F-test in EXAFS fitting of structural models

A. Michalowicz*, K Provost**, S. Laruelle⁺, A. Mimouni* and G. Vlaic×

* Groupe de Physique des Milieux Denses (G.P.M.D.), Université Paris XII, 94010 Créteil Cedex, France and LURE. e-mail : michalov@univ-paris12.fr, fax : 33-1-45-17-15-79
** GPMD, and Laboratoire de Mineralogie et Cristallographie de Paris, Université Pierre et Marie Curie, Place Jussieu, 75252 Paris Cedex 05
+ Laboratoire de Reactivite et de Chimie des Solides, Universite de Picardie Jules Verne,
33, rue Saint Leu, 80039 Amiens, FRANCE
x Dipartimento di Scienze Chimiche, Universita degli Studi di Trieste, Trieste, Italia and Sincrotrone Trieste

The principles of the statistical F-test, used in EXAFS fitting in order to compare different models and eventually choose the more likely, are summarized. Examples, taken in the fields of inorganic chemistry and biochemical applications are discussed. F-test is shown to be very useful to check the statistical limits of a model and to discriminate between physically acceptable models. However, this tool is useless when the choice involves physically unrealistic solutions since it cannot prevent parameter variations out of their physically significant limits.

Keywords : EXAFS fitting, Statistical tests, best fit significance.

1. Introduction

One of the most efficient use of EXAFS modelling is the comparison of different models in order to choose the « best » one and discard, with convincing arguments, the worse one. Since choice by eyes may lead to mis-interpretations of the results, a quantitative method deserves to be proposed. The F-test has been proposed as a statistical test of EXAFS fitting for the first time by Joyner (Joyner, 1987). The main principles of this test is that, if statistical $\Delta\chi^2$ is the good function to be fitted, it is not the good number to use if one needs to compare and decide if two fits are significantly different and which one is the best. The test is based on the comparison of the reduced $\Delta\chi^2$, noted $\Delta\chi^2_v$.

2. The principles.

The test is based on the calculation of the statistical $\Delta\chi^2$

$$\Delta\chi^2 = (Nind/Npts) \sum(\chi_{th.}-\chi_{exp})^2/\epsilon^2 \quad (1)$$

where N_{ind} is the number of independant points in the fitted EXAFS spectrum , N_{pts} the total number of points (when raw data is used $N_{ind} = N_{pts}$), χ_{th} and χ_{exp} respectively the theoretical and the experimental EXAFS signals and ϵ the standard deviation. If the number of fitted parameters is N_{par}, $v = N_{ind} - N_{par}$ is the degree of freedom of the fit. The reduced $\Delta\chi^2$ is then $\Delta\chi^2_v$

$$\Delta\chi^2_v = \Delta\chi^2 / v. \quad (2)$$

The meaning and the way to estimate these quantities is discussed elsewhere (Lytle, S&C report, 1988 and present XAFS conference, 1998).
We want to compare two fits noted $\Delta\chi^2_1$ and $\Delta\chi^2_2$ with respectively the degrees of freedom v_1 and v_2. In order to avoid confusing notation we shall not use $\Delta\chi^2_v$ in the following formulas.

The general definition of the quantity F is :
$$F = (\Delta\chi^2_1 / v_1) / (\Delta\chi^2_2 / v_2) \quad (3).$$
In this notation fit 2 should be the best :
$$\Delta\chi^2_1 / v_1 > \Delta\chi^2_2 / v_2 \quad (4)$$
and thus F >1.

When the fit improvement does not involve any change in the number of fitted parameters one must use only this expression

In order to improve the fit, the number of fitted parameters may be increased : $Npar_1 < Npar_2$ and $v_1 > v_2$.
It is still possible to use equation (4) but it is preferable to test the variation of $\Delta\chi_v$, $(\Delta\chi^2_1-\Delta\chi^2_2)/(v_1- v_2)$, versus its final value $\Delta\chi^2_2/v_2$:

$$F = [(\Delta\chi^2_1-\Delta\chi^2_2)/ (v_1- v_2)] / (\Delta\chi^2_2/v_2) \quad (5).$$
The equation F > 1 is exactly identical to (4). Thus there is no difference between the two cases if one uses only a qualitative comparison of the two fits. However the probabilities calculated from the two sets of variables are different.

Let's call F_{exp} the experimental value of the variable F. The probability that F is greater that F_{exp} obeys to the so-called F probability law $P(F_{exp},v_1, v_2)$ (replace v_1 by v_1-v_2 in the second case). Some values of P for a set of F_{exp} and v_1, v_2 values are tabulated (Bevington, 1992). However it is very easy to calculate these values for any probability. The subroutines can be found in Numerical Recipies (Press, 1986). We have written a code adapted to the study of EXAFS fits (but usable for any fitting study). The code is available on the web of LURE (Orsay, France) :
http://www.lure.u-psud.fr/sections/chimie/xafsmac/xafsmac_main.htm
One should remark that in all the cited books, the tables give the probability P' that F does not exeed F_{exp}. In order to compare to the results given in this work, one should write P'=1-P.
How can these P values be used ?
P is the probability that $F > F_{exp}$. Since F is the ratio of the two reduced $\Delta\chi^2$, which is related to the estimation of the best fit, we can assimilate P to the probability that fit n°2 is better than fit n°1, or, in the second case, the probability that the $\Delta\chi^2$ improvement while increasing the number of parameters is significantly greater than its final value.

It is important to remark that all these formulas can be applied rigorously only when then added parameters are not correlated. In EXAFS it is not the case for the couples (R, ΔEo) or (N, σ). Even parameters belonging to different shells can be significantly correlated. Future developement of other tests, including the treatment of correlated parameters is planned and one should use the F-test as a simple first semi-quantitative approach to the statistical tests. Even if the calculated probabilities should not be used as rigourous numbers, their values can help to estimates the limits of a fitting improvement by adding new parameters.

Moreover, F-test is rigourously usable only if the errors are randomly and normally distributed. When EXAFS fits are dominated by systematic errors, statistical tests are unrelevant.

In this case $\Delta\chi^2_\nu$ is much greater than the statistical value (≈ 1)

The fit should be improved by a serious change in the structural model until $\Delta\chi^2_\nu$ reaches values close to 1 and the F test should be used only to choose between models obeying to these statistical laws.

3. Examples

3.1 Distorded metallic oxide octahedral site

The first example deals with the modelling of the first coordination sphere of metallic oxides, such as WO_3, studied, for example, for their electrochromic properties. The question is to check if this octaedron is regular or distorted. The strategy consists to try first a regular octahedron WO_6, and compare this fit to the one obtained by adding a second shell of oxygens.

While filtering the first EXAFS peak ($\Delta k \approx 10$ Å^{-1} $\Delta R \approx 1.5$) the number of independent points is reduced to $N_{ind}=2*\Delta k*\Delta R/\pi = 9$. Each shell needs 4 fitted parameters (N, σ, R, ΔEo). However fitting 8 parameters with $N_{ind}=9$ for the two shell fit is not very reasonable. It is possible to reduce N_{par} : The total number of neighbours $N_1 + N_2$ can be fixed to 6. It is possible to assume a common Debye-Waller value and it is reasonable to use identical Eo for signals involving the same kind of atoms in the first shell. Energy anisotropy may occur, but within reasonable limits (a few eV). In most cases, a single Eo value is sufficient to get a good fit, and it is a good rule in order to avoid unreasonable variations of this parameter. Within these constraints the number of parameters added for fit 2 is 1 (R_2). Finally, $v_1 = 9-4 = 5$, $v_2 = 4$ and $v_1 - v_2 = 1$. The fitting results are displayed in table 1.

	N	σ (Å)	R (Å)	ΔE_o (eV)	$v, \Delta\chi^2$
One shell	2.3(2)	.03(1)	1.76(1)	-6.4(3)	5, 7.2
shell1	3.9(2)	0.06(1)	1.79(1)	-2.9(1.8)	4, 3.9
shell2	2.1(2)	0.06(1)	2.07(2)	-2.9(1.8)	

Table 1 Comparison between one and two shells fits for WO_3

Both fits are close to the statistical limit. The use of F-test is justified, although the treatment of the parameters correlations may change slightly the quantitative evaluation of the probability.

$\Delta\chi^2_1 = 7.2$, $\Delta\chi^2_2 = 3.9$, $F_{exp} = 3.38$ and $P(F_{exp},5,1)= 86$ %. The two shell fit is found to be significantly better than the fit with only one shell. It is noticeable that this result is consistent with the well known tendency of WO_6 octahedra to undergo strong distortions. One may remark also that the number of oxygens found in fit 1 is too small, as expected for distorted sites wrongly modelled by a single gaussian distribution of distances. This example of the F-test use can be generalized each time a test of site distorsion is needed, for example in Jahn-Teller coordination complexes. If the distorsion is small, the result of the test may be negative.

In the previous fit 2 we have let N_1 vary freely, and we have found $N_1 \approx 4$, $N_2 \approx 2$. It is possible to use the F-test in order to check which is the more likely distribution of oxygens between the short (1.78Å) and the long one (2.06Å). We have compared three fits with the same total number of neighbours 6, and different repartitions : ($N_1 = 4$, $N_2=2$), ($N_1 = 3$, $N_2=3$) or ($N_1 = 5$, $N_2=1$). Thus the number of parameters for the three fits is identical : $v_1 = v_2 = v_3 = 5$. The corresponding $\Delta\chi^2$ are respectively equal to 4.1, 32.8 and 20.6, which is undoubtedly favorable to fit 1 with a probability of 95% compared to fit 3. Such an analysis has been used by one of us in order to find the best structural model in Ce/Zr oxydes studied for their catalytic

properties (Vlaic, 1997). In this case the best fit choice was less obvious and the help of F-test necessary.

3.2 Octahedral site in inorganic biological molecules

The second example is extracted from the study of the cobalt site of a coordination complex involved in biochemistry : Vitamin B_{12}. The local structure of this molecule is composed once more of a non regular octahedron, CoN_5C. Four nitrogens are located in planar macrocyclic ligand and the two resting ligands are axially bound. The question is to check if the Fourier filtered first peak contains a sufficient amount of information (statistically speaking) to be able to fit the two axial distances accurately.

The statistical study is performed in four steps and the fitting results displayed in table 2 :

1) fit the shell with 6 nitrogen ligands (N fixed).

2) compare fit 1 with a fit where the number of nitrogen is free. $F_{exp} = 39$, $P(F_{exp},9,1) = 99.9\%$.

3) compare fit 2 with a two shell fit, one for the equatorial ligands (4N) and one for the axial ligands considered as identical (2N). $F_{exp} = 3.75$, $P(F_{exp},8,2) = 91\%$.

4) compare fit 3 with a 3 shells fit, including the two axial ligands free to vary independently. $F_{exp} = 0.42$, $P(F_{exp},6,2) = 31\%$. Fit 4 is worse than fit 3.

This study leads to a series of remarks:

i) Fit 1 is obviously out of the statistical limit. The use of a statistical test in order to compare this fit with the others is useless. On the contrary, fits 2, 3 and 4 are close to the statistical convergence.

i) In the litterature, step 2 is often missing (Sagit, 1992). Obviously, fit 3 is better than fit 1. But this is not a good proof that 4+2 ligands is better than 6, if step 2 is not included in the discussion. In some cases, octahedral sites can be fitted with only 4 ligands. Such a result is due to a poor data resolution and to the statistics. In such cases missing step 2 should be considered as an over-interpretation of the data.

ii) step 3 seems to improve the fit, but the results obtained for the axial ligands are physically questionable : what is the meaning of a shell of 2 nitrogens with a Debye-Waller coefficient greater than 0.14 Å ? Probably an actual Co-Ligand signal, but distorted by a Fourier filtering truncation effect.

iii) adding a third shell should be definitely discarded. Although $\Delta\chi^2$ is improved, the price paid for this improvement is too high : 3 more parameters.

	N	σ (Å)	R (Å)	ΔE_o (eV)	$v, \Delta\chi^2$
One shell	6 (fixed)	.080(5)	1.900(5)	-1.2(8)	9, 40.0
One shell	3.5(2)	.05(1)	1.89(1)	-2(1)	8, 6.8
shell1	3.75(40)	0.05(1)	1.89(1)	-2.5(1)	
shell2	2.25(40)	0.14(3)	2.31(6)	-2.5(1)	6, 3.02
shell1	4.0(4)	0.05(1)	1.89(1)	-2(1)	
shell2	1.0(4)	0.07(1)	2.31(6)	-2(1)	
shell 3	1.25(40)	0.06(4)	2.22(3)	-2(1)	3, 2.5

Table 2. Fits for a vitamin B_{12} derivative : methyl-Cobalamin

The conclusion of this study is that one should take care of the statistical meaning of their fits. Trying to get more informations than given by the data makes no sense, even if the results seem apparently correct.

3.3 Mixing of two different kinds of atoms in the same shell.

The last example illustrates the danger of using any comparative tool, eyes, F-test, or any other method when one of the fits is obviously physically inacceptable. We have studied the case of a very simple and well known coordination complex ion : $[Co(NH_3)_5Cl]^{2+}$. The crystal structure gives $R_1(Co-N) = 1.97$ Å and R_2 (Co-Cl) =2.27 Å (Messmer, 1968).

We have tried two different fits : fit 1 with free N_1 and N_2, and fit 2 with fixed values of N_1 and N_2 to the known ones (5 and 1). The fitting results are given in table 3.

	N	σ (Å)	R (Å)	ΔE_0 (eV)	$\nu, \Delta\chi^2$
N_1(Cl) and	4.5(1.8)	0.15(2)	2.20(2)	-7(1)	
N_2(N) free	2.5(1.8)	0.05(2)	1.95(6)	-3(2)	**4, 1.4**
N_1(Cl) and	1	0.08(1)	2.29(1)	-2(1)	
N_2 (N) fixed	5	0.07(3)	1.97(3)	-2.1(1)	**6, 5.0**

Table 3 Fitting results for the two shells of $[Co(NH_3)_5Cl]^{2+}$

Obviously fit 1 is the best : the F-test gives him a probability of 95%. It is the wrong one because letting the number of nitrogen to vary, we have permitted the fit to fall in the so-called « false minimum trap » (Michalowicz, 1998), particularly dangerous in this case because of a mixing of two shells (N and Cl) having opposite phases. This solution provides false number of neighbours, false Debye-Waller factors, false distances and doubtful ΔEo for a model based on FEFF phase shifts. Although it is quite easy to avoid this trap (Michalowicz, 1998), it is clear that in such a case, the F-test is totally powerless. This example illustrates the fact that « goodness of fit », even evaluated by quantitative tests, is not an absolute criteria that a model is acceptable. The only valuable limits in EXAFS parameters variations should be physical and chemical limitations. Statistical tests are unable to decide if a « good fit » lies within these limits or not.

4. Conclusion

We have presented some examples illustrating how the F-test can be used in EXAFS modelling when a choice between different physically acceptable models is needed. The test can help in finding which model is the most probable. It can also help to find the limits in the amount of statistical information available, in order to avoid over-interpretation of data. However, just comparing the values of the reduced $\Delta\chi^2$ is sufficient to give qualitatively the right answer. The additional quantitative information given by the complete F-test is the probability of the best model. Since the calculation is very easy to perform, it deserves to be included in fitting codes.

References

Bevington, Data reduction and Error Analysisfor the Physical Sciences, (1992), McGraw Hill inc., p261-266
Joyner R.W., Martin K. J., Meehan P.J., (1987). J. Phys. C **20**, 4005–4012

Lytle F. W., Sayers D. E., and Stern E. A., (1988), "Report of the International Workshop on Standards and Criteria in X-ray Absorption Spectroscopy," *Physica B*, vol. 158, pp. 701-722
Report of the International Workshop on Standards and Criteria in X-ray Absorption Spectroscopy, 1998, Journal of Synchrotron Radiation, XAFS X, Chicago

Messmer G.G., Amma E.R., (1968), Acta Cryst., B24, 417-421

Michalowicz A., Vlaic G., (1998) J. Synchr. Rad., in press, vol 5.

Press W. H., Flannery B. P.,. Teukolsky S. A, and Vetterling W. T., (1986),*Numerical Recipes*. Cambridge: Cambridge University Press,.p 169 of the Fortran version.

Sagit I., Chance M.R., (1992) J. Am. Chem .Soc., 114, 8061

Vlaic G., Fornasiero P., Geremia S., Kaspar J., Graziani M. (1997), (1997), J. Catalysis 168, 386-392