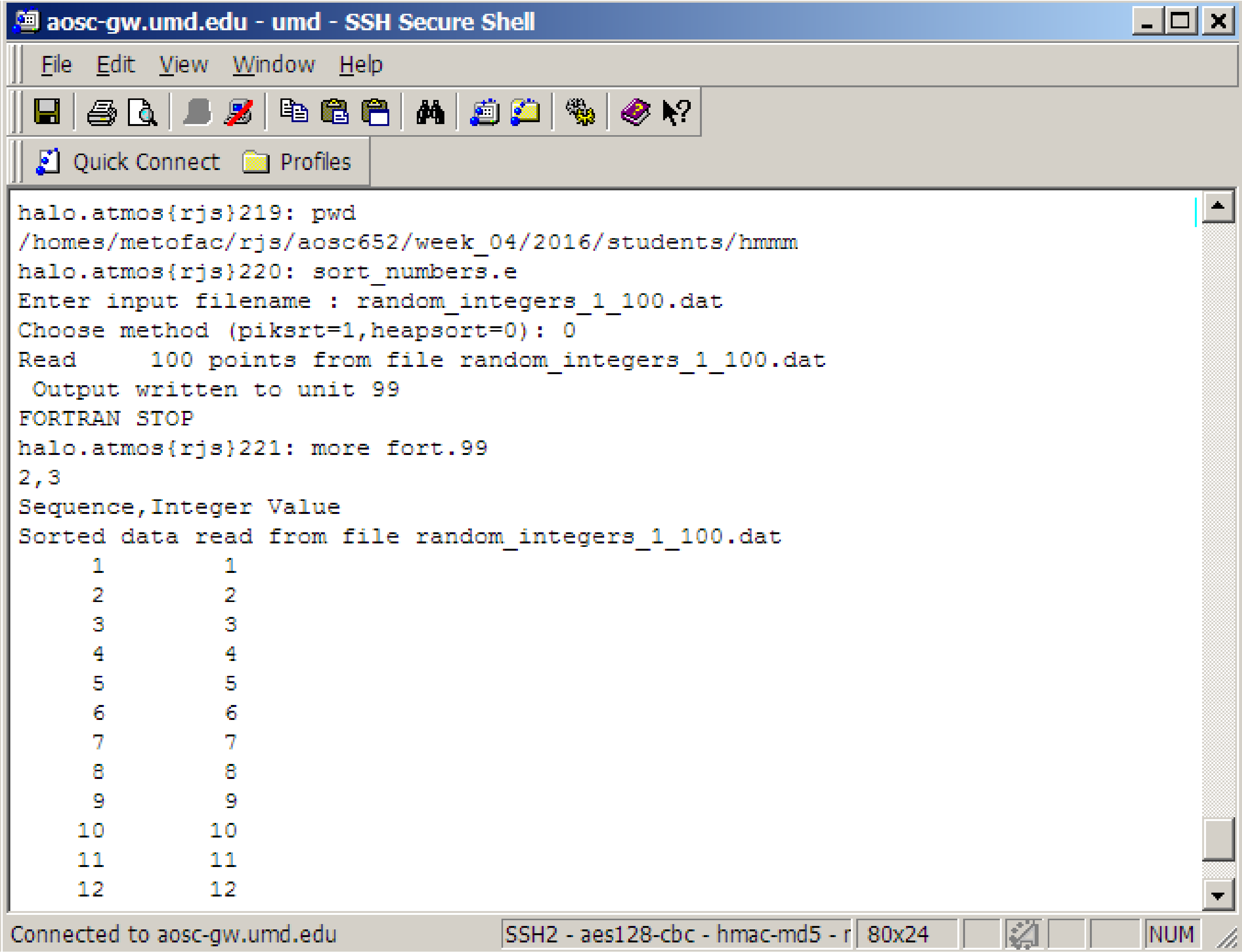
Analysis Methods in Atmospheric and Oceanic Science

AOSC 652

Least Squares Analysis, Statistical Regression, and Spline Fitting: Day 2

28 Sep 2016

AOSC 652: Analysis Methods in AOSC



AOSC 652: Analysis Methods in AOSC

aosc-gw.umd.edu - umd - SSH Secure Shell

File Edit View Window Help

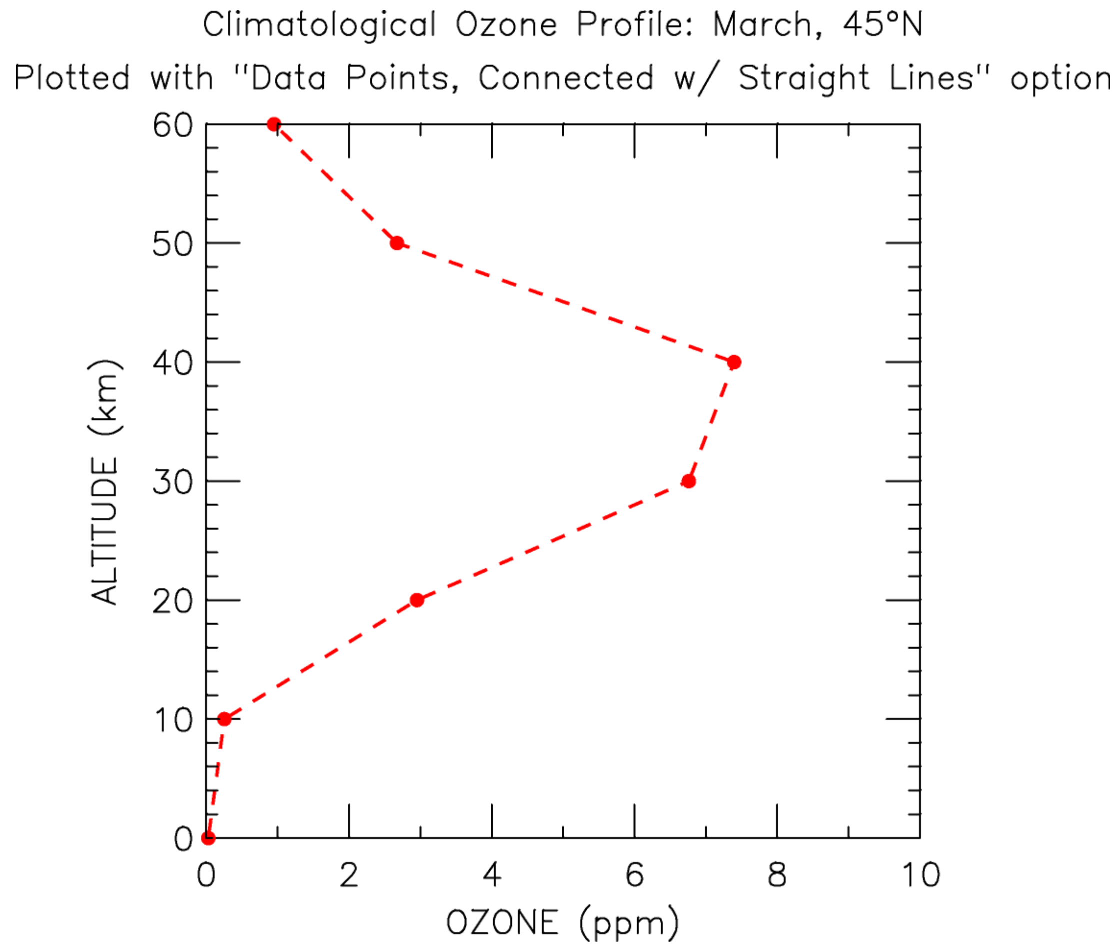
Quick Connect Profiles

```
metosrv8.umd.edu{rjs}186: pwd
/homes/metofac/rjs/aosc652/week_04/2016/students/hmmm
metosrv8.umd.edu{rjs}187: sort_numbers.e
Enter input filename : random_integers_1_100.dat
Choose method (piksrt=1,heapsort=0): 0
Read     100 points from file random_integers_1_100.dat
 Output written to unit 99
FORTRAN STOP
metosrv8.umd.edu{rjs}188: more fort.99
2,3
Sequence,Integer Value
Sorted data read from file random_integers_1_100.dat
     1        31
     2        52
     3        98
     4        15
     5        97
     6        58
     7        34
     8        88
     9        44
    10        35
    11        95
    12        55
```

Connected to aosc-gw.umd.edu | SSH2 - aes128-cbc - hmac-md5 - r | 80x24 | NUM

AOSC 652: Analysis Methods in AOSC

If you were to copy file ~rjs/aosc652/week_05/model_atmosphere.march_45N.dat and plot column 5 (O_3 mixing ratio) vs column 1 (altitude), using the "Data Points, Connected w/ Straight Lines" option, you would get:



Plot made using file ~rjs/aosc652/week_05/stncl.250

Cubic spline fitting:

Every ***two adjacent data points*** have an interpolation given by:

$$f_i(x) = a_i + b_i\, x + c_i\, x^2 + d_i\, x^3 \qquad \text{for } i\text{=1 to N}-1$$

To determine a_i , b_i , c_i , and d_i :

1. Splines must pass through data points
2. Splines must be continuous at interior data points (condition met by use of first derivative)
3. First derivative must by continuous for interior data points (condition met by use of second derivative)
4. To solve for remaining unknowns: section derivative = 0 at first and last data points

Higher order splines can be developed. However, computational difficulty greatly increases by the need to solve larger systems of eqns. Typically gained benefit (visual or physical) is minimal for use of higher order splines.

Analysis Methods for Engineers, Ayyub & McCuen

Cubic spline fitting:

Every ***two adjacent data points*** have an interpolation given by:

$$f_i(x) = a_i + b_i\,x + c_i\,x^2 + d_i\,x^3 \qquad \text{for } i\text{=1 to N−1}$$

To determine a_i, b_i, c_i, and d_i:

1. Splines must pass through data points
2. Splines must be continuous at interior data points (condition met by use of first derivative)
3. First derivative must by continuous for interior data points (condition met by use of second derivative)
4. To solve for remaining unknowns: section derivative = 0 at first and last data points

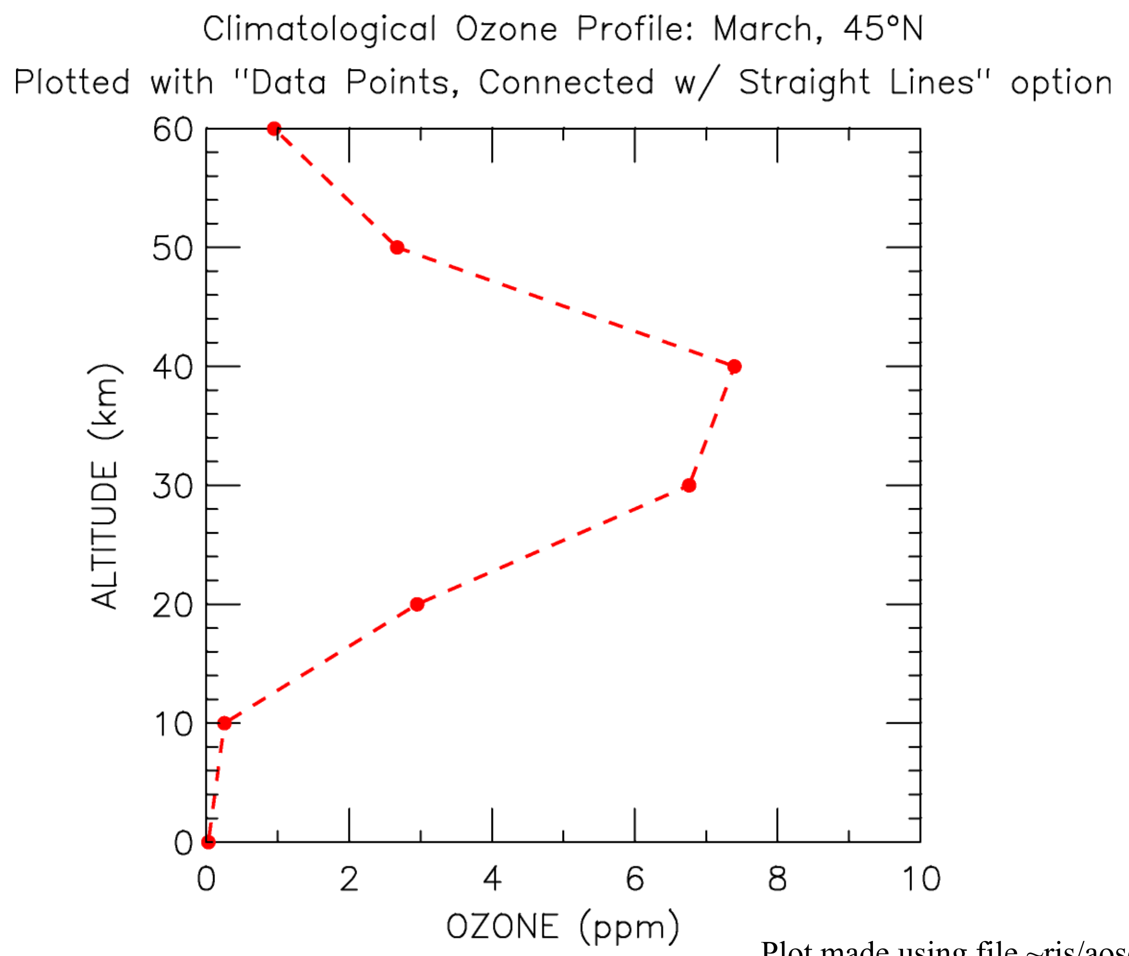
See pages 107 to 111, Press et al.

"hppltd" has a cubic spline option.

However: care must be taken to *correctly specify* whether fit is with respect to horizontal *or* vertical variable.

AOSC 652: Analysis Methods in AOSC

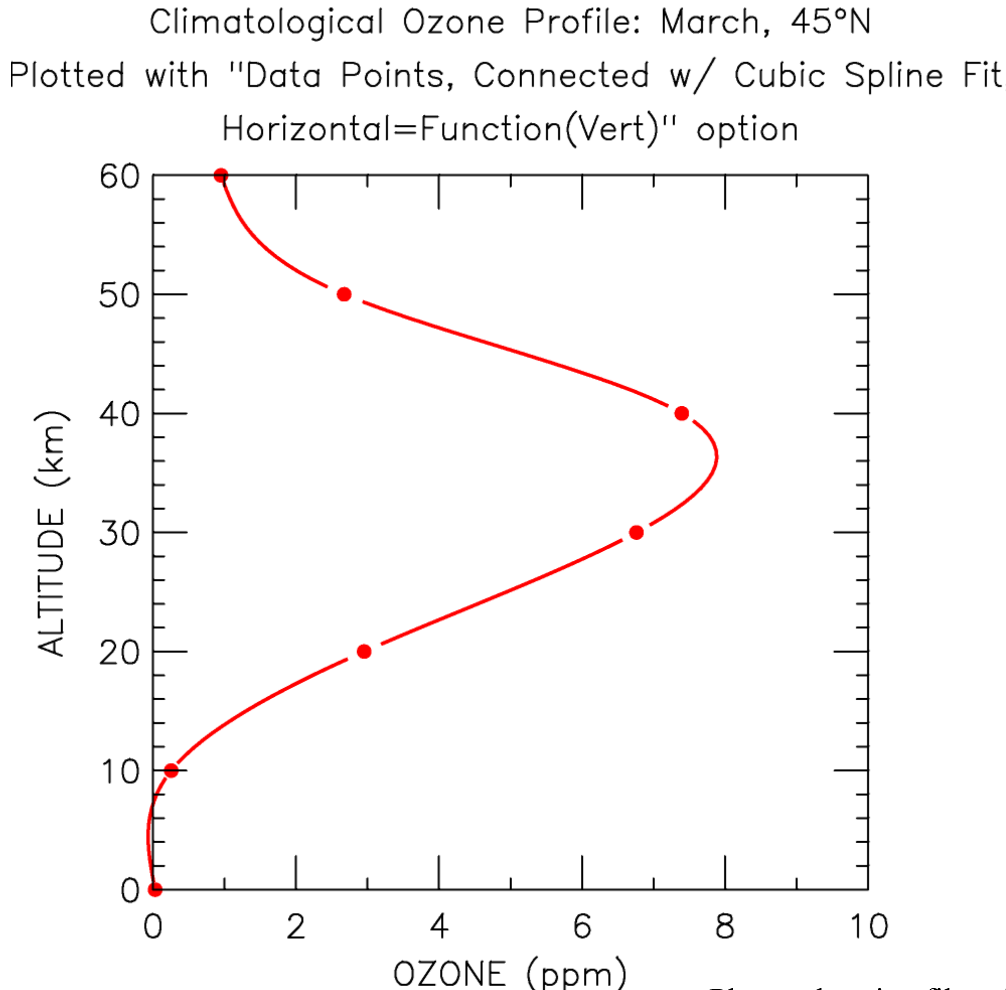
If you were to copy file ~rjs/aosc652/week_05/model_atmosphere.march_45N.dat and plot column 5 (O_3 mixing ratio) vs column 1 (altitude), using the "Data Points, Connected w/ Straight Lines" option, you would get:



Plot made using file ~rjs/aosc652/week_05/stncl.250

AOSC 652: Analysis Methods in AOSC

If you were to copy file ~rjs/aosc652/week_05/model_atmosphere.march_45N.dat and plot column 5 (O_3 mixing ratio) vs column 1 (altitude), using the "Data Points, Connected w/ **Cubic Spline Fit, Horizontal =Function(Vert)**" option, you would get:



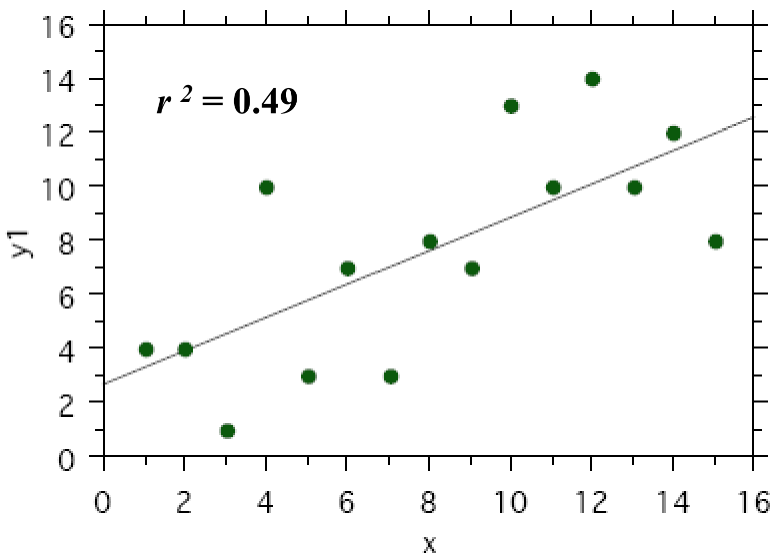
Plot made using file ~rjs/aosc652/week_05/stncl.251

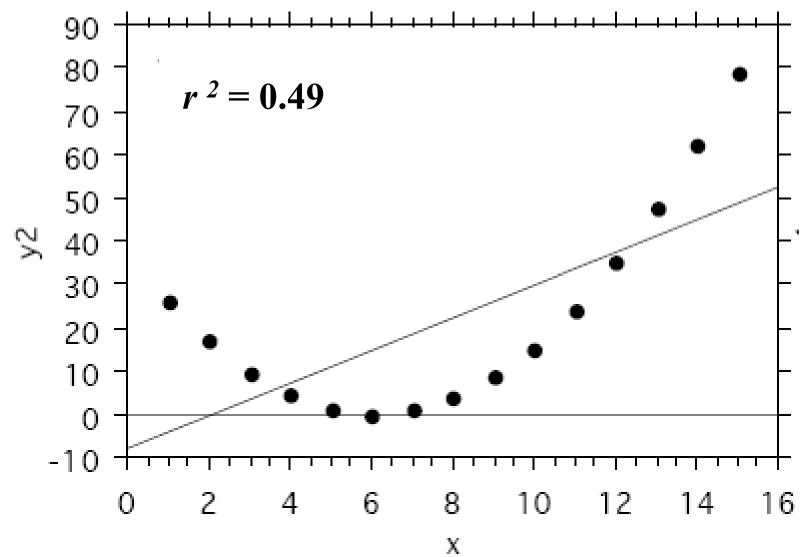
Let's return to Least Squares Fitting:

$y = a + b\,x$: formula for linear least squares

$y = a + b\,x + \mathrm{c}\,x^2$: formula for quadratic least squares

$y = a + b\,x + \mathrm{c}\,x^2 + d\,x^3$: formula for cubic least squares





From Dennis Hartmann's class notes:
http://www.atmos.washington.edu/~dennis/552_Notes_3.pdf

Quadratic Order Least Squares Fitting:

$$y = a + b\,x + c\,x^2$$

Minimize:

$$\sum_{i=1}^{N} (a + b\,x_i + c\,x_i^2 - y_i)^2 \equiv Cost\ Function$$

$$\frac{\partial\ Cost\ Function}{\partial a} = 2\sum_{i=1}^{N}(a + b\,x_i + c\,x_i^2 - y_i) = 0$$

$$\frac{\partial\ Cost\ Function}{\partial b} = 2\sum_{i=1}^{N}(a + b\,x_i + c\,x_i^2 - y_i)\,x_i = 0$$

$$\frac{\partial\ Cost\ Function}{\partial c} = 2\sum_{i=1}^{N}(a + b\,x_i + c\,x_i^2 - y_i)\,(x_i)^2 = 0$$

Quadratic Order Least Squares Fitting:

$$y = a + b\,x + c\,x^2$$

Minimize:

$$\sum_{i=1}^{N} (a + b\,x_i + c\,x_i^2 - y_i)^2 \equiv Cost\ Function$$

$$\begin{pmatrix} \Sigma(1) & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{pmatrix}$$

Minimize:

$$\sum_{i=1}^{N} (a + b\, x_i + c\, x_i^2 - y_i)^2 \equiv Cost\ Function$$

$$\begin{pmatrix} \Sigma(1) & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{pmatrix}$$

Or:

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$$

Minimize:

$$\sum_{i=1}^{N} (a + b\,x_i + c\,x_i^2 - y_i)^2 \equiv Cost\ Function$$

$$\begin{pmatrix} \mathrm{M}_{11} & \mathrm{M}_{12} & \mathrm{M}_{13} \\ \mathrm{M}_{21} & \mathrm{M}_{22} & \mathrm{M}_{23} \\ \mathrm{M}_{31} & \mathrm{M}_{32} & \mathrm{M}_{33} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \mathrm{Z}_1 \\ \mathrm{Z}_2 \\ \mathrm{Z}_3 \end{pmatrix}$$

Or:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \mathrm{I}_{11} & \mathrm{I}_{12} & \mathrm{I}_{13} \\ \mathrm{I}_{21} & \mathrm{I}_{22} & \mathrm{I}_{23} \\ \mathrm{I}_{31} & \mathrm{I}_{32} & \mathrm{I}_{33} \end{pmatrix} \begin{pmatrix} \mathrm{Z}_1 \\ \mathrm{Z}_2 \\ \mathrm{Z}_3 \end{pmatrix}$$

where matrix I is the inverse of matrix M

Can show:

$$I_{11} = \frac{1}{D}\,(M_{22}\cdot M_{33} - M_{32}\cdot M_{23}) \qquad I_{31} = \frac{1}{D}\,(M_{21}\cdot M_{32} - M_{31}\cdot M_{22})$$

$$I_{12} = \frac{-1}{D}\,(M_{12}\cdot M_{33} - M_{32}\cdot M_{13}) \qquad I_{32} = \frac{-1}{D}\,(M_{11}\cdot M_{32} - M_{12}\cdot M_{31})$$

$$I_{13} = \frac{1}{D}\,(M_{12}\cdot M_{23} - M_{13}\cdot M_{22}) \qquad I_{33} = \frac{1}{D}\,(M_{11}\cdot M_{22} - M_{12}\cdot M_{21})$$

$$I_{21} = \frac{-1}{D}\,(M_{21}\cdot M_{33} - M_{23}\cdot M_{31})$$

$$I_{22} = \frac{1}{D}\,(M_{11}\cdot M_{33} - M_{13}\cdot M_{31})$$

$$I_{23} = \frac{-1}{D}\,(M_{11}\cdot M_{23} - M_{13}\cdot M_{21})$$

where:

$$D = M_{11}(M_{22}\cdot M_{33} - M_{32}\cdot M_{23}) - M_{12}(M_{21}\cdot M_{33} - M_{23}\cdot M_{31}) + M_{13}(M_{21}\cdot M_{32} - M_{31}\cdot M_{22})$$

AOSC 652: Analysis Methods in AOSC

Then:

$$a = \left(\mathrm{I}_{11} \cdot \mathrm{Z}_1 + \mathrm{I}_{12} \cdot \mathrm{Z}_2 + \mathrm{I}_{13} \cdot \mathrm{Z}_3 \right)$$

$$b = \left(\mathrm{I}_{21} \cdot \mathrm{Z}_1 + \mathrm{I}_{22} \cdot \mathrm{Z}_2 + \mathrm{I}_{23} \cdot \mathrm{Z}_3 \right)$$

$$c = \left(\mathrm{I}_{31} \cdot \mathrm{Z}_1 + \mathrm{I}_{32} \cdot \mathrm{Z}_2 + \mathrm{I}_{33} \cdot \mathrm{Z}_3 \right)$$

where:

$$\mathrm{Z}_1 = \sum_{i=1}^{N} y_i$$

$$\mathrm{Z}_2 = \sum_{i=1}^{N} x_i \cdot y_i$$

$$\mathrm{Z}_3 = \sum_{i=1}^{N} x_i^2 \cdot y_i$$

AOSC 652: Analysis Methods in AOSC

Please copy FORTRAN code:

~rjs/aosc652/week_05/data_fit.f

and data file:

~rjs/aosc652/week_05/oh_tbl_mtn.dat

to an appropriate location in your directory

We will now use this code to obtain a linear, least squares fit of:

Satellite_Column_OH (column 1) versus
Ground_Based_Column_OH (column 3)

from data in file oh_tbl_mtn.dat

Please copy FORTRAN code:

~rjs/aosc652/week_05/data_fit.f

and data file:

~rjs/aosc652/week_05/oh_tbl_mtn.dat

to an appropriate location in your directory

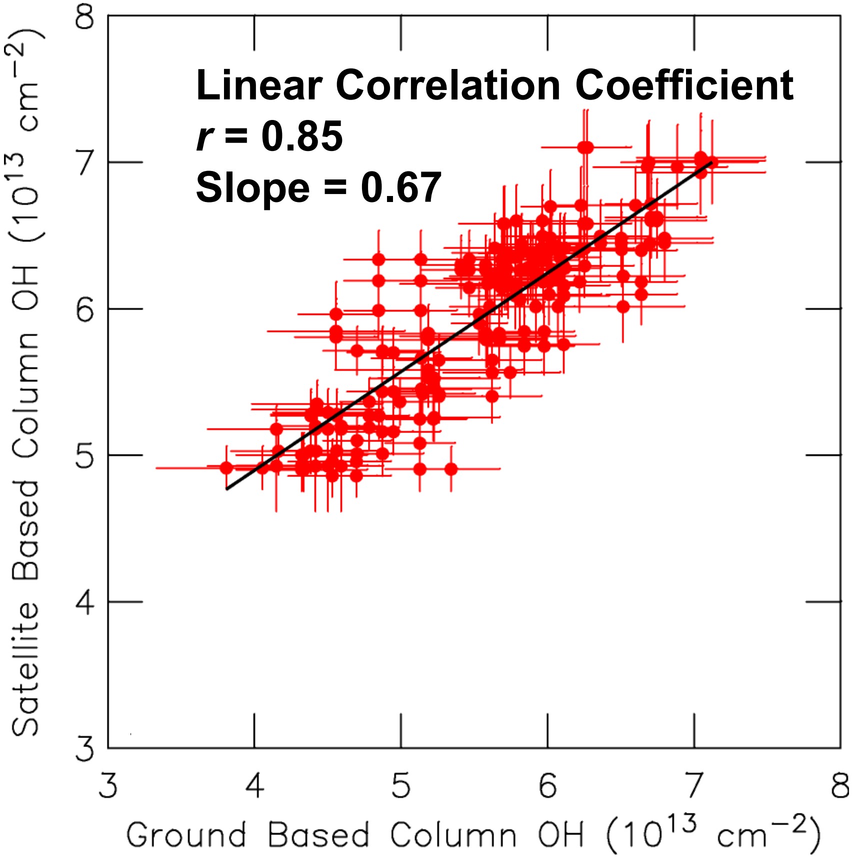
We will now use this code to obtain a linear, least squares fit of:

Satellite_Column_OH (column 1) versus
Ground_Based_Column_OH (column 3)

from data in file oh_tbl_mtn.dat

Now we shall plot Satellite_Column_OH vs Ground_Based_Column_OH and the linear, least squares fit.

AOSC 652: Analysis Methods in AOSC



AOSC 652: Analysis Methods in AOSC

Please copy FORTRAN code:

~rjs/aosc652/week_05/data_fit.f

and data file:

~rjs/aosc652/week_05/oh_tbl_mtn.dat

to an appropriate location in your directory

We will now use this code to obtain a linear, least squares fit of:

Satellite_Column_OH (column 1) versus
Ground_Based_Column_OH (column 3)

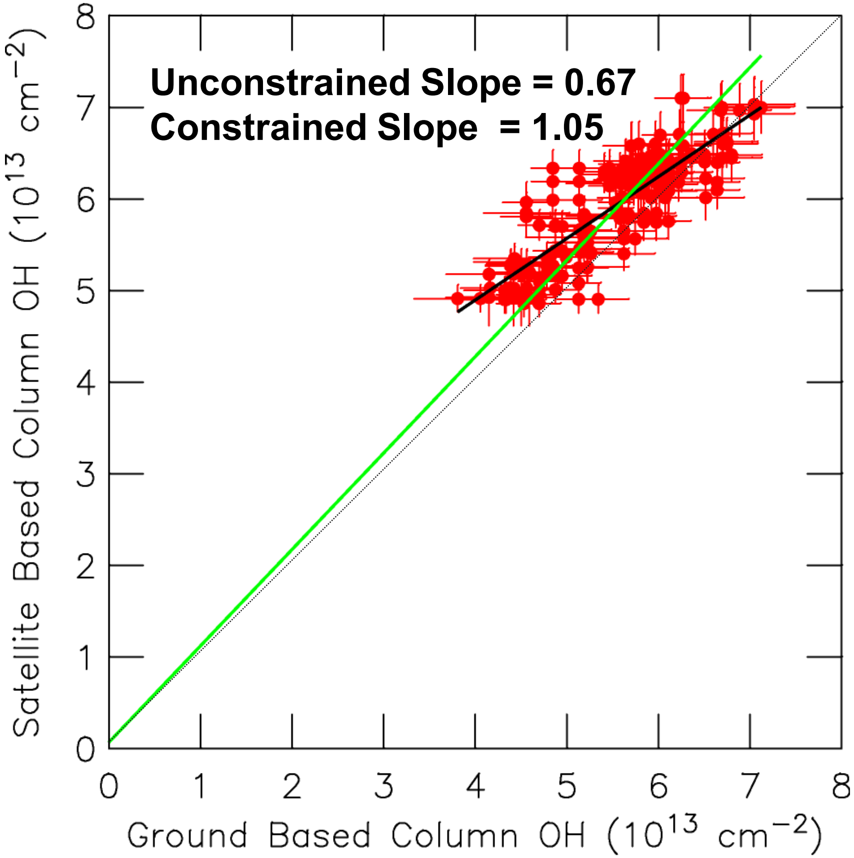
from data in file oh_tbl_mtn.dat

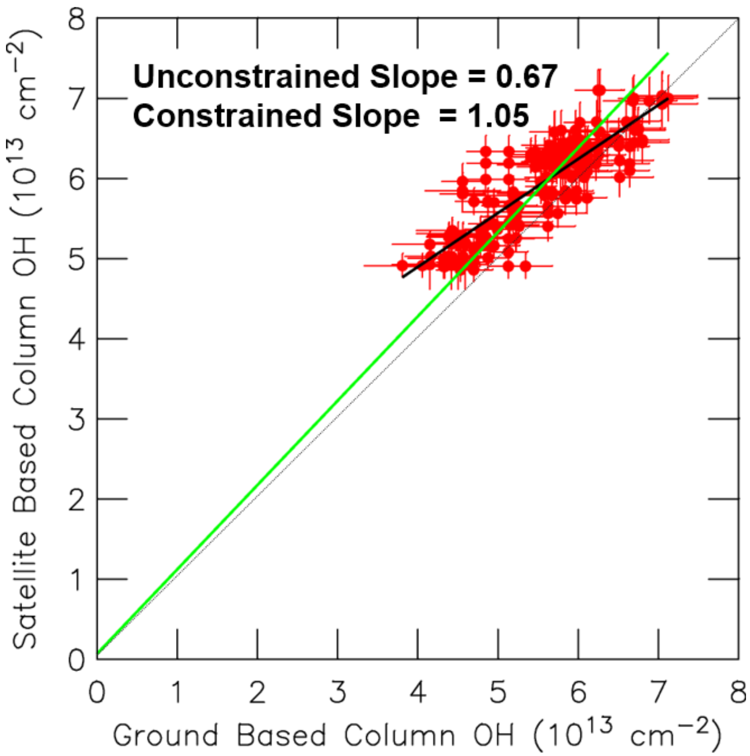
Now we shall plot Satellite_Column_OH vs Ground_Based_Column_OH and the linear, least squares fit.

Now repeat the fit, forcing the fit to go through the origin (0.,0.)

Construct a plot that shows the data and both fits, that has the x-axis and y-axis both starting at the origin.

AOSC 652: Analysis Methods in AOSC





Type of Linear Fit	Slope	Intercept ($10^{13} cm^{-2}$)	r	$\chi^2_{Reduced}$
Standard fit (Line 3)	0.673 ± 0.029	2.203 ± 0.161	0.85	
Fit through 0 (Line 1)	1.065 ± 0.005	0	0.85	

Wang et al., Validation of Aura MLS OH Measurements with FTUVS Total OH Column Measurements at Table Mountain, California, *JGR*, 2008 (Canty & Salawitch are co-authors ☺)

Much thanks to Shuhui Wang, JPL/Caltech, for providing the data from this plot, for our use in this class.

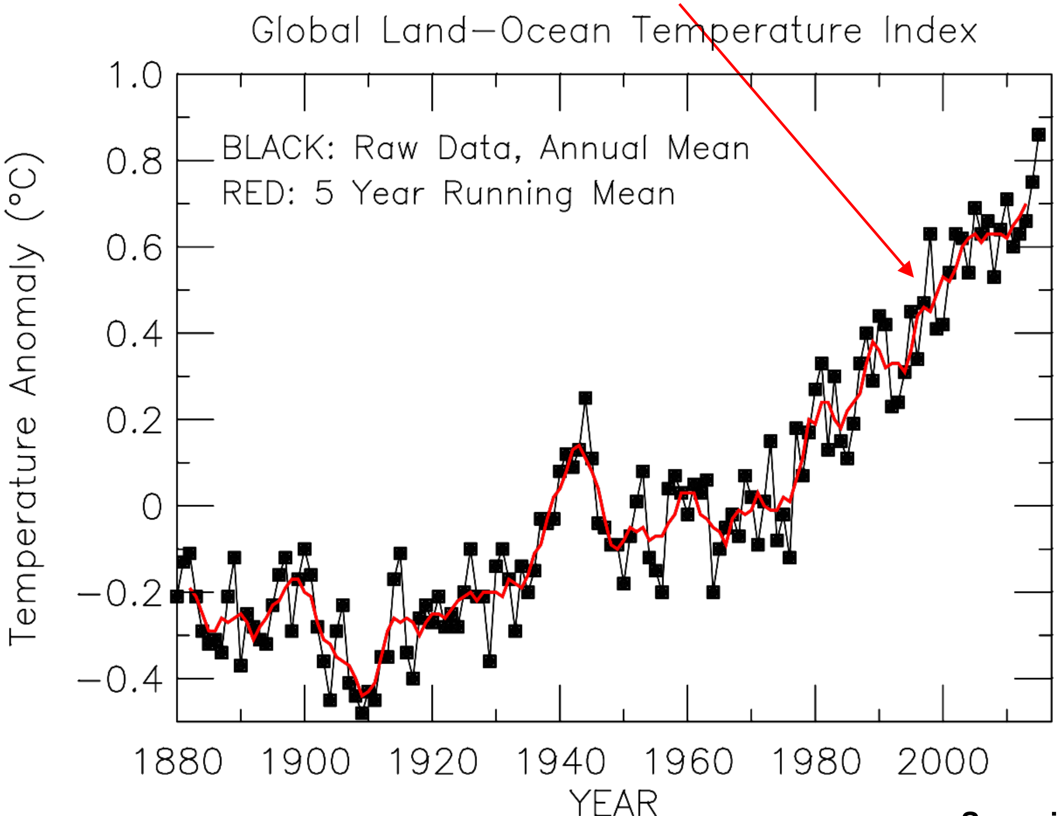
AOSC 652: Analysis Methods in AOSC

Your next assignment is to:

a) complete a FORTRAN subroutine that computes coefficients for a quadratic least squares fit

and

b) apply quadratic least squares fit to the 5 year running mean of the global average temperature anomaly we have been studying



See ~rjs/aosc652/week_04/stncl.501