



CO₂ emission prediction based on carbon verification data of 17 thermal power enterprises in Gansu Province

Wei Shi¹ · Jiapeng Yang¹ · Fuwei Qiao² · Chengyuan Wang³ · Bowen Dong³ · Xiaolong Zhang³ · Sixue Zhao¹ · Weijuan Wang¹

Received: 12 September 2023 / Accepted: 2 December 2023 / Published online: 11 December 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The energy and power industry is an important field for CO₂ emission reduction. The CO₂ emitted by thermal power enterprises is a major cause of global climate change, and also a key challenge for China to achieve the goals of “carbon peaking and carbon neutrality.” Therefore, it is essential to scientifically and accurately predict the CO₂ emissions of key thermal power enterprises in the region. This will guide carbon reduction strategies and policy recommendations for leaders, and also provide a valuable reference for similar regions globally. This study utilizes the factor analysis method to extract the common factors influencing CO₂ emissions based on the carbon verification data of 17 thermal power enterprises in Gansu Province. Additionally, the DISO (distance between indices of simulation and observation) index is employed to comprehensively evaluate three prediction models, namely multiple linear regression, support vector regression, and GA-BP neural network. Ultimately, this study provides a reasonable prediction of CO₂ emissions for the aforementioned enterprises in Gansu Province. The results show that the three common factors obtained by factor analysis, namely energy consumption and output factor, energy quality factor, and energy efficiency factor, can effectively predict the CO₂ emissions from thermal power enterprises. In the three prediction models, GA-BP neural network has the best overall performance with DISO value of 0.95, RMSE value of 11848.236, and MAE value of 7880.543. Over the period 2022–2030, CO₂ emissions from 17 thermal power enterprises in Gansu Province are predicted to increase. Under the low-carbon, scenario baseline, and high-carbon scenarios, the CO₂ emissions will reach 71.58 Mt, 79.25 Mt, and 87.97 Mt, respectively, by 2030.

Keywords CO₂ emissions · Prediction model · Thermal power · Factor analysis · GA-BP neural network

Introduction

Human development is facing major challenges due to global warming and other climate crises caused by greenhouse gas emissions. The primary issues stem from human overreliance on energy sources like fossil fuels, resulting in

a continuous increase in greenhouse gas levels within the atmosphere (Anser et al. 2020). In order to cope with this challenge, reducing CO₂ emissions, improving energy efficiency, and achieving low-carbon development have become urgent tasks facing the international community. In China, the attainment of “carbon peaking and carbon neutrality” objectives within the region holds significant importance for the nation’s progress towards low-carbon development (Chen et al. 2022, Tian and Qi 2023). The energy and power industry is a key area for CO₂ emission reduction, and CO₂ emission reduction in the thermal power industry is crucial for China to achieve “carbon peaking and carbon neutrality.” This study uses measured data, which can more accurately predict the CO₂ emission trends of key thermal power enterprises in the region, compared with statistical data. It provides a scientific basis for leaders to formulate carbon emission reduction policies and measures. On a global scale, it can also serve as a reference for other countries to formulate

Responsible Editor: V.V.S.S. Sarma

✉ Wei Shi
shiwei0716@163.com

¹ College of Geography and Environmental Science, Northwest Normal University, Lanzhou 730070, Gansu, China

² College of Economics, Northwest Normal University, Lanzhou 730070, Gansu, China

³ Gansu Academy of Eco-environmental Sciences, Lanzhou 730000, Gansu, China

CO₂ emission targets and strategies for similar regions or scenarios. The purpose of this study is to provide a theoretical basis and decision-making reference for policymakers to formulate CO₂ emission targets and measures for thermal power generation areas, contributing to the low-carbon development of global thermal power generation and global climate action.

At present, research in the field of CO₂ emission forecasting has produced many valuable academic results. Numerous scholars employ the STIRPAT model, scenario analysis, and various techniques to forecast CO₂ emissions. These approaches are frequently integrated with the grey prediction model (Yin et al. 2023), LEAP model (Cai et al. 2023, Wang et al. 2022b), and system dynamic model (Feng et al. 2013) for assessing CO₂ emission predictions. Zhao et al. (2022) employed the extended STIRPAT model to assess the factors influencing China's CO₂ emissions and subsequently utilized the model's regression outcomes to project CO₂ emissions for the period spanning 2020 to 2040 across six distinct scenarios. The findings indicated that China's CO₂ emissions might reach their peak in 2029, as per the established baseline scenario. Ma et al. (2019) employed the LEAP model to predict CO₂ emissions in the Beijing-Tianjin-Hebei metropolitan areas for highway passenger transport. The results indicate that total regional carbon emissions will continue to increase by 2030. Wang et al. (2022a), Wei et al. (2023), and Rao et al. (2023) applied the STIRPAT model and scenario analysis to forecast the timing of CO₂ emissions peaking in Shanghai, Henan Province, and Hubei Province. Their findings suggest that achieving the carbon peak before 2030 is attainable through a low-carbon and green development approach. In addition, Hamzacebi and Karakurt (2015) used a gray prediction model to forecast that Turkey's energy-related CO₂ emissions will reach 496,404 Mt in 2025. Mirzaei and Bekri (2017) employed system dynamic models to predict that Iran's total CO₂ emissions would reach 985 Mt by 2025. Wakiyama and Kuramochi (2017) adopted scenario analysis and time series regression models to predict CO₂ emissions from electricity in the Japanese residential sector for the period of 2016 to 2030, demonstrating that Japan can surpass its post-2020 emission reduction target. Babbar et al. (2021) predicted the carbon sequestration in Sariska Tiger Reserve for the year 2035 using Markov chain and InVEST model and found that the area will reduce 0.107 Tg of carbon from 2018 to 2035. Handayani et al. (2022) used the LEAP model to analyze the power industry in the Association of Southeast Asian Nations, indicating that greenhouse gas emissions will reach zero emissions by 2050.

Due to the widespread adoption and utilization of artificial intelligence algorithms, an increasing number of researchers have incorporated machine learning techniques, particularly artificial neural networks, into recent studies on predicting CO₂ emissions (Zhang et al. 2020).

Machine learning is presently a prominent area of focus in CO₂ emission prediction research. Through the training and learning of known data, it can predict CO₂ emissions efficiently and accurately (Zhao et al. 2023). It has good adaptability to nonlinear problems and has become a useful tool for predicting CO₂ emissions (Acheampong and Boateng 2019, Zhang et al. 2022a). Several scholars have conducted relevant research on CO₂ emissions in China as a whole or in some regions. Qiao et al. (2020) introduced a novel hybrid algorithm in conjunction with a genetic algorithm to enhance the traditional least squares support vector machine (LSSVM) model. Their study, which relied on CO₂ emission data from developed countries, developing countries, and the global dataset, revealed an upward trend in China's CO₂ emissions from 2018 to 2025. Zhang et al. (2022b) constructed a genetic algorithm (GA) to enhance a BP neural network for forecasting building-related CO₂ emissions in Jiangsu Province. Their findings indicate a future decline in the total CO₂ emissions from buildings in the province. Su et al. (2022) recommended utilizing the monarch butterfly optimization approach to optimize the GM (1,1) model for predicting total primary energy production and CO₂ emissions in Tianjin. The outcomes indicated that Tianjin's CO₂ emissions are projected to peak in 2031 at 65.1009 Mt. Li et al. (2023) applied the random forest model for screening factors affecting CO₂ emissions. Subsequently, they utilized the BP neural network to predict China's CO₂ emissions, revealing that a peak in these emissions can be expected in 2030 under the 14th Five-Year Plan scenario. Some scholars have also conducted carbon-related research in other countries and regions outside China. Ameyaw et al. (2020) predicted the CO₂ emissions of West Africa countries from 2015 to 2030 using a recurrent neural network bidirectional long short-term memory algorithm and found that the future emission levels will be on an upward trend. The scholar (Ağbulut 2022) used three machine learning algorithms, deep learning, support vector machine, and artificial neural network, to predict Turkey's transportation CO₂ emissions, and found that all three algorithms reflect high prediction accuracy. Aryai and Goldsworthy (2023) proposed a particle swarm optimization extreme random tree regression model and showed that it outperforms long- and short-term memory and extreme learning machine for day-ahead prediction of emission intensity in the Australian national electricity market. Seo and Park (2023) used six operating parameters of diesel vehicles to predict their CO₂ emissions based on artificial neural networks, and found that engine torque and fuel/air ratio can improve the accuracy of prediction.

As shown above, numerous scholars have undertaken substantial research of both theoretical significance and practical value in the field of predicting CO₂ emissions from thermal power generation. Nonetheless, current research primarily centers on forecasting CO₂ emission trends at the national,

provincial, or urban levels, with limited attention devoted to CO₂ emission predictions for provincial-level enterprises. Meanwhile, CO₂ emission studies based on measured activity data are also rare and need to be further expanded.

Gansu province, situated in western China, serves as a vital energy hub with an extensive thermal power generation capacity and robust peaking capabilities. These factors are instrumental in ensuring regional power supply stability and fostering economic development (Shi et al. 2022). Thermal power enterprises are the basic unit of the power industry, and also one of the main sources of CO₂ emissions. This study initiates its investigation at the enterprise level within Gansu Province, encompassing a substantial portion of the thermal power enterprises operating in the region. Drawing from the 2021 annual greenhouse gas emission reports and carbon verification data of these enterprises, it employs factor analysis to identify common factors among the indicators influencing CO₂ emissions in the thermal power sector in Gansu Province. Subsequently, the study utilizes the CCHZ-DISO assessment system to meticulously evaluate three prediction models: multiple linear regression, support vector regression (SVR), and genetic algorithm-based backpropagation (GA-BP) neural network. The aim is to rigorously select the optimal prediction model for making accurate projections of CO₂ emissions in thermal power enterprises within Gansu Province. Ultimately, based on the research results, this paper provides targeted carbon reduction strategies and policy recommendations.

Data and methodology

Data sources

The data for this research is sourced from the 2021 Annual Greenhouse Gas Emission Report and Carbon Verification Materials of 17 thermal power enterprises within Gansu Province. In 2021, the total thermal power generation in Gansu Province was 104.472 billion kWh, and the total power generation of the 17 thermal power enterprises involved in this study was 60.474 billion kWh, accounting for 57.89%, which is representative. These units are all key greenhouse gas emission units. According to the requirements of the Department of Ecology and Environment of Gansu Province, they calculated their greenhouse gas emissions in 2021 according to the “Enterprise Greenhouse Gas Emission Accounting Method and Reporting Guide-Power Generation Facilities” and the “Enterprise Greenhouse Gas Emission Reporting Verification Guide (Trial).” At the same time, they also accepted the on-site and document verification by a third-party verification agency organized by the Gansu Academy of Eco-environmental Sciences, ensuring the authenticity, accuracy, and completeness of the data. The

data indicators and their specific sources are presented in Table 1.

Methodology

This paper employs factor analysis to extract the common components of CO₂ emission determinants of thermal power enterprises. These components are then used as input variables for the prediction model, which compares the performance of three models: multiple linear regression, SVR, and the GA-BP neural network for forecasting TEE. To evaluate the overall performance of each model, we employ the DISO index and select the optimal one for CO₂ emission prediction.

Factor analysis

The technique of factor analysis finds extensive application in diminishing the dimensions within a high-dimensional variable space. It employs multiple comprehensive variables to depict the connection among numerous initial variables (Bi et al. 2022, Lu et al. 2020). This study adopts factor analysis to analyze the indicators of factors affecting CO₂ emissions from thermal power enterprises, transforms the data of several indicators into several comprehensive variables by dimensionality reduction, and reveals the dominant factors affecting CO₂ emissions from thermal power enterprises in Gansu Province by extracting common factor.

In the selection of influencing factors, scholars have shown that economic scale, population size, energy structure, power structure, power generation efficiency, energy consumption, power supply standard coal consumption, and other factors have an impact on CO₂ emissions in the power industry (Chen et al. 2021, Wang et al. 2022a). This paper sets the following seven indicators based on the 2021 Annual Greenhouse Gas Emission Report and Carbon Verification Materials of 17 thermal power enterprises in Gansu Province: power generation of enterprise units (PGU), power supply of enterprise units (PSU), fuel consumption by fossil fuel (FCF), lower heating value of fossil fuel (LHV), carbon content per unit calorific value (CCV), coal consumption per unit electricity supply (CCS), and operating hours of enterprise units (OHU). This study develops an indicator framework for factors influencing CO₂ emissions within Gansu Province’s thermal power enterprises. Employing SPSS software, it extracts common factors from multiple indicators related to CO₂ emission influence, aiming to unveil the underlying relationships and patterns among these factors. Concurrently, this research employs the common factors derived from factor analysis as input variables for the prediction model. This strategic approach serves to diminish the intricacy and dimensionality of the prediction model while enhancing its generalizability and stability.

Table 1 Data indicators and specific sources involved in this study

No.	Indicator	Code	Unit	Data source and explanation
1	Power generation of enterprise units	PGU	MWh	2021 Annual Greenhouse Gas Emission Report, enterprise monthly production report for 2021, enterprise production data and emission summary table
2	Power supply of enterprise units	PSU	MWh	2021 Annual Greenhouse Gas Emission Report, enterprise monthly production report for 2021, enterprise production data and emission summary table
3	Fuel Consumption by Fossil Fuel	FCF	t	2021 Annual Greenhouse Gas Emission Report, enterprise monthly production report for 2021, enterprise coal consumption ledger for 2021
4	Lower heating value of fossil fuel	LHV	GJ/t	2021 Annual Greenhouse Gas Emission Report, enterprise fossil fuel combustion emission table
5	Carbon content per unit calorific value	CCV	tC/GJ	2021 Annual Greenhouse Gas Emission Report, enterprise fossil fuel combustion emission table
6	Coal consumption per unit electricity supply	CCS	tce/MWh	2021 Annual Greenhouse Gas Emission Report, enterprise production data and emission summary table
7	Operating hours of enterprise units	OHU	h	2021 Annual Greenhouse Gas Emission Report, enterprise production data and emission summary table
8	Purchased electricity carbon emission	PEE	tCO ₂	2021 Annual Greenhouse Gas Emission Report, enterprise purchased electricity carbon emission table
9	Fuel combustion carbon emission	FCE	tCO ₂	2021 Annual Greenhouse Gas Emission Report, enterprise production data and emission summary table
10	Total enterprise carbon emission	TEE	tCO ₂	TEE=(PEE)Purchased electricity carbon emission + (FCE)Fuel combustion carbon emission

The data of each indicator is obtained from the 2021 Annual Greenhouse Gas Emission Report as a priority, and then supplemented from other data sources

Prediction model construction

In this study, three prediction models of multiple linear regression, SVR, and GA-BP neural network are constructed for enterprise units. The introduction and specific content of the model construction are as follows:

Multiple linear regression model

Multiple linear regression is a regression analysis method that describes the relationship between multiple independent and dependent variables through a linear function, then determines the parameters of the linear function by minimizing the error, and finally uses the obtained regression equation to predict the change in the dependent variable.

Multiple linear regression is a statistical method that employs a linear function to delineate the relationships among multiple independent variables and a dependent variable. It accomplishes this by minimizing errors to ascertain the parameters of the linear function. Subsequently, the derived regression equation is utilized to forecast changes in the dependent variable.

The underlying principle revolves around the assumption of a linear functional relationship between the dependent variable Y and the general variables $x_1, x_2, x_3, \dots, x_k$. This relationship typically follows the general form outlined below:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_k x_k + \mu \quad (1)$$

In the formula, $\beta_j (j = 1, 2, 3, \dots, k)$ is a linear regression coefficient, β_0 is a linear regression constant, μ is a random error, and k represents the number of independent variables. The matrix expression of n random equations expressed by Eq. (2) is as follows:

$$Y = X\beta + U \quad (2)$$

In the formula, Y is the $n \times 1$ vector of the dependent variable data, X is the $n \times k$ matrix of the independent variable data, β is the $k \times 1$ vector of the regression coefficient, U is the $n \times 1$ vector of the random error data.

SVR model

Support vector regression is a machine learning method capable of effectively addressing problems related to function fitting and regression prediction. The core of this model is to find a hyperplane so that the distance from all data to this hyperplane is minimized. The study of finite samples can theoretically obtain the global optimal solution. In this study, the ε -SVR model is selected to predict the CO₂ emissions of thermal power enterprises. The steps are as follows:

For the given training set as in Eq. (3):

$$\begin{cases} T = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_n, y_n)\} \\ x_i \in R^n, y_i \in R, i = 1, 2, \dots, n \end{cases} \quad (3)$$

In the formula, x_i is the input factor, y_i is the expected value, and n is the total number of data samples in the training set.

Set a linear function expression on R^n as:

$$y = \omega \cdot x + b \quad (4)$$

In the formula, ω is the weight vector, $b \in R$, SVR operates on the principle of structural risk minimization, aiming to identify the optimal regression function. This transforms the problem of function estimation into the optimization Eqs. (5) and (6).

$$\min \frac{1}{2} \|\omega\|^2 + C \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (5)$$

$$s.t. \begin{cases} y_i - (\omega \cdot x_i) - b \leq \varepsilon + \xi_i, i = 1, 2, \dots, n \\ (\omega \cdot x_i) + b - y_i \leq \varepsilon + \xi_i^*, i = 1, 2, \dots, n \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases} \quad (6)$$

In the formula, C represents the penalty parameter, ε represents the insensitive coefficient, and ξ_i and ξ_i^* represent the slack variables.

During the solving process, it is common to convert the model into its dual counterpart. The kernel function $K(x_i \cdot y_i)$ is utilized to convert the linear regression problem into a non-linear regression problem within the Hilbert space. The constructed ε -SVR model is as follows:

$$\min \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i \cdot y_j) + \varepsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) - \sum_{i=1}^n y_i(\alpha_i^* - \alpha_i) \quad (7)$$

$$s.t. \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i \geq 0, \alpha_i^* \leq \frac{C}{n}, i = 1, 2, \dots, n \end{cases} \quad (8)$$

Then, the optimal solution is $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \bar{\alpha}_2, \bar{\alpha}_2^*, \dots, \bar{\alpha}_n, \bar{\alpha}_n^*)^T$, and the final regression function is shown in Eq. (9).

$$f(x) = \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i)K(x_i \cdot y_i) + \bar{b} \quad (9)$$

In the formula, \bar{b} is the bias term of the support vector regression model.

GA-BP neural network model

This study employs a GA-BP neural network to predict CO₂ emissions from thermal power enterprises. The GA-BP neural network combines genetic algorithm (GA) and back propagation (BP) neural network techniques. GA optimizes

the initial weights and thresholds of BP, enhancing its convergence speed and accuracy.

MATLAB software is used for GA-BP neural network programming, following the flowchart in Fig. 1: CO₂ emission and common factor data from Gansu Province’s thermal power enterprises are imported. An 80-20 split creates training and test sets, respectively. Both input and output variables for these sets are normalized to the [0,1] interval, facilitating network training and testing. A three-layer BP neural network is constructed: the input layer matches input variables, the hidden layer has nine nodes, and the output layer corresponds to CO₂ emissions. The neural network runs for a maximum of 5000 iterations, with an error threshold of 1e−6 and a learning rate of 0.01. The genetic algorithm optimizes initial BP weights and thresholds over 50 generations with a population size of 5. Parameters are encoded as real numbers between − 1 and 1 with an accuracy of 1e−6. The selection function is a normal geometric selection with a parameter of 0.09; the crossover employs an arithmetic function with a parameter of 2, and the variance function is non-uniform with parameters [2 50 3]. Mean square error serves as the fitness function, and the best individual of each generation advances to the next.

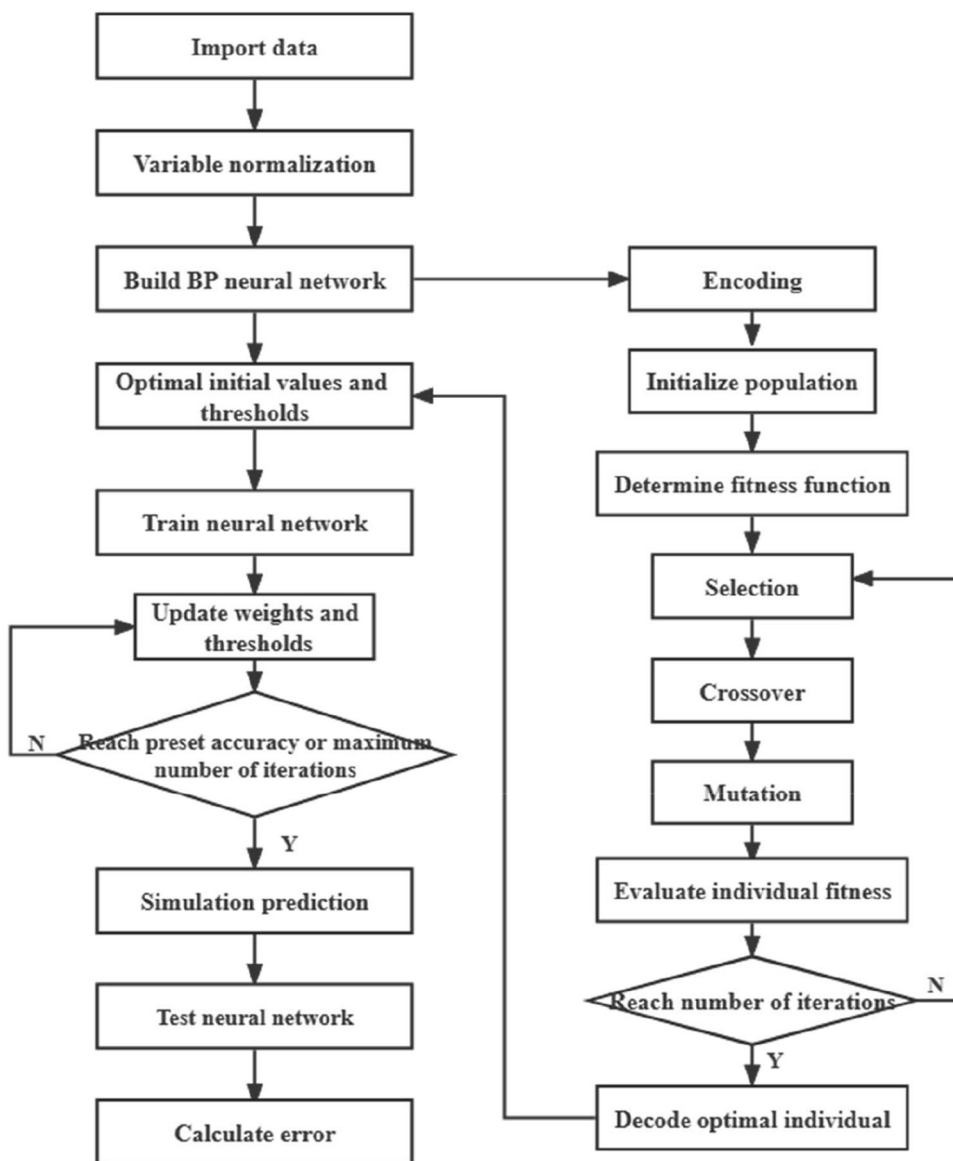
Repetition ends when a preset iteration limit is reached or fitness function value reaches a threshold. The best individual, which includes the best weights and thresholds, is then assigned to the BP neural network. Training begins with the training set to calculate training error. The BP neural network is then tested using the test set, and the test error is computed.

CCHZ-DISO Assessment System

In this study, we employed the CCHZ-DISO (Chen, Chen, Hu, and Zhou distance between indices of simulation and observation) assessment system to assess the overall performance of the three prediction models and identify the most suitable model. The DISO index is a statistical measurement comprehensive index constructed by Hu et al. (2022, 2019), which can simulate the distance between each observation index and quantify the overall performance of the model.

The CCHZ-DISO assessment system can solve the scientific problems of comprehensive quantitative evaluation of big data and models when multiple different system indicators are contradictory, and has very good expansibility and flexibility for the selection of statistical indicators (Zhou et al. 2021). Deng et al. (2021) constructed a comprehensive statistical index DISO using four statistical indicators: bias,

Fig. 1 GA-BP neural network model flow chart



correlation coefficient, root mean square error, and relative standard deviation to measure the performance of soil moisture and heat transfer simulation. In addition, Kalmár et al. (2021) used DISO index and Taylor diagram to evaluate the performance of the RegCM4.5 model in simulating precipitation, indicating the overall performance of DISO index energy simulation, and the results of DISO index evaluation are not much different from those of Taylor diagram.

The calculation of DISO index can be carried out in the following three steps:

Step 1: Calculate the statistical measure of each model for the reference data OBS; the statistical indicators are n , in the form of $(s_i^1, s_i^2, \dots, s_i^n)$, where $i=0, 1, \dots, m$; m is the

number of models. $(s_0^1, s_0^2, \dots, s_0^n)$ is the statistical index value of OBS relative to itself.

Step 2: Normalize all statistical indicators by dividing the difference between the maximum and minimum values.

$$(nors_i^1, nors_i^2, \dots, nors_i^n) = \left(\frac{s_i^1}{p^1}, \frac{s_i^2}{p^2}, \dots, \frac{s_i^n}{p^n} \right) \quad (10)$$

In the formula, $p^j = \max(\max(s_i^j) - \min(s_i^j), |\max(s_i^j)|, |\min(s_i^j)|)$, $i = 0, 1, \dots, m, j = 1, 2, \dots, n$ and the normalized value range is $[-1 \sim 1]$.

Step 3: Use the Euclidean distance between $(nors_i^1, nors_i^2, \dots, nors_i^n)$ to calculate the DISO index:

$$DISO_i = \sqrt{(nors_i^1 - nors_0^1)^2 + (nors_i^2 - nors_0^2)^2 + \dots + (nors_i^n - nors_0^n)^2} \tag{11}$$

When $i = 0$, $DISO_0 = 0$ represents the distance between OBS and itself. The DISO index of the simulation model can be obtained by the formula (11). Enhanced overall performance is associated with a lower DISO index in the model. By quantifying the DISO value, one can conveniently and quantitatively assess the overall performance among different models.

Result analysis

Measurement of common factors of CO₂ emissions

In this study, the influencing factors (PGU, PSU, FCF, LHV, CCV, CCS, and OHU) were used as the original variables for factor analysis using the analysis software SPSS. Table 2 displays the test results for KMO and Bartlett. The KMO value stands at 0.717, and the significance of the Bartlett spherical test is 0.000. These outcomes signify a correlation among variables and affirm the appropriateness of the selected variables for factor analysis. The Pearson correlation coefficient and its significance level between the seven variables are shown by the correlation matrix in Table 3. The table reveals that the correlation coefficient between 15 variable pairs demonstrates statistical significance at a significance level of $p = 0.01$. There is a strong positive correlation between PGU and PSU ($r = 0.999$), while there is no significant linear relationship between CCV and CCS ($r = -0.001$). A significant positive correlation exists among the three variables: PGU, PSU, and

FCF, with a correlation coefficient (r) exceeding 0.9 and a p value (P) less than 0.01, suggesting the presence of common underlying factors. At the same time, the correlation coefficient between the variable of CCS and other variables has not reached a significant level, which may mean that it has no common potential factor with other variables. CCS is a key indicator measuring the energy efficiency level (Lin and Zhu 2020, Wang et al. 2018, Yu et al. 2023) and constituting the carbon verification data of thermal power enterprises. Based on the results of KMO and Bartlett’s test, this study uses principal component analysis to extract common factors from 7 variables including CCS, and performs factor rotation.

From the gravel diagram of the eigenvalues of the common factors (Fig. 2), it becomes evident that component F1 is the most significant common factor. The eigenvalues of the first three common factors F1, F2, and F3 are greater than or equal to 1 (4.014, 1.377, and 1.000, respectively). The first common factor, F1, contributes 57.34%, the second common factor, F2, contributes 19.670%, and the third common factor, F3, contributes 14.281%. In total, these three common factors cumulatively contribute to 91.29%. Therefore, it can be considered that these three common

Table 2 Test results of KMO and Bartlett

KMO measure of sampling adequacy		0.717
Bartlett’s test of sphericity	Approx. chi-squared	5979.143
	df	21
	Sig.	0.000

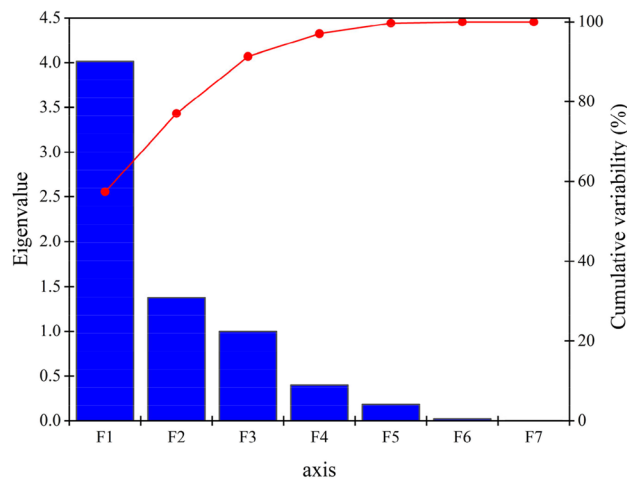


Fig. 2 Gravel diagram of common factor eigenvalues

Table 3 Correlation matrix

Variable	FCF	LHV	CCV	PGU	PSU	CCS	OHU
FCF	1	0.386**	0.381**	0.971**	0.971**	0.005	0.693**
LHV		1	0.804**	0.372**	0.371**	-0.039	0.534**
CCV			1	0.303**	0.299**	-0.001	0.582**
PGU				1	0.999**	-0.006	0.599**
PSU					1	-0.003	0.596**
CCS						1	-0.010
OHU							1

The table with double asterisks (**) indicates $P < 0.01$, with asterisks (*) indicating $P < 0.05$

factors are the representatives of the original data and can participate in further research.

Utilizing the maximum variance method, we achieve orthogonal rotation of the three comprehensive factors, resulting in the rotated component matrix, which is illustrated in Table 4. By considering the load, we can elucidate the connection between F1, F2, and F3 and the original input variables. From the table, it is evident that the load coefficients of the variables FCF, PGU, and PSU on F1 exceed 0.9, while their load coefficients on F2 and F3 are notably lower. This observation suggests that these three variables are primarily influenced by the common factor F1, with limited associations with F2 and F3. The load coefficients of LHV and CCV on F2 are greater than 0.9, while the load coefficients on F1 and F3 are very low, indicating that these two variables are mainly affected by F2 and have little relationship with F1 and F3. In addition, OHU has a higher load on F1 and F2, which is 0.586 and 0.590, respectively, indicating that this variable has a certain correlation with F1 and F2. The common factor F3 contains only one variable of CCS, and its load coefficient is 1.0.

Based on the above analysis and the component score coefficient matrix (Table 5), each common factor is given an appropriate label to reflect the potential concept it represents, as follows: The first common factor F1 is named “energy consumption and output factor,” which reflects the combined level of four variables: FCF, PGU, PSU, and OHU, mainly related to energy use or production. The second common factor of F2 is named “energy quality factor,” which reflects

the combined level of three variables: LHV, CCV, and OHU, and is mainly related to the quality and characteristics of fuels used by thermal power producers. The third common factor F3, is named “energy efficiency factor,” which mainly reflects one variable, CCS, and is related to fuel utilization and energy saving and optimization. Next, the three common factors are used as input variables for the multiple linear regression, SVR, and BP neural network models to construct the CO₂ emission prediction model.

CO₂ emission prediction of thermal power enterprises

CO₂ emission prediction model of thermal power enterprises

In this study, the three common factors F1, F2, and F3 obtained from factor analysis were incorporated as input variables in the prediction model. This was done to reduce model complexity and dimensionality, ultimately enhancing its generalization ability and stability. The results of the three models of multiple linear regression, SVR, and BP neural network are presented as follows:

Multiple linear regression model

Within the framework of the multiple linear regression model, we employ the three common factors as predictors for estimating the total enterprise carbon emission (TEE)

Table 4 Rotated component matrix

Variable	Synthetic factor		
	1	2	3
Fuel consumption by fossil fuel (FCF)	0.961	0.238	0.008
Lower heating value of fossil fuel (LHV)	0.195	0.905	− 0.035
Carbon content per unit calorific value (CCV)	0.143	0.939	0.015
Power generation of enterprise units (PGU)	0.976	0.166	− 0.006
Power supply of enterprise units (PSU)	0.976	0.163	− 0.003
Coal consumption per unit electricity supply (CCS)	0.001	− 0.012	1.000
Operating hours of enterprise units (OHU)	0.586	0.590	0.001

Table 5 Component score coefficient matrix

Variable	Component		
	1	2	3
Fuel consumption by fossil fuel (FCF)	0.324	− 0.071	0.008
Lower heating value of fossil fuel (LHV)	− 0.121	0.487	− 0.021
Carbon content per unit calorific value (CCV)	− 0.149	0.518	0.029
Power generation of enterprise units (PGU)	0.345	− 0.116	− 0.008
Power supply of enterprise units (PSU)	0.346	− 0.118	− 0.005
Coal consumption per unit electricity supply (CCS)	− 0.002	0.009	0.999
Operating hours of enterprise units (OHU)	0.100	0.217	0.008

Table 6 ANOVA table for multiple linear regression model

Model	Sum of squares	df	Mean square	F	Sig.
Regression	4,671,441,256,244.730	3	1,557,147,085,414.910	5525.041	.000b
Residual	137,535,222,624.652	488	281,834,472.592		
Total	4,808,976,478,869.383	491			

of the dependent variables. To evaluate the model’s overall fit, we conducted an analysis of variance (ANOVA), as presented in Table 6. The outcomes indicate a highly statistically significant *F* value of 5525.041, with a *p* value less than 0.001. Figure 3 describes the comparison between the predicted value calculated by the multiple linear regression model and the actual CO₂ emission value. There is a strong linear relationship between the actual value and the predicted value in the figure. The confidence interval represents the range of the prediction error. Most of the points fall within the 95% confidence interval near the diagonal, indicating that the prediction model has high accuracy and calibration. Specifically, the *R*-squared of the model is 0.97148, the correlation coefficient *R* is 0.98564, the root mean square error (RMSE) of the data sample is 16,719.568, and the mean absolute error (MAE) is 12,381.481, which indicates that the regression results are more satisfactory and can be used for the prediction of emissions.

SVR model

In this study, MATLAB software was used to predict the TEE of enterprises based on SVR model with three common factors F1, F2, and F3 as input variables. After normalizing both the training and test set data, we employed the radial basis kernel function as our kernel. The penalty factor *C* was configured at 4.0, the radial basis function parameter *g* was set to 0.8, and the error tolerance *p* was established as 0.01. Figures 4 and 5 show the comparison between the predicted and actual values of CO₂ emissions in the training set and the test set. The diagram reveals that the *R*-squared for both the training and test datasets is 0.98244 and 0.98295, respectively. Additionally, the correlation coefficient *R* stands at 0.99120 and 0.98155 for the training and test sets, respectively, signifying a superior model fitting effect. Meanwhile, both the training and test sets display low values for MAE and RMSE. This points to the SVR model’s accurate predictive capability for CO₂ emissions in Gansu Province’s thermal power enterprises.

Fig. 3 Comparison of predicted and actual CO₂ emissions based on the multiple linear regression model

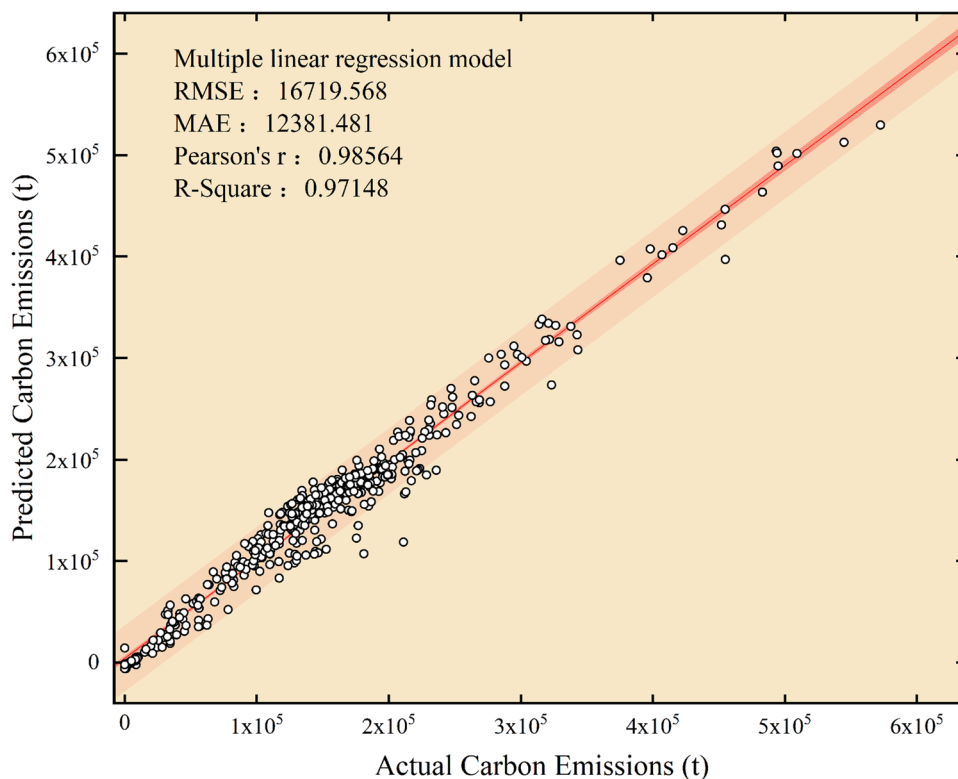


Fig. 4 Comparison of predicted and actual CO₂ emissions for the training set based on the SVR model

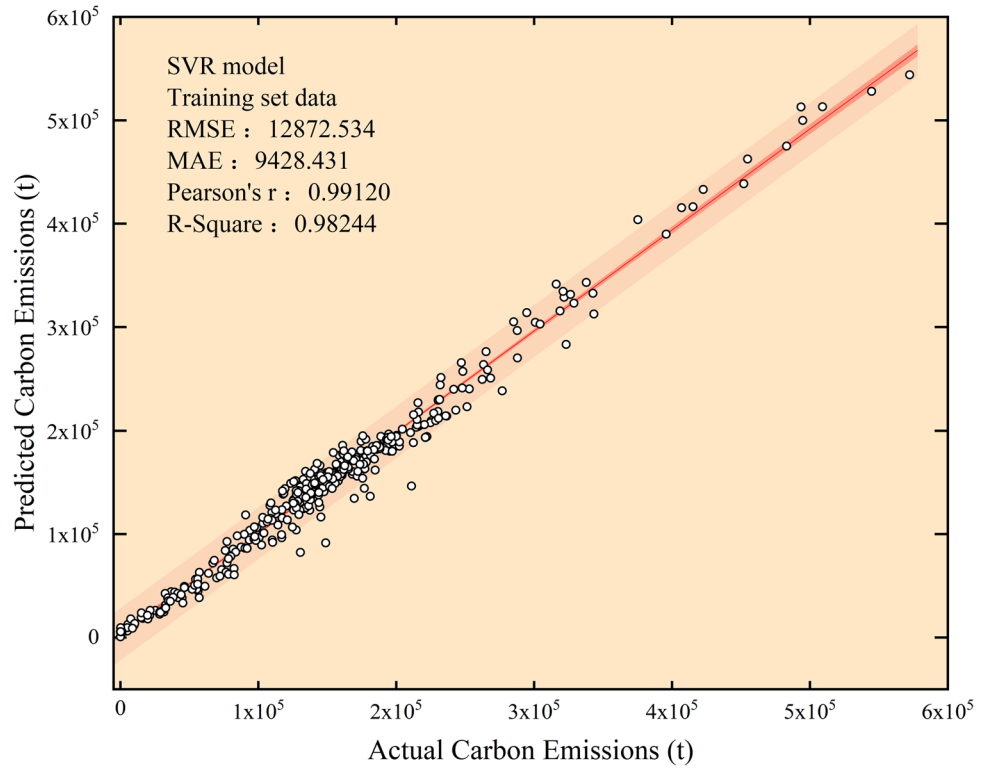


Fig. 5 Comparison of predicted and actual CO₂ emissions for the test set based on the SVR model

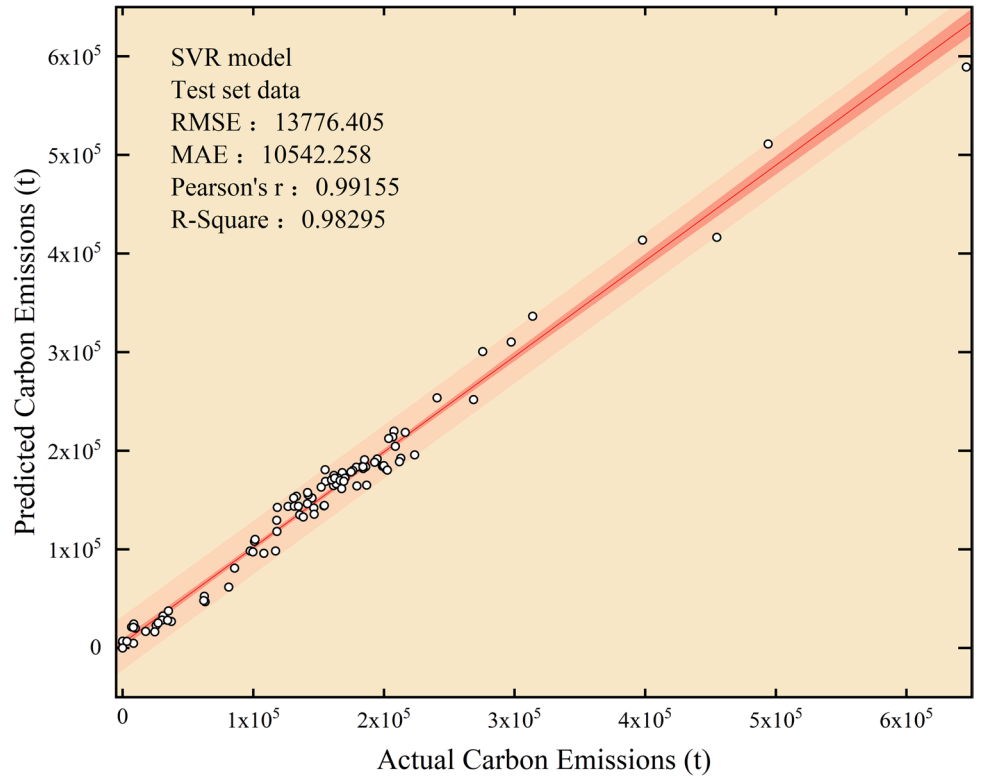


Fig. 6 Comparison of predicted and actual CO₂ emissions for the training set based on the GA-BP neural network

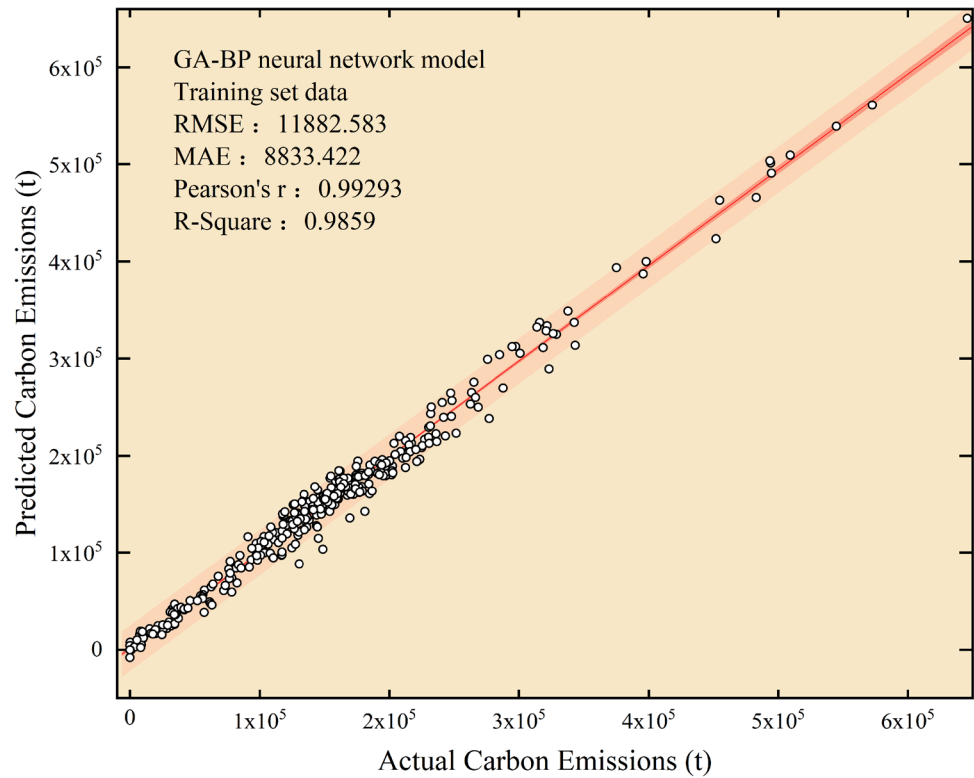
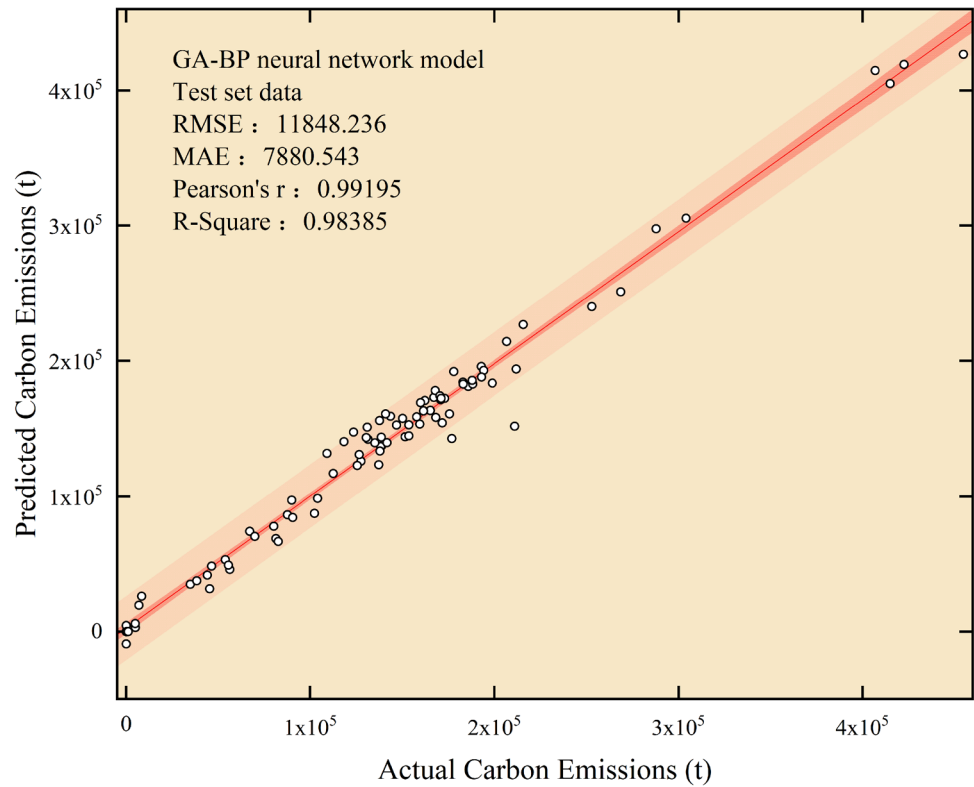


Fig. 7 Comparison of predicted and actual CO₂ emissions for the test set based on the GA-BP neural network



GA-BP neural network model

In this study, MATLAB software was used to predict the TEE of enterprises based on GA-BP neural network model with three common factors F1, F2, and F3 as input variables. The predicted and actual values of CO₂ emissions in the training set and test set are shown in Figs. 6 and 7. The figure illustrates that the GA-BP neural network model demonstrates high prediction accuracy and correlation in both the training and test sets. RMSE and MAE are diminutive, accompanied by Pearson's *r* and *R*-squared values approaching 1. These outcomes point to the model's adeptness in data fitting and its strong generalization capacity. At the same time, the evaluation index of the model on the test set is slightly better than that on the training set, indicating that the model does not have the problem of overfitting or underfitting. This indicates that the GA-BP neural network model serves as an efficient predictor for CO₂ emissions in Gansu Province's thermal power enterprises.

Comparative evaluation of models

In order to comprehensively compare and evaluate the three prediction models of the multiple linear regression model, SVR model, and GA-BP neural network, the CCHZ-DISO assessment system was used in this study. RMSE, MAE, and *R*-squared were selected to calculate the DISO index of each model according to formula (10) and formula (11). The SVR model and GA-BP neural network use the evaluation index of the test set to calculate the DISO index. The model's superior overall performance is indicated by the smaller DISO index value. Refer to Table 7 for the results. The DISO index calculations reveal values of 0.95, 1.18, and 1.41 for the GA-BP neural network, SVR, and multiple linear regression models, respectively. Among these models, the GA-BP neural network exhibits the lowest DISO value, signifying its superior overall performance. Consequently, it will be employed for CO₂ emission prediction.

CO₂ emission prediction based on GA-BP neural network model

CO₂ emission scenario setting

This study will set three CO₂ emission scenario models: low-carbon, benchmark, and high-carbon. In the past five years (2016–2021), the total thermal power generation

in Gansu Province increased from 704.18 to 104.472 billion kilowatt hours, with an annual growth rate of 8.2%, and the total energy consumption of electricity production and supply increased from 491.69 to 680.56 million tons of standard coal, with an annual growth rate of 6.7%. At the same time, “Gansu Province's 14th Five-Year Energy Development Plan” proposes that the scale of thermal power installed capacity will reach 35.58 million KW from 2020 to 2025, with an annual growth rate of 10%. It can be predicted that during the “14th Five-Year Plan” period (2020–2025), the energy consumption and output factor F1 will show a certain growth trend. In addition, during the “15th Five-Year Plan” period (2026–2030), the “Implementation Plan for Carbon Peaking in Gansu Province” pointed out that the region should gradually reduce coal consumption and achieve the goal of peaking CO₂ emissions by 2030. Therefore, the annual growth rate of energy consumption and output factor F1 during the “15th Five-Year Plan” period will be significantly lower than that during the “14th Five-Year Plan” period. Based on the above analysis, on the basis of considering historical data and existing planning, reflecting the policy effects under different scenarios, the annual growth rates of F1 energy consumption and output factors in low-carbon, benchmark, and high-carbon scenarios from 2021 to 2025 are set to 3%, 8%, and 12% respectively, and the annual growth rates of the three scenarios from 2026 to 2030 are set to 1%, 4%, and 6% respectively.

Among the 17 thermal power enterprises in Gansu Province studied in this paper, 12 adopt general bituminous coal or medium and high volatile bituminous coal, and the remaining 5 adopt lignite or mixed coal. The average carbon content per unit calorific value in 2021 is 27.2 TC/TJ. In order to realize the energy saving and carbon reduction plan of Gansu Province and establish the economic system of green low-carbon cycle development, the thermal power enterprises in the province will gradually improve the quality of fuel and reduce the dependence and consumption of low-quality coal (such as lignite and mixed coal). In this paper, the carbon content per unit calorific value of bituminous coal provided by the “Provincial Greenhouse Gas Inventory Guidelines (Trial)” is 26.1 TC/TJ as the reference value in 2030, and the source quality factor growth rate from 2021 to 2030. The annual growth rates of low-carbon, benchmark, and high-carbon scenarios are set to – 1.0%, – 0.5%, and 0%, respectively. In addition, the average power supply coal consumption of the 17 thermal power enterprises studied in

Table 7 Comparison of DISO values for the three models

Rank	Model	RMSE	MAE	<i>R</i> -squared	DISO
1	GA-BP neural network	11,848.236	7880.543	0.98385	0.95
2	SVR model	13,776.405	10,542.258	0.98295	1.18
3	Multiple linear regression	16,719.568	12,381.481	0.97148	1.41

this paper is 347.34 g/kwh in 2021. In order to promote the clean and low-carbon transformation of the power industry and achieve the goal of carbon peak and carbon neutralization on schedule, the “Implementation Plan for the Transformation and Upgrading of National Coal-Fired Power Units” points out that the average power supply coal consumption of national thermal power will be reduced to below 300 g/kwh in 2025. Based on this standard, this paper reflects the differences in technological progress and policy regulation under different scenarios on the basis of conforming to the actual situation and future goals of thermal power enterprises in Gansu Province. The annual growth rates of F3 energy efficiency factors under low-carbon, benchmark, and high-carbon scenarios by 2030 are − 5.2%, − 3.6%, and − 2.0%, respectively. The scenario of CO₂ emission prediction of thermal power enterprises in Gansu Province is set as shown in Table 8 below.

CO₂ emission scenario prediction results

The prediction results of CO₂ emissions under the three scenarios are shown in Fig. 8. The prediction results show that under the baseline scenario, the CO₂ emissions of 17 thermal power enterprises in Gansu Province show a steady growth trend, from 71.17 Mt in 2021 to 79.25 Mt in 2030, an increase of 8.08 Mt, equivalent to 11.36% of the level in 2021, with an annual growth rate of 1.20%. Under the low-carbon scenario, the CO₂ emissions of 17 thermal power enterprises in Gansu Province showed a slow growth trend, from 71.17 Mt in 2021 to 71.58 Mt in 2030, an increase of 0.41 Mt, equivalent to 0.58% of the level in 2021, with an average annual growth rate of 0.06%. In addition, the growth rate of CO₂ emissions in this scenario will further slow down after 2025. The average annual growth rate of CO₂ emissions from 2025 to 2030 is only 0.02%, and there is no significant change in CO₂ emissions. Under the high-carbon scenario, the CO₂ emissions of 17 thermal power enterprises in Gansu Province showed a rapid growth trend, from 71.17 Mt in 2021 to 87.97 Mt in 2030, an increase of 16.80 Mt, equivalent

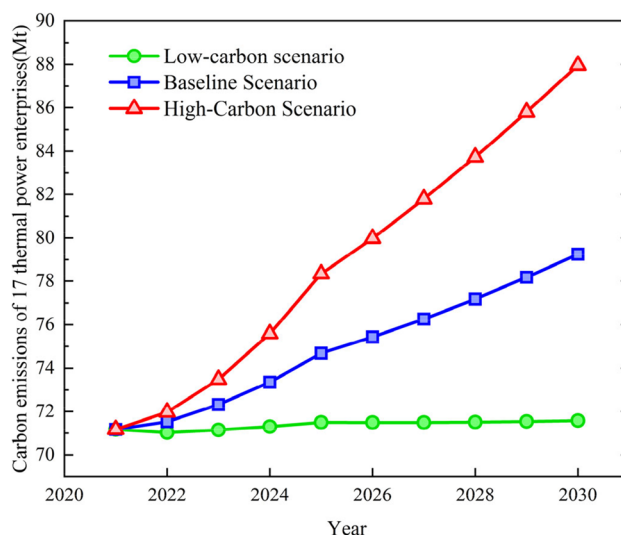


Fig. 8 Prediction results of three CO₂ emission scenarios: low carbon, benchmark, and high carbon

to 23.61% of the level in 2021, with an average annual growth rate of 2.38%.

By comparing the differences between different scenarios, in all years, the CO₂ emissions under the high-carbon scenario are always the highest, followed by the baseline scenario, and the CO₂ emissions under the low-carbon scenario are the lowest. By 2030, the CO₂ emissions of the high-carbon scenario will be 1.23 times that of the low-carbon scenario and 1.11 times that of the baseline scenario. Under the baseline and high-carbon scenarios, the growth rate of energy consumption and output is faster, which may be the main reason for the rapid growth of CO₂ emissions. In contrast, the CO₂ emissions of thermal power enterprises in low-carbon scenarios are relatively low, and the growth rate is slow. This is due to the implementation of measures such as controlling energy consumption and output, optimizing energy quality, and improving energy efficiency, which effectively slows down and controls the growth rate of CO₂ emissions of thermal power enterprises.

Table 8 CO₂ emission prediction scenario setting of thermal power enterprises in Gansu Province

Scenario mode	Years	Annual change rate (%)		
		F1 energy consumption and output factor	F2 energy quality factor	F3 energy efficiency factor
Low-carbon	2021~2025	3.0	− 1.0	− 5.2
	2026~2030	1.0	− 1.0	− 5.2
Baseline	2021~2025	8.0	− 0.5	− 3.6
	2026~2030	4.0	− 0.5	− 3.6
High-carbon	2021~2025	12.0	0	− 2.0
	2026~2030	6.0	0	− 2.0

The overall CO₂ emissions of 17 thermal power generation enterprises in Gansu Province are on the rise during 2022–2030. Although the CO₂ emissions have been effectively controlled under the low-carbon scenario, they have not reached the peak of CO₂ emissions. Due to the limitations of practical reasons, this paper does not predict the future deployment of carbon capture, utilization, and storage technology (CCUS) in thermal power plants in Gansu Province to further reduce CO₂ emissions. CCUS technology is an important CO₂ emission reduction technology. Recent research underscores the pivotal role of CCUS technology in curbing CO₂ emissions from power generation (Cai et al. 2022). Future investigations should delve deeper into applying CCUS technology to mitigate emissions in Gansu Province's thermal power sector, thus advancing the province's power industry towards carbon peak and neutrality goals.

Conclusions and policy implications

Conclusions

In this paper, the factor analysis method is used to measure the common factors of the CO₂ emission influencing factors of 17 thermal power enterprises in Gansu Province. Based on the three CO₂ emission prediction models of multiple linear regression, SVR, and GA-BP neural network, the DISO index is used to evaluate the overall performance of the prediction model. Based on setting three CO₂ emission scenarios of low-carbon, baseline, and high-carbon, the CO₂ emissions of these 17 thermal power enterprises during 2022–2030 were predicted by using GA-BP neural network prediction model. The main conclusions of this study are as follows:

1. Through factor analysis, the three common factors F1 energy consumption and output factor, F2 energy quality factor, and F3 energy efficiency factor can be effectively used to predict the CO₂ emissions of thermal power enterprises, and are in line with the current CO₂ emission verification work of thermal power enterprises.
2. Among the three CO₂ emission prediction models of multiple linear regression, SVR, and GA-BP neural network, the GA-BP neural network model has the best overall performance. Its DISO value is 0.95; RMSE and MAE are 11848.236 and 7880.543, respectively. The second is the SVR model, the DISO value is 1.18, and the overall performance is good. The DISO value of the multiple linear regression model is 1.41, and the overall performance is poor.
3. The overall CO₂ emissions of 17 thermal power enterprises in Gansu Province are on the rise during 2022–2030. Under the baseline and high-carbon scenarios,

the rapid growth of energy consumption and output is the main reason for the growth of CO₂ emissions. It is expected that by 2030, the CO₂ emissions under the baseline scenario will reach 79.25 Mt, and the CO₂ emissions under the high-carbon scenario will reach 87.97 Mt. However, under the low-carbon scenario, CO₂ emissions will reach 71.58 Mt by 2030, with an average annual growth rate of only 0.06%.

Policy implications

1. Reasonable planning of energy consumption and output, optimizing power scheduling. Power generation enterprises should formulate reasonable energy consumption and output plans according to the national and local carbon emission control requirements, and control the growth rate of fuel consumption and power generation. At the same time, the power grid should improve the flexibility and intelligence level of scheduling, accelerate the implementation of the flexibility transformation of existing thermal power units in the province, tap the peak shaving potential of thermal power units, encourage coal-fired units to increase efficient energy storage facilities, and establish a priority scheduling energy system that adapts to the characteristics of wind and light power. In addition, power generation enterprises also need to optimize the operation and start-up mode of the unit, make rational use of all kinds of peak shaving resources in the system, and give full play to the clean and efficient advantages of 600 MW and above large-capacity high-parameter units when undertaking basic load.
2. Improve and ensure the fuel quality of the power plant. Thermal power enterprises in Gansu Province should optimize the fuel structure, phase out low-quality and high-polluting fuels, and increase the proportion of clean and efficient fuel use. At the same time, strengthen fuel quality monitoring and management, improve fuel efficiency and safety. By giving priority to the release of high-quality production capacity of coal mine projects and ensuring the railway capacity of coal cross-regional transportation, the level of guarantee for coal production and transportation needs is improved. Under the same conditions, priority should be given to ensuring the fuel supply of coal-fired generating units with advanced energy efficiency. Give full play to the role of the market, stabilize the sharp fluctuation of the price of electric coal, ensure that the power plant burns the designed coal, and avoid the increase of the actual operation energy consumption of the unit caused by the fluctuation of fuel quality to the greatest extent.
3. Enhancing energy efficiency, increasing policy support for energy-saving and emission-reduction upgrades to power units, and advancing research and demonstra-

tions of CCUS technologies are crucial steps. Gansu Province should encourage thermal power generation companies to elevate their energy conservation and emissions reduction efforts. This can be achieved through the adoption of cutting-edge clean technologies and equipment, as well as by enhancing the operational efficiency and reliability of thermal power generation units. Furthermore, optimizing operational scheduling strategies, establishing a mechanism linking power generation to energy consumption levels, and promoting the increased electricity generation of coal-fired units with low coal consumption are essential objectives. Additionally, expediting the development of a well-rounded auxiliary service market mechanism and pricing system will ensure that peaking units participating in flexibility enhancements receive appropriate benefits.

Author contribution Wei Shi and Jiapeng Yang: formal analysis, data curation, and writing—original draft; Fuwei Qiao and Chengyuan Wang: writing—reviewing and editing; Bowen Dong and Xiaolong Zhang: conceptualization and methodology; Sixue Zhao and Weijuan Wang: supervision and project administration.

Funding This work was supported by the funding of the Key Research and Development Program of Gansu Province, China (grant number 21YF5FA028)

Data availability The data underlying this article cannot be shared publicly due to the sensitive nature of the data and the enterprise's privacy policies.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication We obtained explicit consent from the Gansu Academy of Eco-environmental Sciences to use their data in this research project.

Competing interests The authors declare no competing interests.

References

- Acheampong AO, Boateng EB (2019) 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>
- Ağbulut Ü (2022) Forecasting of transportation-related energy demand and CO₂ emissions in Turkey with different machine learning algorithms. *Sustainable Production and Consumption* 29:141–157. <https://doi.org/10.1016/j.spc.2021.10.001>
- Ameyaw B, Li Y, Annan A, Agyeman JK (2020) West Africa's CO₂ emissions: investigating the economic indicators, forecasting, and proposing pathways to reduce carbon emission levels. *Environ Sci Pollut Res* 27:13276–13300. <https://doi.org/10.1007/s11356-020-07849-7>
- Anser MK, Alharthi M, Aziz B, Wasim S (2020) Impact of urbanization, economic growth, and population size on residential carbon emissions in the SAARC countries. *Clean Techn Environ Policy* 22:923–936. <https://doi.org/10.1007/s10098-020-01833-y>
- Aryai V, Goldsworthy M (2023) Day ahead carbon emission forecasting of the regional National Electricity Market using machine learning methods. *Eng Appl Artif Intell* 123:106314. <https://doi.org/10.1016/j.engappai.2023.106314>
- Babbar D, Areendran G, Sahana M, Sarma K, Raj K, Sivadas A (2021) Assessment and prediction of carbon sequestration using Markov chain and InVEST model in Sariska Tiger Reserve. *India J Clean Prod* 278:123333. <https://doi.org/10.1016/j.jclepro.2020.123333>
- Bi Y, Wu J, Zhai X (2022) Quantitative prediction model of water inrush quantities from coal mine roofs based on multi-factor analysis. *Environ Earth Sci* 81:314. <https://doi.org/10.1007/s12665-022-10432-7>
- Cai L, Duan J, Lu X, Luo J, Yi B, Wang Y, Jin D, Lu Y, Qiu L, Chen S, Zhang H, Wang L (2022) Pathways for electric power industry to achieve carbon emissions peak and carbon neutrality based on LEAP model: a case study of state-owned power generation enterprise in China. *Comput Ind Eng* 170:108334. <https://doi.org/10.1016/j.cie.2022.108334>
- Cai L, Luo J, Wang M, Guo J, Duan J, Li J, Li S, Liu L, Ren D (2023) Pathways for municipalities to achieve carbon emission peak and carbon neutrality: a study based on the LEAP model. *Energy* 262:125435. <https://doi.org/10.1016/j.energy.2022.125435>
- Chen G, Hou F, Li J, Chang K (2021) Decoupling analysis between carbon dioxide emissions and the corresponding driving forces by Chinese power industry. *Environ Sci Pollut Res* 28:2369–2378. <https://doi.org/10.1007/s11356-020-10666-7>
- Chen L, Msigwa G, Yang M, Osman AI, Fawzy S, Rooney DW, Yap P-S (2022) Strategies to achieve a carbon neutral society: a review. *Environ Chem Lett* 20:2277–2310. <https://doi.org/10.1007/s10311-022-01435-8>
- Deng M, Meng X, Lu Y, Li Z, Zhao L, Hu Z, Chen H, Shang L, Wang S, Li Q (2021) Impact and sensitivity analysis of soil water and heat transfer parameterizations in community land surface model on the Tibetan Plateau. *J Adv Mod Earth Syst* 13:e2021MS002670. <https://doi.org/10.1029/2021MS002670>
- Feng YY, Chen SQ, Zhang LX (2013) System dynamics modeling for urban energy consumption and CO₂ emissions: a case study of Beijing, China. *Ecol Model* 252:44–52. <https://doi.org/10.1016/j.ecolmodel.2012.09.008>
- Hamzacebi C, Karakurt I (2015) Forecasting the energy-related CO₂ emissions of Turkey using a grey prediction model. *Energy Sources, Part A: Recov Utiliz Environ Effects* 37:1023–1031. <https://doi.org/10.1080/15567036.2014.978086>
- Handayani K, Anugrah P, Goembira F, Overland I, Suryadi B, Swandaru A (2022) Moving beyond the NDCs: ASEAN pathways to a net-zero emissions power sector in 2050. *Appl Energy* 311:118580. <https://doi.org/10.1016/j.apenergy.2022.118580>
- Hu ZY, Chen X, Zhou QM, Chen DL, Li JF (2019) DISO: A rethink of Taylor diagram. *Int J Climatol* 39:2825–2832. <https://doi.org/10.1002/joc.5972>
- Hu ZY, Chen DL, Chen X, Zhou QM, Peng YZ, Li JF, Sang YF (2022) CCHZ-DISO: a timely new assessment system for data quality or model performance from Da Dao Zhi Jian. *Geophys Res Lett* 49:9. <https://doi.org/10.1029/2022GL100681>
- Kalmár T, Pieczka I, Pongrácz R (2021) A sensitivity analysis of the different setups of the RegCM4.5 model for the Carpathian region. *Int J Climatol* 41:E1180–E1201. <https://doi.org/10.1002/joc.6761>

- Li Y, Huang S, Miao L, Wu Z (2023) Simulation analysis of carbon peak path in China from a multi-scenario perspective: evidence from random forest and back propagation neural network models. *Environ Sci Pollut Res* 30:46711–46726. <https://doi.org/10.1007/s11356-023-25544-1>
- Lin B, Zhu J (2020) Chinese electricity demand and electricity consumption efficiency: do the structural changes matter? *Appl Energy* 262:114505. <https://doi.org/10.1016/j.apenergy.2020.114505>
- Lu Y, Liu K, Xu S, Wang Y, Zhang Q (2020) Identifying flow units by FA-assisted SSOM—an example from the Eocene basin-floor-fan turbidite reservoirs in the Daluhu Oilfield, Dongying Depression, Bohai Bay Basin, China. *J Pet Sci Eng* 186:106695. <https://doi.org/10.1016/j.petrol.2019.106695>
- Ma H, Sun W, Wang S, Kang L (2019) Structural contribution and scenario simulation of highway passenger transit carbon emissions in the Beijing-Tianjin-Hebei metropolitan region, China. *Resour Conserv Recycl* 140:209–215. <https://doi.org/10.1016/j.resconrec.2018.09.028>
- Mirzaei M, Bekri M (2017) Energy consumption and CO₂ emissions in Iran, 2025. *Environ Res* 154:345–351. <https://doi.org/10.1016/j.envres.2017.01.023>
- Qiao WB, Lu HF, Zhou GF, Azimi M, Yang Q, Tian WC (2020) A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer. *J Clean Prod* 244:16. <https://doi.org/10.1016/j.jclepro.2019.118612>
- Rao C, Huang Q, Chen L, Goh M, Hu Z (2023) Forecasting the carbon emissions in Hubei Province under the background of carbon neutrality: a novel STIRPAT extended model with ridge regression and scenario analysis. *Environ Sci Pollut Res* 30:57460–57480. <https://doi.org/10.1007/s11356-023-26599-w>
- Seo J, Park S (2023) Optimizing model parameters of artificial neural networks to predict vehicle emissions. *Atmos Environ* 294:119508. <https://doi.org/10.1016/j.atmosenv.2022.119508>
- Shi W, Tang W, Qiao F, Sha Z, Wang C, Zhao S (2022) How to reduce carbon dioxide emissions from power systems in Gansu Province and analyze from the life cycle perspective. *Energies* 15:3560. <https://doi.org/10.3390/en15103560>
- Su Q, Wang H, Shi X, Li G, Sun K (2022) Research on energy supply and demand forecast and carbon neutralization path based on grey-monarch butterfly optimization model. *J Phys Conf Ser* 2247:012008. <https://doi.org/10.1088/1742-6596/2247/1/012008>
- Tian C, Qi L (2023) Empirical decomposition and peaking path of carbon emissions in resource-based areas. *J Clean Prod* 395:136372. <https://doi.org/10.1016/j.jclepro.2023.136372>
- Wakiyama T, Kuramochi T (2017) Scenario analysis of energy saving and CO₂ emissions reduction potentials to ratchet up Japanese mitigation target in 2030 in the residential sector. *Energy Policy* 103:1–15. <https://doi.org/10.1016/j.enpol.2016.12.059>
- Wang H, Li B, Khan MQ (2022a) Prediction of Shanghai electric power carbon emissions based on improved STIRPAT model. *Sustainability* 14:13068. <https://doi.org/10.3390/su142013068>
- Wang J, Wang R, Zhu Y, Li J (2018) Life cycle assessment and environmental cost accounting of coal-fired power generation in China. *Energy Policy* 115:374–384. <https://doi.org/10.1016/j.enpol.2018.01.040>
- Wang J, Li Y, Zhang Y (2022b) Research on carbon emissions of road traffic in Chengdu city based on a LEAP model. *Sustainability* 14:5625. <https://doi.org/10.3390/su14095625>
- Wei Z, Wei K, Liu J (2023) Decoupling relationship between carbon emissions and economic development and prediction of carbon emissions in Henan Province: based on Tapio method and STIRPAT model. *Environ Sci Pollut Res* 30:52679–52691. <https://doi.org/10.1007/s11356-023-26051-z>
- Yin FF, Bo Z, Yu L, Wang JZ (2023) Prediction of carbon dioxide emissions in China using a novel grey model with multi-parameter combination optimization. *J Clean Prod* 404:13. <https://doi.org/10.1016/j.jclepro.2023.136889>
- Yu X, Dong Z, Zheng D, Deng S (2023) Analysis of critical peak electricity price optimization model considering coal consumption rate of power generation side. *Environ Sci Pollut Res*. <https://doi.org/10.1007/s11356-023-29754-5>
- Zhang S, Huo Z, Zhai C (2022a) Building carbon emission scenario prediction using STIRPAT and GA-BP neural network model. *Sustainability* 14:9369. <https://doi.org/10.3390/su14159369>
- Zhang SS, Huo ZG, Zhai CC (2022b) Building carbon emission scenario prediction using STIRPAT and GA-BP neural network model. *Sustainability* 14:17. <https://doi.org/10.3390/su14159369>
- Zhang Z, Hong WC, Li J (2020) Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm. *IEEE Access* 8:14642–14658. <https://doi.org/10.1109/ACCESS.2020.2966712>
- Zhao HQ, Hu J, Hao F, Zhang HY (2022) Determinants of carbon dioxide emissions and their peaking prospect: evidence from China. *Front Environ Sci* 10:11. <https://doi.org/10.3389/fenvs.2022.913835>
- Zhao Y, Liu R, Liu Z, Liu L, Wang J, Liu W (2023) A review of macroscopic carbon emission prediction model based on machine learning. *Sustainability* 15:6876. <https://doi.org/10.3390/su15086876>
- Zhou QM, Chen DL, Hu ZY, Chen X (2021) Decompositions of Taylor diagram and DISO performance criteria. *Int J Climatol* 41:5726–5732. <https://doi.org/10.1002/joc.7149>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.