An Examination of the Spatial Distribution of Carbon Dioxide and Systematic Errors

Brennan Coffey, Mike Gunson, Christian Frankenberg and Greg Osterman
Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Drive, Pasadena, CA, USA 91109-8099

Abstract:

The industrial period and modern age is characterized by combustion of coal, oil, and natural gas for primary energy and transportation leading to rising levels of atmospheric of CO_2[2]. This increase, which is being carefully measured, has ramifications throughout the biological world. Through remote sensing, it is possible to measure how many molecules of CO_2 lie in a defined column of air[1]. However, other gases and particles are present in the atmosphere, such as aerosols and water, which make such measurements more complicated[1]. Understanding the detailed geometry and path length of the observation is vital to computing the concentration of CO_2[4]. Comparing these satellite readings with ground-truth data (TCCON) the systematic errors arising from these sources can be assessed. Once the error is understood, it can be scaled for in the retrieval algorithms to create a set of data, which is closer to the TCCON measurements[1]. Using this process, the algorithms are being developed to reduce bias, within .1% worldwide of the true value. At this stage, the accuracy is within 1%, but through correcting small errors contained in the algorithms, such as accounting for the scattering of sunlight, the desired accuracy can be achieved[2].

Introduction:

Carbon Dioxide is a vital piece to nearly every process in nature[5]. It is created naturally with volcanic eruptions, and is the primary source of energy for organisms such as plants, which use it in the Citric Acid Cycle to produce glucose. This flow of events indicates that there is a cycle of carbon dioxide namely sources and sinks[2]. Methods of removing carbon dioxide from the atmosphere are known as sinks: the largest contributors to the sinks in the carbon dioxide cycle are plants, and oceans[8]. In addition to those sinks, sources contribute to the equilibrium state by producing the gas; these sources are most commonly known to be the oceans (in a release phase) and plants again[8]. The earth works in cycles and both of these processes add to, and remove from the total carbon dioxide concentration on a global scale. On a greater time scale, the cycles of the earth have been much higher in the past, the last time the levels of carbon dioxide were seen as high as 380 ppm globally was around 15 million years ago (Fig. 1). Data also indicates that the earth has been through many cycles in the past 400 thousand years, global ranges were typically from ~180-300 ppm[5]. From this data it would be reasonable to extrapolate that in the time frame we are in now should be a cooling period of falling levels of CO_2, however that is not the case. Since the industrial revolution and the advent of burning fossil fuels for energy, global concentrations of CO_2 have risen steadily to record levels within a limited time frame[4]. The ramifications of having such a substantial change could be terrible; but we don't know how the environment will react[8]. Our earth, and all of its gases, lives in a tight equilibrium, so tight that we know the mixing ratios of several species perfectly; if the equilibrium were to shift different ratios would be established and that change in reflective properties of the greenhouse gas layer and the water vapor there could cause global temperature shifts and wreak havoc on all ecological and biological systems alike[8].

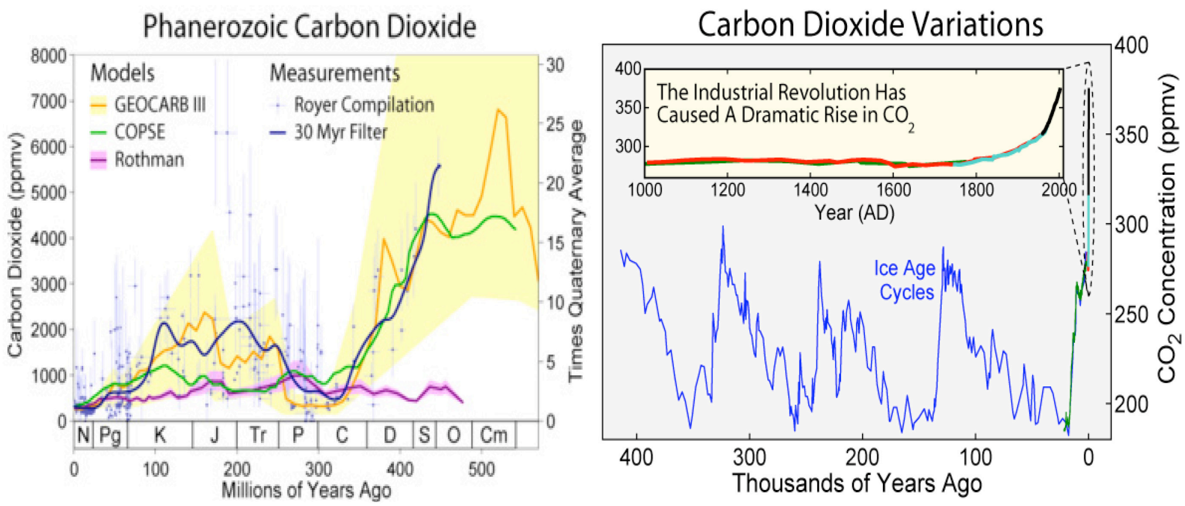


Fig. 1 Left shows the variability of CO_2 as far back as 600 million years ago. The right graph shows the cycles of rising and falling levels of CO_2 in the atmosphere as far back as 450 thousand years ago and suggests a possible correlation to the point that breaks the maximum value of ~300 ppm, as indicated, the trend should be heading downwards and buffer at ~180-300 ppm.

The carbon dioxide cycle (Fig. 2) is the process by which carbon dioxide is recycled in the atmosphere. Natural processes, on an annual basis, release on the order of 770 billion tons of carbon dioxide into the atmosphere. This is to be compared against the 30 billion tons, or 30 gigatons that humans contribute to the atmosphere of CO_2. Nature is excellent when it comes to absorbing the CO_2 in the atmosphere, in fact most of the readings that we have now indicate that the natural processes absorb all of the natural production and close to half of the part that humans have contributed[6]. While humans only contribute about 4% of the total CO_2 in the atmosphere, small changes on this order can have terrible biological and ecological ramifications[2].

The Greenhouse Effect is one of the most common cited effects of higher carbon dioxide concentrations. When UV radiation enters the atmosphere from the sun, some of the radiation is reflected by clouds, and some more is reflected by other particles in the air. However, some of this radiation reaches the ground and the radiation that isn't absorbed by the earth is reflected back to space. The absorbance of the atmosphere is directly proportional to the amount of light that is reflected back to earth. The light continues on this pattern because of the water vapor concentration in the atmosphere. One important fact to understand is that the earth sits in equilibrium, and a slight concentration increase in one of the constituent gases in the atmosphere can cause a shift in some of the others. Higher concentrations of Carbon dioxide could cause a shift in the concentration and ratios of water vapor, which would ultimately facilitate any major ecological changes[8]. Through measuring how the carbon dioxide changes in the atmosphere scientists can hopefully understand more about how these physical processes work, and the extent to which humanity is influencing them[1].

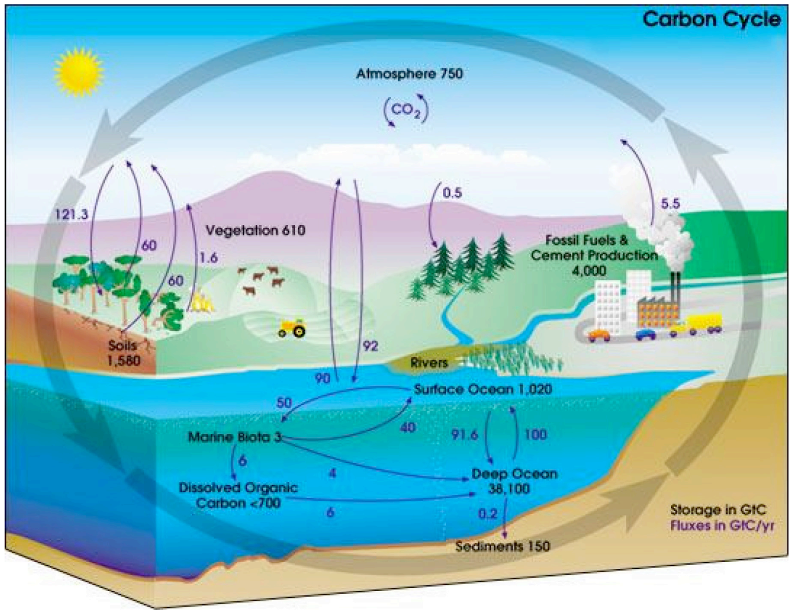


Fig. 2 The Carbon Dioxide Cycle is shown here in addition to relative quantities, in gigatons, of the amount of carbon dioxide emitted per each process.

OCO-2

The mission OCO-2 (Orbiting Carbon Observatory) will address this problem through its space-based measurements, which have the precision and coverage to characterize the spatial and temporal distribution of CO_2[1]. It travels in a 16- day ground tracking repeat orbit, and at 705km altitude, which provides a complete coverage of the sunlit side of the globe[1]. Spectroscopic measurements will be taken of the O_2 column abundance as well as the CO_2 column abundance[1]. Then the abundances will be analyzed in order to determine the mole fraction of CO_2 (X_{CO2})[2]. To test the accuracy of these measurements, we will be comparing them against a ground-based network of readings called the Total Carbon Column Observing Network (TCCON)[6]. TCCON ground stations are static nodes, which read CO_2 levels in their region and then are validated through comparisons between aircraft readings of CO_2 concentration[3]. We will be comparing the values obtained with the retrieval algorithm in the OCO instrument against the ground base TCCON measurement to obtain an accuracy of .1%[2]. This accuracy is vital to understanding the sources and sinks of CO_2 because surface flux of CO_2 is known to vary by no more than ~2% on a continental scale (1000km x 1000km)[3]. Knowing the regional change in CO_2 is between 1 and 2 ppm could improve our understanding of the flux of CO_2. At least one annual cycle is needed to get an idea of what the season variability of CO_2 is[3]. OCO-2 is scheduled for a launch in February 2014.

Measurement Approach

To collect representative values of the X_{CO2}, the OCO-2 instrument will read the reflected light from the earth's surface. Certain molecules have very specific spectral signatures, so when the instrument reads the reflected light, the relative intensity of the reflected light will contain information about how much of a certain molecule is in a dry mole fraction of air[3]. The OCO-2 instrument will measure this intensity to determine the absorbance of light within the column, and from that will be able to determine the abundance of molecules in the measurement region.

The measurement approach is heavily dependent on the path length of the reading; however, there are many variables that must be taken into account in order to get an accurate measurement. Inhomogeneous surface features make the measurements difficult, as the absorption will change based on whether the reading is taken over land or water, forest or desert. In addition to this, some geological features cause high concentrations of optically dense particles called aerosols to obstruct the lights path; it becomes even more complicated when considering the earth's rough surface[6]. These facts coupled with cloud interference make it difficult to get a true reading every time. In order to account for these errors, the OCO-2 instrument utilizes the varying absorption spectra of these gases to get valid readings of CO_2[3]. Its spectra are near the infrared and the varieties of gases are used to measure the amount of Oxygen in the atmosphere and then interpolate the X_{CCO2} from the acquired path length[3]. The OCO-2 instrument uses three different wavelengths selected specifically for their properties of absorbance in their relative parts of the spectrum[3]. The three wavelengths are nominally .765 µm, 1.61 µm and 2.06 µm. The O_2 A-band wavelength is .76 micron, the weak CO_2 wavelength is 1.61 micron and the strong CO_2 wavelength is 2.06 micron. The purpose behind the two different wavelengths to read the CO_2 is just one of the problems that come with taking spectral data of CO_2: over cities, there are other particles, generally large ones, which obstruct and diffract light so we cannot get an accurate reading of the X_{CO2}. The 1.61 micron channel is absorbed readily by CO_2 and optically thick aerosols and clouds so we cannot get any CO_2 data with that alone, but while the 2.06 micron channel, as it is very sensitive to aerosols in the atmosphere, allows scientists to see the absorption of the aerosols is in that column of air is and correct for it in the weak CO_2 channel[6]. Using a combination of this data, the correct X_{CO2} can be calculated. The sensitivity is to a 3km x 3km region on the surface of the earth, this allows the general spatial distribution to be measured rather than being too precise and accounting for sporadic changes.

The OCO-2 Science Validation focused primarily on finding methods to characterize these changes and apply appropriate filters to get as close as possible to the .1% accuracy of the true values[2]. To figure out which variables are correlated with the CO_2 reading, there were several approaches used. The data was first filtered based on mission specifications, and then graphed on a global map to display the regional variability of the readings. The purpose of this method is to see where there are valid data points, on a trend, how high the readings generally are, and what the global variability looks like (Fig. 3).

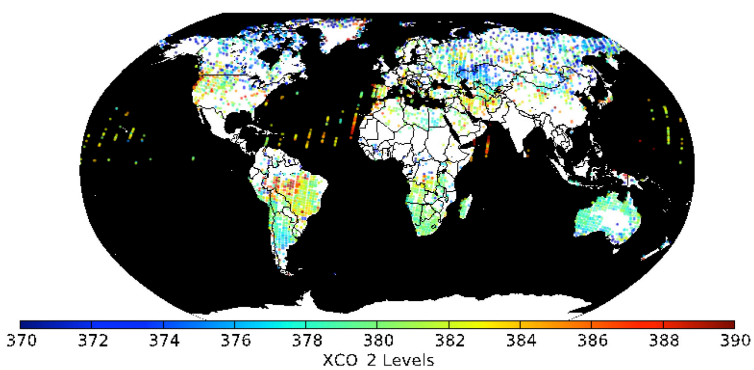


Fig. 3 Shows satellite soundings and the corresponding value of the CO_2. This allows us to see the total distribution of the CO_2 and where there is rejected data, which can be analyzed, further and corrected for.

The extent of the experimentation involved was to generate graphs and analysis tools, which held the data in varying views. As the comparison between CO_2 is all between ground truth stations and satellite readings, it is important to understand the value differences between those two stations and the variables that are correlated with that difference. To assess possible correlations with the GOSAT/ACOS (Global Greenhouse Gas Observation by Satellite) data and the OCO-2 retrieval algorithm, analysis tools were written to compare the change in the readings and compare the difference to a Gaussian distribution. In addition to this statistical analysis, a program was written which used techniques of a multivariate analysis linear regression to attempt to explain any possible correlation between the change in CO_2 and other variables (Fig. 4). See Appendix A for more analyses.

Dependent Variable: Delta XCO_2
Method: Least Squares
Date: Tue, 26 Jul 2011
Time: 11:54:04
obs: 235
variables: 7

variable	coefficient	std. Error	t-statistic	prob.
const	-3.392830	1.997043	-1.698927	0.090697
l2_b28.sza	-0.000069	0.014273	-0.004840	0.996143
l2_b28.aod_total	71.745406	17.512270	4.096865	0.000058
l2_b28.chi2_1	1.980813	1.716121	1.154238	0.249611
l2_b28.aod_low	-236.088732	29.753289	-7.934878	0.000000
l2_b28.aod_high	-106.037324	33.676590	-3.148695	0.001859
l2_b28.lza	-0.047236	0.019511	-2.421021	0.016259

Models stats		Residual stats	
R-squared	0.360408	Durbin-Watson stat	1.442144
Adjusted R-squared	0.343577	Omnibus stat	3.452984
F-statistic	21.412882	Prob(Omnibus stat)	0.177907
Prob (F-statistic)	0.000000	JB stat	3.217943
Log likelihood	-454.337787	Prob(JB)	0.200093
AIC criterion	3.926279	Skew	0.283937
BIC criterion	4.029331	Kurtosis	3.078489

Multivariate Regression for: 'Lamont'

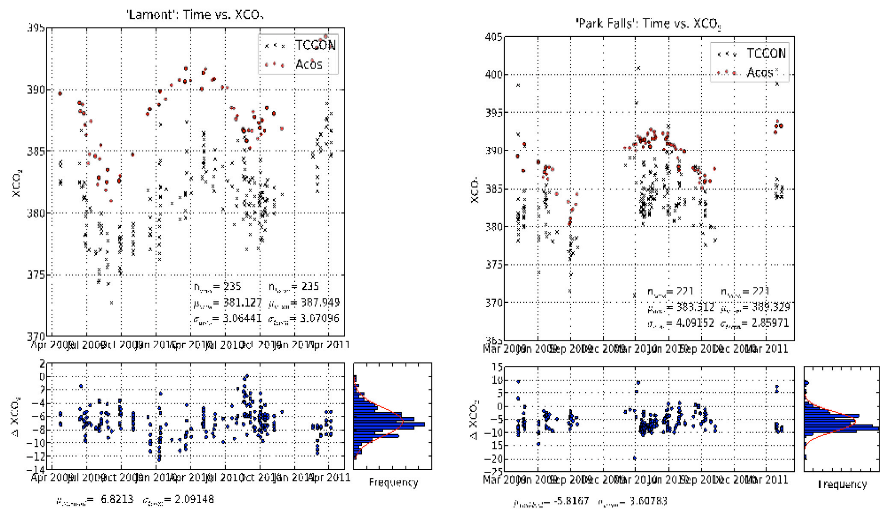


Fig. 4 shows several of the comparisons between TCCON ground stations and ACOS readings. The variability of each graph is indicated by the statistics computed below the lower ΔXCO_2 scatter graph and the histogram with a Gaussian distribution overlaid. The Multivariate regression is also calculated per each data set and graphs possible correlations between ΔXCO_2 and other confounded variables

In addition to creating the tools to analyze the data as shown in Fig. 3, other methods were looked into in order to make sure that the data could be reached reasonably. To create a pool of data around one station, it is extremely expensive from a fiscal and temporal perspective as scanning through many thousand soundings takes a very long time. The technique implemented was querying a database directly which, through specific commands, sorts and filters all of the data around one specific location. This amendment to the analysis aspect saves on the order of one day per graph and thus allows the validation team to access more data, faster. Using a database access system also allows the temporal distribution to be more easily viewed as opening multiple files and reading them takes a long time. A filter can be put in place to only use data points whose timestamp is within a specified range of the TCCON measurement. This point brings up another advantage to using this system: the TCCON measurements are averaged, which provides far more accurate X_{CO2} readings, and there is no interpolation or extrapolation involved as the TCCON and ACOS data points are matched by sounding id's. These techniques are mainly used to evaluate the magnitude of the bias introduced by confounding of a variety of variables. Once, they have been identified, other genetic algorithms can be used to determine how and if they should be scaled.

Using some of these techniques described, with the help of others written by the validation team, the initial error of the retrieval algorithm was ~5%, but that has been reduced to 1%. However, this does not meet the mission specifications, and more systematic error issues are being resolved to bring that error down to the target level of 0.1%. Currently the retrieval algorithms are consistently ~6 ppm below the TCCON values. Other methods of fixing this error then fall to coding and random errors within the code. For example, to reduce some error, the curvature of the earth had to be taken into account for determining relativistic times with respect to nadir readings. This was not accounted for in the original algorithm and the solar zenith angle was consistently off. Correcting this small error is just one of the ways in which the validation team is working on retrieving the best data possible.

References

[1] Crisp, David, Charles E. Miller, and Philip L. DeCola. *NASA Orbiting Carbon Observatory: Measuring the Column Averaged Carbon Dioxide Mole Fraction from Space*. Tech. no. 023508. Vol. 2. Journal of Applied Remote Sensing. Print.

[2] Crisp, David. *OCO-2 Overview*. 14 July 2011. PPT.

[3] Crisp, David. *The Orbiting Carbon Observatory: NASA's First Dedicated Carbon Dioxide Mission*. Rep. 2008. Print.

[4] *OCO*. Web. 02 Aug. 2011. <http://oco.jpl.nasa.gov/observatory/>.

[5] *OCO*. Web. 02 Aug. 2011. <http://oco.jpl.nasa.gov/science/>.

[6] *OCO*. Web. 02 Aug. 2011. <http://oco.jpl.nasa.gov/science/Validation/>.

[7] Riebeek, Holli. "The Carbon Cycle : Feature Articles." *NASA Earth Observatory : Home*. Web. 02 Aug. 2011. <http://earthobservatory.nasa.gov/Features/CarbonCycle/>.

[8] "Greenhouse Gas Emissions | Climate Change | U.S. EPA." *US Environmental Protection Agency*. Web. 04 Aug. 2011. <http://www.epa.gov/climatechange/emissions/index.html>.

Acknowledgements

This work was conducted at the Jet Propulsion Laboratory of the California Institute of Technology under contract from the National Aeronautical and Space Administration.

Appendix A

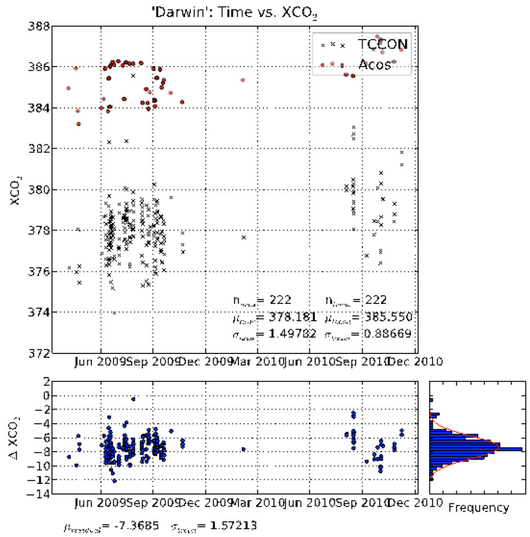
Darwin Data:

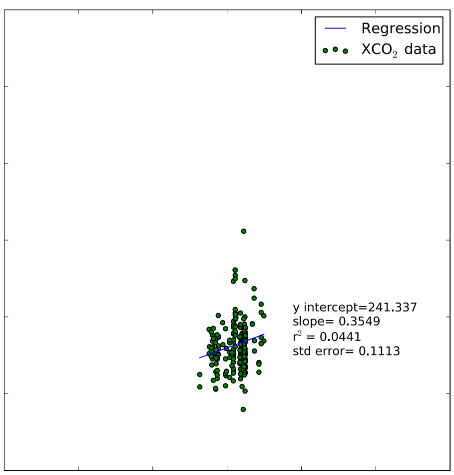
Dependent Variable: Delta XCO_2
Method: Least Squares
Date: Tue, 26 Jul 2011
Time: 15:00:00
obs: 222
variables: 7

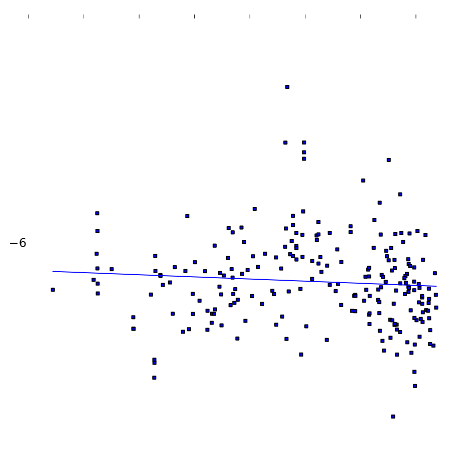
variable	coefficient	std. Error	t-statistic	prob.
const	-11.606268	2.078216	-5.584728	0.000000
l2_b28.sza	-0.014096	0.014209	-0.992022	0.322302
l2_b28.aod_total	-42.409016	21.219680	-1.998570	0.046914
l2_b28.chi2_1	7.227347	1.617477	4.468283	0.000013
l2_b28.aod_low	17.632472	39.322841	0.448403	0.654314
l2_b28.aod_high	28.192298	30.875182	0.913105	0.362210
l2_b28.lza	-0.001482	0.010779	-0.137444	0.890809

Models stats		Residual stats	
R-squared	0.151240	Durbin-Watson stat	1.272398
Adjusted R-squared	0.127553	Omnibus stat	4.265252
F-statistic	6.385104	Prob(Omnibus stat)	0.118526
Prob (F-statistic)	0.000003	JB stat	4.712210
Log likelihood	-397.242884	Prob(JB)	0.094789
AIC criterion	3.641828	Skew	0.158141
BIC criterion	3.749119	Kurtosis	3.639838

Multivariate Regression for: 'Darwin'







This set of graphs shows the output of the multivariate linear regression. These plots were made with the goal of evaluating the magnitude of confounding variables.

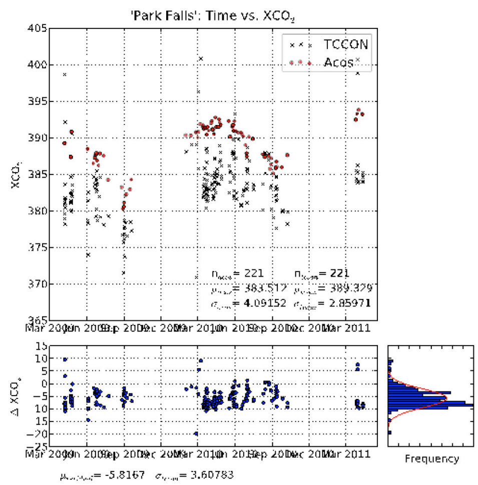
Park Falls:

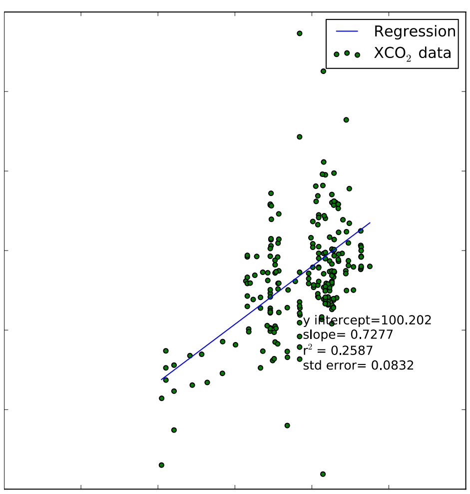
Dependent Variable: Delta XCO_2
Method: Least Squares
Date: Tue, 26 Jul 2011
Time: 15:38:29
obs: 221
variables: 7

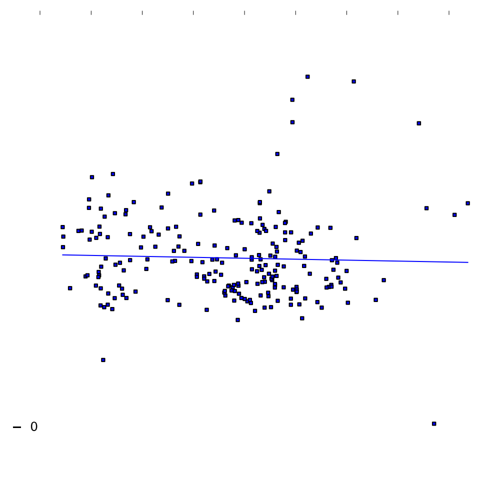
variable	coefficient	std. Error	t-statistic	prob.
const	-5.890139	3.035835	-1.940204	0.053669
l2_b28.sza	0.076233	0.027034	2.819895	0.005255
l2_b28.aod_total	82.341650	18.617458	4.422819	0.000016
l2_b28.chi2_1	1.759884	2.576925	0.682940	0.495384
l2_b28.aod_low	-277.929643	28.657944	-9.698171	0.000000
l2_b28.aod_high	-85.796732	33.255477	-2.579928	0.010551
l2_b28.lza	0.017274	0.023055	0.749250	0.454529

Models stats		Residual stats	
R-squared	0.384337	Durbin-Watson stat	1.494156
Adjusted R-squared	0.367075	Omnibus stat	45.301428
F-statistic	22.265438	Prob(Omnibus stat)	0.000000
Prob (F-statistic)	0.000000	JB stat	470.997198
Log likelihood	-543.553766	Prob(JB)	0.000000
AIC criterion	4.982387	Skew	-0.322265
BIC criterion	5.090021	Kurtosis	10.122752

Multivariate Regression for: 'Park Falls'







This set of graphs shows the output of the multivariate linear regression. These plots were made with the goal of evaluating the magnitude of confounding variables at different ranges.