Acta Cryst. (1995). A**51**, 880–887

R-Free Likelihood-Based Estimates of Errors for Phases Calculated from Atomic Models

BY V. YU. LUNIN AND T. P. SKOVORODA

Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow Region, 142292 *Russia*

(*Received* 6 *April* 1995; *accepted* 22 *May* 1995)

Abstract

Reasonable assumptions about the statistical properties of errors in an atomic model lead to the probability distributions for the values of structure-factor phases. These distributions contain some generally unknown parameters reflecting how large the model errors are. These parameters must be determined properly to give realistic estimates of phase errors. Maximum-likelihood-based estimates suggested by Lunin & Urzhumtsev [*Acta Cryst.* (1984), A**40**, 269–277] are good for models not subjected to refinement but underestimate the errors when being used for refined models. The *R*-free methodology of Brünger [*Nature (London)*, (1992), **355**, 472–474] applied to the likelihood-function calculation allows realistic phase-error estimates to be obtained for both unrefined and refined models. These estimates may be used as an additional indicator in the refinement process.

0. Introduction

It is common practice to use preliminary atomic models to estimate structure-factor phases at different stages of structure determination. Such models may be incomplete, contain positional errors or even consist of pseudo-atoms possessing no structural meaning but just reflecting the electron-density distribution in the studied object (Agarwal & Isaacs, 1977; Lunin *et al.*, 1985; Subbiah, 1991; Lamzin & Wilson, 1993; Wilson & Agard, 1993; Lunin *et al.*, 1995). To combine the phase values calculated from such models with those obtained by other methods or to use them to construct Fourier syntheses, it is necessary to have realistic estimates of phase quality.

Some additional hypotheses about the statistical nature of coordinate errors, missing atoms *etc.* allow one to obtain information about unknown phases in the form of probability distributions, *e.g.* (for acentric reflections)

$$P(\varphi_s) \simeq \exp[(2\alpha_s/\beta_s)F_s^{obs}F_s^{mod}\cos(\varphi_s - \varphi_s^{mod})]. \quad (1)$$

Here, F_s^{obs} and F_s^{mod} are experimentally observed and calculated from the model structure-factor moduli and φ_s^{mod} are calculated from the model phases. Parameters α and β reflect the level of errors in the model and define the expected phase errors. They may be calculated provided distributions of errors are known, but in practical cases they are unknown parameters. The determination of appropriate values for these parameters is the key step when estimating the phase errors. Some methods of estimating these values were compared by Read (1986).

The main idea when estimating the level of model errors is to compare the structure-factor moduli calculated from the model with the experimentally observed ones. To put this idea into more definite shape, a widely used statistical method of likelihood maximization (Cox & Hinkley, 1974) was applied to estimate the parameters in the distributions (1) (Lunin, 1982; Lunin & Urzhumtsev, 1984; Read, 1986). Some other applications of maximum-likelihood methods in crystallography were discussed by Bricogne (1988, 1990). Testing of the method for known structures with independent random positional errors has shown that in these cases the method provides adequate estimates of phase errors, but being applied to atomic models subjected to refinement it has a tendency to predict much smaller phase errors than they really are.

Similar difficulties that arise when the usual crystallographic *R*-factor criterion is applied to structure-factor moduli calculated from refined atomic models were overcome by Brünger (1992, 1993) within the framework of the *R*-free methodology. In this approach, some structure factors are excluded from the refinement process and only these excluded reflections are used to calculate some control criterion value (*e.g.* *R* factor).

In the present paper, we discuss a way to combine these two ideas, namely maximum-likelihood estimates of phase errors and the *R*-free methodology in order to obtain realistic values for the expected phase errors for both models with independent positional errors and those subjected to refinement.

1. Likelihood-based estimates (LB estimates) for phase errors

1.1. *Statistical modeling of phase errors*

To obtain probabilities of different phase values, we must consider the structure involved not as the unique fixed one but as an element of an ensemble of possible structures with defined probabilities for each structure to

occur. In this paper, we suppose that we have a model consisting of M atoms in the positions $\{\mathbf{r}_j^{\rm mod}\}_{j=1}^M$ and possessing the temperature parameters $\{B_j^{\rm mod}\}_{j=1}^M$. We consider as possible structures the ones consisting of:

(*a*) M atoms whose positions and temperature parameters are $\mathbf{r}_j = \mathbf{r}_j^{\rm mod} + \Delta\mathbf{r}_j$, $B_j = B_j^{\rm mod} + \Delta B_j$, where $\Delta\mathbf{r}_j$ and ΔB_j are independent random errors; we suppose here that all $\Delta\mathbf{r}_j$ possess the same radially symmetric probability distribution and all ΔB_j have similar probability distributions too;

(*b*) $m = N - M$ atoms additionally, which are absent in the model and whose positions are supposed to be independently and uniformly distributed in the unit cell; the temperature factors of these atoms are also supposed to be independent random variables.

Under these assumptions, the moduli and phases of structure factors become random variables and we may speak about the joint probability distribution $P(F, \varphi)$ of the modulus and phase for every particular structure factor.

The central limit theorem of the theory of probabilities allows one to calculate for the distribution of a sum of N independent random variables the main term in its expansion into powers of $N^{-1/2}$. As in other papers (Luzzati, 1952; Sim, 1959; Srinivasan & Parthasarathy, 1976; Bricogne, 1984; Read, 1990), it is possible to show in the case considered that

$$P(F, \varphi) = (F/\pi\varepsilon\beta)\exp\{-[F^2 + \alpha^2(F^{\rm mod})^2 - 2\alpha F F^{\rm mod}\cos(\varphi - \varphi^{\rm mod})]/\varepsilon\beta\} \quad (2)$$

for acentric structure factors and

$$P(F, S) = (2\pi\varepsilon\beta)^{-1/2}\exp\{-[F^2 + \alpha^2(F^{\rm mod})^2 - 2\alpha F F^{\rm mod}S]/2\varepsilon\beta\} \quad (3)$$

for centric ones, where $S = \cos(\varphi - \varphi^{\rm mod})$ and takes for centric reflections values 1 or -1 only.

Parameters α and β depend on the reciprocal-vector $\mathbf{s}$ value and reflect the model quality. They are defined as

$$\alpha = \langle\cos(2\pi\mathbf{s}, \Delta\mathbf{r})\rangle_{\Delta\mathbf{r}}\langle\exp(-\Delta B\, s^2/4)\rangle_{\Delta B},$$

$$\beta = \sum_{j=1}^{M} f_j^2(s)\exp(-B_j^{\rm mod}s^2/2)[\langle\exp(-\Delta B\, s^2/2)\rangle_{\Delta B} - \alpha^2] + \sum_{j=M+1}^{N} f_j^2(s)\langle\exp(-B_j s^2/2)\rangle_{B_j}. \quad (4)$$

Here, $f_j(s)$ are atomic scattering factors and multipliers $\varepsilon(\mathbf{s})$ compensate different mean intensities for different types of reflections (they are equal to the number of transposed symmetry matrixes that leave the reciprocal vector $\mathbf{s}$ unchanged).

Owing to radical symmetry of the coordinate error, probability distribution parameters α and β depend on the $s = |\mathbf{s}|$ values only and, for a 'thin' spherical layer in the reciprocal space, we can think of parameters α and β as the same for all reflections. It is worth noting that all the uncertainties in the coordinate- and temperature-parameter errors and the model incompleteness have accumulated in the two parameters α and β. Furthermore, if we suppose that the observed $F^{\rm obs}$ values are present on a relative scale, this will only change the values of α and β parameters, which will contain one more unknown factor, namely the scale factor. So, we do not suppose below that $F^{\rm obs}$ values are reduced to the absolute scale and consider the problem of defining the scale coefficient as part of the more general problem of the determination of the unknown parameters α and β.

The joint probability distribution (2) allows one to obtain the conditional probability distribution for the structure-factor phase φ assuming that the modulus value F has the experimentally obtained value $F^{\rm obs}$:

$$P(\varphi|F = F^{\rm obs}) = \{2\pi I_0[2(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}]\}^{-1} \times \exp\{2(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}\cos(\varphi - \varphi^{\rm mod})\}. \quad (5)$$

This distribution allows one to obtain the usual best phase, $\varphi^{\rm best} = \varphi^{\rm mod}$, and the figure of merit

$$m = \langle\cos(\varphi - \varphi^{\rm mod})\rangle = I_1[2(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}]/I_0[2(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}], \quad (6)$$

or estimate the absolute phase error by its expected value

$$\langle|\varphi - \varphi^{\rm mod}|\rangle = \int_0^{\pi}\varphi\exp\{2(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}\cos(\varphi)\}\,{\rm d}\varphi \times \{\pi I_0[2(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}]\}^{-1}. \quad (7)$$

For centric reflections, the last three formulae take the form

$$P(S|F = F^{\rm obs}) = \{2\cosh[(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}]\}^{-1} \times \exp\{(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}S\} \quad (8)$$

$$m = \tanh[(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}], \quad (9)$$

$$\langle|\varphi - \varphi^{\rm mod}|\rangle = \{(2/\pi)\exp[(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}] \times \cosh[(\alpha/\varepsilon\beta)F^{\rm obs}F^{\rm mod}]\}^{-1}. \quad (10)$$

The main problem in applying these formulae is to find values of α and β parameters that reflect adequately the errors in the model. We discuss below a way to obtain these values.

1.2. *Maximum-likelihood estimates for distribution parameters*

Consider now reflections from some 'thin' spherical layer in the reciprocal space [parameters α and β in (2) and (3) are the same for all such reflections]. The joint

probability distributions allow one to obtain marginal probability distributions of modulus values by integrating (2) or (3) with respect to unknown phases:

$$P(F) = (2F/\varepsilon\beta)\exp\{-[F^2 + \alpha^2(F^{\rm mod})^2]/\varepsilon\beta\} \times I_0[2(\alpha/\varepsilon\beta)F\,F^{\rm mod}] \quad (11)$$

or

$$P(F) = (2/\pi\varepsilon\beta)^{1/2}\exp\{-[F^2 + \alpha^2(F^{\rm mod})^2]/2\varepsilon\beta\} \times \cosh[(\alpha/\varepsilon\beta)F\,F^{\rm mod}]. \quad (12)$$

These distributions contain the same unknown parameters α and β as (5) and (8) do. Let us consider experimental modulus values ($F_{\rm s}^{\rm exp}$) as realizations of the random variables $F_{\rm s}$ distributed in accordance with (11) or (12) and try to find parameter values that are consistent with these realizations.

The maximum-likelihood method suggests as estimates for values of unknown distribution parameters those that maximize 'the probability' of getting for the set of moduli $\{F_{\rm s}\}$ the actually obtained values $\{F_{\rm s}^{\rm exp}\}$. To be more precise, let the joint probability distribution $P_{\rm joint}(\{F_{\rm s}\};\alpha,\beta)$ for a set $\{F_{\rm s}\}$ of moduli depend on some unknown parameters α and β. Then, the maximum-likelihood estimates are those maximizing the likelihood function

$$L(\alpha,\beta) = P_{\rm joint}(\{F_{\rm s}^{\rm exp}\};\alpha,\beta) \quad (13)$$

or, equivalently, its logarithm. Suppose that moduli $\{F_{\rm s}\}$ for different structure factors are mutually independent [or, which leads to the same result, using the 'diagonal approximation' (Bricogne, 1993) for their joint probability distribution], we get the logarithm of the likelihood function in the form

$$\begin{aligned}\ln L(\alpha,\beta) = {} & \sum_{\rm s}\ln\{(2F_{\rm s}/\varepsilon_{\rm s}\beta) \\ & \times \exp\{-[(F_{\rm s}^{\rm obs})^2 + \alpha^2(F_{\rm s}^{\rm mod})^2]/\varepsilon_{\rm s}\beta\} \\ & \times I_0[2(\alpha/\varepsilon_{\rm s}\beta)F_{\rm s}^{\rm obs}F_{\rm s}^{\rm mod}]\} \\ & + \sum_{\rm s}\ln\{(2/\pi\varepsilon_{\rm s}\beta)^{1/2} \\ & \times \exp\{-[(F_{\rm s}^{\rm obs})^2 + \alpha^2(F_{\rm s}^{\rm mod})^2]/2\varepsilon_{\rm s}\beta\} \\ & \times \cosh[(\alpha/\varepsilon_{\rm s}\beta)F_{\rm s}^{\rm obs}F_{\rm s}^{\rm mod}]\}, \end{aligned} \quad (14)$$

where the first sum is extended over acentric reflections and the second over centric ones.

Details of the method used for maximization of this function are discussed in Appendix *A*. We note here only that there may be two different cases depending on the value of Ω defined below [see (34)], which may be considered as the covariance value between the weighted observed and the model intensities. If $\Omega < 0$, the maximum of $L(\alpha,\beta)$ is attained at point $\alpha = 0$, $\beta = B$ and (6) and (9) give zero figures of merit for all the $\varphi_{\rm s}^{\rm mod}$ phases in the considered reciprocal-space layer, *i.e.* the present model does not produce information about the phases of reflections of this layer. If $\Omega > 0$, there exist some nontrivial optimal values for α and β and phases $\varphi_{\rm s}^{\rm mod}$ have nonzero figures of merit. It should be mentioned, too, that parameters α and β for different layers in the reciprocal space are different and must be determined separately for every zone in s^2.

Below we call the likelihood-based (LB) estimates for phase errors the estimates that follow from distributions (5) and (8) with parameters α and β determined from maximization of the likelihood function (14).

1.3. *Testing of methods for prediction of the level of the phase errors*

After the parameters α and β defining distribution (5) or (8) for a particular reflection have been found, we can calculate the expected phase difference between the true phase and the model one. In a test case, when the exact phases (*e.g.* ones calculated from the refined model) are known, we also know the real phase differences $\varphi^{\rm ex} - \varphi^{\rm mod}$. We cannot compare these values immediately to judge how well the parameters α and β were estimated since the phase error is of a statistical nature and its unique value is usually not representative. The more reliable figure is the value of some statistic, *e.g.* averaged phase error for a large group of reflections. In this case, the standard deviation for its value decreases as the inverse square root of the reflection number, therefore for well defined α and β values the averaged expected error must not differ greatly from the averaged difference between $\varphi_{\rm s}^{\rm ex}$ and $\varphi_{\rm s}^{\rm mod}$ phases.

Following the above reasoning in subsequent experiments with the phase-error-level prediction, we divided the interval in s^2 values into 20 equal bins. For every bin, the averaged value of expected phase errors

$$T_k^{\rm pred} = (1/m_k)\sum_{j=1}^{m_k}\langle|\varphi_{{\rm s}_j} - \varphi_{{\rm s}_j}^{\rm mod}|\rangle \quad (15)$$

was calculated and compared with the averaged value of the real phase differences between the exact phases and the model ones:

$$T_k^{\rm real} = (1/m_k)\sum_{j=1}^{m_k}|\varphi_{{\rm s}_j}^{\rm ex} - \varphi_{{\rm s}_j}^{\rm mod}|. \quad (16)$$

Here, k is the bin number; $\{{\rm s}_j\}_{j=1}^{m_k}$ are reciprocal-lattice points of the corresponding layer; the expected values $\langle\,\rangle$ in (15) are calculated in accordance with (7) and (10).

2. Comparison of the LB-estimated and real phase errors

2.1. *Preliminary remarks*

The tests presented in this paper were performed with the protein G structure. This protein was crystallized in space group $P2_12_12_1$ with unit-cell dimensions

34.9 x 40.3 x 42.2 Å. The experimental moduli of resolution up to 1.8 Å and the atomic model containing 564 non-H atoms and 96 water O atoms were kindly supplied for test purposes by E. Dodson. We refer to this model below as the PDB model.

When testing the developed methods with models subjected to refinement, the refinements were aimed at checking the closeness of predicted and real errors in different circumstances rather than at getting an ideal model and were interrupted at intermediate stages.

In the first set of tests, the moduli of the structure factors calculated from the PDB model were considered as 'observed' ones. Water molecules were not included in these calculations.

2.2. *Independent coordinate errors*

Fig. 1 shows the quality of phase-error-level prediction by the distributions (5) and (8) with maximum-likelihood-determined parameters in the cases when independent random shifts were introduced into the model coordinates. These shifts had a Gaussian distribution with zero mean value and the equal variances varied for different tests. Two starting models were tested with the mean absolute values of coordinate errors of 0.39 and 0.79 Å, respectively. In both cases, predicted errors were very close to the real ones.

2.3. *Refined atomic models*

At the next stage of the experiments, the model with 0.79 Å starting mean coordinate error was subjected to reciprocal-space refinement with the use of the *FROG* refinement program (Urzhumtsev, Lunin & Vernoslova, 1989). The refinement consisted of four steps with the upper limits of resolution zones (in s^2 values) of 0.093, 0.185, 0.247, 0.309 $Å^{-2}$, respectively. Three cycles of steepest-descent minimization were made for every zone. Geometrical constraints were not used in this refinement;

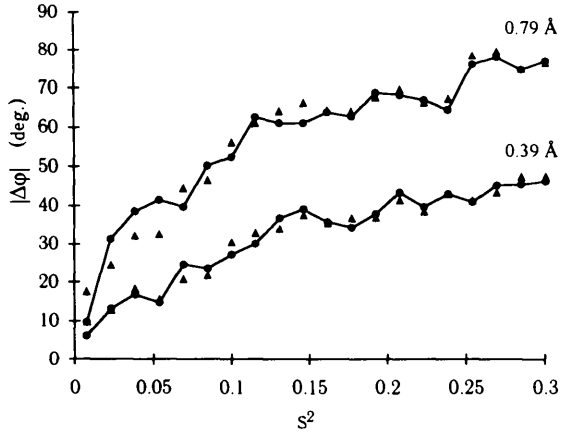

Fig. 1. The averaged values of the LB-estimated (▲) and real (–●–) phase errors for atomic models with independent coordinate shifts. The curves are marked by the mean absolute value of the model coordinate shifts.

however, the stereochemical quality of the model was improved during the refinement. The final value of the *R* factor was 0.217 (in the resolution zone up to 1.8 Å) and the mean coordinate error was reduced to 0.55 Å. Fig. 2 shows the changing of the real and predicted errors with the extension of the resolution zone.

It follows from these tests that, for the models subjected to refinement, LB estimates of phase errors are valid for reflections of the resolution zones not included into the refinement but are substantially less than real errors for reflections used in the refinement process.

3. *R*-free LB estimates

A picture similar in appearance was obtained by Brünger (1992) in his studies of the correlation of *R* factors with the real model quality. He demonstrated that, if all the reflections are used in the refinement, then the *R* factor does not reflect adequately the model quality but can be rather small for incorrect models. He has also shown that the situation changes drastically if some set of reflections (we call it the control group) is excluded from the refinement and the *R* factor is calculated for reflections of this control group only. This *R*-factor value (called by the author the *R*-free factor) reflects the model quality better than the ordinary one. In our tests presented above, reflections of higher-resolution zones not included in the refinement process at the intermediated stages may be considered as such a control group. The test results show that α and β values determined from the these reflections provide us with realistic estimates of phase errors. The extension of this observation is an attempt to exclude from the refinement a number of reflections distributed evenly in the reciprocal space and use these reflections only to estimate α and β parameters.

We call below the *R*-free LB estimates of phase errors those obtained from the distributions (5) and (8) with α and β parameters, which were obtained by maximization of the likelihood function (14) extended over reflections that were not included in the refinement.

4. Comparison of the *R*-free LB-estimated and real phase errors

4.1. *Simulated data*

In this series of tests, the moduli calculated from the PDB model without water molecules were considered as observed ones. Fig. 3 shows the results of the use of the *R*-free LB estimates for refined models. The model with mean coordinate error 0.79 Å was taken as the starting model and then subjected to four steps of refinement in the resolution zones defined above in §2.3. A randomly chosen half of the reflections was used in the refinement process and the other half was used as the control group to estimate α and β parameters in the distributions (5)

and (8). It should be noted that as many as 50% of the reflections were included in the control group to make the results more clear-cut. We also show below a result of the use of a smaller control group. Two strategies of refinement were tried. At the first attempt, geometrical restraints were not applied to the model. A relatively small number of reflections used resulted in a failed refinement. The final model had an R factor of 0.176 for reflections included in the refinement but almost unimproved mean coordinate error, which was equal to 0.72 Å. The mean phase errors calculated from the final model were the same as the starting ones and this was reflected adequately by their R-free LB estimates. In the second run, the geometrical restraints were used in the refinement and this resulted in the tendency of the phases to be improved, which was reflected by the R-free LB estimates.

Fig. 4 shows the real and predicted mean phase errors separately for reflections included and not included in the refinement. It follows from these plots that the distributions (5) and (8) produce correct estimates of phase errors for both types of reflection provided the proper values for α and β parameters are defined. It is worthy of note, too, that the averaged values of phase errors are nearly the same for both types of reflection, that is, the phases of the included reflections are being improved in the refinement process no better than the excluded ones.

4.2. *Experimental data*

The last set of experiments was performed with the experimental F^{obs} values. It should be noted that in this case we cannot perform as clear a comparison of the real and predicted phase errors as before since we do not know the exact phase values. In the previous tests, the 'exact' phases φ_s^{ex} and 'observed' moduli were calculated from the same model and we tried to judge the closeness of φ_s^{mod} and φ_s^{ex} phases by means of calculating (15) and comparing it with (16) calculated with the same phases. Now we try to predict the deviation of φ_s^{mod} phases from some, actually unknown, phases φ_s^{true}, but use as a check criterion (16), where other phases φ_s^{ex} are used. The most we can do is to include in the calculation of φ_s^{ex} phases all the structure atoms, including water molecules, to reduce the differences between φ_s^{ex} and φ_s^{true}.

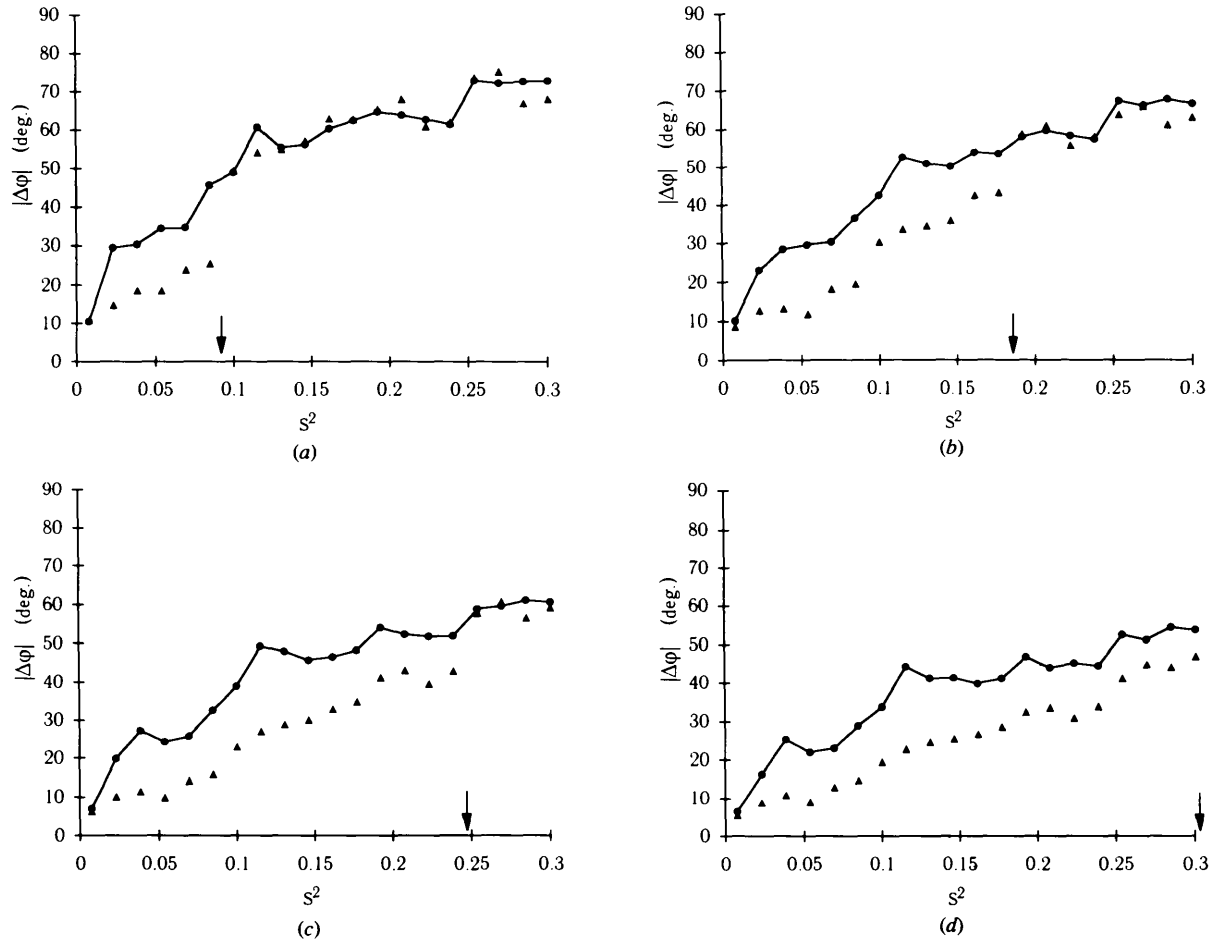

Fig. 2. Changing of averaged values of LB-estimated (▲) and real (–●–) phase errors during refinement. The arrows indicate the upper limit of the resolution zone for reflections included in the refinement.

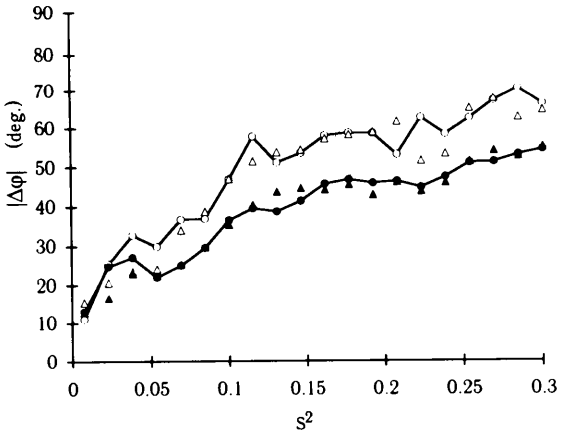

Fig. 3. The averaged values of R-free LB-estimated and real phase errors after unrestrained and restrained refinements with simulated data. The control group contained 50% of the reflections. –○– unrestrained refinement, real errors; △ unrestrained refinement, estimated errors; –●– restrained refinement, real errors; ▲ restrained refinement, estimated errors.

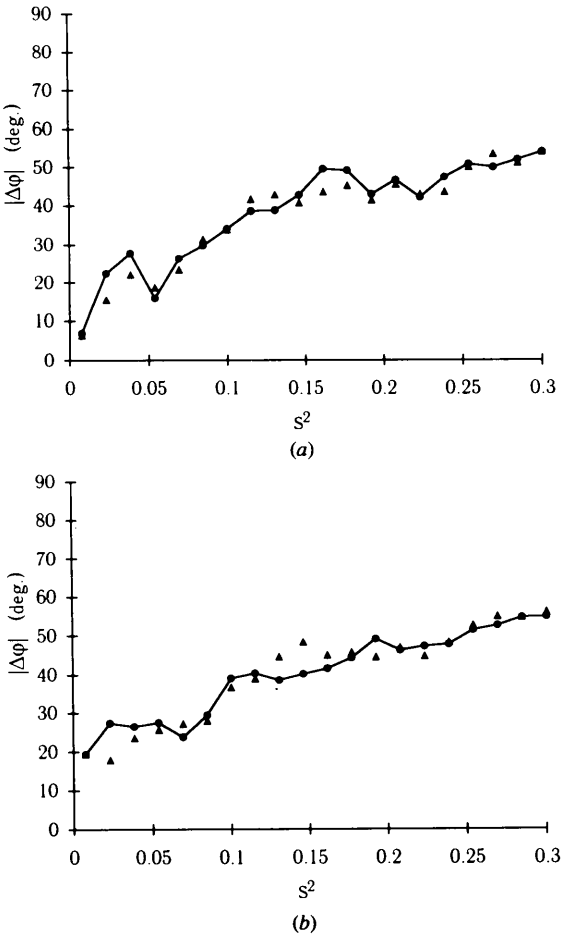

Fig. 4. The averaged values of R-free LB-estimated (▲) and real (–●–) phase errors after refinement with simulated data for different types of reflections: (*a*) the reflections included in the refinement; (*b*) the reflections of the control group. The control group contained 50% of the reflections.

The starting model was the same as before. It is worth noting that water molecules were not included in this model, so both the coordinate errors and the model incompleteness influenced the discrepancies between φ^{ex} and φ^{mod} phases. In this refinement, the experimental F_s^{obs} values were used and the phases φ_s^{ex} in (16) were calculated with the use of all the water atoms. The same refinement protocol has resulted in a model with R factor 0.26 and mean coordinate error 0.40 Å. Fig. 5 shows the real and predicted phase errors for this case.

Fig. 5(*b*) shows similar results when only 10% of the reflections were used as the control group during the refinement process.

5. Discussion

Generally speaking, there are two main reasons why LB estimates, which are adequate for models possessing independent coordinate errors, become invalid for refined models. The first is that the distributions (5) and (8) were

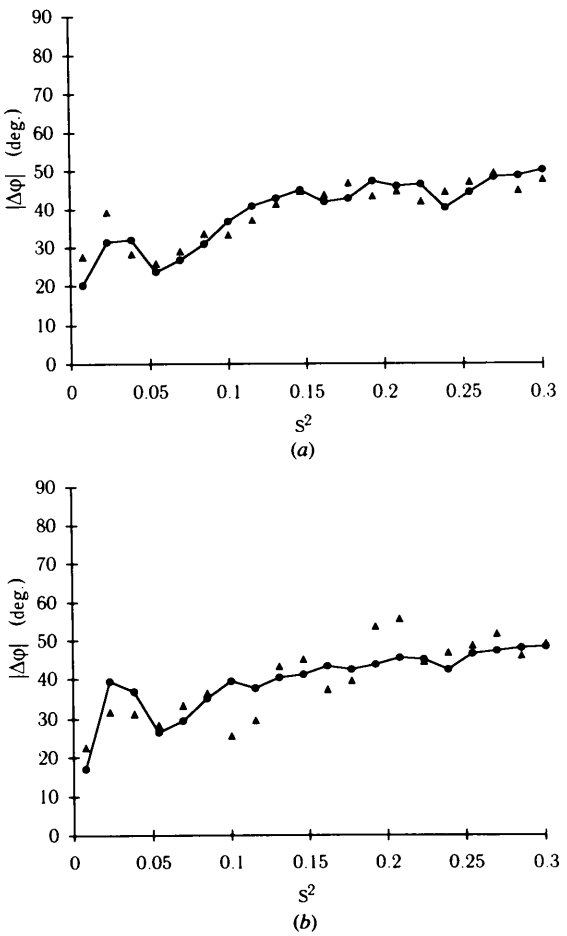

Fig. 5. The averaged values of R-free LB-estimated (▲) and real (–●–) phase errors after refinement with experimental data: (*a*) 50% of reflections were included in the control group; (*b*) 10% of reflections were included in the control group.

derived under the assumption that coordinate shifts are independent, while for a refined model they are constrained by moduli of structure factors included in the refinement. Nevertheless, as follows from Fig. 2, these correlations of the shifts do not prevent us from obtaining the true expected phase-error values for the reflections not participating in the refinement. Furthermore, Fig. 4 shows that (5) and (8) can represent correctly the expected phase errors for both types of reflection (*i.e.* included and excluded from the refinement) if the distribution parameters are chosen properly. So we may still use these distributions to estimate phase errors even for refined models, provided the appropriate values for α and β were found.

The other possible reason why LB estimates failed when being applied to refined models is that we tried to determine α and β parameters from a badly defined likelihood function. The values F_s^{obs} and F_s^{mod} present in (14) were artificially brought close together for the reflections included in the refinement so the corresponding distributions (11) and (12) could not be considered as independent when calculating the likelihood. The exclusion of these terms from the calculations radically altered the quality of the error prediction.

The *R*-free LB estimates reflect not only errors in calculated phases but the model quality too. So these estimates may be used as an additional (to *R*-free factor) tool for realistic judging of a model quality during the refinement process.

This work was supported in part by ISF grant RMZ000 and RFFI grant 94-04-12844.

APPENDIX *A*
Maximum-likelihood estimates for distribution parameters

A1. Maximization problem

Let us consider the reflections belonging to a spherical layer in the reciprocal space and let n_c and n_a be the numbers of centric and acentric reflections in this layer and $n = n_a + n_c$. Let $F_{e,j}$ and $F_{m,j}$ $(j = 1, \ldots, n)$ be the values of experimental and model structure-factor moduli for these reflections. We will use below the following notations:

$$b_j = F_{e,j}F_{m,j}/\varepsilon_j, \tag{17}$$

$$\begin{aligned} A &= [1/(2n_a + n_c)]\sum_{j=1}^{n} w_j F_{m,j}^2/\varepsilon_j, \\ B &= [1/(2n_a + n_c)]\sum_{j=1}^{n} w_j F_{e,j}^2/\varepsilon_j, \\ C &= [1/(2n_a + n_c)]\sum_{j=1}^{n} w_j F_{m,j}F_{e,j}/\varepsilon_j, \\ D &= [1/(2n_a + n_c)]\sum_{j=1}^{n} w_j F_{m,j}^2 F_{e,j}^2/\varepsilon_j^2, \end{aligned} \tag{18}$$

where the weights w_j are equal to 2 for acentric reflections and to 1 for centric ones.

Let us introduce new variables u and v connected with α and β by

$$u^2 = 1/\beta, \quad v^2 = \alpha^2/\beta \quad (u \geq 0, \quad v \in \mathbb{R}^1). \tag{19}$$

In this notation, the problem of maximization of the function (14) is reduced to maximization of the function

$$\begin{aligned} Q(u, v) = 2\ln(u) - Bu^2 - Av^2 \\ + [1/(2n_a + n_c)]\sum_{j=1}^{n} w_j \mu_j(uvb_j), \end{aligned} \tag{20}$$

with respect to u and v values, where the functions $\mu_j(x)$ are defined as

$$\mu_j(x) = \begin{cases} \ln[I_0(2x)] & \text{for acentric reflections,} \\ 2\ln[\cosh(x)] & \text{for centric reflections.} \end{cases} \tag{21}$$

Necessary conditions for the maximum point are in this case

$$\begin{cases} \partial Q/\partial u = 2/u - 2Bu + 2v\Lambda(uv) = 0, \\ \partial Q/\partial v = \quad - 2Av + 2u\Lambda(uv) = 0, \end{cases} \tag{22}$$

where the function $\Lambda(\tau)$ is defined as

$$\Lambda(\tau) = [1/(2n_a + n_c)]\sum_{j=1}^{n} w_j b_j H_j(\tau b_j), \tag{23}$$

$$H_j(x) = \begin{cases} I_1(2x)/I_0(2x) & \text{for acentric reflections,} \\ \tanh(x) & \text{for centric reflections.} \end{cases} \tag{24}$$

Calculating the sum and the difference of the first and second equations in (22) multiplied by u and v, respectively, we obtain

$$\begin{cases} 1 + 2uv\Lambda(uv) = Av^2 + Bu^2 \\ Bu^2 - Av^2 = 1. \end{cases} \tag{25}$$

From the last equation, it is possible to see that

$$\begin{aligned} (Bu^2 + Av^2)^2 &= (Bu^2 - Av^2)^2 + 4ABu^2v^2 \\ &= 1 + 4ABu^2v^2, \end{aligned} \tag{26}$$

so the first equation in (25) results in

$$1 + 2uv\Lambda(uv) = (1 + 4ABu^2v^2)^{1/2}. \tag{27}$$

Let us introduce now a new variable t connected to u and v as

$$t = uv. \tag{28}$$

Equations (25) and (27) mean that at the maximum point t must satisfy the equation

$$G(t) \equiv (1 + 4ABt^2)^{1/2} - 1 - 2t\Lambda(t) = 0; \tag{29}$$

and u and v are equal to

$$v^2 = [(1 + 4ABt^2)^{1/2} - 1]/2A,$$
$$u^2 = [(1 + 4ABt^2)^{1/2} + 1]/2B. \quad (30)$$

The optimal α and β values can be calculated then as

$$\alpha = v/u, \quad \beta = 1/u^2. \quad (31)$$

A2. Mathematical analysis

We exclude from the analysis the singular case when there exists a scale factor λ such that $F_j^{obs} = \lambda F_j^{mod}$ for all the reflections, *i.e.* the model is ideal.

The function $G(t)$ is even, so we can consider it for $t \geq 0$ only.

It is easy to see that $G(t) = 0$ always has the trivial solution $t = 0$, *i.e.* $v = 0$, $u = B^{-1/2}$ or $\alpha = 0$, $\beta = B$.

Using asymptotic formulae for the modified Bessel functions, we can obtain, for small values of t,

$$G(t) = -2(D - AB)t^2 + O(t^4) \quad \text{for } t \to 0 \quad (32)$$

and, for large t,

$$\lim_{t\to\infty}(1/t)G(t) = 2[(AB)^{1/2} - C]. \quad (33)$$

The value of $(AB)^{1/2} - C$ is always positive owing to Cauchy inequality, so $G(t) = 0$ has at least one nontrivial solution if the value

$$\Omega = D - AB \quad (34)$$

is positive.

It is possible to show, too, that the function $Q(u, v)$ tends to $-\infty$ when the point (u, v) tends to infinity. So, the maximum value is attained at an inner point.

In the vicinity of $u = B^{-1/2}$, $v = 0$,

$$Q(u, v) = (-\ln B - 1) - 2B(u - B^{-1/2})^2 + (\Omega/B)v^2 + \ldots, \quad (35)$$

so if $\Omega < 0$ this point is the maximum point, but if $\Omega > 0$ this is a saddle point and the maximum is attained at a point corresponding to the nontrivial solution of $G(t) = 0$.

References

AGARWAL, R. C. & ISAACS, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.
BRICOGNE, G. (1984). *Acta Cryst.* A**40**, 410–445.
BRICOGNE, G. (1988). *Acta Cryst.* A**44**, 517–545.
BRICOGNE, G. (1990). *Acta Cryst.* A**46**, 284–297.
BRICOGNE, G. (1993). *Acta Cryst.* D**49**, 37–60.
BRÜNGER, A. T. (1992). *Nature (London)*, **355**. 472–474.
BRÜNGER, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics*. Imperial College, London, England.
LAMZIN, V. S. & WILSON, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
LUNIN, V. YU. (1982). *The Use of Maximum Likelihood Approach to Estimate Phase Errors in Protein Crystallography*. Pushchino, Russia.
LUNIN, V. YU., LUNINA, N. L., PETROVA, T. E., VERNOSLOVA, E. A., URZHUMTSEV, A. G. & PODJARNY, A. (1995). *Acta Cryst.* Submitted.
LUNIN, V. YU. & URZHUMTSEV, A. G. (1984). *Acta Cryst.* A**40**, 269–277.
LUNIN, V. YU., URZHUMTSEV, A. G., VERNOSLOVA, E. A., CHIRGADZE, YU. N., NEVSKAYA, N. A. & FOMENKOVA, N. P. (1985). *Acta Cryst.* A**41**, 166–171.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
READ, R. J. (1990). *Acta Cryst.* A**46**, 900–912.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
SUBBIAH, S. (1991). *Science*, **252**, 128–133.
URZHUMTSEV, A. G., LUNIN, V. YU. & VERNOSLOVA, E. A. (1989). *J. Appl. Cryst.* **22**, 500–506.
WILSON, C. & AGARD. D. A. (1993). *Acta Cryst.* A**49**, 97–104.

Acta Cryst. (1995). A**51**, 887–897

The Analytical Calculation of Absorption in Multifaceted Crystals

BY R. C. CLARK

Department of Mathematical Sciences, University of Aberdeen, Aberdeen AB9 2TY, Scotland

AND J. S. REID

School of Physics, University of Aberdeen, Aberdeen AB9 2UE, Scotland

(*Received* 20 *April* 1995; *accepted* 5 *June* 1995)

Abstract

The exact analytic method of evaluating the absorption during scattering in multifaceted convex crystals is developed in a way that permits efficient computation. A fast and accurate algorithm is given for finding the Howells polyhedra whose determination is fundamental to the analytic method. The algorithm allows for the evaluation of cases when the sample is only partly illuminated, can be adapted to more general situations