Report

CMIP data is used and updated by various institutes for research regarding climate change. The extension of these data files is .nc which is hard to open using usual apps. Our project focuses on extracting data from such files using python and also plot it on geographical map so as to touch basics of GIS and as supervised machine learning algorithms are trending topics, we have touched linear regression in our project. The concerned file contains wind speed data at 10m height from 1st Jan 1850 to 31st Dec 2014.

First using Anaconda prompt, one has to install 'netCDF4' library, then import that one. Using 'dataset' attribute the primary information regarding the downloaded .nc file was obtained. After, that the variables and their distribution dimensions were extracted with 'data.variables' commands. Here wind speed is distributed over the time, longitudes and latitudes. After, printing all the data we can understand the index of concerned latitude and longitude of concerned geographical region.

To plot the distribution on map again, 'Basemap' toolkit has to be downloaded from open - source websites and again the address of downloaded file has to be pasted in Anaconda prompt using pip to install it. After that, the lower left latitude and longitude as well as upper right latitude and longitude was specified to limit the map to the concerned region only. In the code concerned region is India. Then, the color-scheme, color bar was specified and map plotted for day 2 (index 1) in India. The plot looks like,

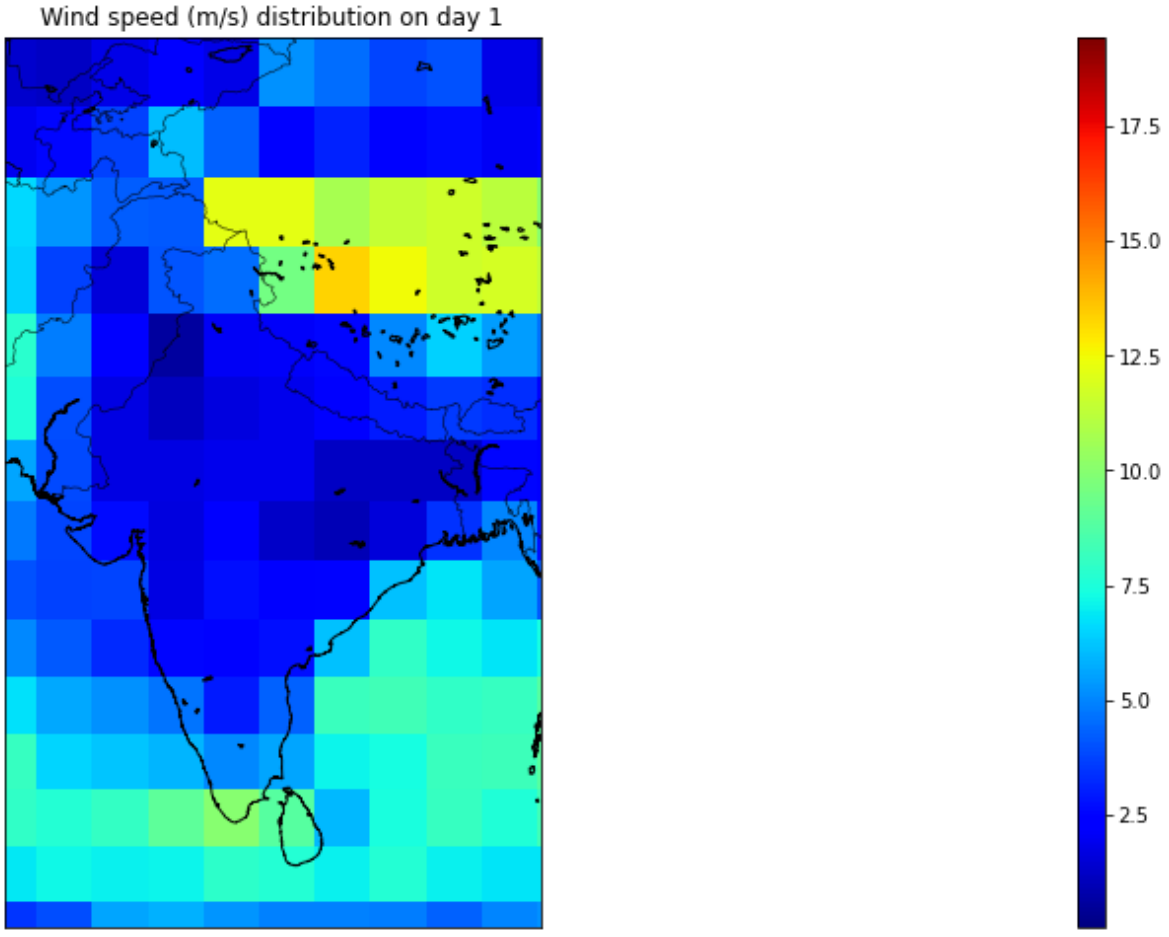

Fig.1 Wind speed distribution in India on 2nd Jan 1850

Now we try to fit a model between longitudes and wind speed for given day. The sklearn.linear_model lib is imported for 'LinearRegression'. The latitude considered is 12.55775611523069 decimal degrees corresponding to south India. As it is known that wind speed increases as we go from shallow to deep waters, longitudes consisting land and sea were taken and windspeed associated with it also were extracted. After fitting the data for linear regression, the obtained R-square value is 0.8726 which is good fit. Slope and intercept of that lines are 0.234 and -9.6369 respectively. Finally, the actual values from CMIP data and predicted values from regression are printed and plotted. The plot looks like follows:

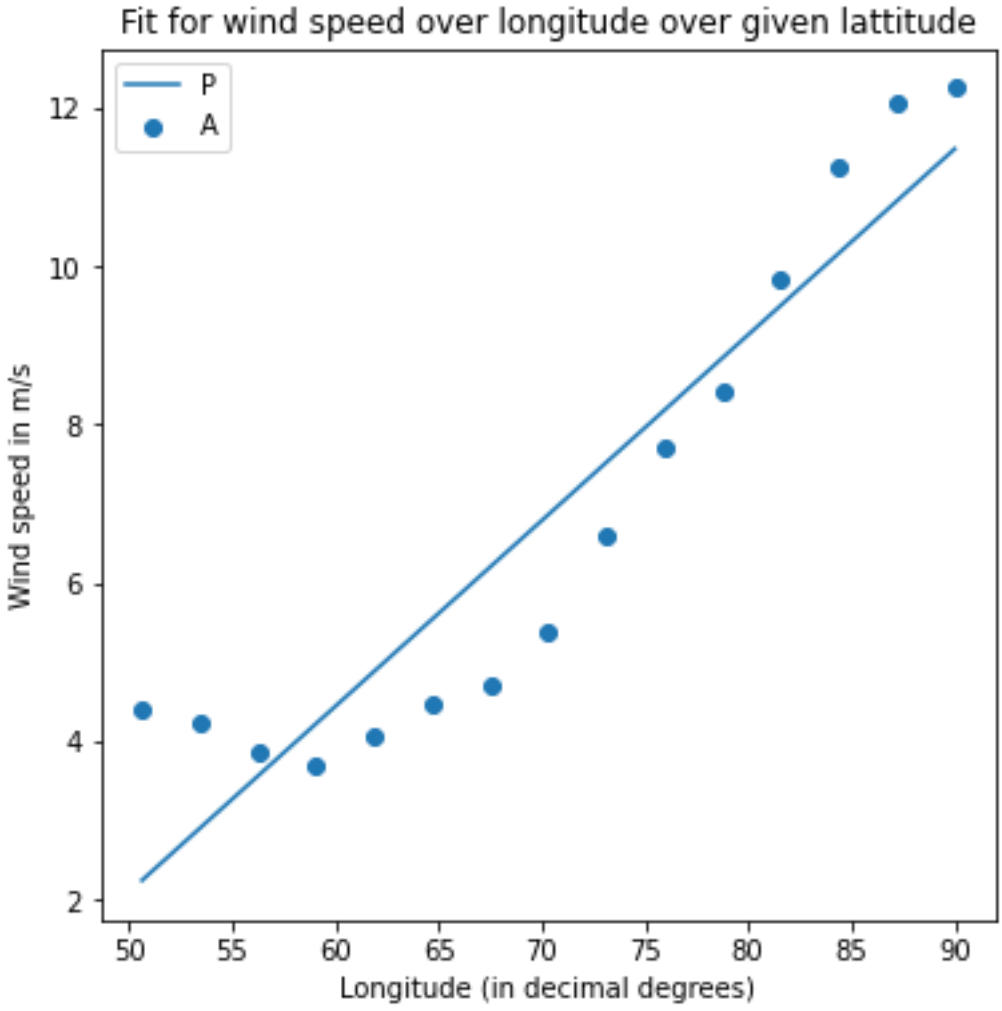

Fig. Regression model for wind speed and longitudes over given latitude.

So, we have studied the how to handle .nc files using python, extract the data and variables and also to plot it on geographical map for GIS analysis. Also, we have studied linear regression fit which can be used to obtain wind speed for intermediate longitudes and also further longitudes within the vicinity. The data presentation techniques using matplot.lib are explored along with basics of data extraction techniques. We can plot such maps for all the days. Further scope consists, monthly mean distribution of wind for various years. And timeseries animation using plotly lib can also be done.