American Mineralogist, Volume 74, pages 685–687, 1989

Applications of Mössbauer goodness-of-fit parameters to experimental spectra: A discussion of random noise versus systematic effects

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

Dyar's (1986) reply to Waychunas (1986), regarding Mössbauer goodness-of-fit parameters and the use of simulated Mössbauer spectra, has largely misinterpreted the results of the Waychunas (1986) paper. Dyar (1986) has mixed systematic effects and random noise in experimental data, making it appear that such disparate effects cannot be separately analyzed. Dyar (1986) has also indicated that the results produced in Waychunas (1986) do not hold for "real" data, in particular, some of her own observations. The results, in fact, will hold for Dyar's (1984) data, for data collected under other conditions, and for other types of spectroscopy dealing with statistically random events.

It should be emphasized that the original discussion (Waychunas, 1986) does not examine all of the data and results of Dyar (1984), some of which are useful for methodology comparisons, but rather only those directly dealing with goodness-of-fit parameters. The Mössbauer spectrum simulations were performed to understand why good fits seem to get worse as the data get better [improved signal-to-noise (S/N) ratio], and why differing laboratories seem to have obtained widely variant MISFIT values despite using identical fitting models on very similar spectral data.

RANDOM NOISE VERSUS SYSTEMATIC EFFECTS

Random noise is a part of all Mössbauer spectra because of the inherent randomness of nuclear decay. In the simulation and in all experimental cases that I have observed, the noise has a Gaussian distribution. (In the case of very small counting rates, Poisson statistics may be more appropriate, but this will not be necessary with the usual Mössbauer experimental setup.) If the noise were not Gaussian in distribution, then the χ^2_R (reduced χ^2) values for good fits would not approach 1.00 (Bevington, 1969), and the utility of χ^2_R as a statistical parameter would be reduced or lost. (It is thus always possible to fit an experimental spectrum with some model fit such that the value of χ^2_R is very close to 1.00.) Other noise may be added to this Gaussian component, but not from spectrometer drift or any of the other effects noted by Dyar. For example, taking the worst case with a widely drifting spectrometer, the changing spectrum will produce altered peak shapes (perhaps simply broadened, perhaps skewed), but the distribution of noise in the background will still be Gaussian. If there is horrendous vibration about the spectrometer, there may be no absorption spectrum observed at all; however, the background noise even in this case will remain Gaussian in distribution. Spectrometer drive variations and sample thickness, size, or texture cannot affect the random-noise distribution at all. The simulations are therefore completely adequate for studying the effects of random noise in an experimental Mössbauer spectrum, regardless of the presence of any systematic effects.

What Dyar (1986) has actually treated when discussing systematic noise are systematic spectral distortion effects rather than a type of noise. For example, cosine smearing distorts the velocity spectrum seen by the absorber with a net effect of line shape change. It cannot affect counting statistics. Hence a badly cosine-smeared spectrum can still be fit (with a proper model) to a χ^2_R value near 1.00. Similarly, a nonlinear drive can be accommodated in the fitting procedure. However, a non-Gaussian distribution of noise could not be so easily accommodated.

The important conclusion to be drawn here is that it will be very difficult to affect the random background of a Mössbauer spectrum in any ordinary experiment and that the background will almost always have a purely Gaussian distribution. This means that random and "systematic" noise can be easily separated in terms of their spectral effects.

SPECTROMETER DRIFT

Dyar (1984, 1986) has attributed the degradation of the Mössbauer fit statistical parameters with experimental run duration to instrumental problems, such as the cosine-smearing factor, baseline inconsistencies, and source problems, as well as to long-term electronic drift and spectrometer nonlinearity and drift. However, only a few of these effects can logically be expected to degrade a spectrum as a function of time, the rest being time independent. More significantly, the real problem with all of these possible causes of spectrum degradation is that none of them can affect the Gaussian counting statistics that are the fundamental reason why χ^2_R worsens (for slightly erroneous and worse models) with increasing run time. It is possible that a poor spectrometer could cause a fit model to be less and less correct with increasing run duration, leading to accelerated divergence of the χ^2_R parameter to larger values. However, Dyar (1984, 1986) has presented only limited evidence or observations to link the divergence of the goodness-of-fit parameters to any experimental effects. As an example, consider Dyar's (1984) Figure 9 and Table 5, which show the effect of

0003–004X/89/0506–0685$02.00

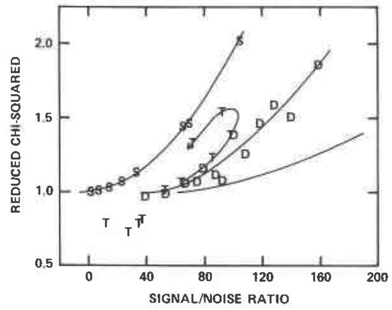

Fig. 1. The χ^2_R parameter as a function of signal-to-noise ratio (S/N) in fits to (S) simulated Mössbauer spectra (Waychunas, 1986) with intentionally incorrect fitting model, (D) run-duration study of Dyar (1984), and (T) sample thickness study of Dyar (1984). Curve through thickness points indicates trajectory with increasing thickness.

varied run duration on χ^2 and MISFIT (M). The S/N change represented by a run duration of from 3 to 60 h is approximately 35 to 156. Over this range in S/N, the simulations in Waychunas (1986) would predict a χ^2 increase of about 125%. This is consistent with a χ^2 variation of from 500 to about 1125, whereas Dyar (1984) observed a change of about 500 to 950. Since Dyar's fitting model is probably better than the intentionally slightly incorrect model used by Waychunas (1986), the variation in χ^2 that she observed is consistent with only the S/N variation, and no experimental effects need be invoked. In order to test this notion further, Gaussian noise could be added to Dyar's spectra, artificially degrading their S/N. The χ^2 versus S/N behavior derived from fits to such spectra would then descend back over the same path as the original observations if only S/N effects caused the χ^2 variation.

MISFIT PARAMETERS

Several of the rationalizations in Dyar (1984) concerning optimum spectral-data collection are incorrectly drawn because of the confusion created by the fact that there are two *different* MISFIT parameters, which behave differently and aberrantly at small S/N values. Dyar (1984) evidently calculated the MISFIT value named M in Waychunas (1986), whereas Dollase (personal communication, 1987) has been calculating the M′ version. The fact that M goes negative at small S/N values is due completely to its formulation. Hence, there is no true significance to the zero-crossing point or MISFIT values close to zero that Dyar (1984) used as an indication of optimal fitting. In fairness to Dyar (1984, 1986), this is not obvious without the simulations described in Waychunas (1986).

APPLICATIONS OF SIMULATION RESULTS TO EXPERIMENTAL SPECTRA

In her 1986 paper, Dyar has brought up a comparison from her 1984 paper that supposedly demonstrates that the simulations in Waychunas (1986) do not agree with real data. Dyar (1986, p. 1266) has stated that "data in my 1984 paper show that χ^2 reaches a minimum as MISFIT approaches zero in spectra with one million baseline counts; the simulation of that experiment by Waychunas holds χ^2 *constant* and shows MISFIT decreasing with increasing baseline counts." Dyar's 1984 paper shows such a plot (page 1134, Fig. 8), and from the behavior of MISFIT, it is clearly the M MISFIT version. The proper comparison is to Figure 3 in Waychunas (1986) though the background is only held constant at 500 000 counts, an unimportant difference. In this figure, the χ^2_R and the M parameter behave in a manner completely consistent with Dyar's plot. Dyar has chosen to consider the inappropriate χ^2_R curve. The two curves for χ^2_R represent a perfect fit (no change in χ^2_R) and a slightly poor model-based fit. Because *all* experimental spectra are at least slightly incorrectly fit, the upper χ^2_R curve generally applies. (Note that the M parameter has also been calculated and plotted for both types of fits, but it is relatively insensitive to the magnitude of the error involved.) Dyar's (1986) statement that the χ^2_R parameter has been held constant in the fits to the simulations using a (rigorously) correct model is also not true. The χ^2_R values all come out equal to 1.00 because of the perfect fits and the purely Gaussian statistics in the spectrum—the reason why a good random-number generator is critical to the simulations.

The best evidence indicating that the results of the simulations can be applied to real spectra is seen in Dyar's (1984) observations themselves. For example, the effect of sample concentration (Fig. 8) and run duration (Fig. 9) are very similar except at the greatest sample concentrations where thickness-broadening effects are significant. How can this be if the spectrum is supposedly affected mainly by thickness effects with increasing sample concentration, but mainly by spectrometer drift or some other hypothetical time-dependent effect with run duration? Why should the same model be inadequate in terms of the goodness-of-fit parameters in the same way for such differing effects? The answer is that both sets of spectra are approximately equally affected by S/N effects over most of the range that Dyar (1984) investigated. It does not matter how the S/N is changed, whether by sample concentration or run duration, the effect is the same.

In order to illustrate this, in Figure 1 the results of the simulation and Dyar's observations are plotted. Dyar's χ^2 values have been converted to χ^2_R for this comparison. Dyar's run-duration data reside along a smooth parabolic curve just as do the simulations. This is indicative of a constant fitting error for both data sets and not a time-dependent change in spectral form. Dyar's varying concentration data start off along the same curve as the run-duration points, but diverge to larger χ^2_R values with increasing sample concentration. The estimated trajectory of these data as sample concentration increases shows both the divergence due to increasing fitting error and reduction of χ^2 due to decreasing S/N as sample absorption drops for the greatest concentration. The simulation points represent a larger (intentional) fitting error than in any of Dyar's fits. Another curve illustrates the effect of a smaller (least-concave parabola) fitting error with varying S/N. By varying the type of fitting model, a whole family of similar curves can be generated. The separation between such curves is due entirely to fitting-model errors, whereas the trend along the curves is due to S/N alone. The curves follow the empirical relation

$$\chi^2_R = 1.00 + [E \times 10^{-4}(S/N)^2],$$

where E is directly related to the magnitude of the fitting error. For the perfect-fit model (which will fit absolutely anything perfectly), the curve flattens to a constant at χ^2_R equal to 1.00 independent of S/N. The χ^2_R parameter is thus dominated by the S/N variation for almost all of Dyar's points rather than by any changing appropriateness of the fit model.

Dyar's data points for the lowest sample concentrations yield χ^2_R values considerably smaller than 1.00. This is presumably due to the magnitude of the uncertainty in χ^2_R for low S/N spectra and the fact that a least-squares fitting program always attempts to minimize χ^2_R. Hence, the lowest rather than the most probable χ^2_R values can be generated. However, I have not observed χ^2_R values as low as these previously.

This analysis does not imply that such factors as the thickness effect have insignificant effects on a spectrum. Dyar's percentage transmission data demonstrate that this occurs (Table 4). It is just that, in the range of S/N that Dyar (1984) examined, any model-dictated fitting error may be dominated by the effects of great variations in the S/N ratio. For a different reason, the MISFIT parameters have reduced utility. They diverge meaninglessly at low S/N, independent of any physical effect, and hence should not be used in the manner that Dyar (1984) described.

Dyar (1986) has stated that her results (those that were obtained by the zero-crossing points of M) are well corroborated by theoretical considerations. It is true that sample thickness leads to a "blackness" or self-absorption effect and that this broadens spectral lines. But the effect has no threshold (Gutlich et al., 1978; Margulies and Ehrman, 1961; Ure and Flinn, 1971) so that the ideal sample concentration is almost zero if no broadening effects are desired. More realistically, an "ideal" sample is constituted by an absorber that maximizes the resonant absorption and minimizes the total nonresonant absorption and the line broadening. Theoretical estimates for minerals depend on the particular recoil-free fractions but are about 5- to 6-mg Fe cm^{-2} (Greenwood and Gibb, 1971; Gutlich et al., 1978). But if the fitting program refines the line widths and line shapes, there should be minimal variation in the goodness-of-fit parameters directly due to thickness effects. Hence, whether one can expect to see thickness effects with a goodness-of-fit parameter at any concentration level depends on the flexibility of the fitting model. If variations in line shape occur, these could be seen by examining the refined line-shape parameters, thereby verifying the presence of the broadening effect. However, since there is no significance to the MISFIT values calculated at the lower S/N ratios, it can only be fortuitous that Dyar has obtained thickness values reasonably similar to theory.

As a further example, consider the assumption that the simulations can be used to predict optimum sample concentration from the behavior of the goodness-of-fit parameters. Since there is no actual physical spectroscopy in the simulations, this is, of course, apocryphal. However, using the background data for 500 000 counts, the optimal absorption as determined by the MISFIT zero-crossing point is about 3%. For silicate minerals, this would be equivalent to a sample concentration of about 5 mg Fe cm^{-2}, remarkably close to the theoretical estimates. This illustrates why Dyar's (1984) result is fortuitous and that most of the degradation she observed in the χ^2 parameter is due to S/N effects.

Finally, one implication that researchers might obtain from Dyar (1984, 1986) is that the S/N effects can be neglected in the analysis of experimental data. Depending on the type of experiment being performed, this assumption can result in poorly tested models and the choice of improper or poor run conditions, namely, sample concentrations that are too small and run durations that are too short. If one does wish to optimize experimental run conditions, both S/N and experimental factors should be considered. Fortunately, S/N effects can be easily modeled and appropriate corrections made, as desired, to MISFIT. The χ^2_R parameter affords a direct indication of model inappropriateness and should not be modified. S/N variations do not actually change the form of the spectrum, only its quality. In contrast, experimental effects such as thickness broadening must be accommodated by a change in the fitting model, since the spectrum is truly altered by such effects.

REFERENCES CITED

Bevington, P.R. (1969) Data reduction and error analysis for the physical sciences. McGraw-Hill, New York.

Dyar, M.D. (1984) Precision and interlaboratory reproducibility of measurements of the Mössbauer effect in minerals. American Mineralogist, 69, 1127–1144.

——— (1986) Practical application of Mössbauer goodness-of-fit parameters for evaluation of real experimental results: A reply. American Mineralogist, 71, 1266–1267.

Greenwood, N.N., and Gibb, T.C. (1971) Mössbauer spectroscopy. Chapman and Hall, London.

Gutlich, P., Link, R., and Trautwein, A. (1978) Mössbauer spectroscopy and transition metal chemistry. Springer-Verlag, Berlin.

Margulies, S., and Ehrman, J.R. (1961) Transmission and line broadening of resonance radiation incident on a resonance absorber. Nuclear Instruments and Methods, 12, 131–137.

Ure, M.C.D., and Flinn, P.A. (1971) A technique for the removal of the "blackness" distortion of Mössbauer spectra. In I.J. Gruverman, Ed., Mössbauer effect methodology, vol. 7, p. 245–262. Plenum Press, New York.

Waychunas, G.A. (1986) Performance and use of Mössbauer goodness-of-fit parameters: Response to spectra of varying signal/noise ratio and possible misinterpretations. American Mineralogist, 71, 1261–1265.

MANUSCRIPT RECEIVED JULY 16, 1987
MANUSCRIPT ACCEPTED JANUARY 20, 1989