Error Reporting Recommendations: A Report of the Standards and Criteria Committee

Adopted by the IXS Standards and Criteria Committee July 26, 2000

1. Introduction

The development of the field of x-ray absorption fine structure over the last 30 years has resulted in a dramatic improvement in the quality of experimental XAFS and data analysis. This more rigorous data analysis also requires more quantitative estimates of the uncertainties in the results of the analysis. The Standards and Criteria committee has developed a series of recommendations to assist the community to determine reliable estimates of uncertainties in experimentally derived results. It also encourages the inclusion in any publication of an explanation or citation of how these uncertainties were obtained. The committee hopes that the adoption of a common terminology and methods of data analysis and error reporting will enhance the ability of x-ray spectroscopists to compare results from different groups, beamlines, and software packages.

The Standards and Criteria Committee of the IXS has been developing these recommendations for many years. In 1988 and 1990, independent workshops were held dealing with this topic and related matters. The reports of these workshops suggested certain procedures to follow for theory, data acquisition, data analysis, and reporting requirements. Subsequently the IXS was formed, and created the Standards and Criteria Committee which focussed on development of recommendations regarding error estimation and reporting. In 1998 and 1999 meetings of the S&C Committee considered drafts of this report. In the 2000 S&C meeting this report was adopted by the S&C committee and submitted to the IXS Executive Committee for consideration of adoption as a policy of the IXS. The report makes several specific recommendations to standardize the reporting of errors. Supplementary references may be found on the IXS home page (`http://ixs.csrri.iit.edu/`) under Education and Dissemination.

The intent of this report is to present to the IXS and its members the general principles for systematizing and standardizing the reporting of the uncertainties associated with experimental results derived from XAFS measurements. The guiding principle is general enough to cover the procedures used by most groups, even if they use methods for assessing errors that differ from those recommended here. The specific recommendations are intended for the majority of groups who use standard codes and who wish to use standardized procedures. The recommendations made here represent the initial attempt by the IXS to standardize error reporting. Further refinements are expected, particularly with regard to the estimation of systematic errors, and so refinements and improvements of the procedures recommended here are to be expected in the future. This is also a somewhat technical document since it is being used to justify the basis for these recommendations. The S&C committee intends to develop a separate document demonstrating how to use these criteria in the context of a particular example of data analysis.

2. General Concepts

The general principle that should apply to all quantitative results derived from XAS measurements represents the starting point for this report and is as follows:

> **General Principle: Reports of all quantitative results that are derived from XAS measurements must be accompanied by an estimate of the uncertainty and a description or a citation that explains the basis for the uncertainty.**

For a few groups that have developed their own methods, particularly if they use analysis methods that are different from the general fitting techniques assumed here, this means that they must explain clearly what they have done or refer readers to the published literature explaining their methods. Most groups are not involved in developing new analytical methods and rely on standard codes to provide their data analysis. It is the intention of the specific recommendations that will be made below to provide these groups with an understanding of and justification for these specific procedures and, also, to encourage the developers of analysis codes to modify their error assessment procedures to be consistent with these recommendations.

The starting point for any error assessment is a model for evaluating the confidence limits. Several closely related formulations have been used by most analytical codes. It is proposed here that confidence limits can be evaluated from a functional of the generic form

$$(\Delta\chi)^2 = W \sum_{i=1}^{N} |Data_i - Model_i|^2/\epsilon_i^2. \tag{1}$$

Here N is the number of data points in the fitting range, W is a dimensionless factor described below, and ϵ_i is the measurement uncertainty for the i-th data point. This equation applies to both non-k-weighted data and k-weighted data, provided the data, model, and errors are weighted in the same manner. The functional (1) is analogous, but is not identical, to the standard statistical χ^2 function. The following essential points should be borne in mind:

1. The points $Data_i$ and $Model_i$ may be represented in E, k, or R-space. In each case the measurement uncertainty ϵ_i should be calculated and normalized accordingly, as discussed below.

2. For R-space fits $W = N_{idp}/N$, where N_{idp} is the number of statistically independent data points, and N is the number of complex data points contained within the range of the fit.

3. For back transformed k- and R-space fits, $W = N_{idp}/N$, and N_{idp} is the number of statistically independent data points, which is approximately given by $N_{idp} = 2\Delta k\Delta R/\pi$, rounded off to the nearest integer. [L. Brillouin, "Science and Information Theory", Academic Press, New York, 1962; E. O. Brigham, "The Fast Fourier Transform", Prentice-Hall, Englewood Cliffs, NJ, 1974]. N_{idp} is referred to by Brillouin as the degrees of freedom in the signal. Here Δk and ΔR are the ranges

in k- and R-space respectively over which there is useful data. If the transform or fit range extend beyond where there is meaningful data above the noise, then the limit for the interval that should be used in calculating the number of independent points is the limit of data. For example, if one has a Fourier transform of the data from an amorphous material that contains one shell whose amplitude (signal) is negligible relative to the noise above 2.5 Å, the number of independent points cannot be increased by extending an R-space filter up to a higher R. Typically, for analysis from Fourier transforms for a single shell, the k-space range is about 10 Å^{-1} and the R-space interval is about 2 Å giving a value for N_{idp} of 13. This is generally viewed as a conservative number and should be used for guidance only in estimating the maximum number of fitting parameters that can be used (see below).

4. For E-space (raw data) fits, W may be taken as 1.0. In this case, intrinsic limitations on the number of adjustable parameters become apparent through analysis of the covariance matrix. In *ab initio* fitting it should be stressed that these intrinsic limits are the same as in k- and r-space fitting, whether Fourier filtered or not. The energy range for the fit determines a Δk, and the number of shells included in the fit determines a ΔR, which together limit the number of parameters that can be determined.

5. The r.m.s. measurement error ϵ may be used in Eq. (1) instead of the individual ϵ_i:

$$\epsilon^2 = \sum_i \epsilon_i^2 / N. \tag{2}$$

6. With $(\Delta\chi)^2$ defined as in Eq. (1), a fit can be considered acceptable when $(\Delta\chi)^2 \sim \nu$, where $\nu = N_{idp} - P$ is the number of degrees of freedom in the fit. P is the number of parameters used in the fit. If the final results are dependent upon several stages of fitting then P should be the maximum number of parameters used in the analysis process. It should be noted stressed that a discretely stepped variable is still counted as a variable.

Confidence limits for the fit parameters can be estimated from Eq. (1) with one of two methods:

a. If the covariance matrix [**C**] for the fit is available, the uncertainty δP_j in the j-th independent parameter P_j may be calculated as

$$\delta P_j = \sqrt{C_{jj}}\,, \tag{3}$$

where C_{jj} is the j-th diagonal element of the covariance matrix.

b. When the covariance matrix is not available, the confidence limit δP_j should be estimated by varying the parameter P_j away from its optimal value while optimizing all other parameters until $(\Delta\chi)^2$ increases by 1.0 above its minimum value. *The importance of optimizing all other parameters while varying P_j cannot be overemphasized.*

The methods (a) and (b) are completely equivalent. In order to obtain consistent results it is essential that the normalization of the $(\Delta\chi)^2$ be precisely as in Eq. (1). The use of any other definition for $(\Delta\chi)^2$ will yield confidence limits that are not directly compatible with the recommendations of this report.

Additional insight on the reliability of the fit can be gained by inspecting the correlation matrix for the fit. The correlation coefficient between the i-th and j-th independent parameters may be calculated as

$$r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}, \tag{4}$$

where the C's are the respective elements of the fit covariance matrix. The correlation coefficient takes values in the range [-1,1], and describes the interdependent effect of the i-th and j-th parameters on $(\Delta\chi)^2$. A positive correlation coefficient indicates that the increase in $(\Delta\chi)^2$ caused by increasing one parameter from its best-fit value can be compensated to some degree by increasing the other parameter. A negative correlation coefficient indicates that the increase in $(\Delta\chi)^2$ caused by increasing one parameter from its best-fit value can be compensated to some degree by decreasing the other parameter. The degree of compensation is roughly equal to the magnitude of r_{ij}. As a general rule-of-thumb, $|r| = 0.3$ and $|r| = 0.7$ are indicative of weakly and strongly correlated parameters, respectively.

Recommendation 1: Use Eq. (1) as a standard definition for $(\Delta\chi)^2$, and Eq. (3) as a standard method for estimating confidence limits for the fit parameters.

When the errors ϵ_i are normally distributed the confidence limits calculated with either definition will correspond to a significance level of 68.3 percent. The use of other statistical confidence levels, e.g., 95 percent, 99.9 percent, or "joint confidence intervals", is also acceptable, in which case Eq. (3) should be modified accordingly. This can be done by multiplying the confidence limit δP_j with an appropriate scaling coefficient, as tabulated for instance in "Numerical Recipes: The Art of Scientific Computing", William Press, et al.; Cambridge University Press, 1986. When the errors are not normally distributed, the precise confidence level corresponding to Eq. (3) should be determined with explicit Monte Carlo simulations if the probability distribution of the error ϵ_i is known.

It is important that criterion (6) noted above, $(\Delta\chi)^2 \sim \nu$, be satisfied. Significant deviation in either direction, i.e., $(\Delta\chi)^2 \ll \nu$ or $(\Delta\chi)^2 \gg \nu$, should be considered as an indication that the estimate for the error ϵ in Eq. (1) is inadequate. Suggestions for dealing with such situations are provided in Section 4.

3. Statistical Errors

The error in any measurement has a statistical and a systematic component. The first step in estimating the total error should always be to estimate the statistical component.

Statistical errors result from photon counting statistics, fluctuations in the position and intensity of the photon beam, mechanical instabilities in the beamline, electronic noise, and other factors. The distinguishing feature of statistical errors is that they vary randomly, both in sign and magnitude. As a result the statistical error averages out to zero, and can be made arbitrarily small in any particular measurement by acquiring more data and/or the use of longer integration times. The estimation of the statistical error is fairly straightforward, and several methods can been used.

1. Subtracting a smoothed function χ' from the background-subtracted experimental χ data. The statistical component of the error may then be calculated for each point as
$$\epsilon_{statistical} = \chi_i - \chi'_i, \tag{5}$$
where the index i in $\epsilon_{statistical}$ has been dropped for clarity. The average statistical error should be estimated from the r.m.s. value of (5) over data segments with similar statistical weight, e.g., over segments with a constant integration time. The smoothed data χ' may be obtained either by smoothing with a low-order polynomial, or with low-pass Fourier filtering.

2. From the r.m.s. amplitude of the R-space transform *in a region devoid of structural features.* A commonly used range is between 15 and 25 Å. If the statistical noise is truly white, the amplitude of its spectrum in R-space can be adequately approximated by a single number, ϵ_R, which is related to the r.m.s. noise amplitude in k-space, ϵ_k, by Parseval's theorem:
$$\epsilon_k = \epsilon_R \sqrt{\frac{\pi(2w+1)}{\delta k(k_{max}^{2w+1} - k_{min}^{2w+1})}}. \tag{6}$$
Here ϵ_R is the r.m.s. noise amplitude in the k-weighted R-space spectrum, ϵ_k is the r.m.s. noise amplitude in the unweighted k-space spectrum, w is the k-weight of the transform, the transform range is $[k_{min}, k_{max}]$, and δk is the spacing of the points in k-space. The above formula assumes that an FFT with equidistant k-space points is used, and the forward and back transforms are normalized by $\sqrt{\delta k/\pi}$ and $\sqrt{\delta r/\pi}$, respectively, which is a common XAFS convention. A similar relationship between the r.m.s. amplitudes of the k-space and r-space noise exists for non-equidistant discrete Fourier transforms, but cannot be expressed in compact closed form. It should be noted that Eq. (6) estimates the r.m.s. noise amplitude over the entire k-range used in the FFT, and it is impossible to account for effects such as the possibly different statistical quality of data segments, e.g., due to differing integration times. It is also not possible to estimate the error point-by-point, as in Eq. (5).

3. On the basis of Poisson statistics from the raw absorption data. For example, the statistical error for transmission data may be calculated on a point-by-point basis from
$$\epsilon^2_{statistical} = 1/N_0 + 1/N, \tag{7}$$
where N_0 and N are the *actual* number of photon counts (not the integrated "counts" from voltage to frequency converters; for more information on determining ion chamber currents see tutorial documents at IXS home page (`http://ixs.csrri.iit.edu/`) that are detected for each data point by the I_0

and I chambers, respectively. Once again the point index i has been dropped from Eq. (7) for clarity. The corresponding formula for fluorescence yield data is

$$\epsilon^2_{statistical} = (N_f/N_0)^2(1/N_0 + 1/N_f). \tag{8}$$

The average statistical error should be estimated from the r.m.s. value of Eq. (7) or Eq. (8) over data segments with similar statistical weight, e.g., over segments with a constant integration time.

4. By processing a number of independent scans (or independent partial sums of scans) in parallel and calculating the spread in the final results. It is convenient to also include the total sum in this process and to estimate the statistical error as

$$\epsilon_i^2 = \frac{1}{M(M-1)} \sum_{j=1}^{M} (\chi_i^{(j)} - <\chi_i>)^2, \tag{9}$$

where M is the number of independent scans (or partial sums) $\chi^{(j)}$ being processed, and $<\chi>$ is the total sum. The normalization pre-factor in Eq. (9) must be $1/M(M-1)$ to ensure that the results of this estimate be representative of the uncertainty in $<\chi>$ (as opposed to the spread within the set of measurements $\{\chi^{(j)}\}$). Only then will the estimate from Eq. (9) be directly comparable to those obtained with Eqs. (5-8).

When post-background subtraction methods are used (Eqs. 5-6, 9), it should be verified that the interpolation often used to place the data on a constant-k grid does not significantly alter the noise level. In addition, some background-removal programs perform smoothing of the data, which could drastically alter the noise levels. The presence of smoothing (either intentional, or as an artifact of the numerical procedures used) can be tested with synthetic data to which a known amount of noise has been added prior to background subtraction.

The number of scans needed for a reliable estimate of the upper limit of $\epsilon_{statistical}$ is fairly low, and may be estimated with standard statistical methods. The following relationships hold for a set of measurements $\{y_1,..,y_M\}$ drawn from a normal distribution with a mean μ and variance σ^2:

$$< y >= \mu \pm \ t_{\alpha/2}\epsilon, \quad B_{1-\alpha/2} \leq s/\sigma \leq B_{\alpha/2}. \tag{10}$$

Here M is the number of measurements, $<y> = \sum_i y_i/M$ is the estimated mean, $s^2 = \sum_i (y_i - <y>)^2/(M-1)$ is the estimated variance, $\epsilon = s/\sqrt{M}$ is the standard error of the set of M measurements, $1-\alpha$ is the desired confidence level, and $t_{\alpha/2}$ and the B's are constants tabulated in many statistical textbooks and in the appendix.

For example, for a set of 4 measurements ($M = 4$) it is possible to state with 95 percent confidence ($\alpha = 0.05$) that

$$\mu =< y > \pm \ 3.18(s/2), \quad 0.27\sigma \leq s \leq 1.76\sigma, \tag{11}$$

i.e., even after four measurements it is possible to place an upper limit of $3.7s$ on the true value of σ with 95 percent confidence. When using Table I (Appendix) in conjunction with Eq. (10) it is important to note that the statistical error $\epsilon_{statistical}$ estimated with Eqs. (5-9) from the average of M scans $<\chi>$ will be approximately equal to $s/\sqrt{M}$, not s. Inspection of Table I in the appendix also indicates that in this particular example there is very little to be gained from performing more than 15 scans *as far as the precision of the estimate for* σ *is concerned.*

Once an estimate for $\epsilon_{statistical}$ is available it can be substituted into Eq. (1), and best-fit parameter values and their confidence limits can be estimated with the procedures outlined in Section 2 with $\epsilon = \epsilon_{statistical}$. If the condition $(\Delta\chi)^2 \sim \nu$ is satisfied it may be assumed that the error ϵ is dominated by the statistical component and the estimated confidence limits adequately represent the uncertainty in the best-fit parameter values. It should be noted that this will generally not be the case for the majority of XAFS data acquired and analyzed today, particularly for high quality transmission data from concentrated standards where the statistical error is small.

The case $(\Delta\chi)^2 \ll \nu$ corresponds to a situation where the calculated $\epsilon_{statistical}$ overestimates of the total error ϵ. Since by definition $\epsilon_{statistical} = \epsilon$, this condition is most likely the result of errors in the implementation of Eqs. (5-9) used by the experimenter. Suggestions for dealing with the case $(\Delta\chi)^2 \gg \nu$ are provided in next section.

4. Systematic Errors

The most common indication that the data analysis is affected by systematic errors is that $(\Delta\chi)^2 \gg \nu$ when $\epsilon = \epsilon_{statistical}$ is used in Eq. (1). This will most likely be the case for the majority of XAFS data acquired and analyzed today. Systematic errors, which are introduced both during acquisition and analysis of EXAFS data, arise from a large number of sources that are discussed in more detail in the 2000 Report of the Standards and Criteria Committee in the section of the Error Analysis Group. Some of the more common sources of acquisition-related systematic errors include sample inhomogeneities, radiation damage, thickness and particle size effects, insufficient suppression of higher harmonics in the monochromatized photon beam, detector nonlinearity, glitches (both monochromator and sample-related), and improper sample alignment. Analysis-related errors include: systematic modifications of the amplitude of the EXAFS oscillations caused by improper pre-edge background subtraction and/or normalization to "unit step height"; imperfect references (both experimental and *ab initio*); improper determination of S_0^2 and/or improper energy-dependent normalization when *ab initio* references are used; and technical errors during pre-processing of the data. While some types of systematic error may be eliminated through good data acquisition and analysis practices (e.g., harmonics, alignment, sample preparation), others are often unavoidable (e.g., imperfect standards, certain types of glitches, inadequate energy-dependent normalization). Further work is needed to estimate the magnitude and distribution of the unavoidable systematic errors, e.g., through round-robin type measurements of well-characterized samples on various beamlines around the world.

A clear distinction needs to be made between identifiable and well-characterized sources of systematic error, such as thickness effects, self-absorption effects, energy-dependent normalization, and inadequate structural models, and poorly understood systematic errors, such as those listed in the previous paragraph. The former sources of error are calculable, must be corrected for, and should not be included in the estimate for ϵ.

At the present time the magnitude of the systematic error in a "typical" XAFS experiment is not known. Planned future activities of the IXS Standards and Criteria committee include round-robin type measurements at various XAFS beamlines around the world and modeling of various analytical procedures. The goal of these activities will be to determine the magnitude and distribution of the major systematic errors. The following procedure is proposed as an interim solution:

1. Confidence limits should be initially estimated with $\epsilon = \epsilon_{statistical}$ as described in Section 2.

2. If, as is often the case, it is found that $(\Delta\chi)^2 \gg \nu$ after step 1, then confidence limits should then be rescaled by a constant factor f, which the experimenter believes is a fair representation of the systematic error in their experiment or data analysis.

3. The values of $(\Delta\chi)^2$, N_{idp}, ν, $\epsilon_{statistical}$ and the scaling factor f should be disclosed in all published results.

Recommendation 2: Systematic errors are important in much of the XAFS data acquired and analyzed today. While standardized methods for estimating the magnitude of systematic errors are not presently available, their effect should not be ignored. One possible method for treating systematic errors is given above. Whatever the method chosen by the experimenter, enough details should be disclosed in all published work to allow an independent evaluation of the reliability of the procedures used to estimate the systematic error.

While not rigorously correct, the solution outlined above is consistent with accepted practices in other fields, e.g., x-ray diffraction, and allows experimenters to provide *some* estimate of the uncertainty in the fit results. The procedure is equivalent to assuming that the systematic error, constant throughout the fit range, is added in quadrature to the statistical error, and scales roughly as $(f^2-1)\epsilon^2_{statistical}$, where f is the scaling factor used in step 3 above.

5. Additional information

Determining the fit quality when the contribution of systematic effects to the total error is significant, e.g., when $(\Delta\chi)^2 \gg \nu$, is not easy. For example, it is not clear how to distinguish fits that are truly bad (in the sense of inadequate models) from those simply dominated by systematic errors. These two situations may be differentiated to some extent by examining an $\mathcal{R}$-factor, defined as

$$\mathcal{R}^2 = 100 \times \frac{\sum_{i=1}^{N} |Data_i - Model_i|^2}{\sum_{i=1}^{N} |Data_i|^2}\ \%. \tag{12}$$

As long as the signal-to-noise ratio (S/N) of the data is good, the $\mathcal{R}$-factor of adequate fits can be expected to be not more than a few percent, which is thought to be the typical accuracy of modern *ab initio* EXAFS codes. It is therefore desirable that analysis codes provide users with both $\mathcal{R}$ (Eq. 13) and S/N.

The signal-to-noise ratio should be estimated by dividing the fit range into regions of similar statistical weight, e.g., regions acquired with a certain integration time, estimating $\epsilon_{statistical}$ for each region with Eqs. (5-9), and calculating a "local" S/N value for each region:

$$(S/N)_m \approx \frac{1}{M}\sqrt{\sum_{i=1}^{M} |Data_i|^2/\epsilon^2_{statistical}}\,. \tag{13}$$

Here M is the number of data points in the corresponding m-th region, and the sum is over data points in that region only. The overall S/N ratio quoted for the entire range should be $\min\{(S/N)_m\}$, the smallest value obtained from Eq. (13).

6. Reporting Requirements

Recommendation 3: Certain information should be provided to allow an independent estimate of the reliability of the confidence limits quoted in published work. The level of disclosure should be appropriate to the importance of the XAFS results to the case being presented. The reporting requirements listed below should be included as appropriate when XAFS is one of the primary experimental methods used to support the conclusions of the work being proposed for publication.

Consistent with the general principle proposed in this report, it is essential that certain information be disclosed in order to adequately describe the basis for reporting of results and estimating errors from XAFS data. Information that should be included, as appropriate, may include:

1. The type of functional being minimized. The definition in Eq. (1) has been proposed in order to develop a more standardized approach for the XAFS community.

2. The function used to model the data.

3. The types of standards for scattering amplitudes and phases, e.g., empirical or *ab initio*. The procedure used to calibrate or check the standards and to determine S_0^2 should be described.

4. The fitting space. This may be E-space (raw absorption data $\mu(E)$), k-space (raw EXAFS data $\chi(k)$ without Fourier filtering), R-space (Fourier-transformed EXAFS data), or back-transformed k-space (Fourier-filtered k-space data, sometimes referred to as q-space).

5. K-weighting of the data, if applicable.

6. Fits in R-space and k-space should provide both the range of the fit and the range of the transform.

7. Number of parameters in the fit, and any externally imposed, should be given. The parameter count should include all parameters varied at any stage of the refinement process, not just those varied during the last stage. An estimate for the maximum number of parameters that can be extracted from the analysis should be included and justified.

8. The best-fit value of $(\Delta\chi)^2$. An $\mathcal{R}$-factor should be identified and its value provided.

9. The procedure used to estimate the confidence limits should be described. The confidence limits represent an estimate of the standard error.

10. Parameter correlations for multi-shell fits, particularly those outside the usual (N,σ^2) and (R,E_0) groups, should be described if they have significantly affected the specific results being presented.

Appendix — Statistical Tables

M	$B_{0.025}$	$B_{0.975}$	$t_{0.025}$	M	$B_{0.025}$	$B_{0.975}$	$t_{0.025}$	M	$B_{0.025}$	$B_{0.975}$	$t_{0.025}$
2	2.23	0.03	12.71	10	1.45	0.54	2.26	18	1.33	0.67	2.10
3	1.92	0.16	4.30	11	1.45	0.57	2.23	19	1.32	0.68	2.09
4	1.76	0.27	3.18	12	1.41	0.59	2.18	20	1.32	0.68	2.09
5	1.67	0.35	2.78	13	1.39	0.61	2.16	21	1.31	0.69	2.08
6	1.61	0.41	2.57	14	1.38	0.62	2.14	22	1.30	0.70	2.07
7	1.55	0.45	2.45	15	1.38	0.64	2.13	23	1.30	0.71	2.07
8	1.52	0.49	2.37	16	1.36	0.65	2.12	24	1.29	0.71	2.06
9	1.48	0.52	2.31	17	1.34	0.66	2.11	25	1.29	0.72	2.06