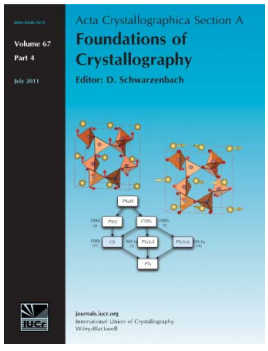
Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673
Editor: **D. Schwarzenbach**

Estimation of the variance in any point of an electron-density map for any space group

Carmelo Giacovazzo, Annamaria Mazzone and Giuliana Comunale

Acta Cryst. (2011). A**67**, 368–382



Acta Crystallographica Section A: Foundations of Crystallography covers theoretical and fundamental aspects of the structure of matter. The journal is the prime forum for research in diffraction physics and the theory of crystallographic structure determination by diffraction methods using X-rays, neutrons and electrons. The structures include periodic and aperiodic crystals, and non-periodic disordered materials, and the corresponding Bragg, satellite and diffuse scattering, thermal motion and symmetry aspects. Spatial resolutions range from the subatomic domain in charge-density studies to nanodimensional imperfections such as dislocations and twin walls. The chemistry encompasses metals, alloys, and inorganic, organic and biological materials. Structure prediction and properties such as the theory of phase transformations are also covered.

Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 2 March 2011
Accepted 27 April 2011

Estimation of the variance in any point of an electron-density map for any space group

Carmelo Giacovazzo,[a,b]* Annamaria Mazzone[a] and Giuliana Comunale[c]

[a]Istituto di Cristallografia, CNR, Via G. Amendola, 122/O 70126 Bari, Italy, [b]Dipartimento Geomineralogico, Università di Bari, 70125 Bari, Italy, and [c]Dipartimento di Chimica, Università della Basilicata, 85100 Potenza, Italy. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

In a recent paper [Giacovazzo & Mazzone (2011). *Acta Cryst.* A**67**, 210–218] a mathematical expression of the variance at any point of the unit cell has been described. The formulas were derived in *P*1 for any type of Fourier synthesis (observed, difference and hybrid) under the following hypothesis: the current phases are distributed on the trigonometric circle about the correct values according to von Mises distributions. This general hypothesis allows the variance expressions to be valid at any stage of the phasing process. In this paper the method has been extended to any space group, no matter whether centric or acentric. The properties of the variance generated by space-group symmetry are described; in particular it is shown that the variance is strictly connected with the *implication transformations*, which are basic for Patterson deconvolution. General formulas simultaneously taking into account phase uncertainty and measurement errors have been obtained, valid no matter what the quality of the model.

1. Symbols and notation

Paper I: Giacovazzo & Mazzone (2011).
$F = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h}\mathbf{r}_j) = |F| \exp(i\varphi)$: structure factor of the target structure.
$F_p = \sum_{j=1}^{p} f_j \exp(2\pi i \mathbf{h}\mathbf{r}'_j) = |F_p| \exp(i\varphi_p)$, where $\mathbf{r}'_j = \mathbf{r}_j + \Delta\mathbf{r}_j$: structure factor of the model structure.
$\mathbf{C}_s \equiv (\mathbf{R}_s, \mathbf{T}_s)$ sth symmetry operator ($\mathbf{C}_s\mathbf{r} \equiv \mathbf{R}_s\mathbf{r} + \mathbf{T}_s$): $\mathbf{R}_s$ and $\mathbf{T}_s$ are the rotational and translational matrices, respectively.
n: number of the symmetry operators for the target and for the model structure.
$\sum_N = \sum_{j=1}^{N} f_j^2$, $\sum_p = \sum_{j=1}^{p} f_j^2$, where p is the number of atoms in the model structure.
$F_q = F - F_p = |F_q| \exp(i\varphi_q)$: structure factor of the *ideal difference structure*.
$E = A + iB = R\exp(i\varphi)$, $E_p = A_p + iB_p = R_p \exp(i\varphi_p)$, $E_q = A_q + iB_q = R_q \exp(i\varphi_q)$, $R = |F|/\sum_N^{1/2}$, $R_p = |F_p|/\sum_N^{1/2}$.
$\rho(\mathbf{r}) = (2/V)\sum_{\mathbf{h}>0} |F_\mathbf{h}| \cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_\mathbf{h})$: general expression of an electron-density map.
$\rho_p(\mathbf{r}) = (2/V)\sum_{\mathbf{h}>0} |F_{p\mathbf{h}}| \cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}})$: electron-density map of the model structure.
$\rho_{\rm obs}(\mathbf{r}) = (2/V)\sum_{\mathbf{h}>0} m_\mathbf{h} |F_\mathbf{h}| \cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}})$: observed electron density when a model is available.
$[\rho(\mathbf{r})]_{\rm N} = (2/V)\sum_{\mathbf{h}>0} R_\mathbf{h} \cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_\mathbf{h})$: electron-density map calculated *via* normalized structure factors.
$\rho_{p\rm N}(\mathbf{r}) = (2/V)\sum_{\mathbf{h}>0} R_{p\mathbf{h}} \cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}})$: electron-density map of the model structure calculated *via* normalized structure factors.
$\rho_{\rm obsN}(\mathbf{r}) = (2/V)\sum_{\mathbf{h}>0} m_\mathbf{h} R_\mathbf{h} \cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}})$: observed electron density when a model is available, calculated *via* normalized structure factors.
$P(\mathbf{u}) = (2/V)\sum_{\mathbf{h}>0} |F_\mathbf{h}|^2 \cos(2\pi\mathbf{h}\mathbf{u})$: Patterson synthesis.

In all the above Fourier syntheses (observed, difference, hybrid) the term of order zero is omitted. Accordingly, the average values of the corresponding maps are always zero. By $\mathbf{h} > 0$ it is meant that the summation is over one half of the reciprocal space (only one member of each Friedel pair is included).

$\mathrm{var}\rho(\mathbf{r}) = \langle[\rho(\mathbf{r})]^2\rangle - [\langle\rho(\mathbf{r})\rangle^2]$: variance of the map ρ in a point $\mathbf{r}$.
$[\mathrm{var}\rho(\mathbf{r})]_{\rm N} = \langle[\rho(\mathbf{r})]_{\rm N}{}^2\rangle - \{\langle[\rho(\mathbf{r})]_{\rm N}\rangle^2\}$: variance of the normalized electron-density map.
$D_i(x) = I_i(x)/I_0(x)$, I_i is the modified Bessel function of order i.
$D = \langle\cos(2\pi\mathbf{h}\Delta\mathbf{r})\rangle$: the average is performed per resolution shell.
$\sigma_A = D(\Sigma_p/\Sigma_N)^{1/2}$.
$\sigma_R^2 = \langle|\mu|^2\rangle/\sum_N$, $\langle|\mu|^2\rangle$ is the measurement error.
$e = 1 + \sigma_R^2$.
$m = \langle\cos(\varphi - \varphi_p)\rangle = I_1(X)/I_0(X)$, where $X = 2\sigma_A R R_p/(e - \sigma_A^2)$.
$s = \sin\theta/\lambda$.
$m_{c\mathbf{h}} = 0.5 + 0.5\tanh(X/2)$.
EDM: electron-density modification.
CORR: correlation between the model and the target electron-density maps.

2. Introduction

An algebraic expression estimating the variance of the electron-density map in a point **r** of the unit cell for observed, difference and hybrid Fourier syntheses was obtained in Paper I, under the hypothesis that a model structure is available and that the space group is $P1$. To carry out the calculations, two assumptions were made: (i) there is no error in the measurements; (ii) each phase $\varphi_{\mathbf{h}}$ is distributed around $\varphi_{p\mathbf{h}}$ according to the von Mises distribution

$$M(\varphi; X, \varphi_p) = [2\pi I_0(X)]^{-1} \exp[X \cos(\varphi - \varphi_p)]. \quad (1)$$

The above hypotheses allowed the calculation of the variance no matter what the correlation between the model and target structures. When CORR = 0 the phases are randomly distributed and equation (1) becomes a flat function; when CORR $\sim$ 1 the phases are definitively fixed and equation (1) coincides with the Dirac delta function. The final expression was

$$\mathrm{var}\rho(\mathbf{r}) = \frac{2}{V^2}\Big\{\sum_{\mathbf{h}>0}(1 - m_{\mathbf{h}}^2)|F_{\mathbf{h}}|^2 - \sum_{\mathbf{h}>0}|F_{\mathbf{h}}|^2 \times [m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})]\cos(4\pi\mathbf{h}\cdot\mathbf{r} - 2\varphi_{p\mathbf{h}})\Big\}. \quad (2)$$

In that paper it was shown that:

(*a*) As a first approximation, the second term is negligible with respect to the first one. Thus the concept of *map variance* was introduced and its main properties (as a function of CORR, of the data resolution *etc.*) were described. The simplified expression

$$\mathrm{var}\rho(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{\mathbf{h}}^2)|F_{\mathbf{h}}|^2 \quad (3)$$

was assumed to be a reliable approximation of equation (2).

(*b*) The ratio $\langle\rho(\mathbf{r})\rangle/(\mathrm{var}\rho)^{1/2}$ was considered as the signal/noise ratio; its properties were related to several modern phasing methods including charge flipping (Oszlányi & Sütő, 2004, 2005, 2007; Palatinus & Chapuis, 2007), the *VLD* algorithm (Burla, Caliandro *et al.*, 2010; Burla, Giacovazzo & Polidori, 2010) and EDM procedures (Shiono & Woolfson, 1992; Refaat & Woolfson, 1993; Giacovazzo & Siliqi, 1997).

(*c*) Expression (3) may be combined with the variance expression calculated by Coppens & Hamilton (1968) (see also Rees, 1976), based on the variance of the observed amplitudes [say $\sigma^2(|F_{\mathbf{h}}|)$]:

$$\sigma^2[\rho(\mathbf{r})] = \frac{4}{V^2}\sum_{\mathbf{h}>0}\sigma^2(|F_{\mathbf{h}}|)\cos^2(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{\mathbf{h}}). \quad (4)$$

Expression (4) was essentially dedicated to establishing the accuracy of the results of the structure analysis after the final least-squares refinement. The combined expression suggested in Paper I was

$$\mathrm{var}\rho(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}[(1 - m_{\mathbf{h}}^2)|F_{\mathbf{h}}|^2 + 2\sigma^2(|F_{\mathbf{h}}|)] \times \cos^2(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{\mathbf{h}}). \quad (5)$$

When the model coincides with the target structure the term representing the average map variance [see equation (3)] vanishes, and the variance coincides with the Coppens & Hamilton term. If CORR $\ll$ 1, then the Coppens & Hamilton contribution becomes negligible and the variance is dominated by expression (3).

In Paper I it was also anticipated that the concept of map variance does not hold for space groups with symmetry higher than $P1$, or, in other words, the variance may vary strongly from point to point as an effect of the space-group symmetry. This paper aims to establish a variance expression valid in all the space groups and to discover its properties, including the possible connections between variance properties and phasing methods.

In §3 we derive the variance expression valid in $P\bar{1}$, in §§4–6 the method is extended to all acentric and centric space groups, in §§7–10 the main properties of the variance are described, in §§11–12 E maps and hybrid Fourier syntheses are considered, in §13 the variance function is drawn for a simple, even if unrealistic, structural example, and in §14 a criterion for the active use of the variance is anticipated.

3. The variance estimate in $P\bar{1}$

We will assume that both φ and φ_p may only have values of 0 or π, that a weight m_c may be associated with the probabilistic relation $\varphi \simeq \varphi_p$. Then

$$\langle\rho(\mathbf{r})\rangle = \frac{2}{V}\sum_{\mathbf{h}>0} m_{c\mathbf{h}}|F_{\mathbf{h}}|\cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}}) = \rho_{\mathrm{obs}}(\mathbf{r}) \quad (6)$$

and

$$\langle\rho^2(\mathbf{r})\rangle = \frac{4}{V^2}\left\langle\sum_{\mathbf{h}>0}|F_{\mathbf{h}}|^2\cos^2(2\pi\mathbf{h}\cdot\mathbf{r})\right\rangle + \frac{4}{V^2}\sum_{\mathbf{h}\neq\mathbf{k}>0} m_{c\mathbf{h}}m_{c\mathbf{k}}|F_{\mathbf{h}}F_{\mathbf{k}}|\cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}}) \times \cos(2\pi\mathbf{k}\mathbf{r} - \varphi_{p\mathbf{k}}). \quad (7)$$

Since the last term in equation (7) is equal to

$$\rho^2{}_{\mathrm{obs}}(\mathbf{r}) - \frac{4}{V^2}\sum_{\mathbf{h}>0} m_{c\mathbf{h}}^2|F_{\mathbf{h}}|^2\cos^2(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{p\mathbf{h}})$$

we obtain

$$\mathrm{var}\rho(\mathbf{r}) = \langle\rho^2(\mathbf{r})\rangle - \langle\rho(\mathbf{r})\rangle^2 = TH_1 + TH_2(\mathbf{r}) \quad (8)$$

where

$$TH_1 = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{c\mathbf{h}}^2)|F_{\mathbf{h}}|^2,$$

$$TH_2(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{c\mathbf{h}}^2)|F_{\mathbf{h}}|^2\cos(4\pi\mathbf{h}\cdot\mathbf{r}).$$

A second useful expression for the variance is

$$\mathrm{var}\rho(\mathbf{r}) = \frac{4}{V^2}\sum_{\mathbf{h}>0}(1 - m_{c\mathbf{h}}^2)|F_{\mathbf{h}}|^2\cos^2(2\pi\mathbf{h}\cdot\mathbf{r}). \qquad (9)$$

It may be useful to notice now that the variance is strictly correlated with the Patterson function. Indeed

$$\mathrm{var}\rho(\mathbf{r}) = \frac{1}{V}[P_w(0) + P_w(2\mathbf{r})]$$

where P_w is a weighted $[w = (1 - m_{c\mathbf{h}}^2)]$ Patterson function. If the model is uncorrelated with the target structure then

$$\mathrm{var}\rho(\mathbf{r}) = \frac{1}{V}[P(0) + P(2\mathbf{r})].$$

The relation between the variance and the Patterson function will be more evident when the space-group symmetry is taken into account.

varρ shows analogies and remarkable differences with the variance calculated for $P1$. Among the analogies we notice:

(a) In each point $\mathbf{r}$ the variance is the sum of two contributions [see equation (8)]: TH_1 (the *constant term*) which does not vary with $\mathbf{r}$, and $TH_2(\mathbf{r})$ (the *variable term*) depending on $\mathbf{r}$.

(b) varρ is expected to be non-negative in any point of the unit cell [see equation (9)].

(c) The variance has three-dimensional periodicity, half that of the electron density. Indeed the variable term assumes the same value in $\mathbf{r}$ and in $\mathbf{r} + \mathbf{u}$, where $\mathbf{u} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$ and u, v, w may be 0 or $\frac{1}{2}$.

The variance in $P\overline{1}$, however, shows a variability larger than in $P1$; in particular it attains strong maxima on the inversion centres [indeed TH_2 is at a maximum if $\cos(4\pi\mathbf{h}\cdot\mathbf{r}) = 1$ for any Miller index, that is when x, y, z are 0 or $\frac{1}{2}$].

To estimate the largest oscillations of varρ we used *Newqb* ($C_{48}H_{40}N_4O_{10}$, $P\overline{1}$; Sheldrick, 1982) as the target structure and we associated with it six models with different values of $\langle\sigma_A\rangle$, where $\langle\sigma_A\rangle$ is the average of the σ_A values corresponding to different resolution shells. In Fig. 1 we plot the values of oscv = $\{\max[\mathrm{var}\rho(\mathbf{r})] - \min[\mathrm{var}\rho(\mathbf{r})]\}/TH_1$ *versus* $\langle\sigma_A\rangle$. In the same figure we also show oscσ *versus* $\langle\sigma_A\rangle$, where oscσ = $\{\max[\sigma_\rho(\mathbf{r})] - \min[\sigma_\rho(\mathbf{r})]\}/(TH_1)^{1/2}$ and $\sigma_\rho(\mathbf{r}) = (\mathrm{var}\rho)^{1/2}$. We notice:

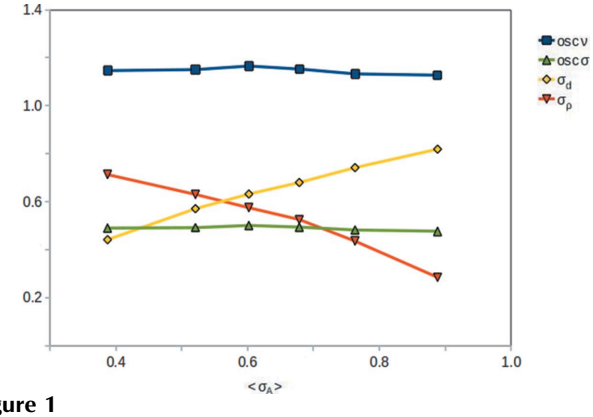


Figure 1
Newqb. oscv (blue line), oscσ (green line), σ_ρ (red line) and σ_d (yellow line) are plotted *versus* $\langle\sigma_A\rangle$. The squares and the triangles correspond to six structure models characterized by different $\langle\sigma_A\rangle$ values.

(a) oscv is constantly close to 1.1 (in $P1$ it was in the interval 0.14–0.30);

(b) oscσ is constantly close to 0.5 (in $P1$ it was in the interval 0.07–0.15).

The above observations suggest that the concept of map variance may be used only in $P1$: when symmetry operators are present in the crystal the variance depends on the point $\mathbf{r}$ in which it is calculated.

In the same Fig. 1 we compare σ_ρ with σ_d, the standard deviation of the pixel intensity distribution of the current electron-density map. As in $P1$ σ_ρ and σ_d are anticorrelated: the first diminishes and the second increases when the model–target correlation increases. The two parameters should not be confused.

A schematic overview of the variance distribution for the model structure characterized by $\langle\sigma_A\rangle = 0.68$ is given in Fig. 2, where we show the projection of $TH_2(\mathbf{r})$ (the variance is quite similar; it differs from TH_2 by a constant term) on the plane (**a**, **b**). As theoretically foreseen, the periodicity of the variance along the unit-cell axes is half that of the cell periods, and maxima are concentrated on the inversion centres; the positive regions of the map are in blue, the negative (corresponding to positive minima of the variance) are in brown.

4. The variance estimate in acentric space groups

Owing to the well known symmetry relationship

$$F_{\mathbf{h}R} = F_{\mathbf{h}}\exp(-2\pi i\mathbf{h}\mathbf{T}) = |F_{\mathbf{h}}|\exp[i(\varphi_{\mathbf{h}} - 2\pi\mathbf{h}\mathbf{T})],$$

the general expression of the electron density for an acentric space group is

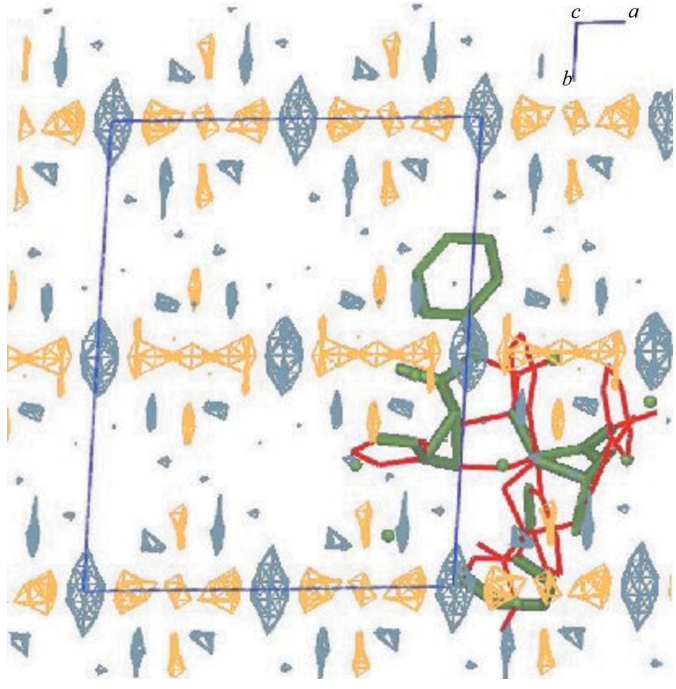


Figure 2
Newqb, $P\overline{1}$ space group: crystal structure in red, model structure in green ($\langle\sigma_A\rangle$ = 0.68). Projection of TH_2 on the plane (**a**, **b**): blue peaks correspond to maxima, regions in brown to negative minima.

$$\rho(\mathbf{r}) = \frac{2}{V}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}| \sum_{s=1}^{n} \cos[\varphi(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r})].$$

The symbol $\sum_{\mathbf{h},\mathrm{ind}}$ indicates that the summation goes over the symmetry-independent reflections; the contribution of the Friedel opposites is contained in the summation. If each phase $\varphi_{\mathbf{h}}$ is assumed to be distributed around $\varphi_{p\mathbf{h}}$ according to the von Mises distribution [equation (1)], then

$$\langle\rho(\mathbf{r})\rangle = \frac{2}{V}\sum_{\mathbf{h},\mathrm{ind}} m_{\mathbf{h}}|F_{\mathbf{h}}| \sum_{s=1}^{n} \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}]$$

and

$$\begin{aligned}\langle\rho(\mathbf{r})\rangle^2 = {}& \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}} m_{\mathbf{h}}^2|F_{\mathbf{h}}|^2 \sum_{s,q=1}^{n} \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ & \times \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_q\mathbf{r}] \\ & + \frac{4}{V^2}\sum_{\mathbf{h}\neq\mathbf{k},\mathrm{ind}} m_{\mathbf{h}}m_{\mathbf{k}}|F_{\mathbf{h}}F_{\mathbf{k}}| \sum_{s,q=1}^{n} \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ & \times \cos[\varphi_p(\mathbf{k}) - 2\pi\mathbf{k}\mathbf{C}_s\mathbf{r}].\end{aligned} \tag{10}$$

Accordingly

$$\begin{aligned}\langle\rho^2(\mathbf{r})\rangle = {}& \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2\Big\langle \sum_{s,q=1}^{n} \cos[\varphi(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ & \times \cos[\varphi(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_q\mathbf{r}]\Big\rangle \\ & + \frac{4}{V^2}\sum_{\mathbf{h}\neq\mathbf{k},\mathrm{ind}} m_{\mathbf{h}}m_{\mathbf{k}}|F_{\mathbf{h}}F_{\mathbf{k}}| \sum_{s,q=1}^{n} \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ & \times \cos[\varphi_p(\mathbf{k}) - 2\pi\mathbf{k}\mathbf{C}_q\mathbf{r}].\end{aligned} \tag{11}$$

Combining equations (10) and (11) gives

$$\begin{aligned}\mathrm{var}\rho(\mathbf{r}) = {}& \langle\rho^2(\mathbf{r})\rangle - \langle\rho(\mathbf{r})\rangle^2 \\ = {}& \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2\Big\langle \sum_{s,q=1}^{n} \cos[\varphi(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ & \times \cos[\varphi(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_q\mathbf{r}]\Big\rangle \\ & - \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}} m_{\mathbf{h}}^2|F_{\mathbf{h}}|^2 \sum_{s,q=1}^{n} \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ & \times \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_q\mathbf{r}],\end{aligned}$$

which, after some calculations, leads to

$$\mathrm{var}\rho(\mathbf{r}) = TH_1 + TH_2(\mathbf{r}) + TD(\mathbf{r}) \tag{12}$$

where

$$TH_1 = \frac{2}{V^2}\sum_{\mathbf{h}>0} (1 - m_{\mathbf{h}}^2)|F_{\mathbf{h}}|^2$$

$$TH_2(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2(1 - m_{\mathbf{h}}^2) \sum_{s\neq q=1}^{n} \cos\{2\pi\mathbf{h}[(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]\}$$

$$TD(\mathbf{r}) = -\frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2[m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})] \times \sum_{s,q=1}^{n} \cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].$$

For $n = 1$ $TH_2(\mathbf{r}) \equiv 0$, and the variance expression described in Paper I for $P1$ is obtained:

$$\begin{aligned}\mathrm{var}\rho(\mathbf{r}) = {}& \frac{2}{V^2}\sum_{\mathbf{h}>0} (1 - m_{\mathbf{h}}^2)|F_{\mathbf{h}}|^2 - \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2[m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})] \\ & \times \cos[2\varphi_p(\mathbf{h}) - 4\pi\mathbf{h}\mathbf{r}].\end{aligned}$$

Equation (12) may be simplified by writing it in a more compact form:

$$\mathrm{var}\rho(\mathbf{r}) = TH(\mathbf{r}) + TD(\mathbf{r}), \tag{13}$$

where

$$TH(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2(1 - m_{\mathbf{h}}^2) \sum_{s,q=1}^{n} \cos[2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]$$

$$TD(\mathbf{r}) = -\frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}|^2[m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})] \times \sum_{s,q=1}^{n} \cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].$$

Since TH_1 does not vary with $\mathbf{r}$ and $TH_2(\mathbf{r}) = TH_2(-\mathbf{r})$, $TH(\mathbf{r})$ is a centrosymmetric function, while $TD(\mathbf{r})$ is non-centrosymmetric.

Another useful form of equation (13) is

$$\begin{aligned}\mathrm{var}\rho(\mathbf{r}) = {}& \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}} (1 - m_{\mathbf{h}}^2) \sum_{s,q=1}^{n} F_{\mathbf{h}\mathbf{R}s}\exp(-2\pi i\mathbf{h}\mathbf{R}_s\mathbf{r})F_{-\mathbf{h}\mathbf{R}q} \\ & \times \exp(2\pi i\mathbf{h}\mathbf{R}_q\mathbf{r}) \\ & - \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}} [m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})] \sum_{s,q=1}^{n} F_{\mathbf{h}\mathbf{R}s}\exp(-2\pi i\mathbf{h}\mathbf{R}_s\mathbf{r})F_{\mathbf{h}\mathbf{R}q} \\ & \times \exp(-2\pi i\mathbf{h}\mathbf{R}_q\mathbf{r}).\end{aligned} \tag{14}$$

Equations (12)–(14) are the formulas we were looking for; they will be used to describe the variance properties.

5. The variance estimate in centric space groups

In accordance with §§3 and 4 the electron density in a centric space group of order n is given by

$$\rho(\mathbf{r}) = \frac{2}{V}\sum_{\mathbf{h},\mathrm{ind}} |F_{\mathbf{h}}| \sum_{s=1}^{n/2} \cos[\varphi(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}], \tag{15}$$

where $n/2$ is the number of symmetry operators not referred by an inversion centre. The limit to the summation is due to the fact that the contribution of the Friedel opposite reflections to $\rho(\mathbf{r})$ is already considered when the cosine, instead of the exponential formulation, is used for defining the electron density. The expected value of $\rho(\mathbf{r})$ is

$$\langle\rho(\mathbf{r})\rangle = \frac{2}{V}\sum_{\mathbf{h},\mathrm{ind}} m_{c\mathbf{h}}|F_{\mathbf{h}}|\sum_{s=1}^{n/2}\cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] = \rho_{\mathrm{obs}}(\mathbf{r}).$$

The same procedure adopted for the acentric space groups leads to the following results:

$$\mathrm{var}\rho(\mathbf{r}) = TH_1 + TH_2(\mathbf{r}) + TD(\mathbf{r}), \tag{16}$$

where

$$TH_1 = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{c\mathbf{h}}^2)|F_{\mathbf{h}}|^2$$

$$TH_2(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|F_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2)\sum_{s\neq q=1}^{n/2}\cos\{2\pi\mathbf{h}[(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]\}$$

$$TD(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|F_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2)\sum_{s,q=1}^{n/2}\cos[2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].$$

Equation (16) may be rewritten as

$$\mathrm{var}\rho(\mathbf{r}) = TH(\mathbf{r}) + TD(\mathbf{r}), \tag{17}$$

where

$$TH(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|F_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2)\sum_{s,q=1}^{n/2}\cos[2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]$$

$$TD(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|F_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2)\sum_{s,q=1}^{n/2}\cos[2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].$$

The variance expression described in §3 for $P\overline{1}$ may be easily obtained from equation (17); indeed in $P\overline{1}$ the relation $TH_2(\mathbf{r}) \equiv 0$ arises because two symmetry operators not related by the inversion centre do not exist. Furthermore, the only pair of subscripts available for the calculation of $TD(\mathbf{r})$ are $s = 1$ and $q = 1$.

A further simplification is possible in centric space groups: since both $\mathbf{C}_s$ and $-\mathbf{C}_s$ belong to the set of symmetry operators (on assuming the origin on an inversion centre), then

$$\mathrm{var}\rho(\mathbf{r}) = 2TH(\mathbf{r}) = \frac{1}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|F_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2) \times \sum_{s,q=1}^{n}\cos[2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]. \tag{18}$$

Equation (18) is the relation we were looking for.

6. Simplified expressions for the variance in acentric and centric space groups

The algebraic formulas obtained in §§4 and 5 for describing the variance components TH_2 and TD involve a summation over symmetry-independent reflections and a double internal summation over the symmetry operators. In order to simplify the analysis of the variance maps we need to rewrite the variance expressions in a more useful form. We notice: (i) $2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r} = 2\pi\mathbf{h}\mathbf{C}_s(\mathbf{I} - \mathbf{C}_s^{-1}\mathbf{C}_q)\mathbf{r} = 2\pi\mathbf{h}\mathbf{C}_s(\mathbf{I} - \mathbf{C}_\tau)\mathbf{r}$ where $\mathbf{C}_\tau = \mathbf{C}_s^{-1}\mathbf{C}_q$; (ii) $2\pi\mathbf{h}\mathbf{C}_s(\mathbf{I} - \mathbf{C}_\tau)\mathbf{r} = 2\pi\mathbf{h}\mathbf{R}_s(\mathbf{I} - \mathbf{C}_\tau)\mathbf{r}$; (iii) the typical expression of the Patterson function quoted in §1 may also be transformed into

$$P(\mathbf{u}) = \frac{2}{V}\sum_{\mathbf{h},\mathrm{ind}}\sum_{s=1}^{n}|F_{\mathbf{h}\mathbf{R}s}|^2\cos(2\pi\mathbf{h}\mathbf{R}_s\mathbf{u}),$$

which explicitly takes the symmetry into account.

As a consequence of the three points above, $TH(\mathbf{r})$ and $TD(\mathbf{r})$, as defined for acentric space groups [see equation (13)], may be rewritten as follows:

$$TH(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}|F_{\mathbf{h}}|^2(1 - m_{\mathbf{h}}^2)\sum_{s=1}^{n}\cos[2\pi\mathbf{h}(\mathbf{I} - \mathbf{C}_s)\mathbf{r}] \tag{19}$$

$$TD(\mathbf{r}) = -\frac{2}{V^2}\sum_{\mathbf{h}>0}|F_{\mathbf{h}}|^2[m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})] \times \sum_{s=1}^{n}\cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{I} + \mathbf{C}_s)\mathbf{r}]. \tag{20}$$

Analogously, the variance expression for centric crystals may be transformed into

$$\mathrm{var}\rho(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}|F_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2)\sum_{s=1}^{n}\cos[2\pi\mathbf{h}(\mathbf{I} - \mathbf{C}_s)\mathbf{r}]. \tag{21}$$

We are now ready to analyse the three components of the variance.

7. The variance component TH_1 and the concept of map variance

TH_1 does not vary with $\mathbf{r}$, is independent of the space-group symmetry and diminishes when CORR increases; it coincides with the map variance defined in Paper I for $P1$. As shown in §4, for a general space group the variance is the sum of a group of three terms – TH_1, $TH_2(\mathbf{r})$ and $TD(\mathbf{r})$. Accordingly, TH_1 has to be considered only the mean value of the variance, when the average is calculated over all the points of the unit cell. We suggest two further interpretations for TH_1:

(*a*) As a figure of merit, able to estimate the average similarity between the target Patterson function $P(\mathbf{u})$ and a weighted Patterson function $P_w(\mathbf{u})$, with model-dependent weights; indeed

$$TH_1 = P(0) - P_w(0),$$

where

$$P_w(0) = \frac{2}{V^2}\sum_{\mathbf{h}>0}m_{\mathbf{h}}^2|F_{\mathbf{h}}|^2$$

is the value of a modified Patterson synthesis with coefficients $m_{\mathbf{h}}^2|F_{\mathbf{h}}|^2$. If the model is uncorrelated with the target then $m_{\mathbf{h}}$ nearly vanishes and TH_1 attains its maximum value. If the model nearly coincides with the target, then $m_{\mathbf{h}}$ approaches unity for all the reflections and TH_1 attains its minimum value.

(*b*) Caliandro, Carrozzini, Cascarano, De Caro, Giacovazzo & Siliqi (2008), under the same assumptions adopted in this paper, derived, for each $\mathbf{h}$ coefficient of a difference Fourier synthesis, the variance expression $\sigma_{q\mathbf{h}} = (1 - m^2)|F_{\mathbf{h}}|^2$. Accordingly TH_1 is nothing but the sum of the variances of the single reflections, considered statistically independent, divided by the squared unit-cell volume:

$$\mathrm{var}\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{h}} \sigma^2_{q\mathbf{h}}. \qquad (22)$$

8. The variance component TH_2 and Patterson deconvolution methods

$TH_2(\mathbf{r})$ is a function varying with $\mathbf{r}$; the distribution of its values in the unit cell depends on the space-group symmetry. The weights $m^2_{\mathbf{h}}$ increase when CORR increases; consequently the average values of the largest peaks in $TH_2(\mathbf{r})$ will decrease when CORR increases.

Before going on in our analysis it may be useful to determine whether TH_2 may be neglected when compared to TH_1. Since TH_1 is independent of $\mathbf{r}$, the simplest way is to calculate, for each pair of target–model structures, the TH_2 map and look for oscTH_2 = max[$|TH_2(\mathbf{r})|/TH_1$]. We will use as target structures a low-symmetry compound (BCDIMP, $P2_1$, chemical content in the asymmetric unit $C_{55}H_{76}N_4O_{37}$; Impellizzeri *et al.*, 2000) and a high-symmetry compound [*i.e.* TOTC, $P6_1$, $C_{33}H_{36}O_6{\cdot}0.2(C_{16}H_{33}OH)$; Sheldrick, 1982]. To take into account the effects of the heavy atoms on the TH_2 map, a third target structure was considered (ERICA, $C_{37}H_{43}FeO_4$, P_1; Othen, 1990). For each target structure four models were simulated, characterized by different values of $\langle\sigma_A\rangle$. The values of oscTH_2 are plotted (blue squares) *versus* the parameter $\langle\sigma_A\rangle$ in Figs. 3–5. We notice: (i) oscTH_2 reaches values up to 0.08 for the light-atom and low-symmetry structure (BCDIMP), 0.46 for the light-atom high-symmetry space group (TOTC), and 0.15 for the structure with a heavy atom (ERICA); (ii) the trend of the oscTH_2 curves *versus* $\langle\sigma_A\rangle$ is not similar in the three figures. The rationale may be the following: both $TH_2(\mathbf{r})$ and TH_1 have the same coefficients [*i.e.* $(1 - m^2_{\mathbf{h}}|F_{\mathbf{h}}|^2)$], and therefore should have a similar trend *versus* $\langle\sigma_A\rangle$. The ratio $TH_2(\mathbf{r})/TH_1$ is therefore more influenced by the specific characteristics of the model rather than the $\langle\sigma_A\rangle$ value.

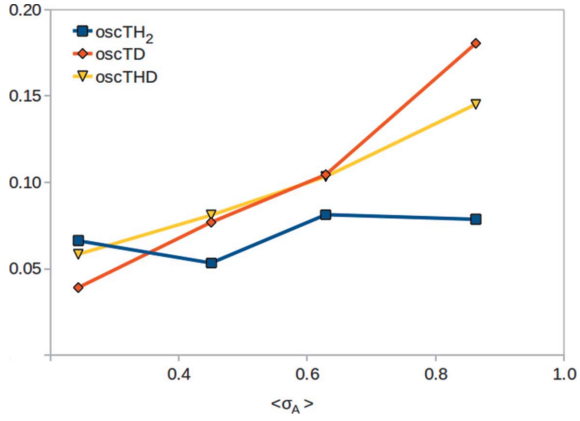


Figure 3
BCDIMP. oscTH_2, oscTD and oscTHD *versus* $\langle\sigma_A\rangle$ for four model structures.

Figure 4
TOTC. oscTH_2, oscTD and oscTHD *versus* $\langle\sigma_A\rangle$ for four model structures.

TH_2 may be closely related to Patterson deconvolution methods. Since early times it was clear that a Patterson map is a sum of structure images as seen from the atomic positions (Wrinch, 1939; Buerger, 1959; Beevers & Robertson, 1950; Garrido, 1950). To derive a single image of the structure from a Patterson map it is necessary to apply a deconvolution process. Quite effective are the *implication transformations* (see Pavelčík, 1988; Pavelčík *et al.*, 1992, and literature quoted therein): they are symmetry operations which transform a region of the Patterson map defined *via* specific symmetry rules into a function whose maxima may coincide with the atomic positions. A typical implication transformation is

$$J_s(\mathbf{r}) = P[(\mathbf{I} - \mathbf{C}_s)\mathbf{r}], \qquad (23)$$

where $(\mathbf{I} - \mathbf{C}_s)\mathbf{r}$ represents any point lying in the Harker (1936) section HS($\mathbf{I}$, $\mathbf{C}_s$) of the Patterson map. From a mathematical point of view, the implication transformation may be accomplished by using the reflexive generalized inverse $\mathbf{D}^*_s$ of the matrix $\mathbf{D}_s = \mathbf{I} - \mathbf{C}_s$ (Ben-Israel & Greville, 1974; Ardito *et al.*, 1985):

$$\mathbf{r} = \mathbf{D}^*_s b(\mathbf{I}, \mathbf{C}_s) + (\mathbf{I} - \mathbf{D}^*_s\mathbf{D}_s)\mathbf{r},$$

where $b(\mathbf{I}, \mathbf{C}_s) = (\mathbf{I} - \mathbf{C}_s)\mathbf{r}$.

If $\mathbf{C}_s$ represents symmetry operators $\overline{1}$, $\overline{3}$, $\overline{4}$, $\overline{6}$ then the three coordinates of $\mathbf{r}$ may be derived without ambiguity; if it

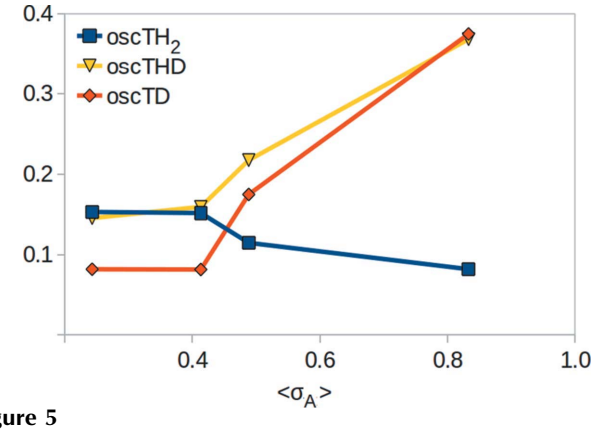


Figure 5
ERICA. oscTH_2, oscTD and oscTHD *versus* $\langle\sigma_A\rangle$ for four model structures.

represents symmetry operators 2, 3, 4, 6 one component of $\mathbf{r}$ remains undefined; if $\mathbf{C}_s$ represents the operator $\overline{2}$, two coordinates remain undefined.

When more symmetry operators coexist the so-called *symmetry minimum function* (SMF) is frequently used, proposed in early times by Buerger (1959) and applied by, among others, Nordman (1966), Richardson & Jacobson (1987) and Sheldrick (1992):

$$\mathrm{SMF}(\mathbf{r}) = \underset{s=2}{\overset{n}{M}} J_s(\mathbf{r}),$$

where the operator M indicates that the lowest value among the $n-1$ (we omit the identity from the set of symmetry operators) functions J_s has to be chosen.

There are several ways of exploiting the information contained in the Harker sections. For example:

(*a*) The Fourier transform of the Harker sections may be used for defining the values of the one-phase semi-invariants (Ardito *et al.*, 1985; Cascarano *et al.*, 1987).

(*b*) By combining the symmetry minimum function with superposition techniques (Pavelčík, 1988; Pavelčík *et al.*, 1992; Sheldrick, 1992, and more recently Caliandro, Carrozzini, Cascarano, De Caro, Giacovazzo, Mazzone & Siliqi, 2008). These last authors were able to solve *ab initio* proteins up to 7890 atoms in the asymmetric unit.

Different uses of the Harker sections imply different types of calculations. When the Harker sections are used for estimating structure semi-invariants or for Patterson deconvolution, only some sections may be considered and the symmetry minimum function should be calculated with respect to them. Indeed (see Ardito *et al.*, 1985 for the algebraic analysis):

(i) An interatomic vector $\mathbf{u} = (\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}$ lying on the Harker section $\mathrm{HS}(\mathbf{C}_s, \mathbf{C}_q)$ may be rewritten as $\mathbf{u} = (\mathbf{I} - \mathbf{C}_q)\mathbf{C}_s\mathbf{r}$. Accordingly, the information contained in sections of type $\mathrm{HS}(\mathbf{C}_s, \mathbf{C}_q)$ is equivalent to that provided by sections of type $\mathrm{HS}(\mathbf{I}, \mathbf{C}_q)$.

(ii) If $\mathbf{C}_s \neq \mathbf{C}_s^{-1}$ then $\mathrm{HS}(\mathbf{I}, \mathbf{C}_s^{-1})$ provides the same information as $\mathrm{HS}(\mathbf{I}, \mathbf{C}_s)$.

(iii) If $\mathbf{C}_q$ is the transform of $\mathbf{C}_v$ by the element $\mathbf{C}_s$ (*i.e.* $\mathbf{C}_q = \mathbf{C}_s\mathbf{C}_v\mathbf{C}_s^{-1}$) then $\mathrm{HS}(\mathbf{I}, \mathbf{C}_q)$ provides the same information as $\mathrm{HS}(\mathbf{I}, \mathbf{C}_v)$.

When Harker sections are used to calculate the variance, all the $\mathrm{HS}(\mathbf{I}, \mathbf{C}_s)$ concur to define it: one should allow s to vary freely from 1 to n. In this case the symmetry minimum function should not be used – it should be replaced by the *symmetry sum function*. Indeed, according to equation (19), $TH(\mathbf{r})$ may be rewritten as

$$TH(\mathbf{r}) = V^{-1} \sum_{s=1}^{n} \{P[(\mathbf{I} - \mathbf{C}_s)\mathbf{r}] - P_w[(\mathbf{I} - \mathbf{C}_s)\mathbf{r}]\}, \qquad (24)$$

where P_w is the Patterson synthesis with coefficients $m_{\mathbf{h}}^2|F_{\mathbf{h}}|^2$. In terms of implication transformations equation (24) may also be rewritten as

$$TH(\mathbf{r}) = V^{-1} \sum_{s=1}^{n} [J_s(\mathbf{r}) - J_{ws}(\mathbf{r})], \qquad (25)$$

where J_{ws} is the implication function corresponding to P_w.

From equations (24) and (25) the following conclusions arise:

(*a*) In acentric space groups the variance, through the component $TH(\mathbf{r})$, is related to the implication transformations $J_s(\mathbf{r})$ directly provided by the Patterson synthesis only at the beginning of the phasing process, when a random model is available. As soon as CORR increases, $J_{ws}(\mathbf{r})$ is no longer negligible and the variance is related to the differences between $J_s(\mathbf{r})$ and $J_{ws}(\mathbf{r})$; it will vanish in the last stages of a successful phasing process.

(*b*) In centric space groups the variance is fully related to the implication transformations because it coincides with $2TH(\mathbf{r})$.

(*c*) When the symmetry minimum function is applied to intersecting Harker sections for defining the atomic positions, it contributes to make the density distribution more concentrated in a specific point of the unit cell. If the symmetry sum function is applied, as for the case of the variance, the density distribution is more dispersed into the unit cell.

Let us use two examples to better describe the properties of the symmetry sum function.

Example (1). Space group $P2$, $\mathbf{C}_1 = \mathbf{I}$, $\mathbf{C}_2 = 2_{[010]}$. According to equation (19)

$$TH(\mathbf{r}) = \frac{2}{V^2} \sum_{\mathbf{h}>0} |F_{\mathbf{h}}|^2 (1 - m_{\mathbf{h}}^2) + \frac{2}{V^2} \sum_{\mathbf{h}>0} |F_{\mathbf{h}}|^2 (1 - m_{\mathbf{h}}^2) \times \cos[2\pi \mathbf{h}(\mathbf{I} - \mathbf{C}_2)\mathbf{r}].$$

TH_1 is constant for any point of the unit cell, TH_2 attains its maxima in all the points $\mathbf{r} = (x, y, z)$ giving rise to maxima in the Harker section $\mathrm{HS}(\mathbf{I}, \mathbf{C}_2)$. Thus, if $(u = 2x, w = 0, w = 2z)$ corresponds to a maximum in the Harker section, then TH_2 will remain constant for all the points (x, y, z), with y free. For example, all the points (x, y, z) have the same TH_2 as the points $(x, 0, z)$: TH_2 is therefore constant along columns parallel to the b axis.

Example (2). Space group $P2_12_12_1$, $\mathbf{C}_1 = \mathbf{I}$, $\mathbf{C}_2 = 2_{1[100]}$, $\mathbf{C}_3 = 2_{1[010]}$, $\mathbf{C}_4 = 2_{1[001]}$. TH_1 is constant for any point of the unit cell. In addition, three sets of columns, parallel to **a**, **b**, **c**, respectively, are originated by maxima in the three Harker sections $\mathrm{HS}(\mathbf{I}, \mathbf{C}_i)$, $i = 2, 3, 4$. The domains where the sets of columns intersect each other are used in different ways according to whether the symmetry minimum function is employed (then the minimum among the density values of the intersecting columns is chosen) or the symmetry sum function is employed (then the sum of the density values of the intersecting columns is used). The variance calculation requires the use of the symmetry sum function.

A particularly important aspect concerns the periodicity of TH_2: we will show that it is related to the Cheshire cell (Hirshfeld, 1968). Let us consider, for a given space group, the *allowed or permissible origins* (see Giacovazzo, 1998, and literature quoted therein): they are those points of the unit cell which, when taken as the origin, maintain the same symmetry operators $\mathbf{C}_s, s = 1, \ldots, n$. They are easily recognizable in *International Tables for Crystallography*, Vol. A (Hahn, 1992), because they have the same 'symmetry environment'. It was

shown by Giacovazzo (1998) that all origins allowed by a fixed functional form of the structure factor are connected by translational vectors $\mathbf{X}_0$ for which

$$(\mathbf{R}_s - \mathbf{I})\mathbf{X}_0 = \mathbf{V},\ s = 1, 2, \ldots, n \qquad (26)$$

where $\mathbf{V}$ is a vector with zero or integer components.

Let us now consider, in the TH_2 expression [see equation (19)], the cosine $\cos[2\pi\mathbf{h}(\mathbf{I} - \mathbf{C}_s)\mathbf{r}]$. It does not change if calculated in the point $\mathbf{r}$ or in $\mathbf{r} + \mathbf{X}_0$, where $\mathbf{X}_0$ is defined by equation (26). Consequently, TH_2 has the same periodicity of a Cheshire cell, *e.g.* in $P2$ the Cheshire cell is defined by ($\mathbf{a}/2$, $\varepsilon\mathbf{b}$, $\mathbf{c}/2$) (ε is an infinitesimal quantity), and in $P2_12_12_1$ by ($\mathbf{a}/2$, $\mathbf{b}/2$, $\mathbf{c}/2$). As a consequence, positions compatible with the Harker sections are repeated in each Cheshire cell by translation, to give the complete TH_2 map [the reader may check Fig. 2, where the Cheshire cell ($\mathbf{a}/2$, $\mathbf{b}/2$, $\mathbf{c}/2$), characteristic of $P\overline{1}$, is clearly recognizable]. It may be useful to recall that, when the symmetry minimum function is used for phasing attempts, this enhanced periodicity (summed to the centric nature of the Patterson map) causes formidable difficulties for the success of the phasing process; adequate filtering algorithms (Burla *et al.*, 2002, 2006) are required to restate the correct space-group symmetry of the target structure and to solve it. The Cheshire periodicity is instead the natural periodicity of TH_2.

Equations (19) and (21) allow us to discover another feature of TH_2, valid for symmorphic space groups, no matter whether they are centric or acentric. The last summation in these equations reduces to $\sum_{s=1}^{n} \cos[2\pi\mathbf{h}(\mathbf{I} - \mathbf{R}_s)\mathbf{r}]$, which attains its maximum value (equal to n) in the origin of the unit cell, where, for acentric crystals,

$$TH_1 = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{\mathbf{h}}^2)|F_{\mathbf{h}}|^2, \quad TH_2(0) = nTH_1;$$

for centric crystals,

$$TH_1 = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{\mathbf{ch}}^2)|F_{\mathbf{h}}|^2, \quad TH_2(0) = nTH_1.$$

Such a TH maximum is independent of the target and of the model structure, and only depends on the space-group symmetry. The question is: is it possible to foresee whether other points in the unit cell exist with the same properties of the origin? The answer is easily obtained from equation (26): indeed for any other permissible origin the sum $\sum_{s=1}^{n} \cos[2\pi\mathbf{h}(\mathbf{I} - \mathbf{R}_s)\mathbf{r}]$ will be equal to n. It may be concluded that in symmorphic space groups all the allowed origins will show TH maxima, and therefore variance maxima, no matter what the model and target structures; the simplest example has been described in §3 and Fig. 2, where the variance maxima on the inversion centres for $P\overline{1}$ are shown.

9. About the variance component $TD(\mathbf{r})$

TD is a function of the point $\mathbf{r}$, is non-centrosymmetric, depends on the space-group symmetry and, through the phase values, on the atomic positions of the model structure. While $TH(\mathbf{r})$ (on average) tends to be maximized when CORR = 0 and minimized when CORR = 1, $TD(\mathbf{r})$ shows a different trend: it tends to vanish both when CORR = 0 and when CORR = 1.

It may be useful to determine whether $TD(\mathbf{r})$ may be neglected when compared to TH_1. The simplest way is to calculate for each pair of target–model structures the $TD(\mathbf{r})$ map and look for

$$\mathrm{osc}TD = \max[|TD(\mathbf{r})|/TH_1].$$

We used the same model–target structures as those described in §8. The values of oscTD are plotted in Figs. 3–5 (red squares) *versus* the parameter $\langle\sigma_A\rangle$. We notice:

(*a*) Unlike TH_2, $TD(\mathbf{r})$ has the characteristic of an observed Fourier synthesis. Indeed it is calculated *via* observed amplitudes, with weights depending on the correlation between model and target structures {*e.g.* $[m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})]$}, and model phases. Thus $TD(\mathbf{r})$ should simultaneously show features of the model and of the target structures, with more emphasis on one or the other structure according to the value of CORR.

(*b*) oscTD reaches values up to 0.18 in BCDIMP; as an effect of the higher symmetry, values up to 0.43 are attained for TOTC, and, as an effect of the heavy atom, values of 0.37 for ERICA.

(*c*) As for TH_2, the trend of the oscTD curves *versus* $\langle\sigma_A\rangle$ is not the same in the three figures (increasing with $\langle\sigma_A\rangle$ in BCDIMP and ERICA, decreasing in TOTC). Indeed both TD and TH_1 tend to be minimized when CORR tends to unity: in these conditions the trend of the ratio $|TD(\mathbf{r})|/TH_1$ is mainly defined by the characteristics of the specific model.

To better understand the properties of $TD(\mathbf{r})$ in any acentric space group we notice:

(i) If $\mathbf{r}$ is an atomic position and $\mathbf{C}_s$ varies over the set of symmetry operators, $(\mathbf{I} - \mathbf{C}_s)\mathbf{r}$ represents the interatomic vectors connecting $\mathbf{r}$ with its symmetry equivalents.

(ii) $(\mathbf{I} - \mathbf{C}_s)\mathbf{r}$ does not change with an origin translation (that allows TH_2 to be related to the Patterson function), while $(\mathbf{I} + \mathbf{C}_s)\mathbf{r}$ changes with it.

(iii) $(\mathbf{I} + \mathbf{C}_s)\mathbf{r}$, when $\mathbf{C}_s$ varies over the set of symmetry operators of the space group, represents the potential interatomic vectors between the atom in $\mathbf{r}$ and the atoms ideally related to it by the symmetry operators $-\mathbf{C}_s$, existing in the related centrosymmetric space group, provided the inversion centre is located at the origin.

On the basis of this interpretation we can apply the implication transformation to the $(\mathbf{I} + \mathbf{C}_s)\mathbf{r}$ positions included in the expression. Accordingly, if $\mathbf{R}_s$ corresponds to a rotation axis of order 3, 4, 6, then the TH density is of the columnar type and the TD density is of the point type (the three coordinates of $\mathbf{r}$ are fixed without ambiguity). If $\mathbf{R}_s$ corresponds to a twofold axis, the TH density is of the columnar type and the TD density is of the planar type, and *vice versa* for a twofold inversion axis.

In agreement with the preceding observation it is easily seen that the periodicities of $TD(\mathbf{r})$ and of $TH_2(\mathbf{r})$ do not coincide. For example, while the periodicity of $TH_2(\mathbf{r})$ in $P2$ reflects the Cheshire cell ($\mathbf{a}/2$, $\varepsilon\mathbf{b}$, $\mathbf{c}/2$), that of $TD(\mathbf{r})$ mostly complies with ($\varepsilon\mathbf{a}$, $\mathbf{b}/2$, $\varepsilon\mathbf{c}$). Indeed the implication transfor-

mation, applied to the point $\mathbf{u}_j = (\mathbf{I} + \mathbf{C}_s)\mathbf{r}_j = (0, 2y_j, 0)$, leads to a density distribution that is constant in the plane $y = y_j$. Accordingly, $TD(\mathbf{r})$ shows a periodicity of $\mathbf{b}/2$, while it is approximately constant in the planes perpendicular to $\mathbf{b}$.

To check if the variance significantly varies from point to point, we plotted in Figs. 3–5 by yellow triangles the values of

$$\mathrm{osc}THD = \max[|TH_2(\mathbf{r}) + TD(\mathbf{r})|/TH_1].$$

It is easily seen that the concept of map variance, assumed to be valid as a first approximation in $P1$, cannot be maintained for higher-symmetry space groups. More precisely, the variance component TH_1 may still be considered the mean value of the map variance, but now the variance significantly varies from point to point.

10. A general expression of the variance valid at any stage of the structure analysis

If CORR is not very high, the contribution to the variance arising from measurement errors is negligible; indeed it depends on $\sigma^2(|F_\mathbf{h}|)$, which is usually quite a small percentage of the $|F_\mathbf{h}|^2$ moduli. At the end of a satisfactory structure refinement, $m_\mathbf{h}^2 \simeq 1$ for nearly all the reflections; then the phases $\varphi_{p\mathbf{h}}$ are distributed about the target values according to Dirac delta functions and the variance arising from the phase uncertainty nearly vanishes: the variance depending on measurement errors is now dominant.

The above situation is ideal and rare; indeed the final structure model may contain some inadequacies (*e.g.* hydrogen positions not well defined, occupancy factors not perfectly fixed, imperfect displacement parameter estimates *etc.*). The usual crystallographic refinement constitutes a model limitation, which may be overcome if favourable conditions exist for multipolar refinement. In this case the phases of some weak reflections, characterized by small $m_\mathbf{h}^2$ values, may still be far away from their correct values: then the variance estimates described in this paper will not be negligible, and may contribute substantially to the total map variance. It may then be useful to combine into a unique variance expression the contribution arising from phase uncertainties and that related to measurement errors.

In §2 we recalled the expression of the variance obtained by Coppens & Hamilton (1968), valid in $P1$ under the condition that model and target phases coincide; then the variance may only originate from measurement errors (say $\Delta|F|$). We want to obtain, under the same conditions, the general expression of the variance valid for any acentric space group (not provided by Coppens & Hamilton). The full procedure described in this paper may be repeated under modified hypotheses. In particular, the electron density to consider is

$$\Delta\rho(\mathbf{r}) = \frac{2}{V}\sum_{\mathbf{h},\mathrm{ind}} \Delta|F_\mathbf{h}| \sum_{s=1}^{n} \cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}],$$

where $\Delta|F_\mathbf{h}| = |F|_\mathrm{obs} - |F|_\mathrm{true}$, and the assumptions are: (i) the same $\Delta|F|$ is associated with symmetry-equivalent reflections; (ii) $\langle\Delta|F|\rangle = 0$, and $\langle\Delta|F_\mathbf{h}|\Delta|F_\mathbf{k}|\rangle = 0$ if $\mathbf{h} \neq \mathbf{k}$; (iii) $\langle\Delta|F_\mathbf{h}|\Delta|F_\mathbf{h}|\rangle = \sigma^2(|F_\mathbf{h}|)$, where $\sigma(|F_\mathbf{h}|)$ is the standard deviation of the observed structure-factor modulus. Since $\langle\Delta\rho(\mathbf{r})\rangle = 0$ we obtain

$$\begin{aligned}\sigma^2[\rho(\mathbf{r})] &= \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}}\sigma^2(|F_\mathbf{h}|)\sum_{s,q=1}^{n}\cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_s\mathbf{r}] \\ &\quad\times\cos[\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}\mathbf{C}_q\mathbf{r}] \\ &= \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}\sigma^2(|F_\mathbf{h}|) + \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}\sigma^2(|F_\mathbf{h}|) \\ &\quad\times\sum_{s\neq q=1}^{n}\cos[2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}] \\ &\quad+\frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}\sigma^2(|F_\mathbf{h}|)\sum_{s,q=1}^{n}\cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].\end{aligned}$$

Such a formula may be combined with equation (12) to obtain a general expression valid at any stage of the crystallographic phasing. In detail

$$\mathrm{var}\rho(\mathbf{r}) = TH_1 + TH_2(\mathbf{r}) + TD(\mathbf{r})$$

where

$$TH_1 = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_\mathbf{h}^2)|F_\mathbf{h}|^2 + \sigma^2(|F_\mathbf{h}|)$$

$$\begin{aligned}TH_2(\mathbf{r}) &= \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}[|F_\mathbf{h}|^2(1 - m_\mathbf{h}^2) + \sigma^2(|F_\mathbf{h}|)] \\ &\quad\times\sum_{s\neq q=1}^{n}\cos\{2\pi\mathbf{h}[(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]\}\end{aligned}$$

$$\begin{aligned}TD(\mathbf{r}) &= -\frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}\{|F_\mathbf{h}|^2[m_\mathbf{h}^2 - D_2(X_\mathbf{h})] - \sigma^2(|F_\mathbf{h}|)\} \\ &\quad\times\sum_{s,q=1}^{n}\cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].\end{aligned}$$

If CORR $\neq 1$, the contribution to the variance provided by measurement errors is negligible.

Let us now extend to all centric space groups the $P\bar{1}$ Coppens & Hamilton (1968) formula for the variance, in such a way that the new expression is valid at any stage of the structure analysis. We obtain

$$\begin{aligned}\mathrm{var}\rho(\mathbf{r}) &= \frac{1}{V^2}\sum_{\mathbf{h},\mathrm{ind}}[|F_\mathbf{h}|^2(1 - m_{c\mathbf{h}}^2) + \sigma^2(|F_\mathbf{h}|)] \\ &\quad\times\sum_{s,q=1}^{n}\cos[2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}].\end{aligned}$$

11. Variance of *E*-Fourier syntheses

In the everyday work of a crystallographer E-Fourier syntheses are often preferred to F syntheses, particularly in the steps dedicated to crystal structure solution. It is also therefore useful to derive the variance for these maps. Luckily the

method described above may be applied without substantial modifications. The corresponding final formulas are:

(*a*) For acentric crystals,

$$\mathrm{var}[\rho(\mathbf{r})]_N = TH_{N1} + TH_{N2}(\mathbf{r}) + TD_N(\mathbf{r}),$$

where

$$TH_{N1} = \frac{2}{V^2}\sum_{\mathbf{h}>0}(1 - m_{\mathbf{h}}^2)|E_{\mathbf{h}}|^2$$

$$TH_{N2}(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|E_{\mathbf{h}}|^2(1 - m_{\mathbf{h}}^2)\sum_{s\neq q=1}^{n}\cos\{2\pi\mathbf{h}[(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}]\}$$

$$TD_N(\mathbf{r}) = -\frac{2}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|E_{\mathbf{h}}|^2[m_{\mathbf{h}}^2 - D_2(X_{\mathbf{h}})] \times \sum_{s,q=1}^{n}\cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{C}_s + \mathbf{C}_q)\mathbf{r}].$$

(*b*) For centric crystals

$$\mathrm{var}[\rho(\mathbf{r})]_N = \frac{4}{V^2}\sum_{\mathbf{h},\mathrm{ind}}|E_{\mathbf{h}}|^2(1 - m_{c\mathbf{h}}^2)\sum_{s,q=1}^{n}\cos[2\pi\mathbf{h}(\mathbf{C}_s - \mathbf{C}_q)\mathbf{r}].$$

An interesting characteristic of the *E*-map variance is the enhanced 'peakiness' with respect to the corresponding *F* maps; this property reflects the larger peakiness of the *E* with respect to the *F* electron-density maps. To give an example, we report in Fig. 6 oscTH_2, oscTD and oscTHD *versus* $\langle\sigma_A\rangle$ calculated for a light-atom acentric structure (BCDIMP) by using *E*-Fourier coefficients and the same four models employed in Fig. 3. We notice: (i) oscTD and oscTHD strongly increase for high $\langle\sigma_A\rangle$ values (up to 0.78, *versus* the value of 0.18, the maximum attained for an *F* synthesis); (ii) the increase of the variance is more evident at high values of $\langle\sigma_A\rangle$, when peaks are in greater contrast with the background.

It may be worthwhile noticing that the variance for *E* maps is expected to be on a much smaller scale than the variance for *F* maps. That does not mean that the phase uncertainty is smaller, but only that usually $\langle|E|^2\rangle \ll \langle|F|^2\rangle$. A correct estimate of the information provided by any map should make use of the variance-based signal-to-noise ratio, as described in §14 (see also Paper I).

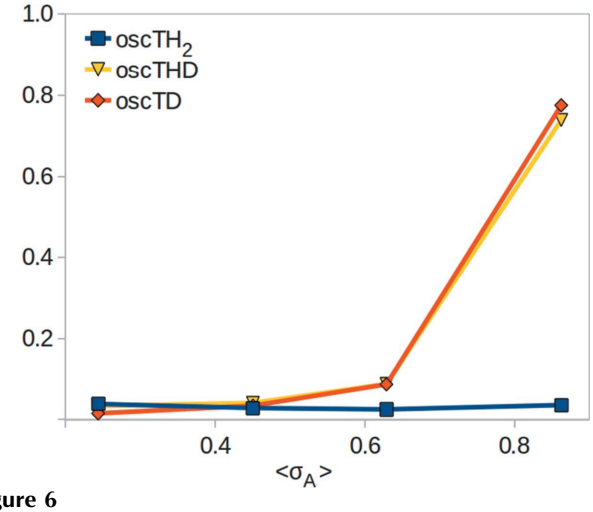


Figure 6
BCDIMP. *E*-Fourier synthesis, oscTH_2, oscTD and oscTHD *versus* $\langle\sigma_A\rangle$ for the same four model structures used in Fig. 3.

12. Variance estimates for hybrid Fourier syntheses

We have shown in Paper I that, if τ and ω are any pair of real numbers, and $\rho_Q(\mathbf{r}) = \tau\rho(\mathbf{r}) - \omega\rho_p(\mathbf{r})$ a general hybrid electron density, its variance in $P1$ is given by

$$\mathrm{var}\rho_Q(\mathbf{r}) = \tau^2\mathrm{var}\rho(\mathbf{r}). \quad (27)$$

Equation (27) is valid in any acentric or centric space group provided var$\rho(\mathbf{r})$ coincides with equation (12) or equation (18), respectively. Owing to its wide use, equation (27) is particularly interesting for the difference electron density, characterized by $\tau = \omega = 1$.

13. Drawing the variance: a simple example

It may be useful to show by a simple example how the variance maxima and minima are related to the target and to the model structures, and how such minima and maxima are influenced by the symmetry. To make things simpler we will consider quite a simple space group, $P2_1$, and a simulated, non-realistic target structure with only three atoms in a nearly empty unit cell, to avoid overlapping of the interatomic vectors in the Patterson map. We will locate three oxygens, O_1, O_2, O_3, in the positions (0.00, 0.10, 0.15), (0.20, 0.20, 0.00) and (0.30, 0.40, 0.35), respectively. The corresponding Harker peaks will be characterized by the following coordinates:

$O_1{-}O_1$: (0.00, 0.50, 0.30), (0.00, 0.50, 0.70),

$O_2{-}O_2$: (0.40, 0.50, 0.00), (0.60, 0.50, 0.00),

$O_3{-}O_3$: (0.60, 0.50, 0.70), (0.40, 0.50, 0.30).

We will first consider the case in which a model is not available (case 1, only observed amplitudes are known); then we will create three ideal model structures with different values of CORR to explore where the variance attains its maxima and minima (cases 2–4).

Case 1. No model is available. This situation corresponds to the following statistical problem: estimate the variance of a hypothetic electron-density map, given the Patterson synthesis. The reader may notice that, even if the electron-density map cannot be calculated because no phase information is available, its variance may be estimated according to the present theory. On assuming that the phases are randomly dispersed between 0 and 2π, $TH_1 + TH_2(\mathbf{r})$ may be computed by fixing $m_{\mathbf{h}} = 0$ for all the reflections. In this case TD (**r**) = 0, and the variance coincides with $TH_1 + TH_2(\mathbf{r}) = V^{-1}P(0) + V^{-1}P[(\mathbf{I} - \mathbf{C}_2)\mathbf{r}]$, which only depends on the observed amplitudes.

The maxima of the variance are expected to lie on the points of the unit cell indicated by the implication transformations, that is along the columns centred on the lines (0.00, *y*, 0.15), (0.20, *y*, 0.00) and (0.30, *y*, 0.35), respectively; the

intensity is constant along the columns. Obviously, symmetry-equivalent columns, displaced according to the periodicity of the Cheshire cell, will also be present. A graphical representation of the variance, more precisely a projection on the plane (**a**, **c**) of the term $TH_2(\mathbf{r})$, is shown in Fig. 7; the variance columns, parallel to **b**, contain the oxygen positions, marked in the figure by red crosses. The Cheshire periodicity and the centric nature of $TH_2(\mathbf{r})$ are clearly recognizable.

The result agrees well with expectations: the variance must be larger along the columns indicated by the implication transformations because their density is expected to vary substantially during the phasing process.

If one of the three oxygen atoms constituting the target structure is replaced by a heavy atom, the columns corresponding to the heavy-atom–heavy-atom Harker vectors should have a larger variance intensity. That agrees well with the expectations. Indeed, during the phasing process, the variation of the density of the pixels belonging to the columns should vary in a dramatic way.

The indications of the variance for any attempt at crystal structure solution are the following: try to locate the atoms of a possible model inside the columns defined by the implication transformations. There the variance is larger, and therefore the electron density is expected to increase. The above example suggests a first conclusion: the variance expression '*knows*' the theory of the implication transformations and suggests locating atoms according to it.

Case 2. Model atoms lie far away from target atomic positions, CORR = 0. We choose a model structure with three oxygen atoms (O_1, O_2, O_3) located at (0.10, 0.10, 0.30), (0.15, 0.30, 0.20) and (0.05, 0.40, 0.40), respectively (see green crosses in Fig. 7). Since $m_{\mathbf{h}} \simeq 0$ for all the reflections, $TD(\mathbf{r}) = 0$ and, as in case 1, $TH(\mathbf{r}) = V^{-1}P(0) + V^{-1}P[(\mathbf{I} - \mathbf{C}_2)\mathbf{r}]$; *e.g.* the variance is not influenced by the specific model structure, provided it is not correlated with the target structure. Accordingly, the variance will attain its maxima exactly along the columns defined for case 1 and vanishes along the columns (0.10, y, 0.30), (0.15, y, 0.20) and (0.05, y, 0.40), containing the model atomic positions. That agrees well with expectations; indeed the variance density is not expected to vary during the phasing process.

Case 3. CORR = 1, model and target atomic positions coincide, $m_{\mathbf{h}}$ close to unity for nearly all the reflections. Then $TH_2 = TH_1 = TD = 0$ and the variance vanishes in all the points of the unit cell. That is exactly what one may expect: the electron density is not allowed to vary any more because the phases are distributed according to Dirac delta functions.

Case 4. CORR is in between 0 and 1, some model atoms lie far away from target atomic positions. This is the most interesting case: we choose for O_1 the same position occupied in the target structure, and for O_2 and O_3 the same positions as in case 2 (see Fig. 8). We calculated for this model $\langle\sigma_A\rangle = 0.63$; $m_{\mathbf{h}}$ will then be close to unity for reflections with large values of $R_{\mathbf{h}}$ and $R_{p\mathbf{h}}$, and close to zero for weak reflections. According to equation (23), the variance is the sum of:

(*a*) The constant term $TH_1 = V^{-1}[P(0) - P_w(0)]$. Its value depends on the correlation between the model and target structures.

(*b*) The variable term $TH_2(\mathbf{r}) = V^{-1}\{P[(\mathbf{I} - \mathbf{C}_2)\mathbf{r}] - P_w[(\mathbf{I} - \mathbf{C}_2)\mathbf{r}]\}$. As in case 1 the component $V^{-1}P[(\mathbf{I} - \mathbf{C}_2)\mathbf{r}]$ attains its maxima on the columns (0.00, y, 0.15), (0.20, y, 0.00)

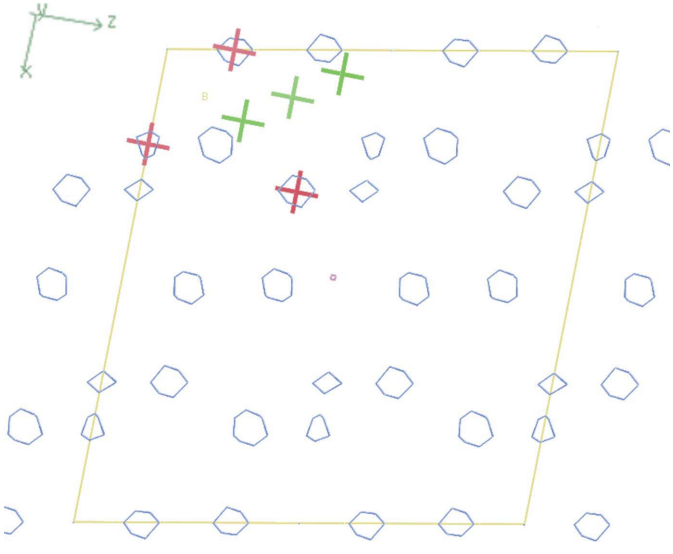


Figure 7
Three O-atom model structure, assigned space group $P2_1$. Projection on the plane (**a**, **c**) of $TH_2(\mathbf{r})$ in the case in which no model structure is available, or a model uncorrelated with the target structure has been fixed (*i.e.* $\langle\sigma_A\rangle = 0$). In both cases $\mathrm{var}\rho(\mathbf{r}) \equiv TH_2(\mathbf{r}) + TH_1$. The red crosses coincide with target atomic positions, the green crosses with the model atomic positions.

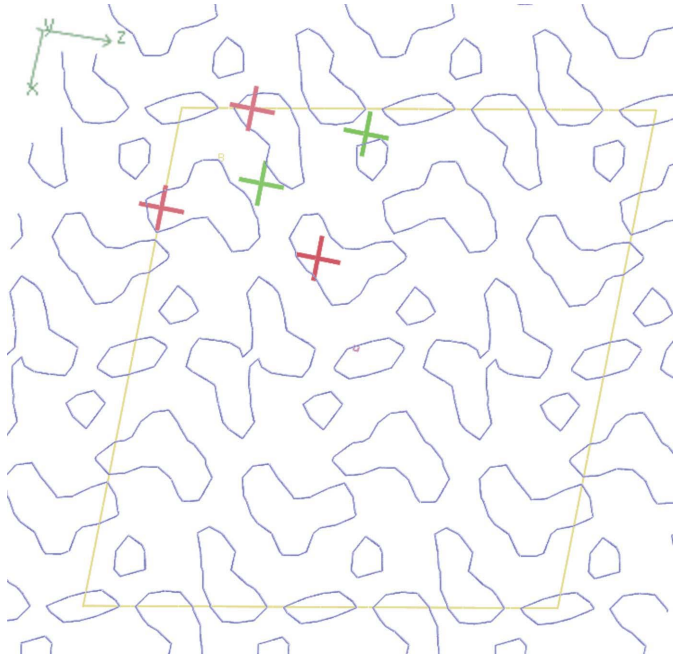


Figure 8
Projection on the plane (**a**, **c**) of $TH_2(\mathbf{r})$ for a three O-atom model and target structure ($\langle\sigma_A\rangle = 0.63$), assigned space group $P2_1$. The red crosses coincide with target atomic positions, the green crosses with the model atomic positions. One atomic position is in common at (0.00, 0.10, 0.15).

and (0.30, y, 0.35), and symmetry equivalents, and is null along the columns (0.15, y, 0.20) and (0.05, y, 0.40).

The component $V^{-1}P_w[(\mathbf{I}-\mathbf{C}_2)\mathbf{r}]$ will subtract density to the columns suggested by the implication transformations by a quantity which increases with CORR. If we consider $|F_\mathbf{h}|^2(1-m_\mathbf{h}^2)$, the coefficient of both TH_1 and $TH_2(\mathbf{r})$, we see that the contribution to the variance of the reflections with the largest observed and calculated amplitudes nearly vanishes (for them $m_\mathbf{h}^2 \simeq 1$). Accordingly, TH_1 and $TH_2(\mathbf{r})$ contribute to the variance mostly through reflections with large observed and small calculated values. The projection of $TH_2(\mathbf{r})$ on the plane (**a**, **c**) is shown in Fig. 9; red crosses and green crosses indicate the atomic positions of the target and of the model structure, respectively. Maxima of $TH_2(\mathbf{r})$ are still on the columns passing through the target atomic positions, but the corresponding density values are weaker. The $TH_2(\mathbf{r})$ values are still vanishing at the model positions far from the target atoms [owing to the algebraic form of the implication functions, $TH_2(\mathbf{r})$ cannot significantly contribute to their density].

(*c*) The variable term $TD(\mathbf{r})$. Let us write

$$TD(\mathbf{r}) = -TD_1(\mathbf{r}) - TD_2(\mathbf{r}) - TD_3(\mathbf{r})$$

and analyse the three components (defined below) separately. We will look for the maxima of TD_1, TD_2, TD_3 by bearing in mind that they correspond to minima of $TD(\mathbf{r})$ and, probably, to variance minima.

(i)

$$TD_1(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}|F_\mathbf{h}|^2[m_\mathbf{h}^2 - D_2(X_\mathbf{h})]\cos[2\varphi_p(\mathbf{h}) - 4\pi\mathbf{hr}]. \tag{28}$$

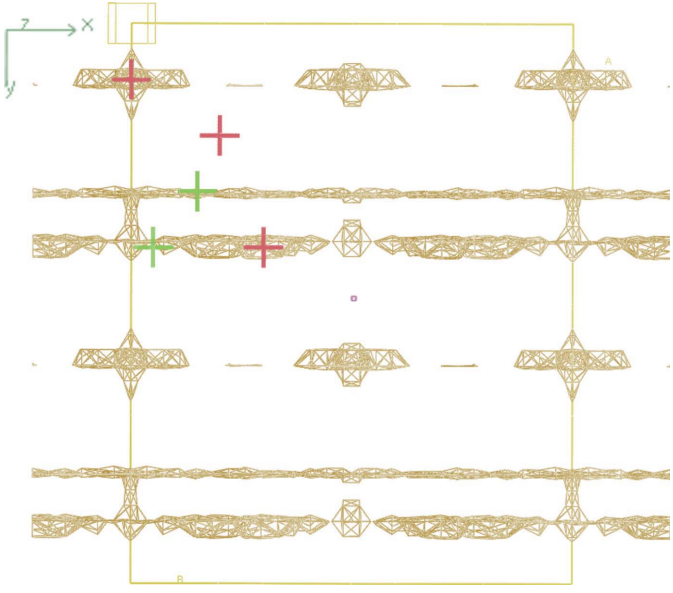


Figure 9
Three O-atom model and target structures ($\langle\sigma_A\rangle$ = 0.63), assigned space group $P2_1$. Projection of $TD(\mathbf{r})$ on the plane (**a**, **b**). The red crosses coincide with target atomic positions, the green crosses with the model atomic positions. One atomic position is in common at (0.00, 0.10, 0.15).

Appendix *A* suggests that maxima of $(2/V^2)\sum_{\mathbf{h}>0}|F_{p\mathbf{h}}|^2\cos[2\varphi_p(\mathbf{h}) - 4\pi\mathbf{hr}]$ should lie at $\mathbf{u} = (\mathbf{r}_{pj1} + \mathbf{r}_{pj2})/2$. Since model phases and target amplitudes are used in equation (28), maxima should stay in correspondence with half-sums of target positional vectors and of model positional vectors, provided $\langle\sigma_A\rangle$ is sufficiently high. Otherwise the maxima may lie far away from the above ideal scheme, even because the weights $[m_\mathbf{h}^2 - D_2(X_\mathbf{h})]$ minimize the contribution from the reflections with the largest observed and calculated amplitudes.

(ii)

$$TD_2(\mathbf{r}) = \frac{2}{V^2}\sum_{\mathbf{h}>0}|F_\mathbf{h}|^2[m_\mathbf{h}^2 - D_2(X_\mathbf{h})]\cos[2\varphi_p(\mathbf{h}) - 4\pi\mathbf{hC}_2\mathbf{r}].$$

The situation is very similar to case (i): the only difference is that $\mathbf{u} = (\mathbf{r}_{pj1} + \mathbf{r}_{pj2})/2$ is replaced by its symmetry equivalent $\mathbf{u} = \mathbf{C}_s(\mathbf{r}_{pj1} + \mathbf{r}_{pj2})/2$.

(iii)

$$TD_3(\mathbf{r}) = \frac{4}{V^2}\sum_{\mathbf{h}>0}|F_\mathbf{h}|^2[m_\mathbf{h}^2 - D_2(X_\mathbf{h})]\cos[2\varphi_p(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{I}+\mathbf{C}_2)\mathbf{r}].$$

In accordance with §9 the implication transformation calculated for the point $(\mathbf{I}+\mathbf{C}_2)\mathbf{r}$ implies maxima distributed along sheets centred on the planes (x, 0.10, z), (x, 0.30, z) and (x, 040, z); 0.10, 0.30, 0.40 are the y coordinates of the model atoms. The $TD_3(\mathbf{r})$ maps, however, cannot behave in strict agreement with the above indications because observed amplitudes and special weights are used.

These circumstances make the map quite dependent on the CORR value, on the specific model structure *etc.*

In Fig. 9 we show the $TD(\mathbf{r})$ projection on the plane (**a**, **b**). It is easily seen that the most evident features of the map are the sheets defined by the TD_3 implication transformation. Accordingly, in Fig. 9 such sheets are marked in brown: they correspond to the TD_3 maxima and therefore to the TD negative minima. To aid readers, we notice that two target atoms happen to lie on such sheets, and one of them overlaps with a model atom.

An overview of the variance distribution may be obtained by looking at the $TH_2(\mathbf{r}) + TD(\mathbf{r})$ map (see Fig. 10); we omitted TH_1 (it is constant) to emphasize the contrast of the map. Accordingly, the negative regions of the $TH_2(\mathbf{r}) + TD(\mathbf{r})$ map (in brown) correspond to the (positive) minima of the variance map. We observe:

(*a*) Since the algebraic form of the implication function hinders TH_2 from providing a significant contribution to the variance density at the model atomic positions, there the variance mainly arises from the TD contribution. According to Fig. 9, the maxima of TD_3 lie on sheets centred on the y coordinate of the model atomic positions. The overall result is the following: $TH_2(\mathbf{r}) + TD(\mathbf{r})$ is negative, and therefore the variance is positive and minimum, on the sheets defined by the TD_3 component.

(*b*) The positive maxima of $TH_2(\mathbf{r}) + TD(\mathbf{r})$ (in blue), and therefore of the variance, mostly coincide with the maxima of TH_2. According to the algebraic form of the implication

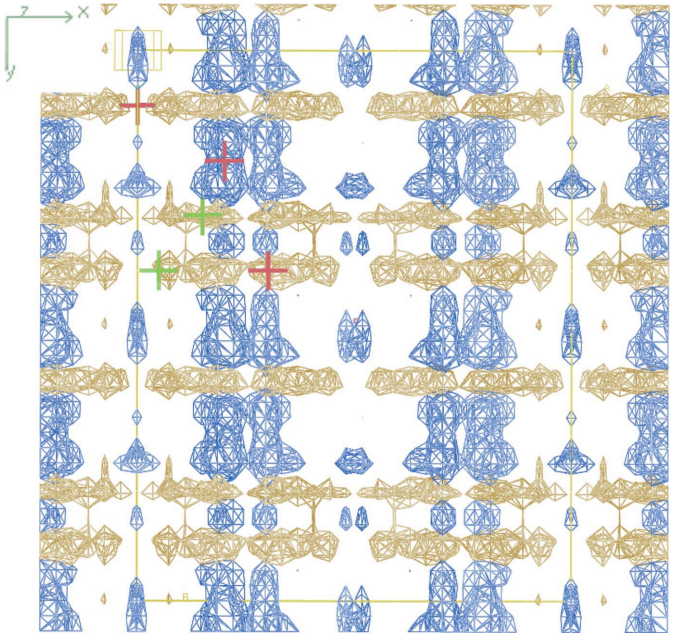


Figure 10
Three O-atom model and target structures, assigned space group $P2_1$. $TH_2(\mathbf{r}) + TD(\mathbf{r})$ map projected on the plane (**a**, **b**): we omitted TH_1 (which is constant) to emphasize the contrast of the map. The most negative regions of the map are in brown, the most positive regions in blue. Red crosses coincide with target atomic positions, green crosses with model atomic positions. One atomic position is in common at (0.00, 0.10, 0.15). The map is representative of the variance: the blue regions correspond to the variance maxima, the brown regions to the variance positive minima.

transformations, the maxima correspond to columns parallel to **b**. The variance density tends to vanish or to become negative where the columns intersect the TD_3 planes; this is the condition for two of the three target atoms.

What is the rationale of a variance density like that shown in Fig. 10? If a model atom does not coincide with a target atom, the corresponding electron density is expected to be small, and it will vanish when the target structure is recovered. Accordingly the variance is expected to be small at such positions.

If we consider the case in which model and target atom coincide, the corresponding electron density will be much larger when the full structure is recovered. Accordingly, the variance is expected to be large at such positions.

But why does the variance present sheets and columns rather than globular domains centred on model or target atoms? This is a kind of indeterminacy which is a consequence of the specific symmetry: *e.g.* intersecting Harker sections would make the variance distribution more concentrated in globes rather than dispersed in sheets.

14. The signal/noise ratio in electron-density maps

The theory described in this paper suggests an interesting question: can we estimate the conditional probability of an electron density $\rho(\mathbf{r})$ in a point $\mathbf{r}$ of the map given the prior knowledge of a structural model? In view of this we should consider $\rho(\mathbf{r})$ as a variable itself, a function of other primitive variables, the distribution of which should be *a priori* known. The answer is simpler if the central limit theorem is invoked. Let us consider the acentric space groups:

(*a*) According to the definition given in §1 the electron density is the sum of a large number of terms, each $\mathbf{h}$ term represented by $(2/V)R_{\mathbf{h}}\cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{\mathbf{h}})$.

(*b*) Each $\mathbf{h}$ term is statistically independent from the others, with the mean value given by $(2/V)m_{\mathbf{h}}R_{\mathbf{h}}\cos(2\pi\mathbf{h}\cdot\mathbf{r} - \varphi_{\rho\mathbf{h}})$ and variance given by

$$\sigma^2_{\mathbf{h}} = \frac{2}{V^2}[|F_{\mathbf{h}}|^2(1 - m^2_{\mathbf{h}}) + \sigma^2(|F_{\mathbf{h}}|)]\sum_{s=1}^{n}\cos 2\pi\mathbf{h}(\mathbf{I} - \mathbf{C}_s)\mathbf{r} - \frac{2}{V^2}\{|F_{\mathbf{h}}|^2[m^2_{\mathbf{h}} - D_2(X_{\mathbf{h}})] - \sigma^2(|F_{\mathbf{h}}|)\} \times \sum_{s=1}^{n}\cos[2\varphi_\rho(\mathbf{h}) - 2\pi\mathbf{h}(\mathbf{I} + \mathbf{C}_s)\mathbf{r}].$$

In accordance with the central limit theorem $\rho(\mathbf{r})$ will have a Gaussian distribution with mean value equal to the sum of the expectation values of each $\mathbf{h}$ term [*i.e.* $\rho_{\rm obs}(\mathbf{r})$], and with variance equal to the sum of the variances of the $\mathbf{h}$ terms [*i.e.* $\mathrm{var}\rho(\mathbf{r}) = \sum_{\mathbf{h}>0}\sigma^2_{\mathbf{h}}$]. We obtain for $\rho(\mathbf{r})$ the conditional probability

$$P[\rho(\mathbf{r})|\rho_p(\mathbf{r})] \simeq \frac{1}{\sigma[\rho(\mathbf{r})](2\pi)^{1/2}}\exp\left(-\frac{1}{2}\left\{\frac{[\rho(\mathbf{r}) - \rho_p(\mathbf{r})]}{\sigma[\rho(\mathbf{r})]}\right\}^2\right), \tag{29}$$

where $\sigma[\rho(\mathbf{r})] = [\mathrm{var}\rho(\mathbf{r})]^{1/2}$.

For centric space groups the distribution [equation (29)] is still valid, but

$$\sigma^2_{\mathbf{h}} = \frac{2}{V^2}[|F_{\mathbf{h}}|^2(1 - m^2_{c\mathbf{h}}) + \sigma^2(|F_{\mathbf{h}}|)]\sum_{s=1}^{n}\cos 2\pi\mathbf{h}(\mathbf{I} - \mathbf{C}_s)\mathbf{r}$$

and $\mathrm{var}\rho(\mathbf{r}) = \sum_{\mathbf{h}>0}\sigma^2_{\mathbf{h}}$.

The reader may notice that in our mathematical treatment we did not consider the effects of the limited data resolution on the conditional distribution [equation (29)]. This subject is beyond the aims of this paper and will be considered in a future work.

Equation (29) prompts two additional interesting questions: may an active variance criterion be useful during the phasing process, in particular in EDM procedures? May a variance criterion, combining uncertainty of the phases with measurement errors and taking full account of the space-group symmetry, be useful at the end of a successful standard (say, by using the spherical-atom hypothesis) crystallographic refinement?

A complete answer to the above questions is outside the aims of this paper, and requires extended calculations in a wide crystallographic area. We limit ourselves to guessing about the potential advantages arising from the active use of a criterion based on the variance.

Let us first briefly discuss the first question. In the absence of any additional prior information (*e.g.* the envelope for proteins) the most common criterion for selecting the pixels to use in the Fourier inversion of a modified electron-density

map is to fix a threshold TRH for the $\rho_{obs}(\mathbf{r})$ densities (all the pixels are set to zero unless their density is larger than TRH). This practice presupposes that the pixel density is the signal and that the variance for any pixel of the unit cell is constant. The theory described in this paper states that the variance varies from point to point in the unit cell, and allows one to calculate, for any type of electron density (observed, difference, hybrid) and for any model (no matter whether poor or not), the corresponding variance values. Then a different criterion may be used to select the pixels to use in the electron-density Fourier inversion: for example, the *signal/noise ratio*, defined as

$$S(\mathbf{r})/N = \rho_{obs}(\mathbf{r})/\sigma[\rho(\mathbf{r})], \tag{30}$$

where $\rho_{obs}(\mathbf{r})$ may be considered the available signal. All the pixel densities are set to zero before the Fourier inversion unless they are larger than a given threshold for $S(\mathbf{r})/N$.

The criterion [equation (30)] allows us to compare the quality of the information contained in an electron-density map with that carried out by the related E map or by the difference or hybrid Fourier synthesis. As an example, the variance in a point $\mathbf{r}$ of a difference electron density is equal to that of the corresponding electron-density synthesis in the same point $\mathbf{r}$. However, it is well known that the quality of the information is quite different. The reason may be easily understood by calculating the $S(\mathbf{r})/N$ ratios for the two syntheses (see Paper I).

A further advantage of the criterion [equation (30)] is that it automatically takes into consideration the quality of the model. For example, if one chooses, in an EDM procedure, to Fourier invert a fixed percentage of pixels (*e.g.* 10%, those corresponding to the largest density values) in all the EDM cycles, very likely different thresholds of $S(\mathbf{r})/N$ are automatically fixed as a consequence of those choices.

Concerning the second question, it is more realistic to think that some residual model misfit exists at any stage of the structure refinement. Therefore, combining the residual phase ambiguity with measurement errors, and taking full account of the space-group symmetry, may provide a supplementary tool for a successful multipolar structure refinement.

15. Conclusions

In the literature formulas estimating the variance in any point of an electron-density map are available; they are only valid in $P1$ and in $P\overline{1}$, disregard the effects of the space-group symmetry and only take into account measurement errors. As a consequence they may be applied only at the end of a successful phase refinement (*e.g.* when aspherical-atom scattering factors are introduced for the study of charge-density distributions), mostly in $P\overline{1}$, when phases are considered correctly fixed.

In this paper new algebraic expressions for estimating the variance in any point of the unit cell are derived. The formulas take into account: (i) the space-group symmetry (it is basically incorrect to disregard it, as the symmetry heavily modifies the variance); (ii) the uncertainty on the current phases. As a consequence our theory may be applied at any stage of the crystal structure determination process, both during the phasing step and at the phase-refinement stage. In other words, the theory is valid for poor- as well as for high-quality structural models.

The new variance expression has been subdivided into three components, the main properties of which have been characterized as follows:

(*a*) A term (TH_1) that does not vary from point to point but depends on the misfit between model and target structures. It fixes the average variance of the map.

(*b*) A second component (TH_2), varying from point to point, strictly connected with the implication transformations, a quite basic tool for the Patterson deconvolution. In some way TH_2 'knows' the theory of the implication transformation: its main task is to fix, *via* weights depending on the misfit between model and target structures, the variance on the points indicated by the implication transformations (*i.e.* the points where the target atoms may potentially stay).

(*c*) A third term (TD), varying from point to point, depending on the model phases and on the observed moduli. Its main task is to fix, *via* weights depending on the misfit between model and target structures, the variance on points related to the model structure.

The variance expression, depending on measurement error and which is available in the literature, has been generalized to all the space groups. Owing to the orthogonality between phase uncertainty (depending on the model) and measurement errors (depending on the experiment) unique expressions of the variance taking into account both phase and measurement errors have been derived for any space group.

APPENDIX *A*

It is well known that

$$\begin{aligned} P(\mathbf{u}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) &= \frac{1}{V^2}\int_V \sum_{\mathbf{h},\mathbf{k}} F_{\mathbf{h}}F_{\mathbf{k}} \exp -2\pi i[\mathbf{kr} + \mathbf{h}(\mathbf{u}+\mathbf{r})]\,\mathrm{d}\mathbf{r} \\ &= \frac{1}{V^2}\sum_{\mathbf{h},\mathbf{k}} F_{\mathbf{h}}F_{\mathbf{k}} \exp(-2\pi i\mathbf{hu}) \int_V \exp -[2\pi i(\mathbf{h}+\mathbf{k})\mathbf{r}]\,\mathrm{d}\mathbf{r}, \end{aligned} \tag{31}$$

where the symbol '*' defines the convolution operation. The integral on the right-hand side of equation (31) vanishes except when $\mathbf{k} = -\mathbf{h}$: accordingly

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 \exp(-2\pi i\mathbf{hu}). \tag{32}$$

To check where $P(\mathbf{u})$ is maximum let us replace in equation (32) $|F_{\mathbf{h}}|^2$ by its algebraic expression. We have

$$P(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} \sum_{j1,\,j2=1}^{N} f_{j1}f_{j2} \exp[2\pi i\mathbf{h}(\mathbf{r}_{j1} - \mathbf{r}_{j2} - \mathbf{u})],$$

which suggests maxima at $\mathbf{u} = \mathbf{r}_{j1} - \mathbf{r}_{j2}$.

Let us now consider

$$\rho_{\rm FF}(\mathbf{u}) = \rho(\mathbf{r}) * \rho(\mathbf{r})$$
$$= \frac{1}{V^2}\int_V \sum_{\mathbf{h},\mathbf{k}} F_{\mathbf{h}}F_{\mathbf{k}} \exp -2\pi i[\mathbf{kr} + \mathbf{h}(\mathbf{u} - \mathbf{r})]\,{\rm d}\mathbf{r}$$
$$= \frac{1}{V^2}\sum_{\mathbf{h},\mathbf{k}} F_{\mathbf{h}}F_{\mathbf{k}} \exp(-2\pi i\mathbf{hu}) \int_V \exp -[2\pi i(\mathbf{k} - \mathbf{h})\mathbf{r}]\,{\rm d}\mathbf{r}.$$

The last integral vanishes except when $\mathbf{k} = \mathbf{h}$: accordingly

$$\rho_{\rm FF}(\mathbf{u}) = \frac{1}{V}\sum_{\mathbf{h}} F_{\mathbf{h}}F_{\mathbf{h}} \exp(-2\pi i\mathbf{hu}) = \frac{2}{V}\sum_{\mathbf{h}>0} |F_{\mathbf{h}}|^2 \cos(2\varphi_{\mathbf{h}} - 2\pi\mathbf{hu}). \tag{33}$$

To check where $\rho_{\rm FF}(\mathbf{r})$ is maximum let us replace in the middle term of equation (33) $F_{\mathbf{h}}F_{\mathbf{h}}$ by its algebraic expression. We have

$$\rho_{\rm FF}(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{h}} \sum_{j1,j2=1}^{N} f_{j1}f_{j2} \exp[2\pi i\mathbf{h}(\mathbf{r}_{j1} + \mathbf{r}_{j2} - \mathbf{u})] \tag{34}$$

which suggests a maximum at $\mathbf{u} = \mathbf{r}_{j1} + \mathbf{r}_{j2}$.

References

Ardito, G., Cascarano, G., Giacovazzo, C. & Luić, M. (1985). *Z. Kristallogr.* **172**, 25–34.

Beevers, C. A. & Robertson, J. H. (1950). *Acta Cryst.* **3**, 164.

Ben-Israel, A. & Greville, T. N. E. (1974). *Generalized Inverses: Theory and Applications.* New York: John Wiley.

Buerger, M. J. (1959). *Vector Space and its Application in Crystal Structure Investigation.* New York: John Wiley.

Burla, M. C., Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Siliqi, D. (2006). *J. Appl. Cryst.* **39**, 527–535.

Burla, M. C., Caliandro, R., Giacovazzo, C. & Polidori, G. (2010). *Acta Cryst.* A**66**, 347–361.

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2002). *Z. Kristallogr.* **217**, 629–635.

Burla, M. C., Giacovazzo, C. & Polidori, G. (2010). *J. Appl. Cryst.* **43**, 825–836.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Mazzone, A. & Siliqi, D. (2008). *J. Appl. Cryst.* **41**, 548–553.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2008). *Acta Cryst.* A**64**, 519–528.

Cascarano, G., Giacovazzo, C., Luić, M., Pifferi, G. & Spagna, R. (1987). *Z. Kristallogr.* **179**, 113–125.

Coppens, P. & Hamilton, W. C. (1968). *Acta Cryst.* B**24**, 925–929.

Garrido, J. (1950). *Comptes Rendus*, **231**, 297–298.

Giacovazzo, C. (1998). *Direct Phasing in Crystallography*. Oxford University Press.

Giacovazzo, C. & Mazzone, A. (2011). *Acta Cryst.* A**67**, 210–218.

Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.

Hahn, Th. (1992). Editor. *International Tables for Crystallography*, Vol. A. Dordrecht: Kluwer Academic Publishers.

Harker, D. (1936). *J. Chem. Phys.* **4**, 381–390.

Hirshfeld, F. L. (1968). *Acta Cryst.* A**24**, 301–311.

Impellizzeri, G., Pappalardo, G., D'Alessandro, F., Rizzarelli, E., Saviano, M., Iacovino, R., Benedetti, E. & Pedone, C. (2000). *Eur. J. Org. Chem.* **6**, 1065–1076.

Nordman, C. E. (1966). *Trans. Am. Crystallogr. Assoc*, **2**, 29–38.

Oszlányi, G. & Sütő, A. (2004). *Acta Cryst.* A**60**, 134–141.

Oszlányi, G. & Sütő, A. (2005). *Acta Cryst.* A**61**, 147–152.

Oszlányi, G. & Sütő, A. (2007). *Acta Cryst.* A**63**, 156–163.

Othen, E. (1990). Part II Thesis, Oxford University, England.

Palatinus, L. & Chapuis, G. (2007). *J. Appl. Cryst.* **40**, 786–790.

Pavelčík, F. (1988). *Acta Cryst.* A**44**, 724–729.

Pavelčík, F., Kuchta, L. & Sivý, J. (1992). *Acta Cryst.* A**48**, 791–796.

Rees, B. (1976). *Acta Cryst.* A**32**, 483–488.

Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.

Richardson, J. W. & Jacobson, R. A. (1987). *Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 310–317. Oxford University Press.

Sheldrick, G. M. (1982). *Testing Structures for Direct Methods*. University of Cambridge, England.

Sheldrick, G. M. (1992). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer.

Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.

Wrinch, D. M. (1939). *Philos. Mag.* **27**, 98–122.