Acta Cryst. (1999). A**55**, 525–532

The joint probability distribution function of structure factors with rational indices. V. The estimates

Carmelo Giacovazzo,[a]* Dritan Siliqi,[a,b] Cristina Fernández-Castaño[c] and Giuliana Comunale[d]

[a]*IRMEC c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy,* [b]*Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania,* [c]*Departamento de Cristalografía, Instituto de Química-Física 'Rocasolano', CSIC, Madrid, Spain, and* [d]*Centro di Geodinamica, Università della Basilicata, Via Anzio, Potenza, Italy. E-mail: c.giacovazzo@area.ba.cnr.it*

(*Received 15 June 1998; accepted 10 November 1998*)

Abstract

The probabilistic formulae [Giacovazzo, Siliqi & Fernández-Castaño (1999). *Acta Cryst.* A**55**, 512–524] relating standard and half-integral index reflections are modified for practical applications. The experimental tests prove the reliability of the probabilistic relationships. The approach is further developed to explore whether the moduli of the half-integral index reflections can be evaluated in the absence of phase information; *i.e.* by exploiting the moduli of the standard reflections only. The final formulae indicate that estimates can be obtained, even though the reliability factor is a constant.

1. Symbols and notation

Papers by Giacovazzo & Siliqi (1998), Giacovazzo, Siliqi, Carrozzini *et al.* (1999), Giacovazzo, Siliqi, Altomare *et al.* (1999) and Giacovazzo, Siliqi & Fernández-Castaño (1999) will be referred as papers I, II, III and IV, respectively. The notation adopted herein is essentially that used in paper IV, to which the reader is referred.

2. Introduction

Papers I and II of this series were dedicated to the intensity statistics of structure factors with rational indices, of which Wilson statistics for standard reflections is a particular case. In papers III and IV, the joint probability distribution functions of structure factors with rational indices were derived. Probabilistic relationships relating the real and imaginary parts of the structure factors were obtained, which encompass those derived *via* the Hilbert-transform method (Mishnev, 1993, 1996; Zanotti *et al.*, 1996). The first part of this paper is devoted to the applicative aspects of our theory.

In the second part of this paper, the probabilistic theory described in paper IV will be further developed to estimate the intensities of the half-integral index reflections in the absence of phase information on the standard reflections. This problem was discussed by Mishnev (1996) in the framework of the Hilbert-transform method; he concluded that, for a usual crystal structure, it is impossible to estimate the intensities of the half-integral index reflections owing to the fact that the casual Fourier transform condition (Toll, 1956; Wu & Ohmura, 1962) is not satisfied. This condition is not necessary in our probabilistic approach; therefore, it is interesting to explore how the problem is solved by the probabilistic techniques. Some applications to real cases will also be described.

3. New simplified formulae

In paper IV, the following conclusive formulae were derived for the three-dimensional canonical case [here, we explicitly set to zero the vanishing cumulants $K_{10}(\mathbf{p})$ and $K_{10}(\mathbf{q})$]:

$$\begin{aligned}\langle A_{\mathbf{p}}|\{A_{\mathbf{q}},B_{\mathbf{q}}\}\rangle &= \sum_{\mathbf{q}}\frac{K_{14}(\mathbf{p},\mathbf{q})}{K_{02}(\mathbf{q})}[B_{\mathbf{q}}-K_{01}(\mathbf{q})]\\ &= \sum_{\mathbf{q}}\left[\frac{\sum_{11}(\mathbf{p},\mathbf{q})}{\sum_2(\mathbf{q})(1-2s^2_{\mathbf{q}})}(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{q}-\mathbf{p}})\right]\\ &\quad\times[B_{\mathbf{p}}-\textstyle\sum_1(\mathbf{q})s_{\mathbf{q}}], \qquad \text{(CPR1)}\end{aligned}$$

$$\begin{aligned}V_{A_{\mathbf{p}}} &= K_{20}(\mathbf{p})-\sum_{\mathbf{q}}\frac{K^2_{14}(\mathbf{p},\mathbf{q})}{K_{02}(\mathbf{q})}\\ &= \tfrac{1}{2}\left\{\textstyle\sum_2(\mathbf{p})-\displaystyle\sum_{\mathbf{q}}\left[\frac{\sum^2_{11}(\mathbf{p},\mathbf{q})}{\sum_2(\mathbf{q})(1-2s^2_{\mathbf{q}})}\right.\right.\\ &\quad\left.\left.\times(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{q}-\mathbf{p}})^2\right]\right\}, \qquad \text{(CPR2)}\end{aligned}$$

$$\begin{aligned}\langle B_{\mathbf{p}}|\{A_{\mathbf{q}},B_{\mathbf{q}}\}\rangle &= K_{01}(\mathbf{p})+\sum_{\mathbf{q}}\frac{K_{23}(\mathbf{p},\mathbf{q})}{K_{20}(\mathbf{q})}A_{\mathbf{q}}\\ &= \textstyle\sum_1(\mathbf{p})s_{\mathbf{p}}\\ &\quad+\displaystyle\sum_{\mathbf{q}}\left[\frac{\sum_{11}(\mathbf{p},\mathbf{q})}{\sum_2(\mathbf{q})}(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{p}-\mathbf{q}})\right]A_{\mathbf{q}}, \qquad \text{(CPR3)}\end{aligned}$$

$$V_{B_\mathbf{p}} = K_{02}(\mathbf{p}) - \sum_\mathbf{q} \frac{K_{23}^2(\mathbf{p},\mathbf{q})}{K_{20}(\mathbf{q})}$$
$$= \tfrac{1}{2}\left\{ \textstyle\sum_2(\mathbf{p})(1-2s_\mathbf{p}^2) - \sum_\mathbf{q}\left[\frac{\sum_{11}^2(\mathbf{p},\mathbf{q})}{\sum_2(\mathbf{q})}(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{p}-\mathbf{q}})^2\right]\right\}. \quad \text{(CPR4)}$$

The coefficients $\sum_1(\mathbf{p})$, $\sum_2(\mathbf{p})$, $\sum_1(\mathbf{q})$, $\sum_2(\mathbf{q})$ and $\sum_{11}(\mathbf{p},\mathbf{q})$ depend on the resolution of $F_\mathbf{p}$ and $F_\mathbf{q}$; therefore, their values have to be calculated for each pair $(\mathbf{p},\mathbf{q})$. Some approximations may be used to reduce the computing time: the most suitable involve the use of normalized structure factors, of which the real and imaginary parts are defined as

$$A_\mathbf{p}^N = A_\mathbf{p}\Big/\textstyle\sum_2^{1/2}(\mathbf{p}), \quad A_\mathbf{q}^N = A_\mathbf{q}\Big/\sum_2^{1/2}(\mathbf{q}),$$
$$B_\mathbf{p}^N = B_\mathbf{p}\Big/\textstyle\sum_2^{1/2}(\mathbf{p}), \quad B_\mathbf{q}^N = B_\mathbf{q}\Big/\sum_2^{1/2}(\mathbf{q}).$$

For example, (CPR1) becomes

$$\langle A_\mathbf{p}^N|\{A_\mathbf{q}^N,B_\mathbf{q}^N\}\rangle = \sum_\mathbf{q}\frac{\sum_{11}(\mathbf{p},\mathbf{q})}{\sum_2^{1/2}(\mathbf{p})\sum_2(\mathbf{q})(1-2s_\mathbf{q}^2)} \times (s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{q}-\mathbf{p}})\textstyle\sum_2^{1/2}(\mathbf{q}) \times \left\{B_\mathbf{q}^N - \left[\sum_1(\mathbf{q})\Big/\sum_2^{1/2}(\mathbf{q})\right]s_\mathbf{q}\right\}.$$

We can now introduce the approximations (which are exact for equal atom structures):

$$\textstyle\sum_{11}(\mathbf{p},\mathbf{q})\Big/\sum_2^{1/2}(\mathbf{p})\sum_2^{1/2}(\mathbf{q}) \approx 1$$

and

$$\textstyle\sum_1(\mathbf{q})\Big/\sum_2^{1/2}(\mathbf{q}) = \sum_j Z_j\Big/\left(\sum_j Z_j^2\right)^{1/2} = (N_\text{eff})^{1/2},$$

where N_eff is the number of atoms in the unit cell for equal-atom structures, and is the 'statistically equivalent number' for structues including different atomic species. Z_j is the atomic number of the jth atom. Then,

$$\langle A_\mathbf{p}^N|\{A_\mathbf{q}^N,B_\mathbf{q}^N\}\rangle \approx \sum_\mathbf{q}\frac{(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{q}-\mathbf{p}})}{(1-2s_\mathbf{q}^2)}[B_\mathbf{q}^N-(N_\text{eff})^{1/2}s_\mathbf{q}]. \quad \text{(CPRN1)}$$

By analogy,

$$V_{A_\mathbf{p}^N} \approx \tfrac{1}{2}\left[1-\sum_\mathbf{q}\frac{(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{q}-\mathbf{p}})^2}{(1-2s_\mathbf{q}^2)}\right], \quad \text{(CPRN2)}$$

$$\langle B_\mathbf{p}^N|\{A_\mathbf{q}^N,B_\mathbf{q}^N\}\rangle \approx (N_\text{eff})^{1/2}s_\mathbf{p} + \sum_\mathbf{q}(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{p}-\mathbf{q}})A_\mathbf{q}^N, \quad \text{(CPRN3)}$$

$$V_{B_\mathbf{p}^N} \approx \tfrac{1}{2}\left[(1-2s_\mathbf{p}^2)-\sum_\mathbf{q}(s_{\mathbf{p}+\mathbf{q}}+s_{\mathbf{p}-\mathbf{q}})^2\right]. \quad \text{(CPRN4)}$$

It may be useful to recall that

$$s_\mathbf{p} = -\pi^3\left|\sin(\pi p_s)\right|(p_1p_2p_3)^{-1},$$
$$s_{\mathbf{p}-\mathbf{q}} = -\pi^3\left|\sin[\pi(p_s-q_s)]\right| \times [(p_1-q_1)(p_2-q_2)(p_3-q_3)]^{-1}.$$

4. About the conditional probability $P(R_\mathbf{p},\varphi_\mathbf{p}|\{R_\mathbf{q},\varphi_\mathbf{q}\})$

The probability distribution functions of paper IV estimate $A_\mathbf{p}$ and $B_\mathbf{p}$ given prior knowledge of the subset $\{A_\mathbf{q},B_\mathbf{q}\}$. It is often useful, in practical cases, to express the probability densities, and therefore the expectation values, in terms of the modulus $|F_\mathbf{p}|$ and of the phase $\varphi_\mathbf{p}$. Let us assume that

(*a*) $P(A_\mathbf{p}|\{A_\mathbf{q},B_\mathbf{q}\})$ and $P(B_\mathbf{p}|\{A_\mathbf{q},B_\mathbf{q}\})$ are given by equations (21) and (22) of paper (IV);

(*b*) $P(A_\mathbf{p},B_\mathbf{p}|\{A_\mathbf{q},B_\mathbf{q}\}) \approx P(A_\mathbf{p}|\{A_\mathbf{q},B_\mathbf{q}\})P(B_\mathbf{p}|\{A_\mathbf{q},B_\mathbf{q}\})$. This is justified by the fact that $A_\mathbf{p}$ and $B_\mathbf{p}$ are not correlated. Indeed, in the canonical case, the cumulants $K_{12}(\mathbf{p})$ and $K_{12}(\mathbf{q})$ are always vanishing, as well as the λ_{12} terms.

Then, denoting

$$A_\mathbf{p} = |F_\mathbf{p}|\cos\varphi_\mathbf{p} \quad \text{and} \quad B_\mathbf{p} = |F_\mathbf{p}|\sin\varphi_\mathbf{p}$$

gives

$$P(|F_\mathbf{p}|,\varphi_\mathbf{p}\big|\{A_\mathbf{q},B_\mathbf{q}\}) \approx (2\pi)^{-1}(V_AV_B)^{-1/2}\exp(-M_A^2/2V_A - M_B^2/2V_B) \times |F_\mathbf{p}|\exp[-|F_\mathbf{p}|^2g_3 - |F_\mathbf{p}|^2g_2\cos 2\varphi + |F_\mathbf{p}|g_1\cos(\varphi_\mathbf{p}-\theta)], \quad (1)$$

where

$$g_3 = (V_A^{-1}+V_B^{-1})/4,$$
$$g_2 = (M_A/V_A)^2 + (M_B/V_B)^2$$

and

$$\tan\theta = (M_B/V_B)/(M_A/V_A).$$

Deriving $\langle|F_\mathbf{p}|\big|\{A_\mathbf{q},B_\mathbf{q}\}\rangle$ and $\langle|\varphi_\mathbf{p}|\big|\{A_\mathbf{q},B_\mathbf{q}\}\rangle$ requires the series expansion of (1) as described in paper I. The final formulae are then easily obtained but are rather cumbersome. We therefore decided to use a different approach. Once (CPRN1)–(CPRN4) have been obtained, we assume that

$$|E_\mathbf{p}|_\text{est}^2 = \langle|E_\mathbf{p}|^2\big|\{|E_\mathbf{q}|,\varphi_\mathbf{q}\}\rangle = \langle|A_\mathbf{p}^N|^2\big|\ldots\rangle + \langle|B_\mathbf{p}^N|^2\big|\ldots\rangle \quad (2)$$

and

$$\langle\varphi_\mathbf{p}|\ldots\rangle = \tan^{-1}[\langle B_\mathbf{p}^N|\ldots\rangle/\langle A_\mathbf{p}^N|\ldots\rangle] \quad (3)$$

with variance

$$V_{|E_\mathbf{p}|} = \sigma_{|E_\mathbf{p}|}^2 = V_{A_\mathbf{p}^N} + V_{B_\mathbf{p}^N}. \quad (4)$$

Table 1. *Test structures and main crystallochemical data*

Structure code	Chemical composition	Space group	Z	$N_{\rm eff}$	Data resolution (Å)	Number of standard reflections
SCHWARZ†	$C_{46}H_{70}O_{27}$	$P1$	1	71.5	0.87	4877
PGE2‡	$C_{20}H_{32}O_5$	$P1$	1	24.6	0.88	1560
NEWQB§	$C_{24}H_{20}N_2O_5$	$P\bar{1}$	4	122.3	1.04	3673
M-FABP¶	$C_{667}N_{170}O_{216}S_3$	$P2_12_12_1$	4	4311.2	2.14	7589

† Data according to B. Schweizer (unpublished), retrieved from a list of test structures provided by G. Sheldrick (University of Göttingen). ‡ Data according to DeTitta *et al.* (1990). § Data according to Grigg *et al.* (1978). ¶ Data according to Zanotti *et al.* (1992).

It will be shown that this simple scheme works quite well in the practical applications.

5. The canonical case: simplified relationships for the centrosymmetrical structures

The conclusive formulae for the estimation of the modulus and phase of the standard and non-standard reflections in space group $P\bar{1}$ seem quite different from those derived in $P1$. It may be worthwhile noting that, in the canonical case, the formulae can be restructured to the same formulation. In $P\bar{1}$, the following relationships hold [see equations (32) and (33) in paper III]:

$$M_{01}(\mathbf{p}) = K_1(\mathbf{p}) + \sum_{\mathbf{q}} K_{11}(\mathbf{p},\mathbf{q})[|F_{\mathbf{q}}| \times \cos(\varphi_{\mathbf{q}} - \pi p_s) - K_1(\mathbf{q})]/K_2(\mathbf{q}), \quad (5)$$

$$V_{01}(\mathbf{p}) = K_2(\mathbf{p}) - \sum_{\mathbf{q}} K_{11}^2(\mathbf{p},\mathbf{q})/K_2(\mathbf{q}), \quad (6)$$

where $|M_{01}(\mathbf{p})|$ is the expected value of $|F_{\mathbf{p}}|$, $V_{01}(\mathbf{p})$ is the associated variance,

$$K_1(\mathbf{p}) = \textstyle\sum_1(\mathbf{p})\Big(\prod_{i=1}^{3} c_{p_i/2}\Big),$$

$$K_2(\mathbf{p}) = \textstyle\sum_2(\mathbf{p})\Big[1 + \Big(\prod_{i=1}^{3} c_{p_i/2}\Big) - 2\Big(\prod_{i=1}^{3} c^2_{p_i/2}\Big)\Big],$$

$$K_{11}(\mathbf{p},\mathbf{q}) = \textstyle\sum_{11}(\mathbf{p},\mathbf{q})\Big[-2\Big(\prod_{i=1}^{3} c_{p_i/2}c_{q_i/2}\Big) + \Big(\prod_{i=1}^{3} c_{(p_i+q_i)/2}\Big) + \Big(\prod_{i=1}^{3} c_{(p_i-q_i)/2}\Big)\Big].$$

The phase $\varphi_{\mathbf{p}}$ was estimated through the relationship

$$P(\varphi_{\mathbf{p}} = \pi p_s|\ldots) \approx \tfrac{1}{2} + \tfrac{1}{2}\tanh[|F_{\mathbf{p}}|M_{01}(\mathbf{p})/V_{01}(\mathbf{p})]. \quad (7)$$

Let us analyse option 1 of the canonical case (*i.e.* $\mathbf{p}$ is a half-integral index reflection, $\mathbf{q}$ a standard index). Then $\varphi_{\mathbf{p}}$ is restricted to $\pm\pi/2$ and, according to (7), $M_{01}(\mathbf{p})\sin(\pi p_s)$ is the expected value of $B_{\mathbf{p}}$. Therefore, according to (5),

$$\begin{aligned}\langle B_{\mathbf{p}}|\ldots\rangle &= M_{01}\sin\pi p_s \\ &\approx \textstyle\sum_1(\mathbf{p})s_{\mathbf{p}} \\ &\quad + \sum_{\mathbf{q}}\Big\{\frac{\sum_{11}(\mathbf{p},\mathbf{q})A_{\mathbf{q}}}{\sum_2(\mathbf{q})}\cos(\pi q_s)\sin(\pi q_s) \\ &\quad \times \Big[\Big(\prod_{i=1}^{3} c_{(p_i+q_i)/2}\Big) + \Big(\prod_{i=1}^{3} c_{(p_i-q_i)/2}\Big)\Big]\Big\} \\ &= \textstyle\sum_1(\mathbf{p})s_{\mathbf{p}} + \sum_{\mathbf{q}}\frac{\sum_{11}(\mathbf{p},\mathbf{q})}{\sum_2(\mathbf{q})}(s_{\mathbf{p}+\mathbf{q}} + s_{\mathbf{p}-\mathbf{q}})A_{\mathbf{q}},\end{aligned} \quad (8)$$

which coincides with (CPR3). Analogously, it is easy to show that the variance (6) is identical to that calculated *via* the relation (CPR4). It may be concluded that the relations (CPR3) and (CPR4) [and therefore (CPRN3) and (CPRN4)], originally derived for the $P1$ case, can be used to estimate the structure factors of the half-integral index reflections in $P\bar{1}$. By analogy, the reader may easily verify that the relations (CPR1) and (CPR2) [and therefore (CPRN1) and (CPRN2)], originally derived for the estimation of the real part of the structure factors in $P1$, can be used without any modification for estimating the structure factors of the standard reflections in the $P\bar{1}$ case.

6. Experimental estimates

In the following calculations, we have used the test structures listed in Table 1. A centric space group has been used to check the procedure of §5. For space groups of symmetry higher than triclinic, the unique standard reflections have been expanded to simulate $P1$ or $P\bar{1}$ symmetry. To compare the estimates with the true values, we have computed, from the published atomic parameters, the normalized structure factors $E_{\rm true}$ of the standard and of the half-integral index reflections.

For M-FABP, in Table 2 we present, for various values of RATIO, the number of half-integral index reflections (NHI) with $|E|_{\rm est}/\sigma_{|E|} >$ RATIO, estimated *via* (2), (3) and (4), along with the corresponding relative average phase error (ERR) and the discrepancy index R given by

$$R = \textstyle\sum\big||E|_{\rm est} - |E|_{\rm true}\big|\big/\sum|E|_{\rm true}.$$

Table 2. *M-FABP: statistical outcome for the estimates of the half-integral index reflections via equations (2), (3) and (4)*

NHI is the number of half-integral index reflections for which $|E|_{\rm est}/\sigma_{|E|}$ > RATIO; ERR and R are the corresponding relative average phase error and the discrepancy index, respectively.

RATIO	NHI	ERR (°)	R
0.0	31472	24.56	0.21
0.7	26101	17.05	0.18
1.5	19999	12.52	0.16
2.2	15017	10.46	0.14
4.4	5546	7.50	0.11
5.9	2711	6.37	0.10
8.1	862	5.15	0.08
9.5	382	5.58	0.08
11.0	169	5.27	0.08
13.2	42	4.00	0.05
15.4	13	4.69	0.04

Table 3. *M-FABP: statistical outcome for the estimates of the standard reflections via (2), (3) and (4) using the true half-integral index reflections as prior information*

NI is the number of standard reflections for which $|E|_{\rm est}/\sigma_{|E|}$ > RATIO; ERR and R are the corresponding relative average phase error and the discrepancy index, respectively.

RATIO	NI	ERR (°)	R
0.0	7589	12.42	0.15
1.4	6367	8.48	0.13
2.8	3957	5.99	0.11
4.2	2138	4.45	0.09
5.6	1151	3.38	0.07
7.0	597	2.70	0.06
8.4	352	2.40	0.06
9.8	187	1.99	0.05
11.2	103	1.78	0.05
12.6	53	1.36	0.05
15.4	17	1.06	0.04

All the observed reflections were used to estimate each half-integral index structure factor. We observe that:

(*a*) $|E|_{\rm est}/\sigma_{|E|}$ is a good ranking parameter which is able to indicate, in the absence of experimental data, which half-integral index reflections are accurately determined;

(*b*) both ERR and R increase with decreasing values of $|E|_{\rm est}/\sigma_{|E|}$. When $|E|_{\rm est}/\sigma_{|E|}$ falls to values smaller than unity, the estimates become unreliable (there are about 11 473 reflections with $|E|_{\rm est}/\sigma_{|E|} < 1.5$, with ERR = 46° and $R = 0.48$).

In Table 3, the reverse case is considered: standard structure factors were estimated from the true half-integral index reflections. For space groups of symmetry higher than triclinic, symmetry-equivalent standard reflections were calculated and their moduli and phases averaged. NI is now the number of standard reflections with $|E|_{\rm est}/\sigma_{|E|}$ larger than RATIO; ERR and R are the corresponding average phase error and discrepancy index, respectively. The trends of Table 2 are confirmed; in particular, the ratio $|E|_{\rm est}/\sigma_{|E|}$ is an efficient reliability parameter.

Table 4. *M-FABP: statistical outcome for the estimates of the standard reflections via (2), (3) and (4) using the previously estimated half-integral index reflections as prior information*

NI is the number of standard reflections for which $|E|_{\rm est}/\sigma_{|E|}$ > RATIO; ERR and R are the corresponding relative average phase error and the discrepancy index, respectively.

RATIO	NI	ERR (°)	R
0.0	7589	4.68	0.09
1.4	6423	2.88	0.08
2.8	3919	1.97	0.07
4.2	2080	1.61	0.06
5.6	1095	1.12	0.05
7.0	589	1.06	0.05
8.4	332	1.16	0.05
9.8	177	1.51	0.06
11.2	101	0.48	0.06
12.6	48	0.47	0.07
15.4	19	0.34	0.03

A point for consideration now arises: the 'back-evaluation' of the standard structure factors from the true half-integral index reflections is not realistic because the experimental values of their moduli are not available. In Table 4, we present the statistical outcome when the estimated half-integral structure factors are used as prior information. The estimates are now more accurate than in Table 3, owing to the fact that the errors in the estimates of the half-integral index reflections are compensated, in the reverse pathway, by the errors in the estimation of the standard reflections. Therefore, to 'back-obtain' small R values for the standard structure-factor estimates does not guarantee that the half-integral reflections are well estimated.

The outcome for the other test structures confirms the trend indicated by Tables 2–4. For brevity, we do not give an extended statistical analysis, but only collect in Table 5 the essential features of SCHWARZ, PGE2 and NEWQB. For each test structure, we present the values of NHI, ERR and R for the subset of reflections for which $|E|_{\rm est}/\sigma_{|E|} < 1$ and for the complementary set ($|E|_{\rm est}/\sigma_{|E|} > 1$). Tables 2–5 suggest that the reliability of the estimates does not decrease with increasing structural complexity. The limited number of tests, however, does not allow us to define unequivocally the factors to which the reliability is sensitive.

7. The probabilistic formula estimating $|F|$

We explore in this section whether it is possible to estimate the structure-factor moduli of the half-integral-index reflections given the moduli of the standard reflections. To provide the reader with a self-consistent

Table 5. *SCHWARZ, PGE2, NEWQB: statistical outcome for the half-integral index reflections estimated via (2), (3) and (4)*

NHI is the number of half-integral index reflections for which $|E|_{\rm est}/\sigma_{|E|} > 1$ or $|E|_{\rm est}/\sigma_{|E|} < 1$; ERR and R are the corresponding relative average phase error and the discrepancy index, respectively.

Structure code	NHI	RATIO	ERR (°)	R
SCHWARZ	701	< 1.0	64.04	0.62
	4277	> 1.0	21.43	0.22
PGE2	539	< 1.0	49.43	0.53
	1085	> 1.0	15.94	0.20
NEWQB	666	< 1.0	75.14	0.76
	3033	> 1.0	13.53	0.28

premise, we recall the basic relations obtained in paper IV. We denoted by

$$P(\mathbf{X}) = P(X_1, X_2, \ldots, X_{2n+2})$$

the joint probability distribution

$$P(A_{\mathbf{p}}, B_{\mathbf{p}}, A_{\mathbf{q}_1}, B_{\mathbf{q}_1}, \ldots, A_{\mathbf{q}_n}, B_{\mathbf{q}_n}),$$

where

$$X_1 = A_{\mathbf{p}} \quad \text{and} \quad X_2 = B_{\mathbf{p}}$$

indicate the variables we wish to estimate. The variable X_j represents A_j or B_j according to the value of j, *i.e.* odd and even values of j correspond to the A and B variables, respectively. Then,

$$P(\mathbf{X}) = (2\pi)^{-(n+1)}(\det \boldsymbol{\lambda})^{1/2} \exp(-\tfrac{1}{2}\bar{\mathbf{T}}\boldsymbol{\lambda}\mathbf{T}), \qquad (9)$$

where

$$\mathbf{T} = (d_1, d_2, \ldots, d_{2n+2}),$$
$$d_j = K_j - X_j,$$
$$\boldsymbol{\lambda} = \mathbf{K}^{-1}$$

and $\mathbf{K}$ is the (symmetrical) variance–covariance matrix:

$$\mathbf{K} = \begin{vmatrix} K_{11} & K_{12} & \ldots & K_{1,2n+2} \\ K_{21} & K_{22} & \ldots & K_{2,2n+2} \\ \vdots & \vdots & \ddots & \vdots \\ K_{2n+2,1} & K_{2n+2,2} & \ldots & K_{2n+2,2n+2} \end{vmatrix}.$$

In accordance with equation (18) of paper IV, the explicit expression for (9) is

$$P(\mathbf{X}) = (2\pi)^{-(n+1)}(\det \boldsymbol{\lambda})^{1/2} \exp\Big(-\tfrac{1}{2}\sum_{j=1}^{2n+2} \lambda_{jj} d_j^2 - \sum_{j=2}^{2n+2} \lambda_{1j} d_1 d_j - \sum_{j=3}^{2n+2} \lambda_{2j} d_2 d_j - \sum_{j_1 > j_2 = 3}^{2n+2} \lambda_{j_1 j_2} d_{j_1} d_{j_2} \Big), \qquad (10)$$

where λ_{ij} are the elements of the matrix $\boldsymbol{\lambda}$. The distribution (10) was used to estimate X_1 and X_2 given prior knowledge of the set $\{A_{\mathbf{q}}, B_{\mathbf{q}}, \mathbf{q} = 1, \ldots, n\}$. In this paper, we are interested in estimating $|F_{\mathbf{q}}| = |A_{\mathbf{q}}^2 + B_{\mathbf{q}}^2|^{1/2}$ when the only available information is the set of moduli $\{|F_{\mathbf{q}}|, \mathbf{q} = 1, \ldots, n\}$. In this situation, it is convenient to return to the more explicit notation

$$P(A_1, B_1, A_2, B_2, \ldots, A_{n+1}, B_{n+1})$$

and to introduce the change of variables

$$A_j = |F_j| \cos\varphi_j, \quad B_j = |F_j| \sin\varphi_j, \quad j = 1, \ldots, n+1.$$

Then (10) may be written as

$$\begin{aligned} &P(|F_1|, \ldots, |F_{n+1}|, \varphi_1, \ldots, \varphi_{n+1}) \\ &= (2\pi)^{-(n+1)}(\det \boldsymbol{\lambda})^{1/2}|F_1| \ldots |F_{n+1}| \\ &\quad \times \exp\Big[-\tfrac{1}{2}\sum_{j=1}^{n+1} \lambda_{2j-1,2j-1}(|F_j|\cos\varphi_j - K_{A_j})^2 \\ &\quad -\tfrac{1}{2}\sum_{j=1}^{n+1} \lambda_{2j,2j}(|F_j|\sin\varphi_j - K_{B_j})^2 \\ &\quad -\sum_{j=2}^{n+1} \lambda_{1,2j-1}(|F_1|\cos\varphi_1 - K_{A_1})(|F_j|\cos\varphi_j - K_{A_j}) \\ &\quad -\sum_{j=1}^{n+1} \lambda_{1,2j}(|F_1|\cos\varphi_1 - K_{A_1})(|F_j|\sin\varphi_j - K_{B_j}) \\ &\quad -\sum_{j=2}^{n+1} \lambda_{2,2j-1}(|F_1|\sin\varphi_1 - K_{B_1})(|F_j|\cos\varphi_j - K_{A_j}) \\ &\quad -\sum_{j=2}^{n+1} \lambda_{2,2j}(|F_1|\sin\varphi_1 - K_{B_1})(|F_j|\sin\varphi_j - K_{B_j}) \\ &\quad -\sum_{j_1 > j_2 = 3}^{2n+2} \lambda_{2j_1-1,2j_2-1}(|F_{j_1}|\cos\varphi_{j_1} - K_{A_{j_1}}) \\ &\quad \times (|F_{j_2}|\cos\varphi_{j_2} - K_{A_{j_2}}) \\ &\quad -\sum_{j_1 > j_2 = 3}^{2n+2} \lambda_{2j_1-1,2j_2}(|F_{j_1}|\cos\varphi_{j_1} - K_{A_{j_1}}) \\ &\quad \times (|F_{j_2}|\sin\varphi_{j_2} - K_{B_{j_2}}) \\ &\quad -\sum_{j_1 > j_2 = 3}^{2n+2} \lambda_{2j_1,2j_2-1}(|F_{j_1}|\sin\varphi_{j_1} - K_{B_{j_1}}) \\ &\quad \times (|F_{j_2}|\cos\varphi_{j_2} - K_{A_{j_2}}) \\ &\quad -\sum_{j_1 > j_2 = 3}^{2n+2} \lambda_{2j_1,2j_2}(|F_{j_1}|\sin\varphi_{j_1} - K_{B_{j_1}}) \\ &\quad \times (|F_{j_2}|\sin\varphi_{j_2} - K_{B_{j_2}})\Big]. \end{aligned} \qquad (11)$$

From here on, we will neglect in (11) the terms not involving $|F_1|$ or φ_1 because they do not contribute to the estimation of $|F_1|$ (see paper IV). Except for very low resolution reflections, in the canonical case the following assumptions can be made:

$$\begin{aligned} K_{A_j} \approx K_{B_j} &\approx 0 && \text{for} \quad j = 1, \ldots, n+1, \\ \lambda_{12} &= 0, \\ \lambda_{1,2j-1} &= 0 && \text{for} \quad j \neq 1, \\ \lambda_{2,2j} &= 0 && \text{for} \quad j \neq 1, \\ \lambda_{2j-1,2j-1} &\approx \lambda_{2j,2j} && \text{for} \quad j = 1, \ldots, n+1, \\ \lambda_{1,2j} &\approx -\lambda_{2,2j-1} && \text{for} \quad j = 1, \ldots, n+1. \end{aligned}$$

Then (11) reduces to

$$\begin{aligned} &P(|F_1|, \ldots, |F_{n+1}|, \varphi_1, \ldots, \varphi_{n+1}) \\ &\quad = (2\pi)^{-(n+1)} (\det \boldsymbol{\lambda})^{1/2} |F_1| \ldots |F_{n+1}| \\ &\qquad \times \exp\bigg(-\tfrac{1}{2} \sum_{j=1}^{n+1} \lambda_{2j,2j} |F_j|^2 - |F_1| \cos\varphi_1 \\ &\qquad \times \sum_{j=2}^{n+1} \lambda_{1,2j} |F_j| \sin\varphi_j \\ &\qquad + |F_1| \sin\varphi_1 \sum_{j=2}^{n+1} \lambda_{1,2j} |F_j| \cos\varphi_j \bigg). \end{aligned}$$

Let us now integrate over φ_{n+1} by applying the relation (Abramowitz & Stegun, 1972)

$$\int_0^{2\pi} \exp(\xi_1 \sin\varphi + \xi_2 \cos\varphi)\, \mathrm{d}\varphi = 2\pi I_0[(\xi_1^2 + \xi_2^2)^{1/2}],$$

where I_0 is the modified Bessel function. We obtain

$$\begin{aligned} &P(|F_1|, \ldots, |F_{n+1}|, \varphi_1, \ldots, \varphi_n) \\ &\quad = (2\pi)^{-n} (\det \boldsymbol{\lambda})^{1/2} |F_1| \ldots |F_{n+1}| \\ &\qquad \times \exp\bigg(-\tfrac{1}{2} \sum_{j=1}^{n+1} \lambda_{2j,2j} |F_j|^2 - |F_1| \cos\varphi_1 \\ &\qquad \times \sum_{j=2}^{n} \lambda_{1,2j} |F_j| \sin\varphi_j - |F_1| \sin\varphi_1 \\ &\qquad \times \sum_{j=2}^{n} \lambda_{1,2j} |F_j| \cos\varphi_j \bigg) I_0(|F_1| \lambda_{1,2n+2} |F_{n+1}|). \end{aligned}$$

The subsequent integration over $\varphi_1, \ldots, \varphi_n$ results in

$$\begin{aligned} &P(|F_1|, \ldots, |F_{n+1}|) \\ &\quad \approx (\det \boldsymbol{\lambda})^{1/2} |F_1| \ldots |F_{n+1}| \exp\bigg(-\tfrac{1}{2} \sum_{j=1}^{n+1} \lambda_{2j,2j} |F_j|^2 \bigg) \\ &\qquad \times \prod_{j=2}^{n+1} I_0(|F_1| \lambda_{1,2j} |F_j|). \end{aligned} \tag{12}$$

We can now calculate the conditional probability

$$\begin{aligned} &P(|F_1| \big| |F_2|, \ldots, |F_{n+1}|) \\ &\quad \approx L^{-1} |F_1| \exp\left(-\tfrac{1}{2} \lambda_{22} |F_1|^2 \right) \prod_{j=2}^{n+1} I_0(|F_1| \lambda_{1,2j} |F_j|), \end{aligned} \tag{13}$$

where

$$L = \int_0^\infty |F_1| \exp\left(-\tfrac{1}{2} \lambda_{22} |F_1|^2 \right) \prod_{j=2}^{n+1} I_0(|F_1| \lambda_{1,2j} |F_j|)\, \mathrm{d}|F_1|$$

is a normalization factor which does not depend on $|F_1|$ and may be calculated by numerical methods.

The use of (13) is time consuming if large values of n are used for estimating $|F_1|$ and if the calculations must be repeated for numerous $|F_1|$. Since the argument of I_0 is generally small, the approximation $I_0 \approx \exp(x^2/4)$ may be used. Then

$$P(|F_1| \big| |F_2|, \ldots, |F_{n+1}|) \approx 2g|F_1| \exp(-g|F_1|^2), \tag{14}$$

where

$$g = \tfrac{1}{2} \lambda_{22} - \tfrac{1}{4} \sum_{j=2}^{n+1} \lambda_{1,2j}^2 |F_j|^2. \tag{15}$$

Using standard methods, the following relationships arise from (14):

$$\langle |F_1| \big| |F_2|, \ldots, |F_{n+1}| \rangle \approx (\pi^{1/2}/2) g^{-1/2} \tag{16}$$

and

$$\langle |F_1|^2 \big| |F_2|, \ldots, |F_{n+1}| \rangle \approx g^{-1}. \tag{17}$$

The conditional distribution (14) may be conveniently expressed in terms of the pseudo-normalized structure factors $E_j = F_j / m_j^{1/2}$, where m_j is the value of $\sum_2$ for the jth reflection. We obtain

$$P(|E_1| \big| |E_2|, \ldots, |E_{n+1}|) \approx 2(gm_1)|E_1| \exp(-gm_1|E_1|^2), \tag{18}$$

from which

$$\begin{aligned} |E_1|_{\text{est}} &\equiv \langle |E_1| \big| |E_2|, \ldots, |E_{n+1}| \rangle \approx (\pi^{1/2}/2)(gm_1)^{-1/2}, \\ |E_1|^2_{\text{est}} &\equiv \langle |E_1|^2 \big| |E_2|, \ldots, |E_{n+1}| \rangle \approx (gm_1)^{-1} \end{aligned}$$

and

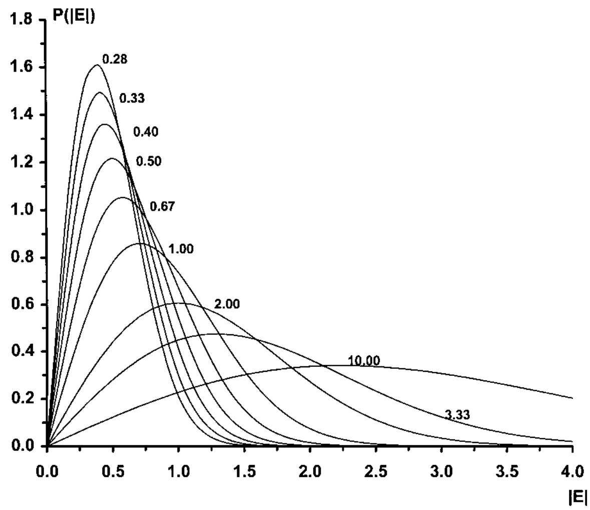

Fig. 1. The distribution (18) for selected values of $(gm_1)^{-1}$.

$$\begin{aligned} V_{|E_1|} &= \sigma^2_{|E_1|} \\ &= \langle |E_1|^2 \big| |E_2|, \ldots, |E_{n+1}| \rangle - \langle |E_1| \big| |E_2|, \ldots, |E_{n+1}| \rangle^2 \\ &\approx (1-\pi/4)(gm_1)^{-1} \\ &\approx (1-\pi/4)|E_1|^2_{\rm est}. \end{aligned}$$

We observe the following.

(*a*) Equation (18) reduces to the acentric Wilson distribution when the contribution arising from the summation on the right-hand side of (15) is neglected. Then $\lambda_{22} = 2\big/\sum_2$ and $gm_1 = 1$.

(*b*) If both the terms on the right-hand side of (15) are considered, $|E_1|^2_{\rm est}$ varies proportionally to $(gm_1)^{-1}$, while the variance of the estimate is large for large values of $|E_1|^2_{\rm est}$. This behaviour may be easily understood by considering Fig. 1, where the distribution (18) is shown for selected values of $(gm_1)^{-1}$: the curves are flatter for larger values of $|E_1|^2_{\rm est}$. This is undesirable behaviour.

(*c*) In accordance with point (*b*), $|E_1|_{\rm est}/\sigma_{|E_1|}$ is a constant:

$$|E_1|_{\rm est}/\sigma_{|E_1|} = [\pi/(4-\pi)]^{1/2} \approx 1.91.$$

8. Experimental tests

The results obtained in paper IV suggest that the calculation of $(gm_1)^{-1}$ as given by (15) may be accomplished without inverting the matrix **K**. Since

$$\begin{aligned} \lambda_{22}^{-1} &= \tfrac{1}{2}\left\{ \textstyle\sum_2(\mathbf{p}) - \displaystyle\sum_{\mathbf{q}} \left[\frac{\sum_{11}^2(\mathbf{p},\mathbf{q})}{\sum_2(\mathbf{q})} (s_{\mathbf{p+q}} + s_{\mathbf{q-p}})^2 \right] \right\} \\ &\approx m_1 e/2, \end{aligned}$$

where

$$e = 1 - \sum_{\mathbf{q}} (s_{\mathbf{p+q}} + s_{\mathbf{q-p}})^2,$$

and, since

$$\begin{aligned} \lambda_{1,2j} &= -(K_{1,2j}/K_{2j,2j})\lambda_{22} \\ &= -\left[\textstyle\sum_{11}(\mathbf{p},\mathbf{q}_j)(s_{\mathbf{p+q}_j} + s_{\mathbf{q}_j-\mathbf{p}}) \big/ \sum_2(\mathbf{q}_j) \right] \lambda_{22} \\ &= -2m_1^{-1/2} m_j^{-1/2} e^{-1} (s_{\mathbf{p+q}_j} + s_{\mathbf{q}_j-\mathbf{p}}), \end{aligned}$$

then

$$gm_1 = e^{-1} - e^{-2} \sum_{\mathbf{q}} (s_{\mathbf{p+q}} + s_{\mathbf{q-p}})^2 |E_{\mathbf{q}}|^2. \qquad (19)$$

Because of numerical approximations, gm_1, as given by (15) or (19), may be too close to zero; hence its reciprocal $(gm_1)^{-1}$ may become too large. To avoid numerical instabilities, we introduce into (19) the approximation $(1-x)^{-1} \approx 1 + x$. Then (19) becomes

Table 6. *Discrepancy index for the NREFL half-integral reflections according to equation (20)*

NREFL is the number of half-integral index reflections whose moduli have been estimated *via* equation (20) and *R* is the corresponding discrepancy index.

Structure code	NREFL	*R*
PGE2	1624	0.48
SCHWARZ	4978	0.42
NEWQB	3699	0.65
M-FABP	38640	0.53

$$\begin{aligned} |E_1|^2_{\rm est} &= (gm_1)^{-1} \\ &= e\Big[1 + e^{-1} \sum_{\mathbf{q}} (s_{\mathbf{p+q}} + s_{\mathbf{q-p}})^2 |E_{\mathbf{q}}|^2\Big] \\ &\approx 1 + \sum_{\mathbf{q}} (s_{\mathbf{p+q}} + s_{\mathbf{q-p}})^2 (|E_{\mathbf{q}}|^2 - 1). \qquad (20) \end{aligned}$$

Equation (20) has been applied to the test structures listed in Table 1 and the results are shown in Table 6 [NREFL is the number of half-integral index reflections whose moduli have been estimated *via* (20) and *R* is the corresponding discrepancy index]. It may be observed that the *R* index is relatively high for all the test structures (compare it with the *R* indices in Tables 2–5, obtained when the phase information is available), but it should be acceptable if a reliability index was available to pick up the most accurate estimates. Unfortunately, this is impossible, owing to the fact that $|E|_{\rm est}/\sigma_{|E|}$ is expected to be a constant for all the reflections. Such a result may be compared with the conclusion of Mishnev (1996) arising from the Hilbert-transform method: while the Hilbert-transform method is unable to estimate half-integral reflections, the probabilistic approach is less restrictive but does not provide a criterion for picking up the most reliable estimates.

We thank the referees for useful revisions and suggestions.

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover Publications.
DeTitta, G. T., Langs, D. A., Edmonds, J. W. & Duax, W. L. (1980). *Acta Cryst.* B**36**, 638–645.
Giacovazzo, C. & Siliqi, D. (1998). *Acta Cryst.* A**54**, 957–970.
Giacovazzo, C., Siliqi, D., Altomare, A., Cascarano, G., Rizzi, R. & Spagna, R. (1999). *Acta Cryst.* A**55**, 322–331.
Giacovazzo, C., Siliqi, D., Carrozzini, B., Guagliardi, A. & Moliterni, A. G. G. (1999). *Acta Cryst.* A**55**, 314–321.
Giacovazzo, C., Siliqi, D. & Fernández-Castaño, C. (1999). *Acta Cryst.* A**55**, 512–524.
Grigg, R., Kemp, J., Sheldrick, G. M. & Trotter, J. (1978). *J. Chem. Soc. Chem. Commun.* pp. 109–111.
Mishnev, A. F. (1993). *Acta Cryst.* A**49**, 159–161.

Mishnev, A. F. (1996). *Acta Cryst.* A**52**, 629–633.
Toll, J. S. (1956). *Phys. Rev.* **104**, 1760–1770.
Wu, T.-Y. & Ohmura, T. (1962). *Quantum Theory of Scattering.* New York: Prentice-Hall.
Zanotti, G., Fogale, F. & Capitani, G. (1996). *Acta Cryst.* A**52**, 757–765.
Zanotti, G., Scapin, G., Spadon, P., Veerkamp, J. H. & Sacchettini, J. C. (1992). *J. Biol. Chem.*, **267**, 18541–18550.