**DETECTION OF GLOBAL WARMING IN SOUTHERN TEXAS**

**Diego Bass1**

M.E. 5013 – Big Data Analysis in Extreme Environments

The University of Texas at San Antonio

San Antonio, Texas, 78249

Global warming is defined as the gradual increase in the overall temperature of the earth’s atmosphere, dominantly caused by everyday pollution from warehouses, production plants, power plants, automobiles, etc. The rise in pollution output throughout the world has brought out great concern from environmentalists, however, there has been equal dispute over the topic of global warming. The scope of this project is to investigate whether or not global warming can be detected effecting southern Texas climate conditions. To conduct this study, data was collected pertaining to the following: 1) pollution measurements spanning from Corpus Christi, Laredo, Mission, and Brownsville, Tx, 2) measured solar irradiance in Edinburg, Tx (conducted by UT Pan American), and 3) measured avg. air temperatures in Edinburg, Tx. With these data sets, statistical analysis will develop average data vs. time plots, histograms with gaussian distributions overlayed, correlation analysis, linear regression analysis, and satellite images of current meteorological data. From this analysis, it was shown that the pollution data obtained (Ozone) carried a non-existent correlation to meteorological and solar irradiance data when taken in an over time total data analysis. However, when analyzing data from a yearly mean, Ozone (O3) contained a moderate correlation with average daily Solar radiation measured, and average air temperature measured. When plotted with respect to a yearly mean, the average daily temperature per year increased by 1.5 °C from an average of 23 °C in 2010 to 24.5 °C in 2020. Due to the detected gradual increase in average temperature over the course of ten years, along with its moderate correlation with ozone data, it can be concluded that global warming is present in southern Texas, as shown in the data collected. The cause of the gradual increase in overall temperature can not be directly linked to the quantity of ozone present. Since the correlation between pollution and temperature was not as strong, it would be recommended to investigate other pollutants to conclude if pollution has a great effect on Texas’ climate.

**I. Introduction**

Global warming detection in southern Texas can be determined by the use of the following collected datasets: 1) pollution measurements spanning from Corpus Christi [1], Laredo [2], Mission [3], and Brownsville [4], Tx, 2) measured solar irradiance in Edinburg, Tx (conducted by UT Pan American) [5], and 3) measured avg. air temperatures in Edinburg, Tx [6]. With the data collected, statistical analysis (such as yearly, monthly, and weekly mean) will develop average data vs. time plots, histograms with gaussian distributions overlayed, correlation analysis, linear regression analysis, and satellite images of current meteorological data.

**II. Methodology**

**Pollution Data**

Raw pollution data was downloaded in parts based upon time stamp and location. As separated csv (comma separated values) files, the data was combined into one large csv and converted into a pandas data frame. Characteristics of the data frame were found such as the size (145,287 rows x 11 columns) and info. The data frame will be grouped by location and their date and time. Each location will be analyzed for their yearly, monthly, and weekly average. Analysis will then be conducted on the data frame as a whole (including all locations) for its yearly, monthly, and weekly average.

**Solar Irradiance Data**

Raw data was collected and downloaded in parts based upon time period. The csv files were combined and converted into a pandas data frame. Characteristics of the dataset showed that the size to be 50,281 rows x 8 columns worth of data. Data included global horizontal, direct normal, diffused horizontal irradiance (measured in W/m­­­­2), average air temperature, and average wind speeds at the location of measurement (Edinburg, Tx). From inspecting the data, average air temperature and wind speed were missing considerable amount of data, therefore were excluded from analysis. The total data frame was separated into respective forms of irradiance, in which those datasets were analyzed for yearly, monthly, and weekly mean.

**Meteorological Data**

Similar to the above datasets, raw data was collected and downloaded in parts per year of a specific location (Edinburg, Tx). After combining the data into one csv file, the data was converted into a pandas data frame. Data was inspected and formatted properly for analysis. Inspection led to finding the size of the data frame (4,018 rows x 11 columns) containing information on maximum, minimum, mean daily temperature (°C), daily pressure (mm), and daily solar radiation (MJ/m2). The data frame was analyzed for each of its parameters over a yearly, monthly, and weekly average.

**Total Data**

Utilizing each dataset after proper formatting and initial analysis was performed, histograms, correlations, and linear regressions were conducted comparing the entire collection of data.

**III. Results**

**Pollution Data**

After proper formatting, data analysis included visualizing data. The data with no averaging can be displayed in figure A1, where it is noticeable that measurements were varied and dense. Figure A2 displays the data after taking averages of the data per week of each month and year. Drastic improvement is apparent in the visualization and readability of the data. The overall dataset when taking the mean per week of each month and year is shown in figure 1.

Chart, histogram

Description automatically generated

Figure 1. Overall dataset (mean)

**Solar Irradiance Data**

After formatting data into pandas, visual inspection of each parameter was performed before and after taking averages of the data per week of each month and year. The data are shown in figures A3 through A8. The data included average global horizontal, direct normal, and diffused horizontal irradiance measured in W/m2 taken from Edinburg, Tx.

**Meteorological Data**

Upon formatting data into pandas, raw data was averaged in similar fashion to the datasets above. This was done for consistency and ease of analyzing the datasets together when looking into correlations. The maximum, minimum and mean temperature can be shown in figure 2.

Chart

Description automatically generated

Figure 2. Average Temperature over time

Average solar radiation taken at the same time as the temperature data was also analyzed and formatted as shown in figure A9. As figure 2 can be indefinite to see data trends, yearly averages were taken for maximum and mean temperature, shown in figure 3.

Chart, line chart

Description automatically generated

Figure 3. Yearly mean of Temperatures

Displayed in figure 3, it is notable that from the year of 2010, there is an increasing trend of average temperature measured in Edinburg, Tx. From 2010 to 2020, there is an approximate increase of 1.5 °C in average daily temperature while there is an increase of 1.75 °C in average maximum temperatures. Both data sets contain a dominate increasing trend pattern.

**Total Data**

In order to visualize how the dataset of each parameter was distributed, probability density functions (histograms) were utilized. All interested parameters were analyzed, but figure 4 will only display data from ozone pollution.

Chart, histogram

Description automatically generated

Figure 4. Histogram of Ozone pollution

Based on the figure above, the gaussian distribution (normal distribution) outlined the data fairly with some points of data outside of the limits of the function. The calculated mean of the data was found to be 0.0252 ppm with a standard deviation of ± 0.0123 ppm. The following histogram analysis can be found within the python notebook. Notably the average daily temperature, figure A10, contained a mean value of 23.99 °C with a standard deviation of ± 6.56 °C.

Correlation analysis will determine if multiple data sets contain similar trends (whether the data increases or decreases at the same time). Figure 5 displays correlation amongst all interested parameters to display which datasets contain similar trend patterns.

Chart

Description automatically generated

Figure 5. Data Correlation over entire Time Series

The above figure outlines several important details: 1) Ozone quantity has no similarities to any of the parameters, 2) average global horizontal irradiance contains moderate correlation with daily solar radiation data, temperature data, and average direct normal irradiance data trends. Therefore, whenever each of those datasets increase or decrease, the other parameters also increase or decrease in a similar fashion.

Looking to organize data based upon yearly averages, figure 6 displays the new correlations analyzed between each of the datasets. With formatting the data within yearly means, data trends have now changed. Ozone (O3) now contains moderate correlation with measured daily solar radiation, and temperature. Average global horizontal irradiance contains similar amount of correlation with the previous parameters mentioned in figure 5. Therefore, on a yearly basis from 2015 to 2020, the amount of measured ozone contained similar trends to the measured amount of daily solar radiation, and daily and maximum temperatures.

Graphical user interface, chart

Description automatically generated

Figure 6. Yearly average Correlation Analysis

Seeing that ozone contains a possible correlation, regression analysis was to be performed to determine if a function can be fitted upon data comparisons. Weekly averages were taken to obtain reasonable amount of data points for regression analysis. Figure 7 displays the weekly average data of ozone vs. solar radiation measured.

Chart, scatter chart

Description automatically generated

Figure 7. Weekly comparison of Ozone vs. Solar Radiation

Displayed by figure 7, the there was no definite trend between ozone pollution data when compared to solar radiation data weekly averages. Yearly data does not contain enough data points to define any regression analysis. When plotting global horizontal weekly mean vs. weekly average temperature, there was significant in data trends. These trends are displayed with its respective linear fit in figure 8. The plot contained an R2 value of 0.86, therefore matching fairly well.

Chart, scatter chart

Description automatically generated

Figure 8. Comparison plot of Weekly mean of Global Horizontal vs. Temperature

Further analysis included utilization of Metpy to build satellite images of current meteorological climates in Texas. Unable to merge data that was extracted with the above analysis with the satellite images, the connection between the two methods of analysis were unsuccessful. The satellite images can be found in figure A11, which is a comprehensive plot containing satellite, rtma, and metar data.

**IV. Conclusions**

Global warming is defined as the gradual increase in the overall temperature of the earth’s atmosphere. Based upon the data evaluated above, it is notable that on a yearly basis, both the maximum and daily temperature measured increased from the year of 2010 to 2020. Based on the 10-year data trend, it is anticipated that the increase in temperatures will continue in the future. Based upon correlation data, since the amount of measured temperature is closely related to the amount of measured solar irradiance, it can also be anticipated that solar irradiance will be greatly influenced by the changes of temperature. With the trend of increasing temperatures in southern Texas is visible, by definition, global warming is present. However, the cause of the rising temperatures cannot be determined definitely. Within a yearly basis, the measured average of ozone contained moderate correlation to solar radiation as well as measured maximum and average temperatures. However, when ozone was plotted against the moderately correlated parameters, no possible regression could be made. Therefore, the pollution data that was obtained for southern Texas was not adequate for conclusive analysis. In order to obtain conclusive analysis in the future, data containing a variety of pollutants should be obtained over a wider time period, allowing more data to be evaluated within the correlation analysis and regression analysis. With more pollutant data, a conclusive result will be obtained to determine whether greenhouse emissions are the root cause of increasing temperatures in Texas.

**References**

[1] Corpus Christi. (n.d.). Retrieved May 11, 2021, from <https://openaq.org/#/location/1974>, <https://openaq.org/#/location/1973>

[2] Laredo. (n.d.). Retrieved May 11, 2021, from <https://openaq.org/#/location/1321>

[3] Mission. (n.d.). Retrieved May 11, 2021, from <https://openaq.org/#/location/1813>

[4] Brownsville. (n.d.). Retrieved May 11, 2021, from <https://openaq.org/#/location/833>

[5] Andreas, A. (n.d.). UTPASRL daily data. Retrieved May 11, 2021, from <https://midcdmz.nrel.gov/apps/daily.pl?site=UTPASRL&start=20110901&yr=2021&mo=4&dy=13>

[6] National Centers for Environmental Information (NCEI). (n.d.). Quality controlled datasets. Retrieved May 11, 2021, from <https://www.ncdc.noaa.gov/crn/qcdatasets.html>

**Appendix**

Chart

Description automatically generated

Figure A1. Sample of Raw Pollution Data

Chart, line chart

Description automatically generated

Figure A2. Mean Pollution Data (manipulated fig. A1)

Chart, histogram

Description automatically generated

Figure A3. Hourly Global Horizontal data (Edinburg, Tx)

Chart

Description automatically generated

Figure A4. Global Horizontal Data (mean)

Chart, histogram

Description automatically generated

Figure A5. Hourly Direct Normal data

Chart, bar chart, histogram

Description automatically generated

Figure A6. Direct Normal data (mean)

Chart, histogram

Description automatically generated

Figure A7. Hourly Diffuse Horizontal data

Chart, line chart

Description automatically generated

Figure A8. Diffuse Horizontal data (mean)

A picture containing text, needle

Description automatically generated

Figure A9. Solar Radiation (mean)

Chart, histogram

Description automatically generated

Figure A10. Histogram of Daily Temperature over time

A picture containing text

Description automatically generated

Figure A11. Comprehensive plot of current Texas climate