Acta Cryst. (1994). A**50**, 703–714

A Bayesian Approach to Extracting Structure-Factor Amplitudes from Powder Diffraction Data

By D. S. Sivia and W. I. F. David

*ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX*11 0*QX, England*

(*Received* 25 *October* 1993; *accepted* 25 *March* 1994)

Abstract

This paper presents a method for the reliable extraction of structure-factor amplitude information from the least-squares integrated-intensity refinement of powder diffraction data. The inevitable overlap of Bragg reflections can lead to strongly correlated reflection intensities that can, in turn, produce unrealistic negative intensity estimates. A Bayesian method is presented that tackles the problem of highly correlated positive and negative intensities. The results indicate that accurate structure-factor amplitudes may be recovered even in regions of a powder diffraction pattern where overlap is almost complete.

1. Introduction

Structure determination from powder diffraction data alone is a substantial crystallographic challenge. Powder diffraction data are, in general, of poorer statistical quality than single-crystal data. More significantly, however, the collapse of three dimensions of diffraction data on to the one dimension of a powder diffraction pattern leads to inevitable peak overlap. Much attention has been given to the problems of overlapping integrated intensities (David, 1987, 1990; Jansen, Peschar & Schenk, 1992; Estermann, McCusker & Baerlocher, 1992; Bricogne, 1991; Gilmore, Henderson & Bricogne, 1991). It is, however, clear that a poor evaluation of the integrated intensities in a powder diffraction pattern will always lead to poor results no matter how good the algorithms used in the analysis of overlapped reflections. In this paper, attention is focused on the reliable extraction of structure-factor amplitudes. It is assumed that the unit cell has already been derived using, for example, auto-indexing techniques and that, therefore, the Bragg-peak positions can be accurately determined. The approach is based upon the Pawley (1981) method and involves the least-squares fitting of the diffraction pattern to separate integrated intensities. The parameters that are usually varied in such a procedure include background parameters, cell parameters (determining peak positions), peak-width parameters (determining peak shape) and the integrated intensities (which are directly proportional to peak area). The procedure works extremely well if peak overlap is either non-existent or exact (in the case, for example, of 511 and 333 cubic Bragg reflections). When substantial overlap occurs, intensity values can become highly correlated. Although the summed area of a group of Bragg peaks will be well determined, the individual intensities can vary wildly between negative and positive values that are substantially larger in magnitude than the overall 'clump' intensity itself. This is clearly wrong and has been taken to represent a fundamental weakness of the least-squares method. Pawley (1981) was aware of the problem and ingeniously introduced into the least-squares analysis additional slack constraint terms that had the tendency to force individual intensities to be close to the mean value of a 'clump' of intensities. In practice, this approach does not completely obviate the problem of highly varying intensity values. An elegant tech-

©1994 International Union of Crystallography
Printed in Great Britain – all rights reserved

ISSN 0108-7673 ©1994

nique proposed by Le Bail, Duroy & Fourquet (1988) has been found to surmount the problems of wild intensity variations and has as a result become the preferred method of integrated-intensity extraction from powder diffraction data. A major benefit of this approach is its computational efficiency, in that no large least-squares matrix need be inverted. The method also gives inherently positive intensities if the background is correctly estimated. This appears at first sight to give significant advantages over other methods. It does, however, suffer from a number of limitations that include difficulties in evaluating the estimated standard deviations of and statistical correlations between intensity values. This paper sets out to provide a rigorous statistical formulation of the extraction of structure-factor amplitudes from overlapped reflections in a powder diffraction pattern. The approach is based upon the least-squares approach and thus is based upon the work of Pawley (1981) rather than that of Le Bail *et al.* (1988). The main departure from the work of Pawley is the method of dealing with highly correlated overlapping reflections. Rather than employing an empirical restraint algorithm, this paper shows that there is a perfectly natural probabilistic method of including the requirement of positive intensities within a Bayesian context.

An algorithm is developed by considering firstly, in § 2, the assumptions of a conventional approach to the question: 'Given that A^2 is measured to be $I = I_0 \pm \sigma_I$, what is the expected value of A?' The conventional answer is well defined as long as $\sigma_I \ll I_0$ but breaks down when $I_0 \ll \sigma_I$ and in particular when I_0 is negative. The Bayesian approach is developed for a single peak from a consideration of the shortcomings and assumptions of a conventional analysis (§ 3). § 4 deals with the complication of overlapping reflections while § 5 considers further aspects of prior knowledge about the distribution of peak intensities. In § 6, the algorithm is applied to a section of a neutron powder diffraction pattern of benzene obtained from the high-resolution powder diffractometer HRPD at ISIS.

2. A conventional analysis

Experimentally measured diffraction data are related linearly to the intensity of a Bragg peak, which, in turn, depends on the square of the structure-factor amplitude. The linear relationship means that Bragg-peak intensities can be estimated from diffraction data in a straightforward manner. A simple procedure often used in the estimation of 'observed' peak intensities in the Rietveld method involves adding up the measured counts in the region of a Bragg peak and subtracting from it the average of the background in its neighbourhood. When applied in an iterative manner, this is essentially the technique developed by Le Bail *et al.* (1988). A more statistically rigorous estimate of the intensity, however, can be obtained by adopting a procedure that involves the fitting of a suitable profile and background function to the Bragg peak. The first detailed Bayesian fitting procedure was developed by Oatley & French (1982).

Whether an adding or fitting procedure is used, the result (having been corrected for multiplicity factors) tends to be stated as

$$I = I_0 \pm \sigma_I, \tag{1}$$

where I_0 is the best estimate of the intensity I and σ_I is a measure of its uncertainty. In powder diffraction data, the values of the background and the intensity of the Bragg peaks are usually found to have very little correlation between them; even if their estimates were strongly linked, any doubt about the magnitude of the background would be reflected in a corresponding larger value of the error bar σ_I. The crystal structure is, of course, related to the amplitudes (and phases) of the structure factors rather than to their intensities. The unnormalized structure-factor amplitude A is related to this reduced intensity I by

$$I = A^2. \tag{2}$$

(Throughout this paper, A denotes an un-normalized structure-factor amplitude. It is related to the true structure-factor amplitude $|F|$ by a multiplicative term that involves the reflection multiplicity, m_j, the appropriate Lorentz–polarization correction and a scale, the magnitude of which is determined by the number of counts in the diffraction pattern.) It thus seems reasonable that the best estimate of the amplitude A_0 is given by the square root of the intensity $I_0^{1/2}$. The error in the amplitude σ_A can be estimated simply by differentiating (2), squaring both sides and taking (the square roots of) the expectation values:

$$\sigma_A = \langle(\delta A)^2\rangle^{1/2} = [\langle(\delta I)^2\rangle/(4A^2)]^{1/2} = \sigma_I/(2I_0^{1/2}). \tag{3}$$

Unfortunately, this elementary analysis breaks down when I_0 is negative, as often occurs for weak reflections and strongly overlapping reflections. The clue to the resolution of this difficulty lies in the examination of the assumptions that are implicit in the seemingly innocuous analysis presented above.

3. The Bayesian approach

Consider (1). The parameters I_0 and σ_I are usually numbers that are returned by a fitting program. I_0 represents the value of the intensity that the Bragg peak would have in order to minimize the discrepancy between measured and predicted data. It is the *maximum-likelihood* estimate because it is the value of the intensity that maximizes the *likelihood function*

or the conditional probability distribution function (p.d.f.), prob ({data}|I). In fact, (1) is shorthand for saying that the likelihood function is approximately Gaussian with mean I_0 and standard deviation σ_I:

$$\text{prob}\,(\{\text{data}\}|I) \propto \exp[-(I-I_0)^2/2\sigma_I^2]. \qquad (4)$$

If the data are themselves subject to independent and additive Gaussian noise, then it can easily be shown that I_0 is given by the familiar least-squares procedure. If the data are Poisson with very low counts, however, then the least-squares estimate becomes a poor approximation to the (logarithm of the) likelihood function (Antoniadis, Berruyer & Filhol, 1990).

The quantity that is required, however, is not the likelihood function but the *posterior probability*: prob (I|{data}). The words 'maximum-likelihood estimate' can be misleading because they seem to suggest that I_0 is the most likely value of the intensity. This is not so: I_0 is the intensity that would make the observed data most probable; this is logically quite different from the most probable intensity, given the observed data. (An everyday example that highlights this important distinction is the difference between the probability of rain given that there are clouds overhead and the probability of clouds overhead given that it is raining. The two probabilities are clearly not the same!) Fortunately, these two distinct conceptual entities are related to each other through Bayes's theorem (Jeffreys, 1939; Jaynes, 1983):

$$\text{prob}\,(I|\{\text{data}\}) \propto \text{prob}\,(\{\text{data}\}|I)\ \text{prob}\,(I). \qquad (5)$$

The posterior p.d.f. is thus related to the likelihood function through the *prior probability* prob (I). The prior p.d.f. represents what is known about the intensity before the data have been analysed. One clear piece of prior information is that the intensity is positive; this information can be encoded by the simple assignment:

$$\text{prob}\,(I) = \begin{cases} \text{constant} & \text{for } I \geq 0 \\ 0 & \text{otherwise} \end{cases}. \qquad (6)$$

A discussion about alternative, more optimal, prior p.d.f.'s is presented in § 5. This naive assignment is, however, almost always adequate; it is the form that is used throughout §§ 3 and 4.

In order to estimate the amplitudes of the structure factors, the posterior p.d.f. prob (A|{data}) is required. The following coordinate transformation must, therefore, be performed:

$$\text{prob}\,(A|\{\text{data}\}) = \text{prob}\,(I|\{\text{data}\})\left|\frac{\mathrm{d}I}{\mathrm{d}A}\right|. \qquad (7)$$

The *Jacobian* term on the right-hand side is given by the derivative of (2) as $2|A|$. The posterior p.d.f. for the amplitudes can then be written as

$$\text{prob}\,(A|\{\text{data}\}) \propto |A| \exp[-(A^2-I_0)^2/2\sigma_I^2], \qquad (8)$$

where all the normalization factors that do not depend explicitly on A have been omitted. Note that the coordinate transformation automatically incorporates the simple prior of (6) because real values of the amplitude can only map to positive intensities. The p.d.f. of (8) is also symmetric, so that prob (A) = prob ($-A$). This is not unexpected since square roots are being used. By definition, the amplitude cannot be negative, so that only the positive solution need be considered.

The strange-looking expression (8) can be approximated by a more familiar form, in the usual way, by expanding its logarithm, $L = \ln[\text{prob}\,(A|\{\text{data}\})]$, as a quadratic Taylor series about its maximum:

$$L(A) \simeq L(A_0) + \tfrac{1}{2}(A-A_0)^2 \left.\frac{\mathrm{d}^2L}{\mathrm{d}A^2}\right|_{A_0} + \ldots, \qquad (9)$$

where

$$\left.\frac{\mathrm{d}L}{\mathrm{d}A}\right|_{A_0} = 0.$$

This enables the formulation of a good approximation for the posterior p.d.f. for the amplitude with a Gaussian of mean A_0 and standard deviation σ_A:

$$\begin{aligned}\text{prob}\,(A|\{\text{data}\}) &= \exp(L) \\ &\simeq [1/\sigma_A(2\pi)^{1/2}] \\ &\quad \times \exp[-(A-A_0)^2/2\sigma_A^2], \qquad (10)\end{aligned}$$

where the relevant parameters are given by

$$A_0 = \tfrac{1}{2}[2I_0 + (4I_0^2 + 8\sigma_I^2)^{1/2}]^{1/2} \qquad (11)$$

and

$$\begin{aligned}\sigma_A &= \left(-\left.\frac{\mathrm{d}^2L}{\mathrm{d}A^2}\right|_{A_0}\right)^{-1/2} \\ &= \{(1/A_0^2) + [2(3A_0^2 - I_0)/\sigma_I^2]\}^{-1/2}. \qquad (12)\end{aligned}$$

It can easily be shown that these expressions for the parameters of the Gaussian p.d.f. of (10) reduce to those of the conventional analysis, given in § 2, in the limit $I_0 \gg \sigma_I$. For those reflections where the intensity is poorly determined by the data, the best estimates of the corresponding amplitudes and their reliabilities can be very different. These results are illustrated graphically with a few examples.

Consider first of all the case of a well determined reflection, $I = 9$ (1), as illustrated in Fig. 1(*a*). It confirms that, in this instance, the conventional result for the amplitudes agrees almost perfectly with the Bayesian solution. For a reflection that is poorly determined by the data, $I = 1$ (9), Fig. 1(*b*) shows an interesting divergence. The Bayesian solution of (8), which is plotted as a full line, indicates that the amplitude could easily be 0 but is very unlikely to be bigger than 6; the optimal estimate is about 2.6. In contrast, the conventional result, shown as a dashed

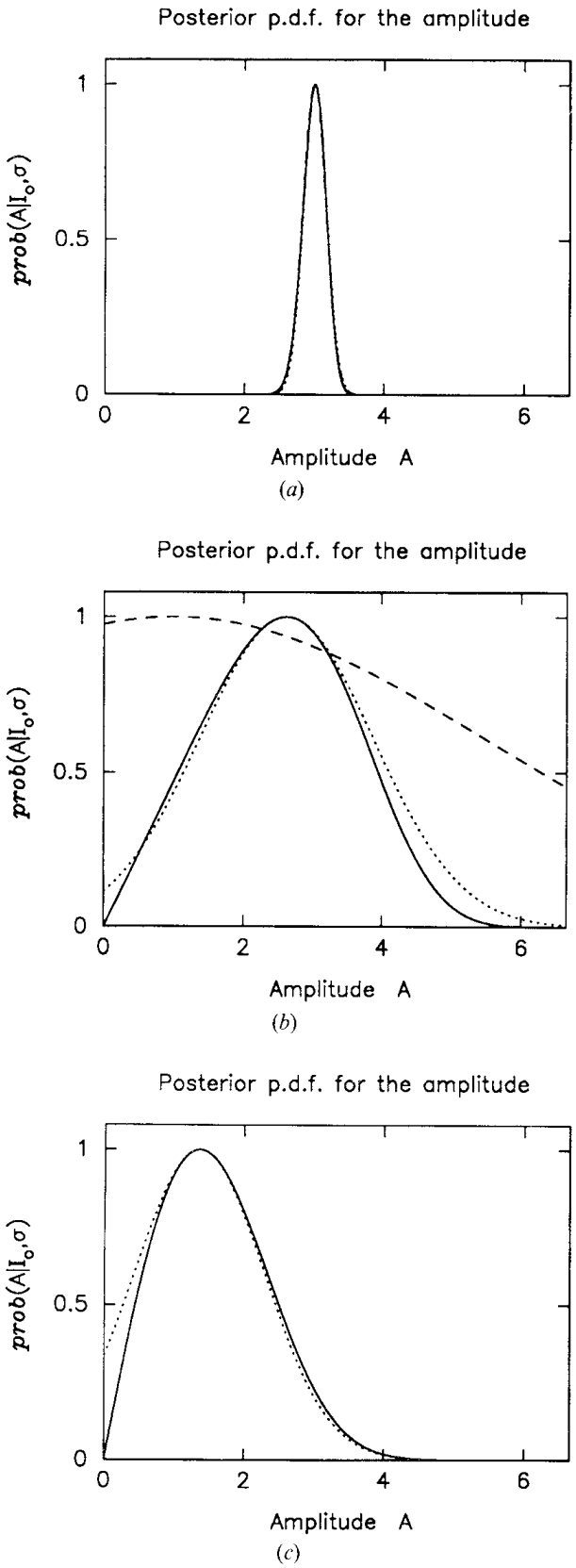

Fig. 1. The posterior probabilities for amplitudes given the best-fit intensities (mean and estimated standard deviation): (*a*) $I = 9\ (1)$; (*b*) $I = 1\ (9)$; (*c*) $I = -20\ (9)$. The full line is the Bayesian solution (with a naive positivity prior), the dotted line is the Gaussian approximation to the Bayesian solution and the dashed line is the 'conventional' solution (where it exists).

line, gives an estimate of 1 for the amplitude and does not exclude values of up to about 12 as being unreasonable. It can also be seen that the Gaussian p.d.f. of (10), (11) and (12), plotted as a dotted line, is a good approximation to the Bayesian solution. Finally, the most striking case of all occurs when the estimate of the intensity is negative; for example: $I = -20\ (9)$. Fig. 1(*c*) shows that there is a perfectly respectable Bayesian solution, whereas the conventional analysis breaks down completely.

4. Extension to overlapping peaks

The Bayesian analysis presented above, for estimating the amplitudes of individual (isolated) reflections, extends quite naturally to the case of overlapping Bragg peaks. This additional complication of overlap occurs frequently in powder diffraction.

Again, the usual procedure is to start by estimating the intensities of the reflections with a suitable fitting program (since these are linearly related to the data). Because of the overlap, however, the likelihood function of (4) has to be generalized to a multivariate Gaussian:

$$\text{prob}\,(\{\text{data}\}|\{I\}) \propto \exp\left[-\tfrac{1}{2}\sum_{i,j}(I_i - I_{0i})(C_I^{-1})_{ij}(I_j - I_{0j})\right], \tag{13}$$

where I_j is the intensity of the jth reflection, with a best-fit value of I_{0j}, and C_I^{-1} is the inverse of the *covariance matrix* for the intensities:

$$(C_I)_{ij} = \langle (I_i - I_{0i})(I_j - I_{0j}) \rangle. \tag{14}$$

As before, Bayes's theorem states that this likelihood function must be multiplied by the prior p.d.f. prob $(\{I\})$ to obtain the posterior p.d.f. for intensities: prob $(\{I\}|\{\text{data}\})$. Once again, the simple prior of (6) is used, so that it is zero if any of the intensities are negative. This is automatically incorporated in the coordinate transformation to the amplitudes: $I_j = A_j^2$. The Jacobian term in (7) is now given by the determinant of the diagonal matrix:

$$\partial I_j / \partial A_i = 2A_i \delta_{ij}, \tag{15}$$

where δ_{ij} is the Kronecker delta. Thus, the posterior p.d.f. for the amplitudes becomes

$$\text{prob}\,(\{A\}|\{\text{data}\}) \propto \left(\prod_j |A_j|\right) \exp\left[-\tfrac{1}{2}\sum_{i,j}(A_i^2 - I_{0i}) \times (C_I^{-1})_{ij}(A_j^2 - I_{0j})\right]. \tag{16}$$

This p.d.f. can also be approximated by a multivariate Gaussian by expanding its logarithm, $L = \ln[\text{prob}\,(\{A\}|\{\text{data}\})]$, as a quadratic Taylor series

about the maximum (at $\{A_0\}$):

$$\text{prob}\,(\{A\}|\{\text{data}\}) \simeq K \exp\left[-\tfrac{1}{2}\sum_{i,j}(A_i - A_{0i}) \times (C_A^{-1})_{ij}(A_j - A_{0j})\right], \quad (17)$$

where K is a normalization constant and C_A^{-1} is the inverse of the covariance matrix for the amplitudes [like (14) but with A's instead of I's]:

$$(C_A^{-1})_{ij} = -\left.\frac{\partial^2 L}{\partial A_i \partial A_j}\right|_{\{A_0\}}, \quad \text{where} \quad \left.\frac{\partial L}{\partial A_j}\right|_{\{A_0\}} = 0. \quad (18)$$

In practice, it is better to find the optimal solution $\{A_0\}$ by minimizing L, rather than by trying to solve the set of nonlinear simultaneous equations for its gradient to be zero. We believe that the posterior p.d.f. of (16) has a single maximum (in the positive hyper-octant of the amplitudes), which can be found quite efficiently by a simplex search (Nelder & Mead, 1965) followed by a Newton–Raphson refinement (Press, Flannery, Teukolsky & Vetterling, 1986). This is illustrated graphically by the following examples.

Fig. 2(*a*) shows the posterior p.d.f. for the intensities of two overlapping reflections. The maximum likelihood estimate of the intensities assumed is $I_1 = 9\,(3)$, $I_2 = 4\,(2)$, with a relative cross correlation of -80%. The corresponding posterior p.d.f. for the two amplitudes is shown in Fig. 2(*b*); the Gaussian approximation to this, given by (17) and (18), is shown as a dotted line. A more striking example of this analysis can be seen in Fig. 3, where the data come from a neutron powder diffraction experiment conducted on High-Resolution Neutron Powder Diffractometer (HRPD) at the ISIS spallation neutron facility. The maximum-likelihood estimate for the intensities of the $\bar{4}04$ and the $\bar{5}21$ reflections is $I_1 = -5\,(25)$, $I_2 = 14\,(11)$, with a relative cross correlation of -97%. The corresponding posterior p.d.f.'s for the intensities and amplitudes are given in Figs. 3(*a*) and (*b*), respectively. Although probability theory warns that there is a large amount of uncertainty with regard to the amplitudes of the reflections, it is still quite surprising to find that the optimal estimate is given by $A_1 = 3.1$ and $A_2 = 2.9$. A cursory inspection of the best-fit intensities, for example, would tend to suggest that A_2 should be greater than A_1. This counterintuitive result is confirmed by a Rietveld refinement of the crystal structure, which gives the amplitudes as $A_1 = 3.1$ and $A_2 = 2.6$.

5. Discussion

In § 3, a very simple assignment was used for the prior p.d.f. for the intensity of a reflection. This section addresses the question of whether there is a better alternative to this naive choice. The prior p.d.f. is supposed to reflect all that is known about the intensity before analysis of the current data. Equation (6) certainly incorporates the prior knowledge that the intensity is positive but, unfortunately, it does a little more than just that.

Suppose, for example, it had been decided to estimate the amplitude directly from the data. Then, a (nonlinear) fitting program would provide the likelihood function prob $(\{\text{data}\}|A)$. To translate this into the posterior p.d.f. for the amplitude, Bayes's theorem states that the prior p.d.f. prob (A) must now be assigned. Since (6) incorporates knowledge about positivity, the same prior could be used with the caveat that the I's are replaced with A's. Sadly, this would not lead to the same result as the analysis of § 3! The source of the inconsistency lies in the form of the prior, since (6) for I's does not represent the same information as (6) for A's. This can be seen

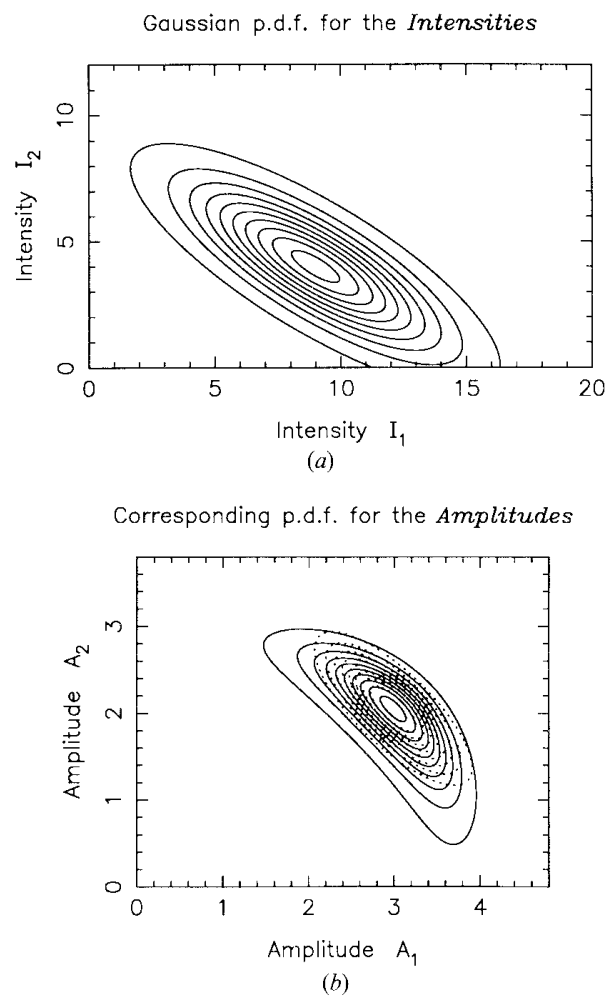

Fig. 2. (*a*) The Gaussian probability distribution function (p.d.f.) representing the best-fit intensities of two overlapping reflections, $I_1 = 9\,(3)$ and $I_2 = 4\,(2)$ with a relative cross correlation of -80%. (*b*) The corresponding p.d.f. for the amplitudes. The dotted contours are the multivariate Gaussian approximation to the posterior probability distribution function (solid contours) for the amplitudes.

by writing out the coordinate transformation of (7) explicitly for the priors:

$$\text{prob}\,(A) = \text{prob}\,(I)|A|. \qquad (19)$$

Jaynes (1983) pointed out that if a p.d.f. is required that expresses complete *ignorance* about such a scale parameter, other than positivity, then the same (form of) prior should be assigned irrespective of whether one is dealing with A, I or any other power of A. This requirement can only be satisfied if a *Jeffreys prior* is used:

$$\text{prob}\,(x) \propto 1/x, \quad \text{for } x > 0, \qquad (20)$$

which is equivalent to saying the prior p.d.f. should be uniform (a constant) for the logarithm of the intensity (or amplitude). Before the effect of using the Jeffreys prior on the results of the analysis in § 3 is illustrated, the possibility of using *informative* priors is considered.

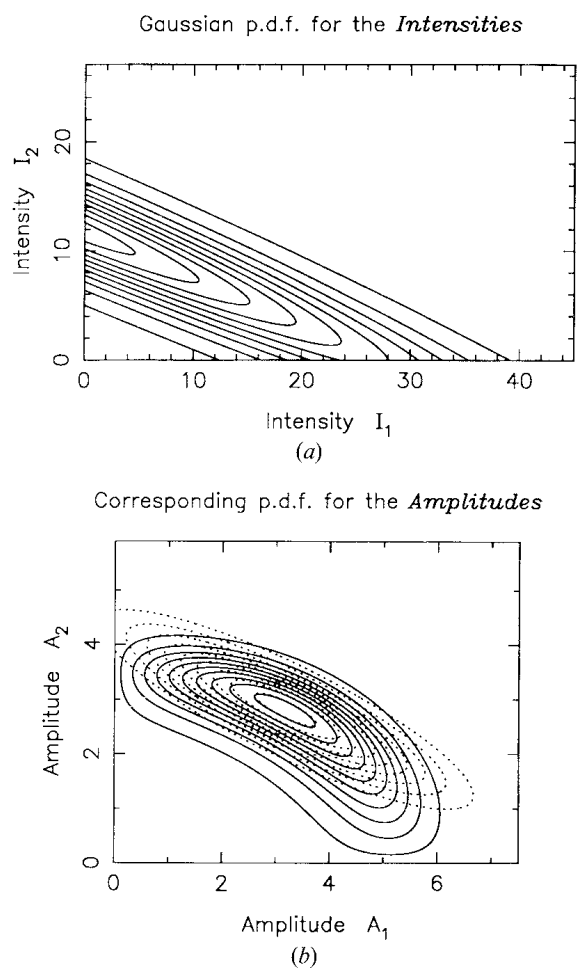

Fig. 3. (*a*) The Gaussian probability distribution function (p.d.f.) representing the best-fit intensities of two overlapping $\bar{4}04$ and $\bar{5}21$ reflections, $I_1 = -5.1$ (259) and $I_2 = 13.9$ (113) with a relative cross correlation of -97%. (*b*) The corresponding p.d.f. for the amplitudes. The dotted contours are the multivariate Gaussian approximation to the posterior probability distribution function (solid contours) for the amplitudes.

The Jeffreys prior arises from claiming complete ignorance about the scale (of the size) of the parameter involved. Suppose, however, that the average value of the intensity for a given reflection is known: $\langle I \rangle = \mu$. Then, the principle of maximum entropy (Jaynes, 1983) states that the following p.d.f. should be assigned:

$$\text{prob}\,(I|\mu) \propto \exp\,(-I/\mu), \quad \text{for } I \geq 0. \qquad (21)$$

This is, in fact, just the *Wilson prior* for an acentric reflection (Wilson, 1949; French & Wilson, 1978). If in addition it was known that the data were from a centrosymmetric crystal, then the Wilson prior appropriate for a centric reflection could be used:

$$\text{prob}\,(I|\mu, \text{centric}) \propto (I^{-1/2}) \exp\,(-I/2\mu,), \quad \text{for } I \geq 0. \qquad (22)$$

Note that both the Wilson priors revert to the Jeffreys form if μ is not known; details are given in the Appendix.

A few graphical examples are used to illustrate the effect of using these alternative priors. The solid line in Fig. 4(*a*) shows the posterior p.d.f. for the amplitude of a reflection that results from using the naive prior of (6) together with the likelihood information $I = 9$ (1); this is, of course, the same as Fig. 1(*a*). Also plotted in Fig. 4(*a*) are the corresponding posterior p.d.f.'s for Jeffreys prior (dotted line) and the two Wilson priors given $\langle I \rangle = 20$ (acentric = dashed; centric = dots and dashes). It demonstrates that, for a well determined reflection, where $I_0 >> \sigma_I$, the strength of the evidence of the data leads to the same conclusions irrespective of the prior state of knowledge. The conditions for Fig. 4(*b*) are the same as those of Fig. 4(*a*) except that $\langle I \rangle = 1$ (instead of 20). The posterior p.d.f.'s for the Wilson priors have changed slightly but are still largely determined by the evidence of the current data. The case for reflection that is poorly determined by the experimental measurements does, however, give different results.

Fig. 4(*c*) shows the results for $I = 1$ (9), with $\langle I \rangle = 20$ for the Wilson priors. Although the posterior probabilities for the naive and the Wilson priors are different in detail, they are all consistent to within the large uncertainties they represent. If, however, $\langle I \rangle$ is 1, as in Fig. 4(*d*), then the posterior p.d.f.'s for the Wilson priors become significantly narrower than that for the naive prior. Although the prior knowledge that $\langle I \rangle = 20$ is more inconsistent with the likelihood information than $\langle I \rangle = 1$, the latter has a greater effect on the posterior probability. This shows that the prior information that a reflection is expected to be weak is far more powerful than knowledge that it is expected to be strong. This is easily confirmed by noticing that the Wilson prior of (21) goes asymptotically to the form of the naive

prior of (6) as the expectation value of the intensity becomes very large.

For the case of Figs. 4(*c*) and (*d*), the Jeffreys prior leads to quite a different posterior p.d.f. for the amplitudes as compared with the naive or Wilson priors. There are two reasons why the use of a Jeffreys prior is not recommended. Firstly, it has a simple pole singularity at the origin that necessitates the imposition of a lower cut-off $I_{\min}$ for it to be handled properly. Although the lower cut-off has little effect for well measured reflections, the value of $I_{\min}$ affects significantly the 95% confidence interval for the amplitudes of poorly determined intensities. Secondly, any subsequent use of the inferred amplitudes and their estimated reliabilities to solve the structure of the crystal tends to involve a least-squares minimization. Since a least-squares analysis will implicitly assume that the posterior p.d.f. for the amplitudes is Gaussian, it will often be a bad approximation for the Jeffreys prior (or, indeed, for the Wilson centric prior).

This last point actually leads us to recommend the naive prior of (6) as the best one from a practical viewpoint. If the intensity of the reflection is well determined by the data, then the form of the prior is largely irrelevant. If it is poorly determined, then the benefits of using the most appropriate prior can easily be wiped out by the crudity of the subsequent Gaussian approximation to a very non-Gaussian posterior probability. Even though the naive positivity prior may not be optimal, it is almost always adequate since the resultant posterior probability tends to be suffciently broad to encompass the other options. If a different prior is to be used, then the Wilson prior of (21) is a good one because it retains all the practical benefits of the simple prior of (6).

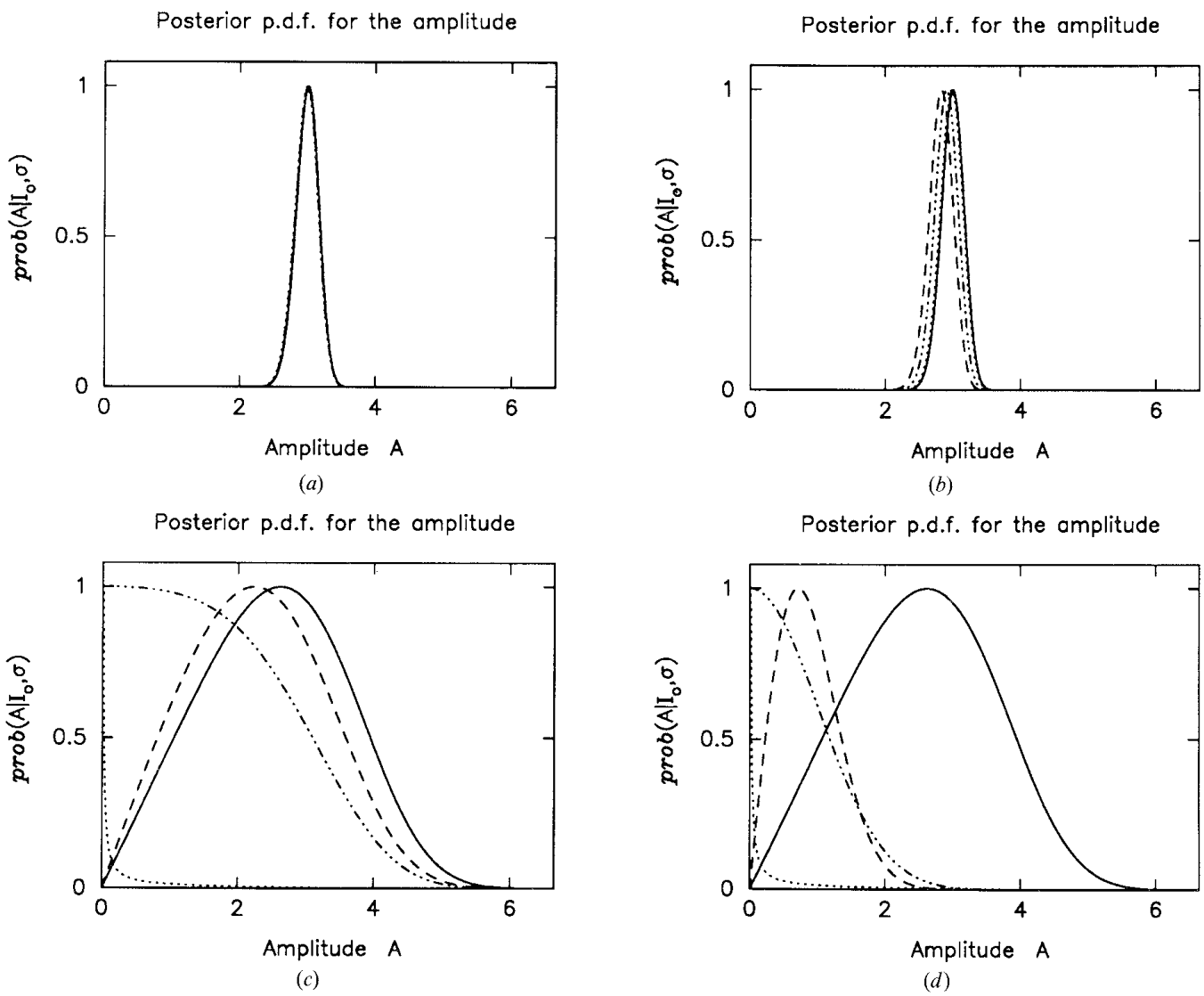

Fig. 4. The posterior probabilities for amplitudes given the best-fit intensities (mean and estimated standard devaiation): (*a*) and (*b*) $I = 9\ (1)$, (*c*) and (*d*) $I = 1\ (9)$. The full line is the Bayesian solution (with a naive positivity prior), the dotted line is the corresponding posterior for the Jeffreys prior. The dashed and dash–dotted lines correspond to the Wilson acentric and centric priors, respectively. In (*a*) and (*c*), for the Wilson priors, $\langle I \rangle$ is taken to be 20; in (*b*) and (*d*), $\langle I \rangle = 1$.

Finally, the discussion of practicality above has a bearing on the question: 'Is it better to refine crystal structures on intensities or amplitudes?' The real answer is that it should not matter at all if the analyses are carried out in a consistent fashion. In practice, however, the known positivity of the intensities is a difficult constraint to impose directly since one has to deal with p.d.f.'s that are truncated Gaussians. As has been seen, this awkward p.d.f. in intensity space maps on to a well behaved Gaussian-like p.d.f. in amplitude space. This suggests that, in cases of structure determination and the construction of Fourier maps, it should be better to use the inferred amplitudes since then the ubiquitous least-squares procedure automatically incorporates the positivity constraint.

6. Worked example and concluding remarks

In this section, the Bayesian analysis outlined above is applied to a high-resolution neutron powder diffraction data set of benzene collected on the HRPD at ISIS. The observed, calculated and difference data used in this example are shown in Fig. 5 and range from 0.91 to 1.11 Å. Despite the narrow d-spacing range used here (the full data set lies between 0.70 and 2.02 Å), there are 140 crystallographically distinct reflections. The data were initially refined using the Rietveld method to determine the correct scale factor for normalizing the structure factors. Integrated intensities were subsequently extracted using the Pawley method. Of the 140 reflections, 108 were nonoverlapping; the remaining 32 were overlapped in 16 pairs. The largest separation between pairs of completely overlapped (*i.e.* unresolvable) reflections was 0.00003 Å ($\Delta d/d = 3 \times 10^{-5}$); the smallest separation between reflections that were deemed to be not completely overlapping was 0.00005 ($\Delta d/d = 5 \times 10^{-5}$). This very small separation is an order magnitude less than the peak width; not surprisingly, this leads to extremely high correlations between the refined peak intensities. This is illustrated graphically in Fig. 6 which shows the magnitude of the percentage correlation matrix. The refined parameters are 140 peak intensities, grouped into 124 clumps, and two parameters for a straight-line background; cell constants and peak-width parameters were fixed at previously refined values. The dominance of the diagonal and near-diagonal elements indicates that only neighbouring intensities are correlated with one another. From the top two rows and the final two columns, it is also noticeable that there is little correlation between the individual peak intensities and the two background parameters (the percentage correlations are around 15% in each case). This lack of correlation between sharp Bragg peaks and slowly varying backgrounds is a general feature of the refinement of powder diffraction data.

Results from two selected ranges are presented in Table 1. It is clear that, when correlations are less than 80%, the standard least-squares analysis gives presentable results. However, for higher cross correlations, the standard Pawley method leads to problems. Take, for example, the 455 and 027 reflections, which are −100% correlated. In effect, the only good refined quantity is the sum of the two intensities. This is evident from refined intensity values. The conventional approach gives an estimated value for the structure-factor amplitude, A, of 7.0 (11) for 455 and an undetermined value for 027. Despite the high

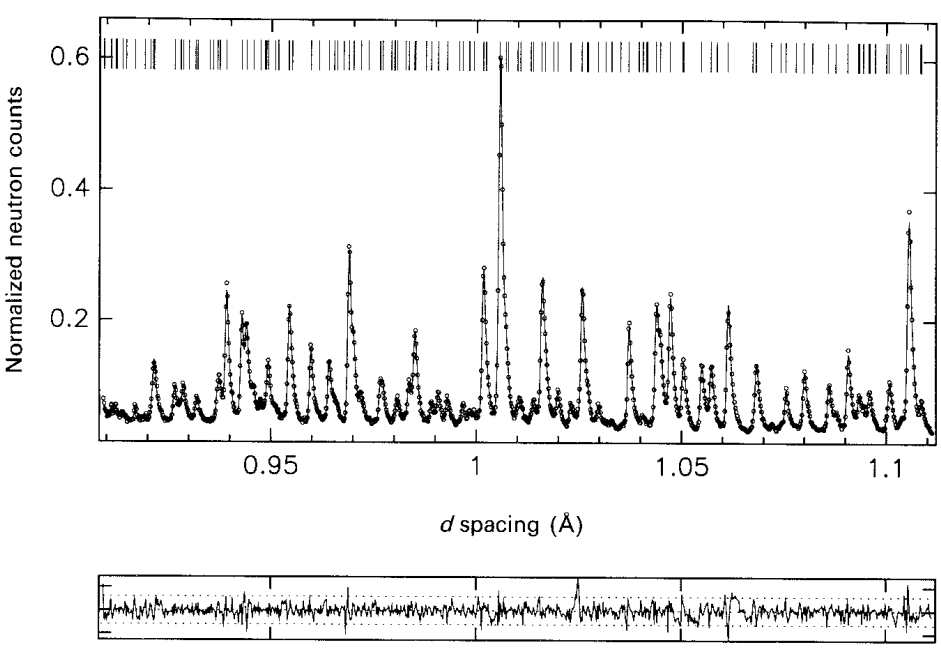

Fig. 5. Observed (dots) and calculated (full line) diffraction pattern of benzene obtained on the High Resolution Powder Diffractometer, HRPD, at ISIS. The refined data range fom 0.91 to 1.11 Å and contain 140 reflections that are marked by tick marks. The quality of fit is good (reduced $\chi^2 = 2.16$ for 126 parameters and 1004 observations) as is evidenced by the difference/estimated standard deviation plot at the bottom of the figure. (The dotted lines denote a difference/estimated standard deviation of ±3.)

Table 1. *Selected extracted structure-factor amplitudes from benzene*

Columns 1–3 contain the Miller indices; column 4 contains the d spacing in Å. Columns 5 and 6 contain the intensities refined using a standard Pawley-type least-squares analysis, along with their estimated standard deviations. Columns 7–12 contain the percentage cross correlations between neighbouring intensities. For example, 426 and 193 are −51% correlated; 0,10,0 and 027 are +64% correlated. Columns 13 and 14 contain the structure-factor amplitudes and errors obtained using the conventional analysis outlined in § 2. Columns 15 and 16, on the other hand, contain the structure-factor amplitudes and errors obtained using the Bayesian approach outlined in this paper. The subsequent five columns contain the percentage cross correlations between neighbouring amplitudes. The final column is the true (signed) structure factor from the known crystal structure of benzene.

h	k	l	d spacing	I_0	σ_I							$I_0^{1/2}$	$\sigma_{I^{1/2}}$	A	σ_A						F_{calc}
4	2	6	0.93482	**−0.26**	**0.85**	−51	6	−2	2	0	0	–	–	**0.61**	**0.36**	−21	0	0	0	0	**−0.57**
1	9	3	0.93567	**0.40**	**0.92**	−30	10	−9	1	−1	0	**0.63**	**0.73**	**0.77**	**0.38**	0	0	0	0	0	**−0.95**
0	10	0	0.93676	**63.54**	**7.16**	−68	64	−7	6	0	0	**8.08**	**0.44**	**8.66**	**0.33**	−37	20	0	0	0	**−8.43**
4	5	5	0.93727	**48.27**	**15.39**	−100	16	−14	0	0	0	**6.95**	**1.11**	**3.96**	**0.56**	−87	0	0	0	0	**4.29**
0	2	7	0.93733	**−56.86**	**29.09**	−17	15	0	0	0	0	–	–	**2.36**	**1.59**	0	0	0	0	0	**0.09**
1	7	5	0.93892	**42.62**	**12.13**	−100	2	0	0	0	0	**6.53**	**0.93**	**6.45**	**0.89**	−99	0	0	0	0	**6.53**
3	6	5	0.93898	**47.06**	**12.01**	−2	0	0	0	0	0	**6.86**	**0.88**	**6.85**	**0.84**	0	0	0	0	0	**−6.77**
6	3	4	0.94268	**71.07**	**1.12**	−36	2	0	0	0	0	**8.43**	**0.07**	**8.43**	**0.07**	0	0	0	0	0	**8.32**
1	1	7	0.94386	**24.69**	**0.55**	−100	−23	1	0	0	0	**4.97**	**0.06**	**4.97**	**2.49**	−100	0	0	0	0	**6.99**
5	3	5	0.94386	**24.69**	**0.55**	−23	1	0	0	0	0	**4.97**	**0.06**	**4.97**	**2.48**	0	0	0	0	0	**1.66**
4	0	6	0.95402	**57.37**	**20.09**	−96	77	−9	0	0	0	**7.57**	**1.33**	**4.30**	**0.89**	−60	−19	24	0	0	**−2.42**
6	6	1	0.95410	**−32.31**	**16.95**	−91	21	0	0	0	0	–	–	**1.79**	**1.20**	−49	7	0	0	0	**0.73**
7	3	2	0.95423	**81.48**	**8.22**	−42	0	0	0	0	0	**9.03**	**0.46**	**8.09**	**0.25**	−50	0	0	0	0	**−8.65**
3	8	3	0.95498	**7.28**	**1.22**	−1	1	0	0	0	0	**2.70**	**0.23**	**2.82**	**0.21**	0	0	0	0	0	**−2.23**
2	9	2	0.95948	**39.34**	**8.16**	−99	9	0	0	0	0	**6.27**	**0.65**	**6.13**	**0.60**	−98	0	0	0	0	**6.27**
0	8	4	0.95954	**20.04**	**16.11**	−10	0	0	0	0	0	**4.48**	**1.80**	**4.88**	**1.48**	0	0	0	0	0	**4.35**
3	7	4	0.96151	**2.59**	**0.69**	−8	1	−1	0	0	0	**1.61**	**0.21**	**1.64**	**0.21**	0	0	0	0	0	**1.61**
6	6	0	0.96394	**82.33**	**1.76**	−34	8	−1	0	0	0	**9.07**	**0.10**	**9.08**	**0.10**	0	0	0	0	0	**9.04**

correlation and negative refined intensity, the Bayesian analysis gives a meaningful result. The final values, however, must be treated with caution. Firstly, there is still a high correlation (−87%) between the two amplitude values. The positivity constraint imposed by the Bayesian analysis generally reduces cross correlation but by no means eliminates it. Secondly, the refined amplitude for 027 is substantially higher than the true value. This will always be the case when peaks are severely overlapped and the resulting errors are large. In general, strong peaks may be accurately determined; weak peaks are often overestimated although the associated error is usually sufficient to cover the discrepancy. Where peak overlap is complete (*e.g.* 117 and 535), no advantages are offered by the Bayesian analysis.

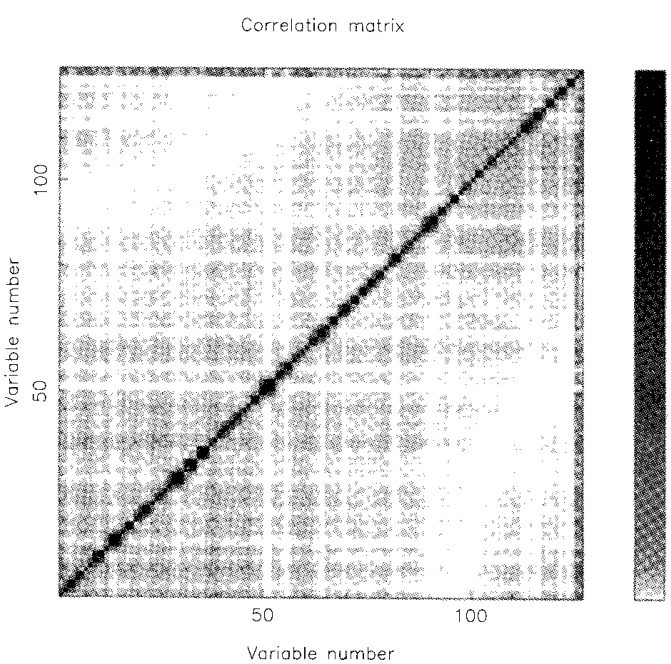

Fig. 6. A two-dimensional grey-scale representation of the magnitude of the percentage correlation matrix for the 126 refined parameters; 124 of these are associated with the 140 Bragg intensities and the remaining two are the background parameters (indicated by the two top rows and rightmost columns).

The results of the top and bottom parts of Table 1 are shown graphically in Figs. 7 and 8, respectively. In both diagrams, (*a*) represents the outcome of the least-squares analysis; both show occurrences of negative intensities in highly overlapped regions. The results of the subsequent Bayesian analysis are in close agreement with the calculated intensities given by the refinement of the crystal structure; this is illustrated in parts (*b*) and (*c*) of Figs. 7 and 8.

The Bayesian presentation in this paper is consistent with that given earlier by French & Wilson (1978). The main difference is that they only considered the case of isolated reflections, whereas in this paper the analysis is generalized to include situations of peak overlap common in powder diffraction. This extension is made easier by an alternative choice for approximating the posterior p.d.f: in this paper, it is described as a Gaussian, the parameters of which are given by the maximum of the posterior and its curvature at that point; French & Wilson also approximate to a Gaussian distribution but estimate its mean and variance from the first and second moments of the posterior. If the posterior p.d.f. were

really a Gaussian, the two procedures would give identical results; in practice, they give very similar results. The advantage of the current procedure is that it leads to a more efficient algorithm, which generalizes easily to the case of correlated intensities. For example, the answer can be calculated analytically for isolated reflections; by contrast, French & Wilson have to calculate the moments numerically by interpolating from a stored table.

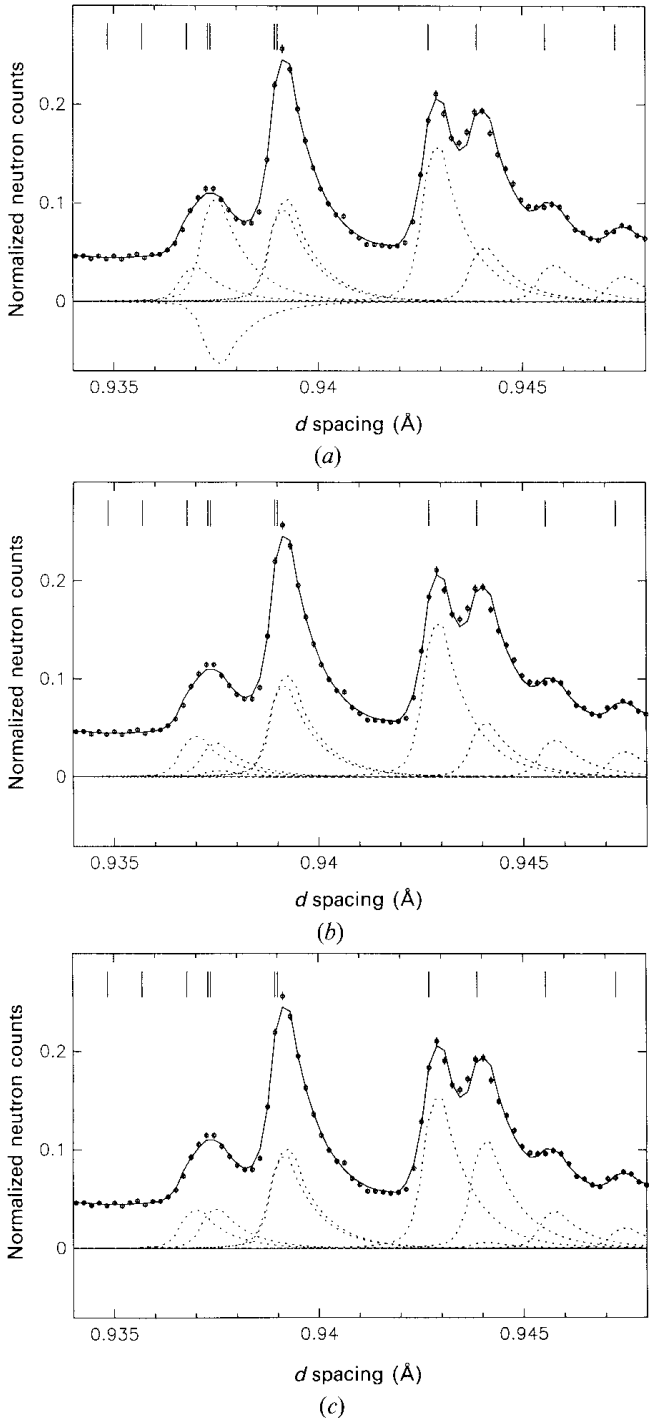

Fig. 7. The observed and calculated powder diffraction pattern of benzene between 0.934 and 0.948 Å. (*a*) The results of least-squares fitting using the Pawley method. (*b*) The square of individual inferred structure-factor amplitudes from the subsequent Bayesian analysis. (*c*) The calculated intensities given by the refinement of the crystal structure. [The apparent disagreement at 0.944 Å is because in (*a*) and (*b*) there are two completely overlapping reflections whose intensities are presumed to be equal.]

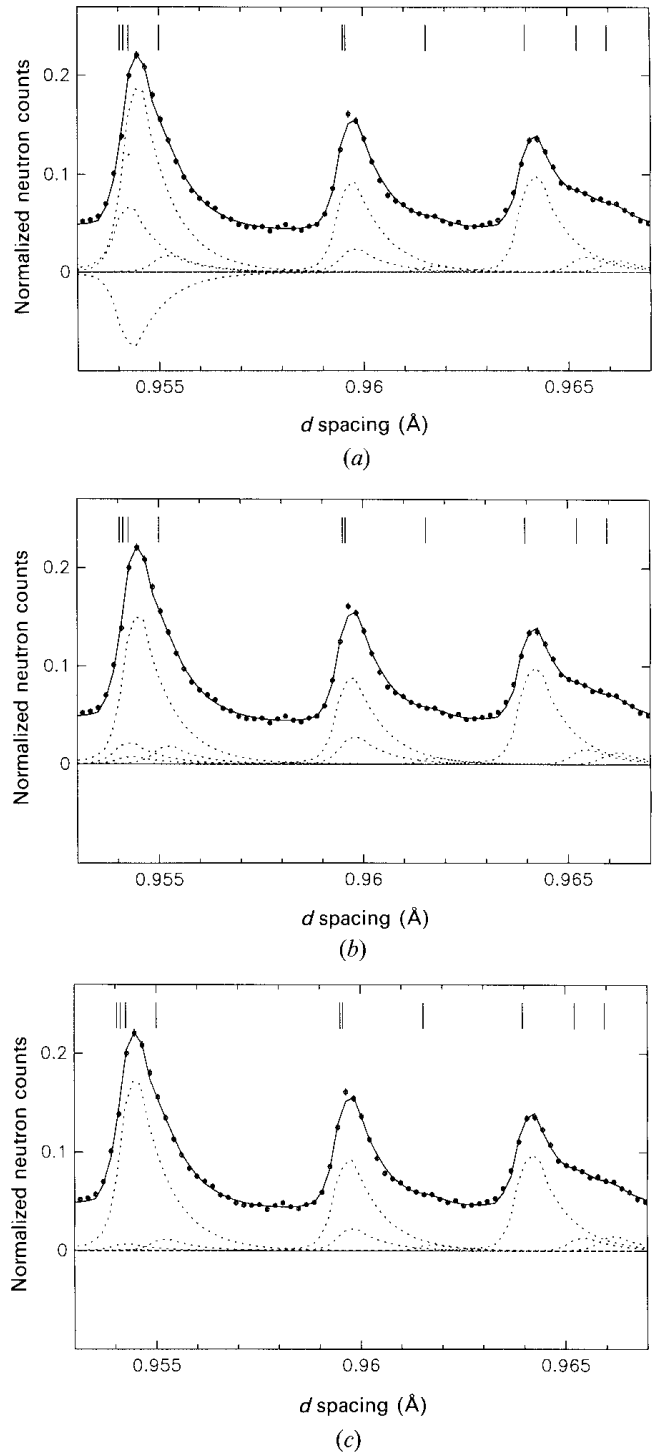

Fig. 8. The observed and calculated powder diffraction pattern of benzene between 0.953 and 0.967 Å. (*a*) The results of least-squares fitting using the Pawley method. (*b*) The square of individual inferred structure-factor amplitudes from the subsequent Bayesian analysis. (*c*) The calculated intensities given by the refinement of the crystal structure.

APPENDIX

Maximum-entropy derivation of the Wilson distributions and their relationship to the Jeffreys prior

In the discussion in § 5, it was mentioned that complete ignorance about a scale parameter X is expressed through the assignment of a Jeffreys (1939) prior:

$$\text{prob}\,(X) \propto 1/X, \quad \text{for } X > 0. \qquad (A1)$$

This peculiar form looks less strange once it is realized that it is equivalent to a p.d.f. that is uniform with respect to $\log(X)$. Technically, a suitable range between $X_{\min}$ and $X_{\max}$ should be defined so that $(A1)$ can be normalized; in practice, however, it often does not need to be specified explicitly as the posterior p.d.f. is usually well behaved even when the limits of 0 and ∞ are considered. If the results depend strongly on the range then, in that case, probability theory is warning that the prior knowledge is at least as important as the data! A Jeffreys prior would therefore be assigned for the magnitude of a structure factor if there is complete ignorance about its scale; in this case, it doesn't matter whether the amplitude A or the intensity I is used, since it is the requirement for such consistency that leads to $(A1)$. This still leaves the question of which p.d.f. should be used if there is some cogent information.

Jaynes (1983) has suggested that the principle of maximum entropy (MaxEnt) should be used to assign p.d.f.'s when one is given some *testable* information. That is to say, prob (X) should be chosen by maximizing its entropy S:

$$S = -\int \text{prob}\,(X) \ln[\text{prob}\,(X)/m(X)]\,\mathrm{d}X, \qquad (A2)$$

where $m(X)$ is a suitable measure over the space of possibilities for X, subject to normalization and any other testable constraints available. In the context of Wilson distributions, let it be assumed that the average value of μ of the intensity of a reflection is known; in terms of prob (I), this is given by

$$\langle I \rangle = \int_0^\infty I\,\text{prob}\,(I)\,\mathrm{d}I = \mu, \qquad (A3)$$

or, in terms of prob (A), can be written as

$$\langle A^2 \rangle = \int_0^\infty A^2\,\text{prob}\,(A)\,\mathrm{d}A = \mu. \qquad (A4)$$

For the measure, let a simple uniform assignment be made with respect to the real and imaginary parts of the structure factor; this means that $m(X,Y)$ is a constant in the Argand plane, where the complex structure factor Z is given by $Z = X + iY$. Since the probability that the amplitude of a general reflection will lie between A and $A + \delta A$ is proportional to the area, $2\pi A\delta A$, of a thin ring of radius A, this gives a measure $m(A)$ that is proportional to A. If the reflection is also known to be centric, then the corresponding probability is proportional to the element of area $2\delta A$; consequently, $m(A)$ is a constant. With a change of variable from A to I (where $I = A^2$), a uniform measure can be assigned in both cases:

$$\left.\begin{matrix} m(A) \\ m(I) \end{matrix}\right\} = \text{constant} \begin{cases} \text{for a centric reflection} \\ \text{otherwise} \end{cases} \qquad (A5)$$

By the method of Lagrange multipliers, the entropy in $(A2)$ can be maximized, with a constant measure, subject to the constraint of $(A3)$ or $(A4)$ as appropriate; this yields the following MaxEnt assignments for the p.d.f.'s:

$$\text{prob}\,(I|\mu) = (1/\mu)\exp(-I/\mu), \quad \text{for } I > 0 \qquad (A6)$$

and

$$\text{prob}\,(A|\mu,\ \text{centric}) = (2/\pi\mu)^{1/2}\exp(-A^2/2\mu), \quad \text{for } A > 0. \qquad (A7)$$

Equation $(A6)$ is, of course, just the Wilson distribution for an acentric reflection; $(A7)$ is its centric counterpart and can be recognized as such when it is transformed into a p.d.f. for the intensity:

$$\text{prob}\,(I|\mu,\ \text{centric}) = (2\pi\mu I)^{-1/2}\exp(-I/2\mu), \quad \text{for } I > 0. \qquad (A8)$$

Now, suppose the magnitude of the average intensity was not known (because the data were not on an absolute scale, for example). Then rather than the conditional p.d.f.'s of $(A6)$ and $(A8)$, the *marginal* distributions prob (I) and prob $(I|\text{centric})$ would be required. For the case of the general reflection, this means that the following integral must be evaluated:

$$\text{prob}\,(I) = \int_0^\infty \text{prob}\,(I,\mu)\,\mathrm{d}\mu = \int_0^\infty \text{prob}\,(I|\mu)\,\text{prob}\,(\mu)\,\mathrm{d}\mu, \qquad (A9)$$

where the product rule of probability theory has been used to express the joint p.d.f. prob (I, μ) in terms of the assignment of $(A6)$ and a prior for μ. Since the average value of the intensity is a scale parameter, the Jeffreys prior of $(A1)$ should be used to represent complete *a priori* ignorance about its size: prob $(\mu) \propto 1/\mu$. Carrying out the integration yields

$$\text{prob}\,(I) \propto 1/I, \quad \text{for } I > 0. \qquad (A10)$$

For the centric case, it is easiest to do the marginal integral in amplitude space:

$$\text{prob}\,(A|\text{centric}) = \int_0^\infty \text{prob}\,(A,\mu|\text{centric})\,\mathrm{d}\mu$$

$$= \int_0^\infty \text{prob}\,(A|\mu,\text{centric})\ \text{prob}\,(\mu|\text{centric})\mathrm{d}\mu. \quad (A11)$$

With the MaxEnt assignment of ($A7$) and a Jeffreys prior for prob $(\mu|\text{centric})$, because the knowledge that a reflection is centric says nothing about its expected value, the integral of ($A11$) yields

$$\text{prob}\,(A|\text{centric}) \propto 1/A, \quad \text{for } A > 0. \quad (A12)$$

Equation ($A12$) is, of course, equivalent to ($A10$) with a change of variables.

In conclusion, it can be seen that the Wilson distributions can easily be derived with the maximum-entropy principle. The acentric case arises from the imposition of a constraint on the expected value of the intensity of a general reflection, whereas the centric p.d.f. requires the additional knowledge that the structure factor must be real. If there is complete uncertainty as to the absolute scale of the data then, as expected, both the Wilson distributions revert to the form of the Jeffreys prior.

References

ANTONIADIS, A., BERRUYER, J. & FILHOL, A. (1990). *Acta Cryst.* **A46**, 692–711.

BRICOGNE, G. (1991). *Acta Cryst.* **A47**, 803–829.

DAVID, W. I. F. (1987). *J. Appl. Cryst.* **20**, 316–319.

DAVID, W. I. F. (1990). *Nature (London)*, **346**, 731–734.

ESTERMANN, M. A., MCCUSKER, L. B. & BAERLOCHER, C. (1992). *J. Appl. Cryst.* **25**, 539–543.

FRENCH, S. & WILSON, K. (1978). *Acta Cryst.* **A34**, 517–525.

GILMORE, C. G., HENDERSON, K. & BRICOGNE, G. (1991). *Acta Cryst.* **A47**, 830–841.

JANSEN, J., PESCHAR, R. & SCHENK, H. (1992). *J. Appl. Cryst.* **25**, 231–236, 237–243.

JAYNES, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, edited by R. D. ROSENKRANTZ. Dordrecht: Reidel.

JEFFREYS, H. (1939). *Theory of Probability*. Oxford Univ. Press.

LE BAIL, A., DUROY, H. & FOURQUET, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.

NELDER, J. A. & MEAD, R. (1965). *Comput. J.* **7**, 308–313.

OATLEY, S. & FRENCH, S. (1982). *Acta Cryst.* **A38**, 537–549.

PAWLEY, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes: the Art of Scientific Computing*. Cambridge Univ. Press.

WILSON, A. J. C. (1949). *Acta Cryst*, **2**, 318–321.