposes a Bayesian network-based uncertainty modeling in decision-making system for energy consumption management. By analyzing the uncertainty factors in the energy consumption data, a Bayesian network model is constructed to predict and analyze the energy consumption. And the uncertainty factors are used as decision variables to construct the energy consumption management decision-making system based on Bayesian network. The experimental results show that the uncertainty model and decision-making system constructed in this paper have more favorable performance compared with other benchmark methods, and exhibit smaller measurement errors in experimental tests. At the same time, the application of this paper's decision-making system for energy consumption management of building complexes can significantly reduce management costs, and obtain the double benefits of reducing energy consumption and saving costs.

Keywords: bayesian networks, uncertainty modeling, decision-making systems, energy management

1. Introduction

In today's society, the energy problem is becoming increasingly severe, and the energy management and optimization of buildings, as a large consumer of energy, has become an important research topic. And with the continuous progress of science and technology, intelligent algorithms bring new ideas and methods for building energy management [1-3].

Green building firstly emphasizes the strategic idea of saving energy, not polluting the environment, maintaining ecological balance, and embodying sustainable development, and its purpose is energy saving and environmental protection [4-5]. Building energy consumption is generally the energy

✉ Corresponding author.
E-mail address: 13605682145@163.com (J. Xu).
Received 25 February 2024; Revised 10 June 2024; Accepted 05 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-205

consumption of a building during the use of the building, including heating, cooling, hot water, cooking, lighting, electrical appliances power consumption, elevator, sewage, cleaning power consumption, etc. [6-9]. Saving energy requires the adoption of new intelligent technologies in green buildings, using the "wisdom" of intelligent systems to minimize energy consumption, and specific applications in intelligent lighting systems, air-conditioning, ventilation, refrigeration, heating intelligent systems, elevator group control systems, etc. [10-13]. And the building energy management system for the core of the building intelligent system, should support the intelligent integration of CNC integration platform technology architecture, combined with computer technology, network technology, communication technology, automatic control technology, all the intelligent, security and energy consumption equipment in the building (or building group) to carry out comprehensive and effective monitoring and management [14-17]. It also enriches the comprehensive use function of the building and improves the efficiency of property management, ensures that all relevant equipment in the building complex is in a highly efficient, energy-saving, and optimal operating state, and thus provides a safe, comfortable, convenient, and efficient working environment for the staff in the building [18-21].

This paper constructs the uncertainty model of building group energy management based on Bayesian network by introducing the Bayesian network model and analyzing the uncertainty factors in building group energy management. The causal relationship between the uncertainty factors is analyzed, and the uncertainty factors in the building model and operation status are taken as the main decision variables to construct the decision-making system of energy consumption management based on Bayesian network. Five typical building clusters are selected as the experimental research objects of this paper, and their electric energy consumption data are decomposed to quantify the uncertainty data in the energy consumption data. The method proposed in this paper is compared with the benchmark methods such as linear regression, random forest, and support vector regression to highlight the superior performance of this paper's model for energy management of building clusters under uncertainty conditions. Using the energy management cost as a utility measure, we compare the changes in the energy management cost of five building complexes before and after using the decision-making system in this paper, to demonstrate the effectiveness of the energy management decision-making system constructed based on Bayesian network in this paper.

2. Overview of Bayesian networks

Bayesian networks have a sound and mature theoretical foundation for probabilistic reasoning [22], which can not only effectively quantify the uncertainty of information and represent the polymorphism of variables, but also intuitively express the complex causal relationships between random variables in the form of graphs, and have been widely used in reasoning and learning in various fields.

2.1. Representation of Bayesian networks

Bayesian network (BN) is one of the important inference models in probabilistic graphical models [23]. Its essence is a directed acyclic graph (DAG) with probabilistic information consisting of nodes and directed edges, denoted as $BN = \{(N, E), \theta\}$. A simple two-layer BN model is shown in Fig. 1.

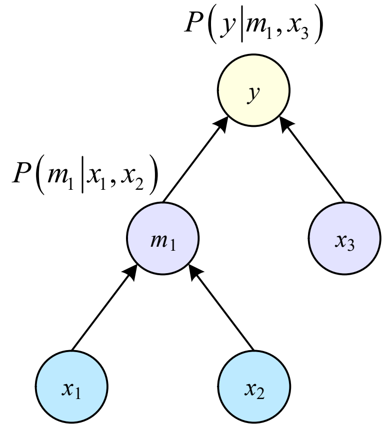

Fig. 1. Bayesian network model

N is a node representing a random variable, and for any node $x_i \in N$, it may have multiple states, i.e., the state space of x_i . $A_{x_j} = \{A_1, A_2, \cdots, A_j\}$ Dependencies between random variables are represented by directed edges E between the nodes, and for a given directed edge $x_i \to x_j$, x_i is said to be the parent node of x_j, and x_j is the child node of x_i. For those nodes that have no parent nodes are called root nodes, those that have no children are called leaf nodes, and the rest are called intermediate nodes.Conditional independence of the nodes in a Bayesian network is naturally embedded in the DAG. Typically, the set of parent nodes of a given node is denoted as $pa(x_i)$ and the set of non-descendant nodes is denoted as $nd(x_i)$, then given $pa(x_i)$, x_i and $nd(x_i)$ are conditionally independent, viz:

$$P(x_i \mid pa(x_i), nd(x_i)) = P(x_i \mid pa(x_i)) \tag{1}$$

The θ in $BN = \{(N, E), \theta\}$ represents the Conditional Probability Table (CPT) defined on the child node, which stores the joint probability distribution of the parent nodes directly associated with that child node, quantifying the dependency between that node and its parent. Therefore, once the prior probability of the root node and the conditional probability tables of the rest of the nodes are given, the joint probability distribution with all nodes included can be computed:

$$P(X) = P(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} P(x_i \mid pa(x_i)) \tag{2}$$

2.2. Inference in Bayesian networks

After completing the construction of the BN topology, the setting of the conditional probability table for each sub-node, and the input of the prior probability for the root node, the Bayesian full probability formula can be utilized to perform the inference layer by layer from the bottom to the top, and the edge probability distribution of the leaf node (the fusion node) can be obtained in the end. Assuming that the state space of leaf node y in the BN model is $A_y = \{A_1^y, A_2^y, \ldots, A_j^y\}$, and the model contains a total of L intermediate nodes m_i and n root nodes x_i, and the state space of each node is $A_{x_1} = \{A_1^{x_j}, A_2^{x_j}, \ldots, A_k^{x_j}\}$, the edge probability that the leaf node is in a certain state A_j is:

$$p\left(y=A_j^y\right)=\sum_{x_1,\ldots x_x,m_1,\ldots,m_L} p\left(x_1,x_2,\ldots,x_n,m_1,m_2,\ldots,m_L,y=A_j^y\right)$$
$$=\sum_{pa(y)} P\left(y=A_j^y \mid pa(y)\right) \tag{3}$$
$$\prod_{i=1}^{L}\sum_{\pi(m_i)} P\left(m_i \mid pa\left(m_i\right)\right)\prod_{i=1}^{n} p\left(x_i=A_k^{x_i}\right)$$

where y is the leaf node in the BN model. $x_{i=1,..,n}$ is each root node. $m_{l=1,..,L}$ is each intermediate node. $A_k^{x_i}$ indicates that the ith root node is in the kth state in the state space. $P(y \mid pa(y))$ and $P\left(m_l \mid pa\left(m_l\right)\right)$ are conditional probability tables storing the joint probability distributions between these child nodes and the directly associated parent node $pa(\cdot)$, reflecting the dependencies between the nodes.

2.3. Learning in Bayesian networks

The parameter learning tool used in this paper is Bayesian estimation [24]. Let a Bayesian network contain n nodes (including root, intermediate and leaf nodes), and for convenience, the set of their states is denoted by $X=\left\{x_1,x_2,\ldots,x_j,\ldots,x_n\right\}$, where x_j denotes the current state of the jth node. Based on the premise of independent homogeneous distribution, if this Bayesian network is observed m times at different times, a sample set $D=\left\{D_1,D_2,\ldots,D_m\right\}$ will be obtained, where $D_i=\left\{x_1^i,x_2^i,\ldots,x_n^i\right\}$ is the set of states of all nodes at the ith observation.

In parameter learning for BN using Bayesian estimation, θ in $BN=\{(N,E),\theta\}$ is treated as a random variable, such that the prior knowledge about θ is a probability distribution. And what is to be computed is the posterior probability distribution $P(\theta \mid D)$ for θ after $D=\left\{D_1,D_2,\ldots,D_m\right\}$ is observed.

When the prior probability distribution and the posterior probability distribution belong to the same class, the prior distribution and the posterior distribution are said to be conjugate distributions. If the prior and the posterior are conjugate distributions, the posterior distribution can be easily calculated when the prior distribution is known. Because of this property, the prior probability distribution is often assumed to be a Dirichlet distribution in Bayesian parametric learning. The Dirichlet distribution is a probability distribution of a multivariate continuous random variable, which is conjugate to the multinomial distribution and is often used as the prior probability distribution of the multinomial distribution. Therefore, it is assumed that the distribution of the parameter θ to be estimated, $P(\theta)$, is a Dirichlet distribution:

$$P(\theta)=Dir\left(\theta \mid \alpha_1,\ldots,\alpha_r\right)=\frac{\Gamma(\alpha)}{\prod_{k=1}^{r}\Gamma\left(\alpha_k\right)}\prod_{k=1}^{r}\theta_k^{\alpha_{n-1}} \tag{4}$$

where $\alpha=\sum\alpha_k$; $\alpha_k>0$, $k=1,\ldots,r$; $\alpha_1,\ldots,\alpha_r$ are hyperparameters of the Dirichlet distribution for a priori pseudocounts. $\Gamma(\cdot)$ is the gamma function satisfying $\Gamma(x+1)=x\Gamma(x)$. Thus, the probability of observing the sample is:

$$P(D)=\int p(\theta)p(D\mid\theta)d\theta=\int\frac{\Gamma(\alpha)}{\prod_{k=1}^{r}\Gamma(\alpha_k)}\prod_{k=1}^{r}\theta_k^{\alpha_k-1}\times\prod_{k=1}^{r}\theta_k^{v_k}d\theta \\ =\frac{\Gamma(\alpha)}{\Gamma(\alpha+M)}\prod_{k=1}^{r}\frac{\Gamma(\alpha_k+M_k)}{\Gamma(\alpha_k)} \tag{5}$$

In Bayesian estimation methods, in order to comply with the maximum entropy principle, in the absence of prior distributions, it is generally assumed that the prior of the parameters obeys a uniform distribution. Specifically, if there is no information about the prior distribution of the parameters, the hyperparameters are set $\alpha_1=\alpha_2=\cdots=\alpha_r$. In this section, the Dirichlet distribution is used as the prior distribution of the parameters, and under the assumption that the parameters are independent, then each variable x_i and its parent variable $pa(x_i)=j$ obey the following Dirichlet distribution:

$$P(\theta\mid D)=\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_{ij}}\theta_{ijk}^{m_{jk}+\alpha_{jk}-1} \tag{6}$$

where θ_{ijk} denotes the conditional probability when $x_i=k$ and $pa(x_i)=j$. Thus, the value of θ_{ijk} can be calculated using the following expression:

$$\theta_{ijk}=\frac{\alpha_{ijk}+m_{ijk}}{\sum_{k=1}^{r_i}\left(m_{ijk}+\alpha_{ijk}\right)} \tag{7}$$

where m_{ijk} is the frequency of occurrence of node x_i in sample set D when $x_i=k$ and $pa(x_i)=j$ are present. α_{ijk} is a hyperparameter that serves as an additional virtual sample set D' to the original sample set D. After combining D and D', the value of θ_{ijk} can be computed by means of maximum likelihood estimation. In particular, if the samples in sample set D are weighted, the frequency of eligible samples is calculated by multiplying the corresponding weights.

3. Analysis of uncertainties in building stock energy management

Building stock energy management is usually used to predict building stock energy consumption and energy saving potential fini, but due to the existence of many uncertainties in building energy use activities, some energy parameters have random, fuzzy, and uncertain properties, and there are still some obstacles in the current building stock energy management in simulating and calculating the uncertainty of the energy management of the existing buildings. The key to study the uncertainty of building energy management lies in the method of quantifying the uncertainty parameters. In this paper, through the field investigation and the study of related literature, we organize to get the uncertainty in building complex energy management mainly includes the following aspects.

3.1. Uncertainty in building physical modeling

First, the actual meteorological conditions around the building (e.g., dryness, humidity, temperature, solar radiation, wind speed, etc.) deviate from the average meteorological parameters in the local meteorological databases used in energy management, and the actual meteorological parameters are often characterized by random fluctuations. Secondly, it is difficult to fully align the physical model of the building with its own physical form structure. Finally, because the building is in a dynamic state of operation will be affected by the fluctuation of some random parameters, the actual operating characteristics of the system are different from those in the standard or architectural design, such as the use of the building, occupancy and operating hours, etc., which is not a fixed mode of operation. Sources of uncertainty in some of the thermophysical parameters of the building itself lead to uncertainty in the predicted results. Asynchrony between the actual operating characteristics of different system types within the building. These factors lead to different outputs from the simulation of the energy management model than expected, and the prediction of the energy consumption of the regional building stock can deviate from the actual situation due to these uncertainties.

3.2. Uncertainty of operational status

Scenario uncertainty in building energy management refers to the asynchrony between the external environment of the building, the building's operational status and the operation of the energy system, as well as the mismatch with the actual status. The regional building cooling and heating loads are highly correlated with the actual energy demand, and parameters such as regional meteorological parameters, hour-by-hour room occupancy rate of building personnel, energy use habits of building personnel, and hour-by-hour utilization rate of equipments have stochastic attributes. Some energy parameter input conditions for building energy simulation are fixed, but the actual data of these parameters have random fluctuation compared with the data used in the model. Energy modeling parameters that fall into this category include the thermal properties of building envelope materials, internal heat sources, HVAC system property parameters, and their operation and control setup parameters. These parameters often deviate from their design and operation specifications and model settings in real buildings because current energy simulation software cannot set dynamic and randomly fluctuating parameter attributes for a given energy parameter boundary condition. In addition, the efficiency of energy systems and equipment gradually decreases over their life cycle, increasing the level of uncertainty in energy management.

3.3. Uncertainty in the model itself

Model uncertainty comes from the discrepancy between the model itself and the actual state. The modeling inadequacy is reflected in the possible selection of different energy management models, which have advantages, disadvantages and differences in predicting energy consumption. Various building energy management models approximate the physical heat transfer phenomena occurring in buildings by abstracting complex phenomena into simplified models. At present, although there are some high simulation single building energy management models, for regional building clusters, there is no relevant software that can simulate and predict the complex regional building cluster energy consumption problems so far. Therefore, in order to establish a regional building energy management model, it is necessary to simplify the model for regional building groups, in which some factors with weak influence will be weakened or ignored, and the high-precision model obviously describes the actual state more accurately than the low-precision model. On top of the traditional building energy consumption management model, quantitative uncertainty analysis is performed based on the actual engineering experience equations to make up for the shortcomings of the model in these aspects and quantify the uncertainty range of the model.

4. Bayesian network-based modeling of energy consumption uncertainty in building clusters

The Bayesian network structure uses a visualization network to condense to represent the influence relationship between uncertainty factors. Using GeNIe software, the uncertainty factors in the energy management of building complexes are set up as network nodes, including outdoor meteorological parameters, thermal performance parameters of the envelope, and statistical parameters of the building envelope in the building model, cooling and heating conformity, fresh air loads, and operating conditions and states in the operational state, as well as uncertainty factors such as simplification of the energy calculation, assumptions of the model boundaries, and the differences between the model and the actual building in the model . The causal relationship between each network node is analyzed through expert knowledge to form a chained mesh graphical structure with transfer relationship. And according to the published cutting-edge research results to increase or decrease the link between the network nodes, to amend the network structure, after standardizing the data, to get the conditional probability distribution of the network nodes, to combine the expert knowledge with and its learning, to guarantee the scientific and accuracy of the Bayesian network.

4.1. Bayesian network model data processing

In this paper, the Bayesian model is constructed with the uncertainty factors in the energy management of building complexes, experts are invited to rate the selected uncertainty factors, and the data are standardized for the expert ratings, and the standardized data are divided into two-dimensional hierarchical levels by using the theory of the risk matrix method, with the horizontal coordinate representing the reasonableness of the uncertainty factors and the vertical coordinate representing the importance of the uncertainty factors. The 5*5 matrix method defined by the International Organization for Standardization was used to construct the evaluation matrix of uncertainty factors, and the data were divided into three levels, F1, F2 and F3, from low to high. The matrix form is shown in Figure 2, when the sum of the two dimensional scores is less than or equal to 6, the uncertainty factor is in the F1 interval, that is, the factor is F1, and all the uncertainty factor data are standardized in turn, and the rank matrix scores of each uncertainty factor are shown in Table 1.

Table 1. Standardized data state statistics of uncertainty factors

Uncertainty factor	F1	F2	F3	Total
Outdoor meteorological parameters(X1)	12	36	52	100
Thermal performance parameters of containment structure(X2)	0	21	79	100
Containment structure statistics parameters(X3)	19	64	17	100
Building shape coefficient(X4)	11	24	65	100
Building usage situation(X5)	2	38	60	100
Cold load(X6)	49	51	0	100
Heat load(X7)	0	3	97	100
Fresh air load(X8)	0	29	71	100
Operating condition and condition(X9)	5	82	13	100
Control and setting parameters(X10)	22	75	3	100
HVAC system(X11)	1	43	56	100
Lighting system(X12)	29	38	33	100
Simplified energy consumption calculation(X13)	0	42	58	100
Model boundary hypothesis(X14)	9	73	18	100
Model and actual building differences(X15)	63	12	25	100

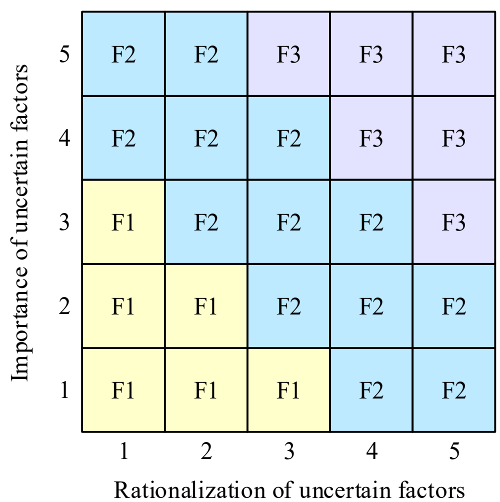

Fig. 2. Risk model matrix form

4.2. *Bayesian network structure learning*

The correlation and independence between the nodes are counted to determine the relationship between the nodes. When there is a correlation between the nodes, there is an edge between the nodes and vice versa there is no edge. Fisher'Z test is used for continuous sample data and χ^2 test or G test is used for discrete sample data. The second category is the search scoring method, which optimizes the model according to the search algorithm and scoring criterion, the scoring function measures the fit of the sample data, the higher the score the better the network structure. Commonly used scoring functions are minimum description length scoring function, Bayesian scoring function. Commonly used search algorithms are K2 algorithm, BDeu scoring algorithm.

The Bayesian network model construction software used in this paper is GeNIe, which is easy to operate, has a simple and intuitive interface, has a variety of built-in inference algorithms, and supports the import of a variety of data formats, such as the commonly used Excel saved csv. format, which makes data analysis more convenient. Using GeNIe software for network construction and analytical reasoning, the processed data are shown in Table 2.

Table 2. Expert grading standardization

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
3	3	1	2	3	1	2	3	2	2	2	3	3	3	2
3	3	2	3	3	1	2	3	2	2	2	3	2	2	3
3	3	2	3	3	2	3	2	1	2	2	3	2	2	2
3	2	1	3	3	1	2	2	2	3	1	3	3	3	2
3	2	1	2	2	1	2	2	2	2	2	2	3	3	3
3	3	2	2	2	2	2	3	2	2	1	2	2	2	3
2	2	2	2	2	2	2	3	2	3	2	3	3	3	2
3	3	1	3	3	1	2	2	2	2	2	2	2	2	2
3	2	1	3	3	1	3	2	2	2	2	3	3	3	3
2	3	2	3	2	2	2	2	1	3	1	3	2	2	2
3	3	2	3	3	1	2	2	2	3	2	3	3	2	2
3	2	1	2	2	1	2	2	1	2	2	2	3	3	2
2	2	2	2	3	2	2	3	1	2	2	3	3	3	2
3	2	2	3	3	2	3	3	1	2	2	3	2	2	3
2	2	1	2	2	1	2	3	1	2	1	3	2	2	2
3	2	2	3	3	1	2	2	2	2	1	2	3	2	3
3	3	1	3	2	2	3	3	2	3	2	2	2	2	2
2	3	2	2	2	1	3	2	2	2	2	3	3	3	2
3	2	1	2	2	2	3	3	1	2	1	2	3	2	3
3	3	2	3	3	1	3	2	1	3	1	3	3	2	3
2	2	2	2	2	2	3	2	2	2	2	2	2	3	2
2	2	1	2	2	2	3	3	1	2	2	3	3	3	2
3	2	1	3	2	2	3	3	1	2	1	2	2	3	3
2	3	1	2	3	2	3	3	2	2	1	3	3	3	3
2	3	2	3	2	2	3	2	1	2	2	3	3	2	3

In order to improve the modeling efficiency and simplify the network structure, it is necessary to use expert knowledge to initially determine the causal relationship between uncertainty factors. After exporting the data to the software and generating Bayesian nodes to determine the level of uncertainty factors, the building model, the operation status, and the model's own factors are distributed to the corresponding time line, and the factor relationship pointing function that comes with the application software preliminarily determines the causality between the factors and the factors that don't have a relationship, and they are connected by a directed arrow line. The preliminary structure learning network model is shown in Figure 3, the relationship between the factors in the network structure and the reality of the indicators may be different, mainly because the machine learning algorithms run, usually can only find the Bayesian network structure of the same level of the category of a factor, can not accurately determine the causal relationship of the undirected edges, resulting in the analysis of the results of the limitations. It is necessary to optimize the network model in combination with expert recommendations to remove the arrows that obviously do not have a relationship. The final composition is a cyclic network structure with the modeling of uncertainty factors in energy management of building complexes as the purpose of the study, and the uncertainty factors in the building model, the operation state and the model itself as the variables.

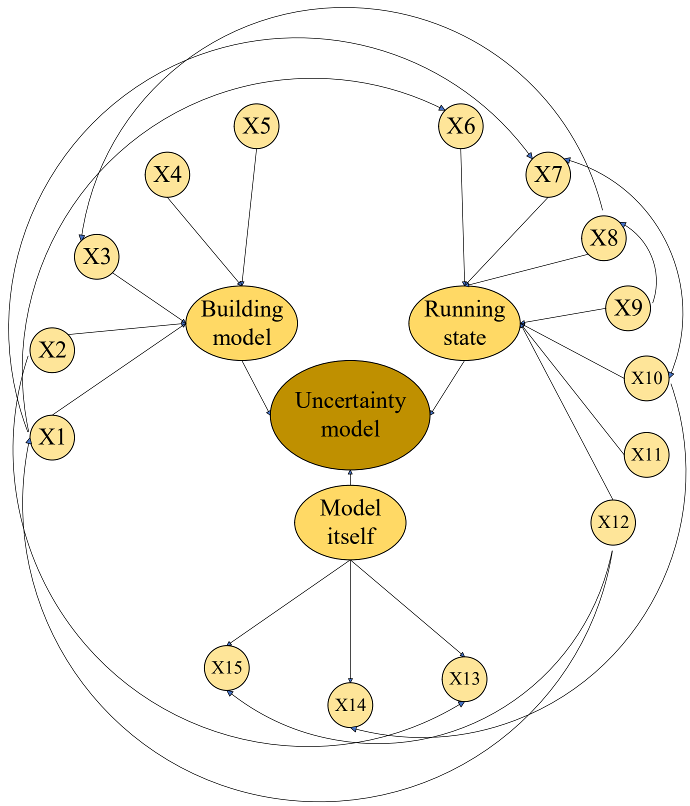

Fig. 3. Bayesian network structure learning model

4.3. Bayesian network parameter learning

After determining the structure of the Bayesian network, it is necessary to learn the parameters of each node in the network, and clarify the quantitative relationship of the interconnection between the factors. In this paper, we use the questionnaire to collect the completed data to provide data support for the probability distribution of the nodes, with the help of Bayesian estimation to represent the conditional probability learning problem in the Bayesian network.

Bayesian network parameter learning is divided into two main stages:

1) Add the initial probability of node variables. The original factor state has two state columns, manually add a line of state columns, named F1, F2, F3 from top to bottom, set the initialization probability of each node are set to 1/3, there are associated factors will automatically update the state columns to establish a new link.

2) Associate the state with standardized data. Import the standardized processed csv.format data, the software will automatically associate the status with the data and updates. Manually match the status bar with the numbers, F1 status with 1, F2 status with 2, F3 status with 3. After determining the data for parameter learning, the result is shown in Figure 4.

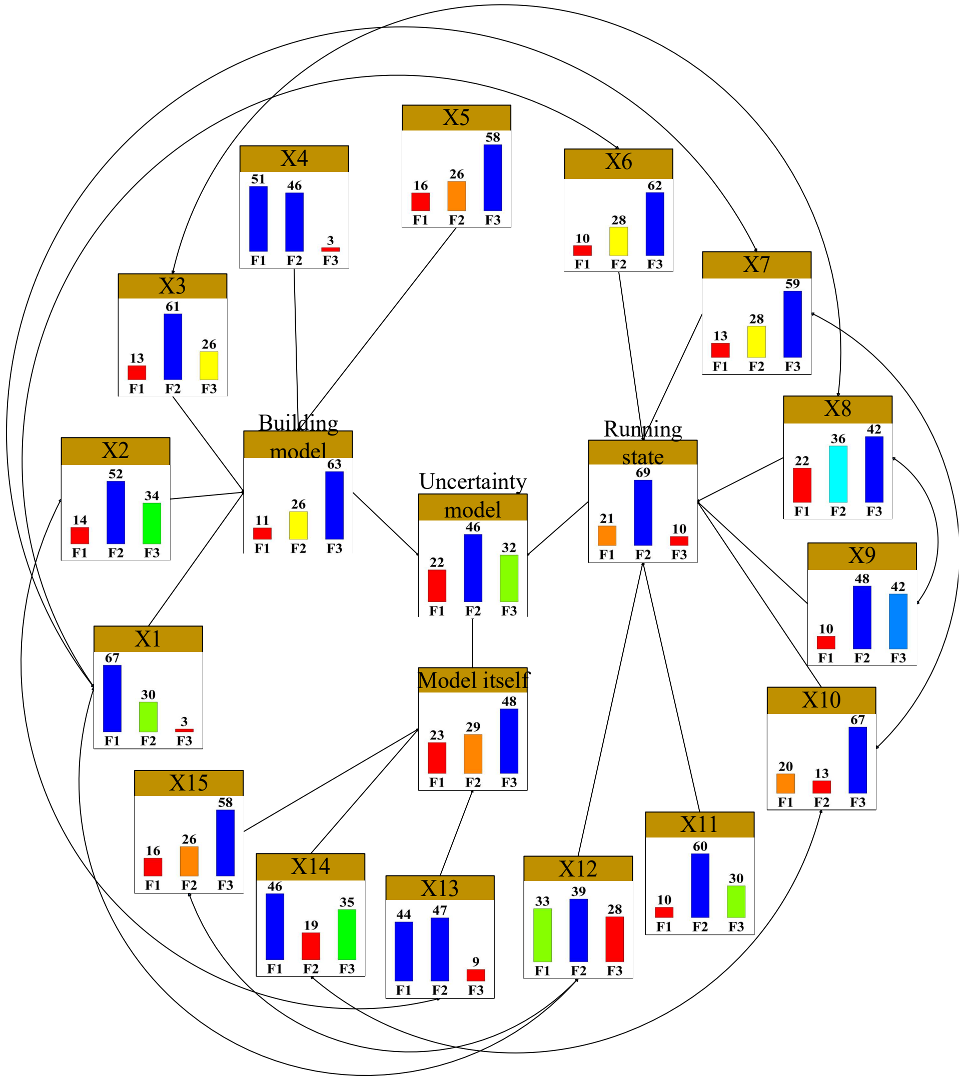

Fig. 4. Bayesian network parameter learning model

5. Bayesian network-based decision-making system for energy consumption management

Decision analysis discusses two main roles, the decision maker and the decision analyzer. The decision maker gives preferences for different outcomes and the decision analyzer needs to enumerate the possible actions and outcomes and then get the preferences from the decision maker to decide the best course of action. With the introduction of Bayesian networks, it has become possible to build systems that generate reliable probabilistic reasoning from evidence. In this paper, the elements and theory of the decision-making system for energy management in a building complex are shown in Fig. 5.

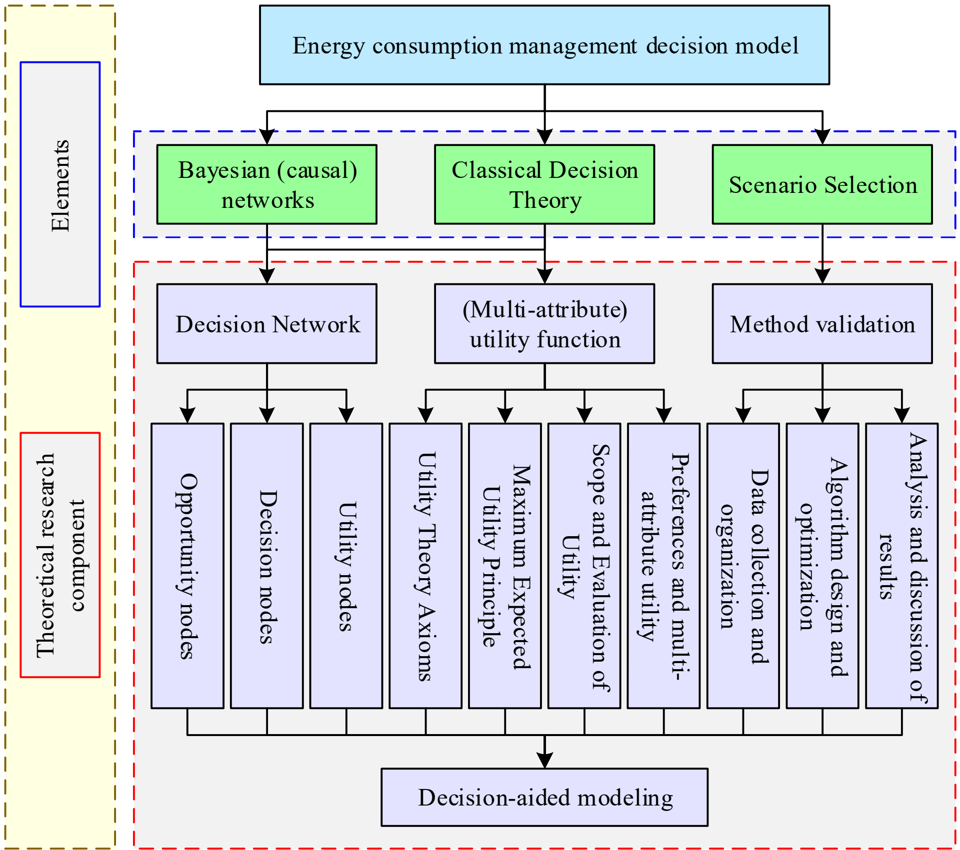

Fig. 5. Energy consumption management decision model

In this paper, the decision-making system is constructed through the following steps:

1) Create a causal model. Identify the decision variables for energy management and then draw arcs between them indicating what change in one variable causes a change in another and how that change can be mitigated by the program.

2) Simplify to a qualitative decision model. Since this model is used in this paper to formulate programmatic decisions, it is able to simplify the model by removing variables that are not involved in programmatic decisions.

3) Assigning probabilities. Probabilities can come from data sets, literature studies, or subjective assessments by experts. Assessing the probability of an outcome given a cause condition can utilize statistical methods.

4) Assigning utility. Create a range from the best to the worst outcome and give each outcome a numerical value, then place the outcomes within that range as well. If there are exponentially many outcomes, combine them using a multivariate utility function.

5) Validate and improve the model. In order to evaluate the system, a set of correct decision results is needed.

6) Perform sensitivity analysis. This step tests which variables the model is sensitive to small changes in.

5.1. Decision-making networks

Decision networks combine Bayesian networks with actions and additional node types for utility. A decision network represents the current state of the decision maker, its possible actions, the states that the decision maker's actions can produce, and the utility of the states. Three types of nodes are included:

Chance nodes: represent random variables, as represented in a Bayesian network. Each chance node is linked to a conditional distribution that is indexed by the state of the parent node.

Decision node: represents the node at which the decision maker has a choice of action. It is assumed that this paper deals with a single decision node.

Utility node: represents the utility function of the decision maker. The utility node takes as parent nodes all the variables that can be described by the outcome state that directly affects the utility.

5.2. Probability distribution

The decision network is constructed on the basis of the energy management that has already been implemented, with the aim of making an analysis, providing appropriate solutions and evaluating them for the next energy management. In this paper, only uncertainty factors are considered as decision variables, which are determined by information from field surveys, reviewing construction design drawings, and asking residents around the building complex.

In this paper, only the probability and utility assignments of the two uncertainty decision variables, building model and operation status, are considered. Firstly, the probability allocation is carried out, and the probability can come from database, literature research or subjective estimation of experts, and this paper adopts the method of data statistics. The probability of occurrence of each uncertainty factor in the building model in the data set is obtained as shown in Fig. 6. For example, the probability of occurrence of outdoor meteorological parameter factor among each uncertainty factor in all case data is 25.75%, which is obtained statistically in turn.

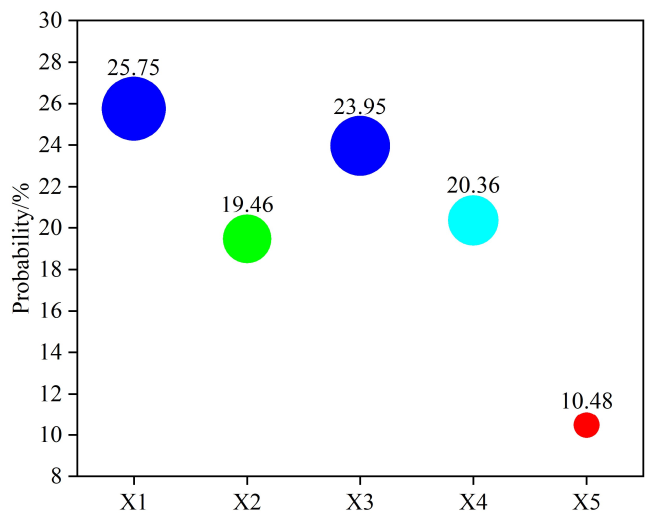

Fig. 6. Building model probability

The probability of occurrence of each uncertainty factor in the operating condition is shown in Fig. 7. For example, the probability of occurrence of the uncertainty factor of cold load is 15.34%, which is obtained statistically in order.

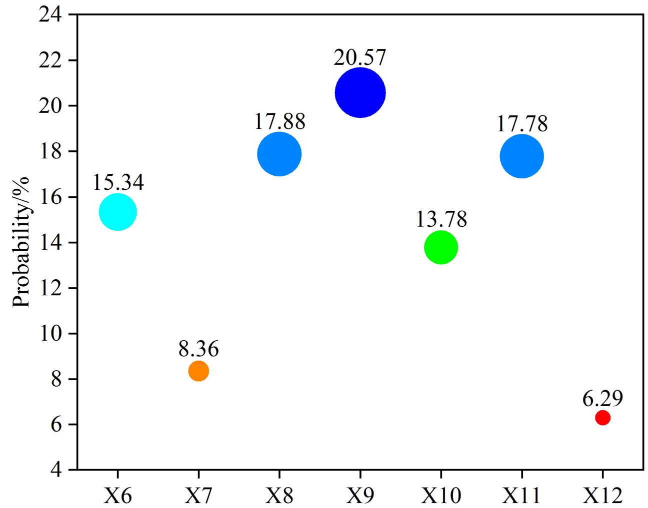

Fig. 7. Operating status probability

Next, utilities are assigned, and when the number of possible outcomes is small, they can be enumerated or evaluated individually. The utility is assigned to each scenario under the two variables of building model and operational status, and a utility table is created, where the utility values are used to calculate the expected utility of the energy management scenarios. And the preferences need to be obtained from the decision maker.

5.3. Utility Functions

In this paper, the maximum expected utility calculation for multivariate problems with preferences includes two cases.

1) Containing deterministic preferences (with preferences). This paper begins with the deterministic case; for a deterministic environment, the decision maker has a value function U. The fundamental regularity that arises from the deterministic preference structure is known as preference independence. If variable $S_l,\cdots,S_n$ is mutually preference independent, the preference behavior of this decision maker can be described as maximizing a function of equation (8):

$$U = \sum_i U_i \tag{8}$$

Each U_i is a value function involving only variable S_i, called an additive value function. An additive function is an extremely natural way to describe a value function and is valid in many real-world situations.

2) Preferences that involve uncertainty (no preferences). If uncertainty arises, utility independence is involved, and the basic concept extends preference independence to cover probability distributions of outcomes: if preferences between probability distributions of outcomes are independent of the specific values of the variables in variable set S_2 on the variables in variable set S_l, then variable set S_1 is utility independent with respect to variable set S_2. Mutual utility independence is satisfied when each of its subsets is utility independent of the rest of the variables, implying that The behavior of the decision maker can be described by the multiplicative utility function equation (9):

$$\begin{aligned} U = {} & k_1U_1 + k_2U_2 + k_3U_3 + k_1k_2U_1U_2 + k_2k_3U_2U_3 \\ & + k_3k_1U_3U_1 + k_1k_2k_3U_1U_2U_3 \end{aligned} \tag{9}$$

In the formula U_i represents the utility of each variable and the constant k_i represents the degree of preference for the variable, taking values between 0 and 1. It contains only single-variable utility functions and constants, each single-variable utility function can be developed independently of the others, and this combination will be guaranteed to produce the correct overall preference.

6. Experiments and analysis of results

In this section, the paper describes the dataset, data preprocessing, decomposing the data, the proposed method and comparing the methods, and finally dissecting the experimental results to confirm that the method in this paper has higher accuracy.

6.1. Experimental data and processing

6.1.1. Experimental data. The dataset used for the experiments in this paper collects meter data from 1748 building complexes covering a two-year time period, 2022 and 2023, where the frequency of electricity, heating, cooling water, steam, and irrigation water is measured on an hourly basis. In this section of the study, two years of hourly electricity consumption data for five different buildings were analyzed. Table 3 shows the architectural information of the five buildings, where Min, Max, Mean and Std denote the minimum energy consumption, maximum energy consumption, average energy consumption and standard deviation of energy consumption in KW, respectively, and Fig. 8 shows the hourly purchased power consumption curves for each building complex, which shows different power consumption curves for each complex, presenting an uncertainty and non-linear pattern.

Table 3. Building information

Name	Type	Area(Ft2)	Min	Max	Mean	Std
Wolf	College classroom	73864	10.12	120.17	53.46	13.82
Bull	Public services	65438	3.47	153.24	39.48	12.97
Robin	College dormitory	63962	0.29	68.93	32.64	11.46
Fox	Office	25701	0.48	84.32	28.93	14.35
Rat	Public services	26987	0.37	97.89	35.17	11.52

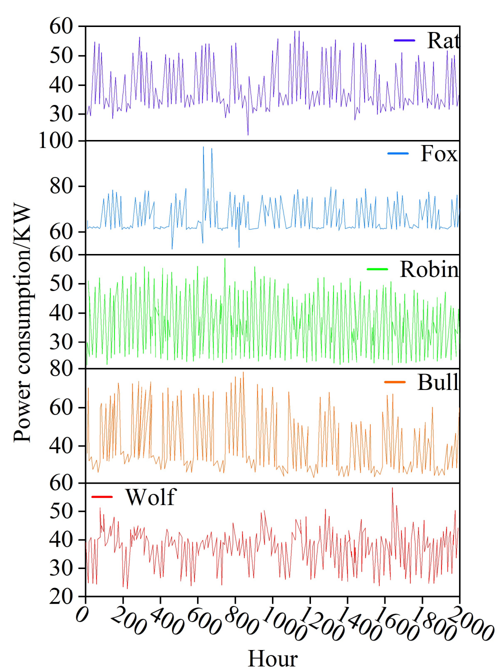

Fig. 8. The amount of electricity used in five different buildings

6.1.2. Data processing. In real-world application scenarios, data is dirty and all kinds of dirty data are mixed in. Therefore, data cleaning is crucial. In the scenario studied in this paper, the missing values, outliers and duplicates that may exist in the building energy consumption data are cleaned to ensure the completeness and accuracy of the data. After the data processing is completed, 70% of the data from each building is used as a training set to train the model, and the remaining 30% is used as a test set to test the performance of the model.

6.2. Data disaggregation

In the data decomposition stage of this research method, this study uses a Bayesian network based uncertainty model and decision system to decompose the raw data into multiple intrinsic mode functions (IMFs) and one residual, Figure 9 shows the result of Robin decomposition using this paper's method, decomposed into individual IMF components and one residual. It is easy to see that the raw data of this building can be more linear and smooth. It is easy to see that the regularity of the components is more obvious compared to the original data, which improves the effectiveness of energy management.

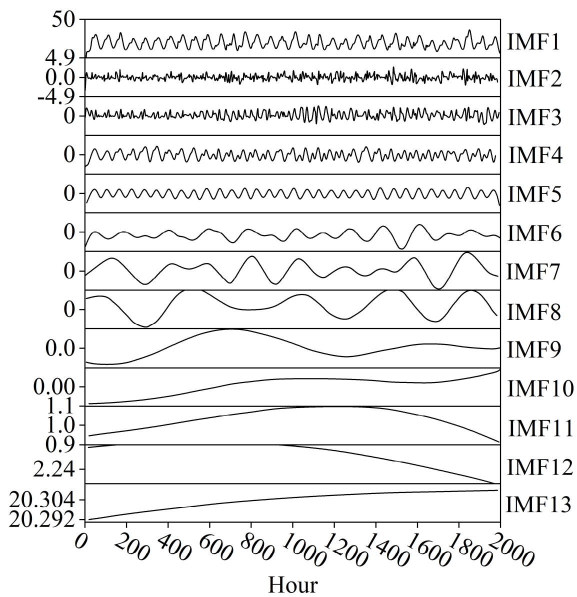

Fig. 9. Robin's decomposition result

6.3. Experimental results

In this section of the study, the Bayesian network based uncertainty modeling and decision making system is used in building stock energy management and compared with other benchmark methods such as Linear Regression (LR), Random Forest (RF), Support Vector Regression (SVR), Long and Short-Term Memory Network (LSTM), Informer, PSO-Informer, EEMD- Informer. Mean absolute percentage error (MAPE), mean square error (RMSE), and mean absolute error (MAE) were used as evaluation metrics to measure the performance of the model, and Table 4 shows the calculation results of different methods. The bold font on each building group indicates the best result for each criterion, and according to the information in the table, it can be obtained that this paper's method has higher accuracy and lower error. Taking the building Wolf as an example, the MAPE, RMSE and MAE of this paper's method are reduced by 13.91%, 9.42%, and 9.80%, respectively, compared to the next best performing EEMD-Informer model. And compared to the worst-performing LR model, this paper's method reduces 67.71%, 69.68%, and 67.21% in the three metrics, respectively, showing advanced performance superiority and accuracy.

Table 4. Performance results of different methods(%)

Name	Index	LR	SVR	RF	LSTM	Informer	PSO-Informer	EEMD-Informer	Proposed Method
Wolf	MAPE	13.04	13.61	12.48	10.83	8.74	8.19	4.89	**4.21**
	RMSE	9.83	8.31	8.57	8.13	6.89	6.71	3.29	**2.98**
	MAE	6.74	5.27	5.79	5.28	4.67	4.29	2.45	**2.21**
Bull	MAPE	15.48	9.84	10.79	9.98	8.69	8.57	9.02	**8.37**
	RMSE	9.27	6.08	6.59	6.47	5.89	5.81	4.67	**4.36**
	MAE	6.39	3.74	4.24	3.83	3.29	3.31	3.06	**2.83**
Robin	MAPE	87.48	84.36	83.58	79.86	29.57	27.14	20.28	**17.32**
	RMSE	5.36	5.28	4.98	4.89	3.85	3.79	2.98	**2.76**
	MAE	3.19	2.94	2.99	2.83	2.59	2.51	2.17	**1.96**
Fox	MAPE	21.96	15.31	20.03	14.76	10.92	10.73	8.69	**7.32**
	RMSE	8.47	5.79	6.38	5.29	4.67	4.58	3.47	**3.18**
	MAE	4.76	3.58	4.23	3.45	2.89	2.87	2.23	**2.01**
Rat	MAPE	69.78	63.71	58.47	62.38	49.36	47.48	29.84	**25.37**
	RMSE	8.94	8.39	8.36	8.27	6.48	6.39	3.71	**3.58**
	MAE	5.49	5.34	5.73	5.66	4.38	4.15	2.49	**2.17**

In order to more clearly indicate the degree of improvement of the methods in this paper, three percentage improvement metrics are used in this study. The percentage of improvement in MAPE, RMSE and MAE between the two methods can be calculated as respectively:

$$P_{MAPE}=\left|\frac{MAPE_1-MAPE_2}{MAPE_1}\right| \tag{10}$$

$$\mathrm{P}_{RMSE}=\left|\frac{RMSE_1-RMSE_2}{RMSE_1}\right| \tag{11}$$

$$P_{MAE}=\left|\frac{MAE_1-MAE_2}{MAE_1}\right| \tag{12}$$

Table 5 shows the percentage improvement of this paper's method compared to other benchmark methods. Based on the results of the model runs listed in Table 4 and Table 5 (PM for Proposed Method), it can be concluded that the model in this paper performs better than other benchmark methods. This indicates that the model in this paper is able to decompose uncertain and nonlinear data into deterministic smooth data that is more suitable for model training, which can improve the performance of the model.

Table 5. Percentage improvement of this method compared to other methods

Name	Index	PM & LR	PM & SVR	PM & RF	PM & LSTM	PM & Informer	PM & PSO-Informer	PM & EEMD-Informer
Wolf	P_{MAPE}	67.71	69.07	66.27	61.13	51.83	48.6	13.91
	P_{RMSE}	69.68	64.14	65.23	63.35	56.75	55.59	9.42
	P_{MAE}	67.21	58.06	61.83	58.14	52.68	48.48	9.80
Bull	P_{MAPE}	45.93	14.94	22.43	16.13	3.68	2.33	7.21
	P_{RMSE}	52.97	28.29	33.84	32.61	25.98	24.96	6.64
	P_{MAE}	55.71	24.33	33.25	26.11	13.98	14.50	7.52
Robin	P_{MAPE}	80.2	79.47	79.28	78.31	41.43	36.18	14.60
	P_{RMSE}	48.51	47.73	44.58	43.56	28.31	27.18	7.38
	P_{MAE}	38.56	33.33	34.45	30.74	24.32	21.91	9.68
Fox	P_{MAPE}	66.67	52.19	63.45	50.41	32.97	31.78	15.77
	P_{RMSE}	62.46	45.08	50.16	39.89	31.91	30.57	8.36
	P_{MAE}	57.77	43.85	52.48	41.74	30.45	29.97	9.87
Rat	P_{MAPE}	63.64	60.18	56.61	59.33	48.60	46.57	14.98
	P_{RMSE}	59.96	57.33	57.18	56.71	44.75	43.97	3.50
	P_{MAE}	60.47	59.36	62.13	61.66	50.46	47.71	12.85

6.4. Analysis of the effectiveness of decision-making systems

The Bayesian network-based energy management decision-making system is applied to the experimental building complex to compare the energy costs before and after the implementation of the decision-making system. The changes in the energy management costs of the five buildings before and after the implementation of the decision are expressed in 2023 and 2024, respectively, as shown in Fig. 10. From the figure, it can be seen that the energy management costs of each building cluster show a significant decrease after using the energy management decision-making system proposed in this paper. Among them, the average value of the energy management cost of the Fox building cluster decreased from 9.88 million yuan to 5.53 million yuan, with the largest decrease of 78.66%. The energy management costs of the other building clusters all showed decreases of different magnitudes. Meanwhile, through the monitoring of indoor environmental quality, it is found that the indoor temperature, humidity and other indicators are kept within the comfort range, which meets the requirements of personnel comfort and equipment operation. This shows that the Bayesian network-based energy management decision modeling system can effectively reduce the energy cost and obtain the maximum utility while ensuring the normal use function of the building and the indoor environmental quality.

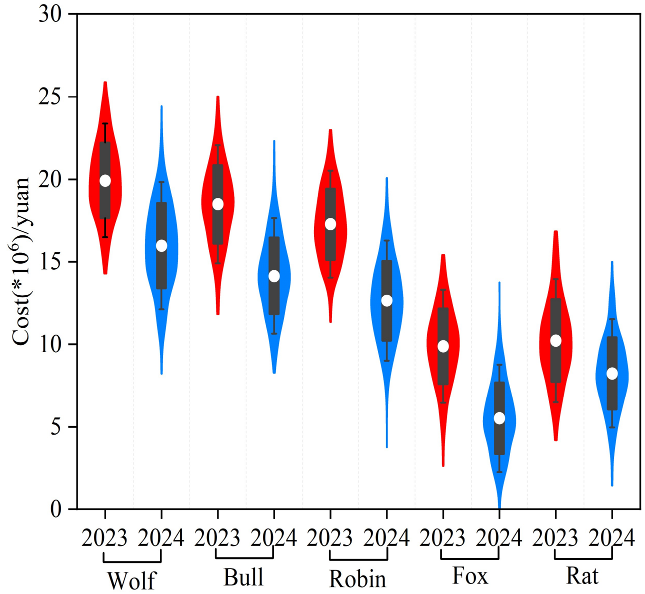

Fig. 10. Decision system application effect

7. Conclusion

In this study, uncertainty modeling based on Bayesian network is proposed for the uncertainty problem in intelligent energy management of building complexes. The Bayesian network model is constructed by analyzing the uncertainty factors in the energy management of building complexes to realize the scientific decision-making of energy management of building complexes. The results of the experimental research show that the method of this paper has significant performance improvement in the management of energy consumption of building complexes compared with other aspects. And after adopting the decision-making system constructed in this paper, the maximum reduction of the cost of energy consumption management of the building complex can reach 78.66%. Effectively meet the requirements of this paper's decision-making system of energy consumption management utility, for the intelligent development of building group energy consumption management to provide a scientific and useful reference.

References

[1] Chen, Z., Xiao, F., Guo, F., & Yan, J. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. Advances in Applied Energy, 9, 100123.

[2] Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T., & Guan, X. (2021). A review of deep reinforcement learning for smart building energy management. IEEE Internet of Things Journal, 8(15), 12046-12063.

[3] Molina-Solana, M., Ros, M., Ruiz, M. D., Gómez-Romero, J., & Martín-Bautista, M. J. (2017). Data science for building energy management: A review. Renewable and Sustainable Energy Reviews, 70, 598-609.

[4] Olubunmi, O. A., Xia, P. B., & Skitmore, M. (2016). Green building incentives: A review. Renewable and Sustainable Energy Reviews, 59, 1611-1621.

[5] Sharma, M. (2018). Development of a 'Green building sustainability model'for Green buildings in India. Journal of cleaner production, 190, 538-551.

[6] Li, C., Ding, Z., Zhao, D., Yi, J., & Zhang, G. (2017). Building energy consumption prediction: An extreme deep learning approach. Energies, 10(10), 1525.

[7] Seyedzadeh, S., Rahimian, F. P., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: a review. Visualization in Engineering, 6, 1-20.

[8] Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., ... & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews, 82, 1027-1047.

[9] Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews, 74, 902-924.

[10] Vandenbogaerde, L., Verbeke, S., & Audenaert, A. (2023). Optimizing building energy consumption in office buildings: A review of building automation and control systems and factors influencing energy savings. Journal of Building Engineering, 107233.

[11] Wang, X., Feng, W., Cai, W., Ren, H., Ding, C., & Zhou, N. (2019). Do residential building energy efficiency standards reduce energy consumption in China?–A data-driven method to validate the actual performance of building energy efficiency standards. Energy Policy, 131, 82-98.

[12] Yang, X., Zhang, S., & Xu, W. (2019). Impact of zero energy buildings on medium-to-long term building energy consumption in China. Energy Policy, 129, 574-586.

[13] Ashouri, M., Fung, B. C., Haghighat, F., & Yoshino, H. (2020). Systematic approach to provide building occupants with feedback to reduce energy consumption. Energy, 194, 116813.

[14] Marinakis, V., & Doukas, H. (2018). An advanced IoT-based system for intelligent energy management in buildings. Sensors, 18(2), 610.

[15] Khan, S. U., Khan, N., Ullah, F. U. M., Kim, M. J., Lee, M. Y., & Baik, S. W. (2023). Towards intelligent building energy management: AI-based framework for power consumption and generation forecasting. Energy and buildings, 279, 112705.

[16] Mischos, S., Dalagdi, E., & Vrakas, D. (2023). Intelligent energy management systems: a review. Artificial Intelligence Review, 56(10), 11635-11674.

[17] Arun, S. L., & Selvan, M. P. (2017). Intelligent residential energy management system for dynamic demand response in smart buildings. IEEE Systems Journal, 12(2), 1329-1340.

[18] Boodi, A., Beddiar, K., Benamour, M., Amirat, Y., & Benbouzid, M. (2018). Intelligent systems for building energy and occupant comfort optimization: A state of the art review and recommendations. Energies, 11(10), 2604.

[19] Hannan, M. A., Faisal, M., Ker, P. J., Mun, L. H., Parvin, K., Mahlia, T. M. I., & Blaabjerg, F. (2018). A review of internet of energy based building energy management systems: Issues and recommendations. IEEE access, 6, 38997-39014.

[20] Aguilar, J., Garces-Jimenez, A., R-moreno, M. D., & García, R. (2021). A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings. Renewable and Sustainable Energy Reviews, 151, 111530.

[21] Shah, A. S., Nasir, H., Fayaz, M., Lajis, A., & Shah, A. (2019). A review on energy consumption optimization techniques in IoT based smart building environments. Information, 10(3), 108.

[22] Francesco Giuseppe Ciampi,Andrea Rega,Thierno M.L. Diallo,Francesco Pelella,Jean Yves Choley & Stanislao Patalano. (2024). Energy consumption prediction of industrial HVAC systems using Bayesian Networks. Energy & Buildings114039-.

[23] Yujie Ouyang,Yunfeng Han,Zeyu Wang & Yifei He. (2024). Underwater Network Time Synchronization Method Based on Probabilistic Graphical Models. Journal of Marine Science and Engineering(11),2079-2079.

[24] John Mart V DelosReyes & Miguel A Padilla. (2025). Obtaining a Bayesian Estimate of Coefficient Alpha Using a Posterior Normal Distribution. Educational and psychological measurement 00131644241311877.