reflections and the average $|E|$ value may be used as a criterion to select those matrices which on refinement are expected to yield good phases. However, Fig. 3 shows a few matrices of high average $|E|$ value not giving good starting sets at all.

Remarkably, the number of unobserved reflections does not play a significant role. The table shows very good starting sets obtained from matrices containing a large number of unobserved reflections.

The influence of the number of symmetry-equivalent reflections is not very clear. Fig. 4 contains no evidence of good matrices being found for particular values of the ratio between dependent and independent reflections only.

Clearly, the quality of the best matrices produced using the new algorithm far exceeds those from earlier attempts. For all three structures, large starting sets – over 30 reflections – could be generated with very low average phase errors (see Table 1).

Although memory requirements are considerable (2–3 Mbyte), the construction algorithm is very fast. Using a MicroVAX II, the construction takes 1–3 min only.

It is interesting to note that not until the determinant was maximized did the phases of the starting set bear any relation whatsoever to the true phases. This illustrates the validity of the generalized maximum determinant rule (Tsoucaris, 1970; Karle, 1970; Heinerman *et al.*, 1979).

### References


Graaff, R. A. G. de & Vermin, W. J. (1982). *Acta Cryst.* A**38**, 464–470.

Heinerman, J. J. L., Kroon, J. & Krabbendam, H. (1979). *Acta Cryst.* A**35**, 101–105.

Karle, J. (1970). *Proc. Natl Acad. Sci. USA*, **75**, 2545–2548.

Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181–187.

Main, P. (1975). In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 165–175. Copenhagen: Munksgaard.

Taylor, D. J., Woolfson, M. M. & Main, P. (1978). *Acta Cryst.* A**34**, 870–883.

Tsoucaris, G. (1970). *Acta Cryst.* A**26**, 492–499.

Vermin, W. J. & de Graaff, R. A. G. (1978). *Acta Cryst.* A**34**, 892–894.

Yao Jia-xing (1981). *Acta Cryst.* A**37**, 642–644.


*Acta Cryst.* (1990). A**46**, 692–711

# Maximum-Likelihood Methods in Powder Diffraction Refinements


By Anestis Antoniadis and Jacques Berruyer

*University of St Etienne, Department of Mathematics*, 23, *Rue du Docteur Paul Michelon*, 42100 *St Etienne CEDEX, France*

and Alain Filhol

*Institut Max von Laue–Paul Langevin*, 156*X*, *Centre de Tri*, 38042 *Grenoble CEDEX, France*


(*Received* 15 *November* 1989; *accepted* 4 *April* 1990)


## Abstract

The validity of least-squares procedures commonly used nowadays for the analysis of single-crystal, X-ray and neutron diffraction data is examined. An improved methodology that rests on sound statistical theory is proposed and turns out to be a fruitful way to consider any crystallographic refinement. A maximum-likelihood estimation procedure is developed for Poisson regression models. Measures of the goodness of fit (other than the $R$ factor), generalized residuals and diagnostic plots are described. Confidence regions and intervals are also discussed. A set of measures of the influence of data on the fit and the parameter estimates is obtained for Poisson statistics. Finally, the effect of under or over dispersion of the data randomness with respect to a true Poisson distribution is considered and model-independent estimates of this dispersion are discussed.


### General notation and symbols frequently used

| | |
|---|---|
| $\eta, \theta, \ldots$ | Lower case greek *italics* denote scalar parameters. |
| $\boldsymbol{\eta}, \boldsymbol{\theta}, \ldots$ | Lower case greek **bold** denote column-vector parameters. |
| $Y, X, \ldots$ | Upper case *italics* normally denote real random variables. |
| $y, x, \ldots$ | Lower case *italics* normally denote observed values of real random variables (realizations). |
| $\boldsymbol{Y}, \boldsymbol{X}, \ldots$ | **Bold** upper case *italics* normally denote column random vectors with corresponding components $Y_i, X_i, \ldots$. |

0108-7673/90/080692-20$03.00 

| | |
|---|---|
| $y, x, \ldots$ | **Bold** lower case *italics* normally denote observations of random vectors with corresponding components $y_i, x_i, \ldots$. |
| $\mathbf{X}, \mathbf{A}, \ldots$ | **Bold** upper case roman letters normally denote matrices. |
| ${}^t\mathbf{X}$ | The transpose of matrix $\mathbf{X}$. |
| $\mathbf{X}^{-1}$ | The inverse of matrix $\mathbf{X}$. |
| $\hat{Y}, \hat{\boldsymbol{\eta}}, \ldots$ | A hat over a symbol or expression denotes a sample estimate of the corresponding parameter. |
| $\tilde{\boldsymbol{\beta}}$ | Also denotes a sample estimate but is used only for the minimum $\chi^2$ estimate of $\boldsymbol{\beta}$. |
| $\mathbb{E}_\theta(Y)$ | The expectation (population mean) of a random variable whose probability distribution is characterized by the parameter $\theta$. Whenever the subscript is obvious it is omitted. |
| $\mathrm{Var}_\theta(Y)$ | The variance of the random variable $Y$ for a probability distribution characterized by the parameter $\theta$. |
| $\sigma_\theta(Y)$ | The standard deviation of the random variable $Y$. |
| $\mathbb{E}_\theta(\boldsymbol{Y})$ | The expectation (population mean) of a random vector $\boldsymbol{Y}$. It is a column vector with components the expectations of the $Y_i$'s. |
| $\mathrm{Var}_\theta(\boldsymbol{Y})$ | The variance-covariance matrix of $\boldsymbol{Y}$. |
| $\mathbb{R}^p$ | The set of $p$-dimensional column vectors. |
| $\mathcal{O}(n^{-\alpha})$ | A quantity whose behavior is similar to that of $M/n^\alpha$ when $n \to \infty$, with $M$ a finite constant |
| $o(n^{-\alpha})$ | A quantity $M_n$ such that $\lim_{n\to\infty} M_n/n^\alpha = 0$. |
| $\Pr(A)$ | The probability of the event $A$. |

## 1. Introduction

From the data (counts with a *a priori* Poisson-like distribution) of a diffraction experiment to the final model (the crystallographic structure), there are usually several steps such as: information summary (data reduction), a set of corrections and a least-squares refinement of the parameters of the appropriate model. These steps may be performed independently or all together within a single refinement procedure, as in the case of the Rietveld profile refinement method, a very popular method for powder diffraction analysis, originally developed by Rietveld (1969).

In this method, the step-scan diffraction pattern is directly fitted, point by point, to a model pattern in the form of a background contribution and a set of diffraction peaks. The latter contribution is derived from a set of structural (positional and displacement), textural and instrumental parameters. Assuming a known functional shape for each of the $p$ Bragg peaks in the pattern and following Rietveld's notation, an observed step count $y_i$ for a counting time $T$ at an angle $t_i$ may be considered as the realization of a Poisson-distributed random variable $Y_i$ (randomness by counting statistics errors) whose mean $\mathbb{E}(Y_i) = n_i$ can be written as

$$\eta_i = \mathbb{E}(Y_i) = \sum_{k=1}^{p} T.\, I_k f_k(t_i; \theta_k) + T.\, B(t_i), \quad (1)$$

where $B(t)$ is a smooth function of $t$ representing the background per unit of time and $f_k(t; \theta_k)$ models the shape of the $k$th reflection. The intensity $I_k$ and the parameters $\theta_k$ depend usually on a set of structural parameters, say $\boldsymbol{\beta}$, which are of direct interest.

The problem is thus to find a set of parameters $\hat{\boldsymbol{\beta}}$ explaining as well as possible the data through the postulated model (1). This is usually done by means of a weighted least-squares procedure. However, when studies are made for the validity of the results (*e.g.* Hill & Madsen, 1984) a number of problems appears such as unexpectedly large values of goodness-of-fit statistics in some data sets, unrealistic estimates of standard deviations in the parameters *etc.* This leads to the practical conclusion that *improving the data statistics further and further* (*larger counting times*) *may lead to* '*worse*' *results*, which, abruptly said, is apparently in contradiction with classical statistical theory.

The present work is devoted to a complete discussion of existing statistical methods used for the analysis of powder or single-crystal diffraction data (fitting of rocking curves) and concerns a class of statistical regression models that allow the experimentalist to extract the best possible information from his experiment.

Estimation proceeds by defining a measure of discrepancy between the data and a corresponding set of fitted values generated by the model. In what follows, the maximum-likelihood principle is used to obtain estimates of the parameters in a regression model when the experimental errors are assumed to follow a Poisson or a Poisson-like distribution. A crucial point is concerned with the actual maximization of the likelihood function, which in many cases has to be performed using numerical methods, often of iterative character. A common way for the calculation of the maximum-likelihood estimates is to use the *method of scoring* (to be seen later) due to Fisher (1922). It is shown, in our context, that the method of scoring can be performed as an iterative reweighted least-squares procedure which differs from frequently used standard least-squares fitting algorithms.

Minimization with respect to $\boldsymbol{\beta}$ of the following weighted sum of squares

$$\chi^2 = \sum_{i=1}^{n} [y_i - \eta_i(\boldsymbol{\beta})]^2 / \eta_i(\boldsymbol{\beta}) \quad (2)$$

leads to a weighted least-squares estimate of the unknown parameters. In Rietveld's refinement

method (known also as *minimum modified* $\chi^2$ *method* in the statistical literature), the estimator $\hat{\boldsymbol{\beta}}$ of the unknown parameter column vector $\boldsymbol{\beta}$ is the one that minimizes

$$\chi^2_{\text{mod}} = \sum_{i=1}^{n} [y_i - \eta_i(\boldsymbol{\beta})]^2 / y_i \qquad (3)$$

instead of (2). When the counts are large and the structural model is 'smooth' enough [*i.e.* when $\boldsymbol{\eta}(\boldsymbol{\beta})$ as a function of $\boldsymbol{\beta}$ is at least second-order differentiable], it can be proved that the least-squares and the modified minimum $\chi^2$ estimates have similar behavior.

When $T$ is large, then $\mathbb{E}(Y_i)$ is large for all $i$ and the distribution of $Y_i$ may be approximated by a Gaussian random variable whose mean and variance is $\eta_i$. Further assumption that the measured profiles $y_i$ are realizations of stochastically independent $Y_i$'s leads to a likelihood function given by (see *e.g.* Rao, 1973)

$$\frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} \eta_i^{-1/2} \exp[-\tfrac{1}{2}(y_i - \eta_i)^2 / \eta_i]. \qquad (4)$$

In using as a measure of discrepancy the deviance $-[\ell(\boldsymbol{\eta}, y) - \ell(y, y)]$ where $\ell(y, y)$ is the maximum log-likelihood achievable for an exact fit in which fitted values equal data, one obtains for the previous Gaussian approximation [assuming also that $\ln(\eta_i / y_i)$ are negligible],

$$-[\ell(\boldsymbol{\eta}, y) - \ell(y, y)] \simeq \sum_{i=1}^{n} (y_i - \eta_i)^2 / \eta_i.$$

Hence, minimization of the deviance with respect to the parameter set $\boldsymbol{\beta}$ is approximately equivalent to weighted least-squares refinement. Generally, because $\ell(y, y)$ does not depend on the parameters, minimizing the deviance or maximizing the log-likelihood $\ell(\boldsymbol{\eta}, y)$ are equivalent problems. Therefore, if the approximation that the errors of the observations are independent and normally distributed is reasonable, a properly weighted least-squares analysis leads to the same estimates of the parameters as R. A. Fisher's maximum-likelihood method (Fisher, 1922).

When the minimum number of counts in any profile point is large (large sample sizes), an ideal situation, the minimum $\chi^2$ method (Rietveld's least-squares algorithm) will produce estimates of the fit parameters with equivalent properties (asymptotic unbiasedness, first-order efficiency *etc.*) to those of the estimates obtained by applying the maximum-likelihood method. However, we demonstrate in this paper that this is no longer true for moderate to small sample sizes and the maximum-likelihood method is then the more accurate of the two.

As already stated above, a habitual source of concern to users of the Rietveld refinement method (Albinati & Willis, 1982) is that estimated standard deviations of the parameters (e.s.d. for short) are often unreliable, with a goodness-of-fit index unacceptably large (overdispersion). Our framework includes a class of regression families that allow the analyst to model this overdispersion and to include it as part of the fit. It also provides some informative diagnostic measures for the fit. These measures turn out to be useful in identifying observations that are not well explained by the model, as well as those dominating some important aspect of the fit. Computational issues and applications of our method in the analysis of real and simulated data are presented.

## 2. Maximum likelihood and least squares

This section emphasizes the basic statistical notions that will be used later. The theory developed here is not original to this article. Readers of the literature on least-squares refinement and maximum-likelihood estimation will find most of the ideas familiar, though stated from a different point of view.

In the refinement of a crystal structure, one assumes a structural model with variable parameters, the values of which are chosen so as to achieve the 'best' agreement between the calculated and the observed data. More generally, the aim in model fitting is to replace the observed data set $y$ with a set of fitted values $\hat{y}$ derived from a model, these fitted values being as 'close' as possible to the data values. To do so requires some measure of discrepancy to be defined between the components $y_i$ of $y$ and the $\hat{y}_i$'s, and this definition requires certain assumptions to be made about the variation in the $y_i$'s that is not accounted for by the model. Standard weighted least-squares refinement chooses $S^2 = \sum_i w_i (y_i - \hat{y}_i)^2$ as the measure of discrepancy.

This formula has two implications: firstly, the simple summation of the individual terms $w_i(y_i - \hat{y}_i)^2$, each depending on only one observation, implies that the measurements are independent in some sense; secondly, the use of the weight factors, $w_i$, implies that the observations may be of varying precision which may or may not depend on the components of $\hat{y}$. If we model the unaccounted variation in statistical terms, the first property becomes stochastic independence, and the second property is interpreted by requiring that the variance of the distribution of deviations be proportional to the inverse of the weights. Moreover, the use of $S^2$ implies the normal (or Gaussian) frequency distribution for each component $\varepsilon_i = Y_i - \eta_i$ of the residual variation $\varepsilon$, in which the frequency of $Y_i$ at $y_i$ given $n_i$ is proportional to

$$\exp[-w_i(y_i - \eta_i)^2], \qquad (5)$$

where $1/2w_i$ is the variance of the distribution.

We can look at (5) in two ways. If we regard it as a function of $y_i$ for fixed $\eta_i$, (5) specifies the probability distribution of the observations. Alternatively, for

a given observation $y_i$ we may consider (5) as a function of $n_i$ giving the relative plausibility of different values of $\eta_i$ for the particular observed value $y_i$ of $Y_i$. In the latter form it becomes the likelihood function whose maximization with respect to $\eta_i$ leads to the maximum-likelihood estimation of $\eta_i$. Thus, if the distribution is normal, the methods of maximum likelihood and least squares give identical estimates. However, for non-normal distributions, sums of squares will no longer be appropriate measures of goodness-of-fit and least-squares estimation may be inappropriate.

The method to be developed in this paper considers both problems (least squares and maximum likelihood) but is particularly concerned with the maximum-likelihood principle. This will be clear from the subsequent discussion. Meanwhile, it is worth clarifying the various aspects of maximum-likelihood theory and its advantages over least squares.

Although the method of maximum likelihood (hereafter abbreviated MML) dates back to Bernoulli (1861), it is generally agreed that Fisher (1922) rediscovered it and set the stage for its general acceptance in the statistical world. Efron (1982) provides an excellent discussion on the maximum-likelihood principle. MML, as used in practice, is a theory for making specific point and interval estimates for unknown parameters, while in fact it is a data summarization process. More precisely, the maximum-likelihood method can be described as follows: given a family $\mathcal{F}$ of probability densities for $\boldsymbol{Y}$ characterized by a population parameter $\boldsymbol{\theta}$ in a parameter set $\Theta$ (*i.e.* our prior belief on the model that has generated the observed data),

$$\mathcal{F} = \{f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\},$$

we observe the data $\boldsymbol{y}$ for $\boldsymbol{Y}$. Let $\hat{\boldsymbol{\theta}}$ be the value of $\boldsymbol{\theta}$, assumed to exist, which maximizes the probability density $\boldsymbol{y} \to f_{\boldsymbol{\theta}}(\boldsymbol{y})$. The *maximum-likelihood summary* of the data, abbreviated MLS, is the density function (model) corresponding to $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$,

$$\text{MLS: } \hat{f} = f_{\hat{\boldsymbol{\theta}}}.$$

There are two important points here:

(*a*) The parameter '$\boldsymbol{\theta}$' as used here is only a name, and plays no role in the summarization process. Any other way of naming the members of $\mathcal{F}$ results in the same MLS $\hat{f}$, given the same data $\boldsymbol{y}$.

(*b*) The MLS is not a number or a vector, it is a probability distribution. We are summarizing a data set by a probability distribution.

Next, suppose that $\gamma(f)$ is a parameter (function of the unknown probability mechanism) we wish to estimate. The *maximum-likelihood estimate*, MLE for short, is the corresponding function of $\hat{f}$,

$$\text{MLE: } \hat{\gamma} = \gamma(\hat{f}). \tag{6}$$

Thus, once we have calculated the MLS $\hat{f}$, we have available the MLE for every possible parameter $\gamma(f)$ while this is not the case for least-squares estimation: take, for example, the estimation of $e^{-\lambda}$ when observing counts generated by a Poisson random variable of parameter $\lambda$. The least-squares estimate of $e^{-\lambda}$ is easily shown to be the relative frequency of zero counts observed within the sample, while the maximum-likelihood estimate is $e^{-\hat{\lambda}}$ where $\hat{\lambda}$ denotes the sample mean.

This automatic way in which the maximum-likelihood method produces estimates for even very complicated parameters reflects well the distinction between MLE and standard least-squares estimates and justifies MML's popularity. Of course, for the same reason, MLE can be non-optimal if the experimentalist has one specific estimation problem in mind (the reader may refer again to the Poisson example just cited above).

Finally note that the extended class of models presented hereafter includes as an important example the class of generalized linear models (GLIM) introduced by Nelder & Wedderburn (1972) and which are typically used to analyze linear exponential family regression models in biomedical sciences.

## 3. Maximum-likelihood estimation for Poisson regression models

In this section, we introduce a method of regression analysis for Poisson distributed data by fitting nonlinear regression models to the Poisson means using maximum-likelihood theory. Specifically, we consider the regression situation, where we observe independent Poisson variates $Y_1, Y_2, \ldots, Y_n$ with density functions proportional to

$$\Pr(Y_i = y_i) = f(y_i; \theta_i) = (y_i!)^{-1} \exp\{[y_i\theta_i - b(\theta_i)]\}, \quad i = 1, \ldots, n, \tag{7}$$

with $\theta_i = \ln[\eta_i(\boldsymbol{\beta})]$, where $\eta_i(\boldsymbol{\beta})$ denotes the $i$th Poisson mean, $n$ is the sample size and $\boldsymbol{\beta}$ is the $p$-dimensional parameter vector characterizing the theoretical pattern. In order to obtain the MLE for the unknown parameter vector $\boldsymbol{\beta}$ we must maximize with respect to $\boldsymbol{\beta}$ either the likelihood function or, which is more convenient, the logarithm $\ell[\boldsymbol{\eta}(\boldsymbol{\beta}), \boldsymbol{y}]$ of the likelihood function, given by

$$\ell[\boldsymbol{\eta}(\boldsymbol{\beta}), \boldsymbol{y}] = \sum_{i=1}^{n} \{y_i \ln[\eta_i(\boldsymbol{\beta})] - \eta_i(\boldsymbol{\beta})\}. \tag{8}$$

The maximum-likelihood equations for $\boldsymbol{\beta}$ based on (8) are quite simple. The *score vector* is defined by

$$\partial\ell[\boldsymbol{\eta}(\boldsymbol{\beta}), \boldsymbol{y}]/\partial\boldsymbol{\beta} = {}^t[\partial\ell/\partial\beta_1, \ldots, \partial\ell/\partial\beta_p], \tag{9}$$

where the superscript $^t$ denotes transposition.

Let $\mathbf{X}(\boldsymbol{\beta})$ denote the $n \times p$ matrix $\partial\boldsymbol{\eta}/\partial\boldsymbol{\beta}$, which we call the *local design matrix*. This design matrix

depends on the unknown parameters in a nonlinear fashion. We will assume that $\mathbf{X}(\boldsymbol{\beta})$ has full rank $p$ for all possible values of $\boldsymbol{\beta}$ (*i.e.* the model is identifiable). The maximum-likelihood estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by solving the score vector for zero. Hence, the likelihood equations take the form

$$\sum_{i=1}^{n} \frac{\partial \eta_i(\boldsymbol{\beta})}{\partial \beta_j}\left(\frac{y_i}{\eta_i(\boldsymbol{\beta})}-1\right) = \{{}^t\mathbf{X}(\boldsymbol{\beta})\mathbf{V}^{-1}(\boldsymbol{\beta})[\mathbf{y}-\boldsymbol{\eta}(\boldsymbol{\beta})]\}_j = 0, \qquad j=1,\ldots,p \tag{10}$$

or, in matrix notation,

$${}^t\mathbf{X}(\boldsymbol{\beta}) \,.\, \mathbf{u} = 0,$$

where $\mathbf{u}$ is the $n$-dimensional vector $\partial \ell/\partial \boldsymbol{\eta}$ with components given by $[y_i/\eta_i(\boldsymbol{\beta})-1]$.

Here, $V(\boldsymbol{\beta})$ denotes the $n \times n$ variance-covariance matrix of the observed data random vector $\boldsymbol{Y}$, which is diagonal by the independence assumption with diagonal entries the weights $\eta_i$.

Since the ML equations are generally nonlinear with respect to the unknown parameters, Fisher's method of scores (see *e.g.* Rao, 1973) is used to develop an algorithm to find a root $\hat{\boldsymbol{\beta}}$ of (10). The algorithm is defined as a Newton-Raphson-type algorithm, where the matrix of second derivatives of the log-likelihood function $\ell$ is replaced by a suitable approximation.

More precisely, the *standard Newton-Raphson method* for the iterative solution of (10) calls for evaluating $\mathbf{u}$, $\mathbf{X}(\boldsymbol{\beta})$ and the second derivatives of $\ell$ for an initial value of $\boldsymbol{\beta}$ and for solving the linear equations

$$(-\partial^2 \ell/\partial \boldsymbol{\beta}\,{}^t\partial \boldsymbol{\beta}')(\boldsymbol{\beta}^* - \boldsymbol{\beta}) = {}^t\mathbf{X}(\boldsymbol{\beta})\mathbf{u} \tag{11}$$

for an updated estimate $\boldsymbol{\beta}^*$. This procedure is repeated until convergence. Equation (11) is derived from the first two terms of a Taylor-series expansion for $\partial \ell/\partial \boldsymbol{\beta}$.

Note that

$$\left(\frac{-\partial^2 \ell}{\partial \boldsymbol{\beta}\,{}^t\partial \boldsymbol{\beta}}\right) = -\sum_{i=1}^{n} \frac{\partial \ell}{\partial \eta_i}\frac{\partial^2 \eta_i}{\partial \boldsymbol{\beta}\,{}^t\partial \boldsymbol{\beta}} - {}^t\mathbf{X}(\boldsymbol{\beta})\frac{\partial^2 \ell}{\partial \boldsymbol{\eta}\,{}^t\partial \boldsymbol{\eta}}\mathbf{X}(\boldsymbol{\beta}).$$

The matrix $-\partial^2 \ell/\partial \boldsymbol{\beta}\,{}^t\partial \boldsymbol{\beta}$, denoted by $\mathcal{J}(\boldsymbol{\beta})$, is called the observed information for $\boldsymbol{\beta}$ while $\mathcal{J}(\boldsymbol{\eta}) = -\partial^2 \ell/\partial \boldsymbol{\eta}\,{}^t\partial \boldsymbol{\eta}$ is the observed information for $\boldsymbol{\eta}$. By standard arguments one has

$$\mathbb{E}(\partial \ell/\partial \eta_i) = 0, \qquad i=1,\ldots,n$$

and it follows that

$$\mathcal{I}(\boldsymbol{\eta}) = \mathbb{E}[\mathcal{J}(\boldsymbol{\eta})] = \mathbb{E}(-\partial^2 \ell/\partial \boldsymbol{\eta}\,{}^t\partial \boldsymbol{\eta}) = \mathrm{Var}\,[\mathbf{u}(\boldsymbol{\eta})]$$

which is positive-definite.

*Fisher's scoring method* is defined as Newton-Raphson iteration with replacement of the Hessian $\mathcal{J}(\boldsymbol{\beta})$ by its expectation $\mathcal{I}(\boldsymbol{\beta})$ (at the current parameter values $\boldsymbol{\beta}$), whereas another method, called the *linearization method*, is defined by replacing the Hessian by ${}^t\mathbf{X}(\boldsymbol{\beta})\mathcal{J}(\boldsymbol{\eta})\mathbf{X}(\boldsymbol{\beta})$. With any of these approximations (11) becomes

$$[{}^t\mathbf{X}(\boldsymbol{\beta}) \,.\, \mathbf{A}_{\boldsymbol{\beta}} \,.\, \mathbf{X}(\boldsymbol{\beta})](\boldsymbol{\beta}^* - \boldsymbol{\beta}) = {}^t\mathbf{X}(\boldsymbol{\beta})\mathbf{u}, \tag{12}$$

where $\mathbf{A}_{\boldsymbol{\beta}}$ is either $\mathcal{I}(\boldsymbol{\eta})$ or $\mathcal{J}(\boldsymbol{\eta})$.

Rather than handle the numerical solution of (12) directly, note that they have the form of normal equations for a weighted least-squares regression: $\boldsymbol{\beta}^*$ solves the minimization of

$${}^t[\mathbf{A}_{\boldsymbol{\beta}}^{-1}\mathbf{u} + \mathbf{X}(\boldsymbol{\beta})(\boldsymbol{\beta}^* - \boldsymbol{\beta})]\mathbf{A}_{\boldsymbol{\beta}} \times [\mathbf{A}_{\boldsymbol{\beta}}^{-1}\mathbf{u} + \mathbf{X}(\boldsymbol{\beta})(\boldsymbol{\beta}^* - \boldsymbol{\beta})],$$

that is, it results from regressing $\mathbf{A}_{\boldsymbol{\beta}}^{-1}\mathbf{u} + \mathbf{X}(\boldsymbol{\beta})\boldsymbol{\beta}$ onto the columns of $\mathbf{X}(\boldsymbol{\beta})$ using weight matrix $\mathbf{A}_{\boldsymbol{\beta}}$. It is this treatment of the scoring method *via* least squares that we used for the maximum-likelihood method. To stabilize the numerical method, we do not accept the current $\boldsymbol{\beta}^*$, but rather use it to define a direction in which the likelihood increases. Thus, we used the iteration

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} + \alpha_r[{}^t\mathbf{X}(\boldsymbol{\beta}^{(r)}) \,.\, \mathbf{A}_{\boldsymbol{\beta}^{(r)}} \,.\, \mathbf{X}(\boldsymbol{\beta}^{(r)})]^{-1} \times {}^t\mathbf{X}(\boldsymbol{\beta}^{(r)})\mathbf{u}_r, \tag{13}$$

with the step length $\alpha_r > 0$ chosen by a linear search algorithm [Goldstein's (1965) test]. Of course we are not able to show in general that the iterative procedure will converge or, if it converges, that the obtained maximum is unique. This depends on the choice of the initial estimates.

Let us give now a final remark about the maximum-likelihood algorithm and its analogy with a weighted least-squares regression. In the Poisson regression model, the **weights are determined by the fit**, which should not be confused with a classical weighted least-squares problem where the **weights determine the fit**. For those readers who remain doubtful of this subtle difference in interpretation of the weights, we suggest moving the last point in Fig. 1 further and further to the right and observing the consequences.

We end this section by discussing a *convergence criterion* for the ML algorithm, since in the majority of refinement programs there exist only stopping criteria for convergence. Indeed, a vital part of any nonlinear maximum-likelihood algorithm is the test for convergence to the maximum solution. Such a test, or convergence criterion, consists of an indicator calculated at each iteration and a tolerance level such that convergence is declared when the indicator falls below the tolerance level. Many authors (see *e.g.* Bard, 1974; Draper & Smith, 1966; Ralston & Jennrich, 1978) writing about nonlinear least-squares refinement recommend relative change in the sum-of-squares convergence criteria, but they also state that there is no known criterion that is absolutely satisfac-

tory. Bates & Watts (1981) propose an orthogonality convergence criterion for classical nonlinear least-squares models. All but the last of these are more correctly described as *termination criteria*, since they merely indicate whether further iterations might be useful; they are not convergence criteria since they do not necessarily indicate whether a local maximum has been reached. On the other hand, orthogonality is an absolute indicator of convergence. Motivated by the Bates & Watts criterion, a similar convergence criterion has been developed in our maximum-likelihood theory for Poisson data.

Since the maximum-likelihood-estimation procedure described above is computationally equivalent to a weighted least-squares refinement, it follows also from (10) that a solution to the likelihood equations corresponds to a point at which the vector of 'residuals' $y-\eta(\beta)$ is orthogonal to the vector space spanned by the columns of $\mathbf{X}(\beta)$, in the geometry determined by the variance-covariance matrix of $Y$. An important consequence of this orthogonality is that the residual vector has zero projection onto the tangent plane spanned by the columns of $\mathbf{X}(\beta)$, and so we used the length of this projection as an indicator for convergence and we developed a meaningful tolerance level based on statistical considerations (a relative offset of the tangent plane confidence disc for the unknown parameters). Further details are discussed in Antoniadis & Berruyer (1990).

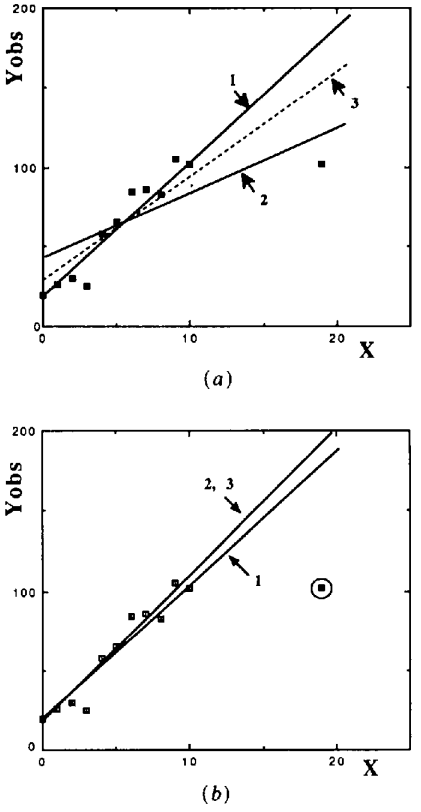


Fig. 1. Scatter plots of a simulated Poisson data set and two fitted regression line models, one by weighted least-squares refinement and the other by MLE; 1: true model, 2: WLS, 3: ML. In (*b*) the last point (circled) was not included in the fit.

## 4. Large-sample inference and model adequacy

### 4.1. *Asymptotic theory and tests*

This section addresses potential mis-specification of the nonlinear predictor $\eta(\beta)$ in generalized Poisson nonlinear regression models. Our approach is based on results from large-sample likelihood theory. We provide a brief review of the necessary results from this theory while setting out notation. The reader is referred to Cox & Hinkley (1974) for a more complete discussion.

It is helpful to distinguish two types of asymptotic situations: that where $n\to\infty$ and that where each $Y_i$ becomes approximately normal (for the latter case we will refer to an index $T\to\infty$). In the context of diffraction-pattern analysis, the $n$-asymptotics are equivalent to a small profile step width, while $T$-asymptotics are related to large-step counting times. We will focus first on the commonly occurring situation in which $n$ is large, regardless of the size of $T$.

Central to asymptotic likelihood arguments is the score vector defined by (9). Under some mild requirements, the central limit theorem applies and it can be shown that $\partial\ell[\eta(\beta), y]/\partial\beta$ is asymptotically Gaussian with mean zero and covariance matrix $\mathcal{I}(\beta)$. Standard limit calculations and Taylor expansions can then be employed to show that the MLE $\hat{\beta}$ is asymptotically multivariate normal with mean $\beta$ and covariance matrix $\mathcal{I}(\beta)^{-1}$.

Tests of hypotheses about $\beta$ and interval estimation are based on this result. For example, as we shall see, confidence regions can be specified using the fact that

$$^t(\hat{\beta}-\beta)\mathcal{I}(\beta)(\hat{\beta}-\beta)$$

has asymptotically a $\chi^2_p$ distribution with $p$ degrees of freedom. All the above results can be modified by replacing the theoretical information matrix $\mathcal{I}(\beta)$ with $\mathcal{I}(\hat{\beta})$ or even the observed one $\mathcal{J}(\hat{\beta})$ while retaining the $\chi^2$ result.

A second class of likelihood statistics is that based on the likelihood ratio and its asymptotic distribution, which has the advantage of yielding inferences that are independent (to some extent) of the arbitrary parametrization used. We shall be primarily concerned with that formed from the logarithm of a ratio of likelihoods, called the *deviance*.

Given $n$ profile points we could fit models to them containing up to $n$ parameters. The simplest model, the null model, has one parameter representing a common mean $\eta$ for all the $y_i$'s; it corresponds to a diffraction pattern with only a flat background contribution. At the other extreme the full model has $n$

parameters, one per observation, and the $n_i$'s derived from it match the data exactly. In practice the null model is too simple and the full model is uninformative. However, the full model provides us with a baseline for measuring the discrepancy for an intermediate model with $p$ parameters ($0 < p < n$). More precisely, the maximum log-likelihood achievable in a full model with $n$ parameters is $\ell(y, y)$. The discrepancy of a fit is proportional to twice the difference between the maximum log-likelihood achievable and that achieved by the model under investigation. In the Poisson regression case, the discrepancy can be written

$$D(y, \hat{\eta}) = 2\{\ell(y, y) - \ell(\hat{\eta})\}$$
$$= 2\left\{\sum_{i=1}^{n} [y_i \ln(y_i/\hat{\eta}_i) - (y_i - \hat{\eta}_i)]\right\},$$

known as the deviance for the current model, and called sometimes the $G^2$ statistic (see Bishop, Fienberg & Holland, 1975). By standard asymptotic arguments one can show that the deviance for a correct Poisson model without systematic error is approximately distributed as a $\chi^2$ with $n-p$ degrees of freedom, where $p$ is the dimension of the fitted model. Thus, the deviance can and will be used for goodness-of-fit purposes.

In profile-fitting methods, frequently one wishes to choose between two models which differ essentially in the number of parameters used to describe the pattern. A model with fewer restraints, that is with a greater number of parameters, can usually be made to fit the data better than can a more restrained model (when the parameters in a model are a subset of these in another model we say that the two models are nested); the crystallographer is thus often tempted to add more and more parameters to obtain better fits. It is therefore important to obtain a convenient method which allows one to decide whether the increase in the number of parameters leads to a significant improvement in the agreement between the observed and calculated patterns. This can be done by *hypothesis-testing* procedures or *model-selection* criteria.

The deviance can be used when comparing a series of nested models. For example, let $H_i: \boldsymbol{\theta} \in \Omega_i$ $(i = 1, 2)$ be two nested hypotheses $\Omega_1 \supseteq \Omega_2$ of dimension $p_1$ and $p_2$ respectively. Let $D_i$ $(i = 1, 2)$ be the deviance for model $H_i$ $(i = 1, 2)$. By a suitable large-sample argument we have that, asymptotically, under $H_2$, $D_1$ and $D_2 - D_1$ have independent $\chi^2$ distributions with degrees of freedom $n - p_1$ and $p_1 - p_2$, respectively, and are asymptotically independent of $\hat{\boldsymbol{\theta}}$. An approximate test for $H_2$ under $H_1$ may then be based on the Fisher–Snedecor distribution of the $F$ statistic

$$F = \frac{(D_2 - D_1)/(p_1 - p_2)}{D_1/(n - p_1)}.$$

Tests of this form have already been proposed for other types of distributions (see *e.g.* Mardia, 1972, p. 154; Jensen, 1981) and have been applied in crystallography by Hamilton (1965).

Another viable alternative to hypothesis testing is the application of model-selection criteria involving the deviance. Model-selection criteria take account simultaneously of both the goodness-of-fit (likelihood) of a model and the number of parameters used to achieve that fit. The criteria we consider can be represented as special cases of criteria such as those introduced by Akaike (1974*a*, *b*) or Schwarz (1978). They take the form of a penalized likelihood function, that is, the deviance of the current model plus a penalty term, which increases with the number of parameters. All these criteria take the form

$$Q_k = D_k + \alpha(n)m(k), \tag{14}$$

where $D_k$ is the deviance for the $k$th model, $\alpha(n)$ represents the cost of fitting an additional parameter and $m(k)$ is the number of independent parameters in the nonlinear predictor $\eta$. Akaike's information criterion (AIC) is of the form (14) with $\alpha(n) = 2$ for all $n$, while in Schwarz's criterion one has $\alpha(n) = \ln(n)$. Since, for $n$ greater than 8, $\ln(n)$ exceeds 2, Schwarz's criterion favors models with fewer parameters than does Akaike's. Application of both these criteria to some real and simulated examples are given at the end of this paper.

### 4.2. *Measuring the overall goodness of fit*

Fitting a model to data may be regarded as a way of replacing a set of data values $y$ by a set of fitted values $\hat{y}$ derived from a model involving in general a relatively small number of parameters. Measures of discrepancy (or goodness of fit) may be formed in various ways, but we shall be primarily concerned with the ones usually used in crystallography as well as those based on the deviance.

In pattern decomposition methods on diffraction data, assuming that the relevant error distribution is Poisson, and after the usual fitting is completed, model adequacy is generally examined *via* the Pearson $\chi^2$ statistic, which takes the form

$$\chi^2 = \sum_{i=1}^{n} (y_i - \hat{\eta}_i)^2 / V_{\hat{\eta}_i}(Y_i), \tag{15}$$

where $V_{\hat{\eta}_i}(Y_i) = \hat{\eta}_i$ is the estimated variance for the $i$th observation. If the model is correct, then the above statistic is approximately distributed like a $\chi^2$ distribution with $n - p$ degrees of freedom. If the value of this Pearson statistic departs significantly from its expectation $n - p$, then it can be concluded that either the Poisson assumption is inappropriate, or that the theoretical model for diffraction peaks is incomplete or incorrect. In any case, inference drawn from badly fitted models should be viewed with caution.

While in the field of profile fitting $\chi^2$ is widely accepted as a reasonable measure of goodness of fit, other statistics for describing the goodness of fit, not commonly used in the statistical literature, are still very popular among crystallographers. Thus, the statistical behavior of, for example, the conventional agreement factors $R_p$ and $R_{wp}$ are (see *e.g.* Young, Prince & Sparks, 1982) worth a closer look:

$$R_p = \frac{\sum_{i=1}^{n} |y_i(\text{obs.}) - (1/c)y_i(\text{calc.})|}{\sum_{i=1}^{n} y_i(\text{obs.})}$$

$$R_{wp} = \left\{ \frac{\sum_{i=1}^{n} w_i[y_i(\text{obs.}) - (1/c)y_i(\text{calc.})]^2}{\sum_{i=1}^{n} w_i[y_i(\text{obs.})]^2} \right\}^{1/2},$$

where $w_i$ is the $i$th weight and $c$ is the scale factor. In these expressions the data $y_i$ need not be profile points in a powder diffraction experiment, but can be any experimentally accessible quantities. In our opinion these statistics are poor measures of goodness of fit, especially in the case of profile fitting. To support this statement, consider for example a diffraction pattern without any Bragg reflexion and with a flat background, that is, suppose that the observed data $y_i$ are independent observations from the same Poisson distribution with parameter $\mu$. Under the Poisson assumption, the statistic $R_{wp}$ becomes

$$R_{wp} = \left\{ \frac{\sum_{i=1}^{n} [y_i(\text{obs.}) - \hat{\mu}]^2/\hat{\mu}}{\sum_{i=1}^{n} [y_i(\text{obs.})]^2/\hat{\mu}} \right\}^{1/2},$$

where $\hat{\mu}$ is any consistent estimator of the variance $\mu$ of $Y_i$. It is not difficult to see that $R_{wp}$ behaves like $1/(1+\mu)$ and therefore can be made arbitrarily small when $\mu$ is large. The agreement factors are more appropriate when they were used as building blocks for the $\mathcal{R}$ test, as suggested by Hamilton (1964), but then it turns out that the likelihood ratio tests of the previous section are more powerful.

When modeling using likelihood principles, the deviance is a more adequate statistic as a measure of the overall goodness of fit. The Pearson statistic (15) is often much more nearly $\chi^2$ than is that of the deviance (*e.g.* see Larntz, 1978) but we point out that the statistic having a more nearly $\chi^2$ distribution is not directly connected to being the better measure of overall lack of fit. This seems to be an obvious issue that it is easy to overlook. We feel that what is presented during the analysis of simulated examples makes a strong case for the superiority of the deviance as such a measure, once the deviance is corrected in order that its probability distribution is closer to a $\chi^2$ one. More precisely, for a linear model with Gaussian errors, it is known that the deviance $D(y, \hat{\eta})$ has exactly a $\chi^2$ distribution with $(n-p)$ degrees of freedom when the postulated theoretical model with $p$ parameters is correct. In the Poisson case this result holds only asymptotically and only for very large $n$ and $T$. Definition of a modified deviance by

$$D^*(y, \hat{\eta}) = c^{-1} D(y, \hat{\eta}),$$

where $c = \mathbb{E}(D)/(n-p)$, allows $\mathbb{E}(D^*)$ to be better approximated by $(n-p)$ than is $\mathbb{E}(D)$; this implies a better approximation of the distribution of $D^*$ by the $\chi^2_{n-p}$ distribution. For Poisson counting statistics we have (*e.g.* see Antoniadis & Berruyer, 1990), if terms up to the order $n^{-3/2}$ and $\eta_i^{-2}$ are ignored:

$$\mathbb{E}[D(y, \hat{\eta})] = n - p - \varepsilon_p + 1/6 \sum_{i=1}^{n} \frac{1}{\hat{\eta}_i}, \qquad (16)$$

where $\varepsilon_p$ is a correction term, usually of order $n^{-1}$, depending on the model.

### 4.3. *Generalized residuals and diagnostic plots*

Plots of residuals and of functions of residuals are particularly useful for identifying patterns in the data that may suggest overdispersion or bias due to misspecifications of one or more components in the assessed model. Diagnostic measures are also invaluable aids for a thorough detection of influential data points. For linear and nonlinear Gaussian models, these procedures are well documented in books such as those by Belsley, Kuh & Welsh (1980) and Cook & Weisberg (1982). In this section we consider their extension to the Poisson nonlinear regression case and we discuss their use in identifying individual poorly fitting observations.

Most of the asymptotic results pertaining to individual case diagnostics require $T$ to be large. In particular, this is the case for the distribution of residual deviance and of residuals as defined subsequently. The aim is to consider residuals that are approximately normally distributed. We will consider first some general recipes for calculating the residuals $R(y_i, \eta_i)$, treating the $\eta_i$ as known and turn subsequently to the effect of replacing $\eta_i$ by the fitted values $\hat{\eta}_i$. When treating the means $\eta_i$ as known, but more or less arbitrary, we will drop the subscripts.

For Poisson random variables the major possibilities are the following:

($a$) linear or 'Pearson' residuals,

$$R_L(Y, \eta) = [Y - \mathbb{E}_\eta(Y)]/\sigma_\eta(Y) = (Y - \eta)/\eta^{1/2},$$

where $\mathbb{E}_\eta$ and $\sigma_\eta$ denote the mean and standard deviation of a scalar Poisson random variable with parameter $\eta$;

($b$) transformed linear residuals,

$$R_t(Y, \eta) = \{t(Y) - E_\eta[t(Y)]\}/\sigma_\eta[t(Y)],$$

where the transformation $t(\cdot)$ is either $t(y)=y^{2/3}$ or $t(y)=(y+3/8)^{1/2}$; and

(*c*) deviance residuals,

$$R_D(Y,\eta)=\text{sign}\,(\hat{\eta}-\eta)\{2[\ell(Y,\hat{\eta})-\ell(Y,\eta)]\}^{1/2},$$

where $\hat{\eta}$ is the maximum-likelihood estimator (MLE) of $\eta$.

The two cases in (*b*) correspond to the choice of $t(\cdot)$ to make the $T$-asymptotic skewness of $t(Y)$ zero in the hope of achieving asymptotic normality, and alternatively, the choice of $t(\cdot)$ to make the $T$-asymptotic variance of $t(Y)$ constant and equal to 1/4. The residuals based on the former choice are called *Anscombe residuals*, while the second choice defines the so-called *variance-stabilizing residuals*; see Efron (1982) for a fuller discussion.

Once the maximum-likelihood estimates $\hat{\eta}_i$ of the responses $\eta_i$ are fitted, the above residuals, standardized to have unit asymptotic variance, are given by

$$\hat{R}(Y,\hat{\eta}_i)=R(Y,\hat{\eta}_i)/(1-h_i)^{1/2}, \tag{17}$$

where $h_i$ is the $i$th diagonal element of the so-called *hat matrix* $\mathbf{H}$ defined by

$$\mathbf{H}=\mathcal{I}(\hat{\boldsymbol{\eta}})^{1/2}\mathbf{X}(\hat{\boldsymbol{\beta}})[{}^t\mathbf{X}(\hat{\boldsymbol{\beta}})\,.\,\mathbf{A}_{\hat{\boldsymbol{\beta}}}\,.\,\mathbf{X}(\hat{\boldsymbol{\beta}})]^{-1} + {}^t\mathbf{X}(\hat{\boldsymbol{\beta}})\mathcal{I}(\hat{\boldsymbol{\eta}})^{1/2}; \tag{18}$$

and $R$ is $R_L$ or $R_t$ or $R_D$.

All the diagnostics proposed in this section are functions of $h_i$ and $R(y_i,\hat{\eta}_i)$. The $h_i$ belong to the interval $[0,1]$ and are essentially measures of case leverage. Large values of $h_i$ are useful in detecting potentially influential points for the fit. Informative displays of the above residuals include plots of $h_i$ or $R(y_i,\hat{\eta}_i)$ *versus* the index $i$. Such plots are particularly valuable when trying to decide whether an unacceptably large deviance or $\chi^2$ is due to a small number of outlying observations (due to a local misspecification in the model) or to a more general lack of fit or to overdispersion. In either case, large individual components indicate observations poorly accounted for by the model.

The above quantities, however, cannot adequately measure the effect on the many components of the fitted model. A common technique to assess influence of individual observations on the estimation of the unknown parameters or their standard deviation is by case deletion. In fact, many of the influence measures of linear regression are based on the differences $\hat{\boldsymbol{\beta}}_{(i)}-\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}_{(i)}$ denotes the least-squares estimate of $\boldsymbol{\beta}$ after deletion of the $i$th case from the data. Two such influence measures are *Cook's distance* ($C_i$) and the *likelihood distance* ($LD_i$) measures. Using a local asymptotically linear approximation of the Poisson nonlinear model by a linear model (see Antoniadis & Berruyer, 1990), an approximation to Cook's distance and likelihood distance are respectively given by the following expressions:

$$C_i=\frac{\hat{h}_i}{p(1-\hat{h}_i)}R_L(y_i,\hat{\eta}_i)^2 \tag{19}$$

and

$$LD_i=n\ln\{[p/(n-p)]C_i+1\}. \tag{20}$$

These measures will be used in the analysis of some realistic simulated examples provided later.

### 4.4. *Confidence regions and confidence intervals*

In practice, the estimated values of the parameter vector $\boldsymbol{\beta}$ will not be equal to the true values because of random errors in the data. A problem that is often overlooked in studies employing nonlinear least-squares techniques for parameter estimation is confidence-region estimation. More precisely, since $\hat{\boldsymbol{\beta}}$ is a random vector, it may be possible to indicate with some specific confidence level $(1-\alpha)$ in what region $\text{CR}_\alpha(\boldsymbol{Y})$ about $\hat{\boldsymbol{\beta}}$ we might reasonably expect $\boldsymbol{\beta}$ to be. Such regions are known as $100(1-\alpha)\%$ confidence regions. The present section addresses the problem and presents the available mathematical techniques for the evaluation of such confidence regions and intervals.

Mathematically, a joint confidence region for all the parameters is defined by the image of a function

$$\text{CR}_\alpha:\mathbf{Y}\rightarrow \text{a region in } \mathbb{R}^p$$

that satisfies $\Pr[\boldsymbol{\beta}\in\text{CR}_\alpha(\boldsymbol{Y})]\geq 1-\alpha$, that is which covers the true value of the unknown parameter with a probability at least $1-\alpha$.

Several methods for finding confidence regions exist and they are all equivalent in the very large sample case. Methods that, for all functions $\boldsymbol{\eta}(\boldsymbol{\beta})$ and confidence levels $1-\alpha$, are statistically guaranteed to contain the true value $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$ $100(1-\alpha)\%$ of the time are called exact; all other methods are called approximate.

The method (linearization method) analyzed in this section is exact for Gaussian linear models. For Poisson nonlinear models it is only approximate but has the advantage that the resulting confidence regions and intervals are simple and inexpensive to compute and that it produces bounded convex confidence regions. For a description of exact methods such as the lack-of-fit method for example, see Antoniadis & Berruyer (1990).

Linearization methods assume that $\boldsymbol{\eta}(\boldsymbol{\beta})$ can be adequately approximated by an affine or linear function at the vicinity of the least-squares solution $\boldsymbol{\eta}(\hat{\boldsymbol{\beta}})$. Under such an assumption, the linear least-squares confidence region for the true parameter vector $\boldsymbol{\beta}^*$ consists of those values of $\boldsymbol{\beta}$ for which:

$${}^t(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\mathbf{A}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\leq c_\alpha, \tag{21}$$

where the matrix **A** is either the observed information matrix $\mathcal{J}(\hat{\boldsymbol{\beta}})$ for $\boldsymbol{\beta}$ (see § 3.1) or Fisher's information matrix $\mathcal{I}(\hat{\boldsymbol{\beta}})$ and where $c_\alpha$ is such that

$$\Pr(\chi^2 \leq c_\alpha) = 1 - \alpha.$$

An alternate, though misguided, approach to construction of confidence regions is to consider the components $\beta_j$ one at a time. This approach ignores the covariance structure of the $p$ variables $\hat{\beta}_j$, $j = 1, \ldots, p$ and leads to the intervals:

$$\beta_i \in [\hat{\beta}_i - \hat{\sigma}(\hat{\beta}_i) z_\alpha, \hat{\beta}_i + \hat{\sigma}(\hat{\beta}_i) z_\alpha]$$

where $z_a$ is the $100(1-\alpha)$ percentile point of a standard normal variate and $\hat{\sigma}(\hat{\beta}_i)$ is the square root of the $i$th diagonal entry of the matrix $\mathcal{J}(\hat{\boldsymbol{\beta}})$. Although, prior to sampling, the $i$th interval above has (approximately) probability $1-\alpha$ of covering $\beta_j^*$, we do not know what to assert, in general, about the probability of all intervals containing their respective $\beta_j^*$'s. This probability is not $1-\alpha$ (see Donaldson & Schnabel, 1987).

If we adopt the attitude that all of the separate confidence statements should hold simultaneously with a specified high probability $1-\alpha$ then we should consider simultaneous confidence intervals. These are defined as shadows of the $p$-dimensional confidence ellipsoid for $\boldsymbol{\beta}$. These shadows may be obtained by projecting the $p$-dimensional ellipsoid on each co-ordinate axis (see Fig. 2).

For $p > 3$, we cannot graph the joint confidence region for $\boldsymbol{\beta}$. However, it is often informative to investigate simultaneous confidence regions for the components of $\boldsymbol{\beta}$ in pairs, even if they share the same weaknesses as the individual confidence intervals. In matrix notation, if we set

$$\underset{(p\times 2)}{\mathbf{L}} = [\ell_i \ \ell_j]$$

so that

$${}^t\mathbf{L}\boldsymbol{\beta} = \begin{bmatrix} {}^t\ell_i \\ {}^t\ell_j \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} {}^t\ell_i\boldsymbol{\beta} \\ {}^t\ell_j\boldsymbol{\beta} \end{bmatrix},$$

where $\ell_i$ denotes the $i$th $p$-dimensional basis vector,

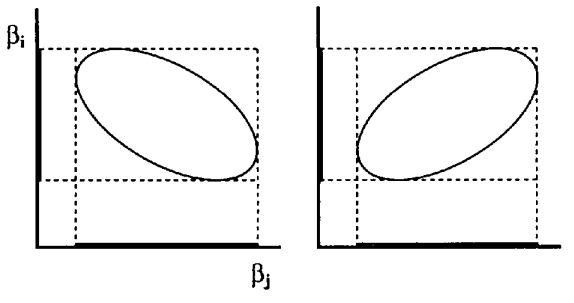


Fig. 2. Two confidence ellipses yielding the same simultaneous individual intervals (and hence the same rectangles) for the parameters $\beta_i$ and $\beta_j$.

then ${}^t\mathbf{L}\boldsymbol{\beta}$ lies in the two-dimensional ellipse

$${}^t({}^t\mathbf{L}\boldsymbol{\beta} - {}^t\mathbf{L}\hat{\boldsymbol{\beta}})({}^t\mathbf{L}\mathbf{A}^{-1}\mathbf{L})^{-1}({}^t\mathbf{L}\boldsymbol{\beta} - {}^t\mathbf{L}\hat{\boldsymbol{\beta}}) \leq c_\alpha, \quad (22)$$

if and only if $\boldsymbol{\beta}$ lies in the $p$-dimensional ellipsoid defined by (21). This result shows that joint two-dimensional confidence ellipses are shadows (projections) of the $p$-dimensional ellipsoid for $\boldsymbol{\beta}$ and therefore are generally wider than the correct $(1-\alpha)$% two-dimensional regions. To shed some light on how much wider they can be, displays of the slice through the center of the true ellipsoid by the plane spanned by the vectors $\ell_i$ and $\ell_j$ are very useful (see Fig. 8).

## 5. Data dispersion and quasi-likelihood

### 5.1. *Over- or underdispersion with respect to the Poisson distribution*

A common problem with counting data is that, even with a very good explanatory structural model for the mean, the fits obtained are poor. This is reflected in large (small) residual deviances and adjusted residuals which have a variance greater (smaller) than 1. This indicates that, conditional upon the explanatory peaks and background included in the final model, the variance of the observations are not of the same order as their means, as they should be in the Poisson case. Such data are frequently described as being *over- or underdispersed.*

In fact, while the basic physical phenomenon involved in diffraction experiments (photon or neutron emission) is Poissonian in nature, the recorded counts may exhibit a different statistical behavior due to a combined effect of the counting chain response and of data corrections. In other words, response data as presented for analysis may have been aggregated or scaled, or the usual assumption of independence may be incorrect, *i.e.* the data are correlated, or simply important explanatory 'variables' are incorrectly excluded from the regression relationship (systematic effects).

The method to be developed in the following sections is particularly concerned with these problems since over/underdispersion data do not satisfy the basic assumptions of our maximum-likelihood approach and also since the lack of an explicit dispersion parameter in standard least-squares algorithms can be shown to be, at least partially, responsible for some odd results.

The regression methods described in the previous sections were particular applications of the theory of maximum-likelihood methods in *curved exponential families.* Consideration of such distributions instead of the Poisson ones for the observed counts allows a generalization of MLE algorithm, and then an extra 'dispersion parameter' appears in a natural way.

### 5.2. *Estimating overdispersion with quasi-likelihood*

A quasi-likelihood method has been proposed by Wedderburn (1974) for the estimation of parameters in generalized linear regression models (GLIM) when there is some assumed relationship between the mean and variance of each observation but not necessarily a fully specified likelihood.

Let us remark first that an interesting property of the algorithm we used in fitting diffraction data is that the distributional assumption of the errors enters only through the variance function of that distribution. Thus given the assumption of Poisson counts, the fitting algorithm uses only the fact that $V(\mu)=\mu$. Wedderburn's method originally proposed for linear models (discussed by McCullagh & Nelder, 1983, Ch. 8) can therefore be extended to the non-linear case. We will require only that the variance function of our observations is known up to a multiplicative constant, so that $V(\mu)=\varphi\mu$ where $\varphi$ is the dispersion parameter. As with ordinary quasi-likelihood, $\varphi$ is implicitly assumed to be functionally independent of $\mu$. When $\varphi>1$ we clearly have overdispersion with respect to the Poisson distribution (underdispersion when $\varphi<1$). The log quasi-likelihood is of the form $\mathscr{L}[\boldsymbol{\mu}(\boldsymbol{\beta}), \varphi]$, with $\varphi>0$, where

$$\mathscr{L}[\boldsymbol{\mu}(\boldsymbol{\beta}), \varphi]=\sum_{i=1}^{n}\{-\tfrac{1}{2}\ln(2\pi\varphi y_i) - \varphi^{-1}[y_i\ln(y_i/\mu_i)-(y_i-\mu_i)]\}. \quad (23)$$

The quasi-likelihood estimates of the components of $\boldsymbol{\beta}$ are obtained by maximizing $\mathscr{L}(\boldsymbol{\mu}, \varphi)$, so the computation of $\hat{\boldsymbol{\beta}}$ may be done exactly as in the previous sections. The estimator $\hat{\boldsymbol{\beta}}$ is still consistent and asymptotically normal with covariance matrix $\varphi\hat{\mathbf{V}}$ where $\hat{\mathbf{V}}$ denotes the ordinary covariance matrix of the maximum-likelihood estimator of $\boldsymbol{\beta}$ in the standard Poisson regression model. In other words, asymptotically $\varphi$ acts as a simple scaling factor.

When $\varphi$ is unknown, which is generally the case in practice, $\varphi$ is estimated by $D(\hat{\boldsymbol{\mu}}, \mathbf{y})/(n-p)$ where $D(\boldsymbol{\mu}, \mathbf{y})$ is the deviance, and this estimate of $\varphi$ is used in the computation of standard errors of the components of $\hat{\boldsymbol{\beta}}$.

None of the statistical ideas presented and formalized here are new; they have been used many times in diffraction data analyses (see *e.g.* Sakata & Cooper, 1979; Young, Prince & Sparks, 1982). The basic objection for such an estimation of the overdispersion parameter $\varphi$ is that it presupposes that all major systematic effects have been accounted for by the model and it is therefore much safer first to identify and isolate the major systematic effects and only then to assign the remainder to residual or unexplained variation.

In practice, when analyzing real data, the $\chi^2$ statistic or the deviance are much larger or smaller than that predicted by Poisson sampling, indicating rejection of both structural model and Poisson distribution. This leaves us with little information still about what distribution actually generated the data. In the following section we propose an alternative way of estimating the parameter $\varphi$. Our estimator does not rely on the assumption that all the systematic effects have been accounted for by the model and is therefore more robust.

### 5.3. *Non-parametric estimation of the dispersion parameter* $\varphi$

In this subsection, a nonparametric estimator of the parameter $\varphi$ is proposed. It is based on local linear fitting and on the fact that, despite the presence of the parameter $\varphi$, the observations are Poisson-like distributed. More precisely, we shall assume hereafter that the square-root transform (see § 4.3) is asymptotically a variance-stabilizing transformation for the observed variables $Y_i$, *i.e.* for moderate to large counts $Y_i^{1/2}$ (which differs little from the transformation prescribed in § 4.3)* is asymptotically distributed as a normal variable with mean $\mu_i^{1/2}$ and variance $\sigma^2=\varphi/4$. It then follows that the transformed data $y_i^{1/2},\ldots,y_n^{1/2}$ become a regression model

$$X_i=Y_i^{1/2}=\psi(t_i)+\varepsilon_i \qquad (i=1,\ldots,n), \quad (24)$$

where the residuals $\varepsilon_i$ are independent random variables with expectation zero and variance $\varphi/4$. In order to estimate the residual variance independently of any parametric model for the function $\psi$, the estimator will only depend on some weak assumptions on the smoothness of $\psi$.

Our proposal for obtaining residual variance nonparametrically is based on pseudo-residuals $\tilde{\varepsilon}_i$. Pseudo-residuals $\tilde{\varepsilon}_i$ are obtained by taking continuous triples of design points $t_{i-1}$, $t_i$, $t_{i+1}$, joining the two outer observations by a straight line and then computing the difference between this straight line and the middle observation $X_i$:

$$\tilde{\varepsilon}_i=\frac{t_{i+1}-t_i}{t_{i+1}-t_{i-1}}X_{i-1}+\frac{t_i-t_{i-1}}{t_{i+1}-t_{i-1}}X_{i+1}-X_i$$
$$=a_iX_{i-1}+b_iX_{i+1}-X_i \qquad (i=2,\ldots,n-1). \quad (25)$$

From the properties of these pseudo-residuals we are lead to the following definition of an estimate of the residual variance $\hat{\sigma}^2$:

$$\hat{\sigma}^2=(n-2)^{-1}\sum_{i=2}^{n-1}c_i^2\tilde{\varepsilon}_i^2, \quad (26)$$

where $c_i^2=(a_i^2+b_i^2+1)^{-1}$ for $i=2,\ldots,n-1$. Indeed, for a function $\psi$ assumed to be twice differentiable, it is not difficult to see that $\mathbb{E}(\hat{\varepsilon}_i^2)=$

* When $(Y_i+3/8)^{1/2}$ is used instead of $Y_i^{1/2}$ the asymptotic behavior is reached more rapidly.

Table 1. *Maximum-likelihood estimation of the parameters of a simulated pattern with low Poisson counts*

The first row gives the values of the true parameters and the following ones display the parameter values estimated by several minimization procedures (Fig. 3). The true values were used as starting values; for each peak Int is the integrated intensity, Pos is the position and FWHM is the full width at half maximum. The second row reports the estimated parameters by the maximum-likelihood method while the third row provides the estimated values by standard non-weighted least-squares refinement since the modified minimum $\chi^2$ method diverges due to zero counts (standard errors are given in parentheses). The $R$ factors are 0·77 and 0·78 for the MLE and the least-squares refinement respectively; the goodness of fit is 100% for both.

A simulated weak pattern

| | | Peak 1 | | | Peak 2 | | | Peak 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | Background | Int | Pos | FWHM | Int | Pos | FWHM | Int | Pos | FWHM |
| Simulated | 0·1 | 0·50 | 10·5 | 0·333 | 0·2 | 11·0 | 0·333 | 0·3 | 12·3 | 0·333 |
| Maximum likelihood | 0·15 (9) | 0·32 (14) | 10·49 (6) | 0·32 (14) | 0·20 (14) | 11·02 (12) | 0·40 (29) | 0·30 (16) | 12·26 (9) | 0·52 (25) |
| Least squares | 0·16 (8) | 0·14 (7) | 10·46 (3) | 0·14 (7) | 0·27 (15) | 10·99 (12) | 0·54 (30) | 0·38 (25) | 12·30 (21) | 0·92 (68) |

$(a_i^2+b_i^2+1)\sigma^2+O(n^{-2})$, and this justifies our choice for $\hat{\sigma}^2$ as an asymptotically unbiased estimator of $\sigma^2$.

The precise assumptions needed for obtaining the asymptotic results are:

($a$) there are no multiple measurements at any design point;

($b$) $\max|t_i-t_{i-1}|=O(1/n)$;

($c$) the function $\psi$ is differentiable.

It is easy to see that the above conditions are sufficient for asymptotic unbiasedness of $\hat{\sigma}^2$. If the bias term is disregarded and if moreover the function $\psi$ is Lipsitch of order greater than $\frac{1}{4}$, then $\hat{\sigma}^2$ is also asymptotically normal (see Antoniadis & Berruyer, 1990). The bias problem and the smoothness condition on $\psi$ is further discussed in § 6.2.

For models that are overdispersed or mis-specified, such an estimator might be useful for model selection and model checking. We have undertaken some simulations to study primarily the bias and secondly the validity of residual variance estimator as a tool for detecting and estimating over- or underdispersion counting data. The results provide useful information and are reported in the data-analysis section.

## 6. Statistical analysis of some simulated examples

Although the strength of any data-analysis method will be judged from its use on 'real' data, these experiences do not easily tell us how correct the data reduction is. We have therefore produced artificial data with known intensities and counting errors and used these to illustrate the statistical methodology described in the previous sections and to assess the performance of the maximum-likelihood estimators. The next subsection discusses the simulation study.

### 6.1. *The simulation study*

In this section we outline the design of the simulation experiments used to examine how the input data sets and the sample size affect the performance of the maximum-likelihood statistics.

The simulated patterns were constructed as a sum of a polynomial background and a given number of one-dimensional distribution functions (Gauss, Cauchy, Lorentz or pseudo-Voigt). Several values for the average background, for the sample size ($n$) (number of data points) and peak parameters were chosen in various ranges. At each setting, described in the captions of the tables that follow, Poisson counts were generated according to the assumed structural model for their means. The pseudo-random Poisson counts were created from uniform variates using a simulation method by Antoniadis, Berruyer & Filhol (1985). The obtained data were analyzed by our MLE procedure or by the modified minimum $\chi^2$ method.

6.1.1. *MLE and minimum $\chi^2$ fitting.* Our first examples are designed to discuss the sensitivity of the MLE and modified minimum $\chi^2$ estimators to weak or strong intensities.

The results corresponding to a low count pattern are summarized in Table 1 and in Fig. 3. The assumed background being only 0·1, a number of zero counts generated by the Poisson process were artificially reset to 0·00001 in order to prevent overflow problems while computing. As starting values for the parameters we used their true values. The method of maximum likelihood was applied directly to the data with several observed cell values of 0·00001, and produced nonzero estimates for such cells. The modified minimum $\chi^2$ method suffers from the drawback that either it is not defined or it diverges when several data points are equal to zero and are replaced by some 'small' positive value as in our example. For comparison purposes we thus used a standard non-weighted least-squares procedure for fitting. Note in Table 1 and Fig. 3 the good agreement between the theoretical parameters used in the simulation and those obtained by the MLE procedure. However, the goodness-of-fit statistic [see (16)] based on the deviance is quite sensitive to the fact that the counts are low and its asymptotic $\chi^2$ approximation is probably not accurate. Nevertheless, since the parameter e.s.d.'s are not based on this statistic but only on the structural model and the Poisson assumption they are quite reliable.

Since the modified minimum $\chi^2$ method (MMCS) cannot fit data with observed cell values of zero, we now examine an example with non-zero but low cell counts. For the results displayed in Fig. 4 we simulated a pattern made with a very low intensity peak embedded in a weak (non-zero) background noise.

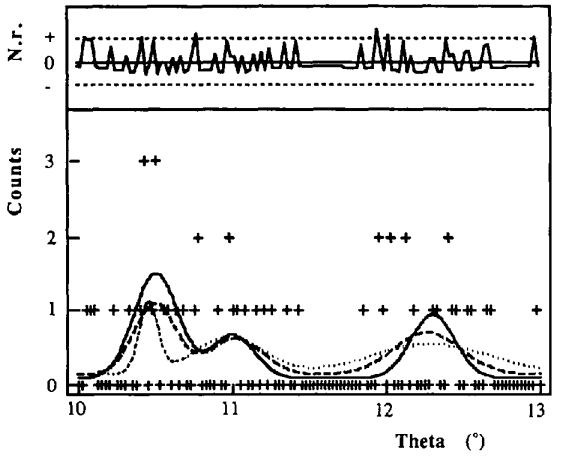


Fig. 3. A weak simulated diffraction pattern with three Gaussian peaks of equal FWHM (see Table 1). The thick black curve is the input pattern, the long-dashed one is the result of the MLE fit and the short-dashed one represents the fit by non-weighted least-squares refinement. The oscillating graph shows the normalized residuals $R_L$ (*N.r.*) within their 95% confidence band. We are faced here with data having very bad statistics and, for example for the third peak, the hazard of this simulation grouped strong counts on the left side of the peak. The discrepancy between the fitted and the theoretical curves may seem large at first sight but is within the statistical uncertainty for the maximum-likelihood method. This is not true for least-squares refinement.

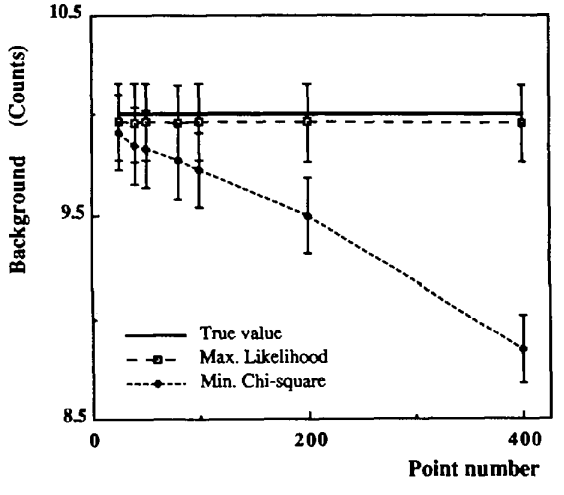


Fig. 4. Bias difference for the estimated background parameter between the modified minimum $\chi^2$ method and maximum-likelihood method. The two methods were applied to the same data set representing a weak pattern with a low-intensity peak. The horizontal axis represents the number of sampled data points. The peak was practically unobservable when the counts were recorded with a small step width (400 points) and became discernible when the cells were grouped by groups of 16 points. One can see that the larger the number of sampled points the greater the MMCS estimating bias for the background. The two methods give similar results for the peak intensity in this example.

Table 2 and Fig. 5 present another simulation with higher counts. A horizontal background and four peaks of different shapes were used. Only pseudo-Voigt-function peak shapes were used for the fit in order to check the stability of the algorithm in an asymptotic situation. Recall that the mixing parameter of a Gaussian peak when represented as pseudo-Voigt is 0·0, and for a Cauchy is 1·0. Thus, the algorithm performs very well. Note also the inadequacy of the $R_{wp}$ factor as an agreement index when the counts are high and the accuracy of the $\chi^2$ approximation for the deviance statistic in this case.

6.1.2. *Model-selection criteria.* The next set of simulations was generated in order to analyze the model-selection criteria introduced in § 4. The data set was analyzed twice to allow us to examine the effect of increasing the number of parameters during the fit and to check the accuracy of the selection criteria [AIC and *F*-test (see § 4.1)]. The results are reported in Table 3, and the corresponding graphs are shown in Figs. 6 and 7. As one can see, the AIC criterion (as well as Schwarz criterion) balances quite well the underfitting and overfitting risks by optimally adjusting the bias in the log-likelihood ratio when the maximum-likelihood estimates are used. It behaves consistently with the true model for large sample sizes. The sensitivity of the model-selection criteria as well as that of the *F*-test still holds for low counts. The fit in Fig. 7 was performed on a data set simulated under similar conditions to that displayed in Table 3, but with entries divided by a factor of ten.

Certainly, from a mathematical point of view, consistency is an attractive asymptotic property to expect from a model selection procedure, but any such consideration presupposes that there exists a 'true' order of a model. In the case of real data, the concept of true order is not known and one has to be cautious with model selection procedures.

6.1.3. *Confidence regions.* We present now an example of confidence-region estimation with a real data set. We found that the standard methods (see § 4.4) for approaching this problem were inadequate. Fig. 8 displays the confidence regions at 95% confidence level for the background parameter and the intensity of the peak obtained by fitting the data displayed in Fig. 9 when all other parameter estimates are held fixed at their maximum-likelihood values.

Table 2. *Maximum-likelihood estimation of the parameters of a simulated pattern with high Poisson counts*

The pattern is shown in Fig. 5 and has 201 scan steps over a $\theta$ scan range 10–14°. The values of the true parameters are given in the first row. The second row reports the estimated parameters while the third one provides their standard errors. $R$ factor = 0·047, goodness of fit = 24·72%. For this estimation we used a pseudo-Voigt function for each peak. Their estimated mixing parameters (with their respective e.s.d.'s are: 0·04 (0·06), 0·46 (0·06), 0·99 (0·07) and 0·55 (0·06), while their expected values are 0, 0·5 (approximately), 1 and 0·5. Thus, the observed difference between the fitted and expected values is not statistically significant.

A simulated 'strong' pattern

| | | Peak 1 | | | Peak 2 | | | Peak 3 | | | Peak 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameters | Background | Int | Pos | FWHM | Int | Pos | FWHM | Int | Pos | FWHM | Int | Pos | FWHM |
| Simulated | 100·0 | 100·0 | 10·5 | 0·1 | 100·0 | 11·5 | 0·1 | 100·0 | 12·5 | 0·1 | 100·0 | 13·5 | 0·1 |
| Estimated | 98·45 | 99·8 | 10·5 | 0·097 | 101·13 | 11·5 | 0·106 | 100·57 | 12·5 | 0·100 | 103·46 | 13·5 | 0·100 |
| E.s.d. | 1·41 | 2·28 | 0·0007 | 0·002 | 2·4 | 0·001 | 0·003 | 2·3 | 0·001 | 0·005 | 2·4 | 0·001 | 0·003 |

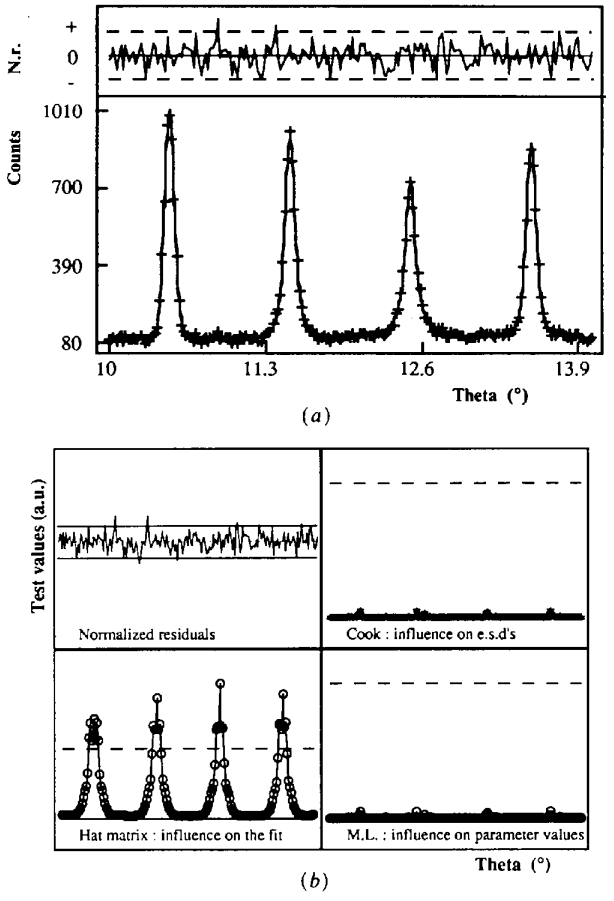


Fig. 5. A strong simulated diffraction pattern with four peaks of equal width (FWHM = 0·1) and intensities but different peak shapes (Gauss, Lorentz, Cauchy and pseudo-Voigt with mixing parameter 0·5). (*a*) The input and the fitted pattern cannot be distinguished on this plot. The fitted model assumes that all peaks are pseudo-Voigt in order to check the efficiency of the MLE algorithm in the asymptotic situation (large counts). The estimation results are reported in Table 2. Here again the oscillating graph shows the $R_L$ normalized residuals within their 95% confidence band. (*b*) The various diagnostic plots described in § 4.3 for the fits in (*a*). The broken lines indicate the location of the statistical threshold levels. Here the hat-matrix test clearly indicates that the four peaks are too narrow (insufficient sampling) and thus data points near peak tops strongly influence the fit quality. However, in that specific case, the other tests say that we can nevertheless trust the obtained parameter values and their e.s.d.'s.

One can see clearly that, since the estimated parameters are highly correlated, the standard statistical methods for obtaining confidence squares for a pair of fit parameters are significantly bad for predicting realistic confidence regions. Indeed, for each point marked □ in Fig. 8, the corresponding pattern was obtained and the corresponding normalized residuals were calculated. The results are displayed sequentially in Fig. 9. Point 1 is the only one lying within the multidimensional confidence ellipsoid. Points 2 and 3 lie on its projection onto the parameter plane. Point 4 is unambiguously outside the confidence ellipsoid. As one can see, the only point producing a correct fit is the one lying within the true multidimensional confidence ellipsoid.

The results presented here show how poor the standard confidence rectangles can be in some cases. Users will continue to use such regions, however, because it is readily available in software packages and provides a concise representation of the information needed to assess the precision of the estimated parameters individually. The conclusion is that more reliable methods for confidence-region estimation procedures should be used when the precision of the estimates is an important element for future analyses.

### 6.2. *Estimating dispersion on simulated and real data*

In this subsection we have undertaken a simulation to study the validity of (26) of § 5 for estimating the $\varphi$ factor in overdispersed count data.

Three different models described in the caption of Table 4 were used. For each model a Gaussian random noise with mean 0 and standard deviation 1 was added to the profile function, since the task of the simulation was to check the behavior of the variance estimator and not the validity of the square-root transform for Poisson data.

With the notation of § 5, let $\mathbf{C}$ denote the $(n-2)\times(n-2)$ diagonal matrix with elements $C_{i,i} = c_{i+1}$, and let $\mathbf{A}$ be $(n-2)\times n$ tridiagonal with elements $A_{i,i} = a_{i+1}$, $A_{i,i+1} = -1$, $A_{i,i+2} = b_{i+1}$ and $\mathbf{D} = {}^t\mathbf{AC}^2\mathbf{A}$. It is easily seen that the estimator $\hat{\sigma}^2$ defined by (26) is of the form

$$\hat{\sigma}^2 = {}^t\mathbf{XDX}/[\mathrm{tr}(\mathbf{D})], \qquad (27)$$

Table 3. *The parameters of a simulated pattern with* 121 *profile steps with high Poisson counts (see Figs.* 6 *and* 7*) for checking the model-selection criteria*

Four Gaussian peaks were used for the model.

Model selection for a 'high-count' pattern

| | | Peak 1 | | | Peak 2 | | | Peak 3 | | | Peak 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Background | Int | Pos | FWHM | Int | Pos | FWHM | Int | Pos | FWHM | Int | Pos | FWHM |
| Parameters | 100 | 500 | 10·4 | 0·30 | 500 | 11·0 | 0·40 | 200 | 12·0 | 0·20 | 50 | 11·82 | 0·15 |

where $\mathbf{X}$ denotes the vector of data points and tr denotes the trace of a matrix. The divisor, tr($\mathbf{D}$) in (27), is a necessary consequence of the condition that when the mean $\mu(t)$ of the $X_t$'s is a straight line then $\hat{\sigma}^2$ must be an unbiased estimator of $\sigma^2$. Expression (27) makes relatively easy the calculation of the moments of the estimator $\hat{\sigma}^2$, which can be calculated for a general $\mu$. We have

$$\hat{\sigma}^2 = {}^t(\boldsymbol{\mu}+\boldsymbol{\varepsilon})\mathbf{D}(\boldsymbol{\mu}+\boldsymbol{\varepsilon})/\mathrm{tr}(\mathbf{D})$$
$$= ({}^t\boldsymbol{\mu}\mathbf{D}\boldsymbol{\mu} + 2\,{}^t\boldsymbol{\mu}\mathbf{D}\boldsymbol{\varepsilon} + {}^t\boldsymbol{\varepsilon}\mathbf{D}\boldsymbol{\varepsilon})/\mathrm{tr}(\mathbf{D})$$

and hence, by equation (15.47) of Kendall & Stuart (1977 p. 382),

$$\mathbb{E}(\hat{\sigma}^2) = \sigma^2 + {}^t\boldsymbol{\mu}\mathbf{D}\boldsymbol{\mu}/\mathrm{tr}(\mathbf{D}) \qquad (28)$$

and, since $\boldsymbol{\varepsilon}$ and ${}^t\boldsymbol{\varepsilon}\mathbf{D}\boldsymbol{\varepsilon}$ are uncorrelated,

$$\mathrm{Var}(\hat{\sigma}^2) = [4\sigma^2\,{}^t\boldsymbol{\mu}\mathbf{D}^2\boldsymbol{\mu} + 2\sigma^4\mathrm{tr}(\mathbf{D}^2)]/[\mathrm{tr}(\mathbf{D})]^2. \qquad (29)$$

In our subsequent discussion we will concentrate on the bias of our estimator given by (28). Whenever the function $\boldsymbol{\mu}$ is smooth, *i.e.* differentiable with bounded derivatives, the estimator should have a bias of the order $n^{-1}$ where $n$ denotes the sample size. For a flat background with equidistant profile steps, this is the case and the estimator behaves well. For an equidistant design and normally distributed residuals it is not difficult to see that expression (29) leads to a variance for $\hat{\sigma}^2$ equal to $35n^{-1}\sigma^4/9$ and indeed that is what happens as one can note from the results in the first column of Table 4. Unfortunately, the bias becomes important for narrow peaks (small number of points per peak). We decided therefore to correct this bias by estimating the function $\boldsymbol{\mu}$ itself by nonparametric methods (not any assumption on the functional form of $\boldsymbol{\mu}$ except its 'smoothness'). The estimator of $\boldsymbol{\mu}$, say $\hat{\boldsymbol{\mu}}$, so obtained was used to correct $\hat{\sigma}^2$ by its estimated bias ${}^t\hat{\boldsymbol{\mu}}\mathbf{D}\hat{\boldsymbol{\mu}}/\mathrm{tr}(\mathbf{D})$.

So far, we have always assumed that the profile function $\boldsymbol{\mu}$ could be expressed by an analytic formula, which holds true across the entire observed range of

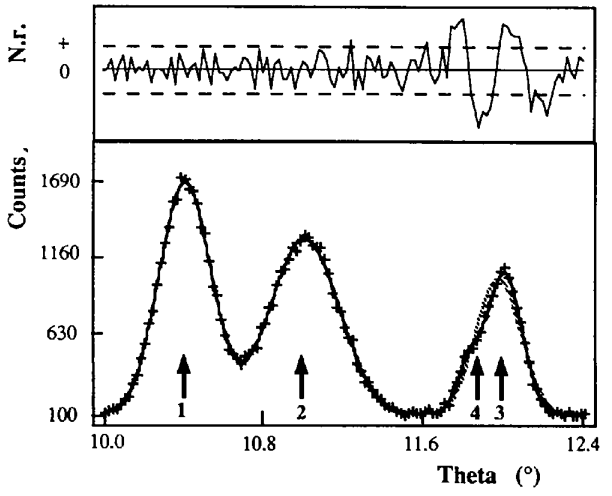


Fig. 6. A strong simulated diffraction pattern ($n = 100$) with four Gaussian peaks. When three peaks ($p_2 = 10$) were fitted (this figure), the corresponding selection statistics were: AIC = 275·16, Schwarz = 301·21; with four peaks correctly fitted ($p_1 = 13$), one obtains AIC = 111, Schwarz = 144·86 and the $F$ statistic = 58·07, with $p_1 - p_2 = 3$ and $n - p_1 = 87$ degrees of freedom, shows a significant improvement in the fit at a 95% confidence level.

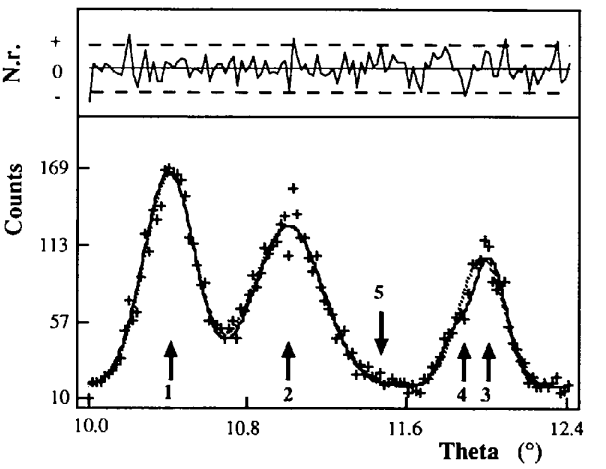


Fig. 7. A weak simulated diffraction pattern with four Gaussian peaks. It looks similar to that displayed in Fig. 6 but was produced with the intensities of Table 3 divided by a factor of ten. When three peaks were fitted (peaks 1, 2, 3) the corresponding selection statistics were: AIC = 146·83, Schwarz = 174·74; with four peaks fitted (peaks 1, 2, 3, 5; 5 is mis-specified) one obtains AIC = 156·77, Schwarz = 193·12 and the $F$ statistic = 0·284, with $p_1 - p_2 = 3$ and $n - p_1 = 87$ degrees of freedom, shows a non-significant improvement in the fit at a 95% confidence level. When the four peaks (peaks 1, 2, 3, 4) are modeled correctly, the criteria drop to AIC = 133·16, Schwarz = 170·03 and the $F$ statistic = 6·73, with $p_1 - p_2 = 3$ and $n - p_1 = 87$ degrees of freedom, now shows a significant improvement with respect to the three-peak fit at a 95% confidence level [$6{\cdot}73 > 3{\cdot}23 = F^{-1}_{3,87}(0{\cdot}95)$].

the explanatory variable $t$. Nonparametric estimation of $\mu$ by local smoothing avoids this assumption. In our case, a smooth curve has been fitted to the data using 'supersmoother', a local smoothing algorithm developed by Friedman & Stuetzle (1981) and discussed in Friedman (1984).

For the data generated according to the entries of Table 4, cross validation led to the choice of a smoothing window size $L=3$ resulting in the bias-corrected estimates of overdispersion reported in the last two columns of Table 4. As one can see from the results of Table 4, the bias-corrected estimates $\hat{\sigma}^2$ of $\sigma^2$ are satisfactory for the broad peak example even for relatively small sample size $n$ and improve as the sample size gets larger. For a 'narrow' peak and a moderate sample size ($n=100$), the corrected overdispersion parameter is still distorted upwards. This is a predictable phenomenon since local linear smoothers tend to 'cut corners' near a bend in the regression curve $\mu$.

In practice, however, the square-root transform on real overdispersed Poisson data makes the transform pattern much smoother, thus producing estimates that are not likely to be much distorted. The above techniques were applied to some real neutron powder diffraction diagrams with moderate to high resolution and it was found that overdispersion factors up to 2 are not rare in real experiments.

### 6.3. *Comparison of minimum $\chi^2$ and maximum likelihood in Rietveld refinements*

The advantages of ML as compared to minimum $\chi^2$ (weighted least squares) for profile refinement of Poissonian diagrams, such as those recorded in powder diffraction experiments, is easy to demonstrate

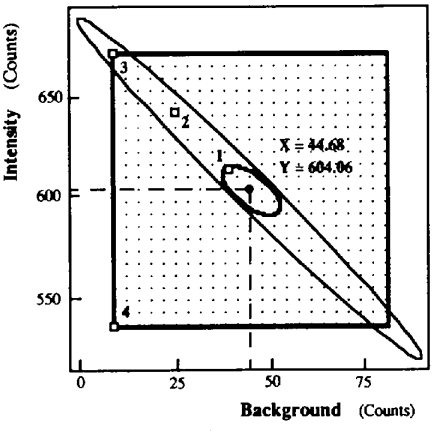


Fig. 8. The shaded rectangular region corresponds to the domain bounded by the individual confidence intervals at their 95% level, the large ellipsoidal region corresponds to the shadow of the confidence ellipsoid on the background-intensity plane and the small ellipsoid is the intersection of the confidence ellipsoid with the plane passing through the maximum-likelihood estimate and parallel to the background-intensity plane.

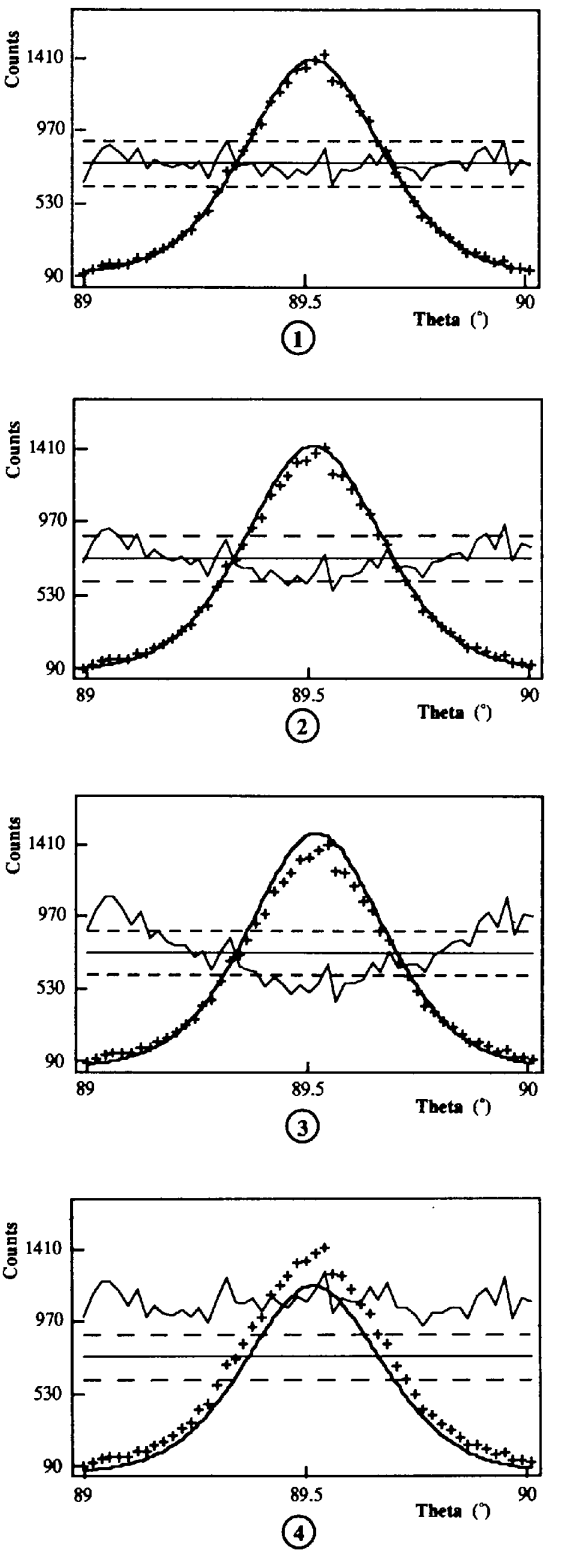


Fig. 9. A real diffraction pattern with one peak; the fitted model assumes that the peak is pseudo-Voigt. The corresponding confidence ellipsoid for the background and the intensity is shown in Fig. 8. Each graph in this figure reports the corresponding model and residuals for points 1, 2, 3 and 4 within the confidence ellipsoid graph of Fig. 8.

Table 4. *Estimation of the variance of a simulated Gaussian sample with standard deviation* $\sigma = 1$

The first column gives the number of data points; the next columns report the average of the dispersion estimates (and their standard errors in parentheses) obtained by equation (26) over 100 simulated data sets. For the flat background column, a single constant signal of 100 was perturbated by a Gaussian noise; for the broad-peak column, a well sampled symmetric smooth peak, centered at the middle of the pattern interval [10, 20], of integrated intensity 500 and of full width at half maximum (FWHM) 4, was added to the background. For the last column the peak was chosen narrower (maximum intensity 1000, FWHM = 1), in order to have less points within the peak range.

Gaussian noise $\sigma = 1$

| Number of data points | Flat background | Broad peak | Narrow peak |
|---|---|---|---|
| 100 | 1·005 (45) | 0·982 (37) | 2·900 (35) |
| 250 | 1·003 (15) | 1·001 (15) | 0·961 (14) |
| 500 | 0·992 (7) | 0·992 (7) | 0·945 (6) |
| 1000 | 0·999 (4) | 0·999 (3) | 0·980 (4) |

when the fitted diagram is not too complicated and when the fitted parameters are the background and peak parameters themselves. In the case of the Rietveld method, diagrams with sometimes more than a thousand strongly overlapping peaks are processed in one row. The refinement no longer bears on the background and peak parameters but on a much extended set: a complicated combination of the latter parameters through a structural model of the sample compound and simple functions modeling the instrument response.

Rodriguez-Carvajal & Menarde (1989) have undertaken an impressive simulation work from known crystallographic structures and we are pleased to be allowed to give here a summary of their results prior to publication. The method used is the following:

(*a*) From given structural parameters (unit-cell dimensions, nature and position of atoms, thermal parameters) and given instrumental parameters (peak shape, resolution curve, background-to-peak scaling factor, counting time) a theoretical (*i.e.* deterministic) diagram $Y_i$(th.) is computed.

(*b*) $Nr$ realizations of the corresponding simulated diagram are generated, which closely mimic true measured ones, through the use of a generator of pseudo-random numbers according to the Poisson law (Antoniadis *et al.*, 1985): $y_i(\text{sim.}) = P[\,Y_i(\text{th.})]$.

(*c*) This process is repeated for $N_t$ increasing values of the counting time and thus $N_r \times N_t$ simulated patterns are produced.

(*d*) Each diagram is refined through the Rietveld method by using either the classical weighted least-squares algorithm or the maximum-likelihood one. The refined model is either the true or a biased one (*e.g.* inadequate peak shapes).

The values of the refined parameters are then compared to the true values (bias and dispersion) and

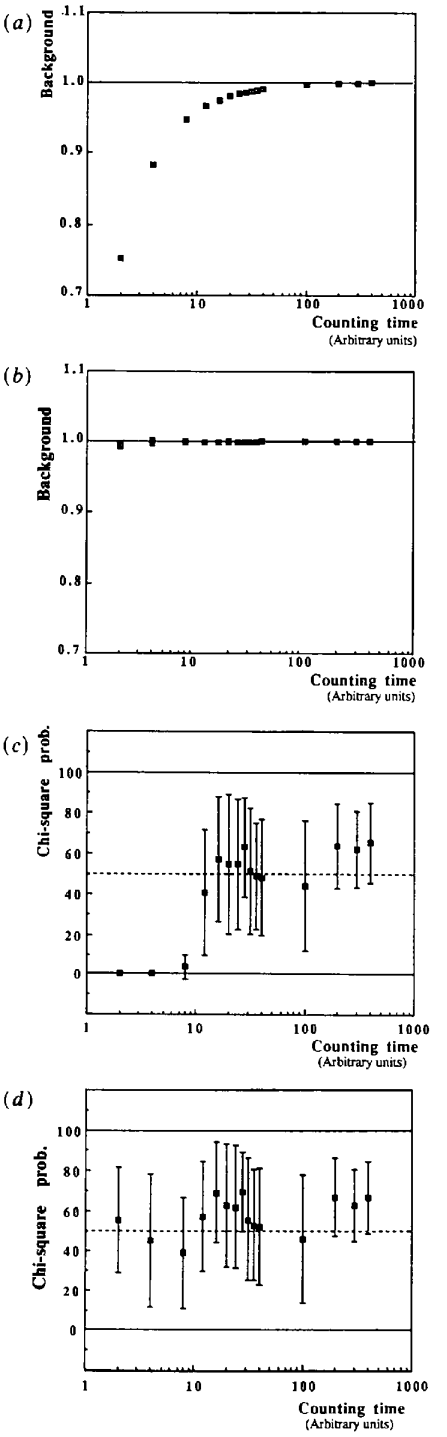


Fig. 10. A set of diffraction patterns of rutile ($TiO_2$) have been simulated for increasing counting times. Each simulation was repeated ten times. All diagrams were fitted using both the WLS (modified minimum $\chi^2$) and ML methods. (*a*) and (*b*) The background value is equal to 1. The scatter plots show the WLS estimates (*a*) and the ML estimates (*b*). The experimental error bars are within the size of the markers. (*c*) and (*d*) The $\chi^2$ goodness-of-fit probabilities for the fits obtained by WLS (*c*) and by ML (*d*). A probability between 20 and 80% indicates a reasonable fit. The error bars reflect the scatter of the values computed for ten independent simulations.

the computed variances are compared to the empirical ones.

The first set of simulations is for the diffraction patterns of rutile ($TiO_2$) (Sabine & Howard, 1982) with instrumental parameters typical of the high-resolution neutron diffractometer D2B of the ILL. The results are summarized in Fig. 10, and discussed below.

(*a*) The refined model is unbiased: both minimization techniques give the correct answer within error bars if the counting time is large enough. In the opposite case, weighted least squares (WLS) strongly underestimates the background value and the $\chi^2$ probability drops to zero while ML always gives the correct answer.

(*b*) The refined model is biased (given peak shape: pseudo-Voigt with mixing parameter $m = 0{\cdot}5$; fitted peak shape: Gauss or Lorentz): some fitted parameters clearly exhibit a bias (positive or negative depending on the use of Gaussian or Lorentzian peak shapes), others are barely affected; again both minimization techniques give similar results if the counting time is large enough and in addition the $\chi^2$ probability is very small. In the opposite case, WLS again strongly underestimates the background value and, for both algorithms, the probability of the $\chi^2$ test increases when the statistical fluctuations become large enough to 'hide' the bias.

Thus, for the case of this simple structure refined by the Rietveld method, we may conclude that:

The ML and the minimum $\chi^2$ methods give similar results for parameters most pertinent to the crystallographer (*i.e.* atomic parameters); this fully explains the extensive use made up to now of the latter minimization algorithm.

Agreement between the average computed (theoretical) and the observed (empirical) variances is good.

When the model is biased, increasing the counting time does not change the bias on the fitted parameters.

The statistical test most sensitive to a bias in the model is the $\chi^2$ test while the various $R$ factors (Hamilton, 1965), familiar to the crystallographer, are much less clear.

A second set of simulations refers to the complicated structure of bis(3-acetylamino-1,2,4-triazole-$O,N^4$)diaquacopper(II) sulfate pentahydrate (Biagini Cingi, Manotti Lanfredi, Tiripicchio, Haasnoot & Reedijk, 1989) [31 non-H atoms; 1700 reflections in the $\sin(\theta_{Bragg})/\lambda$ range from 0·0544 to 0·6226 Å$^{-1}$ ($\lambda = 1{\cdot}6$ Å)]. The fits were performed with the modified minimum $\chi^2$ algorithm only. The main findings of this latter study were that: high enough counting times are crucial even if the model is unbiased; and the profile bias in the model has large effects on all fit parameters (including atom coordinates which were fairly insensitive in the $TiO_2$ case).

In other words, since with real diagrams the profile bias - among others - is rarely absent, the main limitation of the Rietveld method is not the use of minimum $\chi^2$ instead of maximum likelihood but systematic errors. For example, the use of normalized residuals instead of the mere $(y_{obs} - y_{cal})$ may not seem a decisive improvement when the discrepancy between all observed and fitted peaks is large because of the lack of appropriate analytical functions.

Thus, the improvement brought by ML to the complicated case of structural fits through the Rietveld method will be, in most cases, hidden behind the dominating effects of a collection of, more or less, unavoidable systematic errors. However, even though the gain may not seem immediate, firstly, ML is basically more correct and is no more time consuming, secondly, it may handle dispersion and/or detector-response corrections in a natural way (Antoniadis & Berruyer, 1990). On the contrary, when the interest is in profile fitting itself, the gain with ML is obvious, especially for low counting rates and because of the set of diagnostic tests which may be designed.

## 7. Summary and discussion

The diffraction pattern data-analysis methods developed in this paper rely on the maximum-likelihood method. This is the natural analog of least-squares refinement for the Poisson situation, and indeed for all generalized forms of data fitting. The method provides estimates of the parameters and also estimated standard errors. The basic idea is simply to choose as estimates the values of the unknown parameters which maximize the probability density of the observed data. It turns out that the maximum-likelihood estimate of the parameters is also the minimizer of the Kullback–Leibler distance (see *e.g.* Rao, 1973) from the observed points $y_i$ to the fitted values $\eta_i(\boldsymbol{\beta})$. The main point to retain from this is that any error distribution generates its own Kullback–Leibler distance function (deviance) and hence its own analogy with least-squares fitting. The maximum-likelihood estimates are found by iterative search algorithms.

The common practice of correcting in diffraction studies the values of the parameter e.s.d.'s by multiplying them by a goodness-of-fit index in standard least-squares minimization procedures has been justified and made precise by the use of the quasi-likelihood principle for count data. When the model that is fitted is incomplete or incorrect, the goodness-of-fit index usually departs significantly from unity and the real parameter e.s.d.'s are in doubt in a statistical sense. Thence, in the present paper, we propose an estimate of the over- or underdispersion factor in quasi-likelihood models whose form does not rely on any assumptions about the theoretical models for the structure of diffraction peaks.

However, the simulation study has shown that this estimator is generally biased with its bias depending on how smooth is the theoretical profile function representing the peak shapes. Its interpretation and its uses require therefore some experience, especially in some delicate cases. A model that could conveniently model dispersion as well as mean response in a Poisson regression situation should be preferable for more practical uses. A theory of such models (double Poisson family models) is actually under study. Without pretending to be a fully developed theory, preliminary examples show the potential of this theory for 'robustifying' diffraction data analyses. The results will be reported elsewhere.

Whether or not quasi-maximum likelihood fits are better than least-squares refinement is certainly debatable in asymptotic situations but it is clear that for low count patterns they can be more powerful revealing the limitations of ordinary fitting least-squares procedures.* Moreover, if $g(\boldsymbol{\eta})$ is any continuous reparametrization of $\boldsymbol{\eta}$, the maximum-likelihood estimate of $g(\boldsymbol{\eta})$ is exactly $g(\hat{\boldsymbol{\eta}})$, a property which is not shared by the modified minimum $\chi^2$ estimators. Finally, one may prove that the minimum $\chi^2$ estimates are not always consistent (see Davis, 1985). The difference among the two estimators is most perceived in some given asymptotic situations where the counting time $T$ and the number of data points $n$ become larger at a certain rate. This is clearly illustrated in Fig. 4, where one can see how biased a modified minimum $\chi^2$ can be.

A more serious concern about our analyses involves the large values of the goodness-of-fit statistics in some data sets with large sample sizes, a problem that has been also addressed by Hill & Madsen (1984, 1986). When the sample size is large almost any structural model yields a highly significant $\chi^2$ or deviance value and the experimentalist receives little guidance as to which structural model actually fits the data better. The deviance, introduced in previous sections, is an effective device for preliminary data analysis, particularly when the experimentalist has many structural models under review, since it supplements the usual tests with a quantitative measure of the size of the discrepancy between the statistical model and observed data. The aim is to see if the discrepancy, although highly significant, is small enough in order that the model can be considered as providing a satisfactory approximation to the data. Another sensitive measure of the progress of a refinement is the Durbin–Watson $d$ statistic introduced first in Rietveld refinements by Hill & Flack (1987). The $d$ statistic, defined in our case through the generalized residuals of § 4.3, quantifies the serial correlation between adjacent residuals and provides a convenient means of assessing the reliability of the derived values of the parameter e.s.d.'s.

Our main conclusion is to permit *a priori* a model as general as possible, then use the data, together with the likely background, to suggest possible modifications to the hypothesized model. The best sort of mechanism of this type seems to be an accurate residual analysis which considers the adequacy of the modified model for each individual bin. This could solve many of the practical problems encountered when fitting long-counting-time or large-sample-size data.

In our approach, in view of the narrow profile step widths, we used the bin midpoints as 'continuous' data $\mu$ with the appropriate theoretical densities (Gauss, Cauchy, Lorentz *etc.*) as the basis of the likelihood. Such an approach ignores the grouping effect in the data-collection method and this midpoint approximation for low count data with fewer and larger bins could rise to significantly different results. Estimation of the parameters in such a situation requires numerical evaluation of the integrals of the theoretical peaks over the bins. This approach will be pursued and compared to the present 'continuous' approximation in a future work.

A final word about the computer program used for obtaining the figures and results of the previous section. The program, called *ABFfit*, is available either on DEC-VAX/VMS or on a Apple Macintosh computer. Convenience features include an integrated random simulator, the choice of several different profile functions and graphical initializations for the MLE iterative algorithm. Both programs are used successfully in several laboratories.

The authors are very indebted to G. McIntyre (Institut Laue–Langevin, Grenoble) who checked the English of the typescript and are also grateful to a referee for his very careful reading of the paper and his pertinent comments.

* *Note added in proof*: A recent Monte Carlo and theoretical study (Antoniadis & Hammersley, 1990) has demonstrated that WLS fits weighted by the inverse of the data suffer from a systematic bias which does not exist with MLE.

## References

AKAIKE, H. (1974*a*). *IEEE Trans. Autom. Control*, **AC-19**, 716–723.

AKAIKE, H. (1974*b*). *Int. Fed. Autom. Control*, **3**, 1877–1884.

ALBINATI, A. & WILLIS, B. T. M. (1982). *J. Appl. Cryst.* **15**, 361–374.

ANTONIADIS, A. & BERRUYER, J. (1990). *Least-Squares in Linear and Nonlinear Regression Models. Lecture Notes in Statistics.* Institut Laue–Langevin, Grenoble, France. To be published.

ANTONIADIS, A., BERRUYER, J. & FILHOL, A. (1985). *Simulation d'un Spectre avec un ou Plusieurs Pics de Bragg.* Tech. Rep. ILL No. 85AN19T. Institut Laue–Langevin, Grenoble, France.

ANTONIADIS, A. & HAMMERSLEY, A. (1990). In preparation.

BARD, Y. (1974). *Nonlinear Parameter Estimation.* New York: Academic Press.

BATES, D. M. & WATTS, D. G. (1981). *Technometrics*, **23**, 179–182.

BELSLEY, D. A., KUH, E. & WELSH, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.
BERNOULLI, D. (1861). *Biometrika*, **48**, 3-13.
BIAGINI CINGI, M., MANOTTI LANFREDI, A. M., TIRIPICCHIO, A., HAASNOOT, J. G. & REEDJIK, J. (1989). *Acta Cryst.* C**45**, 601-604.
BISHOP, Y. M., FIENBERG, S. E. & HOLLAND, P. W. (1975). *Discrete Multivariate Analysis.* Massachussets: MIT Press.
COOK, R. D. & WEISBERG, S. (1982). *Residuals and Influence in Regression.* London: Chapman & Hall.
COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics.* London: Chapman & Hall.
DAVIS, L. J. (1985). *Ann. Stat.* **13**, 947-957.
DONALDSON, J. R. & SCHNABEL, R. B. (1987). *Technometrics*, **29**, 67-83.
DRAPER, N. R. & SMITH, H. (1966). *Applied Regression Analysis.* New York: Wiley.
EFRON, B. (1982). *Ann. Stat.* **10**, 340-357.
FISHER, R. A. (1922). *Philos. Trans. R. Soc. London Ser. A*, **222**, 309-360.
FRIEDMAN, J. (1984). *A Variable Span Smoother.* Stanford Tech. Rep. No. 5. Department of Statistics, Stanford Univ., California, USA.
FRIEDMAN, J. & STUETZLE, W. (1981). *J. Am. Stat. Assoc.* **76**, 817-823.
GOLDSTEIN, A. (1965). *SIAM J. Control*, **3**, 147-151.
HAMILTON, W. C. (1964). *Statistics in Physical Sciences.* New York: Ronald Press.
HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502-510.
HILL, R. J. & FLACK, H. D. (1987). *J. Appl. Cryst.* **20**, 356-361.
HILL, R. J. & MADSEN, I. C. (1984). *J. Appl. Cryst.* **17**, 297-306.
HILL, R. J. & MADSEN, I. C. (1986). *J. Appl. Cryst.* **19**, 10-18.
JENSEN, J. L. (1981). *Scand. J. Stat.* **8**, 193-206.
KENDALL, M. G. & STUART, A. (1977). *The Advanced Theory of Statistics*, 4th ed. Vol. 1. London: Griffin.
LARNTZ, K. (1978). *J. Am. Stat. Assoc.* **73**, 253-263.
MCCULLAGH, P. & NELDER, J. A. (1983). *Generalized Linear Models.* London: Chapman & Hall.
MARDIA, K. V. (1972). *Statistics of Directional Data.* New York: Academic Press.
NELDER, J. A. & WEDDERBURN, R. M. W. (1972). *J. R. Stat. Soc. A*, **135**, 370-384.
RALSTON, M. L. & JENNRICH, R. I. (1978). *Technometrics*, **20**, 7-14.
RAO, C. R. (1973). *Linear Statistical Inference and its Applications.* London, New York: Wiley.
RIETVELD, H. M. (1969). *J. Appl. Cryst.* **2**, 65-71.
RODRIGUEZ-CARVAJAL, J. & MENARDE, M. (1989). Institut Laue-Langevin, Grenoble, France. Private communication.
SABINE, T. M. & HOWARD, C. J. (1982). *Acta Cryst.* B**38**, 701-702.
SAKATA, M. & COOPER, M. J. (1979). *J. Appl. Cryst.* **12**, 554-563.
SCHWARZ, G. (1978). *Ann. Stat.* **6**, 461-464.
WEDDERBURN, R. M. W. (1974). *Biometrika*, **61**, 439-447.
YOUNG, R. A., PRINCE, E. & SPARKS, R. A. (1982). *J. Appl. Cryst.* **15**, 357-359.

# SHORT COMMUNICATIONS

*Contributions intended for publication under this heading should be expressly so marked; they should not exceed about* 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

*Acta Cryst.* (1990). A**46**, 711-713

## Physical-property tensors and tensor pairs in crystals.

By S. Y. LITVIN* and D. B. LITVIN, *Department of Physics, The Pennsylvania State University, The Berks Campus, PO Box* 7009, *Reading, PA* 19610-6009, *USA*

(*Received* 8 *October* 1989; *accepted* 30 *April* 1990)

**Abstract**

The form of physical-property tensors of rank 0, 1 and 2 invariant under the 32 crystallographic point groups and their subgroups are tabulated. This constitutes the basis for the tensorial classification of domain pairs in ferroic crystals which is given *via* a group theoretical classification of the corresponding physical-property tensor pairs. We tabulate this classification of tensor pairs for all physical-property tensors of rank 0, 1 and 2, and domain point-group symmetry.

### 1. Introduction

A ferroic crystal contains two or more equally stable domains of the same structure but of different spatial orientation. These domains can coexist in a crystal and may be distinguished by the values of components of certain macroscopic tensorial physical properties of the domains (Aizu, 1973; Newnham, 1974; Newnham & Cross, 1974; Wadhawan, 1982). Aizu (1970; see also Cracknell, 1972) has given a tensorial classification of ferroic crystals based on a rank 1 physical-property tensor's ability to distinguish some or all of the domains. This method of classification of ferroic crystals was extended by Litvin (1984) to an arbitrary physical-property tensor and used to determine the tensorial classification of non-magnetic crystals for all physical-property tensors of rank less than or equal to four (Litvin, 1985).

In the study of the mutual relationships between domains, the simplest object one can consider is a pair of domains. A classification of domain pairs *via* a tensorial classification of corresponding tensor pairs of a *full* physical-property tensor characterizing the domains, where each domain is characterized by a unique form of the physical-property tensor, was introduced by Janovec (1972). This

* Mailing address: 1701 Bern Road, Apartment B2, Wyomissing, PA 19610, USA.

0108-7673/90/080711-03$03.00 