Prediction and Evaluation Model Based on Carbon Emission with RF: A Case Study of Six Coastal City Clusters in China

Kaiyuan Zheng (✉ kzyzheng196@163.com)

Tsinghua University

Ying Zhang

Tsinghua University

Yuntao Guan

Tsinghua University

Research Article

Keywords: Carbon emission, Random Forest, Prediction, Evaluation, Urban adaptivity, Coastal city clusters

Posted Date: February 8th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1268862/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Prediction and evaluation model based on carbon emission with RF: A case study of six coastal city clusters in China

Kaiyuan Zheng¹, Ying Zhang¹, Yuntao Guan^{1*}

(¹ Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China; kyszeng196@163.com (K.Z.); 459657865@qq.com (Y.Z.); guanyt@tsinghua.edu.cn (Y.G.)

*Correspondence: guanyt@tsinghua.edu.cn

Abstract: Aiming at predicting the problems of social economics, environmental pollution, climate change and marine disasters caused by carbon emission, a predicting model based on carbon emission with the Random Forest (RF) model was constructed. Meanwhile, a novel urban adaptivity evaluation system is put forward considering the above four domains of indicators, hence the two models are integrated. Six coastal city clusters of China are selected as study area and the result of RF model shows that northern city clusters suffer more pollutant loads due to their heavy industry layout, southern cities generally have higher GDP while they're more vulnerable toward extreme weather and marine disasters. The result of evaluation system indicates that northern city clusters have higher urban adaptivity due to their balance between economics and pollution as well as less vulnerability on climate change because of relatively high latitude. On the contrary, southern cities are supposed to focus on environmental pollution and tropical storms to pursue superior compatibility.

Keywords: Carbon emission; Random Forest; Prediction; Evaluation; Urban adaptivity; Coastal city clusters

1. Introduction

Global warming has been a globally prominent issue due to the massive emission of carbon dioxide (Mardani et al. 2019) and other Greenhouse Gas (GHG), which attracts substantial awareness with its effect in recent years. The sharply increasing carbon emission in the past decades has become the spotlight of all countries around the world (Wang et al. 2019; Zhang et al. 2018). Therefore, the boosting attention on global warming due to the extensive use of fossil energy sources have led several countries to search for the correlation between carbon emission and other key issues (Mardani et al. 2020; Fan et al. 2021; Mardani et al. 2020; Gu et al. 2021; Rahman et al. 2020), such as social economics, pollution control, climate change and marine disasters, etc. Thus numerous studies carried out these years focused on the relationship between the aforementioned 4 domains and carbon emission (Caglar 2020; Xie et al. 2020). Of which carbon emission has always been the output result of other driving factors, while it can also be treated as an input variable for predicting other indicators by contrary, hence we carried on this research.

As for social economics, China has imposed strict policies in dealing with the complex problem between carbon emission and social progress, working on boosting its economic prosperity as well as reducing carbon emission and environmental pollution by improving energy utilization efficiency. These works mainly concentrate on the relationship between social economics, environmental pollution and energy consumption, while neglecting the influence of carbon emission (Wei et al. 2019; Zhao et al. 2019). Besides, the situation of aquatic pollution and air pollution is severe in many cities of China, and the amount of urban environmental pollutants and carbon emission varies in a similar tendency. Hence studies are also supposed to be carried on to quantify the relationship between them (Guan et al. 2022).

What's more, climate change is another concern closely related to carbon emission, even can be regarded as the most predominant issue that influences people's daily life directly. This directly leads to the increasing frequency of extreme temperature and precipitation due to massive carbon emission, so the relationship among them is also to be explored (Lu et al. 2008). Furthermore, coastal cities are facing unprecedented risks of marine disasters due to carbon emissions, like the increasing possibility of cyclones or storms, and steady rising of sea level, while only a few studies focused on carbon flows that integrated sea and land through natural and anthropogenic processes (Stom et al. 2005; Fan et al. 2020). Therefore, for quantifying the relationship between carbon emission and marine disasters, it's essential to incorporate this factor into the comprehensive evaluation system.

With the above 4 domains related to carbon emission, it's necessary to adopt appropriate methods to comprehensively predict all the above results related to carbon emission. Numerous models and algorithms have been utilized in this area, like gray prediction theory (Zhao et al. 2018; Yang et al. 2016), time series prediction methods (Gerami et al. 2018; Chang et al. 2018), linear regression prediction (Qi et al. 2017; Afshari et al. 2017) and non-linear regression prediction (Wang et al. 2016; Takahashi et al. 2016). These models mainly use historical statistical data as the basis for prediction, thus have high requirements for the quality of information and materials (Li et al. 2021). Compared with these conventional methods, Machine Learning (ML) algorithms are supposed to solve statistical prediction problems more efficiently and perform better while dealing with complex and tedious data. Random Forest (RF) has gained remarkable popularity for its ability to handle multi-dimensional classification and regression problems with excellent accuracy and limited possibility of overfitting (Wen et al. 2022). Many studies have proved that RF outperforms other popular ML algorithms in various application scenarios (Uddin et al. 2019) and is supposed to be superior in model interpretation and faster to train, thus it's chosen to be applied in this research.

The existed researches mainly focus on predicting a single factor related to carbon emission, while lacking a comprehensive system that incorporates all 4 domains. Therefore, we construct an urban adaptivity evaluation system involved with the area of social economics, environmental pollution, climate change and marine disasters, which consists of numerous indicators. By training the input data and exporting the output results with RF, we assess the prediction results with such an evaluation system. Six city clusters around coastal areas of the Chinese mainland are selected and analyzed due to their prosperous economy, high population density, and developed industry, thus we furtherly explore the relationship between carbon emission and other factors therein

2. Materials and Methods

2.1 Study area

Coastal provinces of east China are regarded as the most prosperous areas within this country, which mainly ranges from 20°N to 40°N in latitude and from 110°E to 130°E in longitude. With mild temperature and substantial precipitation, these provinces host nearly 44.52% population and 42.53% GDP of this country. Developed industry boosts the progress of local urbanization and modernization while also producing large quantities of GHG emissions as well as environmental pollutants. The problems related to carbon emission are involved in the area of environmental protection, climate change and marine disasters, etc., which causes great pressure upon the sustainable development of coastal cities. Therefore, the Chinese government is working on improving energy utilization efficiency and cutting off carbon emissions while boosting the prosperity of the economy, which promotes the necessity of predicting futural situations and providing planning assistance.

The aforementioned six city clusters are distributed in the prosperous areas of China (as shown in Fig. 1 and Table 1): Mid-southern Liaoning City Cluster (MLCC), Beijing-Tianjin-Hebei City Cluster (BTHCC), Shandong Peninsula City Cluster (SPCC), Yangtze River Delta's City Cluster (YRDCC), Southeastern Fujian City Cluster (SFCC) and Pearl River Delta's City Cluster (PRDCC). Required data is obtained from Urban Economy Development & Environmental Pollution Panel Data of 295 cities from 1996 to 2017. With the prediction and analysis among various aspects of these city clusters, we may conclude certain suggestions for futural development.

Fig. 1 Study area of six coastal city clusters

Table 1 Composition of six coastal city clusters

Name of city clusters	Cities contained
Mid-southern Liaoning City Cluster (MLCC)	Shenyang, Dalian, Anshan
Beijing-Tianjin-Hebei City Cluster (BTHCC)	Beijing, Tianjin, Shijiazhuang, Tangshan
Shandong Peninsula City Cluster (SPCC)	Jinan, Qingdao, Yantai
Yangtze River Delta's City Cluster (YRDCC)	Shanghai, Suzhou, Hangzhou, Nanjing, Ningbo, Wuxi
Southeastern Fujian City Cluster (SFCC)	Quanzhou, Fuzhou, Xiamen
Pearl River Delta's City Cluster (PRDCC)	Guangzhou, Shenzhen, Foshan, Dongguan, Zhuhai

2.2 Random Forest

Machine learning has been an ideal tool in data analysis, especially for uncovering the relationship among large datasets. Due to the inherent characteristics and superior classification performance, RF is regarded as the best ML algorithm in dealing with the inner relationship among data. This model was developed by Breiman as the improvement of the classification and regression tree (CART) to improve its predicting accuracy. This method realizes better predictive performance without obviously extending computational expense. It has a random selection process consisting of inputting variables and data,

generating certain classification and regression trees (Fang et al. 2021). With the above mechanism, the RF algorithm integrates all the decision trees in which each tree depends on a random vector that is independent and identically distributed (as shown in Fig. 2).

Fig. 2 Diagram of Random Forest algorithm

According to the detailed diagram of RF algorithm and the regression classification principle of CART, we defined several following probabilities based on the decision tree principles. Of which $N(t)$ is regarded as the sample number of the decision tree training sample data set L at node t , $N_j(t)$ represents the number of sample data at node t that belong to the j th category w_j , and n represents the total number of training sample data.

1) The sample probability p at node t based on L is estimated as

$$p(t) = \frac{N(t)}{n} \quad (1)$$

2) With the dataset L , the sample probability p at the corresponding node t contained in the category w_j can be expressed as

$$p(w_j|t) = \frac{N_j(t)}{N(t)} \quad (2)$$

3) We assume that t_L and t_R represent the left and right sub-nodes of t , respectively, thus the corresponding sample probabilities based on data set L (p_L and p_R) can be demonstrated as

$$p_L = \frac{p(t_L)}{p(t)} \quad (3)$$

$$p_R = \frac{p(t_R)}{p(t)} \quad (4)$$

Based on the proportion of the sample data belonging to each category at node t , the corresponding category at t can be finally determined. Considering the relationship between input variables and output variables, we set carbon emission as the input data and treat data involved with social economy, environmental pollution, climate change and marine disasters as output data. Then to train the RF model with the above variables and predict the corresponding results to construct an evaluation system.

2.3 Urban adaptivity evaluation system

As discussed above, we classify the input data into four categories, which involve the area of social economics, environmental pollution, climate change and marine disasters. With each category, we set some indicators and assign corresponding weights attached to them, thus building an urban adaptivity evaluation system (as shown in Fig. 3). The urban adaptivity evaluation system comprehensively considers the influence of all aforementioned aspects, represents the harmony and balance among social development, environmental protection and other areas. Therefore, it's supposed to be the objective function that reflects the overall performance of city clusters under a certain time scale.

For the area of social economics, GDP of each city is selected as the single representative indicator; the indicator of environmental pollution contains the emission of sewage, sulfide dioxide (SO₂) and smoke dust; climate change covers six subareas: annual mean temperature, annual highest temperature, annual lowest temperature, annual relative humidity, annual total sunshine duration and annual total precipitation; as for marine disasters, it's mainly concerned with the increment of sea level and the influence of tropical storm.

In view of the variety of city numbers within each city cluster, we utilize the average value of each indicator as the input data, to offset the influence of it. The relative weight of each indicator within 4 domains is obtained based on the consideration of their relative importance compared with other indicators (Kaushik et al., 2021; Mou et al., 2021). All the data will be normalized with Eq. (5) to be analyzed in the same order of magnitude, then imported into the risk evaluation system to calculate eventual urban adaptivity value.

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

where x_i is the i th value of dataset, $x_{\min}$ is the minimum value of dataset, $x_{\max}$ is the maximum value of dataset and y_i is the normalized value of x_i .

Fig. 3 Scheme of urban adaptivity evaluation system

$$A_i = \sum_{j=1}^n w_{1j} \left(\sum_{k=1}^m w_{2k} y_k \right) \quad (6)$$

where i is the index of city cluster, j is the domain index of evaluation system, n is the domain number of evaluation system, k is the indicator index of domain j , m is the indicator number of domain j , A_i denotes the urban adaptivity of city cluster i , w_{1j} denotes the relative weight of domain j , w_{2k} denotes the relative weight of indicator k , y_k denotes the normalized value of indicator k .

3. Results and discussion

3.1 Analysis of prediction results

In terms of the training sets and prediction sets of social economics, we classify them into two categories in the following charts: the training sets are contained in the light green region and the prediction sets are shown in the darker green region. By collecting the carbon emission data from 1997 to 2018 and predicting the following five years (as shown in Fig. 4), we regard them as the input data of RF algorithm. According to various input datasets, we set corresponding training years and following five years as predicting years so as to work out the variation tendency and mechanism of them.

Fig. 4. Carbon emission from 1997 to 2022 of six city clusters

3.1.1 Social economics

The GDP data acquired from 2010 to 2017 is supposed as training set and that of the following five years is to be predicted (as shown in Fig. 5). The overall developing trend of six city clusters is keeping increasing, while certain city clusters show different performance. Among which, GDP of the MLCC and SFCC always remains at a relatively low level with certain variation, while an uncommon point lies in the boosting trend of SFCC that appears to drop a lot in the prediction sets. Meanwhile, although BTHCC and YRDCC lead the increasing trend and show a solid economic foundation, they both lower down developing pace after 2017, which is quite different from that of the actual data, as well. Considering the lack of input data during the process of training and the decrease of carbon emission of prediction datasets, RF can solely output a relatively low level of GDP. It's originated from the shortcoming of input data, but can also be understood.

Fig. 5 GDP variation from 2010 to 2022 of six city clusters

3.1.2 Environmental pollution

The area of environmental protection incorporated three indicators: annual wastewater discharge, annual sulfur dioxide discharge and annual smoke dust discharge (as shown in Fig. 6). Overall, the discharge amount of pollutants tent to be more stable during the prediction period, with different city clusters performing variously towards each indicator.

As for wastewater discharge (as shown in Fig. 6(a)), YRDCC always ranks first compared with its counterpart due to the developed industry in Yangtze River Delta that produces large quantities of sewage;

on the contrary, wastewater discharge in the Shandong Peninsula keeps lowest all the time. Besides, the discharge amount in other city clusters has a similar variation trend and prediction values.

All city clusters behave more complex under the condition of sulfur dioxide discharge, which is shown in Fig. 6(b). The discharge amount of BTHCC top all city clusters because of its substantial heavy industry plants, with YRDCC and MLCC following behind. What's more, the developing trend of them all get more stable in the prediction datasets, as well.

BTHCC and MLCC have the most discharge amount of smoke and dust rather than other city clusters (as shown in Fig. 6(c)), which may be relative to the industrial structure of north China. Oppositely, the emission amount in south China is nearly half of that in BTHCC and MLCC, which is totally different from the analysis result of wastewater discharge.

Fig. 6 Environmental pollutants discharge from 2010 to 2022 of six city clusters, annual wastewater discharge in (a), annual sulfur dioxide discharge in (b) and annual smoke dust discharge in (c).

3.1.3 Climate change

Indicators in climate change are concerned with temperature, precipitation, sunshine duration and humidity. All the factors are selected to comprehensively represent a variation of the six city clusters in climate change and integrated to assess futural performance therein.

The annual mean temperature varies much among city clusters, showing the geographical ingenuity from north to south across China clearly (as shown in Fig. 7(a)). Nevertheless, the variation of the highest and lowest temperature of six clusters can be more obvious, especially for the aspect of the highest temperature. During the winter period, the lowest temperature (as shown in Fig. 7(d)) keeps decreasing with the increase of latitude both in the training set and predicting set, and this indicator is always below zero in MLCC, BTHCC, SPCC and YRDCC, with that of SFCC and PRDCC remains

higher than zero. While the highest temperature (as shown in Fig. 7(c)) can be more similar among city clusters due to the overall high temperature across China when it comes to summer: nearly all city clusters have ever reached 40°C in recent 10 years except MLCC, and the highest value appears in various cities during the time series; also the developing trend hasn't shown apparently statistically logic, which is to be discussed.

Annual total precipitation (as shown in Fig. 7(b)) has a similar developing trend with annual highest temperature, changes randomly both in the training set and predicting set, making it hard to predict accurately. However, this indicator has a distinct characteristic in view of the value itself, increasing with the latitude descends, and south China (mainly over 1500mm) owns much more substantial rainfall than that of north China (mainly below 900mm).

Sunshine duration is another indicator that ranges spatially because of the geographical variety among six city clusters (as shown in Fig. 7(e)). Generally speaking, north cities have longer sunshine duration hours that may last for 2400~2600 hours yearly, while south cities mainly remain no more than 1600 hours each year. The data in the prediction period fits with this trend well, thus it's considered to be an acceptable result.

Correspondingly, relative humidity also relates to geographical characteristics (as shown in Fig. 7(f)). This indicator decreases from south (mainly over 75%) to north (mainly below 55%), which matches with the distribution of total precipitation well because of the inherent connection between the two variables.

Fig. 7 Climate indicators from 2004 to 2021 of six city clusters, annual mean temperature in (a), annual total precipitation in (b), annual highest temperature in (c), annual lowest temperature in (d), annual total sunshine duration in (e) and annual relative humidity in (f).

3.1.4 Marine disasters

Tropical storms and sea level increase are incorporated in the area of marine disasters. The frequency and intensity of tropical storms are shown in Fig. 8(a) and Fig. 8(b), which are mainly concerned with cities of south China, especially for city clusters in Fujian Province (SFCC) and Guangdong Province (PRDCC). These city clusters nearly have 3 times of tropical storms per year with the average annual intensity between 4 and 5, while northern cities seldom experience tropical storms and the intensity is always limited within 1.

In terms of annual variation of sea level, random distribution of variation also appears both in the training set and predicting set, each city cluster has its positive and negative value under this condition. With the progress of time series, all city clusters tend to face higher sea levels, which imposes certain pressure upon city development.

Fig. 8 Marine disaster indicators from 2000 to 2022 of six city clusters, frequency of tropical storms in (a), intensity of tropical storms in (b) and annual variation of sea level in (c).

3.1.5 Geographical visualization of prediction results

Based on our statistical prediction results and analysis above, the aforementioned 4 evaluation domains that related with the development of city clusters are supposed to describe their comprehensive development and represent their characteristics from temporal and spatial dimensions. Fig. 9 shows the geographical analyzing results for the 4 domains, classifies all city clusters into six levels according to their performance on the basis of the training set and predicting set. Therefore, we'll discuss the overall condition of the six city clusters with their geographical features so as to work out the coherent bond among them.

Fig. 9 Geographical analyzing results of six city clusters, social economics level in (a), environmental protection level in (b), climate change level in (c) and marine disaster level in (d).

Fig. 9 (a) demonstrates the social economics layout of six city clusters, in which YRDCC, BTHCC and PRDCC show their advantage over other regions. Considering the economic foundation of Shanghai, Jiangsu, Zhejiang, Beijing and Guangdong, these provinces nearly lead Chinese development all the time, so this ranking result has a certain reference for regional planning.

Compared with other areas, SPCC shows best performance in terms of environmental protection (as shown in Fig. 9(b)), the discharge amount of wastewater, sulfur dioxide and smoke dust in Shandong Peninsula are all significantly less than that of other city clusters (as shown in Fig. 6). On the contrary, environmental pollutants in BTHCC and YRDCC are listed as the last two items of the ranking system, which is closely bound with the industry layout of these two city clusters. BTHCC hosts numerous steel plants, electricity power plants and non-ferrous metal metallurgy plants with developed coal energy around Beijing, Tianjin and Hebei, these plants produce substantial gaseous pollutants like sulfur dioxide, nitrogen dioxide, nitrogen monoxide, carbon monoxide and hydrogen sulfide, etc., which directly causes the prominent emission of adverse gases in north China. YRDCC owns quantities of electronics plants, textile plants and printing plants, these factories will discharge quite a lot of wastewater to treatment plants yearly; meanwhile, the municipal sewage pipeline network is well-developed in east China due to the prosperous economy there, which assists to collect wastewater effectively, so YRDCC reaches highest in terms of sewage discharge (as shown in Fig. 6(a)).

Among the four categories involved in climate change, i.e. temperature, precipitation, sunshine duration, relative humidity, all city clusters show various performances that are related to their spatial location. PRDCC located in the lowest latitude and MLCC located in the highest latitude tend to react sharply in response to carbon emission therein (as shown in Fig. 9(c)). Considering their relative coastal location, the variation of temperature and precipitation plays a significant role during the process of climate change. For those cities in between, the situation of climate change can be milder especially for

SPCC. The limited pollutants discharge obviously helps Shandong Province alleviate the climate change effect caused by carbon emission.

Owing to their relatively high latitude, MLCC, BTHCC and SPCC are less inclined to be affected by tropical storms, thus their marine disasters are apparently less frequent than that of southern city clusters (as shown in Fig. 9(d)). Nevertheless, those cities located in sub-tropical monsoon climate and tropical monsoon climate are more vulnerable and exposed under the risk of tropical storms and sea level, which leads to the high level of their marine disasters eventually.

3.2 Analysis of evaluation results

3.2.1 Urban adaptivity of predicting set

Based on our description of the urban adaptivity evaluation system, the adaptivity of all the city clusters is calculated in the predicting period according to their indicators of the four domains, which is also visualized in Fig. 10 to display their urban adaptivity level, respectively.

It's easy to draw the conclusion that MLCC and BTHCC show the best performance from 2018 to 2021 and SPCC keeps the lowest urban adaptivity all the time. We assume that the difference among northern city clusters is bound with their industry layout and economic level, although SPCC discharges the least amount of pollutants, it's still not considered to be an ideal representative that balances the economy and environment due to its shortage in GDP. Meanwhile, southern city clusters commonly behave poor which may be related to their exposure to marine disasters.

The urban adaptivity evaluation system aims to quantitatively describe the corresponding relation between a city's prosperity and their pressure under the framework of climate change and environmental pollution, thus it's regarded as a standard that represents this balance. With this system, we can furtherly analyze the comprehensive performance of six city clusters based on their statistical data.

Fig. 10 Urban adaptivity level from 2018 to 2021 of six city clusters, adaptivity level of 2018 in (a), adaptivity level of 2019 in (b), adaptivity level of 2020 in (c), adaptivity level of 2021 in (d).

3.2.2 Urban adaptivity in average

Urban adaptivity of six city clusters are presented in Table S1 and its normalized value is showed in Fig. 11. Meanwhile, Fig. 12 describes the average of urban adaptivity and its ranking result, which contains the dataset between 2010 and 2021, so it reveals the comprehensive evaluation result of the six city clusters. Among which BTHCC and YRDCC lead the ranking result, which fits the reality to a certain extent. The complex problem between social development and environmental protection seems to find its ideal solution there.

Fig. 11 Heatmap of urban adaptivity from 2018 to 2021

Fig. 12 Geographical map of urban adaptivity level on average.

3.3 Error analysis

Due to lack of monitoring and statistical data, only GDP is selected as the validation indicator to analyze the error of prediction results. We use the data set between 2018 and 2020 as validation set, and the Mean Relative Error (MRE) is chosen to demonstrate the performance of RF (as shown in Table 2).

Table 2 Mean Relative Error of prediction results within six city clusters.

City cluster	MLCC	BTHCC	SPCC	YRDCC	SFCC	PRDCC
MRE/%	8.20	13.08	40.29	26.89	73.29	42.64

The overall error analysis result is not as ideal as expected, especially for city clusters in south China, such as SFCC and PRDCC. For the predicting mechanism of RF algorithm, it can only calculate based on statistical data, while the boosting trend of GDP with decreasing carbon emission is opposite with its historical data, so the prediction result has an apparent relative error.

4. Conclusions

With consideration of the effect of carbon emission on social economics, environmental pollution, climate change and marine disasters, we set indicators of the above 4 domains and construct a predicting model based on carbon emission with Random Forest algorithm. Six coastal city clusters of China (MLCC, BTHCC, SPCC, YRDCC, SFCC and PRDCC) are selected as the study area. By inputting the training set under certain time series, we obtain the prediction set of the following 5 years, thus evaluating the urban adaptivity of the six city clusters. With the above modeling and evaluation process, this paper puts forward the following conclusions:

1) BTHCC and YRDCC show an obvious advantage over other city clusters in terms of social economics, while they also suffer heavy pollution in gaseous pollutants emission and wastewater discharge, respectively, which is bound with their industry layout and structure. MLCC receives substantial annual pollutants loads as well, while it's less vulnerable to climate change and marine disasters due to its relatively high latitude. Similarly, SPCC also varies limitedly in temperature,

precipitation and sea level, while the low level of GDP still affects its overall evaluation although its apparent achievement in environmental protection. SFCC and PRDCC are more likely to be influenced by climate change, which can be directly presented in the annual highest temperature and frequency of tropical storms, etc. Therefore, each city cluster has its own problem to be solved, which is necessary to attach more importance to local policy-makers.

2) The characteristics of six city clusters are closely related to their geographical location and industry foundation. The traditional economic dominant provinces within Yangtze River Delta and Pearl River Delta show their advantage over other areas, thus they keep leading the list of GDP. Meanwhile, quantities of heavy industry plants distributed in north China also cause substantial pollutants discharge therein, which is quite different from that of south China. City clusters near the South China Sea are significantly affected by the monsoon yearly, so climate change and marine disasters are required to be furtherly emphasized in these cities.

3) Urban adaptivity considers the compatibility between social development and environmental protection. Therefore, we construct an urban adaptivity evaluation system that comprehensively considers all aspects of the above 4 domains to provide guidelines for future urban planning based on RF prediction results. Northern city clusters perform better due to their less exposure to climate change and marine disasters, and southern cities are supposed to focus more on the balance between economics and environment. The assessment result of six city clusters is significant, it may assist local policy-makers to design cities more properly in terms of industry structure and layout.

Appendix A

Table S1 Urban adaptivity of six city clusters from 2010 to 2021

Year	MLCC	BTHCC	SPCC	YRDCC	SFCC	PRDCC
2010	0.195	0.246	0.016	0.532	0.148	0.323
2011	0.520	0.582	-0.010	0.597	0.356	0.387
2012	0.570	0.519	-0.028	0.508	0.413	0.515
2013	0.704	0.537	0.243	0.359	0.299	0.397
2014	0.610	0.459	0.027	0.518	0.585	0.534
2015	0.504	0.374	0.119	0.401	0.651	0.378
2016	0.278	0.554	0.621	0.414	0.433	0.439
2017	0.339	0.621	-0.012	0.248	0.383	0.565
2018	0.417	0.632	0.013	0.546	0.342	0.427
2019	0.655	0.403	-0.012	0.453	0.603	0.435
2020	0.581	0.490	0.180	0.477	0.412	0.405
2021	0.523	0.632	0.041	0.538	0.519	0.503
Average	0.491	0.504	0.100	0.466	0.429	0.442

Author Contributions: Conceptualization and methodology, K.Z.; LID simulation and analysis, K.Z.; writing—original draft preparation, K.Z.; writing—review and editing, Y.Z.; supervision, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable .

Competing interests The authors declare no competing interests.

References

- Abbas Mardani, Huchang Liao, Mehrbakhsh Nilashi, Melfi Alrasheedi, Fausto Cavallaro. A multi-stage method to predict carbon dioxide emission using dimensionality reduction, clustering, and machine learning techniques, *Journal of Cleaner Production*, 2020, 275, 122942.
- Abbas Mardani, Yee Van Fan, Mehrbakhsh Nilashi, Robert E. Hooker, Seckin Ozkul, Dalia Streimikiene, Nanthakumar Loganathan, A two-stage methodology based on ensemble Adaptive Neuro-Fuzzy Inference System to predict carbon dioxide emission, *Journal of Cleaner Production*, 2019, 231, 446-461.
- Abdullah Emre Caglar. The importance of renewable energy consumption and FDI inflows in reducing environmental degradation: Bootstrap ARDL bound test in selected 9 countries, *Journal of Cleaner Production*, 2020, 264, 121663.
- Afshari A, Friedrich LA. Inverse modeling of the urban energy system using hourly electricity demand and weather measurements, Part 1: Black-box model. *Energy Build*, 2017, 157, 126–138.
- Chang C, Lin T. Estimation of carbon dioxide emissions generated by building and traffic in Taichung city. *Sustainability*. 2018, 1.
- Fan, B., Li, Y. & Pavao-Zuckerman, M. The dynamics of land-sea-scape carbon flow can reveal anthropogenic destruction and restoration of coastal carbon sequestration. *Landscape Ecol*, 2021, 36, 1933–1949.
- Gerami MH, Rabbaniha M. Forecasting the anchovy Kilka fishery in the Caspian sea using a time series approach. *Turk J Fish Aquat Sci*. 2018, 18, 1288–1292.
- Hao Gu, Meng Yang, Chong-shi Gu, Xiao-fei Huang. A factor mining model with optimized random forest for concrete dam deformation monitoring, *Water Science and Engineering*, 2021, 14, 330-336.
- Qichang Xie, Xingyu Wang, Xiaoping Cong, How does foreign direct investment affect CO2 emission in emerging countries? New findings from a nonlinear panel analysis, *Journal of Cleaner Production*, 2020, 249, 119422.
- Qi RH, Li DJ, Zhang LZ. Performance investigation on polymeric electrolyte membrane-based electrochemical air dehumidification system. *Appl Energy*. 2017, 208, 1174–1183.
- Stoms DM, Davis FW, Andelman SJ, Carr MH, Gaines SD, Halpern BS, Hoenicke R, Leibowitz SG, Leydecker A, Madin EM, Tallis H, Warner RR. Integrated coastal reserve planning: making the land-sea connection. *Front Ecol Environ*, 2005, 8, 429–436.
- Li, Y., Dong, H. & Lu, S. Research on application of a hybrid heuristic algorithm in transportation carbon emission. *Environ Sci Pollut Res*. 2021, 28, 48610–48627.
- Takahashi T, Thornton B, Sato T. Temperature based segmentation for spectral data of laser-induced plasmas for quantitative compositional analysis of brass alloys submerged in water. *Spectrochim Acta Part B At Spectrosc*. 2016, 124, 87–93.
- Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inf Decis Making*. 2019, 19, 281.
- Wang CC. Modelling of the compressive strength development of cement mortar with furnace slag and desulfurization slag from the early strength. *Constr Build Mater*. 2016, 128, 108–117.

Wang, J., Yang, F. & Chen, K. Regional carbon emission evolution mechanism and its prediction approach: a case study of Hebei, China. *Environ Sci Pollut Res*, 2019, 26, 28884–28897.

Weiyang Fan, Yu Hao, An empirical research on the relationship amongst renewable energy consumption, economic growth and foreign direct investment in China, *Renewable Energy*, 2020, 146, 598-609.

Wei Liang, Ming Yang, Urbanization. economic growth and environmental pollution: Evidence from China, *Sustainable Computing: Informatics and Systems*, 2019, 21, 1-9.

Xiaoyuan Guan, Xinman Zhu, Xiaojing Liu, Carbon Emission, air and water pollution in coastal China: Financial and trade effects with application of CRS-SBM-DEA model, *Alexandria Engineering Journal*, 2022, 61, 1469-1478.

Xuyang LU, Genwei CHENG, Feipeng XIAO, Jihui FAN, Modeling effects of temperature and precipitation on carbon characteristics and GHGs emission in Abies fabric forest of subalpine, *Journal of Environmental Sciences*, 2008, 20, 339-346.

Yang JW, Xiao XP, Mao SH, Rao CJ. Grey couple prediction model for traffic flow with panel data characteristics. *Entropy*. 2016, 18, 454.

Yifan Wen, Ruoxi Wu, Zihang Zhou, Shaojun Zhang, Shengge Yang, Timothy J. Wallington, Wei Shen, Qinwen Tan, Ye Deng, Ye Wu, A data-driven method of traffic emissions mapping with land use random forest models. *Applied Energy*. 2022, 305, 117916.

Yuan Fang, Xiaoqing Lu, Hongyang Li, A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage, *Journal of Cleaner Production*. 2021, 328, 129657.

Zhang J, Jiang HQ, Liu GY, Zeng WH. A study on the contribution of industrial restructuring to reduction of carbon emission in China during the five Five-Year Plan periods. *J Clean Prod*, 2018, 17, 629–635.

Zhao ZD, Hu CZ. Grey prediction models for the standard limit of vehicle noise. *Proc Inst Mech Eng Part D J Automob Eng*. 2018, 232, 973– 979.

Zhao, Z. Zhao, H. Zhang, The impact of growth, energy and financial development on environmental pollution in China: New evidence from a spatial econometric analysis, *Energy Econ*. 2019, 104506.

Zia Ur Rahman, Shoukat Iqbal Khattak, Manzoor Ahmad, Anwar Khan. A disaggregated-level analysis of the relationship among energy production, energy consumption and economic growth: Evidence from China, *Energy*, 2020, 194, 116836.