TRIED TPA-03 Report

Carl Robinson - 29th Oct 2017

V – 2nd Part : Temperature and CO2

0) Task, data and methods

Task: To conduct an initial data analysis of temperature and carbon dioxide (CO_2) measurements, which were taken as part of a climatological study in the south of France between 1950 and 2009. The relationship between these two variables will be explored.

Methods: The methods employed in this analysis include linear regression, correlation coefficients, and correction of global tendencies and offsets. Data is represented visually using line graphs and scatter plots.

Data:
The temperature data file consists of:

- 11 Columns
  - Year
  - Month
  - 9 temperature measurements in degrees celsius, one for each of nine geographical locations in the northern hemisphere, arranged in order from north to south (Reykjavik, Oslo, Paris, New York, Tunis, Alger, Beyrouth, Atlan27N40W, Dakar)
- 348 Rows
  - One month from Jan 1982 to Dec 2010 (29 years)

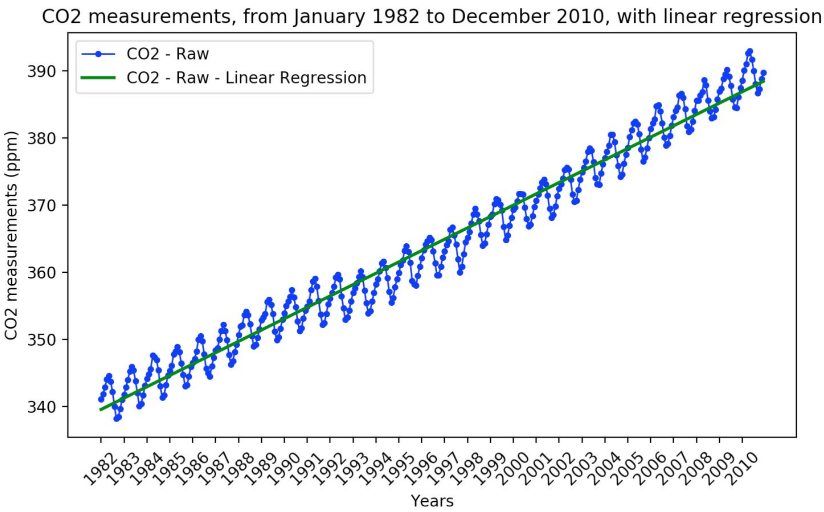
If we take the cities as instances, and the months as the variables, each column represents a chronological series of temperature data for a city.

The CO2 data file consists of:

- 3 Columns
  - Year
  - Month
  - CO2 measurement in parts per million (ppm) for a single geographical location, mount Mauna Loa in Hawaii
- 348 Rows
  - One month from Jan 1982 to Dec 2010 (29 years)

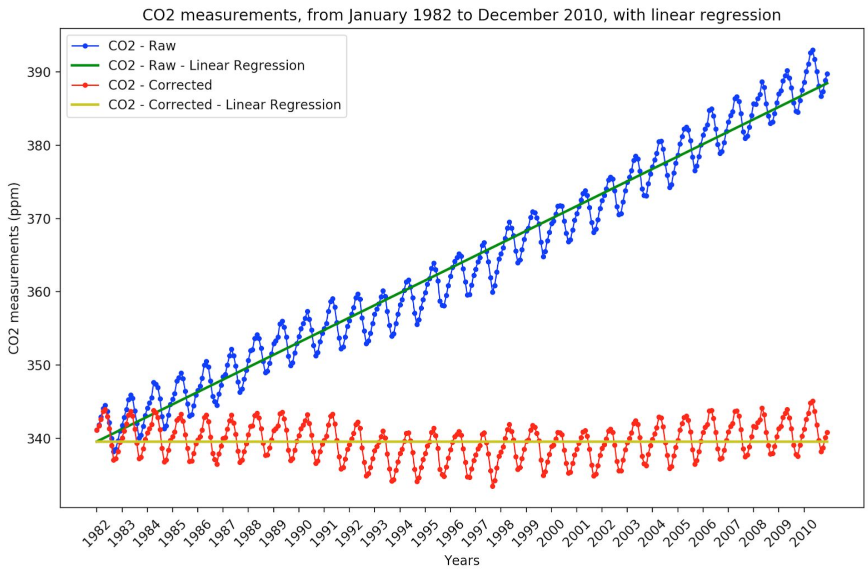
If we take mount Mauna Loa as a single instance, and the months as the variables, the column represents a chronological series of CO2 data for this location.

1) Raw C02 data



- There is significant seasonal variability in CO2 levels, as seen by the oscillating rise and fall each year.
- There is a clear global tendency for CO2 to increase over the years, as seen by the upwards trend from ~340 ppm in 1982 to ~390 ppm in 2010. This global tendency can is modeled by a linear regression line: $y = B1*x + B0$, where B0 = 339.431, B1 (gradient) = 1.690
- The global rate of change in CO2 seems to slow down between 1991 and 1995, as can be seen by the majority of CO2 values during this period falling beneath the linear regression line. However, the rate of change accelerates again after 1995.

2) Corrected C02 data



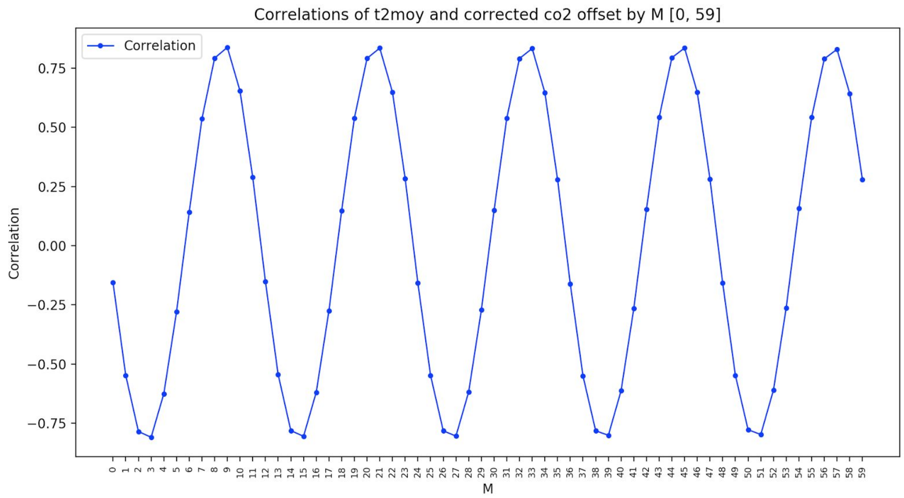
- The global tendency of CO2 to increase is much stronger than the global rise in temperature. It would dominate any comparisons made between CO2 and temperature, preventing any seasonal analysis.
- This global tendency must therefore be attenuated. To do this, we subtract the value predicted by the linear regression line from the raw CO2 data: $CO2\_corrected = CO2\_raw - (B1 * time\_in\_years)$
- The linear regression of the corrected data is flat (B1 = 0), as the global tendency has been eliminated.
- The seasonal variation of the corrected CO2 data can now be analysed alongside monthly temperature data.

3) Offset corrected C02 data

3.1) Correlations of average monthly temperature and corrected co2 offset by M [0, 59]

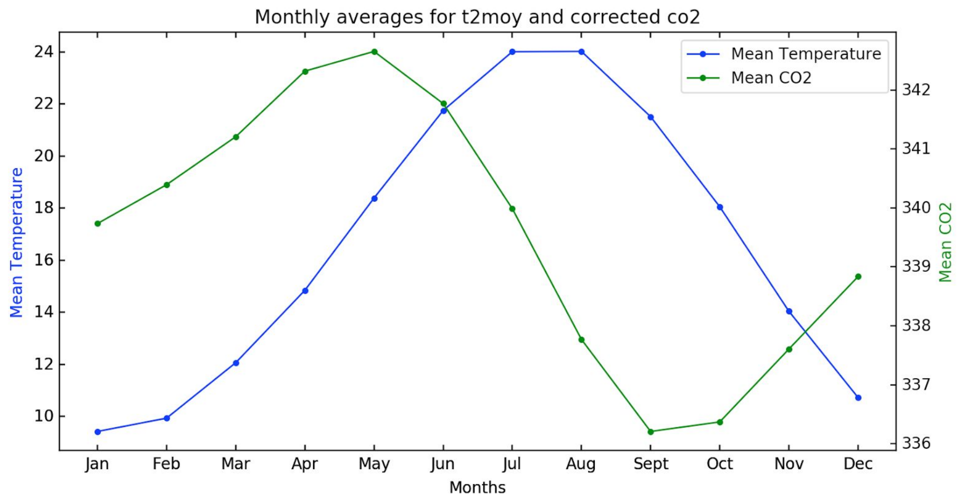
- As both CO2 concentration and temperature levels vary on a seasonal basis, it is important to determine whether there is any offset between them.
- We can use the correlation coefficient to measure the correlation between the two variables at each month, as this will tell us when temperature and CO2 values rise and fall together, and when they do not.

- A value of 1.0 indicates a perfect linear relation (the variables are proportional to one another, and rise/fall together in synchrony), a value of -1.0 indicates they are inversely related (where one rises and the other falls), and a value of 0 indicates that no relationship exists.
- We take the correlation coefficient of the two time series when offset in 60 monthly increments (M), and plot these against the offset in the graph above.



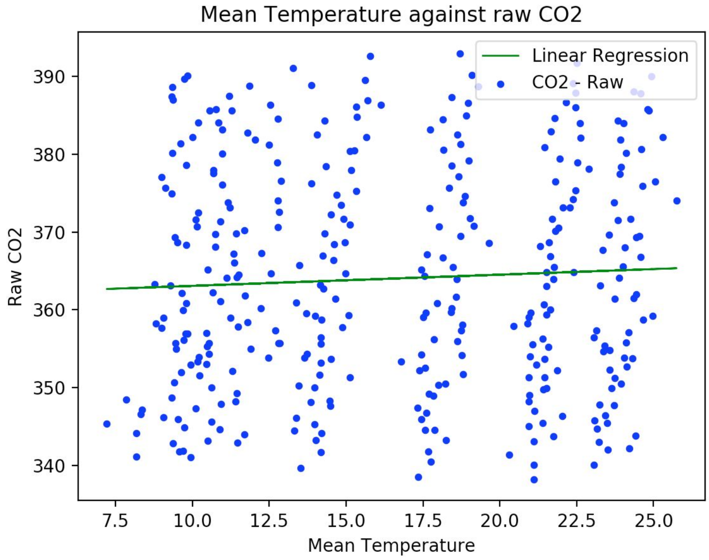
- We see that the correlation values oscillate significantly from highly correlated to anti-correlated, on a yearly cycle. The two variables are most anti-correlated when the correlation value is at its minimum value of -0.809, which occurs when M = {3, 15, 27, 39, 51}.
- Taking the modulo of these numbers with 12 (the number of months in a year) gives M = {3, 3, 3, 3, 3}. This shows that for all years, an **offset of 3 months** consistently produces the strongest anti-correlation.

3.2) Monthly average temperatures and corrected CO2 values



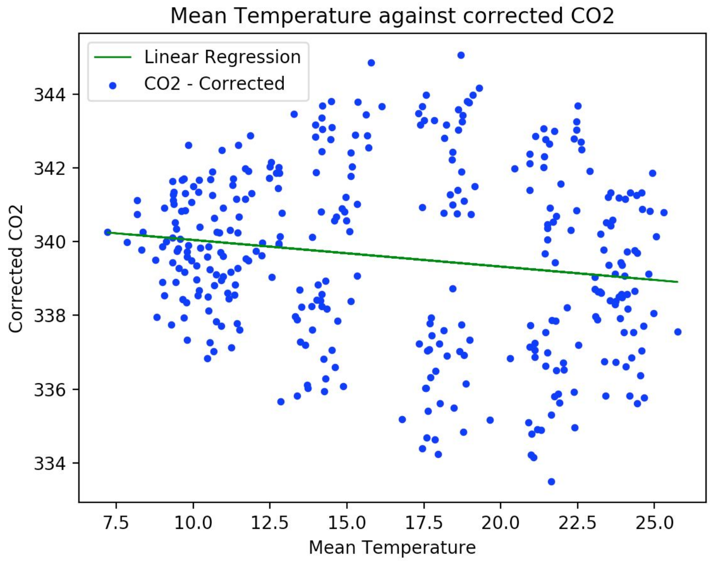
- This graph clearly shows the seasonal oscillations of both variables, and the 3 month offset between the temperature peak (highest point on the graph) and the CO2 peak.
- We can see that in winter (Jan/Feb) the temperature is low and the CO2 is rising, but as the temperature rises in Spring (Mar/Apr/May) the resulting vegetation growth consumes much of the CO2, leading to a sharp fall in CO2 the summer months after the peak in May. As the temperature starts to fall in September, the CO2 levels bottom out into the trough. However, the CO2 begins to rise again in October, due to much of the vegetation dying off in autumn.
- It can be concluded that the CO2 levels are governed by the temperature changes. There is a 3 month time-lag between a rise in temperature and the corresponding drop in CO2. The two variables can therefore be said to be anti-correlated.
- To achieve the greatest anti-correlation in the data, we just need to add 3 months to the green CO2 time series to shift it left by 3 months, so that the peak of one graph coincided with the trough of the other.

4) Scatter diagrams



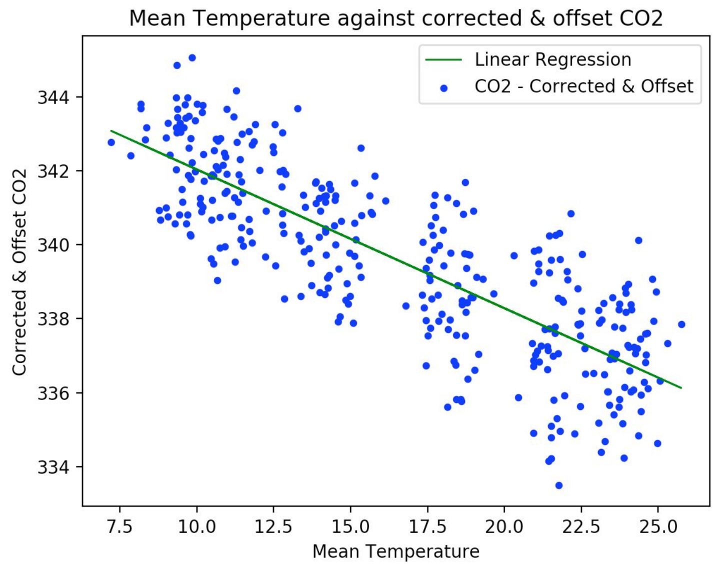
Raw CO2 / Mean Temperature

- Plotting the raw CO2 concentrations against the mean temperature values shows the points roughly arranged in columns. These columns correspond to the mean temperatures in each month.
- There is a huge variation in the CO2 concentrations for each of these columns, which dominates the graph. This makes any comparison with temperature impossible.
- The linear regression line must be considered unreliable, as it has a positive gradient (B1 = 0.144), whereas we would expect a negative gradient given the anti-correlated pattern seen in the previous graph.
- A very low correlation coefficient of 0.0527 confirms our intuition that this representation is unreliable, as we know the two variables are anti-correlated. This low correlation coefficient is as a result of the strong global tendency for CO2 to increase with respect to the mean temperature.



Corrected CO2 / Mean Temperature

- Correcting for the global tendency in CO2 increase reveals the seasonal variation that wish to analyse.
- The points on the graph are arranged horizontally by mean temperature as before, but now they are split vertically by CO2 concentration. The result is a loop of points corresponding to the months in the year.
- The split in CO2 values is most noticeable in the spring and autumn months which have moderate mean temperatures, highlighting the 3-month correlation offset between the two variables caused by the vegetation growth.
- The linear regression line now has negative gradient (B1 = -0.072), indicating a slight anti-correlation.
- The correlation coefficient is now also negative (-0.156), which is consistent with the anti-correlated pattern we detected earlier.



Corrected & Offset CO2 / Mean Temperature

- Applying the 3-month offset to the data eliminates the CO2 response time-lag, which aligns the phase of the two oscillations and thereby maximises the anti-correlation between the variables.
- The pattern of points now displays an inverse linear relationship, and the linear regression line has a strong negative gradient (B1 = -0.375).
- The large negative correlation coefficient of -0.809 confirms that the two variables are strongly anti-correlated.

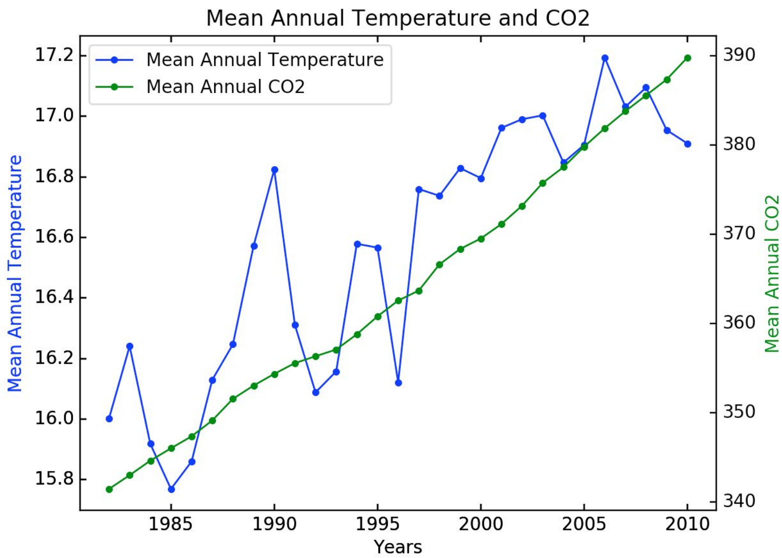
The results can be summarised as follows:

	B0	B1 (Gradient)	Correlation Coefficient
Raw CO2	361.629	0.144	0.0527
Corrected CO2	340.770	-0.072	-0.156
Corrected & Offset CO2	345.792	-0.375	-0.809

- This table highlights the importance of data preprocessing. The raw data seemed to indicate very little correlation between variables due to both the global tendency to increase, and the time-lag phenomenon. The corrected data improved on this, but still suffered from the time-lag phenomenon.
- By correcting and offsetting the data we finally obtained a representation suitable for comparing the seasonal variation of CO2 with the mean monthly temperature.

5) Comparison of temperature change with CO2 change

- To analyse how the global change in mean temperature and CO2 are related, we need to eliminate the seasonal variation. We do this by calculating a single mean temperature value for each year across all 9 cities, and a single mean CO2 value for each year. This gives us one temperature value and one CO2 value for each of the 29 years between 1982 and 2010.
- The two time-series follow each other closely when plotted on a graph, illustrating that the two variables are strongly correlated on a global level. A large positive correlation coefficient of 0.863 supports this finding.



- It can be noted that mean temperature is highly variable from one year to the next, while mean CO2 increases every steadily year after year.

Conclusions

- On a seasonal level, the mean monthly temperatures and monthly CO2 concentrations are highly anti-correlated. The CO2 level is offset by 3-months from the seasonal temperature changes which cause it to rise/fall, and has a strong global tendency to increase each year, both of which needed to be eliminated in order to discover its true relationship mean temperature.
- On a global level, the mean yearly temperatures and yearly CO2 concentrations are highly correlated, rising together. This is consistent with our understanding of the effects of climate change.
- Preprocessing of data is crucial if the correct relationships between variables are to be found.