Performance and use of Mössbauer goodness-of-fit parameters: Response to spectra of varying signal/noise ratio and possible misinterpretations

GLENN A. WAYCHUNAS
Center for Materials Research, Stanford University, Stanford, California 94305, U.S.A.

ABSTRACT
Simulated Mössbauer spectra of Lorentzian form have been fit with proper and improper models over a wide range of signal/noise (S/N) ratios in order to evaluate the characteristics of the goodness-of-fit parameters $\chi^2$ and MIFIT as a function of sample concentration and run duration. Results demonstrate a marked difference in behavior. $\chi^2$ may be reduced, suggesting an improved fit, if an incorrectly fit spectrum is degraded in quality. This is not the case with MIFIT. However, MIFIT may be defined in two slightly differing ways, each of which “diverges,” to large negative and positive values, respectively, at small S/N ratios. Failure to allow for these characteristics can lead to fundamental misinterpretations of an effect like “spectrometer drift” or the estimation of optimal data collection conditions. The application of $\chi^2$, MIFIT, and $\Delta$ MIFIT parameters to fit evaluations in mineralogical spectroscopy is reviewed in light of the differing parameter responses.

INTRODUCTION
Recently Dyar (1984) has presented a comparison of Mössbauer spectroscopy facilities in an attempt to establish optimal run parameters and system-system disparities. Unfortunately, she failed to analyze the effects of varied signal/noise (S/N) ratio independently on the performance of the goodness-of-fit parameters themselves. As will be shown, changes in S/N can have large effects on the value of $\chi^2$ or MIFIT for identical fit models. Hence her results in several cases require re-evaluation.
Dyar also neglected to utilize the $\Delta$ MIFIT parameter in the evaluation of her spectral fits. As has already been shown elegantly by Ruby (1973), and which will be briefly demonstrated in this paper, the great utility of the MIFIT parameter resides in comparison with its uncertainty, i.e., $\Delta$ MIFIT.
In the present work the goodness-of-fit parameters are defined and tested by application to the fitting of simulated Lorentzian Mössbauer spectra with correct and incorrect models.

GOODNESS-OF-FIT PARAMETERS
One of the crucial problems in evaluating any type of fit to spectral data is a proper answer to the question “How well is the model confirmed by the experimental data points?” This question follows from direct application of the scientific method, whereas the question “What model when applied to the data gives the best fit by some criterion?” is more generally the question that is answered.
The first question allows for testing of a (presumably) proper model, whereas the second simply looks for the best-fit model without testing its significance.
The general goodness-of-fit parameter that is calculated, $\chi^2$, has difficulties when used with Mössbauer spectra since it is affected by the magnitude of the background count and, as will be demonstrated, can actually be reduced by degrading the quality of a spectrum to which an incorrect model is being fit. Hence $\chi^2$ is not a good parameter for comparison of spectra collected at differing laboratories with differing experimental arrangements. Furthermore, as has been discussed by Law (1973), evaluation of $\chi^2$ alone is insufficient to determine the appropriateness of the fitting model. The testing should always be done with independent confirming evidence.
To improve the situation, Ruby (1973) suggested a new goodness-of-fit parameter which he called “MIFIT.” This parameter appears to be much more insensitive to spectral quality than $\chi^2$. Moreover, comparison between the calculated MIFIT value and the uncertainty in the MIFIT ($\Delta$ MIFIT) can yield a clear answer to the first question raised above.
$\chi^2$ is defined as

$$\chi^2 = \sum_{i=1}^{N} \frac{(Y_c(I) - Y_d(I))^2}{Y_d(I)}, \quad (1)$$

where $N$ is the number of data points, and the $Y_c(I)$ and the $Y_d(I)$ are the calculated and observed spectral values, respectively.
The reduced $\chi^2$ ($\chi^2_r$) is equal to $\chi^2$ divided by the number of degrees of freedom of the least-squares fit, i.e., the number of data points minus the number of varied parameters, $N - n$:

$$\chi^2_r = \chi^2 / (N - n). \quad (2)$$
Table 1. Results of least-squares fits to simulated Mössbauer spectra

<table>
<thead>
<tr>
<th>Spectrum background count</th>
<th>Absorption</th>
<th>Type of fit model</th>
<th>(M) (%)</th>
<th>(\Delta M) (%)</th>
<th>(M') (%)</th>
<th>(\Delta M') (%)</th>
<th>(\chi^2)</th>
<th>S/N**</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 000</td>
<td>3.3%</td>
<td>100% L</td>
<td>-0.977</td>
<td>-1.031</td>
<td>13.622</td>
<td>16.074</td>
<td>1.00</td>
<td>7.38</td>
</tr>
<tr>
<td>200 000</td>
<td>3.3%</td>
<td>100% L</td>
<td>-0.246</td>
<td>-0.259</td>
<td>3.848</td>
<td>1.701</td>
<td>1.00</td>
<td>14.76</td>
</tr>
<tr>
<td>500 000</td>
<td>3.3%</td>
<td>100% L</td>
<td>-0.099</td>
<td>-0.104</td>
<td>1.583</td>
<td>0.436</td>
<td>1.00</td>
<td>23.34</td>
</tr>
<tr>
<td>1 000 000</td>
<td>3.3%</td>
<td>100% L</td>
<td>-0.049</td>
<td>-0.052</td>
<td>0.800</td>
<td>0.175</td>
<td>1.00</td>
<td>33.00</td>
</tr>
<tr>
<td>4 000 000</td>
<td>3.3%</td>
<td>100% L</td>
<td>-0.012</td>
<td>-0.013</td>
<td>0.202</td>
<td>0.036</td>
<td>1.00</td>
<td>66.00</td>
</tr>
<tr>
<td>10 000 000</td>
<td>3.3%</td>
<td>100% L</td>
<td>-0.005</td>
<td>-0.005</td>
<td>0.081</td>
<td>0.014</td>
<td>1.00</td>
<td>104.36</td>
</tr>
<tr>
<td>50 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>-0.768</td>
<td>-1.085</td>
<td>14.028</td>
<td>16.774</td>
<td>1.01</td>
<td>7.38</td>
</tr>
<tr>
<td>200 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>-0.110</td>
<td>-0.286</td>
<td>4.052</td>
<td>5.769</td>
<td>1.03</td>
<td>14.76</td>
</tr>
<tr>
<td>500 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>0.014</td>
<td>0.123</td>
<td>1.726</td>
<td>0.472</td>
<td>1.07</td>
<td>23.34</td>
</tr>
<tr>
<td>1 000 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>0.052</td>
<td>0.068</td>
<td>0.917</td>
<td>0.194</td>
<td>1.14</td>
<td>33.00</td>
</tr>
<tr>
<td>4 000 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>0.075</td>
<td>0.024</td>
<td>0.294</td>
<td>0.044</td>
<td>1.45</td>
<td>66.00</td>
</tr>
<tr>
<td>10 000 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>0.078</td>
<td>0.013</td>
<td>0.166</td>
<td>0.020</td>
<td>2.02</td>
<td>104.36</td>
</tr>
<tr>
<td>50 000</td>
<td>0.33%</td>
<td>100% L</td>
<td>-11.343</td>
<td>-10.715</td>
<td>59.055</td>
<td>273.51</td>
<td>1.00</td>
<td>2.33</td>
</tr>
<tr>
<td>50 000</td>
<td>1.0%</td>
<td>100% L</td>
<td>-1.135</td>
<td>-1.163</td>
<td>14.946</td>
<td>19.059</td>
<td>1.00</td>
<td>7.00</td>
</tr>
<tr>
<td>50 000</td>
<td>10%</td>
<td>100% L</td>
<td>-0.010</td>
<td>-0.011</td>
<td>1.177</td>
<td>0.032</td>
<td>1.00</td>
<td>70.00</td>
</tr>
<tr>
<td>50 000</td>
<td>0.33%</td>
<td>85% L 15% G</td>
<td>-10.463</td>
<td>-11.079</td>
<td>59.928</td>
<td>279.66</td>
<td>1.00</td>
<td>2.33</td>
</tr>
<tr>
<td>50 000</td>
<td>1.0%</td>
<td>85% L 15% G</td>
<td>-0.915</td>
<td>-1.222</td>
<td>15.376</td>
<td>19.686</td>
<td>1.01</td>
<td>7.00</td>
</tr>
<tr>
<td>50 000</td>
<td>3.3%</td>
<td>85% L 15% G</td>
<td>0.014</td>
<td>0.123</td>
<td>1.726</td>
<td>0.472</td>
<td>1.07</td>
<td>23.34</td>
</tr>
<tr>
<td>50 000</td>
<td>10%</td>
<td>85% L 15% G</td>
<td>0.074</td>
<td>0.021</td>
<td>0.267</td>
<td>0.039</td>
<td>1.47</td>
<td>70.00</td>
</tr>
</tbody>
</table>

* L = Lorentzian, G = Gaussian.
** S/N = signal/noise ratio—see text for definition.

The MISFIT parameter is defined in terms of two variations on \( \chi^2 \). The distance \( D \) is defined as

\[
D = N(x^2 - 1), \tag{3}
\]

which, on substituting for \( x^2 \), becomes

\[
D = \left( \frac{N}{N - n} \sum_{i=1}^{n} \left[ \frac{Y_c(I) - Y_d(I)}{Y_d(I)} \right]^2 \right) - N. \tag{4}
\]

The distance is the discrepancy or distance of the fit from the observed data.

Ruby (1973) neglected the size of \( n \) in formulating \( \chi^2 \) so that \( D \) is simply

\[
D = \sum_{i=1}^{n} \left[ \frac{Y_c(I) - Y_d(I)}{Y_d(I)} \right]^2 - 1. \tag{5}
\]

The second variant on \( \chi^2 \) is the signal \( S \), defined as the summation of the spectral data above (or below) some baseline. The definition of the baseline, \( Y_0 \), is nontrivial, but neglecting this for the moment, the definition of \( S \) is

\[
S = \sum_{i=1}^{n} \left[ \frac{Y_c(I) - Y_d(I)}{Y_d(I)} \right]^2 - 1. \tag{6}
\]

MISFIT is the ratio of \( D/S \) from Equations 5 and 6.

The uncertainty in \( M \) is given by (Bevington, 1969):

\[
\sigma_M = \left( \frac{\sigma_{D^2}}{D^2} + \frac{\sigma_{S^2}}{S^2} - 2 \frac{\sigma_{D} \sigma_{S}}{DS} \right)^{1/2} M, \tag{7}
\]

which is approximated by

\[
\sigma_M \approx \left( \frac{\sigma_{D^2}}{D^2} + \frac{\sigma_{S^2}}{S^2} \right)^{1/2} M = \left( \frac{\Delta D}{D} \right)^2 + \left( \frac{\Delta S}{S} \right)^2 \tag{8}
\]

Replacement of the \( \Delta D \) and \( \Delta S \) values yields (Ruby, 1973)

\[
\Delta M = \left( \frac{1}{S} \right) \left[ N(1 + M^2) + 4D(1 + M) \right]^{1/2}. \tag{9}
\]

Another version of MISFIT that Ruby (1973, Eq. 7a) used for several examples is derived for cases of small absorption.

This formulation cannot produce negative values, is simpler to calculate, and has beneficial properties similar to those of the first MISFIT parameter. It can be written in the form

\[
M' = \left( \sum_{i=1}^{n} \frac{[Y_c(I) - Y_d(I)]^2}{Y_d(I)} \right)^{-1}. \tag{10}
\]

In order to display the differing characters of these goodness-of-fit parameters, synthetic Mössbauer spectra were generated with pure Lorentzian form and were fit with the known (generating) peak-shape functions and with incorrect functions. The spectra generation was carried out over a range of S/N that suitably mimicked the effect of varied run duration and sample concentration for typical Mössbauer experiments.

**Spectrum synthesis**

The synthetic Mössbauer spectra were generated by the Mössbauer fitting program MOSFIT (Dollase, 1971, unpub.). The original set of data points comprise smooth Lorentzian curves with fixed background count and absorption fraction. The effect of varied S/N is added by scaling of the spectrum absorption, appropriate offsetting of the background count, and addition of random statistical noise.
consistent with the magnitude of each resulting datum point according to the relation

\[ A_{\text{vel}} + (A_{\text{vel}} R_{\text{norm}}) = A'_{\text{vel}} \]

where

\[ A_{\text{vel}} = (F_{\text{vel}} C_{\text{off}}) - C_{\text{off}}. \]

Here \( F_{\text{vel}} \) is the Lorentzian Mössbauer spectral line profile as generated, \( C_{\text{off}} \) is a scaling factor to produce the desired absorption effect, \( C_{\text{off}} \) is a factor that offsets the background to be consistent with the absorption effect, \( A_{\text{vel}} \) is the scaled simulated Mössbauer spectrum, \( R_{\text{norm}} \) is a normally distributed random number with mean of 0 and standard deviation of 1, and \( A'_{\text{vel}} \) is the simulated Mössbauer spectrum with random statistical noise.

The use of a sufficiently random pseudorandom number generator was necessary to produce \( \chi^2 \) values close to 1.00. The generator was tested rigorously to assure number sequences that closely approximated a normal distribution.

Synthetic spectra consisted of two equal Lorentzian lines summed to produce a doublet spectrum. Two sets of synthetic spectra were generated, one with varying background count and fixed sample absorption simulating the effect of run duration, and the second with fixed background count and varying sample absorption simulating the effect of sample concentration.

All spectra parameters are listed in Table 1.

**Spectrum fitting**

Two types of fitting were performed. In the first type, all spectra were fit with 100% Lorentzian line shapes emulating the conditions of generation. Cycling during the refinement process was continued until there were at least two cycles without parameter change. Seven parameters were varied: position, area and width of each peak, and the background count.

In the second type, the spectra were fit with lines of 85% Lorentzian–15% Gaussian character, representing an incorrect model. The lineshape was produced by a sum of Lorentzian and Gaussian functions often referred to as a pseudo-Voigt profile (Hecq, 1981). As in the first type, the same seven parameters were varied, and the same requirement on fit convergence was used to determine the number of refinement cycles.

All of the goodness-of-fit parameters were calculated for each refined fit, as well as the standard deviations of these and all varied parameters.

**RESULTS**

The goodness-of-fit parameters resulting from the two types of fits to the first set of synthetic spectra are plotted in Figures 1 and 2 as a function of background count and the reciprocal of the background count, respectively.

Considering the Lorentzian-to-Lorentzian fit first, the value of \( \chi^2 \) is found to be constant at 1.00 as should be expected for a random background and perfectly fit data. The \( \text{MISFIT} \) parameter \( M \) is found to increase from a negative value of \(-0.977\%\) at the 50 000-count background level to \(-0.0049\%\) at 10 000 000 background counts. \( M' \), in contrast, decreases from \(3.62\) to \(0.14\%\) over the same range. Extrapolation to infinite background counts, in Figure 2, indicates that both MISFIT parameters would be zero in this limit.

In the case of the 85% Lorentzian–15% Gaussian fit, the \( \chi^2 \) values increase with improvement in the S/N ratio at greater background counts. The \( \text{MISFIT} \) parameter \( M \) increases with increasing background count, but rises to finite positive values instead of extrapolating to zero at infinite background count. \( M' \) decreases with increasing background count and extrapolates to the same finite positive value as \( M \) with infinite background count.

The goodness-of-fit parameters resulting from the two types of fits to the second set of synthetic spectra are plotted in Figure 3 as a function of percent absorption at fixed (500 000) background count. The characteristic behavior of each of the parameters is similar to that seen in Figure 1, except that the larger range of S/N explored produces wider parameter variation. For comparison, the complete set of results is plotted in Figure 4 as a function of S/N. Note that for this plot the signal is defined as the number of counts below background for the maximum spectral absorption, which is the usual spectral definition of signal, and the noise is the magnitude of the square root of the number of counts.
Fig. 3. misfit and $x^2$ versus Mössbauer effect absorption for fits to simulated Mössbauer spectra. Symbols as in Fig. 1.

The importance of having a goodness-of-fit parameter other than $\chi^2$ is evidenced by the behavior of $\chi^2$ when an incorrect fitting model is used. In this case, $\chi^2$ will continuously increase the longer a spectrum is collected. Hence, a poor fitting model may appear better than another superior model if the data are sufficiently poor. Since fitting models for mineralogical spectra are almost always at least as incorrect as those used in the present simulation because of the usual neglect of sample site variations, strain, blackness effects, cosine error, and so on, the trend in $\chi^2$ seen for the synthetic spectra is generally applicable. Thus $\chi^2$ values must be used cautiously for model testing.

On the other hand, if several runs of varying duration can be performed, the rate of change in $\chi^2$ with background count affords some estimate of model validity.

The **misfit** parameters are less sensitive to S/N effects over the usual range of Mössbauer experiments than $\chi^2$, but are markedly sensitive at low S/N ratios. In particular, the negative values produced are artifacts of the formulation of **misfit** and do not represent improved fits. However, the extrapolation of **misfit** values to infinite signal/noise (or background counts) produces an absolute measure of fit quality.

The negative values can lead to misinterpretations. Dyar (1984) collected spectral data as a function of run duration which display an identical trend in $M$ as the incorrect fit model presented in this work. Dyar attributed the effects of increasing $M$ with run duration to spectrometer drift, but the full magnitude of the **misfit** change she observed is consistent with the dependence of $M$ on S/N alone. Thus variation in $M$ cannot be used to test for spectrometer drift if effects on spectral quality are small.

Dyar (1984) also concluded that the optimum run duration for Mössbauer spectra occurred for $M$ near zero. This conclusion is based on the false assumption that $M$ is approximately zero for well-fit high-quality data. It is, in fact, true only for perfect fits in the limit of infinite S/N, both of which are not achieved in practice.

What is true is that spectrum quality must constantly improve, though at a slower and slower rate, with run duration. If spectrometer drift occurs, this is ordinarily manifested by alteration of spectral line shapes. Line-shape changes can be specified by fitting procedures or by Fourier deconvolution methods.

Another methodology consideration treated by Dyar (1984) is that of the optimum sample concentration. This was evaluated using the zero point in plots of **misfit** versus sample concentration for fixed run durations. For the reasons already outlined, this method should not be used to determine optimum experimental conditions. The value obtained by Dyar (1984) for a silicate sample is fortuitously close to theoretical estimates of ideal sample concentration, whereas her value for a sulfide sample is much too small.

**USE OF $\Delta M$ AND THE $Y_0$ PROBLEM**

Although the foregoing discussion indicates anomalies in the character of $M$ and $M'$, these are not severe enough to obviate their virtues. The real utility in using **misfit** lies in the comparison of **misfit** and $\Delta$ **misfit**. Since $\Delta$ **misfit** indicates the uncertainty in $M$ (and this is due mainly to the quality of the data) while $M$ is generally much less sensitive to data quality, it reveals the degree to which the fit model is tested. For example, in the 85% Lorentzian–15% Gaussian fits to the synthetic spectra the value of $M$ and $\Delta M$ for the 500 000-count 3.3% absorption spectrum fit are 0.014 and 0.123%, respectively. An $M$ of 0.014% represents an excellent fit, but a $\Delta M$ of 0.123% indicates that the noise in the data would permit of $M$ values up to 0.37% larger (3 standard deviations). Hence the model is not severely tested, and the low $M$ value is, by itself, deceptive. In the case of the analogous 10 000 000-
count spectrum, $M$ and $\Delta M$ are 0.078% and 0.013%, respectively. The $M$ value is now “on target” and indicates a good fit with a small but definite flaw. The $\Delta M$ value indicates high-quality data and, since it is considerably smaller than $M$, good model testing.

This leads to two requirements for testing of any fitting models: (1) minimization of $\chi^2$ and misfit and (2) $\Delta M$ much smaller than $M$.

A model that reduces $\chi^2$ or $M$ by 50% over some other model (on the same data) is only really superior to the other if $\Delta M$ is significantly smaller than 50% of $M$. Ruby (1973) provided several examples of the use of $M$ and $\Delta M$ in deciphering fit and data quality.

A problem arises in calculating misfit accurately due to the difficulty in obtaining proper $Y_0$ values. $Y_0$ is known exactly for synthetic spectra, but if refined in a least-squares procedure on experimental data, the $Y_0$ value obtained is dependent on the part of the background that is refined. For example, if a sharp absorption occurs on a broad flat background and only data points far from the absorption are used to refine $Y_0$, then the $Y_0$ value obtained is a good representation of the true value. Since this is difficult in practice, $Y_0$ is usually refined with data points incorporating some of the absorption “tail.” This tends to shift $Y_0$ slightly toward the top of the absorption feature. In Mössbauer spectra $Y_0$ is thus reduced and $S$ is smaller than the true value. Further, the shape of the absorption tail will obviously affect the refined $Y_0$ value. Hence $Y_0$ and any peak shape parameters will be correlated.

These concerns illustrate part of the difficulty in defining and utilizing a general goodness-of-fit parameter. Despite their weaknesses, however, it is clear that misfit and $\Delta$ misfit provide a useful alternative to $\chi^2$. The safest application of these parameters appears to be in combination where possible.

**Conclusions**

(1) The general goodness-of-fit parameter, $\chi^2$, has reduced value in judging fit quality since it is strongly dependent on data quality if an improper fit model is being utilized. (2) misfit is less sensitive than $\chi^2$ to data quality but not insensitive to it. It has anomalous behavior for fits to low-quality data that must be taken into account when using it alone for judging fit perfection. (3) $M$ comparison with $\Delta M$ is of high value in testing fit models. Use of $M$ or $\chi^2$ alone are most useful if spectra from the same sample are available with varied S/N ratios. (4) Attention should be paid to the proper refinement of $Y_0$ in order to estimate $M$, $\Delta M$, and $S$ correctly.

**Acknowledgments**

The author is grateful to W. A. Dollase (UCLA) and S. L. Ruby (SLAC/SSRL) for helpful discussions. P. A. Flinn (Intel Corporation and Stanford) provided an excellent random-number generator program along with a thorough testing procedure for it. This work was supported by the NSF-MRL program through the Center for Materials Research at Stanford University.

**References**


Manuscript received May 1, 1985
Manuscript accepted May 16, 1986