Factors affecting climate change and analysis*

Changes in carbon dioxide are the main factor

Dongli Sun

April 17, 2024

This paper studies the factors that cause climate change. I will analyze and forecast future climate change based on the data I have collected and combined with the paper. In the original data, I found that climate change is related to changes in carbon dioxide emissions, population density and forest area. I think that the main climate change problem is the change of carbon dioxide, so I will directly relate and expand the relationship between these three. However, I did not find a connection between them in the original data. But I'm going to give a paper on why there's a causal relationship.

1 Introduction

The data for this study of climate change come from forest land for the worldbank(World Bank 2020a), the green gas emission(World Bank 2020c) and population density(World Bank 2020b). The original data simply recorded their individual trends, but there was no complete juxtaposition discussion. I found in the original data that carbon dioxide and population density are increasing trends, while the area of forests is decreasing year by year. I think there is a certain connection between them, and I try to show the changes between them and analyze and generate new data, which I have obtained from the world bank, and present the data between 2000 and 2020. There are many categories of climate change in the database, but I mainly intercept carbon dioxide and population density and forest density as the main research objects. At the same time, I found that these factors and climate change can be two-way results. Factors that cause climate change to get worse, and then climate change causes those factors to increase dramatically, creating this kind of extreme weather. Extreme weather events occur frequently, and rising sea levels and melting glaciers have a very negative impact on local animals. In addition, climate change affects not only the living environment of animals but also human beings. Therefore, I will analyze the data of climate change factors

*Code and data are available at: https://github.com/DongliSun/Track-climate-change.git.

and integrate them to watch their trends, predict the future development and put forward some suggestions for sustainable development. The regional sections of this article are as follows: Section 2 presents the data used for analysis and discovery, including visualization of variables of interest, Section 3 presents a simple linear regression model to prove and predict the relationship between carbon dioxide, forest area and population density, Section 4 shows the interpretation of the model and other findings from analyzing the data. Section 5 discusses the implications of the findings as well as the weaknesses of the paper and its next steps for further research on the topic.

2 Data

2.1 source

The data for this study of climate change come from forest land percentagefor the world-bank(World Bank 2020a), the green gas emission(World Bank 2020c) and population density(World Bank 2020b) can be viewed and I downloaded the csv to facilitate the generation and analysis of data. This website is free and open to the public and can be downloaded and viewed directly. I will use R to complete the whole paper series and data analysis. The database includes data on carbon dioxide emissions from climate change, loss of forest area and trends in population density. I use these variables to simulate, test and build parts of the model for a second analysis. I will show the variation between different variables, but the main method of proof is to prove through the support of the paper. I chose the data between 2000 and 2020, and cleaned and tested them.

I will use R studio(R Core Team 2023) for analysis and data visualization in these processes. For tidyverse package (Wickham et al. 2021c), I will access other important R packages. The testthat package (Wickham 2021b) is mainly used for unit testing functions and code in R packages. The knitr package (Xie 2021) is used to generate dynamic reports from R code. dplyr package (Wickham et al. 2021a) is used to manipulate and clean data, readr package (Wickham et al. 2021b) is used to read and import data, here package (Müller 2021) is used to create a specific save file path, The ggplot2 package (Wickham 2021a) is used to create data visualizations, and the modelsummary package (Arel-Bundock 2021a) is used to create summary tables. The WDI software package (Arel-Bundock 2021b) is primarily used to obtain global Development data on economic, social and environmental aspects from the World Bank's World Development Indicators database. The rstanarm software package (Goodrich et al. 2022) is used to implement the estimation of Bayesian statistical models, especially generalized linear models, which are widely used in the social sciences and biostatistics.

2.2 methodology

I choose carbon dioxide emission, reduction of forest area and population density because I think they are related to each other. For example, with the increase of population density, carbon dioxide emission is bound to be much higher than before, because population needs to consume energy, and energy will produce carbon dioxide. In the same way, forests are important for absorbing carbon dioxide in the loss of forest area, and if the area is reduced, there will be more carbon dioxide. Similarly, I think the change of carbon dioxide is the main cause of climate change, but there are many small factors that can cause the change of carbon dioxide, so I give a section of the paper on the trend of their change to prove my point of view. The following is an analysis of the image display after explore data. The initial phase is about the structure of the data set. But the data sets are individual and contained within, basically every single one of them.

First of all, The data found from the World Bank on forest coverage (World Bank 2020a), the greenhouse gas emissions (World Bank 2020c), and population density (World Bank 2020b) indicate significant trends over the past decade. I only intercepted data from 2000 to 2020, which means we need to clear the data to ensure that I get the total number of years and data I want. I show it in the data run, so I change the different variables between group_by() to better compute and run. Further processing the data out of the image shows that this visualization can help us to more directly understand the trend between them, such as the first graph, the X-axis is the year, and the Y-axis is the related value. The data we calculated will be shown on the chart. Then we make a detailed discrete graph of them to represent them.

I will give the cleaning data for the metholody. As shown in the line grapgh of Figure 1. It shows the changes in the coverage between 2000 and 2020. Since 2000, the forest coverage has begun to decline. We can clearly observe that in 2005, the coverage rate dropped to about 8310%, but from then to about 2006, it kept rising. And then it keeps going down. In 2010, the forest coverage rate was about 8355%, and then from 2011 to 2013, the forest coverage rate continued to decline, directly decreased by about 8280%. If this trend is followed, I can predict that his value will definitely be lower than 8280%.

The third line chart is about carbon dioxide emissions Figure 2, which have been rising continuously since 2000, but began to decline between 2018 and 2019. I can guess that most people are isolated at home due to the emergence of 2019-COVID-19, and many factories are not working, so this may lead to the reduction of carbon dioxide. Because less energy is used. The previous data on the decline and rise of forest cover may be responsible for the continuous increase in carbon dioxide emissions.

The line graph is about population density Figure 3. It is obvious that we can learn from the figure that the population density only has a downward trend between 2001 and 2002, but keeps an upward trend after that, and will reach a maximum of 378per km^2 in 2020.

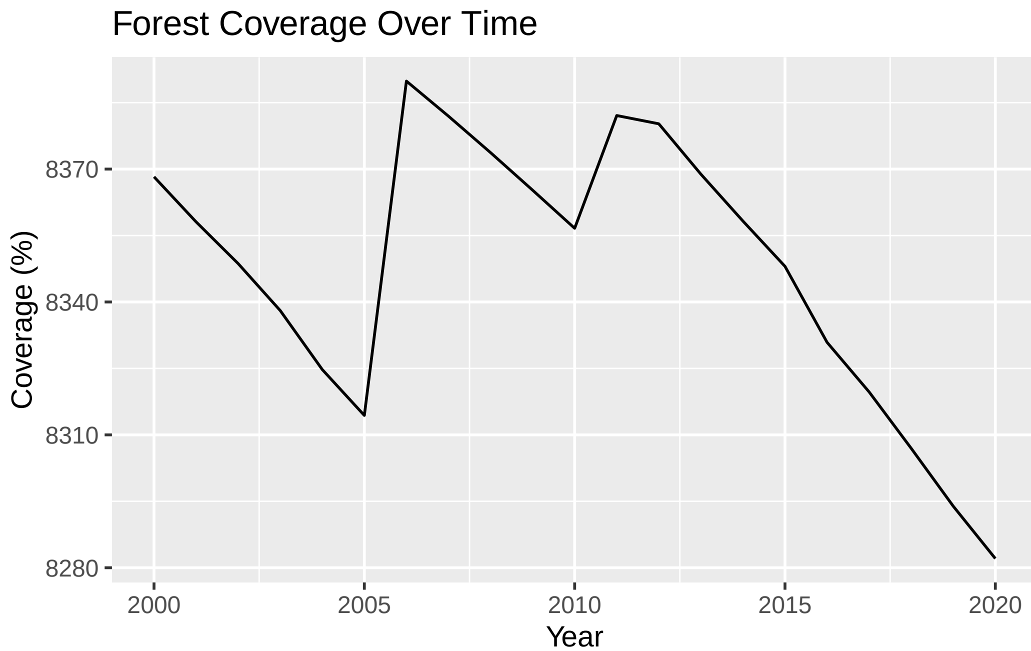

Figure 1: Forest Coverage Over Time

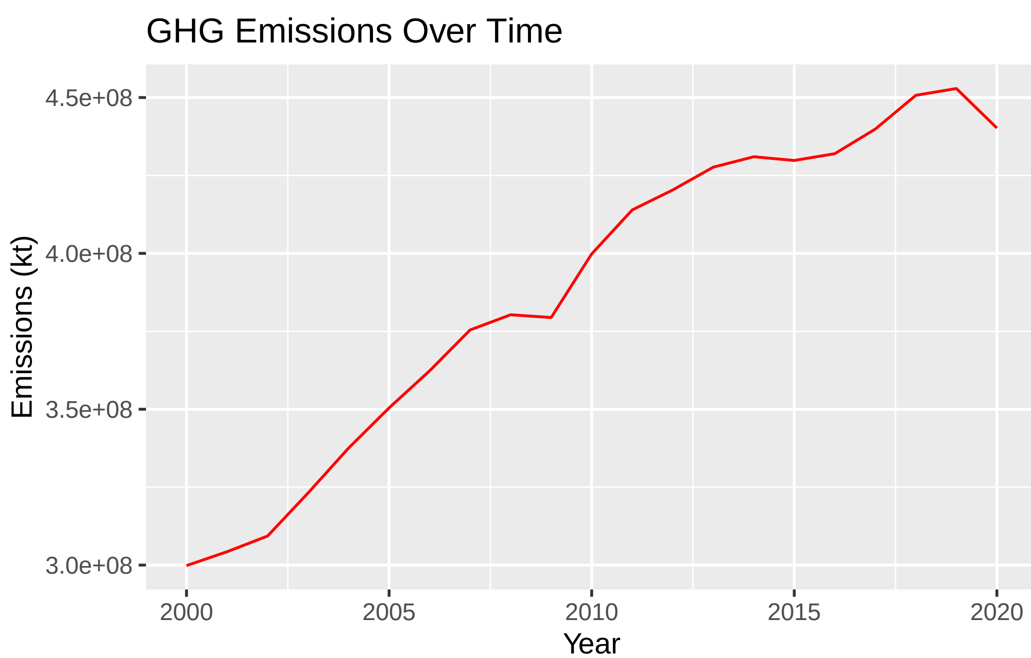

Figure 2: GHG Emissions Over Time

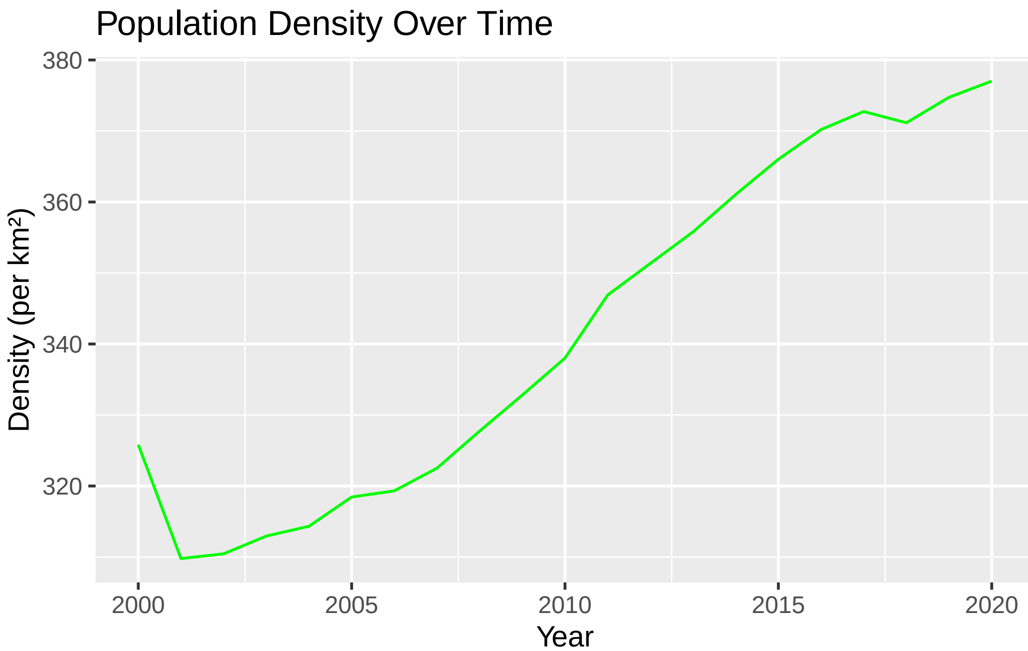

Figure 3: Population Density Over Time

3 Model

3.1 Model set up

We consider the following linear regression model to explore the relationship between temperature and CO2 emissions and forest cover and population density: In the proposed model:

$$
\begin{aligned}
T_i \mid \mu_i, \sigma^2 &\sim \text{Normal}(\mu_i, \sigma^2) \quad \text{(Conditional distribution of temperature changes)} \\
\mu_i &= \alpha + \beta_1 \cdot \text{CO2\_Emissions}_i + \beta_2 \cdot \text{Forest\_Coverage}_i + \beta_3 \cdot \text{Population\_Density}_i \quad \text{(Linear prediction} \\
\alpha &\sim \text{Normal}(0, \tau^2) \quad \text{(Prior distribution of the intercept)} \\
\beta_1, \beta_2, \beta_3 &\sim \text{Normal}(0, \tau^2) \quad \text{(Prior distributions of the slopes)} \\
\sigma^2 &\sim \text{Inverse-Gamma}(\alpha_0, \beta_0) \quad \text{(Prior distribution of the variance)}
\end{aligned}
$$

In this model:

- T_i represents the temperature change, assumed to be normally distributed around a mean μ_i determined by the predictors, with variance σ^2.
- μ_i is the expected temperature change given the levels of CO2 emissions, forest coverage, and population density.

- α, β_1, β_2, and β_3 are the coefficients representing the intercept and the effects of CO2 emissions, forest coverage, and population density, respectively.
- The independent variables of the model are:
  - CO2_Emissions$_i$: the CO2 emissions, expected to influence temperature change.
  - Forest_Coverage$_i$: the percentage of forest coverage, expected to influence temperature change.
  - Population_Density$_i$: the population density per square kilometer, expected to influence temperature change.
- The dependent variable of the model is the temperature change T_i.

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect that temperature will change linearly with changes in forest cover and population density and CO2. We assume that if the decrease of forest cover and the increase of population density and the increase of CO2 emissions lead to the increase of temperature. We then used known data on these variables to estimate the relationship between temperature (dependent variable) and forest cover and population density and CO2 emissions (independent variable). The decrease of forest cover leads to the increase of carbon dioxide, which leads to the increase of greenhouse gases, and finally to global warming. The density of population will consume energy and lead to a higher increase of carbon dioxide, so I think that carbon dioxide is a transitional value to temporarily replace the change of temperature, because the relationship between the three variables is reasonable. In this model, we focus on the key variables in the data and predict the increase or decrease of temperature emissions based on changes in forest cover and population density and CO2 emissions. This model is used to generate inferential statistics about CO2 emissions by filling in gaps in existing data to cover diversity in different regions of the globe with simulation data. The establishment of the model can help us analyze and predict the temperature change and play a qualitative role in the future forecast, and help some countries to implement environmentally friendly policies.

4 Results

```
Call:
lm(formula = temperature_change ~ co2_emissions + forest_coverage +
    population_density, data = climate_data)

Residuals:
```

	(1)
(Intercept)	0.217
	(0.676)
co2_emissions	0.049
	(0.001)
forest_coverage	−0.095
	(0.011)
population_density	0.009
	(0.003)
Num.Obs.	100
R2	0.952
R2 Adj.	0.951
AIC	161.1
BIC	174.1
Log.Lik.	−75.561
RMSE	0.52

```
     Min       1Q   Median       3Q      Max 
-1.24569 -0.32696  0.02832  0.33516  1.26605 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)         0.216952   0.676088   0.321  0.74899    
co2_emissions       0.049445   0.001169  42.306  < 2e-16 ***
forest_coverage    -0.095378   0.010946  -8.714 8.58e-14 ***
population_density  0.008565   0.002806   3.053  0.00293 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5258 on 96 degrees of freedom
Multiple R-squared:  0.9522,	Adjusted R-squared:  0.9507 
F-statistic:   637 on 3 and 96 DF,  p-value: < 2.2e-16
```

As shown in the model summary (see “Model Summary”), it shows the effect of co2_emissions, forest_coverage, and population density on temperature change of a response variable. It can be clearly observed that CO2_emissions and population density have a positive effect, which means that if CO2_emissions increases by one unit, the temperature change will increase to

0.049 units. The coefficient of forest_coverage is -0.095, which brings a negative correlation, which means that if the forest cover increases, the temperature predicted by the model will decrease, and if the forest cover area decreases, the temperature predicted by the model will increase.

Moreover, we can learn from the observation of R2 that the match degree between our current model and the observed model is very high, and it is very close to 1, which proves that his observation of the variable is accurate.

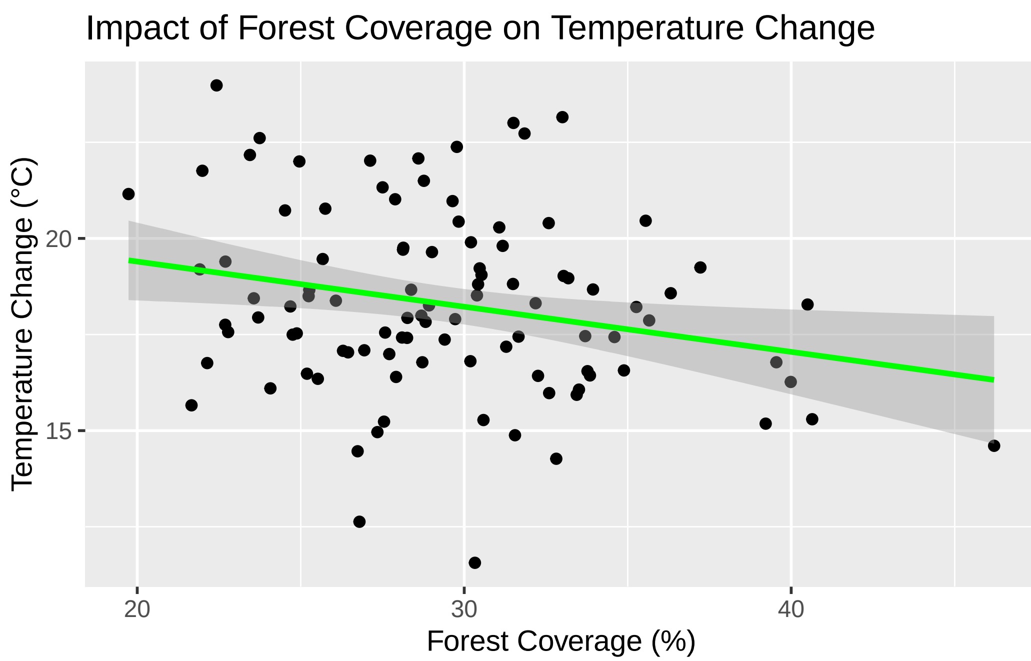

Figure 4: Impact of Forest Coverage on Temperature Change

These three graphs can make us more clearly understand the relationship between temperature and population density Figure 6, forest coverage rate Figure 4 and CO2@fig-co2-impact emission. For in-depth analysis of the model summary (see "Model Summary"), we can learn from Figure 4 that the negative slope of the green regression line can be seen. The slope of this line represents the average decrease in temperature change for every percentage point increase in forest cover. The confidence interval is shown as a grey area, indicating the uncertainty of our regression estimate. Although the regression line indicates an overall trend, the distribution of the scatter points also suggests that the relationship between temperature change and forest cover may be influenced by other unobserved variables.

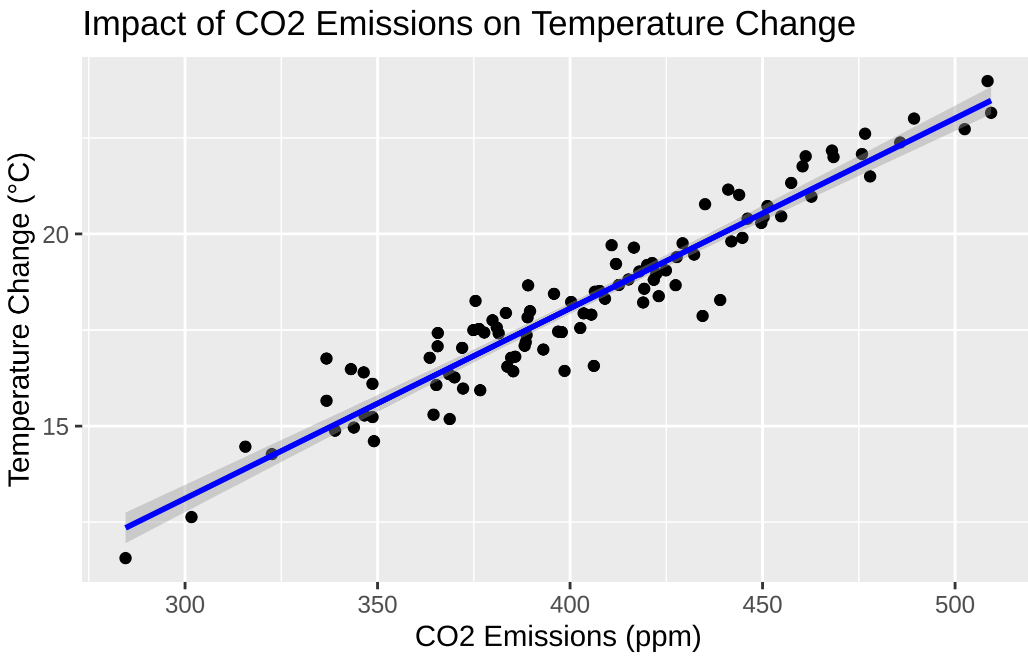

Figure 5: Impact of CO2 Emissions on Temperature Change

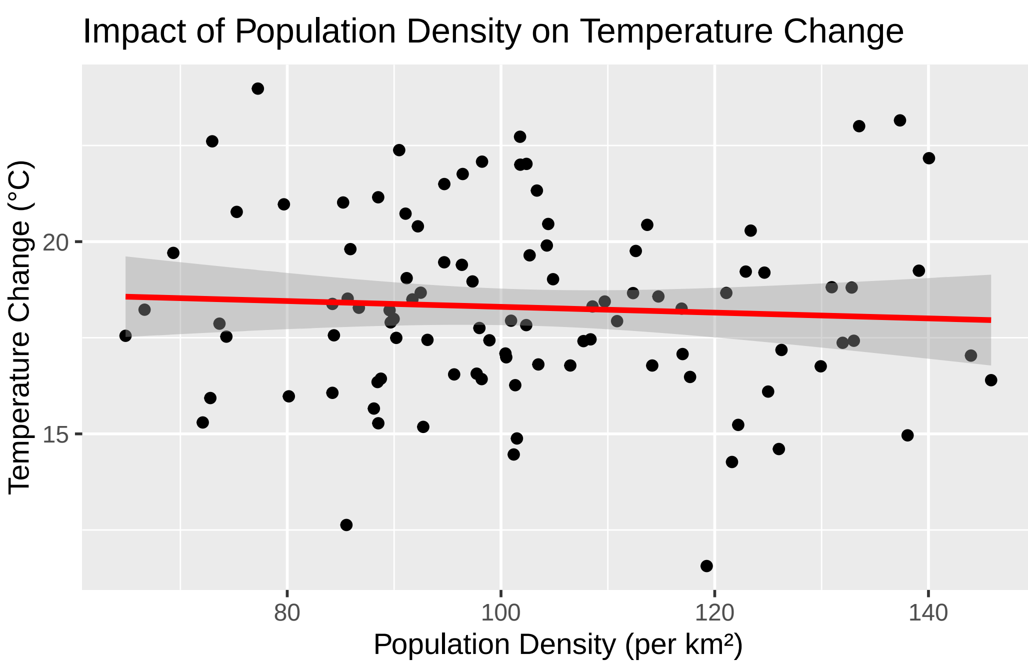

Figure 6: Impact of Population Density on Temperature Change

For the CO2@fig-co2-impact presentation, the blue areas about the blue line indicate how confident we are in our estimate of the regression line; The narrower the interval, the higher our confidence in the estimate. From this graph we can observe that the data points are roughly distributed around the regression line, indicating that CO2 emissions are an important predictor of temperature change. In addition, the scattered points are closely distributed around the regression line, indicating a good fit between the model and the data. This is also consistent with the high R2 value in the model summary provided earlier.

The red line is the trend line obtained by linear regression analysis, which shows the relationship between population density and temperature change. As can be seen from the figure, with the increase of population density, the temperature change also has a slight increasing trend. The gray area represents the 95% confidence interval, which shows that if we repeat the study, there is a 95% chance that the regression line will fall within this interval. The width of the confidence interval indicates the uncertainty of the forecast: the wider the interval, the greater the uncertainty of our estimate of the trend line.

5 Discussion

5.1 Forest cover and population density as well as carbon dioxide emissions affect temperature changes and thus climate change

We can learn from the above about the correlation between forest cover and population density as well as carbon dioxide emission and temperature. Then, we can start from these three factors and analyze the data. We can easily see that the change of the temperature is caused by the change of the three variables. And in this carbon dioxide emissions produce a very important factor.

5.2 The significance of this study

The significance of our research is to protect the environment, especially in the current environment is so bad, to understand some factors that will affect climate change and protect the diversity and integrity of the ecology. Understanding the climate change has not only biological impacts but also economic impacts, hitting industries that depend on nature such as agriculture and water resources. Many farmers suffer from it and suffer a lot of economic losses. So understanding climate change is a very important factor.

5.3 Weaknesses and next steps

5.3.1 Weaknesses

In fact, I only intercepted a part of the factors, and did not analyze all the factors of climate change. There are still many variables that affect climate change or are related to temperature and the three variables. In fact, in the beginning, the csv I downloaded did not show the annual sum values, which I downloaded from the website. The data during this period is correct, but I only have 21 variables per dataset, so I think the sample may be too small.

5.3.2 Next Steps

I need to increase the number of samples and download them completely to simulate and clarify the data. In this process, I think we can explore new topics and research based on my existing articles. Second, I think we can use more models to test the correctness of data than just linear models. We can learn more about the factors that will affect climate change and implement collection and prediction.

references

Arel-Bundock, Vincent. 2021a. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://CRAN.R-project.org/package=modelsummary.

———. 2021b. *WDI: World Development Indicators (World Bank)*. https://CRAN.R-project.org/package=WDI.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

Müller, Kirill. 2021. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley et al. 2021a. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley. 2021a. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley et al. 2021b. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Wickham, Hadley. 2021b. *Testthat: Unit Testing for r*. https://CRAN.R-project.org/package=testthat.

Wickham, Hadley et al. 2021c. *Tidyverse: Easily Install and Load the Tidyverse*. https://CRAN.R-project.org/package=tidyverse.

World Bank. 2020a. "Forest Area (% of Land Area)." https://data.worldbank.org/indicator/AG.LND.FRST.ZS?end=2020&start=2000&view=chart.

———. 2020b. "Population Density (People Per Sq. Km of Land Area)." https://data.worldbank.org/indicator/EN.POP.DNST?end=2020&start=2000.

———. 2020c. "Total Greenhouse Gas Emissions (Kt of CO2 Equivalent)." https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE?end=2020&start=2000.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.