



OPEN

Research on the community electric carbon emission prediction considering the dynamic emission coefficient of power system

Hui Yu^{1,4}, Yang Yang¹, Bin Li¹, Bowen Liu^{1,4}✉, Yuanhu Guo¹, Yunqi Wang¹, Zhongfu Guo¹ & Ronghua Meng^{2,3,4}

Based on the counted power system emission factors of North China Power Grid, a community carbon emissions sample database is constructed. The support vector regression (SVR) model is trained to forecast the power carbon emissions, which is optimized by genetic algorithm (GA). A community carbon emission warning system is designed according the results. The dynamic emission coefficient curve of the power system is obtained by fitting the annual carbon emission coefficients. The time series SVR carbon emission prediction model is constructed, while the GA is improved to optimize its parameters. Taking Beijing Caochang Community as an example, a carbon emission sample database is generated based on the electricity consumption and emission coefficient curve to train and test the SVR model. The results show that the GA–SVR model fits well with the training set and the testing set, and the prediction accuracy of the testing set reaches 86%. In view of the training model in this paper, the carbon emission trend of community electricity consumption in the next month is predicted. The carbon emission warning system of the community is designed, and the specific strategy of community carbon emission reduction is proposed.

In the background of prominent global climate issues and increasing heat island effect, low-carbon development has become the requirement of the times. With the acceleration of China urbanization process, the urban residents continue to expand. Industrial clusters have accelerated development, and energy consumption and carbon emissions have been increasing continuously. The problem of high carbon emissions in cities has become particularly important. The community, especially the high-density community, is the main unit of the city. With the implementation of policies such as coal to electricity, the electricity carbon emissions become increasingly the main community carbon emissions. Prediction of the community electricity carbon emissions shows great significance to promote low-carbon life of residents and build low-carbon cities.

In recent years, as the largest developing country, China has a large economy and a high proportion of population. China consumes a lot of fossil fuels and its carbon emissions are growing rapidly. The total carbon emissions increased continuously after the total carbon emissions of China exceeded the United States firstly in 2008. China has become the largest carbon emitter since its total carbon emissions increased doubled the United States in 2019. To solve the problem of global warming and reduce greenhouse gas emissions, low carbon development has become the consensus of the world. In 2014, China first proposed the 2030 carbon peak plan in the *U.S.–China Joint Declaration on Climate Change*, and announced China's carbon peak and carbon neutrality goals in 2020. At the same time, in January 2022, the State Council issued the *14th Five-Year Plan for Energy Conservation and Emission Reduction Comprehensive Work Plan*, which improved policy mechanisms, deployed key projects, and clarified the near-term emission reduction goals of China.

In order to serve the national goal of carbon emission reduction, carbon emission prediction models and solutions have become the focus of research by scholars at home and abroad. At present, there are more forecasts^{1–3} for carbon emissions within a certain industry. In the power industry, carbon emissions are usually forecasted from economic and power consumption aspects. Researchers have carried out a series of studies based on

¹State Grid Beijing Urban District Power Supply Company, Beijing 100032, China. ²Hubei Key Laboratory of Construction and Management in Hydropower Engineering, China Three Gorges University, Yichang 443002, Hubei, China. ³Intelligent Manufacturing Innovation Technology Center, China Three Gorges University, Yichang 443002, Hubei, China. ⁴These authors contributed equally: Hui Yu, Bowen Liu and Ronghua Meng. ✉email: liubowen@Outlook.com

the influencing factors of carbon emissions^{4,5}, regression models^{6,7}, system dynamics models^{8–10} and nonlinear models^{11,12}. For example, in terms of influencing factors, He et al.¹³ pointed out that carbon emissions from the power sector account for more than 40% of the total emissions in China. They examined the influencing factors of carbon emissions from the power sector in China at both national and provincial levels, with economic growth being the main driver and power consumption intensity, thermal power generation energy intensity and power mix being the main inhibiting factors. Sun Wei et al.¹⁴ used stochastic frontier analysis to screen factors affecting carbon emission intensity from the perspective of carbon emission efficiency, and constructed a carbon emission intensity prediction model based on factor analysis and extreme learning machine. McKibbin et al.¹⁵ adopted an approach developed using the G-Cubed multi-country model in which economic structure and emission outcomes were determined simultaneously. The framework for emission forecast should focus on the time-variant sources of economic growth and structure of the global economy.

The construction of prediction model is also the focus of research on carbon emissions: based on linear fitting prediction, mainly including IPAT equation, carbon emission environment Kuznets curve, grey prediction, and other models. In the early 1970s, Ehrlich et al.¹⁶ proposed the IPAT equation, which used mathematical equations to quantify the relationship between economic growth and resource environment. This paper applied IPAT equation to analyze the factors affecting carbon emissions and to predict carbon emissions. By 1990s, Grossman et al.¹⁷ found that there was an inverted U type relationship between economic growth and environmental quality. They proposed the carbon emission environmental Kuznets curve, which has received much attention from scholars. However, the application of the method by researchers shows that, it is currently difficult to explain the contradictory conclusions generated by the study, which reflects the complex relationship between carbon emissions and the economy; on the other hand, the defects of the environmental Kuznets curve theory itself cannot be ignored¹⁸. For the complex relationship between various factors in power systems, methods to deal with multi-factor problems have been proposed one after another. Grey prediction models based on grey system theory¹⁹ are widely used in several fields. Several scholars^{20–22} have used grey prediction models to forecast carbon emissions. The grey prediction model is widely used, easy to calculate, less data sample requirement, higher accuracy, and suitable for short- and medium-term forecast. For long-term forecast, the uncertainty of future perturbation factors will make the forecast accuracy lower. Xiong et al.²³ developed a new multivariate grey model based on linear time-varying discrete parameters. In this model, a linear time-varying function was introduced into the traditional model to dynamically optimize the fixed parameters that can only be used for static analysis. This method does not change the essence of linear prediction, although it improves some prediction accuracy.

In recent years, with the development of big data technology, machine learning methods began to be widely used in load prediction and state estimation, and achieved good application results, so some scholars began to use machine learning^{24–26} for carbon emission prediction. Support vector regression (SVR) is also widely used in time series prediction problems, and it performs superior generalization ability and can better avoid local optima through parameter optimization. Wang et al.²⁷ proposed an online SVR model to predict air pollutant levels in forward time series. Melahat et al.²⁶ used deep learning (DL), support vector machine (SVM) and artificial neural network (ANN) algorithms to predict carbon emissions in the electricity production sector in Turkey, respectively. Li et al.²⁸ used genetic algorithm (GA) to optimize the weights and thresholds of the SVR to predict CO₂ emissions in Beijing from 2016 to 2020 by scenario analysis. Saleh et al.²⁹ established an SVR model with energy consumption such as electric energy and coal combustion as input variables, which directly affect the increase of CO₂ emissions. The SVRs have been used successfully in prediction of time series problems in many ways.

Although a series of studies have been conducted by scholars on the carbon emissions prediction, few articles have been published on the carbon emissions prediction for community electricity. Usually, the carbon emissions of community electricity consumption are mainly related to the community electricity consumption and the carbon emission factor of electricity supply. Carbon emission coefficient³⁰ is the number of carbon emission mass per unit of energy produced during the combustion or use of each energy source. In the existing literature, most of them use fixed carbon emission factors, or guideline factors issued by the state. In fact, the carbon emission factor of electric energy varies dynamically depending on the energy ration of power production by the power supply company. Especially for community, city, and national electricity carbon emission forecast, the small changes of carbon emission coefficients have a great impact on the total carbon emission due to the large consumption of electric energy. The traditional forecast methods have some problems such as unstable regression and difficult to determine the influencing factors, which will reduce the scientific accuracy of forecast. After research, it is found that the changes of carbon emissions and power supply carbon emission coefficients in a certain region are time-series correlated, so the historical data of carbon emissions can be used to predict the future carbon emissions.

Now, China has proposed the concept of smart community. The community will fully implement the construction requirements of new electric power system, complete electric energy substitution, replace coal with electricity and gas with electricity. The electric energy becomes the main energy source of the community. To carry out the promotion of terminal electric energy substitution products, residents' clean energy and reliable electricity use will be effectively guaranteed. The well-being of regional people's livelihood will be improved in an orderly manner. Therefore, the prediction of carbon emission for community electric energy can effectively represent the overall carbon emission of the community, and is increasingly meaningful for community carbon emission reduction.

Based on the complex nonlinear carbon emission prediction system, this paper intends to calculate the annual carbon emission coefficient according to the power production of North China Power Grid (NCPG). The dynamic carbon emission factor curve of electricity will be fitted. Considering both the dynamic carbon emission coefficient on the supply side and the electricity consumption on the demand side, a sample database of carbon emissions in the community will be built. The GA optimization SVR model (GA-SVR) is improved and designed to predict the carbon emissions of community electricity consumption.

Carbon emission sample database

Different power supply networks have different ratios of energy generation; thus, the same carbon emission factor is inappropriate. Even if the energy ratio of the same power supply network is different in different years, the carbon emission factor is dynamic. This paper collects the energy consumption of NCPG from 2011 to 2020, calculates the annual carbon emission factors, and fits the carbon emission factors after 2021.

Statistical of annual power system emission coefficient. According to Power system emission coefficient calculation tool³¹, the calculation method of carbon dioxide baseline emission factor OM in China power grid. Based on the total net power generations, fuel types and fuel consumptions of all power plants in the power system, the formula (1) is as follows:

$$EF_{grid,OMsimple,y} = \frac{\sum_i (FC_{i,y} \times NCV_{i,y} \times EF_{CO_2,i,y})}{FC_y} \quad (1)$$

$EF_{grid,OMsimple,y}$ is the simple electricity marginal emission factor OM (tCO_2/MWh) of the power system when the research object in Year y . EG_y is the net total generation of the power system in Year y . That is, the total electricity (MWh) supplied to the grid by all units other than the must operating cost. $FC_{i,y}$ is the total fuel consumption (mass or volume unit) of the above units in Year y . $NCV_{i,y}$ is the average low calorific value (GJ/mass or volume unit), while the $EF_{CO_2,i,y}$ is the CO_2 emission factor of fuel i in Year y , (tCO_2/G). Then, i is the type of fossil fuel consumed by the power system in year y , while y is the year that the data are available for project submission. In the case of power exchange between power grids, the power exchange between power grids will produce net receiver party. The simple electricity marginal emission factor is equal to the weighted average value of the unit electricity emission factor of the local power plant and the unit electricity receivers. The emission factor of unit electricity of local power plants is calculated according to the above formula, and the net receive electricity adopts the simple marginal emission factor of transferred power grid.

The data of power generation, fuel consumption for power generation and low calorific value of power generation fuel for OM are originated respectively from *China Energy Statistical Yearbook*³². The data of auxiliary power consumption rate are derived from *China Electric Power Yearbook*. The data of power exchange between power grids are from *Compilation of electric power industry statistics*. The CO_2 emission factors of fuel are derived from Table 1.4 in Chapter I of the *Guidelines for the preparation of the 2006 IPCC national inventory Energy Volume*. The lower limit of 95% confidence interval of each fuel emission factor is determined according to the principle of conservatism.

Establishment of dynamic carbon emission coefficient regression model. According to the principle of least squares, the fitting curve objective is constructed as formula (2):

$$\min_{\varphi} \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n (\varphi(x_i) - y_i)^2 \quad (2)$$

where $\varphi(x_i)$ represents the ordinate of the fitting model function, while y_i expresses the ordinate of the real scatter point. The smallest interpolation quadratic sum of the ordinates from the function model and scattered points will be considered as the best fit degree.

The polynomial fitting model is set as formula (3):

$$y = a_0 + a_1x + \dots + a_kx^k \quad (3)$$

The carbon emission curve is obtained by fitting the calculated annual carbon emission coefficient. Then the daily carbon emission coefficient will appear.

Calculation of carbon emission. According to the electricity consumption and carbon emission coefficient curve of the community, daily carbon emission can be obtained from the following formula (4).

$$CE = y_{OM} \times E \quad (4)$$

where CE is the carbon emission, y_{OM} is the carbon emission coefficient curve, and E is the community electricity consumption.

Methods

Due to the continuous construction of communities and changes in resident living habits, the sample data beyond a certain period for carbon emission forecast will inevitably lead to inaccurate prediction. The training sample size is insufficient to support the learning process of prediction model, when the recent sample data are considered merely. The SVR³³ introduced in this paper is a novel small-sample learning method that can solve the small training sample problems. Due to its wide range of weight parameters, poor training results will appear with empirical parameters, while it will be time-consuming by iterative methods.

In this paper, GA is introduced to synchronously optimize the penalty factor C and radial basis function parameter g in SVR to obtain the global optimal solution. Compared with traditional optimization methods, GA³⁴, based on biological evolution, is a modern intelligent algorithm that simulates the fittest survival in the biological world and the natural genetic mechanism, which has a strong global search ability. Its calculation

speed is fast, with high robustness and strong expansibility. The GA is easy to combine with SVR model, which speeds up the solution process of this problem.

Basic principles of genetic algorithms. GA is stochastic global search and optimization methods developed to mimic the mechanisms of biological evolution in nature. The model simulates the natural selection and genetic mechanism of Darwinian biological evolution process. To find the optimal solution by following the law of survival of the fittest. Keep the excellent individuals and eliminate the bad individuals during the optimization process. The GA optimizes the penalty parameter C and kernel function parameter g of the SVR model to improve its prediction accuracy, aiming at minimizing the root mean square error (RMSE) of the predicted value and the actual value of the sample. The fitness function formula is:

$$fit = \sqrt{\frac{1}{N} (T_{sim} - T_{train})^2} \quad (5)$$

where T_{sim} represents the predicted value of the training set, T_{train} means the actual value of the sample, and N represents the number of samples in the training set.

Basic principle of SVR. The SVR is a supervised learning algorithm, which is widely used in discrete prediction and other fields. For a given data set $\{x_i, y_i\}_{i=1}^N$, N is the sample number in the training set; x_i and y_i are input and output variables respectively. SVR maps the original space of the input data to a higher dimensional feature space through the nonlinear Gaussian kernel function. In the feature space, the problem is transformed into the construction of the optimal linear plane $f(x) = w \cdot \varphi(x) + b$ for fitting data. Where b is offset and w is weight. Therefore, by minimizing the vector norm w^2 to find the smoothest function $f(x)$, the maximum allowable error of the predicted value of each training data is ε , as Fig. 1. Then the Lagrange multiplier is used to rephrase the constrained optimization problem as a dual problem. For each constraint, the quadratic programming is used to determine, and then the offset of the optimal weight is calculated to obtain the predicted value. In this way, the previous regression problem can be replaced by a constrained optimization problem. The formulae of objective and constraints are as follows:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\delta_i + \delta_i^*) \quad (6)$$

$$\begin{cases} y_i - w \cdot \varphi(x) - b \leq \varepsilon + \delta_i \\ w \cdot \varphi(x) + b - y_i \leq \varepsilon + \delta_i^* \\ \delta_i \geq 0 \\ \delta_i^* \geq 0 \end{cases}, i = 1, 2, \dots, N \quad (7)$$

In mathematic formulae (6) and (7), where C is the penalty coefficient, δ_i and δ_i^* are both relaxation variables, and ε is the loss function. In addition, this model is converted into a pairwise function by multiplier solving method with the Lagrange function. The nonlinear function can be expressed as formula (8):

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (8)$$

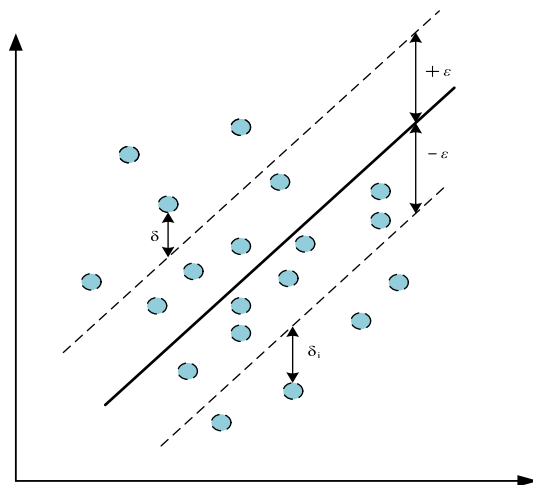


Figure 1. Schematic diagram of support vector product regression.

where α_i and α_i^* denote Lagrange multipliers and $K(x_i, x_j)$ denotes the kernel function. The SVR will have different learning and generalization performance by different kernel functions. The linear kernel function, polynomial kernel function and radial basis function (RBF) are commonly used by SVR. RBF performs better than linear kernels on dealing with nonlinear problems and has fewer hyperparameters than polynomial kernels. Therefore, RBF is widely considered to be an ideal function on handling complex and multidimensional samples. In this paper, RBF is chosen as the kernel function, as formula (9).

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (9)$$

where σ denotes the bandwidth, which is sufficient to handle a limited number of samples.

Algorithm implementation. The improved GA–SVR algorithm is implemented as follows:

Step 1 the obtained carbon emission data are normalized, and the SVR model kernel function RBF and parameters are selected.

Step 2 parameters C and g of the SVR model are considered as variables, and optimized through the improved GA.

Step 3 train and test the SVR model as the decode process of GA.

Step 4 calculate the fitness function of the GA and save the optimal solution. Judge whether the segment satisfies the stopping condition. If the condition is not satisfied, continue the GA operation to Step 2, while the condition is satisfied, Step 5 will be executed.

Step 5 forecast 1 month carbon emission values based on known annual carbon emission data.

The flow chart of GA–SVR is described in Fig. 2.

Results

Caochang Community is in Dongcheng District, Beijing, China. It is from Qianmen-East Street in the west to Caochang-shitiao in the east, and starts Xi-damochang Street in the north to Lianguang Street in the south, which covers an area of 586,000 square meters. The population is 9000, including 4817 resident customers and 359 non-residents. In 2021, the power consumption of this area is 14.772 million kwh, and the load of residential customers is account for about 68%. Caochang area is evaluated as A+, which belongs to Tong'an and Anhangxing Districts. The total length of the transmission line is 18.1 km, and the cabling rate achieves 100%. It has 47 shutters and 87 box transformers. The coverage rate of line automation reaches 100%, while the coverage of equipment automation ascends to 92.73%.

The north has higher carbon emissions than the south in China, especially to warm by heating in winter. Large coal consumption leads to high carbon emissions. Caochang Community realizes the electric energy substitution in the city firstly, which becomes the first demonstration block to replace coal and gas with electricity. Since 2007, the region has completed the 'coal to electricity', 'gas to electricity' work successively. The proportion of regional electric energy in terminal energy consumption is increasing. Since 2018, it has completed the construction of regional hutong overhead lines before others and promoted the three modernizations of power equipment boxes. All these measures are devoted to build a green, low-carbon, clean and livable model of old urban areas. After 2019, 52 Symbiosis Institutes have been connected to power, and have comprehensively promoted terminal power alternative products. The clean and reliable of resident electricity consumption are effectively guaranteed. The

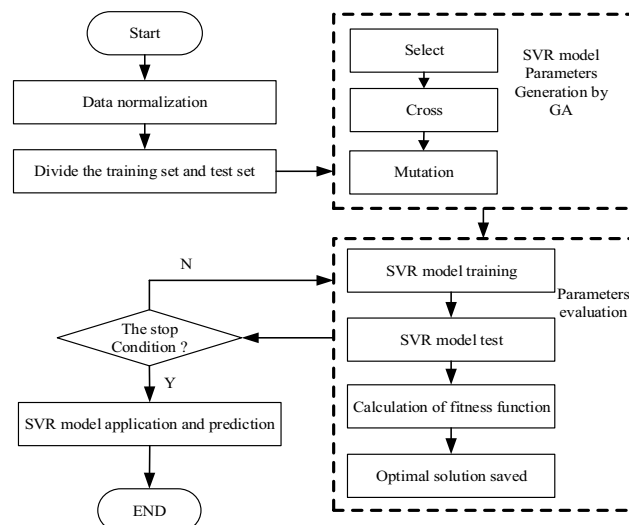


Figure 2. Flow chart of GA–SVR algorithm.

improvement of regional resident livelihood is promoted in an orderly manner. So, the carbon emission prediction of electric energy can effectively represent the overall carbon emission of the community.

Taking Caochang Community in Beijing as an example, the carbon emission sample database is generated based on the electricity consumption and emission factor curve for GA-SVR model's training and testing.

Dynamic carbon emission coefficient curve regression fitting. According to the statistical method of this paper, the annual emission factor of the power system is calculated. The linear fitting is performed by Matlab to obtain the curve shown in Fig. 3.

The curve formula of carbon emission factor obtained through linear fitting is as formula (13):

$$y_{OM} = -1.547e - 05x^5 + 0.1554x^4 - 624.3x^3 + 1.254e + 06x^2 - 1.26e + 09x + 5.063e + 11 \quad (10)$$

The square root of the error $R^2 = 0.9172$, which is greater than 0.9, performs an excellent fitting effect.

From the fitting curve of the annual emission coefficient of the power system, in the process of continuous fluctuation, the coefficient declined steadily. Emission factors from 2011 to 2012 show a high situation. Since 2013, the annual carbon emission factors turn to a significant downward trend. However, during 2018 and 2019, the emission factors increased significantly due to the use of blast furnace gas and converter gas with higher fuel emission factor. The above raw materials are by-products of the production process of steel enterprises. Although the gas improves the carbon emission factor of electric energy during power generation, it forms an industrial ecological chain with the steel metallurgical industry, which can improve the added value of gas.

Introduction of evaluation metrics. In order to evaluate the prediction effect of GA-SVR model, four evaluation metrics are introduced as follows. Root means square error (RMSE), mean absolute error (MAE), mean bias error (MBE) and R-Square (R^2). RMSE is used to measure the deviation between the observed value and the true value. MAE and MBE can better reflect the prediction error. R^2 can reflect the fitting ability from a statistical point of view. The formulae of these evaluation metrics are as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{m=1}^N (y_m - \hat{y}_m)^2} \quad (11)$$

$$MAE = \frac{1}{N} \sum_{m=1}^N |y_m - \hat{y}_m| \quad (12)$$

$$MBE = \frac{1}{N} \sum_{m=1}^N (y_m - \hat{y}_m) \quad (13)$$

$$R^2 = 1 - \frac{\sum_{m=1}^N (y_m - \hat{y}_m)^2}{\sum_{m=1}^N (y_m - \bar{y})^2} \quad (14)$$

Prediction result analysis of GA-SVR model. In order to analyze the operation effect of the algorithm and the fairness of multi-algorithm comparison, the following algorithms are all implemented by Matlab2020(a).

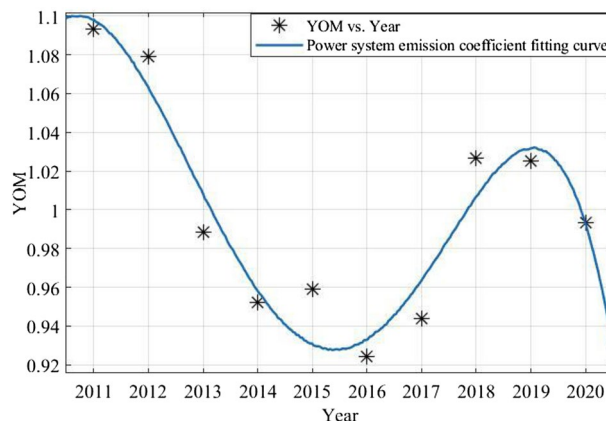


Figure 3. Power system emission coefficient fitting curve.

The computer system is Windows 10 with Intel (R) Core (TM) i7-10700 CPU @ 2.90 GHz configuration. Set the parameter values and ranges for GA and SVR. Referring to the parameter setting of similar problems in the existing literature, different parameters are set to conduct the Taguchi experiments. According to the experiment results, when the population size is 10 and the number of iterations is 25, the satisfied results can be obtained with appropriate running time. In SVR model, the training set scale is 0.7, the remaining is testing set. The range of parameter C and g are both taken (0,100). The following Fig. 4 is the fitness variation curve of parameter optimization.

The optimal penalty parameters C and g are obtained as 22.6580 and 8.9441 respectively, which is optimized by GA. Analyze the prediction results of the GA-SVR model with the training set and the testing set, the comparison curves are shown in Figs. 5 and 6.

From the two comparison curves of Figs. 5 and 6, the fitting accuracy of the raining set prediction results by GA-SVR performs excellent. The curve characteristics perform basically consistent. The RMSE is 1.951 merely, while the RMSE of testing set is 14.3466. The R^2 of the training set achieves 0.9960, while the R^2 of testing set exceeds 0.85. Although testing set prediction curve deviates from the expected value, the curve change trend can reflect the change of the true value.

Prediction results analysis of compared algorithms. In order to verify the validity of the prediction model in this paper, a SVR with random parameters ($C=4$, $g=0.8$), the BP neural network (BP) and Random Forest (RF) models are introduced to compare with GA-SVR. Three metrics R^2 , MAE and MBE are utilized to evaluate the prediction effect of the models. The results are as follows.

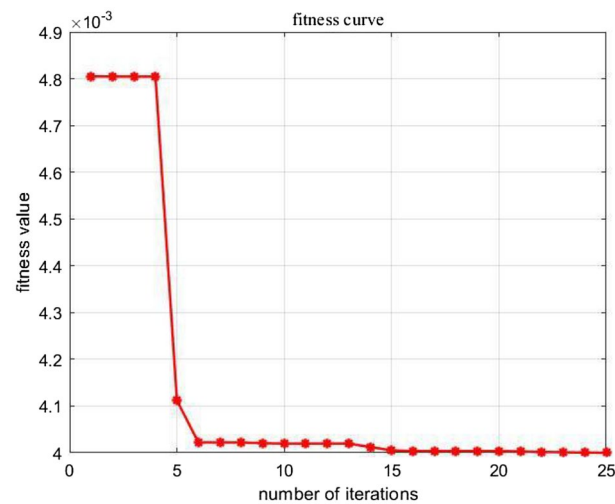


Figure 4. Fitness iteration curve.

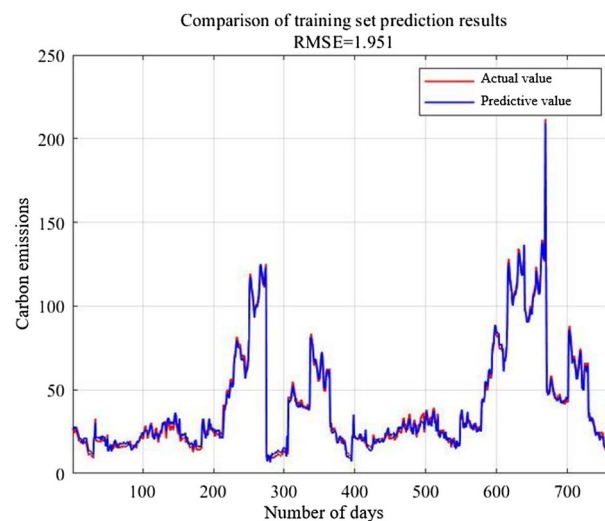


Figure 5. Prediction comparison of training set.

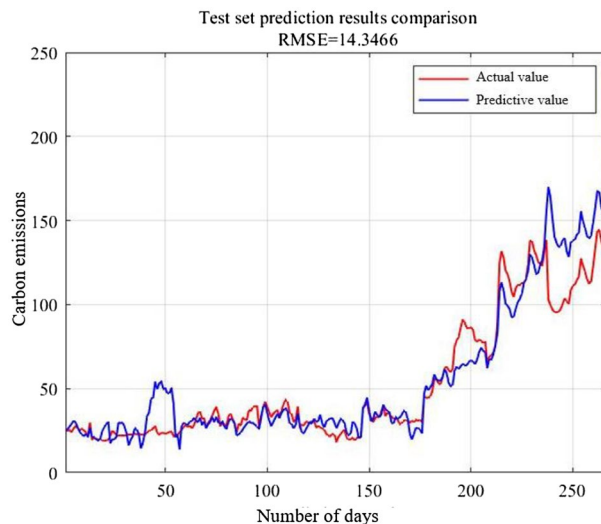


Figure 6. Prediction comparison of testing set.

In Table 1, the improved GA-SVR model performs excellent on the carbon emissions prediction of electricity consumption in urban communities. The fit correlation coefficients between prediction values and actual values achieves 0.9960 and 0.8601 respectively for the training set and the testing set. They both reaches more than 85%. The community electricity carbon emission forecast is accurate. The training and testing set of the improved algorithm fit more accurately than the other three comparison algorithms.

For the stability comparison, the four comparison algorithms, GA_SVR, SVR, BP and RF, are independently run for 20 times. The evaluation indicators are obtained as shown in Table 2. The horizontal direction represents the worst values (min), the best values (max), the average values (mean) and the standard deviations (std) of the prediction accuracy R^2 of the training set and the test set.

It can be seen from the statistical results in Table 2 that GA_SVR algorithm performs best in the solution process. When the determined parameters of SVR, the solution is the most stable, but the result metrics are not very well. The GA_SVR solution process performs both well and stable, while GA is introduced to improve the parameters of SVR. The optimal value, the worst value and the average value perform more outstanding than other compared algorithms.

Since electricity consumption data may be affected by uncertainties such as the epidemic, the carbon emissions of community electricity consumption have shown a significant upward trend in a certain period. So, the data uncertainty is strong. After the improvement, the fitting accuracy is greatly improved compared with the random parameter SVR.

Model	Training set			Testing set		
	R^2	MAE	MBE	R^2	MAE	MBE
GA-SVR	0.9960	1.7800	-0.0742	0.8601	9.2341	2.9112
SVR	0.6974	9.9612	-1.2350	0.7284	13.9492	-6.9973
BP	0.5412	14.1563	-2.5174	0.5410	15.8081	-11.9257
RF	0.9328	4.7816	0.0648	0.7073	13.8018	-7.3491

Table 1. Comparison of prediction algorithms.

Model	Comparison index of training set R^2				Comparison index of testing set R^2			
	Min	Max	Mean	Std	Min	Max	Mean	Std
GA-SVR	0.9882	0.9963	0.9916	0.0027	0.8235	0.8661	0.8488	0.0115
SVR	0.6958	0.6958	0.6958	0.0000	0.6899	0.6899	0.6899	0.0000
BP	0.3468	0.7106	0.5483	0.0967	0.3627	0.6050	0.5157	0.0631
RF	0.9238	0.9325	0.9283	0.0023	0.6655	0.6982	0.6882	0.0072

Table 2. Comparison index performance of R^2 . The optimal values of each evaluation index are in [bold].

Future carbon emissions forecast. After substituting the annual emission factor curve of the power system from 2019 to 2021 and the average daily electricity consumption of residents into the trained GA-SVR model. The trend comparison between the actual and predicted carbon emissions (tCO₂e) of the community in January 2022 is obtained, and the prediction accuracy R² is 84.33%.

As shown in Fig. 7, the prediction results are consistent in the actual carbon emission data of electric power. In the Caochang Community Energy Big Data Management System, the electricity consumption of the community can be calculated in real time. After the prediction model is modified according to the latest data, the carbon emission data in the next 30 days can be estimated with a high accuracy. It performs guiding significance for community carbon emission early warning within 1 month.

Carbon emission early warning mechanism and suggestions for community residents emission reduction. *Design of carbon emission early warning mechanism.* As the implementation of the replacing coal with electricity policy, the electricity carbon emissions increase for a period, which is mainly due to the increase in electricity consumption. However, with the stabilization of electricity consumption and the control of carbon emissions from power supplies, the effect of carbon emissions reduction will become more and more significant.

According to the carbon emission prediction results of community residents, the energy consumption analysis, energy saving and efficiency, and convenient power connection will be provided. If the forecast carbon emissions exceed a certain threshold compared with the same period of previous years, the previous month and the previous day, a carbon emission early warning will be issued to remind residents³⁵.

As shown in the Fig. 8, carbon emissions in 1 month can be predicted. If the forecast monthly carbon emissions and the forecast results of a certain day exceed the threshold of the early warning, carbon emission warning will be issued to residents of the community and the rational use of electricity will be reminded.

Recommendations of community carbon reduction strategy. Northern communities prefer to keep warm with coal. With the implementation of gas to electricity, coal to electricity policy, community electricity carbon emis-

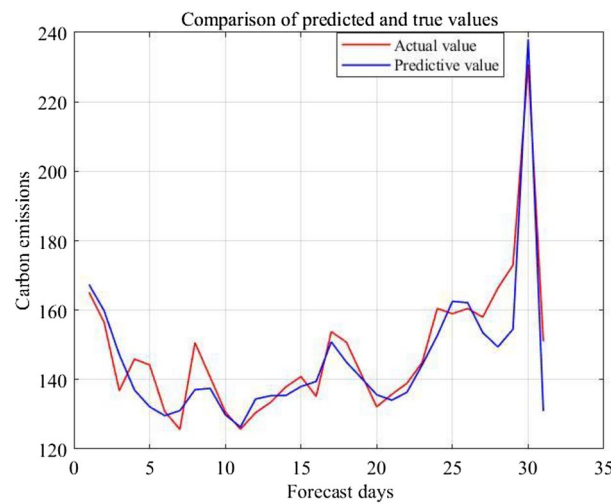


Figure 7. Carbon emission forecast and actual values for January 2022.

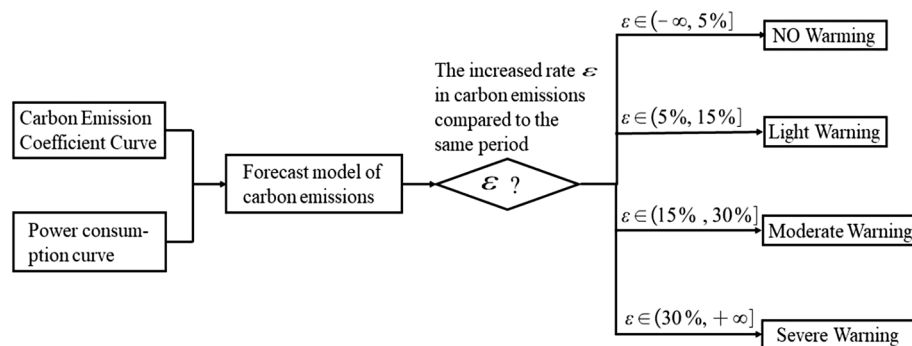


Figure 8. Community carbon emission warning process.

sion has grown significantly. On the premise of carbon emission reduction of NCPG power supply, the community will also achieve carbon emission reduction from the following perspectives:

New energy as a power supply has a lot of improvement space. In the process of building a new power system, solar, wind and other new energy power generation will gradually become the main power source. For example, the construction of intelligent micro-network parking lot, the exploration of variable charging integrated box-type transformer, DC V2G charging pile and other technical applications. All these will provide basic guarantee for regional green transportation and facilitate resident travel.

Implement intelligent transportation strategy. Select Xixinglong Street for street lamp reconstruction and layout intelligent control street lamps. Install photovoltaic panel lampshade, intelligent control device, wireless charging, WIFI, video surveillance, information announcement screen and other value-added functions. Realize intelligent lighting of street lamps, self-use, complementary control of municipal electricity, and promote the parallel of intelligent transportation and green transportation. To solve the problem of charging electric bicycles for common people, two centralized charging sheds will be selected in North Lucao Garden and Xixinglong Street for photovoltaic transformation. Provide convenient, safe and green charging experience for community residents, and guide residents to change their travel mode.

Strengthen the promotion of green and low-carbon life. Improve community residents' awareness of low carbon environmental protection from various aspects, such as food, clothing, housing and transportation, which will promote carbon emission reduction in life.

Conclusion

This paper calculates the emission coefficient of the power system, and constructs a carbon emission fitting curve. Based on the community electricity consumption data and the dynamic carbon emission coefficient, a rapid carbon emission prediction model is constructed and trained. Although the carbon emission prediction has achieved ideal results, the research in this paper hasn't take into account the national policies and the key factors that affect the carbon emission change of electric energy. There is still much room to optimize the results of the testing sets. In future, the above factors would be considered for further research.

Data availability

All data generated or analyzed during this study are included in supplementary file Code&data.

Received: 7 November 2022; Accepted: 6 March 2023

Published online: 05 April 2023

References

1. Yu, Y. S., Jeong, J. W., Chon, M. S. & Cha, J. A study on prediction of real-driving CO₂ emissions for OBD of light-duty diesel vehicle. *Trans. Korean Soc. Autom. Eng.* **27**, 12. <https://doi.org/10.2139/ssrn.4088435> (2019).
2. Yang, H. J. & Oconnell, J. F. Short-term carbon emissions forecast for aviation industry in Shanghai. *J. Clean. Prod.* **275**, 122734. <https://doi.org/10.1016/j.jclepro.2020.122734> (2020).
3. Zhang, F. *et al.* Prediction of China's copper material flows under carbon emissions projections for the shared socioeconomic pathways. *Resour. Conserv. Recycl.* **182**, 106340. <https://doi.org/10.1016/j.resconrec.2022.106340> (2022).
4. Zhao, M. *et al.* Decomposing the influencing factors of industrial carbon emissions in Shanghai using the LMDI method. *Energy* **35**(6), 2505–2510. <https://doi.org/10.1016/j.energy.2010.02.049> (2010).
5. Wang, S., Wang, Y. X., Zhou, C. X. & Wang, X. L. Projections in various scenarios and the impact of economy, population, and technology for regional emission peak and carbon neutrality in China. *Int. J. Environ. Res. Public Health* **19**(19), 12126. <https://doi.org/10.3390/ijerph191912126> (2022).
6. Xu, X. & Liao, M. Prediction of carbon emissions in China's power industry based on the mixed-data sampling (MIDAS) regression model. *Atmosphere* **13**(3), 423. <https://doi.org/10.3390/atmos13030423> (2022).
7. Fang, D. B., Zhang, X. L., Yu, Q., Jin, T. C. & Tian, L. A novel method for carbon dioxide emission forecasting based on improved Gaussian processes regression. *J. Clean. Prod.* **173**, 143–150. <https://doi.org/10.1016/j.jclepro.2017.05.102> (2018).
8. Zeng, M. *et al.* Pre-integrated forecasting method research of urban electricity consumption based on system dynamics and econometric model. *J. Appl. Sci.* **13**, 4732–4737. <https://doi.org/10.3923/jas.2013.4732.4737> (2013).
9. Dyson, B. & Chang, N. B. Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste Manag.* **25**, 669–679. <https://doi.org/10.1016/j.wasman.2004.10.005> (2005).
10. Venkatesan, A. K., Ahmad, S., Johnson, W. & Batista, J. R. Systems dynamic model to forecast salinity load to the Colorado River due to urbanization within the Las Vegas Valley. *Sci. Total Environ.* **409**, 2616–2625. <https://doi.org/10.1016/j.scitotenv.2011.03.018> (2011).
11. Acheampong, A. O. & Boateng, E. B. Modelling carbon emission intensity: Application of artificial neural network. *J. Clean. Prod.* **225**, 833–856. <https://doi.org/10.1016/j.jclepro.2019.03.352> (2019).
12. Ahmed, Z., Wang, Z. & Ali, S. Investigating the non-linear relationship between urbanization and CO₂ emissions: An empirical analysis. *Air Qual. Atmos. Health* **12**(8), 945–953. <https://doi.org/10.1007/s11869-019-00711-x> (2019).
13. He, Y. *et al.* Factors influencing carbon emissions from China's electricity industry: Analysis using the combination of LMDI and K-means clustering. *Environ. Impact Assess. Rev.* **93**, 106724. <https://doi.org/10.1016/j.eiar.2021.106724> (2021).
14. Sun, W. & Huang, C. Predictions of carbon emission intensity based on factor analysis and an improved extreme learning machine from the perspective of carbon emission efficiency. *J. Clean. Prod.* **338**, 130414. <https://doi.org/10.1016/j.jclepro.2022.130414> (2022).
15. McKibbin, W. J., Pearce, D. & Stegman, A. Long term projections of carbon emissions. *Int. J. Forecast.* **23**(4), 637–653. <https://doi.org/10.1016/j.ijforecast.2007.10.004> (2007).
16. Ehrlich, P. R. & Holdren, J. P. Impact of population growth. *Science (New York)* **171**, 3977. <https://doi.org/10.1126/science.171.3977.1212> (1971).
17. Grossman, G. M. & Krueger, A. B. Environmental impacts of a North American free trade agreement. <https://doi.org/10.3386/w3914> (1991).
18. Zhong, M. C. & Zhang, X. G. A critical review of environmental Kuznets curve theory and research. *China Popul. Resour. Environ.* **20**(02), 62–67. <https://doi.org/10.3969/j.issn.1002-2104.2010.02.011> (2010).

19. Zhao, A. W. & Li, D. Grey projection of carbon emissions in China. *Pract. Underst. Math.* **42**(04), 61–69. <https://doi.org/10.3969/j.issn.1000-0984.2012.04.009> (2012).
20. Shen, X. Y. Analysis and prediction of factors influencing carbon emissions in China. *Stat. Decis. Mak.* **19**, 90–92. <https://doi.org/10.13546/j.cnki.tjyj.2010.19.016> (2010).
21. Yang, K. L. & Zhang, Z. Y. Analysis of carbon emission forecast and countermeasures for low carbon economy development in Tianjin city—based on improved GM(1,1) model. *J. Chongqing Univ. Technol. (Nat. Sci.)* **28**(02), 111–114. [https://doi.org/10.3969/j.issn.1674-8425\(z\).2014.02.024](https://doi.org/10.3969/j.issn.1674-8425(z).2014.02.024) (2014).
22. Sun, Y. Q. & Li, Q. Grey correlation analysis and prediction of carbon emission influencing factors in Hunan Province. *J. Changsha Univ. Sci. Technol. (Soc. Sci. Ed.)* **26**(1), 65–69. <https://doi.org/10.3969/j.issn.1672-934X.2012.01.011> (2012).
23. Xiong, P. P. *et al.* Forecasting carbon emissions using a multi-variable GM (1, N) model based on linear time-varying parameters. *J. Intell. Fuzzy Syst.* **41**(6), 6137–6148. <https://doi.org/10.3233/JIFS-202711> (2021).
24. Liu, Z. K., Jiang, P., Wang, J. Z. & Zhang, L. F. Ensemble system for short term carbon dioxide emissions forecasting based on multi-objective tangent search algorithm. *J. Environ. Manag.* **302**(1), 113951. <https://doi.org/10.1016/j.jenvman.2021.113951> (2022).
25. Ren, F. & Long, D. H. Carbon emission forecasting and scenario analysis in Guangdong Province based on optimized fast learning network. *J. Clean. Prod.* **317**, 128408. <https://doi.org/10.1016/j.jclepro.2021.128408> (2021).
26. Bakay, M. S. & Ağbulut, Ü. Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *J. Clean. Prod.* **285**, 125324. <https://doi.org/10.1016/j.jclepro.2020.125324> (2021).
27. Wang, W. J., Men, C. Q. & Lu, W. Z. Online prediction model based on support vector machine. *Neurocomputing* **71**(4–6), 550–558. <https://doi.org/10.1016/j.neucom.07.020> (2008).
28. Li, J., Zhang, B. & Shi, J. Combining a genetic algorithm and support vector machine to study the factors influencing CO₂ emissions in Beijing with scenario analysis. *Energies* **10**(10), 1520. <https://doi.org/10.3390/en10101520> (2017).
29. Saleh, C., Dzakiyullah, N. R. & Nugroho, J. B. Carbon dioxide emission prediction using support vector machine. *IOP Conf. Ser. Mater. Sci. Eng.* **114**(1), 012148. <https://doi.org/10.1088/1757-899X/114/1/012148> (2016).
30. Ang, B. W. & Su, B. Carbon emission intensity in electricity production: A global analysis. *Energy Policy* **94**, 56–63. <https://doi.org/10.1016/j.enpol.2016.03.038> (2016).
31. <https://www.mee.gov.cn/ywgz/ydqhbh/wsqtz/202012/W020201229610353816665.pdf>.
32. <http://www.stats.gov.cn/was5/web/search?channelid=288041&andsen=%E4%B8%AD%E5%9B%BD%E8%83%BD%E6%BA%90%E7%BB%9F%E8%AE%A1%E5%B9%B4%E9%89%B4>.
33. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**(3), 273–297. <https://doi.org/10.1007/BF00994018> (1995).
34. Iba, K. Reactive power optimization by genetic algorithm. In *Power Industry Computer Application Conference*. <https://doi.org/10.1109/pica.1993.291017> (IEEE, 1994).
35. Han, J. & Xirong, Z. The index system of establishment of carbon emission warning for our country. In *Proceedings of the Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE2011) November 19–20, Melbourne, Australia*. https://doi.org/10.1007/978-3-642-25194-8_73 (Springer, 2011).

Acknowledgements

This work is supported by Natural Science Foundation of Hubei Province (Grant No. 2022CFC033), Hubei Key Laboratory of Construction and Management in Hydropower Engineering (China Three Gorges University) Open Fund (Grant No. 2020KSD15), and National Natural Science Foundation of China (52075292).

Author contributions

H.Y.: Conceptualization, Funding acquisition, Project administration; Y.Y.: Data curation, Investigation; B.L.: Data curation, Investigation; B.L.: Formal analysis, Investigation, Methodology, Visualization; Y.G.: Investigation, Methodology; Y.W.: Investigation, Project administration; Z.G.: Formal analysis, Visualization; R.M.: Methodology, Software, Visualization. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-31022-y>.

Correspondence and requests for materials should be addressed to B.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023