Bruce Ravel
Molecular Environmental Science Group
Bioscience Division
Argonne National Laboratory
http://cars9.uchicago.edu/~ravel
bravel@anl.gov

1. Scientific background
2. Feff without Atoms
3. One-shell/one-data-set fit
4. Multiple k-weight fitting
5. Multiple data set fitting
6. Bringing it all together

Scientific Background

Polyvinyl chloride – PVC – is a type of rigid plastic used for water and sewage transport in the United States and elsewhere. In the US, home building codes require copper pipe for bringing water into homes and PVC for carrying water and sewage out of home.

The manufacture of PVC uses organic tin species, mostly dimethyl tin, as a stabilizing agent. Over time, organic tin species can leach out of PVC an into municipal water supplies. My collaborator, Chris Impelliteri, at the US Environmental Protection Agency is studying this leaching process. The organic tin species used as stabilizers evolve during the manufacturing process, so we are using XAS to identify and characterize the tin species present in commercial PVC pipes.

To start, we have begun building a library of organic (methyl tin, butyl tin, phenyl tin, tricyclohexyl tin, etc) and inorganic (metallic tin, tin oxide, tin chloride) standard compounds. To have confidence that I can interpret spectra measured on PVC samples, I want to carefully analyze the standards. In this document, I show the analysis of two methyl tin species.

As well as being a real-world example of using Artemis, this will serve to introduce several important concepts, including multiple k-weight fitting, multiple data set fitting, and the concept of constraining parameters across data sets.

ARTEMIS
EXAFS DATA ANALYSIS
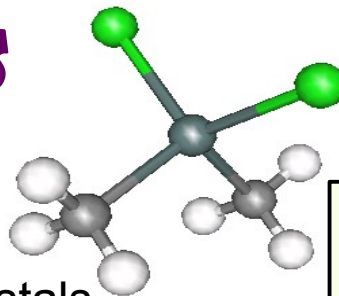
Methyltin compounds are molecules, not crystals. Atoms won't help in this case. Using Google, I eventually found a Protein Data Bank file containing coordinates for dimethyltin dichloride (shown above). By cutting and pasting and adding some information that Feff expects, I have a suitable `feff.inp` file.

```
TITLE dimethyltin dichloride

HOLE 1      1.0

*              mphase,mpath,mfeff,mchi
CONTROL    1        1       1      1
PRINT      1        0       0      0

RMAX         6.0
NLEG         4

POTENTIALS
*     ipot     Z    element
      0      50     Sn
      1      17     Cl
      2       6     C
      3       1     H

ATOMS
*     x              y           z
   -0.027    2.146     0.014    2
    0.002   -0.004     0.002    0
    1.042   -0.716     1.744    2
   -2.212   -0.821     0.019    1
    1.107   -0.765    -1.940    1
    0.996    2.523     0.006    3
   -0.554    2.507    -0.869    3
   -0.537    2.497     0.911    3
    0.532   -0.365     2.641    3
    1.057   -1.806     1.738    3
    2.065   -0.339     1.736    3
```

```
COMPND      5261536
HETATM     1   C1   LIG     1      -0.027    2.146     0.014   1.00   0.00
HETATM     2   SN2  LIG     1       0.002   -0.004     0.002   1.00   0.00
HETATM     3   C3   LIG     1       1.042   -0.716     1.744   1.00   0.00
HETATM     4   CL4  LIG     1      -2.212   -0.821     0.019   1.00   0.00
HETATM     5   CL5  LIG     1       1.107   -0.765    -1.940   1.00   0.00
HETATM     6   1H1  LIG     1       0.996    2.523     0.006   1.00   0.00
HETATM     7   2H1  LIG     1      -0.554    2.507    -0.869   1.00   0.00
HETATM     8   3H1  LIG     1      -0.537    2.497     0.911   1.00   0.00
HETATM     9   1H3  LIG     1       0.532   -0.365     2.641   1.00   0.00
HETATM    10   2H3  LIG     1       1.057   -1.806     1.738   1.00   0.00
HETATM    11   3H3  LIG     1       2.065   -0.339     1.736   1.00   0.00
END
```

Note that the absorber does not need to be at (0,0,0), nor does it need to be the first item in the list.

METHYLTIN DATA

Shown here are data collected on dimethyltin dichloride and on monomethyl trichloride. Each has methyl (CH_3^{1-}) and chloride ligands. They differ only in the numbers of ligands, as shown to the right. The difference in numbers of C and Cl scatterers leads to the differences in data shown below.

A SIMPLE FITTING MODEL

ARTEMIS
EXAFS DATA ANALYSIS
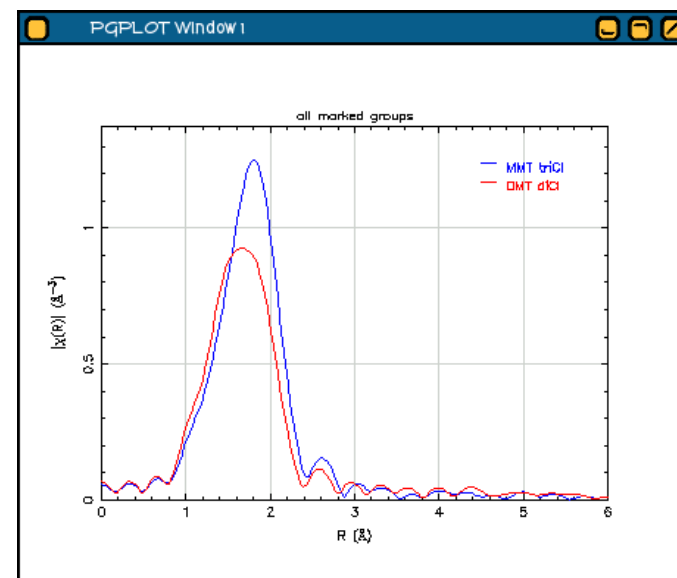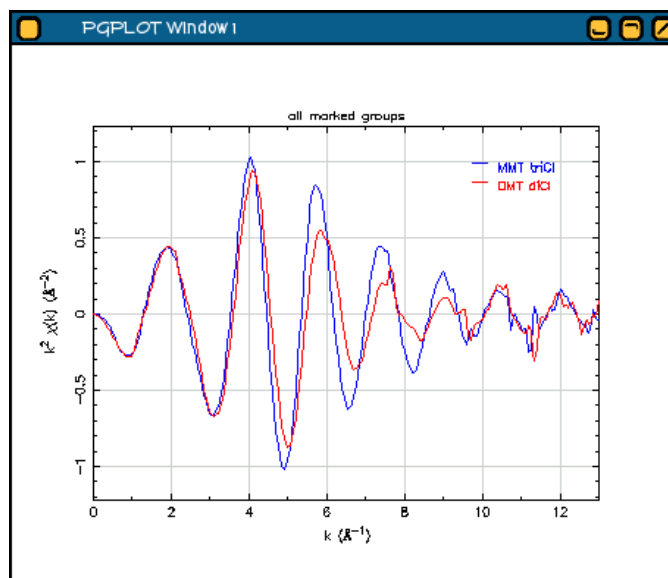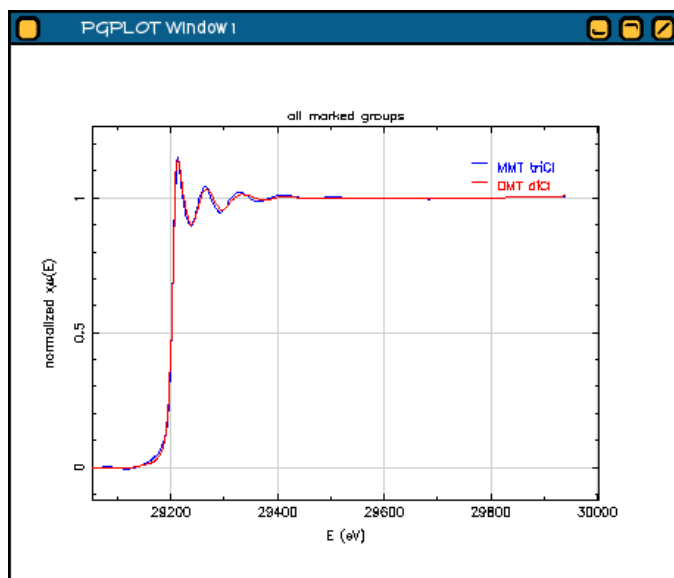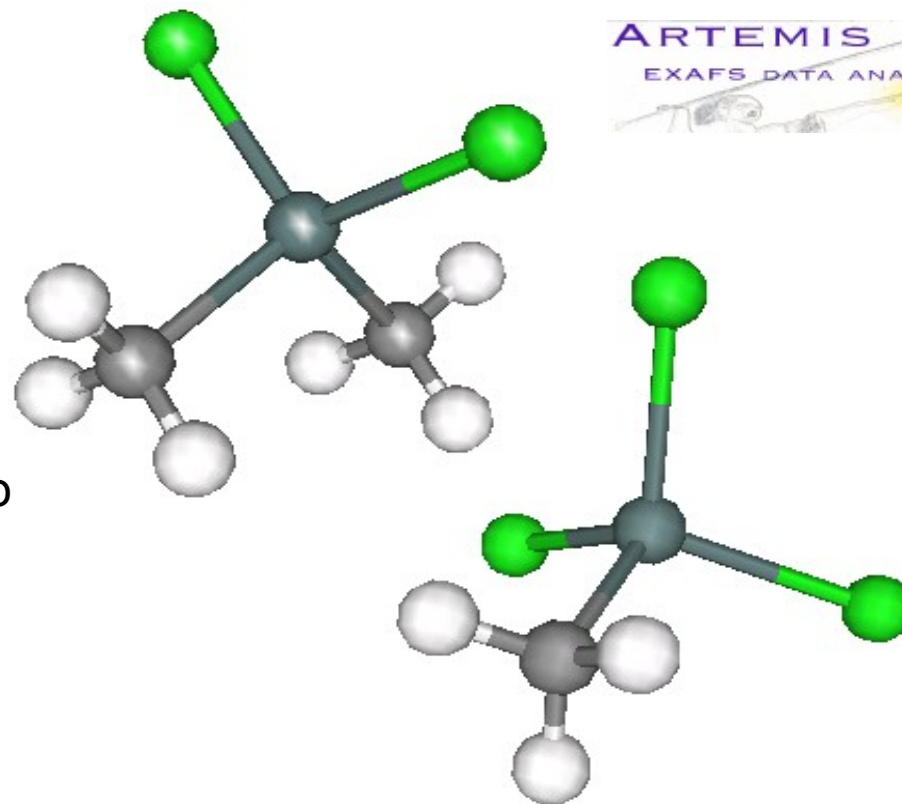
Along with an amplitude parameter (which will represent S_0^2) and an energy shift parameter, I will need parameters for ΔR and σ^2 for each kind of scatterer. Since C and Cl bond differently and have different masses, we cannot reasonably share ΔR or σ^2 parameters between these two scatterers.

AN ARTEMIS EXAMPLE: METHYL TIN
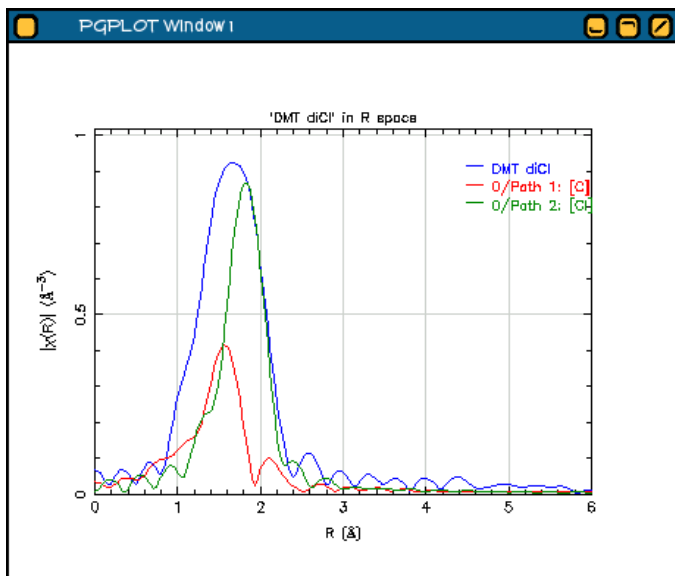
Result of Simple Fit

The results of this fit are something of a disaster. Although the plot of the fit looks OK, the results for the parameters are very problematic. The amplitude term is much larger than one, error bars are huge, σ^2 are huge, correlations between parameters are quite high, and there are very high correlations between terms that we expect not to be highly correlated (such as `delr_c` and `amp`).

The probable cause of this is the use of 6 fitting parameters in the presence of only 7 independent points.

k-range: [2:10.5]
R-range: [1:2.4]

$$\frac{2\,\Delta k\,\Delta R}{\pi} \approx 7$$

```
Guess parameters
   amp            =      3.2938350    +/-        1.3585440    (1.0000)
   enot           =      4.1733440    +/-        3.5366130    (0.0000)
   delr_c         =      0.1517810    +/-        0.1323330    (0.0000)
   ss_c           =      0.0526320    +/-        0.0338130    (0.0030)
   delr_cl        =     -0.0000870    +/-        0.0286490    (0.0000)
   ss_cl          =      0.0180140    +/-        0.0054830    (0.0030)

Correlations between variables:
      ss_cl and amp          -->   0.9366
    delr_cl and enot         -->   0.8658
     delr_c and amp          -->   0.7703
      ss_cl and delr_c       -->   0.6870
       ss_c and amp          -->   0.6514
       ss_c and enot         --> -0.5696
      ss_cl and ss_c         -->   0.5290
    delr_cl and delr_c       -->   0.4313
       ss_c and delr_c       -->   0.3387
     delr_c and enot         -->   0.2752
All other correlations are below 0.25
```
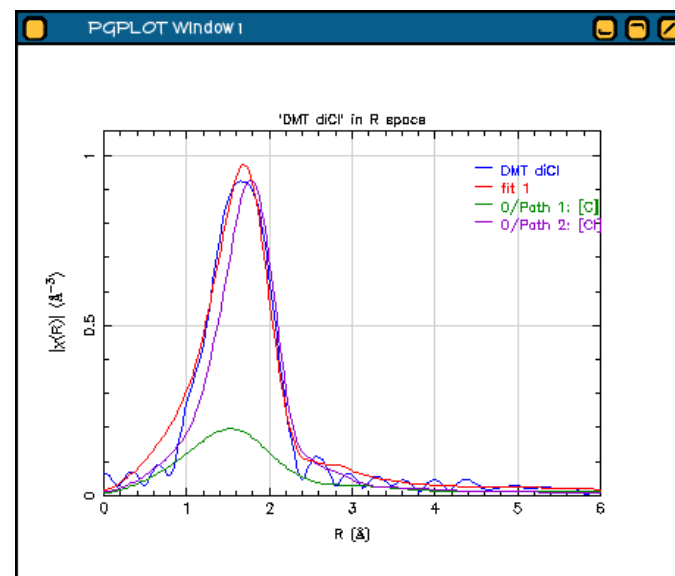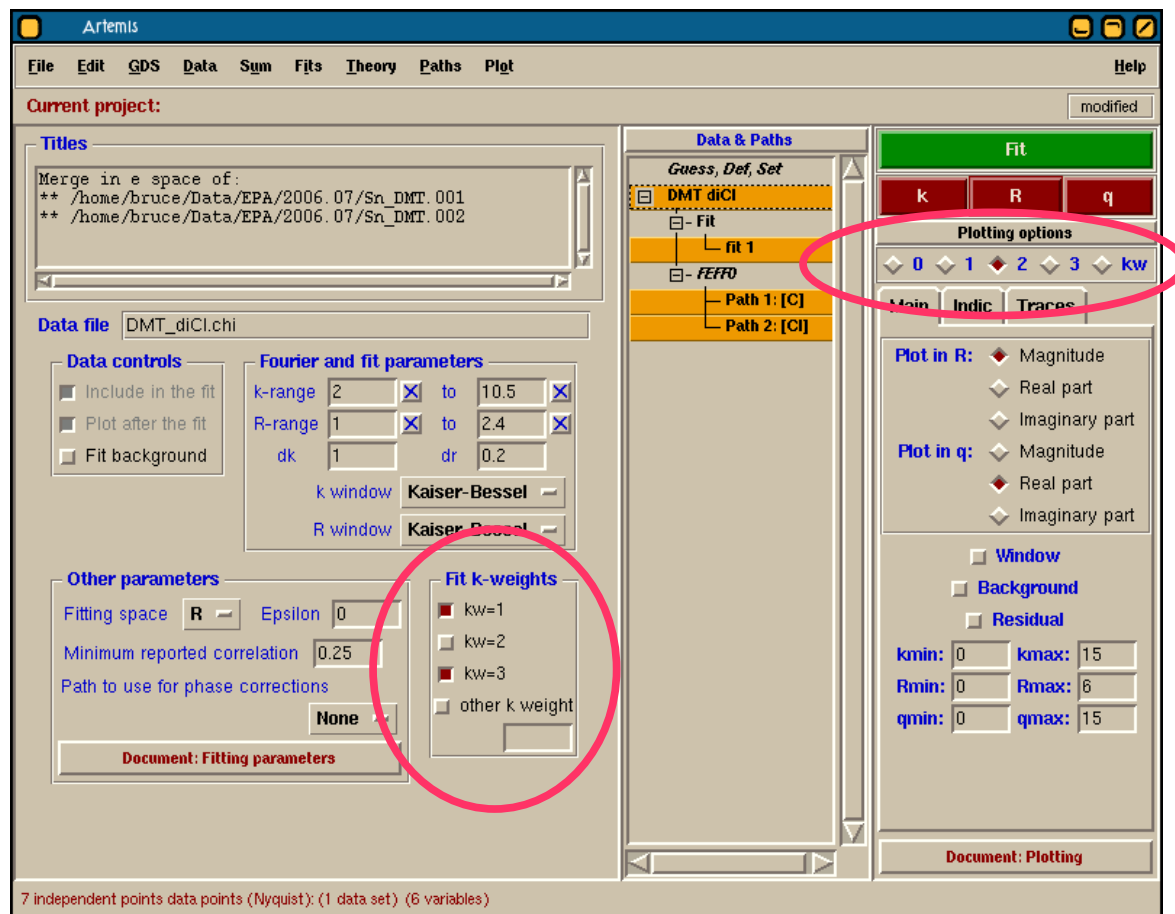
AN ARTEMIS EXAMPLE: METHYL TIN

MULTIPLE K−WEIGHT FITTING

Note that Artemis has two sets of controls for k-weighting. The buttons in the upper right corner control what k-weighting is used when plotting data. The buttons in the middle of the window control what k-wighting is used to evaluate the fit.

It is possible to use multiple k-weightings when evaluating the fit. This is done simply by checking more that one value.

The fit is optimized by varying the fitted parameters such that the chi-square expression shown below is minimized. In a multiple k-weight (MKW) fit, chi-square is evaluated for each k-weight and the sum of those chi-squares is minimized.

$$\chi^2 = \frac{N_{data}}{(N_{ip} \cdot \epsilon)} \sum_{i=rmin}^{rmax} \left[\Re\left(\chi_d(R_i) - \chi_t(R_i)\right)^2 + \Im\left(\chi_d(R_i) - \chi_t(R_i)\right)^2 \right]$$

AN ARTEMIS EXAMPLE: METHYL TIN

RATIONALE FOR MKW FITTING

Different parameters effect different regions of the data. That is, a poor value of a parameter causing a misfit between the data and theory will cause more misfit in some regions of the data than in others. For instance, the wrong value of   will result in a much larger misfit at high k due to its dependence upon k^2 in the EXAFS equation. Similarly, a wrong value for the energy shift will cause considerably more misfit at low k since its effect goes roughly as $1/k$.

Different values of k-weight emphasize different regions of the spectrum. A k-weight of 3 puts more emphasis at high-k while a k-weight of 1 tends to favor low-k.

By using multiple k-weights, we (in some sense) distribute the importance of the evaluation of chi-square over the entire k range. In this way, we make better use of the data available.

|	
S_0^2	same at all k
ΔR	high k, goes as k
σ^2	high k, goes as k^2
ΔR	low k, goes as $1/k$

Simple model + MKW

This is a considerable improvement. The amplitude term is a bit large, but consistent with 1. The other parameters are all of reasonable sizes. The correlations are a bit large, and there remains a bit of correlation between phase and amplitude terms. All in all, this is progress.

```
Guess parameters
  amp           =      1.2949470   +/-       0.2802790   (1.0000)
  enot          =      4.3574080   +/-       2.4561970   (0.0000)
  delr_c        =     -0.0538830   +/-       0.0328090   (0.0000)
  ss_c          =      0.0031840   +/-       0.0048510   (0.0030)
  delr_cl       =      0.0229030   +/-       0.0224470   (0.0000)
  ss_cl         =      0.0062180   +/-       0.0033170   (0.0030)


Correlations between variables:
       ss_cl and ss_c       -->  0.8857
     delr_cl and enot       -->  0.8655
       ss_cl and amp        -->  0.8430
      delr_c and enot       -->  0.8213
     delr_cl and delr_c     -->  0.8129
        ss_c and amp        -->  0.7890
      delr_c and amp        -->  0.7703
All other correlations are below 0.25
```


'DMT diCl' in R space
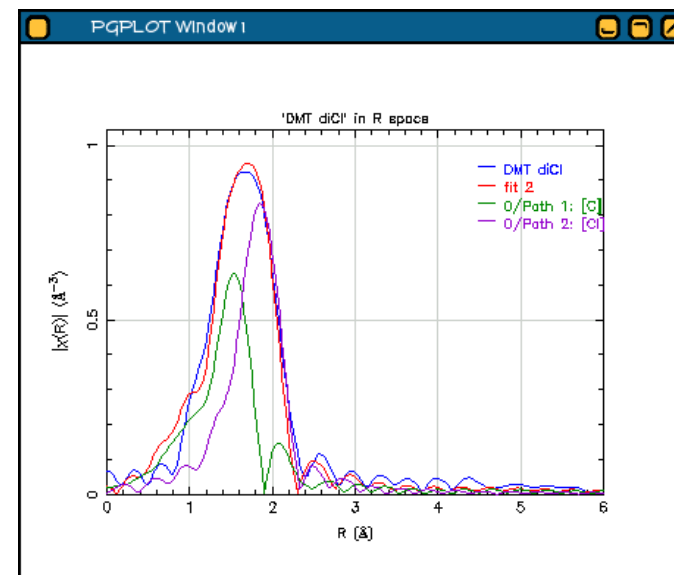
SETTING UP A MULTIPLE DATA SET FIT

I assert that the Sn-Cl and Sn-C bonds in mono- and dimethyl tin should behave the same. I can thus co-refine these two data sets using the same parameter set for both data, with the caveat that I set I use three Cl and 1 C scatterer for the monomethyl tin data.

When reading in the second data set, Artemis responds with the prompt show above. This is asking if you want to replace the first data set or to import a second, thus starting a multiple data set (MDS) fit. In the case, click "New". This results in a new entry in the Data and Paths List.

In the Data menu, use the "Clone a Feff calculation" option to copy the Feff calculation to the new data set. Finally, change the N path parameters for the path under the monomethyl tin to 1 and 3 for the C and Cl paths.
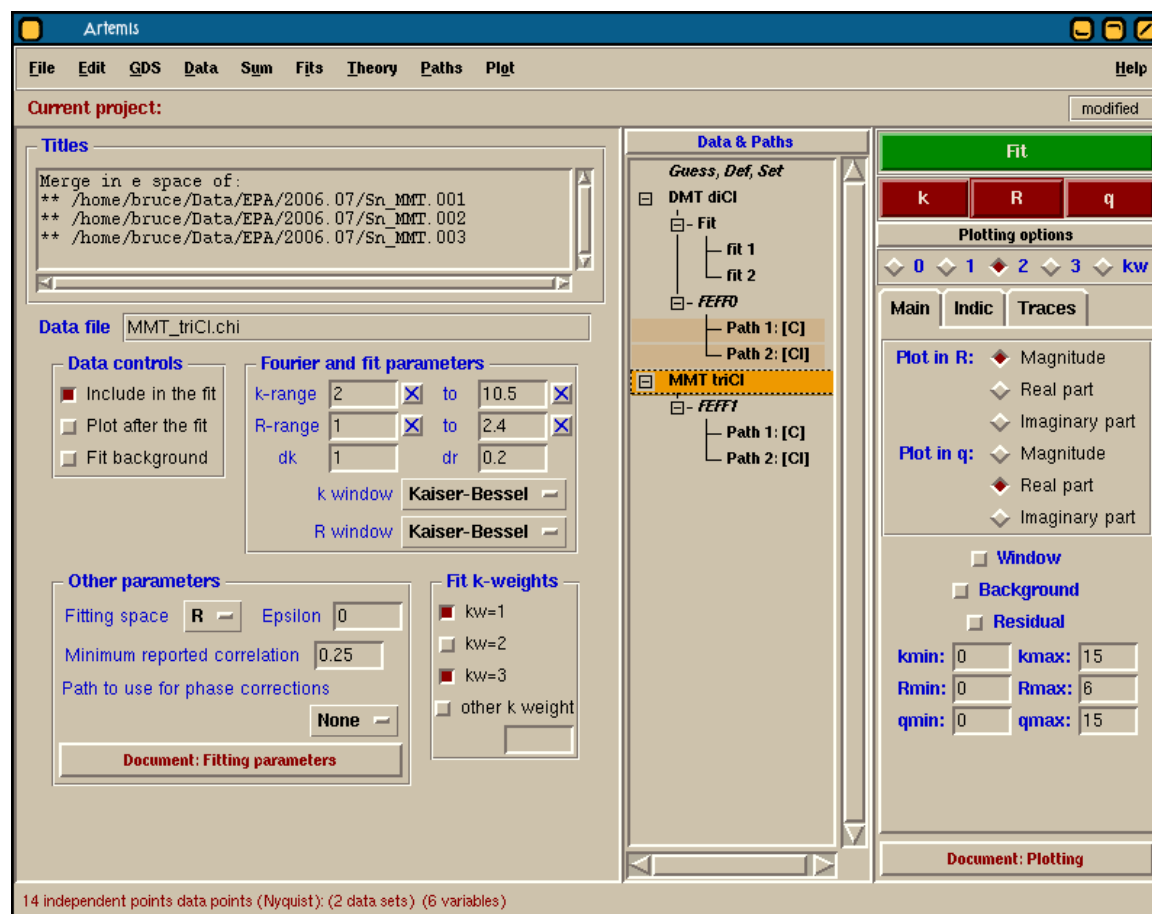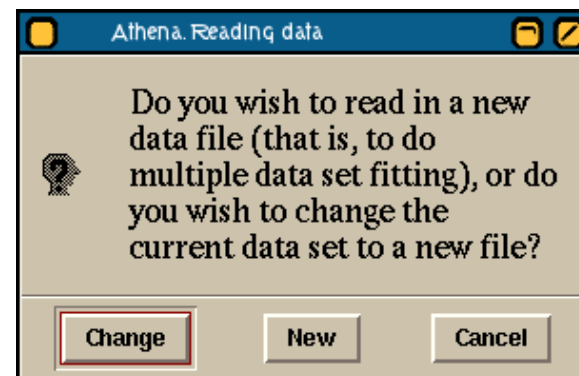
RESULT OF THE MDS FIT

This is an improvement. The error bars are smaller and, except for the slightly large amplitude, the values are all reasonable. The two correlations between *R* parameters and the amplitude remain, but this can be understood. The C and Cl paths are a bit out of phase. The extent to which they are out of phase depends on their *R* values. Also the extent to which they are out of phase determines how much they destructively interfere. It is for that reason that the *R* and amplitude values are correlated.

Should we worry about the amplitude value? Perhaps, but first we should verify that the data normalization was performed correctly in Athena, that the sample was adequately homogeneous, and that we trust Feff's energy loss model for Sn.

```
Guess parameters
   amp       =     1.3937930  +/-    0.1664850   (1.0000)
   enot      =     4.3480760  +/-    1.3725100   (0.0000)
   delr_c    =    -0.0445910  +/-    0.0205190   (0.0000)
   ss_c      =     0.0038470  +/-    0.0028630   (0.0030)
   delr_cl   =     0.0344070  +/-    0.0120300   (0.0000)
   ss_cl     =     0.0072070  +/-    0.0017110   (0.0030)


Correlations between variables:
    delr_cl and enot      -->   0.8697
      ss_cl and amp       -->   0.8453
      ss_cl and ss_c      -->   0.8364
    delr_c  and enot      -->   0.7974
    delr_cl and delr_c    -->   0.7790
    delr_c  and amp       -->   0.7703
      ss_c  and amp       -->   0.7495
All other correlations are below 0.25
```
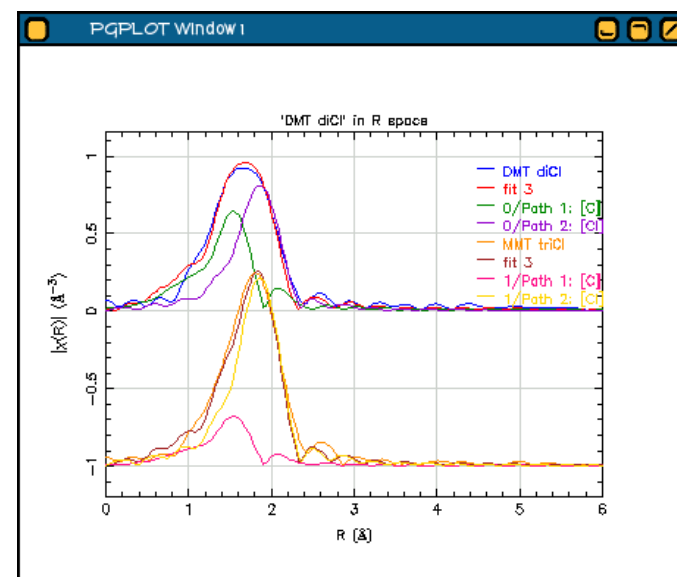
WHY MDS IS A GOOD IDEA

Multiple data set fitting is one of the most powerful tools provided by Artemis and Ifeffit. Many experiments can benefit by MDS fitting. Any experiment for which an extrinsic parameter is varied can be benefit. Some example include:

- temperature series, in which σ^2 values should behave like an Einstein (or single frequency) oscillator
- pressure series, in which σ^2 values are related to a Gruneissen parameter
- concentration or doping sequences
- certain chemistry changes, such as pH

Of course, the caveat that the phase or state of the sample cannot change through the series.

For any physical parameter for which you think it is reasonable that it remain the same throughout the sequence, your analysis will benefit by MDS fitting. MDS fitting provide more information to the fit without adding a commensurate number of new parameters. In the methyl tin example, I **doubled** the information content without adding a single new parameter!

The trick that made this work is Artemis & Ifeffit's ability to encode constraints on the parameters. In this case, I constrained the parameters for the C and Cl scatterers to be the same for all data sets. The math expression capability allows you to create arbitrarily sophisticated constraints.

BRINGING IT ALL TOGETHER

In this demonstration we saw examples of
- multiple k-weight fitting
- multiple data set fitting
- constraints between parameters and between data sets

These advanced features can be combined with other capabilities, including
- multiple Feff calculations
- the use of math expressions to build sophisticated constraints
- incorporating imprecise prior knowledge using restraints.

All of these tool are available in Artemis and Ifeffit and using them wisely allows you to make the best possible use of your EXAFS data. .

For more information on this and other topics related to Artemis, Ifeffit, and EXAFS analysis, check out http://xafs.org. And follow the links at xafs.org to the Ifeffit mailing list. The mailing list is a great place to ask questions about the software and about absorption spectroscopy in general. The discussion has a very high signal-to-nose ratio and includes contributions from a a large number of EXAFS experts.

AN ARTEMIS EXAMPLE: METHYL TIN