The influence of experimental and model uncertainties on EXAFS results

Hermann Rossner[a] and Hans Krappe[a]

[a]Hahn-Meitner-Institut, Glienicker Str. 100, D-14109 Berlin, Germany.
Email:rossner@hmi.de

We analyze EXAFS oscillations in k-space with the FEFF code to obtain main-shell distances R_ν and mean-square displacement parameters σ^2_i for all single and multiple scattering paths i in the shells ν up to a maximum shell radius R_{max}. To quantify the uncertainty in the determination of these model parameters we take into account experimental errors and uncertainties connected with background subtraction, with the approximate handling of the electronic many-body problem in FEFF, and with the truncation of the multiple scattering series. The impact of these uncertainties on the R_ν and σ^2_i is investigated in the framework of Bayesian methods. We introduce an *a priori* guess of these model parameters and consider two alternative strategies to control the weight of the *a priori* input relative to that of the experimental data. We can take a model parameter space of up to 250 dimensions. Optionally we can also fit the coordination numbers N_j ($j \leq \nu$) and the skewness of the distribution of the R_ν besides the R_ν and σ^2_i. The method is applied to 10K Cu K-edge and 300K Au L_3-edge data to obtain model parameters and their *a posteriori* error correlation matrices.

Keywords: **error analysis, Bayesian approach, FEFF**

1. Introduction

Our method of analyzing extended X-ray absorption fine-structure (EXAFS) data is described in detail by Krappe & Rossner (2000) and thus will be only briefly sketched in this article. The well-known ill-posed nature of the inverse scattering problem requires the introduction of an *a priori* probability distribution for the model parameters to be obtained by the fit. Using Bayes' theorem, the *a priori* information is combined with the experimental information to obtain an *a posteriori* probability. It measures the degree to which the *a priori* assumption has to be modified on the basis of the data obtained by the measurement.

Following the work of Turchin & Nozik (1969) we consider two types of conditions on the weight of the *a priori* data relative to the experimental data in the fit. The first condition gives to the *a priori* data the largest weight compatible with the experimental data within a one-standard-deviation margin. With the *a priori* data to be considered below, we find that even the largest weight is compatible with the data, which means that the latter do not contain information that requires a modification of the *a priori* assumptions. The second condition associates the most probable weight to the *a priori* data. It is then possible to define the subspace $\mathfrak{R}$ of the whole model-parameter space where the data determine the fit, and the complimentary space where the fit is only controlled by the *a priori* assumptions. The error correlations between all fitted parameters are determined by the matrix elements of the regularized variance matrix.

We use a formula for X-ray absorption on a polycrystalline or amorphous sample which includes the third cumulants $C_{3\,i}$ of the peaks in the pair-distribution function (Stern, 1988)

$$\chi(k) = \frac{S_0^2}{k} \sum_{i=1}^{p} N_i \frac{\left| f_{eff}\left(k, R_i\right) \right|}{R_i^2} e^{-2k^2\sigma_i^2 - \frac{2R_i}{\lambda(k)}}$$
$$* \sin\left\{ 2k\left[R_i - c_i\right] + \phi_i\left(k\right) - \frac{4}{3} C_{3i} k^3 \right\}$$

with corrections $c_i = 2\sigma^2_i\,(R_i^{-1} + \lambda^{-1})$ to the mean shell radii. The lengths of multiple scattering paths are functions of the shell radii, the single scattering R_i. We therefore take as independent model parameters S_0^2, E_0, the shell radii R_i, all σ^2_i, and optionally the coordination numbers N_i and third cumulants $C_{3\,i}$ for the single scattering paths, and build the model-parameter vector **x**. The remaining quantities λ, $f_{eff}(R_i)$, $\phi(R_i)$ are functions of k and the model parameters **x**. They form a vector **y**. The systematic uncertainties in these quantities, the truncation error of the multiple scattering series, the experimental errors and the uncertainty of the *a priori* data will be modeled by Gaussians. Bayes' theorem then yields also a Gaussian for the *a posteriori* probability distribution of the components of **x**. The mean values $\langle \mathbf{x} \rangle$ follow from solving the normal equation $(Q+A)\langle\mathbf{x}\rangle = \mathbf{b}$ and the *a posteriori* variances are given by $(Q+A)^{-1}$, where Q is the information matrix, containing the derivatives of χ with respect to the model parameters, the experimental errors and the uncertainties of the model, A is the inverse variance matrix of the *a priori* data, and the vector **b** depends on the average values of the experimental data.

2. Data Analysis

2.1 10K Cu data

The K-edge copper data of Newville (1994) were analyzed within the k range 0.1-19.9 Å^{-1} in steps of 0.05 Å^{-1} (see Fig. 1). Considering the envelope of the Fourier filtered experimental data with an r-space window of 20-25 Å, an absolute statistical error of $\Delta\chi_{exp}$=0.001 has been assigned to these data. As the *a priori* model-parameter set we chose the 10 K fcc lattice constant a=3.6032 Å (Zabinski et al., 1995), the correlated Debye model with θ_{Debye}=315 K for the σ^2_i, the third cumulants $C_{3\,i}$ of all pair-distribution functions were assumed to vanish, the Fermi energy E_0 was set to 8979 eV, and the many-body amplitude reduction factor S_0^2=1. The FEFF7 code (Zabinski et al., 1995) is used with a filter of the curved wave-amplitude ratio of 4%, a cluster radius of R_{max}=8 Å, and a maximum leg number nleg=5 for multiple scattering paths. The FEFF code yields the coordination numbers N_i of the ideal fcc crystal, the electron mean free path $\lambda(k)$, the scattering amplitudes $f_i\,(R_i,k)$, and phases $\phi_i\,(R_i,k)$, for 10 single and 59 multiple scattering paths. The fit parameters that represent the components of the vector **x** were defined as S_0^2, E_0, the 10 single scattering half path lengths R_i, (i=1,...,10), all mean-square displacement parameters σ^2_i, (i=1,...,69), and the 10 single scattering third cumulants $C_{3\,i}$, (i=1,...,10). The coordination numbers N_i are not included in this fit, and the 59 multiple scattering half path lengths R_i, (i=11,...,69), were related to the single scattering values assuming an ideal fcc crystal structure. All third cumulants of the multiple scattering paths were kept

zero. Thus, the **x** vector had 91 components. The 139 components of the **y** vector were composed of $\lambda(k)$, $f_i(R_i,k)$, and $\phi_i(R_i,k)$, with $i=1,...,69$, and the model errors associated with them were assumed to be $\Delta\lambda/\lambda=10\%$, $\Delta f_i/f_i=7\%$, and $\Delta\phi_i=0.07$ rad. The truncation error, which is caused by the 4% threshold of the curved wave amplitude ratio was determined as described by Krappe & Rossner (2000), and is shown in Fig.2 as dash-dotted line. The resulting effective error, defined by the root mean square sum of the experimental error (long dashed line), the model errors, and the truncation error, is displayed by the solid line. The short dashed line represents an additional systematic error to account for deficiencies of the theory at small k ($k\leq3$ Å^{-1}) and for the truncation in r space at 8 Å. Its exponential tail approximates the envelope of the Fourier filtered $\chi_{exp}(k)$ with an r-space window of 8-25 Å.

Figure 2
k-dependence of the effective error (solid line), the experimental error (dashed line), the truncation error (dash-dotted line), and the many-body model uncertainty (dotted line) for Cu.
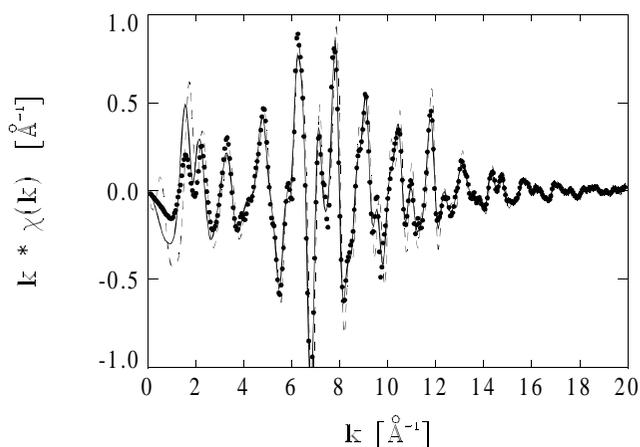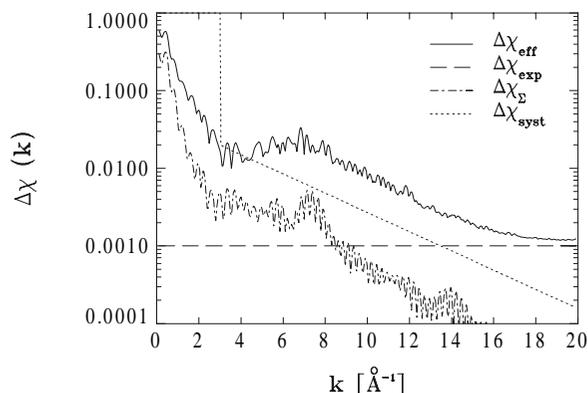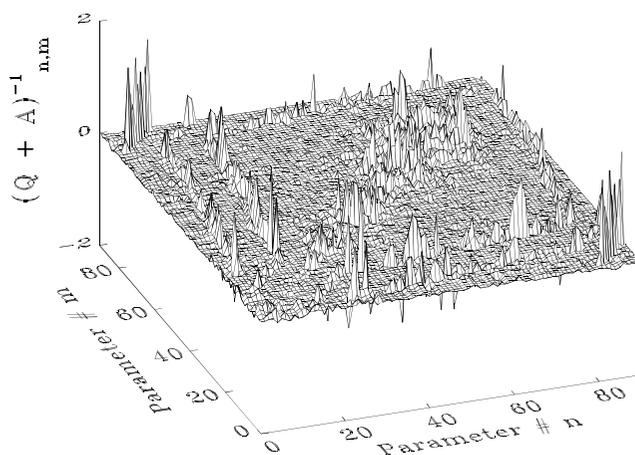

Figure 1
EXAFS oscillation of experimental data (dots), the *a priori* data (thin line), and the *a posteriori* data (thick line) for Cu.

Figure 3
Matrix elements of the regularized variance matrix for Cu; diagonal elements are suppressed.

Using the *a priori* parameter set and the uncertainty estimate described above, the second Turchin condition was applied, resulting in a $\Re$-space of dimension 30. This can be interpreted as an effective number of independent parameters $N_{eff}=30$, which is much smaller than the number of independent data points $N_d\approx(2/\pi)\Delta k\Delta R=58$ traditionally used in Fourier analysis. In a second analysis of the same input data as before we determined – in the spirit of the second Turchin condition – two independent regularization parameters for two sets of model parameters. The first parameter group corresponds to the set S_0^2, E_0, $R_1,...,R_{10}$, the second to the parameter set $\sigma^2_1,...,\sigma^2_{69}$, $C_{3\ 1},...,C_{3\ 10}$. The *a posteriori* variance matrix of **x** is shown in Fig. 3, where the diagonal elements are left out. The parameter sequence starting with n=0 is: S_0^2, E_0, $R_1,...,R_{10}$, $\sigma^2_1,...,\sigma^2_{69}$, $C_{3\ 1},...,C_{3\ 10}$. Close to the diagonal line strong error correlations between the mean-square displacement parameters σ^2_i and σ^2_j ($i\approx j$) of the multiple scattering paths are clearly seen. Also the correlations between the multiple scattering σ^2_i's and the single scattering R_j's and σ^2_j's ($i\approx j$) are represented by significant ridges. The strongest error correlations exist between the single scattering R_i's and the third cumulants. Correlations between R_i and σ^2_i for single scattering paths are small as shown by the weak ridges parallel to the diagonal line.
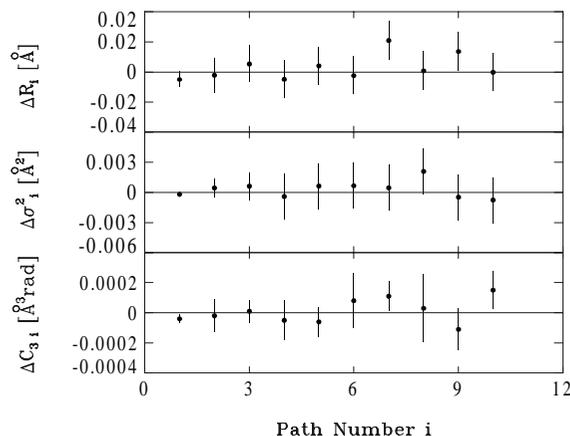

Figure 4
Deviations from their *a priori* values for the single scattering path values of R_i, σ^2_i, and $C_{3\ i}$ and the corresponding *a posteriori* errors for Cu.

The fit results for the single scattering path values of R_i, σ^2_i, and $C_{3\;i}$ are shown in Fig. 4 together with their *a posteriori* errors. They are plotted as deviations from the *a priori* values and indicate no significant deviations from the *a priori* estimate. We conclude that the experimental data are consistent with the information we had before the measurement

2.2   300K Au data

The same data-analysis procedure has been applied to the L_3-edge gold data of Newville (1994), measured at 300 K and analyzed in the k range 0.2-15.4 Å^{-1} in steps of 0.05 Å^{-1}. From the envelope of the Fourier-filtered experimental data with an r-space window of 20-25 Å we estimate an absolute statistical error of $\Delta\chi_{exp}$=0.0002. Again the FEFF7 code was used with a 4% amplitude threshold, R_{max}=8 Å, nleg=5, S_0^2=1, E_0=11918 eV, a(fcc)=4.07825 Å, θ_{Debye}=180 K, and all third cumulants set to zero. The **x** vector now has 45 components : S_0^2, E_0, seven single scattering and 22 multiple scattering paths with half path lengths R_i (i=1,...,7), mean-square displacement parameters σ^2_i (i=1,...,29), and third cumulants $C_{3\;i}$ (i=1,...,7). The lengths of the multiple scattering paths were again computed from the single scattering values. The same model uncertainties were assumed as in the previous case.

Application of the second Turchin condition gives an effective number of independent parameters N_{eff}=12, which, again, is much smaller than the number of independent data points $N_d \approx (2/\pi)\Delta k \Delta R$=40. The fitting procedure yields deviations from the *a priori* estimates for C_3, R and σ^2 values of the single scattering paths as shown in Fig. 5. Contrary to the 10 K Cu data a significant deviation of the $C_{3\;1}$ value is observed for the 300 K Au data. This nonzero third cumulant results in an improved description of the experimental data, leaving the fcc crystal parameter essentially unchanged, but modifying the Debye-Waller factor σ^2_1 slightly from 0.0087 Å^2 to 0.0077±0.0002 Å^2.

3.   Conclusion

We showed that the Bayesian approach combined with Turchin's conditions is a powerful tool in EXAFS data-analysis. We have demonstrated that in cases where either the *a priori* information is very good and/or the model uncertainties are large such that the first Turchin condition would not lead to a modification of the *a priori* parameters, new information about the crystal structure can be provided by the second Tuchin condition which puts a stronger weight on the measured data. Following the error propagation from the measured input data to the output of the data analysis in a systematic way, it is shown how the *a posteriori* mean values and variances depend on the probability distributions of the experimental data, the *a priori* information, and the model uncertainties. Since in our approach the normal equations are regularized by the *a priori* data we can handle models with a large parameter space, and applying Turchin's conditions those parameters are identified which are sensitive to the experimental data.

References

Krappe, H.J., Rossner, H.H. (2000). *Phys. Rev.* B **61**, 6596-6610.
Newville, M. (1994). Ph.D. thesis, University of Washington
Stern, E.A. (1988). In: D.C. Koningsberger and R. Prins, eds., *X-Ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*, page 3. Wiley Interscience Publication
Turchin, V.F., Nozik, V..Z. (1969). *Izv. Atmospheric and Oceanic Physics* **5**, 29-38.
Zabinski, S.I., Rehr, J.J., Ankudinov, A., Albers, R., & Eller, M.J. (1995). *Phys. Rev.* B **52**, 2995-3009
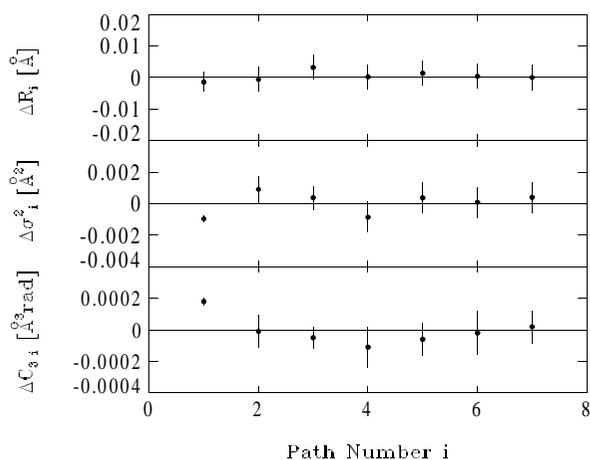
Figure 5
Deviations from their *a priori* values for the single scattering path values of R_i, σ^2_i, and $C_{3\;i}$ and the corresponding *a posteriori* errors for Au.