International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2626-2638

https://doi.org/10.70517/ijhsa463220

Grid Spatial Load Forecasting Method Based on Big Data Technology and Optimization of Medium-voltage Distribution Network Wiring Mode

Yuan Ma[1,*], Jialin Liu[1], Can Chen[1], Guangda Xu[1] and Xinsheng Ma[1]
[1] Electric Power Research Institute, State Grid Jibei Electric Power Co., Ltd., Beijing 100045, China
Corresponding authors: (e-mail: dhkj45@163.com).

Abstract Traditional linear regression is difficult to capture complex load changes, has low prediction accuracy, and lacks systematicity in distribution network wiring optimization, which affects power supply efficiency. This paper combines big data with machine learning to propose a grid load forecasting and wiring optimization solution. First, after processing based on the Apache Hadoop framework, real-time data access is performed through Kafka, and real-time analysis and calculation are performed using Spark Streaming. Random forest is used for load forecasting, and data access efficiency is optimized through consumer subscription and asynchronous processing. Grafana is used to monitor over-limit alarms to ensure accurate predictions. Then, load and geographic data are integrated, and K-Means clustering is applied to identify high-load areas. The GWR (Geographically Weighted Regression) model is constructed to evaluate the impact of spatial characteristics on load. Finally, based on the distribution network wiring model of load data, the node electrical parameters are set and the wiring scheme is optimized using genetic algorithm. The experimental results show that the load forecast MAE (Mean Absolute Error) is reduced by 21.58%, and the loss is reduced by 33.16% after the wiring mode is optimized. The comprehensive method based on big data effectively improves the load forecasting accuracy and distribution network optimization efficiency, providing an important reference for the development of smart grids.

Index Terms Smart Grid, Load Forecasting, Big Data Technology, Machine Learning Algorithms, Wiring Mode Optimization

I. Introduction

Faced with the continued growth of global energy demand, especially in the context of accelerated urbanization and the rising proportion of renewable energy, the stability and efficiency of the power system have become increasingly prominent. The dynamic characteristics of power loads become complex and changeable due to weather changes, economic fluctuations and user behavior patterns [1]-[3]. The traditional linear regression model, based on the simplified assumption of a linear relationship between load and related factors, fails to fully consider the influence of these complex factors, resulting in limited prediction accuracy [4]-[6]. The research on wiring mode optimization of medium-voltage distribution network also faces challenges. It mainly focuses on in-depth analysis of a single mode and lacks a comprehensive and systematic optimization framework, so the power supply efficiency still needs to be improved [7], [8]. In order to enhance the operating efficiency of the power system, an innovative solution must be explored to improve the accuracy of load forecasting and optimize the wiring pattern of the distribution network.

In the field of load forecasting, many scholars have tried to apply a variety of models and methods to meet the challenges. For example, Li J [9] used the ensemble empirical mode decomposition algorithm to decompose the load data into components of different frequencies, then used the multivariate linear regression method and this LSTM to predict the low-frequency and high-frequency components respectively, and finally obtained better prediction results. In addition, Li G [10] proposed a multi-step forecasting method based on phase space reconstruction and support vector machine methods, aiming to improve the accuracy of medium-term load forecasting. Ngoc T T [11] improved the prediction effect of load demand data through grid search algorithm. In order to solve the problem of drastic changes in short-term electricity consumption and increasing data complexity, Lv L [12] applied a hybrid model that combined variational mode decomposition, LSTM neural network, seasonal factor elimination and error correction technology. Although these models have advantages in terms of computational efficiency and simplicity of implementation, they still have certain limitations when dealing with complex practical application scenarios.

Traditional methods are difficult to capture nonlinear and sudden load changes, especially when the load fluctuates greatly or is affected by multiple external factors, and their prediction results often deviate from reality. In addition, there are relatively few studies on the optimization of distribution network wiring modes. Existing studies mainly rely on empirical rules and lack systematicity and scientificity [13]-[15]. This lack of comprehensive consideration of the wiring mode optimization strategy has led to increased power supply pressure on the power system during peak loads and may cause safety hazards in power supply [16]-[19]. Therefore, it is urgent to develop a new research framework to comprehensively improve the effects of load forecasting and wiring mode optimization.

In order to cope with the complex challenges faced by the power system, some scientific researchers are actively exploring the use of big data and machine learning technology to innovate load forecasting methods, using cutting-edge algorithms such as random forests and support vector machines to build load forecasting models. These algorithms have shown significant advantages in handling nonlinear relationships, and their prediction performance has been significantly improved compared with traditional methods [20]-[22]. These machine learning models rely on their deep learning capabilities on massive amounts of data to more precisely depict the correlation between complex inputs and outputs. Despite this, although current research has made certain progress in the field of load forecasting, it often ignores the detailed analysis of the spatial dimension, especially the potential impact of geographical factors on load dynamic changes, which to a certain extent limits the further improvement of forecast accuracy [23]-[25]. At the same time, in the optimization exploration of distribution network connection modes, most of the existing research still remains at the analysis level of a single mode, lacking systematic, data-driven scientific optimization strategies. Most decisions still rely on experience and judgment, and fail to fully tap the potential of data resources [26], [27]. In view of this, this paper is committed to integrating big data technology, machine learning and spatial analysis methods to implement a comprehensive optimization strategy for grid spatial load forecasting and medium-voltage distribution network wiring mode [28]-[30], aiming to comprehensively improve the accuracy of load forecasting. properties, and provide scientific basis for optimal decision-making of wiring modes.

The core research goal of this paper is to apply a new paradigm of grid spacial load forecasting driven by big data and simultaneously optimize the wiring mode of the medium-voltage distribution network. To achieve this goal, this paper follows a series of rigorous research steps. Starting from data integration and preprocessing, the deep laws of load changes are captured through machine learning modeling [31]-[33]. Real-time data stream processing technology is used to ensure that the prediction model can flexibly respond to the ever-changing power demand and enhance the response speed and adaptability of the system. With the help of spatial analysis technology, the intrinsic relationship between regional characteristics and load changes is deeply explored, thereby further improving the spatial resolution and accuracy of the prediction [34], [35]. At the level of wiring mode optimization, genetic algorithms are adopted. This series of research results not only provides valuable references for the planning and construction of smart grids, but also lays a solid foundation for promoting the modernization and transformation of power systems and responding to future energy challenges. Through the in-depth exploration of this study, this paper hopes to open up new horizons for the power industry, stimulate the vitality of technological innovation and management optimization, and promote the sustainability of power supply and the efficiency of resource allocation, contributing to the construction of a greener and smarter power system.

II. Grid Spacial Load Forecasting and Wiring Mode Optimization

II. A. Data Integration and Processing

II. A. 1) Data Collection and Preprocessing

The historical load data comes from the power company database, covering the hourly data of the past five years, including seasonal and holiday electricity consumption, to reflect the seasonal changes in load. Meteorological data is obtained through the National Meteorological Administration and online platforms, including key variables such as temperature and humidity, which have an important impact on power load. Geographic information data is provided by GIS (Geographic Information System). Combined with power load data, the impact of geographical features on load is analyzed. User behavior data relies on real-time records of smart meters and is aggregated and stored by power companies. This paper uses the Apache Hadoop framework to process data. First, MapReduce is used to clean the data and remove missing and outliers; the mean interpolation is used for missing values; outliers are identified by Z-score. Hive SQL is used to remove duplicates to ensure that load records are unique. The Z-score method is used for data standardization to eliminate the impact of dimensions and prepare for model training.

II. A. 2) Data Integration

The Apache Spark's DataFrame API (Application Programming Interface) is used to merge the processed load data, meteorological data, geographic information data, and user behavior data, and connect them according to timestamps and regional IDs to ensure data consistency and integrity. During the data integration process, new features are built for forecasting needs, and features such as "temperature change rate" and "humidity change rate" are extracted from meteorological data, and features such as "peak hour power consumption ratio" are generated based on user behavior data. These new features are designed to improve the model's ability to explain load changes and enhance the accuracy of forecasts.

After data integration, large-scale data that has been processed multiple times requires efficient storage and management. This paper chooses to use HDFS (Hadoop Distribute File System) as the main data storage system to ensure high availability and scalability of data. At the same time, Hive is used for metadata management to facilitate subsequent data query and analysis. During data storage, this paper regularly backs up and synchronizes data to prevent data loss. At the same time, a data access permission management mechanism is established to ensure the security of sensitive data.

II. A. 3) Real-time Data Processing

In order to enhance the accuracy and immediate response capability of load forecasting, this study builds a real-time data stream processing architecture and adopts Apache Kafka and Apache Spark Streaming technologies to realize real-time data processing. In the data access link, Kafka efficiently aggregates real-time load data, meteorological information and user behavior data with its excellent high throughput and low latency characteristics. Subsequently, in the data processing stage, Spark Streaming is used for real-time analysis and calculation, precisely capturing load change trends through sliding window technology. In addition, the real-time processing results can be immediately fed back to the load forecasting model, driving the dynamic update of the model to ensure that it is highly adaptable to the rapidly changing load environment. When new data arrives, the system intelligently triggers the update mechanism and automatically adjusts the model parameters to precisely reflect the latest load characteristics.

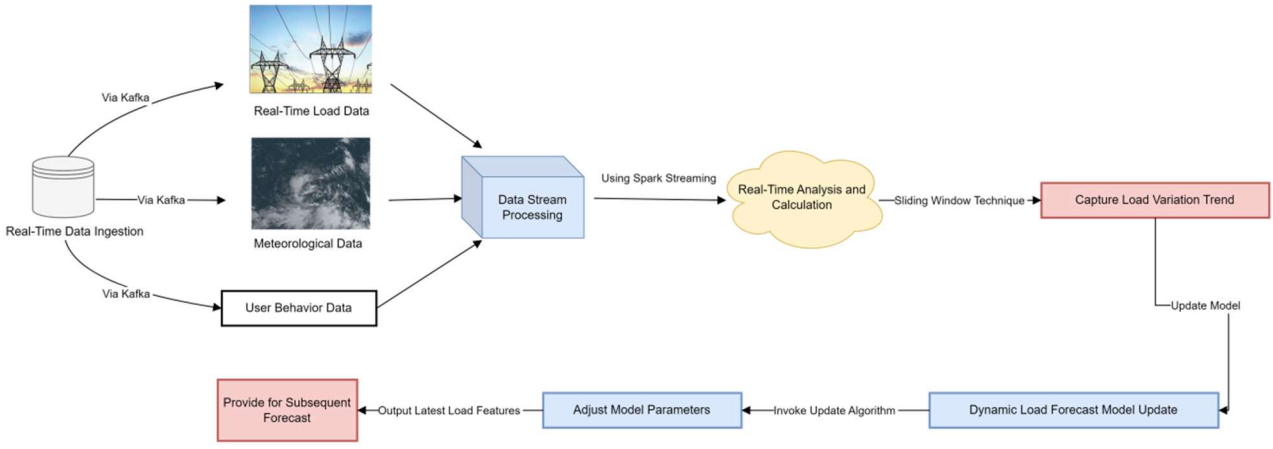

Figure 1: Real-time data stream processing system

Figure 1 shows the components of the real-time data stream processing system. Real-time data access is performed through Kafka, including load data, meteorological data, and user behavior data. All data are aggregated in the data stream processing stage, and Spark Streaming is used for real-time analysis and calculation. The sliding window technology is applied to timely capture the load change trend. The processing results are used to dynamically update the load forecasting model, and the model parameters are automatically adjusted by calling the update algorithm.

II. B.Machine Learning Modeling

II. B. 1) Random Forest Model

Before model training, the data set is divided into a training set (70%), a test set (20%), and a 10% validation set (10%). Stratified sampling is used to ensure balanced distribution. The random forest model is initialized using sk learn ensemble, Random Forest Regressor. After setting the hyperparameters, feature importance is evaluated

and the model is streamlined through feature_importances_. GridSearchCV is combined with cross-validation to optimize the hyperparameters and select the combination with the smallest MSE.

II. B. 2) LSTM Network Model

The training set is scaled to [0, 1] by Min-Max. Keras is used to build the input layer, LSTM layer, and fully connected layer of the LSTM model. The LSTM layer has multiple layers and 50 units, and retains time series information. The fully connected layer uses the ReLU activation function to output the load forecast value. Keras Tuner optimizes the hyperparameters and evaluates the performance through the validation set MSE.

II. B. 3) Selection of Two Models

After all models are trained and evaluated, this paper compares the performance indicators of the random forest model and the LSTM model, and selects the model with better performance as the final load forecasting model and the other model as an auxiliary model. The forecast results of the two models are visualized, and the accuracy of the forecast is intuitively displayed by comparing the actual load with the predicted load.

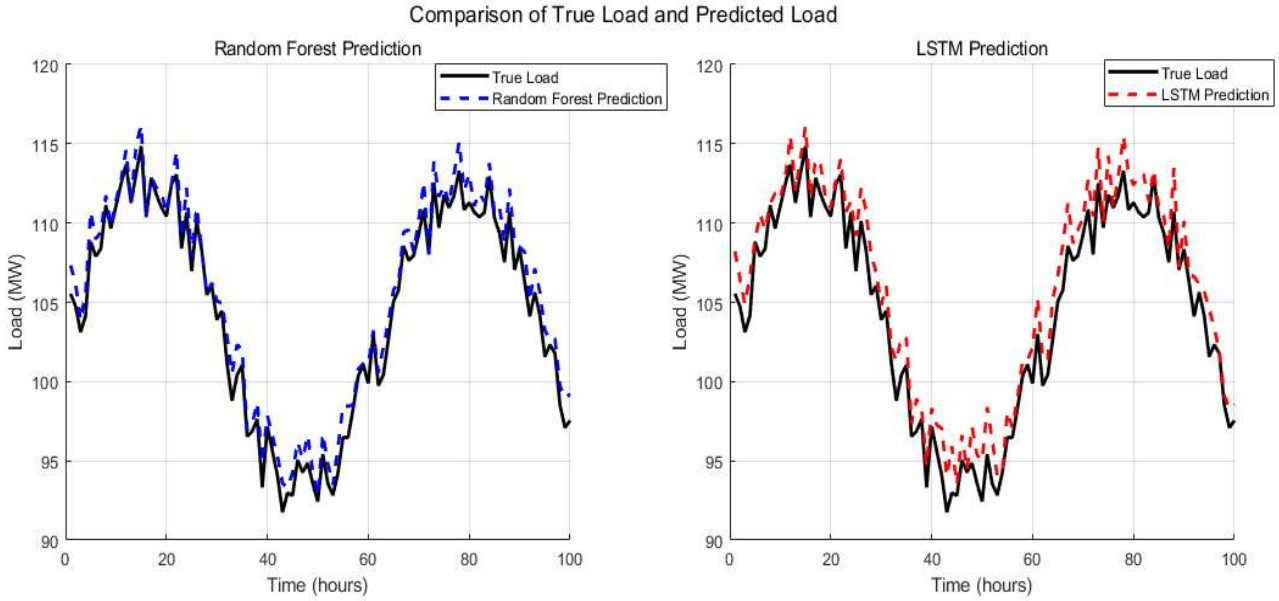

Figure 2: Model prediction results for load

Figure 2 shows the load forecast of the random forest model on the left and the forecast of the LSTM model on the right. The horizontal axis is time and the vertical axis is load (MW). In the left figure, the black solid line represents the natural fluctuation of the true load, which fluctuates regularly with a 100-hour cycle. The blue dotted line is the random forest prediction value. The overall trend is consistent with the true load, but there are deviations at certain points in time, indicating that the model has certain limitations in capturing sudden changes. The LSTM model diagram on the right of Figure 2 also uses a solid black line to represent the true load, and a red dotted line to represent the LSTM prediction result. The LSTM model is relatively sensitive in responding to short-term load changes, but in multiple time periods, the predicted values slightly deviate from the true values, showing the model's shortcomings in processing volatile data. This comparison highlights the advantages of the random forest model in overall prediction accuracy. Ultimately, based on the results of cross-validation and the evaluation of the test set, this paper determines the random forest model as the best load forecasting model and the LSTM model as an auxiliary model.

II. C.Real-time Data Stream Processing

II. C. 1) Construction of Data Stream Monitoring System

In order to realize the monitoring and processing of real-time load data, this study uses Apache Kafka as the data stream processing platform. The specific steps include: first, a Kafka cluster is built; one or more Kafka proxy nodes are deployed to ensure high availability and powerful data processing capabilities; Zookeeper is used for cluster management and coordination. Secondly, multiple topics are created in Kafka, `real_time_load` and `environment_data`, which are used to receive real-time load data and environmental change data respectively.

Using the Kafka command-line tool, these topics are created and the appropriate number of partitions and replicas are set to improve data throughput and fault tolerance. Finally, a data producer application is developed to use Kafka's Producer API to send load data and environmental data to the corresponding topics in real-time. The data producer obtains real-time data from power monitoring equipment and meteorological monitoring stations, and sends it to Kafka after formatting it in JSON format to ensure the accuracy and timeliness of the data stream.

After the real-time data is accessed, the system needs to input it into the trained load forecasting model. The specific steps are as follows: first, Kafka's Consumer API is used to develop a data consumer application and subscribe to the aforementioned topics in real-time to receive newly generated load data and environmental change data. The received JSON format data is parsed to extract relevant features, such as timestamp, load value, temperature, humidity, etc. Before inputting the real-time data into the model, the data needs to be normalized and missing values are processed. The same Min-Max scaling method as in the training phase is used to scale the data to the [0, 1] range. In load forecasting, it is usually necessary to analyze the relationship between load data and environmental data. The covariance formula is as shown in Formula 1:

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y}) \tag{1}$$

The correlation coefficient formula is as Formula 2:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{2}$$

In Formulas 1 and 2: X and Y are two variables, load and temperature. x_i and y_i are the i-th observations. $\bar{x}$ and $\bar{y}$ are the means of X and Y. σ_X and σ_Y are the standard deviations of X and Y. $Cov(X,Y)$ is the covariance, which measures the common variation trend of the two variables. $\rho(X,Y)$ is the correlation coefficient, which ranges from [-1,1] and indicates the linear correlation between the two variables.

II. C. 2) Load Forecast Update

After the real-time data processing is completed, the data is immediately injected into the pre-trained load forecasting model to perform the immediate load forecasting task. This process starts with loading the saved model through the consumer application, using the joblib or pickle module to achieve persistent storage and efficient loading of the model, ensuring that the model can quickly enter the working state. Subsequently, the processed real-time data is input into the loaded model, and the load forecast value is quickly generated by calling its predict function. This step ensures the system's rapid response and accurate prediction of real-time data changes. The prediction results and their timestamp information are properly stored in the MongoDB database. In order to improve data access efficiency and flexibility, SQLAlchemy is used as a bridge for database operations to simplify data management.

In order to intuitively display the real-time load forecast, the system builds a visual dashboard, and the Grafana tool is used to communicate with the database in real-time to ensure the instant update and accuracy of the data. The dashboard interface is carefully designed to intuitively display key indicators such as real-time load forecast trends, actual load conditions, and environmental variables. In addition, the system has a built-in threshold monitoring mechanism. Once the predicted load reaches the preset safety limit, the system immediately triggers an alarm to notify relevant personnel to respond quickly.

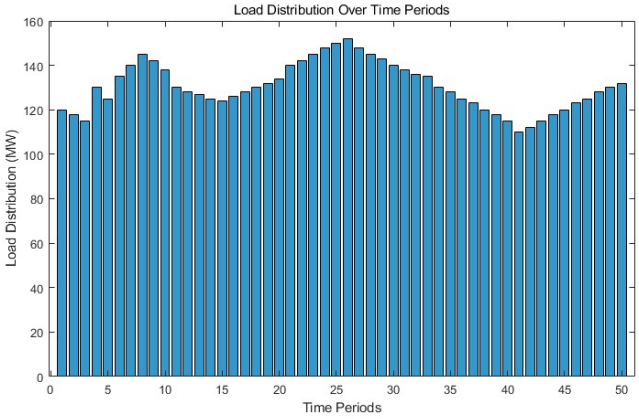

Figure 3: Load distribution within a time period

Figure 3 shows the load distribution within 50 time periods. The load value fluctuates between 110MW and 152MW, reflecting the change trend of power load over time. The load shows a gradual upward trend in the early time period, reaching 145MW. The load climbs again at the 15th time point, reaching a maximum of 152MW, showing the peak period of power demand. This change pattern may be related to the power demand in daily life, such as the difference between working days and weekends, morning and evening peaks and valleys, indicating the load pressure of the system during peak periods and idle resources during low peak periods. By monitoring these data in real- time, the system can manage power resources and ensure the stability of power supply.

In order to ensure the stability and real-time performance of the system under high load conditions, system performance optimization must be carried out. By adjusting the configuration of the Kafka consumer, the `max.poll.records` parameter is set to achieve batch processing, improve data processing efficiency, and reduce the delay of each processing. The asynchronous processing mode is adopted to ensure the decoupling between real-time data processing and load forecasting, and the asynchronous consumption of real-time data is realized by using message queues, which not only improves the throughput of the system, but also enhances the flexibility and response speed of processing. In order to enhance the robustness of the system, Kafka's retry mechanism and dead letter queue are implemented to ensure that when an exception occurs during data processing, the system can automatically recover and reprocess the unsuccessful data, ensuring the reliability and stability of the system in complex environments.

II. D.Spatial Analysis Application

II. D. 1) Spatial Feature Identification

Before conducting spatial exploration, the integration and preprocessing of load data and geographic information data are indispensable steps. Based on the regional ID, this paper integrates the previously processed load data with geographic information data to construct a comprehensive data set covering multiple information such as load value, geographic coordinates, and population density. This integration lays a solid foundation for subsequent spatial exploration. Using the Geopandas library, this paper converts the integrated data into GeoDataFrame format for easy spatial analysis, and ensures that the geographic coordinate system of all data is unified to WGS84 (World Geodetic System-1984 Coordinate System) to ensure the accuracy of geospatial processing. In the process of data merging, if missing values are encountered, this paper adopts interpolation or mean substitution strategies to ensure the integrity of the data.

This study uses GIS technology and combines the K-Means clustering algorithm to divide the region into several clusters, and determines the optimal number of clusters through the elbow rule. At the same time, this paper conducts time series analysis on the load data of each region to gain in-depth insights into its seasonal and trend changes.

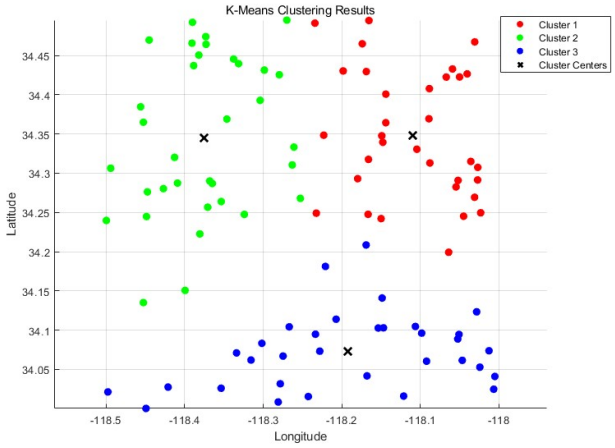

Figure 4: Spatial analysis of load data by K-Means clustering algorithm

In Figure 4, the load data is spatially analyzed using the K-Means clustering algorithm. Figure 4 shows the longitude and latitude distribution of 100 data points, each representing the load situation of a region. The latitude ranges from 34 degrees to 34.5 degrees, and the longitude ranges from -118 degrees to -118.5 degrees. By setting 3 clusters, the algorithm divides these regions into three categories, marked with red, green and blue, showing regions with similar load characteristics. The cluster center marks the center of each cluster. The degree of load

concentration can be identified through visualization, and the load differences between different regions can be understood, thereby providing data support for load management and power grid optimization.

II. D. 2) Spatial Regression Model Construction

In order to quantify the role of spatial characteristics in load forecasting, this study uses the Geographically Weighted Regression (GWR) Model for modeling and analysis. First, the GWR model is constructed, and the load value is set as the response variable, while geographic characteristics such as population density, distance to major power supply facilities, and land use type are used as explanatory variables. With the help of the GWR library in Python, the corresponding geographic and load data are input to ensure that the model can effectively capture spatial heterogeneity. Subsequently, the least squares method is used to estimate the model parameters and precisely calculate the effects of each explanatory variable in different geographical locations. In this process, the model generates a unique set of local regression coefficients for each geographical location to reflect the spatial impact of the explanatory variables on load forecasting.

When deeply analyzing the results of the GWR model, this paper first spatially visualizes the regression coefficients and maps them to the geographic space with the help of GIS tools to intuitively show the degree of influence of explanatory variables on load in different regions. Through map analysis, key spatial features are identified to provide strong support for decision making. In addition, this paper also conducts spatial heterogeneity analysis to explore the complex relationship between load and explanatory variables in each region.

II. E. Wiring Mode Optimization

II. E. 1) Construction of Medium-voltage Distribution Network Wiring Model

In the first step of wiring mode optimization, the medium-voltage distribution network wiring model is constructed according to the load forecast results and the current status of the power grid. First, the basic information of the existing power grid is collected through the geographic information system GIS, which includes the capacity, load distribution, user distribution and line parameters of each substation.

Table 1: Basic information of each substation in the medium-voltage distribution network

Substation ID	Substation Capacity (MVA)	Load Distribution (kW)	Number of Users	Line Type
1	20	1500	100	ACSR 25mm²
2	30	2500	150	ACSR 35mm²
3	15	1200	80	ACSR 30mm²
4	25	1800	120	ACSR 40mm²
5	10	800	60	ACSR 25mm²
6	18	1300	90	ACSR 30mm²
7	22	1600	110	ACSR 35mm²
8	12	900	70	ACSR 30mm²
9	20	1400	95	ACSR 25mm²
10	30	2700	175	ACSR 40mm²

Table 1 shows the basic information of each substation in the medium-voltage distribution network, including substation number, capacity, load distribution, number of users, and line type used. These data provide data support for load forecasting and wiring mode optimization. Among them, substation 1 has a capacity of 20 MVA and a load distribution of 1500 kW, serves 100 users, and uses ACSR 25mm² conductors, indicating that it has a strong power supply capacity. In contrast, substation 5 has a smaller capacity of only 10 MVA, but its load distribution is 800 kW, serving 60 users, showing a lower power supply pressure. Overall, the load utilization rate of the station is high and its power supply is stable. At the same time, the type of ACSR (Aluminum Conductor Steel Reinforced) conductor used in each substation also reflects different load carrying capacities. Through data identification of load concentration areas, the foundation is laid for optimizing the wiring scheme of the medium-voltage distribution network.

Next, the power system modeling software is used to design and construct the wiring model of the medium-voltage distribution network. The model incorporates each node substation, distribution box, user terminal and line into it, and truly reflects the topological structure of the power grid. According to the collected data, the electrical parameters set include: ACSR is selected as the conductor type; the cross-sectional area is 50 mm²; the line length is 1000 m; the resistance value is calculated to be 0.564 Ω; the inductive reactance is 125.6 Ω; the rated voltage is 10 kV; the rated current is about 288.68 A. The precise setting of these parameters ensures that the model can accurately reflect the electrical characteristics of the power grid.

II. E. 2) Optimization Algorithm Selection and Application

Next, this paper uses genetic algorithm (GA) to optimize the wiring mode and find the best wiring scheme. First, the basic framework of the genetic algorithm is built, including individual encoding, fitness function, selection, crossover and mutation operations. In this process, this paper uses binary encoding to represent the various line configurations in the wiring scheme to ensure that the encoding form is concise and easy to handle.

Subsequently, the fitness function is designed to evaluate the advantages and disadvantages of different wiring schemes. The fitness function mainly considers three aspects. The first is the power supply efficiency, which is evaluated by calculating the energy loss of the entire distribution network. The smaller the loss, the higher the fitness score. The second is the load balance, which evaluates the load balance of each node. The more balanced the load, the higher the fitness score, avoiding local overload. The final is to consider the performance of the wiring scheme under fault conditions, which can maintain power supply in the event of a fault, with a higher fitness score. The parameters of the genetic algorithm are set, including population size, crossover probability, and mutation probability. The setting of these parameters combines experience and previous research results to determine the appropriate range, thereby improving the convergence speed and optimization effect of the algorithm. Through the above steps, this paper constructs a systematic genetic algorithm framework that can effectively optimize and analyze the wiring mode of the medium-voltage distribution network.

In the optimization process, this paper executes the genetic algorithm to find the best wiring scheme. The specific steps are as follows: first, multiple wiring schemes are randomly generated as the initial population of the genetic algorithm. Each scheme is composed of different line configurations to ensure the diversity of the population. Next, the fitness value is calculated for each individual scheme and evaluated using the previously designed fitness function. By calculating the fitness value, the better individuals are selected.

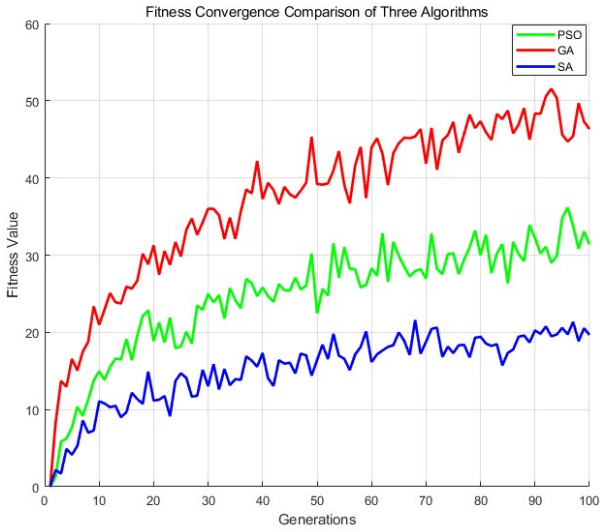

Figure 5: Comparison of fitness convergence of three algorithms

Figure 5 shows the fitness convergence comparison of GA, Particle Swarm Optimization (PSO) and Simulated Annealing (SA) in load forecasting. GA reaches a fitness of 46 in the 100th iteration, showing the fastest convergence speed and the highest fitness value, indicating its ability to quickly find the optimal solution in a short generation. In contrast, PSO converges slowly, while SA converges the slowest. Overall, GA is superior to PSO and SA, has stronger global search capabilities and optimization efficiency, and is suitable for load forecasting tasks in smart grid distribution networks.

In the selection operation, this paper adopts methods such as roulette selection to select better individuals from the current population to form a new population to maintain the inheritance of excellent genes. The selected individuals are crossover and mutation operations are performed to generate new individuals. The crossover operation can be achieved by single-point crossover or multi-point crossover, while the mutation operation can be achieved by randomly changing the state of a certain line. The whole process is iterated multiple times, including fitness evaluation, selection, crossover and mutation, until the preset stop condition is reached and the fitness converges or the maximum number of iterations is reached.

III. Evaluation Indicators and Calculation Methods

III. A. MAE of Load Forecasting

In order to evaluate the accuracy of load forecasting, a data set of the predicted values generated by the model and the actual load values is collected to obtain the MAE (Mean Absolute Error) value. The smaller the MAE value, the higher the accuracy of the forecasting model.

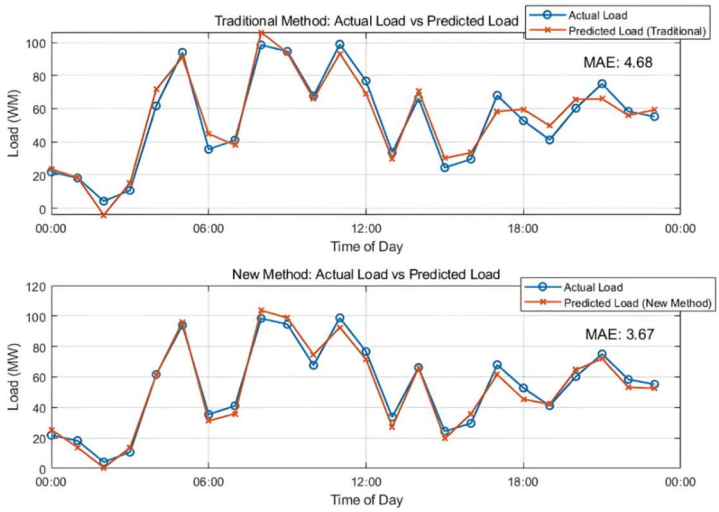

Figure 6: Prediction effect of load data

The upper figure of Figure 6 shows the comparison between the predicted load of the traditional method and the predicted load of the method in this paper. The actual load data is the recorded 24-hour load value, which ranges from 0MW to 100MW. According to the calculation, the MAE of the traditional method is 4.68MW, which reflects that the method has a large deviation in capturing load changes. The lower figure of Figure 6 shows the prediction results of the proposed method, which makes the predicted load curve closer to the actual load curve. The predicted value of the new method in this paper has a smaller error than the actual value. The results show that the MAE is 3.67MW. This method reduces the error by 21.58%, indicating that the proposed method has achieved significant improvements in load forecasting.

III. B. Accuracy of Load Forecasting Model

The datasets of predicted values and actual values are prepared to ensure that the number of data points is consistent between the two, and the mean square error (MSE) is calculated. The smaller the RMSE value, the higher the prediction accuracy of the model.

Table 2: Related error analysis

Time Point	Actual Load (kW)	Predicted Load (kW)	Prediction Error (kW)	Square Error (kW^2)
0:00	50	52	2	4
1:00	48	46	-2	4
2:00	55	58	3	9
3:00	53	51	-2	4
4:00	49	50	1	1
5:00	56	54	-2	4
6:00	60	61	1	1
7:00	62	60	-2	4
8:00	58	59	1	1
9:00	54	53	-1	1
10:00	65	66	1	1
11:00	70	68	-2	4
Total				38

Table 2 shows the evaluation results of the load forecasting model, including the comparison between actual load and predicted load and the related error analysis. The time points in Table 2 are from 00:00 to 11:00. The actual load data fluctuates between 48 and 70 kW, while the predicted load is between 46 kW and 68 kW, indicating that the forecasting model performs better in certain periods. The prediction error column shows that the error is between ±3 kW. The square error reflects the degree of deviation of the model in a specific period by squaring the prediction error at each time point. According to the error calculation in the table, the total square error is 38 kW² and the RMSE is 1.78kw, indicating that the overall prediction accuracy of the model is good.

III. C. R^2 Determination Coefficient

The R^2 determination coefficient is used to evaluate the explanatory power of the load forecasting model for data variation. The closer the R^2 value is to 1, the stronger the model's ability to explain data variation is, reflecting the model's fitting effect and prediction ability.

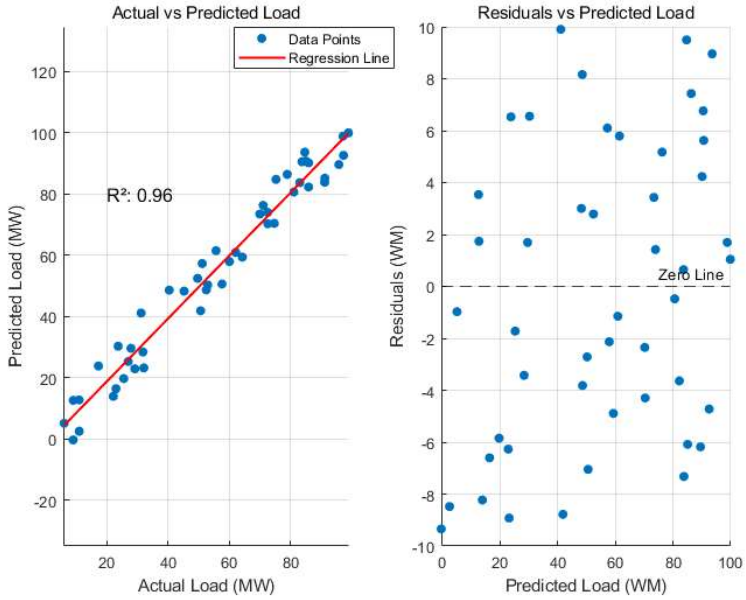

Figure 7: Evaluation results of the load forecasting model

Figure 7 presents the evaluation results of the load forecasting model. The left side is a scatter comparison chart of actual load and predicted load. The horizontal axis represents the actual load value and the vertical axis represents the predicted load value. The density of data points intuitively reflects the fit of the model. The red regression line in the figure clearly reveals the close linear relationship between the predicted value and the actual value, with R^2 as high as 0.96, demonstrating the model's powerful analytical power and accurate prediction performance for data fluctuations. The residual graph on the right shows the error distribution between the predicted load and the actual load. The horizontal axis corresponds to the predicted load and the vertical axis is the residual. The error distribution is balanced and has no significant systematic deviation, which further confirms the robustness of the model.

III. D. Wiring Efficiency Improvement Ratio

In order to evaluate the effect of wiring mode optimization, this paper first needs to calculate the loss of the power supply system and define the loss before and after optimization. The calculation of loss is usually based on the actual load and line characteristics, and is estimated by multiplying the current and resistance. On this basis, this paper calculates the efficiency improvement ratio by comparing the difference in loss before and after optimization.

Figure 8 shows the change in power loss over a 12-month period before and after the wiring pattern optimization. The horizontal axis represents time (months) and the vertical axis represents power supply loss. The loss data before optimization is relatively high, all around 150MW, and shows a certain fluctuation, reflecting that the system is less efficient at this stage. In contrast, the loss data after optimization is significantly reduced, with losses around 100MW and a small fluctuation, indicating that the optimization measures effectively improve the power supply efficiency, and the loss is reduced by 33.16% after 12 months of optimization.

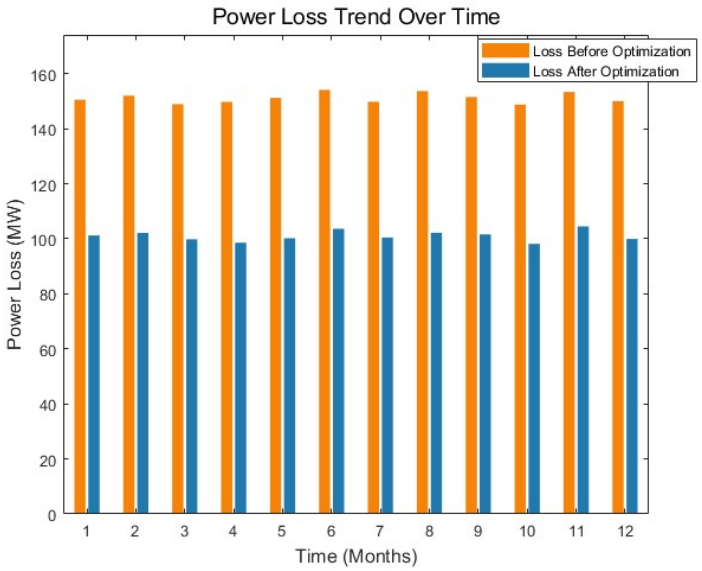

Figure 8: Change in power supply loss

III. E. Economic Benefit Evaluation

When evaluating the economic benefits after optimizing the wiring mode, the cost changes before and after optimization are analyzed by collecting and comparing the data of power supply cost and maintenance cost. The specific methods include calculating the difference in total cost, expressing the economic benefit in percentage, conducting life cycle cost analysis at the same time, considering factors such as equipment depreciation and energy efficiency improvement, and evaluating the additional income brought by the improvement of power supply efficiency, such as reducing power outage losses. Through these comprehensive analyses, the actual economic benefits of wiring mode optimization are quantified.

Table 3: Changes in various costs before and after wiring mode optimization

Cost Category	Cost Before Optimization (Units: Yuan)	Cost After Optimization (Units: Yuan)	Cost Difference (Units: Yuan)	Lifecycle Cost (Units: Yuan)
Power Supply Cost	150,500.75	120,250.50	-30,250.25	400,000.00
Maintenance Cost	50,300.20	30,100.40	-20,199.80	100,500.00
Equipment Depreciation	20,800.10	15,200.75	-5,599.35	60,000.00
Energy Efficiency Cost	10,600.00	7,500.85	-3,099.15	30,800.00
Power Outage Loss	25,400.00	5,300.25	-20,099.75	15,200.00
Total	257,601.05	178,352.75	-79,248.3	606,500.00

Table 3 shows the changes in various costs before and after the wiring mode optimization, highlighting the economic benefits of optimization measures. The power supply cost is reduced by 30,250.25 yuan; the maintenance cost is reduced from 50,300.20 yuan to 30,100.40 yuan; the equipment depreciation is also significantly reduced, from 20,800.10 yuan to 15,200.75 yuan, showing the effectiveness of optimization in various expenditures; the energy efficiency cost is reduced from 10,600.00 yuan to 7,500.85 yuan, a decrease of 3,099.15 yuan; the power outage loss drops from 25,400.00 yuan to 5,300.25 yuan, a relatively large decrease, which shows that the optimization measures have significantly improved the power supply efficiency and reduced the economic losses caused by the power outage. Taken together, the total cost drops from 257,601.05 yuan to 178,351.75 yuan, a difference of 79,248.3 yuan, and the overall decrease is 30.76%.

IV. Conclusions

This paper combines big data technology and advanced machine learning algorithms to apply a grid spatial load forecasting method and a medium-voltage distribution network wiring mode optimization scheme. By integrating multi-source data such as historical load, meteorology and geography, and applying random forest and long short-term memory network (LSTM) models, this paper significantly improves the accuracy of load forecasting and

significantly reduces MAE. In addition, after optimizing the wiring mode, power supply losses are significantly reduced and economic benefits are significantly improved. Although this study has achieved positive results in load forecasting and wiring mode optimization, there are still problems such as strong data dependence and insufficient model interpretability. Future research can further explore multi-model fusion, real-time data analysis technology, and applications in different scenarios to improve the robustness and applicability of the model and provide more comprehensive support for the development of smart grids.

Funding

This research was funded by the Science and Technology Project of State Grid Corporation of China, Research on Cooperative Control Technology for Distribution Network Balance Zones Considering Flexible Regulation of Power Sources and Loads, grant number B3018K230008.

References

[1] Ali M, Adnan M, Tariq M, et al. Load forecasting through estimated parametrized based fuzzy inference system in smart grids[J]. IEEE Transactions on Fuzzy Systems, 2020, 29(1): 156-165.
[2] Udo W S, Kwakye J M, Ekechukwu D E, et al. Smart grid innovation: machine learning for real-time energy management and load balancing[J]. International Journal of Smart Grid Applications, 2024, 22(4): 405-423.
[3] Wu Q, Chen X, Zhou Z, et al. Deep reinforcement learning with spatio-temporal traffic forecasting for data-driven base station sleep control[J]. IEEE/ACM transactions on networking, 2021, 29(2): 935-948.
[4] Udo W S, Kwakye J M, Ekechukwu D E, et al. Smart grid innovation: machine learning for real-time energy management and load balancing[J]. International Journal of Smart Grid Applications, 2024, 22(4): 405-423.
[5] Si C, Xu S, Wan C, et al. Electric load clustering in smart grid: Methodologies, applications, and future trends[J]. Journal of Modern Power Systems and Clean Energy, 2021, 9(2): 237-252.
[6] Omitaomu O A, Niu H. Artificial intelligence techniques in smart grid: A survey[J]. Smart Cities, 2021, 4(2): 548-568.
[7] Wang C, Wang Y, Ding Z, et al. A transformer-based method of multienergy load forecasting in integrated energy system[J]. IEEE Transactions on Smart Grid, 2022, 13(4): 2703-2714.
[8] Jahangir H, Tayarani H, Gougheri S S, et al. Deep learning-based forecasting approach in smart grids with microclustering and bidirectional LSTM network[J]. IEEE Transactions on Industrial Electronics, 2020, 68(9): 8298-8309.
[9] Li J, Deng D, Zhao J, et al. A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network[J]. IEEE Transactions on Industrial Informatics, 2020, 17(4): 2443-2452.10.1109/TII.2020.3000184
[10] Li G, Li Y, Roozitalab F. Midterm load forecasting: A multistep approach based on phase space reconstruction and support vector machine[J]. IEEE Systems Journal, 2020, 14(4): 4967-4977.
[11] Ngoc T T, Le Van Dai C M T, Thuyen C M. Support vector regression based on grid search method of hyperparameters for load forecasting[J]. Acta Polytechnica Hungarica, 2021, 18(2): 143-158.
[12] Lv L, Wu Z, Zhang J, et al. A VMD and LSTM based hybrid model of load forecasting for power grid security[J]. IEEE Transactions on Industrial Informatics, 2021, 18(9): 6474-6482.
[13] Aprillia H, Yang H T, Huang C M. Statistical load forecasting using optimal quantile regression random forest and risk assessment index[J]. IEEE Transactions on Smart Grid, 2020, 12(2): 1467-1480.
[14] Li Z, Li Y, Liu Y, et al. Deep learning based densely connected network for load forecasting[J]. IEEE Transactions on Power Systems, 2020, 36(4): 2829-2840.
[15] Ali S, Riaz S, Liu X, et al. A Levenberg–Marquardt based neural network for short-term load forecasting[J]. Computers, Materials and Continua, 2023, 75(1): 1783-1800.
[16] Janjua J I, Ahmad R, Abbas S, et al. Enhancing smart grid electricity prediction with the fusion of intelligent modeling and XAI integration[J]. International Journal of Advanced and Applied Sciences, 2024, 11(5): 230-248.
[17] Jalali S M J, Ahmadian S, Khosravi A, et al. A novel evolutionary-based deep convolutional neural network model for intelligent load forecasting[J]. IEEE Transactions on Industrial Informatics, 2021, 17(12): 8243-8253.
[18] Obst D, De Vilmarest J, Goude Y. Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France[J]. IEEE transactions on power systems, 2021, 36(5): 4754-4763.
[19] Zhang W, Quan H, Gandhi O, et al. Improving probabilistic load forecasting using quantile regression NN with skip connections[J]. IEEE Transactions on Smart Grid, 2020, 11(6): 5442-5450.
[20] Ge Q, Guo C, Jiang H, et al. Industrial power load forecasting method based on reinforcement learning and PSO-LSSVM[J]. IEEE transactions on cybernetics, 2020, 52(2): 1112-1124.
[21] Sharma S, Majumdar A, Elvira V, et al. Blind Kalman filtering for short-term load forecasting[J]. IEEE Transactions on Power Systems, 2020, 35(6): 4916-4919.
[22] Wang J, Chen X, Zhang F, et al. Building load forecasting using deep neural network with efficient feature fusion[J]. Journal of Modern Power Systems and Clean Energy, 2021, 9(1): 160-169.
[23] Ozer I, Efe S B, Ozbay H. A combined deep learning application for short term load forecasting[J]. Alexandria Engineering Journal, 2021, 60(4): 3807-3818.
[24] Zhao W, Li T, Xu D, et al. A global forecasting method of heterogeneous household short-term load based on pre-trained autoencoder and deep-LSTM model[J]. Annals of Operations Research, 2024, 339(1): 227-259.
[25] Von Krannichfeldt L, Wang Y, Hug G. Online ensemble learning for load forecasting[J]. IEEE Transactions on Power Systems, 2020, 36(1): 545-548.
[26] Wen X, Shen Q, Zheng W, et al. AI-driven solar energy generation and smart grid integration a holistic approach to enhancing renewable energy efficiency[J]. International Journal of Innovative Research in Engineering and Management, 2024, 11(4): 55-66.

[27] Wang Y, Chen J, Chen X, et al. Short-term load forecasting for industrial customers based on TCN-LightGBM[J]. IEEE Transactions on Power Systems, 2020, 36(3): 1984-1997.
[28] Hein K, Xu Y, Wilson G, et al. Coordinated optimal voyage planning and energy management of all-electric ship with hybrid energy storage system[J]. IEEE Transactions on Power Systems, 2020, 36(3): 2355-2365.
[29] Afrasiabi M, Mohammadi M, Rastegar M, et al. Deep-based conditional probability density function forecasting of residential loads[J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3646-3657.
[30] Sun C, Ning Y, Shen D, et al. Graph Neural Network-Based Short-Term Load Forecasting with Temporal Convolution[J]. Data Science and Engineering, 2024, 9(2): 113-132.
[31] Dudek G, Pełka P, Smyl S. A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(7): 2879-2891.
[32] Badr M M, Mahmoud M M E A, Fang Y, et al. Privacy-preserving and communication-efficient energy prediction scheme based on federated learning for smart grids[J]. IEEE Internet of Things Journal, 2023, 10(9): 7719-7736.
[33] Lee Y G, Oh J Y, Kim D, et al. Shap value-based feature importance analysis for short-term load forecasting[J]. Journal of Electrical Engineering & Technology, 2023, 18(1): 579-588.
[34] Fan G F, Guo Y H, Zheng J M, et al. A generalized regression model based on hybrid empirical mode decomposition and support vector regression with back-propagation neural network for mid-short-term load forecasting[J]. Journal of Forecasting, 2020, 39(5): 737-756.
[35] Hammad M A, Jereb B, Rosi B, et al. Methods and models for electric load forecasting: a comprehensive review[J]. Logist. Sustain. Transp, 2020, 11(1): 51-76.