Advanced Topics in EXAFS Analysis

Bruce Ravel

Synchrotron Methods Group, Ceramics Division
Materials Science and Engineering Laboratory
National Institute of Standards and Technology

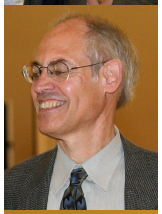
APS XAFS Summer School
July 6–10, 2009

Copyright

Acknowledgements

Matt Newville, of course. Without IFEFFIT none of us would be having this much fun

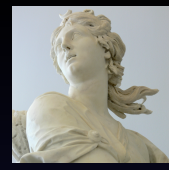
Shelly Kelly, bug finder extraordinare and progenitor of several examples in this talk



John Rehr and his group. If we didn't have fun with FEFF, we wouldn't have fun with IFEFFIT

Ed Stern, for teaching us so well and for getting all this XAS stuff started in the first place

- The many users of my software: without years of feedback and encouragement, my codes would suck way more than they do
- The folks who make the great software I use to write my codes: Perl, Perl/Tk, Emacs, The Emacs Code Browser, Subversion
- The folks who make the great software used to write this talk: LaTeX, Beamer, Avogadro, The Gimp, Gnuplot

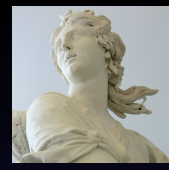


This Talk

This is NOT the introductory talk

- I assume you are a veteran of many XAS experiments and that you already have your own data that you care about.
- I assume you are familar with the EXAFS equation.
- I assume you understand XAS data processing and have done some EXAFS analysis.
- Some familiarity with IFEFFIT or ARTEMIS will help.

The audience for this talk is interested in advanced techniques which will improve their use of their EXAFS data.



The "Multiples"

Multiple k-weight Corefinement of the data using *multiple* values of k-weighting in the Fourier Transform

Multiple Feff Calculations Using *multiple* runs of the FEFF program to generate the theory used in your fitting model

Multiple Data Sets Corefinement of *multiple* data sets – this may be data measured at multiple edges, multiple temperatures, etc.

Constraints Between Parameters At the heart of an EXAFS fitting model are the relationships imposed between fitting parameters

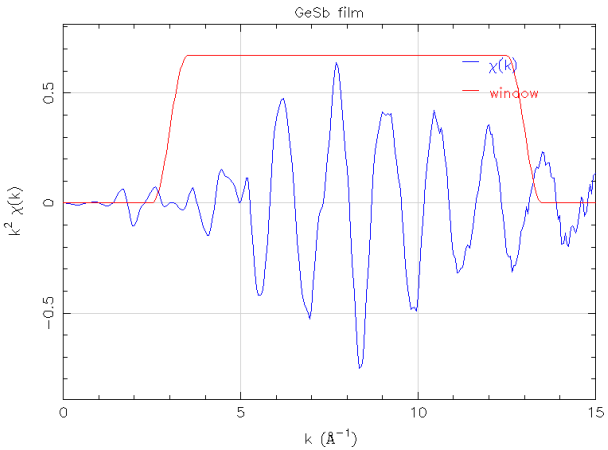
Restraints on Parameters Application of imperfect knowledge to influence the evaluation of a fit

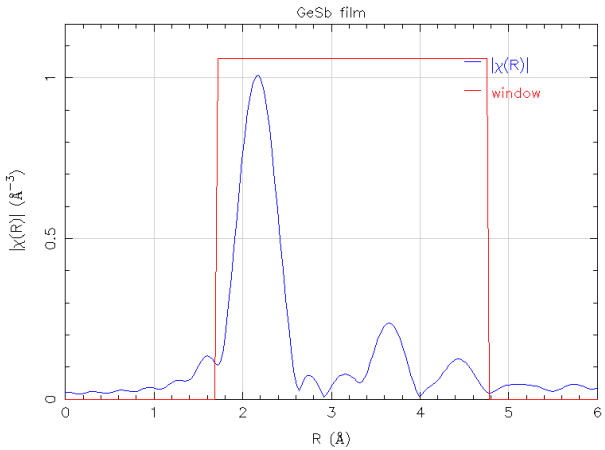
Using the "multiples"

All of these are implemented in ARTEMIS.

Information Content of EXAFS (I)

Sometimes, we have *beautiful* data. This is the merge of 5 scans on a 50 nm film of GeSb on silica, at the Ge edge and measured in fluorescence at NSLS X23a2.





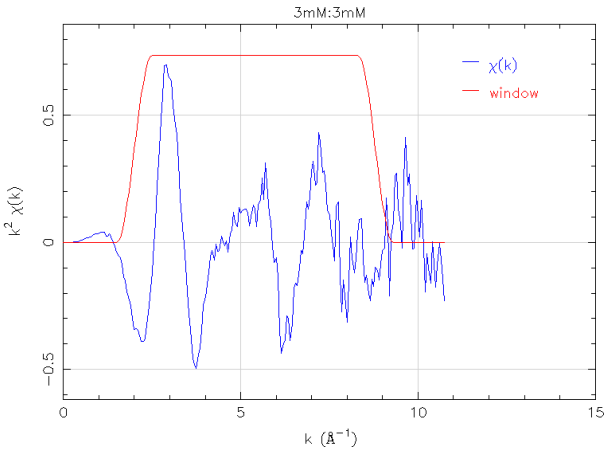
Here, I show a Fourier tranform window of [3 : 13] and I suggest a fitting range of [1.7 : 4.7]. Applying the Nyquist criterion:

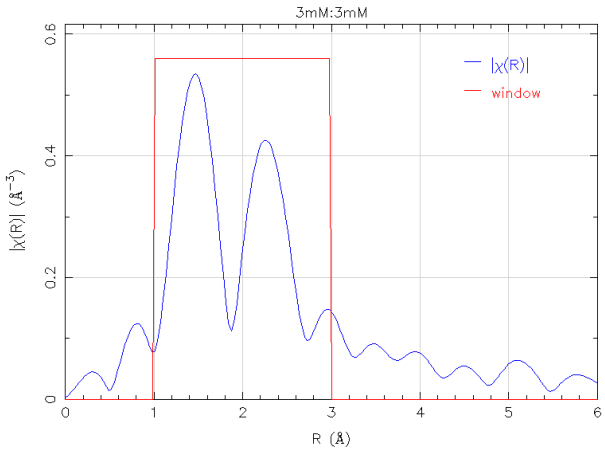
$$N_{idp} \approx \frac{2\Delta k \Delta R}{\pi} \approx 19$$

This gives us an upper bound of the information content of that portion of the EXAFS spectrum.

Information Content of EXAFS (II)

Sometimes, we have less-than-beautiful data. This is the merge of 42 scans on a solution containing 3 mM of Hg bound to a synthetic DNA complex, measured in fluorescence at APS 20BM.





Here, I show a Fourier tranform window of [2 : 8.8] and I suggest a fitting range of [1 : 3]. Applying the Nyquist criterion:

$$N_{idp} \approx \frac{2\Delta k \Delta R}{\pi} \approx 8$$

This talk discusses strategies for dealing with severely limited information content.

What is This Nyquist Criterion?

Given that we apply Fourier analysis to $\chi(k)$, we can treat EXAFS as a signal processing problem. If

- The signal is ideally packed **and**
- The error in the fitting parameters is normally distributed **and**
- We understand and can enumerate all sources of error **and**
- We know the theoretical lineshape of our data **then**

$$N_{idp} \approx \frac{2\Delta k \Delta R}{\pi}$$

where, for EXAFS, Δk is the range of Fourier transform and ΔR is the range in R over which the fit is evaluated.

Unfortunately ...

None of those conditions really get met in EXAFS. N_{idp} is, at best, an upper bond of the actual information content of the EXAFS signal.

Statistical Parameters: Definitions

IFEFFIT uses a Levenberg-Marquardt non-linear least-squares minimization, a standard χ^2 fitting metric, and a simple definition of an R-factor:

$$\chi^2 = \frac{N_{idp}}{\epsilon N_{data}} \sum_{i=min}^{max} \left[\mathrm{Re}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 + \mathrm{Im}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 \right] \quad (1)$$

$$\chi^2_\nu = \frac{\chi^2}{\nu} \quad (2)$$

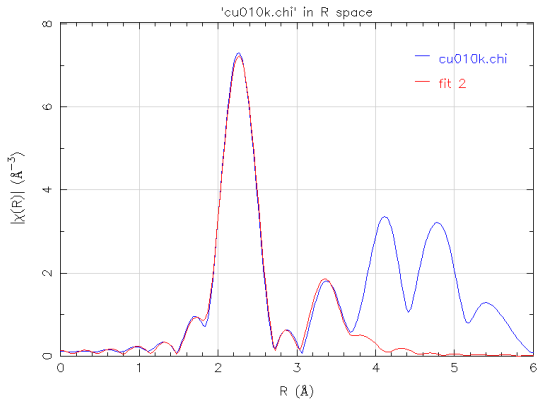
$$\nu = N_{idp} - N_{var} \quad (3)$$

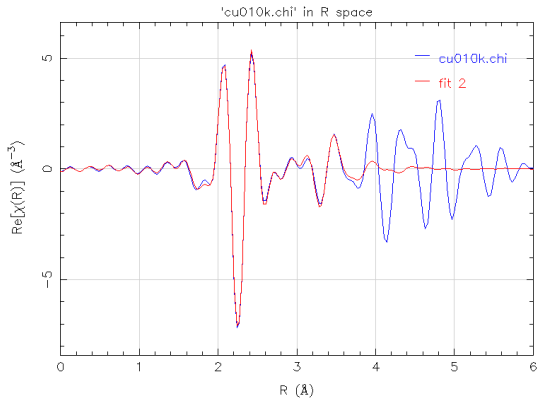
$$\epsilon = \text{measurement uncertainty}$$

$$\mathcal{R} = \frac{\sum_{i=min}^{max} \left[\mathrm{Re}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 + \mathrm{Im}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 \right]}{\sum_{i=min}^{max} \left[\mathrm{Re}\left(\chi_d(r_i)\right)^2 + \mathrm{Im}\left(\chi_d(r_i)\right)^2 \right]} \quad (4)$$

An Obviously Good Fit

Here is a fit to the first two shells of copper metal at 10 K





This is an unambigously good fit:

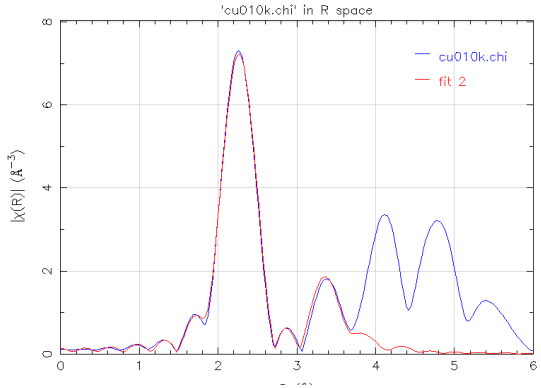
$\mathcal{R}$	0.0025
N_{idp}	16
ν	12
S_0^2	0.95(3)
E_0	5.98(36) eV
a	3.6072(26) Å
Θ_D	505(16) K

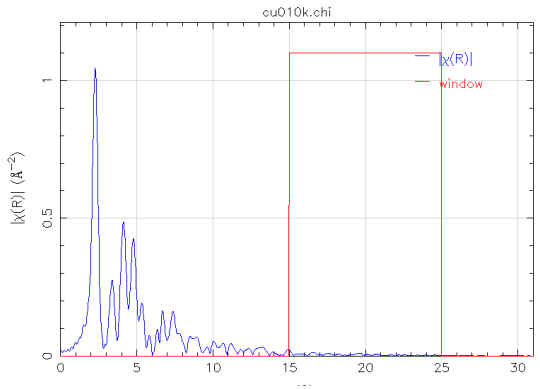
Yet $\chi^2 = 32.03$!

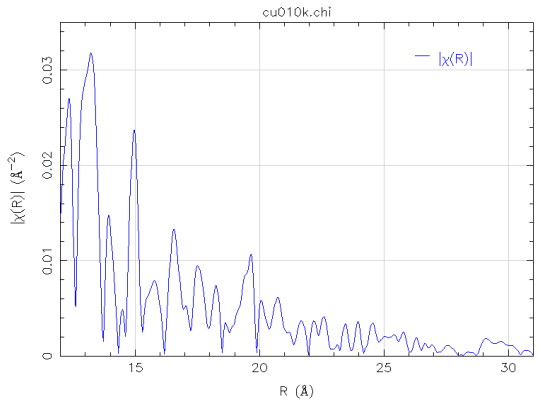
What's goin' on here?

Why is χ^2 for an obviously good fit so much larger than 1?

Statistical Parameters: Fit Evaluation







The determination of measurement uncertainty is, perhaps, a bit hokey in IFEFFIT. It is the average of the signal between 15 Å and 25 Å in the Fourier transform – a range that probably does not include much signal above the noise.

Is that signal between 15 Å and 25 Å in copper metal? Perhaps....

In any case, this method ignores the following:

- Approximations and errors in theory
- Sample inhomogeneity
- Detector nonlinearity
- Gremlins ;-)

Statistical Parameters: Interpretation I

OK then ... what is the implication of ϵ never being evaluated correctly by IFEFFIT?

1. χ^2_ν is always somewhere between big and enormous.
2. χ^2_ν is impossible to interpret for a *single* fit.
3. χ^2_ν **can** be used to compare different fits. A fit is improved if χ^2_ν is significantly smaller.
4. Error bars are taken from the diagonal of the covarience matrix. If χ^2_ν is way too big, the error bars will be way too small. The error bars reported by IFEFFIT have been scaled by $\sqrt{\chi^2_\nu}$.
5. Thus the error bars reported by IFEFFIT are of the "correct" size if we assume that the fit is a "good fit".

Statistical Parameters: Interpretation II

How do we know if a fit is "good"?

- The current fit is an improvement over the previous fit if χ^2_ν is sufficiently smaller.
- You should be suspicious of a fit for which N_{var} is close to N_{idp}, i.e. a fit for which ν is small.
- All variable parameters should have values that are physically defensible and error bars that make sense.
- The results should be consistent with other things you know about the sample.
- The R-factor should be small and the fit should closely overplot the data. (That was redundant. ☺)

Interpreting Error Bars

The interpretation of an error bar depends on the meaning of the parameter.

- A fitted σ^2 value of, say, 0.00567 ± 0.00654 is troubling. That result means suggests that σ^2 is quite ill-determined for that path and not even positive definite. Yikes!
- On the other hand, a fitted E_0 value of, say, 0.12 ± 0.34 is just fine. E_0 can be positive or negative. A fitted value consistent with 0 suggests you chose E_0 wisely back in ATHENA.

Outside Knowledge

Because the information content of the XAS measurement is so limited, we are forced to incoporate knowledge from other measurements into our data analysis and its interpretation.

- Other XAS measurements — for instance, the "chemical transferability" of S_0^2
- Diffraction tells us structure, coordination number, bond lengths, etc.
- Things like NMR, UV/Vis, and IR can tell us about the ligation environment of the absorber
- Common sense:
  - $R_{NN} \ncong 0.5\,\text{Å}$, $R_{NN} \ncong 4.0\,\text{Å}$
  - $\sigma^2 \not< 0\,\text{Å}^2$
- ... and anything else your (physical ‖ chemical ‖ biological ‖ whatever) intuition tells you

The Path Expansion

IFEFFIT is used to evaluate the EXAFS equation:

$$\chi(k,\Gamma) = \mathrm{Im}\left(\frac{(N_\Gamma S_0^2) F_\Gamma(k)}{2\,kR_\Gamma^2} e^{i(2kR_\Gamma + \Phi_\Gamma(k))} e^{-2\sigma_\Gamma^2 k^2} e^{-2R_\Gamma/\lambda(k)} \right) \quad (5)$$

$$\chi_{\mathrm{theory}}(k) = \sum_\Gamma \chi(k,\Gamma)$$

$$R_\Gamma = R_{0,\Gamma} + \Delta R_\Gamma \quad (6)$$

$$k = N\sqrt{(E_0 - \Delta E_0)} \quad (7)$$

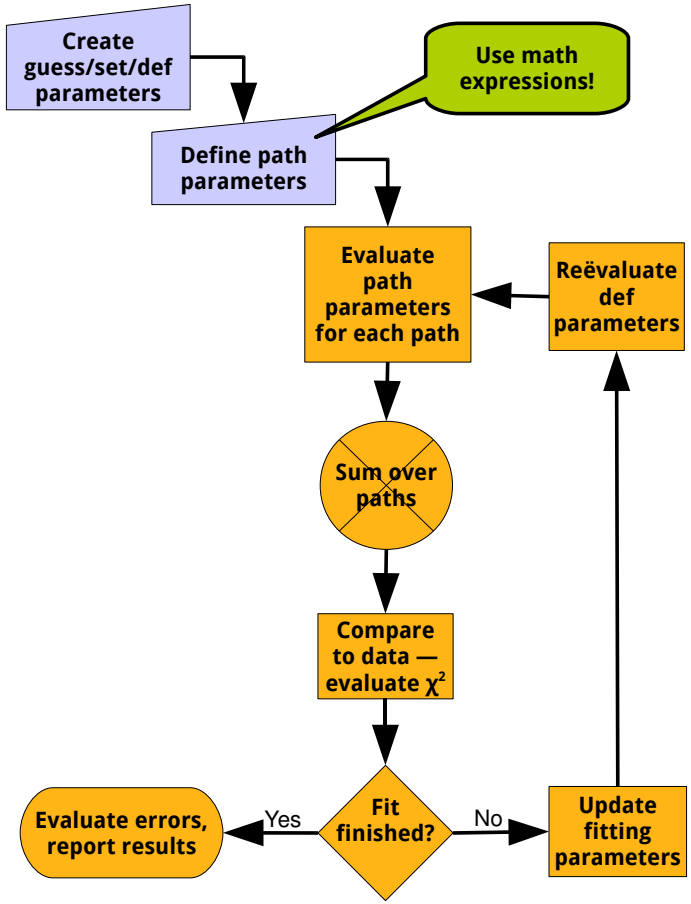
$\chi_{\mathrm{theory}}(k)$ is the function that is fit to data by varying the fitting parameters using theory from FEFF (the terms in brown).

In IFEFFIT the terms in light blue are not themselves the fitting parameters. They are written in terms of the actual fitting parameters.

Flow control in Ifeffit

Every trick in this talk exploits the fact that IFEFFIT introduces this layer of abstraction between the path parameters and the parameters of the fit.

Virtually any clever idea you have for describing your data can be expressed using IFEFFIT's math expressions.



k-Dependence of Different Parameters

Let's look at the EXAFS equation again:

$$\chi(k,\Gamma) = \mathrm{Im}\left(\frac{(N_\Gamma S_0^2) F_\Gamma(k)}{2\, kR_\Gamma^2} e^{i(2kR_\Gamma + \Phi_\Gamma(k))} e^{-2\sigma_\Gamma^2 k^2} e^{-2R_\Gamma/\lambda(k)} \right)$$

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

S_0^2	same at all k
ΔR	high k, goes as k
σ^2	high k, goes as k^2
ΔE_0	low k, goes as $\frac{1}{k}$

By using multiple k-weights, we hope to distribute the sensitivity of the evaluation of χ^2 over the entire k range and to make better use of the data available.

Evaluating A Multiple k-weight Fit

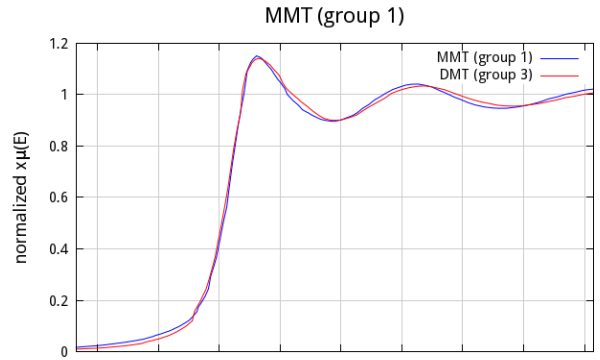
To evaluate an MKW fit, a χ^2 is evaluated for each value of k-weighting used in the fit.

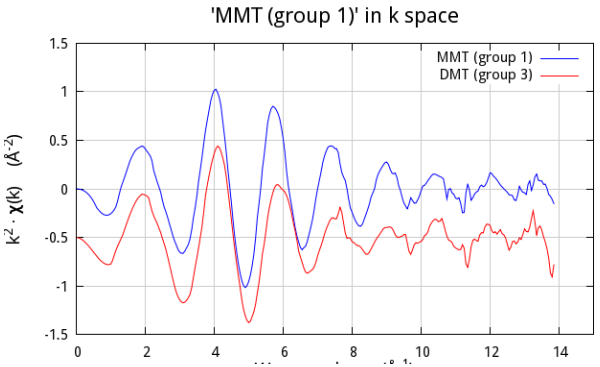
$$\chi^2_k = \frac{N_{idp}}{\epsilon N_{data}} \sum_{i=min}^{max} \left[\mathrm{Re}\left(\chi_{d,k}(r_i) - \chi_{t,k}(r_i)\right)^2 + \mathrm{Im}\left(\chi_{d,k}(r_i) - \chi_{t,k}(r_i)\right)^2 \right]$$

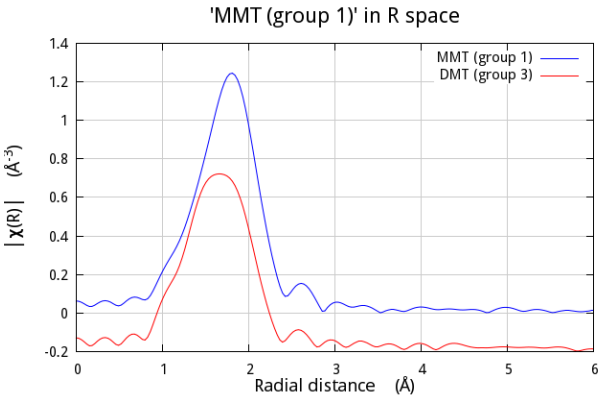
$$\chi^2_{total} = \sum_{\text{all k-weights}} \chi^2_k \tag{8}$$

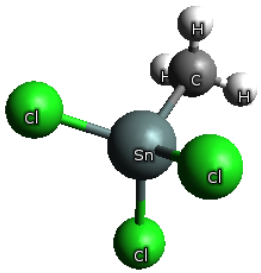
It's that simple!

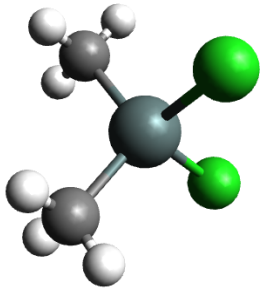
Example: Methyltin in Solution





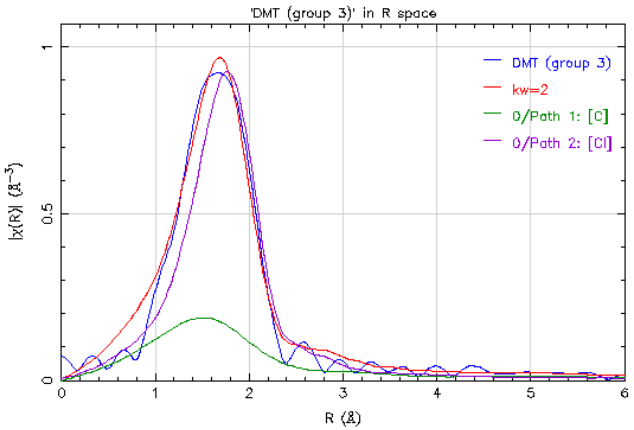






```
TITLE     dimethyltin dichloride
HOLE      1       1.0
CONTROL   1       1       1       1
PRINT     1       0       0       0
RMAX      6.0
POTENTIALS
*    ipot    Z  element
       0   50   Sn
       1   17   Cl
       2    6   C
       3    1   H
ATOMS
*   x         y         z
  -0.027    2.146     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
```

Example: Dimethyltin Fit with kw=2




```
Fitting statistics
  Independent points           : 7.26171
  Number of variables          : 6
  Chi-square                   : 2360.88
  Reduced Chi-square           : 1871.16
  R-factor                     : 0.01194
  Measurement uncertainty (k)  : 0.00019
  Measurement uncertainty (R)  : 0.00279

Guess parameters +/- uncertainties
  amp     =  3.2982190 +/- 1.4696810
  enot    =  3.9554520 +/- 3.9062430
  delr_c  =  0.1479970 +/- 0.1487560
  ss_c    =  0.0551050 +/- 0.0394380
  delr_cl = -0.0012190 +/- 0.0316440
  ss_cl   =  0.0179940 +/- 0.0059550

Correlations between variables:
         amp and ss_cl   -->  0.9375
        enot and delr_cl -->  0.8746
         amp and delr_c  -->  0.7643
      delr_c and ss_cl   -->  0.6863
         amp and ss_c    -->  0.6666
        enot and ss_c    --> -0.5834
```

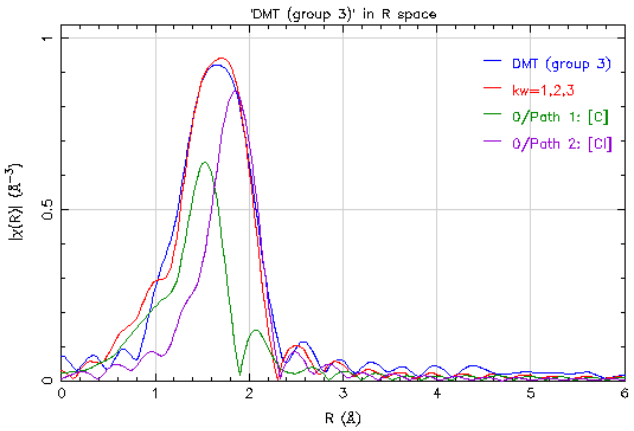
The fit *looks* ok, but it's actually kind of a mess.

- The S_0^2 value is way too big
- The σ^2 values are quite large
- One correlation is disturbingly high

The problem?

Severe information limits!

Example: Dimethyltin Fit with kw=1,2,3




```
Fitting statistics
  Independent points           : 7.26171
  Number of variables          : 6
  Chi-square                   : 4977.34
  Reduced chi-square           : 3944.89
  R-factor                     : 0.01395
  Measurement uncertainty (k)  : 0.00063
  Measurement uncertainty (R)  : 0.00110

Guess parameters +/- uncertainties
  amp     =  1.2735360 +/- 0.2885410
  enot    =  4.1254840 +/- 2.5937630
  delr_c  = -0.0573240 +/- 0.0371250
  ss_c    =  0.0028010 +/- 0.0055860
  delr_cl =  0.0198180 +/- 0.0250350
  ss_cl   =  0.0058720 +/- 0.0037930

Correlations between variables:
       ss_cl and ss_c    -->  0.8776
     delr_cl and enot    -->  0.8755
       ss_cl and amp     -->  0.8528
      delr_c and enot    -->  0.8331
        ss_c and amp     -->  0.8117
     delr_cl and delr_c  -->  0.7947
```

This is much better.

- S_0^2 and σ^2 are more like what we anticipate
- The correlations are a bit more comforting

Problems remain

The information is still strained, but MKW certainly helps!

Absorbing Atoms in Multiple Environments

Consider situations like these:

1. A crystal with the absorbing atom in *multiple* lattice positions
2. A metalloprotein with *multiple*, inequivalent active sites
3. An adsorbed metallic species that might be in *multiple* ligation environments
4. A physical mixture of *multiple* species, e.g. dirt
5. A thin film with *multiple* layers

A Feff input file has one-and-only-one absorbing site

A single FEFF input file and a single FEFF run cannot possibly be used to describe any of those situations. How can we make progress?

$YBa_2Cu_3O_7$: Multiple Lattice Positions

In $YBa_2Cu_3O_7$, copper occupies 2 sites. Site 1 is in a four-fold planar configuration. Site 2 is near the center of a square pyramid. The unit cell has 1 Cu1 and 2 Cu2 positions.

```
title YBCO: Y Ba2 Cu3 O7
space = P M M M
rmax  = 7.2   a=3.817 b=3.882  c=11.671
core  = cu1
atoms
! At.type   x        y       z       tag
   Y       0.5      0.5     0.5
   Ba      0.5      0.5     0.1839
   Cu      0        0       0       cu1
   Cu      0        0       0.3546  cu2
   O       0        0.5     0       O1
   O       0        0       0.1589  O2
   O       0        0.5     0.3780  O3
   O       0.5      0       0.3783  O4
```

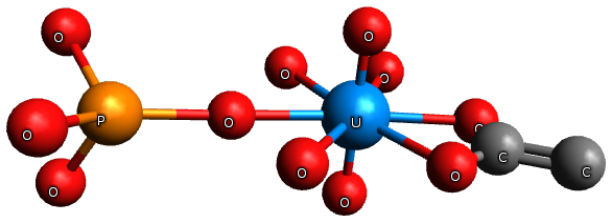
This is handled naturally in IFEFFIT after running FEFF twice:

```
guess  s0sqr = 0.9
path {
   file    site1/feff0001.dat
   label   1st path, site 1
   s02     s0sqr / 3
 }
## all subsequent paths for site 1
## have s02 * 1/3
```

```
path {
   file    site2/feff0001.dat
   label   1st path, site 2
   s02     2 * s0sqr / 3
 }
## all subsequent paths for site 2
## have s02 * 2/3
```

Uranyl Ion Absorbed to Biomass

A uranyl solution brought into equillibrium with biomass will proportionate in a pH-dependent manner among hydroxyl, phosphoryl, and carboxyl ligands.



- Uranyl species tend to have 5 or 6 equatorial O's
- Phosphoryl ligands are monodentate
- Carboxyl ligands are bidentate
- Hydroxyl's just dangle

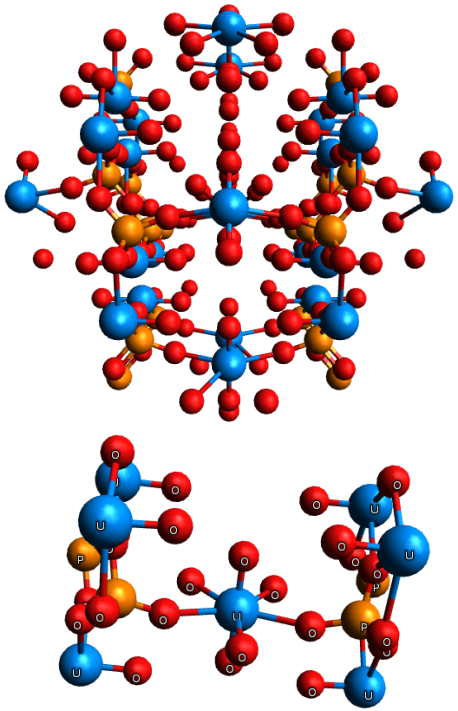
- Use crystalline triuranyl diphoshate tetrahydrate for the phosphoryl component
- Use crystalline sodium uranyl triacetate for the carboxyl component
- Use weights as fitting parameters to determine proportionation

There is no way to find a feff.inp file for that!

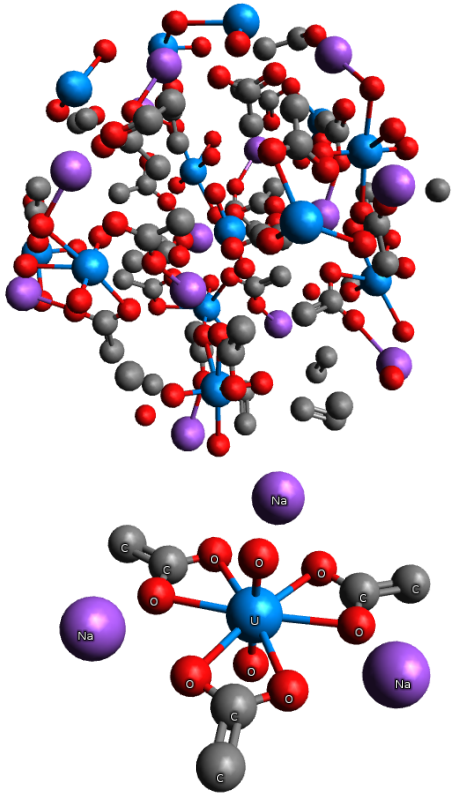
S. Kelly, et al. *X-ray absorption fine-structure determination of pH dependent U-bacterial cell wall interactions*, Geochim. Cosmochim. Acta (2002) **66**:22, 3875–3891

Using Crystal Analogs as Feff Structures

Triuranyl diphoshate tetrahydrate contains a monodentate U-P moiety.



Sodium uranyl triacetate contains a bidentate U-C moiety.



The moral of this story ...

The structure used in the FEFF calculation doesn't need to be "perfect". Close is usually good enough to get started.

Evaluating A Multiple Feff Calculation Fit

To evaluate an MFC fit, paths from each FEFF calculation are used in the sum over paths used to compute the theoretical $\chi(k)$.

$$\chi_{th}(k) = \sum_{\Gamma}^{\text{(all structures)}} \sum_{\Gamma}^{\text{(all included paths)}} \chi_{\Gamma}(k)$$

$$\chi^2 = \frac{N_{idp}}{\epsilon N_{data}} \sum_{i=min}^{max} \left[\text{Re}\left(\chi_d(r_i) - \chi_{th}(r_i)\right)^2 + \text{Im}\left(\chi_d(r_i) - \chi_{th}(r_i)\right)^2 \right]$$

Again, it's that simple!

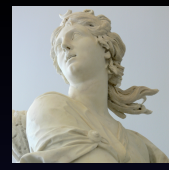
An Ensemble of Related Data Sets

Consider situations like these:

1. You have data at *multiple* temperatures using a cryostat and/or a furnace
2. You have data at *multiple* pressures from a high pressure cell
3. You have powders/films/solutions of *multiple* stoichiometries
4. You have data at *multiple* edges of the same material

Some parameters may be related across data sets

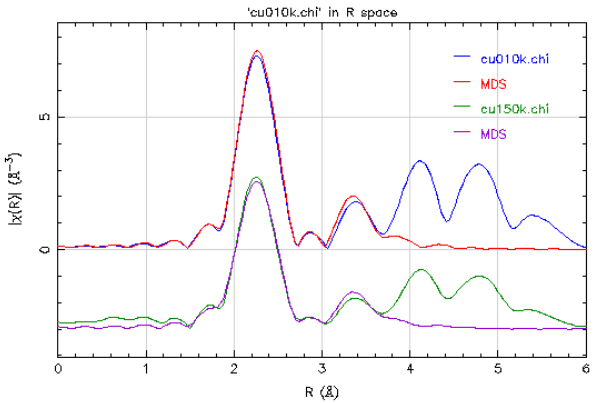
Corefining related data sets will dramatically increase the information content of the fit — each data set **is** an independent measurement — while not equivalently increasing the number of fitting parameters.



Example: Multiple Temperatures

I can corefine the copper metal data measured at many temperatures. Here is the fit we saw earlier, extended to include 10 K **and** 150 K data:

- S_0^2 and E_o are the same for both data sets
- All σ^2 parameters for all paths at both temperatures are computed from *one* variable Debye temperature.
- A linear dependence in temperature is assumed for the lattice expansion coefficient — $\alpha(T) = m \cdot T + b$



Twice as many independent points, only one more parameter!
The more data, the better!

Evaluating A Multiple Data Set Fit

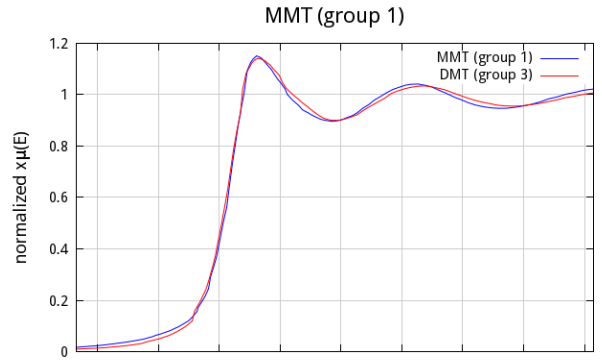
To evaluate an MDS fit, a χ^2 is evaluated for each data set

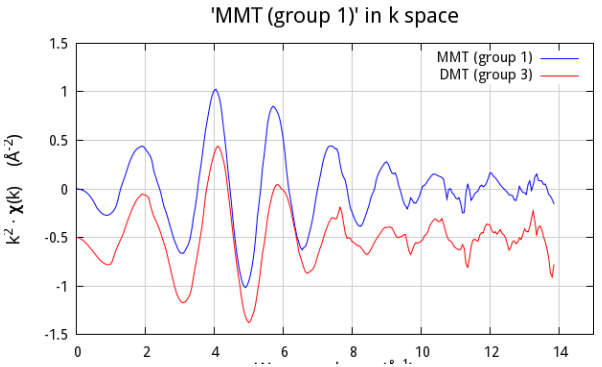
$$\chi^2_D = \frac{N_{idp}}{\epsilon N_{data}} \sum_{i=min}^{max} \left[\mathrm{Re}\left(\chi_{d,D}(r_i) - \chi_{t,D}(r_i)\right)^2 + \mathrm{Im}\left(\chi_{d,D}(r_i) - \chi_{t,D}(r_i)\right)^2 \right]$$

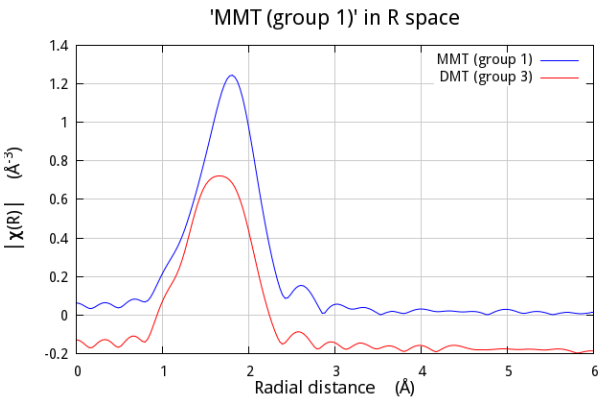
$$\chi^2_{total} = \sum_{\text{all data sets}} \chi^2_D \tag{9}$$

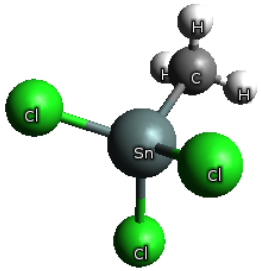
Yet again, the evaluation of the fitting metric is a trivial extension of the simplest case.

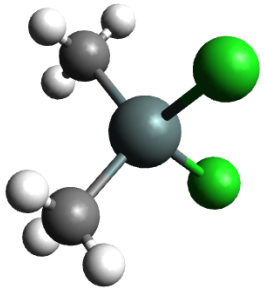
Example: Methyltin in Solution









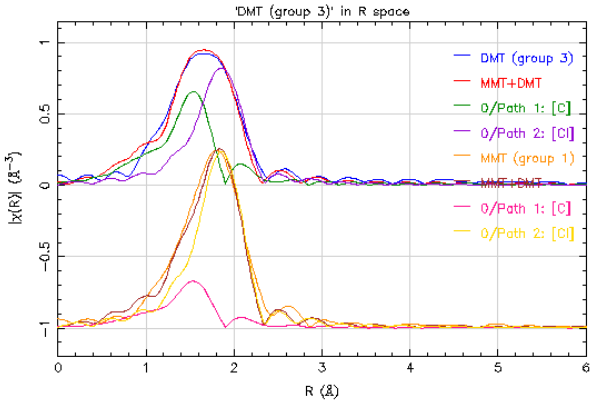


```
TITLE     dimethyltin dichloride
HOLE      1       1.0
CONTROL   1       1       1       1
PRINT     1       0       0       0
RMAX      6.0
POTENTIALS
*    ipot    Z  element
       0   50   Sn
       1   17   Cl
       2    6   C
       3    1   H
ATOMS
*   x          y          z
  -0.027     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
```

Example: Stoichiometry

I can corefine forms of methyltin, remembering that monomethyl tin has 1 Sn–C ligand and 3 Sn-Cl, while dimethyl tin has 2 and 2.

- S_0^2 and E_o are the same for both data sets
- I *assert* that the σ^2's for the Sn–C and Sn–Cl ligands are the same for di- and monomethyltin.
- Similarly, I *assert* the bond lengths are the same.

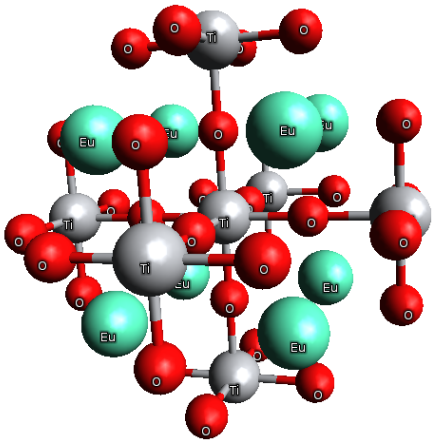


Twice the information, same number of parameters!

The simple assertion that the ligands are invarient between these samples adds considerable depth to the fitting model. Is this assertion correct? That is easily tested by lifting constraints on σ^2 and ΔR are comparing the fit results.

Example: Two Edges

$EuTiO_3$ is a regular cubic perovskite:

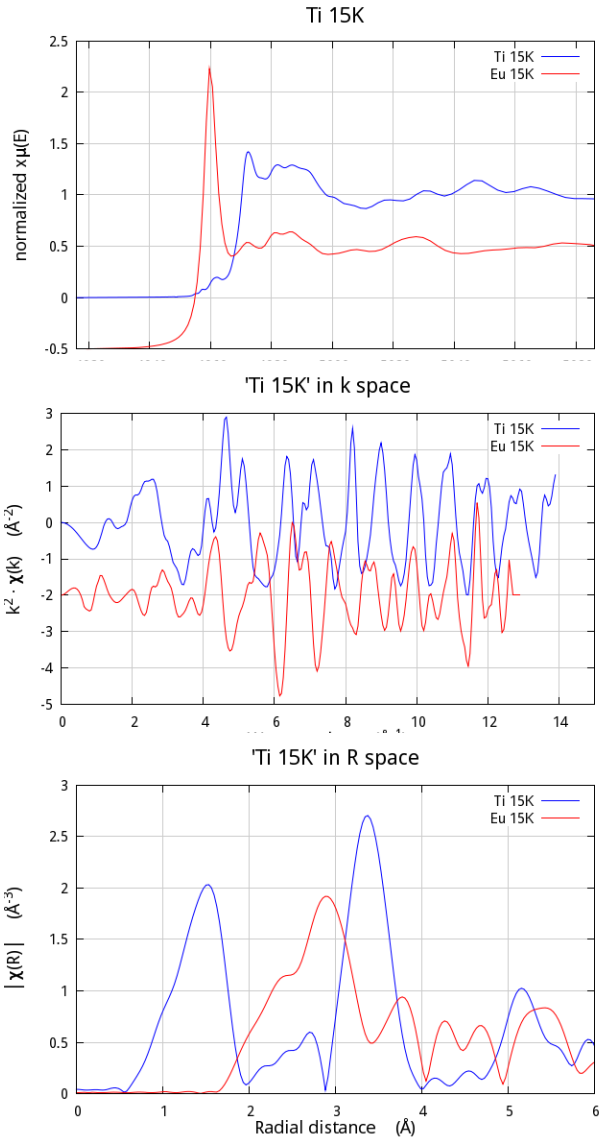


The data from the two edges share:

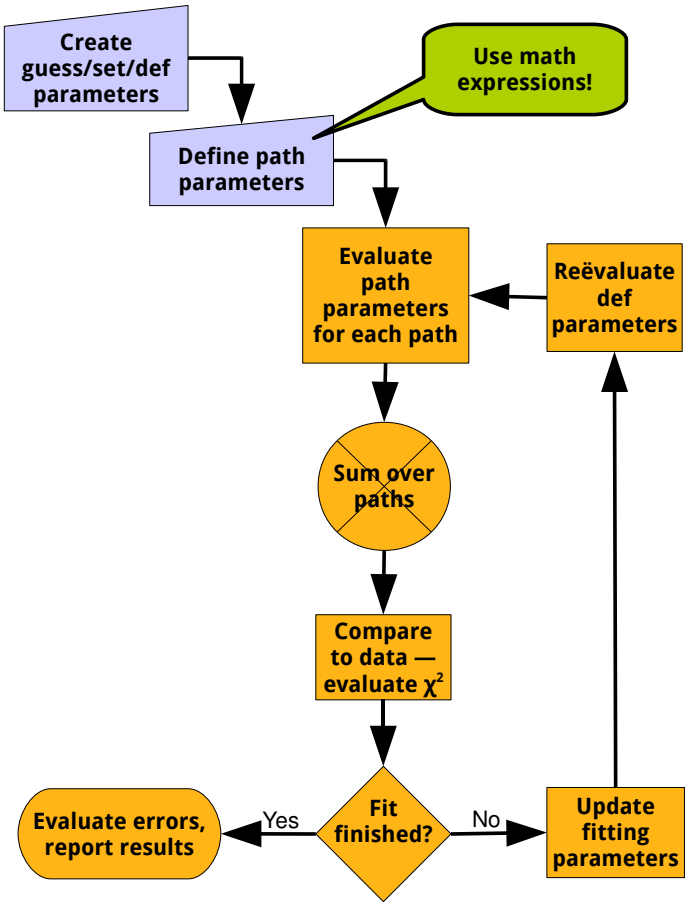
- A lattice constant
- An Eu–Ti σ^2

Other parameters must be independent for the two edges.

I have data from 15 K to 500 K, so I can combine *multiple* temperatures, *multiple* edges, **and** *multiple* FEFF calculations!



Building EXAFS Models



All of IFEFFIT's magic happens in the blue steps. The effective use of MFC or MDS fitting, begins with clever model building.

The Need for Constraints

Let's look at the EXAFS equation yet again:

$$\chi(k,\Gamma) = \mathrm{Im}\left(\frac{(N_\Gamma S_0^2) F_\Gamma(k)}{2\, kR_\Gamma^2} e^{i(2kR_\Gamma + \Phi_\Gamma(k))} e^{-2\sigma_\Gamma^2 k^2} e^{-2R_\Gamma/\lambda(k)} \right)$$

For every path used in the fit, you must somehow evaluate N, S_0^2, σ^2, ΔR, E_0.

That's 5 parameters per path, but even for the beautiful GeSb data we had fewer than 20 independent points.

Are we doomed?

The Simplest Constraints

Although each of N, S_0^2, σ^2, ΔR, E_0. must be evaluated for each path, they are not necessarily independent parameters for each path.

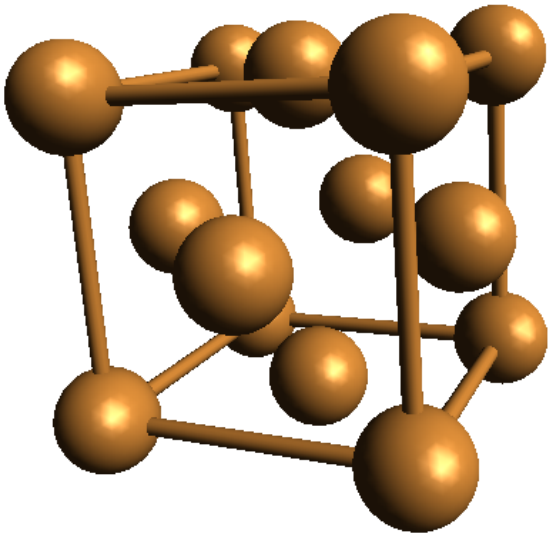
Consider the copper metal we have already seen in this talk:

- S_0^2 This is a parameter of the central atom and has something to do with the relaxation of electrons around the core-hole. In copper metal, S_0^2 is the same for all paths.
- E_0 In a single data set, single FEFF calculation fit, this parameter is used to align the wavenumber grids of the data and theory. In copper metal, E_0 is the same for all paths.

S_0^2 and E_0 represent the simplest kind of constraint — parameters that are the same for each path.

Slightly More Interesting Constraints

Copper metal also demonstrates simple constraints between paths involving math expressions:



ΔR As a highly symmetric, cubic metal, a volume expansion coefficient can be used to describe all path lengths. $\Delta R = \alpha * R_{eff}$

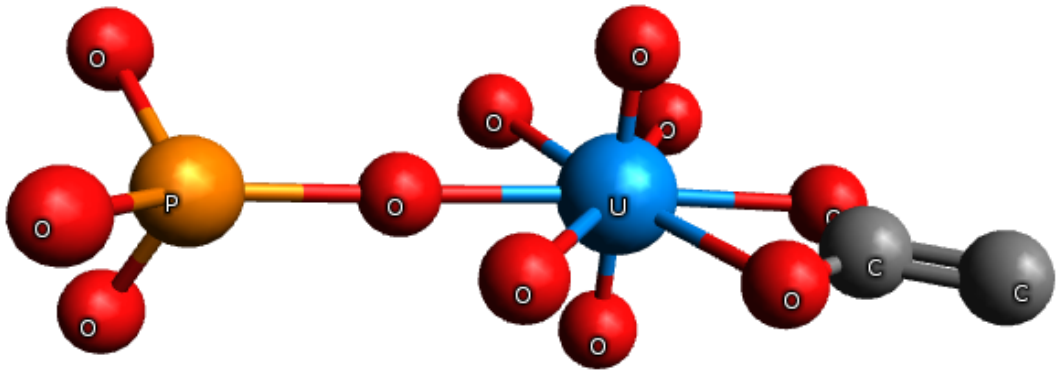
σ^2 As a monoatomic metal, the mean square deviations in each path length can be described by the Debye temperature. $\sigma^2 = \mathrm{debye}(T, \Theta_D)$

Path geometry

IFEFFIT is clever enough to use the correct values for R_{eff} (the path length used in the FEFF calculation) and the reduced mass as path parameter math expressions are evaluated for each path.

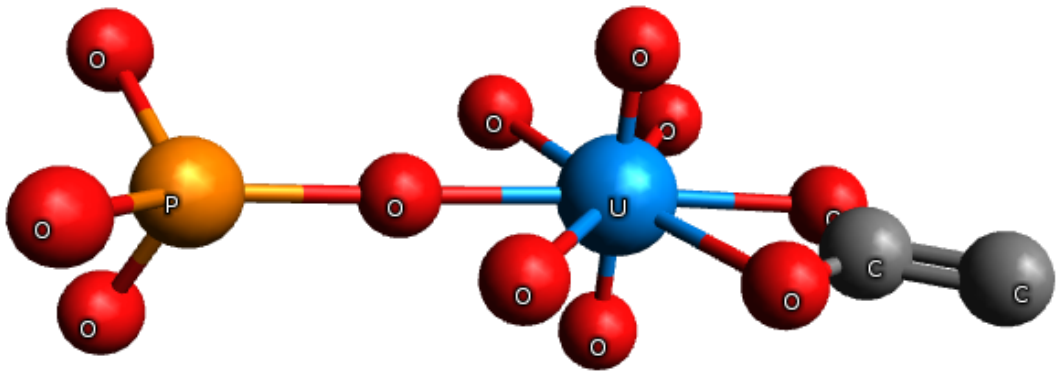
Model Building For Fun and Profit (I)

The uranyl problem requires multiple FEFF calculations. Making effective use of those calculations requires interesting constraints.



- The equatorial oxygen associated with a phosphoryl ligand is shorter than for a carboxyl ligand
- The posphoryl ligand is monodentate, thus $N_P = N_{short}$
- The carboxyl is bidentate, thus $N_c = N_{long}/2$
- If we assert that there are 6 equatorial oxygen atoms, then $N_{short} + N_{long} = 6$

Model Building For Fun and Profit (II)



Uranyl coordination parameters

```
set      n_eq      = 6
guess    n_short   = 3
def      n_long    = n_eq - n_short
def      n_p       = n_short
def      n_c       = n_long / 2

guess    ss_short  = 0.003
def      ss_long   = ss_short
```

We have described the coordination numbers and σ^2 for the equatorial oxygen atoms with a minimal number of guesses. The constraints on σ^2 and N_{eq} can be lifted easily by switching a set or a def to a guess.

Using Imperfect Knowledge

IFEFFIT allows the incoproration of imprecise prior knowledge by adding restraints in quadrature to the fitting metric.

$$\chi^2_{data} = \frac{N_{idp}}{\epsilon N_{data}} \sum_{i=min}^{max} \left[\mathrm{Re}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 + \mathrm{Im}\left(\chi_d(r_i) - \chi_t(r_i)\right)^2 \right]$$

$$\chi^2 = \chi^2_{data} + \sum_j \left[\frac{\lambda_{0,j} - \lambda_j}{\delta\lambda_j}\right]^2 \qquad (10)$$

λ_0	prior knowledge
λ	fitted value
$\delta\lambda$	confidence

The meaning of $\delta\lambda$

As $\delta\lambda \to \infty$, a restraint becomes unimportant.
As $\delta\lambda \to 0$, you admit no prior knowledge.

Restraints: A simple example

Suppose you have reason to believe that $0.6 < S_0^2 < 1.0$.

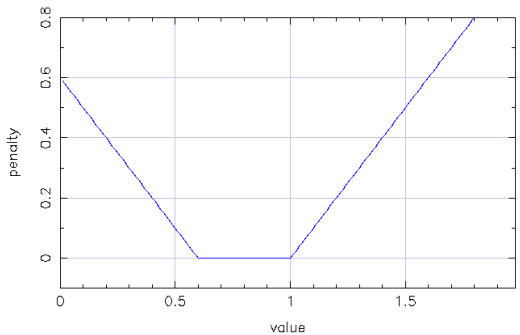
Enforce this with "hard-wall" boundaries

```
guess S0sqr = 0.8
path(1, S02 = max(0.6, min(1.0, S0sqr) ) )
```

If the fitted value of S_0^2 strays out of bounds, error bars cannot be properly calculated.

Apply a restraint, added in quadrature with χ^2_{data}

```
guess    S0sqr = 0.8
set      scale = 2000
restrain S0sqr_res = scale * penalty(S0sqr, 0.6, 1.0)
path(1, S02 = S0sqr)
```



S_0^2 is encouraged to stay in bounds to avoid a penalty to χ^2, but error bars can be properly evaluated when S_0^2 strays.

Restraints: A simple example (continued)

The assumption of "chemical transferability" of S_0^2 may be suspect, particularly if the known standard used to determine S_0^2 is prepared differently from the unknown.

Restrain S_0^2 to be like the standard

```
guess     S0sqr = 0.9
set       S0sqr_known = 0.876
set       scale = 2000
restrain  S0sqr_res = scale * (S0sqr - S0sqr_known)
```

Again, `S0sqr_res` is added in quadrature with χ^2.

How big should the `scale` be?

I don't have a good answer. The square root of χ^2 evaluated without the restraint seems to be a good size. In the end, it depends upon how much trust you place on the restraint.

Restraints: Bond Valence Sums

The Bond Valence relates the valence of an ion to its ligand bond lengths using empirical parameters as determined by Brown and Altermatt, Acta Cryst. B41 (1985) pp. 244–247:

$$V_i = \sum_{j=1}^{N} \exp\left(\frac{R'_{ij} - R_{ij}}{0.37}\right) \tag{11}$$

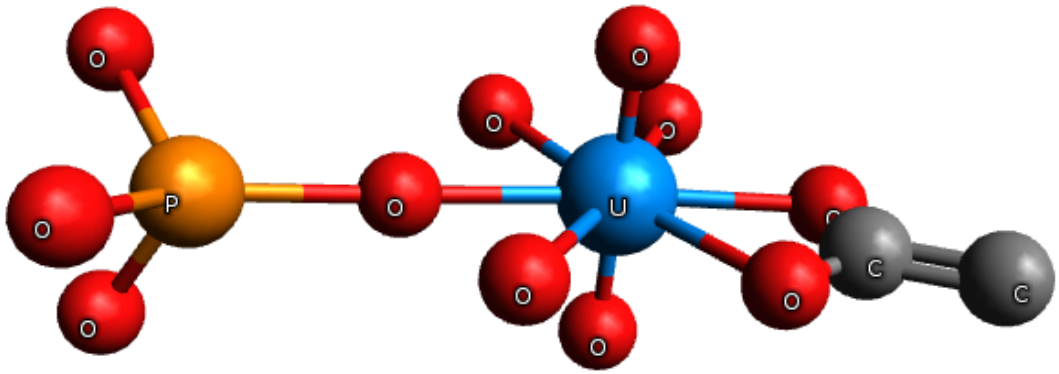
0.37 and R'_{ij} are empirical parameters, R'_{ij} is different for each kind of pair, i.e. Fe–S, Ni–O, etc. Tetrahedral coordination involves different distances and valence than octahedral.

Octahedral iron example

```
set       valence = 2
set       rij     = 1.734  # R'ij for Fe(2+)-O
set       rnot    = 2.14
set       scale   = 1000
guess     delr    = 0.0
restrain  bvs     = valence - 6 * exp( (rij - (rnot+delr)) / 0.37 )
```

Model Building For Fun and Profit (III)

From a literature survey, we know that the short and long equatorial oxygen bonds tend to be about 2.32 Å and 2.45 Å in uranyl complexes.



Uranyl coordination parameters

```
guess      r_short     = 2.32
guess      r_long      = 2.45
restrain   r_short_res = scale * penalty(r_short, 2.30, 2.34)
restrain   r_long_res  = scale * penalty(r_short, 2.43, 2.47)
```

These restraints encourage those distances to stay near their imprecisely known values.

The "Multiples"

Multiple *k*-weight Corefinement of the data using *multiple* values of *k*-weighting in the Fourier Transform

Multiple Feff Calculations Using *multiple* runs of the FEFF program to generate the theory used in your fitting model

Multiple Data Sets Corefinement of *multiple* data sets – this may be data measured at multiple edges, multiple temperatures, etc.

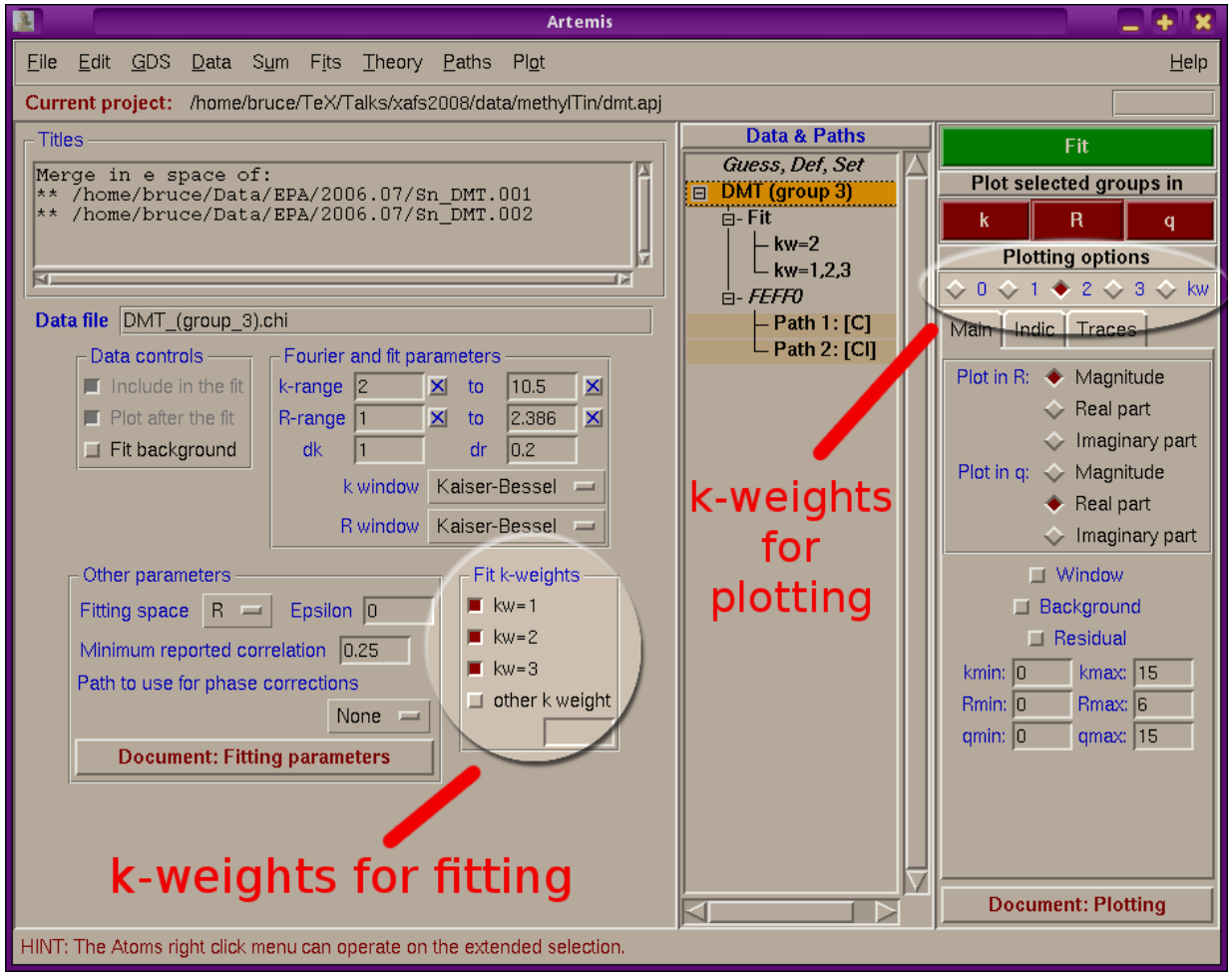
Constraints Between Parameters At the heart of an EXAFS fitting model are the relationships imposed between fitting parameters

Restraints on Parameters Application of imperfect knowledge to influence the evaluation of a fit

Using the "multiples"

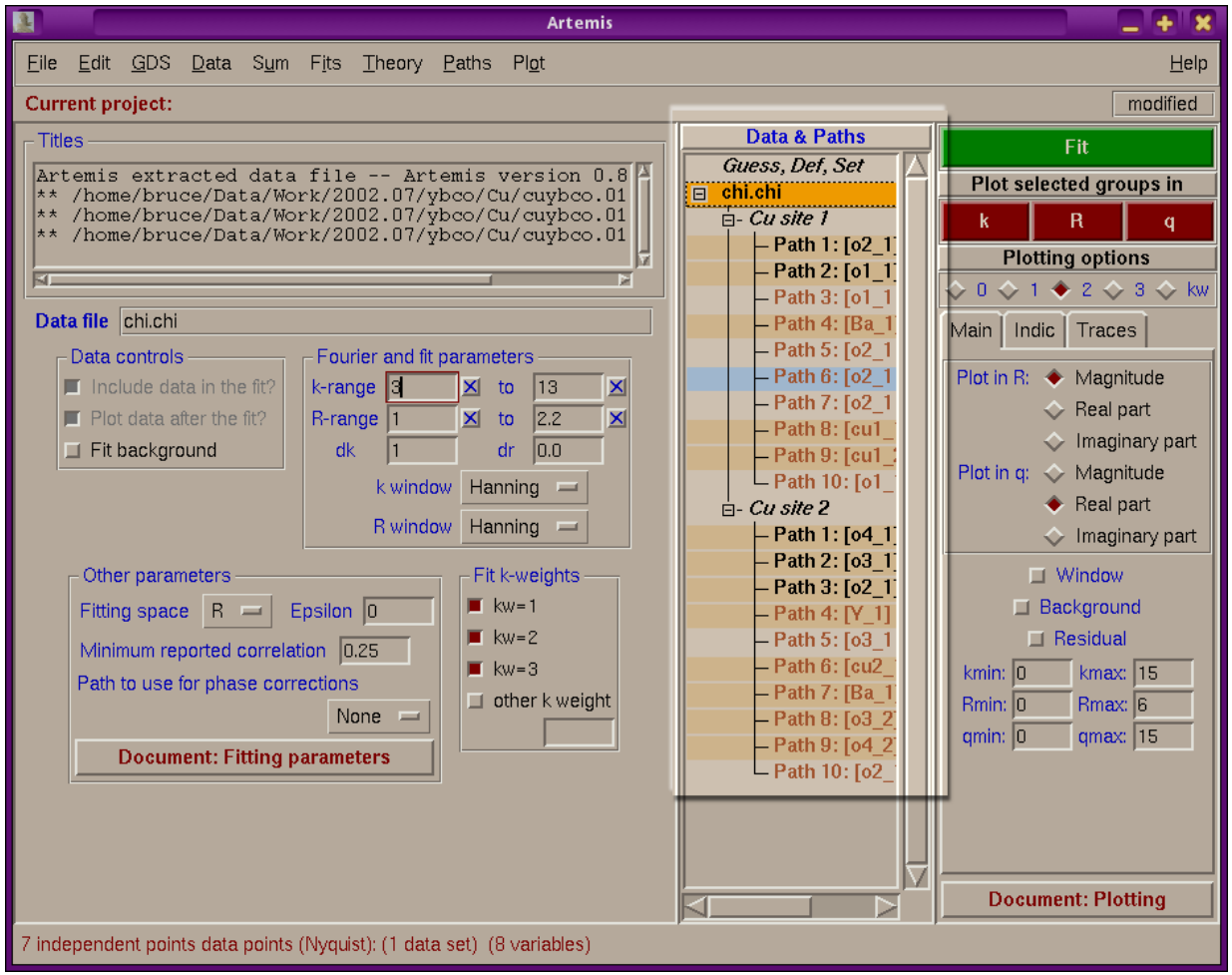
All of these are implemented in ARTEMIS.

Artemis: Multiple *k*-Weights



Simply click any/all of the fitting *k*-weight buttons.
Do not confuse them with the buttons that control k-weighting for fits.

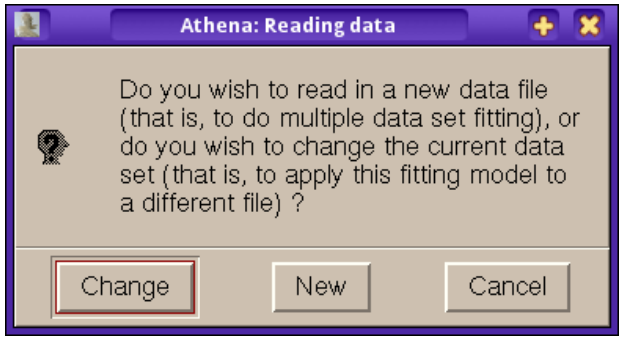
Artemis: Multiple Feff Calculations



Use as many FEFF calculations as you need, run FEFF, import and include whichever paths you need. (*Nota bene:* there is an out-of-the-box limit of 128 paths in IFEFFIT.)

Artemis: Multiple Data Sets (I)

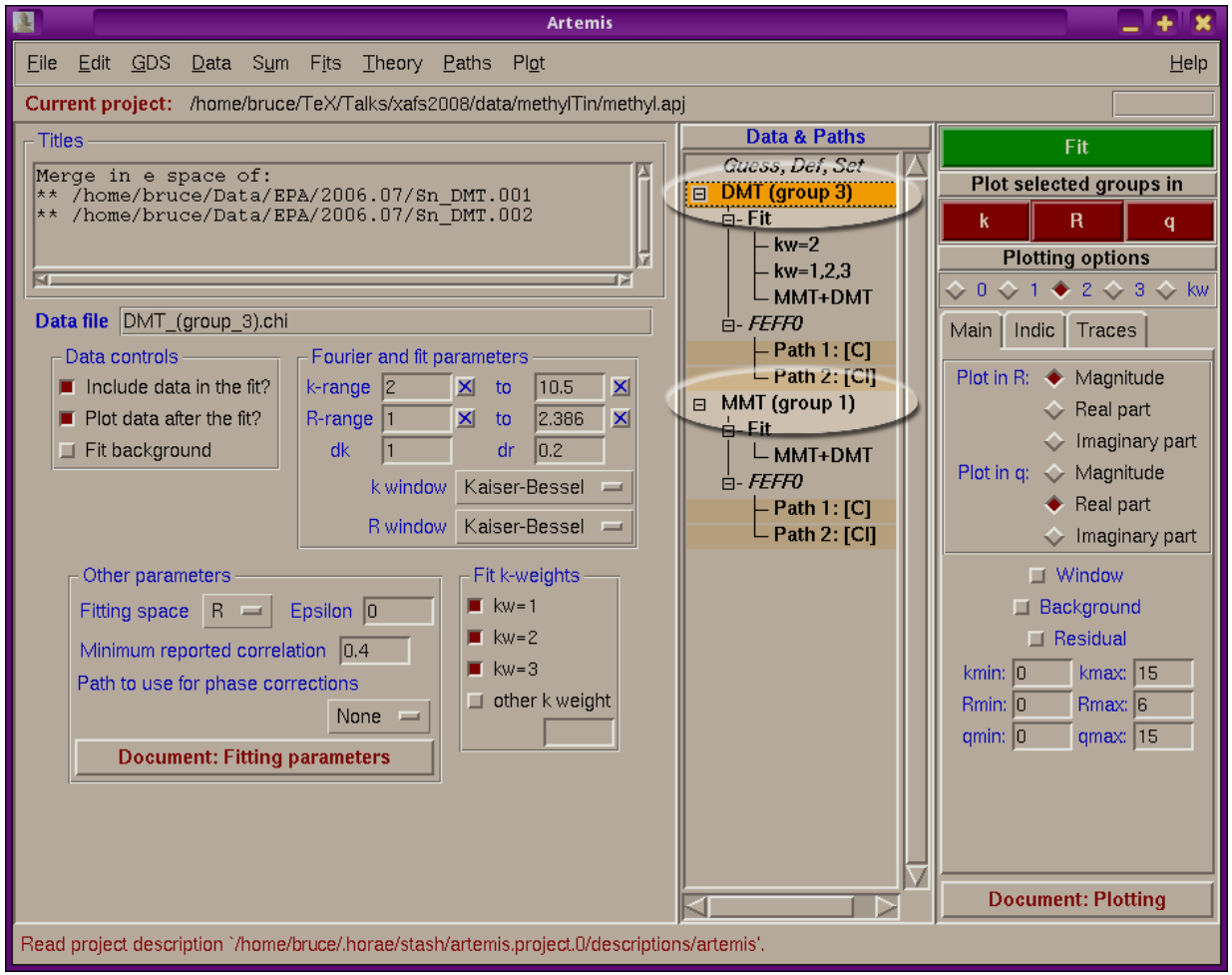
If you have already imported data into ARTEMIS, you will be presented with this dialog the next time you import data



Change With this option, you replace the current data set with a new one, allowing you to apply the current fitting model to a different data set.

New With this option, a new data set is added to the list of data and paths, allowing you to perform a *multiple* data set fit.

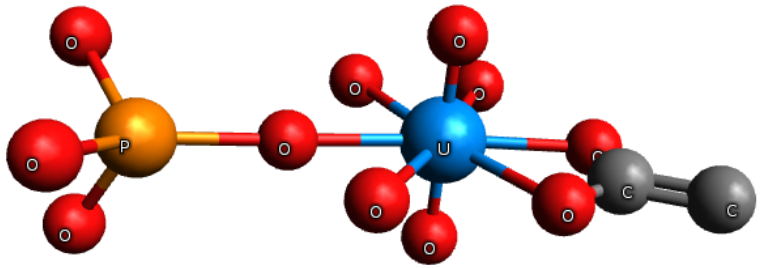
Artemis: Multiple Data Sets (II)



Use as many data sets calculations as you need. Each data set needs one or more paths associated with it. (*Nota bene:* there is an out-of-the-box limit of 10 data sets in IFEFFIT.)

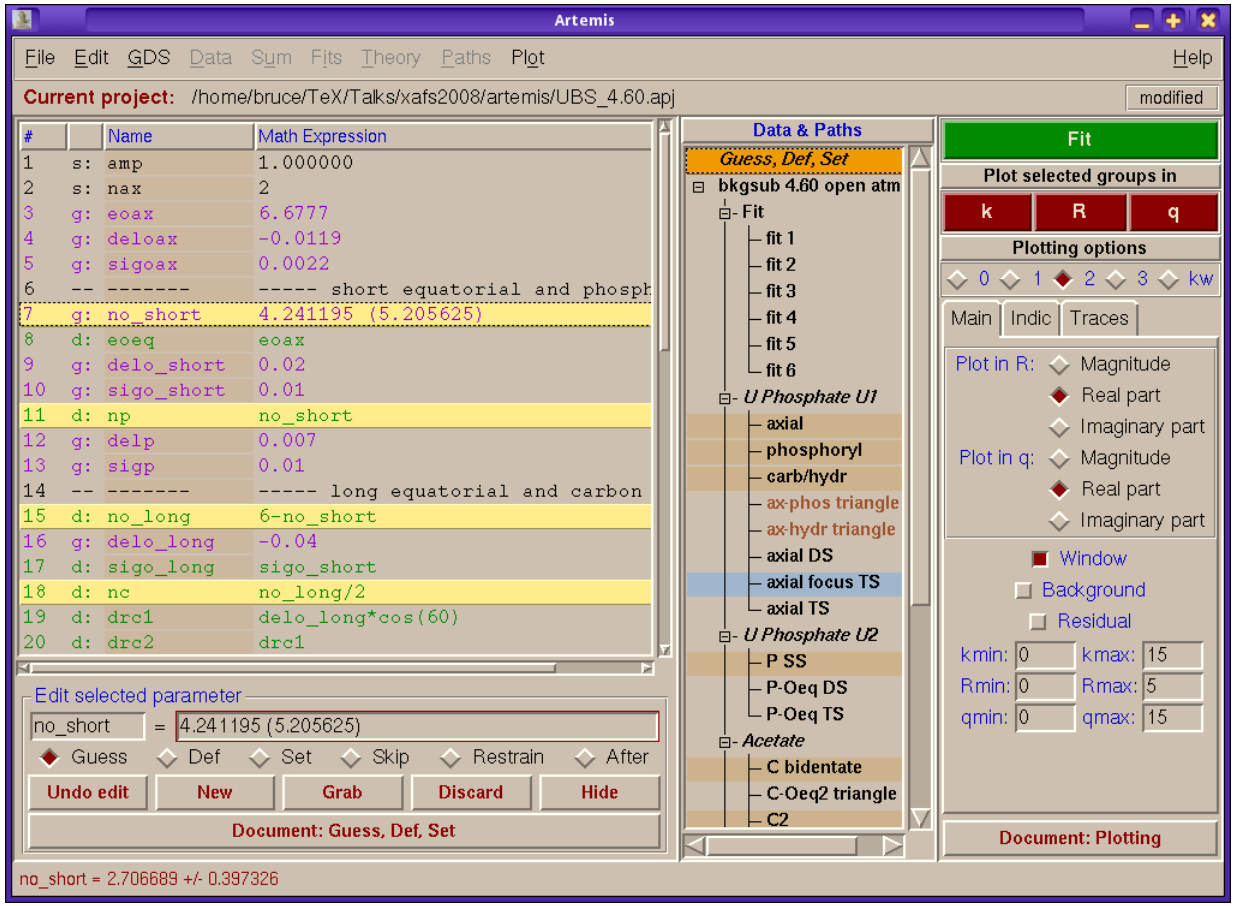
Artemis: Specifying Constraints (I)

In the uranyl example we saw earlier, the oxygen distances associated with the mono- and bidentate ligands are at different distances.



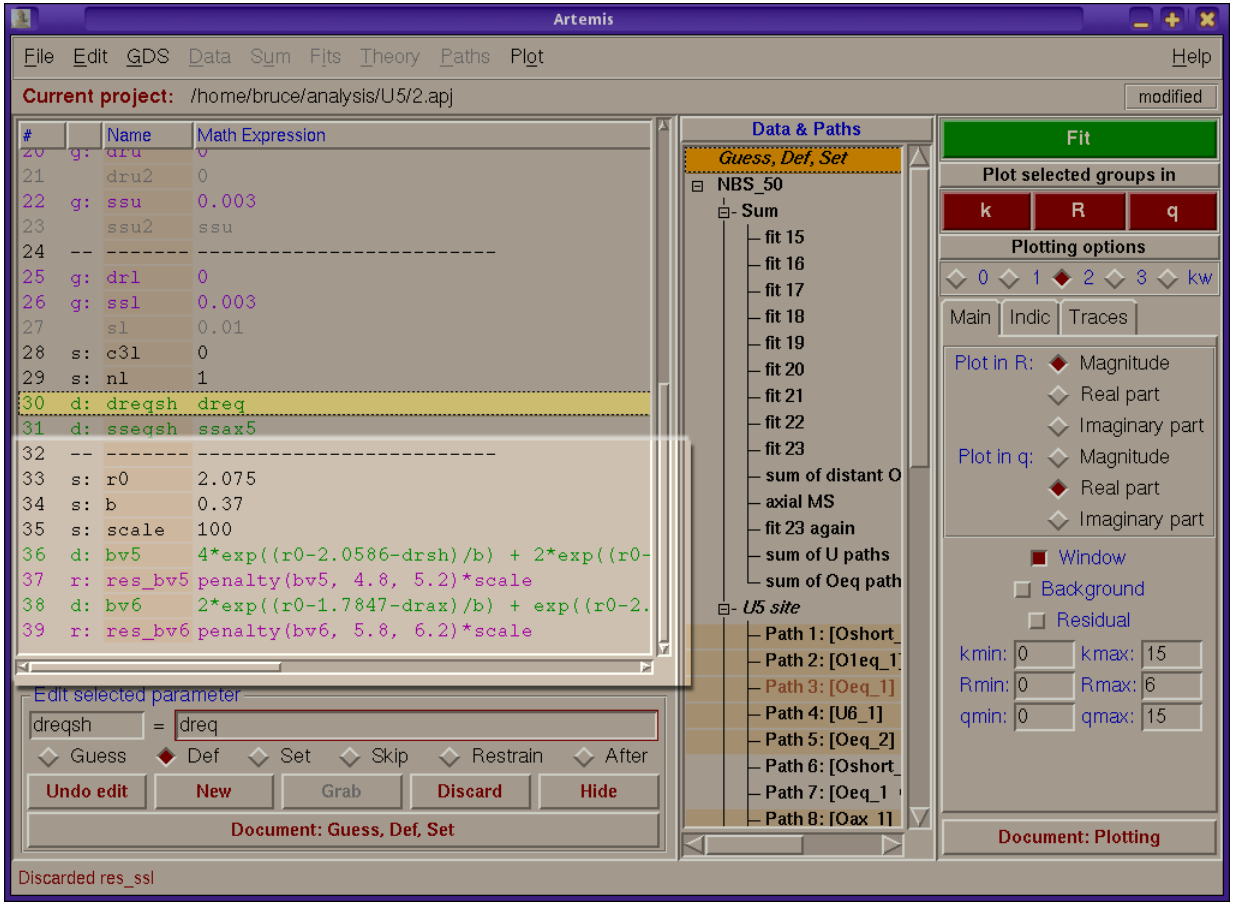
The number of P atoms is equal to the number of shorter equatorial oxygen atoms, while the number of C atoms is half the number of longer equatorial oxygens.

Artemis: Specifying Constraints (II)



Constraints are implemented as def parameter math expressions.
(Path parameters can also be expressed as math expressions.)
IFEFFIT's math expressions are quite expressive.

Artemis: Specifying Restraints



Restraints are managed on the Guess, Def, Set page, like any other parameter and will be properly used in the fit. A restraint depends upon a def or guess parameters – something that changes during the fit.

Now you know all the tricks!

Your assignment:

Use 'em all!

Do great data analysis!