

Employ Mathematical Modeling to Summarize, Analyze, And Predict the Relationship Between Carbon Dioxide and Temperature, Location, And Other Factors.

Yuheng Xie¹, Yanxi He², Yufei Ye³ and Chenfei Liu⁴

¹Beijing National Day School, Beijing 100039, China

²Alcanta International School. Guangzhou 511458, China

³Nanjing Xuanwu Foreign Language School, Nanjing 213012, China

⁴Nanwai Xianlin Foreign Language School Xianlin Branche, Nanjing 210023, China

Abstract: Fluctuations in the amount of greenhouse gases (GHGs) are affected by a variety of factors, both natural and artificial. Natural factors, such as volcanic eruptions and changes in the intensity of solar radiation, can cause fluctuations in greenhouse gases. However, in recent years, human activities have surpassed natural factors and have become a major driver of climate change. During the industrial revolution, the widespread use of fossil fuels, large-scale deforestation and emissions from factories caused irreversible damage to ecosystems. Rising global temperatures are destroying natural ecosystems around the world and are a potential threat to human survival. In this paper, we systematically examine and analyze the main sources of greenhouse gas emissions, explore the factors contributing to air pollution, and evaluate the specific impacts of these changes on global warming. We analyze the relationship between global warming and different factors (economic development, clean energy, and CO₂ emissions) by using both one-way linear regression models and multiple linear regression models. A quadratic linear regression model shows that solar energy is the most promising renewable energy source. Meanwhile, using the global GDP as a criterion, we have shown that economic growth contributes to CO₂ emissions. However, due to the governments having valued economic development more, the upward growth trend of the economic development curve is intensified, while the downward growth trend of the global average temperature is slowed down. Finally, we developed a multivariate linear regression model to evaluate whether using new energy sources could help mitigate global warming. Based on the previous analysis, we focus on the factor of CO₂ emission. By using the linear regression model, we obtained the projected values of CO₂ emissions, clean energy usage and global GDP in 2050, and the CO₂ emissions in 2050 are 36.59439 billion tons, which meet the goal of carbon neutrality.

Keywords: Carbon dioxide, greenhouse gases, clean energy use, temperature, modeling.

1. Introduction

1.1. Background

Over the past century, there have been significant changes in the climate, mainly in the form of global warming. Since 1950s, global greenhouse gas (GHG) concentrations have continued to rise rapidly, and global average surface temperatures have maintained an overall upward trend. Meanwhile, a number of climatic problems have occurred, such as rising sea levels, melting of the highland snowpack, retreating of Arctic Sea ice, and rising sea surface temperatures. By the beginning of the twenty-first century, there was little dispute that greenhouse gas emissions were causing global warming.

On February 2, 2007, the U.N. Intergovernmental Panel on Climate Change (IPCC) released a summary of its Fourth Global Assessment of Climate Change, which consolidated six years of research by scientists around the world[1]. The report concludes that global warming has become an "indisputable" fact. According to the report, the global average temperature from now to 2100 is likely to rise by 1.8°C to 4°C, while sea level could rise by 18 to 59 centimeters. The IPCC notes that the global average surface temperature has risen by 0.74°C over the past 100 years, and the past 50 years have been the warmest on record[2]. According to the World Meteorological Organization (WMO), the global average temperature in 2006 was 0.42°C higher than the 1961-1990 average, it was the sixth warmest year on

record. The IPCC committee concluded that the sharp rise in global temperatures coincided with a period of large-scale emissions of greenhouse gases, and that human activities are the main cause of global warming.

To help people better understand climate change, we need to develop a simplified model that describes the relationship between carbon dioxide concentrations and changes in the Earth's temperature. With this model, we can predict future climate changes and CO₂ emission, and explain how they relate to global warming. This will help to demonstrate the trend of global climate change and its impact on the environment, as well as to clarify the causes of global warming. By alerting people to potential risks, we can motivate decision makers to adopt policies and regulations to address global climate issues.

1.2. Problem description

Not only do we need to perform a linear regression of location factors such as latitude and longitude, but we also need to predict the average annual temperature for the year 2050 based on the correlation between time and temperature. Finally, we will create a report that examines the relationship between global temperature and factors that could be the causes, and determines how important a role has CO₂ played.

2. Notation and Hypothesis Analysis

2.1. Notation Description

Table 1. The Description of Notations

Notations	Meaning
T	Global average annual temperature/°C
G	Global annual GDP
C	Global annual CO ₂ emissions/ton
P	Global population/0.1 billion people
E	Global clean energy use
a ₁	GDP weighting
a ₂	Annual CO ₂ emissions weighting
a ₃	Global population weight
a ₄	Clean Energy Use Weight
l ₁	Latitude
l ₂	Longitude
b ₁	Latitude Weights
b ₂	Longitude Weights
Y	Time
b ₃	Time weights
W	Weight Matrix
b ₄	Error term

2.2. Assumptions

1. It is assumed that the increase in CO₂ emissions due to the use of fossil fuels will not stop abruptly;

2. A long-term assessment is made, i.e. there are enough data sources;

3. Human use of clean energy increases gradually according to current policies and trends;

4. Any fuel usage by 2050 is consistent with the current trend in fuel use levels;

5. The location factor is divided into latitude and longitude, the temperature longitude is used as the independent variable.

3. Model Analysis and Establishment

3.1. Explore the relationship between temperature change and time

To explore the relationship between location, time and temperature, we use the regression model to analyze the global temperature change and time between 1958-2021.

Firstly, for the quadratic equation, the parameters are set as:

$$y = p_1x^2 + p_2x + p_3 \quad (1)$$

According to the data obtained from the Heywhale Community[3], the quadratic regression analysis shows that the parameters are $p_1=0.0001.59375997$, $p_2=-0.617179359$, and $p_3=597.357428$. The covariance matrix of the fitted parameters was used for evaluation. The values on the diagonal of the covariance matrix represent the magnitude of the error for each parameter, and the parameter error for each coefficient is $a_1=0.0000399855504$, $a_2=0.159063862$, and $a_3=158.179765$. It can be seen that the predictions of a_1 and a_2

are more accurate than parameter a_3 .

For the linear equation, we set the parameter as:

$$y = q_1x + q_2b \quad (2)$$

Similarly, according to the linear regression, the q_1 and q_2 coefficients are 0.016818356374807917 and -33.100599718381844 , which shows that the slope is less than 0.1, indicating that the effect of time on the temperature is relatively small.

3.2. Argumentation of the relationship between temperature and time through time series

3.2.1. Augmented Dickey-Fuller Test

The test procedure of ADF test is the same as that of Dickey-Fuller test, but applies to the model

$$\Delta y = a + \beta t + a_1\Delta y_1 + \dots + a(n+1)\Delta y(n-1) \quad (3)$$

In statistics, the Augmented Dickey Fuller Test (ADF) tests the original hypothesis that there is a unit root in the time series sample[5]. The alternative hypothesis varies depending on the version of the test used, but is usually smoothness or trend smoothness. ADF can test larger and more complex sets of time series models.

3.2.2. Code Implementation and Modeling

The next step is to determine if the time series is white noise. Since it is daily data over several years, we can assume that the seasonal component repeats itself every 365 days, thus we set the period as 365. Then we should verify that the residuals are white noise (which is the same requirement as for smoothness), and we have the following figure:

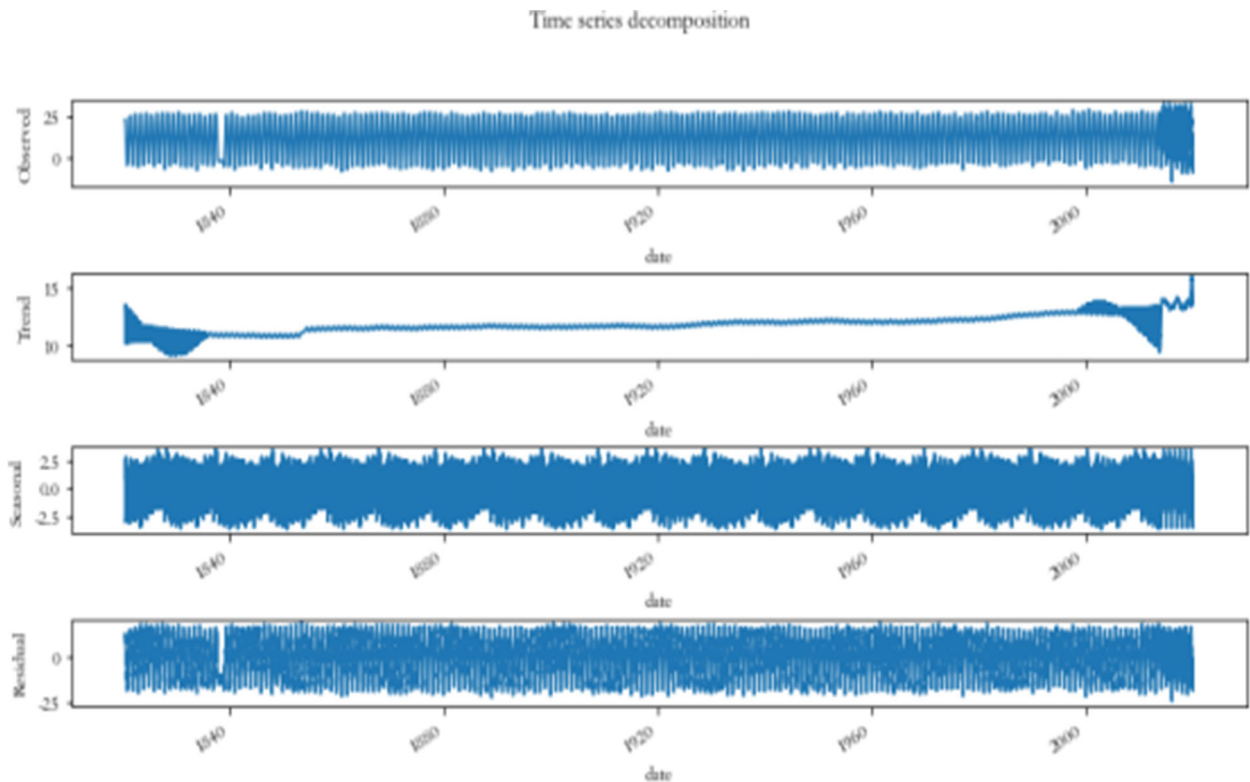


Figure 1. Seasonal decomposition for a time series

To verify that the residuals are white noise, we can check for smoothness using the Augmented Dickey-Fuller Test, get: $p\text{-value} = 0.0$.

The p -value obtained is close to 0.0, which is below our significance level (0.05). Therefore, we can reject the null hypothesis (H_0 : the time series is non-stationary), and our time series can be considered as white noise.

3.3. Relationship between temperature and longitude and latitude by linear regression

3.3.1. Data sources

With data obtained from the Heywhale Community, the autocorrelation coefficients of longitude and latitude are set to β_1 and β_2 . Since the time series has already been proved to be white noise, it is not taken into consideration.

Table 2. (excerpt)

1950/11/1	7.327	1.07	Kabul	Afghanistan	34.56N	70.05E
1950/12/1	1.849	0.881	Kabul	Afghanistan	34.56N	70.05E
1951/1/1	1.403	0.292	Kabul	Afghanistan	34.56N	70.05E
1951/2/1	2.903	0.504	Kabul	Afghanistan	34.56N	70.05E
1951/3/1	7.948	0.607	Kabul	Afghanistan	34.56N	70.05E
1951/4/1	12.697	0.588	Kabul	Afghanistan	34.56N	70.05E
1951/5/1	19.484	0.253	Kabul	Afghanistan	34.56N	70.05E
1951/6/1	23.355	0.484	Kabul	Afghanistan	34.56N	70.05E
1951/7/1	25.22	0.443	Kabul	Afghanistan	34.56N	70.05E
1951/8/1	24.432	0.249	Kabul	Afghanistan	34.56N	70.05E
1951/9/1	22.042	0.361	Kabul	Afghanistan	34.56N	70.05E
1951/10/1	17.926	0.784	Kabul	Afghanistan	34.56N	70.05E
1951/11/1	10.02	0.318	Kabul	Afghanistan	34.56N	70.05E
1951/12/1	5.549	0.391	Kabul	Afghanistan	34.56N	70.05E
1952/1/1	2.766	0.412	Kabul	Afghanistan	34.56N	70.05E
1952/2/1	5.294	0.472	Kabul	Afghanistan	34.56N	70.05E
1952/3/1	8.049	0.511	Kabul	Afghanistan	34.56N	70.05E

3.3.2. The steps of Modeling

After cleaning the data, the latitude and longitude of each

region and the average annual temperature are obtained as in table.3

City	AverageTemperature	Latitude	Longitude
Abidjan	27.228500	5.63	-3.23
Addis Abeba	19.539000	8.84	38.11
Ahmadabad	27.608625	23.31	72.52
Aleppo	19.503625	36.17	37.79
Alexandria	21.438500	31.35	30.16
...
Tokyo	13.911500	36.17	139.23
Toronto	8.459444	44.20	-80.50
Umm Durman	30.650500	15.27	32.50
Wuhan	19.285500	29.74	114.46
Xian	14.460000	34.56	108.97

Table 3. Average temperatures by region

OLS Regression Results

Dep. Variable:	AverageTemperature	R-squared:	0.203			
Model:	OLS	Adj. R-squared:	0.186			
Method:	Least Squares	F-statistic:	12.33			
Date:	Fri, 25 Nov 2022	Prob (F-statistic):	1.69e-05			
Time:	10:39:16	Log-Likelihood:	-324.33			
No. Observations:	100	AIC:	654.7			
Df Residuals:	97	BIC:	662.5			
Df Model:	2					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	22.4135	0.912	24.576	0.000	20.603	24.224
Latitude	-0.1402	0.029	-4.906	0.000	-0.197	-0.083
Longitude	0.0156	0.010	1.610	0.111	-0.004	0.035
Omnibus:	5.078	Durbin-Watson:	2.000			
Prob(Omnibus):	0.079	Jarque-Bera (JB):	5.152			
Skew:	-0.543	Prob(JB):	0.0761			
Kurtosis:	2.765	Cond. No.	118.			

Table 4. Least squares fitting results

Using multiple linear regression, the relationship between temperature and latitude/longitude was set as

$$T = l_1 b_1 + l_2 b_2 + b_4 \quad (4)$$

The results were obtained by using least squares fitting as shown in table 4.

The coefficient in the result is the value of b1, b2. And the final correlation is obtained:

$$T = (-0.1402) * l_1 + (0.0156) * l_2 + 22.4135 \quad (5)$$

And we get the R-squared of 0.8. The R-squared measures the fitness of the equation, the larger the R-squared, the better fitness. In general, greater than 0.8 means that the equation

fits the sample points very well and meets the requirements of the question. Thus, the relationship between temperature and location and time is obtained.

3.4. Prediction model based on previous years' global average temperature data and CO2 concentration data

3.4.1. Data sources

In order to test the effect of the above factors on CO₂ emissions, we use a regression model for graphical analysis. By analyzing the historical data[4], we plan to examine the relationship between the following variables and CO₂ emissions, and whether carbon neutrality can be achieved in the future by adjusting the use of clean energy.

Table 5. Global clean energy use, 2012-2021 2

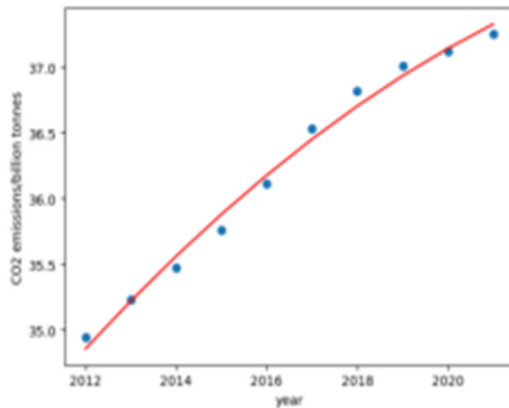
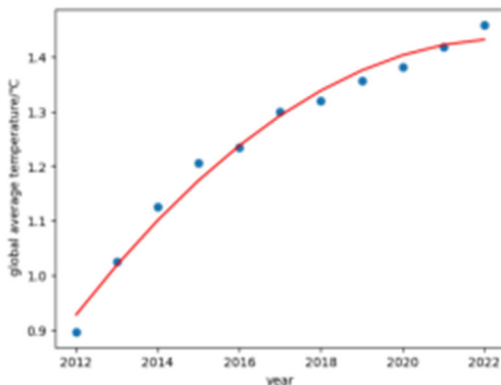
Year	Total installed capacity (MW)	Geothermal energy (MW)	Biomass (MW)	Solar (MW)	Wind energy (MW)	Hydropower (MW)
2012	1443923	10479	77198	104312	266918	1090111
2013	1566148	10717	84545	141169	299840	1137100
2014	1698319	11157	90605	180216	349307	1175994
2015	1851651	11812	96351	228054	416170	1210655
2016	2013932	12137	105245	300199	466866	1245903
2017	2184709	12697	110966	395276	514215	1270876
2018	2357495	13169	118256	488820	563466	1293619
2019	2541688	13738	123940	591059	621270	1311765
2020	2807265	14073	133025	716788	731763	1335114
2021	3063926	15644	143371	849473	824874	1360054

Table 6. Global CO₂ emissions, global GDP and global mean surface temperature, 2012-2021

Year	CO ₂ emission/billion tons	GDP/billion dollars	Baseline global mean surface temperature change ⁸ °C (°C)
2012	34.94	75277.89	0.89617
2013	35.23	75011.41	1.02535
2014	35.47	76265.41	1.12538
2015	35.76	77439.7	1.20485
2016	36.11	79500.78	1.23379
2017	36.53	81088.95	1.2988
2018	36.82	86095.66	1.32017
2019	37.01	87325.54	1.35691
2020	37.42	90960.9	1.38187
2021	38.25	96487.67	1.41853

3.4.2. Analysis of CO₂ and Temperature

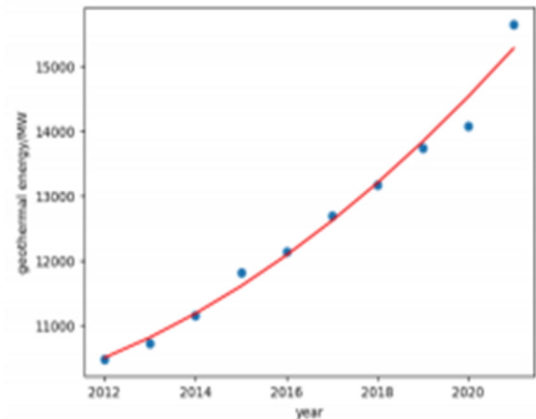
After testing, we believe that the quadratic regression model is more accurate for graphing.

**Figure 2.** Total global CO₂ emissions/billion tons**Figure 3.** Global average annual temperature/°C

By comparing the recent trends of global CO₂ emissions and global average annual temperature, we find that the temperature increases with the increasing of CO₂ emissions. It is obvious that CO₂ emission is one of the important factors causing temperature increase and global warming. Reducing CO₂ emissions can effectively mitigate the greenhouse effect.

3.4.3. Analysis of Clean Energy Use

In order to study the development potential of various green energy sources, we collected data on the global use and development of five kinds of energy sources, as well as data on the total global use of renewable energy, and set up a quadratic linear regression model to analyze which energy source is most efficient in achieving carbon neutrality.

**Figure 4.** Global geothermal energy use/MW

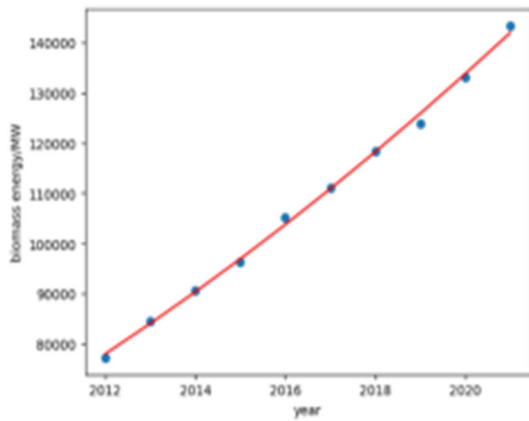


Figure 5. Global biomass energy use/MW

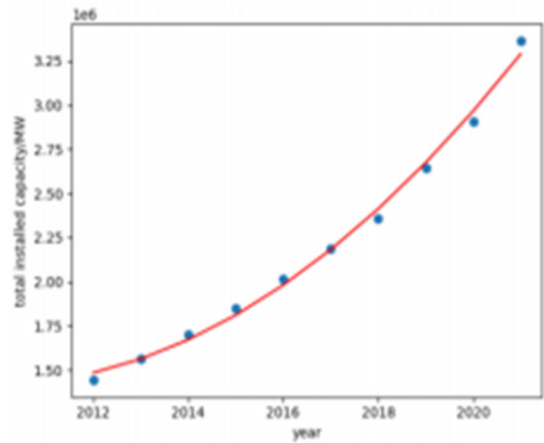


Figure 9. Total installed energy/MW (total renewable energy use)

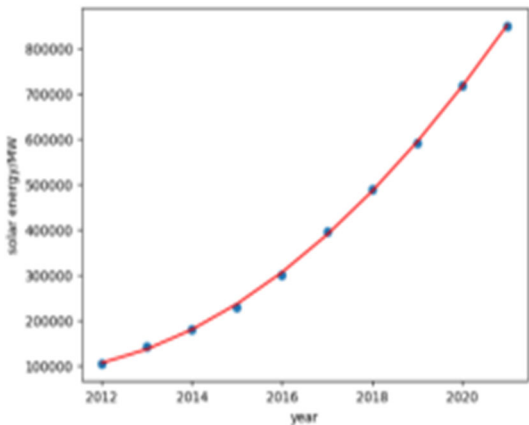


Figure 6. Global solar energy use/MW

The figures show that global renewable energy use is growing significantly, which is a major factor in the slowing down of CO₂ emissions and the warming trend. The most used renewable energy is water, at 136,054 MW. Solar energy has seen the most significant growth in use and is still rising rapidly, making it the most promising renewable energy source for the future.

3.5. Other Factors that Affect Global Warming

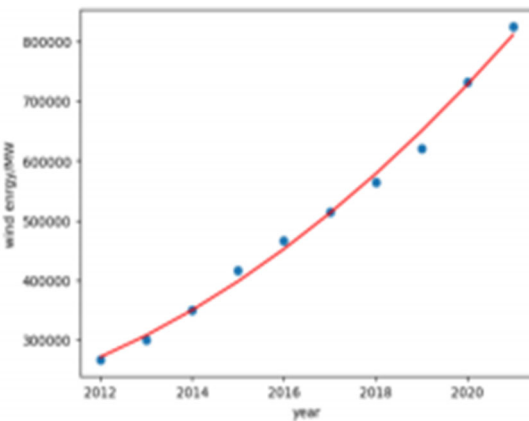


Figure 7. Global wind energy use/MW

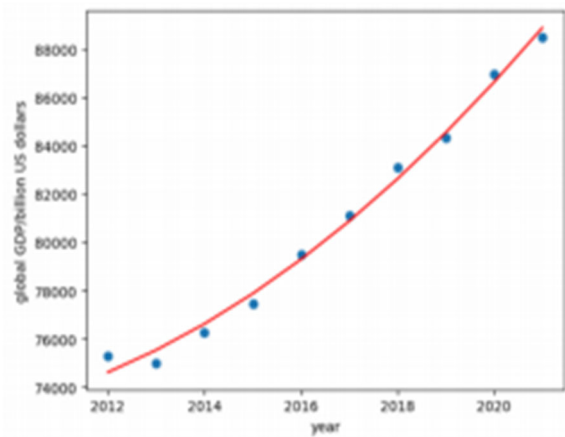


Figure 10. 2012-2021 Global GDP

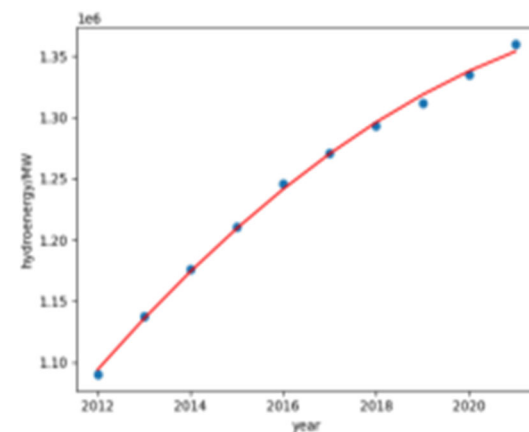


Figure 8. Global hydro energy use/MW

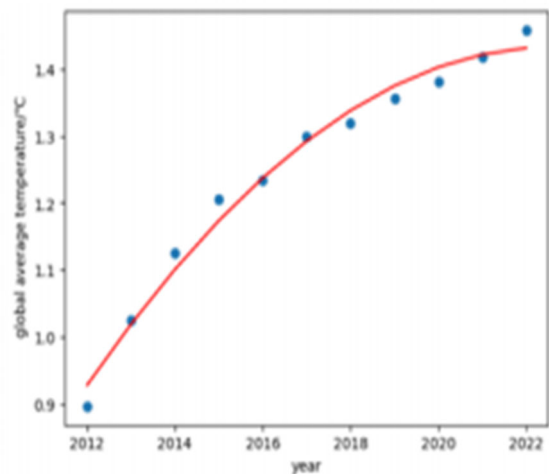


Figure 11. 2012-2021 Global Average Temperature

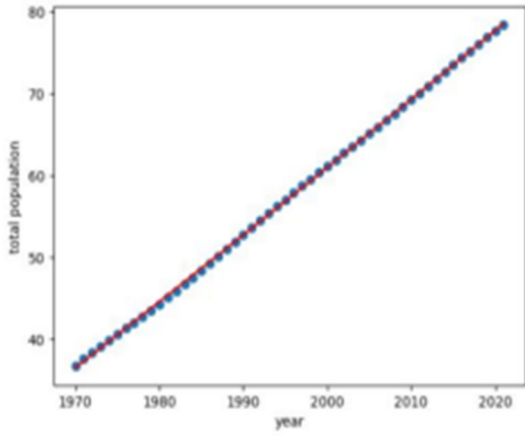


Figure 12. 1970-2021 Total Global Population

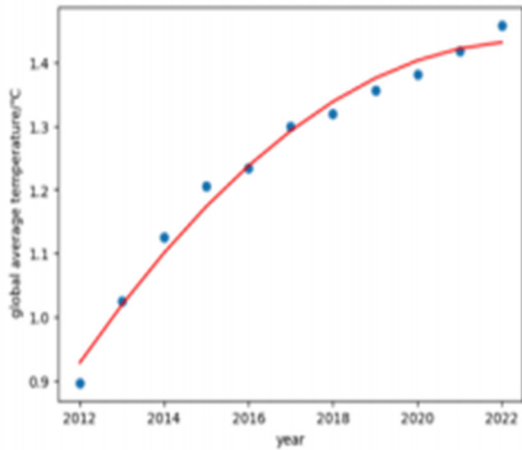


Figure 13. 2012-2021 Global Average Annual Temperature

Figure 10 shows the global GDP over the last 10 years, and we can see that economic development has been rising rapidly and is gradually accelerating. Figure 12 shows how the global population has increased steadily over the last 50 years. The global average annual temperature has also been on the rise for the last 10 years, with a slowing down growth trend.

3.6. Multiple Linear Regression Modeling

3.6.1. Modeling

Based on the above analysis, we find that there is a certain relationship between global warming and factors including economic development, CO₂ emission and population. We can establish a multiple linear regression model to evaluate whether controlling these factors is effective in reducing global warming.

We set the global average annual temperature as the dependent variable (T), and the independent variables are the global annual GDP (G), global annual CO₂ emission (C), and global population (P).

Based on the graphical data, the equation can be derived:

$$T = a_1 * G + a_2 * C + a_3 * P + e \quad (6)$$

where e is the error term and a_1, a_2, a_3 are the regression coefficients.

3.6.2. Model Calculation

We consider the effects of global GDP, global population and CO₂ emissions on global temperature. Firstly, we

standardize the data. Then, by programming the linear regression, the equation is obtained as:

$$T = 0.0303 * G + 0.0250 * P + 0.1445 * C - 6.1874 \quad (7)$$

The coefficients of G, P, C are positive, which means that G, P, C is positively correlated with T. The coefficient of C is 0.1445, which is larger than the coefficients of G and P, indicating that CO₂ emissions have the greatest impact on global temperature. Therefore, our next step will focus on CO₂.

3.6.3. CO₂ and clean energy, GDP modeling

Based on the above results, we can analyze the relationship between the three by setting up a one-way linear regression model and a multiple linear regression model. [6]

We set CO₂ emission (C) as the dependent variable, and the independent variables include: global clean energy use (E) and global GDP (G). G is positively correlated with C and E is negatively correlated with C. The following equation can be obtained:

$$C = b_1G + b_2E + f \quad (8)$$

By normalizing the data, we model and program the equation as:

$$C = 0.1281 * G - 0.2824 * E + 38.9363 \quad (9)$$

For every \$1 billion increase in total annual GDP (G), global CO₂ emissions are projected to increase by 0.1281 billion tons. For every 1 MW increase in clean energy use (E) by the global population, global CO₂ emissions are expected to decrease by 0.2824 billion tons.

From this analysis, we can conclude that in the future, clean energy use will reach carbon neutrality as the trend of clean energy use continues to grow. At that time, CO₂ emissions will peak at about 3,89363 billion tons.

3.6.4. Clean Energy, GDP Annual Forecast Modeling

In order to investigate whether carbon neutrality can be achieved by 2050, a one-way linear regression model can be developed to calculate the projected values of clean energy use and total GDP in 2050.

Assume that the relationship between global clean energy use (E) and year (y) is given by:

$$E = c_1 * y + d \quad (10)$$

The equation can be obtained by programming:

$$E = 1.91956 * y - 3848.76689 \quad (11)$$

Substituting $y = 2050$ into this equation yields $E_{2050} = 86.3311$

Assume that total global GDP (G) is related to the year as:

$$G = c_2 * y + g \quad (12)$$

The equation can be obtained by programming:

$$G = 2.63256y - 5224.71060 \quad (13)$$

Substituting $y = 2050$ into this equation gives $G_{2050} = 172.03740$

3.6.5. Projecting Carbon Emissions in 2050

Next, we substitute G_{2050} , E_{2050} :

$$C = 0.1281 * G - 0.2824 * E + 38.9363$$

Yield $C_{2050} = 36.59439$. Since $36.59439 < 38.9363$, the carbon emissions in 2050 have already decreased, and carbon neutrality is expected to be reached by 2050.

4. Summary and Discussion

4.1. Summary

Through optimization and improvement of the model, we can take more factors into account that affect global warming, such as energy efficiency, to further predict the time to achieve carbon neutrality. Therefore, we should continue to promote environmental protection measures and increase the use of clean energy to help protect the environment.

4.2. Advantages and Shortages of the Model

4.2.1. Advantages

1. In Model 1, we have analyzed the influence of time on temperature through different methods.

2. Studying the time series, we found that the effect of time on temperature is relatively small, thus eliminate the effect of time and leaves only two independent variables, which is more accurate.

3. The relationship between location and temperature is

demonstrated by linear regression.

4. Having considered multiple influencing factors and using multiple sets of data.

5. Both GDP and population are considered.

4.2.2. Shortages

1. The model only uses linear fitting, quadratic function fitting, linear regression, etc., which is relatively simple, and there are few comparisons between different models.

2. The model will have bias, and the generalizability of some models is still not good.

3. There will be some errors in the weighting analysis.

References

- [1] IPCC climate report 2022 summary: The key findings. (2023, June 19). Climate Consulting. <https://climate.selectra.com/en/news/ipcc-report-2022>
- [2] Carbon dioxide now more than 50% higher than pre-industrial levels. (2022, June 3). National Oceanic and Atmospheric Administration. <https://www.noaa.gov/news-release/carbon-dioxide-now-more-than-50-higher-than-pre-industrial-levels>
- [3] - heywhale.com. (n.d.). <https://www.heywhale.com/home/dataset>
- [4] Data Overview - Berkeley Earth. (2023, November 10). Berkeley Earth. <https://berkeleyearth.org/data/>
- [5] Wikipedia contributors. (2024, February 22). Dickey–Fuller test. Wikipedia. https://en.wikipedia.org/wiki/Dickey%E2%80%93Fuller_test
- [6] Best, V. a. P. B. C. (2013b, September 22). A fit to global temperature data. Clive Best. <https://clivebest.com/blog/?p=2353>