INTERNATIONAL TABLES FOR CRYSTALLOGRAPHY

ISSN 1574-8707
it.iucr.org

Volume I, X-ray Absorption Spectroscopy and Related Techniques
ISBN: 978-1-119-43394-1

Chapter 5.7

Keywords: X-ray absorption spectroscopy; X-ray fluorescence; χ^2_r; significance.

Goodness-of-fit measures in XAS, χ^2_r calibrations and limitations, and hypothesis testing

Christopher T. Chantler*

School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia. *Correspondence e-mail: chantler@unimelb.edu.au

This chapter is concerned with statistical analysis following the collection of data and the pooling of equivalent and inequivalent data sets, the use of statistical measures of inference and therefore the use of hypothesis testing and significance testing.

1. Introduction

This section develops and follows concepts in Bunker (2010), Newville (2024*a*), Bunker (2024) and Chantler (2024), and leads to and relates to the next few chapters (Booth, 2024*a*,*b*; Newville, 2024*b*). χ^2 is used as the key input to estimates of F in the statistical F-test for hypothesis testing; that is, in the assessment of preferred models of local structure theory that are applied to experimental data, which model, theory or nanostructure agrees well or significantly better than a different, poorer model or theory (Section 2)?

In evaluating parameter uncertainty, minimum radial separations, which can be measured, deduced or fitted, and necessary minimum uncertainties in radii or other parameters, some basic statistics can reveal the limitations of some current approaches and point towards the use of the experimental data to answer these questions (Section 3). This addresses which measures of goodness of fit χ^2_r can or cannot determine, and why.

Section 4 gives a brief evolution of the approaches and challenges in implementing statistical analysis for X-ray absorption fine-structure (XAFS) data and especially in the definition and meaning of the point uncertainty σ in the formulae for χ^2. It also discusses and encourages the use of χ^2_r and the potential for determining routinely small separations of inner-shell radii and paths.

Section 5 contrasts this basis in statistical analysis with the limitations of the conventional use of alternative measures of goodness of fit using Nyquist-like interpretations.

Section 6 discusses the limitations of current conventional definitions of the number of independent data points ($N_{\rm indp}$) in equations for the goodness-of-fit measure and recommends the usage of χ^2 for this purpose. Section 7 develops this by discussing some of the limitations of the current usage of $N_{\rm indp}$ in the formulae for goodness-of-fit measures. Section 8 relates back to the measures of goodness of fit and their meaning at the beginning of this chapter and explains how the other sections in the chapter permit a well-defined and uniform approach to statistical inferences of all sorts, including hypothesis testing and goodness of fit of unknown structures..

Related chapters

Volume I: 4.6, 4.7, 5.1, 5.2, 5.3, 5.6, 5.8, 5.9, 5.12, 5.13, 5.14, 5.15

2. Hypothesis testing: the *F*-test

Comparison of two different possible models in XAFS data analysis can be achieved using the statistical *F*-test to evaluate the statistical significance. The *F*-test checks the limitation of the models to discriminate between the two models and is not applicable to physically unrealistic solutions. A quantitative method of distinguishing better models avoids misinterpretations caused by eye observation methods. Effectively, this is a core of hypothesis testing, although the well-defined χ^2, $\chi^2_{\rm r}$, $\Delta\chi^2$ and $\Delta\chi^2_{\rm r}$ with the number of degrees of freedom provide the basic input for such comparisons. Hence, the *F*-test is a specific interpretation and use of χ^2 and $\chi^2_{\rm r}$ with a specific cutoff chosen according to a suitable hypothesized distribution function. Some questions that apply to the use of each of these interrelated measures are as follows.

(i) What is a good χ^2 to suggest a physical and reasonable model? (The answer in general is the number of degrees of freedom.)

(ii) What is a good $\chi^2_{\rm r}$ to suggest a physical and reasonable model? (The answer is 1.)

(iii) What is a good χ^2 difference to suggest significant improvement of a model (given a change in the number of degrees of freedom, *i.e.* a change in the number of fitting parameters)? (The answer, with several caveats, is $\Delta\chi^2 = 1$.)

(iv) What is a good $\chi^2_{\rm r}$ difference to suggest significant improvement of a model (given a change in the number of degrees of freedom, *i.e.* a change in the number of fitting parameters)? (A generally incorrect answer is $\Delta\chi^2_{\rm r} = 1$.)

For the latter two questions the *F*-test is particularly useful and insightful (Streltsov *et al.*, 2018).

In 1987, Joyner and coworkers demonstrated the *F*-test as a statistical test in EXAFS data analysis (Joyner *et al.*, 1987). This method has been used in EXAFS data analysis, claiming that the *F* value must follow the *F*-distribution law (Filipponi, 1995; Michalowicz *et al.*, 1999; Klementev, 2001). *F* can be given from a linear regression problem,

$$F_{\upsilon_1-\upsilon_2,\upsilon_2} = \frac{\chi_1^2 - \chi_2^2}{\chi_2^2}\frac{\upsilon_2}{\upsilon_1 - \upsilon_2}, \qquad (1)$$

where χ_1^2 and χ_2^2 are the goodness of fit for model 1 and model 2, respectively, $\upsilon = N_{\rm ind} - N_{\rm para}$ are the degrees of freedom of the fit, $N_{\rm ind}$ is the total number of data points and $N_{\rm para}$ is the number of fitted parameters. The number of fitted parameters should be increased in order to improve the fit ($N_{\rm para1} < N_{\rm para2}$, so that $\upsilon_1 > \upsilon_2$). In XAFS data analysis, the *F*-test should represent a reliable approximation when the *F* value exceeds α, a particular percentage point of the *F* distribution ($F_{\upsilon_1-\upsilon_2,\upsilon_2} \gg F_{\upsilon_1-\upsilon_2,\upsilon_2,\alpha}$). A benchmark of α, or the '*p*-value', is 0.05 for the *F*-test, explaining that the improvement in the fit is two standard deviations above the noise (for the assumption of normally distributed data).

There are numerous (related) statistical measures to assess whether a model is 'true, 'valid', 'plausible' or 'preferred'. In the case of XAFS, the model includes the theory, the proposed structure, the number and nature of the parameters, the values of the refined parameters and the experimental uncertainties. Conventionally for the *F*-test a '*p*-value' is assessed relative to the likelihood of a significant improvement. This has a conventional cutoff value for a perceived significance, but other cutoffs can be defended in particular circumstances. Both $\chi^2_{\rm r}$ and the *F*-test are derived from the definition and measurement of χ^2. Bayesian methods of several types are widespread (essential?) in XAFS analysis, and indeed χ^2 is a Bayesian method. In the simplest form, a Bayesian inference is made that some particular theory can model the experiment. A second Bayesian inference is made that the XAFS relates to a particular structure or molecule, perhaps approximately presented by a set of radii and bond angles. Within, for example, that space, χ^2 and the *F*-test are well defined. Multiple optional hypotheses or Bayesian inferences, if you will, will suggest nanostructure 1 versus nanostructure 2, or even theory 1 versus theory 2, with either the same number of (independent) fitting parameters or with different numbers. The *F*-test will then be able to distinguish between them and assess the significance of the improvement.

Two other methods that are often cited in statistical selection between different models include the Bayesian information criterion (BIC; Schwarz, 1978) and the Akaike information criterion (AIC; Akaike, 1974). $\mathrm{BIC} = k\ln(n) - 2\ln(\hat{L})$ and $\mathrm{AIC} = 2k - 2\ln(\hat{L})$, where $\hat{L}$ is the maximized value of the likelihood function of the model *M*, *i.e.* $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function, x is the set of observed data, n is the number of data points in x, the number of observations or equivalently the sample size, and k is the number of parameters estimated by the model. They can both be derived from Bayesian statistics but with different prior probability distributions. Both are useful and valuable, and there are current debates and discussions on these and other modified functionals. They are claimed to be good and useful heuristics (*i.e.* not definitive). AIC tends to overfit; that is, to prefer models with more (independent?) parameters. Neither is particularly reliable for sparse data sets or a large number of (independent?) model parameters. Recent reviews have claimed that if the goal is prediction, AIC and leave-one-out cross-validations are preferred. If the goal is selection, inference or interpretation, BIC or leave-many-out cross-validations are preferred. Both can produce conclusions far from the true cause of the data or the 'true model' (Burnham & Anderson, 2004; Vrieze, 2012; Ding *et al.*, 2018). In the context of the discussion here, we should consider the links between these and least-squares and maximum-likelihood approaches. For a Gaussian or normal distribution model, in terms of the residual sum of squares $\{\mathrm{RSS} = \sum_{i=1}^{n}[y_i - f(x_i;\hat{\theta})]^2$; within least-squares fitting the maximum-likelihood estimate for the variance of the residual distribution of a model is the reduced χ^2, $\chi^2_{\rm r} = \mathrm{RSS}/\nu$; $\nu = n - m$ is the degrees of freedom for n observations and m unknowns or independent parameters (k)$\}$ the BIC can be derived as $\mathrm{BIC} = n\ln(\mathrm{RSS}/n) + k\ln(n) = \chi^2 + k\ln(n)$ (Priestley, 1981; Kass & Raftery, 1995). Similarly, similar assumptions applied to AIC yield $\ln(\hat{L}) = -(n/2)\ln\chi^2 + C$, ergo $\mathrm{AIC} = 2k + n\ln(\chi^2) + C$, so that different models would differ by $\Delta\mathrm{AIC} = 2k + n\ln(\chi^2)$ as per conventional least squares

Table 1
Illustration demonstrating that poorly estimated experimental uncertainties can artificially influence the reported uncertainty of the model parameters, yielding misleading parameter estimates, for example $\sigma(R_j)$; yet $\sigma(R_j)(\chi^2_r)^{1/2}$ remains a robust measure of the parameter uncertainty. Here, we model a single sine-wave frequency with noise (Fig. 1).

$1\sigma(\chi)$ uncertainty	Estimated 1σ uncertainty	χ^2_r	$\sigma(R_j)$ (Å)	$\sigma(R_j)(\chi^2_r)^{1/2}$ (Å)
0.20	1.0	0.0513	0.0185	0.00419
0.20	0.50	0.141	0.00937	0.00352
0.20	0.10	4.25	0.00188	0.00388
0.20	0.05	17.7	0.000944	0.00397
0.20	0.025	69.0	0.000453	0.00376

Table 2
A simple illustration demonstrating that collecting more accurate experimental data allows parameters in the model to be determined to greater precision. We highlight the parameter R_j, as it applies to the discussion presented. Here, we model a single sine-wave frequency with noise (Fig. 1).

1σ uncertainty	χ^2_r	$\sigma(R_j)$ (Å)
1.0	1.09	0.0218
0.50	1.24	0.00971
0.33	1.04	0.00583
0.20	1.19	0.00368
0.10	0.705	0.00183
0.05	1.21	0.000923
0.0025	0.911	0.000459

(Burnham & Anderson, 2002); yet again the significance level is not defined. Perhaps more importantly, much of these derivations assumes constant uncertainty, which is usually not the case, and is not the case in XAFS. This discussion argues for using χ^2_r or an assessment of $\Delta\chi^2$ or indeed the F-test, but does not indicate or directly suggest the level for significance. This chapter focusses on the most widely used approaches in XAFS analysis, including the known limitations thereof. However, we point to other chapters for more information on alternative measures.

3. Important notions of analysis for XAFS; minimum atomic radial separations from data

There is a real and pervading question as to what information content can be gained from a spectrum, with or without uncertainties, and how this can define structure with a higher level of accuracy and insight. We illustrate this in the current section with pure sine waves and a finite k-range of fitting approximately matching that of our current data sets (Trevorah *et al.*, 2020). We omit mean free path and thermal broadening but include a defined uniform Gaussian noise $\sigma(\chi)$ in the spectrum, point-wise as each data point is defined to be an independent data point. We then 'estimate', as part of the preparation for fitting the spectrum, a given noise or uncertainty estimate and fit accordingly (Fig. 1). If we correctly 'estimate' the same scale of Gaussian noise as the data, it should be no surprise that a model fit yields a χ^2_r value close to unity. Conversely, also as expected, if our 'estimate' is five times too large, or ten times too small, then the estimated χ^2_r is increased or decreased by this error squared, and the uncertainty on a particular parameter, say $\sigma(R_j)$, is scaled inversely to this. This confirms that in the presence of a poor uncertainty estimate the robust estimate of a parameter is $\sigma(R_j)(\chi^2_r)^{1/2}$, as known from standard statistical analysis (Table 1).

The spectrum can be very noisy, yet it can determine the radius of a shell or atomic scatterer as 2.000 ± 0.002 Å (Fig. 1), with perhaps a 3% uncertainty on amplitude and a similar uncertainty on phase or phase offset while fitting over a k-range from 0 to 10 Å^{-1}. In this illustration, the parameters are made similar to real data presented later; one standard deviation uncertainties and noise are estimated as equivalent to $\sigma(\chi) \simeq 0.2$ and $\chi^2_r \simeq 1$.

Conversely, as the experimental data uncertainty improves (Table 2), as long as the prediction of uncertainty matches the data uncertainty, the uncertainty on a particular parameter, such as especially the fitted determined shell radius, is given to higher and higher accuracy from 2.00 ± 0.02 Å down to, for example, 2.0000 ± 0.0005 Å. In other words, the data are the key, and the uncertainty in the data points is the key. This determines the accuracy of structural parameters. Collecting data with smaller uncertainties allows the determination of (XAFS) model parameters with greater precision. This should be a fairly intuitive result. In other words, there is a correct estimate of uncertainty which allows statistical analysis to probe hypotheses of the model, theory, structure, bond distances and shells, and other physically meaningful parameters. The idea, which should be paramount, is to let the data dictate the limit of hypothesis testing and insight.

4. Notional minimum atomic radii separations from variants of the Nyquist theorem

Now let us consider the Nyquist-like prescriptions commonly discussed and used across the XAFS community. The key argument is that two components from two radial shells (with similar atomic number and scattering) will beat depending

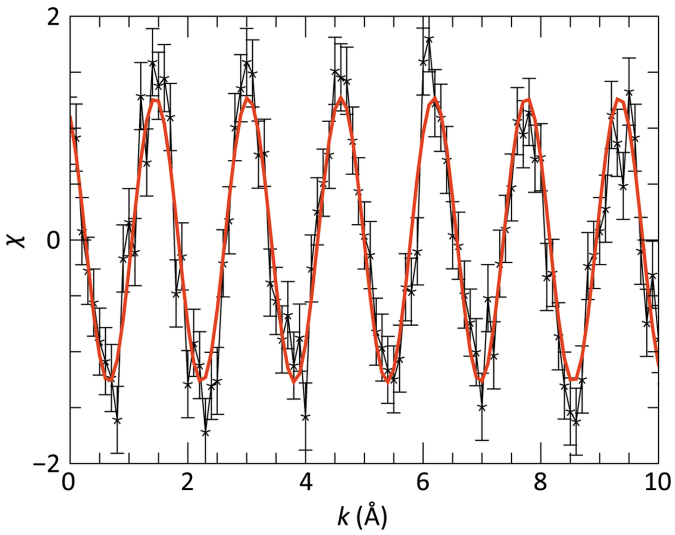

Figure 1
A simple pure sine-wave simulation including point-wise normally distributed noise (black line) and fit (red line). This is discussed in relation to normal statistical analysis and signal processing in Tables 1 and 2 and the text.

upon their phase, and their combined amplitude will decrease with k until a minimum occurs. If one has sufficient data over a sufficient k-range, then one can determine the radial separation, the phase kink or offset and the overall amplitude. From the requirement to reach or measure the minimum, we have the standard criterion (Lee *et al.*, 1981)

$$\Delta R = \frac{\pi}{2\Delta k}. \qquad (2)$$

Hence, if $k_{\max} = 15$ Å^{-1} or $\Delta k_{\rm range} = 15$ Å^{-1} then $\Delta R \simeq 0.1$ Å; similarly, if $\Delta k_{\rm range} = 7.5$ Å^{-1} then $\Delta R \simeq 0.2$ Å. This has regularly been presented as a fundamental limit of XAFS analysis or other Fourier-transform data collection. Depending upon the transform convention, one can report a minimum as a factor of 2π less than this. However, this value or estimate should be more like the minimal change that cannot be ignored, rather than the minimal change that can be detected. Lee *et al.* (1981) correctly state that with an arbitrarily good signal and noise, the resolution of different distances can be increased, although with correlation of parameters and noise this could be more limited. Thermal broadening, for example, damps the sine waves and increases with k, and so can be correlated with the beat from the two shells. This is included in (standard) correlated least-squares fitting analysis. In general, with good experimental data or well-defined uncertainties, these correlations can be overcome and separate closely spaced radial shell distances can be distinguished.

We illustrate this in Fig. 2. Here, we model two nearby frequency sine waves, closely matching our experimental data, with added Gaussian noise as in the earlier illustration. This mimics two bonding radii equivalent to the best fit of our experimental data model. With correctly defined uncertainties and noise, it is straightforward to determine separate radial distances below this 'Nyquist' or aliasing limit. It is absolutely not in conflict with signal processing; rather, it is the consequence of signal processing. In the figure, even a short fitting range of k can identify separate nearby shell radii to high accuracy, some 100 times more accurate than the separation, as long as the data quality is sufficient. If the uncertainties or noise are too large, or if the parameters are not independent but are highly correlated, then the limit is weaker and the resolution of, for example, two shells is weaker. An understanding of non-Gaussian distributions and cumulants can also confirm this finding.

5. A definition of σ in χ for hypothesis testing and significance

Much activity around the 1990s emphasized the need to fit spectra to allow structural insight, although the measures used varied quite widely and with non-uniform results. O'Day *et al.* (1994) introduced a goodness-of-fit measure but did not incorporate uncertainties or the standard deviation of the experimental data. They stated that 'there is currently no accepted method for determining these errors'. Similarly, Filipponi & Di Cicco (1995) commented that 'any XAFS report should be accompanied by a detailed analysis of the statistical errors due to random noise in the raw spectra'. However, 'general procedures to estimate errors ... are still not well established'.

There have been attempts to estimate uncertainty for XAFS data. Dent *et al.* (1992) used a piecewise polynomial to extract residual noise hopefully free of any structure, and equivalently used Fourier filtering to remove the dominant structure to hopefully yield a noise spectrum. These are recursive methods and depend upon an ideal fit of any structure using empirical means in order to derive the variance and noise that would allow the structure to be determined.

Filipponi (1995) commented that the uncertainties in the fitted XAFS parameters should be given by the spread of such parameters resulting from variance from an ensemble of experimental spectra. However, he comments 'unfortunately, only a single measurement is usually available'. He then provides three prescriptions for evaluating the noise distribution based upon an assumption of normal distribution of

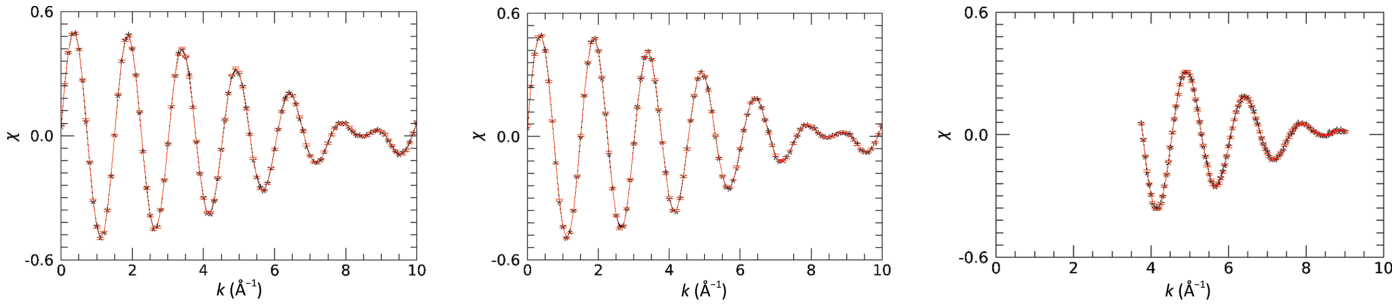

Figure 2
Fitting two nearby bond radii with noise to accuracies far below the 'Nyquist' interpretation so long as the data points have high and known accuracy, and uncertainties are maintained and propagated. Separate radii $R_1 = 2.161$ Å and $R_2 = 1.966$ Å are estimated correctly to within one standard error (s.e.) with an accuracy of 0.001 Å (*a*, *b*) or even 0.005 Å (*c*) for a short-ranged spectrum. The input noise and the corresponding estimate uncertainty are $\sigma(\chi) = 0.005$ and $\chi^2_{\rm r} \simeq 1$. The thermal broadening does not significantly impact the accuracy of the determination of parameters. Amplitudes are correctly fitted within one s.e. uncertainties of 0.4% (*a*, *b*) or 4% (*c*) and phases are correctly fitted to within one s.e. uncertainties of 0.012 radians (*a*, *b*) or 0.08 radians (*c*). Normal least-squares fitting is well able to correctly separate radial distances differing by, for example, 0.195 Å.

errors with assumptions of the magnitudes of these multivariate distributions.

He suggests that a Metropolis Monte Carlo algorithm may be used to sample the parameter probability distribution. When applied to experimental data this will result in a sequence of independent sets of parameter values, each of which produces best fits of the experimental spectrum. The spread then represents the statistical uncertainty. This is again a *post facto* representation and depends upon the initial determination of uncertainty. Finally, statistical errors can also be estimated from an assumption of perfect structural determination followed by a noise analysis of the residual, a little like that of Dent *et al.* (1992).

The *GNXAS* software (Filipponi, Di Cicco *et al.*, 2024; Filipponi, Natoli *et al.*, 2021) estimates the noise in energy space. After fitting the XAFS structure, an error bar for each data point is generated by first fitting a polynomial of degree $q < M$ over M data points, and the residual square difference divided by $M - q$ forms an estimate of the noise in the data. Repeating this along the spectrum allows an uncertainty to be estimated at each point via interpolation (Westre *et al.*, 1995; Filipponi & Di Cicco, 1995; Filipponi, 1995).

An alternative approach is employed by the *IFEFFIT* package (Newville, 2001*b*; Newville & Ravel, 2024), which estimates an uncertainty of experimental X-ray absorption spectroscopy (XAS) spectra as a function of wavenumber $\chi(k)$ based upon a Fourier transform of R-space background against theoretical models produced via the *FEFF*6 or *FEFF*8*L* package. *IFEFFIT* is also the foundation for other software used in XAFS analysis, which often provide the benefit of a graphical user interface (GUI), such as the *ARTEMIS* and *ATHENA* packages (Ravel & Newville, 2005, 2024). The measure of model agreement in *IFEFFIT*, ψ_r^2, is calculated as

$$\psi_r^2 = \frac{\psi_k^2}{N_{\rm indp} - N_{\rm var}}, \tag{3}$$

$$\psi_k^2 = \frac{1}{\varepsilon_k^2}\frac{N_{\rm indp}}{N_{\rm pts}}\sum_{i=1}^{N_{\rm pts}} k_i^w[\chi_{\rm data}(k_i) - \chi_{\rm th}(k_i)]^2 \tag{4}$$

or alternatively

$$\psi_R^2 = \frac{1}{\varepsilon_R^2}\frac{N_{\rm indp}}{N_{\rm pts}}\sum_{i=1}^{N_{\rm pts}}(\chi_{\rm data}[r_i) - \chi_{\rm th}(r_i)]^2, \tag{5}$$

where $N_{\rm indp}$ is an effective estimated 'number of independent points' in the XAFS spectra given by the Nyquist formula

$$N_{\rm indp} = \frac{2\Delta k \Delta R}{\pi} \tag{6}$$

for a fit range of Δk and ΔR in k-space and R-space, respectively (Stern, 1993).

ε_R estimates the uncertainty in the spectrum, which is calculated as the root mean square of the Fourier-transformed data in a region at high R. Parseval's theorem allows the conversion of this parameter into k-space, where w is the power of the k-weighted spectrum (Newville *et al.*, 1999),

$$\varepsilon_k = \varepsilon_R\left[\frac{\pi(2w+1)}{\delta k(k_{\rm max}^{2w+1} - k_{\rm min}^{2w+1})}\right]^{1/2} \tag{7}$$

for data-point k-spacing δk. However, since most sources of noise are not taken into account, ε_k and ε_R are underestimated, the error bars are too small and ψ_r^2 is overly large, often 500–2000, compared with a more ideal propagated χ_r^2.

In an attempt to remedy this, the fit is often re-evaluated using a somewhat arbitrary user-defined constant ε_k or ε_R to yield a 'good fit $\psi_r^2 \simeq 1$' (Calvin, 2013). This assumes that the final fit is perfect in order to define the uncertainties, and is therefore of limited use for hypothesis testing. The use of any such uniform error affects the fit since experimental uncertainties are non-uniform in $k^w\chi(k)$ or $\chi(r)$ space. Without measuring the uncertainties experimentally, this skews the fit towards data points that actually have a large error and away from those with a small measured uncertainty. For example, *IFEFFIT* and *EFEFFIT* often and usually fit in k-space, usually in $k^2\chi, k^3\chi$ or 'simultaneously all $k\chi, k^2\chi, k^3\chi$'. A lot of people publish using *IFEFFIT*. *IFEFFIT* interpolates the data (Chantler, 2024), while *EFEFFIT* interpolates the theory. *ATHENA*/*ARTEMIS* can fit on $k^2\chi$, $k^3\chi$ or 'simultaneously all $k\chi$, $k^2\chi$, $k^3\chi$', but very often users use the fits on transformed R-space. They can also filter and back-transform into 'Q' space. For example, *GNXAS* and related approaches very often transform to R, filter and back-transform to 'Q' before fitting.

Commenting that estimates of statistical precision are critical, Chantler *et al.* (1999) made a series of ten considerations of key limitations of accuracy in X-ray absorption measurements to be addressed. This was followed by a detailed statistical analysis of noise and variance in synchrotron X-ray measurements and in ion-chamber detection (Chantler *et al.*, 2000*a*,*b*). This explicitly measured numerous contributions to variance and precision. Previous authors had also investigated some of these details for absorption. This led to the X-ray extended-range technique (Chantler *et al.*, 2001).

6. The number of independent data points

Two measures of data quality or extent need to be clearly separated: the number of (independent) data points in an XAS measurement across the energy or k-range, N or $N_{\rm idp}$, and the 'effective number of independent parameters', $N_{\rm ipar}$, which can be fitted (well) in a least-squares analysis. In a step-scan experiment where each measurement is made independently of the next, each data point is independent and the total in, for example, a fitting k-range is N or $N_{\rm ipd}$. The number of parameters actually fitted (not constrained) in a model is then $N_{\rm par}$. There may be some particular systematic uncertainties in common, but the counting uncertainties and variances are independent. To get from raw or 'raw' pre-processed $[\mu/\rho]$ versus E data to a χ versus k spectrum involves a variety of possible operations which can change the real or apparent number of independent points. Trevorah *et al.* (2019) explain how to preserve the number of independent points and what

this means. However, it is common to interpolate data, apply several background and spline subtractions *etc.*, and each process can change the correlations of adjacent points but not the original noise and variance. It is possible in a fast continuous scan to have a detector response function (not just the dead time) that is too short for the experimental data to be truly independent. Hence, we strongly recommend considering the raw data statistic and variance and propagating this for the independent points, rather than interpolating onto a uniform grid where correlations will locally ensue (Schalken & Chantler, 2018).

One can ask which data points are relevant or useful to determine, for example, the first-shell radius and which contribute to a particular fitting parameter. For example, pre-edge data do not contribute to measuring the first-shell bond length and only the data transformed into χ versus k space contribute to the shell radii or any other standard XAFS parameter determination. Equally, very high $k > 25$ Å^{-1} data do not normally contribute significantly to any XAFS fitting parameter. However, the data points within a fitting window remain independent, unless for example they have been heavily interpolated. When interpolating either to a fine or finer χ versus k grid to, for example, 0.04 or 0.10 Å^{-1}, or when transforming and interpolating to, for example, a 0.04 Å spacing χ versus R grid, one is adding no new independent data points and no additional information content on any parameter; indeed, one is usually removing information content (Schalken & Chantler, 2018).

$N_{\rm indp}$ is an alternate measure of the 'effective estimated number of independent points' in the XAFS spectrum given by the Nyquist formula

$$N_{\rm indp} = \frac{2\Delta k \Delta R}{\pi} \tag{8}$$

for a fit range of Δk and ΔR in k-space and R-space, respectively (Lee *et al.*, 1981). In practice Δk is estimated or defined as the range of k being fitted (within some Hanning window, for example) or the range of k being used to create the transform of the experimental data into R-space. Similarly, ΔR is estimated as the Fourier-filtered or fitted range used in the transform into R-space. Alternatively,

$$N_{\rm indp} = \frac{2\Delta k \Delta R}{\pi} + I, \tag{9}$$

where I is claimed to be unity (Lin *et al.*, 1991) or 2 (Stern, 1993).

The idea as presented is that this is the maximum number of independent parameters that can be fitted for this data set. Naturally, if two parameters are not independent, but are 100% correlated, then one can only ever fit one or the other, and most XAFS parameters have significant correlation matrices with other parameters, so that this number can be seen as an overestimate. Krappe & Rossner (1999, 2000) saw the need to replace $N_{\rm indp}$ with a more useful and relevant measure to define the effective size of the parameter space. Krappe & Rossner (2002) and Rehr *et al.* (2005) concluded that the actual fitting space in such transformed fitting is commonly significantly less than even the lower estimate above, which in one example appeared to correspond to $I = -5$.

Many data are fitted in k-space, whether using χ, $k\chi$, $k^2\chi$, $k^3\chi$ *etc.* as the fitted function. The theory is determined and determinable over all R-space and can then be transformed into k-space on an arbitrary or regular grid. The theoretical determination can be limited to a number of paths and hence a range of radii, but from theory alone one can consider $\Delta R \simeq \infty$, which would imply from the formula that any number of parameters can be fitted up to the number of data points, or the number of independent data points if extensive interpolation, correlation or preprocessing is performed. The number of parameters which can be fitted, and their uncertainty, depends upon the uncertainty of the data, their spacing and their relevance to the parameters to be fitted.

In current usage, these estimates of the maximum number of independent parameters $N_{\rm indp}$ should be considered as empirical heuristics. The correct value, differing by a possibly large factor, should be found from freeing the most significant near-independent parameters one by one until the correlation matrix and array of uncertainties prove that the data set is not able to reliably determine the next (independent) parameter. In other words, the covariance matrix should explicitly indicate the ability to fit each parameter with whatever experimental and modelling correlations there may be, and it also gives strong indications when too many parameters are being fitted, so long as experimental uncertainties are used in the analysis. This is in sympathy with the least-squares covariance matrix, maximum-entropy correlation matrix and Bayesian approaches.

7. Definition of $\chi^2_{\rm r}$

Rather than being used as an heuristic guideline to the maximum limit of parameters, $N_{\rm indp}$ is also commonly used in the definition of $\chi^2_{\rm r}$ (Lee *et al.*, 1981; Stern, 1993; Newville, 2001*a*,*b*). $\chi^2_{\rm r}$ is the χ^2 per degree of freedom,

$$\chi^2_{\rm r} = \frac{\chi^2}{N_{\rm idp} - N_{\rm par}}. \tag{10}$$

In our current example, we have $N_{\rm idp} \simeq 122$ and $N_{\rm par} \simeq 4$. Conversely, an incorrect prescription which distorts hypothesis testing and relative model agreement is

$$\chi^2_{\rm r} = \frac{\chi^2}{N_{\rm indp} - N_{\rm par}}, \tag{11}$$

where $\Delta k \simeq 6.25$ Å and ΔR might be 2 Å, so $N_{\rm indp}$ might be estimated as 9 (or 10 or 11, or less following the above alternatives), so the denominator becomes $\sim$5 and $\chi^2_{\rm r}$ will be very high and also highly sensitive to adding parameters, leading some to recommend additional scaling, whereas the denominator should represent the data and the estimate of independent parameters should be given by the least-squares covariance matrix or equivalently the maximum-entropy

method. Some have commented that these sorts of (unknown) errors mean that χ^2_r is not useful in the normal sense (Ravel, 2016) or should be normalized, yielding only a relative local measure (Calvin, 2013) and almost invalidating it for hypothesis testing.

Stern (1993) stated the need for the denominator to be increased. In the example of lead metal fitting, Stern *et al.* (1991) needed to increase N_{indp} by 2 relative to earlier work to permit the fitting of additional parameters, apparently successfully. Their comment recognised that it was important and necessary that the denominator allow more parameters without reaching the singularity, but if the correct denominator had been used this would not have been necessary. Whilst Rehr *et al.* (2005) seem to correct this denominator error in their results section, this has not been applied by others and by major software packages. Meanwhile, Filipponi (1995) has pointed out the inadequacy of the use of N_{indp} as a measure or as a determinant of χ^2_r, noting that it is at odds with standard statistical analysis. Amongst other details, he notes the importance of the *F*-test.

8. Conclusion

At the heart of this chapter, but also any error-analysis, fitting and hypothesis testing, is the need to define and propagate the individual data uncertainties, so that appropriate measures of goodness of fit and hypothesis testing can be made (Schalken & Chantler, 2018). The denominator should always be the number of independently measured data points (for example ∼122 or ∼1000) minus the number of fitted parameters (for example ∼4). This chapter recommends the usage of χ^2 and χ^2_r including, where appropriate, estimation of the *F*-test for significance and hypothesis testing of one model or theory compared with another, or of one additional parameter, and whether this is a significant improvement (empirically, and preferably physically).

It is possible to make useful recommendations to (i) add fitting parameters that are as close to being independent as possible and avoid simultaneously fitting redundant or highly correlated parameters, (ii) constrain or define additional parameters *a priori* if possible, (iii) avoid interpolation and try to define and propagate the original raw data uncertainty and variance, (vi) fit additional parameters until the correlation matrix and the uncertainties demonstrate that the information content is not adequate for additional fitting parameters and (v) at all points use the denominator for χ^2_r as $N_{idp} - N_{par}$, not $N_{ipar} - N_{par}$, $N_{indp} - N_{par}$ or some other measure.

References

Akaike, H. (1974). *IEEE Trans. Autom. Contr.* **19**, 716–723.
Booth, C. (2024*a*). *Int. Tables Crystallogr. I*, ch. 5.8, 672–675.
Booth, C. (2024*b*). *Int. Tables Crystallogr. I*, ch. 5.9, 676–677.
Bunker, G. (2010). *Introduction to XAFS: A Practical Guide to X-ray Absorption Fine Structure Spectroscopy*. Cambridge University Press.
Bunker, G. (2024). *Int. Tables Crystallogr. I*, ch. 5.2, 636–638.
Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed., pp. 261–304. New York: Springer-Verlag.
Burnham, K. P. & Anderson, D. R. (2004). *Sociol. Methods Res.* **33**, 261–304.
Calvin, S. (2013). *XAFS for Everyone*. Boca Raton: CRC Press.
Chantler, C. T. (2024). *Int. Tables Crystallogr. I*, ch. 5.6, 659–663.
Chantler, C. T., Barnea, Z., Tran, C. Q., Tiller, J. B. & Paterson, D. (1999). *Opt. Quantum Electron.* **31**, 495–505.
Chantler, C. T., Tran, C. Q., Barnea, Z., Paterson, D., Cookson, D. J. & Balaic, D. X. (2001). *Phys. Rev. A*, **64**, 062506.
Chantler, C. T., Tran, C. Q., Paterson, D., Barnea, Z. & Cookson, D. J. (2000*a*). *X-ray Spectrom.* **29**, 449–458.
Chantler, C. T., Tran, C. Q., Paterson, D., Cookson, D. J. & Barnea, Z. (2000*b*). *X-ray Spectrom.* **29**, 459–466.
Dent, A. J., Stephenson, P. C. & Greaves, G. N. (1992). *Rev. Sci. Instrum.* **63**, 856–858.
Ding, J., Tarokh, V. & Yang, Y. (2018). *IEEE Signal Process. Mag.* **35**, 16–34.
Filipponi, A. (1995). *J. Phys. Condens. Matter*, **7**, 9343–9356.
Filipponi, A. & Di Cicco, A. (1995). *Phys. Rev. B*, **52**, 15135–15149.
Filipponi, A., Di Cicco, A. & Natoli, C. R. (2024). *Int. Tables Crystallogr. I*, ch. 6.12, 787–790.
Filipponi, A., Natoli, C. R. & Di Cicco, A. (2024). *Int. Tables Crystallogr. I*, ch. 6.11, 782–786.
Joyner, R., Martin, K. J. & Meehan, P. (1987). *J. Phys. C Solid State Phys.* **20**, 4005–4012.
Kass, R. E. & Raftery, A. E. (1995). *J. Am. Stat. Assoc.* **90**, 773–795.
Klementev, K. V. (2001). *J. Synchrotron Rad.* **8**, 270–272.
Krappe, H. J. & Rossner, H. (1999). *J. Synchrotron Rad.* **6**, 302–303.
Krappe, H. J. & Rossner, H. H. (2000). *Phys. Rev. B*, **61**, 6596–6610.
Krappe, H. J. & Rossner, H. H. (2002). *Phys. Rev. B*, **66**, 184303.
Lee, P. A., Citrin, P. H., Eisenberger, P. & Kincaid, B. M. (1981). *Rev. Mod. Phys.* **53**, 769–806.
Lin, S.-L., Stern, E. A., Kalb, A. J. & Zhang, Y. (1991). *Biochemistry*, **30**, 2323–2332.
Michalowicz, A., Provost, K., Laruelle, S., Mimouni, A. & Vlaic, G. (1999). *J. Synchrotron Rad.* **6**, 233–235.
Newville, M. (2001*a*). *J. Synchrotron Rad.* **8**, 96–100.
Newville, M. (2001*b*). *J. Synchrotron Rad.* **8**, 322–324.
Newville, M. (2024*a*). *Int. Tables Crystallogr. I*, ch. 5.1, 631–635.
Newville, M. (2024*b*). *Int. Tables Crystallogr. I*, ch. 5.13, 690–694.
Newville, M., Boyanov, B. I. & Sayers, D. E. (1999). *J. Synchrotron Rad.* **6**, 264–265.
Newville, M. & Ravel, B. (2024). *Int. Tables Crystallogr. I*, ch. 6.13, 791–795.
O'Day, P. A., Rehr, J. J., Zabinsky, S. I. & Brown, G. E. J. (1994). *J. Am. Chem. Soc.* **116**, 2938–2949.
Priestley, M. B. (1981). *Spectral Analysis and Time Series*, p. 375. San Diego: Academic Press.
Ravel, B. (2016). *Artemis Manual.* https://bruceravel.github.io/demeter/documents/Artemis/forward.html.
Ravel, B. & Newville, M. (2005). *Phys. Scr.* **2005**, 1007.
Ravel, B. & Newville, M. (2024). *Int. Tables Crystallogr. I*, ch. 6.1, 723–727.
Rehr, J. J., Kozdon, J., Kas, J., Krappe, H. J. & Rossner, H. H. (2005). *J. Synchrotron Rad.* **12**, 70–74.
Schalken, M. J. & Chantler, C. T. (2018). *J. Synchrotron Rad.* **25**, 920–934.
Schwarz, G. E. (1978). *Ann. Statist.* **6**, 461–464.
Stern, E. A. (1993). *Phys. Rev. B*, **48**, 9825–9827.
Stern, E. A., Līvņš, P. & Zhang, Z. (1991). *Phys. Rev. B*, **43**, 8850–8860.
Streltsov, V. A., Ekanayake, R. S., Drew, S. C., Chantler, C. T. & Best, S. P. (2018). *Inorg. Chem.* **57**, 11422–11435.

Trevorah, R. M., Chantler, C. T. & Schalken, M. J. (2019). *IUCrJ*, **6**, 586–602.
Trevorah, R. M., Chantler, C. T. & Schalken, M. J. (2020). *J. Phys. Chem. A*, **124**, 1634–1647.
Vrieze, S. I. (2012). *Psychol. Methods*, **17**, 228–243.
Westre, T. E., Di Cicco, A., Filipponi, A., Natoli, C. R., Hedman, B., Solomon, E. I. & Hodgson, K. O. (1995). *J. Am. Chem. Soc.* **117**, 1566–1583.