ISSN: 1813-4890

Based on ARIMA and LightGBM's time fusion Transformer: EsemGGDPpred Time Series Prediction Model

Xuefen Liu, Zebin Ma and Yonglin Zou

School of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang 524088, China.

Abstract

This paper explored the potential positive effects of GGDP in mitigating climate change.We constructed a time series prediction model based on ARIMA (p=2, d=1, q=1) and LightGBM's time fusion Transformer: EsemGGDPpred Time Series Prediction Model, to predict the expected global impact of adopting GGDP on climate mitigation. Among the three categories of developed countries, developing countries and underdeveloped countries, four typical countries (the United States, China, Germany and Kenya) were selected for analysis, and a comprehensive analysis was carried out from the aspects of policy and national economic level. Our study nicely illuminates the potential positive impact of substitution, showing that adoption of GGDP can be effective in slowing or reducing the growing trend of emissions of greenhouse gases and air pollutants, such as CO2 and NO2, for both developed and developing countries, but for resource-dependent In less developed countries, the effect of GGDP is not obvious.

Keywords

GGDP; EsemGGDPpred; Time Series Forecasting; Sustainable Development.

1. Introduction

For any country, economic development is very important. However, economic devel-opment inevitably consumes resources, and economic development often has a negative im-pact on the environment. GDP is an important macroeconomic indicator reflecting econom-ic development, but it does not reflect these negative effects of economic development on resources and the environment. Green GDP, on the other hand, is the cost of resource deple-tion and environmental losses caused by economic development deducted from GDP. There-fore, it reflects the interaction between the economy and the environment to a certain extent and is one of the important indicators reflecting sustainable development.[1]

The Brundtland Commission's 1987 report to the United Nations, Our Common Future, explicitly defined sustainable development as "Sustainable Development that means meet the needs of the present without compromising the ability of future generation to meet their own needs."[2] As the theme of this era, many indicators are needed to measure the extent of sustainable development. As one of the newly proposed indicators, GGDP frequently ap-pears in newspapers and on television. Can it replace GDP as the main indicator of a coun-try's economic health? In response to the above, it is necessary to build a mathematical model consisting of factors influencing the calculation of GGDP to quickly analyze and de-termine the main criteria for measuring the health of a country's economy.

To solve this evaluation problem, we divided it into 2 tasks:

Task 1: In order to be able to calculate GGDP accurately, we need to select a method that has been proposed to calculate GGDP and that has a significant impact on climate mitigation.

Task 2: Building on the calculations given in Task 1, establish a more stable and concise model using GGDP as the primary indicator of a country's economic health to estimate the expected global impact on climate mitigation.

2. Literature Review

This issue is about proposing "Green" GDP (GGDP), where "green" refers to the inclusion of environmental and sustainability perspectives and factors, and exploring whether it is feasible to replace GDP with GGDP. In recent years, because GDP does not take into account the natural environment, experts and scholars believe that it may not be a good indicator of a country's true economic health. Therefore, "Green" GDP (GGDP) is widely discussed.

First, as early as 1971, MIT scholars proposed an ecological demand indicator to quantify the relationship between economic growth and resource and environmental pressures, and then scholars in various countries proposed "economic welfare scales" that combine environmental and sustainable development with GDP, but they are rather general. In 1993, the United Nations established SEEA-1993 [3], which has been continuously developed and amended.

Second, in [4], Li et al. used the GGDP accounting system promoted by the State Environmental Protection Administration of China, i.e., GGDP = GDP - environmental pollution cost - resource consumption cost, which calculates GGDP and can reflect the degree of harmony between economic growth and nature conservation, but the environmental pollution cost and resource consumption costs have a broader meaning and are not accurate enough when selecting indicators. In [5], Neve et al. used the adjusted net savings accounting system proposed by the World Bank, i.e., GGDP = net national savings minus energy depletion, mineral depletion, net forest depletion, and carbon dioxide and particulate matter emission damages.

Finally, in [6], Deng et al. selected indicators such as sewage treatment cost, waste treatment cost, and other pollution treatment cost in the GGDP calculation method, and indicators such as land depletion, water depletion, and energy depletion in the resource consumption cost indicators, which provide guidance for us in the selection of indicators.

3. Models Introduction

3.1. GGDP Accounting Methodology

Based on the existing statistics and combined with the proposed international GGDP accounting method, this paper proposes the calculation formula of GGDP accounting method, which can well reflect the degree of harmony between economic growth and nature conservation, as seen in Eq. (1).

$$\text{GGDP} = \text{NNS} - \text{ED} - \text{MD} - \text{NFD} - CO_2 \text{ and PMED} \tag{1}$$

Where GDP refers to Gross Domestic Product, ED refers to Energy Depletion, which mainly includes depletion of land resources and water, MD refers to Mineral Depletion, NFD refers to Net Forest Depletion, which mainly includes the value of forest area increased or decreased, and finally, CO_2 and PMED refer to carbon dioxide and Particulate Matter Emission Damages.

3.2. EsemGGDPpred Time Series Prediction Model

For the EsemGGDPpred Time Series Prediction Model, our model has two modules, a feature engineering module and a time series prediction module based on integrated machine learning.

3.2.1. Feature Engineering Modules

This module is mainly used for data collection and data preprocessing. The data sources for this paper are shown in Table 1.

ISSN: 1813-4890

Table 1 Data and Database Websites

Database Names	Database Websites
Economy & Growth	https://data.worldbank.org.cn/indicator/
Environment	https://stats.oecd.org/Index.aspx?DataSetCode=ENV_GIDDB#

Based on the global coverage, we selected the data of GDP and GGDP of four countries, China (Asia), Germany (Europe), the United States (America), and Kenya (Africa), and to better show the trend of GDP and GGDP development in the four countries, we collected the data of the four countries from 1990 to 2018 and presented them in the form of line graphs, as shown in Figure 1.

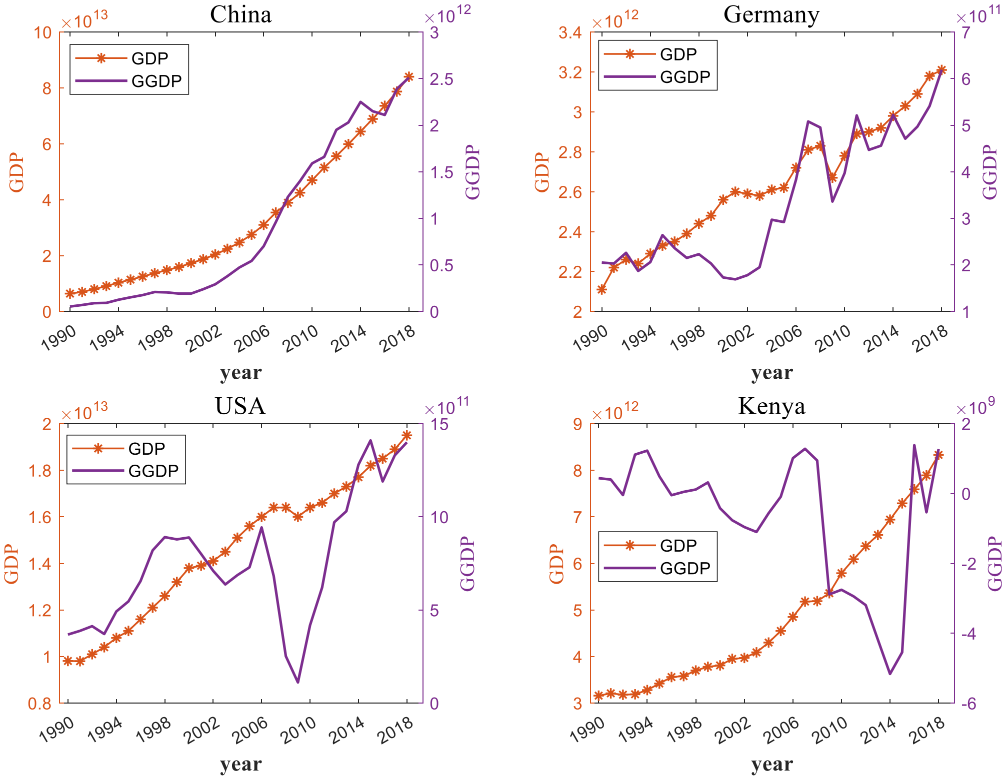

Fig. 1 GDP and GGDP trends for four countries, 1990-2018

Considering the line graph we made, which shows the development trend of GDP and GGDP of four countries from 1990 to 2018, we can easily conclude that the GDP of all four countries has been steadily increasing year by year, but the GGDP of Germany and the U.S. have seen a period of slippage, while the value of GGDP of Kenya even fell to negative values at one point.

For data preprocessing, we first unify all time formats into pandas.timestamp format, then process missing value data, and finally perform sliding window processing on the data.

Then we deal with the missing values. First, visually analyze the missing values. The more blank values, the more serious the missing values are. Missing values for the four countries are plotted below. The missing densities of economic and climate data for these four countries are shown in Figure 2.

ISSN: 1813-4890

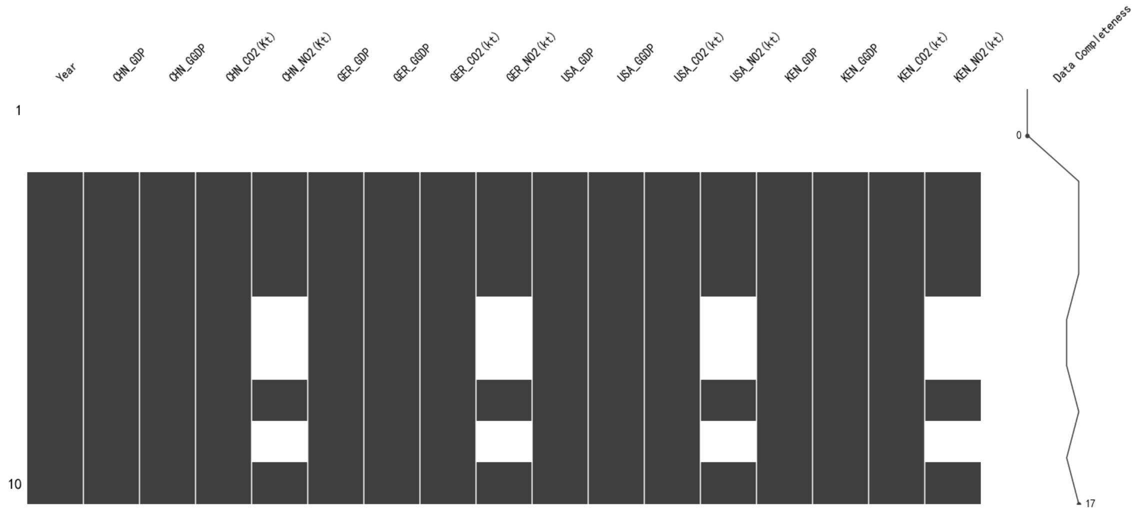

Fig. 2 Density of correlation data before processing missing values

From the figure 2, it can be seen that there are missing NO2 values in all four countries, and the data function is fitted using the third spline interpolation method in this paper.

Cubic Spline Interpolation can be used to smooth the curve. It is defined as follows: For the interval $[p,q]$, $(p=1, q=10), p=S_0<S_l<\cdots<S_n=q$, these $\mathrm{n}+1$ nodes and the function values at these points $f(S_i)=y_i\ \ (i=0,1,\cdots,n), y_i$ is the dependent variable we need to study. if the function $g(S)$ satisfies three conditions, then $g(S)$ is the cubic spline interpolation function of $f(S)$ with respect to n nodes.

Then by this fitting function, the missing values are filled in and the data density plot after filling in the missing values is shown in Figure 3.

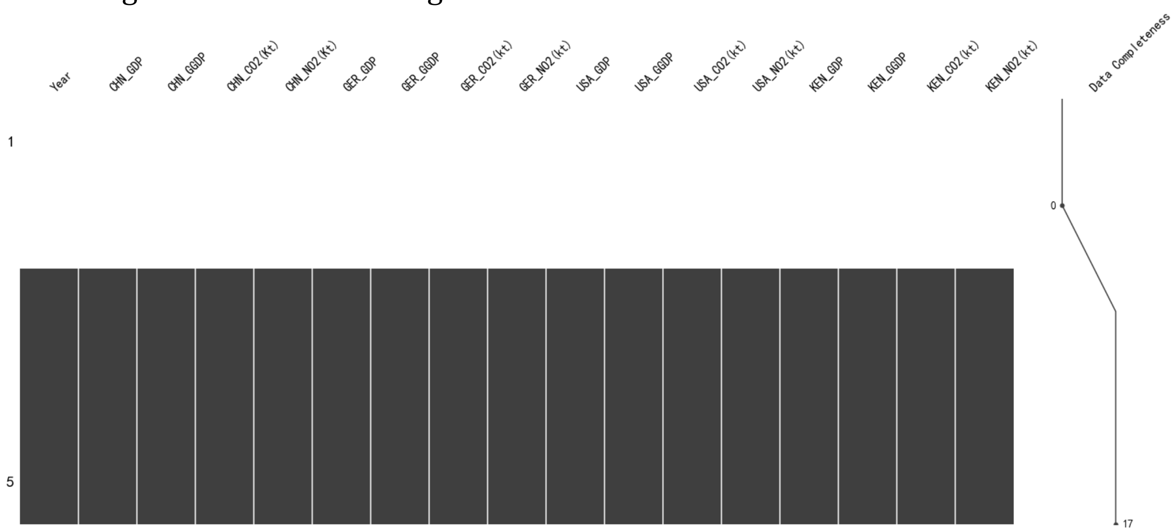

Fig. 3 plot of correlation data density after processing missing values

Finally we apply time sliding window processing to the data.Time sliding window processing is a technique used in time series analysis to create fixed length overlapping windows that slide along the time axis to generate a new set of data points, this technique can help capture local patterns or trends in time series data. By defining the size of the sliding window, defining the span, and generating the sliding window, i.e., completing the sliding window processing of the data, the data is ready to be used for model training.

3.2.2. AMIRA

For the time series model, our model has three modules, namely, statistical time series prediction module, boosted tree time series prediction module and deep learning time series prediction module. For the statistical time series prediction module, we use ARIMA; for the

ISSN: 1813-4890

boosted tree time series prediction module, we use XGBoost model; for the deep learning time series prediction module, we use Temporal Fusion Transformer. Due to space limitation, we only give a brief introduction to the ARIMA model.

ARIMA (Autoregressive Integrated Moving Average) is a time series model used for forecasting. It is a class of three parameter models that capture autocorrelation, trend, and seasonality in data with parameters p, d, and q.

The general form of the ARIMA model is shown in Eq. (2).

$$\mathbf{ARIMA}(\mathbf{p},\mathbf{d},\mathbf{q}) = \mathbf{AR}(\mathbf{p}) + \mathbf{I}(\mathbf{d}) + \mathbf{MA}(\mathbf{q}) \tag{2}$$

where: **AR** (**p**) is autoregressive component of order p, **I**(**d**) is differential component of order d, and **MA** (**q**) is moving average component of order q.

The autoregressive order (p) indicates how many past values of the dependent variable should be considered to predict the next value. This parameter can be used to capture trends in GGDP changes. For example, if the value of p is set to 2, the model will consider the GGDP values of the previous two years to predict this year's GGDP; the order of integration (d) indicates the number of times the data need to be differentiated to make it stationary. A stationary time series has a constant mean and variance. The integration order can be used to capture any seasonality or trends in GGDP data that may lead to discontinuity; the moving average order (q) indicates the number of past forecast errors that should be used to predict the next value, and this parameter can be used to capture the effects of random fluctuations in GGDP data that may not be explained by the autoregressive and integration components.

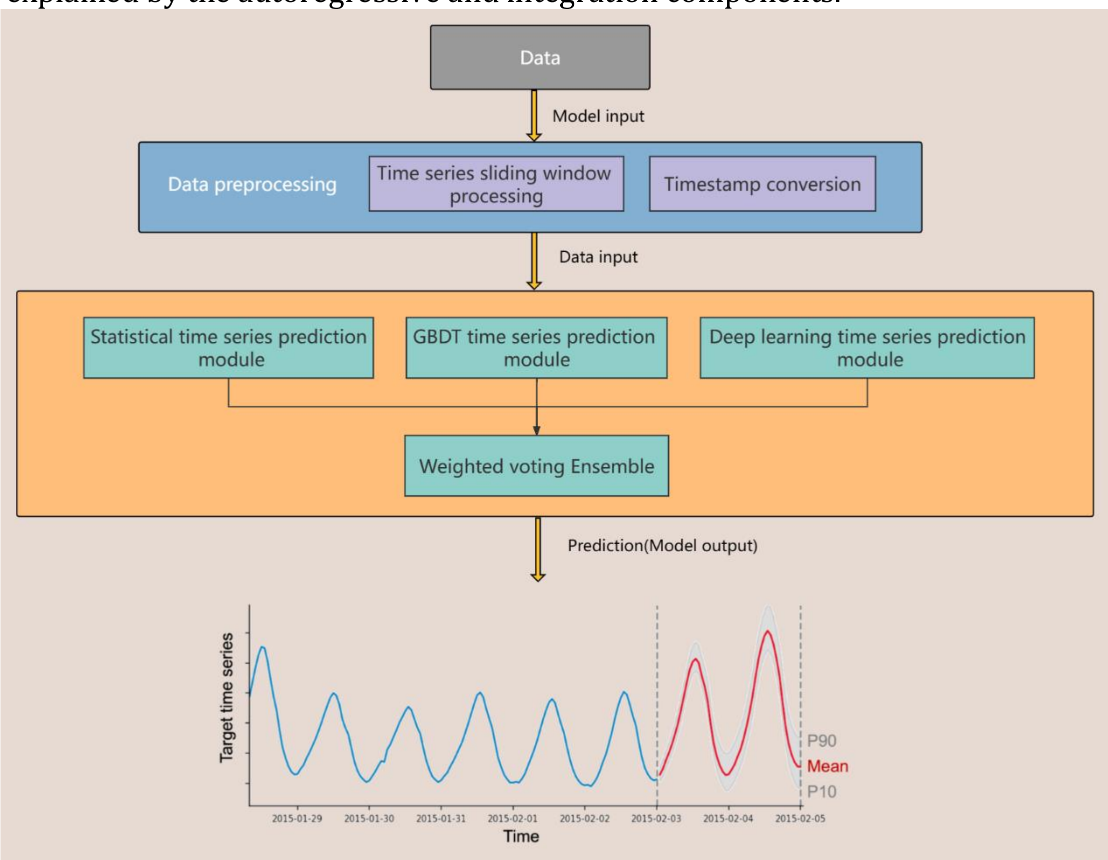

Fig. 4 The structure of the timing model constructed in this paper

3.2.3. Weighted Voting Esemble

For the prediction results of each model, this paper uses a weighted voting method to integrate each model. In this approach, the model trains multiple sub-classification models and then combines the prediction results of all models by weighted voting, while the weight assigned to each model is based on the model's performance. The integration method combines the prediction results of the three modules, which helps to improve the accuracy and robustness of the predictions and reduces the sensitivity of the models to outliers and noise in the time series data.

3.3. Results and Analysis

After the data is constructed, we put the data into the model for training and get the following results. Base data represents the real value, while prediction represents the model prediction value.

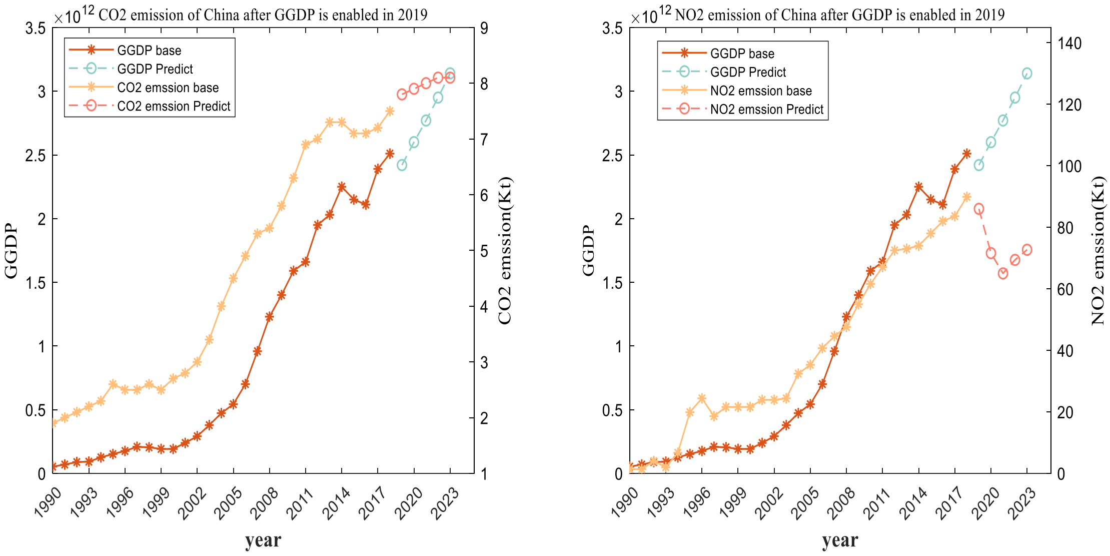

Fig. 5 CO2 emission, NO2 emission and GDP growth curve of China

Observing the CO2 emission, NO2 emission and GGDP growth curve of China, we can see that the trend of CO2 emissions growth in China has slowed down and NO2 emissions have decreased year by year since the adoption of GGDP as a measure in 2019.

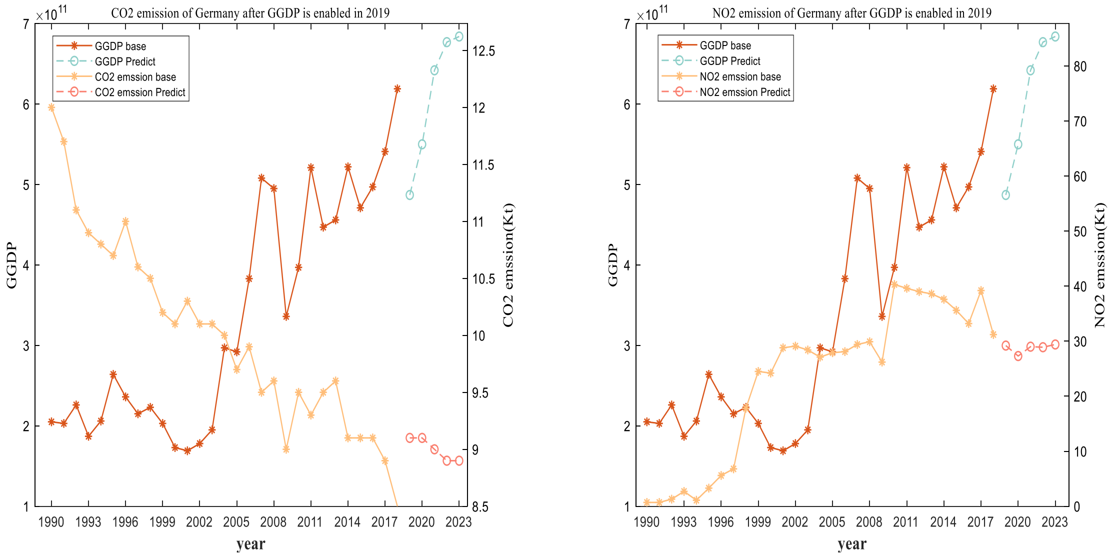

Fig. 6 CO2 emission, NO2 emission and GDP growth curve of Germany

Observing the CO2 emission, NO2 emission and GGDP growth curve of Germany, it can be seen that GGDP causes a decline in 2019 while CO2 emissions increase until 2020, when GGDP values return to normal and grow steadily, with CO2 emissions decreasing year by year and NO2 emissions stabilizing year by year.

ISSN: 1813-4890

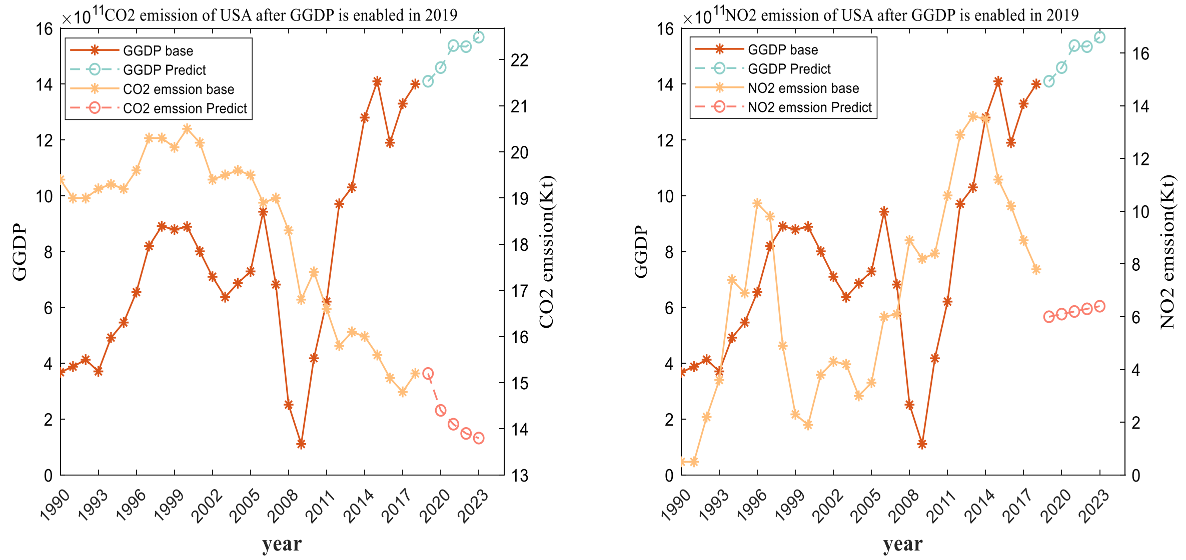

Fig. 7 CO2 emission, NO2 emission and GDP growth curve of USA

Observing the CO2 emission, NO2 emission and GGDP growth curve of USA, it shows that after adopting GGDP as a measure in 2019, the U.S. continues to see an increase in GGDP from 2019 to 2023, with CO2 emissions decreasing year by year and NO2 emissions stabilizing after declining since 2019.

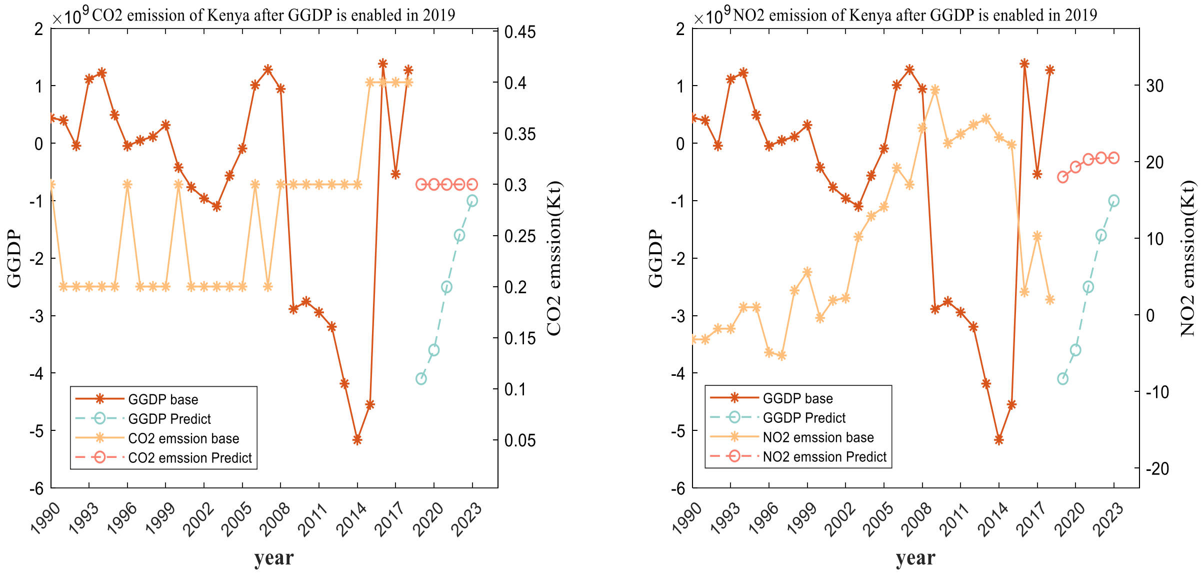

Fig. 8 CO2 emission, NO2 emission and GDP growth curve of Kenya

Finally, we observe the CO2 emission, NO2 emission and GGDP growth curve of Kenya, where a substantial decline in GGDP occurs from 2019 to 2023 after GGDP was adopted as a measure in 2019, with NO2 emissions declining but still rising year by year.

4. Conclusion

Analysis of changes in GGDP, NO2, and CO2 indicators for four countries – China, the U.S., Germany, and Kenya – helps provide insight into changes in global GGDP, as they represent most groups of countries in the world (developed, developing, and less developed). China and the U.S., as major carbon emitters, have experienced slower growth in their CO2 emissions, lower nitrogen dioxide emissions, and continued increases in GGDP since the adoption of GGDP as a measure. This is due to the growing focus on clean energy and environmental policies in both countries, setting an example for the world to follow. Germany's transition to a green, low-carbon economy has also been largely successful, as evidenced by the country's adoption of the GGDP as a measure. Although the value of carbon emissions has fluctuated in relation to GGDP growth, it is still broadly improving. Finally, Kenya, as a sample of less developed countries, has a declining trend in GGDP and carbon emissions are still increasing year by year. This is due to the fact that the country, as a resource-dependent underdeveloped country, still needs to consume a large amount of natural resources to develop its economy. Overall, changes in economic growth and pollutant emissions in these typical countries reflect the trend towards sustainable global economic development, and there are still major challenges in promoting the transition to a green and low-carbon economy in countries around the world.

References

[1] Qiu Guoxin. An objective view of GDP and green GDP[J]. Mall Modernization, 2007(12S):2.

[2] Environmentdevelopment W C O . Our Common Future. 1987.

[3] Noritoshi ARIYOSHI. The System for Integrated Environmental and Economic Accounting (SEEA):[J]. Studies in Regional Science,1993,26(1)

[4] Li Jinlei. Green development accounting of Daihai watershed based on GGDP, GEP and GEEP [D]. Inner Mongolia University,2021.

[5] MARTIN NEVE; BERTRAND HAMAIDE (2017). Environmental Kuznets Curve with Adjusted Net Savings as a Trade-Off Between Environment and Development. Australian Economic Papers, 56(1), 39–58.

[6] Deng Lu, Li Xiaodong, He Chunxia et al. A study on green GDP accounting: Xinjiang 2003-2009 as an example[J]. Anhui Agricultural Bulletin (second half of the month),2012,18(16):30-34.