About estimation of fitted parameters' statistical uncertainties in EXAFS. Critical approach on usual and Monte Carlo methods

Emmanuel Curis[a]* and Simone Bénazeth[a,b]

[a]Laboratoire de Biomathématique, Faculté de Pharmacie, Université René Descartes, 4 avenue de l'Observatoire, 75006 Paris, France, and [b]LURE, Bâtiment 209D, Centre Universitaire Paris-Sud, 91405 Orsay CEDEX, France. E-mail: emmanuel.curis@univ-paris5.fr

An important step in X-ray absorption spectroscopy (XAS) analysis is the fitting of a model to the experimental spectra, with a view to obtaining structural parameters. It is important to estimate the errors on these parameters, and three methods are used for this purpose. This article presents the conditions for applying these methods. It is shown that the usual equation $\Sigma = 2H^{-1}$ is not applicable for fitting in R space or on filtered XAS data; a formula is established to treat these cases, and the equivalence between the usual formula and the brute-force method is evidenced. Lastly, the problem of the nonlinearity of the XAS models and a comparison with Monte Carlo methods are addressed.

Keywords: EXAFS; statistical error estimation; parameter fitting; Monte Carlo; covariance matrix.

1. Introduction

The last step during the analysis of X-ray absorption spectroscopy (XAS or EXAFS) spectra is the fitting of a theoretical model to the experimental data. During this process, values are proposed for the various parameters describing this model, that is, the description of the local spatial environment of the absorbing atom (nearest neighbour distances, numbers and nature; thermal motions, through the use of the Debye–Waller factor).

The question of how much we can trust the obtained values then naturally arises. This question is, in fact, twofold; we may wonder about the validity of the model or, assuming that the model is correct, we may wonder about the precision of the fitted values related to this model. Despite the fact that these two aspects are somehow related, they lead to a different analysis and hence can be studied separately. In this work, we will only deal with the second problem: once a model is selected, what information is available on the quality of the estimation obtained after fitting? The fundamentals of the statistical tools used to answer this question can be found in several textbooks (*e.g.* Saporta, 1990; Neuilly, 1993; Protassov, 1999).

From a statistical point of view, the answer to this problem is contained in the variance–covariance matrix, Σ, of the vector of all the fitted parameters; the square roots of the diagonal terms of this matrix give the standard deviation of the parameters and the off-diagonal terms give the statistical correlations between the parameters. Hence, the aim is to obtain this matrix from the experimental data. Since this covariance matrix depends not only on the experimental data but also on the method applied to determine the parameter values, we will first recall briefly the various methods encountered in X-ray absorption analysis and will present the formalism used throughout this article.

To obtain the covariance matrix, a few different methods have been proposed by various authors in the field of the XAS. Three of these methods are summarized in the report from the Standards and Criteria Committee (2000): (i) use of the analytical equation $\Sigma_{\Theta} = 2H^{-1}$, where H is the Hessian of the fit ('analytical method'), (ii) use of the isocontour levels of the minimized quantity (the so-called 'brute-force method') and (iii) use of the Monte Carlo methods. Another method is also proposed, based on a Bayesian approach (Krappe & Rossner, 1999; Rossner & Krappe, 2001; Klementev, 2001); since this method is built on completely different hypotheses, we will not discuss it here.

All these methods rely on various assumptions, which may or may not be verified during a common analysis of the experimental data. To our knowledge, no study has been undertaken about the validity of these assumptions for real cases, and only very few studies have mentioned the comparison of these methods to simplified (Filiponi, 1995) or theoretical (Ghigna *et al.*, 2001) cases. The aim of this work is to analyse the assumptions involved in the application of the different methods, and then to compare the various methods (when applicable).

We will first detail all the assumptions on which the most frequently used method relies ('analytical method') and we will present extensions of this method to remove some of these

limitations. We will then deal with the brute-force method and show that it relies on the same assumptions as the previous one. Lastly, we will present and compare various Monte Carlo methods. Details of the demonstrations are reported in Appendices A–C.

2. Position of the problem

We want to obtain, from the experimental data, the values of q parameters, denoted θ_1 to θ_q. These parameters form the q components of a vector θ. The experimental data form a set of N values, $y_1 \ldots y_N$, functions of a known quantity x. Hence, one proposes for these values a model $y_i = f(x_i, \theta_1, \ldots, \theta_q)$. Using vector notation, this model can be rewritten $\mathbf{y} = f(\mathbf{x}, \theta)$. Hence, we need to 'invert' the function f so that we may write $\theta = g(\mathbf{x}, \mathbf{y})$. The problem is to select a correct g function. In general, its analytical form is unknown and it remains implicitly defined, as we will see later.

Practically, the measured values are not exact; a statistical error adds to their true value (and, in some cases, a systematic error; we will not discuss this here, since it is equivalent to the case of an incorrect model, which we will discuss later). Hence, each measured value y_i is considered the realization of a random variable Y_i (in a similar way, x_i is the realization of a random variable X_i). Using vector notation, the experimental vector $\mathbf{y}$ is a realization of the random vector $\mathbf{Y}$, as $\mathbf{x}$ is a realization of $\mathbf{X}$.

Consequently, the set of parameters deduced from an experiment, θ, is itself a realization of the random vector Θ, defined as $\Theta = g(\mathbf{X}, \mathbf{Y})$. Knowledge of the precision of the obtained values is then equivalent to knowledge of the random vector Θ, and in particular of its covariance matrix Σ_Θ. These properties are dependent on the properties of the experimental vectors, $\mathbf{X}$ and $\mathbf{Y}$ (hence dependent on the experimental conditions), and also on the properties of the g function (hence dependent on the method used to obtain the parameters).

When the measurements are repeated m times for each x value, m random vectors, $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$, can be defined, and the relation between the fitted parameters and the experimental data becomes $\Theta = g(\mathbf{X}, \mathbf{Y}_1, \ldots, \mathbf{Y}_m)$.

In most cases, the known quantity (x) is controlled by the user and measured with great precision, and hence its associated statistical uncertainty may be neglected. In that case, the random vector $\mathbf{X}$ is completely defined by the experimental values $\mathbf{x}$.

The properties of the random vector $\mathbf{Y}$ are strongly dependent on the nature of the experiment, and hence we will deal with this aspect later, when applying this model to XAS data analysis.

We then have to study the effect of the function g. As stated previously, this function is not defined by an analytical (explicit) formula. Two methods are mainly used to define g: least squares (including many variants) and maximum likelihood.

2.1. Least-squares methods

Least-squares methods are well known; we will just recall here the principle and the matricial and analytical formulations, with $\mathbf{Y}_{\text{th}}$ the random vector of the N theoretical values predicted by the model. The interested reader is referred to statistical textbooks (for instance, Draper & Smith, 1981) for more information on these methods.

The simplest least-squares method is based on the minimization of the Euclidean distance between the experimental vector and its model; the parameter values minimize the quantity

$$\xi_{\text{s}}^2 = {}^{\text{t}}(\mathbf{Y} - \mathbf{Y}_{\text{th}})(\mathbf{Y} - \mathbf{Y}_{\text{th}}) = \sum_{i=1}^{N}\left[y_i - f(x_i, \Theta)\right]^2.$$

To include the experimental uncertainties, weighted ('statistical') least-squares are introduced,

$$\xi_{\text{w}}^2 = {}^{\text{t}}(\mathbf{Y} - \mathbf{Y}_{\text{th}}) S_{\mathbf{Y}}^{*-1} (\mathbf{Y} - \mathbf{Y}_{\text{th}}) = \sum_{i=1}^{N}\left\{\left[y_i - f(x_i, \Theta)\right]/s_i^*\right\}^2,$$

where $S_{\mathbf{Y}}^{*-1}$ is the covariance matrix estimated from the experiment; its diagonal terms are the squares of the experimental standard deviations s_i^*. Note that the analytical formula is correct only if the points are uncorrelated (that is, $S_{\mathbf{Y}}^{*-1}$ is diagonal), but the matricial formulation does not require this assumption.

Other weighting schemes may be used, giving the generalized least-square methods,

$$\xi_{\text{g},w}^2 = {}^{\text{t}}(\mathbf{Y} - \mathbf{Y}_{\text{th}})W(\mathbf{Y} - \mathbf{Y}_{\text{th}}) = \sum_{i=1}^{N} w^2(x_i)\left[y_i - f(x_i, \Theta)\right]^2,$$

where $W = [w(x_i)\,w(x_j)]_{i,j}$ is a symmetric definite-positive matrix. The analytical form given here is valid only for diagonal W matrices; if W is not diagonal, then the analytical formulation becomes

$$\xi_{\text{g},w}^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[y_i - f(x_i, \Theta)\right] w(x_i)\,w(x_j) \left[y_j - f(x_j, \Theta)\right].$$

Simple least squares and weighted least squares are in fact special cases of the generalized least-squares method, with, respectively, $W = I$ and $W = S_{\mathbf{Y}}^{*-1}$. Hence, results can be derived directly for the generalized form.

In all cases, the function g is defined as the function that associates to the observed experimental values, $\mathbf{y}$, the parameters vector, θ, that minimizes ξ^2. In general, the analytical expression of g is not known.[1]

In the case of m repetitions of the measure, averages can be used instead of the experimental values (and, for weighted least squares, the estimated standard deviation of the average, $s_{\text{m},i}^* = s_i^*/m_i^{1/2}$).

---

[1] In fact, it may happen in some cases that this definition of θ is not precise enough, since many different sets can minimize ξ^2. It is then necessary to add information to this definition, such as 'we use the global minimum'. We will not discuss such cases here, since they do not change the result of the analysis, and we will always assume that g is completely defined.

2.2. Maximum-likelihood method

The maximum-likelihood method is based on a completely different idea; the assumption is that the observed experimental values are the most likely to be observed. The searched parameter set is then the one that gives the highest probability of observing the measured experimental data.

This method gives the best estimations of the parameters (that is, generally, the estimations of minimal variance, hence of minimal standard deviation). In particular, when the number of repeated measures, m, becomes high, the properties of the random vector Θ are optimal; it is asymptotically Gaussian ('normal'), unbiased and of minimal variance (Saporta, 1990). However, when m is small, the properties remain unknown.

Details of the derivation of the maximum-likelihood formula, in the case of Gaussian experimental values, are presented in Appendix A. The result is the function given below, which is a special case of generalized least squares with $W = \Sigma_{\mathbf{Y}}^{-1}$; this method will be called maximum-likelihood least squares,

$$
\begin{aligned}
\xi_{\mathrm{ML}}^2 &= {}^{\mathrm{t}}(\mathbf{Y} - \mathbf{Y}_{\mathrm{th}}) \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \mathbf{Y}_{\mathrm{th}}) \\
&= \sum_{i=1}^{N} \left\{ [y_i - f(x_i, \Theta)]/(\sigma_i/m_i^{1/2}) \right\}^2.
\end{aligned}
$$

The hypotheses to apply this formula are:

(i) The experimental distribution is normal. This condition is satisfied in X-ray spectroscopy (Curis & Bénazeth, 2001).

(ii) The experimental points are independent for the analytical formulation. This statement is true only before Fourier filtering.

(iii) The model is exact (there is no systematic error of any kind). This hypothesis is only a first approximation in XAS analysis.

(iv) The standard deviation does not depend on fitted parameters. Rigorously, this statement is not true, since absorption follows a Poisson law, for which standard deviation and average are equal; practically, other sources of statistical errors are more important and make this effect negligible.

If the experimental uncertainties were perfectly known, σ_i and s_i^* (or $\Sigma_{\mathbf{Y}}$ and $S_{\mathbf{Y}}^*$) could be assimilated. In that case, we recognize the weighted ('statistical') least-squares estimator $\xi_{\mathrm{w}}^2 = \xi_{\mathrm{ML}}^2$, and these two methods are then equivalent. However, this point is seldom satisfied in practice.

2.3. Application to XAS analysis

In the case of XAS analysis, different kinds of fit are developed, corresponding to various chosen definitions for x and y: fit of a complete absorption spectrum ($x = E$, $y = \mu$); fit of EXAFS oscillations, before or after filtering ($x = k$, $y = \chi$); and fit of the Fourier transform $[x = R, y = \mathcal{F}(k^p \chi)]$. To avoid heavy formulation, we will present formulas only in the case of EXAFS oscillations or in the general case with y and x, but the computations and the results are valid for any other kind of fit.

As mentioned previously, we will assume high precision of the x values. This assumption does not mean that these values are exact, since various systematic effects can interfere with

their determination (uncertainty in the edge energy, E_0, for instance), but only that no statistical component exists in this error. Hence, this assumption does not prevent the use of parameters in the model to correct x values, as is done, for instance, by introducing the ΔE_0 parameter.

As shown by Curis & Bénazeth (2001), a Gaussian ('normal') distribution can be assumed for the random vector $\mathbf{Y}$. Hence, it is completely defined by its expectation value and its covariance matrix, $\Sigma_{\mathbf{Y}}$. Both are estimated either directly, averaging m experimental spectra, or by the error propagation model developed by Curis & Bénazeth (2000). Let us recall that this covariance matrix is diagonal only for a spectrum before Fourier transformation, an important point for the end of this study.

2.3.1. Use of least squares. We use here a direct transposition of the formula presented above, taking into account the fact that we are working on the averaged spectrum (for which the standard deviation on the ith point is then $s_i^*/m_i^{1/2}$); hence we will present the formulas only for generalized least squares. The most frequently used weighting scheme for generalized least squares is a k^p weighting, with p an integer (usually 1, 2 or 3), for the EXAFS oscillations; no special weighting is used for other fits. The corresponding minimized quantities are

$$
\xi_{\mathrm{g},p}^2 = \sum_{i=1}^{N} \left[k_i^p \chi(k_i) - k_i^p f(k_i; \theta_1, \ldots, \theta_q) \right]^2.
$$

Note that, for $p = 0$, simple least squares are obtained ($\xi_{\mathrm{g},0}^2 = \xi_{\mathrm{s}}^2$). Moreover, there is an equivalence between weighted least squares and $\xi_{\mathrm{g},p}^2$ if, and only if, a constant value s_{m} exists, so that $s_{\mathrm{m},i}^* = s_i^*/m_i^{1/2} = k_i^p s_{\mathrm{m}}$. This equivalence can be achieved, for a given experiment, only for a single value of p (most often, $p = 0$) and if each point is recorded the same number of times; however, examples of non-constant uncertainties are well known (Vaarkamp, 1998), and in such cases there is no relation between any of the $\xi_{\mathrm{g},p}^2$ and ξ_{s}^2 values.

2.3.2. Use of maximum likelihood. Rigorously, maximum likelihood is not used in XAS analysis, because none of the maximum-likelihood conditions are satisfied. However, in the case of the fit of an unfiltered spectrum with constant error bars, the only unsatisfied assumption is that there is no systematic error. Hence, it is practically equivalent to simple least squares, ξ_{s}^2.

3. The bases and the limits of the 'analytical' method

For convenience, an analytical formula is widely used to obtain the covariance matrix, $\Sigma_{\Theta} = 2H^{-1}$, where Σ_{Θ} is the covariance matrix of the fitted parameters vector and H is the Hessian (second derivatives matrix) of the minimized value, ξ_{w}^2 or $\xi_{\mathrm{g},p}^2$ in general; the parameters are implicitly supposed to follow a normal distribution. However, the derivation of Σ_{Θ} can only be performed if the model is linear; hence, rigorously, this method cannot be applied to XAS, where the models are in most cases nonlinear.

Nevertheless, it is well known that this formula may give a close approximation if the model is nearly linear [see, for the

research papers

case of XAS, Filiponi (1995)]. We will discuss later under which conditions this approximation may be used; however, an important point must be stressed: even if the model is linear, the $\Sigma_\Theta = 2H^{-1}$ formula cannot always be used. Hence, before applying this method to XAS, these conditions must be checked.

The derivation of the formula, presented in Appendix B, allows us to give the limits for using the formula:

(i) Maximum-likelihood least squares (ξ^2_{ML}) are used, which require that the experimental covariance matrix is perfectly determined.

(ii) There is no constraint, nor restraint, while fitting; each parameter can reach any value. This is not the case in XAS, since for physical reasons almost all parameters must be positive. However, as far as this criterion is not introduced as a constraint but only as a final check of the model, this condition is satisfied.

(iii) The model is linear.

The practical realization of the fitting procedure in all the known software adds a limitation; the weighting matrix Σ_Y is expected to be diagonal, and hence the experimental points must be independent. Such is not the case for Fourier-transformed data or filtered spectra, but it can be assumed for unfiltered spectra (Curis & Bénazeth, 2000). Other programming details, such as the algorithm used to minimize the least squares, do not change the conditions of use of the formula as far as the second point (no constraint) is concerned.

The first point (perfect knowledge of the covariance matrix) is not achieved, and the estimated matrix $S^{*-1}_{\mathbf{Y}}$ is used instead. Therefore, in practice, the statistical least squares ξ^2_w are applied. Practically, this does not change the formula; however, it changes the law of the Θ vector, since $S^{*-1}_{\mathbf{Y}}$ itself is random. It can be shown that for a single parameter the exact law used to build the confidence intervals is a Student law (Saporta, 1990); however, considering the high number of experimental points used in classical XAS spectra, this Student law can be well approximated by a Gaussian law. Hence, if ξ^2_w is used, we can apply the $\Sigma_\Theta = 2H^{-1}$ formula.

3.1. Generalization of the 'analytical' method

If any least-squares method other than the statistical least-squares method is used, the $\Sigma_\Theta = 2H^{-1}$ formula is wrong even in the linear approximation.

For simple least squares ($\xi^2_s = \xi^2_{g,0}$), if the error bars are constant and the experimental points independent, the formula can easily be adapted to $\Sigma_\Theta = 2s^{*2}H^{-1}$, where s^* is the standard deviation estimated from all the data points (best estimation of the real standard deviation, σ) (see Appendix B).

For any other case (and especially the use of a k^p weighting scheme when fitting, or any Fourier transform fit), there is no other way than using the complete formula obtained in Appendix B,

$$\Sigma_\Theta = 4(H^{-1}FW)\Sigma_{\mathbf{Y}}{}^t(H^{-1}FW).$$

3.2. Linear approximation of a nonlinear model

The most evident unsatisfied hypothesis in XAS analysis is the linearity of the model. Indeed, the theoretical model of the XAS oscillations is definitely not a linear function of the structural parameters. Hence, to what extent can the previous results give acceptable answers?

Roughly, the model linearization can solve the problem. Close to the minimum, the model function f can be approximated by its first-order Taylor expansion f_{lin}. The new model obtained in this way is then linear, and the previous results may be applied.

Nevertheless, this simple approach is not enough. The properties of the parameters vector Θ depend on the implicit function g and not directly on the model function f. When replacing the model function f by its linearized form f_{lin}, the function g is also replaced by a new implicit function, g_{lin}. However, there is no evidence that this new function g_{lin} constitutes a good approximation of the real function g, even if f_{lin} is a reasonable approximation of f. Hence, there is no evidence that the results of the linear model are still applicable, in the most general case.

Practically, one may assume that g_{lin} is a good approximation of g if the minimized estimator and the model are 'regular' enough, having not too many local minima, initial values for the parameters close to the real ones and quite precise (small uncertainties) experimental values. However, these conditions are not always satisfied.

Let us assume the linear approximation anyway. The results for the linear model can then be applied, but for the linearized model, as far as the other conditions stated previously are verified. In particular, to apply the usual formula $\Sigma_\Theta = 2H^{-1}$, the H matrix to use is the Hessian of the linearized model (H_{lin}) and not the Hessian of the complete model (H_0). These two matrices are related,

$$H_{lin} = H_0 + 2\left[\sum_{i=1}^{N} \varepsilon^*_i \frac{\partial^2 f}{\partial\theta_b \partial\theta_c}(k_i; \theta^*_1, \ldots, \theta^*_q)\right]_{1 \le b, c \le q},$$

where $\theta^*_1, \ldots, \theta^*_q$ is the parameter set obtained after fitting, and $\varepsilon^*_i = y_i - f(k_i, \theta^*_1, \ldots, \theta^*_q)$ is the ith individual residual. Practically, the second matrix is often negligible, since the residuals are expected to be small and to present various signs (if the model is correct), but the behaviour of the second derivatives is more difficult to foresee. Note also that most of the minimization algorithms involving the linearized matrix (including the simple Levenberg–Marquardt method) compute the linearized Hessian matrix instead of the complete one (Press *et al.*, 1993). Nevertheless, the implementation of the linear result and, especially, of the Hessian matrix before its inversion must account for this result.

4. The bases and the limits of the 'brute force' method

When the use of the previous method is questionable, some authors propose the building of a contour map for the minimized function, ξ^2. Then, for a given parameter, the limits for its confidence interval are obtained from the values increasing

this minimized function by a certain amount [usually $\Delta\xi^2/(N-q)=1$ for a so-called '1σ' confidence interval, with the underlying idea of a 65% confidence interval with a Gaussian distribution]. We will establish now that, in fact, this method relies on the same hypotheses as the use of the $\Sigma_\Theta = 2H^{-1}$ formula and hence presents the same limitations.

The confidence region is defined from the distribution function of the parameters vector, f_Θ. This region is centred on the observed value of the vector and its boundaries are defined by the equation $f_\Theta(\theta) = C$, where C is a constant determined by the confidence degree we want to obtain on the region; the boundaries are *isodensity* surfaces. Hence, the isocontours of ξ^2 around the minimum – which are practically computed – are related to the (statistical) confidence intervals only if these isocontour lines are also isodensity lines. Let us consider when this is the case, first for the linear model to simplify the formulation.

For a linear model, the ξ^2 function has a quadratic form. The function represents a q-dimensional paraboloid, and its Hessian does not depend on the point in this q-dimensional space. Hence, the isocontours of ξ^2 around the minimum are defined by the equation ${}^t\theta\, F\, W\, {}^tF\, \theta = C$; they are q-dimensional ellipsoids (see Appendix C).

If the experimental distribution is Gaussian, the parameters vector is a q-dimensional Gaussian vector. Hence, the isodensity curves centred on the expectation value are also q-dimensional ellipsoids, defined by the equation $(1/2)\,{}^t\theta\,\Sigma_\Theta^{-1}\theta = C$ (where C is not necessarily the same constant as above).

These two sets of equations will give identical regions if and only if $\Sigma_\Theta^{-1} = F\,W\,{}^tF = H/2$. In other words, this is the case when $\Sigma_\Theta = 2H^{-1}$, which is true only if the conditions stated above are satisfied. In other cases, the isocontours of the minimum and the isodensity surfaces are ellipsoids with different axes or scales; hence there is no relationship between the confidence intervals and the intervals built on the isocontour levels of the minimized function. Fig. 1 presents a simple case of another weighting scheme for a two-parameter linear model, showing the different orientation of the ellipses obtained by the two schemes.

When the model is nonlinear, in general we cannot derive the analytical equations defining the isodensity levels; there is no reason to still observe ellipses, even if the isodensity levels asymptotically tend to ellipses (since least squares are asymptotically Gaussian), and the isocontours will not be ellipses either [see, for instance, Bates & Watts (1988)]. However, it can be shown that, for maximum-likelihood least squares, the isocontours and isodensity levels still have the same shape and orientation (Bates & Watts, 1988), but there is no simple way to associate confidence levels to these confidence intervals. For other least squares, there is still no connection between isocontour and isodensity regions.

In conclusion, for nonlinear least squares, the resulting interval is not a confidence interval, since we cannot associate to this interval a probability to contain the real value; worst, for joint intervals the calculated interval may have a different shape from the true confidence interval. Nevertheless,

building these isocontour lines can give hints about the validity of the linear approximation; the more distorted they are compared with true ellipsoids, the less valid is the linear approximation (and the worse are the associated confidence probabilities). If confidence intervals are not sought, this method gives a rough value for the error bars, but if probabilities, or more precise values of error bars, are required, another method must be used.

5. The bases and the limits of the 'Monte Carlo' method

The two previous sections showed that neither method gives the expected answer as far as nonlinear models are involved, as is the case in XAS. The third method, the Monte Carlo method, will remove this limitation of using only linear models. This method, although rarely used, has already been presented by several authors (Filiponi, 1995; Ellis, 1995; Ghigna *et al.*, 2001). However, except in the last case, we found no detailed analysis of the results or of the Monte Carlo model choice. Ghigna *et al.* (2001) present the simple case of theoretical spectra fitted by a theoretical model and do not take into account the case of real experimental data.

We discuss here the various Monte Carlo methods that can be used, their strengths and weaknesses, and the information that can be extracted from a Monte Carlo simulation.

5.1. Principle of the Monte Carlo methods

In the methods presented in the first part of this article, we only obtain a single realization of the random vector of the fitted parameters, Θ: the set of parameters obtained after fitting the model to the experimental data. The properties of this vector can be determined only through analytical methods, but these can be used only in the linear case. Nevertheless, we can imagine obtaining these properties through statistical tools, which are much more generic but which have the disadvantages of being non-exact and of needing a large number of realizations of the studied vector to
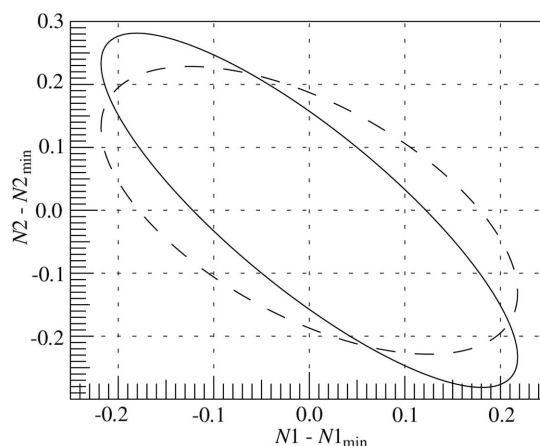

Figure 1
Example of isocontour (dashed) and isodensity (continuous) levels for a generalized linear least-squares model (see Appendix C). Curves have been rescaled to evidence the different orientation of the joint confidence regions.

obtain significant results. Hence, to apply these tools, we need many fit results.

The Monte Carlo method is developed to obtain this collection of fit results without having to repeat the same experiment many times; pseudo-experimental spectra are generated by a computer and then fitted by exactly the same model to obtain a new realization of the random vector Θ. If we generate M spectra, we will obtain M sets of parameters, θ_1 to θ_M. From these realizations, the application of the usual statistical tools will allow us to estimate the expectation, the covariance matrix or even the distribution function of Θ. The estimation of the expectation will give the 'true' value of the parameter; the diagonal of the estimated covariance matrix will give the error bar on this parameter; the off-diagonal terms of this matrix will give the correlations between the parameters; and the distribution function will allow us to assign an estimated probability to the confidence interval we want to define for the parameter.

To apply the Monte Carlo method, we must generate 'experimental' spectra. To achieve this aim, two questions must be answered: firstly, how will these spectra be generated, and, secondly, how many spectra do we need to generate? We will now study these two questions.

5.2. Which Monte Carlo method?

We can imagine several ways to generate pseudo-experimental spectra, which will lead to different variants of the Monte Carlo method. All these methods can be used for different purposes. Basically all the methods can be split into two classes, but they are all based on the generation of random values, which requires the knowledge of their distribution law.

5.2.1. Monte Carlo in experimental space. This class of Monte Carlo methods is founded on the generation of pseudo-experimental values. The most direct approach is to generate random values directly for the experimental vector $\mathbf{Y}$.

If nothing is known about that law, a possible method is to mix the values coming from a set of experimental spectra to create a new spectrum (for instance, point 1 comes from spectrum 1, point 2 from spectrum 3 *etc.*). This method is called the bootstrap method (Saporta, 1990), and its only working hypothesis is that the experimental data points are independent (hence it cannot be applied to Fourier transforms or filtered spectra). However, this method has the disadvantage of overestimating the error bars on the parameters, because the fit is realized on a single experimental spectrum and not on the average. Hence, if m spectra were recorded, we expect an order of about $m^{1/2}$ between this method and the others. For this reason, and since we have information on the experimental distribution in XAS, we did not consider this variant.

Conversely, if everything is known about the law of $\mathbf{Y}$, the method is very powerful; this is the approach used by Ghigna *et al.* (2001). However, in this case we in fact fit a theoretical model on a theoretical spectrum with theoretical error bars; this approach is useful for theoretical comparisons between different methods but is useless in daily analysis.

Intermediate variants assume that the family of laws for $\mathbf{Y}$ is known theoretically but its parameters are experimentally determined. Two cases are of interest in XAS, since it is known that the experimental distribution is Gaussian (Curis & Bénazeth, 2001), and thus completely determined by the knowledge of its expectation and covariance matrix:

(i) The expectation of $\mathbf{Y}$ is known (it is a theoretical model), and the covariance matrix is the experimental covariance matrix. This case assumes that the model is correct, and hence it is well suited to studying questions concerning the quality of a model, which is not the subject of this article. However, we will see later that in the case of an incorrect model the deviation from linearity is much stronger. Since, in real cases, the model cannot be perfect because of various systematic errors, this deviation may be important but would be neglected by this method. Hence, it is not well suited to studying the uncertainties on the fitted parameters, and we will not consider it further in this article.

(ii) The expectation of $\mathbf{Y}$ is the averaged experimental spectrum, and the covariance matrix is the experimental covariance matrix. This case can work even if the model is not perfect but, as we will see later, this approach is not suited to studying the quality of the model. We will consider only this method in this article.

A completely different approach to generate values of $\mathbf{Y}$ was proposed by Press *et al.* (1993): it consists in the use of the law of Θ to generate values of θ, from which (using the model) a spectrum is modelled, then noise is added. This method requires knowledge of the distribution of Θ, which is one of the questions we ask in this article, and hence is not suitable here and will not be considered.

5.2.2. Monte Carlo in parameter space. If the law of $\mathbf{Y}$ is not known or if the computations become too time-consuming (since each generated spectrum requires a minimization), an alternative method is to explore the parameter space by generating random values of the parameters and computing ξ^2 for each set, without minimization. Various methods are available to achieve this aim: (i) direct generation of parameters using their distribution law, as carried out above; (ii) the Markov–Metropolis Monte Carlo method (Filiponi, 1995), which defines a constrained random walk in the parameter space that converges toward the minimum; and (iii) the reverse Monte Carlo method (Gurman & McGreevy, 1990), which generates directly atomic positions in the structural model. Note that the last two methods are, in fact, minimization methods and that a set of parameter values is a side product.

These three methods share the same limitations. Firstly, they require a distribution law for generating the values, but this distribution is not known. This shortfall does not matter when the study is limited to minimization but may lead to incorrect results if one wants to consider the diverse values obtained all along the minimization path, seen as a representative set of the parameter values. Secondly, all these methods study the behaviour of ξ^2 as a function of the parameters and hence give information on the ξ^2 isocontour regions. However, these methods do not consider the fact that

ξ^2 changes for each different experimental set; in other words, these methods do not consider the effect of statistical randomness on the function g. Hence, they do not give any information on the isodensity regions when they are not the same as the isocontour regions; these methods have the same limitation as the 'brute force' method.

Because of these two limitations, these methods are not suitable for the present study and will not be considered further.

5.2.3. Properties of the selected Monte Carlo method.
Because the fits are applied on the averaged spectrum, an average spectrum should be generated each time, using the error on the average, $\sigma_i/m_i^{1/2}$. However, since the error itself is determined from the experiment, the error is also random; if this randomness is neglected, the uncertainties of the parameters are underestimated.

The correct way to handle this problem would be the use of a three-step procedure, using the known law for an estimated covariance matrix to first generate errors, then apply these errors to generate m_i values and then average them. However, this process would be very time-consuming. Since the number of repetitions is small in XAS (each spectrum is recorded around three or four times, in general), we adopted a quicker way; we use the experimental error bars to generate averaged spectra, hence introducing an $m_i^{1/2}$ factor, which should compensate for the uncertainty in the error bars or at worst should overestimate errors (since we are then roughly equivalent to the bootstrap method, from this point of view).

Another problem is the fact that the expectation is not perfectly known; to the real expectation μ_{true}, an error term is added and the expectation becomes $\mu_{\text{true}} + \delta$. Hence, when generating a new spectrum, its values are $y_i = \mu_{i,\text{true}} + \delta_i + \delta'_i$ (where δ_i is fixed and δ'_i varies from one generated spectrum to another), and we minimize in fact

$$\xi_s^2 = \sum_{i=1}^{N}\big[\chi_{\text{true}}(k_i) + \delta_i + \delta'_i - f(k_i, \theta_1, \ldots, \theta_q)\big]^2.$$

The presence of the extra δ_i term produces two effects. Firstly, since the model is now slightly different, the obtained parameters are slightly biased, *i.e.* their expectation is slightly different from the true one. However, since the nonlinearity of the model itself implies that a bias exists (even in the linearized form of the model), we can assume this effect to be negligible.

Secondly, we can write

$$\xi_s^2 = \xi^2 + \sum_{i=1}^{N}\delta_i^2 - 2\sum_{i=1}^{N}\delta_i\varepsilon_i,$$

where ξ^2 is the true residual (exactly corresponding to that minimized for the experimental spectrum) and $\varepsilon_i = y_i - f(k_i, \theta_1, \ldots, \theta_q)$ is the individual residual. Hence, the minimized residual is biased compared with the true one; since the last term remains small, the Monte Carlo residual is always higher than the true one. Consequently, the study of the distribution of the residual cannot be achieved by this method,

and so it is not possible to discuss the quality of the model from the value of the residual.

5.2.4. How many spectra must we generate?.
After selecting the method, we need to determine how many spectra must be generated and fitted in order to obtain a good estimation of the Θ random vector properties.

The lowest value is obtained from the properties of the expectation estimation by the arithmetic average; the precision is around $\sigma/M^{1/2}$, where M is the number of generated spectra. More precisely, if we assume a Gaussian ('normal') distribution for the parameter, the interval that has a 99% likelihood of containing the expectation is $L = 4.66\sigma/M^{1/2}$ wide, assuming that σ is perfectly known (Dagnelie, 1973). If we want to obtain for L and σ values of the same magnitude, we need at least 22 spectra (and 2200 if we want an interval width around $\sigma/10$). In reality, σ is not known and the distribution is not Gaussian, and hence this value is a lower bound. If we want to estimate correctly the correlation terms, we need even more values. Hence, we suggest a minimum of 100 spectra.

The more numerous the spectra we generate, the more precise are the estimated values; this is a basic principle of statistics. Nevertheless, several reasons (including time) give an upper-level bound for the number of spectra.

On the one hand, it is not very useful to give a precision much higher than σ, the uncertainty in the studied parameter, which is a consequence of the experiment. The previous result shows then that it is not useful to generate more than about 1000 spectra, except in special cases when a careful study of the distribution law of the parameters is undertaken.

On the other hand, the estimates given by the Monte Carlo approach are biased, because of the method itself (see above) but also – and mainly – because of the nonlinearity of the model. This bias remains small, as will be exemplified later and as shown by Ghigna *et al.* (2001), but its existence implies that it is useless to give very precise values for the parameter expectations, since these expectations are not exactly the 'true' values of the parameters.

Note that replacing the complete model by a linearized model, as is assumed in the classical approaches (see §3.2), also introduces a bias for the parameters.

In conclusion, for a routine analysis, around 100 spectra are enough; if one is interested in the distribution of the parameters values, it is better to generate around 1000 spectra.

6. Results of the Monte Carlo method: theoretical spectra

Ghigna *et al.* (2001) studied the Monte Carlo simulation results for the case of a model fitted on a theoretical spectrum with added noise. In that paper the existence of the bias was studied but no comparison was made with the use of the $\Sigma_{\Theta} = 2H^{-1}$ formula. In this section we will study these two points; two examples will be presented.

All the computations presented were carried out using the *LASE* software (Curis, 2001). This software, freely available from http://xlase.free.fr, includes tools for the analysis of fit

results by both the Monte Carlo method and the 'covariance-matrix' method (linear approximation of the model). All the fits were realized with the maximum-likelihood estimator (which is, here, the real maximum likelihood since errors are perfectly known). Because we are also studying the distribution, we generated 1000 spectra for the Monte Carlo simulation.

Theoretical spectra were created, for simple structural models, starting from phases and amplitudes computed with the *FEFF* software (http://leonardo.phys.washington.edu/feff). Gaussian noise was added to these spectra.

6.1. Single-layer model: model for aqueous Zn^{2+}

This model is among the most simple we can imagine; it consists of a single layer of backscattering atoms, with six O atoms placed at the apexes of a regular octahedron, of which the centre is occupied by the zinc cation. The distance between the Zn and O atoms is 1.95 Å; we selected a Debye–Waller factor of 5×10^{-3} Å^2. The variables fitted were the first-shell distance (with an initial value of 2 Å) and the Debye–Waller factor (with an initial value of 0.01 Å^2). We used an error constant with k, with a magnitude of 0.01 Å. This error term is quite important; we selected it to magnify the eventual differences between the Monte Carlo and linear approximation methods.

The results are presented in Table 1. For such a simple model there is no significant difference between the results obtained by the two methods, especially when keeping in mind the difficulty of precisely estimating the correlation terms (especially in the linear case); for this quantity, only the order of magnitude is significant.

Table 1
Results for a simple model ZnO$_6$ by the classical method and the Monte Carlo simulation.

The lower part of the correlation matrix corresponds to the results of the linear approximation, the upper part to the results of the Monte Carlo simulation. Debye–Waller factors are scaled by 10^3.

Variable	Value		Correlation	
	Linear	Monte Carlo	R	σ^2_{DW}
R (Å)	1.9500 ± 0.0012	1.9500 ± 0.0012	1	−0.15
σ^2_{DW} (Å^2)	5.00 ± 0.37	4.99 ± 0.39	−0.14	1

Besides the numerical values, the distribution laws obtained for each parameter by the Monte Carlo method are all Gaussian [they all pass through the Kolmogorov test for normality (Lilliefors, 1967), at the 1% confidence level]. This result confirms the validity of the linear approximation, which has a Gaussian distribution.

6.2. Multi-shell model: model of solid cisplatin

The second example involves a more complex model, including four scattering paths (three single-scattering and one multiple-scattering paths). This models allows a good reproduction of the platinum L_{III}-edge XAS for solid cisplatin, [Pt(NH$_3$)$_2$Cl$_2$]. The errors used for this model were experimentally deduced from spectra recorded at LURE (Orsay, France), on the D21 beamline. Fitted variables were, for each shell, its path length and its Debye–Waller factor; a global edge-energy correction, ΔE_0, was added.

The results obtained by the two methods are presented in Table 2. The two methods give some very similar values for the parameters and for their error bars. The results are somehow different for the correlation terms, but, as before, this result may be simply a consequence of the difficulty of obtaining precise values for these quantities.

However, the analysis of the parameter value distributions shows that the linear approximation reaches its limits of validity; whereas it predicts a Gaussian distribution for all the fitted parameters, the use of the Kolmogorov test shows that, at the 5% confidence level, two parameters do not follow a Gaussian distribution: ΔE_0 and the length of the multiple-scattering path, R_{diag}. Nevertheless, since the values estimated for the parameters and their error bars are correct with the linear method, this difference is not so important, except if precise confidence intervals are needed.

7. Results of the Monte Carlo method: experimental spectra

The previous results show that, in the ideal case of a correct model, no significant difference appears between the results obtained by the Monte Carlo method and those obtained by the linear approximation – except for the determination of confidence intervals in some complex cases. Hence, as far as

Table 2
Results for a cisplatin model by the classical method and the Monte Carlo (MC) simulation.

The lower part of the correlation matrix corresponds to the results of the linear approximation, the upper part of the Monte Carlo simulation. Distances are in Å. Debye–Waller factors are scaled by 10^3 and expressed in Å^2. ΔE_0 is in eV.

Variable	Value (linear)	ΔE_0	R_N	σ^2_N	R_{Cl}	σ^2_{Cl}	R_{Pt}	σ^2_{Pt}	R_{diag}	σ^2_{diag}	Value (MC)
ΔE_0	$8 \times 10^{-5} \pm 0.061$	1	−0.88	0.10	−0.81	−0.42	0.36	0.07	−0.73	−0.29	0.01 ± 0.08
R_N	2.0000 ± 0.0016	−0.88	1	0.01	0.57	0.56	−0.29	−0.04	0.66	0.13	1.9998 ± 0.0016
σ^2_N	3.3000 ± 0.1636	0.41	−0.06	1	−0.40	−0.01	−0.15	−0.04	−0.12	0.02	3.302 ± 0.150
R_{Cl}	2.3300 ± 0.0011	−0.81	0.56	−0.64	1	0.20	−0.25	−0.07	0.57	0.20	2.3299 ± 0.0010
σ^2_{Cl}	3.4000 ± 0.0929	−0.16	0.38	−0.02	−0.04	1	−0.15	−0.06	0.31	0.06	3.395 ± 0.098
R_{Pt}	3.3800 ± 0.0057	0.46	−0.39	0.03	−0.32	−0.06	1	−0.09	0.12	−0.21	3.3802 ± 0.0055
σ^2_{Pt}	13.300 ± 0.9063	0.08	−0.02	−0.01	−0.08	−0.01	−0.04	1	−0.04	0.38	13.30 ± 0.8984
R_{diag}	4.3300 ± 0.0023	−0.68	0.61	−0.31	0.54	0.13	0.07	−0.02	1	0.02	4.3297 ± 0.0025
σ^2_{diag}	14.000 ± 0.3775	−0.24	0.16	−0.11	0.19	−0.05	−0.22	0.42	0.13	1	14.105 ± 0.3762

the estimator used for the minimization is the maximum-likelihood estimator (that is, as far as the formula $\Sigma_\Theta = 2H^{-1}$ is exact for a linear model), the linear approximation can be applied.

Nevertheless, under a real analysis, we never work with a correct model, because of the approximation in the theory itself and because of the systematic errors. In particular, we must use a small number of scattering paths to reproduce the experiment, to limit the number of fitted parameters, and to avoid instability problems during the fit procedure related to the influence of numerous and small contributions that tend to cancel each other. In other words, we must forget some paths in this model building, and hence this model is basically incorrect. We can then wonder if the linear approximation is still applicable.

A second problem, which is somehow a consequence of the previous one, is the fact that in some cases the maximum-likelihood estimator (weighted least squares) does not produce an acceptable model, whereas generalized least squares with a k^p weighting gives an acceptable one; this fact is, in some way, a consequence of neglecting terms, because the k^p weighting lowers the importance of the low-k values, where the paths are the more numerous. In this case, the $\Sigma_\Theta = 2H^{-1}$ formula is definitely not applicable and only the Monte Carlo method is valid (as for fits on spectra with non-independent points – fits on Fourier transforms, for instance), since we do not know any software that implements the complete formula given in §3.1

7.1. Fit with weighted least squares

We studied the experimental spectra of an aqueous Zn^{2+} solution (0.5 M zinc chloride, recorded on beamline D21, LURE, Orsay, France) at the Zn K-edge, with the structural model presented previously fitted on the total XAS oscillations. This model is known to be incomplete, since multiple-scattering paths are known to contribute to this spectrum (Kuzmin *et al.*, 1997). However, the aim here is to compare results obtained from the Monte Carlo method and from the linear method, and not to obtain precise values for fitted parameters; hence it is a good example for this study.

Numerical results for this case are presented in Table 3. The conclusions are the same as for the theoretical study; the linear approximation and the Monte Carlo method lead to the same results for the parameters and their errors; the distribution laws study indicates a normal distribution (at the 1% confidence level, with the Kolmogorov test), which is in agreement with a linear approximation.

7.2. Fit with generalized least squares

Since the linear approximation cannot give valid results, there is no need to make a comparison between the two techniques. However, when weighted least squares does not give a good result, that fact suggests interference between various paths which,

individually, are not very important; hence it becomes very difficult to obtain convergence towards the true values. We will show here how a careful 'in depth' use of the Monte Carlo results may help, in this case, to select between paths that are significant and paths that are not.

We will use the example of the solid *cis*-diammine-1,1-cyclobutanedicarboxoplatinum(II) {'carboplatin'; [Pt(NH$_3$)$_2$(OOC)$_2$C$_4$H$_6$]} Pt L_{III}-edge model. The spectra were recorded at LURE, on the D44 beamline. Details of the model are presented elsewhere (Curis *et al.*, 2001); we will present here only the final result. This model includes four scattering paths (three single-scattering paths and one multiple-scattering path). The fit was performed with a k^2 weighting, which corresponds to the use of the $\xi_{g,2}^2$ estimator; this method gives a good reproduction of the experimental EXAFS and of its Fourier transform (Fig. 2a). Monte Carlo simulation was performed with experimental error bars, and a set of 1000 spectra was generated.

The results for the Monte Carlo method are presented in Table 4; for comparison, we also give the results obtained by the linear approximation with weighted least squares (ξ_w^2 estimator). For clarity, correlation terms are not reported.

The first comment to be made about these values is that, for the third path, the uncertainty obtained by the linear model for the path length is unacceptable. With the other estimator, this uncertainty is much smaller and, while still important, is acceptable. The same effect is observed for the Debye–Waller value.

For other parameters, uncertainties are smaller with the linear model. This is understandable, since the linear model is quite close to the maximum-likelihood method, and hence will give the estimators with the smallest variance.

Table 3
Comparison of results from analytical and Monte Carlo methods for a fit on experimental Zn spectra.

Distances (R) are in Å, Debye–Waller factors (σ_{DW}^2) in Å^2, and the edge-energy correction (ΔE_0) in eV. For the correlation matrix, the upper-diagonal part presents the Monte Carlo results and the lower-diagonal part the analytical results.

Variable	Value Linear	Value Monte Carlo	Correlation ΔE_0	Correlation R	Correlation σ_{DW}^2
ΔE_0	2.84 ± 0.04	2.84 ± 0.03	1	-0.38	-0.21
R	2.056 ± 0.005	2.059 ± 0.004	-0.37	1	-0.003
σ_{DW}^2	14.49 ± 0.07	14.49 ± 0.08	-0.22	-0.003	1

Table 4
Results for a solid carboplatin model: paths and values with the linear model and the Monte Carlo model.

ΔE_0 was also fitted, giving -2.97 (12) eV with the linear model and -2.3 (2) eV with the Monte Carlo method. Distances are in Å. Debye–Waller factors are scaled by 10^3 and are in Å^2. Monte Carlo values are closer to the known structure and more physical.

Path	R, linear	R, Monte Carlo	R, cryst.	σ^2, linear	σ^2, Monte Carlo
1: Pt N/O Pt	2.0258 ± 0.0007	2.028 ± 0.002	2.025	1.87 ± 0.05	2.00 ± 0.14
2: Pt C$_{COOH}$ Pt	2.6599 ± 0.0382	2.94 ± 0.06	2.882	52.7 ± 5.78	20 ± 60
3: Pt C$_1$ Pt	2.7175 ± 2.3938	3.16 ± 0.14	3.019	500 ± 300	80 ± 18
4: Pt N Pt O Pt	4.0321 ± 0.0035	4.067 ± 0.006	4.050	0.50 ± 0.42	2.5 ± 1.4

Table 5
Results for a solid carboplatin model: paths and values with the linear model and the Monte Carlo model.

ΔE_0 was also fitted, giving -2.97 (7) eV with the linear model and -2.2 (2) eV with the Monte Carlo method. Distances are in Å. Debye–Waller factors are scaled by 10^3 and are in Å^2.

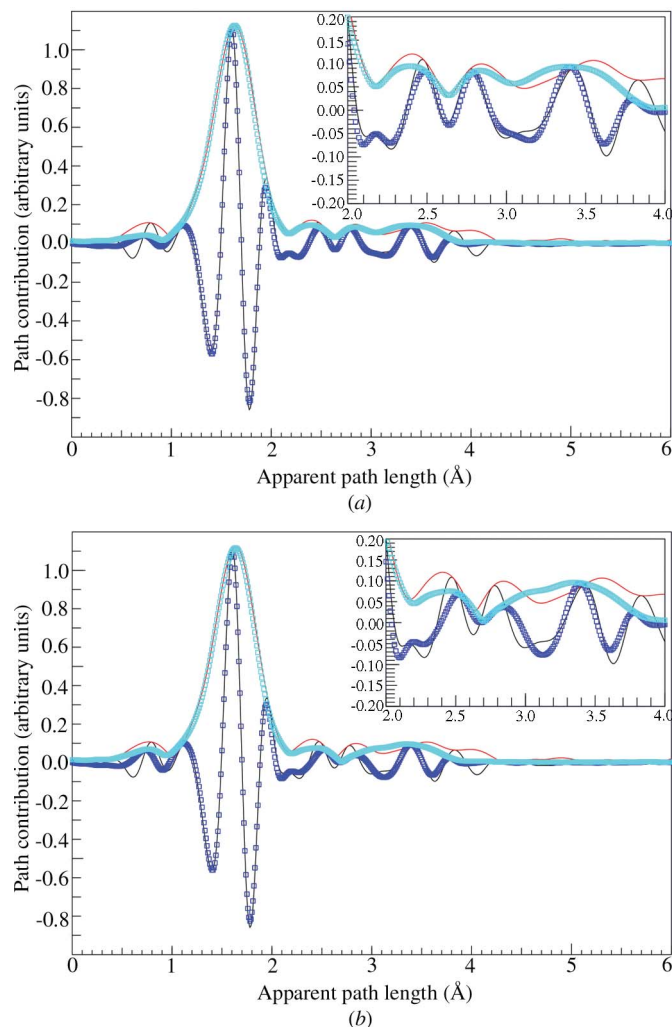| Path	R, linear	R, Monte Carlo	R, cryst.	σ^2, linear	σ^2, Monte Carlo
1: Pt N/O Pt	2.0258 ± 0.0007	2.028 ± 0.002	2.025	1.87 ± 0.05	2.00 ± 0.14
2: Pt C_{COOH} Pt	2.6507 ± 0.0072	2.97 ± 0.04	2.882	52.7 ± 4.73	18 ± 7
4: Pt N Pt O Pt	4.0321 ± 0.0025	4.065 ± 0.006	4.050	0.50 ± 0.41	3.1 ± 1.4

Figure 2
Model of the solid carboplatin EXAFS spectrum (Fourier transform: module and imaginary part). Continuous lines: experiment; squares: model. (*a*) Complete model; (*b*) three-shell model. An inset showing the region where the two models differ is given for each model.

The second comment concerns the values obtained for the third-path parameters and for $\sigma^2_{DW,2}$. Why do we obtain such high uncertainties? Analysis of the 1000 values obtained for these parameters, after the Monte Carlo simulation, shows that, in fact, about half of the fits give physically absurd values for these parameters, suggesting that the paths involved are negligible ($\sigma^2_{DW} = 0.5 \text{ Å}^2$). This result may arise from two factors; either the model is false or the importance of these

paths is low, and hence, because of the statistical noise, it is difficult to evidence them.

To select between these two explanations, we realized a model without the third scattering path. This model is acceptable but does not reproduce the 2.5 Å features of the Fourier transform as well as the first model (Fig. 2*b*, and especially inset); analysis of the Monte Carlo method results show the same problem of high uncertainties for the parameters of the remaining path (Table 5; for comparison, results for the linear model are given; the high uncertainty for $\sigma^2_{DW,2}$ suggests rejection of the model as unphysical). However, analysis of the Monte Carlo results for that problematic path evidences an interesting property (Fig. 3): the joint distribution for the couple (R_2, $\sigma^2_{DW,2}$) is bimodal. Furthermore, the two modes of this distribution correspond to the values obtained for parameters from paths 2 and 3 in the 'complete' model, close to the crystallographic values.

These results suggest an explanation: each single-scattering path on the C atoms gives a small contribution to the signal and, because only the distance is slightly modified between these two paths, their contributions are quite similar. Because of the added noise, when the Monte Carlo spectra are generated it becomes difficult to distinguish between the two contributions. Hence, while fitting the partial model, some fits give the first path contribution, while others give the second, and therefore we obtain the bimodal distribution. While fitting the complete model, sometimes both paths are recognized, but sometimes only one is and the other becomes negligible by the forcing of its Debye–Waller factor to unacceptably high values; this phenomenon leads to very high uncertainties.

These results hence allow us to conclude that both paths must be present in the model, and hence the complete model is valid; however, because of the statistical noise, it is not possible to obtain the correct values for the Debye–Waller factors. This result can be obtained from a Monte Carlo study
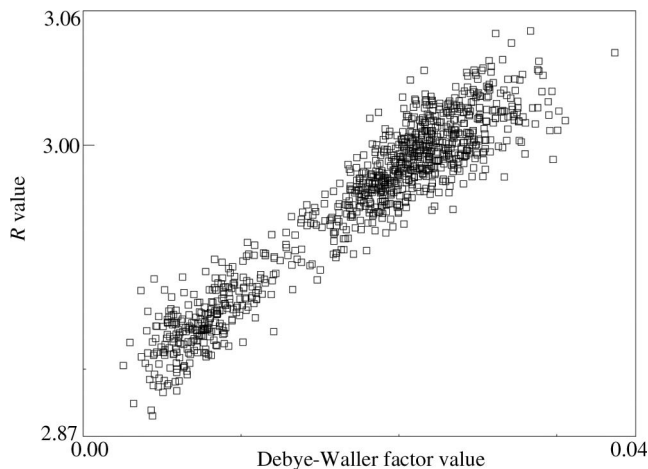

Figure 3
Joint distribution of the (R, σ^2) values for the Pt C_{COOH} Pt path. There are clearly two modes in this distribution.

alone, which shows the effect of noise on the fit; the linear model just signals that a problem exists.

8. Conclusion

In this study, we recalled the different limitations of three of the methods applied for the estimation of the error bars on the fitted parameters in EXAFS analysis. We also showed that the classical formula $\Sigma_\Theta = 2H^{-1}$ and the 'brute force' method rely on the same hypotheses. If one of these hypotheses is not satisfied, the use of the formula is meaningless. In particular, this formula is useless when fitting a Fourier transform or a filtered spectrum (points are not independent) or when using a k^p weighting scheme (it is not the maximum-likelihood estimator). However, the Monte Carlo method can give results that are statistically meaningful in all cases.

Our work indicates the best way to use this method; the most appropriate variant is the use of the experimental average and error bars to generate the spectra; for a typical analysis it is enough to generate around 100 spectra.

When maximum-likelihood least squares are used, one must address the question of the linear approximation validity. The results in this study suggest that, as long as the model is not too distant from reality, it is acceptable to apply the linear approximation, except maybe in the case of quite complex models, especially if one wants to derive true confidence intervals.

The last example also showed that the Monte Carlo method can provide more information than the linear method, when the results are carefully analysed, for the case of complex models. Hence, even when the linear approximation can be used, it may be interesting to perform a Monte Carlo analysis.

The main drawback of the method is the time it takes to obtain a result. For simple models, with recent increases in computer power, the process is quite rapid (for the Zn models it takes about half an hour on a 900 MHz computer for 1000 spectra); for complex models, it is much longer (about one night in the worst case we studied, on the same computer for 1000 spectra). However, this method may reveal interesting features and the linear model may sometimes reach its limit of validity.

Hence, our advise would be to perform a Monte Carlo analysis when possible, or whenever there is a doubt about the results of a linear model; the linear model should be reserved for quick estimation of error bars, without statistical meaning or in very simple cases, and only when the conditions mentioned here are met.

APPENDIX *A*
Derivation of the maximum-likelihood formula

The principle of the derivation of this formula is well known [see, for instance, Saporta (1990) or Press *et al.* (1993)]; however, it is classically derived on non-repeated measures. Therefore, we present here the derivation in the case of repeated measures (not always with the same number of

repetitions), in order to obtain the formula in the more general case. We will also emphasize at which derivation step each of the above-mentioned hypotheses occurs; these hypotheses will be written in bold face in this appendix.

Let us consider measures for N different values of the x variable, with subscript i. For each of these values, we have m_i different measures (y_i values), with subscript j. Most often, m_i is the same for each point (*i.e* does not depend on i), but some authors propose experimental protocols in which only a few experimental points are repeated, to estimate the experimental uncertainties (Filiponi, 1995). Each experimental measure is then doubly subscripted: $Y_{i,j}$. We want to maximize the probability of simultaneously observing all these values, that is the quantity V ('likelihood'), defined as

$$V \prod_{i=1}^{N} \prod_{j=1}^{m_i} \mathrm{d}y_{i,j} = p\left(\bigcap_{i=1}^{N} \bigcap_{j=1}^{m_i} Y_{i,j} \in \left[y_{i,j}, y_{i,j} + \mathrm{d}y_{i,j} \right] \right).$$

Since $0 < V < 1$ (it is a probability) and the logarithm is an increasing function, maximizing V is equivalent to maximizing $\ln V$, and hence equivalent to minimizing $-\ln V$; this last form is commonly used for convenience. Let us now assume that **experimental points are** (statistically) **independent**. One may then write

$$-\ln V - \sum_{i=1}^{N} \sum_{j=1}^{m_i} \ln \mathrm{d}y_{i,j} = -\sum_{i=1}^{N} \sum_{j=1}^{m_i} \ln p\left(Y_{i,j} \in \left[y_{i,j}, y_{i,j} + \mathrm{d}y_{i,j} \right] \right)$$

$$= -\sum_{i=1}^{N} \sum_{j=1}^{m_i} \ln p_{i,j} - \sum_{i=1}^{N} \sum_{j=1}^{m_i} \ln \mathrm{d}y_{i,j},$$

where $p_{i,j}$ is a measure of the probability that $Y_{i,j}$ takes the value $y_{i,j}$. To determine this probability, it is necessary to know the properties of the experimental vector **Y**. Let us assume a **normal distribution** with parameters $\mu_{g,i}$ and $\sigma_{g,i}$; then

$$p_{i,j} = \frac{1}{\sigma_{g,i}(2\pi)^{1/2}} \exp\left[-\frac{\left(y_{i,j} - \mu_{g,i} \right)^2}{2\sigma_{g,i}^2} \right].$$

Let us now assume that $\mu_{g,i} = f(x_i, \theta_1, \ldots, \theta_q)$; in other words, that **the model is correct**. One can then write the maximum-likelihood estimator as

$$-\ln V = (\ln 2\pi/2) \sum_{i=1}^{N} m_i + \sum_{i=1}^{N} m_i \ln \sigma_{g,i}$$

$$+ (1/2) \sum_{i=1}^{N} \sum_{j=1}^{m_i} \left\{ \left[y_{i,j} - f(x_i, \theta_1, \ldots, \theta_q) \right] / \sigma_{g,i} \right\}^2.$$

The first term is constant; it does not interfere with the minimization, and we will not recall it later. Hence, with the usual definition of the expectation and variance estimators (average value and mean-square deviation), the last term can be rewritten to evidence the average spectrum. We then obtain, calling V' the likelihood without the constant term,

$$-\ln V' = \sum_{i=1}^{N}\left[m_i \ln \sigma_{\mathrm{g},i} + (m_i - 1)s_i^{*2}/\sigma_{\mathrm{g},i}^2\right]$$
$$+ \sum_{i=1}^{N}\left[\frac{y_i - f(x_i, \theta_1, \ldots, \theta_q)}{\sigma_{\mathrm{g},i}/(m_i)^{1/2}}\right]^2.$$

The first term depends only on the variance of the normal law and on its experimental estimation. If we assume that **the standard deviation does not depend on the fitted parameters**, this term is constant and can be neglected, and we can consider only the last term.

APPENDIX B
Derivation of the covariance-matrix method formulas

Textbooks present the derivation of this well known formula only for the linear 'statistical' least-squares case [see, for instance, Press *et al.* (1993)] and do not state all the hypotheses. We will present here the derivation in the most general case of generalized least squares (since simple least squares, statistical least squares and maximum-likelihood least squares are special cases with, respectively, $W = I$, $W = S^{*-1}$ and $W = \Sigma^{-1}$); hypotheses will be in bold face when introduced.

If the model f is **linear for the parameters**, one can always write $f(x_i, \theta_1, \ldots, \theta_q) = \theta_1 f_1(x_i) + \cdots + \theta_q f_q(x_i)$, where the q functions $f_1, \ldots, f_q$ are the basis functions of the model. Hence, with vector notation, we have $\mathbf{Y}_{\mathrm{th}} = {}^{\mathrm{t}}F\,\theta$, where F is a $q \times N$ matrix that contains the values of the basis functions for each of the x_i points. Hence, the least-square estimators can be rewritten, with matrix notation, as a quadratic form,

$$\xi^2 = {}^{\mathrm{t}}(\mathbf{Y} - {}^{\mathrm{t}}F\Theta)W(\mathbf{Y} - {}^{\mathrm{t}}F\Theta).$$

If there is **no constraint**, the minimum of the quadratic form is obtained when its first derivative cancels, which occurs when the following equation is satisfied,

$$F\,W\,\mathbf{Y} = F\,W\,{}^{\mathrm{t}}F\,\Theta.$$

This system is linear; it has a unique solution. Noticing that the second derivative of this quadratic form is $H = 2F\,W\,{}^{\mathrm{t}}F$, this solution is

$$\Theta = 2H^{-1}\,F\,W\,\mathbf{Y} = g(\mathbf{Y}).$$

This expression gives the analytical definition of the function g, which can then be studied. Since g is linear, the relation between the covariance matrix Σ_Θ of Θ and the covariance matrix $\Sigma_{\mathbf{Y}}$ of $\mathbf{Y}$ is straightforward,

$$\Sigma_\Theta = 4(H^{-1}F\,W)\,\Sigma_{\mathbf{Y}}\,{}^{\mathrm{t}}(H^{-1}F\,W) = 4H^{-1}F\,W\,\Sigma_{\mathbf{Y}}\,{}^{\mathrm{t}}W\,{}^{\mathrm{t}}F\,{}^{\mathrm{t}}H^{-1}.$$

In general, it is not possible to simplify this formula. However, if **one uses maximum-likelihood least squares**, we have $W = \Sigma_{\mathbf{Y}}^{-1}$ and the previous expression then simplifies and gives the classical result, $\Sigma_\Theta = 2H^{-1}$.

If simple least squares are used ($W = I$), experimental points are independent and *the error is constant*, a second simple case arises: $\Sigma_{\mathbf{Y}} = \sigma^2 I$. After simplification, we then obtain $\Sigma_\Theta = 2\sigma^2 H^{-1}$.

In any of these cases, since g is linear and $\mathbf{Y}$ is Gaussian, the parameters vector Θ is itself Gaussian **as far as the $H^{-1}FW$ term is perfectly known**. This is not the case if W, for instance, is deduced from experiment, especially in the case $W = S^{*-1}$.

APPENDIX C
Obtaining isocontour and isodensity curves

We will consider only the linear case, with the notations and results introduced in Appendix B.

C1. Isocontour curves

Isocontour curves are defined by the equation $\xi^2 = C$, where C is an arbitrary constant; since we are interested in isocontours of ξ^2 around the minimum, $\theta_{\min}$, the quadratic form can be rewritten

$$\xi^2(\delta\theta) = {}^{\mathrm{t}}[\mathbf{Y} - {}^{\mathrm{t}}F(\theta_{\min} + \delta\theta)]\,W\,[\mathbf{Y} - {}^{\mathrm{t}}F(\theta_{\min} + \delta\theta)].$$

By developing this equation, and noting that ${}^{\mathrm{t}}[\mathbf{Y} - {}^{\mathrm{t}}F(\theta_{\min})]\,W\,[\mathbf{Y} - {}^{\mathrm{t}}F(\theta_{\min})]$ is constant and that all terms equal their transpose (they are numbers), the definition equation then becomes

$$C = \xi^2(\delta\theta) = {}^{\mathrm{t}}\delta\theta\,F\,W\,{}^{\mathrm{t}}F\,\delta\theta + 2\,{}^{\mathrm{t}}\theta_{\min}\,F\,W\,{}^{\mathrm{t}}F\,\delta\theta - 2\,{}^{\mathrm{t}}\mathbf{Y}\,W\,{}^{\mathrm{t}}F\,\delta\theta.$$

Now, if we replace $\theta_{\min}$ by its expression (see Appendix B), and note that W and $F\,W\,{}^{\mathrm{t}}F$ are symmetric, we can see that the last two terms are equal. Finally, the isocontour curves are defined by the equation

$$C = \xi^2(\delta\theta) = {}^{\mathrm{t}}\delta\theta\,F\,W\,{}^{\mathrm{t}}F\,\delta\theta.$$

C2. Isodensity curves

The parameters vector Θ is Gaussian, with expectation $\theta_{\min}$ and covariance matrix Σ_Θ. Since this matrix is non-singular, the distribution function (density of probability) of Θ is then

$$f_\Theta(\theta) = \left[(2\pi)^q|\Sigma_\Theta|\right]^{1/2}\exp\left(-{}^{\mathrm{t}}\delta\theta\Sigma_\Theta^{-1}\delta\theta/2\right).$$

Since the pre-exponential factor is constant, we obtain $f_\Theta(\theta) = C$, which defines the boundaries of confidence regions on $\theta_{\min}$, if the term in the exponential is constant. Hence

$$\,{}^{\mathrm{t}}\delta\theta\Sigma_\Theta^{-1}\delta\theta = C.$$

C3. Example

We will consider for illustration a two-parameter linear model: a two-shell model for the first peak of the cisplatin Pt L_{III}-edge spectrum, with only the number of N, N_1, and of Cl, N_2, atoms fitted *via* k^2-weighted least squares but with error bars that are constant with k; hence, we are out of the maximum-likelihood model. Since we are only interested in comparing the shapes of the regions obtained by the two formulas, the precise value of the constant C is of no interest. The values for the F matrix were chosen from *FEFF*-computed theoretical spectra for cisplatin; points were

assumed to be independent and to have a constant error of 0.01. Note that the shape of the region does not depend on the experimental data values (only its location).

Both ${}^t\delta\theta \Sigma_\Theta^{-1} \delta\theta = C$ and $C = \xi^2(\delta\theta) = {}^t\delta\theta\, F\, W\, {}^t F\, \delta\theta = {}^t\delta\theta H \delta\theta / 2$ will lead to ellipses in the (N_1, N_2) plane, differing in orientation. The orientation is defined by the eigenvectors of the matrix (which give the main axes of the ellipses).

We first give the two covariance matrices, obtained from the exact formula and from the inversion of the Hessian,

$$\Sigma_\Theta = \begin{pmatrix} 0.0475 & -0.0508 \\ -0.0508 & 0.0791 \end{pmatrix}$$

and

$$2\sigma^2 H^{-1} = \begin{pmatrix} 0.0017 & -0.0010 \\ -0.0010 & 0.0019 \end{pmatrix}.$$

As can be seen, the use of the classical formula when unadapted can strongly underestimate the uncertainties. Fig. 1 gives the ellipses obtained for each of these 'covariance' matrices; the different orientations of the main axes between the true confidence region and the region obtained by the brute-force method are clearly observed.

References

Bates, D. M. & Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*, ch. 6, pp. 200–231. New York: Wiley.

Curis, E. (2001). PhD thesis, Université Paris-Sud, France.

Curis, E. & Bénazeth, S. (2000). *J. Synchrotron Rad.* **7**, 262–266.

Curis, E. & Bénazeth, S. (2001). *J. Synchrotron Rad.* **8**, 264–266.

Curis, E., Provost, K., Nicolis, I., Bouvet, D., Bénazeth, S., Crauste-Manciet, S., Brion, F. & Brossard, D. (2001). *New J. Chem.* **24**, 1003–1008.

Dagnelie, P. (1973). *Théorie et Méthodes Statistiques*, Vol. 1. Gembloux, Belgium: Presses Agronomiques de Gembloux.

Draper, N. R. & Smith, H. (1981). *Applied Regression Analysis*, 2nd ed. New York: Wiley.

Ellis, P. J. (1995). Phd thesis, University of Sydney, Australia.

Filiponi, A. (1995). *J. Phys. Condens. Matter*, **7**, 9343–9356.

Ghigna, P., Di Muri, M. & Spinolo, G. (2001). *J. Appl. Cryst.* **34**, 325–329.

Gurman, S. J. & McGreevy, R. L. (1990). *J. Phys. Condens. Matter*, **2**, 9463–9473.

Klementev, K. V. (2001). *Nucl. Instrum. Methods Phys. Res. A*, **470**, 310–314.

Krappe, H. J. & Rossner, H. H. (1999). *J. Synchrotron Rad.* **6**, 302–303.

Kuzmin, A., Obst, S. & Purans, J. (1997). *J. Phys. Condens. Matter*, **9**, 10065–10078.

Lilliefors, H. W. (1967). *J. Am. Stat. Assoc.* **62**, 399–402.

Neuilly, M. (1993). *Modélisation et Estimation des Erreurs de Mesure*. Paris: Lavoisier.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1993). *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press.

Protassov, K. (1999). *Probabilités et Incertitudes dans l'Analyse des Données Expérimentales*. Presses Universitaires de Grenoble.

Rossner, H. & Krappe, H. (2001). *J. Synchrotron Rad.* **8**, 261–263.

Saporta, G. (1990). *Probabilités, Analyse des Données et Statistique*. Paris: Éditions Technip.

Standards & Criteria Commitee (2000). *Error Reporting Recommendations*. Report of the International XAFS Society Standards and Criteria Commitee. http://ixs.iit.edu/subcommittee_reports/sc/SC00report.pdf.

Vaarkamp, M. (1998). *Catal. Today*, **39**, 271–279.