Habib University

Data Science

Project Report

Exploring Carbon/Ecological Footprint and Bio-capacity Trends

Author:
Bilal Mohiuddin
(bm03751)

Author:
Rizwan Niaz
(rk03548)

Author:
Ajlal Bawani (bm03751)

Plagiarism Declaration: *I confirm that this assignment is my own work, is not copied from any other person's work (published or unpublished), and has not previously submitted for assessment either at Habib University or elsewhere.*

December 3, 2020

Contents

1 Introduction

Using data from The National Footprint Accounts, this project will attempt to visualize trends in the ecological resource use and resource capacity of nations from 1961 to 2014. The data-set provides data in global hectares, as well as per capita. A breakdown of ecological footprint as well as bio-capacity is provided for all countries included. The Ecological Footprint for a country is a measure of how much area of biologically productive land and water an individual or activity requires to produce all the resources it consumes. The bio-capacity, on the other and, measures the ecosystem's capacity to produce biological materials used by people. Carbon footprint, in ecological footprint terms, is measured by the biologically productive area necessary for absorbing the CO_2 produced[1]. With out project, we hope to explore the Ecological Footprint breakdown for different countries and observe how the distribution of the variables included in the dataset has changed for countries over time.

2 Questions to be Investigated

2.1 How Have Carbon Footprint trends Changed over the Time Period

One of the key trends we wanted to explore was how Carbon emissions have changed over time for different nations. The introduction of machinery powered through fossil fuels has lead to an increase in CO_2 production over time. We wanted to observe how Carbon Footprint levels have changed since 1961 and how changes in those levels themselves have varied.

2.2 How has the share of the global carbon footprint increased over time

By answering this questions we aimed to find out which countries contributed the most to CO_2 production in the past and which countries do so now. This would help identify countries most responsible for current climate change problems and how CO_2 production in each of these countries compares to the other, and to itself in the past.

2.3 How Has the Distribution of Land and Water Bodies changed Over Time

This question aims to see how the distribution of land between the various different ecological classifications has changed over time. It should come as no surprise that the amount of land left to forests has diminished over time, but interestingly enough the area used for crop lands and grazing lands has also changed significantly. By answering this question we aim to see how the change in total area for all of these classifications has changed over time, in order to get a deeper understanding of how urbanization has effected them.

2.4 How are land measurements related to Ecological footprint and each other

The Ecological Footprint for a country in a specific year is based on a number of factors. Some of these being land available for farming, as well as land being used for urbanisation. With our analysis we aim to find out which of these factors contribute to Ecological Footprint, and how much they do so. We will also explore how each of these variables are related to one another.

2.5 Trends in Bio-Capacity and Ecological Footprint

Since our data-set provides information for the Bio-Capacity of a country as well, we will utilise this data by observing how Bio-capacity has been spread out between different countries over the time period covered by our dataset. This will allow us to gauge the sustainability of our current activities through comparison with Ecological Footprint data.

3 Analysis/Results

3.1 Reliability of Data

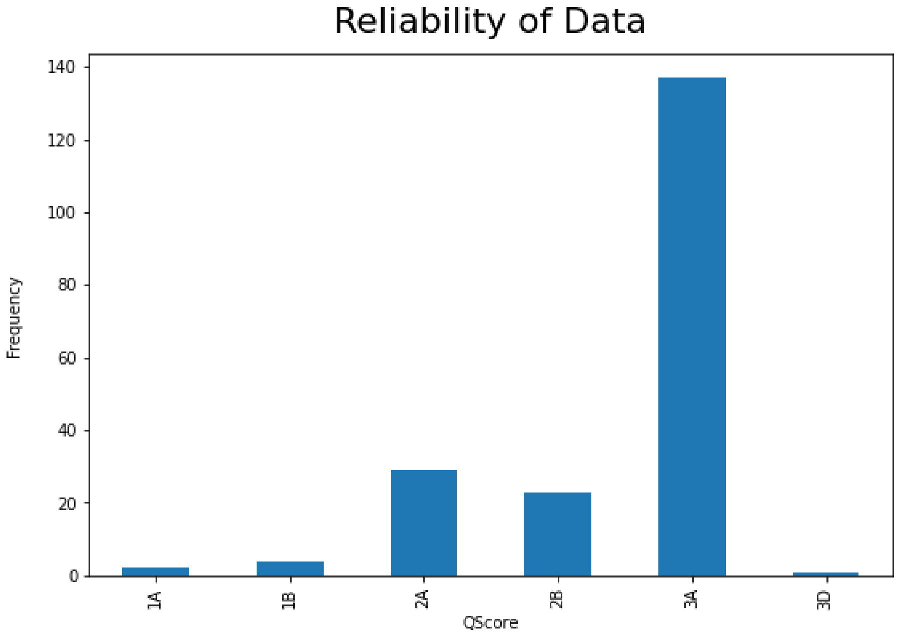

Figure 1: Reliability of Data

Each country in the National Footprint and Bio-capacity Accounts is given a quality score comprised of two elements, time series score [1-3] and latest year score [A-D][2]. Since the National Footprint and Bio-Capacity accounts use data from multiple data-sets for a very large time period, the data is bound to contain some errors and not always be reliable. Hence, in this dataset, each row is assigned a "QScore" to express the reliability of data recorded.

Using the bar chart drawn above,we can conclude that most of the data included in the dataset is reliable since it falls into the *3A* QScore category. There is also a large chunk of *2A* and *2B* datasets available. Such data has generally unreliable values for Ecological Footprint and Bio-Consumption and hence, we have dealt with them accordingly during the pre-processing stage of our analysis.

3.2 Carbon Footprint

3.2.1 World Carbon Footprint Trends

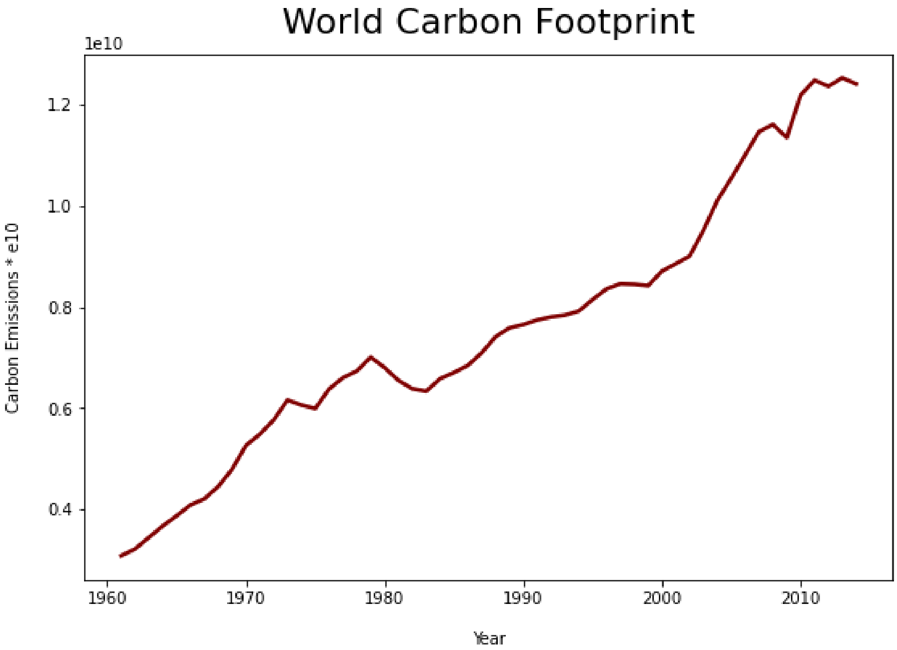

Figure 2: Carbon Footprint Trends Throughout the Years

Figure 2 shows how Carbon Footprint levels have changed for the world from 1969 till 2014. While there appears to be a mostly linear increase at first, there are periods of sudden decrease around the late 1970's and early 1980's. Levels have consistently increase since then,although at a noticeably smaller rate. There is a slight boom in Carbon levels in the early 2000's which could be explained by an increase in infrastructural facilities as technological advancements were gradually made. More recently however, Carbon Footprint levels appear to have stagnated. This may be due to more developed countries having reached a stable financial state where they can afford to switch over to non-renewable resources for generating power. The introduction of Eco-friendly transport facilities, as well as treaties drawn up to reduce greenhouse gas emissions, may also be a contributing factor.

3.2.2 A Comparison - 1970 & 2014

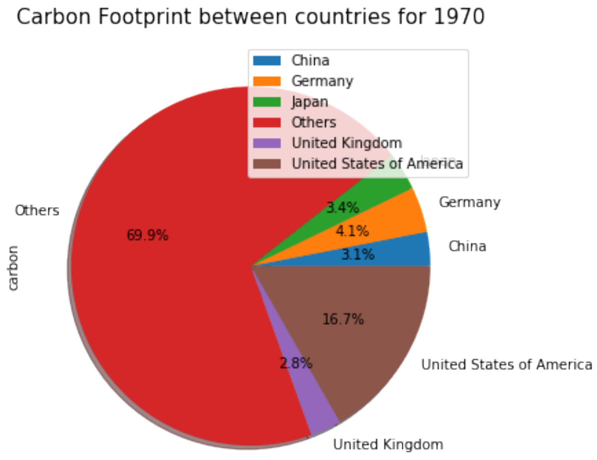

Figure 3: Top 5 countries for Carbon Footprint - 1970

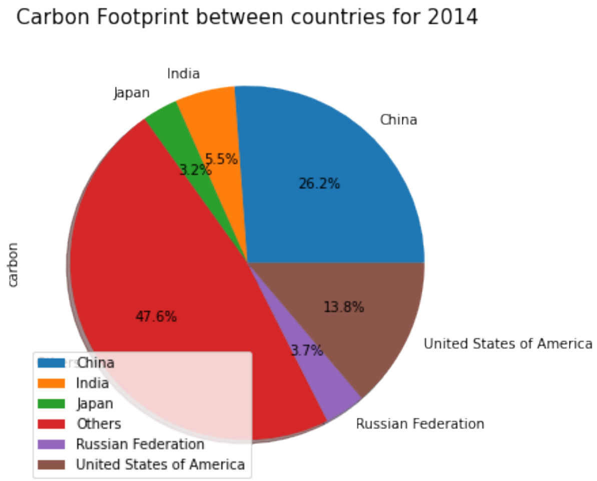

Figure 4: Top 5 countries for Carbon Footprint - 2014

The pie charts drawn above show the top 5 countries that produced the most Carbon Footprint in 1970 and 2014. As can be seen above, the US has the highest footprint in 1970, while China has the highest one in 2014. However, the US still produces about 13.8% of total carbon emission in 2014. India, due to its large population and increasing technological improvements, is one of the top 5 countries in 2014, with about 5.5% of the total carbon footprint. Another interesting detail to note her is that the value for *Other* countries has fallen off from 69.9% in 1970, to 47.6% in 2014. This indicates that the leading countries are now using a much larger share of available carbon resources than before.

3.3 Ecological Footprint and BioCapacity

3.3.1 Total Ecological Footprint and BioCapacity

Since Ecological footprint is a measure of how much land is necessary is necessary for a country to produce all the resources it requires for its own consumption and Biocapacity is a measure of how much productive land the country has available to it, we can see that the two values are linked.

Ecological deficit is the difference between a countries BioCapacity and its Ecological footprint. It is a measure of how much more land the country needs in order to make up for the effect it has had on the environment.

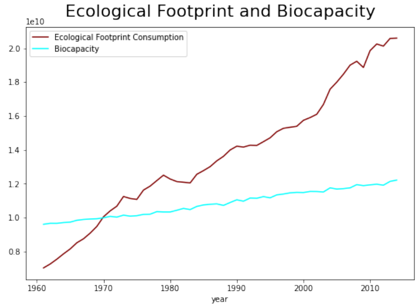

Figure 5: Ecological footprint and Biocapacity

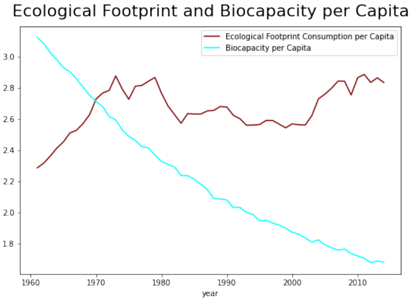

Figure 6: Ecological footprint and Biocapacity per capita

Above we can see the graphs of Ecological Footprint and Bio-Capacity of the entire world along side each other. The Ecological footprint of the entire world can be thought of as the amount of land needed to undo the effect human consumption has on the planet. While Bio-Capacity would be the total amount of productive land available in the entire world.

Looking at figure 5, we can see that before 1970, the total Ecological deficit of the world was negative, and after 1970 the Ecological deficit has steadily increased. This means that after 1970, the total amount of land in the earth was no longer enough to sustain it indefinitely.

Figure 6 is the Bio-Capacity and Ecological footprint per Capita. Ecological footprint per Capita is the amount of land needed to undo the effect a single person has on the environment, while Bio-Capacity per Capita is the amount of land that is actually available for a single person. We can see once again that before 1970, the Bio-Capacity per Capita was greater then the Ecological footprint per Capita.

An interesting thing to note about this graph is that the Ecological footprint per Capita has not increased much since 1970, this means that the average amount of land needed for consumption by a single person has not changed, and therefore the increase in the total ecological footprint is due to increasing population. In the same vein, Bio-capacity per Capita has decreased even though the total Bio-capacity of the world has seen a steady increase, meaning the decrease in the per Capita measurement is due to a large increase in population since the 1970's

3.3.2 Distribution of Ecological Footprint and Bio-Capacity

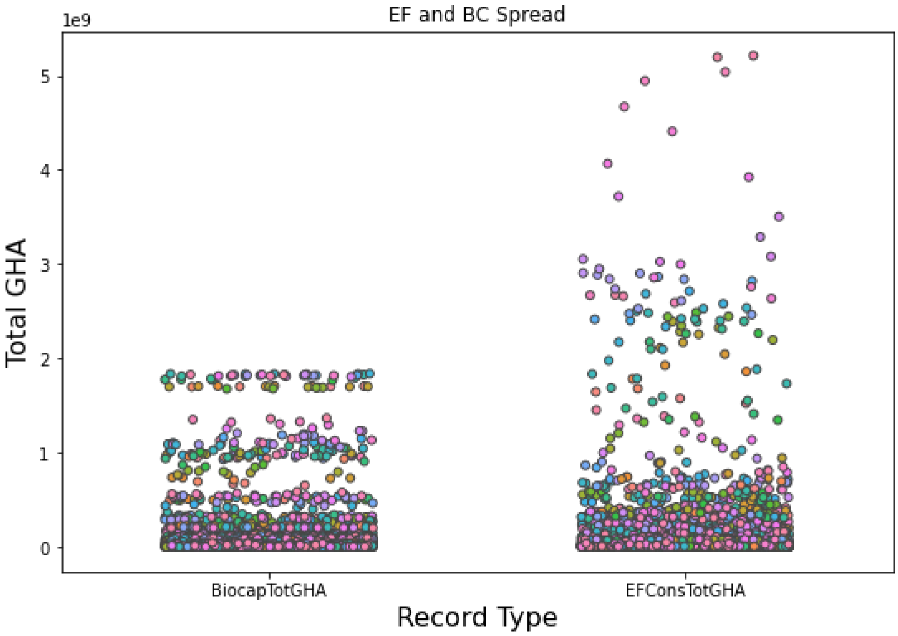

Figure 7: Distribution of EF and BC

The plot drawn above shows values for Ecological Footprint and Bio-Capacity for different countries over the time period of our dataset. Pink colored dots are used for most recent years while orange ones are used for years near 1970. The plot supports the line graph drawn above about how values for both, Ecological Footprint and Bio-Capacity, has increased over the years. However, the increase in Ecological Footprint has been far greater than that for Bio-Capacity, which is indicative of the rate at which we are using up resources to generate power. Values for Bio-Capacity are a lot more concentrated towards the bottom as compared to Ecological Footprint, whose values seem to be slightly more spread out towards higher levels.

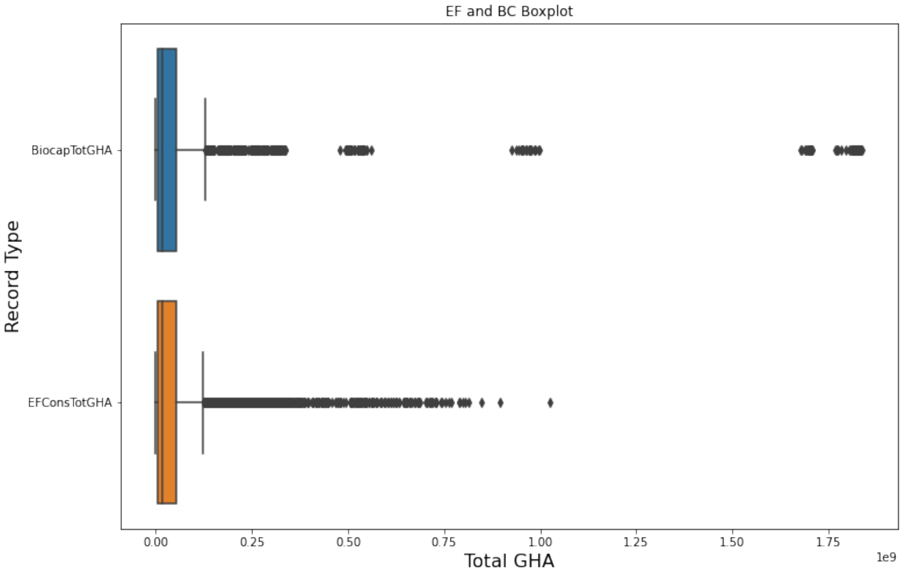

Figure 8: Boxplot for EF and BC

The above statements are supported by the box-plot drawn above. An important thing to note about this plot though, is that countries with unusually high Ecological Footprint values, such as China and the US, have been deliberately omitted. As a result, we can observe that the Bio-Capacity provided by countries with lower Ecological Footprint values is actually higher in some cases than their Ecological Footprint.

3.3.3 Distribution of Bio-Capacity over countries

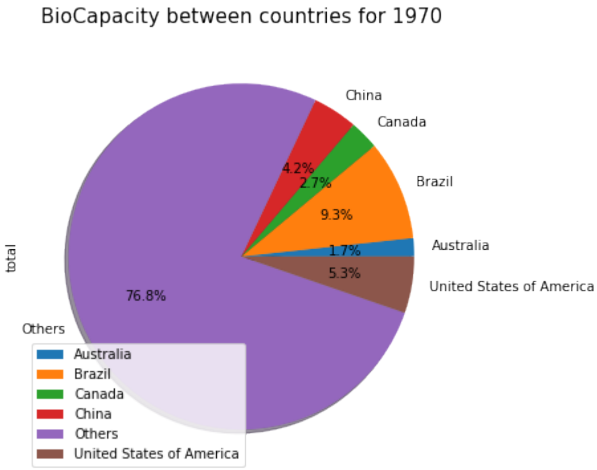

Figure 9: Top 5 countries for Bio-Capacity - 1970

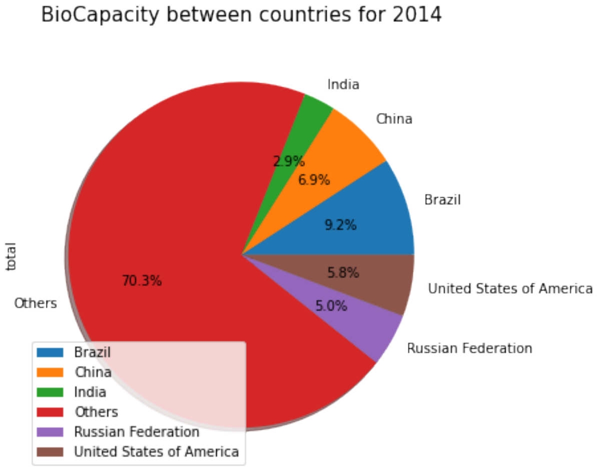

Figure 10: Top 5 countries for Bio-Capacity - 2014

The pie charts drawn above show the top 5 countries that have had the most Bio-Capacity in 1970 and in 2014. It is interesting to note that Brazil has remained quite consistent in terms of the Bio-Capacity being produced. China has increased its percentage, however, its total Ecological Footprint has increased as well. There has only been a slight increase of 6% between the total Bio-Capacity resources for all of these top 5 countries compared to the rest of the world.

3.4 Division of Total Land and Their Correlations

Over time, the distribution of the total area of land between the various land classifications has changed. The following heatmaps illustrate the correlations between the sizes of the various land classifications.

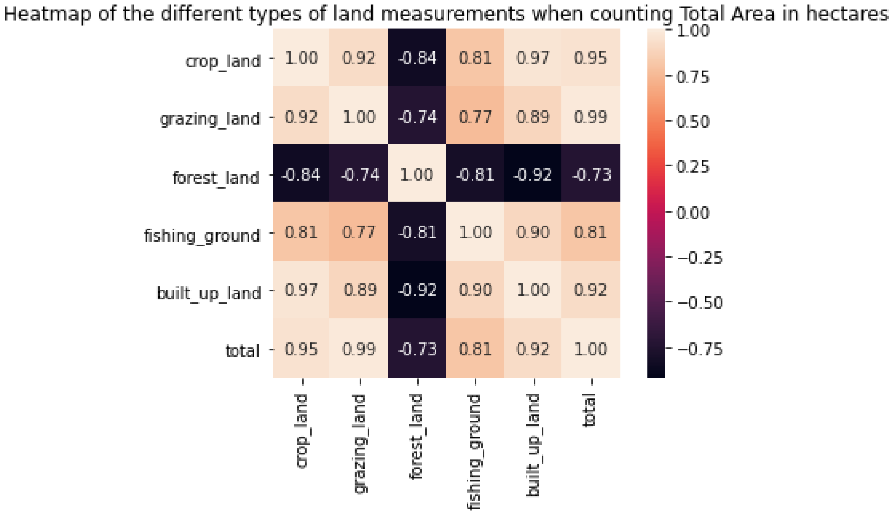

Figure 11: Heatmap of the correlations between the sizes of all land classifications

Figure 9 shows us that all of the various land classifications have positive correlations, except for Forest land. This makes sense as the other classifications often need to have large tracts of forested land cleared away in order to make way for their construction. Another interesting observation is that all of the other classifications except forests have positive correlations with the total area of land available. This means that as time goes by the total area of land keeps increasing along with the other classifications, *except forests*

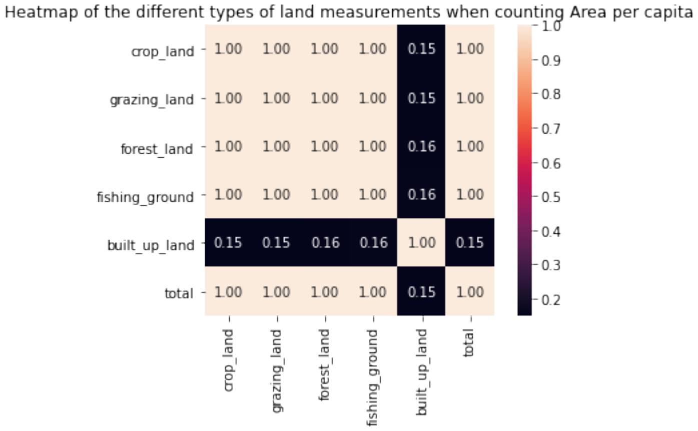

Figure 12: Heatmap showing Area per Capita for the sizes of the various land classifications

The above heatmap shows that the Area per Capita of of all the land classifications except Built up Land have a perfect 1 to 1 correlation with each other. Built up Land is the land needed to house buildings and other urban structures. The following graph gives us a closer look at the trends of these quantities.

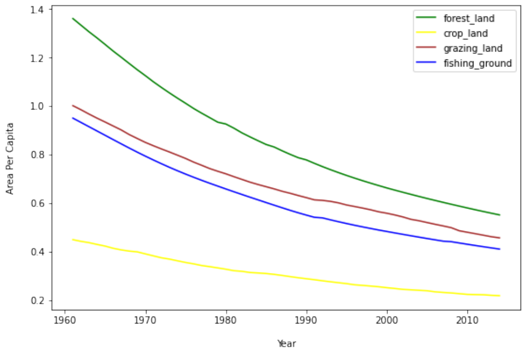

Figure 13: All areas

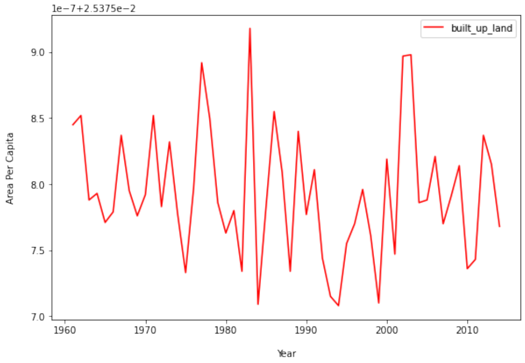

Figure 14: Built Up Land

From the two figures above, we can see very clearly that the Area Per Capita of all of the quantities except for Built Up Land have seen a steady decrease. Buil Up Land on the other hand shown a lot of variance over the years.

We shall now look at the change in total area to understand these trends even better

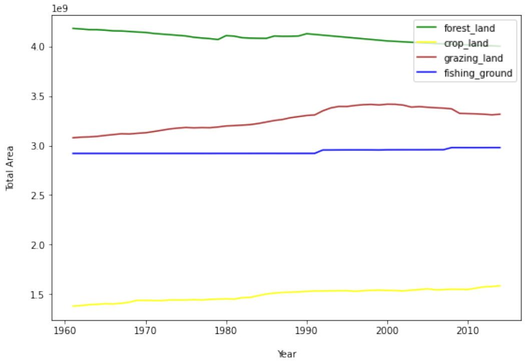

Figure 15: All areas in hectares

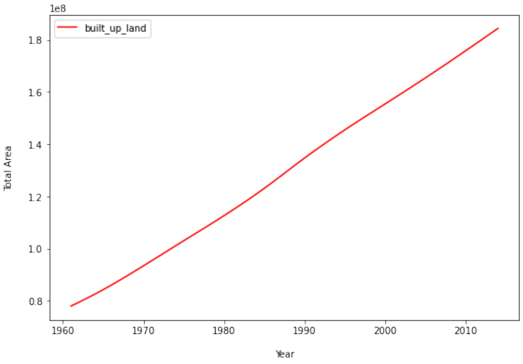

Figure 16: Built Up Land in hectares

The above two graphs show an interesting trend, Grazing Land, Crop Land, and Fishing Ground have shown a slight increase in total area over the years but have remained largely unchanged, while Forest Lands have shown a steady decrease, this lines up with our results in Figure 9. Figure 14 is also interesting, as it shows that Built up land is the only category that has seen a completely linear increase over time. This means that the fluctuations in the Area of Built up Land per Capita seen in Figure 12 is most likely caused by a rapid change in the total world population during this time span. It should also be noted that Built Up Land is one order of magnitude smaller then the other categories, so its per Capita measurement is more likely to be effected by large changes in population

3.5 Comparing Ecological Footprint of consumption and Production

Ecological footprint of Consumption of a country is the amount of land necessary to maintain the amount of resources *consumed* by that country, Ecological Footprint of Production is the amount of land necessary to maintain the amount of resources *produced by that country to be consumed by other countries*

This is an important distinction to make, as some less developed countries might have a large ecological footprint because they produce resources that are consumed by people in more developing countries; If a small developing country uses a large chunk of its land in order for the resources to be used by more developed first world countries, then the Ecological Footprint will be counted in the Ecological Footprint of Consumption of the developed country, and in the Ecological Footprint of Production of the developing country.

The following heatmap shows the correlation between Ecological footprint of Consumption of the various land categories and Carbon footprint per Capita for Pakistan

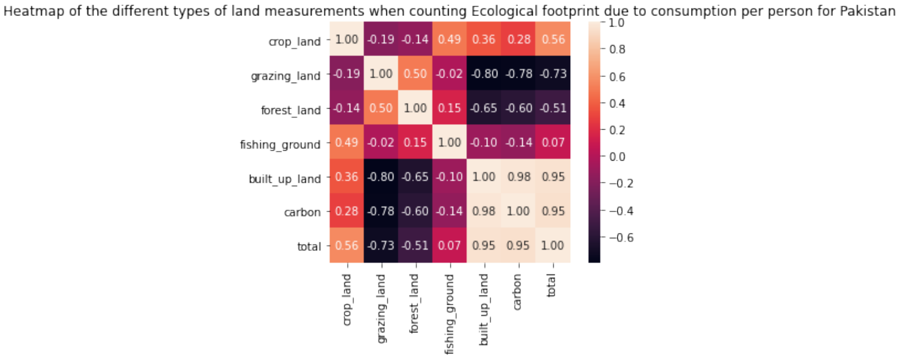

Figure 17: Heatmap of the correlations between Ecological footprint of consumption for the land categories and Carbon emissions in Pakistan

A few things to note from the above heatmap is that in the case of Pakistan, the land categories Crop, grazing and Forest lands seem to be competing with each other. Another thing to note would be that Built up land has an almost perfect 1 to 1 correlation with the carbon footprint.

Now let us look at the Ecological footprint of production per capita

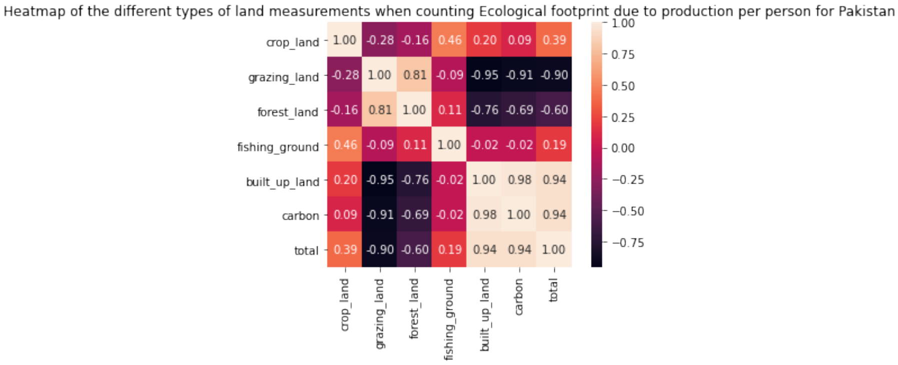

Figure 18: Heatmap of the correlations between Ecological footprint of Production for the land categories and Carbon emissions in Pakistan

As we can see, Figures 15 and 16 are quite similar, meaning that on average a citizen of Pakistan consumes resources from other countries in about the same way as they produce for others.

for example, Pakistani's consumption of Crop Land has a negative correlation with their consumption of Forest Land, i.e If they use more Crop resources the less they use Forest resources, at the same time, if they produce more crop resources for other countries the less they produce forest resources.

Now let us look at the ecological footprints of consumption and production for the entire world.

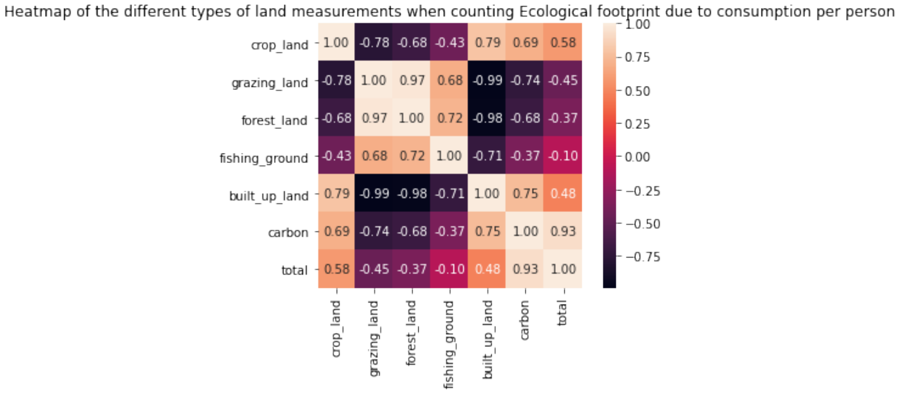

Figure 19: Heatmap of the correlations between Ecological footprint of consumption for the entire world

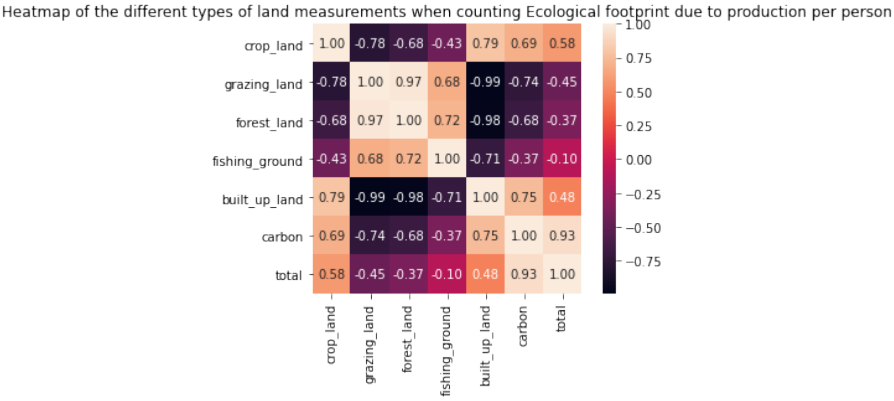

Figure 20: Heatmap of the correlations between Ecological footprint of production for the entire world

As we can see, the above two heatmaps are completely identical, This makes sense, as when looked at as a whole, the world produces the same amount of resources as it consumes.

4 Model Fitting

4.1 ARIMA Forecast Model

Our dataset is a time series dataset with a small number of datapoints for the record type of each country (At most 51 datapoints). After researching online, It was determined that the ARIMA forecast model might be appropriate.

In order to determine the amount of lag necessary for the model Auto-correlation graphs were made on the Carbon footprint variable with the Ecological Footprint of consumption record type.

The following graphs were obtained:

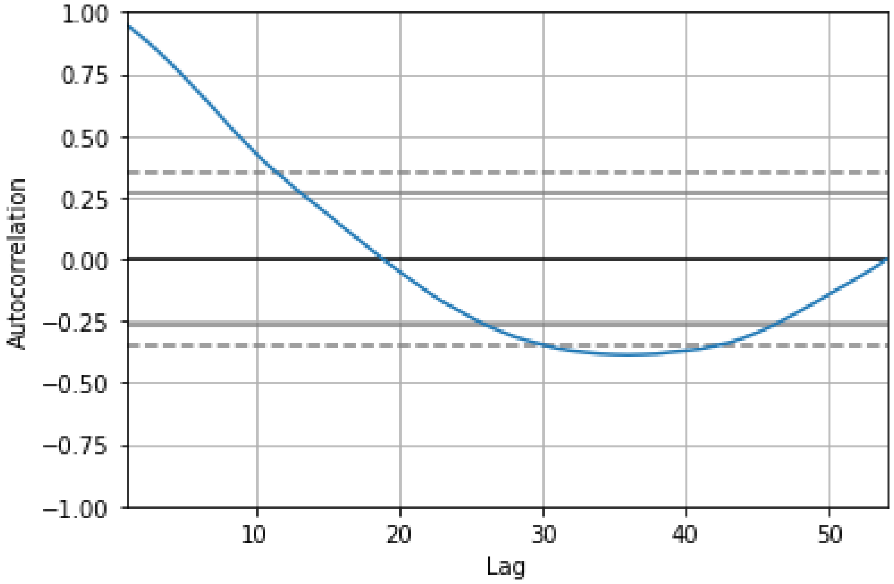

Figure 21: Autocorrelation for Pakistan

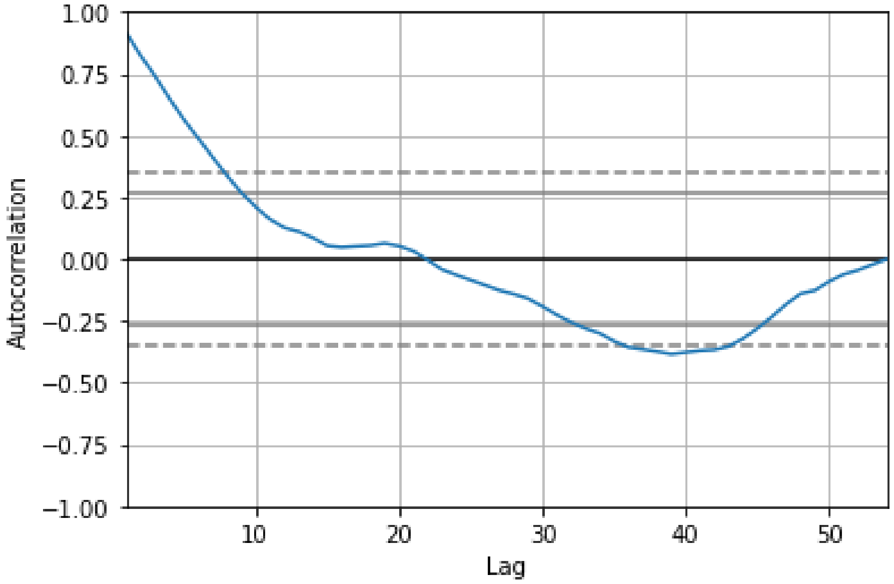

Figure 22: Autocorrelation for the US

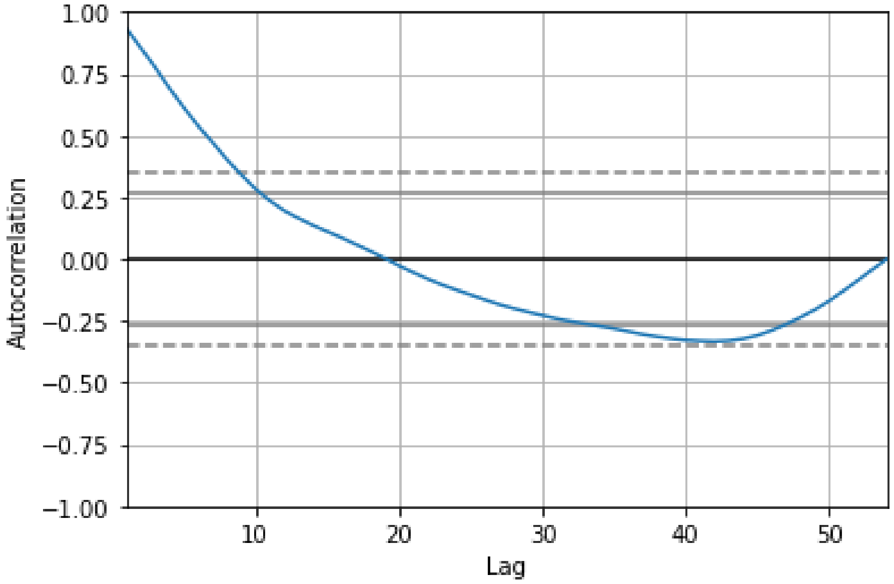

Figure 23: Autocorrelation for China

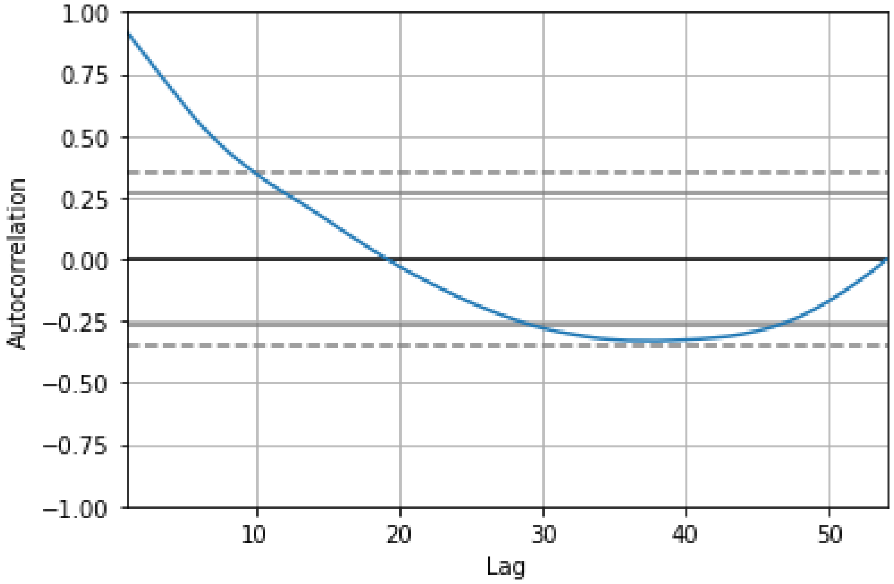

Figure 24: Autocorrelation for India

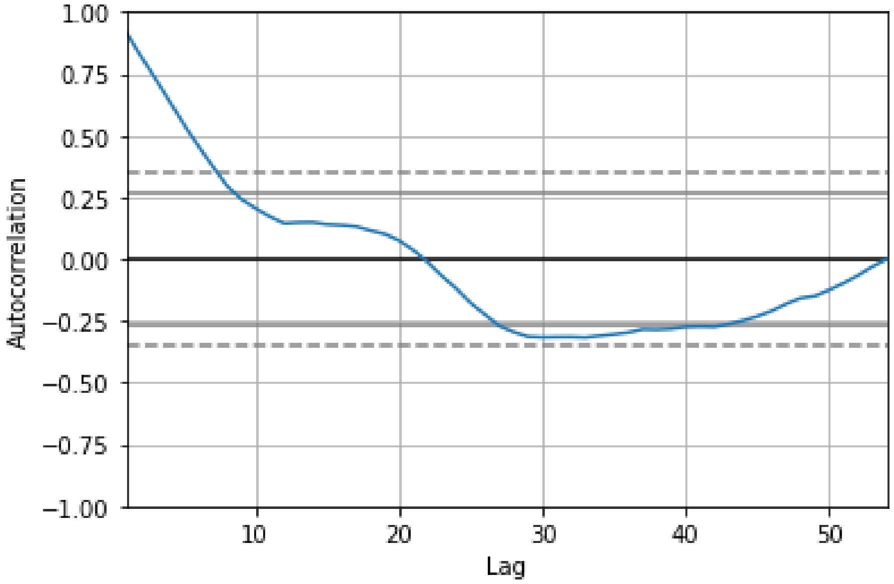

Figure 25: Autocorrelation for Japan

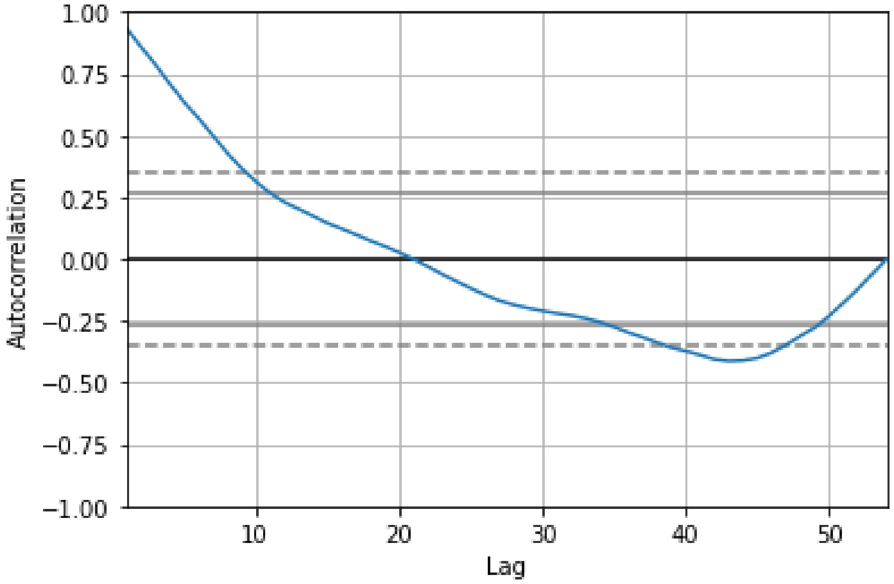

Figure 26: Autocorrelation for the entire world

using these the value of p for the ARIMA models for each of the chosen countries was determined and a model was trained on the data from the years 1961-1995.

This was then tested on the data from 1996-2014. The values of d and q were set to 1 and 0

The following graphs, along with the Mean Squared Error, were obtained.

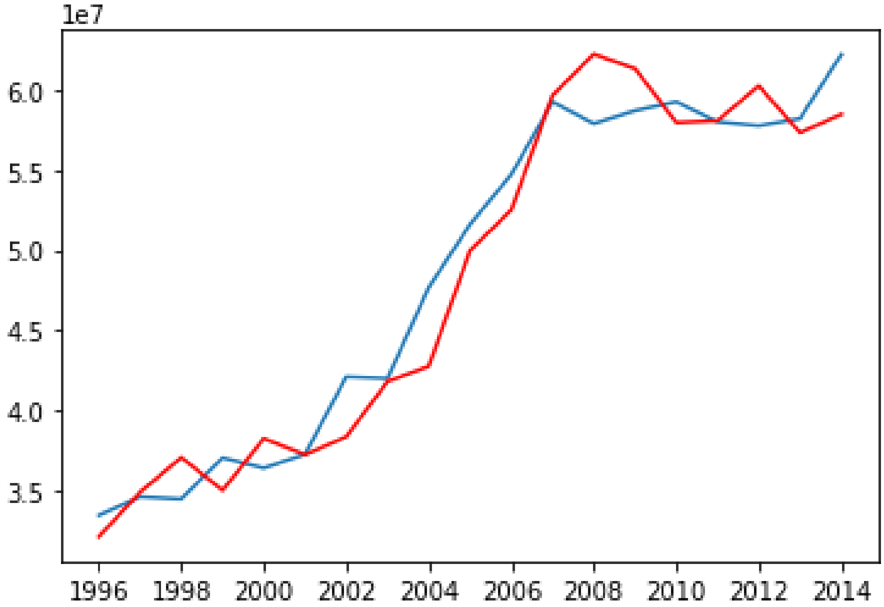

Figure 27: model test results for Pakistan. MSE = 5797294420973.212

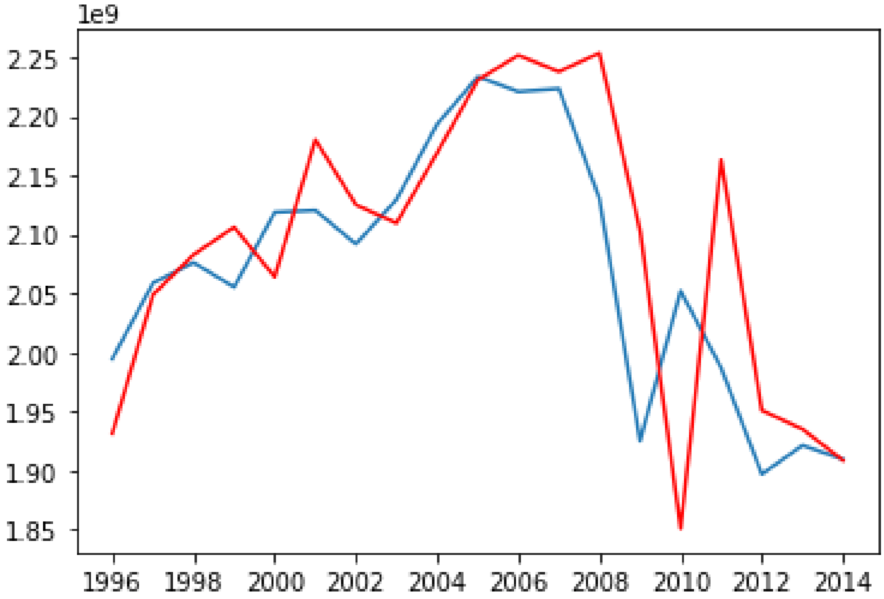

Figure 28: model test results for the US. MSE = 7276197595155971.000

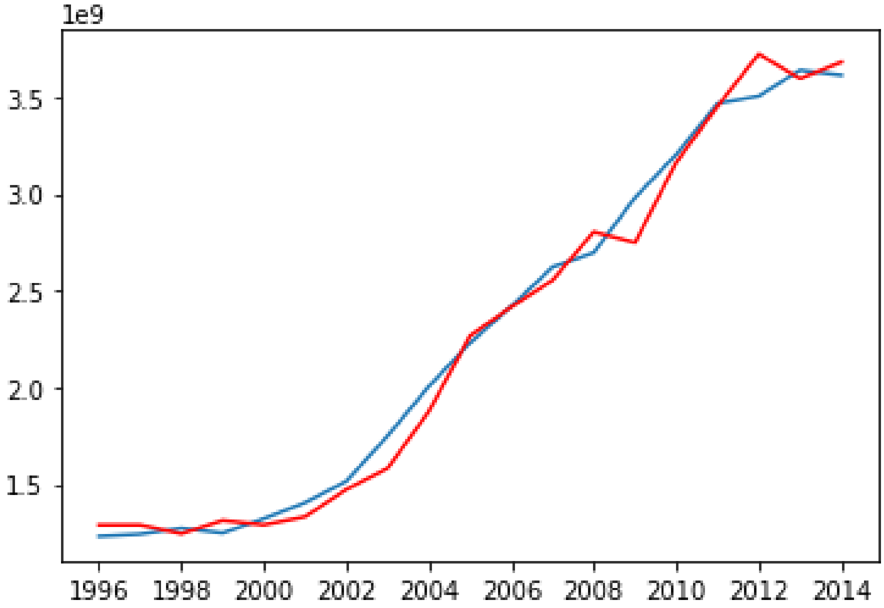

Figure 29: model test results for China. MSE = 10074353800781176.000

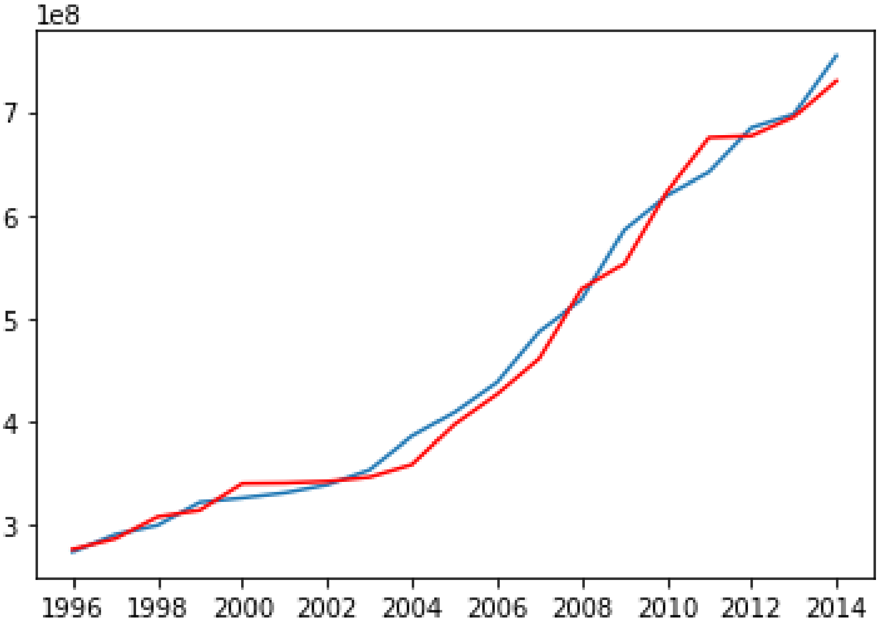

Figure 30: model test results for India. MSE = 273209806743620.094

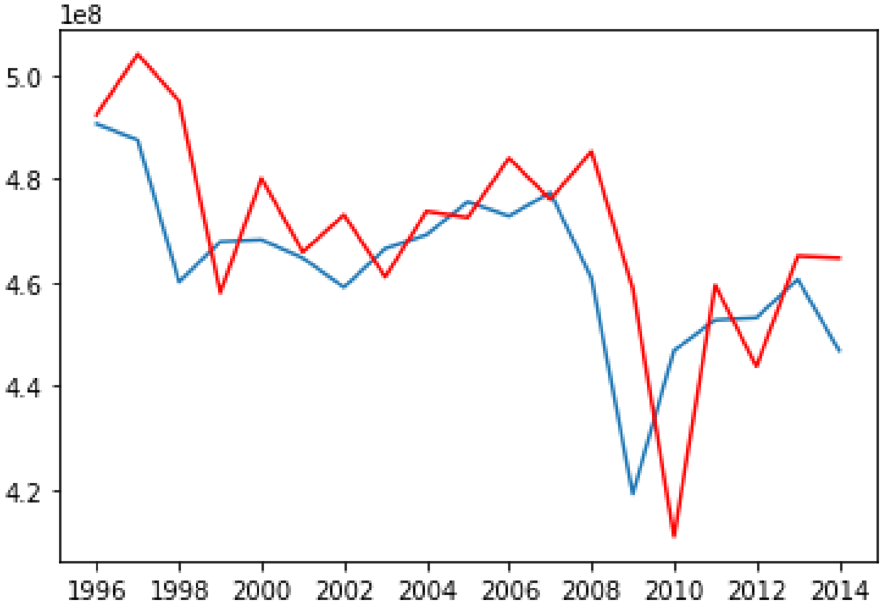

Figure 31: model test results for Japan. MSE = 317717437912347.062

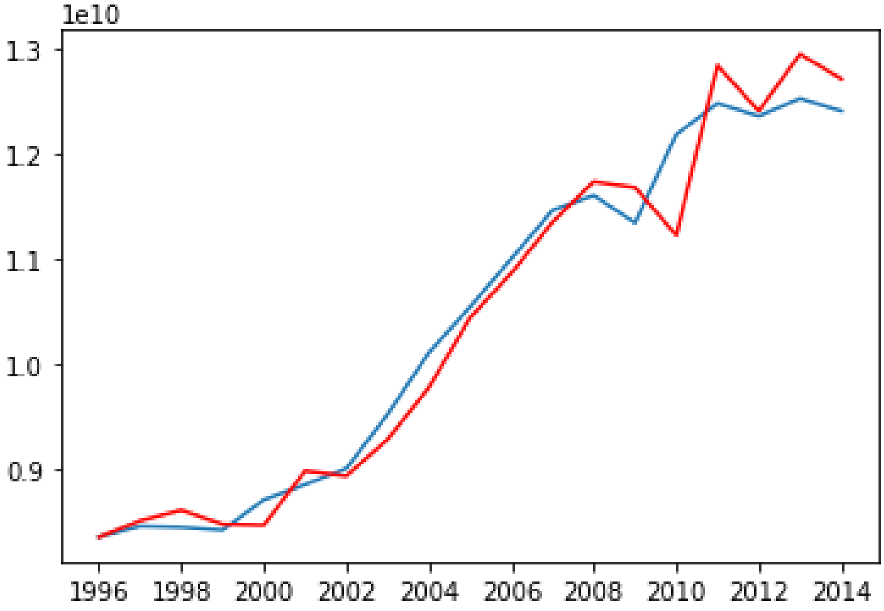

Figure 32: model test results for the world. MSE = 93097287582145408.000

As can be seen from the graphs above, the models fits quite well onto the carbon footprint for China, India, and the World. This indicates that carbon footprint trends for each of these have largely followed the trend and remained predictable over the time period of the dataset. There are slight variations for actual values for Pakistan, The US, and Japan. However, the model does accurately follow the trend by actual values for these countries, although values may have been shifted to appear at later years.

4.2 Time Series Split - KNN & RF

Along with using the ARIMA model, we also went a different route and tried to use a time series split to train different models onto our dataset. In order to make sure the values of Carbon Footprint predicted by our model were affected the greatest by the carbon footprint of recent years, additional columns for carbon footprint from 2 years before were added to each cell. Additionally, columns for change in carbon footprint values were also added. In doing so, we ensured that carbon footprint levels were affected by more relevant years.

Our algorithm consisted of applying a time series split to our data to make sure the model is trained a wide combination of years. The models used were Random Forest Regressor and K Neighbors Regressor. After applying the time series split it was found that the K Neighbors Regressor Model worked better. The Mean and Standard Deviation of both models was as follows:

- K Neighbors Regressor
  - **Mean:** 0.833240
  - **Standard Deviation:** 0.183631
- Random Forest Regressor
  - **Mean:** 0.646453
  - **Standard Deviation:** 0.717099

In order to improve the performance of our algorithm, we performed grid searching to improve our hyper-parameters. Once this was done, we carried out feature selection to find out which variables were the most important in our model. The following graph was obtained for feature importance.

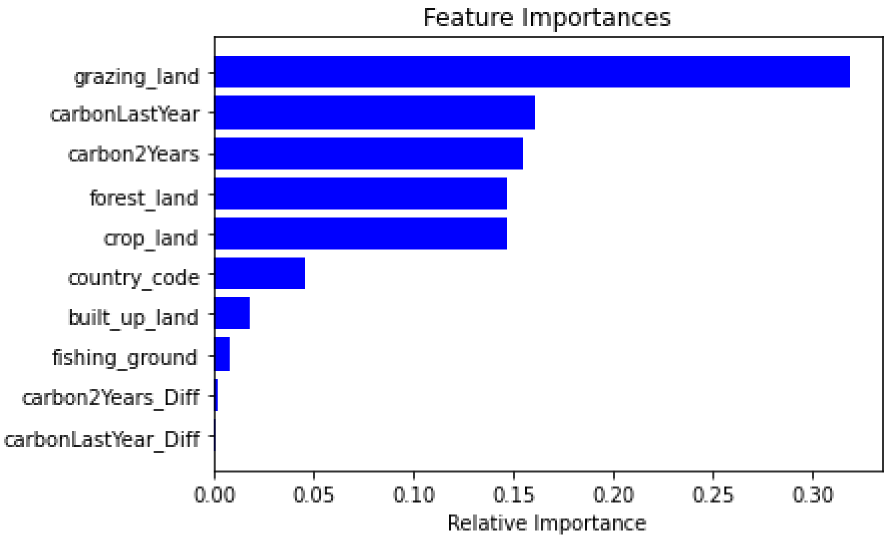

Figure 33: Feature Importance

Removing the last 4 features with less importance in our model caused the performance of our graph to increase.

References

[1] About The Data https://data.footprintnetwork.org/#/abouttheData

[2] Data Quality Scores https://www.footprintnetwork.org/data-quality-scores/