Chapter 4.6

Data acquisition and determination of precision and uncertainty

Christopher T. Chantler*

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

This chapter will discuss standard and advanced data-collection techniques in X-ray absorption spectroscopy and the determination of precision, accuracy and uncertainty using such approaches. In general, any independent measures of precision and accuracy should be expected from data collection and pre-analysis, and any conclusions drawn from the data should be based upon them. Hence, fitting, modelling and local or nanostructural determination should be based upon independent uncertainties, however developed and even if some measures or systematics are unavailable. The key difference in application between processing without uncertainty or precision and processing using independent estimates of precision or uncertainty is between a fingerprinting approach or confirmation of the possibility of a proposed structure (without uncertainty) and a quantitative hypothesis testing to lead to structural information and determination. Since this chapter is intended to be short, it will point towards discussions in other parts of this volume and to some key literature discussions.

1. Introduction

We take as given standard error propagation and statistical analysis, as discussed for example in undergraduate textbooks on statistical analysis (Bevington, 1969), *Numerical Recipes* (Press *et al.*, 2007) and standard least-squares fitting codes and theory (Ito, 1993; Newville, 2024). We should add to this an understanding of Bayesian statistics or *a priori* probabilities (Gregory, 2005; James, 2006; Krappe *et al.*, 2024), reverse Monte Carlo techniques (Timoshenko & Kuzmin, 2024), wavelet-transform discussions, discussions of normal and non-normal and asymmetric distribution functions, and the consequences of these for confidence levels, the percentile certainty of the mean or standard deviation. The next chapter (Chantler, 2024*a*) will discuss absolute measurements and their methodologies, consequences, limitations and applications. This chapter is focused on relevance to X-ray absorption spectroscopy (XAS). Mathematical resources and texts, physics and chemistry advanced treatises are directly relevant and germane, whilst first-year undergraduate laboratory manuals in physics, chemistry mathematics or biosciences are often grossly inadequate.

2. Reproducibility and uncertainty

The simplest measure of data reproducibility is to repeat the measurement under equivalent conditions and to assess the pointwise variance and hence the reproducibility and precision of the measurement. The issue of noise and signal and statistics is raised in Abe *et al.* (2018). This is commonly used in many XAS and X-ray absorption fine-structure (XAFS) measurements. Often due to time, sample or experimental

Related chapters

constraints, a datum of the spectrum may only be measured once, in which case the value, mean, standard deviation and standard error cannot be determined. Two repeats define the mean, but with little insight. Usually, three or more repeated measurements are a minimum to begin to estimate a variance, standard deviation, standard error and precision.

A good recommendation is to collect ten repeated measurements under 'identical' conditions, at which point the variance, reproducibility and precision can be determined and anomalies can be investigated. In general XAS, this can be accomplished in one of two ways: either (i) using 'repeated rapid scans' with continuous or stepwise scanning, assuming that the underlying energies of each step are consistent and that any lag or settling on the monochromator are at least consistent and preferably very small, or (ii) using 'step scans', where repeated measurements are made at each (energy) step. As long as the monochromator energy and crystals settle quickly, then these measurements will all be at the same energy and hence the variance of that point will relate to the same energy (Streltsov *et al.*, 2018; Trevorah *et al.*, 2020; Sier *et al.*, 2020). Option (i) includes 'slew' scanning, QEXAFS scanning modalities and energy-dispersive XAS measurements. Energy-dependent systematics will be distinct in these various approaches.

Any variance defined by the repetitions can include deliberate or accidental variation of experimental parameters. For example, the location on the sample, the angle to the detector, the slit size for the beam, a set of apertures or attenuating filters and any variation of the energy of the beam on the sample during the measurements can all vary with energy or time and yield a systematic shift of the data values. Fluctuation or variation of bandwidth and polarization of the beam or any harmonic content in the monochromated synchrotron beam will yield a variation in signal. If these are not characterized, then the variation during the repeated measurements will yield a larger variance and a weaker precision. If these are characterized (Best & Chantler, 2024; Chantler, 2024a,b) then the variance after correction should be smaller and the standard deviation and standard error, and hence the precision, should be improved. In general, this will yield more precise and more insightful data sets. Sayers (2000) recommends, as a general principle, that

> 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 that uncertainty.

For X-ray transmission experiments, the fundamental datum is the set of $\{E, I_{\text{upstream}}, I_{\text{downstream}}\}_i$, with subscript $i = 1, N$ for N repeated measurements per energy E.

This generates the individual variances $\sigma^2(I_{\text{upstream}})$, $\sigma^2(I_{\text{downstream}})$. The square root of this, σ, is then the standard deviation $\sigma_{\text{s.d.}}$, an estimate of the measurement variability. The standard error $\sigma_{\text{s.e.}}$ is then an estimate of the precision and the determination of the consistency of the data set to a single value is an estimate of the uncertainty of the mean (Chantler, Tran, Paterson, Barnea *et al.*, 2000; Chantler, Tran, Paterson, Cookson *et al.*, 2000).

The standard mean of a set of measurements $\{x_i\}$ is

$$\langle x \rangle = \mu(x) = \sum_{i=1}^{N} \frac{x_i}{N}; \quad \sigma^2(x) = \sum_{i=1}^{N} \frac{[x_i - \mu(x)]^2}{N}.$$

For a finite number of observations these sample variances may yield an estimate of the population variances as

$$s^2(x) = \sum_{i=1}^{N} \frac{[x_i - \mu(x)]^2}{N-1} = \sigma^2(x) \frac{N}{N-1},$$

reflecting the degrees of freedom. At this level, for any variate including the upstream or downstream counts, the variances are unweighted. A general formula for the propagation of errors (*i.e.* uncertainties) for a bivariate function $x = f(u, v)$ involves the covariance $\sigma_{uv}^2 = \langle [u - \mu(u)][v - \mu(v)] \rangle$ from which, if the derivatives are well behaved,

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\delta x}{\delta u}\right)^2 + \sigma_v^2 \left(\frac{\delta x}{\delta v}\right)^2 + 2\sigma_{uv}^2 \left(\frac{\delta x}{\delta u}\right)\left(\frac{\delta x}{\delta v}\right).$$

If the variates u and v are uncorrelated then $\sigma_{uv}^2 = 0$.

In (X-ray) transmission, the simple key processed datum for further analysis is often considered to be $x(E) = (I_{\text{downstream}}/I_{\text{upstream}})(E)$, for example for monitor and detector ion chambers. In this case, if uncorrelated, then following the general formula percentage or relative errors add in quadrature,

$$\frac{\sigma_x^2}{x^2} = \frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} - 2\frac{\sigma_{uv}^2}{uv}.$$

Since the mass attenuation coefficient $[\mu/\rho]$, the linear attenuation coefficient μ per unit density ρ, is related to the local sample thickness t and $[\mu/\rho][\rho t]$ is given by the variate $y = -\ln(I_{\text{downstream}}/I_{\text{upstream}}) = -\ln(x)$, then the uncertainty in y corresponds to the relative uncertainty in x: $\sigma_y = \sigma_c/x$. For such bivariates of u and v, the (linear) correlation coefficient is defined as $R_{uv} = \sigma_{uv}^2/\sigma_u \sigma_v$.

In practice, the upstream and downstream detector signals may be uncorrelated. However, they are often highly correlated in a positive or a negative manner, with a correlation coefficient that is either unity, +1 (for example dominated by the X-ray beam flux), or −1 (for example dominated by absorption in the upstream detector), and indeed is affected by other contributions to the correlation of signals between the detectors (Chantler, Tran, Paterson, Barnea *et al.*, 2000; Chantler, Tran, Paterson, Cookson *et al.*, 2000). It may be obvious that the preferred data collection will have a very high, positive correlation between the detector signals, and that in this case the best approximation to the variance (square of the precision) of the ratio is given by

$$\sigma_{x=I_{\text{downstream}}/I_{\text{upstream}}}^2 = \sum_{i=1}^{N} \frac{[x_i - \mu(x)]^2}{N-1}.$$

That is, the variance of the ratio is the preferred, most precise and simplest processing route (Chantler, Tran, Paterson, Barnea *et al.*, 2000; Chantler, Tran, Paterson, Cookson *et al.*, 2000). Conversely, if the linear correlation coefficient is

618   Christopher T. Chantler · Data acquisition

Int. Tables Crystallogr. I (2024). ch. 4.6, 617–623

negative, then the (quite poor but preferred, and most precise) simple estimate for the variance of the ratio is given by

$$\frac{\sigma^2_{x=u/v=I_{\text{downstream}}/I_{\text{upstream}}}}{x^2} = \frac{\sigma^2_u}{u^2} + \frac{\sigma^2_v}{v^2}.$$

Normally, the covariance and linear correlation coefficient can be well determined from ten repeated measurements (Chantler, 2009). The use of the ratio with matched linear detectors can also address uncompensated Bragg glitches (Abe *et al.*, 2018).

A strong recommendation is that where the correlation between monitor and detector signals is not strongly positive, the data set should often be discarded. The experiment should be reoptimized and the data should be remeasured. However, the discussion above explains how to estimate precision in these cases.

The precision, and standard deviation, is the estimate of the consistency of the repeated measurements, of the second moment of the distribution function and hence of the square root of the variance. The related standard error is the estimate of the agreement with the population mean based upon the variance, $\sigma_{\text{s.e.}} = \sigma_{\text{s.d.}}/N^{1/2}$.

Once the variance, standard deviation and standard error have been obtained for one data subset, they can be combined with different sample observations to propagate the uncertainty or precision to a derived data variate (Fig. 1). Since the components already have defined uncertainties at that time,
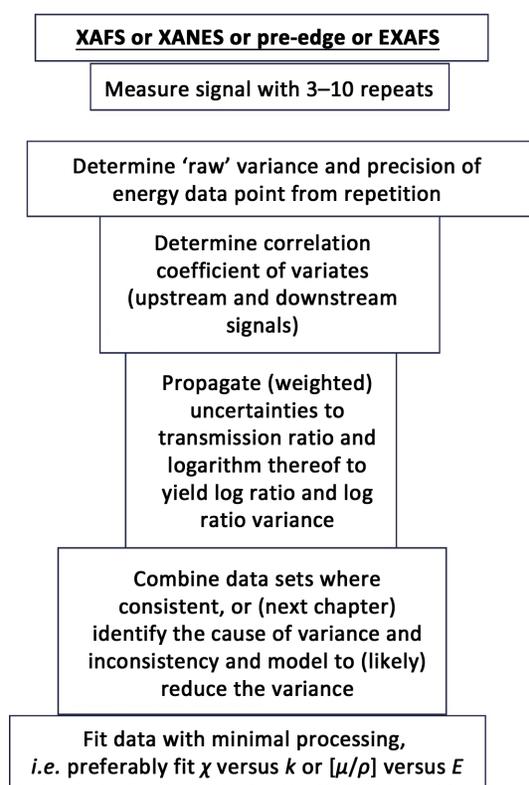

Figure 1
Principles of data collection and pre-processing for XAS and determination of precision.

further propagation should use weighted uncertainties, whether relating to the consistency and the precision or relating to the accuracy and overall uncertainty. If the individual measurements are consistent within the stated individual uncertainties, then the weighted mean $\mu_{x,w}$ and the corresponding weighted uncertainty $\sigma_{x,w}$ are given by

$$\langle x \rangle = \mu_{x,w} = \frac{\sum\limits_{i=1}^{N} \dfrac{x_i}{\sigma_i^2}}{\sum\limits_{i=1}^{N} \dfrac{1}{\sigma_i^2}}$$

and

$$\sigma^2_{x,w} = \frac{1}{\sum\limits_{i=1}^{N} \dfrac{1}{\sigma_i^2}};$$

conversely, if the individual measurements are not consistent within the stated individual uncertainties, then $\sigma_{x,w}$ is estimated by

$$\sigma^2_{x,w} = \frac{\sum\limits_{i=1}^{N} \dfrac{(x_i - \mu_{x,w})^2}{\sigma_i^2}}{\sum\limits_{i=1}^{N} \dfrac{1}{\sigma_i^2}}$$

(Tran *et al.*, 2005; Chantler, 2009). Clearly, in the limit of uniform uncertainties, these estimates converge to the unweighted estimate. The formula for consistent independent measurement usually predicts a very small final uncertainty (precision), whereas the inconsistent formula usually implies some systematic uncertainty, which is uncharacterized or uncorrected for, and typically provides a larger estimate. Sometimes these can be used as lower and upper estimates of data-point uncertainty. Under anomalous but well defined conditions, the former estimate can reach zero and hence be unreliable even as an underestimate (Chantler, 2009).

Combining data sets can be tested using Student's or sample *t*-tests (Snedecor & Cochran, 1989), which reveal whether there is a known or unknown possible systematic error between the subsets, which should perhaps be investigated to avoid excessive variance. The *F*-test, comparing hypothesis and model testing, will be discussed more explicitly in Chantler (2024*c*).

Combining data subsets is important to define a pooled variance and precision, or to investigate a known or unknown systematic error, and to determine a higher level of accuracy. When two subsets of data {*A*}, {*B*} (perhaps with a different process variable such as aperture, filter, time, temperature) are to be pooled and are assumed to have the same variance but possibly different means, then Student's *t*-test may be applied to investigate this question, using the standard error of the difference of the means from the pooled variance,

$$\sigma_{\text{s.e.}} = \left[\frac{\sum\limits_{i \in A}(x_i - \mu_{x_A})^2 + \sum\limits_{i \in B}(x_i - \mu_{x_B})^2}{N_A + N_B - 2} \left(\frac{1}{N_A} + \frac{1}{N_B} \right) \right]^{1/2},$$

with

$$t = \frac{\mu_{x_A} - \mu_{x_B}}{\sigma_{\text{s.e.}}}$$

and

$$A(t|\nu) = \int_{-t}^{+t} p(t')\,\mathrm{d}t' = 1 - I_x\left(\frac{\nu}{2}, \frac{1}{2}\right)$$

for computation or looking up the probability using the incomplete beta function I_x. The 'p-value', $1 - A(t|\nu)$, is then the probability that $|t|$ could be this large or larger by chance, given the number of degrees of freedom $\nu = N_A + N_B - 2$. In the case of two subsets with 'known' or computed different variances $\sigma^2(x_A)$ and $\sigma^2(x_B)$, then one computes

$$t = \frac{\mu_{x_A} - \mu_{x_B}}{\left[\sigma^2(x_A)/N_A + \sigma^2(x_B)/N_B\right]^{1/2}}$$

distributed (approximately) as the Student's t with the number of degrees of freedom

$$\nu = \frac{\left[\sigma^2(x_A)/N_A + \sigma^2(x_B)/N_B\right]^2}{\frac{\left[\sigma^2(x_A)/N_A\right]^2}{N_A - 1} + \frac{\left[\sigma^2(x_B)/N_B\right]^2}{N_B - 1}}.$$

The sample sets are called 'paired' only if the values in the two samples show a one-to-one correspondence, otherwise they are not paired or consistent (Skaik, 2015).

The form of the probability distribution function (normal, Gaussian, Lorentzian, asymmetric or numerous other observed distributions) has a large impact upon the significance of individual or propagated standard deviations and variances. A Lorentzian distribution function has long tails so that 3–6 standard error deviations can be quite common, whereas normal or Gaussian distributions vanish quickly so that three standard error deviations are major deviations of great significance. Peculiar distribution functions found in sample-growth analysis of, for example, roughness, such as top-hat and triangular distribution functions, have distribution widths which do not add in quadrature: rather, a width or σ can be identical to the input widths (i.e. $\sigma_A \oplus \sigma_B \rightarrow \sigma_A$) or can add linearly (i.e. $\sigma_A \oplus \sigma_B \rightarrow \sigma_A + \sigma_B$) (Glover et al., 2009).

There is a large body of literature on statistical inference, and a good brief summary for XAFS analysis in given in Bunker (2010). The direction commended in general here is sometimes called a frequentist approach, which is sometimes equated with a least-squares analysis, although not in any way to dismiss careful and detailed Bayesian approaches to XAFS analysis, which have been used with great insight (Krappe & Rossner, 1999, 2000, 2002, 2004; Klementev, 2001a,b).

The basic Bayesian approach is to define a conditional probability $P(\theta|\mathbf{X})$ of a theoretical set of results θ representing theoretical variables including structural and pre-analysis variables, using a priori constraints, restraints or information $\mathbf{X}$,

$$P(\theta|\mathbf{X}) = \frac{P(\mathbf{X}|\theta)P(\theta)}{\int P(\mathbf{X}|\theta)P(\theta)\,\mathrm{d}\theta},$$

where $\mathbf{X}$ is a vector of measurements and $P(a|b)$ is the conditional probability: the probability of a being true if b is true. Note that the Bayesian approach in general equates to the least-squares approach in the simple limit. Following Bunker (2010),

> viewing probability as an expression of our state of knowledge, rather than a frequency of events, this equation gives us a way to determine the probability that a given set of parameters has a range of values if we have some initial (prior) knowledge of the possibilities and we have additional information from measurements.

Certain assumptions made in standard statistics may not be valid, including the statistical independence of data points and the normal distribution of experimental errors. Hence questions such as this (Are the data points statistically independent? Are experimental errors normally distributed? Are the a priori assumptions true?) should be measured for the relevant data set.

> In most EXAFS data analysis the statistical independence of data points is compromised because fitting is usually performed to (heavily) processed data, including interpolation which correlates the information content and errors of these data points, even if fitting in k-space

(Schalken & Chantler, 2018). This is addressed in particular in *mu2chi* and *eFEFFIT* analysis (Schalken & Chantler, 2018; Trevorah et al., 2019, 2020), for example.

A key recommendation is therefore to analyse and fit data where the experimental uncertainty is well defined, or as well defined as possible, and with little processing such as interpolation, background subtraction and spline removal, and to especially be careful of transformation from k-space to R-space and filtering or back-transformation to Q-space with concomitant propagation of unknown additional systematic errors and uncertainty. Additionally, be careful and avoid directly fitting in $k^n\chi$-space without considering the scaling of uncertainties from $[\mu/\rho]$ versus E to χ versus k space to $k^n\chi$ versus k space (Chantler, 2024c).

In particular, by looking at the unprocessed and raw data (Pettifer & Cox, 1983; Pettifer et al., 2005), i.e. with no interpolation (Schalken & Chantler, 2018; Trevorah et al., 2019) and no removal of a particular signal such as pre-edge background, we can model the experiment and derived parameters with perhaps minimal impact upon the introduction of systematic errors into the data set.

All XAFS analysis and XANES analysis is Bayesian in the sense that an a priori structure is defined by theory or by fingerprint and with this a priori set of assumptions, certain specific parameters are minimized, usually in a least-squares sense. However, often the weightings given are uniform (unweighted) whether the data and uncertainties are unscaled or scaled by k^n. Additionally, the processing environments of *IFEFFIT*, *ARTEMIS*, *LARCH* and other typical packages have a great facility for inserting constraints or restraints on parameters and structures, which are precisely Bayesian constraints, hopefully based on sound chemical and physical principles. Perhaps this is the key comment about all Bayesian constraints: if the a priori constraints are physical and known

620   Christopher T. Chantler · Data acquisition

Int. Tables Crystallogr. I (2024). ch. 4.6, 617–623

to be true then the method is often superior, more insightful and more revealing than an otherwise simple least-squares or probabilistic approach. Conversely, if the assumptions or suggested structures are false then the Bayesian approach will not be helpful. A key requirement for insight is the determination, before the fit, of realistic uncertainties which can be measured by precision or preferably, if possible, by accuracies where as many physical systematics as possible have been addressed prior to the structural analysis.

Bayesian inference relates directly to XANES and principal component analysis (PCA). Good inference follows: if each reference component for PCA is experimentally or theoretically justified and calibrated, and if the sample is known to be a direct mixture of the reference materials, then PCA can use this *a priori* information to give a beautiful and insightful result, with good accuracy. Conversely, if the unknown is known not to be a mixture of the reference compounds, but maybe representative oxidation states, representative Fermi levels or representative local geometries, then the insight and conclusions may be false, which would be poor 'Bayesian inference'.

Examples of Bayesian hypotheses, which may be true or false in XAS analysis, include the following.

(i) The data set is affected by dark-current offsets.

(ii) The data are not affected by dark-current offsets.

(iii) The data include significant harmonic contamination for some energies.

(iv) The data do not include harmonic contamination for any energies.

(v) The detectors are perfectly matched and uniformly efficient.

(vi) The detectors need a blank measurement without the sample to characterize the attenuation and scattering from upstream and downstream air paths, windows and optics.

(vii) The data are affected by fluorescence in transmission.

(viii) The data are affected by self-absorption and absorption in fluorescence.

(ix) The background and background subtraction is of a particular functional form around the edge or at energies well above the edge.

(x) $S_0^2 = 1.0$.

(xi) $S_0^2 \leq 1$, $S_0^2 > 0.5 \ldots$

For any data set, a known systematic error will distort the spectrum and either impair the fit and χ_r^2 or lead to a false minimum or conclusion. Conversely, if the physics and distortion of the spectrum are predictable and defined as an *a priori* constraint, then the distortion can be measured and corrected for. Analysis may begin simply, that is by assuming some local nanostructure and thermal parameters. If a known systematic error is present (for example distortion due to dark-current offsets) then an *a priori* Bayesian correction for this should reduce the variance between repeated data points and define a higher precision, and yield a more meaningful and insightful determination of local nanostructure. In the Bayesian sense, this would be accounting for known knowns and known unknowns (see the Venn diagram in Fig. 2).

When a systematic might be present or not, a physical model for this can detect the presence of the systematic. By including such a Bayesian *a priori* hypothesis, this can prove the existence of the systematic (or its non-existence). In the Bayesian sense this allows one to test for unknown knowns; that is, to test for the presence or existence of systematics
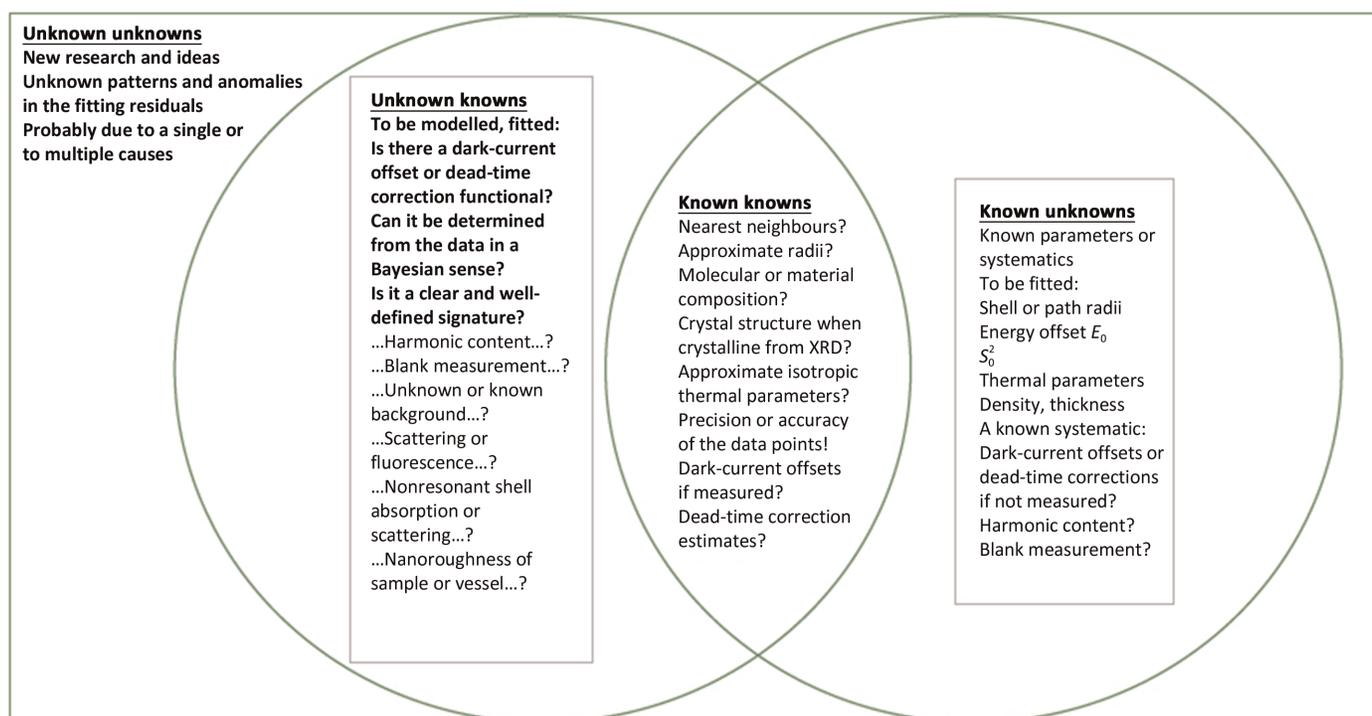

Figure 2
Principles of Bayesian processing for XAS analysis.

which are not known to be present. By mapping the functional form of a known defect or systematic of a sample or experiment, we can correct the data set in a minimal way with minimal distortion (this links directly into the next chapter). Again, the test is the significant reduction of variance in the data, increased precision and reduced χ_r^2.

Finally, Bayesian analysis can look for any pattern of variance or variability which can reveal and prove the existence of an as yet unknown systematic: an unknown unknown. By analysing the pattern, we can hopefully identify the systematic from the data. In general, analysis should apply the knowledge given by the data set with Bayesian *a priori* knowledge of the form of the sample and with knowledge of experimental defects and systematics, known to be present or unknown, to separate the precision from the accuracy in preparation for the fitting of local structure, unknown parameters and any important parameters for scientific insight (for example harmonics or dark-current offsets *etc.*).

The discussion here has stayed general, but has primarily addressed transmission data. Different detection modalities (Chantler, 2024*d*) have different uncertainties and distributions thereof to lead to variance, precision or systematics. For example, fluorescence detection by multiple pixels, say 36-element or 100-element pixel detectors, can generate full spectra for each pixel, although often and usually these are quickly pre-processed and compressed to counts within an energy discriminator region of interest (RoI). Irrespective, the (raw) spread for the pixels for a given incident energy is often 20–30% of the signal, so the variance is so large as to preclude further detailed analysis. This spread is also energy dependent and usually increases away from the edge, especially if some normalization has been made near the edge. Much of this can be addressed by *SeaFFluX* (Trevorah *et al.*, 2019, 2020), with some earlier work also being insightful (Goulon *et al.*, 1982; Tröger *et al.*, 1992; Pfalzer *et al.*, 1999; Booth & Bridges, 2005; Barnea *et al.*, 2011; Chantler, Barnea *et al.*, 2012; Chantler, Rae *et al.*, 2012). Sometimes the signal is scaled and flattened, yielding a reduced variance of the multiple-pixel data. Hence, the apparent precision might locally improve dramatically by such empirical methods, yet the systematic causing the original variance (for example self-absorption) may remain uncorrected and unaddressed, so that for example the fitting of the local structure may be heavily distorted. Conversely, if the physics underlying the variance and systematic are understood (for example if the functional form for self-absorption correction is included in the data analysis) then the precision should be significantly improved and the systematic error and uncertainty should be minimized, therefore yielding a more accurate and insightful determination of local nanostructure.

As a summary, the observed precision is determined from the variance of repeated measurements. Any undetermined systematic effect in the data or merged data will generally lead to a wider variance and a lower precision. This will then lead to a lower χ_r^2 and a weaker ability to model the data for physically unknown structural or other parameters. By addressing such sources of datum deviation, that is by addressing local structure (known knowns) and known

systematics (known unknowns) and possible systematics which may or may not be present (unknown knowns), whether by explicit Bayesian methods or by data analysis of orthogonal signatures for the physics of the effects, we can produce data sets of very good precision and insightful uncertainty for further processing and analysis. This sort of analysis can also reveal new systematics that have never been considered before (unknown unknowns), which can then be investigated. Thus, in general the variance yields an overestimate of the ideal precision and reproducibility of the data, but the two measures (variance and precision) approach one another if the systematic causes of variance are addressed and corrected for. The insight can then yield high-quality and even *ab initio* determination of parameters and uncertainties.

3. Conclusion

The recommendation here is to work towards ten repeated measurements under 'identical' conditions, at which point the variance, reproducibility and precision can be determined and anomalies can be investigated. The recommendation is to provide the data-point uncertainty for the whole spectrum for each detector.

The correlation coefficient between detectors and especially between the detector and monitor is then able to be measured, which tests the performance of the beam optic. Where the correlation between the monitor and detector is not strongly positive, particularly in transmission measurements, the data set should often be discarded, the experiment should be reoptimized and the data should be remeasured. However, this chapter also explains how to optimally provide a measure of precision in these difficult but commonly encountered cases.

The variance and correlation of the detectors can then be propagated (see the next chapter) to $([\mu/\rho][\rho t])_s$ for transmission measurements or to $a_i[\mu/\rho]_{pe}^*[\rho t]$ for fluorescence measurements.

A key recommendation is to analyse and fit data where the experimental uncertainty is well defined with little processing such as interpolation, background subtraction and spline removal, and to be careful of transformation from k-space to R-space, filtering or back-transformation to Q-space. Avoid directly fitting in $k^n\chi$-space without considering the scaling of uncertainties from $[\mu/\rho]$ versus E to χ versus k space to $k^n\chi$ versus k space (Chantler, 2024*c*).

The next chapter discusses the accuracy of an absolute measurement; that is, the ability of the measurements to obtain the true value of the variate (population mean), especially in the presence of systematic effects in the measurement (Chantler, 2024*a*). This chapter is directed towards obtaining measures of variance, standard deviation, standard error and hence precision for raw data points or preprocessed data $x(E)$ and not to discuss the use of these in fitting or in fitting packages; hence, we direct readers to the extensive sets of chapters on these (Parts 5 and 6), to the next chapter on systematics and accuracy (Chantler, 2024*a*) and to the chapter discussing the use of higher accuracy data-collection strategies (Best & Chantler, 2024).

622   Christopher T. Chantler · Data acquisition

Int. Tables Crystallogr. I (2024). ch. 4.6, 617–623

References

Abe, H., Aquilanti, G., Boada, R., Bunker, B., Glatzel, P., Nachtegaal, M. & Pascarelli, S. (2018). *J. Synchrotron Rad.* **25**, 972–980.

Barnea, Z., Chantler, C. T., Glover, J. L., Grigg, M. W., Islam, M. T., de Jonge, M. D., Rae, N. A. & Tran, C. Q. (2011). *J. Appl. Cryst.* **44**, 281–286.

Best, S. P. & Chantler, C. T. (2024). *Int. Tables Crystallogr. I*, ch. 3.14, 375–394.

Bevington, P. R. (1969). *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill.

Booth, C. H. & Bridges, F. (2005). *Phys. Scr.* **2005**, 202.

Bunker, G. (2010). *Introduction to XAFS: A Practical Guide to X-ray Absorption Fine Structure Spectroscopy.* Cambridge University Press.

Chantler, C. T. (2009). *Eur. Phys. J. Spec. Top.* **169**, 147–153.

Chantler, C. T. (2024a). *Int. Tables Crystallogr. I*, ch. 4.7, 624–630.

Chantler, C. T. (2024b). *Int. Tables Crystallogr. I*, ch. 3.43, 558–563.

Chantler, C. T. (2024c). *Int. Tables Crystallogr. I*, ch. 5.7, 664–671.

Chantler, C. T. (2024d). *Int. Tables Crystallogr. I*, ch. 2.8, 88–99.

Chantler, C. T., Barnea, Z., Tran, C. Q., Rae, N. A. & de Jonge, M. D. (2012). *J. Synchrotron Rad.* **19**, 851–862.

Chantler, C. T., Rae, N. A., Islam, M. T., Best, S. P., Yeo, J., Smale, L. F., Hester, J., Mohammadi, N. & Wang, F. (2012). *J. Synchrotron Rad.* **19**, 145–158.

Chantler, C. T., Tran, C. Q., Paterson, D., Barnea, Z. & Cookson, D. J. (2000). *X-ray Spectrom.* **29**, 449–458.

Chantler, C. T., Tran, C. Q., Paterson, D., Cookson, D. J. & Barnea, Z. (2000). *X-ray Spectrom.* **29**, 459–466.

Glover, J. L., Chantler, C. T. & de Jonge, M. D. (2009). *Phys. Lett. A*, **373**, 1177–1180.

Goulon, J., Goulon-Ginet, C., Cortes, R. & Dubois, J. M. (1982). *J. Phys. Fr.* **43**, 539–548.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences.* Cambridge University Press.

Ito, K. M. (1993). Editor. *Encyclopedic Dictionary of Mathematics*, 2nd ed. Cambridge: MIT Press.

James, F. (2006). *Statistics Methods in Experimental Physics*, 2nd ed. Singapore: World Scientific.

Klementev, K. V. (2001a). *J. Phys. D Appl. Phys.* **34**, 209–217.

Klementev, K. V. (2001b). *J. Phys. D Appl. Phys.* **34**, 2241–2247.

Krappe, H. J., Holub-Krappe, E., Konishi, T. & Rossner, H. H. (2024). *Int. Tables Crystallogr. I*, ch. 5.15, 702–704.

Krappe, H. J. & Rossner, H. 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.

Krappe, H. J. & Rossner, H. H. (2004). *Phys. Rev. B*, **70**, 104102.

Newville, M. (2024). *Int. Tables Crystallogr. I*, ch. 5.13, 690–694.

Pettifer, R. F. & Cox, A. D. (1983). *EXAFS and Near-Edge Structure*, edited by A. Bianconi, L. Inoccia & S. Stipcich, pp. 66–72. Berlin, Heidelberg: Springer-Verlag.

Pettifer, R. F., Mathon, O., Pascarelli, S., Cooke, M. D. & Gibbs, M. R. J. (2005). *Nature*, **435**, 78–81.

Pfalzer, P., Urbach, J.-P., Klemm, M., Horn, S., denBoer, M., Frenkel, A. & Kirkland, J. (1999). *Phys. Rev. B*, **60**, 9335–9339.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2007). *Numerical Recipes*, 3rd ed. Cambridge University Press.

Sayers, D. E. (2000). *Error Reporting Recommendations: A Report of the Standards and Criteria Committee*. http://ixs.iit.edu/subcommittee_reports/sc/err-rep.pdf.

Schalken, M. J. & Chantler, C. T. (2018). *J. Synchrotron Rad.* **25**, 920–934.

Sier, D., Cousland, G. P., Trevorah, R. M., Ekanayake, R. S. K., Tran, C. Q., Hester, J. R. & Chantler, C. T. (2020). *J. Synchrotron Rad.* **27**, 1262–1277.

Skaik, Y. (2015). *Pak. J. Med. Sci.* **31**, 1558–1559.

Snedecor, G. W. & Cochran, W. G. (1989). *Statistical Methods*, 8th ed. Ames: Iowa State University Press.

Streltsov, V. A., Ekanayake, R. S., Drew, S. C., Chantler, C. T. & Best, S. P. (2018). *Inorg. Chem.* **57**, 11422–11435.

Timoshenko, J. & Kuzmin, A. (2024). *Int. Tables Crystallogr. I*, ch. 5.14, 695–701.

Tran, C. Q., Chantler, C. T., Barnea, Z., de Jonge, M., Dhal, B. B., Chung, C. T. Y., Paterson, D. & Wang, J. (2005). *J. Phys. B At. Mol. Opt. Phys.* **38**, 89–107.

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.

Tröger, L., Arvanitis, D., Baberschke, K., Michaelis, H., Grimm, U. & Zschech, E. (1992). *Phys. Rev. B*, **46**, 3283–3289.

Int. Tables Crystallogr. I (2024). ch. 4.6, 617–623

Christopher T. Chantler · Data acquisition   **623**