Journal of Intelligent Systems and Internet of Things
Vol. 16, No. 02, PP. 102-116, 2025
DOI: https://doi.org/10.54216/JISIoT.160208

Research on the Evaluation Method of Energy Sustainable Development Indicator System Based on Genetic Algorithm and Local Support Vector Regression

Qian Chen[1, 2,*]

[1]Interdisciplinary Studies College, Payap University, Chiangmai 50100, Thailand

[2]School of Business and Public Administration, Guangxi University of Foreign Languages, Nanning, Guangxi 530222, China

Email: QianChencq@outlook.com

Abstract

With the acceleration of modern urbanisation, the demand for energy by the state and society is increasing. In order to maintain the sustainable availability of energy, it is necessary to establish an energy sustainability indicator system. To address this issue, this paper proposes an innovative evaluation method for energy sustainability indicator system, which aims to provide a multi-scale and more comprehensive assessment of energy sustainability indicators, as well as to ensure the accuracy and reliability of the evaluation results. This paper proposes to use genetic algorithm and local support vector regression (SGA-LSVM) to optimise the projective fuzzy clustering solution model (PPFCM), to establish a new evaluation method of energy sustainability index system based on genetic algorithm and local support vector regression. Based on this method, energy sustainability in different regions is analyzed according to three indicators: energy supply side, demand side and affordability, and the validity of this evaluation method is tested. The study found that, in terms of zoning: the eastern region is in the lead in energy demand side, energy supply side and energy affordability, and the western region has a rising trend in recent years; in terms of population density: the indices of energy demand side, energy supply side and energy affordability of densely populated areas are much higher than the rest of the areas compared to the sparsely populated areas, and the difference between the indices of energy supply side and energy affordability of the sparsely populated and moderately populated areas and the difference between the indices is not significant. The energy supply-side index is slightly higher than that of the medium-population area; Economy and carbon emission: due to China's focus on environmental protection, carbon emission is kept within a stable range while the economy is developing rapidly. By $PC \geq 0.80$, $PE \geq 0.45$ and $XB \leq 0.1$, it shows that the method of evaluating the energy sustainable development index system using the fuzzy projection-seeking clustering energy sustainable development evaluation model based on genetic algorithm and local vector regression optimization is reliable.

Received: January 23, 2025 Revised: February 19, 2025 Accepted: March 12, 2025

Keywords: Genetic algorithm; Local support vector regression; Projection fuzzy clustering solution model; Energy sustainable development index system

1. Introduction

Energy, as the most important and valuable entity of the 21st century, is deeply embedded in the way of life of modern civilization. Almost no areas, from households to industry, from transport to communications, can be supported without energy. However, as global energy demand continues to grow, the current energy paradigm faces many challenges that can be socially, economically and environmentally harmful. In order to achieve sustainable development, we must revisit and reassess the existing energy system to ensure that it meets the needs of human society without causing irreversible damage to the Earth's environment. Therefore, sustainability assessment of energy systems and exploring more environmentally friendly, efficient and sustainable energy solutions have become important issues of our time [1].

With the growing global interest in sustainable development, several attempts have been made to work towards a comprehensive sustainability assessment of energy systems. In order to more scientifically measure the environmental impacts, economic viability and social acceptability of energy systems, researchers have developed a variety of assessment models. For example, the hierarchical analysis method (AHP) as a more mature multi-criteria decision-making method, the decision-making process of AHP is based on subjective logical judgement and objective methodological calculations, and the results of decision-making are comprehensive, systematic and scientific [2]. Data envelopment analysis (DEA) is very suitable for evaluating the efficiency of multi-input-multi-output systems, because the optimal weights obtained by using the actual data of the decision-making units are the weights of each input and output, and there is no need to construct any assumptions on the calculation of the weights during the evaluation process, so the evaluation results have a strong objectivity as a whole [3]. Principal Component Analysis (PCA), greatly simplifies the indicators used, effectively solves the problem of overlapping information between indicators, and the weights of the new indicators are not subjectively determined, but are determined according to the magnitude of the contribution of each new factor [4]. The Approximate Ideal Value Sorting Method (TOPSIS) also does not have strict requirements on the number of samples, data distribution and the number of indicators, so it makes full use of the original data and has less information loss [5]. Resource indicators are primarily used to measure the efficiency of energy services in terms of resource utilization, with a particular focus on their resource intensity, which helps us to understand the depletion of natural resources in the energy production process. Economic indicators, on the other hand, focus on assessing the economic efficiency and impact of energy services, including cost-effectiveness, return on capital investment, and contribution to the local economy. In addition, social indicators are used to measure the impact of energy services on society, which covers aspects such as job creation, improvement in living standards and diversification of energy supply.

In response to the various problems in the current energy sustainability assessment, this paper proposes an innovative solution, i.e., the use of genetic algorithms and local support vector regression to assess energy sustainability indicators, resulting in a novel evaluation method. Different from previous methods, the method closely focuses on the state of energy sustainable development and constructs a comprehensive system of indicators for capacity assessment. On this basis, the model is tested by introducing historical measured data, which seeks to minimize the deviation between the evaluation and prediction results of the energy sustainable development indicator system and the measured data, thus improving the calculation accuracy of the energy sustainable development assessment model. In the end, the optimized energy sustainable development assessment model can be used to accurately predict the country's energy sustainable development situation and provide scientific guidance for the corrective measures related to the country's energy sustainable development.

2. Indicator system for sustainable energy development

2.1. Evaluation index system construction

Energy sustainability is an intricacy concept that encompasses multiple disciplines, so assessing it is a cumbersome and subject to double counting. In addition, the field lacks reliable and accurate evaluation models. De Vries et al. suggest that energy sustainability assessments must meet requirements such as reliability, accuracy, and comparability. When analyzing energy sustainability, a careful examination of each assessment criterion is essential [6]. In other words, it is important to avoid factors in the assessment that do not have a clear purpose or objective, these factors need to be fully considered and controlled when conducting the assessment to ensure the scientific and objective nature of the results.

In the process of academic exploration, this paper, based on the rich research results on the evaluation system of sustainable energy development, further deepens and innovatively constructs a comprehensive and multi-dimensional evaluation system of sustainable energy development. The system aims to comprehensively consider the performance of the entire national economy in sustainable energy development, focusing not only on energy use in the field of living consumption, but also analyzing the energy use in the field of production and consumption in depth. Specifically, the system is designed with three first-level indicators, namely energy supply, demand and energy affordability, and as many as 35 second-level indicators in order to assess each dimension in more detail. The construction of this system provides us with a powerful tool for understanding and analyzing the status and trends of sustainable energy development in a comprehensive and in-depth manner.

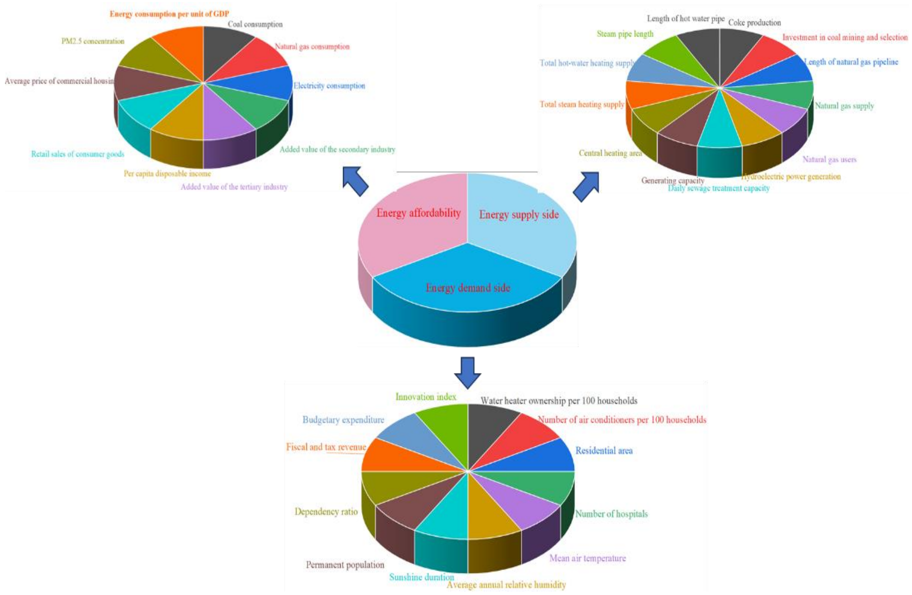


Figure 1. System of evaluation indicators

2.2. Evaluation index model construction

2.2.1. Projective tracing model

The reason for choosing the projection-tracing model is: projection tracing (PP) is a method constructed by Friedman for processing and analyzing high-dimensional complex data the model does not require much data, and it is suitable for analyzing and processing high-dimensional non-normal non-linear data [7-8]. In this paper, the comprehensive evaluation index system for sustainable development of renewable energy based on the 'DPSIR' framework contains multiple indicators, according to the distribution, state is unknown, and there is no obvious linear relationship between the indicators. Therefore, the raw data of this study are complex and difficult to handle high-dimensional data. The projection-tracing algorithm has a relatively large advantage in dealing with this kind of data. The assessment of energy sustainability involves multi-dimensional indicators such as social, economic, energy and environmental indicators, The projection tracing model can perform efficient data downgrading sharpening, helping us to extract key information from massive data and to analyses and process it effectively. The steps to achieve this are as follows:

Step 1: Calculate the projected value of the assessment object $Z(i)$

$$Z(i)=\sum_{j=1}^{m}a(j)x_{ij} \tag{1}$$

Where, $a(j)$ is the projection direction on indicator j (where, $\sum_{j=1}^{m}a(j)^2=1$), j=1,2,3 ......, m. $Z(i)$ is the ith projection volume, x_{ij} is the eigenvalue of i on indicator j, i=1,2,......, n.

Step 2: Calculate the inter-cluster distance $S(z)$

$$S(z)=\sqrt{\frac{\sum_{i=1}^{n}(z(i)-E(z))^2}{n-1}} \tag{2}$$

where the intercluster distance represents the dispersion of the data projection points. $E(z)$ is the expectation values of the projection volume.

Step 3 Calculate the in-group density $D(z)$

$$D(z)=\sum_{i=1}^{n}\sum_{j=1}^{m}\left[(R-d_{ij})-u(R-d_{ij})\right] \tag{3}$$

$$u(t)=\begin{cases}1; t ^{3}\ 0\\0; t<0\end{cases} \tag{4}$$

where R is the width of the density window (empirical value); $u(t)$ is the unit leap function; d_{ij} is the distance between the sample data; $D(z)$ is the intraclass density of the sample projection points, i.e., the relative distance of the projected values within the density window.

Step 4 Construct the projection indicator function $Q(a)$

$$\max\{Q(a)=S(z)D(z)\} \tag{5}$$

As shown in the above equation, when the distance between clusters $S(z)$ is larger and the density within clusters is smaller, the result is more accurate when the objective function $D(z)$ is maximised.

2.2.2. Fuzzy clustering model

The reason for choosing the fuzzy clustering model is that a standardised system for energy sustainability assessment has not yet been established, and therefore it can have a subjective impact on assessing the sustainability of energy at the multi-indicator level. The research object selected for this paper involves multiple years with high variability. From the data itself, the model can achieve clustering in a standardless situation and automatically assess the level of energy sustainability [9-10].

In order to avoid the subjective selection of the clustering centre, this paper chooses the objective function of fuzzy C-mean clustering in the fuzzy clustering model, and for the high-dimensional samples, the objective function is expressed by the generalized Euclidean distance sum of squares. The main formula is:

$$\min\{F(u_{hi},s_{jh})\}=\sum_{i=1}^{n}\sum_{h=1}^{C}\left[u_{hi}{}^{2}\left\|r_{ij}-s_{jh}\right\|\right]^{2} \tag{6}$$

where n is the number of countries; m is the number of indicator dimensions; c is the number of categories to be classified; r_{ij} is the standardised eigenvalue of the ith country for the jth indicator. u_{hi} is the degree of affiliation of country x in category h, s_{jh} is the clustering center of indicator j in category h. Both can be obtained by the Lagrangian method:

$$s_{ij}=\frac{\sum_{i=1}^{n}u_{hj}{}^{2}r_{ij}}{\sum_{i=1}^{n}u_{hj}{}^{2}} \tag{7}$$

$$u_{hj}=\left[\sum_{k=1}^{C}\frac{\sum_{j=1}^{m}\left[\left(r_{ij}-s_{jh}\right)\right]^{2}}{\sum_{j=1}^{m}\left[\left(r_{ij}-s_{jk}\right)\right]^{2}}\right]^{-1} \tag{8}$$

2.2.3. Fuzzy Projective Tracer Clustering Model (FCM)

Although Projective Trace Model (PP) and Fuzzy Clustering Model (FCM) are widely used in data processing and evaluation, classification, etc., they inevitably have their own model limitations and shortcomings, as shown in the following table.

Table 1: The advantages and disadvantages of the two models

Method	Merit	Shortcoming
Projection tracing model (PP)	Reduce the computational complexity, and have little data requirements	Subjective judgement determines the outcome of the assessment
Fuzzy clustering model (FCM)	It is not limited by the evaluation criteria of the sample	More sample-dependent and for sample data from multiple sources, its computational complexity is high and easy to converge locally

Traditional projection-seeking clustering models require a density window width value, which is directly related to the validity of the assessment results. The method currently used in the existing literature is based on trial calculations or empirical judgements. However, this lacks theoretical basis, is subjective and has limitations. In addition, the assessment criteria for energy sustainability have not yet been standardized across countries, and the traditional method of relying on assessment criteria to classify the level of development is not highly applicable in the energy sector. The fuzzy clustering model does not rely on artificial assessment criteria, but it has strict data requirements.vg

2.3 Projection fuzzy clustering solution model

In order to reduce the subjective influence of evaluation criteria and parameter selection and reduce the computational complexity, this paper adopts the fuzzy clustering model's objective function F instead of the projective tracing model's in-group density D. After the dimensionality reduction process of the projective tracing, the clustering samples are reduced from multi-dimensional to uni-dimensional, i.e., j=1, m=1, which reduces the complexity of the algorithm. At the same time, the clustering centre value of the fuzzy clustering model is used to automatically cluster and assess the national energy sustainable development level. Overall,

Then the projection tracking model and fuzzy clustering model are integrated to construct the evaluation function of energy sustainable development index, effectively retaining the advantages of both while avoiding the subjective factor of parameter selection, achieving the complementary fusion of the two, and improving the accuracy and objectivity of the assessment of sustainable energy development. The final projected fuzzy clustering solution model (PPFCM) is obtained after fusion, as shown in the following figure.

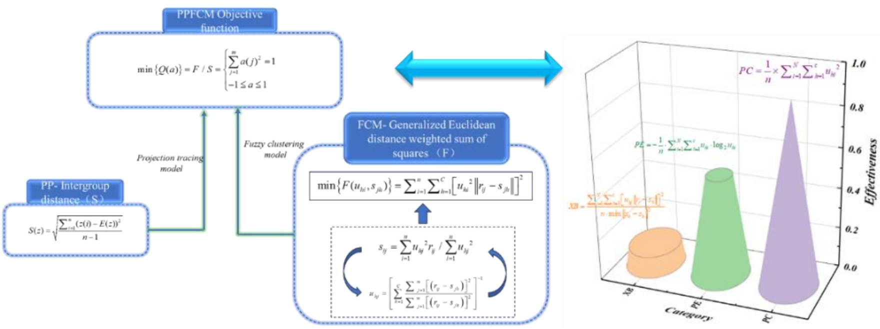


Figure 2. Projection fuzzy clustering solves the model and validates the index

At the same time, in order to ensure the reasonableness of the evaluation results of the combined fuzzy projection seeking clustering model, this paper introduces the clustering validity index to verify the evaluation results, and the following clustering validity index is used for evaluation. In Fig. 2, XB, PE and PC are used to test the indicators. In addition, r represents the ith data. The smaller the values of XB and PE, the better the evaluation results; the closer the value of PC is to 1, the better the evaluation results.

3. Energy evaluation optimization based on SGA-LSVM algorithm

3.1 Quadratic optimization of evaluation model based on genetic algorithm

When solving the complex nonlinear optimization problem of projective fuzzy clustering model, researchers often face many challenges. Although there are many algorithms to choose from, such as particle swarm optimization [11], gray Wolf optimization [12] and gravity search algorithm [13], there are often some problems in the practical application of these algorithms, such as premature convergence, slow computing speed and easy to fall into local optimal. These problems are especially prominent in the case of many optimization variables and the need for frequent binary encoding and decoding. Therefore, in order to solve this problem more effectively, this paper chooses a more mature and widely used genetic algorithm as the basis for solving it. By simulating the natural selection and chromosome exchange mechanism in Darwinian biological evolution, genetic algorithm realizes a global optimization search, and the standard genetic algorithm is based on natural selection and natural gene mechanism, which has become a powerful and widely used optimization algorithm [14].

In the process of constructing the evaluation method of the energy sustainable development indicator system, the assignment of the values of its dimensions and the setting of its affiliation function mainly depends on the experience of the evaluation experts, so the accuracy of the model is largely constrained by the level of experts.

For more complex evaluation methods, when assigning values to the models, the problems of wrong assignment and omission are difficult to avoid completely, which leads to the lack of reliability of the assessment methods. In order to ensure the accuracy of the calculation of the energy sustainability indicator system, it is necessary to carry out secondary optimization of the capacity assessment model [15-16]. Given that it is difficult to obtain more accurate weight values by solely relying on assessment experts, this paper uses historical statistical data to optimize the model. The analysis shows that when the theoretical capacity calculated based on the capacity assessment model is closer to the actual capacity, it means that the weighting coefficients are more accurate, and the accuracy of the capacity assessment model is higher.

In the energy sustainability index assessment model, the optimal weighting solutions of the influencing factors are randomly distributed within the given constraints, which are difficult to locate directly and accurately. In contrast, genetic algorithms can comprehensively explore the constraint space of the parameters to find the optimal solution through the strategy of multi-point parallel search when solving the problem. In addition, genetic algorithms also have the ability to deal with single-objective or multi-objective function problems with complex constraints. Therefore, it is appropriate to apply genetic algorithm to solve the weights of discrete production line capacity evaluation model, and its specific operation flow is shown in Fig. 3. The weights of the influencing factors are defined as unknowns and the main steps of the algorithm are as follows:

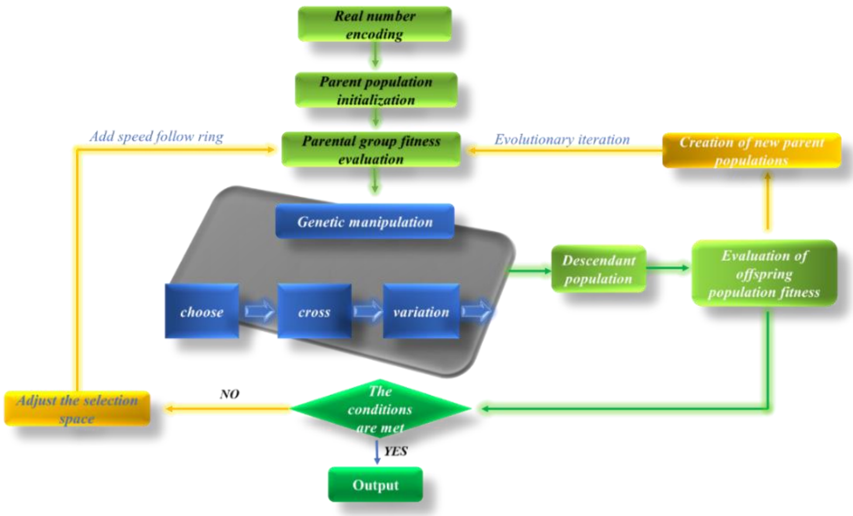


Figure 3. SGA algorithm

(1) Generating the initial population

Within the framework of the genetic algorithm, the first task is to identify the core independent variable that reflects the optimization objective. Based on the previous analyses, we selected the weights as the design parameters. Further, in order to find the optimal values of these design parameters, it is necessary to define the constraint ranges of the variables. According to the previous section, the initial weights $w_i(i=1,2,...,6)$ of each influence factor in the capacity assessment model can be obtained, and according to the empirical value, $\left\{\left[\frac{w_{0i}}{2},\frac{3w_{0i}}{2}\right],i=1,2,...,6\right\}$ is selected as the range of variables for initialization of the population, i.e. the algorithm is given a set of solutions to the stochastic problem, and the individuals with high fitness are selected according to the fitness function given to form an initial population $\left\{Z_{ij},j=1,2,...,n\right\},(i=1,2,...,6)$.

(2) Adaptation Function

From the above analysis, it can be inferred that as the theoretical capacity calculated by the capacity evaluation model is closer to the actual capacity, the accuracy of the model will be higher. Since it is difficult to set the ideal weighting coefficients by human experience, we choose the historical capacity error of a discrete production line as the optimization target. It is known that the actual capacity of the production line in a certain period of time is M_p, while the capacity evaluation model can derive the corresponding theoretical capacity $M_p(z)$. The smaller the difference between the two, the more accurate the construction of the capacity evaluation model is. Accordingly, the objective function is defined as:

$$\begin{cases} \begin{pmatrix} f_1(z) \\ f_i(z) \\ f_m(z) \end{pmatrix} = \begin{pmatrix} \min\|M_{p1} - M_{p1}(z)\| \\ \min\|M_{pi} - M_{p1}(z)\| \\ \min\|M_{pm} - M_{p1}(z)\| \end{pmatrix} \\ s.t. z = \left(z_i \middle| z_i \in \left[\frac{w_{0i}}{2}, \frac{3w_{0i}}{2} \right], i = 1, 2, \ldots, 6 \right) \end{cases} \tag{9}$$

(3) Genetic operator operation

Individuals in the population are selected to construct a new generation of population through selection operation, mutation operation and crossover operation. In the selection stage, a roulette wheel selection mechanism is used, where the probability of selecting an individual is positively correlate. In the mutation stage, a basic positional mutation strategy is used. That is, one or more genes in the coding sequence of an individual are randomly selected for mutation based on the mutation probability. In the crossover phase, a crossover point in the coding sequence of an individual is randomly determined and some of the genes of two paired individuals are exchanged at that crossover point.

(4) Termination conditions

In the case of multiple uncertainties in the production site, the predicted value of the model is often difficult to fully match the historical measurement data. This means that even with the optimal weighting scheme, it is still difficult to accurately predict the convergence of the objective function. In order to meet this challenge and prevent the algorithm results from being limited to the local optimal solution within the constrained interval, this paper innovatively chooses to set a sufficient number of genetic algebras Ni as the termination condition of the algorithm. This decision determines the degree of detail of the algorithm search to some extent. If the genetic algebra Ni is set too small, the algorithm may not be able to fully explore the solution space, and thus miss the global optimal solution. Conversely, if Ni is set too high, the algorithm can be stuck in a lengthy loop, resulting in computational inefficiency. Therefore, in practical applications, we usually need to reasonably determine the number of genetic algebras Ni based on empirical values to ensure that the algorithm can both converge in an acceptable time and find the global optimal solution with a high probability.

3.2 Localized support vector regression

3.2.1 Support Vector Regression

SVR algorithm is an important extension of support vector machine algorithm, which is specially used to solve regression problems. The core idea is to map the eigenvector of the original sample set from a low-dimensional space to a higher-dimensional space. In this new high-dimensional space, the algorithm further performs linear regression operations on the mapped sample set. Through this process, SVR algorithms can more effectively capture complex relationships in the data, thus providing predictions that are more accurate and fitting results [17-18]. It is defined as follows.

For a given set of samples: $S = \left\{ (x_i, y_i) \middle| x_i \in R^n, y_i \in R \right\}_{i=1}^{L}$. Where x_i denotes the eigenvector of the ith sample, y_i denotes the regression value of the ith sample, and L denotes the number of samples. Using the nonlinear mappingφ: $R^n \to R^m (m \geq n)$, Mapping the feature vector x of each sample from the original space R^n to the high dimensional space R^m. The mapped feature vector is denoted as φ(x). Therefore, the mapped sample set is denoted as: $S_\varphi = \left\{ (\varphi(x_i), y_i) \middle| \varphi(x_i) \in R^m, y_i \in R \right\}_{i=1}^{L}$ The sample set S_φ is then fitted in the feature space R^m using a linear function, the specific linear function form is defined as:

$f(x) = w \cdot \varphi(x) + b$ (10)

where: w denotes linear weights ($w \in R^m$), denotes bias ($b \in R$).

Based on the statistical learning theory and the description of the references, the support vector regression method determines the regression function equation (10) by polarising it into the following objective function equation (11).

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{L} \left| f(x_i) - y_i \right|_e \tag{11}$$

where:C is the penalty coefficient and C > 0. $\left| f(x_i) - y_i \right|_e$ is called the ε-insensitive loss function, which is defined as follows:

$$\left| f(x) - y \right|_e = \max\left\{0, \left| f(x) - y \right| - e\right\} \tag{12}$$

Therefore, the final objective optimisation function can be transformed into:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{L}(x_i + x_i^*) \tag{13}$$

Constraints:

$$s.t. \begin{cases} y_i - f(x_i) \le e + x_i \\ -y_i + f(x_i) \le e + x_i^*; i = 1, 2, ..., L \\ x_i, x_i^* \ge 0 \end{cases} \tag{14}$$

Where: e is the maximum error allowed in the regression, which controls the number of support vectors and generalization ability, the larger e is, the fewer support vectors there are. x_i and x_i^* both denote slack variables.

According to the dyadic principle, the optimization problem corresponding to the above equation can be solved using the Lagrange multiplier method and the KKT condition (Karush-Kuhn-Tueker), and the optimal regression hyperplane is:

$$f(x) = w \cdot j(x) + b = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x) + b \tag{15}$$

where a_i , a_i^* is a Lagrange multiplier and $0 \le a_i, a_i^* \le C$; x_i is a support vector satisfying $y_i - f(x_i) = e$ or $-y_i + f(x_i) = e$; b is a constant, which can be derived from the support vector x_i ; $K(x_i, x) = (j(x_i) \cdot j(x))$ is a kernel function.

3.2.2 Local support vector regression

When constructing regression model, support vector machine usually considers all training samples, and this global processing method may ignore the local change information between samples in some cases. To overcome this limitation, local support vector machines (LSVMS) came into being. Because of the traditional support vector machine, it introduces the local learning algorithm skillfully, so that the regression model not only has the global perspective, but also contains the idea of localization. This model can effectively capture the local variation trend of the sample data and reflect the intrinsic characteristics of the data more accurately, thus significantly improving the prediction accuracy of the model. In this way, local support vector machines provide new ideas and methods for dealing with complex regression problems [19].

In 2007, Cheng et al. proposed a new local support vector machine (LSVM) [20] called LSVM based on the similarity between the training samples and the test samples. LSVM uses a similarity function $s(x_i, x_j^*)$ to represent the similarity between the training samples x_j^* and the test samples x_i . According to the different values of the similarity function σ, two variants of LSVM can be generated, when σ taking a real number between [0,1], the LSVM obtained is called SLSVM (Soft Localized Support Vector Machine, SLSVM); when σ is a binary function, the LSVM obtained is called HLSVM (Hard Localized Support Vector Machine, HLSVM), at this time the similarity function expression is:

$$s(x_i, x_j^*) = \begin{cases} 1; x_i \hat{I}\ k\text{-nearest neighbour of } x_j^* \\ 0; \text{ if not} \end{cases} \tag{16}$$

where is the distance function used in the calculation of the x_j^* nearest neighbours.

Cheng et al. used the proposed LSVM for classification problems and achieved better results. In this paper, HLSVM is used for regression to obtain Hard Localized Support Vector Regression (HLSVR) based on HLSVM, and the steps for constructing the regression model are as follows: (1) Determine the value of K; (2) Select K nearest-neighbor samples for each test sample; (3) For the selected K nearest-neighbor samples, use the support vector machine to perform regression modelling; (4) Use the established support vector regression model to predict the test sample; (5) Perform (2) to (4) for each test sample until the prediction of all test samples is completed.

Compared with the standard SVR, using HLSVR to predict the test samples can make full use of the local information of the samples and select the samples with greater similarity to the test samples to participate in the construction of the model, which can effectively improve the prediction accuracy; and HLSVR can reduce the number of samples participating in the construction of the model, which reduces the time of constructing a single model [21].

4. Analysis of energy sustainability index and model validity test

4.1 Evaluation of indicators in each region

Firstly, from the basic situation of the energy sustainable development index of 30 provinces, autonomous regions and municipalities in China from 2005 to 2018 (see Fig. 4), it can be seen that the total indexes of Shandong, Beijing, Liaoning, Jiangsu and Guangdong are in the top five. Guangxi, Gansu, Jiangxi, Guizhou, Ningxia, Hainan, and

In China, the level of sustainable energy development in different regions shows significant differences. Due to various factors, the sustainable energy development in Qinghai region is relatively backward, which leads to the local residents easily falling into the predicament of energy poverty and fuel poverty. This dilemma is closely related to the relatively low level of economic development in central and western China, the imperfect construction of energy infrastructure, and the need to further improve financial investment and energy supply. At the same time, Shandong, Liaoning, Hebei, Shanxi, Heilongjiang and other provinces have shown clear comparative advantages in energy supply. With rich resource endowments and sound industrial bases, these provinces have been able to provide primary energy such as coal, natural gas, electricity and heat in a stable and continuous manner, as well as related energy services, providing strong support for economic and social development. Beijing, Jiangsu, Guangdong, Zhejiang, Shanghai and other provinces, regions and cities, on the other hand, have a higher level of development on the energy demand side and energy affordability, especially in the economically developed areas along the eastern coast, where there are more large and medium-sized cities with higher levels of income among residents, dense populations, sound energy infrastructures, and strong financial resources, and where there is a much stronger demand from residents for green and clean, energy-saving, low-carbon, healthy and efficient high-quality energy services, and a much greater demand for green and clean, energy-saving, low-carbon, healthy and efficient energy services, and a much greater demand in achieving They have a clear layout for achieving green development, carbon emission reduction and even carbon neutrality, and are able to take the lead in implementing the new development concept of ‘innovation, coordination, greenness, openness, and sharing’, and turn their economic advantages into sustainable development advantages in a timely manner.

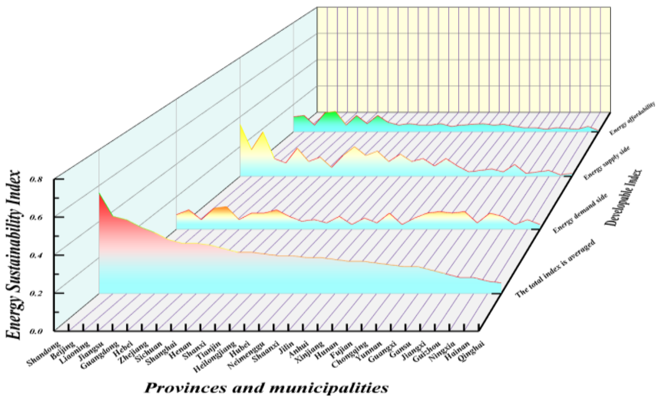


Figure 4. Mean value of the energy sustainability index

Second, this paper further analyses the temporal trend of energy development in major regions of the country from the perspective of administrative subdivisions (see Figure 5). From 2005 to 2018, the level of energy development in China's major regions fluctuated and increased. Although the energy development of each region converges, the process of change is different, and the ranking of energy development level is also different. First, the energy development indexes of North China, East China, and Northeast China continued to lead during this 14-year period. The energy sustainable development index of North China and East China is relatively stable; the energy sustainable development index of Northeast China declined continuously from 2005 to 2009, and then fluctuated after 2009. Second, the energy sustainability index of Central China, South China and Southwest China is in the middle. The energy sustainable development index of Central China and Southwest China fluctuated and increased from 2005 to 2018, and the energy sustainable development index of South China continued to increase from 2005 to 2011, and then fluctuated and decreased after 2011. Third, the degree of energy development in Northwest China is in the last echelon, and its Energy Sustainability Index has shown a fluctuating upward trend during the 14-year period, but there is stillroom for further improvement.

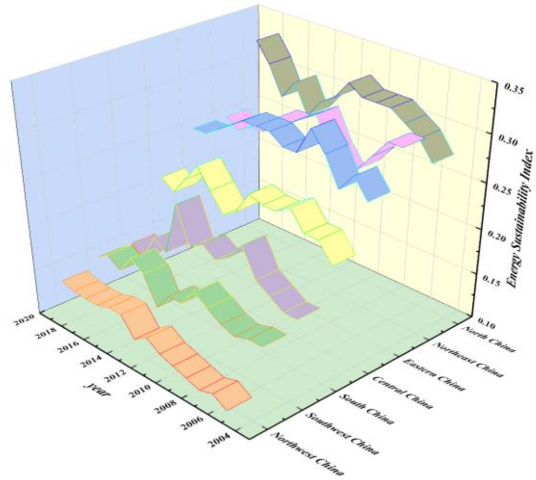
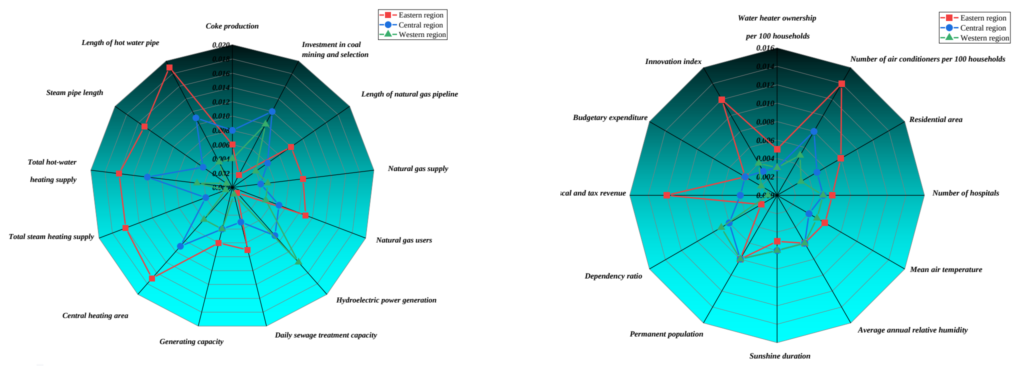


Figure 5. Energy development index of China's seven major administrative regions

4.2 Analysis of indices in a multidimensional perspective

4.2.1 Geographical perspective



(a) Energy supply side

(b) Energy demand side

Figure 6. Energy supplyand demand side exponential distribution

According to figure 6, on the energy supply side, economic development on the southeast coast has facilitated infrastructure development, resulting in a stable, high-quality energy supply. The central region, on the other hand, has advantages in the coal mining industry and coke production, especially the coal resource endowment of Shanxi and other provinces. While the western region is at a disadvantage in other energy supplies, it has great potential for clean and green energy in the future due to its abundant hydro resources, especially in Sichuan and Guizhou. On the energy demand side, the eastern region is in the lead in terms of equipment ownership, local government support, population size, etc., especially in terms of air conditioner ownership, innovation index and fiscal revenue, etc., and Energy infrastructure construction is better than the Midwest, Although the it have advantages in terms of air-conditioning ownership and resident population, there is a large gap in energy infrastructure and equipment renewal compared with the eastern regions.

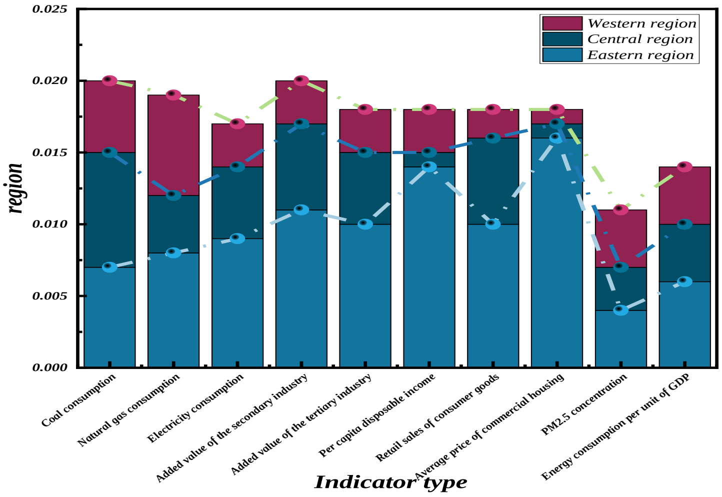
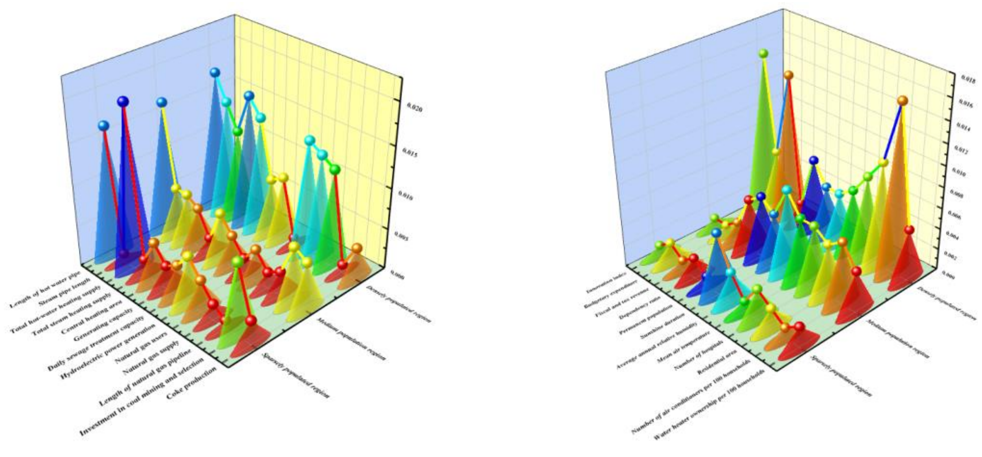


Figure 7. Energy affordability exponential distribution

According to Figure 7, in terms of energy affordability, the eastern region has a stronger economy, with leading indicators such as average price of commercial property, retail sales, and per capita income, demonstrating higher energy consumption capacity. In terms of environmental constraints, the PM2.5 concentration and energy consumption per unit of GDP are lower in the eastern, central and western regions, showing the effectiveness of energy saving and emission-reduction policies and reflecting the process of decarburization and modernization of China's energy consumption structure.

4.2.2 Population density perspective



(a) Energy supply side

(b) Energy demand side

Figure 8. Energy supplyand demand side exponential distribution

According to Figure 8, according to the population density of Chinese provinces, autonomous regions and municipalities, the densely populated areas are Beijing, Shanghai, Tianjin, Jiangsu, Shandong, Guangdong, Henan and Zhejiang, the sparsely populated areas are Heilongjiang, Gansu, Inner Mongolia, Xinjiang and Qinghai, and the rest of the provinces, autonomous regions and municipalities are medium-population areas. According to Figure 7, from the energy demand side, China's densely populated areas have a comparative advantage in terms of air-conditioning ownership per 100 households, innovation index, and fiscal tax revenues, and their larger population size and higher level of urbanization directly drive a relatively strong demand for energy services. On the other hand, the sparsely populated areas in the northwest and southwest of China have relatively low energy demand indices, and their energy demand is relatively small. From the energy supply side, densely populated areas in China have relatively high index values for the length of hot water pipelines, centralized heating area, length of natural gas pipelines, natural gas supply volume, natural gas-using population, and daily wastewater treatment capacity, and these areas have a higher level of energy management and energy security. Densely populated areas have a greater demand for clean energy, such as water and electricity, and therefore have a better supply of renewable energy than sparsely populated areas, and a higher level of optimization of industrial structure.

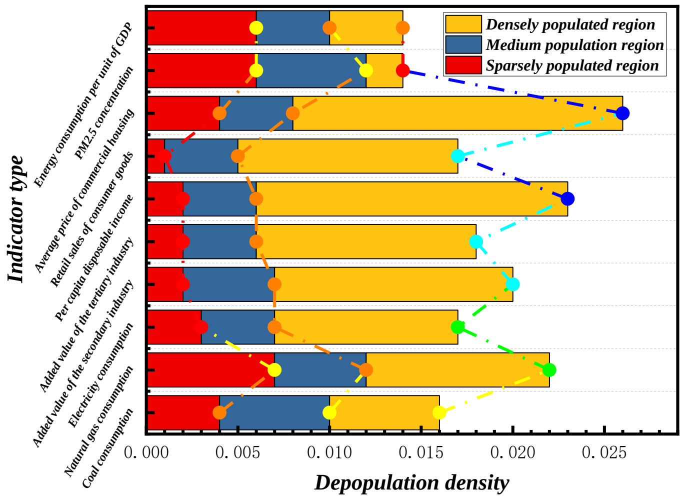


Figure 9. Energy affordability exponential distribution

In terms of energy affordability, there are obvious differences in the regional characteristics of the level of payment for domestic and commercial energy use, green cleanliness, and energy security coefficients at different population densities. Densely populated areas in China lead in primary and secondary energy consumption, value added of secondary and tertiary industries, average price of commercial houses, retail sales of consumer goods, and other tertiary indicators, which are also determined by the strong economic strength and high urbanisation level of densely populated areas. On the contrary, the energy affordability of medium and sparsely populated districts decreases in the same order.

4.2.3 Economic and Carbon Emissions Perspectives

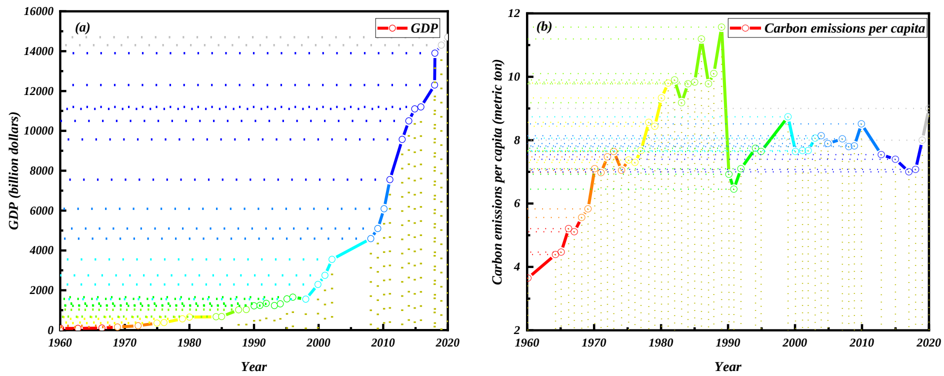


Figure 10. Economic and carbon emissions perspectives

From the comparison in Figure 10 above, it is easy to see that our economic growth has not been accompanied by a rise in our carbon emissions. As an industrialized country, China's oil and gas industry and manufacturing industry emit large amounts of greenhouse gases, but at the same time, we are also subject to the restrictions of the Kyoto Protocol and must fulfil the task of reducing emissions, and we have indeed invested huge financial and human resources in improving energy efficiency, reducing emissions and protecting the environment over the past half-century. This is because our country has always attached importance to economic growth while controlling carbon emissions. In developing and utilizing hydropower, our government has paid great attention to its environmental effects, limiting the negative impacts of production activities through administrative regulations or taxation.

Since the strengthening of environmental management, the quality of our environment has improved significantly. For example, more than half have reduced the number of pollutants such as phosphorus and mercury in the water, or the number of harmful substances in the atmosphere has been significantly reduced, which has led to an increase in the visibility of the air. The urban areas of our capital are well planned, with green belts everywhere except in buildings and streets, and the city has a large number of vehicles, but the sound of honking is largely absent. Factories are generally required to build outside the city, rolling black smoke never appear in the city, not only that, the city's domestic waste will be a huge rubbish treatment plant for unified disposal, China upholds the principle of secondary use, waste into treasure, to a certain extent, greatly reducing the pressure on the environment. China's beautiful natural landscape and open policy attracts many tourists every year, and its environmental policy is also the object of many countries have to follow, indicating that China's energy in the economy and the environment has a greater sustainability of development, that is, has a greater space for development. It also shows that China's energy has a better development prospect.

4.3 Model validity test

In order to further, test the validity of the integrated assessment results of the projection seeking and fuzzy clustering energy sustainability indicator model based on genetic algorithm with local support vector regression optimization. According to the formula in Fig. 3, the clustering validity of the model after each year's data processing is obtained, as shown in Table 2 and Fig. 11, according to the calculated PC, PE and XB are greater than 0.8, greater than or equal to 0.4 and less than or equal to 0.12, and the calculated results are kept in a relatively stable state. It shows that the evaluation of energy sustainable development index is relatively stable and effective.

Table 2: Clustering validity indicators

index	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
PE	0.45	0.45	0.49	0.46	0.49	0.45	0.45	0.48	0.45	0.48	0.49	0.51
PC	0.83	0.82	0.81	0.82	0.81	0.8	0.83	0.82	0.82	0.82	0.92	0.96
XB	0.09	0.09	0.10	0.09	0.09	0.10	0.10	0.09	0.09	0.10	0.12	0.11

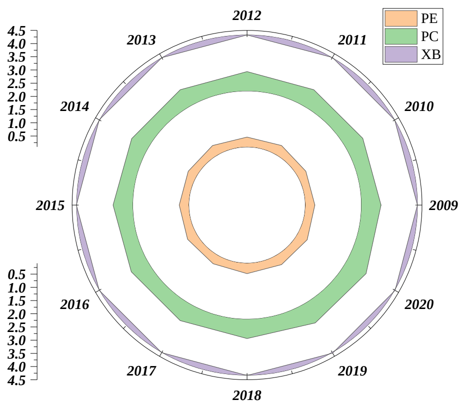


Figure 11. Numerical change plot of clustering effectiveness indicators

5. Conclusion

This study analyses the projection tracing model, fuzzy clustering model, establishes the fuzzy projection tracing clustering energy sustainable development evaluation model and its operation and analysis process by combining the above two models, and combines the genetic algorithm with the local support vector regression method to optimize the energy sustainable development evaluation index model, and innovatively puts forward a set of comprehensive index evaluation methods. The specific conclusions are as follows:

1. This paper is based on statistical projection tracing as well as fuzzy comprehensive evaluation method and so on. On this basis, the analytical methods suitable for comprehensive evaluation of sustainable energy development - projection tracing and fuzzy clustering methods - are selected, and they are reasonably combined to establish a projection fuzzy clustering model, and the model based on genetic algorithm and local vector regression is selected to find the optimal solution, so as to construct the energy sustainable development assessment model.

2. In the comprehensive situation, in terms of zoning: the eastern region is in the leading position in energy demand side, energy supply side and energy affordability, while the western region is lagging behind, although there is an upward trend in recent years; in terms of population density: the indices of energy demand side, energy supply side and energy affordability in densely populated areas are much higher than the rest of the regions, while the indices of energy demand side of sparsely populated areas are lower than that of the moderately populated areas, and the indices of energy supply side and energy affordability are lower than that of the moderately populated areas. The difference between the energy supply side and energy affordability indexes and the medium-population zone is not big, in which the energy supply side index is slightly higher than that of the medium-population zone; in terms of economy and carbon emission: in the stage of rapid development, due to China's focus on environmental protection, the carbon emission is kept in a stable range while the economy develops rapidly.

3. Through the test of PC, PE, XB and other three aspects, as $PC \geq 0.80$,$PE \geq 0.45$, $XB \leq 0.1$, and each operation results in the indicators are maintained in a relatively stable value. It shows that the method of evaluating the energy sustainable development indicator system using the fuzzy projection seeking clustering energy sustainable development evaluation model based on genetic algorithm and local vector regression optimization is reliable.

References

[1] A. R. Azzam, D. Ibrahim et al., “Sustainability assessment of energy systems: A novel integrated model,” *J. Clean. Prod.*, vol. 2019, pp. 1098–1116, 2019.

[2] Y. Cheng, H. Wu, H. Yang, et al., “Fuzzy comprehensive status evaluation on electric energy metering device and test strategy research,” *Electr. Meas. Instrum.*, vol. 12, p. 003, 2012.

[3] K. H. Chen, H. Y. Yang, J. M. Lee, et al., “The impact of energy prices on energy consumption and energy efficiency: Evidence from Taiwan,” *Energy Eff.*, vol. 9, no. 6, pp. 1–21, 2016.

[4] H. Doukas, A. Papadopoulou, N. Savvakis, et al., “Assessing energy sustainability of rural communities using principal component analysis,” *Renew. Sustain. Energy Rev.*, vol. 16, no. 4, pp. 1949–1957, 2012.

[5] A. Afsordegan, M. Sánchez, N. Agell, et al., “Decision making under uncertainty using a qualitative TOPSIS method for selecting sustainable energy alternatives,” *Int. J. Environ. Sci. Technol.*, vol. 13, no. 6, pp. 1419–1432, 2016.

[6] J. De Vries, T. Vinken, L. Hamelin, and I. D. Boer, “Comparing environmental consequences of anaerobic mono- and co-digestion of pig manure to produce bio-energy – A life cycle perspective,” *Bioresour. Technol.*, vol. 125, pp. 239–248, 2012.

[7] S. Espezua, E. Villanueva, C. D. Maciel, et al., “A projection pursuit framework for supervised dimension reduction of high-dimensional small sample datasets,” *Neurocomputing*, vol. 149, pp. 767–776, 2015.

[8] J. A. Malpica, J. G. Rejas, and M. C. Alonso, “A projection pursuit algorithm for anomaly detection in hyperspectral imagery,” *Pattern Recognit.*, vol. 41, no. 11, pp. 3313–3327, 2008.

[9] Y. He, J. Zhou, P. Kou, et al., “A fuzzy clustering iterative model using chaotic differential evolution algorithm for evaluating flood disaster,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10060–10065, 2011.

[10] S. Lu, Y. Shang, and Y. Li, “A research on the application of fuzzy iteration clustering in the water conservancy project,” *J. Clean. Prod.*, vol. 151, pp. 356–360, 2017.

[11] F. L. Li, J. Guo, et al., “Topology optimization of particle swarm optimization,” *Adv. Swarm Intell.*, vol. 12, no. 3, pp. 142–149, 2014.

[12] E. Emary, H. M. Zawbaa, et al., “Feature subset selection approach by gray-wolf optimization,” *Afro-Eur. Conf. Ind. Adv.*, vol. 17, no. 19, pp. 1–13, 2015.

[13] N. Siddique, H. Adeli, et al., “Gravitational search algorithm and its variants,” *Int. J. Pattern Recognit. Artif. Intell.*, 2016.

[14] A. Anwaar, A. Ashraf, et al., “Genetic algorithms: Brief review on genetic algorithms for global optimization problems,” *Hum.-Centered Cogn. Syst.*, pp. 80–85, 2022.

[15] World Energy Council, *Pursuing Sustainability: 2010 Assessment of Country Energy and Climate Policy*, World Energy Council Publication, 2010.

[16] P. Ekanayake, P. Moriarty, and D. Honnery, “Equity and energy in global solutions to climate change,” *Energy Sustain. Dev.*, vol. 26, pp. 72–78, 2015.

[17] L. Zhu, Q. H. Wu, et al., “Mathematical morphology-based short-term wind speed prediction using support vector regression,” in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Eur.*, 2014.

[18] W. Ling, W. L. Lu, et al., “Fuzzy rules extraction based on output-interval clustering and support vector regression for forecasting,” *J. Intell. Fuzzy Syst.*, vol. 27, no. 5, pp. 2563–2571, 2014.

[19] M. T. Tran-Nguyen, L. D. Bui, et al., “Decision trees using local support vector regression models for large datasets,” *J. Inf. Telecommun.*, vol. 4, no. 1, pp. 17–35, 2020.

[20] H. Cheng, P. N. Tan, and R. Jin, “Localized support vector machine and its efficient algorithm,” in *Proc. SIAM Int. Conf. Data Mining*, Minneapolis, MN, USA, 2007, pp. 461–466.

[21] H. Yang, K. Huang, I. King, et al., “Local support vector regression for time series prediction,” *Neurocomputing*, vol. 72, nos. 10–12, pp. 2659–2669, 2009.