Applied Energy

An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization

--Manuscript Draft--

Manuscript Number:	APEN-D-22-00949R1
Article Type:	Research Paper
Keywords:	Electricity load forecast; Fuzzy information particles; Combination optimization strategy; Point-interval prediction system
Corresponding Author:	Jianzhou Wang, Ph.D. Dongbei University of Finance and Economics Dalian, CHINA
First Author:	Kang Wang
Order of Authors:	Kang Wang
	Jianzhou Wang, Ph.D.
	Bo Zeng
	Haiyan Lu
Abstract:	During an era of rapid growth in electricity demand throughout society, accurate forecasting of electricity loads has become increasingly important to guarantee a stable power supply. Nevertheless, historical models do not address the structure of the data itself, and a single model cannot accurately determine the nonlinear characteristics of the data. This would not allow for accurate and stable predictions. With the aim of filling this gap, this paper proposes an innovative intelligent power load point-interval forecasting system. The system discretizes the time series, then performs efficient dimensionality reduction by fuzzification, and multi-level optimization of five benchmark deep learning models by the proposed multi-objective optimization algorithm, and finally analyzes the uncertainty of the prediction results. Experiments comparing the developed prediction system with other models were conducted on three datasets, and the prediction results were discussed for validation from multiple perspectives. The simulation results show that the proposed model has superior prediction accuracy, robustness and uncertainty analysis capability, and can provide accurate deterministic prediction information and fluctuation interval analysis to ensure the long-term safety and stability and operation of the grid.

Dear Editors:

On behalf of my co-authors, we thank you very much for giving us an opportunity to revise our manuscript, we appreciate editor and reviewers very much for their positive and constructive comments and suggestions on our manuscript entitled “An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization”. (ID: APEN-D-22-00949).

We have studied reviewer’s comments carefully and have made revision which marked in red in the paper. We have tried our best to revise our manuscript according to the comments. Attached please find the revised version, which we would like to submit for your kind consideration.

We would like to express our great appreciation to you and reviewers for comments on our paper. Looking forward to hearing from you.

Thank you very much for your attention and consideration!

Dr. Jianzhou Wang
Corresponding author

Dear editors and reviewers:

Thank you very much for e-mailing us the comments raised by the respected reviewers. The manuscript **No. APEN-D-22-00949 "An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization"** has been revised taking into account all of the helpful comments and suggestions. The details of the comments raised, the answers and the actions taken are presented here. All of the changes made in the revised manuscript have been highlighted in red color. We appreciate for respected editors/reviewers' warm work earnestly, and hope that the correction will meet with approval. We look forward to hearing from you.

Best regards,

Jianzhou Wang

List of Responses are as following:

Comment raised by respected Reviewer 1:

Comment: In the manuscript, Wang et al. propose a new system for power load forecasting, which performs fuzzy granular dimensionality reduction on the data in the data preprocessing stage and optimizes the forecasting results of the benchmark model using a multi-objective optimization algorithm, which effectively improves the forecasting accuracy and stability, and analyzes the forecasting results in terms of determinism and uncertainty. The conclusions drawn by the authors are supported by several experiments. This is a very good and well-thought-out paper on a topic of interest to researchers in related fields. This manuscript is acceptable with minor revisions. My detailed comments are provided below.

Response: Thank you very much for your positive evaluation and valuable advice on our study, and it is quite helpful for improving the quality of our paper. And our manuscript has been carefully revised according to your suggestions.

****

Comment 1: It is well known that the introduction section is intended to integrate all the efforts made by scientists for this prediction problem. In the introduction section of this paper, the four main methodological sections on prediction are well described, but in the integrated model section only data preprocessing methods for data denoising techniques are mentioned, while there is a lack of literature review on fuzzy granulation and dimensionality reduction techniques used in the data preprocessing phase of this paper. Therefore, I suggest adding some literature reviews on such methods, which are not limited to the integrated model, but can be separate discussions.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. We have added a literature review on the use of information granulation as a data pre-processing method to the statement of data denoising techniques based on your suggestion (in p.7). Please see the revised manuscript.

References:

[1] Ding, S., Zhang, X., An, Y., & Xue, Y. (2017). Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification. *Pattern Recognition*, *67*, 32–46. https://doi.org/10.1016/j.patcog.2017.02.011
[2] Velázquez-Rodríguez, J. L., Villuendas-Rey, Y., Yáñez-Márquez, C., López-Yáñez, I., & Camacho-Nieto, O. (2020). Granulation in Rough Set Theory: A novel perspective. *International Journal of Approximate Reasoning*, *124*, 27–39. https://doi.org/10.1016/j.ijar.2020.05.003
****

Comment 2: The purpose of the experiment needs to be stated at the beginning of the experimental section. There are several places in this paper where the purpose of the experiment is not clearly stated and sentences are redundant. In particular, lines 9-16 on page 17 and lines 50-60 on page 30.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. We modified the text of the first paragraph(in section 4.3.1, section 5.2) by removing the redundant sentences and adding the purpose of the discussion on Experiment 1 and the Improvement ratio of the indexes. Please see the revised manuscript.

Comment 3: The ", and finally uncertainty analysis." in the summary is grammatically incorrect and could be changed to ", and finally analyzes the uncertainty of the prediction results.".

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we have revised the ", and finally uncertainty analysis." to ", and finally analyzes the uncertainty of the prediction results.". Please see the revised manuscript.

****

Comment 4: There are several instances in the text where the initials are not indicated. For example, in Table 1 and 5.1, 'where loss function'.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we checked the initials in the full text and corrected them (because the distribution is fragmented, no specific location is specified here). Please see the revised manuscript.

****

Comment 5: The format of 4.3.2 is not very aesthetically pleasing, and the format of the upper and lower paragraphs of Figure 3 could be appropriately adjusted.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. We have adapted the format of 4.3.2 to your suggestions. Please see the revised manuscript.

****

Comment 6: "The developing FMICM: MAPE, MAE, RMSE, SDE" in 4.3.1(a) is not indicated as a two-step prediction result.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we added " When making a two-step prediction" before " The developing FMICM: " to indicate that it is a two-step prediction. Please see the revised manuscript.

****

Comment raised by respected Reviewer 2:

Comment: This paper proposes an innovative intelligent power load point-interval forecasting system. Many techniques have been used. The topic is interesting. The methods are sound. Some comments are given as follows:

Response: Thank you very much for your positive evaluation and valuable advice on our study, and it is quite helpful for improving the quality of our paper. And our manuscript has been carefully revised according to your suggestions.

****

Comment 1: Why multi-objective Dingo optimization algorithm (MODOA) is used for optimization? Other multi-objective optimization algorithms (e.g., population extremal optimization) may be better candidates to solve the multi-objective optimization problem. The authors are suggested to add some comments to highlight the motivation of using MODOA. The authors can refer to: https://doi.org/10.1016/j.renene.2019.05.024;http://dx.doi.org/10.1016/j.ins.2015.10.010.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we have modified the fourth article of "The drawbacks of these methods are summarized" (in p.4), which summarizes the shortcomings of existing hybrid models in terms of low prediction accuracy and long running time of multi-objective models. Therefore, we propose a new weight optimization algorithm, MODOA, which echoes the fourth article of "The main contributions and innovations of this study", highlighting that the motivation for using MODOA is to further improve the prediction results. In addition, we designed Experiment 3 to compare the proposed MODOA with three existing multi-objective optimization algorithms (MOGOA,MODA,MOALO) and proved that MODOA has the best prediction results. In addition, we refer to the following three papers in [3][41][42]. Please see the revised manuscript.

References:

[1] Chen, M.-R., Zeng, G.-Q., & Lu, K.-D. (2019). Constrained multi-objective population extremal optimization based economic-emission dispatch incorporating renewable energy resources. *Renewable Energy*, *143*, 277–294. https://doi.org/1

0.1016/j.renene.2019.05.024

[2] Zeng, G.-Q., Chen, J., Li, L.-M., Chen, M.-R., Wu, L., Dai, Y.-X., & Zheng, C.-W. (2016). An improved multi-objective population-based extremal optimization algorithm with polynomial mutation. *Information Sciences*, *330*, 49–73. https://doi.org/10.1016/j.ins.2015.10.010

[3] Zeng, G.-Q., Chen, J., Dai, Y.-X., Li, L.-M., Zheng, C.-W., & Chen, M.-R. (2015). Design of fractional order PID controller for automatic regulator voltage system based on multi-objective extremal optimization. *Neurocomputing*, *160*, 173–184. https://doi.org/10.1016/j.neucom.2015.02.051

****

Comment 2: This work proposes the novel system for power load forecasting system. It is good. Can this system extend to other forecasting problems (e.g., wind speed and traffic flow). The authors can refer to: DOI 10.1109/TVT.2019.2952605; DOI 10.1109/JIOT.2019.2913176. Please give some explanation.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. Since the proposed point-interval prediction system can perform deterministic prediction analysis and volatility prediction analysis on time series data with randomness, volatility, periodicity and diversity characteristics, and the proposed system has high prediction stability, the proposed point-interval prediction system can be extended to other prediction problems with time series nonlinear characteristics, such as wind speed prediction, air pollution prediction and traffic flow prediction.We have added this part of the explanation(in section 5.6). In addition, we refer to the following two papers in [21][22]. Please see the revised manuscript.

References:

[1] Chen, M.-R., Zeng, G.-Q., Lu, K.-D., & Weng, J. (2019). A Two-Layer Nonlinear Combination Method for Short-Term Wind Speed Prediction Based on ELM, ENN, and LSTM. *IEEE Internet of Things Journal*, *6*(4), 6997–7010. https://doi.org/10.1109/JIOT.2019.2913176

[2] Zhao, F., Zeng, G. Q., & Lu, K. di. (2020). EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization. *IEEE Transactions on Vehicular Technology*, *69*(1), 101–113. https://doi.org/10.1109/TVT.2019.2952605

****

Comment 3: Different parameters will influence the performance of forecasting models. How to ensure the system achieve the best performance? How to obtain the suitable parameters of all considered models? How to ensure a fair comparison?

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. The models that require parameter tuning in the proposed prediction system are DOA-BPNN,TCN,GRU,DBN,MODOA. The parameter determination of the single model is obtained by the adjustment parameters experiment before the experiment, and the following table lists the adjustment parameters process of TCN.

Embedding size		Batch size		Epochs finetune	
[128,128,128,64]	4.2536%	10	4.2444%	300	4.3679%
[128,128,64,64]	4.1772%	20	**4.1258%**	400	4.2355%
[128,128,64,32]	4.2946%	30	4.624%	500	**4.1258%**
[128,64,32,32]	4.4423%	40	4.4529%	600	4.493%
[128,64,32,16]	**4.1258%**	50	4.1545%	700	4.2498%

We have added notes on the parameter determination in the Notes section of Table 2. In addition, we added sensitivity analysis and convergence analysis in the Discussion section (in sections 5.4 and 5.5), which mainly discusses the parameter sensitivity and convergence of MODOA. Once the parameters of all models are determined, the evaluation index data for the final comparison is the average of five experiments (in Section 4.3) in order to ensure a fair comparison among the models, since the results of each run of each model are different. Please see the revised manuscript.

****

Comment 4: The authors give the results of step 1, step 2 and step 3. What are the main difficulty forecasting different steps?

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper.

The multi-step forecasting method used in this paper is direct multi-step forecasting, and let the delayed forecasting period be h, then the independent variable is $X_1, X_2, X_3, \cdots, X_n$ and the dependent variable is X_{n+h}. One-step forecasting is done when $h=1$, and multi-step forecasting is done when $h \neq 1$.

The advantage of direct multi-step forecasting is its simplicity and avoidance of error accumulation due to recursive multi-step forecasting. The disadvantage is that it requires increased difficulty in model training. The difficulty of multi-step prediction is mainly the need to choose a multi-step prediction method that is suitable for the data studied in this paper. In this paper, the direct multi-step forecasting method is finally chosen through experiments, and the following table lists some of the experimental data(evaluation index is MAPE).

GRU		Step1	Step2	Step3
Site1	DMSF	5.0158	7.0318	**6.9224**

	RMSF	5.0158	**6.8243**	8.2496
Site2	DMSF	3.1566	**4.6607**	**5.0823**
	RMSF	3.1566	4.8029	5.9112
Site3	DMSF	4.9088	**6.4856**	**8.8005**
	RMSF	4.9088	6.7922	9.2300

TCN		Step1	Step2	Step3
Site1	DMSF	4.1258	7.3553	**7.1392**
	RMSF	4.1258	**7.1455**	8.5473
Site2	DMSF	3.2850	**5.2595**	**5.6436**
	RMSF	3.2850	5.4194	6.5177
Site3	DMSF	4.8582	**6.7052**	**8.2805**
	RMSF	4.8582	7.0556	8.7006

Note: DMSF refers to direct multi-step forecasting, RMSF refers to recursive multi-step forecasting.

****

Comment 5: Please read the whole paper again and correct the possible typos.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we re-examined the paper and corrected relevant grammatical and word errors. Please see the revised manuscript.

****

Comment raised by respected Reviewer 3:

Comment: The paper "An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization" proposed an integrated electricity load forecasting system based on data pre-processing and multi-objective optimization. The article adopted fuzzy granulation technique to preprocess data from a new perspective, creatively optimizes BPNN in the model prediction link, proposes a new multi-objective optimization algorithm to fuse multiple benchmark model prediction results, and conducted detailed experiments and discussions on the prediction results. Based on the rich literature cited, the article provided a logical and clear discussion, which has both theoretical research value and practical application value. The following points are prepared for the authors' reference.

Response: Thank you very much for your positive evaluation and valuable advice on our study, and it is quite helpful for improving the quality of our paper. And our manuscript has been carefully revised according to your suggestions.

****

Comment 1: The conclusion should be distinguished from the abstract and results in a way that makes the reader interested in this area of research or the methods used in this article after enjoying the article and makes the reader excited about the prospects of the research. The conclusion written by the authors contains only a summary of the experiments and discussions of the article, without expressing their own insights. It is suggested that the summary of the results should be shortened from the existing conclusions, plus the merits of the proposed model and their own views on the future prospects of power load forecasting and this system.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we modified the conclusion and empirical analysis sections to include in the empirical analysis section that the proposed prediction system has high prediction stability, so that the proposed point-interval prediction system can be extended to other prediction problems with time-series nonlinear characteristics, such as wind speed prediction, air pollution prediction, and traffic flow prediction. Improved aspects of the proposed forecasting system are added in the conclusion section. This will make the reader excited about the prospects of the study. Please see the revised manuscript.

****

Comment 2: For the sake of the standardization of the paper writing, the names about using models should be the same, especially for detail omissions, such as FIG-TCN, which is written as FIG_TCN in some places, and it is recommended that such names are connected with a uniform symbol.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we have unified the symbols by changing the type of " FIG-TCN " to " FIG_TCN ". Please see the revised manuscript.

Comment 3: It is as if the symbols referred to by the upper and lower limits are missing in "Let be the upper limit of the first prediction interval and be the lower limit of the first prediction interval, and the formula of evaluation index is shown in Table 4.".

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. It is an oversight on our part here that we added the symbolic representation to this sentence. The sentence should read " Let $\mathbf{FIL}_i^{\alpha}$ be the upper limit of the first prediction interval and $\mathbf{FIU}_i^{\alpha}$ be the lower limit of the first prediction interval". Please see the revised manuscript.

****

Comment 4: Errors regarding several initial letters. (1) There are several in Table 1. (2) Line 33 on page 11. (3) Lines 9 and 37 on page 29.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we have reviewed the entire text regarding words and grammar and corrected the errors. Please see the revised manuscript.

****

Comment 5: In the abstract, authors mentioned that "Nevertheless, historical models do not address the structure of the data itself, and a single model cannot accurately determine the nonlinear characteristics of the data." Please clarify the meaning of this sentence.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. The historical model for time series prediction is mainly a single model based on a conventional statistical model [1] and an artificial intelligence model [2]. The electric load data has the structure of periodic and nonlinear characteristics [3]. Since the conventional statistical model cannot accurately predict the nonlinear characteristics of the data, the artificial intelligence model can predict the nonlinear characteristics, but the prediction accuracy is not high and the robustness is not strong, so we use an artificial intelligence model hybrid strategy to predict for the nonlinear features of the data.

References:

[1] Pappas, S. Sp., Ekonomou, L., Karampelas, P., Karamousantas, D. C., Katsikas, S. K., Chatzarakis, G. E., & Skafidas, P. D. (2010). Electricity demand load forecasting of the Hellenic power system using an ARMA model. Electric Power Systems Research, 80(3), 256–264. https://doi.org/10.1016/j.epsr.2009.09.006

[2] Li, X., Ma, X., Xiao, F., Xiao, C., Wang, F., & Zhang, S. (2022). Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). *Journal of Petroleum Science and Engineering*, *208*, 109309. https://doi.org/10.1016/j.petrol.2021.109309

[3] Bo, H., Nie, Y., & Wang, J. (2020). Electric Load Forecasting Use a Novelty Hybrid Model on the Basic of Data Preprocessing Technique and Multi-Objective Optimization Algorithm. *IEEE Access*, *8*, 13858–13874. https://doi.org/10.1109/ACCESS.2020.2966641

****

Comment 6: Please add related experiments to analyze the sensitivity and convergence of the proposed model.

Response: Thank you for the valuable advice and it is quite helpful for improving the quality of our paper. According to your suggestions, we added the sensitivity analysis (in section 5.4) and convergence analysis (in section 5.5) of the proposed system.

In section 5.4, for MODOA, the parameters set are Search Number of Individuals, Maximum iterations Number and ArchiveMaxSize. We analyze the stability of the proposed prediction system with respect to changes in parameter values by varying one of the parameters by the control variables method, given that the other two parameters remain unchanged. Section 5.5 The convergence of MODOA is verified and MODOA has a high convergence rate, it can reach convergence in fewer iterations, which further proves the feasibility of its prediction system. Please see the revised manuscript.

****

An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization

Kang Wang[a], Jianzhou Wang[a,*], Bo Zeng[b], Haiyan Lu[c]

[a] *School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*
[b] *School of Management Science and Engineering, Chongqing Technology and Business University, Chongqing, 400067, China*
[c] *School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia*

** Corresponding author. Address: School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*
E-mail address: wjz@lzu.edu.cn

Abstract

During an era of rapid growth in electricity demand throughout society, accurate forecasting of electricity loads has become increasingly important to guarantee a stable power supply. Nevertheless, historical models do not address the structure of the data itself, and a single model cannot accurately determine the nonlinear characteristics of the data. This would not allow for accurate and stable predictions. With the aim of filling this gap, this paper proposes an innovative intelligent power load point-interval forecasting system. The system discretizes the time series, then performs efficient dimensionality reduction by fuzzification, and multi-level optimization of five benchmark deep learning models by the proposed multi-objective optimization algorithm, and finally analyzes the uncertainty of the prediction results. Experiments comparing the developed prediction system with other models were conducted on three datasets, and the prediction results were discussed for validation from multiple perspectives. The simulation results show that the proposed model has superior prediction accuracy, robustness and uncertainty analysis capability, and can provide accurate deterministic prediction information and fluctuation interval analysis to ensure the long-term safety and stability and operation of the grid.

Keywords: *Electricity load forecast;Fuzzy information particles;Combination optimization strategy; Point-interval prediction system;*

1.Introduction

Electricity is the linchpin of the energy system to achieve carbon neutrality, and effectuating the "bi-carbon" goal and implementing a novel electricity system is a tremendously challenging and pioneering strategic and systemic project. Only by embedding a more flexible and interconnected power system can we achieve global electrification when the conditions are right[1]. The future carbon-neutral world will be highly dependent on electricity for energy supply, and electricity will become the pillar of the entire energy system and help society achieve sustainable development. Thereupon, with the development of technology and society, electric power resources become an increasingly important part of human production and life[2].

However, as people's electricity consumption continues to increase and the price of raw materials rises, considerable countries from all around the world are experiencing a shortage of electricity resources[3]. For the sake of avoiding the shortage of electricity resources triggered by short term surges of electricity consumption and unnecessary load loss and investment decisions, short-term electricity load forecasting has become an indispensable part of the national electricity and energy system[4].In summary, accurate power forecasting helps ensure the utilization of electricity, which is critical to the availability and sustainability of the distribution. On the contrary, the lack of accurate forecasting may lead to poor decision making and result in significant losses to the power system[5]. Load forecasting is divided into short-term forecasting for real-time control, medium-term forecasting for energy system operation, and long-term forecasting for extended planning studies. For example, long-term power load forecasts such as predicting annual peak loads for the next few years are used to optimize expansion decisions, while short-term load forecasts(**STLF**) are used for economic dispatch or unit mix studies, such as forecasting load conditions for the next few hours[6]. In order to obtain effective forecasting results, electric load forecasting has been studied intensively. We can broadly classify these forecasting methods into four categories: physical models, conventional statistical models, artificial intelligence models, and hybrid models[7].

The main physical models are the new-generation building energy simulation program (**EnergyPlus**) [8], real-time combined heat and power operational strategy using a hierarchical optimization algorithm[9]. Building operation data are obtained through EnergyPlus and mathematical models related to the physical system are represented. Real-time combined heat and power operational strategy using a hierarchical optimization algorithm considers the transient response of the building and combines the hierarchical CHP optimal control algorithm to achieve a real-time integrated system of electrical load information by running parallel simulations of two transient building models. Nevertheless, as a result of using simulation tools, the physics-based approach is usually difficult to obtain mathematical expressions for various building energy mechanisms and is not effective for short-term predictions.

Conventional statistical models can be used for load forecasting and speculation based on the available and relatively complete historical statistics, which are mechanically processed and organized using certain mathematical methods to reveal the regular links between the variables concerned. Statistical models mainly include ordinary regression models[10], auto-regressive moving average model(**ARMA**)[11] and Auto-regressive integrated moving average model(**ARIMA**)[12]. Since electricity load data have multiple non-linear components, conventional linear regression model treatments either become inaccurate or too complex to be used in practice. Most of the papers are comparing linear regression models with new models to show the advantages of the new model. Pombeiro et al.[13] proposed a nonlinear model based on fuzzy systems and neural networks, which compared with the linear model yielded a much higher prediction accuracy of the new model. Liu et al.[14] developed an autoregressive moving average model by combining it with a generalized auto-decreasing conditional heterogeneity process, and Cayir Ervurald et al.[15] proposed an integrated genetic algorithm(**GA**) and autoregressive moving average(**ARMA**) method for forecasting, obtaining lower error percentages than ARMA. Sharma et al.[16] used a blind Kalman filter algorithm and an autoregressive integrated moving average model to solve the problem of short-term load forecasting. However, because machine learning time series models have fewer parameters and better computational efficiency, artificial intelligence models have better forecasting accuracy than conventional statistical models in most cases.

Incidentally, with the rapid development and widespread use of artificial intelligence algorithms, many researchers have effectively used artificial intelligence methods to predict electric loads. These methods include support vector machines

(**SVM**)[17], artificial neural networks (**ANN**)[18], fuzzy logic models[19], and deep learning models[20,21,22]. Barman et al.[23] proposed the **GWO-SVM** model based on support vector machines (SVM) with gray wolf optimizer (**GWO**) to predict load data, which eventually achieved higher accuracy. Liang et al.[24] proposed general regression neural network (**GRNN**) combined with fruit fly optimization algorithm (**FOA**) for short-term load forecasting. Chen et al.[25] propose a kind of fresh short-term electric load forecasting method **EMD-Mixed-ELM** based on empirical mode decomposition (**EMD**) and extreme learning machine (**ELM**), which obtained higher forecasting accuracy. Xie et al.[26] proposed a **PSO-ENN** model combining **ENN** and particle swarm optimization, which improved the load forecasting accuracy of ENN. López et al.[27] proposed a new hybrid method of **STLF** based on symbiotic empirical mode decomposition (**EEMD**), beam neural network (**WNN**) and particle swarm optimization (**PSO**), and their results verified the higher accuracy of the proposed model. Hu et al.[28] proposed a short-term electricity load prediction model based on a hybrid **GA-PSO-BPNN** algorithm, which improved the prediction accuracy of **BPNN**. Memarzadeh et al.[29] proposed Short-term electricity load by a new optimal **LSTM-NN** based prediction algorithm, which improves the prediction accuracy.Li et al.[30] proposed a novel framework to improve the prediction accuracy using bi-directional gated recurrent unit (**Bi-GRU**) and sparrow search algorithm (**SSA**). Zhu et al.[31] used time convolutional neural network (**TCN**) to predict time series data and obtained higher prediction than the existing single predictor accuracy. Mehdi Bendaoud et al.[32] used Generative Adversarial Network (**GAN**) to introduce STLF and proposed a conditional Generative Adversarial Network (**cGAN**) architecture to improve the prediction accuracy. Artificial intelligence algorithms generally outperform time series models because of the strong nonlinear predictive capability of artificial intelligence models.

With further research, it has been found that noise in the electric load data affects the final prediction, which is why data preprocessing techniques such as empirical mode decomposition (**EMD**)[33], ensemble empirical mode decomposition (**EEMD**)[34], complete ensemble empirical mode decomposition (**CEEMDAN**)[35], wavelet threshold denoising[36], singular spectrum analysis (**SSA**)[37], and variational modal decomposition (**VMD**)[38]. In addition to using data denoising techniques, there are other data preprocessing methods, such as Shifei Ding et al.[39]proposed a weighted linear support vector machine (GWLMBSVM) based on information granulation, which uses information granulation to divide the data into several particles and classify the particles for prediction. José LuisVelázquez-Rodríguez et al.[40]propose a parametric granulation of particles in rough set theory that can effectively deal with the study of hybrid information systems. In recent years, it has been noticed that load forecasting should focus not only on accuracy but also on the stability of forecasting, so multi-objective optimization algorithms have been proposed[41,42]: Yang et al.[43] proposed a new STLF combining data denoising and prediction model based on bivariate empirical mode decomposition (**BEMD**), multivariate multiscale reciprocal entropy (**MMPE**) and tree structure parzen estimation (**TPE**) algorithms to optimize **LSTM**. Bo et al.[44] used singular spectrum analysis (**SSA**) for data preprocessing, and then proposed a multi-objective evolutionary algorithm based on genetic algorithm to discuss decomposition in detail (**MOEA/D**). Wang et al.[45] used a data decomposition strategy to process the raw data and then combined the single model by multi-objective locust algorithm (**MOGOA**) to greatly improve the prediction of power load forecasting accuracy.

Evaluation of the previous literature shows that the aforementioned prediction methods have some inherent drawbacks. Table 1 shows the advantages and disadvantages of the above model.

The drawbacks of these methods are summarized as follows:

(1) As simulation instruments are employed, it is often physically difficult to obtain mathematical expressions for the various building energy mechanisms, and they are not effective for short-term predictions.

(2) Conventional statistical models are more suitable for linear data. For electric load data with high noise and non-linear factors, conventional linear regression model processing either becomes inaccurate or too complex to be used in practice.

(3) Although artificial intelligence models are applicable to nonlinear data and reduce prediction accuracy, they are relatively data-dependent, easily fall into local optimum, and have long running time owing to slow convergence speed.

(4) The data denoising technique in the hybrid model ignores the importance of information leakage from the denoising method, leading to the optimization of abnormal prediction accuracy. Meanwhile, the existing multi-objective optimization algorithms are not strong in optimizing the balance between prediction accuracy and prediction stability and have a long running time.

Based on the above literature analysis, in this paper, we first propose to optimize the weights and thresholds of the back propagation neural network (**BPNN**) [46] using an iterative update strategy[47]. Then a new integrated power load point-interval forecasting system that combines multiple artificial intelligence techniques is proposed, aiming to improve the deterministic and volatility analysis performance. The system abstracts the original high-dimensional time series granularly into low-dimensional time series, uses five artificial intelligence algorithms to perform deterministic analysis on the time series after data scale compression, then optimizes the deterministic analysis results from multiple perspectives by the proposed MODOA, and finally analyzes the predicted fluctuations to derive the uncertainty interval estimates.

The main contributions and innovations of this study are as follows:

(1) As a novel integrated power load point-interval forecasting system is proposed. The system can simplify the complexity of calculation while improving the accuracy and stability of forecasting; fluctuation analysis is added to the deterministic analysis, and experiments show that the proposed uncertainty analysis results have better interval scores and interval center deviations.

(2)In the data processing stage, the low-level, fine-grained raw ultra-short-term power load data are granulated and abstracted into high-level, coarse-grained low-dimensional time series, and the constructed information grains can portray and reflect the structural features of the time series data, reducing the total amount of data input to the model and effectively improving the accuracy of short-term forecasting.

(3)A new feedback-based weight-threshold optimization algorithm for BPNNs with neural networks is proposed. An evolutionary update technique and a stochastic strategy are used to intelligently optimize the weights and thresholds of the BPNN containing two hidden layers, which improves the problems of slow convergence and low accuracy of peak traffic prediction BPNN and improves the prediction accuracy of the BPNN.

(4) By developing a Multi-objective Dingo Optimization Algorithm for multi-level optimization of the benchmark model, the prediction stability is improved while pursuing prediction accuracy. In addition, the newly proposed MODOA has better prediction performance and faster running speed compared with other weight optimization algorithms in the market.

This paper is organized as follows. Section II presents the specific methodological theory of the invented model, and Section III describes the main components of the integrated power load point-interval forecasting system. In order to illustrate the capabilities of the developed prediction system, four different experiments are conducted in Section IV. Specifically, Section 4.1 describes the dataset used in this study, Section 4.2 presents the multidimensional evaluation metrics for point-interval forecasting, and Section 4.3 discusses and analyzes the experimental results of the developed FMICM compared to other models. Section V gives a discussion of the proof of the proposed prediction system and empirical analysis of the power load forecasting is also given. Finally, conclusions are presented in Section VI. Additionally, the main structure of this study is shown in Figure 1.

2.Methodology

This chapter introduces the main techniques used in the integrated power load point-interval forecasting system, i.e., signal fuzzy processing technique, multi-objective combined optimization algorithm (MODOA), and volatility analysis technique.

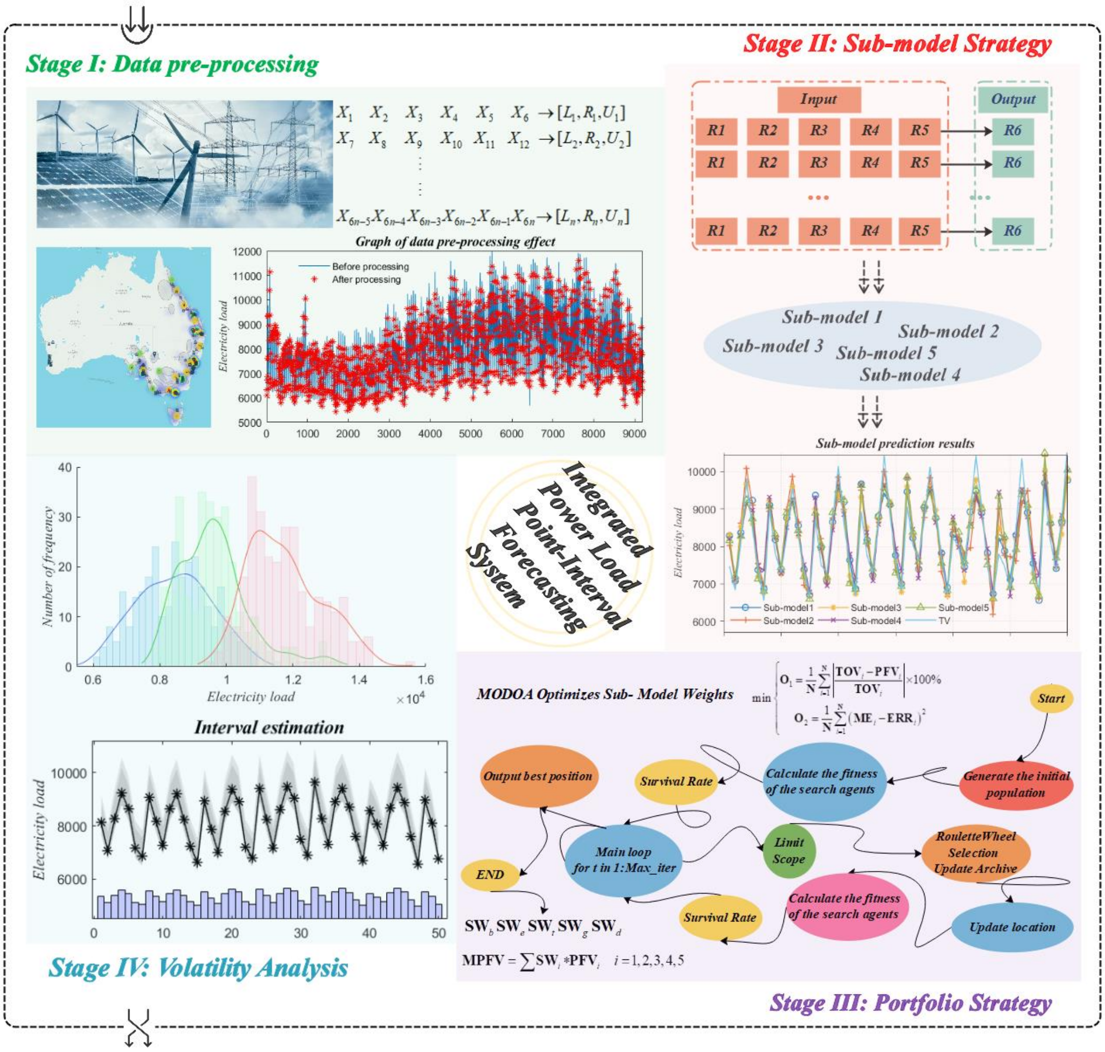

Figure 1 Flow chart of the proposed integrated load forecasting model

Table 1
Evaluation of existing load forecasting models

Models	Refs.	Variables	Results	Advantages	Disadvantages
Physical model					
EnergyPlus	D.B. Crawley et al.(2001)	Power load、Solar thermal、Photovoltaic	It can simulate time series based on DOE-2 and BLAST.	The physical model is easy to operate and does not need a lot of data training	It is difficult to obtain the mathematical expression of energy mechanism, the effect of short-term prediction is poor.
TRANSYS	Yun K et al.(2011)	Power load	The transient response of a building combined with hierarchical CHP optimal control algorithm to forecast data.		
Conventional statistical model					
General regression	VincenzoBianco et al.(2009)	Electricity consumption	Developed regressions are congruent with the official projections.	It has a deep theoretical basis and has higher prediction accuracy than the physical model.	They cannot be well adapted to nonlinear series due to prior linear assumptions.
ARMA	S.Sp.Pappas et al.(2010)	Power load	The reliability, effectiveness and applicability of ARMA in power load data forecasting are proved		
ARIMA	RehanJamil (2020)	Hydropower consumption	The model can predict power load in real time, and has high accuracy and calculation efficiency		
Artificial intelligence algorithm					
SVM	P.Shine et al.(2019)	Electricity consumption	The potential effectiveness of the SVM as a macro-level simulation forecast tool for dairy farm electricity consumption.	Artificial intelligence model has strong nonlinear prediction ability, so the prediction accuracy of artificial intelligence algorithm is generally better than that of time series model.	Relatively dependent on data, it is easy to fall in to local optimization, and the convergence speed is slow, resulting in long running time.
ANN	Si Chen et al.(2021)	Power load	The proposed regression reduces RMSE by 35%, while the ANN with fuzzy hours based reduces RMSE by 42%.		
PSO-ENN	Kun Xie et al.(2020)	Power load	The prediction accuracy of ENN is improved		
LSTM	Gholamreza Memarzadeh et al.(2021)	Power load	Enhance the accuracy and stability of prediction		
Bi-GRU	Xuechen Li et al.(2022)	Oil rate	The observations show that the proposed method performs better than the others in terms of accuracy and robustness.		
TCN	Ruijin Zhu et al.(2020)	Wind power data	TCN shows higher forecasting accuracy than existing predictors such as SVM,MLP, LSTM, and GRU.		
Hybrid model					
BEMD-MMPE-LSTM	Dongchuan Yang et al.(2022)	Power load	A decomposition–ensemble model is proposed for interval-valued load forecasting and it outperforms other model under study.	Integrating the advantages of various models and algorithms can not only improve the prediction accuracy, but also improve the prediction stability.	Because the optimization algorithm used is relatively backward, the prediction accuracy can be further improved
SSA-MOEA	He Bo et al.(2020)	Power load	Effectively improves the efficiency of the power load forecast and adds a new feasible scheme for smart network planning.		
CEEMDAN-MOGOA	Jianzhou Wang et al.(2022)	Power load	Not only has the best performance, but also provides effective technical support for power grid operation scheduling.		

2.1 Signal Fuzzy Processing Technique

Fuzzy information granulation (FIG) is used to construct information grains by creating fuzzy sets on each subsequence formed by the time series after the discretization operation[48]. Fuzzy information granulation mainly includes window division and information fuzzification, the core of which is to complete the fuzzification process after window creation. [49].

The window division is to convert the time series $\bar{\mathbf{T}} = \{\bar{\mathbf{T}}_1, \bar{\mathbf{T}}_2, \cdots, \bar{\mathbf{T}}_\gamma\}$ into the granular time series $\bar{\bar{\mathbf{\Theta}}} = \{\bar{\bar{\mathbf{\Theta}}}_1, \bar{\bar{\mathbf{\Theta}}}_2, \cdots, \bar{\bar{\mathbf{\Theta}}}_\varsigma\}$ after information granulation. By setting the time granularity $\hat{\mathbf{E}}$ to divide $\bar{\mathbf{T}} = \{\bar{\mathbf{T}}_1, \bar{\mathbf{T}}_2, \cdots, \bar{\mathbf{T}}_\gamma\}$ into $\underline{\mathbf{H}}$ subseries $\bar{\bar{\mathbf{\Theta}}} = \{\bar{\bar{\mathbf{\Theta}}}_1, \bar{\bar{\mathbf{\Theta}}}_2, \cdots, \bar{\bar{\mathbf{\Theta}}}_\varsigma\}$, where $\underline{\mathbf{H}} = \gamma / \hat{\mathbf{E}}$ and the $\boldsymbol{\eta}$-th subseries is $\bar{\bar{\mathbf{\Theta}}}_{\boldsymbol{\eta}} = \left[\bar{\bar{\mathbf{T}}}_1^{(\boldsymbol{\eta})}, \bar{\bar{\mathbf{T}}}_2^{(\boldsymbol{\eta})}, \cdots, \bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\boldsymbol{\eta})}\right]$.

$$\{\bar{\mathbf{T}}_1, \bar{\mathbf{T}}_2, \cdots, \bar{\mathbf{T}}_\gamma\} \Rightarrow \left\{\left[\bar{\bar{\mathbf{T}}}_1^{(1)}, \bar{\bar{\mathbf{T}}}_2^{(1)}, \cdots, \bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(1)}\right], \cdots, \left[\bar{\bar{\mathbf{T}}}_1^{(\underline{\mathbf{H}})}, \bar{\bar{\mathbf{T}}}_2^{(\underline{\mathbf{H}})}, \cdots, \bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\underline{\mathbf{H}})}\right]\right\} \tag{1}$$

The information granulation of the time series $\bar{\mathbf{T}} = \{\bar{\mathbf{T}}_1, \bar{\mathbf{T}}_2, \cdots, \bar{\mathbf{T}}_\gamma\}$ is to construct the information particles $\tilde{\mathbf{\Gamma}} = \{\tilde{\mathbf{\Gamma}}'_1, \tilde{\mathbf{\Gamma}}'_2, \cdots, \tilde{\mathbf{\Gamma}}'_\varsigma\}$ using the fuzzy method for each of the $\underline{\mathbf{H}}$ subsequences $\bar{\bar{\mathbf{\Theta}}} = \{\bar{\bar{\mathbf{\Theta}}}_1, \bar{\bar{\mathbf{\Theta}}}_2, \cdots, \bar{\bar{\mathbf{\Theta}}}_\varsigma\}$ formed by the discretization operation.

Definition 1: Suppose $\mathbf{Z}$ is a given theoretical domain, then a fuzzy subset $\mathbf{\Lambda} = \{\chi, \mathbf{\Omega}(\chi) \mid \chi \in \mathbf{Z}\}$ on $\mathbf{Z}$. Where $\mathbf{\Omega}(\chi): \chi \to [0,1]$ represents the affiliation function of $\mathbf{\Lambda}$. If two fuzzy subsets $\mathbf{\Phi}$ and $\mathbf{\Xi}$ are equal, denoted $\mathbf{\Phi} = \mathbf{\Xi}$, when and only when they have the same affiliation function, i.e., $\hat{\mathbf{\Omega}}'_{\mathbf{\Phi}}(\chi) = \hat{\mathbf{\Omega}}''_{\mathbf{\Xi}}(\chi)$.

In this paper, the triangular fuzzy particles are chosen to construct the information grain and its affiliation function is as follows[50]:

$$\mathbf{A}_{Tf}(\mathbf{x}) = \begin{cases} \dfrac{\mathbf{x} - \mathbf{I}_{Tf}}{\mathbf{K}_{Tf} - \mathbf{I}_{Tf}}, \mathbf{I}_{Tf} \le \mathbf{x} \le \mathbf{K}_{Tf} \\ 0, \mathbf{x} < \mathbf{I}_{Tf} \cup \mathbf{x} > \mathbf{N}_{Tf} \\ \dfrac{\mathbf{N}_{Tf} - \mathbf{x}}{\mathbf{N}_{Tf} - \mathbf{K}_{Tf}}, \mathbf{K}_{Tf} < \mathbf{x} \le \mathbf{N}_{Tf} \end{cases} \tag{2}$$

Where $\mathbf{x}$ is the variable in the theoretical domain, $\mathbf{I}_{Tf}$, $\mathbf{K}_{Tf}$, $\mathbf{N}_{Tf}$ are the three parameters of the triangular type fuzzy example affiliation function, which correspond to the lower boundary, average level and upper boundary of the window after fuzzy particleization, respectively[51].

Fuzzy sets get rid of the either-or duality in classical set theory, and extend the value domain of the affiliation function from the binary $\{0,1\}$ to the multi-valued interval $[0,1]$, which is a kind of extension of set theory. Information fuzzification is the fuzzification of each information grain, and the fuzzification of a single sub-window $\bar{\bar{\mathbf{\Theta}}}_{\boldsymbol{\mu}}$ generates multiple fuzzy sets $\tilde{\mathbf{\Gamma}}'_{\boldsymbol{\mu}} = \left[\tilde{\mathbf{\Gamma}}''_{\boldsymbol{\mu};1}, \tilde{\mathbf{\Gamma}}''_{\boldsymbol{\mu};2}, \tilde{\mathbf{\Gamma}}''_{\boldsymbol{\mu};3}\right]$.

Considering the single-window problem, $\bar{\bar{\mathbf{\Theta}}}_{\boldsymbol{\mu}} = \left[\bar{\bar{\mathbf{T}}}_1^{(\boldsymbol{\mu})}, \bar{\bar{\mathbf{T}}}_2^{(\boldsymbol{\mu})}, \cdots, \bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\boldsymbol{\mu})}\right]$ should first be viewed as a window for fuzzification. The task of fuzzification is to build a triangular fuzzy particle TFP on $\bar{\bar{\mathbf{\Theta}}}_{\mu} = \left[\bar{\bar{\mathbf{T}}}_1^{(\mu)}, \bar{\bar{\mathbf{T}}}_2^{(\mu)}, \cdots, \bar{\bar{\mathbf{T}}}_{\hat{\mathrm{E}}}^{(\mu)} \right]$, who can reasonably explain the fuzzy concept $\mathbf{M}$ of $\bar{\bar{\mathbf{\Theta}}}_{\mu}$. The fuzzy particle $\tilde{\mathbf{\Gamma}}'_{\mu} = \left[\tilde{\mathbf{\Gamma}}''_{\mu;1} = \hat{\mathbf{I}}^{\mu}_{Tf}, \tilde{\mathbf{\Gamma}}''_{\mu;2} = \hat{\mathbf{K}}^{\mu}_{Tf}, \tilde{\mathbf{\Gamma}}''_{\mu;3} = \hat{\mathbf{N}}^{\mu}_{Tf} \right]$ can be constructed by the relevant parameters in the determined affiliation function (2) of the triangular fuzzy particle.

2.2 Multi-objective Dingo Optimization Algorithm

MODOA is a location update strategy for multilevel optimization, which finds the individual that makes the multi-objective function optimal by Pareto search. Therefore it mainly consists of two parts: location update and pareto search. The pseudo-code of the developed MODOA is shown in Algorithm 1.

(a) Location Update

Herna´n Peraza-Va´zquez proposed the Dingo Optimization Algorithm based on the predatory behavior of Australian wild dogs, the dingo is Australia's dingo is the most dangerous animal in Australia, the top local carnivore in Australia. Due to its small size, the dingo will select weak or dying objects, and when out hunting the dingo usually attacks in groups, they cooperate with each other, some attacking from behind some flanking, surround the prey in a perimeter and start chasing it until they are exhausted.With this inspiration Dingo Optimization Algorithm divides the considered hunting strategies into Group Attack, Persecution, Scavenger, and Dingoes' Survival Rates. The calculation formula is Equation (3)-(7). The definitions and theories related to the study are given below.

Definition 2: Group Attack. When attacking large animals,the dingo usually attacks in groups, surrounds its prey and starts chasing until it is captured. If the first instantaneous random number $\tilde{\mathbf{I}}'_r$ is smaller than the set random number $\bar{\mathbf{K}}_r$ and the second instantaneous random number $\tilde{\mathbf{I}}''_r$ is smaller than the set random number $\bar{\bar{\mathbf{\Lambda}}}_r$, i.e. $\mathbf{IF}: \tilde{\mathbf{I}}'_r < \bar{\mathbf{K}}_r \cap \tilde{\mathbf{I}}''_r < \bar{\bar{\mathbf{\Lambda}}}_r$,then the group attack strategy is applied.

To begin with, calculate the search agent subset $\overrightarrow{\overline{\overline{\mathbf{N}}}}{}^{\psi;\mathrm{v}}_{\zeta}$. If $\hat{\hat{\mathbf{\Delta}}}{}^{(\psi)}_{\zeta} \notin \mathbf{Q}$ is satisfied, where $\hat{\hat{\mathbf{\Delta}}}{}^{(\psi)}_{\zeta}$ is a random number, then $\hat{\hat{\mathbf{\Delta}}}{}^{(\psi)}_{\zeta}$ is stored to $\mathbf{Q}$, i.e. $\mathbf{Q}^{\Delta}(\lambda) = \hat{\hat{\mathbf{\Delta}}}{}^{(\psi)}_{\zeta}$.Cycle $\mathbf{V}$ times after $\mathbf{Q} = \left[\tilde{\mathbf{Q}}^{\langle 1 \rangle}_{\hat{\hat{\Delta}}}, \tilde{\mathbf{Q}}^{\langle 2 \rangle}_{\hat{\hat{\Delta}}}, \cdots, \tilde{\mathbf{Q}}^{\langle \mathrm{v} \rangle}_{\hat{\hat{\Delta}}} \right]$ contains $\mathbf{V}$ different numbers, the search agent subset $\overrightarrow{\overline{\overline{\mathbf{N}}}}{}^{\psi;\mathrm{v}}_{\zeta} = \left[\overrightarrow{\mathbf{N}}{}^{\psi;1}_{\zeta}, \overrightarrow{\mathbf{N}}{}^{\psi;2}_{\zeta}, \cdots, \overrightarrow{\mathbf{N}}{}^{\psi;\mathrm{v}}_{\zeta} \right]$ that is the location of the set $\mathbf{Q} = \left[\tilde{\mathbf{Q}}^{\langle 1 \rangle}_{\hat{\hat{\Delta}}}, \tilde{\mathbf{Q}}^{\langle 2 \rangle}_{\hat{\hat{\Delta}}}, \cdots, \tilde{\mathbf{Q}}^{\langle \mathrm{v} \rangle}_{\hat{\hat{\Delta}}} \right]$, that is, $\overrightarrow{\mathbf{N}}{}^{\psi;\theta}_{\zeta} = \overrightarrow{\mathbf{P}'}{}^{(\psi;\theta)}_{\zeta}$.The location update formula of group attack policy is:

$$\overrightarrow{\mathbf{P}''}{}^{(\psi)}_{\zeta+1} = \tilde{\mathbf{M}}' \times \sum_{\mathrm{v}=1}^{\kappa} \left[\overrightarrow{\overline{\overline{\mathbf{N}}}}{}^{\psi;\mathrm{v}}_{\zeta} - \overrightarrow{\mathbf{P}'}{}^{(\psi)}_{\zeta} \right] \Big/ \boldsymbol{\kappa} - \overrightarrow{\mathbf{P}'}{}^{(*)}_{\zeta} \tag{3}$$

Among them, $\overrightarrow{\mathbf{P}''}{}^{(\psi)}_{\zeta+1}$ is the new position of a search agent (indicates dingoes' movement). $\boldsymbol{\kappa}$ is a random integer between $[2, Sizepop/2]$,where sizepop is the total size of the population of dingoes. $\overrightarrow{\mathbf{P}'}{}^{(\psi)}_{\zeta}$ is the current search agent. $\overrightarrow{\mathbf{P}'}{}^{(*)}_{\zeta}$ is the best search agent found from the previous iteration, and $\tilde{\mathbf{M}}'$ is a random number uniformly generated in the interval of [− 2, 2].

Definition 3: Persecution. When attacking small animals, wild dogs usually attack individually and chase until they are caught. If the first instantaneous random number $\tilde{\mathbf{I}}'_r$ is smaller than the set random number $\bar{\mathbf{K}}_r$ and the second instantaneous random number $\tilde{\mathbf{I}}''_r$ is greater than the set random number $\bar{\bar{\mathbf{\Lambda}}}_r$, i.e. $\mathbf{IF}: \tilde{\mathbf{I}}'_r < \bar{\mathbf{K}}_r \bigcap \tilde{\mathbf{I}}''_r > \bar{\bar{\mathbf{\Lambda}}}_r$,then the persecution strategy is applied.

Here, we use the random number $\breve{\mathbf{\Delta}}^{(\chi)}_{\zeta} \neq \boldsymbol{\xi}$ in the group attack strategy to determine the location $\overrightarrow{\vec{\mathbf{P}}'_{\psi}\left(\breve{\mathbf{\Delta}}^{(\chi)}\right)}$, with the random number $\breve{\mathbf{E}}''$ and $\tilde{\mathbf{M}}'$ in the group attack strategy.The location update formula of persecution is:

$$\overrightarrow{\vec{\mathbf{P}}''^{(\psi)}_{\zeta+1}} = \overrightarrow{\vec{\mathbf{P}}'^{(*)}_{\zeta}} + \tilde{\mathbf{M}}' \times e^{\breve{\mathbf{E}}''} \times \left(\overrightarrow{\vec{\mathbf{P}}'_{\psi}\left(\breve{\mathbf{\Delta}}^{(\chi)}\right)} - \overrightarrow{\vec{\mathbf{P}}'^{(\psi)}_{\zeta}} \right) \tag{4}$$

In which, $\overrightarrow{\vec{\mathbf{P}}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent. $\overrightarrow{\vec{\mathbf{P}}'^{(\psi)}_{\zeta}}$ is the current search agent. $\overrightarrow{\vec{\mathbf{P}}'^{(*)}_{\zeta}}$ is the best search agent found from the previous iteration, and $\tilde{\mathbf{M}}' \in [-2,2]$ $\breve{\mathbf{E}}'' \in [-1,1]$.

Definition 4: Scavenger. When dingo smells a dead small animal on the ground nearby during his daily walk, this behavior is called scavenger in this section. If the first instantaneous random number $\tilde{\mathbf{I}}'_r$ is greater than the set random number $\bar{\mathbf{K}}_r$, i.e. $\mathbf{IF}: \tilde{\mathbf{I}}'_r > \bar{\mathbf{K}}_r$,then the scavenger strategy is applied.

We also use a random number strategy to determine the location $\overrightarrow{\bar{\bar{\mathbf{P}}}'_{\psi}\left(\hat{\mathbf{\Delta}}^{(\chi)}\right)}$,The location update formula of scavenger is:

$$\overrightarrow{\bar{\bar{\mathbf{P}}}''^{(\psi)}_{\zeta+1}} = \frac{1}{2}\left[e^{\breve{\mathbf{E}}''} * \overrightarrow{\bar{\bar{\mathbf{P}}}'_{\psi}\left(\hat{\mathbf{\Delta}}^{(\chi)}\right)} - (-1)^{\tilde{\mathbf{H}}'''} \times \overrightarrow{\bar{\bar{\mathbf{P}}}'^{(\psi)}_{\zeta}} \right] \tag{5}$$

In which, $\overrightarrow{\bar{\bar{\mathbf{P}}}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent. $\overrightarrow{\bar{\bar{\mathbf{P}}}'^{(\psi)}_{\zeta}}$ is the current search agent, and $\breve{\mathbf{E}}'' \in [-1,1]$ $\tilde{\mathbf{H}}''' \in \{0,1\}$.

Definition 5: Dingoes' Survival Rates. In addition to the above three location update strategies, DOA also considers the survival rate of dingo. The location update formula of dingoes' Survival Rates is:

$$\mathbf{S\hat{r}}^{\psi}_{\xi}(\delta) = \frac{\overline{\mathbf{XF}}^{\psi}_{\xi} - \mathbf{IF}^{\psi}_{\xi}(\delta)}{\overline{\mathbf{XF}}^{\psi}_{\xi} - \overline{\mathbf{NF}}^{\psi}_{\xi}} \tag{6}$$

Among them, $\overline{\mathbf{XF}}^{\psi}_{\xi}$ and $\overline{\mathbf{NF}}^{\psi}_{\xi}$ are the worst and the best fitness value in the current generation, respectively, whereas $\mathbf{IF}^{\psi}_{\xi}(\delta)$ is the current fitness value of the $\delta-$th search agent. When the survival rate of dingo is lower than 0.3,i.e. $\mathbf{S\hat{r}}^{\psi}_{\xi}(\delta) < 0.3$,the location update formula becomes:

$$\overrightarrow{\tilde{\mathbf{P}}''^{(\psi)}_{\zeta+1}} = \overrightarrow{\tilde{\mathbf{P}}'^{(*)}_{\zeta}} + \frac{1}{2} \times \left[\overrightarrow{\tilde{\mathbf{P}}'_{\psi}\left(\vec{\mathbf{\Delta}}^{(\chi)}_{1}\right)} - (-1)^{\bar{\bar{\mathbf{H}}}'''} * \overrightarrow{\tilde{\mathbf{P}}'_{\psi}\left(\vec{\mathbf{\Delta}}^{(\chi)}_{2}\right)} \right] \tag{7}$$

In which, $\overrightarrow{\tilde{\mathbf{P}}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent, $\overrightarrow{\tilde{\mathbf{P}}'^{(*)}_{\zeta}}$ is the best search agent found from the previous iteration, and $\bar{\bar{\mathbf{H}}}''' \in \{-1,1\}$. Since the survival rate is not passed, this formula uses two random number locations $\vec{\mathbf{\Delta}}_1^{(\chi)}$ and $\vec{\mathbf{\Delta}}_2^{(\chi)}$, which means that the two random numbers position $\overrightarrow{\tilde{\mathbf{P}}'_\psi\left(\vec{\mathbf{\Delta}}_1^{(\chi)}\right)}$ and $\overrightarrow{\tilde{\mathbf{P}}'_\psi\left(\vec{\mathbf{\Delta}}_2^{(\chi)}\right)}$ are used to generate new locations according to the generated.

(b) Pareto search

Definition 6: When multiple objectives $\tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)=\left[\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\cdots,\mathbf{O}_k^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right]$ in the objective function that need to be optimized, and these objectives are usually conflicting, the problem of finding a set of vectors $\vec{\mathbf{V}}=\left[\vec{\varpi}_1,\vec{\varpi}_2,\cdots,\vec{\varpi}_\rho\right]$ such that $\tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)=\left[\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\cdots,\mathbf{O}_k^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right]$ is maximized or minimized is called a multi-objective optimization problem.In mathematical terms, a multi-objective optimization problem can be written as:

$$\min\left(\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\cdots,\mathbf{O}_k^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right)$$
$$s.t.\quad \begin{cases}\Theta'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\le 0\\ \Theta''_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)=0\end{cases}\tag{8}$$

Where the integer k is the target number and $\left\{\Theta'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\Theta''_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right\}$ is the constraint function.

The purpose of constructing a multi-objective optimization algorithm is to compensate for the shortage of pursuing only accuracy due to the single optimization algorithm, so the multi-objective function constructed in this paper includes the mean absolute percentage error (MAPE), which pursues accuracy, on the one hand, and the residual variance (RV), which pursues prediction stability, on the other hand.

$$\min\begin{cases}\mathbf{O}_1=\dfrac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left|\dfrac{\mathbf{TOV}_i-\mathbf{PFV}_i}{\mathbf{TOV}_i}\right|\times 100\%\\ \mathbf{O}_2=\dfrac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left(\mathbf{ME}_i-\mathbf{ERR}_i\right)^2\end{cases}\tag{9}$$

Where, $\mathbf{TOV}_i$ denotes the i-th actual observation value, $\mathbf{PFV}_i$ denotes the i-th PF forecast value, $\mathbf{ME}_i$ is the average of the error $\mathbf{ERR}_i=\mathbf{TOV}_i-\mathbf{PFV}_i$ of the i-th true value $\mathbf{TOV}_i$ and the i-th predicted value $\mathbf{PFV}_i$.

The single-objective optimization algorithm does not apply to multi-objective optimization problems.

Proof: Suppose $\vec{\mathbf{\Gamma}}=\left[\vec{\gamma}_1,\vec{\gamma}_2,\cdots,\vec{\gamma}_\rho\right]$ and $\vec{\mathbf{\Pi}}=\left[\vec{\lambda}_1,\vec{\lambda}_2,\cdots,\vec{\lambda}_\rho\right]$ are two sets of solutions. $\mathbf{If:}\ \exists\vec{\mathbf{\Gamma}},\vec{\mathbf{\Pi}}\ s.t.\ \mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{\Gamma}}\right)<\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{\Pi}}\right)\cap\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{\Gamma}}\right)>\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{\Pi}}\right)$,according to the single-objective optimization problem solution, only $\mathbf{O}_1^{\mathbf{M}}$ will be sorted and the optimal solution will be $\vec{\mathbf{\Gamma}}=\left[\vec{\gamma}_1,\vec{\gamma}_2,\cdots,\vec{\gamma}_\rho\right]$, which is not in line with the principle of multi-objective optimization.

For single-objective optimization problems, the maximum value of the derived objective function can be directly selected as the optimal solution at this stage. However, for multi-objective optimization problems, there is usually a tendency of mutual constraints between different objective functions, which may improve the performance of one objective often at the expense of the performance of other objectives, so for multi-objective optimization problems, the solution is usually a set of non-inferior solutions-Pareto solution set.

Definition 7: Given a multi-objective optimization problem $\min \tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, let $\vec{\mathbf{V}}^{*}=\left[\vec{\varpi}_{1}^{*},\vec{\varpi}_{2}^{*},\cdots,\vec{\varpi}_{\rho}^{*}\right]\in\mathbf{\Omega}$,if $\exists\vec{\mathbf{V}}=\left[\vec{\varpi}_{1},\vec{\varpi}_{2},\cdots,\vec{\varpi}_{\rho}\right]\in\mathbf{\Omega}$ such that the following conditions are satisfied:

For any subgoal function $\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ of $\tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ there exists $\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\vec{\mathbf{V}}^{*}\right)\le\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, while there exists at least one subgoal function $\tilde{\mathbf{O}}_{\boldsymbol{\varphi}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ such that $\tilde{\mathbf{O}}_{\boldsymbol{\varphi}}^{\mathbf{M}}\left(\vec{\mathbf{V}}^{*}\right)<\tilde{\mathbf{O}}_{\boldsymbol{\varphi}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, then we say that $\vec{\mathbf{V}}^{*}=\left[\vec{\varpi}_{1}^{*},\vec{\varpi}_{2}^{*},\cdots,\vec{\varpi}_{\rho}^{*}\right]$ is a strong pareto optimal solution.

Definition 8: Given a multi-objective optimization problem $\min \breve{\mathbf{O}}'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, let $\vec{\mathbf{V}}'_{*}=\left[\vec{\varpi}'_{*;1},\vec{\varpi}'_{*;2},\cdots,\vec{\varpi}'_{*;\rho}\right]\in\mathbf{\Omega}$,if $\exists\vec{\mathbf{V}}'=\left[\vec{\varpi}'_{1},\vec{\varpi}'_{2},\cdots,\vec{\varpi}'_{\rho}\right]\in\mathbf{\Omega}$ such that the following conditions are satisfied:

For any subgoal function $\breve{\mathbf{O}}'^{\tau}_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ of $\breve{\mathbf{O}}'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ there exists $\breve{\mathbf{O}}'^{\tau}_{\mathbf{M}}\left(\vec{\mathbf{V}}'_{*}\right)\le\breve{\mathbf{O}}'^{\tau}_{\mathbf{M}}\left(\vec{\mathbf{V}}'\right)$, then we say that $\vec{\mathbf{V}}'_{*}=\left[\vec{\varpi}'_{*;1},\vec{\varpi}'_{*;2},\cdots,\vec{\varpi}'_{*;\rho}\right]$ is a weak pareto optimal solution.

Definition 9: Suppose there are **N** sets of position vectors $\overrightarrow{\mathbf{M\tilde{s}}}=\left[\vec{\mathbf{M}}_{1},\vec{\mathbf{M}}_{2},\cdots,\vec{\mathbf{M}}_{\mathbf{N}}\right]$ in the archive,where $\vec{\mathbf{M}}_{\boldsymbol{\sigma}}=\left[\vec{\varpi}_{\boldsymbol{\sigma}}^{(1)},\vec{\varpi}_{\boldsymbol{\sigma}}^{(2)},\cdots,\vec{\varpi}_{\boldsymbol{\sigma}}^{(\rho)}\right]$, and each set of position vectors corresponds to an adaptation function $\overrightarrow{\mathbf{H\tilde{s}}}=\left[\vec{\mathbf{H}}_{1},\vec{\mathbf{H}}_{2},\cdots,\vec{\mathbf{H}}_{\mathbf{N}}\right]$,where $\vec{\mathbf{H}}_{\boldsymbol{\sigma}}=\left[\vec{\mathbf{H}}_{\boldsymbol{\sigma}}^{(1)},\vec{\mathbf{H}}_{\boldsymbol{\sigma}}^{(2)}\right]$.

We obtain $\overrightarrow{\mathbf{P\tilde{r}}}=\left[\tilde{\mathbf{R}}_{1},\tilde{\mathbf{R}}_{2},\cdots,\tilde{\mathbf{R}}_{\mathbf{N}}\right]$ by pareto ranking $\overrightarrow{\mathbf{H\tilde{s}}}=\left[\vec{\mathbf{H}}_{1},\vec{\mathbf{H}}_{2},\cdots,\vec{\mathbf{H}}_{\mathbf{N}}\right]$ from best to worst, then $\tilde{\mathbf{E}}_{\zeta}=\tilde{\mathbf{R}}_{\zeta}\Big/\sum_{\lambda=1}^{\mathbf{N}}\tilde{\mathbf{R}}_{\lambda}$, $\boldsymbol{\zeta}=1,2,\ldots,\mathbf{N}$ is the probability of being eliminated.This method is known as roulette selection method, also known as proportional selection method.

In the iterative loop, by finding out the group strong pareto solution $\bar{\mathbf{S}}^{*}$,it needs to be filed into $\overrightarrow{\mathbf{A\tilde{r}}}=\left[\vec{\mathbf{\Lambda}}_{1},\vec{\mathbf{\Lambda}}_{2},\cdots,\vec{\mathbf{\Lambda}}_{\delta}\right]$, if the following occurs:

$$\mathbf{If:}\exists\delta:\left\{\forall\eta:\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\bar{\mathbf{S}}^{*}\right)\le\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\vec{\mathbf{\Lambda}}_{\delta}\right)and\exists\eta:\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\bar{\mathbf{S}}^{*}\right)<\tilde{\mathbf{O}}_{\eta}^{\mathbf{M}}\left(\vec{\mathbf{\Lambda}}_{\delta}\right)\right\} \tag{10}$$

Then file $\bar{\mathbf{S}}^{*}$ to $\overrightarrow{\mathbf{A\tilde{r}}}=\left[\vec{\mathbf{\Lambda}}_{1},\vec{\mathbf{\Lambda}}_{2},\cdots,\vec{\mathbf{\Lambda}}_{\delta}\right]$,i.e. $\overrightarrow{\mathbf{A\tilde{r}}}\left(\vec{\mathbf{\Lambda}}_{\delta+1}\right)=\bar{\mathbf{S}}^{*}$.If the $\overrightarrow{\mathbf{A\tilde{r}}}$ storage reaches its limit, $\bar{\mathbf{S}}^{*}$ is substituted for $\overrightarrow{\mathbf{A\tilde{r}}}\left(\vec{\mathbf{\Lambda}}_{\boldsymbol{\mu}}\right)$ using the roulette selection method.

2.3 Volatility Analysis Technique

Nonparametric kernel density estimation simulates the true probability distribution curve without using a priori knowledge of the data distribution. Therefore, it is a non-parametric method suitable for power load interval forecasting studies.We propose the improved kernel density estimation method (IKDE) in this paper, based on the point prediction results obtained from FMICM.

Definition 10: Suming that $\mathbf{\Omega}(\boldsymbol{\psi})$ is the probability density function,

$\boldsymbol{\Xi}(\boldsymbol{\psi})=\int_{-\infty}^{\boldsymbol{\psi}}\boldsymbol{\Omega}(\zeta)d\zeta$ is the cumulative distribution function.As $\boldsymbol{\Xi}_n(\boldsymbol{\chi})=\frac{1}{\mathbf{T}}\sum_{\iota=1}^{\mathbf{T}} l_{\boldsymbol{\psi}_\iota}\le\boldsymbol{\chi}$:

$$\boldsymbol{\Omega}\left(\tilde{\boldsymbol{\psi}}_\phi\right)=\lim_{\delta\to0}\frac{\boldsymbol{\Xi}\left(\tilde{\boldsymbol{\psi}}_\phi+\delta\right)-\boldsymbol{\Xi}\left(\tilde{\boldsymbol{\psi}}_\phi-\delta\right)}{2\delta}=\frac{1}{2\mathbf{HT}}\sum_{\iota=1}^{\mathbf{T}} l_{\tilde{\boldsymbol{\psi}}_\phi-\delta\le\boldsymbol{\psi}\le\tilde{\boldsymbol{\psi}}_\phi+\delta} \tag{11}$$

Rewrite Equation (11) as $\boldsymbol{\Omega}\left(\tilde{\boldsymbol{\psi}}_\phi\right)=\left[\sum_{\iota=1}^{\mathbf{T}}\mathbf{K}\left(\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi/\mathbf{H}\right)\right]\Big/2\mathbf{HT}$, Call it kernel density estimation.

To avoid information leakage, this paper uses the error percentage of the optimization set to fit the kernel probability density function. The error percentage $\overrightarrow{\mathbf{Er}}$ is calculated from the true value $\overrightarrow{\mathbf{To}}=\left[\mathbf{TOV}_1^{\mathbf{M}},\mathbf{TOV}_2^{\mathbf{M}},\cdots,\mathbf{TOV}_{\boldsymbol{\varphi}}^{\mathbf{M}}\right]$ and the predicted value $\overrightarrow{\mathbf{P\tilde{f}}}=\left[\mathbf{PFV}_1^{\mathbf{M}},\mathbf{PFV}_2^{\mathbf{M}},\cdots,\mathbf{PFV}_{\boldsymbol{\varphi}}^{\mathbf{M}}\right]$, which is $\overrightarrow{\mathbf{Er}}=\left[\mathbf{ERR}_1^{\mathbf{M}},\mathbf{ERR}_2^{\mathbf{M}},\cdots,\mathbf{ERR}_{\boldsymbol{\varphi}}^{\mathbf{M}}\right]$, $\mathbf{ERR}_{\boldsymbol{\eta}}^{\mathbf{M}}=\left(\mathbf{TOV}_{\boldsymbol{\eta}}^{\mathbf{M}}-\mathbf{PFV}_{\boldsymbol{\eta}}^{\mathbf{M}}\right)\Big/\mathbf{PFV}_{\boldsymbol{\eta}}^{\mathbf{M}}$.

The setting of bandwidth and the selection of kernel functions directly affect the smoothness and fit of the density curve in the NKDE algorithm, which in turn affects the accuracy of the calculation. Since the kernel function has little effect on the final impact, the Gaussion kernel is chosen here and its kernel probability density function is:

$$\boldsymbol{\Omega}(\boldsymbol{\psi})=\frac{1}{\sqrt{2\pi}\mathbf{TH}}\sum_{\iota=1}^{\mathbf{T}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi}{\mathbf{H}}\right)^2} \tag{12}$$

Where the bandwidth H is optimized based on the error term of the optimized data set using the location update strategy, which is superior to the mean-squared error algorithm.After the optimal probability density curve of the error is known, the confidence interval $\left[\bar{\mathbf{G}}'_{\alpha/2},\bar{\mathbf{G}}'_{1-\alpha/2}\right]$ of the error at $1-\alpha$ confidence level is calculated using the integral equation (13), which satisfies both $\mathbf{P}(\bar{\mathbf{G}}'_{\alpha/2}<\bar{\boldsymbol{\psi}}'<\bar{\mathbf{G}}'_{1-\alpha/2})=1-\alpha$.

$$\begin{aligned}\int_{-\infty}^{\bar{\mathbf{G}}'_{\alpha/2}}\frac{1}{\sqrt{2\pi}\mathbf{TH}}\sum_{\iota=1}^{\mathbf{T}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi}{\mathbf{H}}\right)^2}d\boldsymbol{\psi}&=\alpha/2\\ \int_{-\infty}^{\bar{\mathbf{G}}'_{1-\alpha/2}}\frac{1}{\sqrt{2\pi}\mathbf{TH}}\sum_{\iota=1}^{\mathbf{T}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi}{\mathbf{H}}\right)^2}d\boldsymbol{\psi}&=1-\alpha/2\end{aligned} \tag{13}$$

For a given confidence level, the final interval prediction formula can be derived from the obtained error confidence intervals as $[\mathbf{PFV}_i\times\left(1+\bar{\mathbf{G}}'_{\alpha/2}\right),\mathbf{PFV}_i\times\left(1+\bar{\mathbf{G}}'_{1-\alpha/2}\right)]$.

3. Main structure of integrated electric load point-interval forecasting system

The integrated electric load point-interval forecasting system proposed in this thesis is an electric load point-interval forecasting system integrating data pre-processing module, model forecasting module, combined optimization module and uncertainty analysis module, which improves forecasting accuracy and forecasting stability. This system first decomposes the original ultra-short-term power load data into a series of information grains to reduce the total amount of data input to the model and improve the forecasting accuracy. Secondly, the prediction accuracy and stability of different models for different data are different, from which five AI benchmark models are selected in this module: DOA-BPNN, Extreme Learning Machine (ELM), Time Convolutional Neural Network (TCN), Gated Recurrent Unit (GRU), and Deep Belief Network (DBN), according to which the electric load data are trained to derive

***Algorithm 1* : MODOA**

Input: Predicted values of five models $\left\{\mathbf{PFV}_b, \mathbf{PFV}_e, \mathbf{PFV}_t, \mathbf{PFV}_g, \mathbf{PFV}_d\right\}$

Output: Optimal weight $\left\{\mathbf{SW}_1, \mathbf{SW}_2, \mathbf{SW}_3, \mathbf{SW}_4, \mathbf{SW}_5\right\}$

Initialization of parameters and Archive
Generate the initial population
for *i: 1 ≤ i ≤ SA(SearchAgents)*
    Evaluate the corresponding fitness function ***GHF****ᵢ*
    if *mindominates maxdominates*
    /*calculation **minfitness min_position maxfitness max_position */**
    end if
end for
$\widehat{\mathbf{Sr}}_{\xi}^{\Psi}(\delta) = \left(\overline{\mathbf{XF}}_{\xi}^{\Psi} - \mathbf{IF}_{\xi}^{\Psi}(\delta)\right) \Big/ \left(\overline{\mathbf{XF}}_{\xi}^{\Psi} - \overline{\mathbf{NF}}_{\xi}^{\Psi}\right)$
while *iteration < Max Number of Iterations* **do**
    for *i: 1 ≤ i ≤ SA(SearchAgents)*
        Evaluate the corresponding fitness function ***GHFᵢ***
        if *dominates* **/***Select the minimum fitness value and position***/**
        end if
    end for
    Archive = UpdateArchive(*Archive*,*GHF*,*GHP*)
    if *Archive_member_no>ArchiveMaxSize*
        Ranking = RankingProcess(*Archive*)
    end if
    for *r: 1 ≤ i ≤ SA(SearchAgents)*
        if *random < P* **then**
            if *random < Q* **then**
                $\overrightarrow{\mathbf{P}''^{(\Psi)}_{\zeta+1}} = \tilde{\mathbf{M}}' \times \sum_{\mathbf{v}=1}^{\boldsymbol{\kappa}}\left[\overrightarrow{\overline{\mathbf{N}}^{\Psi;\mathbf{v}}_{\zeta}} - \overrightarrow{\mathbf{P}'^{(\Psi)}_{\zeta}}\right] \Big/ \boldsymbol{\kappa} - \overrightarrow{\mathbf{P}'^{(*)}_{\zeta}}$ **/***Group Attack***/**
            else
                $\overrightarrow{\mathbf{P}''^{(\Psi)}_{\zeta+1}} = \overrightarrow{\mathbf{P}'^{(*)}_{\zeta}} + \tilde{\mathbf{M}}' \times e^{\breve{\mathbf{E}}''} \times \left(\overrightarrow{\mathbf{P}'_{\Psi}\left(\breve{\mathbf{\Delta}}^{(\chi)}\right)} - \overrightarrow{\mathbf{P}'^{(\Psi)}_{\zeta}}\right)$ **/***Persecution***/**
            end if
        else
            $\overrightarrow{\overline{\overline{\mathbf{P}}}''^{(\Psi)}_{\zeta+1}} = \left[e^{\breve{\mathbf{E}}''} * \overline{\overline{\mathbf{P}}}'_{\Psi}\left(\widehat{\mathbf{\Delta}}^{(\chi)}\right) - (-1)^{\tilde{\mathbf{H}}''} \times \overrightarrow{\overline{\overline{\mathbf{P}}}'^{(\Psi)}_{\zeta}}\right] \Big/ 2$ **/***Scavenger***/**
        end if
        if *survival(r) <= 0.3*
            reture $\overrightarrow{\mathbf{SW}_*}$
        end if
        Evaluate the corresponding fitness function ***GHFᵢ***
        if *dominates* /*Select the minimum fitness value and position*/
        end if
    end for
end while
reture $\overrightarrow{\mathbf{SW}_*}$

the base prediction value of each model.Then,based on the evolutionary computation technique of population intelligence and omission strategy, a new multi-level optimization algorithm is proposed to integrate each model and finally obtain the point prediction values of the system. At last, the residual distribution is fitted by fluctuation

Table 2

FMICM uses the model's parameters to set values

Required Model	Parameters		Value
FIG	MF	Types of affiliation functions	triangle
	w	Number of windows for granulation	6
DOA-BPNN	S_A	Number of individuals to be optimized	30
	M_{iter}	Maximum number of iterations	100
	l_r	BPNN's Learning rate	0.1
	E_p	BPNN's Training times	100
	G	BPNN's Error accuracy	0.00004
TCN	Embedding size of the convolutional layers in the residual block		[128,64,32,16]
	Kernel size		[3,3,3,3]
	Dilation rate		[1,2,4,8]
	Batch size		20
	Epochs finetune		500
GRU	Spatial Dimension in GRU		[64,32,16,1]
	Batch size		1
	Epochs finetune		200
DBN	Batch size		128
	Epochs finetune		2000
MODOA	S_A	Search Number of Individuals	100
	M_{iter}	Maximum iterations Number	200
	A_m	ArchiveMaxSize	500
	P	Random numbers in algorithms	0.5
	Q	Random numbers in algorithms	0.7
IKDE	L_b	Lower limit of bandwidth	0.01
	U_b	Upper limit of bandwidth	0.1
	S_A	Search Number of Individuals	6
	M_{iter}	Maximum iterations Number	10

Note: The above parameters were obtained by pre-experiments.The ELM parameters used in this paper are obtained by looping through the global optimal solution, so there is no fixed parameter value.

analysis, and confidence intervals are calculated and coupled with the system point prediction values to obtain the final uncertainty prediction results.Details of the parameters of the model used by FMICM are shown in Table 2.

4. Experiments and Analysis

To validate the predictive performance of the developed integrated system, this thesis conducts experiments using three sets of electricity load data from March 2020 to November 2021 in New South Wales, Australia. The computer facility used for the experiments in this section of the study is matlab2018a with Windows 10 Home Edition, python3, with a 2.5GHz Intel(R) Core(TM) i5-7300HQ CPU.

4.1. Material

NEM operates in New South Wales, the Australian Capital Territory, Queensland, South Australia, Victoria and Tasmania as both a wholesale electricity market and a physical electricity system. Aemo also operates the retail electricity market that supports the wholesale market. The three datasets used in this paper are NEM statistics of the electricity load in New South Wales from March 2020 to November 2021, with one data point taken every half hour. Specifically, each dataset is a seven-month cycle of load data with 9,000 data points and partitioned to 1500 data points.The first 70% of the data is used as a training set to train individual models, 70% to 90% of the data is used to optimize the weights of each model, and the last 10% of the data is used to measure the predictive capabilities of the proposed system. In addition, the specific characteristics of the data set are shown in Table 3.

Table 3
The details of the three datasets utilized

Dataset	**Samples**	**Numbers**	**Statistical Indicator(MW)**			
			Max	**Min**	**Mean**	**Std.**
Site1	Training	6440	11980.08	5384.58	7722.83	1199.55
	Optimizing	1840	11908.24	6101.04	8578.39	1294.57
	Testing	920	11500.53	5630.73	7807.92	1189.47
	All samples	9200	11980.08	5384.58	7902.45	1264.42
Site2	Training	6440	12401.82	5221.13	7190.10	1002.82
	Optimizing	1840	12197.57	5704.44	7660.07	1088.47
	Testing	920	11404.28	5682.96	7408.21	933.01
	All samples	9200	12401.82	5221.13	7305.90	1031.25
Site3	Training	6440	12863.76	5170.46	8053.08	1351.40
	Optimizing	1840	12040.28	5189.86	7691.26	1174.77
	Testing	920	10236.28	4767.17	7005.83	1055.10
	All samples	9200	12863.76	4767.17	7875.99	1330.50

4.2 Evaluation Indicators

4.2.1 Point Forecast

The criterion we use to evaluate how good a point forecast is is to compare its forecast results with our true results and see the size of the difference between the two. In time series forecasting, Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), The Standard Deviation Of Error

(SDE) are our most frequently used and widely used the four evaluation metrics are the most frequently and widely used. In this paper, the above four metrics are selected as the evaluation criteria of the combined model.

Among these four evaluation criteria, MAPE does not only consider the error between the predicted and true values, but also the ratio between the error and the true value, which is one of the commonly used objective functions in some competitions. MAE is the absolute value of the difference between the predicted and true values for each sample, and then summed to find the average. The RMSE has the same properties as the MSE, but the error can be transformed into the same units as the original data. SDE is the standard deviation of the error, which can detect the model prediction stability. Let $\mathbf{TOV}_i$ be the i-th actual observation value and $\mathbf{PFV}_i$ be the i-th point predicted value, and the formula of evaluation index is shown in Table 4.

4.2.2 Interval Forecast

In interval prediction, the commonly used variables are PI coverage probability (PICP) and PI normalized averaged width (PINAW), and the interval score AIS selected in this paper is a tool used to provide comprehensive consideration of coverage probability and normalized averaged width. When the PICP is larger and the PINAW is smaller, the interval prediction result is better, and when the target is not in the PI coverage interval, AIS will give a certain penalty, so the larger the value of AIS

Table 4
Point-interval prediction results evaluation index

Metric	Nomenclature	Equation
MAPE	Mean Absolute Percentage Error	$MAPE=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left\lvert\frac{\mathbf{TOV}_i-\mathbf{PFV}_i}{\mathbf{TOV}_i}\right\rvert\times 100\%$
MAE	Mean Absolute Error	$MAE=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\lvert\mathbf{PFV}_i-\mathbf{TOV}_i\rvert$
RMSE	Root Mean Square Error	$RMSE=\sqrt{\frac{1}{\mathbf{N}}\times\sum_{i=1}^{\mathbf{N}}\left(\mathbf{PFV}_i-\mathbf{TOV}_i\right)^2}$
SDE	The standard deviation of error	$SDE=\sqrt{\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left(\mathbf{ME}_i-\mathbf{ERR}_i\right)^2}$
PICP	PI coverage probability	$PICP=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\tau_i^{\alpha} \qquad \tau_i^{\alpha}=\begin{cases}1 & \mathbf{TOV}_i^{\alpha}\in\left[\mathbf{FIL}_i^{\alpha},\mathbf{FIU}_i^{\alpha}\right]\\ 0 & \mathbf{TOV}_i^{\alpha}\notin\left[\mathbf{FIL}_i^{\alpha},\mathbf{FIU}_i^{\alpha}\right]\end{cases}$
PINAW	PI normalized averaged width	$PINAW=\frac{1}{\mathbf{N}R}\sum_{i=1}^{\mathbf{N}}\left(\mathbf{FIU}_i^{\alpha}-\mathbf{FIL}_i^{\alpha}\right)$
AIS	Average interval score	$AIS=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\gamma_i^{\alpha} \quad \gamma_i^{\alpha}=\begin{cases}-2\alpha\psi_i^{\alpha}-4\left(\mathbf{FIL}_i^{\alpha}-\mathbf{TV}_i\right) & \mathbf{TV}_i<\mathbf{FIL}_i^{\alpha}\\ -2\alpha\psi_i^{\alpha} & \mathbf{FIL}_i^{\alpha}\le\mathbf{TV}\le\mathbf{FIU}_i^{\alpha}\\ -2\alpha\psi_i^{\alpha}-4\left(\mathbf{TV}_i-\mathbf{FIU}_i^{\alpha}\right) & \mathbf{TV}_i>\mathbf{FIU}_i^{\alpha}\end{cases}$
MPICD	Mean PI center deviation	$MPICD=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left\lvert\frac{\mathbf{FIU}_i^{\alpha}+\mathbf{FIL}_i^{\alpha}}{2}-\mathbf{TOV}_i\right\rvert$

Note: MAPE , MAE , RMSE , SDE is the evaluation index of point prediction results, the smaller the value of all four indicators, the better. PICP, PINAW, AIS, MPICD is an indicator to evaluate the good or bad interval prediction results, Except for AIS, all

others are small is better, while AIS is large is better.
indicates the better quality of the prediction interval. If the coverage is the same and there is no difference in width, MPICD plays a role. If two different PIs cover a point, the closer to the midline of PI, the better the quality of PI, also known as the smaller the MPICD the better the prediction interval. Let $\mathbf{FIL}_i^{\alpha}$ be the upper limit of the first prediction interval and $\mathbf{FIU}_i^{\alpha}$ be the lower limit of the first prediction interval, and the formula of evaluation index is shown in Table 4.

4.3 Different experiments and result analysis

During this phase, four different experiments will be planned to investigate the prediction performance of the point or interval of the integrated prediction system developed in this paper. Since the results of each run of the model are different, the final data are averaged over the five results for fair comparison.

4.3.1 Experiment I: Comparison with the prediction model of three-hour interval data without FIG

In Experiment 1, the aim was to verify the enhancement of the predictive power of the fuzzy granulation technique used in the proposed point prediction system. The FMICM was then compared with five single models performed on non-fuzzy granulation data, namely DOABPNN, ELM, TCN, GRU and DBN, and the combined model MODOA, and the prediction results obtained from the experiment are shown in Table 5, and other details of the prediction results are shown below.

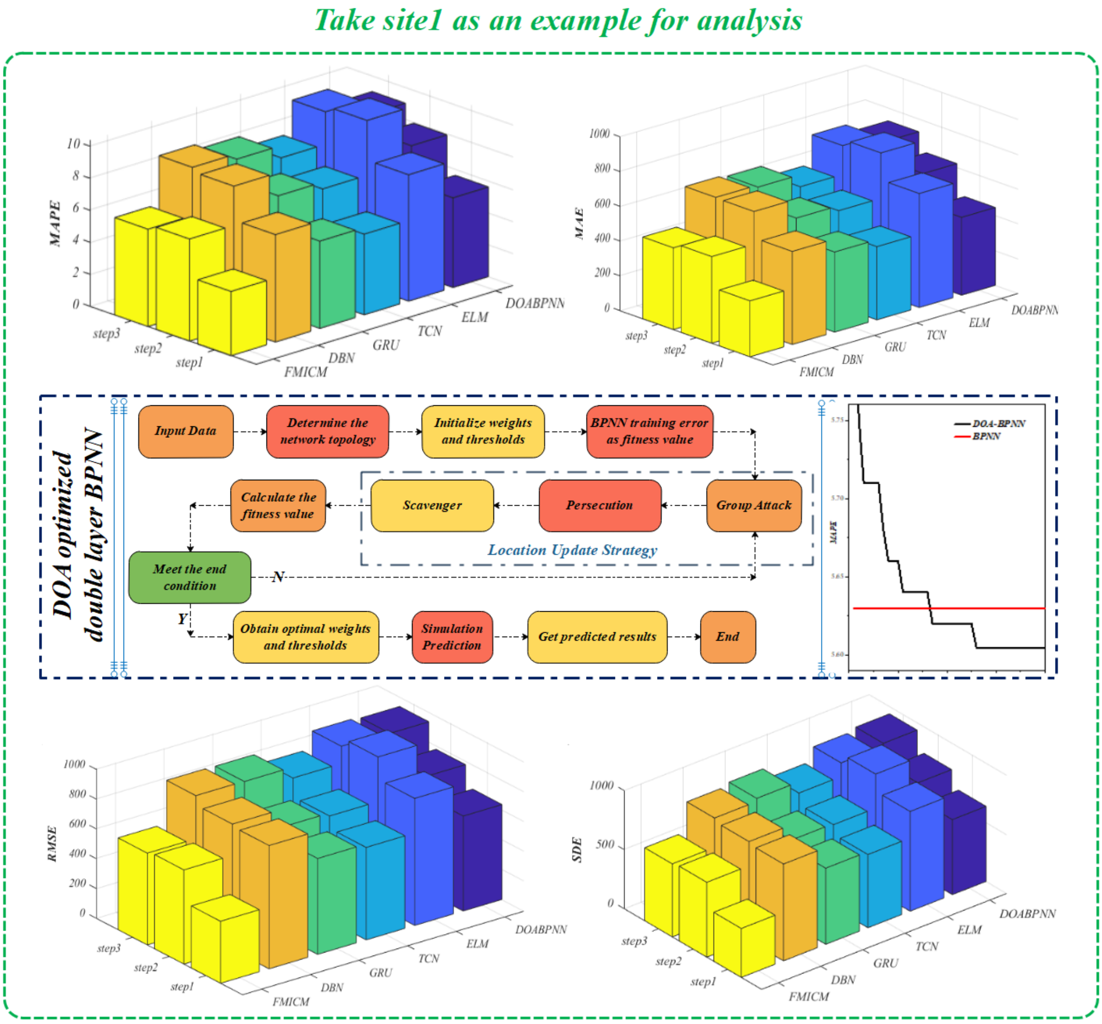

Figure 2 Comparison of the developed model with the single model of site1

(a) For site 1, when making a one-step prediction, the developing system shows a significant improvement in prediction accuracy and stability over the model executed with non-particleized data, $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step1\rangle}=4.0179\%,\mathbf{SDE}_{\mathbf{site1}}^{\langle step1\rangle}=416.638$. When making a two-step prediction,The developing FMICM: $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step2\rangle}=6.385\%,\mathbf{MAE}_{\mathbf{site1}}^{\langle step2\rangle}=498.132,\mathbf{RMSE}_{\mathbf{site1}}^{\langle step2\rangle}=632.713,\mathbf{SDE}_{\mathbf{site1}}^{\langle step2\rangle}=633.951$, which is an improvement com- -pared to both the single model and the combined model performed without fuzzy granular data. While performing the three-step prediction, the accuracy improvement from particleization is more obvious, and the mean absolute percentage error of FMICM is $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step3\rangle}=6.0869\%$, which is $\boldsymbol{\eta}=1.7303\%$ higher than that of the combined model without particleization. In conclusion, for dataset I, the prediction accuracy of the developed integrated system is significantly better than that of the unparticleized model.

(b) For site2, FMICM has the lowest MAPE of $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step1\rangle}=3.0992\%$ when making a one-step prediction.The highest prediction accuracy of the single model with unparticleized data is GRU with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step1\rangle}=3.5191\%$, and FMICM improves the prediction accuracy by $\boldsymbol{\gamma}=0.4199\%$ from both particleization and combination. During two step prediction, FMICM improves more in MAPE, but less in prediction stability. While performing the three-step prediction, the $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step3\rangle}=4.9097\%$ for FMICM. In summary, for dataset two, the FMICM model improved the prediction accuracy of both the single model and the combined model for unparticleized data.

(c) For site3, when making a one-step prediction, the highest prediction accuracy of the unparticleized single model is TCN with $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step1\rangle}=5.3836\%,\mathbf{MAE}_{\mathbf{site3}}^{\langle step1\rangle}=380.518,\mathbf{RMSE}_{\mathbf{site3}}^{\langle step1\rangle}=530.162,\mathbf{SDE}_{\mathbf{site3}}^{\langle step1\rangle}=530.509$. FMICM has improved over all unparticleized models.During two step prediction, the prediction accuracy of the non-particleized models of MODOA_CM,GRU,TCN,DBN,DOA-BPNN,ELM and $\overrightarrow{\mathbf{MAPE}}_{\mathbf{site3}}^{\langle step2\rangle}=\left[7.51\%,8.20\%,8.68\%,8.90\%,9.34\%,9.56\%\right]$ from low to high, respectively. While performing the three-step prediction, the unparticleized combined model has all improved over the single model with $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step3\rangle}=8.39\%$, $\mathbf{MAE}_{\mathbf{site3}}^{\langle step3\rangle}=550.45,\mathbf{RMSE}_{\mathbf{site3}}^{\langle step3\rangle}=686.709,\mathbf{SDE}_{\mathbf{site3}}^{\langle step3\rangle}=688.505$, but not as good as FMICM. In summary, for Dataset III, the combined FMICM model outperformed the unparticleized model in terms of prediction accuracy for any number of prediction steps.

Remark. Through Experiment 1, it was found that the developed FMICM outperformed the single and combined models performed on the unfuzzy granularized data, with the mean MAPE values of $\overrightarrow{\mathbf{MAPE}'}_{\mathbf{M}}=\left[3.9101\%,5.6910\%,6.3293\%\right]$ for the three-step prediction, respectively. In particular, by comparing FMICM with MODOA_CM, it was concluded that the necessity of using fuzzy particleization was effectively verified and FIG could not only improve the prediction accuracy but also the prediction stability. Figure 2 shows the measurements for the three datasets corresponding to Experiment 1.

4.3.2 Experiment II: Comparison with the single model after fuzzy particleization

Experiment 2 aims to verify the superiority of the multi-objective combinatorial optimization algorithm in FMICM, using the multi-objective combinatorial optimization algorithm to optimize the weights of the five single model point prediction results after fuzzy granulation in terms of both prediction accuracy and prediction stability, which is the role of the multi-objective optimization algorithm in FMICM, this experiment obtained five particleized single models DOABPNN,FIG_ELM,FIG_TCN, FIG_GRU,FIG_DBN)and FMICM The prediction results are shown in Table 6, and additional analyses of the experiments performed are described below.

(a) For site1, when making a one-step prediction, the best single-model prediction accuracy is FIG_TCN which $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step1\rangle}=4.1258\%$, and the worst prediction accuracy is FIG_ELM which $\overline{\overline{\mathbf{MAPE}}}_{\mathbf{site1}}^{\langle step1\rangle}=6.0396\%$. The multi-objective optimiza--tion algorithm improves the prediction accuracy and prediction stability of the single model. During two step prediction, FIG_GRU has the highest prediction accuracy in the single model with $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step2\rangle}=7.0318\%,\mathbf{SDE}_{\mathbf{site1}}^{\langle step2\rangle}=694.01$.FIG_DBN has the best

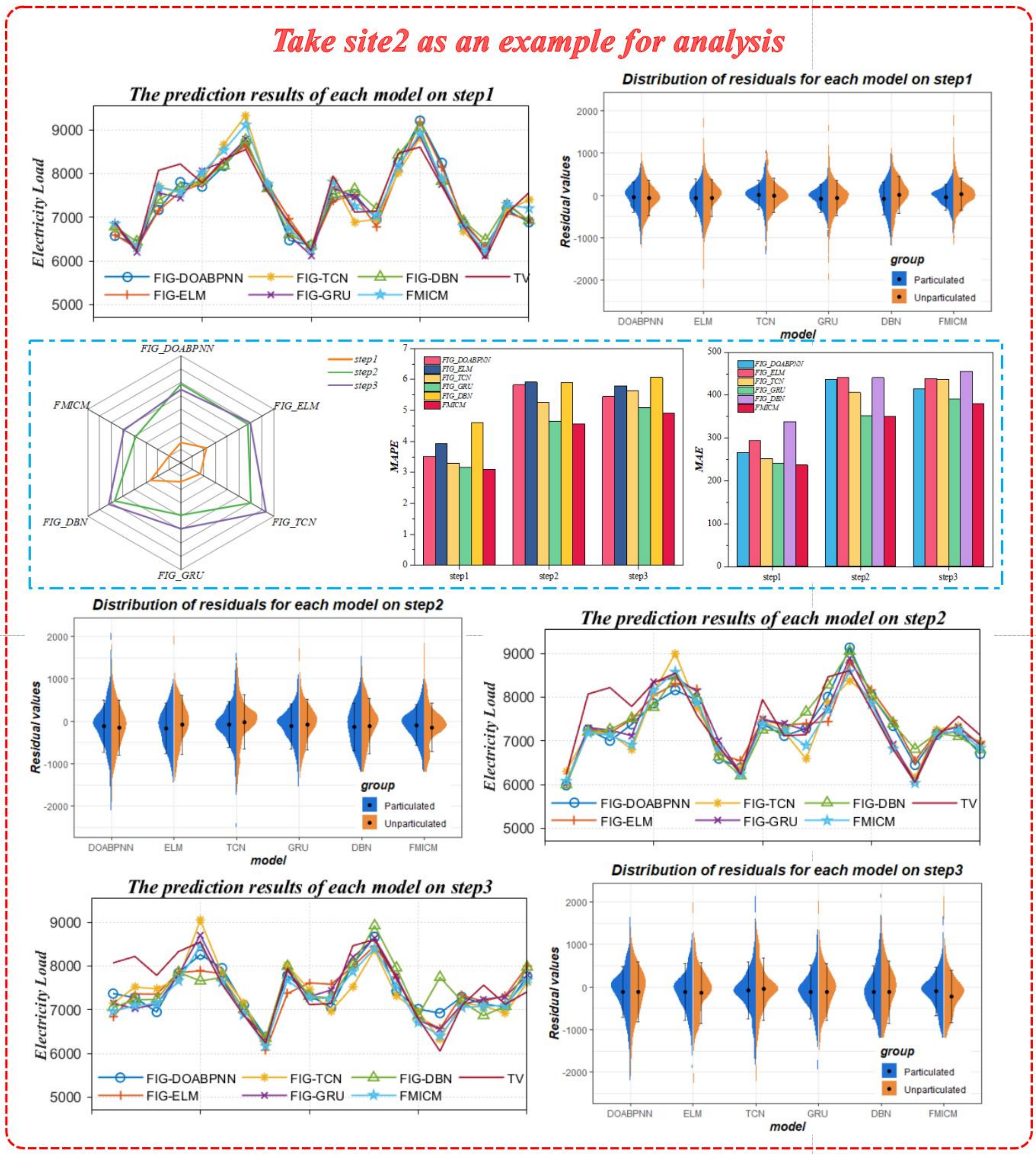

Figure 3 Comparison of point prediction performance of FMICM and different fuzzy post granulation single model

prediction stability with $\overline{\mathbf{MAPE}}_{\mathbf{site1}}^{\langle step2\rangle} = 7.583\%, \overline{\mathbf{SDE}}_{\mathbf{site1}}^{\langle step2\rangle} = 678.47$. While performing the three-step prediction, the prediction advantage of FMICM is more obvious, with MAPE optimizing $\vec{\Gamma} = [1.8212\%, 2.7249\%, 0.9703\%, 0.6468\%, 1.198\%]$ over FIG_DOABPNN,FIG_ELM, FIG_TCN, FIG_GRU, and FIG_DBN, respectively. It can be seen that the multi-objective combined optimization algorithm in FMICM not only improves the prediction accuracy of the single model, but also improves the prediction stability.

(b) For site2, when making a one-step prediction, FMICM has the best prediction in the comparison with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step1\rangle} = 3.0992\%, \mathbf{SDE}_{\mathbf{site2}}^{\langle step1\rangle} = 350.911$ .During two step prediction, FIG_GRU has the lowest MAPE among the single models with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step2\rangle} = 4.6607\%$ .While performing the three-step prediction, FIG_DBN has the worst prediction accuracy with $\overline{\mathbf{MAPE}}_{\mathbf{site2}}^{\langle step3\rangle} = 6.0711\%$, and FIG_GRU has the highest prediction accuracy with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step3\rangle} = 5.0823\%$. It can be concluded that the prediction accuracy of different models changes when the number of prediction steps changes, and the constant is that the prediction effect of FMICM is always higher than that of the single model.

(c) For site 3, when making a one-step prediction, FMICM has the highest prediction accuracy, $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step1\rangle} = 4.6133\%$, followed by FIG_TCN and FIG_GRU with $\overrightarrow{\overline{\mathbf{MAPE}''}}_{\mathbf{M}} = [4.8582\%, 4.9088\%]$. During two step prediction, the best prediction among the single models is FIG_GRU with $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step2\rangle} = 6.4856\%$, $\mathbf{MAE}_{\mathbf{site3}}^{\langle step2\rangle} = 443.829, \mathbf{RMSE}_{\mathbf{site3}}^{\langle step2\rangle} = 563.344, \mathbf{SDE}_{\mathbf{site3}}^{\langle step2\rangle} = 560.792$ .While performing the three-step prediction, FMICM has $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step3\rangle} = 7.9913\%$ and $\mathbf{SDE}_{\mathbf{site3}}^{\langle step3\rangle} = 686.44$, and the prediction accuracy and prediction stability are greatly improved compared with all single models.In summary, the prediction accuracy of different single models in different datasets is different, but the constant is that the prediction accuracy of FMICM is lower than the five single models in all datasets.

Remark. It was found through Experiment 2 that FMICM was lower than different single models in all cases, although different single models had different predictions for different datasets in different prediction steps.It effectively verifies the importance of using MODOA for optimization weights in FMICM. Figure 3 illustrates the comparison between FMICM and the single model after fuzzy granulation using the three-step prediction of site2 as an example.

4.3.3 Experiment III: Comparison with different combinatorial optimization algorithms

Experiment 3 aims to verify the superiority of the multi-objective combinatorial optimization algorithm MODOA, using the common Multi-Objective Grasshopper Optimization Algorithm (MOGOA), Multi-Objective Dragonfly Algorithm (MODA), and Multi-objective Ant Lion Optimizer (MOALO) to optimize the weights of the five models to derive the prediction accuracy and compare with MODOA. The prediction results obtained from the experiments are shown in Table 7, and additional analyses of the experiments performed are described below.

(a) For site1, when making a one-step prediction, the MAPE, MAE, RMSE, and SDE of FMICM are smaller than those of MOGOA, MODA, and MOALO. Among the other three optimization algorithms, the prediction accuracy of MOALO and MOGOA is

Table 5
Point prediction performance evaluation of the developed model versus the unparticleized model

		Step1				Step2				Step3			
		MAPE (%)	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**
Site1	**DOABPNN**	5.6046	448.026	637.542	635.344	7.9493	621.953	776.561	775.191	9.0131	727.162	935.608	937.986
	ELM	7.9007	655.939	853.257	844.484	10.381	810.961	1002.95	989.180	10.015	774.618	946.059	912.349
	TCN	5.0415	421.128	621.389	619.882	6.9572	549.579	703.333	701.069	7.9868	609.293	829.747	797.369
	GRU	5.4580	461.103	644.100	644.587	7.3809	577.946	723.920	710.267	8.7570	678.069	903.301	892.642
	DBN	6.7190	535.295	828.036	819.946	8.8344	686.772	843.464	838.471	9.1004	689.645	905.715	866.442
	MODOA_CM	4.6975	388.392	569.255	570.523	6.5111	521.455	676.519	678.701	7.8172	597.190	800.841	777.130
	Proposed System	**4.0179**	**321.758**	**416.638**	**414.616**	**6.3850**	**498.132**	**632.713**	**633.951**	**6.0869**	**470.232**	**616.445**	**618.279**
Site2	**DOABPNN**	4.1691	313.156	399.813	401.044	6.1514	464.755	589.673	589.824	6.5623	495.733	646.428	648.347
	ELM	4.3381	328.697	424.641	425.869	6.1716	463.347	589.192	590.616	6.6022	498.531	661.531	663.034
	TCN	3.8770	294.044	414.249	415.486	5.5674	416.095	558.469	560.053	6.2079	468.291	649.970	651.963
	GRU	3.5191	263.902	356.898	357.721	5.2193	398.847	528.558	530.173	5.6214	423.709	589.395	591.293
	DBN	4.3528	327.198	416.877	411.459	6.4435	483.789	619.123	621.074	7.3595	550.387	725.378	727.727
	MODOA_CM	3.4443	259.162	355.354	352.189	4.8229	370.430	504.468	495.958	5.4243	417.229	576.385	562.387
	Proposed System	**3.0992**	**236.660**	**350.526**	**350.911**	**4.5679**	**350.803**	**493.272**	**494.553**	**4.9097**	**379.496**	**543.547**	**544.917**
Site3	**DOABPNN**	6.6534	466.879	648.938	649.313	9.3363	625.151	807.708	793.534	10.653	715.515	855.930	838.011
	ELM	7.2865	500.857	646.628	646.959	9.5614	641.999	785.701	786.879	10.162	682.328	822.744	810.251
	TCN	5.3836	380.518	530.162	530.509	8.6810	572.062	745.951	722.123	8.8908	606.923	749.218	748.232
	GRU	5.8537	406.919	533.777	518.127	8.2019	543.124	697.459	650.953	9.2796	624.933	794.546	755.997
	DBN	6.5548	455.728	568.781	559.712	8.8986	595.710	714.186	707.779	9.0915	621.257	745.608	742.357
	MODOA_CM	5.1344	364.044	489.874	491.512	7.5136	504.006	648.172	640.048	8.3931	580.929	686.709	688.505
	Proposed System	**4.6133**	**333.260**	**486.851**	**487.526**	**6.1200**	**430.872**	**557.819**	**558.744**	**7.9913**	**550.450**	**684.534**	**686.440**

Note: The above table shows the point prediction performance results (including MAPE, MAE, RMSE, SDE) using the developed combined prediction models and single models (including DOA_BPNN, TCN, DBN, ELM, GRU) without fuzzy particleization, using data for three-hour intervals.

Table 6

Evaluation of the point prediction performance of the developed models with different post-fuzzy granulation single models

		Step1				Step2				Step3			
		MAPE (%)	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**
Site1	**FIG_DOABPNN**	4.9684	399.265	532.508	522.476	8.2062	654.596	880.329	870.333	6.6247	506.672	653.973	592.987
	FIG_ELM	6.0396	471.739	588.403	519.548	9.1099	709.728	921.726	806.321	9.0080	687.948	884.319	758.503
	FIG_TCN	4.1258	328.608	427.564	417.030	7.3553	571.015	735.764	699.498	7.1392	540.848	728.087	685.458
	FIG_GRU	5.0158	396.544	517.905	475.442	7.0318	538.976	733.077	694.009	6.9224	520.653	768.293	727.059
	FIG_DBN	4.7983	379.957	488.176	463.817	7.5830	576.587	739.802	678.470	7.5961	577.573	765.189	733.006
	Proposed System	**4.0179**	**321.758**	**416.638**	**414.616**	**6.3850**	**498.132**	**632.713**	**633.951**	**6.0869**	**470.232**	**616.445**	**618.279**
Site2	**FIG_DOABPNN**	3.5169	265.705	374.444	375.622	5.8214	436.837	596.627	594.262	5.4540	414.121	573.203	573.727
	FIG_ELM	3.9383	294.740	409.041	410.411	5.9116	441.177	589.239	588.182	5.7952	438.026	596.774	598.634
	FIG_TCN	3.2850	251.480	381.929	383.208	5.2595	406.573	599.519	601.337	5.6436	436.958	664.450	666.480
	FIG_GRU	3.1566	240.894	371.356	370.819	4.6607	352.647	495.158	495.693	5.0823	390.853	545.409	546.343
	FIG_DBN	4.6063	337.851	431.639	431.669	5.9073	440.785	587.588	585.270	6.0711	455.991	609.970	610.172
	Proposed System	**3.0992**	**236.660**	**350.526**	**350.911**	**4.5679**	**350.803**	**493.272**	**494.553**	**4.9097**	**379.496**	**543.547**	**544.917**
Site3	**FIG_DOABPNN**	5.5916	394.488	545.716	547.371	10.147	694.438	845.432	847.188	8.9382	606.866	741.349	734.524
	FIG_ELM	8.3063	600.123	756.186	758.703	12.210	846.379	1086.97	1083.225	10.384	700.653	870.539	863.651
	FIG_TCN	4.8582	354.546	514.617	513.972	6.7052	476.987	626.195	625.655	8.2805	567.912	709.697	710.466
	FIG_GRU	4.9088	352.212	491.731	493.374	6.4856	443.829	563.344	560.792	8.8005	588.317	762.384	742.800
	FIG_DBN	7.2293	502.518	699.202	701.003	9.7904	656.282	808.539	775.814	10.258	684.650	830.367	774.567
	Proposed System	**4.6133**	**333.260**	**486.851**	**487.526**	**6.1200**	**430.872**	**557.819**	**558.744**	**7.9913**	**550.450**	**684.534**	**686.440**

Note: The above table shows the point prediction performance results (including MAPE, MAE, RMSE, SDE) of the developed combined prediction models and the fuzzy particleized single models (including FIG_DOABPNN,FIG_ELM,FIG_TCN,FIG_GRU,FIG_DBN).

higher, and the MAPE of FMICM has different degrees of improvement compared with them. During two step prediction, there is almost no difference in the prediction accuracy of the other three optimization algorithms, while the prediction accuracy of FMICM improves about $\ddot{\boldsymbol{\omega}} = 0.6111\%$ compared with these three algorithms. While performing the three-step prediction, the prediction accuracy of FMICM improves more, and the MAPE of FMICM, MOGOA, MODA, and MOALO are $\overrightarrow{\mathbf{MAPE}}_{\mathbf{site1}}^{\langle step3 \rangle} = [6.087\%, 6.635\%, 6.271\%, 6.652\%]$ .In summary, among the three compared optimization algorithms, for this dataset, MOGOA and MOALO are better at optimizing the first two steps of prediction, and MODA is better at optimizing the three steps of prediction, but neither is as good as not as good as FMICM.

(b) For site2, FMICM has the highest prediction accuracy when making a one-step prediction. The MAPE of the other three optimization algorithms is $\overrightarrow{\mathbf{MAPE'}}_{\mathbf{site2}}^{\langle step1 \rangle} = [3.1401\%, 3.1741\%, 3.1579\%]$. During two step prediction, the prediction accuracy of MODA and MOALO is higher with the exception of FMICM.

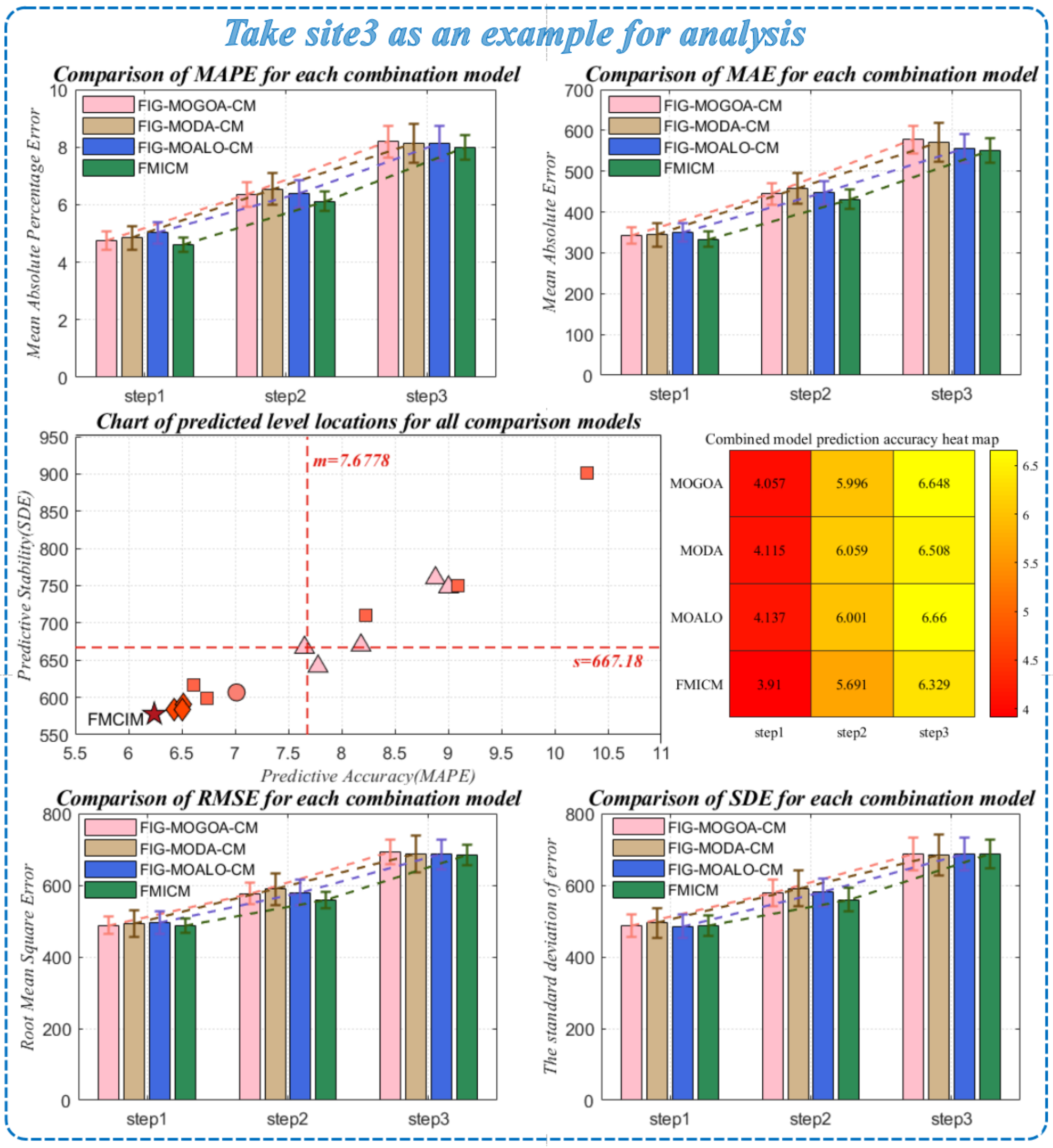

Figure 4 Comparison FMICM with other optimization models of site3

$\overrightarrow{\mathbf{MAPE}''}^{\langle step2\rangle}_{\mathbf{site2}} = [4.6295\%, 4.6367\%], \overrightarrow{\mathbf{SDE}''}^{\langle step2\rangle}_{\mathbf{site2}} = [501.353, 499.359]$. While making the three-step prediction, FMICM has the highest prediction accuracy, MODA the second and MOALO the worst with $\overrightarrow{\mathbf{MAE}''}^{\langle step3\rangle}_{\mathbf{site2}} = [379.496, 390.295, 396.607]$. In summary. FMICM outperformed the three algorithms compared, despite the fact that the other optimization algorithms were sometimes strong and weak in their ability to optimize at different prediction steps.

(c) For site 3, when making a one-step prediction, the optimal of the other three optimization algorithms is MOGOA, $\mathbf{MAPE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 4.7607\%, \mathbf{MAE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 343.376$, $\mathbf{MAE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 487.886, \mathbf{SDE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 487.579$. During two step prediction, the SDE of FMICM is the smallest, followed by MOGOA, MODA and MOALO, $\overrightarrow{\overrightarrow{\mathbf{SDE}}}^{\langle step2\rangle}_{\mathbf{site3}} = [558.744, 578.644, 590.548, 581.334]$. While performing the three-step prediction, The prediction accuracy of FMICM is significantly improved, and its MAE is $\vec{\chi}_{\mathbf{MAE}} = [27.154, 20.665, 6.284]$ compared to MOGOA, MODA and MOALO. In conclusion, the optimization ability of FMICM in site3 is proved, and compared with the previous combined models, the prediction accuracy has been greatly improved compared to the previous combined models.

Remark. Through experiment three, it was found that the weight optimization ability of Multi-Objective Dingo Optimization Algorithm in FMICM model surpassed the known Multi-Objective Grasshopper Optimization Algorithm (MOGOA), Multi - Objective Dragonfly Algorithm (MODA), and Multi-objective Ant Lion Optimizer (MOALO), resulting in sufficient improvement of the final prediction accuracy and effectively validating the importance of MODOA in FMICM. Figure 4 shows how the developed FMICM compares with the combined model using different optimization algorithms.

4.3.4 Experiment IV: Comparison with all model interval estimates

The experiments in this section evaluate the interval estimation results by combining the evaluation metrics AIS for PI coverage probability and PI normalized averaged width and MPICD for evaluating the interval prediction accuracy, with the aim of comparing the developed FMICM model with a single model after fuzzy granulation and different combinations of optimization models to demonstrate that FMICM model is not only the best in point prediction, but also maintains excellent performance in interval estimation. The final test results are shown in Tables 8-9, and the details of this experiment are as follows.

(a) For site1,when making a one-step prediction, the PICP of FIG_ELM is as high as $\tilde{\boldsymbol{\rho}}_{1;1}^{\mathbf{ELM}} = 100\%$, but then the PIAW is as high as $\tilde{\boldsymbol{\omega}}_{1;1}^{\mathbf{ELM}} = 0.4772$, in other words, the high coverage of this model is due to the large PI normalized averaged width. Therefore, we mainly used AIS and MPICD for comparison. With a confidence factor of 95%, the optimal models for AIS in the three-step prediction are FIG_MOGOA_CM,FMICM,FMICM with AIS values of $\vec{\boldsymbol{\Lambda}}'_{\mathbf{FMICM}} = [-237.2, -400.9, -424.4]$. With a confidence factor of 90%, the optimal models for MPICD in the three-step prediction are FMICM,FMICM, FIG_MOGOA_CM, which have MPICD values of $\vec{\mathbf{D}}''_{\mathbf{MPICD}} = [384.94, 574.34, 531.74]$.Therefore, the interval prediction of FMICM in site1 is the best, followed by FIG_MOGOA_CM.

(b) For site2, FMICM performs best in the one-step prediction with $\tilde{\boldsymbol{\Lambda}}_{2;1}^{\mathbf{FMICM}} = -215.9$ and $\tilde{\mathbf{D}}_{2;1}^{\mathbf{FMICM}} = 255.51$ when the confidence coefficient is 95%. The best AIS in the two-step prediction is FMICM and the smallest MPICD is FIG_MODA_CM. The three-step prediction of FIG_GRU has an AIS of $\tilde{\boldsymbol{\Lambda}}_{2;3}^{\mathbf{GRU}} = -279.7$, which is better than the combined model, and the smallest MPICD is FIG_MOGOA_CM with a value of $\tilde{\mathbf{D}}_{2;3}^{\mathbf{MOGOA}} = 400.05$. The results are consistent with the above when the confidence factor is 90%. It is worth mentioning that the interval coverage of the single model here are higher than the combined model. The reason is that the residuals of the single model are larger, resulting in larger intervals obtained from the kernel density estimation curve. In summary, most experiments show that the interval prediction of FMICM is better than other comparative models.

(c) For site3, both AIS and MPICD for the 95% confidence interval of FIG_MOGOA_CM was optimal in the one-step prediction case with $\tilde{\boldsymbol{\Lambda}}_{3;1}^{\mathbf{MOGOA}} = -268.7$ and

$\tilde{\mathbf{D}}_{3;1}^{\mathbf{MOGOA}} = 339.87$. In the two-step prediction case, both AIS and MPICD for the 95% confidence interval of FMICM were optimal with $\tilde{\mathbf{\Lambda}}_{3;2}^{\mathbf{FMICM}} = -255.5, \tilde{\mathbf{D}}_{3;2}^{\mathbf{FMICM}} = 427.1$. FIG-DOABPNN emerges as the best in the three-step prediction with an AIS of $\tilde{\mathbf{\Lambda}}_{3;3}^{\mathbf{D-BPNN}} = -323$, which is better than all types of combined models. When the confidence coefficient is equal to 90%, FMICM performs optimally in all three prediction steps with AIS of $\vec{\mathbf{\Lambda}}''_{\mathbf{FMICM}} = [-499.5, -435.3, -495.4]$, and MPICD of $\vec{\mathbf{D}}''_{\mathbf{FMICM}} = [333.45, 428.44, 550.09]$. In summary, the experiments for dataset three show that the interval prediction of FMICM is better than other comparative models.

Remark. The interval predictions of FMICM were compared with those of eight models by Experiment 4. At $\boldsymbol{\lambda}' = 95\%$ confidence factor, 5/9 experiments proved that FMICM has the best AIS and MPICD. 89% experiments proved that FMICM has higher

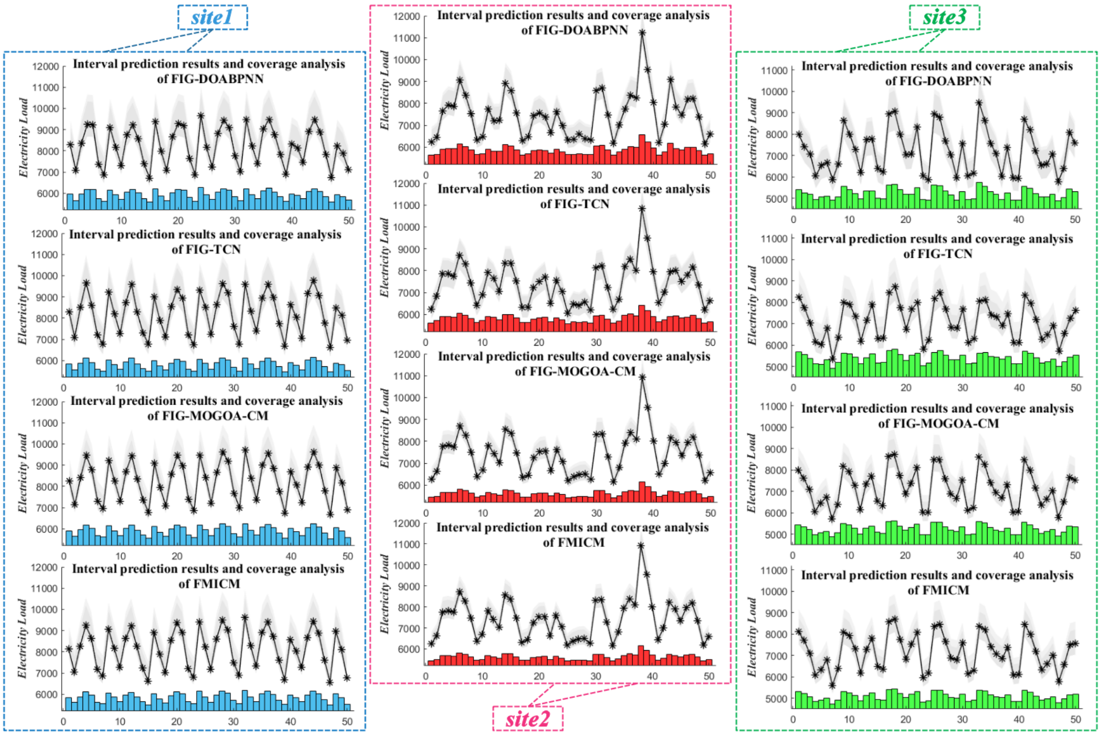

Figure 5 The interval prediction of the developed FMICM with other models

Table 7
Combined model point prediction performance table using different optimization algorithms

		Step1				**Step2**				**Step3**			
		MAPE (%)	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**
Site1	**FIG _MOGOA_CM**	4.2704	338.142	432.497	**405.099**	6.9922	535.377	693.019	638.555	6.6352	496.552	695.786	625.786
	FIG _MODA_CM	4.3205	342.026	438.719	409.726	7.0011	535.801	696.299	640.186	6.2708	474.810	655.924	631.960
	FIG _MOALO_CM	4.2302	335.487	430.500	405.677	6.9951	535.340	693.937	639.094	6.6521	497.727	697.229	625.518
	Proposed System	**4.0179**	**321.758**	**416.638**	414.616	**6.3850**	**498.132**	**632.713**	**633.951**	**6.0869**	**470.232**	**616.445**	**618.279**
Site2	**FIG _MOGOA_CM**	3.1401	239.097	351.361	352.149	4.6525	355.409	504.878	506.311	5.1226	391.791	554.369	552.644
	FIG _MODA_CM	3.1741	241.407	351.748	352.879	4.6295	353.303	499.822	501.353	5.1083	390.295	545.312	546.086
	FIG _MOALO_CM	3.1579	240.004	351.406	352.575	4.6367	353.405	497.981	499.359	5.1994	396.607	558.387	557.081
	Proposed System	**3.0992**	**236.660**	**350.526**	**350.911**	**4.5679**	**350.803**	**493.272**	**494.553**	**4.9097**	**379.496**	**543.547**	**544.917**
Site3	**FIG _MOGOA_CM**	4.7607	343.376	487.886	487.579	6.3440	444.325	576.726	578.644	8.1858	577.604	693.029	685.887
	FIG _MODA_CM	4.8519	344.288	493.998	495.394	6.5449	457.369	588.995	590.548	8.1434	571.115	687.497	**685.29**
	FIG _MOALO_CM	5.0236	350.271	495.783	**485.023**	6.3708	446.596	579.396	581.334	8.1275	556.734	686.473	686.609
	Proposed System	**4.6133**	**333.260**	**486.851**	487.526	**6.1200**	**430.872**	**557.819**	**558.744**	**7.9913**	**550.450**	**684.534**	686.440

Note: The above table shows the point prediction evaluation results (using four metrics MAPE, MAE, RMSE, SDE) of the developed FMICM(Proposed System) optimized using MODOA in combination with models using other three different optimization algorithms in combination (including FIG_MOGOA_CM, FIG_MODA_CM, FIG_ MOALO_CM).

Table 8
Comparison of interval predictions of the development system with other models at a confidence coefficient of 0.95.

α = 0.05		Step1				Step2				Step3			
		PICP (%)	PIAW	AIS	MPICD	PICP (%)	PIAW	AIS	MPICD	PICP (%)	PIAW	AIS	MPICD
Site1	FIG_DOABPNN	87.33	0.2390	-346.1	541.17	85.33	0.3726	-548.1	830.87	95.33	0.3931	-342.2	590.77
	FIG_ELM	1	0.4772	-379.9	511.01	99.33	0.5878	-474.1	729.26	93.33	0.4739	-426.5	777.67
	FIG_TCN	89.33	0.2236	-259.0	439.75	93.33	0.4171	-433.7	673.84	92.67	0.3750	-428.3	613.69
	FIG_GRU	84.67	0.2276	-337.3	506.87	87.33	0.3299	-521.3	670.47	93.33	0.3899	-511.3	592.29
	FIG_DBN	95.33	0.2457	-255.6	431.55	86.67	0.3535	-466.4	728.09	94.67	0.3966	-422.2	613.89
	FIG _MOGOA_CM	92.00	0.2366	**-237.2**	415.53	90.00	0.3332	-416.5	**626.09**	92.67	0.3317	-438.3	555.39
	FIG _MODA_CM	92.67	0.2306	-240.2	411.18	88.67	0.3302	-422.5	626.82	92.67	0.3393	-435.6	560.27
	FIG _MOALO_CM	88.67	0.2101	-279.3	460.88	88.67	0.3283	-423.9	628.74	92.67	0.3367	-435.3	554.08
	Proposed System	92.67	0.2296	-239.6	**408.34**	90.00	0.3413	**-400.9**	626.81	92.67	03414	**-424.4**	**553.54**
Site2	FIG_DOABPNN	94.00	0.1833	-222.0	305.88	94.00	0.3422	-304.8	443.19	96.67	0.3301	-301.1	487.83
	FIG_ELM	94.00	0.2003	-216.1	327.51	95.33	0.3331	-313.8	495.37	96.67	0.3294	-289.5	501.27
	FIG_TCN	94.00	0.1783	-235.8	258.78	90.67	0.2797	-340.9	446.96	94.67	0.3620	-362.9	463.10
	FIG_GRU	92.00	0.1645	-239.1	289.01	94.67	0.3064	-272.0	500.15	97.33	0.3408	**-279.7**	452.73
	FIG_DBN	92.67	0.1909	-222.8	356.32	95.33	0.2873	-310.4	447.15	94.00	0.3131	-285.1	464.06
	FIG _MOGOA_CM	91.33	0.1505	-219.5	266.17	95.33	0.2460	-288.7	362.48	95.33	0.2854	-296.3	**400.05**
	FIG _MODA_CM	92.67	0.1508	-219.8	264.32	94.67	0.2443	-286.2	**361.11**	95.33	0.2848	-287.8	408.78
	FIG _MOALO_CM	90.67	0.1498	-220.4	268.33	94.67	0.2436	-284.4	361.27	94.67	0.2825	-298.8	403.46
	Proposed System	92.00	0.1507	**-215.9**	**255.51**	94.00	0.2431	**-259.1**	390.42	96.00	0.2899	-290.9	401.96
Site3	FIG_DOABPNN	86.00	0.2403	-353.7	408.10	86.67	0.407	-347.3	699.04	99.33	0.4574	**-323.0**	599.50
	FIG_ELM	89.33	0.3548	-337.5	610.36	91.33	0.4997	-581.4	855.78	97.33	0.5418	-409.3	701.07
	FIG_TCN	93.33	0.2644	-273.1	398.16	1	0.4663	-327.4	473.94	1	0.5322	-372.4	568.26
	FIG_GRU	87.33	0.2119	-293.8	360.33	1	0.3647	-256.1	442.90	98.00	0.5202	-397.8	586.99
	FIG_DBN	83.33	0.2929	-471.2	549.93	94.00	0.4243	-317.3	626.12	99.33	0.4987	-351.0	669.06
	FIG _MOGOA_CM	92.00	0.2438	**-268.7**	**339.87**	99.33	0.3788	-266.0	443.35	1	0.5039	-352.5	557.60
	FIG _MODA_CM	92.00	0.2485	-272.7	348.78	99.33	0.3804	-268.1	457.12	1	0.5055	-353.6	557.88
	FIG _MOALO_CM	92.00	0.2415	-269.9	340.13	98.67	0.3697	-260.9	445.61	99.33	0.5127	-361.4	556.51
	Proposed System	90.00	0.2234	-281.7	361.57	99.33	0.3626	**-255.5**	**427.10**	99.33	0.5165	-363.9	**548.99**

Table 9
Comparison of interval predictions of the development system with other models at a confidence coefficient of 0.9.

α = 0.1		Step1				Step2				Step3			
		PICP (%)	PIAW	AIS	MPICD	PICP (%)	PIAW	AIS	MPICD	PICP (%)	PIAW	AIS	MPICD
	FIG_DOABPNN	80.00	0.1845	-556.6	496.33	84.00	0.3257	-831.4	746.20	92.00	0.3234	-634.9	556.88
	FIG_ELM	96.67	0.2979	-511.1	498.87	95.33	0.4699	-804.0	726.26	91.33	0.3893	-747.3	727.72
	FIG_TCN	88.00	0.1903	-415.5	387.31	86.00	0.2878	-691.5	628.67	90.00	0.2968	-696.2	592.99
	FIG_GRU	80.00	0.1786	-515.6	457.67	78.67	0.2468	-794.3	602.68	88.67	0.3018	-781.5	572.76
Site1	FIG_DBN	85.33	0.1727	-420.3	396.73	85.33	0.2883	-687.9	613.10	91.33	0.3228	-688.9	598.23
	FIG _MOGOA_CM	78.67	0.1451	-428.0	385.68	83.33	0.2412	-672.0	**574.34**	90.00	0.2894	-679.7	**531.74**
	FIG _MODA_CM	82.00	0.1556	-410.6	387.88	82.67	0.2364	-683.9	575.23	89.33	0.2815	-681.5	538.40
	FIG _MOALO_CM	0.82	0.1573	-421.0	393.45	82.67	0.2333	-685.0	575.49	90.00	0.2929	-678.8	531.28
	Proposed System	86.00	0.1603	**-403.9**	**384.94**	82.67	0.2373	**-664.9**	575.92	90.00	0.2911	**-671.4**	532.50
	FIG_DOABPNN	90.67	0.1474	-347.7	300.31	90.00	0.2433	-517.7	443.14	92.67	0.2680	-499.9	480.55
	FIG_ELM	84.67	0.1505	-365.9	330.78	86.00	0.2291	-524.9	494.65	87.33	0.2595	-510.5	483.88
	FIG_TCN	91.33	0.1573	-361.8	253.30	85.33	0.2109	-572.8	428.96	90.67	0.2758	-621.1	486.38
	FIG_GRU	88.00	0.1239	-360.3	269.62	88.00	0.2076	-449.6	425.01	92.67	0.2592	**-469.3**	427.80
Site2	FIG_DBN	84.67	0.1612	-370.9	352.56	88.00	0.2179	-496.4	441.08	85.33	0.2536	-517.8	463.25
	FIG _MOGOA_CM	84.67	0.1152	-331.6	253.54	89.33	0.1890	-453.4	361.22	91.33	0.2337	-489.2	421.54
	FIG _MODA_CM	86.00	0.1162	-332.0	253.67	90.00	0.1877	-447.3	**359.33**	92.00	0.2342	-478.5	**411.40**
	FIG _MOALO_CM	85.33	0.1140	-336.6	253.89	90.67	0.1878	-443.6	359.28	90.67	0.2360	-494.1	419.89
	Proposed System	85.33	0.1151	**-330.2**	**252.05**	86.67	0.1861	**-432.9**	381.49	92.00	0.2305	-479.9	411.89
	FIG_DOABPNN	84.00	0.2090	-539.2	396.99	71.33	0.2892	-788.2	696.21	67.33	0.2340	-716.0	615.05
	FIG_ELM	79.33	0.2663	-642.8	604.86	70.00	0.3292	-1084	840.85	80.67	0.3501	-728.1	691.83
	FIG_TCN	81.33	0.1638	-532.9	354.03	92.00	0.3155	-506.2	473.32	90.00	0.3168	-521.7	567.99
	FIG_GRU	79.33	0.1672	-510.8	352.64	90.00	0.2714	-454.2	443.59	86.00	0.2820	-574.7	589.60
Site3	FIG_DBN	80.00	0.2382	-713.0	502.99	78.67	0.3019	-651.8	626.05	76.00	0.2786	-610.8	660.34
	FIG _MOGOA_CM	78..67	0.1538	-502.1	338.86	90.00	0.2771	-448.6	445.50	79.33	0.2662	-517.3	563.2
	FIG _MODA_CM	78.00	0.1545	-516.4	344.20	85.33	0.2573	-460.5	460.69	81.33	0.2679	-512.8	563.21
	FIG _MOALO_CM	78.67	0.1538	-502.8	338.64	90.00	0.2743	-449.5	447.87	80.67	0.2693	-513.2	559.11
	Proposed System	80.67	0.15607	**-499.5**	**333.45**	92.00	0.2788	**-435.3**	**428.44**	90.00	0.3031	**-495.4**	**550.09**

interval prediction accuracy at $\lambda''=90\%$ confidence factor. Additional individual experiments showed that FIG_ MOGOA_CM and FIG_MODA_CM had better MPICD. It can be concluded from the interval prediction tests that the developed FMICM model proved to have excellent interval prediction performance in most of the experiments at the significance level in the experiments. Figure 5 shows how the developed FMICM compares to the eight models in terms of interval prediction.

5. Discussion

In this section, we further analyze the prediction results of four experiments, including the following four main components: Diebold-Mariano (DM)-test, improvement ratio of the indexes, forecasting effectiveness test, sensitivity analysis, convergence analysis and the empirical power load analysis. The detailed testing procedures are described below.

5.1 Diebold-Mariano (DM)-test

Since there are only a few data in the test set in the experiment, comparison of the prediction results can only indicate that the combined model proposed in this sample works better, and the data sampling is not good enough to cause this situation, in order to determine whether it is a fluke caused by the situation, the difference between model A and model B needs to be calculated statistically to be significant, that is a DM test.

Definition 1: Suppose the predicted values of the two models to be compared are $\overrightarrow{\mathbf{Pf}}_1=\left[\mathbf{P}\hat{\mathbf{f}}_1^{(1)},\mathbf{P}\hat{\mathbf{f}}_1^{(2)},\cdots,\mathbf{P}\hat{\mathbf{f}}_1^{(\phi)}\right]$ and $\overrightarrow{\mathbf{Pf}}_2=\left[\mathbf{P}\hat{\mathbf{f}}_2^{(1)},\mathbf{P}\hat{\mathbf{f}}_2^{(2)},\cdots,\mathbf{P}\hat{\mathbf{f}}_2^{(\phi)}\right]$, and the true values are $\overrightarrow{\mathbf{To}}_*=\left[\mathbf{T}\hat{\mathbf{o}}_*^{(1)},\mathbf{T}\hat{\mathbf{o}}_*^{(2)},\cdots,\mathbf{T}\hat{\mathbf{o}}_*^{(\phi)}\right]$. From this, the prediction error of the two models to be compared can be calculated as $\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_1=\left[\mathbf{E}\hat{\mathbf{r}}_1^{(1)},\mathbf{E}\hat{\mathbf{r}}_1^{(2)},\cdots,\mathbf{E}\hat{\mathbf{r}}_1^{(\phi)}\right]\mid\mathbf{E}\hat{\mathbf{r}}_1^{(\boldsymbol{\alpha})}=\mathbf{P}\hat{\mathbf{f}}_1^{(\boldsymbol{\alpha})}-\mathbf{T}\hat{\mathbf{o}}_*^{(\boldsymbol{\alpha})}$, $\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_2=\left[\mathbf{E}\hat{\mathbf{r}}_2^{(1)},\mathbf{E}\hat{\mathbf{r}}_2^{(2)},\cdots,\mathbf{E}\hat{\mathbf{r}}_2^{(\phi)}\right]\mid\mathbf{E}\hat{\mathbf{r}}_2^{(\boldsymbol{\alpha})}=\mathbf{P}\hat{\mathbf{f}}_2^{(\boldsymbol{\alpha})}-\mathbf{T}\hat{\mathbf{o}}_*^{(\boldsymbol{\alpha})}$.

Based on the above preparatory work, the null hypothesis and alternative hypothesis are presented.

$$\begin{aligned}H_0:\mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_1^{(\boldsymbol{\mu})}\right)\right]-\mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_2^{(\boldsymbol{\mu})}\right)\right]=0\\ H_1:\mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_1^{(\boldsymbol{\mu})}\right)\right]-\mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_2^{(\boldsymbol{\mu})}\right)\right]\neq 0\end{aligned}\tag{1}$$

Where the loss function $\boldsymbol{\Omega}(\vec{\chi})$ is calculated as $\boldsymbol{\Omega}(\vec{\chi})=\vec{\chi}^2$, and the constructed DM test statistic is:

$$\mathbf{DM}=\frac{\sum_{\boldsymbol{\mu}=1}^{\boldsymbol{\Pi}}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_1^{(\boldsymbol{\mu})}\right)-\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_2^{(\boldsymbol{\mu})}\right)\right]}{\boldsymbol{\Pi}\sqrt{S^2/\boldsymbol{\Pi}}}\tag{2}$$

Where S^2 refers to the variance of $\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_1^{(\boldsymbol{\mu})}\right)-\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_2^{(\boldsymbol{\mu})}\right)$. The DM test theory assumes that the distribution of the DM test statistic satisfies the standard normal distribution when the significance level is set to α, so the rejection domain is $W=\left\{|\mathbf{DM}|>\left|z_{\alpha/2}\right|\right\}$. When the DM statistic falls into the rejection domain, the original hypothesis is rejected, that is, there is a significant difference between the two prediction models,otherwise when $|\mathbf{DM}|\le\left|z_{\alpha/2}\right|$, there is no reason to reject the original hypothesis, which means that there is no statistically significant difference in the predictive power of the two models.

The DM test computes the predictive validity of this integrated system point estimate and further validates the performance of the combined model against statistical ideas. The test results are shown in Table 10, and other details are shown below.

(a) Comparison with the single model, when the significance level is set to $\alpha' = 0.05$, it can be seen that the majority of DM values are greater than $\bar{\mathbf{z}} = 1.96$, rejecting the original hypothesis that the developed point prediction system is better than the single model before fuzzy particleization. Setting the significance level to $\alpha' = 0.05$, DOABPNN,ELM,TCN,GRU, and DBN in the single model had DM test pass rates of $\overrightarrow{\mathbf{PR}'} = [[67\%, 100\%, 56\%, 56\%, 100\%]$. When the significance level was set to $\alpha'' = 0.1$,The DM test pass rates of DOABPNN, ELM, TCN, GRU, and DBN were $\overrightarrow{\mathbf{PR}''} = [100\%, 100\%, 78\%, 56\%, 100\%]$. In summary, the single model before fuzzy particleization is significantly different from FMICM. Since the DM values are all greater than 0, it indicates that the point prediction effect of FMICM is better than that of the single model before fuzzy particleization, which verifies the conclusion drawn in Experiment 1.

(b) Compared to the single model after fuzzy particleization, $\mathbf{PR}' = 73\%$ of the data passed the test when the significance level was $\alpha' = 0.05$, with FIG_DBN passing all of them. The test pass rate for FIG_DOABPNN was $\mathbf{PR}'_{\mathbf{BPNN}} = 56\%$ at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{BPNN}} = 67\%$ at $\alpha'' = 0.1$. The test pass rate for FIG_ELM was $\mathbf{PR}'_{\mathbf{ELM}} =$ 89% at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{ELM}} = 89\%$ at $\alpha'' = 0.1$. The pass rate of FIG_GRU is $\mathbf{PR}'_{\mathbf{GRU}} = 33\%$ at a= $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{GRU}} = 44\%$ at $\alpha'' = 0.1$ .The pass rate of FIG_TCN is $\mathbf{PR}'_{\mathbf{TCN}} = 56\%$ at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{TCN}} = 67\%$ at $\alpha'' = 0.1$.The pass rate of FIG_DBN is $\mathbf{PR}'_{\mathbf{DBN}} = 100\%$ at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{DBN}} = 100\%$ at $\alpha'' = 0.1$. In summary, most of the models completely passed the DN test, and some of them failed the DM test due to the different data sets. Overall, the DM values of the single model after fuzzy particleization were all greater than 0 unlike FMICM, indicating that the point prediction of FMICM was better than that of the single model after fuzzy particleization, which verified the conclusion reached in Experiment 2.

(c) Compared with different optimization models, the DM test pass rate of the three optimization combination models in site1 is $\mathbf{PR}'' = 89\%$ when the significance level is $\alpha'' = 0.1$, and only the DM value of FIG_MOALO_CM is $\overline{\mathbf{PR}''}_{\mathbf{MOALO}} = 1.4498$. Most of the DM values in site2 are less than 1 and do not pass the test. Step2 in site3 all pass the significance level of $\alpha' = 0.05$ DM test, while the other step predictions did not pass the test. However, it seems that FMICM is significantly different from the three optimization models, and the DM values are all greater than 0. This indicates that the prediction effect of FMICM is better than the other three combined optimization models, which verifies the conclusion drawn in Experiment 3.

5.2 Improvement ratio of the indexes

After the DM test, it can be concluded that the proposed FMICM has significant differences with the single model, the single model after fuzzy particleization and different combined optimization models, in addition, based on the DM value greater than zero can be deduced that FMICM is better than other models.Therefore, the DM test can only qualitatively infer that FMICM is superior to other models, but quantitatively analyze it. Therefore, this section proposes to conduct the indicator improvement rate test with the purpose of further quantitatively indicating the superiority of FMICM based on the DM test to specifically improve MAPE is an important evaluation index to measure the prediction effect of time series data, so MAPE is used as the indicator improvement rate index in this paper.The calculation formula of indicator improvement rate is shown in Equation (5).

$$\mathbf{IR}_{\mathbf{MAPE}} = \left| \frac{Compared_{MAPE} - FMICM_{MAPE}}{Compared_{MAPE}} \right| \times 100\% \tag{5}$$

The point predictions of the developed integrated system were tested against a single model, a single model after fuzzy particleization, and different combined optimization models for metric improvement rates,and the final test results are shown in Table 11, the details are as follows:

(a) The proposed FMICM model was compared with the single model, where the most improved model was ELM with IR of $\vec{\mathbf{I}}_{\Delta} = [38.13\%, 33.49\%, 28.74\%]$ for the three prediction steps, and the least improved models were TCN and GRU with an average index improvement rate of $\mathbf{I}_{\Delta}^{(tcn)} = 18.3522\%$ for TCN and $\mathbf{I}_{\Delta}^{(gru)} = 18.6554\%$ for GRU, also side by side, it shows the high prediction accuracy of these two models. In general, the average index improvement rate of FMICM for a single model is around $\Delta\varpi = 25\%$, which is a large improvement. It indicates that FIG plays a role in improving the prediction accuracy in the point prediction of the system.

(b) Compared with the single model after fuzzy particleization, the average index improvement rates of FIG_DOABPNN, FIG_ELM, FIG_TCN, FIG_GRU, FIG_DBN is $\vec{\mathbf{I}}_{\Delta} = [17.85\%, 30.28\%, 8.85\%, 7.69\%, 24.69\%]$.Therefore, the highest improvement rate is FIG_ELM , with a three-step predicted average index improvement rate of $\vec{\mathbf{I}}_{\Delta}'^{(elm)} = [33.08\%, 34.17\%, 23.58\%]$.The lowest improvement rate is FIG_GRU with multi-step predicted average indicator improvement rates of $\vec{\mathbf{I}}_{\Delta}''^{(gru)} = [9.24\%, 5.61\%, 8.22\%]$. Also the average index improvement rate for FIG_DOABPNN is $\mathbf{I}_{\Delta}^{(bp)} = 17.85\%$, for FIG_TCN is $\mathbf{I}_{\Delta}^{(tcn)} = 8.85\%$, and for FIG_DBN is $\mathbf{I}_{\Delta}^{(dbn)} = 24.69\%$. Overall, the average index improvement rate of FMICM for the single model after fuzzy particleization is $\delta = 17.87\%$, which is a large improvement. This also reflects that MODOA can improve the prediction accuracy in the system.

(c) The index improvements relative to the FIG_MOGOA_CM, FIG_MODA_CM, and FIG_MOALO_CM are $\vec{\mathbf{I}}_{\Delta}^{*} = [4.35\%, 4.40\%, 4.99\%]$. It can be seen that the optimization capability of the MODOA algorithm has been improved to different degrees compared with the other three multi-objective optimization algorithms. In summary, the index improvement rate test shows that the accuracy of the proposed integrated system point prediction is significantly improved over the single model,which also reflects that the combined model can improve the prediction accuracy. The significant improvement in comparison with the unparticleized single model indicates that FIG is important for accuracy improvement. The significant improvements for different combined models indicate that MODOA is better than other optimization algorithms.

5.3 Forecasting Effectiveness

In addition to the accuracy of the forecast results, the size of the difference between the forecast results and the true values, the skewness and kurtosis of the distribution of the forecast results, should also be considered in point forecasting. Forecasting effectiveness is then an indicator to verify this. The calculation principle is as follows.

Define $\mathbf{W}_n = 1 - |\gamma_n|$ as the prediction accuracy, where

$$\gamma_n = \begin{cases} -1 & (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n < 1 \\ (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n & -1 \le (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n < 1 \\ 1 & (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n > 1 \end{cases} \tag{6}$$

Based on the prediction accuracy $\mathbf{W}_n$ can calculate the k-order prediction effective element, which is calculated as follows.

$$\psi^{(k)} = \sum_{n=1}^{N} \vartheta_n \gamma_n^{(k)}, \sum_{n=1}^{N} \vartheta_n = 1 \tag{7}$$

Here, ϑ_n denotes that the probability distribution at a point in time is discrete. Since we do not have access to prior information about the probability distribution, we identify it as 1 and set ϑ_n as $\vartheta_n = 1/\mathbf{N}, n = 1, 2, \dots, \mathbf{N}$, $\mathbf{C}$ is a continuous function of the k-order forecasting effectiveness component, $\mathbf{C}\left(\psi^{(1)}, \psi^{(2)}, \cdots, \psi^{(k)}\right)$ is defined as the k-order prediction effective.

This section uses the one-order prediction effective and the two-order prediction effective, the calculation of the one-order predictive validity is described in Equation (8).

$$\mathbf{C}\left(\psi^{(1)}\right) = \psi^{(1)} \tag{8}$$

There is a second-order predictive validity showing the disparity among the expected standard deviations, which is described in Equation (9).

$$\mathbf{C}\left(\psi^{(1)}, \psi^{(2)}\right) = \psi^{(1)}\left(1 - \sqrt{\psi^{(2)} - \left(\psi^{(1)}\right)^2}\right) \tag{9}$$

The proposed combined model was tested for predictive validity with the single model, the single model after fuzzy granulation and the different combined models, and the final test results are shown in Table 12, and the details of this experiment are as follows.

(a) For the one-order prediction effective, the best results were obtained for the newly proposed FMICM model, with the mean values of $\vec{\mathbf{F}}_{\mathbf{site1}}^{\langle 1 \rangle} = [95.98\%, 93.61\%, 93.91\%]$ for the three-step predictive FE of site1. For site2, the highest FE value was obtained for GRU in the one-order model, with an average one-order prediction effective of $\mathbf{F}_{\mathbf{site2}}^{gru\langle 1 \rangle} = 95.21\%$. Site3 had the highest FE value for FIG_TCN in the granulated one-order model, with an average one-order prediction effective of $\mathbf{F}_{\mathbf{site3}}^{tcn\langle 1 \rangle} = 93.38\%$. The other three combined models had one-order prediction effective of $\overrightarrow{\mathbf{Fo}}_{\mathbf{site3}}^{\langle 1 \rangle} = 94.34\%$, $94.44\%, 94.4\%$ on average.

(b) For the two-order prediction effective, the newly proposed FMICM model is still the best with the two-order values of $\vec{\mathbf{F}}_{\mathbf{site1}}^{\langle 2 \rangle} = [92.85\%, 88.70\%, 88.48\%]$, respectively, $\vec{\mathbf{F}}_{\mathbf{site2}}^{\langle 2 \rangle} = [93.92\%, 91.57\%, 90.77\%]$, $\vec{\mathbf{F}}_{\mathbf{site3}}^{\langle 2 \rangle} = [91.05\%, 89.14\%, 87.21\%]$ for the three sites,The model with the smallest two-order value was DBN with a second-order mean of $\mathbf{F}_{*}^{dbn\langle 2\rangle}=86.67\%$, the best performing single model after fuzzy granulation was FIG_TCN with a two-order mean of $\mathbf{F}_{*}^{tcn\langle 2\rangle}=89.04\%$, and the best performing

Table 10
Results of Diebold Mariano (DM) Test

	Site1			Site2			Site3		
	Step1	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**
DOABPNN	1.8285	3.0855	1.9492	1.8386	2.5803	2.3644	4.3238	4.1836	4.7186
ELM	4.9253	5.1667	3.0496	3.3037	2.8385	2.5105	5.5472	3.1028	3.7168
TCN	2.7372	1.0698	1.5163	2.6599	1.8013	1.9331	2.1202	3.5382	2.4108
GRU	2.7163	2.2543	1.4935	0.0700	0.8778	0.6341	2.3599	2.9007	2.6383
DBN	2.5315	4.8844	3.7328	2.6177	3.2896	3.9826	4.0637	3.0177	2.6855
FIG_DOABPNN	3.7172	4.0274	1.0853	1.4010	2.2854	1.1168	2.4077	7.6293	1.8578
FIG_ELM	5.3801	5.2446	4.3166	2.4546	3.4452	1.7648	6.3094	5.6874	4.6303
FIG_TCN	0.6869	2.4974	3.4223	2.9201	3.2966	2.5596	2.3299	3.2760	1.9782
FIG_GRU	3.4683	2.5899	2.6459	1.0499	0.0863	0.0829	0.3367	0.3543	1.7757
FIG_DBN	4.1482	3.5502	3.7996	3.7360	2.9129	2.3073	4.8371	6.7407	4.7167
FIG _MOGOA_CM	1.6671	2.7789	3.1029	0.6244	0.7456	1.1583	0.1472	3.3909	0.5151
FIG _MODA_CM	2.4670	2.7741	2.2995	0.5298	0.4578	0.2030	0.8126	3.4955	0.2044
FIG _MOALO_CM	1.4498	2.7618	3.1363	0.1884	0.3328	1.2941	0.5916	3.6622	0.2665

Note: The table shows the Diebold Mariano (DM) test results for all models in the experiment (single model, single model after granulation, different optimized combination models). The formula of its DM-test is $\mathbf{DM}=\sum_{\boldsymbol{\mu}=1}^{\boldsymbol{\Pi}}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er}}\hat{\mathbf{r}}_1^{(\boldsymbol{\mu})}\right)-\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er}}\hat{\mathbf{r}}_2^{(\boldsymbol{\mu})}\right)\right]\Big/\left(\boldsymbol{\Pi}\sqrt{S^2/\boldsymbol{\Pi}}\right)$.

Table 11
Results of Improvement ratio of the indexes

	Site1			Site2			Site3		
	Step1	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**
DOABPNN	28.31%	19.68%	32.47%	25.66%	25.74%	25.18%	30.66%	34.45%	24.99%
ELM	49.15%	38.49%	39.22%	28.56%	25.99%	25.64%	36.69%	35.99%	21.36%
TCN	20.30%	8.22%	23.79%	20.06%	17.95%	20.91%	14.31%	29.50%	10.12%
GRU	26.39%	13.49%	30.49%	11.93%	12.48%	12.66%	21.19%	25.38%	13.88%
DBN	40.20%	27.73%	33.11%	28.80%	29.11%	33.29%	29.62%	31.23%	12.10%
FIG_DOABPNN	19.13%	22.19%	8.12%	11.88%	21.53%	9.98%	17.50%	39.69%	10.59%
FIG_ELM	33.47%	29.91%	32.43%	21.31%	22.73%	15.28%	44.46%	49.88%	23.04%
FIG_TCN	2.62%	13.19%	14.74%	5.66%	13.15%	13.00%	5.04%	8.73%	3.49%
FIG_GRU	19.90%	9.20%	12.07%	1.82%	1.99%	3.40%	6.02%	5.64%	9.19%
FIG_DBN	16.26%	15.80%	19.87%	32.72%	22.67%	19.13%	36.19%	37.49%	22.10%
FIG _MOGOA_CM	5.91%	8.68%	8.26%	1.30%	1.82%	4.16%	3.10%	3.53%	2.38%
FIG _MODA_CM	7.00%	8.80%	2.93%	2.36%	1.33%	3.89%	4.92%	6.49%	1.87%
FIG _MOALO_CM	5.02%	8.72%	8.50%	1.86%	1.48%	5.57%	8.17%	3.94%	1.68%

Note: The table shows the Results of Improvement ratio of the indexes for all models (single model, single model after granulation, different optimized combination models) in the experiment. The test formula of its IR is $\mathbf{IR}_{\mathbf{MAPE}}=\left|\left(Compared_{MAPE}-FMICM_{MAPE}\right)/Compared_{MAPE}\right|\times 100\%$.

Table 12

Results of Forecasting Effectiveness

	Site1						Site2						Site3					
	Step1		**Step2**		**Step3**		**Step1**		**Step2**		**Step3**		**Step1**		**Step2**		**Step3**	
	OD1	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**	**OD1**	**OD2**
DOABPNN	94.40	88.94	92.05	86.21	90.99	84.42	95.83	92.78	93.85	89.67	93.44	88.66	93.35	87.25	90.66	82.77	89.35	82.07
ELM	92.10	86.75	89.62	82.36	89.99	83.20	95.66	92.48	93.83	89.72	93.40	88.36	92.71	86.94	90.44	83.69	89.84	82.94
TCN	94.96	90.16	93.04	87.67	92.01	84.49	96.12	92.52	94.43	89.72	93.79	88.31	94.62	89.68	91.32	83.63	91.11	84.69
GRU	94.54	89.97	92.62	87.12	91.24	83.46	96.48	93.45	94.78	90.94	94.38	89.60	94.15	89.32	91.80	84.91	90.72	83.53
DBN	93.28	85.57	91.17	84.88	90.90	82.86	95.65	92.48	93.56	89.02	92.64	86.97	93.45	88.58	91.10	84.90	90.91	84.80
MODOA_CM	95.30	90.75	93.49	88.39	92.18	84.96	96.56	93.49	95.18	91.22	94.58	90.17	94.87	90.50	92.49	86.10	91.61	86.31
FIG_DOABPNN	95.03	90.91	91.79	85.06	93.38	87.84	96.48	93.39	94.18	89.29	94.55	90.01	94.41	89.53	89.85	83.12	91.06	84.88
FIG_ELM	93.96	89.52	90.89	83.83	90.99	84.00	96.06	92.74	94.09	89.52	94.20	89.45	91.69	86.18	87.79	79.25	89.62	82.31
FIG_TCN	95.87	92.50	92.64	86.69	92.86	86.13	96.72	93.36	94.74	89.67	94.36	88.76	95.14	90.64	93.29	88.06	91.72	85.63
FIG_GRU	94.98	90.84	92.97	86.47	93.08	85.18	96.84	93.60	95.34	91.36	94.92	90.17	95.09	90.83	93.51	88.57	91.20	84.03
FIG_DBN	95.20	91.36	92.42	86.30	92.40	85.47	95.39	92.06	94.09	89.52	93.93	89.19	92.77	86.36	90.21	83.10	89.74	82.62
FIG _MOGOA_CM	95.73	92.39	93.01	87.16	93.36	86.52	96.86	93.90	95.35	91.24	94.88	90.42	95.24	91.01	93.66	88.64	91.81	86.76
FIG _MODA_CM	95.68	92.27	93.00	87.09	93.73	87.40	96.83	93.88	95.37	91.34	94.89	90.57	95.15	90.70	93.46	88.36	91.86	86.70
FIG _MOALO_CM	95.77	92.45	93.00	87.13	93.35	86.50	96.84	93.89	95.36	91.35	94.80	90.31	94.98	90.37	93.63	88.61	91.87	86.08
Proposed System	**95.98**	**92.85**	**93.61**	**88.70**	**93.91**	**88.48**	**96.90**	**93.92**	**95.43**	**91.57**	**95.09**	**90.77**	**95.39**	**91.05**	**93.88**	**89.14**	**92.01**	**87.21**

Note: The table shows the Results of Forecasting Effectiveness for all models (single model, single model after granulation, different optimized combination models and FMICM) in the experiment. The test formula of its FE is $\mathbf{C}\left(h^{(1)}\right)=\psi^{(1)}$ and $\mathbf{C}\left(\psi^{(1)},\psi^{(2)}\right)=\psi^{(1)}\left(1-\sqrt{\psi^{(2)}-\left(\psi^{(1)}\right)^{2}}\right)$.

combined model was FIG_MODA_CM with a two-order mean of $\mathbf{F}_{*}^{\mathrm{mod}\,a\langle 2\rangle}=89.81\%$. Through the forecasting effectiveness test, it can be concluded that the newly proposed FMICM model performs best in terms of point predictive validity, which means that the point prediction results of FMICM are not only accurate and stable, but also valid, they are closer to the true values in terms of the skewness and kurtosis distribution of the prediction results.

5.4 Sensitivity analysis

To verify the stability of the proposed prediction system, this section sets up the sensitivity analysis of MODOA in the proposed FMICM , and experiments are performed on three datasets with three steps of prediction. For MODOA, the parameters set are Search Number of Individuals S_A , Maximum iterations Number M_{iter} and ArchiveMaxSize A_m. We analyze the stability of the proposed prediction system with respect to changes in parameter values by varying one of the parameters by the control variables method, given that the other two parameters remain unchanged. The sensitivity index $\mathbf{SI}=\sum_{\mathbf{f}=1}^{\mathbf{K}}\sum_{\varsigma=1}^{\mathbf{P}}\left(\mathbf{E}_{\varsigma}^{\langle\mathbf{f}\rangle}-\bar{\mathbf{E}}\right)^{2}\Big/\mathbf{K}\cdot\mathbf{P}$ used, where $\mathbf{P}$ is the number of trials, $\mathbf{K}$ is the number of parameter changes, $\mathbf{E}_{\varsigma}^{\langle\mathbf{f}\rangle}$ is the point prediction evaluation index value MAPE for each trial, and $\bar{\mathbf{E}}$ is the average of the point prediction evaluation index MAPE for all trials.The specific sensitivity analysis data are shown in Table 13, and the details of this experiment are as follows.

It is obvious from the results that all three datasets show the lowest sensitivity of Maximum iterations Number, which means that M_{iter} has the least influence on the prediction results. The sensitivities of the other three parameters are less than 1 in 89% of the data, which means that the values of the three parameters have a low degree of influence on the prediction results, and thus our proposed model is relatively stable.

Table 13

Sensitivity analysis of the proposed model

	Adjusted parameters	Step1	Step2	Step3
Site1	S_A	1.3597	0.6031	0.1811
	M_{iter}	**0.7710**	**0.4339**	**0.0988**
	A_m	1.5155	0.9268	0.4853
Site2	S_A	0.0844	0.3800	0.4030
	M_{iter}	**0.0370**	**0.3544**	**0.1945**
	A_m	0.1168	0.4189	0.2039
Site3	S_A	0.0928	0.6605	0.0415
	M_{iter}	**0.0425**	**0.6233**	**0.0276**
	A_m	0.1789	1.1067	0.0765

Note: In the sensitivity analysis calculation, the Search Number of Individuals was taken as $\overrightarrow{S_A}=[60,80,100,120,140]$, the Maximum iterations Number was taken as $\overrightarrow{M_{iter}}=[100,200,300,400,500]$, and the ArchiveMaxSize was taken as $\overrightarrow{A_m}=[200,300$

$,400,500,600]$. Five experiments were conducted in each round, i.e., $\mathbf{P}=5$.

5.5 Convergence analysis

Stability can be demonstrated after sensitivity analysis of MODOA, and in addition, the convergence of MODOA needs to be verified, and measuring the convergence process of MODOA can verify its computational efficiency. Figure 6 shows the corresponding convergence analysis process for the three data sets, from which it can be seen that MODOA has a high convergence speed and it can come to convergence in fewer iterations, which further proves the feasibility of its prediction system.

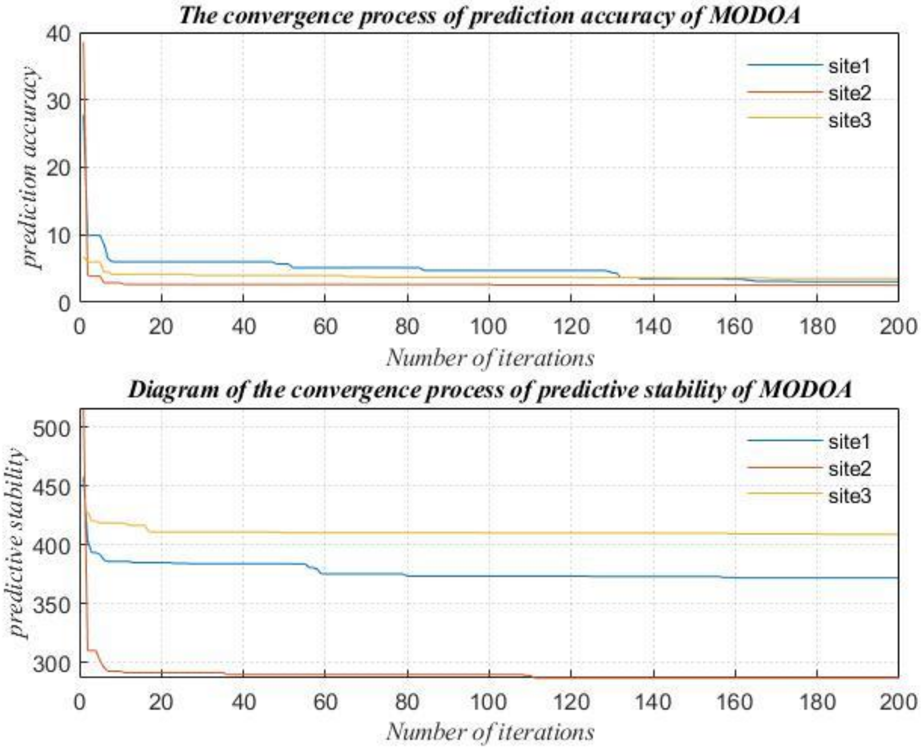

Figure 6 The convergence process of MODOA is shown

5.6 Empirical analysis

Through these checks and tests, the proposed integrated forecasting system is found to have better forecasting accuracy, stability, and effectiveness than the other 14 models. It is able to handle time series data with characteristics of randomness, volatility, periodicity, and diversity, which are affected by various factors such as power load.

(1) Accurate power load forecasting is the most effective way to ensure stable power supply and power quality. When the power generation is insufficient, the output power of generating units can be increased or deployed from other power grids; conversely, if there is excess power generation, the generating units should be shut down or deployed to other power grids, so that the power generation and power consumption can reach a certain dynamic balance. Accurate power load forecasting can help the power sector make timely scientific decisions, reduce costs and ensure the long-term safe and stable operation of the power grid.

(2) Accurate load forecasting can economically and reasonably arrange the start and stop of generating units in the power grid, maintain the safety and stability of the

power grid operation, reduce unnecessary rotation of spare capacity, reasonably arrange the unit maintenance schedule, guarantee the normal production and life of the society, effectively reduce the cost of power generation and improve economic and social benefits. The load forecasting results derived from the combined algorithm are transmitted to the power sector, which facilitates the decision on the future installation of new generating units, the size, location and timing of the installed capacity, the capacity increase and renovation of the power grid, and the construction and development of the power grid.

(3) Since the proposed point-interval prediction system can perform deterministic prediction analysis and volatility prediction analysis on time series data with randomness, volatility, periodicity and diversity characteristics, and the proposed system has high prediction stability, the proposed point-interval prediction system can be extended to other prediction problems with time series nonlinear characteristics, such as wind speed prediction, air pollution prediction and traffic flow prediction.

6.Conclusion

In this era of rapid growth of electricity demand in the whole society, accurate forecasting of power load becomes more and more important to ensure stable power supply as well as power quality. However, the change of electric load is the result of multiple factors, which have complex interconnection, and the load data has strong randomness. Therefore, this paper proposes a novel integrated power load point-interval forecasting system that constructs information grains by building fuzzy sets on subseries formed by discretized time series, which in turn compresses the scale of time series data, simplifies the computational complexity, and effectively improves the accuracy of short-term forecasting; secondly, the MODOA algorithm is used to optimize the five benchmark models in multiple stages to obtain the final point forecasting results, and the fluctuation analysis is performed on the point forecasting results to obtain the uncertain interval forecasting results. The proposed FMICM improves the accuracy and stability of power load data forecasting and expands the application scope of the model.

(1) For point forecasting, FMICM was compared with 14 models in three experiments.FMICM outperformed the single model without fuzzy particleization with the mean MAPE values of $\overrightarrow{\mathbf{MAPE}'}_{\mathbf{M}} = [3.9101\%, 5.6910\%, 6.3293\%]$ for the three-step forecasting.FMICM outperformed all the five single models used for the combination, and compared with FIG_DOABPNN, FIG_ELM, FIG_TCN, FIG_GRU, and FIG_DBN, the average values of MAPE are improved by $\overrightarrow{\mathbf{I}}_{\mathbf{M}}^{(2)} = [1.2752\%, 2.5457\%, 0.5412\%, 0.4748\%, 1.7832\%]$, respectively. The multi-objective dinger optimization algorithm in the FMICM model outperforms the known MOGOA, MODA, and MOALO in terms of weight optimization capability. (2) In terms of interval prediction. The FMICM was compared with eight models. With a confidence factor of 95%, 5/9 experiments showed that FMICM had the best AIS and MPICD, and two additional sets of experiments showed that FIG_GRU and FIG_DOABPNN had a smaller AIS than FMICM. 89% of experiments proved that FMICM had a higher interval prediction accuracy with a confidence factor of 90%, and additional individual experiments showed that FIG_MOGOA_CM and FIG_MODA_CM have better MPICD.

The proposed integrated power load point-interval forecasting system is not only accurate but also effective, which broadens the field of power load forecasting. However, there are still some aspects that need to be improved: (1) Weather conditions such as temperature and humidity can be considered. (2) The peak prediction is added to improve the prediction accuracy.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 71671029)

References

[1] Malekizadeh, M., Karami, H., Karimi, M., Moshari, A., & Sanjari, M. J. (2020). Short-term load forecast using ensemble neuro-fuzzy model. *Energy*, *196*, 117127. https://doi.org/10.1016/j.energy.2020.117127
[2] Fortes, P., Simoes, S. G., Amorim, F., Siggini, G., Sessa, V., Saint-Drenan, Y.-M., Carvalho, S., Mujtaba, B., Diogo, P., & Assoumou, E. (2022). How sensitive is a carbon-neutral power sector to climate change? The interplay between hydro, solar and wind for Portugal. *Energy*, *239*, 122106. https://doi.org/10.1016/j.energy.2021.122106
[3]Chen, M.-R., Zeng, G.-Q., & Lu, K.-D. (2019). Constrained multi-objective population extremal optimization based economic-emission dispatch incorporating renewable energy resources. *Renewable Energy*, *143*, 277–294. https://doi.org/10.1016/j.renene.2019.05.024
[4] Yang, Y., Che, J., Li, Y., Zhao, Y., & Zhu, S. (2016). An incremental electric load forecasting model based on support vector regression. *Energy*, *113*, 796–808. https://doi.org/10.1016/j.energy.2016.07.092
[5] Dudek, G., & Pełka, P. (2021). Pattern similarity-based machine learning methods for mid-term load forecasting: A comparative study. *Applied Soft Computing*, *104*, 107223. https://doi.org/10.1016/j.asoc.2021.107223
[6] Kazemzadeh, M.-R., Amjadian, A., & Amraee, T. (2020). A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. *Energy*, *204*, 117948. https://doi.org/10.1016/j.energy.2020.117948
[7] Li, S., Wang, P., & Goel, L. (2016). A Novel Wavelet-Based Ensemble Method for Short-Term Load Forecasting with Hybrid Neural Networks and Feature Selection. *IEEE Transactions on Power Systems*, *31*(3), 1788–1798. https://doi.org/10.1109/TPWRS.2015.2438322
[8] Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., & Glazer, J. (2001). EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings*, *33*(4), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6
[9] Yun, K., Cho, H., Luck, R., & Mago, P. J. (2011). Real-time combined heat and power operational strategy using a hierarchical optimization algorithm. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, *225*(4), 403–412. https://doi.org/10.1177/2041296710394287
[10] Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy*, *34*(9), 1413–1421. https://doi.org/10.1016/j.energy.2009.06.034
[11] Pappas, S. Sp., Ekonomou, L., Karampelas, P., Karamousantas, D. C., Katsikas, S. K., Chatzarakis, G. E., & Skafidas, P. D. (2010). Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electric Power Systems Research*, *80*(3), 256–264. https://doi.org/10.1016/j.epsr.2009.09.0

06
[12] Jamil, R. (2020). Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. *Renewable Energy*, *154*, 1–10. https://doi.org/10.1016/j.renene.2020.02.117
[13] Pombeiro, H., Santos, R., Carreira, P., Silva, C., & Sousa, J. M. C. (2017). Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. *Energy and Buildings*, *146*, 141–151. https://doi.org/10.1016/j.enbuild.2017.04.032
[14] Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166. https://doi.org/10.1016/j.eneco.2013.02.006
[15] Ervural, B. C., Beyca, O. F., & Zaim, S. (2016). Model Estimation of ARMA Using Genetic Algorithms: A Case Study of Forecasting Natural Gas Consumption. *Procedia - Social and Behavioral Sciences*, *235*, 537–545. https://doi.org/10.1016/j.sbspro.2016.11.066
[16] Sharma, S., Majumdar, A., Elvira, V., & Chouzenoux, E. (2020). Blind Kalman Filtering for Short-Term Load Forecasting. *IEEE Transactions on Power Systems*, *35*(6), 4916–4919. https://doi.org/10.1109/TPWRS.2020.3018623
[17] Shine, P., Scully, T., Upton, J., & Murphy, M. D. (2019). Annual electricity consumption prediction and future expansion analysis on dairy farms using a support vector machine. *Applied Energy*, *250*, 1110–1119. https://doi.org/10.1016/j.apenergy.2019.05.103
[18] Chen, S., Ren, Y., Friedrich, D., Yu, Z., & Yu, J. (2021). Prediction of office building electricity demand using artificial neural network by splitting the time horizon for different occupancy rates. *Energy and AI*, *5*, 100093. https://doi.org/10.1016/j.egyai.2021.100093
[19] Khosravi, A., Nahavandi, S., Creighton, D., & Srinivasan, D. (2012). Interval Type-2 Fuzzy Logic Systems for Load Forecasting: A Comparative Study. *IEEE Transactions on Power Systems*, *27*(3), 1274–1282. https://doi.org/10.1109/TPWRS.2011.2181981
[20] Lv, M., Wang, J., Niu, X., & Lu, H. (2022). A newly combination model based on data denoising strategy and advanced optimization algorithm for short-term wind speed prediction. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-021-03595-x
[21] Chen, M.-R., Zeng, G.-Q., Lu, K.-D., & Weng, J. (2019). A Two-Layer Nonlinear Combination Method for Short-Term Wind Speed Prediction Based on ELM, ENN, and LSTM. *IEEE Internet of Things Journal*, *6*(4), 6997–7010. https://doi.org/10.1109/JIOT.2019.2913176
[22] Zhao, F., Zeng, G. Q., & Lu, K. di. (2020). EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization. *IEEE Transactions on Vehicular Technology*, *69*(1), 101–113. https://doi.org/10.1109/TVT.2019.2952605
[23] Barman, M., & Dev Choudhury, N. B. (2020). A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. *Sustainable Cities and Society*, *61*, 102311. https://doi.org/10.1016/j.scs.2020.102311
[24] Liang, Y., Niu, D., & Hong, W.-C. (2019). Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy*, *166*, 653–663. https://doi.org/10.1016/j.energy.2018.10.119
[25] Chen, Y., Kloft, M., Yang, Y., Li, C., & Li, L. (2018). Mixed kernel bas

ed extreme learning machine for electric load forecasting. *Neurocomputing*, *312*, 90–106. https://doi.org/10.1016/j.neucom.2018.05.068

[26] Xie, K., Yi, H., Hu, G., Li, L., & Fan, Z. (2020). Short-term power load forecasting based on Elman neural network with particle swarm optimization. *Neurocomputing*, *416*, 136–142. https://doi.org/10.1016/j.neucom.2019.02.063

[27] López, C., Zhong, W., & Zheng, M. (2017). Short-term Electric Load Forecasting Based on Wavelet Neural Network, Particle Swarm Optimization and Ensemble Empirical Mode Decomposition. *Energy Procedia*, *105*, 3677–3682. https://doi.org/10.1016/j.egypro.2017.03.847

[28] Hu, Y., Li, J., Hong, M., Ren, J., Lin, R., Liu, Y., Liu, M., & Man, Y. (2019). Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—A case study of papermaking process. *Energy*, *170*, 1215–1227. https://doi.org/10.1016/j.energy.2018.12.208

[29] Memarzadeh, G., & Keynia, F. (2021). Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electric Power Systems Research*, *192*, 106995. https://doi.org/10.1016/j.epsr.2020.106995

[30] Li, X., Ma, X., Xiao, F., Xiao, C., Wang, F., & Zhang, S. (2022). Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). *Journal of Petroleum Science and Engineering*, *208*, 109309. https://doi.org/10.1016/j.petrol.2021.109309

[31] Zhu, R., Liao, W., & Wang, Y. (2020). Short-term prediction for wind power based on temporal convolutional network. *Energy Reports*, *6*, 424–429. https://doi.org/10.1016/j.egyr.2020.11.219

[32] Bendaoud, N. M. M., Farah, N., & ben Ahmed, S. (2021). Comparing Generative Adversarial Networks architectures for electricity demand forecasting. *Energy and Buildings*, *247*, 111152. https://doi.org/10.1016/j.enbuild.2021.111152

[33] Wang, J., Niu, X., Zhang, L., & Lv, M. (2021). Point and interval prediction for non-ferrous metals based on a hybrid prediction framework. *Resources Policy*, *73*, 102222. https://doi.org/10.1016/j.resourpol.2021.102222

[34] He, Y., & Wang, Y. (2021). Short-term wind power prediction based on EEMD–LASSO–QRNN model. *Applied Soft Computing*, *105*, 107288. https://doi.org/10.1016/j.asoc.2021.107288

[35] Niu, X., & Wang, J. (2019). A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. *Applied Energy*, *241*, 519–539. https://doi.org/10.1016/j.apenergy.2019.03.097

[36] Peng, S., Chen, R., Yu, B., Xiang, M., Lin, X., & Liu, E. (2021). Daily natural gas load forecasting based on the combination of long short term memory, local mean decomposition, and wavelet threshold denoising algorithm. *Journal of Natural Gas Science and Engineering*, *95*, 104175. https://doi.org/10.1016/j.jngse.2021.104175

[37] Yang, Q., Deng, C., & Chang, X. (2022). Ultra-short-term / short-term wind speed prediction based on improved singular spectrum analysis. *Renewable Energy*, *184*, 36–44. https://doi.org/10.1016/j.renene.2021.11.044

[38] Qu, C., Zhou, Z., Liu, Z., Jia, S., Wang, L., & Ma, L. (2021). State prediction for marine diesel engine based on variational modal decomposition and long short-term memory. *Energy Reports*, *7*, 880–886. https://doi.org/10.1016/j.egyr.2021.09.185

[39] Ding, S., Zhang, X., An, Y., & Xue, Y. (2017). Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification. *Pattern Recognition*, *67,* 32–46. https://doi.org/10.1016/j.patcog.2017.02.011

[40] Velázquez-Rodríguez, J. L., Villuendas-Rey, Y., Yáñez-Márquez, C., López-Yáñez, I., & Camacho-Nieto, O. (2020). Granulation in Rough Set Theory: A novel perspective. *International Journal of Approximate Reasoning*, *124*, 27–39. https://doi.org/10.1016/j.ijar.2020.05.003

[41] Zeng, G.-Q., Chen, J., Li, L.-M., Chen, M.-R., Wu, L., Dai, Y.-X., & Zheng, C.-W. (2016). An improved multi-objective population-based extremal optimization algorithm with polynomial mutation. *Information Sciences*, *330*, 49–73. https://doi.org/10.1016/j.ins.2015.10.010

[42] Zeng, G.-Q., Chen, J., Dai, Y.-X., Li, L.-M., Zheng, C.-W., & Chen, M.-R. (2015). Design of fractional order PID controller for automatic regulator voltage system based on multi-objective extremal optimization. *Neurocomputing*, *160*, 173–184. https://doi.org/10.1016/j.neucom.2015.02.051

[43] Yang, D., Guo, J., Sun, S., Han, J., & Wang, S. (2022). An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. *Applied Energy*, *306*, 117992. https://doi.org/10.1016/j.apenergy.2021.117992

[44] Bo, H., Nie, Y., & Wang, J. (2020). Electric Load Forecasting Use a Novelty Hybrid Model on the Basic of Data Preprocessing Technique and Multi-Objective Optimization Algorithm. *IEEE Access*, *8*, 13858–13874. https://doi.org/10.1109/ACCESS.2020.2966641

[45] Wang, J., Zhang, L., & Li, Z. (2022). Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm. *Applied Energy*, *305*, 117911. https://doi.org/10.1016/j.apenergy.2021.117911

[46] Hecht-Nielsen. (1989). Theory of the backpropagation neural network. *International Joint Conference on Neural Networks*, 593–605 vol.1. https://doi.org/10.1109/IJCNN.1989.118638

[47] Peraza-Vázquez, H., Peña-Delgado, A. F., Echavarría-Castillo, G., Morales-Cepeda, A. B., Velasco-Álvarez, J., & Ruiz-Perez, F. (2021). A Bio-Inspired Method for Engineering Design Optimization Inspired by Dingoes Hunting Strategies. *Mathematical Problems in Engineering*, *2021*, 1–19. https://doi.org/10.1155/2021/9107547

[48] Pedrycz, W., & Vukovich, G. (2002). Feature analysis through information granulation and fuzzy sets. *Pattern Recognition*, *35*(4), 825–834. https://doi.org/10.1016/S0031-3203(01)00102-9

[49] Mencar, C., & Fanelli, A. M. (2008). Interpretability constraints for fuzzy information granulation. *Information Sciences*, *178*(24), 4585–4618. https://doi.org/10.1016/j.ins.2008.08.015

[50] Li, Y., Tong, Z., Tong, S., & Westerdahl, D. (2022). A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation. *Sustainable Cities and Society*, *76*, 103481. https://doi.org/10.1016/j.scs.2021.103481

[51] Oh, S.-K., Pedrycz, W., & Roh, S.-B. (2009). Hybrid fuzzy set-based polynomial neural networks and their development with the aid of genetic optimization and information granulation. *Applied Soft Computing*, *9*(3), 1068–1089. https://doi.org/10.1016/j.asoc.2009.02.00

List of nomenclature

Abbreviations	
AIS	Average interval score
ARIMA	Auto-regressive integrated moving average model
ARMA	Auto-regressive moving average model
BPNN	Back Propagation Neural Network
CEEMDAN	Complete ensemble empirical mode decomposition
CHP	Combined heat and power
CM	Combine Model
DBN	Deep belief network
DM	Diebold-Mariano test
DOA	Dingo Optimization Algorithm
ELM	Extreme learning machine
ENN	Evolutionary neural networks
FIG	Fuzzy information granulation
FMICM	The proposed combined prediction system
GRU	Gated recurrent unit
IKDE	Improved kernel density estimation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MOALO	Multi-objective Ant Lion Optimizer
MODA	Multi-objective Dragonfly Algorithm
MODOA	Multi-objective Dingo Optimization Algorithm
MOGOA	Multi-Objective Grasshopper Optimization Algorithm
MPICD	Mean prediction interval centre deviation
MSE	Mean Square Error
PF	Point forecast
PI	Prediction interval
PSO	Particle swarm optimization
RMSE	Root Mean Square Error
SDE	The standard deviation of error
SSA	Singular spectrum analysis
TCN	Time convolutional neural network

Symbols	
$\mathbf{ERR}_i^1$ $\mathbf{ERR}_i^2$	Deviation of forecast value and actual value
$\mathbf{FIL}_i^\alpha$	The lower limit of the forecast interval with a confidence level of $1-\alpha$
$\mathbf{FIU}_i^\alpha$	The upper limit of the forecast interval with a confidence level of $1-\alpha$
$\mathbf{MPFV}_i$	The final predicted value of FMICM
$\mathbf{PFV}_i$	Model point forecast value
$\mathbf{TOV}_i$	True observation value
$\mathbf{SW}_i$	Single model weights

Title: An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization

Highlights:

1: A new integrated power load point-interval forecasting system is developed

2:A novel multi-objective optimization algorithm is proposed for multi-level optimization

3:The proposed forecasting system limits the forecast uncertainty.

4:To test the proposed model on Australian electricity load data

An integrated power load point-interval forecasting system based on information entropy and multi-objective optimization

Kang Wang[a],Jianzhou Wang[a,*],Bo Zeng[b], Haiyan Lu[c]

[a] *School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*
[b] *School of Management Science and Engineering, Chongqing Technology and Business University, Chongqing, 400067, China*
[c] *School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia*

** Corresponding author. Address: School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*
E-mail address: wjz@lzu.edu.cn

Abstract

During an era of rapid growth in electricity demand throughout society, accurate forecasting of electricity loads has become increasingly important to guarantee a stable power supply. Nevertheless, historical models do not address the structure of the data itself, and a single model cannot accurately determine the nonlinear characteristics of the data. This would not allow for accurate and stable predictions. With the aim of filling this gap, this paper proposes an innovative intelligent power load point-interval forecasting system. The system discretizes the time series, then performs efficient dimensionality reduction by fuzzification, and multi-level optimization of five benchmark deep learning models by the proposed multi-objective optimization algorithm, and finally analyzes the uncertainty of the prediction results. Experiments comparing the developed prediction system with other models were conducted on three datasets, and the prediction results were discussed for validation from multiple perspectives. The simulation results show that the proposed model has superior prediction accuracy, robustness and uncertainty analysis capability, and can provide accurate deterministic prediction information and fluctuation interval analysis to ensure the long-term safety and stability and operation of the grid.

Keywords: *Electricity load forecast;Fuzzy information particles;Combination optimization strategy; Point-interval prediction system;*

1.Introduction

Electricity is the linchpin of the energy system to achieve carbon neutrality, and effectuating the "bi-carbon" goal and implementing a novel electricity system is a tremendously challenging and pioneering strategic and systemic project. Only by embedding a more flexible and interconnected power system can we achieve global electrification when the conditions are right[1]. The future carbon-neutral world will be highly dependent on electricity for energy supply, and electricity will become the pillar of the entire energy system and help society achieve sustainable development. Thereupon, with the development of technology and society, electric power resources become an increasingly important part of human production and life[2].

However, as people's electricity consumption continues to increase and the price of raw materials rises, considerable countries from all around the world are experiencing a shortage of electricity resources[3]. For the sake of avoiding the shortage of electricity resources triggered by short term surges of electricity consumption and unnecessary load loss and investment decisions, short-term electricity load forecasting has become an indispensable part of the national electricity and energy system[4].In summary, accurate power forecasting helps ensure the utilization of electricity, which is critical to the availability and sustainability of the distribution. On the contrary, the lack of accurate forecasting may lead to poor decision making and result in significant losses to the power system[5]. Load forecasting is divided into short-term forecasting for real-time control, medium-term forecasting for energy system operation, and long-term forecasting for extended planning studies. For example, long-term power load forecasts such as predicting annual peak loads for the next few years are used to optimize expansion decisions, while short-term load forecasts(**STLF**) are used for economic dispatch or unit mix studies, such as forecasting load conditions for the next few hours[6]. In order to obtain effective forecasting results, electric load forecasting has been studied intensively. We can broadly classify these forecasting methods into four categories: physical models, conventional statistical models, artificial intelligence models, and hybrid models[7].

The main physical models are the new-generation building energy simulation program (**EnergyPlus**) [8], real-time combined heat and power operational strategy using a hierarchical optimization algorithm[9]. Building operation data are obtained through EnergyPlus and mathematical models related to the physical system are represented. Real-time combined heat and power operational strategy using a hierarchical optimization algorithm considers the transient response of the building and combines the hierarchical CHP optimal control algorithm to achieve a real-time integrated system of electrical load information by running parallel simulations of two transient building models. Nevertheless, as a result of using simulation tools, the physics-based approach is usually difficult to obtain mathematical expressions for various building energy mechanisms and is not effective for short-term predictions.

Conventional statistical models can be used for load forecasting and speculation based on the available and relatively complete historical statistics, which are mechanically processed and organized using certain mathematical methods to reveal the regular links between the variables concerned. Statistical models mainly include ordinary regression models[10], auto-regressive moving average model(**ARMA**)[11] and Auto-regressive integrated moving average model(**ARIMA**)[12]. Since electricity load data have multiple non-linear components, conventional linear regression model treatments either become inaccurate or too complex to be used in practice. Most of the papers are comparing linear regression models with new models to show the advantages of the new model. Pombeiro et al.[13] proposed a nonlinear model based on fuzzy systems and neural networks, which compared with the linear model yielded a much higher prediction accuracy of the new model. Liu et al.[14] developed an autoregressive moving average model by combining it with a generalized auto-decreasing conditional heterogeneity process, and Cayir Ervurald et al.[15] proposed an integrated genetic algorithm(**GA**) and autoregressive moving average(**ARMA**) method for forecasting, obtaining lower error percentages than ARMA. Sharma et al.[16] used a blind Kalman filter algorithm and an autoregressive integrated moving average model to solve the problem of short-term load forecasting. However, because machine learning time series models have fewer parameters and better computational efficiency, artificial intelligence models have better forecasting accuracy than conventional statistical models in most cases.

Incidentally, with the rapid development and widespread use of artificial intelligence algorithms, many researchers have effectively used artificial intelligence methods to predict electric loads. These methods include support vector machines

(**SVM**)[17], artificial neural networks (**ANN**)[18], fuzzy logic models[19], and deep learning models[20,21,22]. Barman et al.[23] proposed the **GWO-SVM** model based on support vector machines (SVM) with gray wolf optimizer (**GWO**) to predict load data, which eventually achieved higher accuracy. Liang et al.[24] proposed general regression neural network (**GRNN**) combined with fruit fly optimization algorithm (**FOA**) for short-term load forecasting. Chen et al.[25] propose a kind of fresh short-term electric load forecasting method **EMD-Mixed-ELM** based on empirical mode decomposition (**EMD**) and extreme learning machine (**ELM**), which obtained higher forecasting accuracy. Xie et al.[26] proposed a **PSO-ENN** model combining **ENN** and particle swarm optimization, which improved the load forecasting accuracy of ENN. López et al.[27] proposed a new hybrid method of **STLF** based on symbiotic empirical mode decomposition (**EEMD**), beam neural network (**WNN**) and particle swarm optimization (**PSO**), and their results verified the higher accuracy of the proposed model. Hu et al.[28] proposed a short-term electricity load prediction model based on a hybrid **GA-PSO-BPNN** algorithm, which improved the prediction accuracy of **BPNN**. Memarzadeh et al.[29] proposed Short-term electricity load by a new optimal **LSTM-NN** based prediction algorithm, which improves the prediction accuracy.Li et al.[30] proposed a novel framework to improve the prediction accuracy using bi-directional gated recurrent unit (**Bi-GRU**) and sparrow search algorithm (**SSA**). Zhu et al.[31] used time convolutional neural network (**TCN**) to predict time series data and obtained higher prediction than the existing single predictor accuracy. Mehdi Bendaoud et al.[32] used Generative Adversarial Network (**GAN**) to introduce STLF and proposed a conditional Generative Adversarial Network (**cGAN**) architecture to improve the prediction accuracy. Artificial intelligence algorithms generally outperform time series models because of the strong nonlinear predictive capability of artificial intelligence models.

With further research, it has been found that noise in the electric load data affects the final prediction, which is why data preprocessing techniques such as empirical mode decomposition (**EMD**)[33], ensemble empirical mode decomposition (**EEMD**)[34], complete ensemble empirical mode decomposition (**CEEMDAN**)[35], wavelet threshold denoising[36], singular spectrum analysis (**SSA**)[37], and variational modal decomposition (**VMD**)[38]. In addition to using data denoising techniques, there are other data preprocessing methods, such as Shifei Ding et al.[39]proposed a weighted linear support vector machine (GWLMBSVM) based on information granulation, which uses information granulation to divide the data into several particles and classify the particles for prediction. José LuisVelázquez-Rodríguez et al.[40]propose a parametric granulation of particles in rough set theory that can effectively deal with the study of hybrid information systems. In recent years, it has been noticed that load forecasting should focus not only on accuracy but also on the stability of forecasting, so multi-objective optimization algorithms have been proposed[41,42]: Yang et al.[43] proposed a new STLF combining data denoising and prediction model based on bivariate empirical mode decomposition (**BEMD**), multivariate multiscale reciprocal entropy (**MMPE**) and tree structure parzen estimation (**TPE**) algorithms to optimize **LSTM**. Bo et al.[44] used singular spectrum analysis (**SSA**) for data preprocessing, and then proposed a multi-objective evolutionary algorithm based on genetic algorithm to discuss decomposition in detail (**MOEA/D**). Wang et al.[45] used a data decomposition strategy to process the raw data and then combined the single model by multi-objective locust algorithm (**MOGOA**) to greatly improve the prediction of power load forecasting accuracy.

Evaluation of the previous literature shows that the aforementioned prediction methods have some inherent drawbacks. Table 1 shows the advantages and disadvantages of the above model.

The drawbacks of these methods are summarized as follows:

(1) As simulation instruments are employed, it is often physically difficult to obtain mathematical expressions for the various building energy mechanisms, and they are not effective for short-term predictions.

(2) Conventional statistical models are more suitable for linear data. For electric load data with high noise and non-linear factors, conventional linear regression model processing either becomes inaccurate or too complex to be used in practice.

(3) Although artificial intelligence models are applicable to nonlinear data and reduce prediction accuracy, they are relatively data-dependent, easily fall into local optimum, and have long running time owing to slow convergence speed.

(4) The data denoising technique in the hybrid model ignores the importance of information leakage from the denoising method, leading to the optimization of abnormal prediction accuracy. Meanwhile, the existing multi-objective optimization algorithms are not strong in optimizing the balance between prediction accuracy and prediction stability and have a long running time.

Based on the above literature analysis, in this paper, we first propose to optimize the weights and thresholds of the back propagation neural network (**BPNN**) [46] using an iterative update strategy[47]. Then a new integrated power load point-interval forecasting system that combines multiple artificial intelligence techniques is proposed, aiming to improve the deterministic and volatility analysis performance. The system abstracts the original high-dimensional time series granularly into low-dimensional time series, uses five artificial intelligence algorithms to perform deterministic analysis on the time series after data scale compression, then optimizes the deterministic analysis results from multiple perspectives by the proposed MODOA, and finally analyzes the predicted fluctuations to derive the uncertainty interval estimates.

The main contributions and innovations of this study are as follows:

(1) As a novel integrated power load point-interval forecasting system is proposed. The system can simplify the complexity of calculation while improving the accuracy and stability of forecasting; fluctuation analysis is added to the deterministic analysis, and experiments show that the proposed uncertainty analysis results have better interval scores and interval center deviations.

(2)In the data processing stage, the low-level, fine-grained raw ultra-short-term power load data are granulated and abstracted into high-level, coarse-grained low-dimensional time series, and the constructed information grains can portray and reflect the structural features of the time series data, reducing the total amount of data input to the model and effectively improving the accuracy of short-term forecasting.

(3)A new feedback-based weight-threshold optimization algorithm for BPNNs with neural networks is proposed. An evolutionary update technique and a stochastic strategy are used to intelligently optimize the weights and thresholds of the BPNN containing two hidden layers, which improves the problems of slow convergence and low accuracy of peak traffic prediction BPNN and improves the prediction accuracy of the BPNN.

(4) By developing a Multi-objective Dingo Optimization Algorithm for multi-level optimization of the benchmark model, the prediction stability is improved while pursuing prediction accuracy. In addition, the newly proposed MODOA has better prediction performance and faster running speed compared with other weight optimization algorithms in the market.

This paper is organized as follows. Section II presents the specific methodological theory of the invented model, and Section III describes the main components of the integrated power load point-interval forecasting system. In order to illustrate the capabilities of the developed prediction system, four different experiments are conducted in Section IV. Specifically, Section 4.1 describes the dataset used in this study, Section 4.2 presents the multidimensional evaluation metrics for point-interval forecasting, and Section 4.3 discusses and analyzes the experimental results of the developed FMICM compared to other models. Section V gives a discussion of the proof of the proposed prediction system and empirical analysis of the power load forecasting is also given. Finally, conclusions are presented in Section VI. Additionally, the main structure of this study is shown in Figure 1.

2.Methodology

This chapter introduces the main techniques used in the integrated power load point-interval forecasting system, i.e., signal fuzzy processing technique, multi-objective combined optimization algorithm (MODOA), and volatility analysis technique.

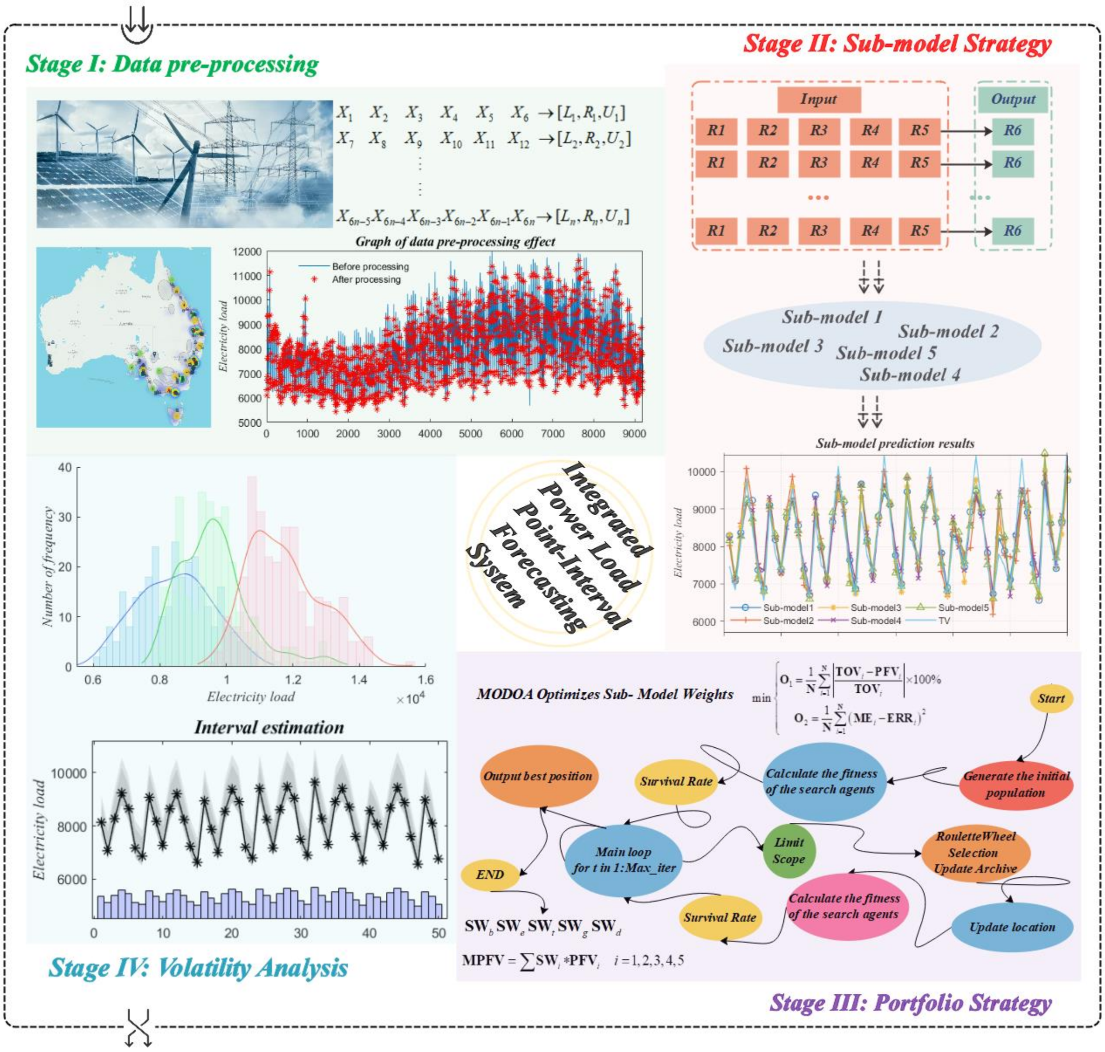

Figure 1 Flow chart of the proposed integrated load forecasting model

Table 1
Evaluation of existing load forecasting models

Models	Refs.	Variables	Results	Advantages	Disadvantages
Physical model					
EnergyPlus	D.B. Crawley et al.(2001)	Power load、Solar thermal、Photovoltaic	It can simulate time series based on DOE-2 and BLAST.	The physical model is easy to operate and does not need a lot of data training	It is difficult to obtain the mathematical expression of energy mechanism, the effect of short-term prediction is poor.
TRANSYS	Yun K et al.(2011)	Power load	The transient response of a building combined with hierarchical CHP optimal control algorithm to forecast data.		
Conventional statistical model					
General regression	VincenzoBianco et al.(2009)	Electricity consumption	Developed regressions are congruent with the official projections.	It has a deep theoretical basis and has higher prediction accuracy than the physical model.	They cannot be well adapted to nonlinear series due to prior linear assumptions.
ARMA	S.Sp.Pappas et al.(2010)	Power load	The reliability, effectiveness and applicability of ARMA in power load data forecasting are proved		
ARIMA	RehanJamil (2020)	Hydropower consumption	The model can predict power load in real time, and has high accuracy and calculation efficiency		
Artificial intelligence algorithm					
SVM	P.Shine et al.(2019)	Electricity consumption	The potential effectiveness of the SVM as a macro-level simulation forecast tool for dairy farm electricity consumption.	Artificial intelligence model has strong nonlinear prediction ability, so the prediction accuracy of artificial intelligence algorithm is generally better than that of time series model.	Relatively dependent on data, it is easy to fall in to local optimization, and the convergence speed is slow, resulting in long running time.
ANN	Si Chen et al.(2021)	Power load	The proposed regression reduces RMSE by 35%, while the ANN with fuzzy hours based reduces RMSE by 42%.		
PSO-ENN	Kun Xie et al.(2020)	Power load	The prediction accuracy of ENN is improved		
LSTM	Gholamreza Memarzadeh et al.(2021)	Power load	Enhance the accuracy and stability of prediction		
Bi-GRU	Xuechen Li et al.(2022)	Oil rate	The observations show that the proposed method performs better than the others in terms of accuracy and robustness.		
TCN	Ruijin Zhu et al.(2020)	Wind power data	TCN shows higher forecasting accuracy than existing predictors such as SVM,MLP, LSTM, and GRU.		
Hybrid model					
BEMD-MMPE-LSTM	Dongchuan Yang et al.(2022)	Power load	A decomposition–ensemble model is proposed for interval-valued load forecasting and it outperforms other model under study.	Integrating the advantages of various models and algorithms can not only improve the prediction accuracy, but also improve the prediction stability.	Because the optimization algorithm used is relatively backward, the prediction accuracy can be further improved
SSA-MOEA	He Bo et al.(2020)	Power load	Effectively improves the efficiency of the power load forecast and adds a new feasible scheme for smart network planning.		
CEEMDAN-MOGOA	Jianzhou Wang et al.(2022)	Power load	Not only has the best performance, but also provides effective technical support for power grid operation scheduling.		

2.1 Signal Fuzzy Processing Technique

Fuzzy information granulation (FIG) is used to construct information grains by creating fuzzy sets on each subsequence formed by the time series after the discretization operation[48]. Fuzzy information granulation mainly includes window division and information fuzzification, the core of which is to complete the fuzzification process after window creation. [49].

The window division is to convert the time series $\bar{\mathbf{T}}=\left\{\bar{\mathbf{T}}_1,\bar{\mathbf{T}}_2,\cdots,\bar{\mathbf{T}}_\gamma\right\}$ into the granular time series $\bar{\bar{\boldsymbol{\Theta}}}=\left\{\bar{\bar{\boldsymbol{\Theta}}}_1,\bar{\bar{\boldsymbol{\Theta}}}_2,\cdots,\bar{\bar{\boldsymbol{\Theta}}}_\varsigma\right\}$ after information granulation. By setting the time granularity $\hat{\mathbf{E}}$ to divide $\bar{\mathbf{T}}=\left\{\bar{\mathbf{T}}_1,\bar{\mathbf{T}}_2,\cdots,\bar{\mathbf{T}}_\gamma\right\}$ into $\underline{\mathbf{H}}$ subseries $\bar{\bar{\boldsymbol{\Theta}}}=\left\{\bar{\bar{\boldsymbol{\Theta}}}_1,\bar{\bar{\boldsymbol{\Theta}}}_2,\cdots,\bar{\bar{\boldsymbol{\Theta}}}_\varsigma\right\}$, where $\underline{\mathbf{H}}=\boldsymbol{\gamma}/\hat{\mathbf{E}}$ and the $\boldsymbol{\eta}$-th subseries is $\bar{\bar{\boldsymbol{\Theta}}}_{\boldsymbol{\eta}}=\left[\bar{\bar{\mathbf{T}}}_1^{(\boldsymbol{\eta})},\bar{\bar{\mathbf{T}}}_2^{(\boldsymbol{\eta})},\cdots,\bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\boldsymbol{\eta})}\right]$.

$$\left\{\bar{\mathbf{T}}_1,\bar{\mathbf{T}}_2,\cdots,\bar{\mathbf{T}}_\gamma\right\}\Rightarrow\left\{\left[\bar{\bar{\mathbf{T}}}_1^{(1)},\bar{\bar{\mathbf{T}}}_2^{(1)},\cdots,\bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(1)}\right],\cdots,\left[\bar{\bar{\mathbf{T}}}_1^{(\underline{\mathbf{H}})},\bar{\bar{\mathbf{T}}}_2^{(\underline{\mathbf{H}})},\cdots,\bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\underline{\mathbf{H}})}\right]\right\} \tag{1}$$

The information granulation of the time series $\bar{\mathbf{T}}=\left\{\bar{\mathbf{T}}_1,\bar{\mathbf{T}}_2,\cdots,\bar{\mathbf{T}}_\gamma\right\}$ is to construct the information particles $\tilde{\boldsymbol{\Gamma}}=\left\{\tilde{\boldsymbol{\Gamma}}_1',\tilde{\boldsymbol{\Gamma}}_2',\cdots,\tilde{\boldsymbol{\Gamma}}_\varsigma'\right\}$ using the fuzzy method for each of the $\underline{\mathbf{H}}$ subsequences $\bar{\bar{\boldsymbol{\Theta}}}=\left\{\bar{\bar{\boldsymbol{\Theta}}}_1,\bar{\bar{\boldsymbol{\Theta}}}_2,\cdots,\bar{\bar{\boldsymbol{\Theta}}}_\varsigma\right\}$ formed by the discretization operation.

Definition 1: Suppose $\mathbf{Z}$ is a given theoretical domain, then a fuzzy subset $\boldsymbol{\Lambda}=\left\{\chi,\boldsymbol{\Omega}(\chi)\,|\,\chi\in\mathbf{Z}\right\}$ on $\mathbf{Z}$. Where $\boldsymbol{\Omega}(\chi):\chi\to[0,1]$ represents the affiliation function of $\boldsymbol{\Lambda}$. If two fuzzy subsets $\boldsymbol{\Phi}$ and $\boldsymbol{\Xi}$ are equal, denoted $\boldsymbol{\Phi}=\boldsymbol{\Xi}$, when and only when they have the same affiliation function, i.e., $\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Phi}}'(\chi)=\hat{\boldsymbol{\Omega}}_{\boldsymbol{\Xi}}''(\chi)$.

In this paper, the triangular fuzzy particles are chosen to construct the information grain and its affiliation function is as follows[50]:

$$\mathbf{A}_{Tf}(\mathbf{x})=\begin{cases}\dfrac{\mathbf{x}-\mathbf{I}_{Tf}}{\mathbf{K}_{Tf}-\mathbf{I}_{Tf}},\mathbf{I}_{Tf}\le\mathbf{x}\le\mathbf{K}_{Tf}\\ 0,\mathbf{x}<\mathbf{I}_{Tf}\cup\mathbf{x}>\mathbf{N}_{Tf}\\ \dfrac{\mathbf{N}_{Tf}-\mathbf{x}}{\mathbf{N}_{Tf}-\mathbf{K}_{Tf}},\mathbf{K}_{Tf}<\mathbf{x}\le\mathbf{N}_{Tf}\end{cases} \tag{2}$$

Where $\mathbf{x}$ is the variable in the theoretical domain, $\mathbf{I}_{Tf}$, $\mathbf{K}_{Tf}$, $\mathbf{N}_{Tf}$ are the three parameters of the triangular type fuzzy example affiliation function, which correspond to the lower boundary, average level and upper boundary of the window after fuzzy particleization, respectively[51].

Fuzzy sets get rid of the either-or duality in classical set theory, and extend the value domain of the affiliation function from the binary $\{0,1\}$ to the multi-valued interval $[0,1]$, which is a kind of extension of set theory. Information fuzzification is the fuzzification of each information grain, and the fuzzification of a single sub-window $\bar{\bar{\boldsymbol{\Theta}}}_{\boldsymbol{\mu}}$ generates multiple fuzzy sets $\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{\mu}}'=\left[\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{\mu};1}'',\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{\mu};2}'',\tilde{\boldsymbol{\Gamma}}_{\boldsymbol{\mu};3}''\right]$.

Considering the single-window problem, $\bar{\bar{\boldsymbol{\Theta}}}_{\boldsymbol{\mu}}=\left[\bar{\bar{\mathbf{T}}}_1^{(\boldsymbol{\mu})},\bar{\bar{\mathbf{T}}}_2^{(\boldsymbol{\mu})},\cdots,\bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\boldsymbol{\mu})}\right]$ should first be viewed as a window for fuzzification. The task of fuzzification is to build a triangular fuzzy particle TFP on $\bar{\bar{\mathbf{\Theta}}}_{\mu} = \left[\bar{\bar{\mathbf{T}}}_1^{(\mu)}, \bar{\bar{\mathbf{T}}}_2^{(\mu)}, \cdots, \bar{\bar{\mathbf{T}}}_{\hat{\mathbf{E}}}^{(\mu)} \right]$, who can reasonably explain the fuzzy concept $\mathbf{M}$ of $\bar{\bar{\mathbf{\Theta}}}_{\mu}$.The fuzzy particle $\tilde{\mathbf{\Gamma}}'_{\mu} = \left[\tilde{\mathbf{\Gamma}}''_{\mu;1} = \hat{\mathbf{I}}^{\mu}_{Tf}, \tilde{\mathbf{\Gamma}}''_{\mu;2} = \hat{\mathbf{K}}^{\mu}_{Tf}, \tilde{\mathbf{\Gamma}}''_{\mu;3} = \hat{\mathbf{N}}^{\mu}_{Tf} \right]$ can be constructed by the relevant parameters in the determined affiliation function (2) of the triangular fuzzy particle.

2.2 Multi-objective Dingo Optimization Algorithm

MODOA is a location update strategy for multilevel optimization, which finds the individual that makes the multi-objective function optimal by Pareto search. Therefore it mainly consists of two parts: location update and pareto search. The pseudo-code of the developed MODOA is shown in Algorithm 1.

(a) Location Update

Herna´n Peraza-Va´zquez proposed the Dingo Optimization Algorithm based on the predatory behavior of Australian wild dogs, the dingo is Australia's dingo is the most dangerous animal in Australia, the top local carnivore in Australia. Due to its small size, the dingo will select weak or dying objects, and when out hunting the dingo usually attacks in groups, they cooperate with each other, some attacking from behind some flanking, surround the prey in a perimeter and start chasing it until they are exhausted.With this inspiration Dingo Optimization Algorithm divides the considered hunting strategies into Group Attack, Persecution, Scavenger, and Dingoes' Survival Rates. The calculation formula is Equation (3)-(7). The definitions and theories related to the study are given below.

Definition 2: Group Attack. When attacking large animals,the dingo usually attacks in groups, surrounds its prey and starts chasing until it is captured. If the first instantaneous random number $\tilde{\mathbf{I}}'_r$ is smaller than the set random number $\bar{\mathbf{K}}_r$ and the second instantaneous random number $\tilde{\mathbf{I}}''_r$ is smaller than the set random number $\bar{\bar{\mathbf{\Lambda}}}_r$, i.e. $\mathbf{IF}: \tilde{\mathbf{I}}'_r < \bar{\mathbf{K}}_r \cap \tilde{\mathbf{I}}''_r < \bar{\bar{\mathbf{\Lambda}}}_r$,then the group attack strategy is applied.

To begin with, calculate the search agent subset $\overrightarrow{\overline{\overline{\mathbf{N}}}}{}_{\zeta}^{\psi;\mathbf{v}}$. If $\hat{\hat{\mathbf{\Delta}}}_{\zeta}^{(\psi)} \notin \mathbf{Q}$ is satisfied, where $\hat{\hat{\mathbf{\Delta}}}_{\zeta}^{(\psi)}$ is a random number, then $\hat{\hat{\mathbf{\Delta}}}_{\zeta}^{(\psi)}$ is stored to $\mathbf{Q}$, i.e. $\mathbf{Q}^{\Delta}(\lambda) = \hat{\hat{\mathbf{\Delta}}}_{\zeta}^{(\psi)}$.Cycle $\mathbf{V}$ times after $\mathbf{Q} = \left[\tilde{\mathbf{Q}}_{\hat{\hat{\Delta}}}^{\langle 1 \rangle}, \tilde{\mathbf{Q}}_{\hat{\hat{\Delta}}}^{\langle 2 \rangle}, \cdots, \tilde{\mathbf{Q}}_{\hat{\hat{\Delta}}}^{\langle \mathbf{v} \rangle} \right]$ contains $\mathbf{V}$ different numbers, the search agent subset $\overrightarrow{\overline{\overline{\mathbf{N}}}}{}_{\zeta}^{\psi;\mathbf{v}} = \left[\overrightarrow{\mathbf{N}_{\zeta}^{\psi;1}}, \overrightarrow{\mathbf{N}_{\zeta}^{\psi;2}}, \cdots, \overrightarrow{\mathbf{N}_{\zeta}^{\psi;\mathbf{v}}} \right]$ that is the location of the set $\mathbf{Q} = \left[\tilde{\mathbf{Q}}_{\hat{\hat{\Delta}}}^{\langle 1 \rangle}, \tilde{\mathbf{Q}}_{\hat{\hat{\Delta}}}^{\langle 2 \rangle}, \cdots, \tilde{\mathbf{Q}}_{\hat{\hat{\Delta}}}^{\langle \mathbf{v} \rangle} \right]$, that is, $\overrightarrow{\mathbf{N}_{\zeta}^{\psi;\theta}} = \overrightarrow{\mathbf{P}'^{(\psi;\theta)}_{\zeta}}$.The location update formula of group attack policy is:

$$\overrightarrow{\mathbf{P}''^{(\psi)}_{\zeta+1}} = \tilde{\mathbf{M}}' \times \sum_{\mathbf{v}=1}^{\kappa} \left[\overrightarrow{\overline{\overline{\mathbf{N}}}}{}_{\zeta}^{\psi;\mathbf{v}} - \overrightarrow{\mathbf{P}'^{(\psi)}_{\zeta}} \right] \Big/ \kappa - \overrightarrow{\mathbf{P}'^{(*)}_{\zeta}} \tag{3}$$

Among them, $\overrightarrow{\mathbf{P}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent (indicates dingoes' movement). $\boldsymbol{\kappa}$ is a random integer between $[2, Sizepop/2]$,where sizepop is the total size of the population of dingoes. $\overrightarrow{\mathbf{P}'^{(\psi)}_{\zeta}}$ is the current search agent. $\overrightarrow{\mathbf{P}'^{(*)}_{\zeta}}$ is the best search agent found from the previous iteration, and $\tilde{\mathbf{M}}'$ is a random number uniformly generated in the interval of [− 2, 2].

Definition 3: Persecution. When attacking small animals, wild dogs usually attack individually and chase until they are caught. If the first instantaneous random number $\tilde{\mathbf{I}}'_r$ is smaller than the set random number $\bar{\mathbf{K}}_r$ and the second instantaneous random number $\tilde{\mathbf{I}}''_r$ is greater than the set random number $\bar{\bar{\mathbf{\Lambda}}}_r$, i.e. $\mathbf{IF}: \tilde{\mathbf{I}}'_r < \bar{\mathbf{K}}_r \cap \tilde{\mathbf{I}}''_r > \bar{\bar{\mathbf{\Lambda}}}_r$,then the persecution strategy is applied.

Here, we use the random number $\breve{\mathbf{\Delta}}_\zeta^{(\chi)} \neq \boldsymbol{\xi}$ in the group attack strategy to determine the location $\overrightarrow{\vec{\mathbf{P}}'_\psi\left(\breve{\mathbf{\Delta}}^{(\chi)}\right)}$, with the random number $\breve{\mathbf{E}}''$ and $\tilde{\mathbf{M}}'$ in the group attack strategy.The location update formula of persecution is:

$$\overrightarrow{\vec{\mathbf{P}}''^{(\psi)}_{\zeta+1}} = \overrightarrow{\vec{\mathbf{P}}'^{(*)}_{\zeta}} + \tilde{\mathbf{M}}' \times e^{\breve{\mathbf{E}}''} \times \left(\overrightarrow{\vec{\mathbf{P}}'_\psi\left(\breve{\mathbf{\Delta}}^{(\chi)}\right)} - \overrightarrow{\vec{\mathbf{P}}'^{(\psi)}_{\zeta}} \right) \tag{4}$$

In which, $\overrightarrow{\vec{\mathbf{P}}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent. $\overrightarrow{\vec{\mathbf{P}}'^{(\psi)}_{\zeta}}$ is the current search agent. $\overrightarrow{\vec{\mathbf{P}}'^{(*)}_{\zeta}}$ is the best search agent found from the previous iteration, and $\tilde{\mathbf{M}}' \in [-2, 2]$ $\breve{\mathbf{E}}'' \in [-1, 1]$.

Definition 4: Scavenger. When dingo smells a dead small animal on the ground nearby during his daily walk, this behavior is called scavenger in this section. If the first instantaneous random number $\tilde{\mathbf{I}}'_r$ is greater than the set random number $\bar{\mathbf{K}}_r$, i.e. $\mathbf{IF}: \tilde{\mathbf{I}}'_r > \bar{\mathbf{K}}_r$,then the scavenger strategy is applied.

We also use a random number strategy to determine the location $\overrightarrow{\bar{\bar{\mathbf{P}}}'_\psi\left(\hat{\mathbf{\Delta}}^{(\chi)}\right)}$,The location update formula of scavenger is:

$$\overrightarrow{\bar{\bar{\mathbf{P}}}''^{(\psi)}_{\zeta+1}} = \frac{1}{2}\left[e^{\breve{\mathbf{E}}''} * \overrightarrow{\bar{\bar{\mathbf{P}}}'_\psi\left(\hat{\mathbf{\Delta}}^{(\chi)}\right)} - (-1)^{\tilde{\mathbf{H}}''} \times \overrightarrow{\bar{\bar{\mathbf{P}}}'^{(\psi)}_{\zeta}} \right] \tag{5}$$

In which, $\overrightarrow{\bar{\bar{\mathbf{P}}}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent. $\overrightarrow{\bar{\bar{\mathbf{P}}}'^{(\psi)}_{\zeta}}$ is the current search agent, and $\breve{\mathbf{E}}'' \in [-1, 1]$ $\quad \tilde{\mathbf{H}}''' \in \{0, 1\}$.

Definition 5: Dingoes' Survival Rates. In addition to the above three location update strategies, DOA also considers the survival rate of dingo. The location update formula of dingoes' Survival Rates is:

$$\widehat{\mathbf{Sr}}^{\psi}_{\xi}(\delta) = \frac{\overline{\mathbf{XF}}^{\psi}_{\xi} - \mathbf{IF}^{\psi}_{\xi}(\delta)}{\overline{\mathbf{XF}}^{\psi}_{\xi} - \overline{\mathbf{NF}}^{\psi}_{\xi}} \tag{6}$$

Among them, $\overline{\mathbf{XF}}^{\psi}_{\xi}$ and $\overline{\mathbf{NF}}^{\psi}_{\xi}$ are the worst and the best fitness value in the current generation, respectively, whereas $\mathbf{IF}^{\psi}_{\xi}(\delta)$ is the current fitness value of the δ−th search agent. When the survival rate of dingo is lower than 0.3,i.e. $\widehat{\mathbf{Sr}}^{\psi}_{\xi}(\delta) < 0.3$,the location update formula becomes:

$$\overrightarrow{\tilde{\mathbf{P}}''^{(\psi)}_{\zeta+1}} = \overrightarrow{\tilde{\mathbf{P}}'^{(*)}_{\zeta}} + \frac{1}{2} \times \left[\overrightarrow{\tilde{\mathbf{P}}'_\psi\left(\vec{\mathbf{\Delta}}_1^{(\chi)}\right)} - (-1)^{\bar{\bar{\mathbf{H}}}''} * \overrightarrow{\tilde{\mathbf{P}}'_\psi\left(\vec{\mathbf{\Delta}}_2^{(\chi)}\right)} \right] \tag{7}$$

In which, $\overrightarrow{\tilde{\mathbf{P}}''^{(\psi)}_{\zeta+1}}$ is the new position of a search agent, $\overrightarrow{\tilde{\mathbf{P}}'^{(*)}_{\zeta}}$ is the best search agent found from the previous iteration, and $\bar{\bar{\mathbf{H}}}''' \in \{-1, 1\}$. Since the survival rate is not passed, this formula uses two random number locations $\vec{\boldsymbol{\Delta}}_1^{(\chi)}$ and $\vec{\boldsymbol{\Delta}}_2^{(\chi)}$, which means that the two random numbers position $\overrightarrow{\tilde{\mathbf{P}}'_\psi\left(\vec{\boldsymbol{\Delta}}_1^{(\chi)}\right)}$ and $\overrightarrow{\tilde{\mathbf{P}}'_\psi\left(\vec{\boldsymbol{\Delta}}_2^{(\chi)}\right)}$ are used to generate new locations according to the generated.

(b) Pareto search

Definition 6: When multiple objectives $\tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)=\left[\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\cdots,\mathbf{O}_k^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right]$ in the objective function that need to be optimized, and these objectives are usually conflicting, the problem of finding a set of vectors $\vec{\mathbf{V}}=\left[\vec{\varpi}_1,\vec{\varpi}_2,\cdots,\vec{\varpi}_\rho\right]$ such that $\tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)=\left[\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\cdots,\mathbf{O}_k^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right]$ is maximized or minimized is called a multi-objective optimization problem.In mathematical terms, a multi-objective optimization problem can be written as:

$$\begin{gathered}\min\left(\mathbf{O}_1^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\mathbf{O}_2^{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\cdots,\mathbf{O}_k^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right)\\ s.t.\ \begin{cases}\Theta'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\le 0\\ \Theta''_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)=0\end{cases}\end{gathered}\tag{8}$$

Where the integer k is the target number and $\left\{\Theta'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right),\Theta''_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)\right\}$ is the constraint function.

The purpose of constructing a multi-objective optimization algorithm is to compensate for the shortage of pursuing only accuracy due to the single optimization algorithm, so the multi-objective function constructed in this paper includes the mean absolute percentage error (MAPE), which pursues accuracy, on the one hand, and the residual variance (RV), which pursues prediction stability, on the other hand.

$$\min\begin{cases}\mathbf{O}_1=\dfrac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left|\dfrac{\mathbf{TOV}_i-\mathbf{PFV}_i}{\mathbf{TOV}_i}\right|\times 100\%\\ \mathbf{O}_2=\dfrac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left(\mathbf{ME}_i-\mathbf{ERR}_i\right)^2\end{cases}\tag{9}$$

Where, $\mathbf{TOV}_i$ denotes the i-th actual observation value, $\mathbf{PFV}_i$ denotes the i-th PF forecast value, $\mathbf{ME}_i$ is the average of the error $\mathbf{ERR}_i=\mathbf{TOV}_i-\mathbf{PFV}_i$ of the i-th true value $\mathbf{TOV}_i$ and the i-th predicted value $\mathbf{PFV}_i$.

The single-objective optimization algorithm does not apply to multi-objective optimization problems.

Proof: Suppose $\vec{\boldsymbol{\Gamma}}=\left[\vec{\boldsymbol{\gamma}}_1,\vec{\boldsymbol{\gamma}}_2,\cdots,\vec{\boldsymbol{\gamma}}_\rho\right]$ and $\vec{\boldsymbol{\Pi}}=\left[\vec{\boldsymbol{\lambda}}_1,\vec{\boldsymbol{\lambda}}_2,\cdots,\vec{\boldsymbol{\lambda}}_\rho\right]$ are two sets of solutions. $\mathbf{If:}\ \exists\vec{\boldsymbol{\Gamma}},\vec{\boldsymbol{\Pi}}\ s.t.\ \mathbf{O}_1^{\mathbf{M}}\left(\vec{\boldsymbol{\Gamma}}\right)<\mathbf{O}_1^{\mathbf{M}}\left(\vec{\boldsymbol{\Pi}}\right)\cap\mathbf{O}_2^{\mathbf{M}}\left(\vec{\boldsymbol{\Gamma}}\right)>\mathbf{O}_2^{\mathbf{M}}\left(\vec{\boldsymbol{\Pi}}\right)$,according to the single-objective optimization problem solution, only $\mathbf{O}_1^{\mathbf{M}}$ will be sorted and the optimal solution will be $\vec{\boldsymbol{\Gamma}}=\left[\vec{\boldsymbol{\gamma}}_1,\vec{\boldsymbol{\gamma}}_2,\cdots,\vec{\boldsymbol{\gamma}}_\rho\right]$, which is not in line with the principle of multi-objective optimization.

For single-objective optimization problems, the maximum value of the derived objective function can be directly selected as the optimal solution at this stage. However, for multi-objective optimization problems, there is usually a tendency of mutual constraints between different objective functions, which may improve the performance of one objective often at the expense of the performance of other objectives, so for multi-objective optimization problems, the solution is usually a set of non-inferior solutions-Pareto solution set.

Definition 7: Given a multi-objective optimization problem $\min \tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, let $\vec{\mathbf{V}}^*=\left[\vec{\varpi}_1^*,\vec{\varpi}_2^*,\cdots,\vec{\varpi}_\rho^*\right]\in\mathbf{\Omega}$,if $\exists\vec{\mathbf{V}}=\left[\vec{\varpi}_1,\vec{\varpi}_2,\cdots,\vec{\varpi}_\rho\right]\in\mathbf{\Omega}$ such that the following conditions are satisfied:

For any subgoal function $\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ of $\tilde{\mathbf{O}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ there exists $\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\vec{\mathbf{V}}^*\right)\le\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, while there exists at least one subgoal function $\tilde{\mathbf{O}}_{\boldsymbol{\varphi}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ such that $\tilde{\mathbf{O}}_{\boldsymbol{\varphi}}^{\mathbf{M}}\left(\vec{\mathbf{V}}^*\right)<\tilde{\mathbf{O}}_{\boldsymbol{\varphi}}^{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, then we say that $\vec{\mathbf{V}}^*=\left[\vec{\varpi}_1^*,\vec{\varpi}_2^*,\cdots,\vec{\varpi}_\rho^*\right]$ is a strong pareto optimal solution.

Definition 8: Given a multi-objective optimization problem $\min \breve{\mathbf{O}}'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$, let $\vec{\mathbf{V}}'_*=\left[\vec{\varpi}'_{*;1},\vec{\varpi}'_{*;2},\cdots,\vec{\varpi}'_{*;\rho}\right]\in\mathbf{\Omega}$,if $\exists\vec{\mathbf{V}}'=\left[\vec{\varpi}'_1,\vec{\varpi}'_2,\cdots,\vec{\varpi}'_\rho\right]\in\mathbf{\Omega}$ such that the following conditions are satisfied:

For any subgoal function $\breve{\mathbf{O}}'^{\tau}_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ of $\breve{\mathbf{O}}'_{\mathbf{M}}\left(\vec{\mathbf{V}}\right)$ there exists $\breve{\mathbf{O}}'^{\tau}_{\mathbf{M}}\left(\vec{\mathbf{V}}'_*\right)\le\breve{\mathbf{O}}'^{\tau}_{\mathbf{M}}\left(\vec{\mathbf{V}}'\right)$, then we say that $\vec{\mathbf{V}}'_*=\left[\vec{\varpi}'_{*;1},\vec{\varpi}'_{*;2},\cdots,\vec{\varpi}'_{*;\rho}\right]$ is a weak pareto optimal solution.

Definition 9: Suppose there are $\mathbf{N}$ sets of position vectors $\overrightarrow{\mathbf{M}\tilde{\mathbf{s}}}=\left[\vec{\mathbf{M}}_1,\vec{\mathbf{M}}_2,\cdots,\vec{\mathbf{M}}_{\mathbf{N}}\right]$ in the archive,where $\vec{\mathbf{M}}_{\boldsymbol{\sigma}}=\left[\vec{\varpi}_{\boldsymbol{\sigma}}^{(1)},\vec{\varpi}_{\boldsymbol{\sigma}}^{(2)},\cdots,\vec{\varpi}_{\boldsymbol{\sigma}}^{(\rho)}\right]$, and each set of position vectors corresponds to an adaptation function $\overrightarrow{\mathbf{H}\tilde{\mathbf{s}}}=\left[\vec{\mathbf{H}}_1,\vec{\mathbf{H}}_2,\cdots,\vec{\mathbf{H}}_{\mathbf{N}}\right]$,where $\vec{\mathbf{H}}_{\boldsymbol{\sigma}}=\left[\vec{\mathbf{H}}_{\boldsymbol{\sigma}}^{(1)},\vec{\mathbf{H}}_{\boldsymbol{\sigma}}^{(2)}\right]$.

We obtain $\overrightarrow{\mathbf{P}\tilde{\mathbf{r}}}=\left[\tilde{\mathbf{R}}_1,\tilde{\mathbf{R}}_2,\cdots,\tilde{\mathbf{R}}_{\mathbf{N}}\right]$ by pareto ranking $\overrightarrow{\mathbf{H}\tilde{\mathbf{s}}}=\left[\vec{\mathbf{H}}_1,\vec{\mathbf{H}}_2,\cdots,\vec{\mathbf{H}}_{\mathbf{N}}\right]$ from best to worst, then $\tilde{\mathbf{E}}_\zeta=\tilde{\mathbf{R}}_\zeta\Big/\sum_{\lambda=1}^{\mathbf{N}}\tilde{\mathbf{R}}_\lambda$, $\zeta=1,2,\ldots,\mathbf{N}$ is the probability of being eliminated.This method is known as roulette selection method, also known as proportional selection method.

In the iterative loop, by finding out the group strong pareto solution $\bar{\mathbf{S}}^*$,it needs to be filed into $\overrightarrow{\mathbf{A}\tilde{\mathbf{r}}}=\left[\vec{\mathbf{\Lambda}}_1,\vec{\mathbf{\Lambda}}_2,\cdots,\vec{\mathbf{\Lambda}}_\delta\right]$, if the following occurs:

$$\mathbf{If}:\exists\delta:\left\{\forall\eta:\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\bar{\mathbf{S}}^*\right)\le\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\vec{\mathbf{\Lambda}}_\delta\right) and \exists\eta:\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\bar{\mathbf{S}}^*\right)<\tilde{\mathbf{O}}_\eta^{\mathbf{M}}\left(\vec{\mathbf{\Lambda}}_\delta\right)\right\} \tag{10}$$

Then file $\bar{\mathbf{S}}^*$ to $\overrightarrow{\mathbf{A}\tilde{\mathbf{r}}}=\left[\vec{\mathbf{\Lambda}}_1,\vec{\mathbf{\Lambda}}_2,\cdots,\vec{\mathbf{\Lambda}}_\delta\right]$,i.e. $\overrightarrow{\mathbf{A}\tilde{\mathbf{r}}}\left(\vec{\mathbf{\Lambda}}_{\delta+1}\right)=\bar{\mathbf{S}}^*$.If the $\overrightarrow{\mathbf{A}\tilde{\mathbf{r}}}$ storage reaches its limit, $\bar{\mathbf{S}}^*$ is substituted for $\overrightarrow{\mathbf{A}\tilde{\mathbf{r}}}\left(\vec{\mathbf{\Lambda}}_{\boldsymbol{\mu}}\right)$ using the roulette selection method.

2.3 Volatility Analysis Technique

Nonparametric kernel density estimation simulates the true probability distribution curve without using a priori knowledge of the data distribution. Therefore, it is a non-parametric method suitable for power load interval forecasting studies.We propose the improved kernel density estimation method (IKDE) in this paper, based on the point prediction results obtained from FMICM.

Definition 10: Suming that $\mathbf{\Omega}(\boldsymbol{\psi})$ is the probability density function,

$\boldsymbol{\Xi}(\boldsymbol{\psi})=\int_{-\infty}^{\boldsymbol{\psi}}\boldsymbol{\Omega}(\zeta)d\zeta$ is the cumulative distribution function.As $\boldsymbol{\Xi}_n(\boldsymbol{\chi})=\frac{1}{\mathbf{T}}\sum_{\iota=1}^{\mathbf{T}} l_{\boldsymbol{\psi}_\iota}\leq\boldsymbol{\chi}$:

$$\boldsymbol{\Omega}(\tilde{\boldsymbol{\psi}}_\phi)=\lim_{\delta\to 0}\frac{\boldsymbol{\Xi}(\tilde{\boldsymbol{\psi}}_\phi+\delta)-\boldsymbol{\Xi}(\tilde{\boldsymbol{\psi}}_\phi-\delta)}{2\delta}=\frac{1}{2\mathbf{HT}}\sum_{\iota=1}^{\mathbf{T}} l_{\tilde{\boldsymbol{\psi}}_\phi-\delta\leq\boldsymbol{\psi}\leq\tilde{\boldsymbol{\psi}}_\phi+\delta} \tag{11}$$

Rewrite Equation (11) as $\boldsymbol{\Omega}(\tilde{\boldsymbol{\psi}}_\phi)=\left[\sum_{\iota=1}^{\mathbf{T}}\mathbf{K}\left(\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi/\mathbf{H}\right)\right]\Big/2\mathbf{HT}$, Call it kernel density estimation.

To avoid information leakage, this paper uses the error percentage of the optimization set to fit the kernel probability density function. The error percentage $\overrightarrow{\mathbf{E\tilde{r}}}$ is calculated from the true value $\overrightarrow{\mathbf{T\tilde{o}}}=\left[\mathbf{TOV}_1^{\mathbf{M}},\mathbf{TOV}_2^{\mathbf{M}},\cdots,\mathbf{TOV}_{\boldsymbol{\varphi}}^{\mathbf{M}}\right]$ and the predicted value $\overrightarrow{\mathbf{P\tilde{f}}}=\left[\mathbf{PFV}_1^{\mathbf{M}},\mathbf{PFV}_2^{\mathbf{M}},\cdots,\mathbf{PFV}_{\boldsymbol{\varphi}}^{\mathbf{M}}\right]$, which is $\overrightarrow{\mathbf{E\tilde{r}}}=\left[\mathbf{ERR}_1^{\mathbf{M}},\mathbf{ERR}_2^{\mathbf{M}},\cdots,\mathbf{ERR}_{\boldsymbol{\varphi}}^{\mathbf{M}}\right]$, $\mathbf{ERR}_{\boldsymbol{\eta}}^{\mathbf{M}}=\left(\mathbf{TOV}_{\boldsymbol{\eta}}^{\mathbf{M}}-\mathbf{PFV}_{\boldsymbol{\eta}}^{\mathbf{M}}\right)\Big/\mathbf{PFV}_{\boldsymbol{\eta}}^{\mathbf{M}}$.

The setting of bandwidth and the selection of kernel functions directly affect the smoothness and fit of the density curve in the NKDE algorithm, which in turn affects the accuracy of the calculation. Since the kernel function has little effect on the final impact, the Gaussion kernel is chosen here and its kernel probability density function is:

$$\boldsymbol{\Omega}(\boldsymbol{\psi})=\frac{1}{\sqrt{2\pi}\mathbf{TH}}\sum_{\iota=1}^{\mathbf{T}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi}{\mathbf{H}}\right)^2} \tag{12}$$

Where the bandwidth H is optimized based on the error term of the optimized data set using the location update strategy, which is superior to the mean-squared error algorithm.After the optimal probability density curve of the error is known, the confidence interval $\left[\bar{\mathbf{G}}'_{\alpha/2},\bar{\mathbf{G}}'_{1-\alpha/2}\right]$ of the error at $1-\alpha$ confidence level is calculated using the integral equation (13), which satisfies both $\mathbf{P}(\bar{\mathbf{G}}'_{\alpha/2}<\bar{\boldsymbol{\psi}}'<\bar{\mathbf{G}}'_{1-\alpha/2})=1-\alpha$.

$$\begin{aligned}\int_{-\infty}^{\bar{\mathbf{G}}'_{\alpha/2}}\frac{1}{\sqrt{2\pi}\mathbf{TH}}\sum_{\iota=1}^{\mathbf{T}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi}{\mathbf{H}}\right)^2}d\boldsymbol{\psi}=\alpha/2\\ \int_{-\infty}^{\bar{\mathbf{G}}'_{1-\alpha/2}}\frac{1}{\sqrt{2\pi}\mathbf{TH}}\sum_{\iota=1}^{\mathbf{T}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{\psi}-\tilde{\boldsymbol{\psi}}_\phi}{\mathbf{H}}\right)^2}d\boldsymbol{\psi}=1-\alpha/2\end{aligned} \tag{13}$$

For a given confidence level, the final interval prediction formula can be derived from the obtained error confidence intervals as $[\mathbf{PFV}_i\times\left(1+\bar{\mathbf{G}}'_{\alpha/2}\right),\mathbf{PFV}_i\times\left(1+\bar{\mathbf{G}}'_{1-\alpha/2}\right)]$.

3. Main structure of integrated electric load point-interval forecasting system

The integrated electric load point-interval forecasting system proposed in this thesis is an electric load point-interval forecasting system integrating data pre-processing module, model forecasting module, combined optimization module and uncertainty analysis module, which improves forecasting accuracy and forecasting stability. This system first decomposes the original ultra-short-term power load data into a series of information grains to reduce the total amount of data input to the model and improve the forecasting accuracy. Secondly, the prediction accuracy and stability of different models for different data are different, from which five AI benchmark models are selected in this module: DOA-BPNN, Extreme Learning Machine (ELM), Time Convolutional Neural Network (TCN), Gated Recurrent Unit (GRU), and Deep Belief Network (DBN), according to which the electric load data are trained to derive

***Algorithm 1* : MODOA**

Input: Predicted values of five models $\left\{ \mathbf{PFV}_b, \mathbf{PFV}_e, \mathbf{PFV}_t, \mathbf{PFV}_g, \mathbf{PFV}_d \right\}$

Output: Optimal weight $\left\{ \mathbf{SW}_1, \mathbf{SW}_2, \mathbf{SW}_3, \mathbf{SW}_4, \mathbf{SW}_5 \right\}$

Initialization of parameters and Archive
Generate the initial population
for *i: 1 ≤ i ≤ SA(SearchAgents)*
    Evaluate the corresponding fitness function ***GHF**ᵢ* → ***GHF_i***
    if *mindominates maxdominates*
    /*calculation **minfitness min_position maxfitness max_position */**
    end if
end for
$\mathbf{S}\widehat{\mathbf{r}}_{\xi}^{\psi}(\delta) = \left(\overline{\mathbf{XF}}_{\xi}^{\psi} - \mathbf{IF}_{\xi}^{\psi}(\delta) \right) \Big/ \left(\overline{\mathbf{XF}}_{\xi}^{\psi} - \overline{\mathbf{NF}}_{\xi}^{\psi} \right)$
while *iteration < Max Number of Iterations* **do**
    for *i: 1 ≤ i ≤ SA(SearchAgents)*
        Evaluate the corresponding fitness function ***GHF_i***
        if *dominates* **/***Select the minimum fitness value and position***/**
        end if
    end for
    Archive = UpdateArchive(*Archive*,*GHF*,*GHP*)
    if *Archive_member_no>ArchiveMaxSize*
        Ranking = RankingProcess(*Archive*)
    end if
    for *r: 1 ≤ i ≤ SA(SearchAgents)*
        if *random < P* **then**
            if *random < Q* **then**
                $\overrightarrow{\mathbf{P}''^{(\psi)}_{\zeta+1}} = \tilde{\mathbf{M}}' \times \sum_{\mathbf{v}=1}^{\mathbf{\kappa}} \left[\overrightarrow{\overline{\mathbf{N}}^{\psi;\mathbf{v}}_{\zeta}} - \overrightarrow{\mathbf{P}'^{(\psi)}_{\zeta}} \right] \Big/ \mathbf{\kappa} - \overrightarrow{\mathbf{P}'^{(*)}_{\zeta}}$ **/***Group Attack***/**
            else
                $\overrightarrow{\vec{\mathbf{P}}''^{(\psi)}_{\zeta+1}} = \overrightarrow{\vec{\mathbf{P}}'^{(*)}_{\zeta}} + \tilde{\mathbf{M}}' \times e^{\breve{\ddot{\mathbf{E}}}''} \times \left(\overrightarrow{\vec{\mathbf{P}}'_{\psi}\left(\breve{\mathbf{\Delta}}^{(\chi)} \right)} - \overrightarrow{\vec{\mathbf{P}}'^{(\psi)}_{\zeta}} \right)$ **/***Persecution***/**
            end if
        else
            $\overrightarrow{\overline{\overline{\mathbf{P}}}''^{(\psi)}_{\zeta+1}} = \left[e^{\breve{\ddot{\mathbf{E}}}''} * \overrightarrow{\overline{\overline{\mathbf{P}}}'_{\psi}\left(\widehat{\mathbf{\Delta}}^{(\chi)} \right)} - (-1)^{\tilde{\mathbf{H}}'''} \times \overrightarrow{\overline{\overline{\mathbf{P}}}'^{(\psi)}_{\zeta}} \right] \Big/ 2$ **/***Scavenger***/**
        end if
        if *survival(r) <= 0.3*
            reture $\overrightarrow{\mathbf{SW}_*}$
        end if
        Evaluate the corresponding fitness function ***GHF_i***
        if *dominates* /*Select the minimum fitness value and position*/
        end if
    end for
end while
reture $\overrightarrow{\mathbf{SW}_*}$

the base prediction value of each model.Then,based on the evolutionary computation technique of population intelligence and omission strategy, a new multi-level optimization algorithm is proposed to integrate each model and finally obtain the point prediction values of the system. At last, the residual distribution is fitted by fluctuation

Table 2

FMICM uses the model's parameters to set values

Required Model	Parameters		Value
FIG	MF	Types of affiliation functions	triangle
	w	Number of windows for granulation	6
DOA-BPNN	S_A	Number of individuals to be optimized	30
	M_{iter}	Maximum number of iterations	100
	l_r	BPNN's Learning rate	0.1
	E_p	BPNN’s Training times	100
	G	BPNN’s Error accuracy	0.00004
TCN	Embedding size of the convolutional layers in the residual block		[128,64,32,16]
	Kernel size		[3,3,3,3]
	Dilation rate		[1,2,4,8]
	Batch size		20
	Epochs finetune		500
GRU	Spatial Dimension in GRU		[64,32,16,1]
	Batch size		1
	Epochs finetune		200
DBN	Batch size		128
	Epochs finetune		2000
MODOA	S_A	Search Number of Individuals	100
	M_{iter}	Maximum iterations Number	200
	A_m	ArchiveMaxSize	500
	P	Random numbers in algorithms	0.5
	Q	Random numbers in algorithms	0.7
IKDE	L_b	Lower limit of bandwidth	0.01
	U_b	Upper limit of bandwidth	0.1
	S_A	Search Number of Individuals	6
	M_{iter}	Maximum iterations Number	10

Note: The above parameters were obtained by pre-experiments.The ELM parameters used in this paper are obtained by looping through the global optimal solution, so there is no fixed parameter value.

analysis, and confidence intervals are calculated and coupled with the system point prediction values to obtain the final uncertainty prediction results.Details of the parameters of the model used by FMICM are shown in Table 2.

4. Experiments and Analysis

To validate the predictive performance of the developed integrated system, this thesis conducts experiments using three sets of electricity load data from March 2020 to November 2021 in New South Wales, Australia. The computer facility used for the experiments in this section of the study is matlab2018a with Windows 10 Home Edition, python3, with a 2.5GHz Intel(R) Core(TM) i5-7300HQ CPU.

4.1. Material

NEM operates in New South Wales, the Australian Capital Territory, Queensland, South Australia, Victoria and Tasmania as both a wholesale electricity market and a physical electricity system. Aemo also operates the retail electricity market that supports the wholesale market. The three datasets used in this paper are NEM statistics of the electricity load in New South Wales from March 2020 to November 2021, with one data point taken every half hour. Specifically, each dataset is a seven-month cycle of load data with 9,000 data points and partitioned to 1500 data points.The first 70% of the data is used as a training set to train individual models, 70% to 90% of the data is used to optimize the weights of each model, and the last 10% of the data is used to measure the predictive capabilities of the proposed system. In addition, the specific characteristics of the data set are shown in Table 3.

Table 3
The details of the three datasets utilized

Dataset	Samples	Numbers	Statistical Indicator(MW)			
			Max	Min	Mean	Std.
Site1	Training	6440	11980.08	5384.58	7722.83	1199.55
	Optimizing	1840	11908.24	6101.04	8578.39	1294.57
	Testing	920	11500.53	5630.73	7807.92	1189.47
	All samples	9200	11980.08	5384.58	7902.45	1264.42
Site2	Training	6440	12401.82	5221.13	7190.10	1002.82
	Optimizing	1840	12197.57	5704.44	7660.07	1088.47
	Testing	920	11404.28	5682.96	7408.21	933.01
	All samples	9200	12401.82	5221.13	7305.90	1031.25
Site3	Training	6440	12863.76	5170.46	8053.08	1351.40
	Optimizing	1840	12040.28	5189.86	7691.26	1174.77
	Testing	920	10236.28	4767.17	7005.83	1055.10
	All samples	9200	12863.76	4767.17	7875.99	1330.50

4.2 Evaluation Indicators

4.2.1 Point Forecast

The criterion we use to evaluate how good a point forecast is is to compare its forecast results with our true results and see the size of the difference between the two. In time series forecasting, Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), The Standard Deviation Of Error

(SDE) are our most frequently used and widely used the four evaluation metrics are the most frequently and widely used. In this paper, the above four metrics are selected as the evaluation criteria of the combined model.

Among these four evaluation criteria, MAPE does not only consider the error between the predicted and true values, but also the ratio between the error and the true value, which is one of the commonly used objective functions in some competitions. MAE is the absolute value of the difference between the predicted and true values for each sample, and then summed to find the average. The RMSE has the same properties as the MSE, but the error can be transformed into the same units as the original data. SDE is the standard deviation of the error, which can detect the model prediction stability. Let $\mathbf{TOV}_i$ be the i-th actual observation value and $\mathbf{PFV}_i$ be the i-th point predicted value, and the formula of evaluation index is shown in Table 4.

4.2.2 Interval Forecast

In interval prediction, the commonly used variables are PI coverage probability (PICP) and PI normalized averaged width (PINAW), and the interval score AIS selected in this paper is a tool used to provide comprehensive consideration of coverage probability and normalized averaged width. When the PICP is larger and the PINAW is smaller, the interval prediction result is better, and when the target is not in the PI coverage interval, AIS will give a certain penalty, so the larger the value of AIS

Table 4

Point-interval prediction results evaluation index

Metric	Nomenclature	Equation
MAPE	Mean Absolute Percentage Error	$MAPE=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left\|\frac{\mathbf{TOV}_i-\mathbf{PFV}_i}{\mathbf{TOV}_i}\right\|\times 100\%$
MAE	Mean Absolute Error	$MAE=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left\|\mathbf{PFV}_i-\mathbf{TOV}_i\right\|$
RMSE	Root Mean Square Error	$RMSE=\sqrt{\frac{1}{\mathbf{N}}\times\sum_{i=1}^{\mathbf{N}}\left(\mathbf{PFV}_i-\mathbf{TOV}_i\right)^2}$
SDE	The standard deviation of error	$SDE=\sqrt{\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left(\mathbf{ME}_i-\mathbf{ERR}_i\right)^2}$
PICP	PI coverage probability	$PICP=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\tau_i^\alpha \qquad \tau_i^\alpha=\begin{cases}1 & \mathbf{TOV}_i^\alpha\in\left[\mathbf{FIL}_i^\alpha,\mathbf{FIU}_i^\alpha\right]\\ 0 & \mathbf{TOV}_i^\alpha\notin\left[\mathbf{FIL}_i^\alpha,\mathbf{FIU}_i^\alpha\right]\end{cases}$
PINAW	PI normalized averaged width	$PINAW=\frac{1}{\mathbf{N}R}\sum_{i=1}^{\mathbf{N}}\left(\mathbf{FIU}_i^\alpha-\mathbf{FIL}_i^\alpha\right)$
AIS	Average interval score	$AIS=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\gamma_i^\alpha \quad \gamma_i^\alpha=\begin{cases}-2\alpha\psi_i^\alpha-4\left(\mathbf{FIL}_i^\alpha-\mathbf{TV}_i\right) & \mathbf{TV}_i<\mathbf{FIL}_i^\alpha\\ -2\alpha\psi_i^\alpha & \mathbf{FIL}_i^\alpha\le\mathbf{TV}\le\mathbf{FIU}_i^\alpha\\ -2\alpha\psi_i^\alpha-4\left(\mathbf{TV}_i-\mathbf{FIU}_i^\alpha\right) & \mathbf{TV}_i>\mathbf{FIU}_i^\alpha\end{cases}$
MPICD	Mean PI center deviation	$MPICD=\frac{1}{\mathbf{N}}\sum_{i=1}^{\mathbf{N}}\left\|\frac{\mathbf{FIU}_i^\alpha+\mathbf{FIL}_i^\alpha}{2}-\mathbf{TOV}_i\right\|$

Note: MAPE , MAE , RMSE , SDE is the evaluation index of point prediction results, the smaller the value of all four indicators, the better. PICP, PINAW, AIS, MPICD is an indicator to evaluate the good or bad interval prediction results, Except for AIS, all others are small is better, while AIS is large is better.

indicates the better quality of the prediction interval. If the coverage is the same and there is no difference in width, MPICD plays a role. If two different PIs cover a point, the closer to the midline of PI, the better the quality of PI, also known as the smaller the MPICD the better the prediction interval. Let $\mathbf{FIL}_i^{\alpha}$ be the upper limit of the first prediction interval and $\mathbf{FIU}_i^{\alpha}$ be the lower limit of the first prediction interval, and the formula of evaluation index is shown in Table 4.

4.3 Different experiments and result analysis

During this phase, four different experiments will be planned to investigate the prediction performance of the point or interval of the integrated prediction system developed in this paper. Since the results of each run of the model are different, the final data are averaged over the five results for fair comparison.

4.3.1 Experiment I: Comparison with the prediction model of three-hour interval data without FIG

In Experiment 1, the aim was to verify the enhancement of the predictive power of the fuzzy granulation technique used in the proposed point prediction system. The FMICM was then compared with five single models performed on non-fuzzy granulation data, namely DOABPNN, ELM, TCN, GRU and DBN, and the combined model MODOA, and the prediction results obtained from the experiment are shown in Table 5, and other details of the prediction results are shown below.

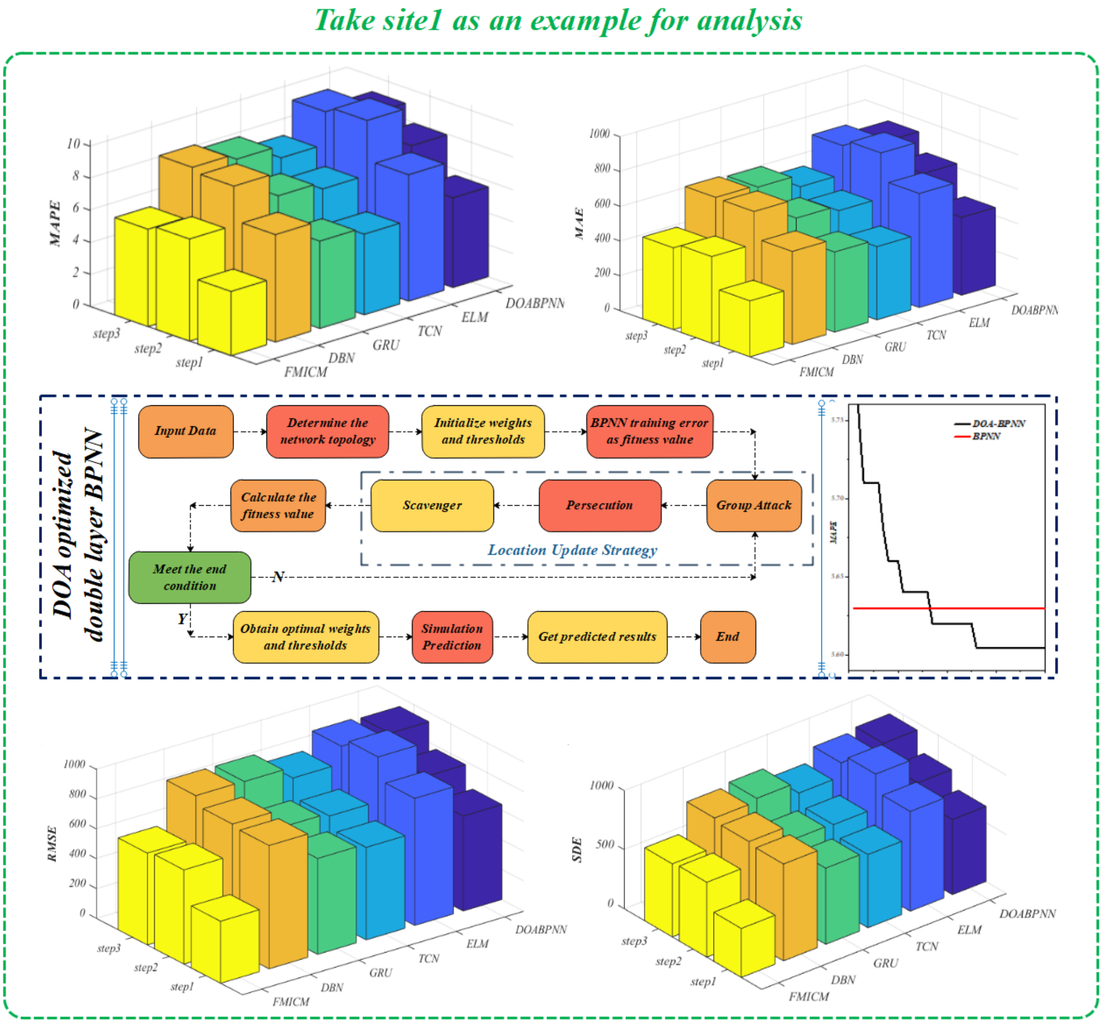

Figure 2 Comparison of the developed model with the single model of site1

(a) For site 1, when making a one-step prediction, the developing system shows a significant improvement in prediction accuracy and stability over the model executed with non-particleized data, $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step1\rangle} = 4.0179\%, \mathbf{SDE}_{\mathbf{site1}}^{\langle step1\rangle} = 416.638$. When making a two-step prediction,The developing FMICM: $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step2\rangle} = 6.385\%, \mathbf{MAE}_{\mathbf{site1}}^{\langle step2\rangle} = 498.132, \mathbf{RMSE}_{\mathbf{site1}}^{\langle step2\rangle} = 632.713, \mathbf{SDE}_{\mathbf{site1}}^{\langle step2\rangle} = 633.951$, which is an improvement com--pared to both the single model and the combined model performed without fuzzy granular data. While performing the three-step prediction, the accuracy improvement from particleization is more obvious, and the mean absolute percentage error of FMICM is $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step3\rangle} = 6.0869\%$, which is $\boldsymbol{\eta} = 1.7303\%$ higher than that of the combined model without particleization. In conclusion, for dataset I, the prediction accuracy of the developed integrated system is significantly better than that of the unparticleized model.

(b) For site2, FMICM has the lowest MAPE of $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step1\rangle} = 3.0992\%$ when making a one-step prediction.The highest prediction accuracy of the single model with unparticleized data is GRU with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step1\rangle} = 3.5191\%$, and FMICM improves the prediction accuracy by $\boldsymbol{\gamma} = 0.4199\%$ from both particleization and combination. During two step prediction, FMICM improves more in MAPE, but less in prediction stability. While performing the three-step prediction, the $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step3\rangle} = 4.9097\%$ for FMICM. In summary, for dataset two, the FMICM model improved the prediction accuracy of both the single model and the combined model for unparticleized data.

(c) For site3, when making a one-step prediction, the highest prediction accuracy of the unparticleized single model is TCN with $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step1\rangle} = 5.3836\%, \mathbf{MAE}_{\mathbf{site3}}^{\langle step1\rangle} = 380.518, \mathbf{RMSE}_{\mathbf{site3}}^{\langle step1\rangle} = 530.162, \mathbf{SDE}_{\mathbf{site3}}^{\langle step1\rangle} = 530.509$. FMICM has improved over all unparticleized models.During two step prediction, the prediction accuracy of the non-particleized models of MODOA_CM,GRU,TCN,DBN,DOA-BPNN,ELM and $\overrightarrow{\mathbf{MAPE}}_{\mathbf{site3}}^{\langle step2\rangle} = [7.51\%, 8.20\%, 8.68\%, 8.90\%, 9.34\%, 9.56\%]$ from low to high, respectively. While performing the three-step prediction, the unparticleized combined model has all improved over the single model with $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step3\rangle} = 8.39\%$, $\mathbf{MAE}_{\mathbf{site3}}^{\langle step3\rangle} = 550.45, \mathbf{RMSE}_{\mathbf{site3}}^{\langle step3\rangle} = 686.709, \mathbf{SDE}_{\mathbf{site3}}^{\langle step3\rangle} = 688.505$, but not as good as FMICM. In summary, for Dataset III, the combined FMICM model outperformed the unparticleized model in terms of prediction accuracy for any number of prediction steps.

Remark. Through Experiment 1, it was found that the developed FMICM outperformed the single and combined models performed on the unfuzzy granularized data, with the mean MAPE values of $\overrightarrow{\mathbf{MAPE}'}_{\mathbf{M}} = [3.9101\%, 5.6910\%, 6.3293\%]$ for the three-step prediction, respectively. In particular, by comparing FMICM with MODOA_CM, it was concluded that the necessity of using fuzzy particleization was effectively verified and FIG could not only improve the prediction accuracy but also the prediction stability. Figure 2 shows the measurements for the three datasets corresponding to Experiment 1.

4.3.2 Experiment II: Comparison with the single model after fuzzy particleization

Experiment 2 aims to verify the superiority of the multi-objective combinatorial optimization algorithm in FMICM, using the multi-objective combinatorial optimization algorithm to optimize the weights of the five single model point prediction results after fuzzy granulation in terms of both prediction accuracy and prediction stability, which is the role of the multi-objective optimization algorithm in FMICM, this experiment obtained five particleized single models DOABPNN,FIG_ELM,FIG_TCN, FIG_GRU,FIG_DBN)and FMICM The prediction results are shown in Table 6, and additional analyses of the experiments performed are described below.

(a) For site1, when making a one-step prediction, the best single-model prediction accuracy is FIG_TCN which $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step1\rangle} = 4.1258\%$, and the worst prediction accuracy is FIG_ELM which $\overline{\overline{\mathbf{MAPE}}}_{\mathbf{site1}}^{\langle step1\rangle} = 6.0396\%$. The multi-objective optimiza--tion algorithm improves the prediction accuracy and prediction stability of the single model. During two step prediction, FIG_GRU has the highest prediction accuracy in the single model with $\mathbf{MAPE}_{\mathbf{site1}}^{\langle step2\rangle} = 7.0318\%, \mathbf{SDE}_{\mathbf{site1}}^{\langle step2\rangle} = 694.01$.FIG_DBN has the best

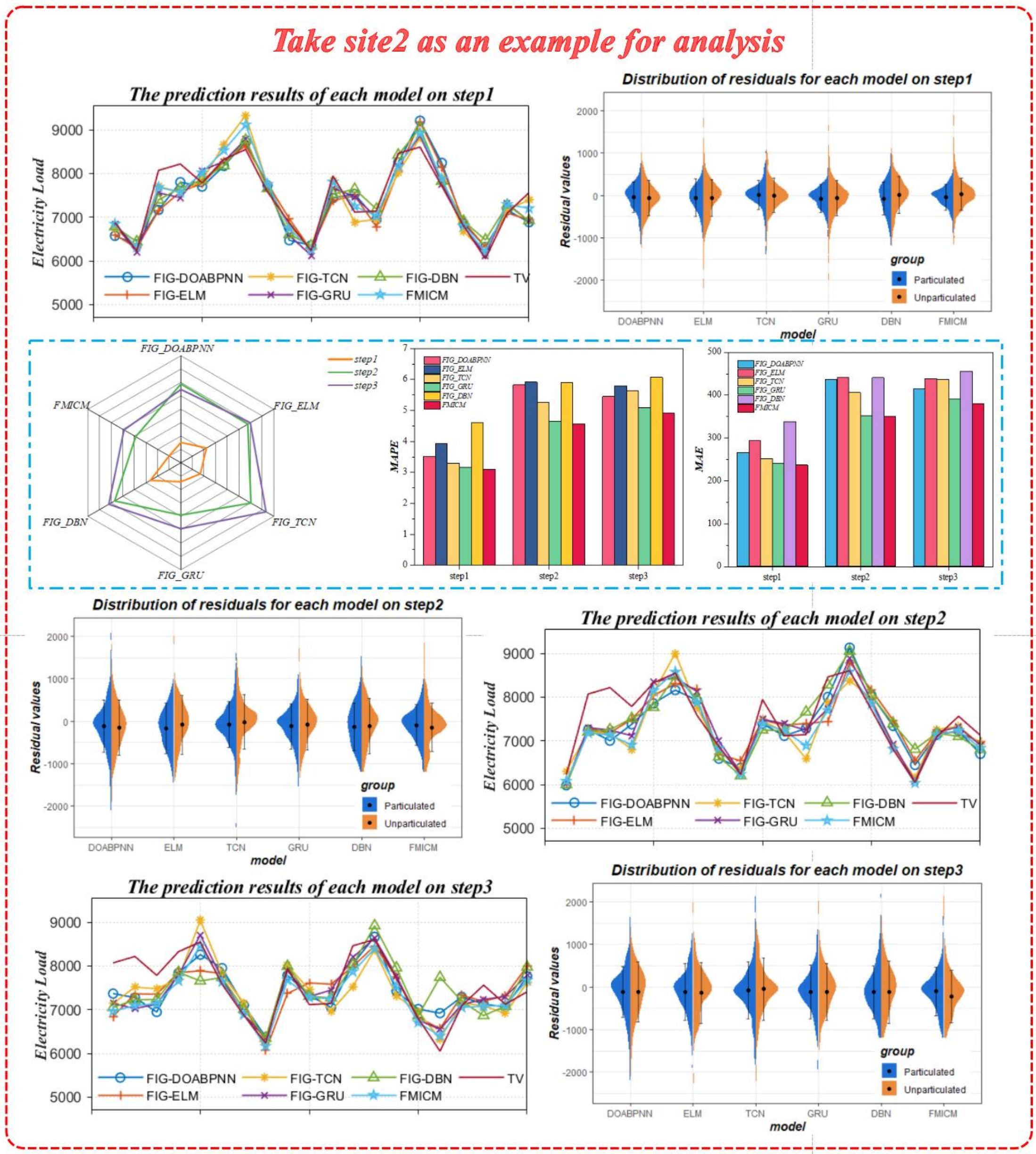

Figure 3 Comparison of point prediction performance of FMICM and different fuzzy post granulation single model

prediction stability with $\overline{\mathbf{MAPE}}_{\mathbf{site1}}^{\langle step2\rangle}=7.583\%, \overline{\mathbf{SDE}}_{\mathbf{site1}}^{\langle step2\rangle}=678.47$. While performing the three-step prediction, the prediction advantage of FMICM is more obvious, with MAPE optimizing $\vec{\mathbf{\Gamma}}=[1.8212\%, 2.7249\%, 0.9703\%, 0.6468\%, 1.198\%]$ over FIG_DOABPNN,FIG_ELM, FIG_TCN, FIG_GRU, and FIG_DBN, respectively. It can be seen that the multi-objective combined optimization algorithm in FMICM not only improves the prediction accuracy of the single model, but also improves the prediction stability.

(b) For site2, when making a one-step prediction, FMICM has the best prediction in the comparison with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step1\rangle}=3.0992\%, \mathbf{SDE}_{\mathbf{site2}}^{\langle step1\rangle}=350.911$ .During two step prediction, FIG_GRU has the lowest MAPE among the single models with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step2\rangle}=4.6607\%$ .While performing the three-step prediction, FIG_DBN has the worst prediction accuracy with $\overline{\mathbf{MAPE}}_{\mathbf{site2}}^{\langle step3\rangle}=6.0711\%$, and FIG_GRU has the highest prediction accuracy with $\mathbf{MAPE}_{\mathbf{site2}}^{\langle step3\rangle}=5.0823\%$. It can be concluded that the prediction accuracy of different models changes when the number of prediction steps changes, and the constant is that the prediction effect of FMICM is always higher than that of the single model.

(c) For site 3, when making a one-step prediction, FMICM has the highest prediction accuracy, $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step1\rangle}=4.6133\%$, followed by FIG_TCN and FIG_GRU with $\overrightarrow{\mathbf{MAPE}''}_{\mathbf{M}}=[4.8582\%, 4.9088\%]$. During two step prediction, the best prediction among the single models is FIG_GRU with $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step2\rangle}=6.4856\%$, $\mathbf{MAE}_{\mathbf{site3}}^{\langle step2\rangle}=443.829, \mathbf{RMSE}_{\mathbf{site3}}^{\langle step2\rangle}=563.344, \mathbf{SDE}_{\mathbf{site3}}^{\langle step2\rangle}=560.792$ .While performing the three-step prediction, FMICM has $\mathbf{MAPE}_{\mathbf{site3}}^{\langle step3\rangle}=7.9913\%$ and $\mathbf{SDE}_{\mathbf{site3}}^{\langle step3\rangle}=686.44$, and the prediction accuracy and prediction stability are greatly improved compared with all single models.In summary, the prediction accuracy of different single models in different datasets is different, but the constant is that the prediction accuracy of FMICM is lower than the five single models in all datasets.

Remark. It was found through Experiment 2 that FMICM was lower than different single models in all cases, although different single models had different predictions for different datasets in different prediction steps.It effectively verifies the importance of using MODOA for optimization weights in FMICM. Figure 3 illustrates the comparison between FMICM and the single model after fuzzy granulation using the three-step prediction of site2 as an example.

4.3.3 Experiment III: Comparison with different combinatorial optimization algorithms

Experiment 3 aims to verify the superiority of the multi-objective combinatorial optimization algorithm MODOA, using the common Multi-Objective Grasshopper Optimization Algorithm (MOGOA), Multi-Objective Dragonfly Algorithm (MODA), and Multi-objective Ant Lion Optimizer (MOALO) to optimize the weights of the five models to derive the prediction accuracy and compare with MODOA. The prediction results obtained from the experiments are shown in Table 7, and additional analyses of the experiments performed are described below.

(a) For site1, when making a one-step prediction, the MAPE, MAE, RMSE, and SDE of FMICM are smaller than those of MOGOA, MODA, and MOALO. Among the other three optimization algorithms, the prediction accuracy of MOALO and MOGOA is

Table 5

Point prediction performance evaluation of the developed model versus the unparticleized model

		Step1				Step2				Step3			
		MAPE (%)	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**
	DOABPNN	5.6046	448.026	637.542	635.344	7.9493	621.953	776.561	775.191	9.0131	727.162	935.608	937.986
	ELM	7.9007	655.939	853.257	844.484	10.381	810.961	1002.95	989.180	10.015	774.618	946.059	912.349
	TCN	5.0415	421.128	621.389	619.882	6.9572	549.579	703.333	701.069	7.9868	609.293	829.747	797.369
Site1	**GRU**	5.4580	461.103	644.100	644.587	7.3809	577.946	723.920	710.267	8.7570	678.069	903.301	892.642
	DBN	6.7190	535.295	828.036	819.946	8.8344	686.772	843.464	838.471	9.1004	689.645	905.715	866.442
	MODOA_CM	4.6975	388.392	569.255	570.523	6.5111	521.455	676.519	678.701	7.8172	597.190	800.841	777.130
	Proposed System	**4.0179**	**321.758**	**416.638**	**414.616**	**6.3850**	**498.132**	**632.713**	**633.951**	**6.0869**	**470.232**	**616.445**	**618.279**
	DOABPNN	4.1691	313.156	399.813	401.044	6.1514	464.755	589.673	589.824	6.5623	495.733	646.428	648.347
	ELM	4.3381	328.697	424.641	425.869	6.1716	463.347	589.192	590.616	6.6022	498.531	661.531	663.034
	TCN	3.8770	294.044	414.249	415.486	5.5674	416.095	558.469	560.053	6.2079	468.291	649.970	651.963
Site2	**GRU**	3.5191	263.902	356.898	357.721	5.2193	398.847	528.558	530.173	5.6214	423.709	589.395	591.293
	DBN	4.3528	327.198	416.877	411.459	6.4435	483.789	619.123	621.074	7.3595	550.387	725.378	727.727
	MODOA_CM	3.4443	259.162	355.354	352.189	4.8229	370.430	504.468	495.958	5.4243	417.229	576.385	562.387
	Proposed System	**3.0992**	**236.660**	**350.526**	**350.911**	**4.5679**	**350.803**	**493.272**	**494.553**	**4.9097**	**379.496**	**543.547**	**544.917**
	DOABPNN	6.6534	466.879	648.938	649.313	9.3363	625.151	807.708	793.534	10.653	715.515	855.930	838.011
	ELM	7.2865	500.857	646.628	646.959	9.5614	641.999	785.701	786.879	10.162	682.328	822.744	810.251
	TCN	5.3836	380.518	530.162	530.509	8.6810	572.062	745.951	722.123	8.8908	606.923	749.218	748.232
Site3	**GRU**	5.8537	406.919	533.777	518.127	8.2019	543.124	697.459	650.953	9.2796	624.933	794.546	755.997
	DBN	6.5548	455.728	568.781	559.712	8.8986	595.710	714.186	707.779	9.0915	621.257	745.608	742.357
	MODOA_CM	5.1344	364.044	489.874	491.512	7.5136	504.006	648.172	640.048	8.3931	580.929	686.709	688.505
	Proposed System	**4.6133**	**333.260**	**486.851**	**487.526**	**6.1200**	**430.872**	**557.819**	**558.744**	**7.9913**	**550.450**	**684.534**	**686.440**

Note: The above table shows the point prediction performance results (including MAPE, MAE, RMSE, SDE) using the developed combined prediction models and single models (including DOA_BPNN, TCN, DBN, ELM, GRU) without fuzzy particleization, using data for three-hour intervals.

Table 6
Evaluation of the point prediction performance of the developed models with different post-fuzzy granulation single models

		Step1				Step2				Step3			
		MAPE (%)	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**
Site1	**FIG_DOABPNN**	4.9684	399.265	532.508	522.476	8.2062	654.596	880.329	870.333	6.6247	506.672	653.973	592.987
	FIG_ELM	6.0396	471.739	588.403	519.548	9.1099	709.728	921.726	806.321	9.0080	687.948	884.319	758.503
	FIG_TCN	4.1258	328.608	427.564	417.030	7.3553	571.015	735.764	699.498	7.1392	540.848	728.087	685.458
	FIG_GRU	5.0158	396.544	517.905	475.442	7.0318	538.976	733.077	694.009	6.9224	520.653	768.293	727.059
	FIG_DBN	4.7983	379.957	488.176	463.817	7.5830	576.587	739.802	678.470	7.5961	577.573	765.189	733.006
	Proposed System	**4.0179**	**321.758**	**416.638**	**414.616**	**6.3850**	**498.132**	**632.713**	**633.951**	**6.0869**	**470.232**	**616.445**	**618.279**
Site2	**FIG_DOABPNN**	3.5169	265.705	374.444	375.622	5.8214	436.837	596.627	594.262	5.4540	414.121	573.203	573.727
	FIG_ELM	3.9383	294.740	409.041	410.411	5.9116	441.177	589.239	588.182	5.7952	438.026	596.774	598.634
	FIG_TCN	3.2850	251.480	381.929	383.208	5.2595	406.573	599.519	601.337	5.6436	436.958	664.450	666.480
	FIG_GRU	3.1566	240.894	371.356	370.819	4.6607	352.647	495.158	495.693	5.0823	390.853	545.409	546.343
	FIG_DBN	4.6063	337.851	431.639	431.669	5.9073	440.785	587.588	585.270	6.0711	455.991	609.970	610.172
	Proposed System	**3.0992**	**236.660**	**350.526**	**350.911**	**4.5679**	**350.803**	**493.272**	**494.553**	**4.9097**	**379.496**	**543.547**	**544.917**
Site3	**FIG_DOABPNN**	5.5916	394.488	545.716	547.371	10.147	694.438	845.432	847.188	8.9382	606.866	741.349	734.524
	FIG_ELM	8.3063	600.123	756.186	758.703	12.210	846.379	1086.97	1083.225	10.384	700.653	870.539	863.651
	FIG_TCN	4.8582	354.546	514.617	513.972	6.7052	476.987	626.195	625.655	8.2805	567.912	709.697	710.466
	FIG_GRU	4.9088	352.212	491.731	493.374	6.4856	443.829	563.344	560.792	8.8005	588.317	762.384	742.800
	FIG_DBN	7.2293	502.518	699.202	701.003	9.7904	656.282	808.539	775.814	10.258	684.650	830.367	774.567
	Proposed System	**4.6133**	**333.260**	**486.851**	**487.526**	**6.1200**	**430.872**	**557.819**	**558.744**	**7.9913**	**550.450**	**684.534**	**686.440**

Note: The above table shows the point prediction performance results (including MAPE, MAE, RMSE, SDE) of the developed combined prediction models and the fuzzy particleized single models (including FIG_DOABPNN,FIG_ELM,FIG_TCN,FIG_GRU,FIG_DBN).

higher, and the MAPE of FMICM has different degrees of improvement compared with them. During two step prediction, there is almost no difference in the prediction accuracy of the other three optimization algorithms, while the prediction accuracy of FMICM improves about $\ddot{\boldsymbol{\omega}} = 0.6111\%$ compared with these three algorithms. While performing the three-step prediction, the prediction accuracy of FMICM improves more, and the MAPE of FMICM, MOGOA, MODA, and MOALO are $\overrightarrow{\mathbf{MAPE}}_{\mathbf{site1}}^{\langle step3 \rangle} = [6.087\%, 6.635\%, 6.271\%, 6.652\%]$ .In summary, among the three compared optimization algorithms, for this dataset, MOGOA and MOALO are better at optimizing the first two steps of prediction, and MODA is better at optimizing the three steps of prediction, but neither is as good as not as good as FMICM.

(b) For site2, FMICM has the highest prediction accuracy when making a one-step prediction. The MAPE of the other three optimization algorithms is $\overrightarrow{\mathbf{MAPE'}}_{\mathbf{site2}}^{\langle step1 \rangle} = [3.1401\%, 3.1741\%, 3.1579\%]$. During two step prediction, the prediction accuracy of MODA and MOALO is higher with the exception of FMICM.

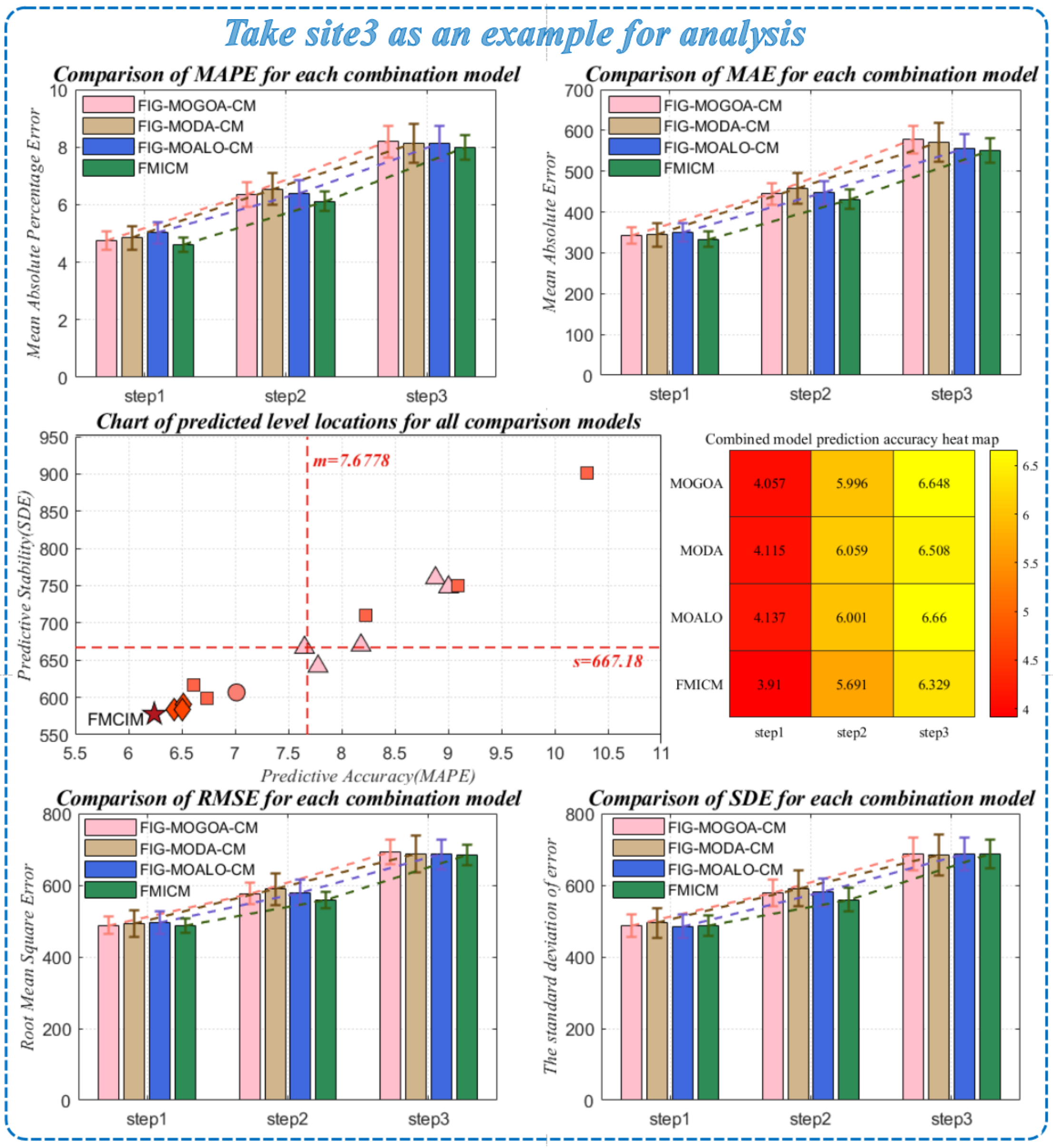

Figure 4 Comparison FMICM with other optimization models of site3

$\overrightarrow{\mathbf{MAPE}''}^{\langle step2\rangle}_{\mathbf{site2}} = [4.6295\%, 4.6367\%]$, $\overrightarrow{\mathbf{SDE}''}^{\langle step2\rangle}_{\mathbf{site2}} = [501.353, 499.359]$. While making the three-step prediction, FMICM has the highest prediction accuracy, MODA the second and MOALO the worst with $\overrightarrow{\mathbf{MAE}''}^{\langle step3\rangle}_{\mathbf{site2}} = [379.496, 390.295, 396.607]$. In summary. FMICM outperformed the three algorithms compared, despite the fact that the other optimization algorithms were sometimes strong and weak in their ability to optimize at different prediction steps.

(c) For site 3, when making a one-step prediction, the optimal of the other three optimization algorithms is MOGOA, $\mathbf{MAPE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 4.7607\%$, $\mathbf{MAE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 343.376$, $\mathbf{MAE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 487.886$, $\mathbf{SDE}^{\rho:\langle step1\rangle}_{\mathbf{site3}} = 487.579$. During two step prediction, the SDE of FMICM is the smallest, followed by MOGOA, MODA and MOALO, $\overrightarrow{\overrightarrow{\mathbf{SDE}}}^{\langle step2\rangle}_{\mathbf{site3}} = [558.744, 578.644, 590.548, 581.334]$. While performing the three-step prediction, The prediction accuracy of FMICM is significantly improved, and its MAE is $\vec{\boldsymbol{\chi}}_{\mathbf{MAE}} = [27.154, 20.665, 6.284]$ compared to MOGOA, MODA and MOALO. In conclusion, the optimization ability of FMICM in site3 is proved, and compared with the previous combined models, the prediction accuracy has been greatly improved compared to the previous combined models.

Remark. Through experiment three, it was found that the weight optimization ability of Multi-Objective Dingo Optimization Algorithm in FMICM model surpassed the known Multi-Objective Grasshopper Optimization Algorithm (MOGOA), Multi - Objective Dragonfly Algorithm (MODA), and Multi-objective Ant Lion Optimizer (MOALO), resulting in sufficient improvement of the final prediction accuracy and effectively validating the importance of MODOA in FMICM. Figure 4 shows how the developed FMICM compares with the combined model using different optimization algorithms.

4.3.4 Experiment IV: Comparison with all model interval estimates

The experiments in this section evaluate the interval estimation results by combining the evaluation metrics AIS for PI coverage probability and PI normalized averaged width and MPICD for evaluating the interval prediction accuracy, with the aim of comparing the developed FMICM model with a single model after fuzzy granulation and different combinations of optimization models to demonstrate that FMICM model is not only the best in point prediction, but also maintains excellent performance in interval estimation. The final test results are shown in Tables 8-9, and the details of this experiment are as follows.

(a) For site1,when making a one-step prediction, the PICP of FIG_ELM is as high as $\tilde{\boldsymbol{\rho}}_{1;1}^{\mathbf{ELM}} = 100\%$, but then the PIAW is as high as $\tilde{\boldsymbol{\omega}}_{1;1}^{\mathbf{ELM}} = 0.4772$, in other words, the high coverage of this model is due to the large PI normalized averaged width. Therefore, we mainly used AIS and MPICD for comparison. With a confidence factor of 95%, the optimal models for AIS in the three-step prediction are FIG_MOGOA_CM,FMICM,FMICM with AIS values of $\vec{\boldsymbol{\Lambda}}'_{\mathbf{FMICM}} = [-237.2, -400.9, -424.4]$. With a confidence factor of 90%, the optimal models for MPICD in the three-step prediction are FMICM,FMICM, FIG_MOGOA_CM, which have MPICD values of $\vec{\mathbf{D}}''_{\mathbf{MPICD}} = [384.94, 574.34, 531.74]$.Therefore, the interval prediction of FMICM in site1 is the best, followed by FIG_MOGOA_CM.

(b) For site2, FMICM performs best in the one-step prediction with $\tilde{\boldsymbol{\Lambda}}_{2;1}^{\mathbf{FMICM}} = -215.9$ and $\tilde{\mathbf{D}}_{2;1}^{\mathbf{FMICM}} = 255.51$ when the confidence coefficient is 95%. The best AIS in the two-step prediction is FMICM and the smallest MPICD is FIG_MODA_CM. The three-step prediction of FIG_GRU has an AIS of $\tilde{\boldsymbol{\Lambda}}_{2;3}^{\mathbf{GRU}} = -279.7$, which is better than the combined model, and the smallest MPICD is FIG_MOGOA_CM with a value of $\tilde{\mathbf{D}}_{2;3}^{\mathbf{MOGOA}} = 400.05$. The results are consistent with the above when the confidence factor is 90%. It is worth mentioning that the interval coverage of the single model here are higher than the combined model. The reason is that the residuals of the single model are larger, resulting in larger intervals obtained from the kernel density estimation curve. In summary, most experiments show that the interval prediction of FMICM is better than other comparative models.

(c) For site3, both AIS and MPICD for the 95% confidence interval of FIG_MOGOA_CM was optimal in the one-step prediction case with $\tilde{\boldsymbol{\Lambda}}_{3;1}^{\mathbf{MOGOA}} = -268.7$ and

$\tilde{\mathbf{D}}_{3;1}^{\mathbf{MOGOA}} = 339.87$. In the two-step prediction case, both AIS and MPICD for the 95% confidence interval of FMICM were optimal with $\tilde{\mathbf{\Lambda}}_{3;2}^{\mathbf{FMICM}} = -255.5, \tilde{\mathbf{D}}_{3;2}^{\mathbf{FMICM}} = 427.1$. FIG-DOABPNN emerges as the best in the three-step prediction with an AIS of $\tilde{\mathbf{\Lambda}}_{3;3}^{\mathbf{D-BPNN}} = -323$, which is better than all types of combined models. When the confidence coefficient is equal to 90%, FMICM performs optimally in all three prediction steps with AIS of $\vec{\mathbf{\Lambda}}''_{\mathbf{FMICM}} = [-499.5, -435.3, -495.4]$, and MPICD of $\vec{\mathbf{D}}''_{\mathbf{FMICM}} = [333.45, 428.44, 550.09]$. In summary, the experiments for dataset three show that the interval prediction of FMICM is better than other comparative models.

Remark. The interval predictions of FMICM were compared with those of eight models by Experiment 4. At $\boldsymbol{\lambda}' = 95\%$ confidence factor, 5/9 experiments proved that FMICM has the best AIS and MPICD. 89% experiments proved that FMICM has higher

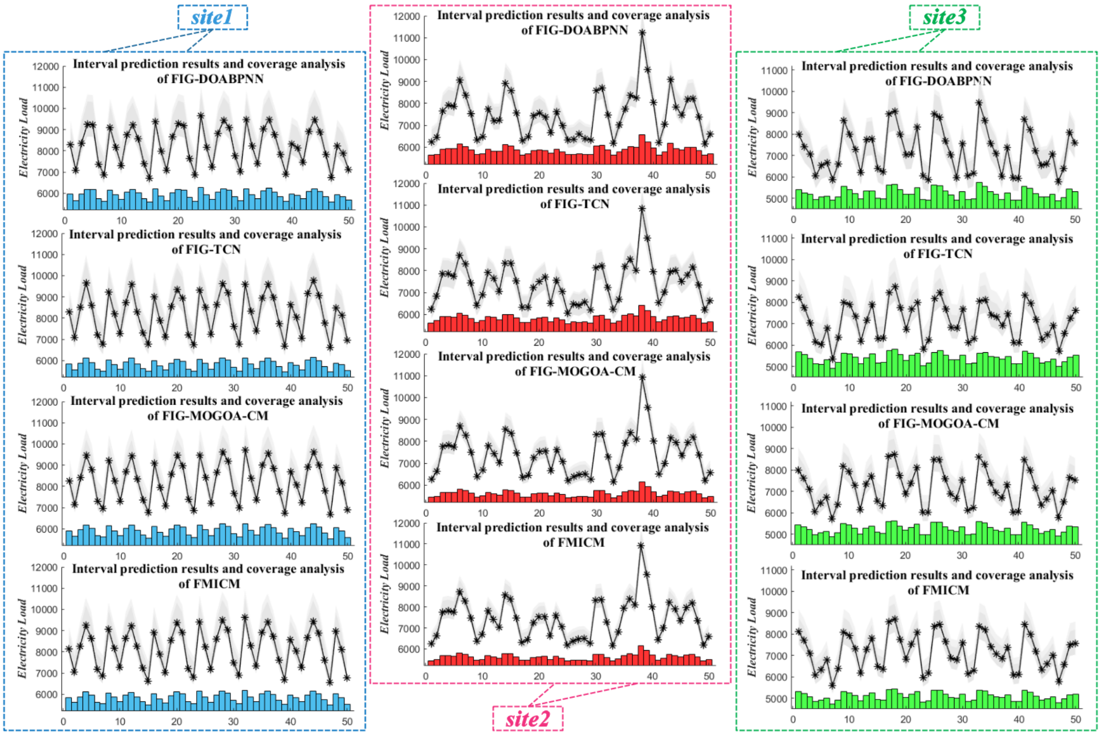

Figure 5 The interval prediction of the developed FMICM with other models

Table 7
Combined model point prediction performance table using different optimization algorithms

		Step1				**Step2**				**Step3**			
		MAPE (%)	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**	**MAPE (%)**	**MAE**	**RMSE**	**SDE**
Site1	**FIG _MOGOA_CM**	4.2704	338.142	432.497	**405.099**	6.9922	535.377	693.019	638.555	6.6352	496.552	695.786	625.786
	FIG _MODA_CM	4.3205	342.026	438.719	409.726	7.0011	535.801	696.299	640.186	6.2708	474.810	655.924	631.960
	FIG _MOALO_CM	4.2302	335.487	430.500	405.677	6.9951	535.340	693.937	639.094	6.6521	497.727	697.229	625.518
	Proposed System	**4.0179**	**321.758**	**416.638**	414.616	**6.3850**	**498.132**	**632.713**	**633.951**	**6.0869**	**470.232**	**616.445**	**618.279**
Site2	**FIG _MOGOA_CM**	3.1401	239.097	351.361	352.149	4.6525	355.409	504.878	506.311	5.1226	391.791	554.369	552.644
	FIG _MODA_CM	3.1741	241.407	351.748	352.879	4.6295	353.303	499.822	501.353	5.1083	390.295	545.312	546.086
	FIG _MOALO_CM	3.1579	240.004	351.406	352.575	4.6367	353.405	497.981	499.359	5.1994	396.607	558.387	557.081
	Proposed System	**3.0992**	**236.660**	**350.526**	**350.911**	**4.5679**	**350.803**	**493.272**	**494.553**	**4.9097**	**379.496**	**543.547**	**544.917**
Site3	**FIG _MOGOA_CM**	4.7607	343.376	487.886	487.579	6.3440	444.325	576.726	578.644	8.1858	577.604	693.029	685.887
	FIG _MODA_CM	4.8519	344.288	493.998	495.394	6.5449	457.369	588.995	590.548	8.1434	571.115	687.497	**685.29**
	FIG _MOALO_CM	5.0236	350.271	495.783	**485.023**	6.3708	446.596	579.396	581.334	8.1275	556.734	686.473	686.609
	Proposed System	**4.6133**	**333.260**	**486.851**	487.526	**6.1200**	**430.872**	**557.819**	**558.744**	**7.9913**	**550.450**	**684.534**	686.440

Note: The above table shows the point prediction evaluation results (using four metrics MAPE, MAE, RMSE, SDE) of the developed FMICM(Proposed System) optimized using MODOA in combination with models using other three different optimization algorithms in combination (including FIG_MOGOA_CM, FIG_MODA_CM, FIG_ MOALO_CM).

Table 8

Comparison of interval predictions of the development system with other models at a confidence coefficient of 0.95.

α = 0.05		**Step1**				**Step2**				**Step3**			
		PICP (%)	**PIAW**	**AIS**	**MPICD**	**PICP (%)**	**PIAW**	**AIS**	**MPICD**	**PICP (%)**	**PIAW**	**AIS**	**MPICD**
	FIG_DOABPNN	87.33	0.2390	-346.1	541.17	85.33	0.3726	-548.1	830.87	95.33	0.3931	-342.2	590.77
	FIG_ELM	1	0.4772	-379.9	511.01	99.33	0.5878	-474.1	729.26	93.33	0.4739	-426.5	777.67
	FIG_TCN	89.33	0.2236	-259.0	439.75	93.33	0.4171	-433.7	673.84	92.67	0.3750	-428.3	613.69
	FIG_GRU	84.67	0.2276	-337.3	506.87	87.33	0.3299	-521.3	670.47	93.33	0.3899	-511.3	592.29
Site1	**FIG_DBN**	95.33	0.2457	-255.6	431.55	86.67	0.3535	-466.4	728.09	94.67	0.3966	-422.2	613.89
	FIG _MOGOA_CM	92.00	0.2366	**-237.2**	415.53	90.00	0.3332	-416.5	**626.09**	92.67	0.3317	-438.3	555.39
	FIG _MODA_CM	92.67	0.2306	-240.2	411.18	88.67	0.3302	-422.5	626.82	92.67	0.3393	-435.6	560.27
	FIG _MOALO_CM	88.67	0.2101	-279.3	460.88	88.67	0.3283	-423.9	628.74	92.67	0.3367	-435.3	554.08
	Proposed System	92.67	0.2296	-239.6	**408.34**	90.00	0.3413	**-400.9**	626.81	92.67	03414	**-424.4**	**553.54**
	FIG_DOABPNN	94.00	0.1833	-222.0	305.88	94.00	0.3422	-304.8	443.19	96.67	0.3301	-301.1	487.83
	FIG_ELM	94.00	0.2003	-216.1	327.51	95.33	0.3331	-313.8	495.37	96.67	0.3294	-289.5	501.27
	FIG_TCN	94.00	0.1783	-235.8	258.78	90.67	0.2797	-340.9	446.96	94.67	0.3620	-362.9	463.10
	FIG_GRU	92.00	0.1645	-239.1	289.01	94.67	0.3064	-272.0	500.15	97.33	0.3408	**-279.7**	452.73
Site2	**FIG_DBN**	92.67	0.1909	-222.8	356.32	95.33	0.2873	-310.4	447.15	94.00	0.3131	-285.1	464.06
	FIG _MOGOA_CM	91.33	0.1505	-219.5	266.17	95.33	0.2460	-288.7	362.48	95.33	0.2854	-296.3	**400.05**
	FIG _MODA_CM	92.67	0.1508	-219.8	264.32	94.67	0.2443	-286.2	**361.11**	95.33	0.2848	-287.8	408.78
	FIG _MOALO_CM	90.67	0.1498	-220.4	268.33	94.67	0.2436	-284.4	361.27	94.67	0.2825	-298.8	403.46
	Proposed System	92.00	0.1507	**-215.9**	**255.51**	94.00	0.2431	**-259.1**	390.42	96.00	0.2899	-290.9	401.96
	FIG_DOABPNN	86.00	0.2403	-353.7	408.10	86.67	0.407	-347.3	699.04	99.33	0.4574	**-323.0**	599.50
	FIG_ELM	89.33	0.3548	-337.5	610.36	91.33	0.4997	-581.4	855.78	97.33	0.5418	-409.3	701.07
	FIG_TCN	93.33	0.2644	-273.1	398.16	1	0.4663	-327.4	473.94	1	0.5322	-372.4	568.26
	FIG_GRU	87.33	0.2119	-293.8	360.33	1	0.3647	-256.1	442.90	98.00	0.5202	-397.8	586.99
Site3	**FIG_DBN**	83.33	0.2929	-471.2	549.93	94.00	0.4243	-317.3	626.12	99.33	0.4987	-351.0	669.06
	FIG _MOGOA_CM	92.00	0.2438	**-268.7**	**339.87**	99.33	0.3788	-266.0	443.35	1	0.5039	-352.5	557.60
	FIG _MODA_CM	92.00	0.2485	-272.7	348.78	99.33	0.3804	-268.1	457.12	1	0.5055	-353.6	557.88
	FIG _MOALO_CM	92.00	0.2415	-269.9	340.13	98.67	0.3697	-260.9	445.61	99.33	0.5127	-361.4	556.51
	Proposed System	90.00	0.2234	-281.7	361.57	99.33	0.3626	**-255.5**	**427.10**	99.33	0.5165	-363.9	**548.99**

Table 9
Comparison of interval predictions of the development system with other models at a confidence coefficient of 0.9.

	α = 0.1	Step1				Step2				Step3			
		PICP (%)	PIAW	AIS	MPICD	PICP (%)	PIAW	AIS	MPICD	PICP (%)	PIAW	AIS	MPICD
	FIG_DOABPNN	80.00	0.1845	-556.6	496.33	84.00	0.3257	-831.4	746.20	92.00	0.3234	-634.9	556.88
	FIG_ELM	96.67	0.2979	-511.1	498.87	95.33	0.4699	-804.0	726.26	91.33	0.3893	-747.3	727.72
	FIG_TCN	88.00	0.1903	-415.5	387.31	86.00	0.2878	-691.5	628.67	90.00	0.2968	-696.2	592.99
	FIG_GRU	80.00	0.1786	-515.6	457.67	78.67	0.2468	-794.3	602.68	88.67	0.3018	-781.5	572.76
Site1	FIG_DBN	85.33	0.1727	-420.3	396.73	85.33	0.2883	-687.9	613.10	91.33	0.3228	-688.9	598.23
	FIG _MOGOA_CM	78.67	0.1451	-428.0	385.68	83.33	0.2412	-672.0	**574.34**	90.00	0.2894	-679.7	**531.74**
	FIG _MODA_CM	82.00	0.1556	-410.6	387.88	82.67	0.2364	-683.9	575.23	89.33	0.2815	-681.5	538.40
	FIG _MOALO_CM	0.82	0.1573	-421.0	393.45	82.67	0.2333	-685.0	575.49	90.00	0.2929	-678.8	531.28
	Proposed System	86.00	0.1603	**-403.9**	**384.94**	82.67	0.2373	**-664.9**	575.92	90.00	0.2911	**-671.4**	532.50
	FIG_DOABPNN	90.67	0.1474	-347.7	300.31	90.00	0.2433	-517.7	443.14	92.67	0.2680	-499.9	480.55
	FIG_ELM	84.67	0.1505	-365.9	330.78	86.00	0.2291	-524.9	494.65	87.33	0.2595	-510.5	483.88
	FIG_TCN	91.33	0.1573	-361.8	253.30	85.33	0.2109	-572.8	428.96	90.67	0.2758	-621.1	486.38
	FIG_GRU	88.00	0.1239	-360.3	269.62	88.00	0.2076	-449.6	425.01	92.67	0.2592	**-469.3**	427.80
Site2	FIG_DBN	84.67	0.1612	-370.9	352.56	88.00	0.2179	-496.4	441.08	85.33	0.2536	-517.8	463.25
	FIG _MOGOA_CM	84.67	0.1152	-331.6	253.54	89.33	0.1890	-453.4	361.22	91.33	0.2337	-489.2	421.54
	FIG _MODA_CM	86.00	0.1162	-332.0	253.67	90.00	0.1877	-447.3	**359.33**	92.00	0.2342	-478.5	**411.40**
	FIG _MOALO_CM	85.33	0.1140	-336.6	253.89	90.67	0.1878	-443.6	359.28	90.67	0.2360	-494.1	419.89
	Proposed System	85.33	0.1151	**-330.2**	**252.05**	86.67	0.1861	**-432.9**	381.49	92.00	0.2305	-479.9	411.89
	FIG_DOABPNN	84.00	0.2090	-539.2	396.99	71.33	0.2892	-788.2	696.21	67.33	0.2340	-716.0	615.05
	FIG_ELM	79.33	0.2663	-642.8	604.86	70.00	0.3292	-1084	840.85	80.67	0.3501	-728.1	691.83
	FIG_TCN	81.33	0.1638	-532.9	354.03	92.00	0.3155	-506.2	473.32	90.00	0.3168	-521.7	567.99
	FIG_GRU	79.33	0.1672	-510.8	352.64	90.00	0.2714	-454.2	443.59	86.00	0.2820	-574.7	589.60
Site3	FIG_DBN	80.00	0.2382	-713.0	502.99	78.67	0.3019	-651.8	626.05	76.00	0.2786	-610.8	660.34
	FIG _MOGOA_CM	78..67	0.1538	-502.1	338.86	90.00	0.2771	-448.6	445.50	79.33	0.2662	-517.3	563.2
	FIG _MODA_CM	78.00	0.1545	-516.4	344.20	85.33	0.2573	-460.5	460.69	81.33	0.2679	-512.8	563.21
	FIG _MOALO_CM	78.67	0.1538	-502.8	338.64	90.00	0.2743	-449.5	447.87	80.67	0.2693	-513.2	559.11
	Proposed System	80.67	0.15607	**-499.5**	**333.45**	92.00	0.2788	**-435.3**	**428.44**	90.00	0.3031	**-495.4**	**550.09**

interval prediction accuracy at $\lambda'' = 90\%$ confidence factor. Additional individual experiments showed that FIG_ MOGOA_CM and FIG_MODA_CM had better MPICD. It can be concluded from the interval prediction tests that the developed FMICM model proved to have excellent interval prediction performance in most of the experiments at the significance level in the experiments. Figure 5 shows how the developed FMICM compares to the eight models in terms of interval prediction.

5. Discussion

In this section, we further analyze the prediction results of four experiments, including the following four main components: Diebold-Mariano (DM)-test, improvement ratio of the indexes, forecasting effectiveness test, sensitivity analysis, convergence analysis and the empirical power load analysis. The detailed testing procedures are described below.

5.1 Diebold-Mariano (DM)-test

Since there are only a few data in the test set in the experiment, comparison of the prediction results can only indicate that the combined model proposed in this sample works better, and the data sampling is not good enough to cause this situation, in order to determine whether it is a fluke caused by the situation, the difference between model A and model B needs to be calculated statistically to be significant, that is a DM test.

Definition 1: Suppose the predicted values of the two models to be compared are $\overrightarrow{\mathbf{Pf}}_1 = \left[\mathbf{Pf̂}_1^{(1)}, \mathbf{Pf̂}_1^{(2)}, \cdots, \mathbf{Pf̂}_1^{(\phi)}\right]$ and $\overrightarrow{\mathbf{Pf}}_2 = \left[\mathbf{Pf̂}_2^{(1)}, \mathbf{Pf̂}_2^{(2)}, \cdots, \mathbf{Pf̂}_2^{(\phi)}\right]$, and the true values are $\overrightarrow{\mathbf{To}}_* = \left[\mathbf{To\hat{}}_*^{(1)}, \mathbf{To\hat{}}_*^{(2)}, \cdots, \mathbf{To\hat{}}_*^{(\phi)}\right]$. From this, the prediction error of the two models to be compared can be calculated as $\overrightarrow{\mathbf{Er\hat{}}}_1 = \left[\mathbf{Er\hat{}}_1^{(1)}, \mathbf{Er\hat{}}_1^{(2)}, \cdots, \mathbf{Er\hat{}}_1^{(\phi)}\right] \mid \mathbf{Er\hat{}}_1^{(\boldsymbol{\alpha})} = \mathbf{Pf̂}_1^{(\boldsymbol{\alpha})} - \mathbf{To\hat{}}_*^{(\boldsymbol{\alpha})}$, $\overrightarrow{\mathbf{Er\hat{}}}_2 = \left[\mathbf{Er\hat{}}_2^{(1)}, \mathbf{Er\hat{}}_2^{(2)}, \cdots, \mathbf{Er\hat{}}_2^{(\phi)}\right] \mid \mathbf{Er\hat{}}_2^{(\boldsymbol{\alpha})} = \mathbf{Pf̂}_2^{(\boldsymbol{\alpha})} - \mathbf{To\hat{}}_*^{(\boldsymbol{\alpha})}$.

Based on the above preparatory work, the null hypothesis and alternative hypothesis are presented.

$$
\begin{aligned}
H_0 &: \mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_1^{(\boldsymbol{\mu})}\right)\right] - \mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_2^{(\boldsymbol{\mu})}\right)\right] = 0 \\
H_1 &: \mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_1^{(\boldsymbol{\mu})}\right)\right] - \mathbf{E}\left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_2^{(\boldsymbol{\mu})}\right)\right] \neq 0
\end{aligned}
\tag{1}
$$

Where the loss function $\boldsymbol{\Omega}(\vec{\chi})$ is calculated as $\boldsymbol{\Omega}(\vec{\chi}) = \vec{\chi}^2$, and the constructed DM test statistic is:

$$
\mathbf{DM} = \frac{\sum_{\boldsymbol{\mu}=1}^{\boldsymbol{\Pi}} \left[\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_1^{(\boldsymbol{\mu})}\right) - \boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_2^{(\boldsymbol{\mu})}\right)\right]}{\boldsymbol{\Pi}\sqrt{S^2/\boldsymbol{\Pi}}}
\tag{2}
$$

Where S^2 refers to the variance of $\boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_1^{(\boldsymbol{\mu})}\right) - \boldsymbol{\Omega}\left(\overrightarrow{\mathbf{Er\hat{}}}_2^{(\boldsymbol{\mu})}\right)$. The DM test theory assumes that the distribution of the DM test statistic satisfies the standard normal distribution when the significance level is set to α, so the rejection domain is $W = \left\{|\mathbf{DM}| > |z_{\alpha/2}|\right\}$. When the DM statistic falls into the rejection domain, the original hypothesis is rejected, that is, there is a significant difference between the two prediction models,otherwise when $|\mathbf{DM}| \le |z_{\alpha/2}|$, there is no reason to reject the original hypothesis, which means that there is no statistically significant difference in the predictive power of the two models.

The DM test computes the predictive validity of this integrated system point estimate and further validates the performance of the combined model against statistical ideas. The test results are shown in Table 10, and other details are shown below.

(a) Comparison with the single model, when the significance level is set to $\alpha' = 0.05$, it can be seen that the majority of DM values are greater than $\bar{\mathbf{z}} = 1.96$, rejecting the original hypothesis that the developed point prediction system is better than the single model before fuzzy particleization. Setting the significance level to $\alpha' = 0.05$, DOABPNN,ELM,TCN,GRU, and DBN in the single model had DM test pass rates of $\overrightarrow{\mathbf{PR}'} = [[67\%, 100\%, 56\%, 56\%, 100\%]$. When the significance level was set to $\alpha'' = 0.1$,The DM test pass rates of DOABPNN, ELM, TCN, GRU, and DBN were $\overrightarrow{\mathbf{PR}''} = [100\%, 100\%, 78\%, 56\%, 100\%]$. In summary, the single model before fuzzy particleization is significantly different from FMICM. Since the DM values are all greater than 0, it indicates that the point prediction effect of FMICM is better than that of the single model before fuzzy particleization, which verifies the conclusion drawn in Experiment 1.

(b) Compared to the single model after fuzzy particleization, $\mathbf{PR}' = 73\%$ of the data passed the test when the significance level was $\alpha' = 0.05$, with FIG_DBN passing all of them. The test pass rate for FIG_DOABPNN was $\mathbf{PR}'_{\mathbf{BPNN}} = 56\%$ at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{BPNN}} = 67\%$ at $\alpha'' = 0.1$. The test pass rate for FIG_ELM was $\mathbf{PR}'_{\mathbf{ELM}} =$ 89% at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{ELM}} = 89\%$ at $\alpha'' = 0.1$. The pass rate of FIG_GRU is $\mathbf{PR}'_{\mathbf{GRU}} = 33\%$ at a= $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{GRU}} = 44\%$ at $\alpha'' = 0.1$ .The pass rate of FIG_TCN is $\mathbf{PR}'_{\mathbf{TCN}} = 56\%$ at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{TCN}} = 67\%$ at $\alpha'' = 0.1$.The pass rate of FIG_DBN is $\mathbf{PR}'_{\mathbf{DBN}} = 100\%$ at $\alpha' = 0.05$ and $\mathbf{PR}''_{\mathbf{DBN}} = 100\%$ at $\alpha'' = 0.1$. In summary, most of the models completely passed the DN test, and some of them failed the DM test due to the different data sets. Overall, the DM values of the single model after fuzzy particleization were all greater than 0 unlike FMICM, indicating that the point prediction of FMICM was better than that of the single model after fuzzy particleization, which verified the conclusion reached in Experiment 2.

(c) Compared with different optimization models, the DM test pass rate of the three optimization combination models in site1 is $\mathbf{PR}'' = 89\%$ when the significance level is $\alpha'' = 0.1$, and only the DM value of FIG_MOALO_CM is $\overline{\mathbf{PR}''}_{\mathbf{MOALO}} = 1.4498$. Most of the DM values in site2 are less than 1 and do not pass the test. Step2 in site3 all pass the significance level of $\alpha' = 0.05$ DM test, while the other step predictions did not pass the test. However, it seems that FMICM is significantly different from the three optimization models, and the DM values are all greater than 0. This indicates that the prediction effect of FMICM is better than the other three combined optimization models, which verifies the conclusion drawn in Experiment 3.

5.2 Improvement ratio of the indexes

After the DM test, it can be concluded that the proposed FMICM has significant differences with the single model, the single model after fuzzy particleization and different combined optimization models, in addition, based on the DM value greater than zero can be deduced that FMICM is better than other models.Therefore, the DM test can only qualitatively infer that FMICM is superior to other models, but quantitatively analyze it. Therefore, this section proposes to conduct the indicator improvement rate test with the purpose of further quantitatively indicating the superiority of FMICM based on the DM test to specifically improve MAPE is an important evaluation index to measure the prediction effect of time series data, so MAPE is used as the indicator improvement rate index in this paper.The calculation formula of indicator improvement rate is shown in Equation (5).

$$\mathbf{IR}_{\mathbf{MAPE}} = \left| \frac{Compared_{MAPE} - FMICM_{MAPE}}{Compared_{MAPE}} \right| \times 100\% \tag{5}$$

The point predictions of the developed integrated system were tested against a single model, a single model after fuzzy particleization, and different combined optimization models for metric improvement rates,and the final test results are shown in Table 11, the details are as follows:

(a) The proposed FMICM model was compared with the single model, where the most improved model was ELM with IR of $\vec{\mathbf{I}}_{\Delta} = [38.13\%, 33.49\%, 28.74\%]$ for the three prediction steps, and the least improved models were TCN and GRU with an average index improvement rate of $\mathbf{I}_{\Delta}^{(tcn)} = 18.3522\%$ for TCN and $\mathbf{I}_{\Delta}^{(gru)} = 18.6554\%$ for GRU, also side by side, it shows the high prediction accuracy of these two models. In general, the average index improvement rate of FMICM for a single model is around $\Delta\varpi = 25\%$, which is a large improvement. It indicates that FIG plays a role in improving the prediction accuracy in the point prediction of the system.

(b) Compared with the single model after fuzzy particleization, the average index improvement rates of FIG_DOABPNN, FIG_ELM, FIG_TCN, FIG_GRU, FIG_DBN is $\vec{\mathbf{I}}_{\Delta} = [17.85\%, 30.28\%, 8.85\%, 7.69\%, 24.69\%]$.Therefore, the highest improvement rate is FIG_ELM , with a three-step predicted average index improvement rate of $\vec{\mathbf{I}}'^{(elm)}_{\Delta} = [33.08\%, 34.17\%, 23.58\%]$.The lowest improvement rate is FIG_GRU with multi-step predicted average indicator improvement rates of $\vec{\mathbf{I}}''^{(gru)}_{\Delta} = [9.24\%, 5.61\%, 8.22\%]$. Also the average index improvement rate for FIG_DOABPNN is $\mathbf{I}_{\Delta}^{(bp)} = 17.85\%$, for FIG_TCN is $\mathbf{I}_{\Delta}^{(tcn)} = 8.85\%$, and for FIG_DBN is $\mathbf{I}_{\Delta}^{(dbn)} = 24.69\%$. Overall, the average index improvement rate of FMICM for the single model after fuzzy particleization is $\delta = 17.87\%$, which is a large improvement. This also reflects that MODOA can improve the prediction accuracy in the system.

(c) The index improvements relative to the FIG_MOGOA_CM, FIG_MODA_CM, and FIG_MOALO_CM are $\vec{\mathbf{I}}^{*}_{\Delta} = [4.35\%, 4.40\%, 4.99\%]$. It can be seen that the optimization capability of the MODOA algorithm has been improved to different degrees compared with the other three multi-objective optimization algorithms. In summary, the index improvement rate test shows that the accuracy of the proposed integrated system point prediction is significantly improved over the single model,which also reflects that the combined model can improve the prediction accuracy. The significant improvement in comparison with the unparticleized single model indicates that FIG is important for accuracy improvement. The significant improvements for different combined models indicate that MODOA is better than other optimization algorithms.

5.3 Forecasting Effectiveness

In addition to the accuracy of the forecast results, the size of the difference between the forecast results and the true values, the skewness and kurtosis of the distribution of the forecast results, should also be considered in point forecasting. Forecasting effectiveness is then an indicator to verify this. The calculation principle is as follows.

Define $\mathbf{W}_n = 1 - |\gamma_n|$ as the prediction accuracy, where

$$\gamma_n = \begin{cases} -1 & (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n < 1 \\ (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n & -1 \le (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n < 1 \\ 1 & (\mathbf{TOV}_n - \mathbf{PFV}_n)/\mathbf{TOV}_n > 1 \end{cases} \quad (6)$$

Based on the prediction accuracy $\mathbf{W}_n$ can calculate the k-order prediction effective element, which is calculated as follows.

$$\psi^{(k)} = \sum_{n=1}^{N} \vartheta_n \gamma_n^{(k)}, \sum_{n=1}^{N} \vartheta_n = 1 \quad (7)$$

Here, ϑ_n denotes that the probability distribution at a point in time is discrete. Since we do not have access to prior information about the probability distribution, we identify it as 1 and set ϑ_n as $\vartheta_n = 1/\mathbf{N}, n = 1, 2, \ldots, \mathbf{N}$, $\mathbf{C}$ is a continuous function of the k-order forecasting effectiveness component, $\mathbf{C}\left(\psi^{(1)}, \psi^{(2)}, \cdots, \psi^{(k)}\right)$ is defined as the k-order prediction effective.

This section uses the one-order prediction effective and the two-order prediction effective, the calculation of the one-order predictive validity is described in Equation (8).

$$\mathbf{C}\left(\psi^{(1)}\right) = \psi^{(1)} \quad (8)$$

There is a second-order predictive validity showing the disparity among the expected standard deviations, which is described in Equation (9).

$$\mathbf{C}\left(\psi^{(1)}, \psi^{(2)}\right) = \psi^{(1)}\left(1 - \sqrt{\psi^{(2)} - \left(\psi^{(1)}\right)^2}\right) \quad (9)$$

The proposed combined model was tested for predictive validity with the single model, the single model after fuzzy granulation and the different combined models, and the final test results are shown in Table 12, and the details of this experiment are as follows.

(a) For the one-order prediction effective, the best results were obtained for the newly proposed FMICM model, with the mean values of $\vec{\mathbf{F}}_{\mathbf{site1}}^{\langle 1 \rangle} = [95.98\%, 93.61\%, 93.91\%]$ for the three-step predictive FE of site1. For site2, the highest FE value was obtained for GRU in the one-order model, with an average one-order prediction effective of $\mathbf{F}_{\mathbf{site2}}^{gru\langle 1 \rangle} = 95.21\%$. Site3 had the highest FE value for FIG_TCN in the granulated one-order model, with an average one-order prediction effective of $\mathbf{F}_{\mathbf{site3}}^{tcn\langle 1 \rangle} = 93.38\%$. The other three combined models had one-order prediction effective of $\overrightarrow{\mathbf{Fo}}_{\mathbf{site3}}^{\langle 1 \rangle} = 94.34\%$, $94.44\%, 94.4\%$ on average.

(b) For the two-order prediction effective, the newly proposed FMICM model is still the best with the two-order values of $\vec{\mathbf{F}}_{\mathbf{site1}}^{\langle 2 \rangle} = [92.85\%, 88.70\%, 88.48\%]$, respectively, $\vec{\mathbf{F}}_{\mathbf{site2}}^{\langle 2 \rangle} = [93.92\%, 91.57\%, 90.77\%]$, $\vec{\mathbf{F}}_{\mathbf{site3}}^{\langle 2 \rangle} = [91.05\%, 89.14\%, 87.21\%]$ for the three sites,The model with the smallest two-order value was DBN with a second-order mean of $\mathbf{F}_{*}^{dbn\langle 2\rangle} = 86.67\%$, the best performing single model after fuzzy granulation was FIG_TCN with a two-order mean of $\mathbf{F}_{*}^{tcn\langle 2\rangle} = 89.04\%$, and the best performing

Table 10
Results of Diebold Mariano (DM) Test

	Site1			Site2			Site3		
	Step1	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**
DOABPNN	1.8285	3.0855	1.9492	1.8386	2.5803	2.3644	4.3238	4.1836	4.7186
ELM	4.9253	5.1667	3.0496	3.3037	2.8385	2.5105	5.5472	3.1028	3.7168
TCN	2.7372	1.0698	1.5163	2.6599	1.8013	1.9331	2.1202	3.5382	2.4108
GRU	2.7163	2.2543	1.4935	0.0700	0.8778	0.6341	2.3599	2.9007	2.6383
DBN	2.5315	4.8844	3.7328	2.6177	3.2896	3.9826	4.0637	3.0177	2.6855
FIG_DOABPNN	3.7172	4.0274	1.0853	1.4010	2.2854	1.1168	2.4077	7.6293	1.8578
FIG_ELM	5.3801	5.2446	4.3166	2.4546	3.4452	1.7648	6.3094	5.6874	4.6303
FIG_TCN	0.6869	2.4974	3.4223	2.9201	3.2966	2.5596	2.3299	3.2760	1.9782
FIG_GRU	3.4683	2.5899	2.6459	1.0499	0.0863	0.0829	0.3367	0.3543	1.7757
FIG_DBN	4.1482	3.5502	3.7996	3.7360	2.9129	2.3073	4.8371	6.7407	4.7167
FIG _MOGOA_CM	1.6671	2.7789	3.1029	0.6244	0.7456	1.1583	0.1472	3.3909	0.5151
FIG _MODA_CM	2.4670	2.7741	2.2995	0.5298	0.4578	0.2030	0.8126	3.4955	0.2044
FIG _MOALO_CM	1.4498	2.7618	3.1363	0.1884	0.3328	1.2941	0.5916	3.6622	0.2665

Note: The table shows the Diebold Mariano (DM) test results for all models in the experiment (single model, single model after granulation, different optimized combination models). The formula of its DM-test is $\mathbf{DM}=\sum_{\mu=1}^{\Pi}\left[\Omega\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_1^{(\mu)}\right)-\Omega\left(\overrightarrow{\mathbf{E}\hat{\mathbf{r}}}_2^{(\mu)}\right)\right]\Big/\left(\Pi\sqrt{S^2/\Pi}\right)$.

Table 11
Results of Improvement ratio of the indexes

	Site1			Site2			Site3		
	Step1	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**	**Step1**	**Step2**	**Step3**
DOABPNN	28.31%	19.68%	32.47%	25.66%	25.74%	25.18%	30.66%	34.45%	24.99%
ELM	49.15%	38.49%	39.22%	28.56%	25.99%	25.64%	36.69%	35.99%	21.36%
TCN	20.30%	8.22%	23.79%	20.06%	17.95%	20.91%	14.31%	29.50%	10.12%
GRU	26.39%	13.49%	30.49%	11.93%	12.48%	12.66%	21.19%	25.38%	13.88%
DBN	40.20%	27.73%	33.11%	28.80%	29.11%	33.29%	29.62%	31.23%	12.10%
FIG_DOABPNN	19.13%	22.19%	8.12%	11.88%	21.53%	9.98%	17.50%	39.69%	10.59%
FIG_ELM	33.47%	29.91%	32.43%	21.31%	22.73%	15.28%	44.46%	49.88%	23.04%
FIG_TCN	2.62%	13.19%	14.74%	5.66%	13.15%	13.00%	5.04%	8.73%	3.49%
FIG_GRU	19.90%	9.20%	12.07%	1.82%	1.99%	3.40%	6.02%	5.64%	9.19%
FIG_DBN	16.26%	15.80%	19.87%	32.72%	22.67%	19.13%	36.19%	37.49%	22.10%
FIG _MOGOA_CM	5.91%	8.68%	8.26%	1.30%	1.82%	4.16%	3.10%	3.53%	2.38%
FIG _MODA_CM	7.00%	8.80%	2.93%	2.36%	1.33%	3.89%	4.92%	6.49%	1.87%
FIG _MOALO_CM	5.02%	8.72%	8.50%	1.86%	1.48%	5.57%	8.17%	3.94%	1.68%

Note: The table shows the Results of Improvement ratio of the indexes for all models (single model, single model after granulation, different optimized combination models) in the experiment. The test formula of its IR is $\mathbf{IR_{MAPE}}=\left|\left(Compared_{MAPE}-FMICM_{MAPE}\right)/Compared_{MAPE}\right|\times 100\%$.

Table 12

Results of Forecasting Effectiveness

	Site1						Site2						Site3					
	Step1		Step2		Step3		Step1		Step2		Step3		Step1		Step2		Step3	
	OD1	OD2	OD1	OD2	OD1	OD2	OD1	OD2	OD1	OD2	OD1	OD2	OD1	OD2	OD1	OD2	OD1	OD2
DOABPNN	94.40	88.94	92.05	86.21	90.99	84.42	95.83	92.78	93.85	89.67	93.44	88.66	93.35	87.25	90.66	82.77	89.35	82.07
ELM	92.10	86.75	89.62	82.36	89.99	83.20	95.66	92.48	93.83	89.72	93.40	88.36	92.71	86.94	90.44	83.69	89.84	82.94
TCN	94.96	90.16	93.04	87.67	92.01	84.49	96.12	92.52	94.43	89.72	93.79	88.31	94.62	89.68	91.32	83.63	91.11	84.69
GRU	94.54	89.97	92.62	87.12	91.24	83.46	96.48	93.45	94.78	90.94	94.38	89.60	94.15	89.32	91.80	84.91	90.72	83.53
DBN	93.28	85.57	91.17	84.88	90.90	82.86	95.65	92.48	93.56	89.02	92.64	86.97	93.45	88.58	91.10	84.90	90.91	84.80
MODOA_CM	95.30	90.75	93.49	88.39	92.18	84.96	96.56	93.49	95.18	91.22	94.58	90.17	94.87	90.50	92.49	86.10	91.61	86.31
FIG_DOABPNN	95.03	90.91	91.79	85.06	93.38	87.84	96.48	93.39	94.18	89.29	94.55	90.01	94.41	89.53	89.85	83.12	91.06	84.88
FIG_ELM	93.96	89.52	90.89	83.83	90.99	84.00	96.06	92.74	94.09	89.52	94.20	89.45	91.69	86.18	87.79	79.25	89.62	82.31
FIG_TCN	95.87	92.50	92.64	86.69	92.86	86.13	96.72	93.36	94.74	89.67	94.36	88.76	95.14	90.64	93.29	88.06	91.72	85.63
FIG_GRU	94.98	90.84	92.97	86.47	93.08	85.18	96.84	93.60	95.34	91.36	94.92	90.17	95.09	90.83	93.51	88.57	91.20	84.03
FIG_DBN	95.20	91.36	92.42	86.30	92.40	85.47	95.39	92.06	94.09	89.52	93.93	89.19	92.77	86.36	90.21	83.10	89.74	82.62
FIG _MOGOA_CM	95.73	92.39	93.01	87.16	93.36	86.52	96.86	93.90	95.35	91.24	94.88	90.42	95.24	91.01	93.66	88.64	91.81	86.76
FIG _MODA_CM	95.68	92.27	93.00	87.09	93.73	87.40	96.83	93.88	95.37	91.34	94.89	90.57	95.15	90.70	93.46	88.36	91.86	86.70
FIG _MOALO_CM	95.77	92.45	93.00	87.13	93.35	86.50	96.84	93.89	95.36	91.35	94.80	90.31	94.98	90.37	93.63	88.61	91.87	86.08
Proposed System	**95.98**	**92.85**	**93.61**	**88.70**	**93.91**	**88.48**	**96.90**	**93.92**	**95.43**	**91.57**	**95.09**	**90.77**	**95.39**	**91.05**	**93.88**	**89.14**	**92.01**	**87.21**

Note: The table shows the Results of Forecasting Effectiveness for all models (single model, single model after granulation, different optimized combination models and FMICM) in the experiment. The test formula of its FE is $\mathbf{C}\left(h^{(1)}\right)=\psi^{(1)}$ and $\mathbf{C}\left(\psi^{(1)},\psi^{(2)}\right)=\psi^{(1)}\left(1-\sqrt{\psi^{(2)}-\left(\psi^{(1)}\right)^{2}}\right)$.

combined model was FIG_MODA_CM with a two-order mean of $\mathbf{F}_{*}^{\mathrm{mod}\,a\langle 2\rangle} = 89.81\%$. Through the forecasting effectiveness test, it can be concluded that the newly proposed FMICM model performs best in terms of point predictive validity, which means that the point prediction results of FMICM are not only accurate and stable, but also valid, they are closer to the true values in terms of the skewness and kurtosis distribution of the prediction results.

5.4 Sensitivity analysis

To verify the stability of the proposed prediction system, this section sets up the sensitivity analysis of MODOA in the proposed FMICM , and experiments are performed on three datasets with three steps of prediction. For MODOA, the parameters set are Search Number of Individuals S_A, Maximum iterations Number M_{iter} and ArchiveMaxSize A_m. We analyze the stability of the proposed prediction system with respect to changes in parameter values by varying one of the parameters by the control variables method, given that the other two parameters remain unchanged. The sensitivity index $\mathbf{SI} = \sum_{\mathbf{f}=1}^{\mathbf{K}} \sum_{\varsigma=1}^{\mathbf{P}} \left(\mathbf{E}_{\varsigma}^{\langle \mathbf{f} \rangle} - \overline{\mathbf{E}}\right)^2 \Big/ \mathbf{K} \cdot \mathbf{P}$ used, where $\mathbf{P}$ is the number of trials, $\mathbf{K}$ is the number of parameter changes, $\mathbf{E}_{\varsigma}^{\langle \mathbf{f} \rangle}$ is the point prediction evaluation index value MAPE for each trial, and $\overline{\mathbf{E}}$ is the average of the point prediction evaluation index MAPE for all trials.The specific sensitivity analysis data are shown in Table 13, and the details of this experiment are as follows.

It is obvious from the results that all three datasets show the lowest sensitivity of Maximum iterations Number, which means that M_{iter} has the least influence on the prediction results. The sensitivities of the other three parameters are less than 1 in 89% of the data, which means that the values of the three parameters have a low degree of influence on the prediction results, and thus our proposed model is relatively stable.

Table 13
Sensitivity analysis of the proposed model

	Adjusted parameters	Step1	Step2	Step3
	S_A	1.3597	0.6031	0.1811
Site1	M_{iter}	**0.7710**	**0.4339**	**0.0988**
	A_m	1.5155	0.9268	0.4853
	S_A	0.0844	0.3800	0.4030
Site2	M_{iter}	**0.0370**	**0.3544**	**0.1945**
	A_m	0.1168	0.4189	0.2039
	S_A	0.0928	0.6605	0.0415
Site3	M_{iter}	**0.0425**	**0.6233**	**0.0276**
	A_m	0.1789	1.1067	0.0765

Note: In the sensitivity analysis calculation, the Search Number of Individuals was taken as $\overrightarrow{S_A} = [60{,}80{,}100{,}120{,}140]$, the Maximum iterations Number was taken as $\overrightarrow{M_{iter}} = [100{,}200{,}300{,}400{,}500]$, and the ArchiveMaxSize was taken as $\overrightarrow{A_m} = [200{,}300$

$,400,500,600]$. Five experiments were conducted in each round, i.e., $\mathbf{P} = 5$.

5.5 Convergence analysis

Stability can be demonstrated after sensitivity analysis of MODOA, and in addition, the convergence of MODOA needs to be verified, and measuring the convergence process of MODOA can verify its computational efficiency. Figure 6 shows the corresponding convergence analysis process for the three data sets, from which it can be seen that MODOA has a high convergence speed and it can come to convergence in fewer iterations, which further proves the feasibility of its prediction system.

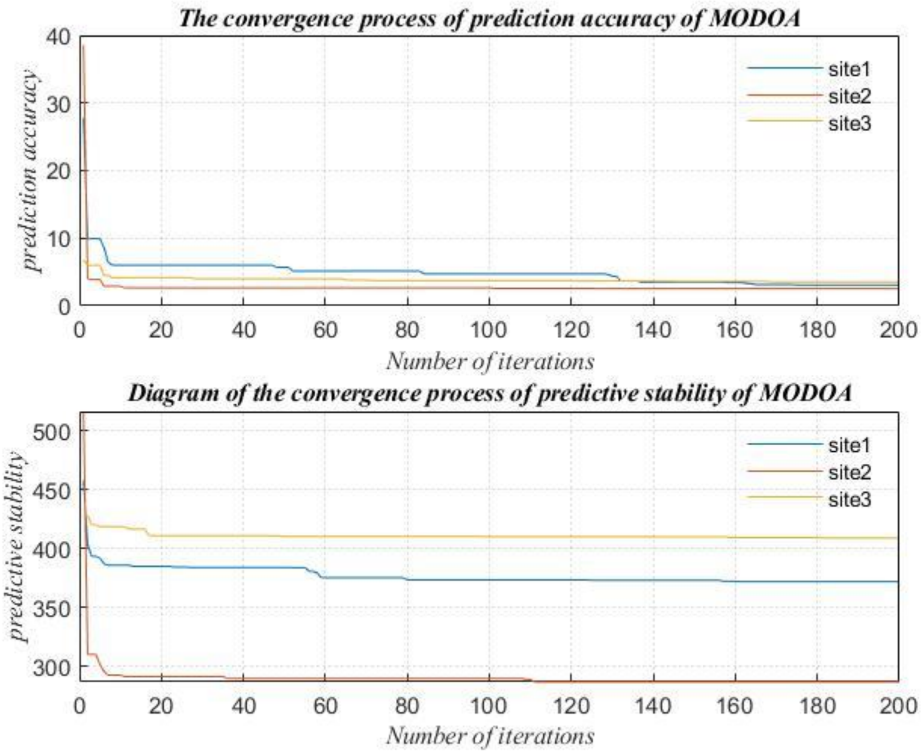

Figure 6 The convergence process of MODOA is shown

5.6 Empirical analysis

Through these checks and tests, the proposed integrated forecasting system is found to have better forecasting accuracy, stability, and effectiveness than the other 14 models. It is able to handle time series data with characteristics of randomness, volatility, periodicity, and diversity, which are affected by various factors such as power load.

(1) Accurate power load forecasting is the most effective way to ensure stable power supply and power quality. When the power generation is insufficient, the output power of generating units can be increased or deployed from other power grids; conversely, if there is excess power generation, the generating units should be shut down or deployed to other power grids, so that the power generation and power consumption can reach a certain dynamic balance. Accurate power load forecasting can help the power sector make timely scientific decisions, reduce costs and ensure the long-term safe and stable operation of the power grid.

(2) Accurate load forecasting can economically and reasonably arrange the start and stop of generating units in the power grid, maintain the safety and stability of the

power grid operation, reduce unnecessary rotation of spare capacity, reasonably arrange the unit maintenance schedule, guarantee the normal production and life of the society, effectively reduce the cost of power generation and improve economic and social benefits. The load forecasting results derived from the combined algorithm are transmitted to the power sector, which facilitates the decision on the future installation of new generating units, the size, location and timing of the installed capacity, the capacity increase and renovation of the power grid, and the construction and development of the power grid.

(3) Since the proposed point-interval prediction system can perform deterministic prediction analysis and volatility prediction analysis on time series data with randomness, volatility, periodicity and diversity characteristics, and the proposed system has high prediction stability, the proposed point-interval prediction system can be extended to other prediction problems with time series nonlinear characteristics, such as wind speed prediction, air pollution prediction and traffic flow prediction.

6.Conclusion

In this era of rapid growth of electricity demand in the whole society, accurate forecasting of power load becomes more and more important to ensure stable power supply as well as power quality. However, the change of electric load is the result of multiple factors, which have complex interconnection, and the load data has strong randomness. Therefore, this paper proposes a novel integrated power load point-interval forecasting system that constructs information grains by building fuzzy sets on subseries formed by discretized time series, which in turn compresses the scale of time series data, simplifies the computational complexity, and effectively improves the accuracy of short-term forecasting; secondly, the MODOA algorithm is used to optimize the five benchmark models in multiple stages to obtain the final point forecasting results, and the fluctuation analysis is performed on the point forecasting results to obtain the uncertain interval forecasting results. The proposed FMICM improves the accuracy and stability of power load data forecasting and expands the application scope of the model.

(1) For point forecasting, FMICM was compared with 14 models in three experiments.FMICM outperformed the single model without fuzzy particleization with the mean MAPE values of $\overrightarrow{\mathbf{MAPE}'}_{\mathbf{M}} = [3.9101\%, 5.6910\%, 6.3293\%]$ for the three-step forecasting.FMICM outperformed all the five single models used for the combination, and compared with FIG_DOABPNN, FIG_ELM, FIG_TCN, FIG_GRU, and FIG_DBN, the average values of MAPE are improved by $\overrightarrow{\mathbf{I}_{\mathbf{M}}^{(2)}} = [1.2752\%, 2.5457\%, 0.5412\%, 0.4748\%, 1.7832\%]$, respectively. The multi-objective dinger optimization algorithm in the FMICM model outperforms the known MOGOA, MODA, and MOALO in terms of weight optimization capability. (2) In terms of interval prediction. The FMICM was compared with eight models. With a confidence factor of 95%, 5/9 experiments showed that FMICM had the best AIS and MPICD, and two additional sets of experiments showed that FIG_GRU and FIG_DOABPNN had a smaller AIS than FMICM. 89% of experiments proved that FMICM had a higher interval prediction accuracy with a confidence factor of 90%, and additional individual experiments showed that FIG_MOGOA_CM and FIG_MODA_CM have better MPICD.

The proposed integrated power load point-interval forecasting system is not only accurate but also effective, which broadens the field of power load forecasting. However, there are still some aspects that need to be improved: (1) Weather conditions such as temperature and humidity can be considered. (2) The peak prediction is added to improve the prediction accuracy.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 71671029)

References

[1] Malekizadeh, M., Karami, H., Karimi, M., Moshari, A., & Sanjari, M. J. (2020). Short-term load forecast using ensemble neuro-fuzzy model. *Energy*, *19 6*, 117127. https://doi.org/10.1016/j.energy.2020.117127

[2] Fortes, P., Simoes, S. G., Amorim, F., Siggini, G., Sessa, V., Saint-Drenan, Y.-M., Carvalho, S., Mujtaba, B., Diogo, P., & Assoumou, E. (2022). How sen sitive is a carbon-neutral power sector to climate change? The interplay betwee n hydro, solar and wind for Portugal. *Energy*, *239*, 122106. https://doi.org/10.10 16/j.energy.2021.122106

[3]Chen, M.-R., Zeng, G.-Q., & Lu, K.-D. (2019). Constrained multi-objective population extremal optimization based economic-emission dispatch incorporating renewable energy resources. *Renewable Energy*, *143*, 277–294. https://doi.org/1 0.1016/j.renene.2019.05.024

[4] Yang, Y., Che, J., Li, Y., Zhao, Y., & Zhu, S. (2016). An incremental elec tric load forecasting model based on support vector regression. *Energy*, *113*, 79 6–808. https://doi.org/10.1016/j.energy.2016.07.092

[5] Dudek, G., & Pełka, P. (2021). Pattern similarity-based machine learning m ethods for mid-term load forecasting: A comparative study. *Applied Soft Compu ting*, *104*, 107223. https://doi.org/10.1016/j.asoc.2021.107223

[6] Kazemzadeh, M.-R., Amjadian, A., & Amraee, T. (2020). A hybrid data mi ning driven algorithm for long term electric peak load and energy demand fore casting. *Energy*, *204*, 117948. https://doi.org/10.1016/j.energy.2020.117948

[7] Li, S., Wang, P., & Goel, L. (2016). A Novel Wavelet-Based Ensemble Me thod for Short-Term Load Forecasting with Hybrid Neural Networks and Featur e Selection. *IEEE Transactions on Power Systems*, *31*(3), 1788–1798. https://do i.org/10.1109/TPWRS.2015.2438322

[8] Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., & Glazer, J. (2001). EnergyPlus: creating a new-generation building energy sim ulation program. *Energy and Buildings*, *33*(4), 319–331. https://doi.org/10.1016/S 0378-7788(00)00114-6

[9] Yun, K., Cho, H., Luck, R., & Mago, P. J. (2011). Real-time combined heat and power operational strategy using a hierarchical optimization algorithm. *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, *225*(4), 403–412. https://doi.org/10.1177/2041296710394287

[10] Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy*, *34*(9), 1413–1421. https://doi.org/10.1016/j.energy.2009.06.034

[11] Pappas, S. Sp., Ekonomou, L., Karampelas, P., Karamousantas, D. C., Kat sikas, S. K., Chatzarakis, G. E., & Skafidas, P. D. (2010). Electricity demand l oad forecasting of the Hellenic power system using an ARMA model. *Electric Power Systems Research*, *80*(3), 256–264. https://doi.org/10.1016/j.epsr.2009.09.0

06
[12] Jamil, R. (2020). Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. *Renewable Energy*, *154*, 1–10. https://doi.org/10.1016/j.renene.2020.02.117
[13] Pombeiro, H., Santos, R., Carreira, P., Silva, C., & Sousa, J. M. C. (2017). Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. *Energy and Buildings*, *146*, 141–151. https://doi.org/10.1016/j.enbuild.2017.04.032
[14] Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, *37*, 152–166. https://doi.org/10.1016/j.eneco.2013.02.006
[15] Ervural, B. C., Beyca, O. F., & Zaim, S. (2016). Model Estimation of ARMA Using Genetic Algorithms: A Case Study of Forecasting Natural Gas Consumption. *Procedia - Social and Behavioral Sciences*, *235*, 537–545. https://doi.org/10.1016/j.sbspro.2016.11.066
[16] Sharma, S., Majumdar, A., Elvira, V., & Chouzenoux, E. (2020). Blind Kalman Filtering for Short-Term Load Forecasting. *IEEE Transactions on Power Systems*, *35*(6), 4916–4919. https://doi.org/10.1109/TPWRS.2020.3018623
[17] Shine, P., Scully, T., Upton, J., & Murphy, M. D. (2019). Annual electricity consumption prediction and future expansion analysis on dairy farms using a support vector machine. *Applied Energy*, *250*, 1110–1119. https://doi.org/10.1016/j.apenergy.2019.05.103
[18] Chen, S., Ren, Y., Friedrich, D., Yu, Z., & Yu, J. (2021). Prediction of office building electricity demand using artificial neural network by splitting the time horizon for different occupancy rates. *Energy and AI*, *5*, 100093. https://doi.org/10.1016/j.egyai.2021.100093
[19] Khosravi, A., Nahavandi, S., Creighton, D., & Srinivasan, D. (2012). Interval Type-2 Fuzzy Logic Systems for Load Forecasting: A Comparative Study. *IEEE Transactions on Power Systems*, *27*(3), 1274–1282. https://doi.org/10.1109/TPWRS.2011.2181981
[20] Lv, M., Wang, J., Niu, X., & Lu, H. (2022). A newly combination model based on data denoising strategy and advanced optimization algorithm for short-term wind speed prediction. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-021-03595-x
[21] Chen, M.-R., Zeng, G.-Q., Lu, K.-D., & Weng, J. (2019). A Two-Layer Nonlinear Combination Method for Short-Term Wind Speed Prediction Based on ELM, ENN, and LSTM. *IEEE Internet of Things Journal*, *6*(4), 6997–7010. https://doi.org/10.1109/JIOT.2019.2913176
[22] Zhao, F., Zeng, G. Q., & Lu, K. di. (2020). EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization. *IEEE Transactions on Vehicular Technology*, *69*(1), 101–113. https://doi.org/10.1109/TVT.2019.2952605
[23] Barman, M., & Dev Choudhury, N. B. (2020). A similarity based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. *Sustainable Cities and Society*, *61*, 102311. https://doi.org/10.1016/j.scs.2020.102311
[24] Liang, Y., Niu, D., & Hong, W.-C. (2019). Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy*, *166*, 653–663. https://doi.org/10.1016/j.energy.2018.10.119
[25] Chen, Y., Kloft, M., Yang, Y., Li, C., & Li, L. (2018). Mixed kernel bas

ed extreme learning machine for electric load forecasting. *Neurocomputing*, *312*, 90–106. https://doi.org/10.1016/j.neucom.2018.05.068

[26] Xie, K., Yi, H., Hu, G., Li, L., & Fan, Z. (2020). Short-term power load forecasting based on Elman neural network with particle swarm optimization. *Neurocomputing*, *416*, 136–142. https://doi.org/10.1016/j.neucom.2019.02.063

[27] López, C., Zhong, W., & Zheng, M. (2017). Short-term Electric Load Forecasting Based on Wavelet Neural Network, Particle Swarm Optimization and Ensemble Empirical Mode Decomposition. *Energy Procedia*, *105*, 3677–3682. https://doi.org/10.1016/j.egypro.2017.03.847

[28] Hu, Y., Li, J., Hong, M., Ren, J., Lin, R., Liu, Y., Liu, M., & Man, Y. (2019). Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—A case study of papermaking process. *Energy*, *170*, 1215–1227. https://doi.org/10.1016/j.energy.2018.12.208

[29] Memarzadeh, G., & Keynia, F. (2021). Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electric Power Systems Research*, *192*, 106995. https://doi.org/10.1016/j.epsr.2020.106995

[30] Li, X., Ma, X., Xiao, F., Xiao, C., Wang, F., & Zhang, S. (2022). Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). *Journal of Petroleum Science and Engineering*, *208*, 109309. https://doi.org/10.1016/j.petrol.2021.109309

[31] Zhu, R., Liao, W., & Wang, Y. (2020). Short-term prediction for wind power based on temporal convolutional network. *Energy Reports*, *6*, 424–429. https://doi.org/10.1016/j.egyr.2020.11.219

[32] Bendaoud, N. M. M., Farah, N., & ben Ahmed, S. (2021). Comparing Generative Adversarial Networks architectures for electricity demand forecasting. *Energy and Buildings*, *247*, 111152. https://doi.org/10.1016/j.enbuild.2021.111152

[33] Wang, J., Niu, X., Zhang, L., & Lv, M. (2021). Point and interval prediction for non-ferrous metals based on a hybrid prediction framework. *Resources Policy*, *73*, 102222. https://doi.org/10.1016/j.resourpol.2021.102222

[34] He, Y., & Wang, Y. (2021). Short-term wind power prediction based on EEMD–LASSO–QRNN model. *Applied Soft Computing*, *105*, 107288. https://doi.org/10.1016/j.asoc.2021.107288

[35] Niu, X., & Wang, J. (2019). A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting. *Applied Energy*, *241*, 519–539. https://doi.org/10.1016/j.apenergy.2019.03.097

[36] Peng, S., Chen, R., Yu, B., Xiang, M., Lin, X., & Liu, E. (2021). Daily natural gas load forecasting based on the combination of long short term memory, local mean decomposition, and wavelet threshold denoising algorithm. *Journal of Natural Gas Science and Engineering*, *95*, 104175. https://doi.org/10.1016/j.jngse.2021.104175

[37] Yang, Q., Deng, C., & Chang, X. (2022). Ultra-short-term / short-term wind speed prediction based on improved singular spectrum analysis. *Renewable Energy*, *184*, 36–44. https://doi.org/10.1016/j.renene.2021.11.044

[38] Qu, C., Zhou, Z., Liu, Z., Jia, S., Wang, L., & Ma, L. (2021). State prediction for marine diesel engine based on variational modal decomposition and long short-term memory. *Energy Reports*, *7*, 880–886. https://doi.org/10.1016/j.egyr.2021.09.185

[39] Ding, S., Zhang, X., An, Y., & Xue, Y. (2017). Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification. *Pattern Recognition*, *67,* 32–46. https://doi.org/10.1016/j.patcog.2017.02.011

[40] Velázquez-Rodríguez, J. L., Villuendas-Rey, Y., Yáñez-Márquez, C., López-Yáñez, I., & Camacho-Nieto, O. (2020). Granulation in Rough Set Theory: A novel perspective. *International Journal of Approximate Reasoning*, *124*, 27–39. https://doi.org/10.1016/j.ijar.2020.05.003

[41] Zeng, G.-Q., Chen, J., Li, L.-M., Chen, M.-R., Wu, L., Dai, Y.-X., & Zheng, C.-W. (2016). An improved multi-objective population-based extremal optimization algorithm with polynomial mutation. *Information Sciences*, *330*, 49–73. https://doi.org/10.1016/j.ins.2015.10.010

[42] Zeng, G.-Q., Chen, J., Dai, Y.-X., Li, L.-M., Zheng, C.-W., & Chen, M.-R. (2015). Design of fractional order PID controller for automatic regulator voltage system based on multi-objective extremal optimization. *Neurocomputing*, *160*, 173–184. https://doi.org/10.1016/j.neucom.2015.02.051

[43] Yang, D., Guo, J., Sun, S., Han, J., & Wang, S. (2022). An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. *Applied Energy*, *306*, 117992. https://doi.org/10.1016/j.apenergy.2021.117992

[44] Bo, H., Nie, Y., & Wang, J. (2020). Electric Load Forecasting Use a Novelty Hybrid Model on the Basic of Data Preprocessing Technique and Multi-Objective Optimization Algorithm. *IEEE Access*, *8*, 13858–13874. https://doi.org/10.1109/ACCESS.2020.2966641

[45] Wang, J., Zhang, L., & Li, Z. (2022). Interval forecasting system for electricity load based on data pre-processing strategy and multi-objective optimization algorithm. *Applied Energy*, *305*, 117911. https://doi.org/10.1016/j.apenergy.2021.117911

[46] Hecht-Nielsen. (1989). Theory of the backpropagation neural network. *International Joint Conference on Neural Networks*, 593–605 vol.1. https://doi.org/10.1109/IJCNN.1989.118638

[47] Peraza-Vázquez, H., Peña-Delgado, A. F., Echavarría-Castillo, G., Morales-Cepeda, A. B., Velasco-Álvarez, J., & Ruiz-Perez, F. (2021). A Bio-Inspired Method for Engineering Design Optimization Inspired by Dingoes Hunting Strategies. *Mathematical Problems in Engineering*, *2021*, 1–19. https://doi.org/10.1155/2021/9107547

[48] Pedrycz, W., & Vukovich, G. (2002). Feature analysis through information granulation and fuzzy sets. *Pattern Recognition*, *35*(4), 825–834. https://doi.org/10.1016/S0031-3203(01)00102-9

[49] Mencar, C., & Fanelli, A. M. (2008). Interpretability constraints for fuzzy information granulation. *Information Sciences*, *178*(24), 4585–4618. https://doi.org/10.1016/j.ins.2008.08.015

[50] Li, Y., Tong, Z., Tong, S., & Westerdahl, D. (2022). A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation. *Sustainable Cities and Society*, *76*, 103481. https://doi.org/10.1016/j.scs.2021.103481

[51] Oh, S.-K., Pedrycz, W., & Roh, S.-B. (2009). Hybrid fuzzy set-based polynomial neural networks and their development with the aid of genetic optimization and information granulation. *Applied Soft Computing*, *9*(3), 1068–1089. https://doi.org/10.1016/j.asoc.2009.02.00

List of nomenclature

Abbreviations	
AIS	Average interval score
ARIMA	Auto-regressive integrated moving average model
ARMA	Auto-regressive moving average model
BPNN	Back Propagation Neural Network
CEEMDAN	Complete ensemble empirical mode decomposition
CHP	Combined heat and power
CM	Combine Model
DBN	Deep belief network
DM	Diebold-Mariano test
DOA	Dingo Optimization Algorithm
ELM	Extreme learning machine
ENN	Evolutionary neural networks
FIG	Fuzzy information granulation
FMICM	The proposed combined prediction system
GRU	Gated recurrent unit
IKDE	Improved kernel density estimation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MOALO	Multi-objective Ant Lion Optimizer
MODA	Multi-objective Dragonfly Algorithm
MODOA	Multi-objective Dingo Optimization Algorithm
MOGOA	Multi-Objective Grasshopper Optimization Algorithm
MPICD	Mean prediction interval centre deviation
MSE	Mean Square Error
PF	Point forecast
PI	Prediction interval
PSO	Particle swarm optimization
RMSE	Root Mean Square Error
SDE	The standard deviation of error
SSA	Singular spectrum analysis
TCN	Time convolutional neural network

Symbols	
$\mathbf{ERR}_i^1$ $\mathbf{ERR}_i^2$	Deviation of forecast value and actual value
$\mathbf{FIL}_i^\alpha$	The lower limit of the forecast interval with a confidence level of $1-\alpha$
$\mathbf{FIU}_i^\alpha$	The upper limit of the forecast interval with a confidence level of $1-\alpha$
$\mathbf{MPFV}_i$	The final predicted value of FMICM
$\mathbf{PFV}_i$	Model point forecast value
$\mathbf{TOV}_i$	True observation value
$\mathbf{SW}_i$	Single model weights

Kang Wang: Methodology, Software, Writing - Review & Editing.
Jianzhou Wang:Conceptualization, Funding acquisition, Supervision
Bo Zeng: Visualization, Validation
Haiyan Lu: Project administration, Formal analysis

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.