# PES University

# Summer Internship at KAnOE

Analysis of Relationship between Forest Cover and Climate and Population in India

TEAM MEMBERS

N L Ramya : 01FB14ECS124

N Sourabh : 01FB14ECS125

Rhea Anand : 01FB14ECS186

Riya John : 01FB14ECS191

ABSTRACT

The study aims to analyze the effect of forests in India on rainfall and temperature. The study attempted to establish that the forest cover is an important factor in determining the amount of rainfall and the average temperature of a district. Today forest cover conversion and deforestation are of global concern so analysis is done on how the population density affects the forest cover. The study also determines the effect of rainfall and temperature on forest cover.

The correlation coefficient is calculated to find the strength of the relationship between the forest cover and the factors - rainfall, temperature and population density. Linear Regression Model (LRM) is used to predict the rainfall, population density and temperature. Using data from 2005 to 2013, forest cover and the variables – rainfall, temperature and population density are predicted for the years 2020, 2025, 2030. The forest cover in India is predicted using Multiple Linear Regression.

INTRODUCTION

The change in forest cover is of global concern as forest is a precious natural resource required to sustain life on Earth. Deforestation is a major cause for climate change and increase in global warming. With the world growing at a very high pace, the increasing need for space is turning out to be an area of concern. And this desperate need has led to felling of trees and clearing of forest cover.

Considering the fact that climate change, a growing critical issue is because of depleting forest cover, the study aims to analyze the complex relationship between forest cover and climate – average rainfall and temperature. Studies related to forest cover change using statistical data help in understanding and proving the phenomena like climate change and global warming.

For better understanding of the impact of forest cover change, factors affecting it must be fully studied. With rapid increases in population, pressure on the natural resource has become intense. Hence the project attempts to establish a critical relationship between forest cover and population density.

The project is aims to analyze for each district in the country and based on the physiographic divisions – Northern Mountains, Northern Plains, Western Plains, Deccan Plateau and Coastal region.

COLLECTION AND MODIFICATION OF DATA

The datasets required for this project were mainly - Forest Cover, Temperature, Rainfall, Population.

The site https://data.gov.in/ provided us with three of these datasets in CSV form namely, Forest cover (2007,2011,2013), Rainfall(2004-2010), Population (1991,2001,2011) with district level granularity. The Temperature data was obtained from Kanoe in grid form for combinations of latitudes and longitudes spanned by India with Avg. values for each month through the years (1951-2013). The common years decided for our study were (2005-2013).

Forest cover for 2005 was found in a forest report in pdf form and was converted to CSV files through the help of online OCR readers and manual corrections. Since forest survey is conducted every 2 years and reports for 2009 weren't available interpolation was done using excel to get values for the missing years .

pdf form reports found on -

http://fsi.nic.in/details.php?pgID=sb\_18

Rainfall data for the years 2010 -2015 was collected from the website Customised Rainfall Information System (CRIS) by the Indian meteorological department. The tables were scraped manually onto CSV files.

Rainfall data for previous five years can be found on -

http://hydro.imd.gov.in/hydrometweb/(S(fw03fs55l50hnqvj0byer3jm))/DistrictRaifall.aspx

The Temperature data was converted to CSV from the txt output of the grid file using python codes.Longitudes and latitudes combinations were mapped to districts. The conversions and approximations required for this process were done using python.

Site used to find the latitudes and longitudes-

http://www.mapsofworld.com/lat\_long/india-lat-long.html

Census of Population is conducted every 10 years . Data was interpolated using excel to find the values for the population density for the years in between.

Distribution of districts according to the type of region they come under was done by scraping information from various sites.

Intersection combination of the obtained datasets was done using excel or python depending on the complexity of the task.

METHODOLOGY

After all the data was compiled and brought to a usable format, graphs were plotted between the forest cover and each of the other factors, namely- rainfall, temperature and population density. Scatter plots with forest cover on the x axis and temperature and rainfall on the y axis was plotted. Also, scatter plots with population density on the x axis and the forest cover on the y axis was also plotted. These graphs were plotted for every year from 2007 to 2013. Separate graphs for all the different physiographic regions of India are plotted for every year. Excel and python was used to plot these graphs.

The correlation coefficient signifies the strength of the relationship between any two variables. The range of possible values for a correlation is between -1 to +1. A positivecorrelation indicates a positive linear association. The strength of the positive linear association increases as the correlation becomes closer to +1. A negative correlation indicates a negative linear association. The strength of the negative linear association increases as the correlation becomes closer to -1.

Analysis ToolPak, an addon in Microsoft Excel was installed to help calculate the correlation coefficient. The correlation between the forest cover and each of rainfall, temperature and population was calculated using this package. The correlation coefficient is calculated for every year from 2007 to 2013. For each year, the correlation coefficient is found for every physiographic region of India.

Regression analysis is used to produce an equation that will predict a dependent variable using one or more independent variables. This equation has the form :*Y = b1X1 + b2X2 + .. + A.* Here, Y is the dependent variable while X1, X2 and so on are independent variables. A is the intercept.

Regression analysis results in a number of number of factors such as R2*,* adjusted R2and p value*.* R2 is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination. The adjusted R2 tells you the percentage of variation explained by only the independent variables that actually affect the dependent variable. The p-value for each term tests the null hypothesis that the coefficient is equal to zero. A low p-value (< 0.05) indicates that you can reject the null hypothesis. In other words, the lower the p value, the greater the chance of the model being accurate.

Regression analysis was performed using ‘R’ programming language. Linear regression was used to predict the rainfall, temperature and population density for the years 2020, 2025 and 2030. The dependent variable was either the rainfall or temperature or the population while the independent variable was the year.

After the values of the temperature, rainfall and population were calculated, multiple linear regression was used to predict the forest cover. The forest cover was chosen as the dependent variable while the other three factors formed the independent variables.

To present and explain all of the obtained results in a user-friendly and an understandable manner, a website was created with the help of Bootstrap, HTML and Javascript. PHP, Ajax and MySql was used in representing the data on the website.

CONCLUSION

The Forest cover data revealed that the forest area has undergone continuous change leading to the loss of 107.2 sq.kmforest in the past 2005–2013. The various forms of human encroachments and global warming have resulted into the loss of forest resources in the area. It was found that the independent variables, the factors considered for the study - rainfall, temperature and population density were associated with forest-cover.

The correlation coefficient indicates the strength of the relationship between any two factors. The average correlation between the rainfall and the forest cover is around 0.5 which shows a positive moderate relationship between them.

It is observed that there is an average correlation of around -0.33 between the forest cover and the temperature which indicates a very slight relationship between them. An average correlation of -0.09 was observed between the forest cover and the population density which tells us that almost no real relationship exists between the two of them. The negative value signifies that the forest cover decreases with an increase in the population and the temperature.

From the regression coefficient, the analysis reveals that increasing population density leads to a decrease in forest cover. Form the analysis on the northern mountains it can be concluded that an increase in population density by 1 unit can lead to decrease in forest-cover by 0.05% and on an average for the entire country it was decrease by 0.001%.

From positive correlation coefficient in most of the districts of India, increase in average rainfall indicates increase in forest-cover. It can be concluded that an increase in rainfall by 1mm across the country will cause 0.011% increase in forest-cover of India.

It is found that there is a varying relationship between forest cover and temperature. In most places of the country increase in temperature indicates decrease in forest cover. However in the Northern Plains, Northern Mountains and Western Plains it indicates increase in forest cover.

It was found that forest cover was affected the most by temperature. If there is an increase in temperature by 1 degree Celsius due to global warming and other factors then there is a possibility of 2.43% decrease of forest-cover in the country. This clearly indicates that global warming has adverse effects on forest cover and it will in turn lead to worse climatic change. It is believed that this project can be useful to protected area managers for forest conservation and thus help in sustainability.

ACKNOWLEDGEMENT

We are thankful to PES University and KAnOE for giving us this golden opportunity to do this project.

We use this opportunity to express our special thanks of gratitude to Dr. Kavi Mahesh, Professor, Knowledge Analytics and Ontological Engineering Centre, PES University, Bangalore. This project was the result of encouragement and guidance provided by him.

We would also like to thank all those people who provided us with their guidance and advice during the project work.

Thank You.