Calculation and Analysis of Green GDP based on EWM and Multiple Linear Regression Model

Xinyi Luo *, Zixu Chen, Lai Wang

School of Information Science, Beijing Language and Culture University, Beijing, China, 100083

* Corresponding author: xk22lxy@126.com

Abstract. With the acceleration of the industrialization process, global problems such as resource shortage and ecological environment deterioration have become increasingly prominent. To uphold the concept of sustainable development, a real, feasible and scientific indicator: green GDP needs to be established urgently. But it is still difficult to promote green GDP on a global scale due to a lack of ideas and technology. In order to calculate and predict GGDP, resource and environmental factors are taken into account. Based on entropy weight method (EWM) and multiple linear regression, an improved global GGDP accounting method is constructed. The future development should pay attention to pollution prevention and ecological environment protection.

Keywords: Green GDP, Entropy Weight Method, Multiple Linear Regression Model.

1. Introduction

As an important symbol of the current national economic accounting system SNA, GDP reflects the total volume and scale of the national economy [1]. However, it only considers the cost of economic operation but excluding resources and environmental issues, This kind of development at the expense of damaging environment and consuming resources is not in line with the concept of sustainable development. Green GDP(GGDP), which highlights environmental and sustainability viewpoints, came into being.

Since 1970s, the United Nations, the World Bank and other international institutions and governments have been making painstaking efforts to construct GGDP as the new measure of economic health[2]. Several important international systems of green national accounts currently include SEEA developed by the United Nations, ENRAP developed by Professor Peskin and applied in the Philippines, SERIEE developed by the European Union, and NAMEA developed by Statistics Netherlands[3], among which SEEA is the most mature and recognized framework[4]. According to national conditions, each country calculated its GGDP differently. In 1992, BEA established IEESA based on SEEA and ISEW. Japan adopted the NAMEA based on SEEA-2003 and developed a hybrid integrated accounting system. Focused on its forest resources, Finland employed NAMEA to calculate green GDP[5]. But none of them have yet been able to calculate GGDP comprehensively.

In order to construct a universal GGDP calculation model and consider the different resource and environment conditions of different countries, the EWM with weight reflecting the degree of importance is used to meet the needs of this study. Based on traditional GDP, combined with EWM, an improved global GGDP accounting method is constructed.In addition, the multiple linear regression is used to quantitatively analyze secondary indicators, such as resources and environment, and further proposing global initiatives.The research path is shown in Figure 1:

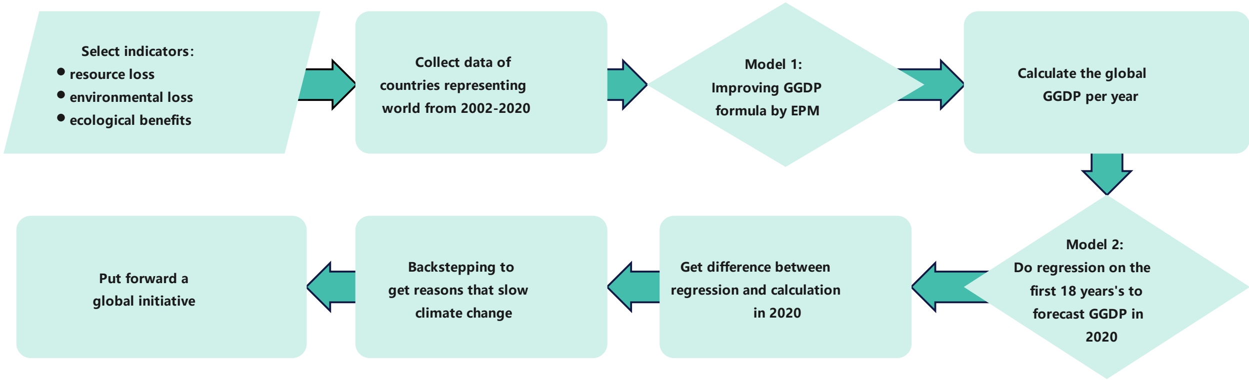

Figure 1. Flow Chart

2. Materials and Methods

2.1. Data Collection and Pre-processing

This study collect the data of 15 indicators reflecting national economic health from 2002 to 2020 from 15 representative countries on all continents such as GDP, resource consumption(coal, oil, natural gas, etc), discharge of waste water and exhaust gas(CO2, SO2, NO2, CH4, etc.), and adjust their basic units to be consistent. They mainly come from the United Nations and the World Bank. These 15 countries are: the U.S., Brazil, Canada, Netherlands, Britain, France, Germany, Italy, Russia, China, Japan, India, Egypt, Africa, South Africa and Australia.The resources of data are of official authority.For data which are not allowed to copy or download, we acquire them via crawler using javascript and python.

Since data collection involves multiple countries and different time points, it is difficult to assure they are fully complete in the procession of collecting.To guarantee the availability of the data and credibility of conclusions,exponential smoothing is adopted.It is suitable for short-term development trend prediction. Compared with the moving average method, which does not consider long-term data, the weighted moving average method gives more weight to recent data, which is also in line with the fact that resource consumption and other data are more sensitive to recent data.

2.2. GGDP computational method

The analysis of natural resource asset accounting in GGDP mainly focuses on two aspects. One is SEEA led by the U.S.and the other is NAMEA represented by Norway. Focusing on the accounting of the physical flow of resources and environment related to the national economic system, both of them provide more effective ways to accurately solve the ecological environmental problems such as greenhouse effect and environmental pollution. In addition, Germany adopts the basic theory and principle of SEEA to reflect the logical connection of pressure, environment and response between human economic activities and the natural environment, and builds GEEA [6].

However, the model still has room for optimization and improvement. Take the Netherlands as an example,although the modal reveals dynamic change relationship between the indicators, the coverage of environmental costs needs to be extended [5], and the study mainly focused on environmental issues in the Netherlands, which is not universal for the calculation of GGDP at a global scale and requires country-specific consideration of natural resources.

SEEA developed by the UN is the most widely accepted calculation system [4], and to apply the theory into practice, it has went through summaries and revisions for several times.So this paper combine it with EWM to determine the calculation of GGDP [7]. The cost accounting of environment and resources is included, and the cost on various kinds of pollutants is considered from two aspects-prevention and control [8] [9].

2.3. Entropy Weight Method

Entropy weight method is an objective weighting method, which avoids the interference of human factors on the weight of indicators. The degree of variation of the indicator is directly proportional to the amount of information it reflects. The smaller the degree of variation of the index, the less information reflected, and the lower its corresponding weight should be, and vice versa [10].

The basic steps of weighting are as follows: firstly, a positive matrix is constructed for all the indicators of each sample and standardized, then the proportion of the ith sample under the jth index is calculated, and then the information entropy of each index is calculated according to formula (1) and normalized to obtain the weight of each index.

$$E_p = -\ln(m)^{-1} \sum_{i=1}^{n} (Y_{ip}) \times \ln(Y_{ip}) \ (0 \le E_p \le 1) \tag{1}$$

Where m is the number of sample and Y_{ip} is data for the i_{th} sample under the same index.

2.4. Multiple Linear Regression

Multiple linear regression can determine the relationship between one dependent variable and more independent variables. It is predicted by fitting a line or surface called the regression model or equation to the data points that minimizes the total error [11]. The equation of the regression is as follows:

$$y = \sum_{i=0}^{n} \beta_i \cdot x_i + \mu \tag{2}$$

Among them, y is the variable to be interpreted, which represents the GGDP values in this study, partial regression coefficient β_i is the weight of the index, x_i is the value of index which is linearly dependent on y separately, μ is the constant to be fitted.

To ensure excellent interpretability and predictability of the regression model, significant independent variables should be selected, and the correlation between them should be lower than the correlation between independent variables and dependent variables. Moreover, it is necessary to conduct hypothesis testing, observe the overall significance of the regression equation through the F-test, calculate the goodness of fit R^2, p value, and the estimated value of error variance s^2, etc., to evaluate the fitting effect of the model.

3. Model construction and solution

3.1. Establishment and solution of improved GGDP model based on EWM

In order to evaluate the overall economic situation rigorously, on the basis of traditional GDP, this study incorporates resource and environmental factors into the calculation of GGDP, and constructs four first-level indicators to measure gross domestic product, resource loss in development, environmental pollution and ecological benefits respectively. EWM is used to assign different weights to each first-level index, and an improved model is established. The calculation formula is as follows:

$$GGDP = \alpha GDP + \beta Loss_{resource} + \gamma Loss_{Environment} + \delta Ecological\ benefit \tag{3}$$

The model is mostly used to compare the relative gap between different countries, so the weight of each index obtained by employing EWM can analyze the amount of information between the data more objectively.

Among them, GDP and $Ecological\ benefit$ are positive indicators, while $Loss_{resource}$ and $Loss_{Environment}$ are negative ones, $\alpha, \beta, \gamma, \delta$ are weights for each indicator. Except for GDP which is calculated by traditional GDP. Other first-level indicators equal to the sum of the corresponding secondary indicators. $Loss_{resource}$ represents the value of the environmental impacts mainly caused by the consumption of coal, oil, and natural gas. In this study, the recovery cost method of unit resource is employed to estimate the loss value of consuming energy resources[12]. The loss value of each resource is equal to the product of the unit price of the resource and its consumption. The environmental damage value $Loss_{Environment}$ consists of two parts: environmental degradation value and environmental damage value. The former mainly accounts for the cost of treating various environmental harmful substances, which belongs to the virtual treatment cost without actual expenditure and is not reflected in the traditional GDP accounting. The latter focuses on the loss of environmental value or property caused by pollution or natural disasters, as well as the loss of depreciation of fixed assets. The calculation of $Ecological\ benefit$ is the same as $Loss_{resource}$, which is measured by the area of forests, farmlands, grasslands, and lakes in each country for a span of 18 years and average unit price of services for each type of ecosystem.

The above four indicators were calculated separately for each country from 2002 to 2020. EWM is used to determine the weight of each index by calculating the information entropy.

3.2. Establishment and solution of Multiple Linear Regression prediction model

According to formula (2), use multiple linear regression to predict the GGDP value in 2020, calculate the difference between it and the real value of GGDP in the same year obtained by the optimization model, and analyze the efforts supposed to make to maintain the same economic growth rate as the current one. In order to save resources and protect the environment, according to the change of global overall data, specific suggestions are put forward.

First, sum the GGDP data(shown in Figure 1) for 15 countries to represent the characteristic data for each continent as well as globally.Then,with the help of Stata, perform the analysis using OLS in multiple linear regression, depending on the sum of GGDP and various indicators from 2002 to 2019, perform significance tests and obtain the coefficients of each secondary index. Finally, apply formula (3) and secondary indexes to predict the world GGDP in 2020, and compare it with the real value.The calculation formula is as follows:

$$I = \frac{p_{GGDP} - a_{GGDP}}{a_{GGDP}} \tag{4}$$

Where p_{GGDP} is predicted value of GGDP and a_{GGDP} is the actual value. The difference between them is the total amount of GGDP expected to be increased. Then, calculate the value of the second-level indicators that need to change in order to maintain the current economic development situation, and make recommendations for mitigating the global climate problem accordingly.

4. Results and analysis

4.1. Analysis of GGDP model

According to the collected data, GGDP of 15 countries is calculated over the years, the results are as shown in Figure 2:

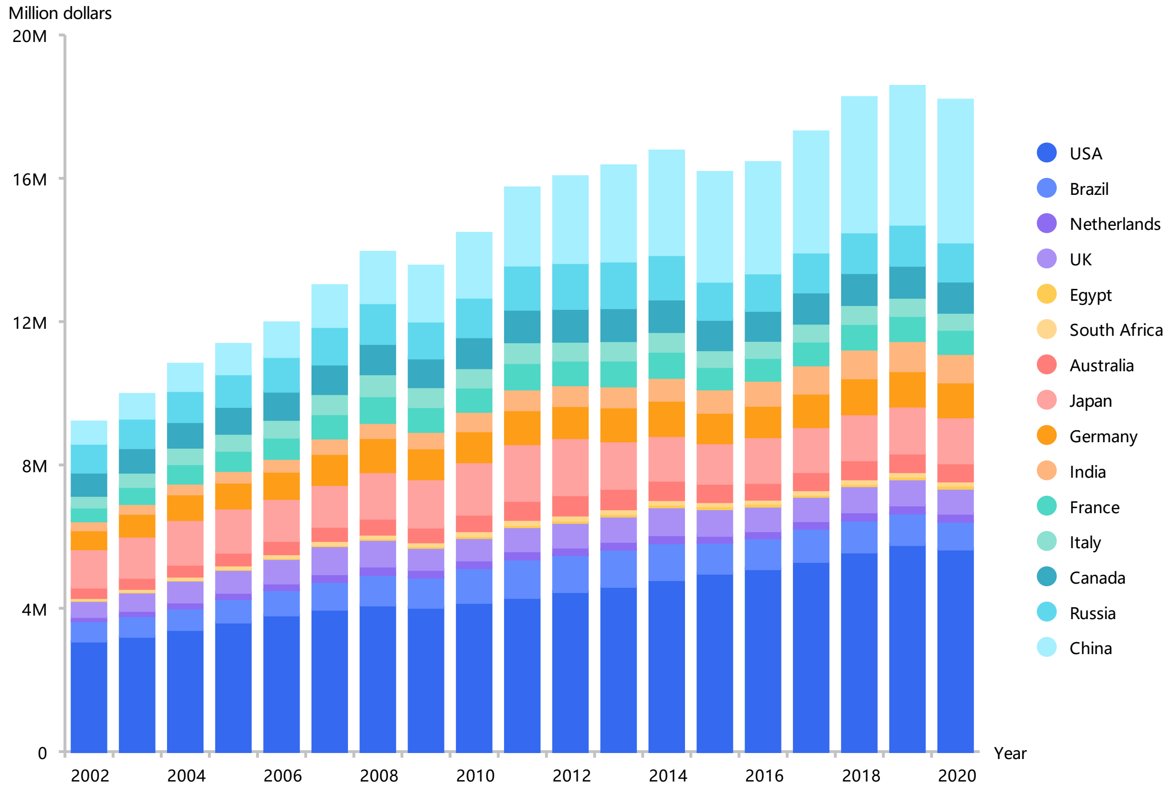

Figure 2. GGDP of 15 countries from 2002 to 2020

Taking the U.S. as the representative data, calculate information entropy of first-level indicators, results are shown in Table 1.Weight of first-level indicators are: α=0.2522, β=0.1414, γ=0.2048, δ=0.4015. (results are preserved four decimal places).

Table 1. information entropy of fist-level indicators

Information Entropy ($E_{p(*)}$)	Value
E_{pGDP}	0.9442
$E_{pResource}$	0.9687
$E_{pEnvironment}$	0.9547
$E_{pEcology}$	0.9111

The index weight reflects the importance of each factor to the overall GGDP evaluation, and the descending order is: ecological benefits, gross domestic product, resource loss and environmental loss. Specifically, efforts should be made in the following aspects.In terms of resources, countries should adopt alternative strategies of energy consumption such as fossil fuels. It is essential to significantly reduce petroleum consumption and popularize clean energy[13], which is beneficial to reduce emissions of noxious gas, such as CO_2, CH_4, PM_{10} and SO_2. In terms of ecosystem, expand the area of farmland and forest not only to improve the ability to consolidate soil and preserve fertilizer but increase the amount of carbon sequestration and oxygen release.As for biological species and global temperature change rate, the annual growth rate is stable,which leads to a larger weight. As human factors are difficult to directly affect it, planting trees and returning farmland to forest are indirect ways to tackle these problems.

4.2. Analysis of Multiple Linear Regression prediction model

The results of significance analysis of multiple linear regression analysis by Stata are shown in the Table 2:

Table 2. significance index value

significance indicators	values
F(15, 2)	618.19
Prob > F	0.0016
R^2	0.9998
Adj R^2	0.9982

As seen from Table.1, the corresponding F-test probability value is 0.0016, which is far less than 0.01, thus rejecting the original hypothesis with 99% certainty. That is, each regression coefficient is significantly not equal to 0. The overall regression is significant. Meanwhile, the regression possesses both high R^2 and adjusted R^2, which also means that it can be well-fitted. Although the heteroskedasticity test reveals the existence of multicollinearity in some of the independent variable indicators, regression is mainly used for forecasting rather than analyzing the importance of the indicators, so ignore this issue.

Table 3 is the coefficient of each second-level index obtained by regression.

Table 3. coefficients and units of second-level indicators

Indicators	Coefficients	Units
coal consumption	2.0167	short tons
oil consumption	-2941.747	megaton
natural gas consumption	10179.77	one billion cubic meters
discharge of waste water	0.8476	kilogram
carbon dioxide emissions	-0.0864	kiloton
methane emissions	-2.0235	kiloton
nitric oxide emissions	1.8454	kiloton
PM10 emissions	-5182.702	mg/m^2
sulfur dioxide emissions	-8.3011	kiloton
farmland area	152.0759	thousand hectare
forest area	7.8702	thousand hectare
meadow area	-143.9362	thousand hectare
inland waters area	-2.1217	thousand hectare
biological species change rate	-2780493	%
global temperature change rate(unsmoothed)	-681858.7	%

Figure 3 shows that the predicted value of GGDP is larger than the actual value, and this difference symbolizes the change that will occur after the transition. Calculate the amount of value each metric would need to change to keep the economy growing at the same or faster rate. It is suggested that coal,oil and natural gas consumption be reduced by 0.338%,0.441% and 0.117% respectively and reduce water consumption by 0.1056 per year. Besides, increase the treatment rate of industrial "three wastes" (waste gas, waste water and solid waste) by 2.7%. Increase the investment in noise, domestic waste and other garbage treatment by 0.147%, 2.043%, 0.527%. On the one hand, plant trees and cultivate forests and conserve water areas.Compared with 2019, 205.55 thousand hectares of forest should be planted, 68.1 hectares of wetland and 173.173 thousand hectares of lake should be added to the protected area.On the other hand, restrict activities that disrupt biodiversity such as hunting.In general, to change the mode of economic growth and vigorously develop circular economy is the only way to alleviate the problem of resource depletion and reduce environmental pollution.

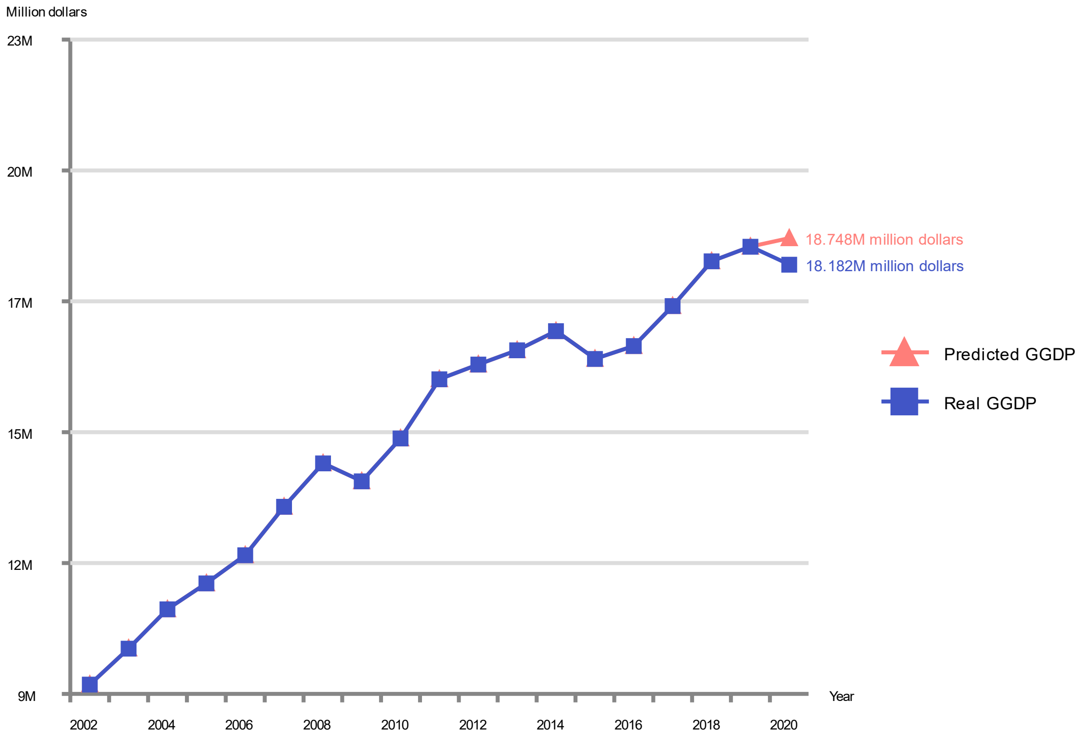

Figure 3. global GGDP value from 2002 to 2020 and forecast value in 2020

5. Conclusion

In this paper, a GGDP calculation and forecasting model is built by using EWM and multiple linear regression, comprehensively considering parameters and collecting data from authoritative sources with high precision. Meanwhile, exponential smoothing is adopted to interpolate missing values, which enhances the accuracy and validity of the model. The experimental results shows that all countries should practice the concept of sustainable development, save energy and reduce emissions, and jointly build a green home. Nevertheless, indicators selected and modelling process need to improve00. The sample data of the model is relatively insufficient. More training examples are required to improve the accuracy of the model in further study.

References

[1] Gao Minxue, Xu Jian, Zhou Jingbo. Comprehensive environmental and economic accounting [M]. Economic Science Press, 2007.

[2] Hu Angang. China: Green Development and Green GDP (1970-2001) [J]. China Science Foundation of China, 2005 (02): 22 - 27.

[3] Gerlagh R, Dellink R, Hofkes M, et al. A measure of sustainable national income for the Netherlands[J]. Ecological Economics, 2002.

[4] Xu Haiping. Study on total factor productivity and national income accounting based on environmental factors [D]. Zhejiang University, 2008.

[5] Daly, H.E., Cobb, J.B. (Eds.), 1989. For the Common Good. Beacon Press, Boston.

[6] Han Dejun, Yang Guangqing, Zheng Anmin. Comparison of typical national natural resource asset accounting mechanism and the construction of Chinese chemical system [J]. Accounting communication, 2022, No.903 (19): 154 - 161.

[7] Wu Chunyou, Guo Lingling, Yu Jingtao. Evaluation model of regional green growth system based on TOPSIS-grey association analysis and empirical [J]. Management Review, 2017, 29 (01): 228 - 239.

[8] Zhu J, Liu X, Jin X, et al. Calculation of Environmental Degradation Cost based on GGDP and BP-Neural Network [C]//Wuhan Zhicheng Times Cultural Development Co., Ltd. Proceedings of 2nd International Conference on Business, Economics, Management Science (BEMS 2020).2020: 7.

[9] Nawapanan Ekkaporn and Kongboon Ratchayuda and Sampattagul Sate. Green GDP Indicator with Application to Life Cycle of Sugar Industry in Thailand [J]. Sustainability, 2022, 14 (2): 918 - 918.

[10] Zhu Yuxin, Tian Dazuo, Yan Feng. Effectiveness of Entropy Weight Method in Decision-Making [J]. Mathematical Problems in Engineering, 2020, 2020: 1 - 5.

[11] Korkmaz Mehmet. A study over the general formula of regression sum of squares in multiple linear regression [J]. Numerical Methods for Partial Differential Equations, 2020, 37 (1): 406 - 421.

[12] Yao Lihui. Research on the construction of a comprehensive green GDP accounting system based on SEEA-2012 —— Take Hunan Province as an example [D]. Hunan: Central South University of Forestry and Technology, 2017.

[13] Hussain Zahid et al. Green Growth, Green Technology, and Environmental Health: Evidence from High-GDP Countries [J]. Frontiers in Public Health, 2022, 9: 816697 - 816697.