Acta Cryst. (1982). B38, 1099–1100

Comparison of the Fits of Two Models to the Same Data Set

BY E. PRINCE

National Measurement Laboratory, National Bureau of Standards, Washington, DC 20234, USA

(*Received* 6 *April* 1981; *accepted* 20 *October* 1981)

Abstract

A frequently encountered problem is the determination of whether one model gives a significantly better fit to a set of data than another. This may be studied by examining the correlation between the differences in the predictions of the models and the corresponding differences between the observed data and the arithmetic means of the predictions. The existence and precision of such correlations may be determined using the techniques of linear regression. The analysis has been applied to a neutron powder diffraction study of the defect structure of nonstoichiometric lithium tantalate.

Introduction

It happens frequently in crystallographic refinement that it is necessary to decide whether one model represents a 'significantly' better fit to the observed data than another. The special case in which the parameters of one model are a subset of the parameters of the other was addressed by Hamilton (1965), who devised a test that may be applied to the ratio of the weighted R indices resulting from refining the two models. This test is derived from statistical tests based on the F distribution (Hamilton, 1964), which is a probability distribution function appropriate to the ratio of two *independent* random variables each of which has a χ^2 distribution. This test is extremely useful for deciding whether the data support a hypothesis that atoms have some nonideal configuration or whether there is evidence for disorder of atomic species at a particular site. In many cases, however, one model is not a subset of the other, and it may not be possible to define two independent χ^2 statistics. Such a case arose in a recent study (Santoro, Roth & Austin, 1982) of nonstoichiometric lithium tantalate by neutron powder diffraction. This paper describes a statistical test that may be used whether or not the statistics are independent and then applies it to the $LiTaO_3$ problem.

Statistical analysis

A general test for comparing two models was introduced by Williams & Kloot (1953). We follow here the analysis given by Himmelblau (1970). It involves determining the slope, λ, of the regression line $Z = \lambda X$, where $Z_i = Y_{oi} - \frac{1}{2}(Y_{1ci} + Y_{2ci})$, and $X_i = (Y_{1ci} - Y_{2ci})$. Here Y_{1cl} and Y_{2cl} are the predicted values of Y at point i for model 1 and model 2 respectively. Thus the test seeks a correlation between the differences in the predictions of the models and the differences between the observed values and the arithmetic means of the predictions. Suppose that model 1 is the correct one, and the measurements have been made with great precision, so that $Y_{ol} = Y_{1cl}$. Then $Z_l = \frac{1}{2}(Y_{1cl} - Y_{2cl})$, and $\lambda = \frac{1}{2}$. A positive slope for the regression line therefore favors model 1, and a negative slope favors model 2. When neither model is a perfect fit and the observations are subject to random fluctuations, $|\lambda|$ will, in general, be less than $\frac{1}{2}$. The hypothesis that the two models represent equally good fits to the data may be tested by establishing a confidence interval for λ and determing whether it does or does not include $\lambda = 0$.

Although this test is applicable to a large variety of problems, the case of immediate interest is the comparison of two models for the defect structure of nonstoichiometric lithium tantalate (Santoro *et al.*, 1982) refined to neutron powder diffraction data by the method developed by Rietveld (1969). In this case the observations are the numbers of counts in an interval of time at a given value of the scattering angle, 2θ. These are random events and are therefore subject to statistical fluctuations having a Poisson distribution. Each observation, therefore, has an estimated variance equal to the total count and an estimated standard deviation equal to the square root of the count. The quantities Z_l should therefore be expressed as fractions of the standard deviations of the observations, Y_{ol}. The expression for Z_l then becomes $Z_l = [Y_{ol} - \frac{1}{2}(Y_{1cl} + Y_{2cl})]/Y_{ol}^{1/2}$. The slope of the regression line is the value of λ that minimizes the quantity

$$f(\lambda) = \sum_{l=1}^{N} (Z_l - \lambda X_l)^2.$$

which is

$$\hat{\lambda} = \sum_{i=1}^{N} Z_i X_i / \sum_{i=1}^{N} X_i^2.$$

The estimated variance of $\hat{\lambda}$ is

$$\hat{\sigma}^2_{\lambda} = \left[\sum_{i=1}^{N} Z_i^2 - \hat{\lambda}^2 \sum_{i=1}^{N} X_i^2\right] \Big/ \left[(N-1)\sum_{i=1}^{N} X_i^2\right],$$

and the $(1-\alpha)$ per cent confidence interval is $\hat{\lambda} \pm \hat{\sigma}_{\lambda} T_{1-\alpha/2}$, where $T_{1-\alpha/2}$ is the value such that

$$\int_{-\infty}^{T} \Phi(t,v)\,dt = 1-\alpha/2$$

and $\Phi(t,v)$ is the density function for Student's t distribution with $v = (N-1)$ degrees of freedom.

Comparison of models for lithium tantalate

Santoro, Roth & Austin (1982) refined two models, one a lithium-vacancy model proposed by Lerner, Legras & Dumas (1968) and the other a stacking-fault model proposed by Nassau & Lines (1970), to a neutron diffraction powder pattern of a lithium tantalate sample with the lithium-deficient composition $9LiTaO_3{:}Ta_2O_5$. The weighted R indices, R_w, for the two models were 0·0950 and 0·0934, respectively, so the model of Nassau & Lines would appear to be preferred. We must, however, address the question 'Is the better fit of the model of Nassau & Lines significant, or is the difference no greater than would normally be observed by chance alone?'

There are 199 points in the powder pattern for which the difference between the predictions of the two models is at least ten counts. The 95% confidence interval for the slope of the regression line for those points is $-0{\cdot}01877 \pm 0{\cdot}01069$. Thus the hypothesis that the two models give equally good fits to the data can be rejected at the 5% confidence level, and the model of Nassau & Lines is a significantly better representation of the data. Fig. 1 shows a scatter plot of the subset of the data that was used in the analysis. The largest differences can be seen to be concentrated in the upper left quadrant.

It must be emphasized that this comparison does not prove that either model is the correct one. It merely shows that the difference between the fits of the two models is improbable and is therefore unlikely to be observed because of chance alone. To infer that the better-fitting model is in fact the correct one requires both that it include all relevant sources of possible bias and that there is no third model that is equally consistent with the constraints of physics and chemistry and fits the data even better. In the case of Rietveld powder refinement, as in the example given here, it is required at least that the data have been carefully examined for such sources of systematic error as preferred orientation, that individual, resolved peaks have been checked to verify the adequacy of the peak shape function, and that an appropriate background function has been included as a set of refined parameters in the model.

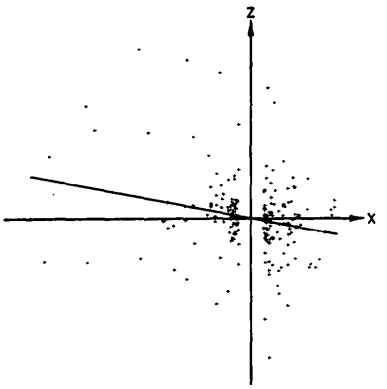

Fig. 1. A scatter plot of the regression of the difference between the observed data and the mean of two models for nonstoichiometric $LiTaO_3$ against the difference between the models.

References

HAMILTON. W. C. (1964). *Statistics in Physical Science*, pp. 84–89. New York: Ronald Press.

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502–510.

HIMMELBLAU, D. M. (1970). *Process Analysis by Statistical Methods*, pp. 216–221. New York: John Wiley.

LERNER, P., LEGRAS. C. & DUMAS, J. P. (1968). *J. Cryst. Growth*, **3**, 231–235.

NASSAU, K. & LINES, M. E. (1970). *J. Appl. Phys.* **41**, 533–537.

RIETVELD, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.

SANTORO, A., ROTH, R. S. & AUSTIN, M. (1982). *Acta Cryst.* B**38**, 1094–1098.

WILLIAMS, E. J. & KLOOT, N. H. (1953). *Aust. J. Appl. Sci.* **4**, 1–17.