Tutorial: Using Graf to fit curves and surfaces to data

David Nicholls, RSAA david.nicholls@anu.edu.au, 30 July 2017

Introduction

This is a brief tutorial on how to fit curves (polynomials, rational polynomials) and surfaces (bi-quadratic) to data. It should give you enough familiarity with the process to explore further.

Background

Graf is a very powerful application for OSX for creating graphs in publication-ready format. It was written (and is being constantly updated) by Ralph Sutherland. It is intuitive and very forgiving, but due to its power, it takes a bit of learning where the controls are and what they do. I don't pretend to know its full capabilities, but it is extremely useful at all levels of expertise.

It reads data from a wide range of text-based files, which you can open in Graf as a data file, ready for fitting and plotting.

There are three types of window in Graf: Data, Plot and Stat, accessed via the OSX Menu bar. The first two of these determine what is plotted and how it is formatted. The Stat window has an array of tools, including line and surface fitting.

What's assumed?

That you know the basics of how to open a text file in Graf and plot it. These steps will be summarised briefly below ("Basics"). *Skip directly to the examples on page 5 if you don't need this.*

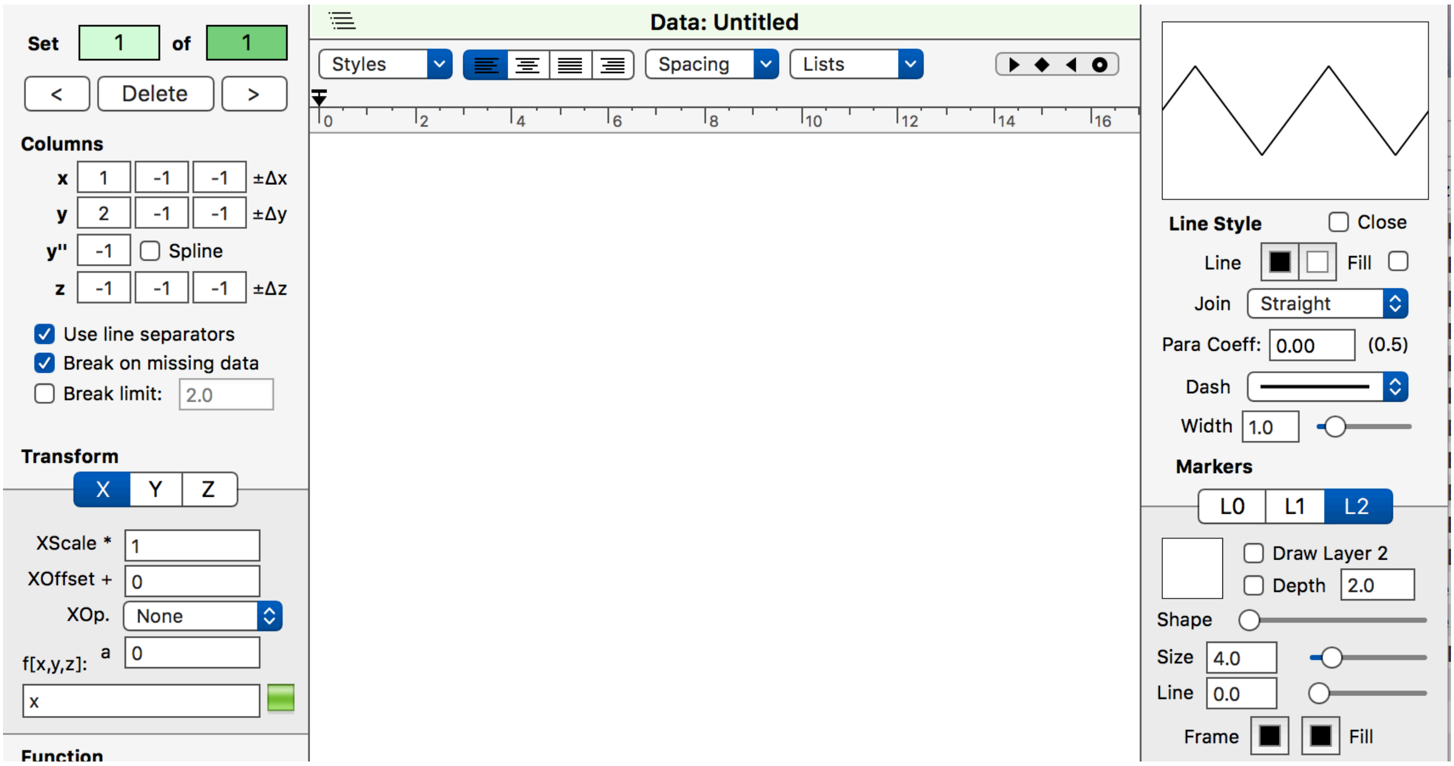
What's in this tutorial?

There are three data files (text, csv) used for the example fits. The tutorial provides step-by-step instructions to follow and screen shots to show the controls to use. Worked results are in the sub-folders.

Basics - the Data window

Graf reads many varieties of text based data files including .CSV and FORTRAN formatted data files, but not compressed or encoded files like .XLSX or FITS. Open the file with Graf, or drag the file onto the Graf dock icon and a new Data window opens which displays the data and any text. Graf interprets any line that starts with an alphabetic character as a comment. Lines commencing with numbers (or spaces, tabs or commas) are treated as data. You can comment out a line by inserting an alpha character (eg # or c) at the beginning. If you accidentally have a heading that begins with a number, Graf will misinterpret it as data, so comment it out with "c", etc.

A new data files looks like this:

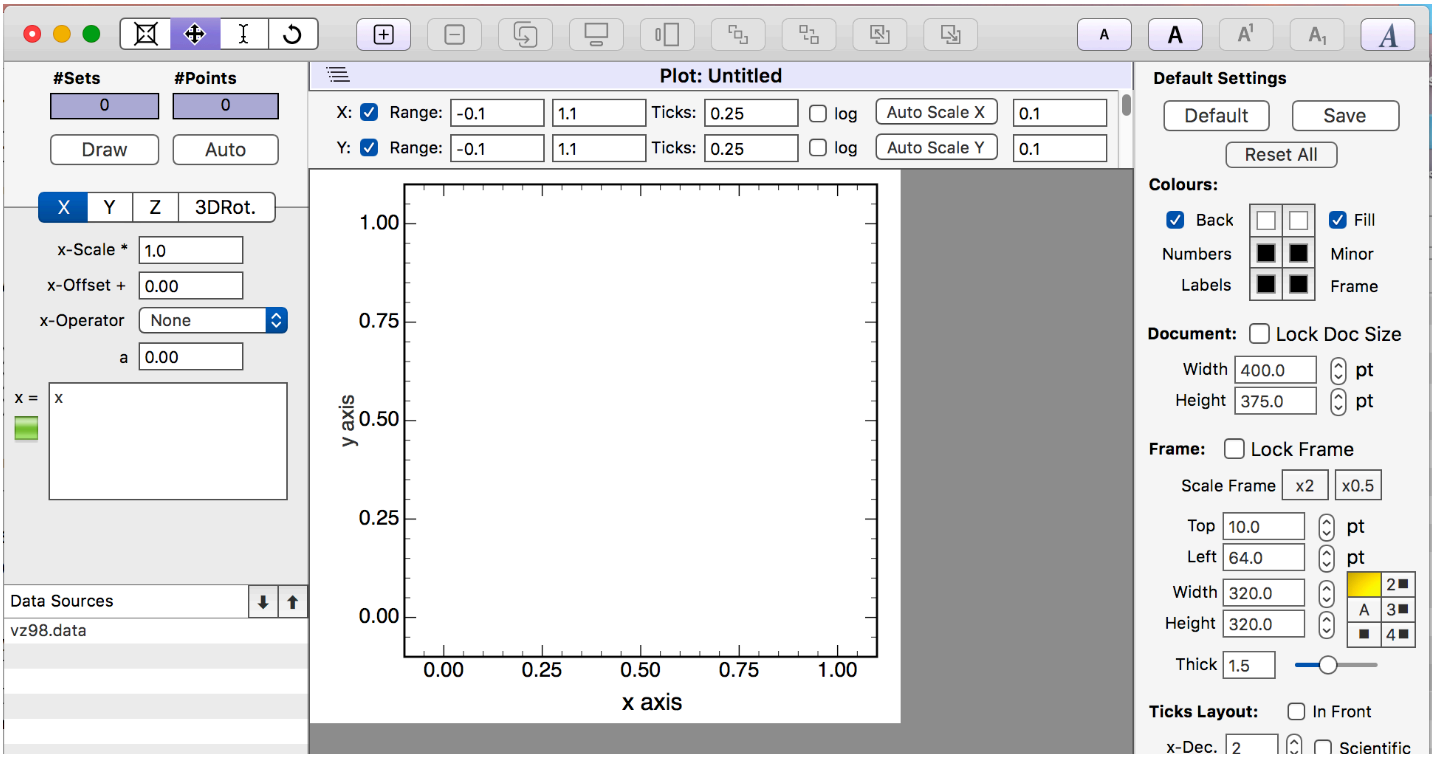


The Data window (left panel) identifies which of the data columns to plot for the x and y axes. It defaults to column 1 for x and column 2 for y (and -1 to ignore any third data column). You can edit the Columns boxes to change this. If there are three columns of data, you can set up two different graphs by creating a second data set, using the ">" button (top left). It defaults to using column 1 for x and column 3 (etc) for the second (etc) data set y coordinate. The <, > buttons flip between these data sets. (Thus the ">" button has two roles, swapping between data sets and creating new ones). For a 3D graph, where you want to explore 3D data, you can set the z value to 3. This is used in exercise 3.

The Data window also includes the settings for the type of plot to be generated (right panel). The default is to draw straight lines between each data point, in black, shown in the small right panel window at top. Click the default "Line" black centred box to change the colour of the plotted line (it opens the OSX colour picker). If you select "None" in the "Join" drop-down, the embedded window goes blank. Click the "Draw Layer 2" check box in the "Markers: section of the right panel and dots appear. Change their colour using the "Fill" box in the "Markers" section.

Basics - the Plot window

When the Data window is open, from the OSX Menu bar choose File > New Plot, and a plot window will open.



In "Data Sources", left panel, select your data file, then click the "Auto" button (Plot window, left panel, top). The plotted data appear, in the format and colour you have chosen in the Data window settings. You can set the range to be plotted, top two lines of the centre panel (this sub-window scrolls, so make sure you have the top part showing).

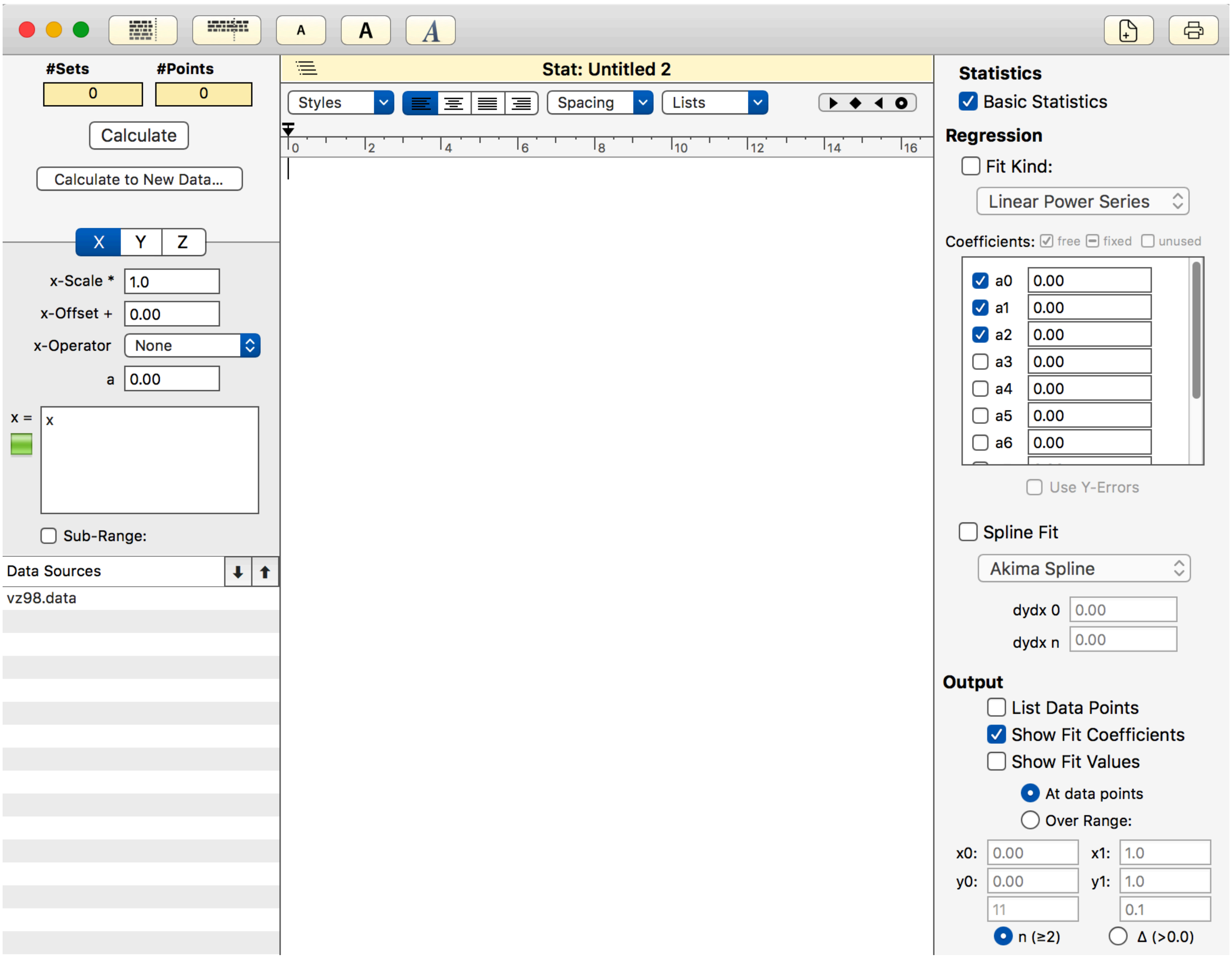
Note: If you want to re-plot, for example after changing line or marker colour in the Data window, use the "Draw" button (left panel, top), to keep the previously set x and y ranges. The Auto button clears any previous range settings.

You can edit the axis titles *in situ* provided you select the edit icon control, third icon in the four-icon box immediately to the right of the OSX file status buttons, top left.

The left axis and bottom axis label centring controls are at the top of the main window, in the centre. You can also drag the labels if you have the "Move" icon selected (left of the "Edit" icon).

Basics - the Stat window

The Stat window has a range of functions, but this tutorial is about line and surface fitting, and these controls appear by default in the right panel. This shows all the controls you will need for this exercise. You open a new Stat window from the OSX Menu bar via File > New Stat .



Examples

First example: a quadratic polynomial fit to a data set

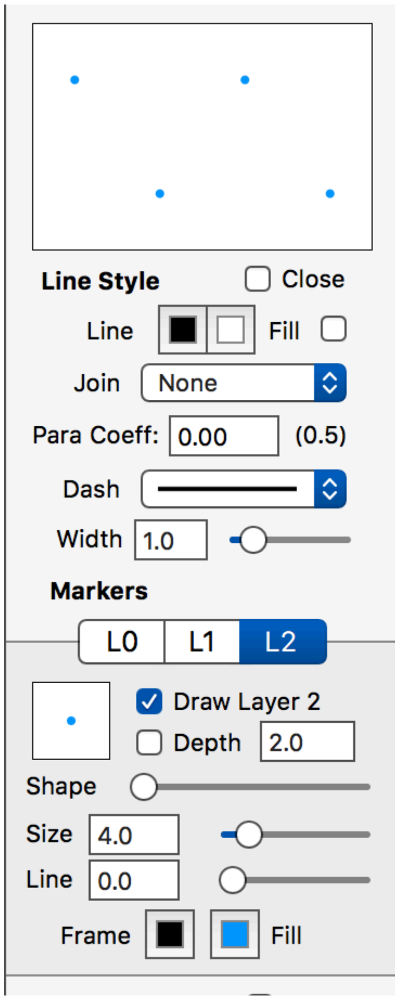
Opening the data file and plotting it

From the tutorial folder in Finder, drag the file named N77_B.txt onto the Graf icon in the dock (open Graf first if you have not retained it in the dock)[1]. The Data window opens, with the Line Style set to the default, "Straight" and the line colour to black. (Three lines are commented out already. Uncomment these if you want to show the points.)

In the Data window:

1. In the right panel (opposite), under Line Style, change the "Join" drop down to "None".
2. Under Markers check "Draw Layer 2". Click the fill box (under Markers, *not* Line Style), and the colour picker window opens. Select mid blue ("Aqua").
3. Save the file as "N77_B.data".

In the OSX Menu bar, click File > New Plot to open a new plot window.



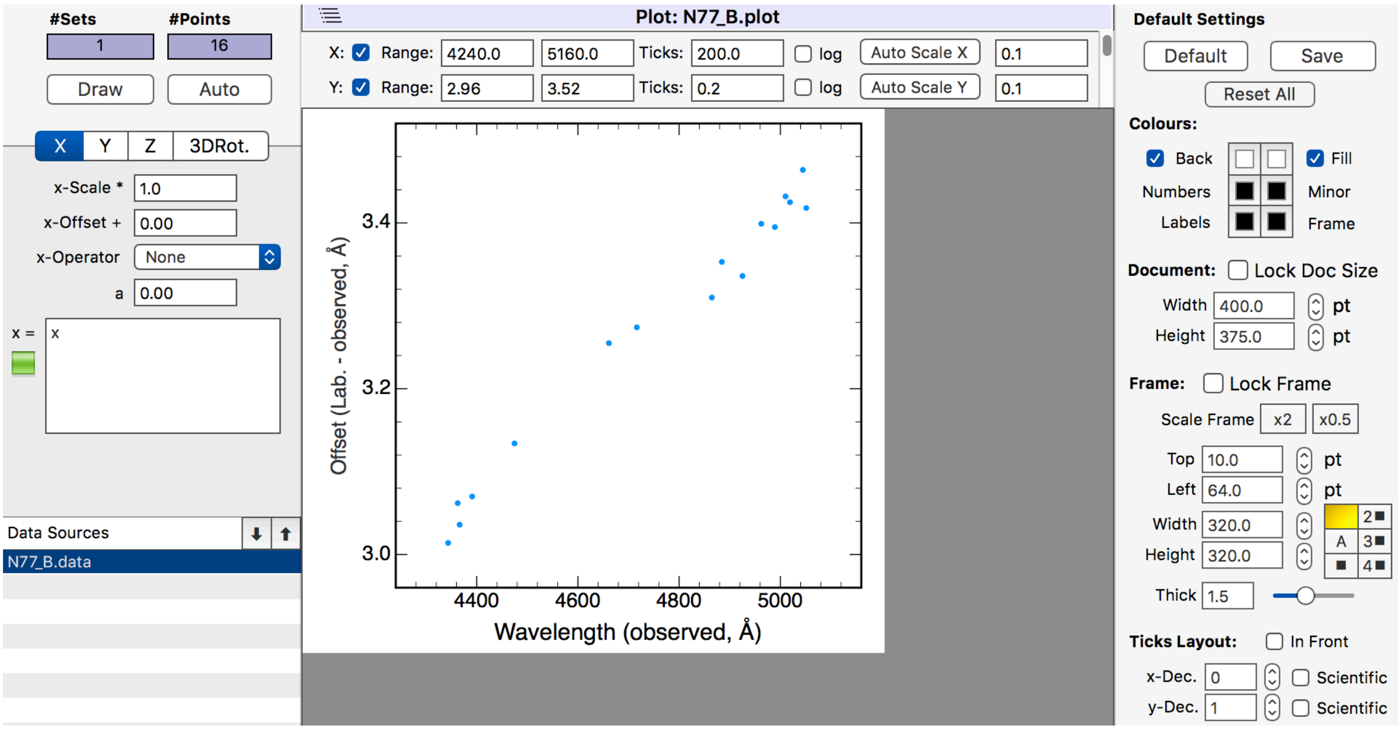
[1] **This file shows data for the spectrum of SMC-N77A, blue channel, wavelength vs. difference between observed and rest frame Lab. wavelengths. The plots shows measurement uncertainties and redshift, and a systematic error in wavelength calibration.

In the Plot window:

4. Select the data file name under Data Sources, left panel. Click the "Auto" button and the data points appear.

5. You can change the axis label as described above, and the axis numerical formatting , in the right panel under "Ticks Layout", but leave this until later. See illustration below.

This is the data we wish to fit with a simple quadratic polynomial. Now for the fun part.



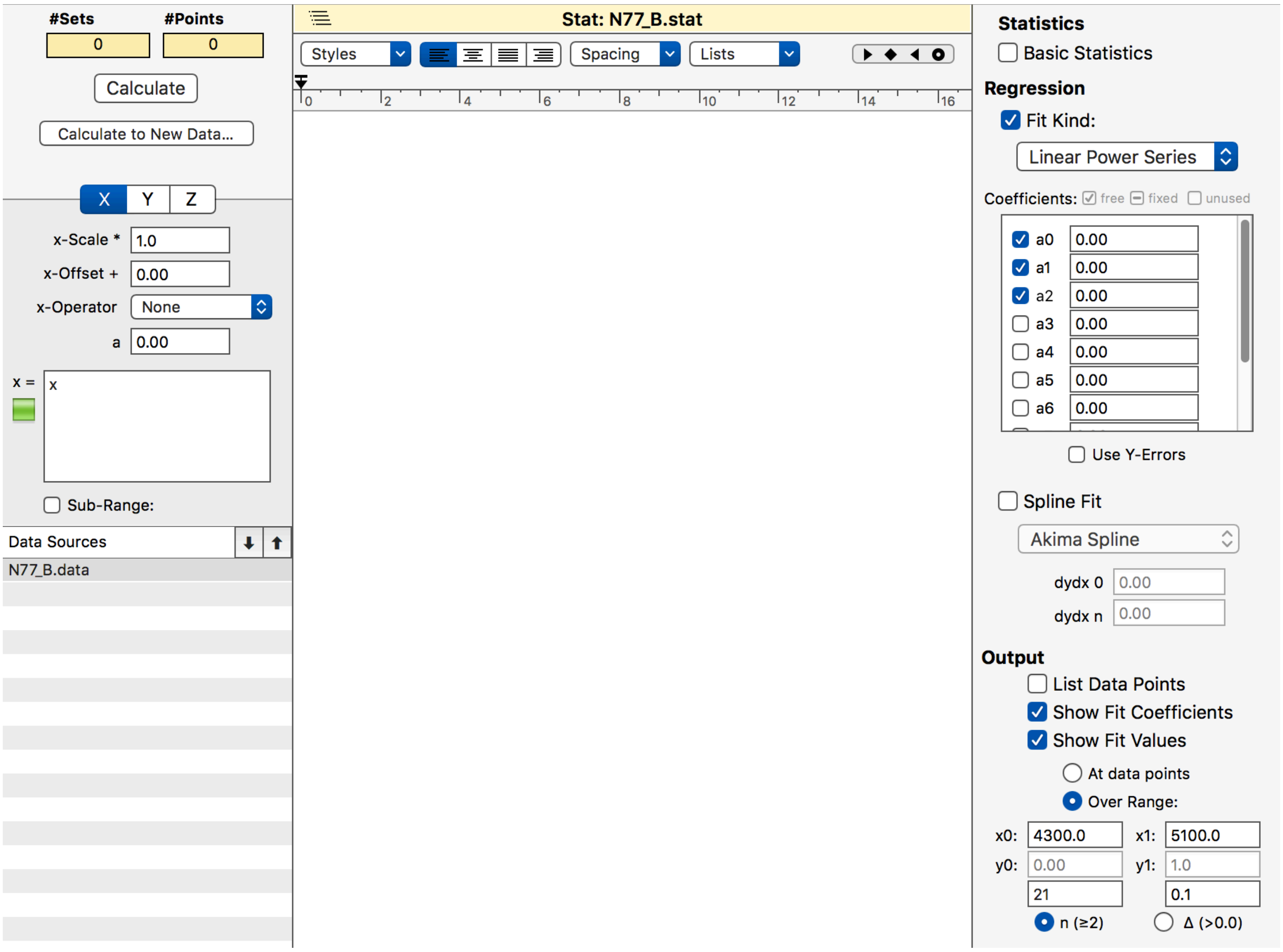
Setting up the Stat window

Open a new Stat window (Menu bar, File > New Stat).

In the Stat window:

5. Select the data file name (left panel under Data Sources).
6. In the right panel, deselect "Basic Statistics" and select "Fit Kind". Leave the default "Linear Power Series".
7. In the right panel, the Coefficients default to quadratic. Each checked coefficient box has three states: off (blank), fixed to value set (dash), and variable (checked). Leave these at default. You can turn on higher order polynomial fits or fix parameters if you want to, later.
8. In the right panel, check "Show Fit Values".
9. Leave "Show Fit Coefficients" selected (default).
10. There are two subsequent choices, "At data points" and "Over Range". For this exercise, select the "Over Range" button.

11. In the x0 box, enter 4300 to set the minimum fit x-value to be calculated, and in the x1 box, enter 5100 to set the maximum fit x-value value to be calculated.
12. Leave the y0 and y1 boxes blank.
13. In the box below y0, enter 21 (or however many points you want calculated—21 is sufficient for a smooth plot).
14. The remaining options at the panel bottom determine how the plotted points are calculated: the default n (≥2) option equi-spaces the points and the Δ (>0.0) sets an increment for plotting x-values. Leave the default selected for this exercise.



Calculating the fit

Next, we calculate a least squares fit to the data using a quadratic polynomial. In the Stat window (left panel, top), click the calculate button. The fit coefficients are shown along with other data relating to the quality of the fit, and at the bottom, the computed fit values. This exercise needs only one click to converge, but, as a general rule, click the "Calculate" button a few times to confirm that the Coefficients (right panel table) stay the same.

The coefficients indicate how well the fit has converged. Ideally, $|a[2]*x^2|$ should be ≲ 10. A much greater value indicates a poor fit or a local, sub optimal minimum.

Now click the "Calculate to New Data" button, and a new window will open, replicating the information in the stat window. Save the new data file as "N77_B.fit.data".

Plotting the fit

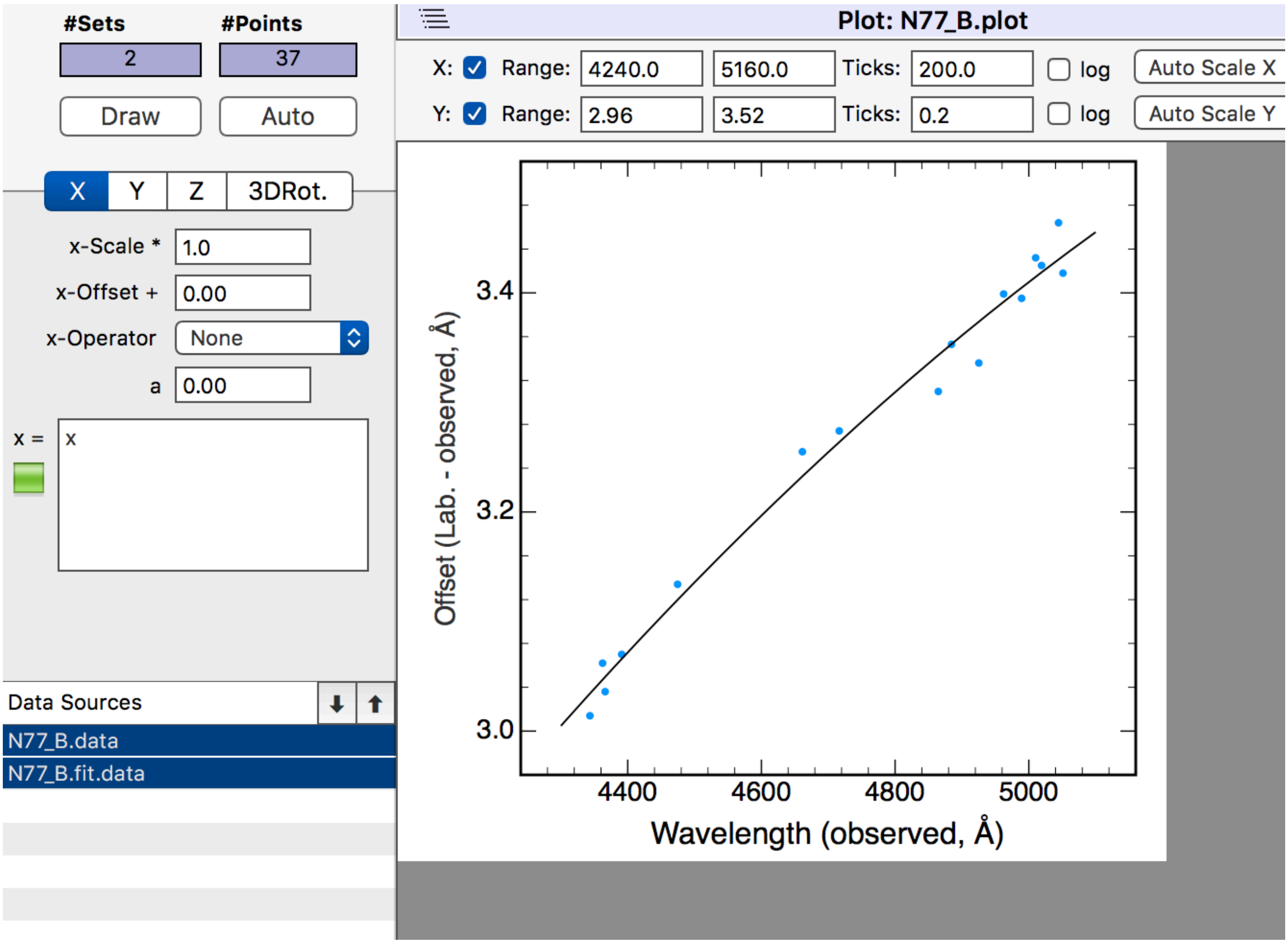
The output from the Stat process is a new Data window. As before, its settings default to a straight line between points, in black.

In the Plot window, the new data file is visible under Data Sources in the left panel. Select both data files and click the "Draw" button (top left). The data and the fit line now both appear on the graph.

You can now edit the axis labels, plot range and coordinate formats as required. It is a good idea always to save the plot and data windows. Save the Plot window as "N77_B.plot".

You can export the result as a PDF: File > Export to PDF.

When you have completed this exercise, take a look at the example files in the sub-folder, Example_1 and compare them with your results. The plot window should look something like:



Second example: a rational polynomial fit to calculated data points

Background

A rational polynomial has the advantage of good stability outside the data range, unlike quadratic and other higher order polynomial fits. It is the ratio of two polynomials, in the case of a quadratic ratio pair, in the form:

$$y = (p_2x^2 + p_1x + p_0)/(q_2x^2 + q_1x + q_0)$$

This example fits a rational polynomial to computed data points for the ratio of the [OIII] 1660+66 line fluxes to the [OIII] 5007 line fluxes, for a range of electron temperatures. The source file is Te.txt, in the main folder.

Graf computes the p and q parameters, with, by default, q_0 set to 1 to simplify computation. This can be changed if necessary by clicking in the "q0" setting box.

In the "Fit Kind" drop down box, Stat window, right panel, this choice is P2/Q2 (i.e., second order in both P and Q). This is a good place to start when you don't know what combination of polynomials might work best.

Step by step procedures

These steps are substantially similar to the first example, until we get to the curve fitting.

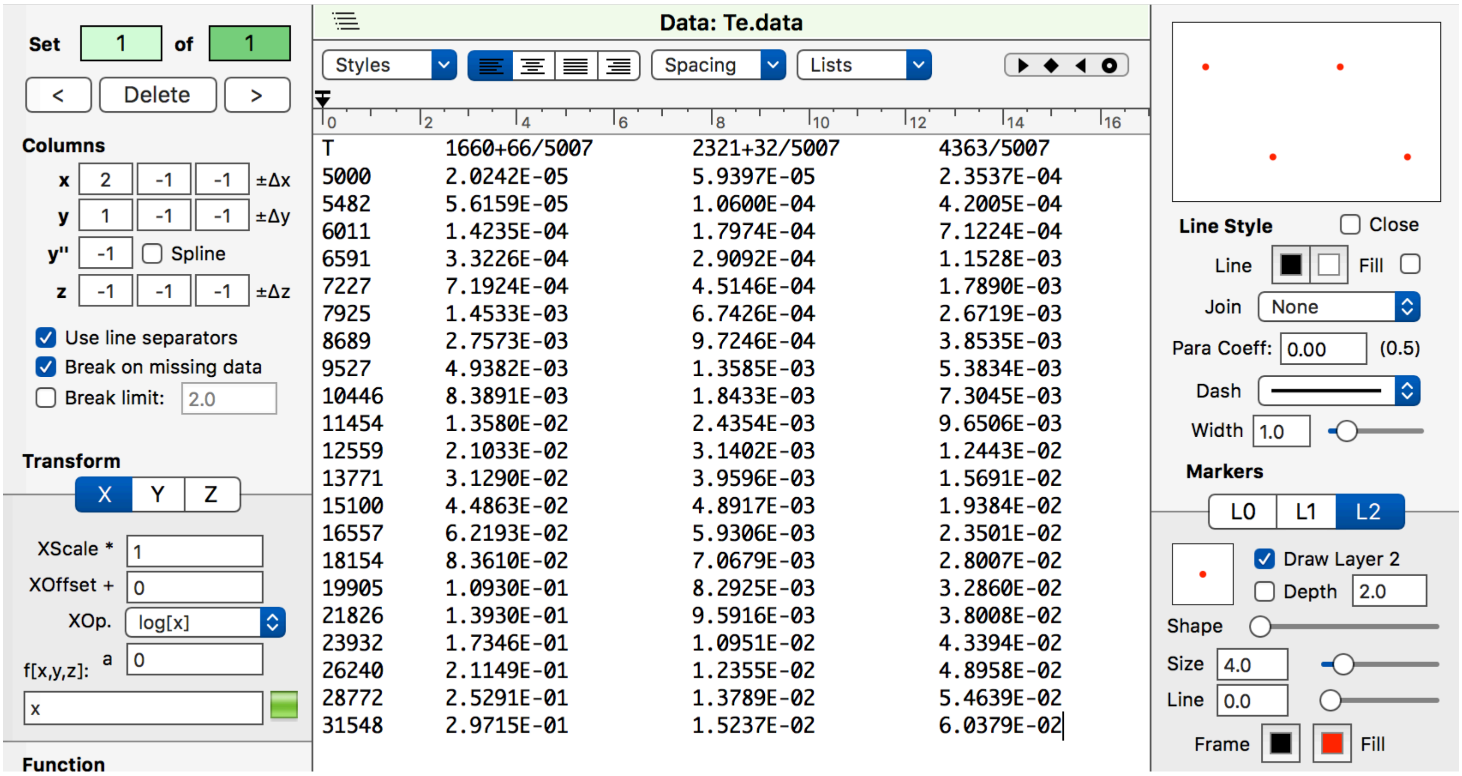
1. Drag the text file Te.txt onto the Graf icon in the dock, as before. A Data window opens. In this case, there are 4 columns, and we'll use only the first two (you can select the others later as practice).
2. Save the Data window as "Te.data" (using "Save As").

In the Data window:

3. We want to plot log(Te) on the y-axis and log(1660+66/5007) (the line flux ratio) on the x-axis. So we need to make two changes to the default setting: the column order and how the data is processed.
4. in the Columns section (left panel, top) set x to 2 and y to 1 (leave the -1 entries unchanged, they handle error values or a z-coordinate, etc., which we're not using here).
5. To plot the log values we need to alter the "Transform" settings (left panel) to pre-process the tabulated linear values. With the "X" segment selected (grey), set the XOp. drop down entry to log(x). Select the "Y" segment, and set the YOp. drop down entry to log(y). Now we're ready to plot.

In the Plot window:

6. In the OSX Menu bar, chose File > New plot. In the Plot window that opens, select the data file in the "Data Sources" section of the left panel, and then click the "Auto" button to set the plotting ranges automatically and draw the graph. A curved line appears, but we want to plot the separate date values.
7. Go back to the Data window, and in the right panel, Line Style, set the "Join" drop down to "None".
8. Under Markers, check the "Draw Layer 2" box.
9. Select the "Fill" box under Markers, and in the colour picker that opens, select the red colour ("Maraschino"). The dots will now show as red in the sub window, right panel Data window. See illustration:



T	1660+66/5007	2321+32/5007	4363/5007
5000	2.0242E-05	5.9397E-05	2.3537E-04
5482	5.6159E-05	1.0600E-04	4.2005E-04
6011	1.4235E-04	1.7974E-04	7.1224E-04
6591	3.3226E-04	2.9092E-04	1.1528E-03
7227	7.1924E-04	4.5146E-04	1.7890E-03
7925	1.4533E-03	6.7426E-04	2.6719E-03
8689	2.7573E-03	9.7246E-04	3.8535E-03
9527	4.9382E-03	1.3585E-03	5.3834E-03
10446	8.3891E-03	1.8433E-03	7.3045E-03
11454	1.3580E-02	2.4354E-03	9.6506E-03
12559	2.1033E-02	3.1402E-03	1.2443E-02
13771	3.1290E-02	3.9596E-03	1.5691E-02
15100	4.4863E-02	4.8917E-03	1.9384E-02
16557	6.2193E-02	5.9306E-03	2.3501E-02
18154	8.3610E-02	7.0679E-03	2.8007E-02
19905	1.0930E-01	8.2925E-03	3.2860E-02
21826	1.3930E-01	9.5916E-03	3.8008E-02
23932	1.7346E-01	1.0951E-02	4.3394E-02
26240	2.1149E-01	1.2355E-02	4.8958E-02
28772	2.5291E-01	1.3789E-02	5.4639E-02
31548	2.9715E-01	1.5237E-02	6.0379E-02

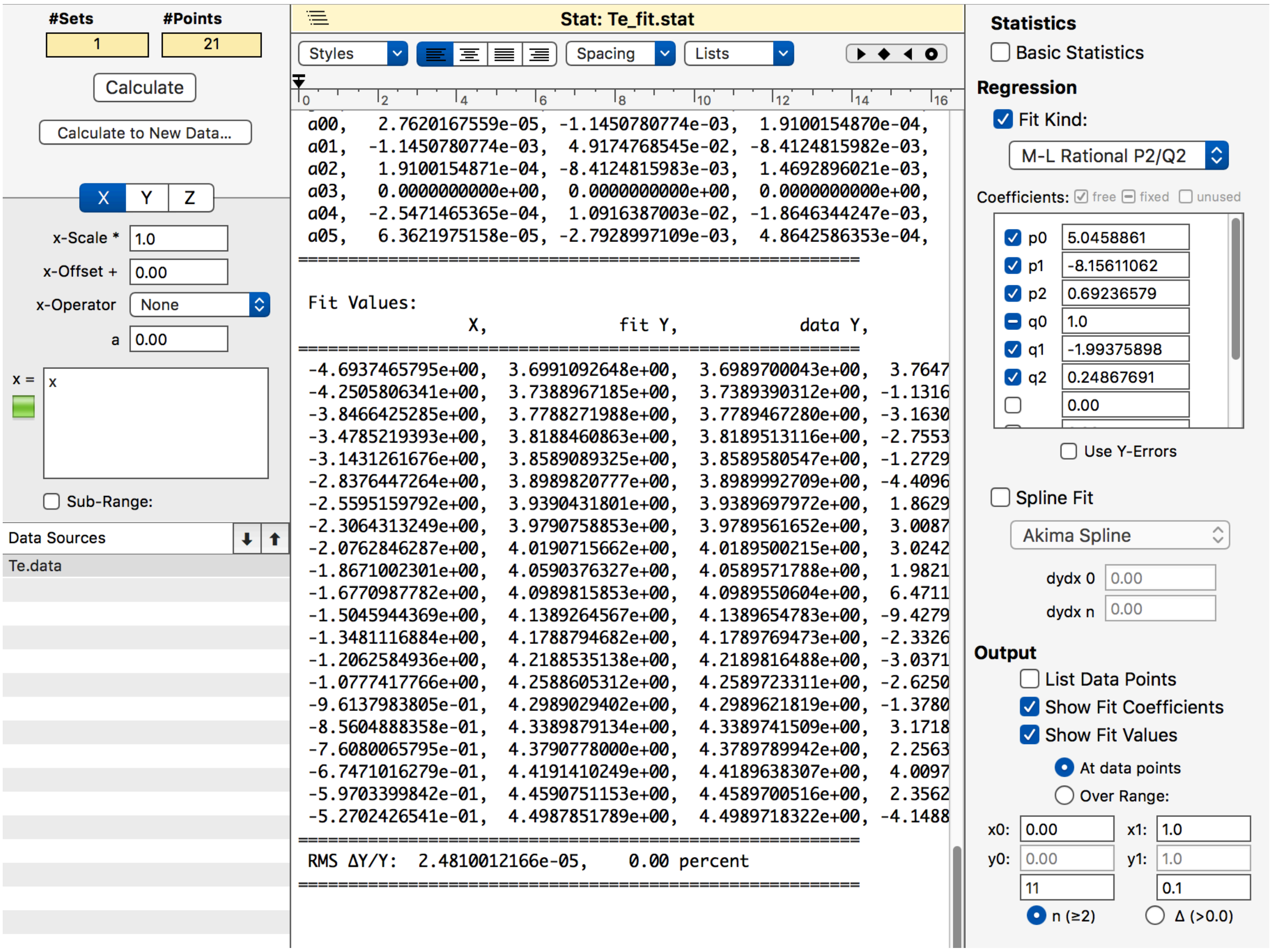
10. Go back to the Plot window and click the "Draw" button at top left. The black line should change to red dots at each data point.

Now we are ready to calculate the rational polynomial fit.

11. In the OSX Menu bar, select File > New stat. A new Stat window opens.

In the new Stat window:

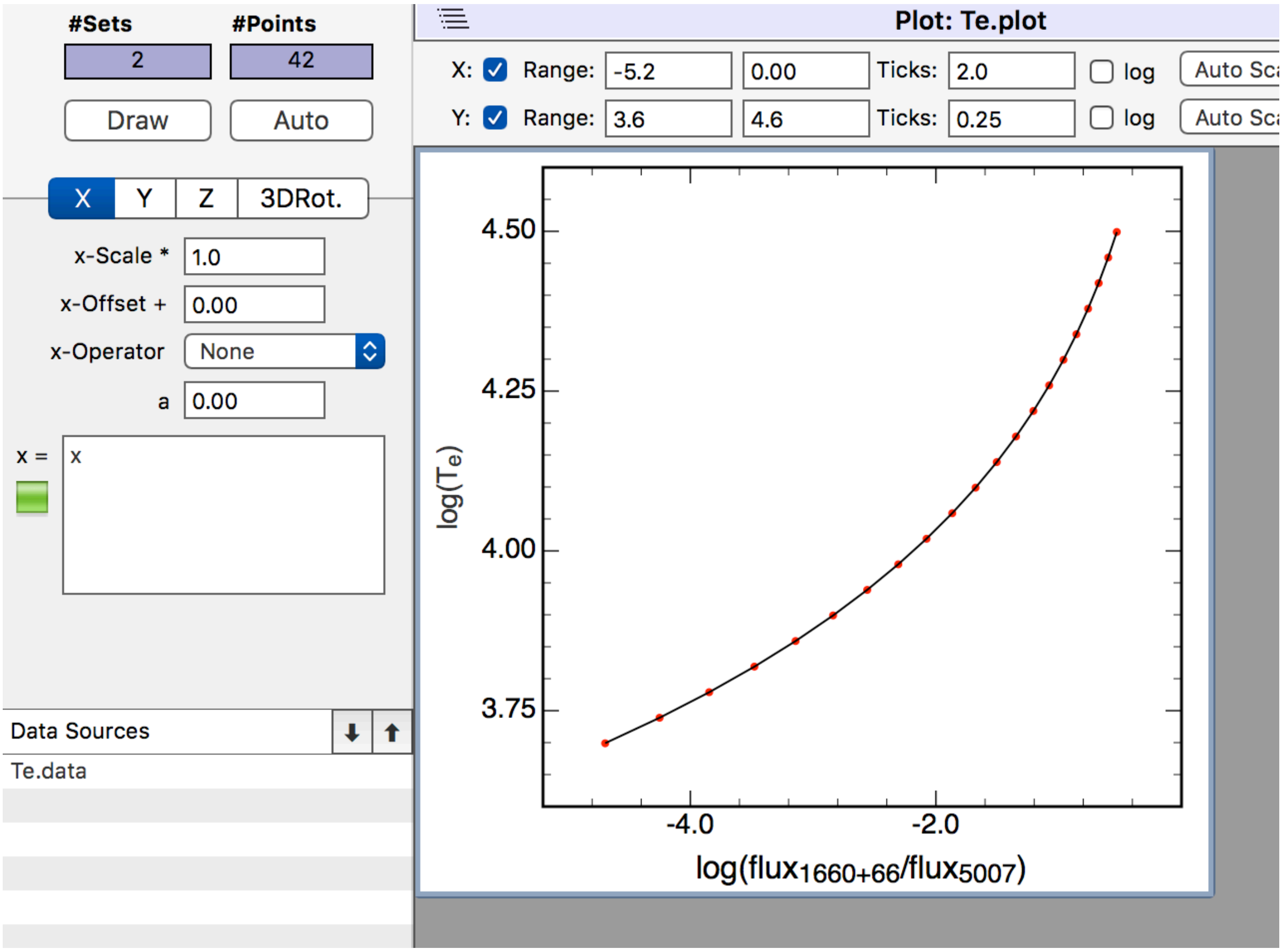
12. In the left panel, select the data file.
13. In the right panel, uncheck "Basic Statistics", check the "Fit Kind" box.
14. In the drop down selection box immediately below that, choose "M-L Rational P2/Q2", to use a quadratic ratio pair.
15. In the right panel, under "Output", check "Show Fit Values".
16. Below that, make sure "At data points" is selected.
17. Now click, at top left the "Calculate" button. This generates the first approximation to the fit. In many cases, the fit will occur on the first iteration, but check: while watching the "Coefficients" table, right panel, click the "Calculate" button two or more times to ensure the coefficient values stop changing. This may take several clicks, but if you can't get the values to stabilise, it will be necessary to change the type of fit to use.
18. In this case, P2/Q2 stabilises quickly, after the second iteration. Now inspect the fit coefficient values. If the fit is good they will be typically between -10 and +10. Very large values means you have found a local minimum but it is unlikely to be the best fit. This also means changing the type of fit. Again, this isn't required for this example.
19. Save the Stat window as "Te_fit.stat".
20. Now click the "Calculate to New Data" button, Stat window, top left. This copies the computed fit data to a data file that can be plotted.
21. Save the new data file as "Te_fit.data".



Now we plot the data and the Rational polynomial fit. In the (already open) Plot window:

22. In the Data Sources sub-window, left panel, select the data file, Te.data, and the Stat output data file, Te_fit.data.
23. Click the "Draw" button, top left. The plotted point values and the fit appear. Because we have left the Stat data window at its default settings, it will appear as a black line, which is what we want. As you can see, the fit is very good.
24. You can use the P2/Q2 parametric equation (top of page 9, above) and the parameter coefficients from the fit data file (Te_fit.data) to give an algebraic expression for calculating Te for any value of the line flux ratio.
25. Edit the plot labels and coordinates as in Example 1.
26. You can export the plot as a PDF as before: Menu bar: File > Export to PDF.

Finally compare your results with the equivalent files in the sub-folder, Example_2. It should look something like this:



Third example, 3D surface fitting

This process follows the same sequence as the first two examples, but uses a third column in the data file. In this case we'll use data extracted from Van Zee (98)'s[2] HII region flux ratios. It is provided in the file vz98.csv, with the (log) line ratios [NII]/[OII], [OIII]/[OII] and R23.

Step be step procedures

1. First, we open the data file: open the file vz98.csv in Graf (or drag onto Graf icon in dock).

In the Data window:

2. Because we have three dimensional data, we need to set the z-coordinate box (left panel, under "Columns"): change the default "-1" to "3". Leave x and y as the default 1 and 2.
3. As earlier, in the right panel, set the Line Style to "None".
4. Under Markers, right panel, check the "Draw Layer 2" box.
5. Click in the "Fill" box in the Markers section, and select mid blue ("Aqua") from the colour picker that opens.
6. Save the file as "vz98.data" ("Save As").
7. Now plot the data. In the OSX Menu bar, select File > New plot.
8. In the Plot window, left panel under Data Sources, select the data file and click "Auto".

The data points appear on the plot as blue dots. At this point it is useful to explore the 3D rotation capabilities of Graf:

In the Plot window:

9. In the left panel near top, select the 3DRot. control next to the XYZ controls. This activates three rotation sliders in a context sub-window.
10. Drag the uppermost slider control to the right so that the "Roll (x)" box above it shows "85". The plot rotates about a horizontal axis
11. Drag the middle slider control back and forth between 0 and 80 (in the "Pitch (y)" box) to see how the data rotates around a vertical axis.
12. Move the sliders back to their default (zero) positions.

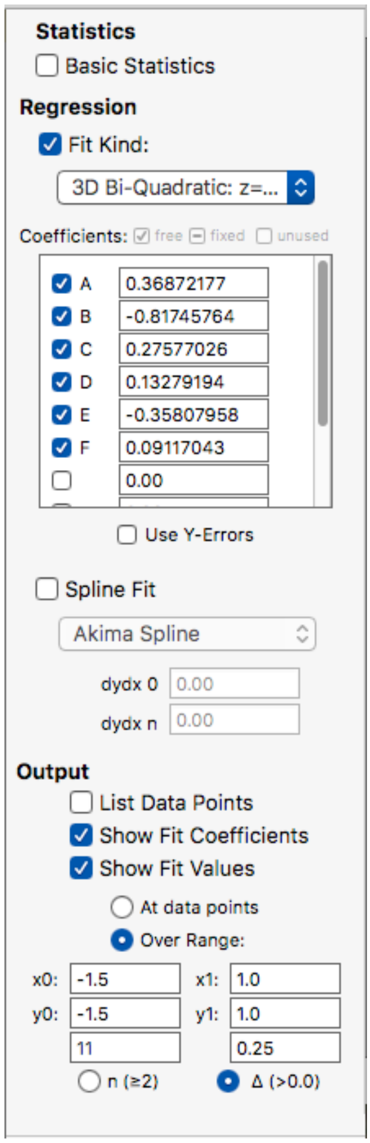
Now to create the surface fit, in this example a bi-quadratic of the form:

$$z = A + Bx + Cy + Dxy + Ex^2 + Fy^2$$

12. Open a new Stat window, and select the vz98.data file under Data Sources, left panel.

In the Stat window:

13. Uncheck"Basic Statistics" (right panel, top) and check the "Fit Kind" box.
14. In the drop down box below "Fit Kind", choose "3D B-Quadratic: z= ..."



[2] van Zee, L. et al., 1998, Astron. J, 116, 2805

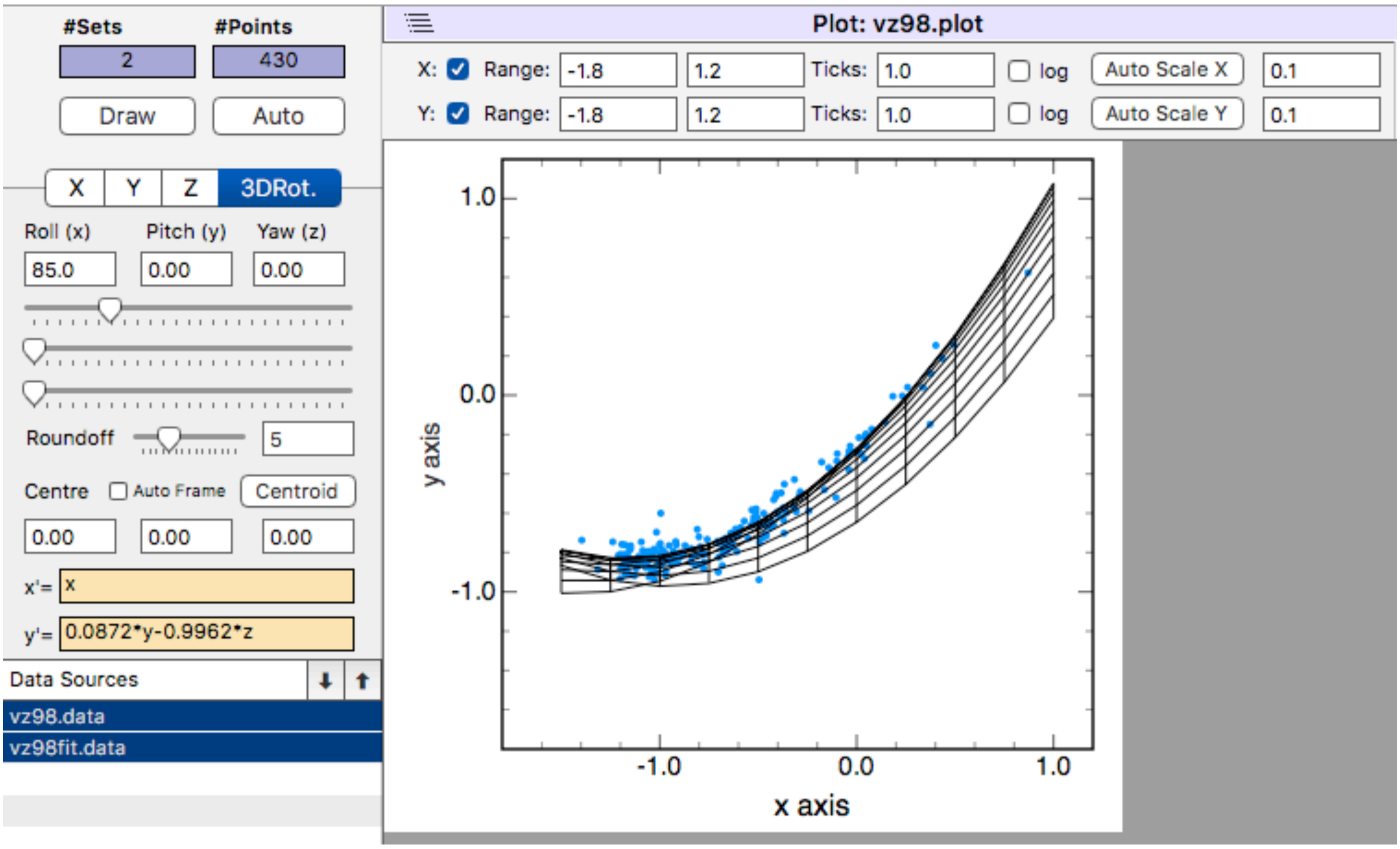
15. Under the "Output" header , right panel, check "Show Fit Values".
16. Select the "Over Range" option below that.
17. In the boxes below this, enter "-1.5" for x0 and y0 and "1.0" for x1 and y1 (no quotes).
18. Select the "Δ (>0.0)" option. Leave the default 0.25 value in the box above that.
19. Now click the "Calculate" button (left panel, top).
20. The right panel should look like the screen shot at right:
21. Finally, click the "Calculate to New Data" button (Left panel, top) to create a new data file with the fit data in it.
22. The new data window opens.

In the new Data window:

23. Set the z coordinate value (Left panel, under "Columns") to 3 to activate 3D plotting. Leave the other values at their defaults.
24. Save this data file as "vz98fit.data".

In the Plot window:

25. Select both files (vz98.data and vz98fit.data) and click "Draw". The plot now shows the data points and a rectangular grid.
25. In the Plot window, left panel near top, select the 3DRot. control next to the XYZ controls.
26. As before (steps 10 and 11, previous page), drag the top slider to 85 and drag the second slider back and forth to see how the surface grid fits the data points.
27. Edit the axis labels if required, This is not necessary unless you want a screen shot, and then you need to use the equations from the x'= and y'= boxes (beige background, left panel) to specify exactly what's being plotted. Here, x and y are the original data in columns 1 and 2 of the Data file (log(NII/OII) and log(OIII/OII)). x' and y' are the transformed coordinates.
28. Save the plot as "vz98.plot".
29. The fitted graph should look something like this:



The bi-quadratic equation and the coefficients A, B, C, D, E, and F are available in the fit data file (vz98fit.data), defining the parametric fit equation.

Conclusion

And that's it. This tutorial has covered briefly how to fit curves and surfaces in Graf. Graf has many more capabilities and nuances, but this should serve as a starting point for exploring further.

DCN
RSAA, ANU
31 July 2017