SC209

ENVIRONMENTAL STUDIES

GROUP 17 PROJECT REPORT

B.TECH. 2018-22

PAR and Climate change

Instructor:
PROF. RANENDU GHOSH

Under the technical guidance of:
SRIKUMAR SASTRY

May 16, 2020

Group Members

- Akshya Garg
  201801028
- Harsh Modi
  201801029
- Nikhil Mehta
  201801030
- Hemant Jain
  201801032
- Mihir Desai **(Group Leader)**
  201801033
- Devanshu Vithlani
  201801034
- Vedant Thakkar
  201801035
- Vishwa Shah
  201801036
- Pushpang Rathore
  201801037
- Chirayu Chaplot
  201801038
- Mahi Patel
  201801039
- Chaitanya Patel
  201801040
- Akarsh Jain
  201801041
- Yash Ghaghada
  201801042
- Darshil Rana
  201801462
- Parthiv Patel
  201801463
- Parth Bhoi
  201801464

Contents

1 Abstract

Executive Summary

Our aim is to use PAR (**P**hotosynthetically **A**ctive **R**adiation) measured by satellites and use its data to create a Machine Learning model to relate it with atmospheric CO_2 and climate change. PAR reaching the earth surface plays a very crucial role in studying and understanding CO_2 variations and global temperature being affected by vegetation canopies. We used two models: 1^{st} (multivariate linear regression) and 2^{nd} (time series forecasting) to establish a relationship between FPAR, CO_2 and global temperature with inputs taken from MODIS products (FPAR, NDVI , GPP, LAI) and NASA GES DISC products (mole fraction CO_2). There is a yearly seasonality observed in the FPAR data. Interestingly enough, FPAR and CO_2 were observed to have inverse seasonality relationships, while FPAR values peak in the month of October, the CO_2 levels are at their lowest for the year. This is explained by a well known phenomenon as Earth "breathing". As expected LAI is observed to be directly correlated with FPAR. There seems to be no clear relationship between seasonality of FPAR and Mean Temperature. We also used these models to forecast future values for the variables.

Keywords: FPAR, LAI, NDVI, GPP, CO_2, Multivariate Linear Regression, Mean Temperature

Problem Statement

The goal is to measure PAR (Photo-synthetically Active Radiation) using satellite data and use ML to explore its relation with atmospheric CO_2 and climate change.

The technical deliverables are the following:

1. Understanding different satellite data and GIS tools.
2. Relating PAR with CO_2 and climate change.
3. Using satellite data and ML to model and predict PAR and possible impacts on CO_2 levels and climate change.

2 Introduction

It is not possible for the plants to utilise the entire spectrum of the incoming radiation for the synthesis of energy and food. It is found that the range from $400nm$ to $700nm$ of the spectrum is useful for the plants to carry out photosynthesis and therefore it is known as **"Photosynthetically Active Radiation(PAR)."**

Researchers have been trying to find a relation between PAR and many other factors such as NDVI (Normalized Difference Vegetation Index), GPP (Gross Primary Production), carbon dioxide, etc.

- **NDVI:** It is a graphical method to assess whether the area of interest consists of live green vegetation or not.
- **GPP:** It is the total rate at which a material is produced, in plant terminology, it is the total amount of carbon dioxide fixed by the plants out of which some of it is used up in respiration.
- **APAR:** Amount of radiation absorbed which is in the range of 400-700nm.
- **FPAR:** It is the fraction of absorbed photosynthetically active radiation, **i.e. FPAR=APAR/PAR.**
- **LAI:** Leaf Area Index is the total leaf area per unit ground area. It is used to monitor growth, assessment of canopies, density.

FPAR is one of the 50 Essential Climate Variables recognized by the UN Global Climate Observing System (GCOS) as necessary to characterize the Earth's climate. So this is the reason why we use FPAR instead of PAR in our calculations further.

It becomes important for us to study these relations and how they affect the process of photosynthesis in plants because it then allows us to know the carbon dioxide levels in the atmosphere, the amount of carbon sequestered, etc.

Our area of interest has been the northern, central parts of Gujarat, some southern parts of Rajasthan, and western Madhya Pradesh. (70°E, 22°N, 75°E, 26°N). Two models have been used to determine various kinds of relationships between different parameters.

The 1st model uses linear regression to establish a relation between NDVI, GPP, and FPAR (Fraction of Absorbed Photosynthetically Active Radiation). It also uses multivariate regression to which the inputs are the three parameters NDVI, GPP, and FPAR and the output is the levels of carbon dioxide. The input CO_2 data was then krigged to create a one to one mapping between the values of FPAR and CO_2 with matching 2D array dimensions.

The 2nd model predicts the future value of carbon dioxide, LAI (Leaf Area Index), Mean temperature, FPAR based on a pre-computed dataset. It uses time series forecasting through a library named Prophet available in R or Python.

3 Data Cleansing and Preparation

3.1 Timeframe of study

Data was collected for the date 01-01-2010 for the regression model.
The time series model is trained on the average values of FPAR on this region from 01-01-2002 to 26-02-2019. The CO_2 was measured at the Mauna Loa Observatory in Hawaii, United States provided by NOAA for the same days.

For FPAR, we had studied data which has temporal resolution of 4 days and for CO_2 data has temporal resolution of 1 day. We chose these years because the data for CO_2 and FPAR was available in these years in both the data sets concurrently: Google Earth Engine MODIS, ERA5 Daily and NASA Earth data GES DISC.

3.2 Area description

[70°E, 22°N, 75°E, 26°N]

This region mostly covers North and central Gujarat and some southern parts of Rajasthan and western parts of Madhya Pradesh. The southern boundary of the area is near the coast of Gulf of Khambhat and the western boundary of the area is near the start of Gulf of Kutch.

In our region of interest, there are mainly three types of climate according to the Koppen climate classification which are:

- Tropical savanna climate (wet and dry) on the southern and eastern sides
- Warm semi-arid climate in most of the central parts
- Warm desert climate on the northern and western sides

For accurate data of CO_2, our region of interest is Mauna Loa Observatory (MLO) on the island of Hawaii. [9]
MLO has activities at five locations on the island of Hawaii, among them the primary observation site is located at [19°32'10" N, 155°34'34" W] coordinates. The observatory is at an altitude of 3400m, which is well suited to measure CO_2 data that is representative of very large areas. Being at a high altitude, it is above the temperature inversion layer.

So,

- undisturbed air (due to high altitude),
- remote location at Hawaii island, and
- minimal influence of vegetation and human activities

are ideal for monitoring and collecting atmospheric elements related with climate change.

3.3 Variables used

- **LAI** (Leaf Area Index) *{unitless}*
- **FPAR** (Fractional Photosynthetically Active Radiation) *{unitless}*
- **CO_2** *{ppm}*
- **NDVI** (Normalized Difference Vegetation Index)
- **Mean Temperature** *{Kelvin}*
- **GPP** (Gross Primary Productivity) *{Mass per unit area per unit time interval}*

3.4 Datasets used

- **FPAR**: MCD15A3H.006 MODIS Leaf Area Index/FPAR 4-Day Global 500m, FPAR absorbed by the green elements of a vegetation canopy [6]
- MODIS Terra Daily NDVI
  Google, MODIS/006/MOD09GA [8]
- **GPP**: MOD17A2H.006: Terra Gross Primary Productivity 8-Day Global 500m, Gross Primary Production. [7]
- **CO_2**: AIRS Science Team/Joao Teixeira (2009), AIRS/Aqua L3 Daily CO_2. [4]
- ERA5 Daily aggregates - Latest climate reanalysis produced by ECMWF / Copernicus Climate Change Service [5]
- Mauna Loa Observatory CO_2 dataset [9]

3.5 How often are these variables collected?

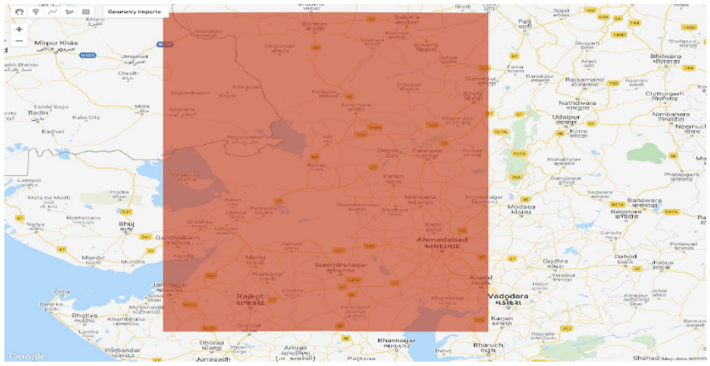
Spatial resolution for CO_2: 2°(lon) x 2.5°(lat)
Spatial resolution for FPAR: 500m x 500m
Spatial resolution for GPP: 500m x 500m
Spatial resolution for NDVI: 500m x 500m

Temporal resolution for CO_2: 1 day
Temporal resolution for FPAR: 4 days
Temporal resolution for GPP: 8 days
Temporal resolution for NDVI: 1 day

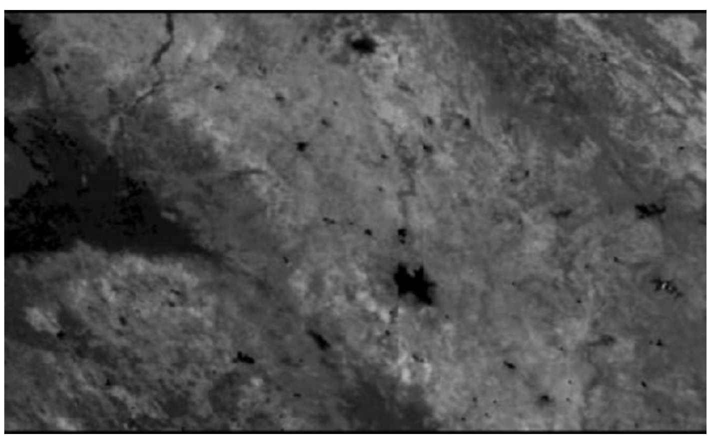
3.6 Data Scraping

1. **FPAR, NDVI and GPP**

   - We highlighted the region of interest in google earth engine code editor.

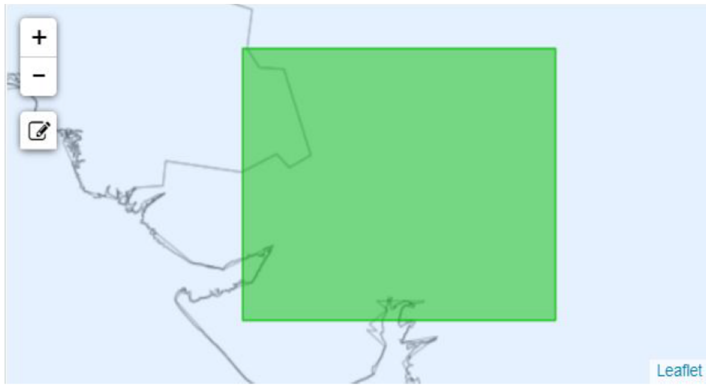
   - We wrote a node.js code with the purpose of scrapping FPAR and LAI data. In the code, we
     - import our region of interest.
     - import the dataset.
     - specify the time period
     - specify the bands related to FPAR for FPAR data.
     - specify the bands related to GPP for GPP data.
     - specify the bands related to NDVI for NDVI data.
   - Upon scrapping we successfully exported the GeoTiff images mapped for each timestamp into google drive.

   - We used the GDAL (Geospatial Data Abstraction Library) library in python to convert the GeoTiff image into a 2D matrix and later into CSV data format for ML modelling purposes.

2. **CO_2**

- In the dataset we selected the coordinates [70:75°E] with 2.50° spatial resolution and [22:26°N] with 2.0° spatial resolution. That means we got 3 points on the latitude and 3 points on the longitude.



- We extracted a netCDF file (.nc extension) from the NASA GIS dataset which contains the value of xCO_2 of these 9 points.
- We converted this netCDF file into a CSV file using python.
- We got data of CO_2 for only 9 points from this dataset (3 points on the latitude and 3 points on longitude), but from the google earth engine MODIS dataset we got the data for FPAR in the 2D matrix of the shape $448 * 737$.
- As the dataset does not have all the pixels unlike the dataset from google earth engine, we interpolated the data of CO_2 using kriging to maintain data coherency across our variables.
- However, we observed while training our model that the predictions we were getting using the aforementioned CO_2 dataset along with the FPAR, were not accurate. So, we used instead CO_2 data of Mauna Loa, which is standardly available in CSV format.

3.7 Correlation Matrix between selected variables

	Mean Temp	**CO_2**	**FPAR**	**LAI**	**ΔCO_2**
Mean Temp	1	0.097957	-0.412452	-0.172790	-0.076492
CO_2	0.097957	1	-0.009282	0.005060	0.045165
FPAR	-0.412452	-0.009282	1	0.934478	0.085144
LAI	-0.172790	0.005060	0.934478	1	0.027801
ΔCO_2	-0.076492	0.045165	0.085144	0.027801	1

The correlation coefficient between two variables represents how changes in one variable affect the other. In formal language if one variable is 'y' and other is 'x' then it represents $\Delta y/\Delta x$ and $|\Delta y/\Delta x|$ value represents the strength of the correlation. Here we can observe from the above table that the strength of correlation between FPAR and xCO_2 is very weak so we can infer that they are not linearly related.

4 Methodology and Results

4.1 Analysis of data

The FPAR, CO_2 and Temperature data we collected had a certain yearly and monthly trend. Looking at the data, the time series model seems to be a good prediction model.

What is Time Series Analysis?
Time series analysis is a statistical technique that deals with time-series data, or trend analysis. Time series data means that data is in a series of particular time periods or intervals. The data is considered in three types:

1. **Time series data:** A set of observations on the values that a variable takes at different times.

2. **Cross-sectional data:** Data of one or more variables, collected at the same point in time.

3. **Pooled data:** A combination of time series data and cross-sectional data.

Using Time Series Analysis to predict FPAR, CO_2, LAI and Mean Temperature values

- The Data that we collected was a Time Series data i.e. the variables (FPAR, CO_2, LAI and Temperature) took different values at different times.

- The collected data showed a weekly and yearly trend as seen in the given graphs. {Figure 1}

- With the aid of weekly and yearly trends, we can see if there exist any pairwise relationship between these values.

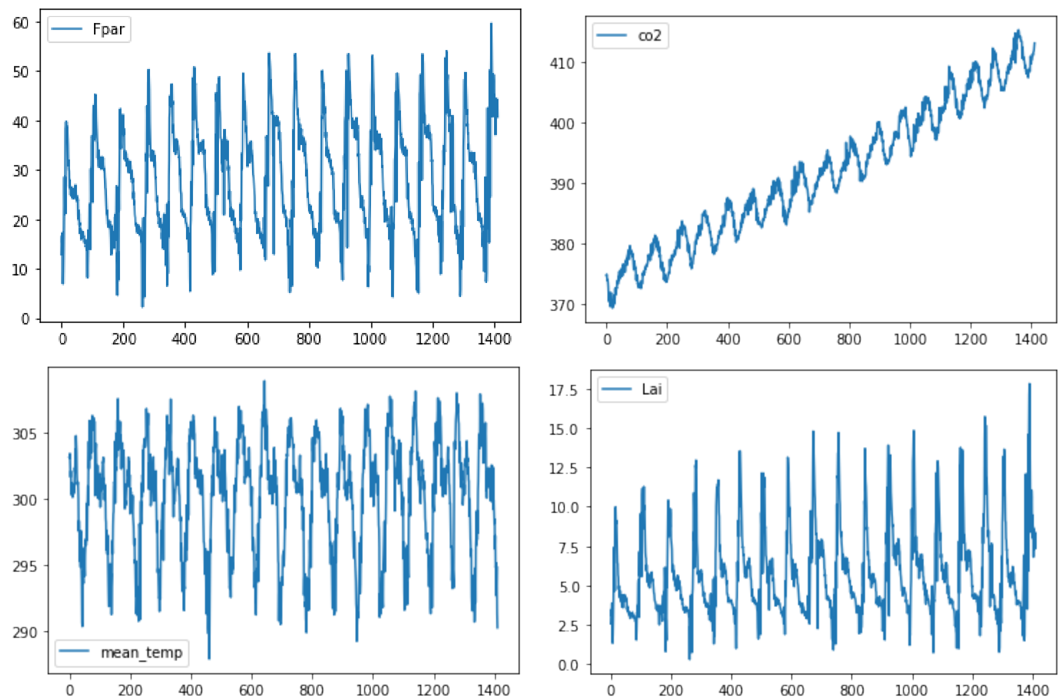


Figure 1: Collected FPAR, CO_2 and Mean Temperature data

4.2 Time Series Model

1. **Requirements of the model** [1][2]

   - The model requires data to be collected on a daily basis or pre-aggregated data over a month with few gaps over the year for a desirable prediction.
   - It strictly requires the data fed to the model to be stationary, that is the average and variance of the data shouldn't change for parts.

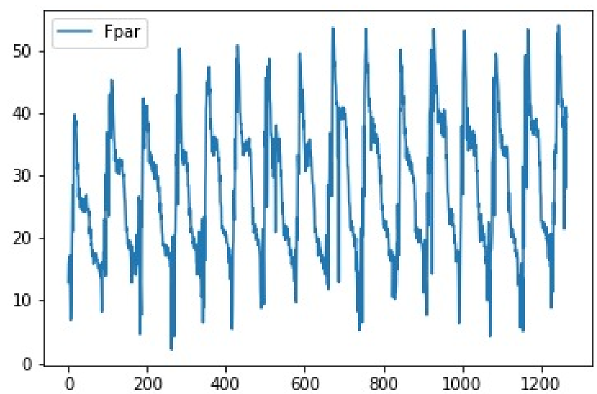
2. **Why do we require stationary data?** [3]
   The quantities we are typically interested in when we perform statistical analysis are:

   - Its expected value
   - Its variance
   - The correlation between values S periods apart for the set of S values.

   We calculate these by taking the mean across many time periods, which is only informative if the expected value is the same across those time periods. In addition, stationary processes avoid the problem of spurious regression.

Note

Sometimes the data collected might not be stationary as shown.



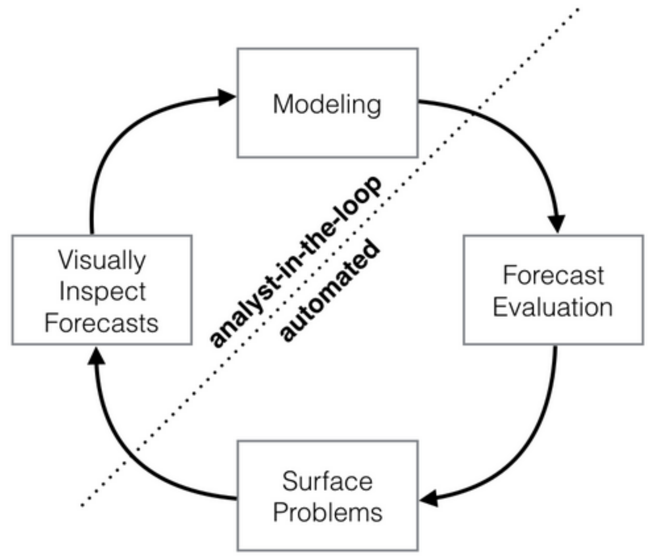
Taking logarithmic values allows us to convert it into stationary data.

3. **What is the Prophet library and why use it?** [10]

We are looking to perform a time series analysis on the obtained data to create a forecasting model that can predict the future values based on past values. This demands for expertise in forecasting and machine learning to yield high-quality results. As it becomes more popular, the demand for such analysis increases which cannot be matched by the availability of expert analysts. Although there exist completely automatic techniques and models, they can be hard to tune and often too inflexible for the user. Prophet is an open-source forecasting tool from Facebook. It simplifies the forecasting process in data science for analysts as it is not easy for everyone to produce high-quality results which have risen in demand with time.

Not all forecasting problems can be solved by the same process and Prophet is optimized to handle data characteristics like seasonalities, holidays at irregular intervals, large outliers or missing observations, historical trend changes and trends with non-linear growth curves which tend to hit a saturation limit. This translates to better business forecasting techniques and planning but can be extended to all time-series analysis as well.

The default forecast library in python has a lot of different techniques and choosing an inefficient model for a particular data can yield poor results. Prophet allows non-experts to customize forecasts intuitively and define characteristics like how aggressively should it follow historical trend changes or inject information about how the forecast will grow or decline. This allows for an analyst-in-the-loop approach which combines human and machine automation capabilities to yield the best possible results.



After inspecting and reading the output, the analyst can fine-tune the model's characteristics to yield a better forecast with the same data. When a problem occurs or poor performance is detected, Prophet surfaces these issues to the analyst to help them understand what went wrong and how to adjust the model based on the feedback.

Prophet uses a procedure for forecasting time series data based on an additive model where non-linear trends are fitted with yearly, weekly, and daily seasonality, plus holiday effects are given as:

$$y(t) = g(t) + s(t) + h(t) + E$$

Here $g(t)$ is the trend function which models non-periodic changes in the value of the time-series, $s(t)$ represents periodic changes (e.g., weekly and yearly seasonality), and $h(t)$ represents the effects of holidays which occur on potentially irregular schedules over one or more days. The error term E represents any idiosyncratic changes which are not accommodated by the model.
This is a similar generalized additive model (GAM) which a class of regression models with potentially non-linear smoothers applied to the regressors. Here, time is the only regressor but we use several linear and nonlinear functions of time as components. GAM allows easy decomposition into components and also adding new components.

4. **The algorithm prophet follows:**

   (i) **Installation:** Install pystan with pip and then using pip to install fbprophet and follow the syntax of python.

   (ii) **Importing the dataset:** The input to Prophet is a data frame with two columns:

   (a) **ds**(datestamp) column: It is of the format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH:MM: SS for a timestamp.

   (b) **y** column: It must be numeric and represents the measurement we wish to forecast like in this model.

   (iii) **To get the extreme end values:** The method head() and tail() is used to get first 5 and last 5 values.

   (iv) **Making the predictions:** Prophet uses Scikit-learn which is a free machine learning library for Python where Prophet class is created and its methods like fit() train the data part of the modelling process. It finds the coefficients for the equation specified by the algorithm:

   (a) **Seasonalities:** Prophet will by default fit daily, weekly and yearly seasonalities. It will also fit daily seasonality for a sub-daily time series and other seasonalities can be added too.

   (b) **Forecasting:** For forecasting, we need to define how far to predict in future and it can be in any form monthly, daily etc. For this, we need to make a data frame for future predictions using make_future_dataframe and then define the period.

   (c) **Predicting:** When the predict function is called, it will assign each row in future a predicted value which it names "yhat" and the range is defined by "yhat_lower" and "yhat_upper". These ranges can be considered as uncertainty levels.

   (v) **Plotting the forecast:** plot() method is used to plot the forecast with time in x-axis. The data is converted into its log to avoid skewness in large data values. The forecast's components are seen by plot_component() method which shows the daily and weekly trends which makes the understanding more clear.

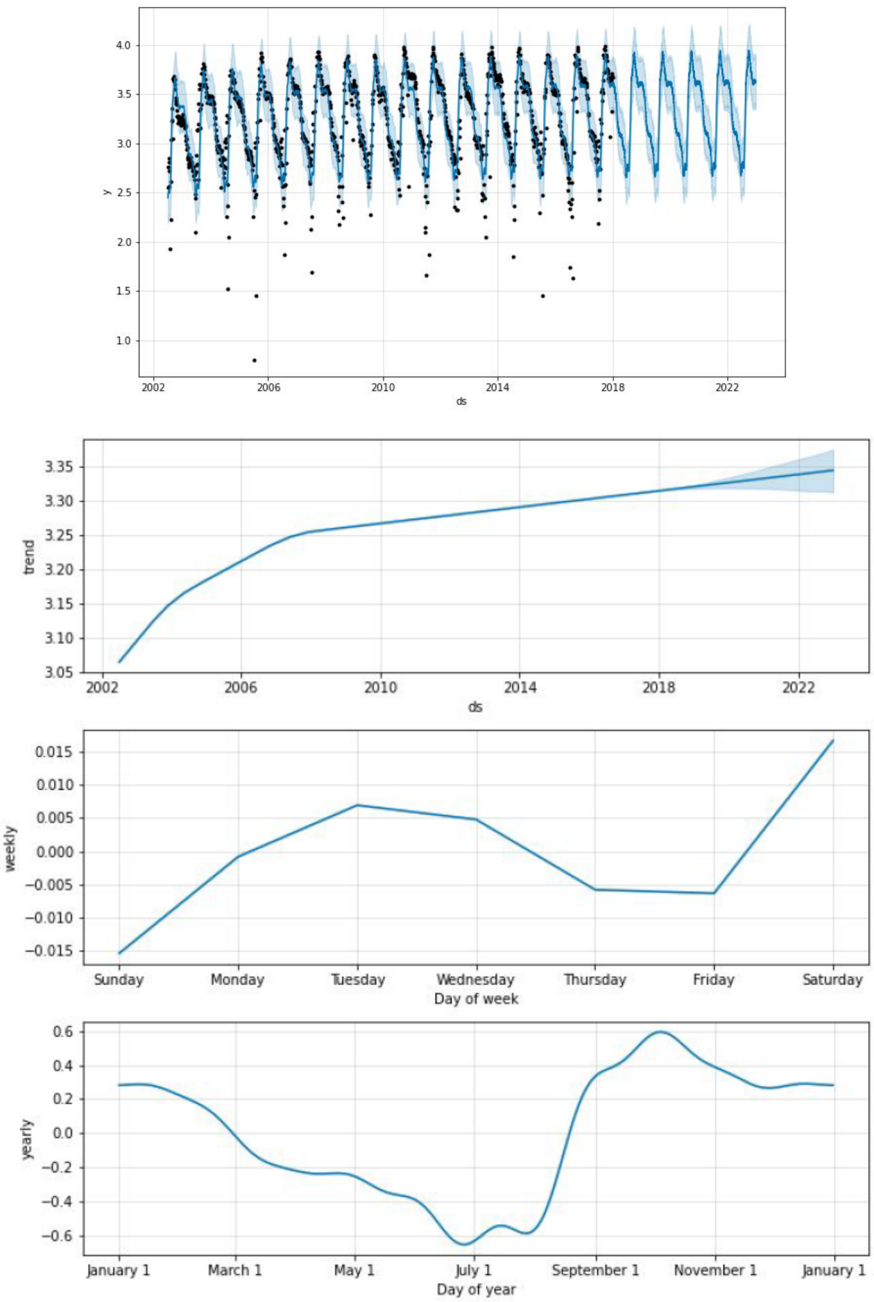


Figure 2: Graph results of FPAR time series analysis

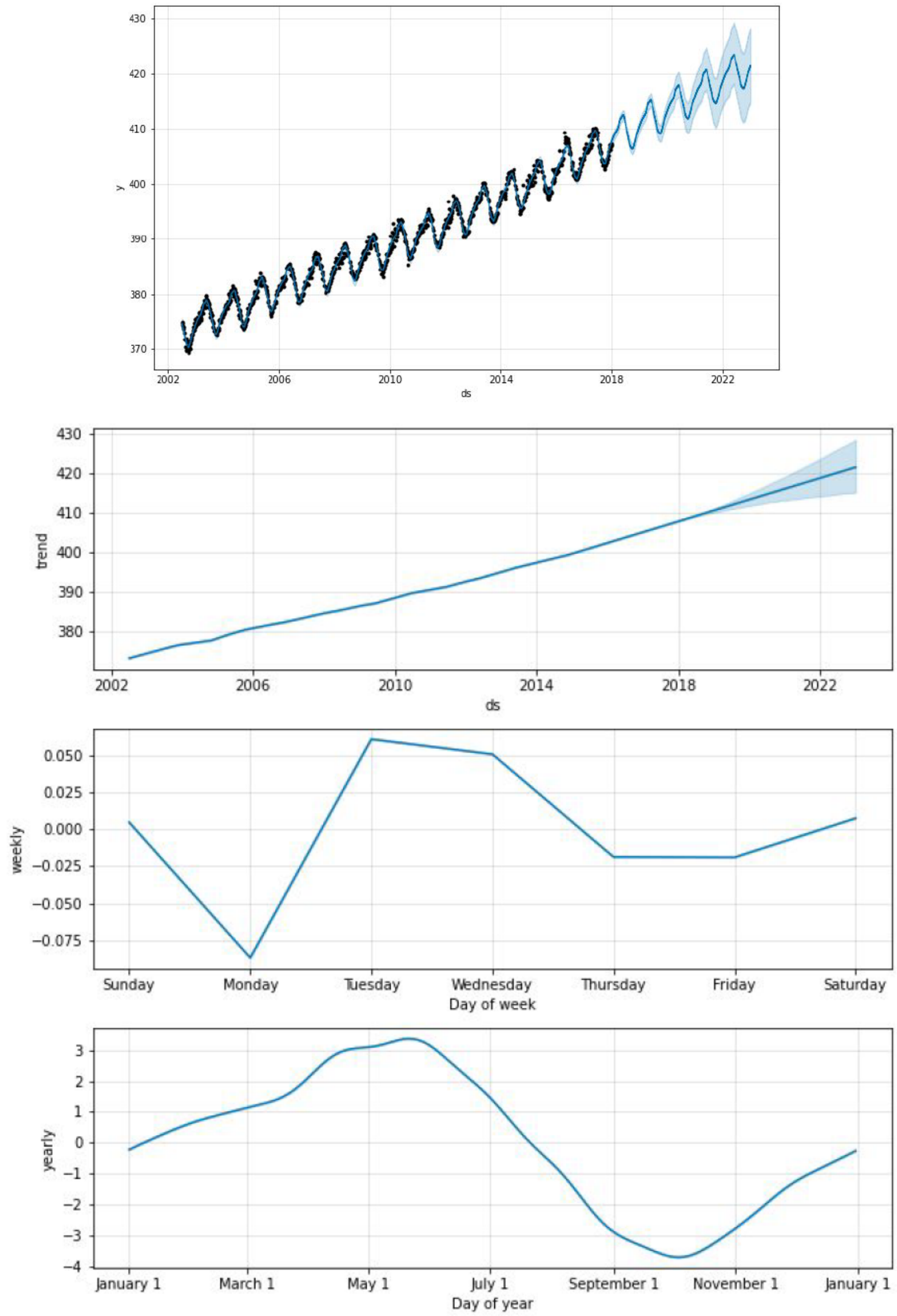


Figure 3: Graph results of CO_2 time series analysis

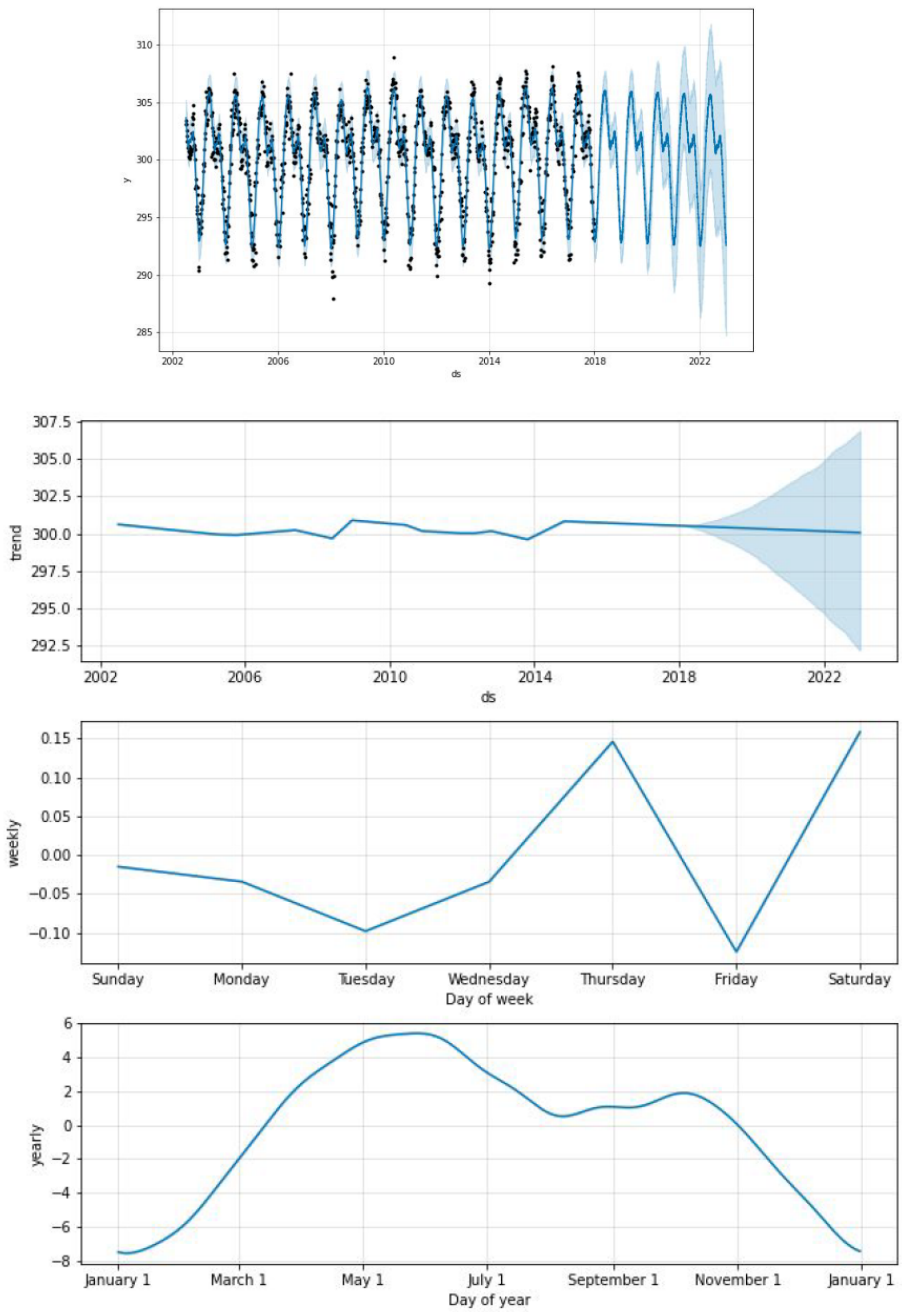


Figure 4: Graph results of mean temperature time series analysis

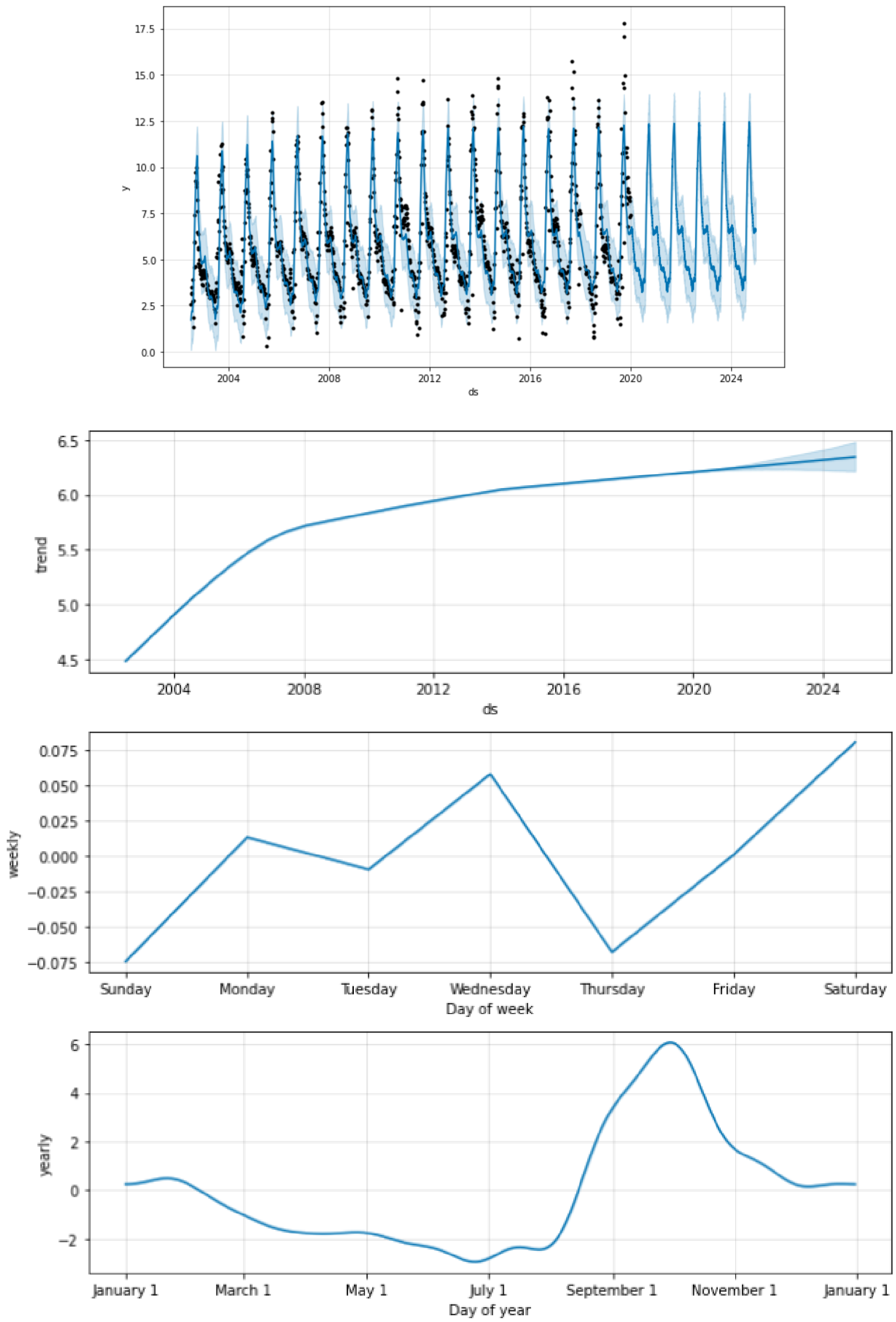


Figure 5: Graph results of LAI time series analysis

4.3 Analysis of the Graphs

- **FPAR**

  1. As seen from the trends of the graphs, the value of FPAR is predicted to increase w.r.t time (in years).
  2. When we look at the monthly trend, its values are at its peak during winter months November, December and its values decrease during the summer months of May, June, July.

- **Mean Temperature**

  1. The Mean Temperature is predicted to not have any significant change over the passage of time (in years).
  2. From the monthly trend, the value of Mean Temperature is at its maximum in months of summer (May-July) and it is minimum during the months of winter(January).

- **CO_2**

  1. The concentration of CO_2 is predicted to increase significantly over the passage of time(in years).
  2. From the monthly trend, the concentration of CO_2 is at its maximum in months of summer(May-July) and at its minimum during the final months of the monsoon season (September-November).

- **LAI**

  1. The Leaf Area Index has an increasing, upward graph over the years.
  2. The value of LAI, like FPAR peaks during the winter months (October-December) and is at its lowest during summer (May-July).

4.4 Comparison between the variables

The relation between FPAR and LAI:

- As expected there is a direct correlation between FPAR and LAI. They both have the same increasing trend which seems to be slowing down and the seasonalities for the weekly as well as yearly data are corresponding as well with each other.

The relation between FPAR and Mean Temperature:

- There seems to be no trend of increase or decrease for Mean Temperature and it seems to remain constant for the years coming though the model is not confident about the prediction due to absence of any clear trend.

- There seems to be no relationship between seasonality of FPAR and Mean Temperature for a week as well as a year.

The relation between FPAR and CO_2:

- We can see that both FPAR and CO_2 have an upward trend (1st plot) as seen from the predictions of the model and will increase in the coming years. The trend for increase of CO_2 has been constant whereas it is seen to be slowing down for FPAR.

- If we observe the yearly seasonality (3rd plot) for both the variables we see an interesting relationship, FPAR almost shows an inverse relationship to the amount of CO_2 for the respective months of a year which suggests that growth of CO_2 has a negative effect on FPAR. We can clearly observe that FPAR peaks in the month of October when the CO_2 levels are at their lowest for the year.

- The observation can be explained as follows: The annual seasonal cycle in atmospheric CO_2 concentrations is well known, largely attributable to the preponderance of mid-latitude, terrestrial vegetation in the northern hemisphere compared to the southern hemisphere.

- When northern hemisphere vegetation is photosynthetically active, which roughly corresponds with emergence through reproduction and primarily occurs during May-September, atmospheric CO_2 is consumed and its concentration declines. Conversely, when northern hemisphere vegetation is senescing or dormant, atmospheric CO_2 levels increase. By analogy with organismic respiration, this oscillatory behaviour is sometimes described as the Earth "breathing."

4.5 Accuracy of the model

Root Mean Squared Error for the models:
(These are average values when predicted on a horizon of 19-180 days.)

Model	**RMS Error**
FPAR	5
CO_2	0.8
Mean Temp.	1.5
LAI	1.3

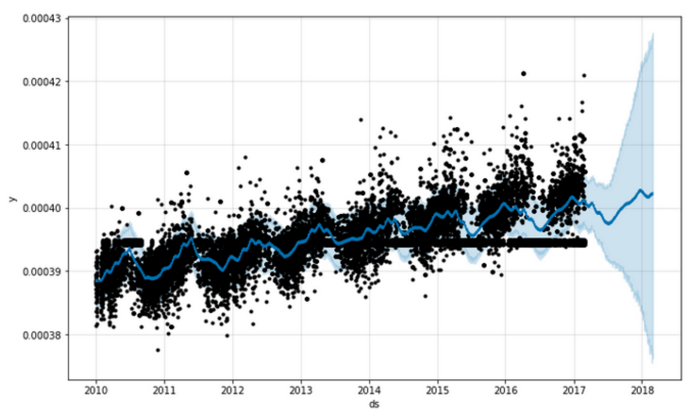
The model seems to be quite confident about the prediction evident from the narrow confidence interval (blue faded region) for FPAR and LAI.
Although, for Mean Temperature and CO_2 the confidence area is not as narrow.

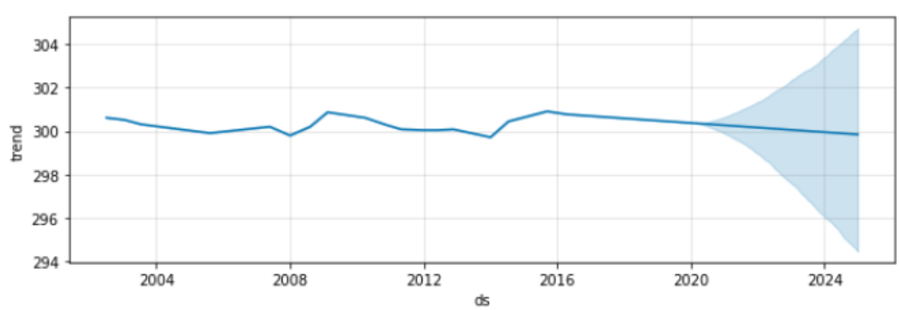
4.6 Time Series Model of CO_2 data taken over Gujarat

(i) **Dataset used:**
NASA GES DISC dataset. [lon:22-26°N and lat:70-75°E]

- Thus we get a total of 9 data points at the intersection of the latitudes and longitudes.
- This Time Series Model analyzes and predicts the CO_2 data at each of these 9 data points.

(ii) **Results and Graphs:**
At all the nine data points the trend observed was almost similar to the graph below:





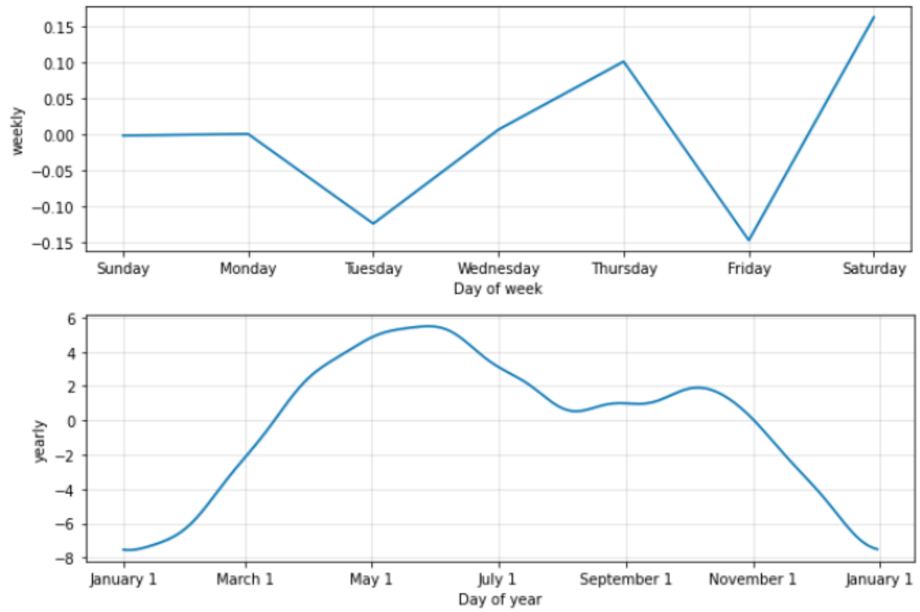


Figure 6: Graph results of CO_2 analysis of Gujarat region

As observed earlier with Mauna Loa dataset, the CO_2 in Gujarat also follows an increasing trend with its peak during the summer months.

Note

The Data used here had a lot of NAN (not-a-number) values, 15000 out of 24000.

At those NAN values, an average of nearby data points was taken, which is not the most accurate value at that intersection point.

Hence the anomaly in the analysis and prediction as seen.

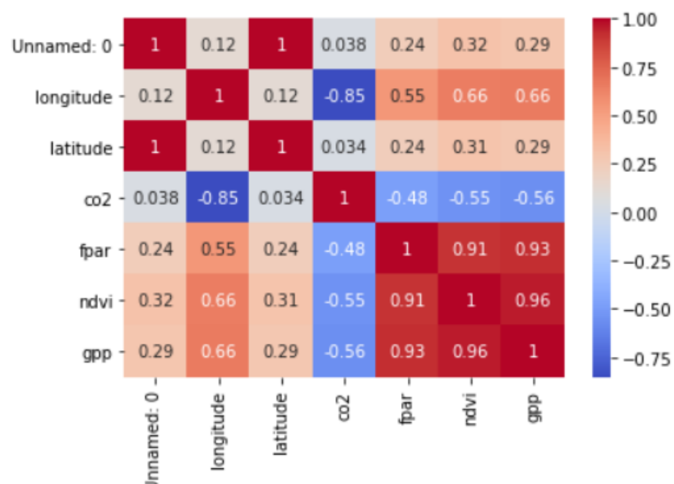
After testing the time series model, we observed a trend between FPAR and CO_2 using regression techniques.

5 Observing linearity between FPAR, CO_2 and other factors

For these Observations, we took other quantities such as NDVI and GPP along with FPAR.

5.1 Datasets used

The dataset of FPAR, NDVI and GPP is taken from the Google Earth Engine which uses the data from the tiff image generated for 01-01-2010 and its dimensions are mapped to the CO_2 values which are taken from NASA GES DISC dataset. The CO_2 values are krigged to the dimensions of the above quantities to map it to a particular pixel or point. To use the data in the regression model, we reshaped the data set available using the .reshape() method to make the array to be two-dimensional, or to have one column and as many rows as necessary. The argument (-1, 1) of .reshape() specifies the same and then the linear regression model was implemented. When the heat map for NDVI, CO_2, FPAR, GPP with respect to their latitude and longitude then it was noticed that the relationship between CO_2 and FPAR is quite less i.e. -0.48 and which is less than 0.5. This shows that FPAR and CO_2 can be related precisely when other factors and quantities are also considered. Whereas the relation between NDVI- FPAR and GPP-FPAR was quite high i.e. 0.93 and 0.91 respectively and this shows that they are directly related to each other.



5.2 Implementation and Methodology

Instead of modelling FPAR and CO_2 directly, we first trained the models using linear regression and found out the relation between FPAR-NDVI, FPAR-GPPand NDVI-GPP. Linear regression is used to determine the extent of the linear relationship of a dependent variable and independent variables. The model was trained and the linear regression line was plotted. To statistically measure how close the data points are to the fitted regression line we used R-squared values.

R-squared = Explained variation / Total variation

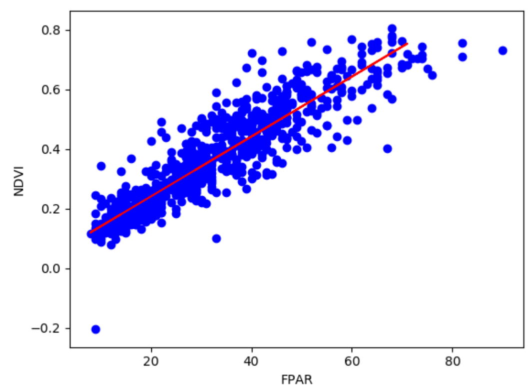
It is also calculated by the ratio of summation of square of the difference of regression line points and their mean and summation of square of the difference of actual points and their mean. It lies between 0 and 1. Higher the R-squared value, better the model fits the data. The R-square values of these corresponding models were quite high and so is the accuracy of the model.

After that, we used a multivariate linear regression model wherein we can find the relation between a dependent variable and more than one independent variable. Here, we trained the linear regression model taking FPAR, NDVI and GPP as the independent variables and CO_2 as the dependent variable.

We calculated the score i.e., the R^2 value of the prediction comparing how accurate the model fits in the data.
But the main reason to train the data was to find out the R^2 value which will indicate how linear the quantities are which will help us to find any trend in the observation. Thus, the model was implemented and further analysed.

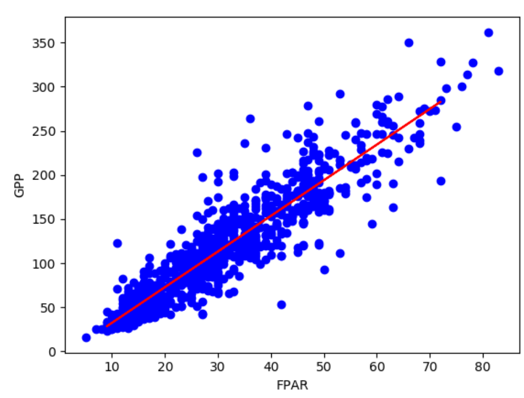
5.3 Results

- **Finding the relation of FPAR with NDVI**
  The data of FPAR was the input that was given to a linear regression model and NDVI values were predicted. We trained the model on some values of the original Dataset.

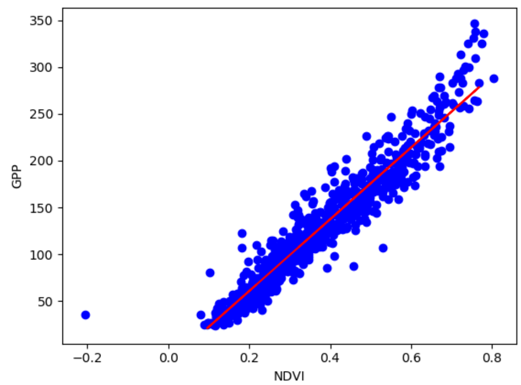


  The blue scatter points are FPAR v.s. NDVI values from the original dataset and the red line is the plot of the predicted values of NDVI for the given test values of FPAR.
  The R^2 score for the relation between FPAR and NDVI is around 0.85. So there is a linear relationship between both of them, which can be interpreted from the graph too.

- **Finding the relation of FPAR with GPP**
  The R^2 score is around 0.84 between FPAR and GPP. The relation between FPAR and GPP is also linear as seen in the graph below.



- **Finding the relation of NDVI with GPP**
  As seen for FPAR v.s. NDVI, we see that NDVI has a linear relationship with GPP too.

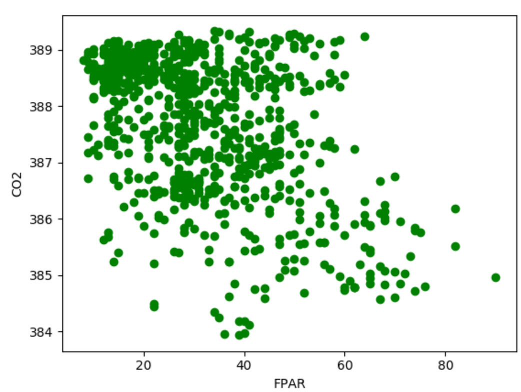


The R^2 score between NDVI and GPP is around 0.95. So we can say all three quantities are linearly proportional to one another.

> **Thus we see that there is a linear relationship between all the three variables FPAR, NDVI and GPP.**
> We can make use of these three quantities together to find the relation with CO_2 in the atmosphere to get better results.

- **Using FPAR, NDVI, GPP to find their relationship with CO_2**

  Plotting the graph of FPAR and CO_2:



We can see that there are more points on the upper left corner of the graph where the value of FPAR is small and for CO_2 in ppm is more. So the general trend observed is that **as the values of FPAR increase, there is a decrease in the values of CO_2.**

The relationship doesn't seem to be linear as there are many exceptions to this general trend which is observed in the graph.

Further, we took more factors/features into consideration, NDVI, FPAR, GPP to check the linearity of the factors.

> **The R^2 value for this multivariate regression model is about 0.32.**
> **The relation between these factors and CO_2 is not linear.**
> This supports the time series model which was discussed earlier.

Even though the model uses FPAR, GPP, and NDVI to observe the relation with CO_2, one of the main factors contributing to this is that the model does not consider the factors such as industrial output, livestock-related and other human influences which contribute to the amount of CO_2 in the atmosphere. This should be one of the reasons for the anomalies that can be seen in the above graph.

6 Conclusion

So far from the analysis and study of CO_2 and FPAR, we have reached the following conclusions:

- CO_2 and FPAR have a general inverse relationship between them; we could confirm this fact from the time series analysis of CO_2 and FPAR that they obey some sort of inverse relation, an increase in the value of FPAR means that more sunlight is available for plants to do photosynthesis, so the amount of CO_2 decreases in the atmosphere.

- From what we have seen in the time series analysis graphs, the major reason for climate change in terms of CO_2 increase are humans only i.e., the CO_2 released from the factories, activities like deforestation are major factors of climate change, we can see that over time the CO_2 concentration is increasing at a constant rate, whereas the FPAR value remains almost constant.
  FPAR does affect the natural CO_2 concentration over time, but it does not explain the constant rise in the CO_2 levels, it is mainly due to human activities.

- We can infer that FPAR does not play any major role in the climate change phenomena as the radiation from the Sun almost remains constant and is independent of any earthly parameters.
  The annual seasonal cycle in atmospheric CO_2 concentrations is well known, as mentioned before, primarily during May-September, atmospheric CO_2 is consumed and its concentration declines. Conversely, atmospheric CO_2 levels increase during the opposite half of the year. By analogy with organismic respiration, this oscillatory behavior is sometimes described as the Earth "breathing."

- To control the amount of CO_2 levels to a natural scale, we have to invent some structure either it be chemical or engineered that could act like carbon sinks, but the natural way of solving the problem is by planting more trees and sustained afforestation that help in depleting the CO_2 offset that is continuously rising.

- As far as the relationship between FPAR and CO_2 is concerned we observed that we could not get any simple relation between FPAR and CO_2 as a regression model so we could assume that the CO_2 data could be polluted by the Industrial emission and pollution by vehicles.

  But as we include other parameters like NDVI and GPP we could get more accurate model than the model which only includes FPAR as a parameter, we observed an increase in the accuracy of our model by a certain amount by including those parameters.

7 References

[1] URL: https://www.statisticssolutions.com/time-series-analysis.

[2] URL: https://docs.microsoft.com/en-us/sql/tutorials/time-series-model-requirements-intermediate-data-mining-tutorial?view=sql-server-2014.

[3] URL: https://stats.stackexchange.com/questions/19715/why-does-a-time-series-have-to-be-stationary.

[4] *AIRS Science Team/Joao Teixeira (2009), AIRS/Aqua L3 Daily CO2 in the free troposphere (AIRS-only) 2.5 degrees x 2 degrees V005, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC).* URL: 10.5067/Aqua/AIRS/DATA335.

[5] *ERA5 Daily aggregates - Latest climate reanalysis produced by ECMWF / Copernicus Climate Change Service.* URL: https://developers.google.com/earth-engine/datasets/tags/era5.

[6] *FPAR: MCD15A3H.006 MODIS Leaf Area Index/FPAR 4-Day Global 500m.* URL: https://doi.org/10.5067/MODIS/MCD15A3H.006.

[7] *GPP: MOD17A2H.006: Terra Gross Primary Productivity 8-Day Global 500m, Gross Primary Production.* URL: https://doi.org/10.5067/MODIS/MOD17A2H.006.

[8] *MODIS Terra Daily NDVI Google, MODIS/006/MOD09GA.* URL: https://doi.org/10.5067/MODIS/MOD09GA.006.

[9] *Mount Loa Observatory data.* URL: https://www.esrl.noaa.gov/gmd/obop/mlo.

[10] Letham B Taylor SJ. *Forecasting at scale.* 2017. URL: https://doi.org/10.7287/peerj.preprints.3190v2.