2.2.3 *Evaluating the correctness of a model*

Two numbers are calculated by FEFFIT to evaluate the goodness of a fit. These are χ^2 statistic and the fraction of misfit, R. [99,108] χ^2 statistic is defined as

$$\chi^2 = \frac{N_I}{N} \sum_{i=1}^{N} \left(\frac{\left| \tilde{\chi}^{\rm th}(r_i) - \tilde{\chi}^{\rm exp}(r_i) \right|}{\epsilon(r_i)} \right)^2 \tag{2.17}$$

where the sum is over all "mere" pairs of points (both real and imaginary parts of the difference are evaluated at each "mere" point) in the region of r-space being fitted. The number of "mere" points in $[r_{\rm rmin}, r_{\rm rmax}]$ is arbitrary as it depends on the grid spacing in k-space and the size of the fast FT array. The prefactor N_I/N in eq. 2.17 is introduced to remove this arbitrariness as the number of independent points, N_I, (eq. 2.13) is independent of k-space grid or array size. [99, 102] The difference of model and data at each point r_i is weighted against the uncertainty *in the difference* (not just the data) at that point, $\epsilon(r_i)$. In FEFFIT a single value of ϵ is used as it is assumed that the uncertainty at point r_i is all due to random noise *in the data*, independent of r_i. The random noise is estimated by the rms value of $\tilde{\chi}^{\rm exp}(r)$ in the range $[15, 25]$ Å, where the XAFS oscillations are assumed to be indistinguishable from the noise (typical p.e. mean free path $\lambda_{\rm p.e.} \sim 10 - 15$ Å). Systematic errors in the data and the theory are not accounted for in χ^2 as those are extremely hard to evaluate. A related, useful quantity calculated in FEFFIT is *reduced* χ^2_ν, given by $\chi^2_\nu = \chi^2/\nu$ with ν being the degrees of freedom in the fit, i.e., $\nu = N_I - N_P$, where N_P is the number of parameters used in the fit.

If the errors are evaluated correctly (i.e., they are mostly random), a good fit should have $\chi^2_\nu \approx 1 \pm \sqrt{2/\nu}$. [95,108] Two fits with different N_P (and hence different ν's) can be compared and if their χ^2_ν differ by more than $2\sqrt{2(1/\nu_1 + 1/\nu_2)}$ (2 times the fluctuation in the difference) the fit with the lower χ^2_ν is significantly better.

Typical values of χ^2_ν found in this thesis, however, are on the order 10-50, indicating that (a) the model is bad, and/or (b) the value of ϵ has been underestimated by not considering systematic contributions to it. In order to check (a) as a possibility, FEFFIT calculates the fraction of misfit, R, given by

$$R = \frac{\sum_{i=1}^{N} \left| \tilde{\chi}^{\rm th}(r_i) - \tilde{\chi}^{\rm exp}(r_i) \right|^2}{\sum_{i=1}^{N} \left| \tilde{\chi}^{\rm exp}(r_i) \right|^2}. \tag{2.18}$$

This gives the fractional misfit, i.e., the ratio of total amplitude between theoretical and experimental curves to the total amplitude under the experimental curve. If R is only a few percent, as always found in this thesis for fits that "look" good, option (a) can be ruled out (*provided the structural parameters obtained are reasonable on physical grounds*) and the reason for $\chi^2_\nu \gg 1$ is the underestimation of ϵ.

If ϵ *were* known (and that would have meant $\chi^2_\nu \approx 1$ for a good fit), the uncertainties in the parameters (one standard deviation) are found by the change needed to increase χ^2 by 1, relative to its minimum value, χ^2_0 (or increase χ^2_ν by $1/\nu$). [99, 108] Since ϵ is underestimated and the fit is good (again, assuming a small R) we can redefine $\epsilon \to \epsilon\sqrt{\chi^2_\nu}$ so that the definition of a good fit having $\chi^2_\nu \approx 1$ is recovered. FEFFIT finds the uncertainties by the change in the parameters that will result in $\chi^2 = \chi^2_0 + 1$ (assumes a "good" fit) but then rescales them by multiplying by $\sqrt{\chi^2_\nu}$. This procedure would *overestimate* the value of these uncertainties as the fit is required to be good; i.e., $\chi^2_\nu \approx 1$, while in reality even if most of the "enhanced" χ^2_ν is due to a bad ϵ a fraction of it could be due to a bad fit, the later effect neglected in the rescaling of uncertainties.

Comparison between different fits by means of χ^2_ν has to be reinterpreted. Since a "good" fit will have $\chi^2_\nu \gg 1$, the standard fluctuation in χ^2_ν has to be renormalized to $\sqrt{2/\nu}\,\chi^2_\nu$. A fit is significantly better than other if the difference in their χ^2_ν is larger than about $2\sqrt{2((\chi^2_{\nu_1})^2/\nu_1 + (\chi^2_{\nu_2})^2/\nu_2)}$ (2 times the fluctuation of the difference), the fit with the lower χ^2_ν being better.

An estimation of the relative sizes of random and systematic errors in ϵ is possible. Initially only random contributions were considered, $\epsilon_{\rm ran}$, while the "correct" estimation of ϵ is given by

$$\epsilon^2 \approx \epsilon^2_{\rm ran} + \epsilon^2_{\rm sys} \approx \epsilon^2_{\rm ran}\chi^2_\nu \Longrightarrow \epsilon_{\rm sys} \approx \epsilon_{\rm ran}\sqrt{\chi^2_\nu - 1} \approx \epsilon_{\rm ran}\chi_\nu \tag{2.19}$$

since $\chi^2_\nu \gg 1$. This indicates that for the data presented in this thesis ($10 \leqslant \chi^2_\nu \leqslant 50$), systematic errors are $\sim 3 - 7$ times bigger than random errors. A pretty severe assumption in deriving eq. 2.19 is that systematic errors are also independent

of r and can be added in quadrature; i.e., they are also normally distributed. This is certainly not the case as systematic errors due to, e.g., a poor background removal will be more significant at low r and errors in the FEFF calculation could perfectly show r-dependence; i.e., could depend on the particular scattering path involved (as an example the use of a uniform interstitial charge density would result in underestimation of charge density along certain paths while it will overestimate it along others).