Modeling Carbon Dioxide Emissions with Green Finance—Using Beijing, Chongqing, and Shanghai as Cases

Junting Wang

Peddie School, Hightstown, NJ 08520, USA
Email: junting.w@outlook.com
Manuscript received June 11, 2024; revised August 2, 2024; accepted October 12, 2024; published November 17, 2024

***Abstract*—The global greenhouse effect has attracted significant attention in recent years. Most of the current models regarding Carbon Dioxide (CO_2) emission apply multiple linear regression analyses on population, Gross Domestic Product (GDP), and energy consumption, disregarding Green Finance Index (GFI) as a driving factor. However, the GFI measures how well a region's financial activities align with environmental goals, which should have an important impact on carbon emissions. We aim to conduct a comprehensive analysis of carbon emissions in China's big cities, highlighting the impact of GFI on carbon emissions and modeling the excessive emissions to be represented in monetary values. Specifically, we picked three cities to serve as case studies—Beijing, Chongqing, and Shanghai. This study establishes an Autoregressive Integrated Moving Average (ARIMA) model to predict future values of the four driving forces (resident population, GDP, energy consumption, GFI); a Back Propagation Neural Network (BPNN) model to predict carbon emissions; and a cost model to analyze the cost related to excessive carbon emissions based on a fictitious scenario inspired by the US's emission goals. The result shows that GFI significantly correlates with lower carbon emissions. Therefore, increasing the GFI is an effective measure to ensure the realization of peak carbon emissions before 2030, which lowers the cost caused by the risk of excessive carbon emission. This study proposes a carbon emission and cost model that can provide a reference for the control of carbon emissions in Beijing, Chongqing, and Shanghai.**

***Keywords*—Autoregressive Integrated Moving Average (ARIMA), Back Propagation Neural Network (BPNN), carbon peak, green finance index**

I. Introduction

Greenhouse Gas (GHG) emissions, particularly Carbon Dioxide (CO_2) has become a pivotal concern worldwide due to its role in global warming. CO_2 emissions, predominantly from industrial and transportation developments, have been persistently rising. Internationally, efforts like the Paris Agreement and various national commitments, such as the European Union and the US's targets for peak and net-zero carbon emissions [1, 2], reflect the urgency to address this issue.

China, as the largest CO_2 emitter globally, accounts for 30.9% of global CO_2 emissions in 2021 [3]. China has witnessed a 215% increase in CO_2 emissions from 2000 to 2021 [4]. To combat climate change, China has set ambitious goals, including a significant reduction in carbon emissions per unit of Gross Domestic Product (GDP) and a substantial increase in renewable energy capacity. Most notably, China is committed to reaching carbon peak by 2030 [5].

In this context, the role of the Green Finance Index (GFI), a measure of a region's financial activities' alignment with environmental objectives, becomes crucial. Current research, however, predominantly employ methods like linear regression and Machine Learning (ML) to predict carbon emissions, often overlooking the potential impact of financial instruments like the GFI [6–10]. The few research regarding GFI and carbon emissions either focused on developed countries or employed traditional rather than ML algorithms [11–13]. Other related research has focused on the costs associated with carbon emissions, employing methods like stochastic modeling, game theory, and neural networks [14–16].

Our study aims to combine these approaches to conduct a comprehensive analysis of carbon emissions in three major Chinese cities: Beijing, Chongqing, and Shanghai. We focus on the influence of four driving forces (GDP, population, energy consumption, and GFI) on carbon emissions. We utilized an Autoregressive Integrated Moving Average (ARIMA) model to predict the driving forces and Back Propagation Neural Networks (BPNN) to model carbon emissions. Additionally, we employed a cost model to estimate the financial implications of exceeding emission targets. The emission target is inspired by the US's emission goals [17] and is set to cut 25% of 2019's carbon emissions by 2030. The overall research process is shown in Fig. 1.

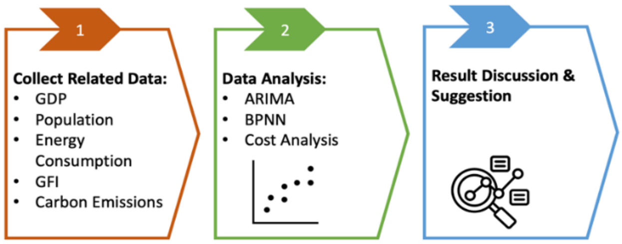


Fig. 1. The overall research process.

II. Methodology

This research first collects, cleans, and visualizes historical data. Then, ARIMA is applied to predict future values of the four driving forces of all three cities in 2020–2030. Thirdly, BPNN is applied to predict carbon emissions of all three cities in 2020–2030. Finally, the cost model is applied to approximate the cost related to controlling carbon emissions to the target value, which is set to 75% of 2019 levels. This modeling process is illustrated in Fig. 2.

A. Data Collection and Visualization

This research covers three case studies (Beijing, Chongqing, and Shanghai) and analyzes many different aspects related to carbon emissions for each one. Therefore, data include population (unit: 10 thousand), GDP (unit: 100 million Yuan), total energy consumption (unit: 10 thousand tons), carbon emissions (unit: 10 thousand tons), and green finance index (value between 0 and 1, higher means more

DOI: 10.7763/IJMO.2024.V14.862

developed green financing).

All data are collected from statistical yearbooks published by local and national governments [18–21]. The dataset, however, is missing certain types of data at certain years. To make sure that data for each city and each category match in all the years used for analysis, the years 2001–2019 are selected and considered as historical data in this research (see Fig. 3).

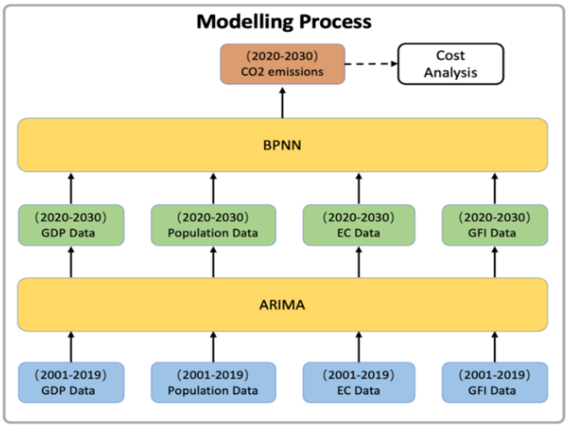
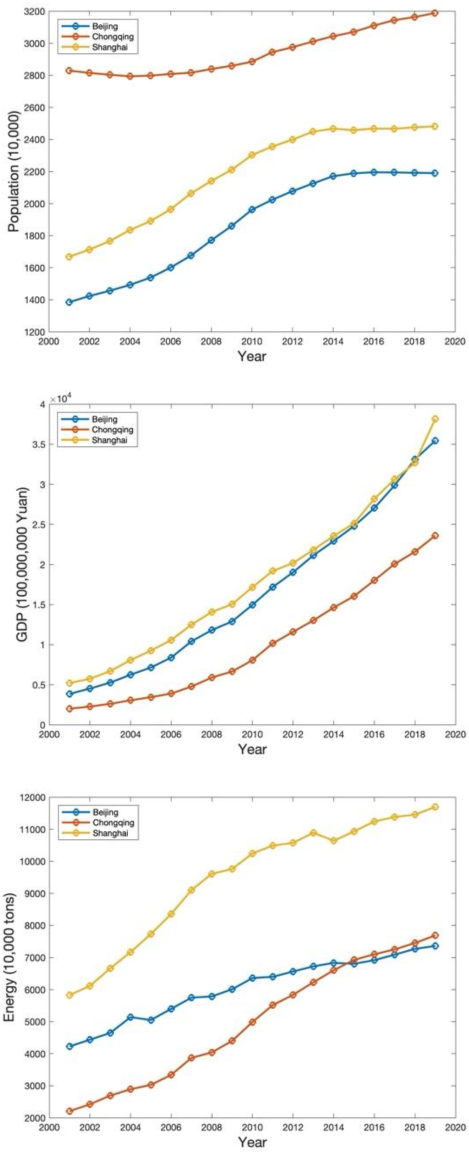


Fig. 2. The modeling process.



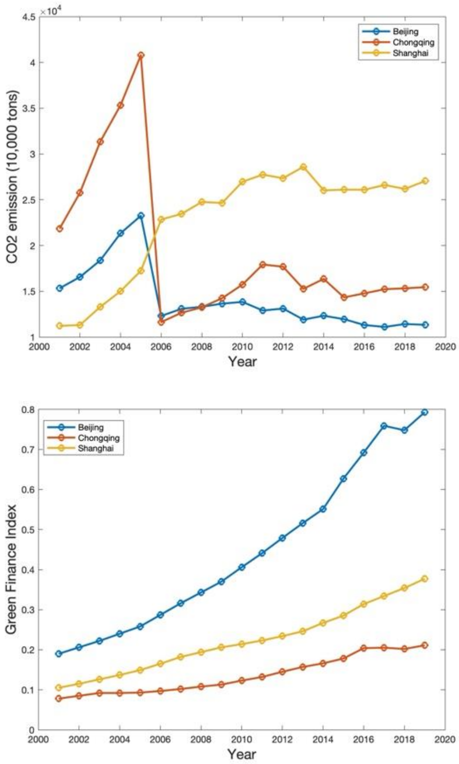


Fig. 3. Visualization of four driving forces and carbon emission.

B. *Predicting Driving Forces Using ARIMA*

ARIMA model is a widely used time series prediction model that consists of three parameters: AR term, I term, and MA term. In Matlab, these parameters correspond to p, d, and q values. The p value is known as the lag order and determines the number of Autoregressive (AR) terms that the model considers, which determines how reliant the model is on past data points. The d value represents the number of integrated terms. The d value should be equal to the number of times a difference calculation needs to be applied to the time series to result in a stationary series. Usually, for a time series with a linear trend, $d = 1$ should be used. The q value is known as the order of moving average and determines the number of Moving Average (MA) terms that the model considers, which determines how much the prediction is directly related to the error or noise calculation of past data points.

Given that X_t represents a time series created from the original dataset that is non-stationary, differencing X_t for d times can result in Y_t, a stationary series. According to the p, d, q values, Y_t can be represented by Eqs. (1) and (2).

$$Y_t = \nabla^{\mathrm{d}} X_t \tag{1}$$

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} \tag{2}$$

1) *Box Jenkins method*

This method is usually used to determine the p, d, q values for an ARIMA model and check if the model is a good fit. It employs a Dickey-Fuller test to identify if the time series is stationary. If not, it is necessary to difference the series d times to make it stationary, thus the d value is obtained. The p and q values are determined by the trailing tendency of the graphs of two functions: Autocorrelation (ACF) and Partial Autocorrelation (PACF). To check if the model is a good fit, researchers visualize the residuals and related information to ensure the residuals are normally distributed and uncorrelated.

Usually, the ACF and PACF graphs don't show a clear trend. Therefore, researchers need to subjectively determine the parameter values mostly from experience. This method also doesn't provide a solution if a mix of AR and MA terms are used.

2) Set up for this research

Given the issues of the Box Jenkins method, this research took a more objective approach. For each of the four driving forces for each city, the 19-year time series is split into train and test data in a 3.75:1 ratio. The training set has data from 2001 to 2015 while the testing set has data from 2016 to 2019. Each training set is treated with 128 different ARIMA models. None of the time series are stationary in this research, so only $d = 1$ and $d = 2$ are considered. The p and q values each range from 0 to 7. A Matlab script is used to automatically run the 128 models to predict the values in 2016–2019. Then, the Root Mean SQUARED error (RMSE) of the predicted values and the testing set is calculated and stored in the matrix. The script keeps track of the lowest RMSE value and its corresponding p, d, q values, these parameter values are recorded for each of the driving forces for each city and will be used to predict the values of corresponding driving forces from 2020 to 2030.

The specific parameter values obtained from the script is shown in Table 1. There are two sets of parameters marked by *, which represents the set doesn't result in the lowest RMSE. This could happen if the lowest achieved RMSE corresponds to an invalid parameter set. An invalid set could mean: 1) both $p = 0$ and $q = 0$; 2) the set can't be applied to the entire time series. In these cases, the valid parameter set that achieved the lowest RMSE is used.

Table 1. The ARIMA parameters used for each of the variables in different cities

City	Population	GDP	Energy Consumption	GFI
Beijing	(3,1,0)	(1,1,4)	(1,1,1)	(0,1,1)*
Chongqing	(5,1,0)	(5,1,0)	(3,1,2)	(6,1,2)
Shanghai	(2,1,0)	(0,1,7)	(3,1,6)*	(0,1,2)

* represents the set doesn't result in the lowest RMSE

To ensure that the ARIMA model with the lowest RMSE is viable in predicting a certain time series, each RMSE value is calculated as a percentage of the mean of the testing set of the corresponding driving force (see Table 2). The error percentages are all very low and thus prove that using an ARIMA model to predict the values of driving forces is viable.

Table 2. RMSE as a percentage of the mean of testing set

City	Population	GDP	Energy Consumption	GFI
Beijing	0.17%	0.90%	0.83%	4.10%
Chongqing	0.24%	0.34%	0.42%	4.71%
Shanghai	0.27%	6.62%	1.19%	1.00%

C. Predicting Carbon Emissions Using BPNN

BPNN is a multilayer feedforward neural network known for its backpropagation process, which means that the model adjusts the weights of a neural network based on information, specifically the error rate, obtained in the previous training epoch. The inputs are collected from the preconnected path and then modeled by the hidden layers using different weights that are usually randomly selected. The output layer then calculates the output from hidden layers' output multiplied by the weight. The gradient of the loss function is calculated by taking the derivative of the loss function by weight and the value of the previous layers are calculated through backpropagation to reduce loss. The process is repeated until the previous sum is updated. In essence, the backpropagation process increases the number of correct output nodes and loss is reduced.

In our study, specifically, the network contains three layers: the input layer includes four nodes, which represent GDP, population, energy consumption, and GFI; the hidden layer has 15 nodes; and the output layer has one node which stands for carbon emissions volume (see Fig. 4).

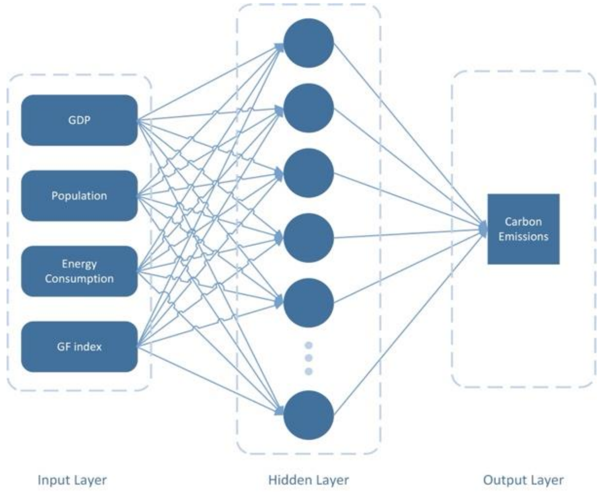


Fig. 4. The BPNN network utilized in this research.

D. BPNN Model Training

Our historical dataset only contains 19 years of carbon emission and driving forces data. Therefore, we applied linear interpolation to expand the 19 data points (yearly) to 216 data points (monthly). The interpolated time series is then split into train and test data in a 5:1 ratio. The training set has data from 2001 to 2016 (months 1–180) while the testing set has data from 2016 to 2019 (months 181–216). The training parameters are the same for each city and are outlined in Table 3.

Table 3. BPNN parameters

Size of hidden layers	Epochs	Target MSE	Learning Rate
15	10000	1e–8	0.001

The validity and generalization performance of the trained model is tested with the relative error of the predictions in the three years of test data (see Table 4 and Fig. 5). The trend of the predicted values for all three cities is consistent with the actual values and the relative errors are all very low. This proves that population, GDP, energy consumption, and GFI are driving forces of carbon emissions and that BPNN has excellent generalization performance.

Table 4. Summary of relative error of BPNN model

City	Beijing	Chongqing	Shanghai
Maximum Relative Error	0.0554	0.0676	0.0437
Average Relative Error	0.0247	0.0364	0.0230

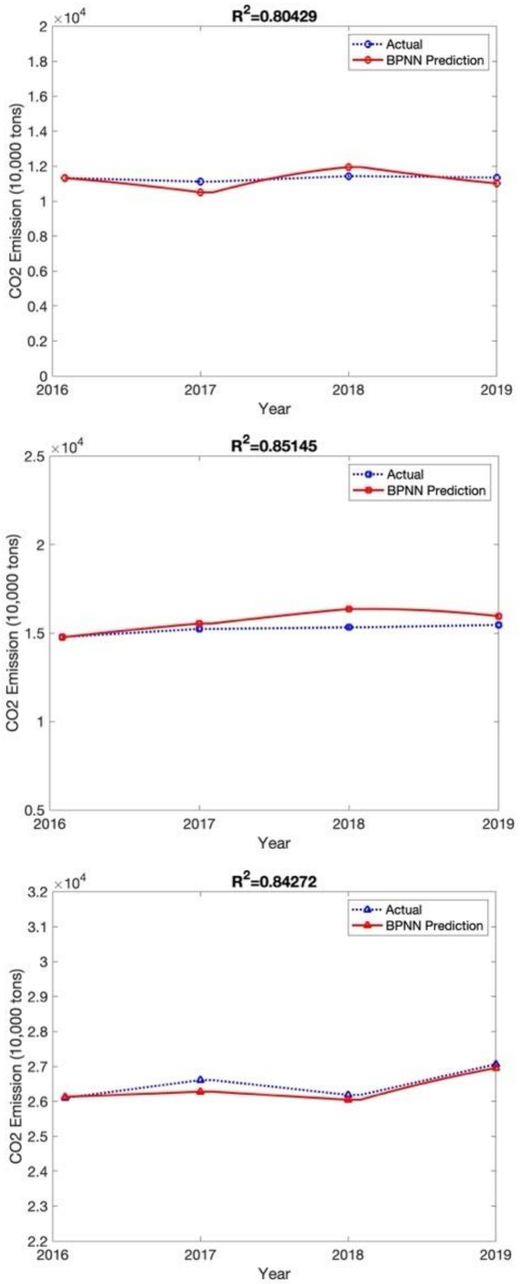


Fig. 5. Actual and predicted CO_2 values for Beijing, Chongqing, and Shanghai, respectively.

The predicted values of the four driving forces in 2020–2030 by the ARIMA model is also interpolated into monthly data points before being inputted into the trained BPNN models to predict the carbon emissions in 2020–2030. To determine the effect of the GFI on carbon emissions, the prediction is conducted under 5 different conditions: no change, GFI × 1.05, GFI × 1.1, GFI × 1.15, and GFI × 1.2.

E. Cost Analysis

This research uses the cost model proposed by Zagheni and Billari [16] to analyze the effect of initial carbon emission on the cost of excessive carbon emissions in Beijing, Chongqing, and Shanghai.

In 1994, Dietz and Rosa [22] proposed a stochastic counterpart of the IPAT and ImPACT identities that model the relation between carbon emissions and population, GDP, and energy consumption (see Eq. (3)). In 2007, Zegheni and Billari [16] further refined the model (see Eq. (4)). The constants in these equations refer to information such as population and could be calculated with constrained linear least square problems.

$$I = aP^bA^cT^de \tag{3}$$

$$\ln(I) = \ln(a) + (b-c)\ln(P) + c\ln(PA) + d\ln(T) + \ln(e) \tag{4}$$

This study analyzes the fictitious scenario where each city is required to import emission credits based on market price if they fail to lower carbon emissions by 25% of 2019 levels by 2030. Given this setup, the potential cost related to carbon emissions for any country can be set to $C(I,t)$, with I being the amount of carbon emissions as the underlying asset and t being the year. This function can be seen as a European option with carbon emissions as an asset. Further assuming that the risk-free rate is constant r, the cost function can be expressed as Eq. (5), where $\bar{I}$ represents the carbon emission threshold (75% of 2019's carbon emissions value).

$$C(I,t) = E\big(\alpha(I_T - \bar{I})e^{-r(T-t)} \mid I_{t0} = I_0 \tag{5}$$

Eq. (6) combines Eq. (5) with the Black-Scholes equation. $N(x)$ is a standard normal distribution accumulation function expressed in Eq. (7) with $d_1(t)$ and $d_2(t)$ expressed in Eqs. (8) and 9, respectively. For variable explanations, see Table 5.

$$C(I,t) = \alpha I\left[\frac{\rho_0 P_i e^{-\rho_0 t}+1}{\left(\frac{P}{P_m}-1\right)e^{-\rho_0 T}+1}\right]^b e^{(c\mu-r)\Delta t}N\big(d_1(t)\big) - \alpha\bar{I}e^{-r\Delta t}N\big(d_2(t)\big) \tag{6}$$

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{w^2}{2}}dw \tag{7}$$

$$d_1(t) = \frac{\ln(I)-\ln(\bar{I})+b\left[\ln\left(P_i e^{-\rho_0 t}+1\right)-\ln\left(P_i e^{-\rho_0 T}+1\right)\right]+c\mu(\Delta t)+\frac{1}{2}c^2\mu^2(\Delta t)}{c\sigma\sqrt{T-t}} \tag{8}$$

$$d_2(t) = d_1(t) - c\sigma\sqrt{T-t} \tag{9}$$

Table 5. Key variable definition

Variable	Definition
Pi	$\frac{P_m-P_0}{P_0}$, Pm = carrying capacity, P0 = initial population
Δt	$T-t$
ρ_0	Population increase rate
b & c	Regression coefficient
μ	Average of random variable

III. Result

A. Driving Forces Prediction

The values that the ARIMA model predicted for all four driving forces from Beijing, Chongqing, and Shanghai in 2020–2030 are visualized alongside historical values in Fig. 6. The GFI prediction for Beijing exceeds 1 for the years 2023 and later, which is impossible in real life since GFI is a value between 0 and 1. Therefore, the values for 2023 and later are adjusted to 1.00 when used with BPNN to predict carbon emissions.

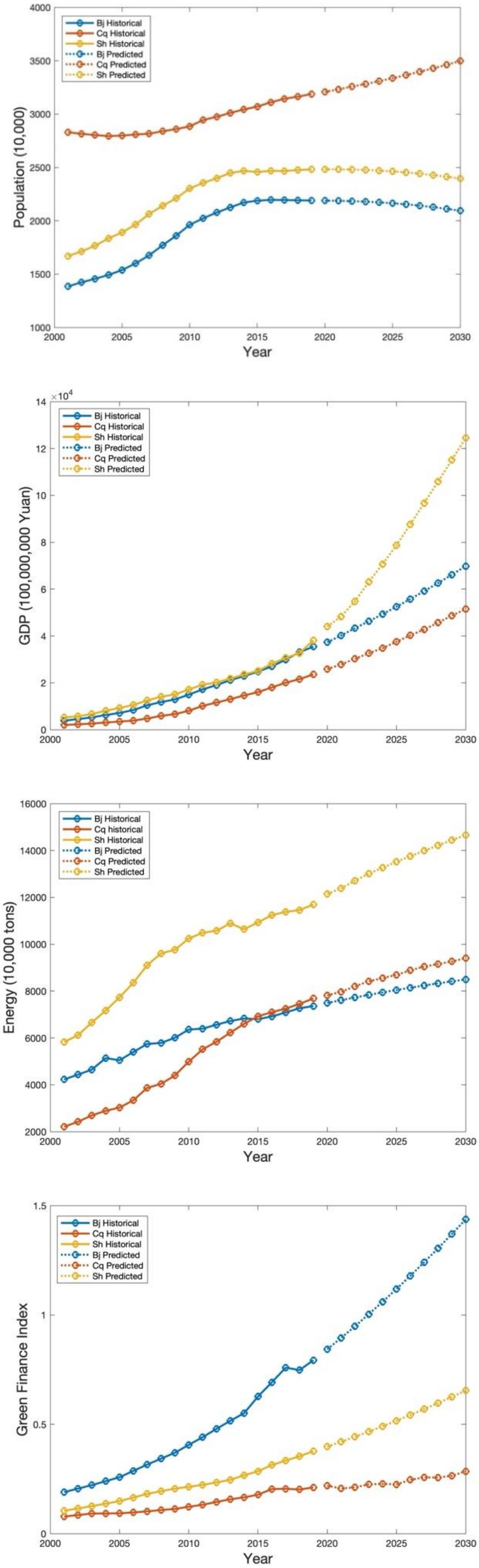


Fig. 6. Visualization of ARIMA prediction values.

B. Carbon Emissions Prediction

The predicted carbon emission values of Beijing is displayed in Fig. 7. GFI modification is not applied to Beijing because many of the predicted GFI values of Beijing already exceed 1, which eliminates room for increasing the GFI values. The predicted carbon emission values of Chongqing and Shanghai with four different GFI modifications are displayed in Figs. 8 and 9, respectively.

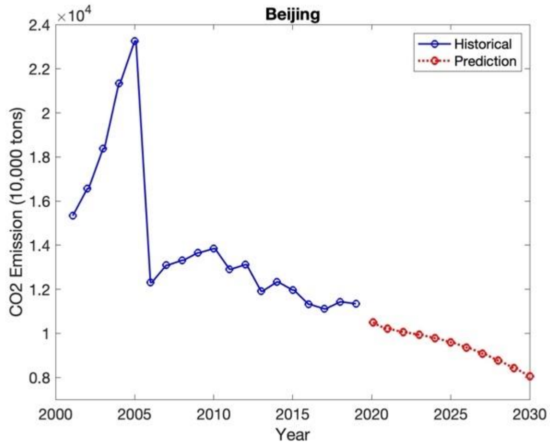


Fig. 7. Beijing's carbon emissions from 2001 to 2030 (past and predicted values).

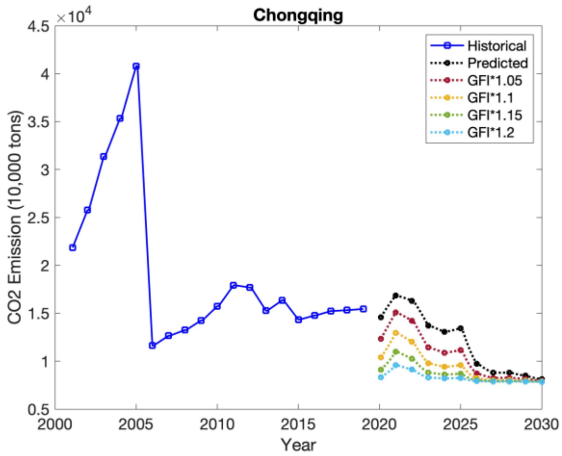


Fig. 8. Chongqing's carbon emissions from 2001 to 2030 (past and 5 predictions).

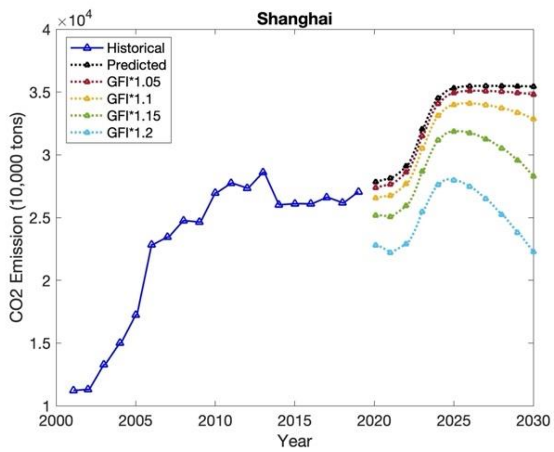


Fig. 9. Shanghai's carbon emissions from 2001 to 2030 (past and 5 predictions).

The non-modified predicted carbon emission values of Beijing and Chongqing show a clear decreasing trend, consistent with reaching carbon peak in 2030. Shanghai's predicted carbon emissions, however, generally shows an increasing trend.

In both Chongqing and Shanghai, higher GFI correlates with lower predicted carbon emissions. Shanghai's carbon emissions trend even changes from increasing to decreasing when GFI values are multiplied by greater than or equal to 1.15. This proves that the GFI highly correlates with low carbon emissions.

C. Cost Analysis

The cost is graphed based on T and I, according to $C(I,t)$. For all three cities, initial carbon emissions I positively correlates with the cost; Beijing's predicted cost negatively correlates with time T while Chongqing and Shanghai's predicted cost positively correlates with time T (see Figs. 10–12).

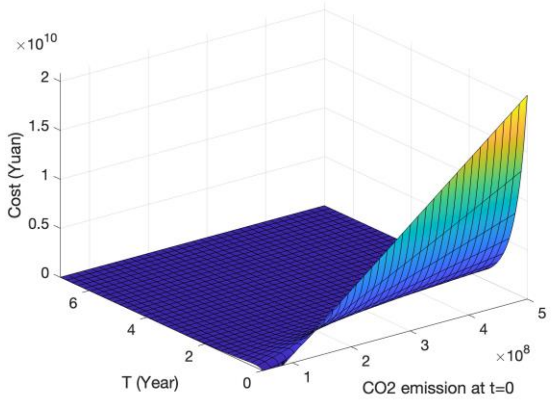


Fig. 10. Cost at time *T* with initial carbon emissions *I* in Beijing.

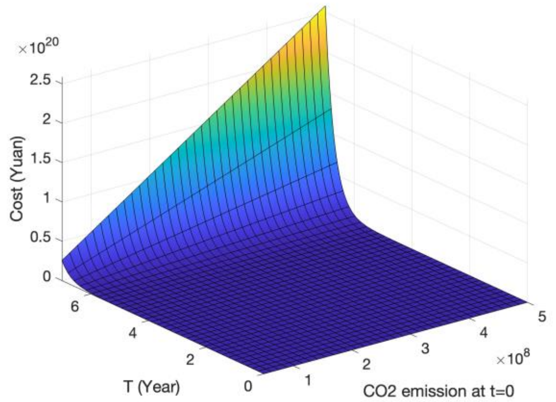


Fig. 11. Cost at time *T* with initial carbon emissions *I* in Chongqing.

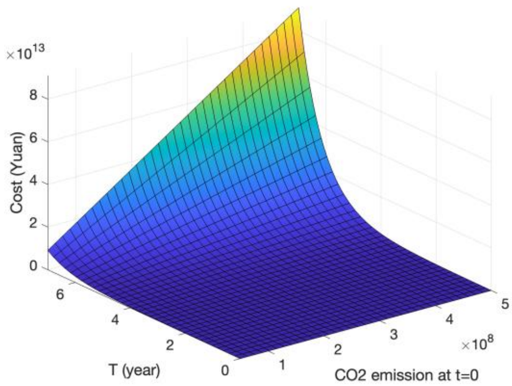


Fig. 12. Cost at time *T* with initial carbon emissions *I* in Shanghai.

IV. Conclusion

This paper explored the driving forces of carbon emissions and the prediction of future carbon emissions from the perspective of governments, specifically focusing on Beijing, Chongqing, and Shanghai as three case studies. It also identifies the cost related to excessive carbon emissions.

This study performed time series prediction using ARIMA; driving force analysis and carbon emissions prediction using BPNN; and cost analysis using equations derived by Zagheni and Billari. All calculations and models are performed in Matlab.

The main findings of this research can be divided into three parts.

First, our training of BPNN shows that using the four driving forces (Population, GDP, Energy Consumption, and GFI) to predict carbon emissions is viable and accurate. Our analysis further shows that GFI has a great and negative impact on carbon emissions, meaning that a higher GFI correlates to low carbon emissions.

Second, the carbon emissions data from 2020–2030 as predicted by our trained BPNN models provide insight into the probability of China reaching carbon peak in 2030. The prediction results from both Beijing and Chongqing show a decreasing trend while the result from Shanghai shows a generally increasing trend.

Third, our cost model results can act as a useful reference for the cost of excessive carbon emissions in the fictitious scenario where the target is set to limit carbon emissions to 75% of 2019 levels by 2030.

This paper provides a comprehensive analysis and prediction of the driving forces and costs related to carbon emissions but has limitations in several areas. Firstly, due to the quantities of publicly available data, this study only took into account the years 2001–2019 as historical data, and only three cities were considered. Future research should expand the quantity of historical data by obtaining data from the 20th century and recent years, as well as expand the number of locations studied to many more cities or even provinces. Secondly, this research assumes the hypothetical condition where each city in China must pay the excessive emissions at a market price if it doesn't lower its carbon emissions to 75% of 2019 levels by 2030. Although this is a practical case, future research could investigate trade agreements or limitations between China and other countries that have certain carbon emission requirements and calculate the cost of excessive carbon emissions based on these real-world trade agreements.

Conflict of Interest

The author declares no conflict of interest.

References

[1] CSO. Net-Zero Emissions Operations by 2050, including a 65% reduction by 2030. Federal Sustainability Plan. [Online]. Available: https://www.sustainability.gov/federalsustainabilityplan/emissions

[2] EU. 2050 long-term strategy. [Online]. Available: https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en

[3] H. Ritchie. (2019). Who emits the most CO_2 today? Our World in Data. [Online]. Available: https://ourworldindata.org/annual-co2-emissions

[4] H. Ritchie and M. Roser. (2020). China: CO_2 Country Profile. Our World in Data'. [Online]. Available: https://ourworldindata.org/co2/country/china

[5] A. Tay, "By the numbers: China's net-zero ambitions," *Population and Environment*, vol. 29, pp. 68–82, 2022.
[6] G. Serafeim and G. V. Caicedo, "Machine learning models for prediction of scope 3 carbon emissions," Harvard Business School Accounting & Management Unit Working Paper 22-080, 2022.
[7] Y. Zhou, J. Zhang, and S. Hu, "Regression analysis and driving force model building of CO_2 emissions in China," *Scientific Reports*, vol. 11, 2021.
[8] L. Amarpuri, N. Yadav, G. Kumar, and S. Agrawal, "Prediction of CO_2 emissions using deep learning hybrid approach: A case study in Indian context," in *Proc. 2019 Twelfth International Conference on Contemporary Computing (IC3)*, 2019.
[9] Q. Nguyen, I. Diaz-Rainey, and D. Kuruppuarachchi, "Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach," *Energy Economics*, vol. 95, 2021.
[10] D. Hou, H. Zhang, L. Li, X. Wang, Y. Lin, and H. Du, "Deep learning-based comparative modeling of carbon emissions projections," in *Proc. 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*, 2021, pp. 241–251.
[11] C. Q. Guo, X. Wang, D. D. Cao, and Y. G. Hou, "The impact of green finance on carbon emission—Analysis based on mediation effect and spatial effect," *Frontiers in Environmental Science*, vol. 10, 2022.
[12] W. Zhang, Z. Zhu, X. Liu, and J. Cheng, "Can green finance improve carbon emission efficiency?" *Environmental Science and Pollution Research*, vol. 29, pp. 68976–68989, 2022.
[13] Z. Fang, C. Yang, and X. Song, "How do green finance and energy efficiency mitigate carbon emissions without reducing economic growth in G7 countries?" *Frontiers in Psychology*, vol. 13, 2022.
[14] J. Qin, Y. Zhao, and L. Xia, "Carbon emission reduction with capital constraint under greening financing and cost sharing contract," *International Journal of Environmental Research and Public Health*, vol. 15, 2018.
[15] F. Zhang and N. Wen, "Carbon price forecasting: A novel deep learning approach," *Environmental Science and Pollution Research*, vol. 29, pp. 54782–54795, 2022.
[16] E. Zagheni and F. C. Billari, "A cost valuation model based on a stochastic representation of the IPAT equation," *Population and Environment*, vol. 29, pp. 68–82, 2007.
[17] FACT SHEET: U.S. Reports its 2025 Emissions Target to the UNFCCC. (2015). [Online]. Available: https://obamawhitehouse.archives.gov/the-press-office/2015/03/31/fact-sheet-us-reports-its-2025-emissions-target-unfccc
[18] National Bureau of Statistics. China Statistical Yearbook. [Online]. Available: [Online]. Available: http://www.stats.gov.cn/sj/ndsj/
[19] Beijing Municipal Bureau of Statistics. Beijing Statistical Yearbook. Available: http://tjj.beijing.gov.cn/
[20] Shanghai Municipal Bureau of Statistics. Shanghai Statistical Yearbook. [Online]. Available: https://tjj.sh.gov.cn/tjnj/index.html
[21] Chongqing Municipal Bureau of Statistics. Chongqing Statistical Yearbook. [Online]. Available: http://tjj.cq.gov.cn/zwgk_233/tjnj/wap.html
[22] T. Dietz and E. A. Rosa, "Rethinking the environmental impacts of population, affluence and technology," *Human Ecology Review*, vol. 1, no. 2, pp. 277–300, 1994.