

Theoretical Evaluation of the Overall Values of Discrepancy Indices for Truncated Data.

I. Normalized R indices for a Centrosymmetric Crystal with Similar Atoms

BY S. PARTHASARATHY AND D. VELMURUGAN

Department of Physics (Crystallography and Biophysics),* University of Madras, Guindy Campus,
Madras – 600025, India

(Received 2 June 1980; accepted 13 January 1981)

Abstract

Theoretical expressions for the local values of six types of normalized R indices applicable to truncated data are derived for a centrosymmetric crystal by taking the model structure to be of the imperfectly related incomplete type. These indices are tabulated as functions of σ_A and y_t . These results can be used to obtain the overall values of the R indices applicable to truncated data which is characterized by a given value of y_t .

1. Introduction

Values of different types of discrepancy indices (*i.e.* R indices) have been worked out in the literature for a number of crystallographic situations, such as crystals with similar atoms, crystals with heavy atoms, crystals with pseudo-symmetry in the atomic distribution, *etc.* (for details see Srinivasan & Parthasarathy, 1976, and the review of Parthasarathy, 1975). These results were obtained under the assumption that all the independent reflections in a given range of $(\sin \theta/\lambda)$ (*e.g.* Cu $K\alpha$ sphere) are available in the data. However, in actual crystals, the observed intensity of a certain percentage of reflections may be negative and many crystallographers truncate the data sets by omitting such reflections as well as those whose net counts are less than one to two times the standard deviation of this value (Stout & Jensen, 1968). Hence, it would be useful to evaluate theoretically the overall values of the various R indices applicable to such *truncated* data, wherein the reflections for which $y_N < y_t$ ($=\sqrt{z_t}$) are omitted.† In this series of papers we shall therefore derive the theoretical expressions of the various R

indices corresponding to different crystallographic situations and show how these results could be used to compute their theoretical *overall values*. Since the various indices behave differently under different situations (Parthasarathy & Parthasarathi, 1975; Parthasarathi & Parthasarathy, 1977; PP, 1977 hereafter), we shall deal with all the R indices in this series.

In this series we shall mostly follow the notations used in chapters 5 and 6 of Srinivasan & Parthasarathy (1976; hereafter SP, 1976). We shall use the abbreviations C and NC to denote the centrosymmetric and non-centrosymmetric cases respectively. We shall denote the joint probability density function (joint p.d.f. hereafter) of the normalized structure amplitudes y_N and y_P^c valid for the complete data by $P(y_N, y_P^c)$ and that valid for the truncated data by $P_t(y_N, y_P^c)$. For convenience we shall refer to $P_t(y_N, y_P^c)$ as the truncated distribution function. The expectation value of $f(y_N, y_P^c)$ for the truncated data will be denoted by $\langle f(y_N, y_P^c) \rangle_t$ and that for the complete data by $\langle f(y_N, y_P^c) \rangle$.

The experimental value of an R index is usually computed by taking all the independent observed reflections as a single group. We refer to the value of the R index thus obtained as the *overall value* of the R index. A number of factors, such as the positional errors in the atoms in the model, lack of similarity of the atoms in the unit cell, *etc.*, cause the value of an R index for a general model to vary with $\sin \theta/\lambda$. The theoretical evaluation of the overall value of an R index is therefore generally carried out in two stages; namely, (i) evaluation of the R index in different narrow regions of $\sin \theta/\lambda$ (within which the atomic scattering factors can be treated as constant) and (ii) evaluation of the overall value by taking care of the factors which change with $\sin \theta/\lambda$ (*e.g.* atomic scattering factors) and by giving proper weights depending on the number of observed reflections in the various ranges. We shall therefore refer to the value of an R index calculated using reflections in a narrow range of $\sin \theta/\lambda$ as a *local value* of the R index in order to distinguish it from its overall value. The overall value of an R index will be

* Contribution No. 545.

† For convenience, we shall refer to the data which consist of all the theoretically possible independent reflections within a given region of $(\sin \theta/\lambda)$ as the *complete data* and the data which exclude the reflections for which $y_N < y_t$ as the *truncated data*.

denoted with a bar over the corresponding symbol, while the local value of the index will be denoted without the bar.

In this paper we shall obtain, for a centrosymmetric crystal, the theoretical expressions for the local values of the six normalized R indices,* *viz.* $R_1(F)$, $R_1(I)$, ${}_B R_1(F)$, ${}_B R_1(I)$, $R_1^f(F)$ and $R_1^f(I)$ and show how these could be used to evaluate the overall values of these indices.

2. Derivation of the theoretical expressions for the overall values of the normalized R indices

2.1. Preliminary results

We shall presently explain the general notation and nomenclature regarding the overall values of different types of R indices computed for truncated data by taking the conventional normalized index $R_1(F)$ as an illustrative example. The corresponding results for the other indices can be similarly obtained.

The overall value of the conventional normalized index $R_1(F)$ applicable to the complete data is defined to be (SP, 1976)

$$\bar{R}_1(F) = \frac{\sum_{hkl} ||F_N| - |F_P^c|/\sigma_1|}{\sum_{hkl} |F_N|}, \quad (1)$$

where the summation is over *all* the theoretically possible independent reflections within a given ($\sin \theta/\lambda$) range† ($0 \leq s \leq S_{\max}$, say), $|F_P^c|$ is the magnitude of the calculated structure factor of the trial structure consisting of P out of N atoms in the unit cell and $|F_N|$ is the magnitude of the structure factor of the crystal structure. Here σ_1^2 is defined to be‡

$$\sigma_1^2 = \langle |F_P^c|^2 \rangle / \langle |F_N|^2 \rangle = \sigma_P^2 / \sigma_N^2. \quad (2)$$

For a structure with similar atoms

$$\sigma_1^2 = P/N, \quad (3)$$

which is a constant independent of $\sin \theta/\lambda$. The overall value of this index computed with the truncated data is

denoted by $[\bar{R}_1(F)]_t$ and is defined to be

$$[\bar{R}_1(F)]_t = \frac{\sum_{hkl}^t ||F_N| - |F_P^c|/\sigma_1|}{\sum_{hkl}^t |F_N|}, \quad (4)$$

where \sum_{hkl}^t means that the summation is over the independent reflections of the *truncated* data in the range 0 to S_{\max} . In our theoretical studies, it is convenient to use the normalized variables y_N and y_P^c defined by

$$y_N = |F_N|/\sigma_N, y_P^c = |F_P^c|/\sigma_P. \quad (5)$$

Making use of (5), we can rewrite (4) as

$$[\bar{R}_1(F)]_t = \frac{\sum_{hkl}^t \sigma_N |y_N - y_P^c|}{\sum_{hkl}^t \sigma_N y_N}. \quad (6)$$

If the reflections are divided into groups based on the values of s , we can rewrite (6) as

$$[\bar{R}_1(F)]_t = \frac{\sum_s \sum_r^t \sigma_{Nsr} |y_N - y_P^c|_{sr}}{\sum_s \sum_r^t \sigma_{Nsr} y_{Nsr}}, \quad (7)$$

where the summation over r is carried out over the reflections in a given range of $\sin \theta/\lambda$ and over s is carried out over the various ranges into which the interval 0 to S_{\max} has been partitioned. Here σ_{Nsr} denotes the value of σ_N for reflection r in the range s . If the $\sin \theta/\lambda$ ranges are sufficiently narrow, we can take the values of σ_{Nsr} for the different reflections in the range s to be the same (σ_{Ns} , say). We can therefore rewrite (7) as

$$[\bar{R}_1(F)]_t = \frac{\sum_s \sigma_{Ns} \sum_r^t |y_N - y_P^c|_{sr}}{\sum_s \sigma_{Ns} \sum_r^t y_{Nsr}}, \quad (8)$$

If n_s is the number of *observed* independent reflections in the range s in the truncated data we can write (8) as

$$[\bar{R}_1(F)]_t = \frac{\sum_s \sigma_{Ns} n_s [\langle |y_N - y_P^c| \rangle_s]_t}{\sum_s \sigma_{Ns} n_s [\langle y_N \rangle_s]_t}, \quad (9)$$

where $[\langle |y_N - y_P^c| \rangle_s]_t$ and $[\langle y_N \rangle_s]_t$ denote respectively the expectation values of $|y_N - y_P^c|$ and y_N for the observed reflections in the range s of the truncated data.

* We shall not consider the indices $R_1^*(F)$ and $R_1^*(I)$ since $R_1^f(F)$ and $R_1^f(I)$ are somewhat better than the former two (PP, 1977).

† We shall use s to stand for $\sin \theta/\lambda$. S_{\max} is the maximum value of s for the given data.

‡ We shall denote $1 - \sigma_1^2$ by σ_2^2 .

Table 1. Definition of the overall values of various normalized R indices for truncated data and related results

Number	Name of the R index	Notation for the overall value of the index	Overall value of the index		Definition of the local value of the unnormalized indices based on the normalized variables y_N and y_P^c
			Definition*	Equivalent expression	
1	Conventional index based on F	$[\bar{R}_1(F)]_t$	$\frac{\sum^t F_N - F_P^c /\sigma_1}{\sum^t F_N }$	$\frac{\sum_s f_s n_s [R_1(y)]_t}{\sum_s f_s n_s}$	$[R_1(y)]_t = \frac{\langle y_N - y_P^c \rangle_t}{\langle y_N \rangle_t}$
2	Conventional index based on I	$[\bar{R}_1(I)]_t$	$\frac{\sum^t I_N - I_P^c/\sigma_1^2 }{\sum^t I_N}$	$\frac{\sum_s f_s^2 n_s [R_1(z)]_t}{\sum_s f_s^2 n_s}$	$[R_1(z)]_t = \frac{\langle y_N^2 - (y_P^c)^2 \rangle_t}{\langle y_N^2 \rangle_t}$
3	Booth-type index based on F	$[_B\bar{R}_1(F)]_t$	$\frac{\sum^t (F_N - F_P^c /\sigma_1)^2}{\sum^t F_N ^2}$	$\frac{\sum_s f_s^2 n_s [_B R_1(y)]_t}{\sum_s f_s^2 n_s}$	$[_B R_1(y)]_t = \frac{\langle (y_N - y_P^c)^2 \rangle_t}{\langle y_N^2 \rangle_t}$
4	Booth-type index based on I	$[_B\bar{R}_1(I)]_t$	$\frac{\sum^t (I_N - I_P^c/\sigma_1^2)^2}{\sum^t I_N^2}$	$\frac{\sum_s f_s^4 n_s [_B R_1(z)]_t}{\sum_s f_s^4 n_s}$	$[_B R_1(z)]_t = \frac{\langle (y_N^2 - (y_P^c)^2)^2 \rangle_t}{\langle y_N^4 \rangle_t}$
5	Fractional-type index based on F	$[\bar{R}'_1(F)]_t$	$\frac{1}{N_0} \sum^t \left \frac{ F_N - F_P^c /\sigma_1}{(F_N + F_P^c /\sigma_1)/2} \right $	$\frac{\sum_s n_s [R'_1(y)]_t}{\sum_s n_s}$	$[R'_1(y)]_t = 2 \left\langle \left \frac{y_N - y_P^c}{y_N + y_P^c} \right \right\rangle_t$
6	Fractional-type index based on I	$[\bar{R}'_1(I)]_t$	$\frac{1}{N_0} \sum^t \left \frac{I_N - I_P^c/\sigma_1^2}{(I_N + I_P^c/\sigma_1^2)/2} \right $	$\frac{\sum_s n_s [R'_1(z)]_t}{\sum_s n_s}$	$[R'_1(z)]_t = 2 \left\langle \left \frac{y_N^2 - (y_P^c)^2}{y_N^2 + (y_P^c)^2} \right \right\rangle_t$

* The summation \sum^t is over the independent reflections of the truncated data. The summation \sum_s is over the different narrow ranges of $\sin \theta/\lambda_0$. N_0 is the number of reflections in the truncated data.

Since $[\langle y_N \rangle_s]_t$ is a property of the structure but not of the model and since for a structure with similar atoms† the p.d.f. of y_N is independent of s (Wilson, 1949), it follows that $[\langle y_N \rangle_s]_t$ will depend only on y_t . Hence we can write $[\langle y_N \rangle_s]_t$ simply as $[\langle y_N \rangle]_t$. We can therefore rewrite (9) as

$$[\bar{R}_1(F)]_t = \frac{\sum_s n_s \sigma_{Ns} [R_1(y)]_t}{\sum_s n_s \sigma_{Ns}}, \quad (10)$$

where $[R_1(y)]_t$ is defined to be

$$[R_1(y)]_t = \frac{[\langle |y_N - y_P^c| \rangle_s]_t}{[\langle y_N \rangle]_t}. \quad (11)$$

We refer to $[R_1(y)]_t$ as the local value of $R_1(y)$ calculated with the *observed* reflections in a particular range s . For equal-atom astructures, $\sigma_N = \sqrt{N}f$. We can therefore rewrite (9) as

$$[\bar{R}_1(F)]_t = \frac{\sum_s f_s n_s [R_1(y)]_t}{\sum_s f_s n_s}. \quad (12)$$

† For a structure with heavy atoms the p.d.f. of y_N involves the parameter σ_1^2 . Since σ_1^2 is a function of s , it follows that $[\langle y_N \rangle_s]_t$ could also depend on s .

It is clear from (12) that we can evaluate the overall value $[\bar{R}_1(F)]_t$ for the truncated data in $0 \leq s < S_{\max}$ if we know, corresponding to the various ranges of s , the values of the scattering factor of the atom, the number of independent reflections in the truncated data and the local values of $[R_1(y)]_t$. Of these, f_s and n_s for a given crystal are known quantities; we shall show in the rest of this paper how the local values $[R_1(y)]_t$ and those for the other normalized R indices can be obtained theoretically.

As stated earlier, the above considerations can be carried over to the other five normalized R indices. The final expressions which are the analogues of (12) are hence listed in Table 1 without derivation.

2.2. Derivation of the joint p.d.f. of y_N and y_P^c for the truncated data

It is seen from columns 5 and 6 of Table 1 that the overall values of the various normalized R indices can be calculated if we obtain the expectation values of simple functions of y_N and y_P^c . Thus the overall values of these R indices can, in principle, be evaluated if we know the joint p.d.f. of y_N and y_P^c for the truncated data, i.e. the function $P_t(y_N, y_P^c)$. It is clear that the function

$P_t(y_N, y_P^c)$ is related to the function $P(y_N, y_P^c)$ by

$$P_t(y_N, y_P^c) = \frac{1}{\beta} P(y_N, y_P^c), \quad y_t \leq y_N < \infty \quad (13)$$

$$0 \leq y_P^c < \infty,$$

where β^* is given by

$$\beta = \int_{y_t}^{\infty} \int_0^{\infty} P(y_N, y_P^c) dy_N dy_P^c. \quad (14)$$

The function $P(y_N, y_P^c)$ for the C case is known to be (SP, 1976)

$$P(y_N, y_P^c) = \frac{2}{\pi \sigma_B} \exp \left\{ -\frac{[y_N^2 + (y_P^c)^2]}{2\sigma_B^2} \right\}$$

$$\times \cosh \left[\frac{\sigma_A y_N y_P^c}{\sigma_B^2} \right],$$

$$0 \leq y_N < \infty, \quad 0 \leq y_P^c < \infty, \quad (15)$$

where

$$\sigma_A^2 + \sigma_B^2 = 1, \quad \sigma_A = \sigma_1 D \quad (16)$$

$$D = \exp \left(-\frac{\pi^3}{4} |H|^2 \langle |\Delta r|^2 \rangle \right), \quad |H| = 2s. \quad (17)$$

Making use of (15) and (16), we have shown {in the Appendix† [see (A-10)]} that β_C is given by

$$\beta_C = \operatorname{erfc} (y_t/\sqrt{2}). \quad (18)$$

From (13) and (15), we obtain

$$P_t(y_N, y_P^c) = \frac{2}{\beta_C \pi \sigma_B} \exp \left\{ -\frac{[y_N^2 + (y_P^c)^2]}{2\sigma_B^2} \right\}$$

$$\times \cosh \left[\frac{\sigma_A y_N y_P^c}{\sigma_B^2} \right],$$

$$y_t \leq y_N < \infty, \quad 0 \leq y_P^c < \infty. \quad (19)$$

2.3. Derivation of the theoretical expressions for the local values of normalized R indices applicable to truncated data

Index $[R_1(y)]_t$. From the definition of $[R_1(y)]_t$ in Table 1, we have

$$[R_1(y)]_t = \frac{1}{\langle y_N \rangle_t} \int_{y_t}^{\infty} \int_0^{\infty} |y_N - y_P^c| P_t(y_N, y_P^c) dy_N dy_P^c, \quad (20)$$

* Since the function $P(y_N, y_P^c)$ for the C and NC cases is different, β will be different for the two cases. Hence, the quantity appropriate to the C and NC cases will be denoted by β_C and β_{NC} respectively.

† This appendix has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 35848 (7 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

where $P_t(y_N, y_P^c)$ is given by (19). Making use of (13), we can rewrite (20) as

$$[R_1(y)]_t = \frac{\langle y_N \rangle}{\beta_C \langle y_N \rangle_t} \left[\frac{1}{\langle y_N \rangle} \int_0^{\infty} \int_0^{\infty} |y_N - y_P^c| P(y_N, y_P^c) \right.$$

$$\times dy_N dy_P^c - \frac{1}{\langle y_N \rangle} \int_0^{y_t} \int_0^{\infty} |y_N - y_P^c|$$

$$\times P(y_N, y_P^c) dy_N dy_P^c \left. \right]. \quad (21)$$

The first term within the square brackets in (21) represents $\langle |y_N - y_P^c| \rangle / \langle y_N \rangle$, which by definition is $R_1(y)$. It has been shown that (SP, 1976)

$$R_1(y) = [2(1 + \sigma_A)]^{1/2} + [2(1 - \sigma_A)]^{1/2} - 2. \quad (22)$$

Making use of the known result that $\langle y_N \rangle = \sqrt{2/\pi}$ (Wilson, 1949) and the expressions for β_C [see (18)] and $\langle y_N \rangle_t$ [see (A-11)], we can rewrite (21) as

$$[R_1(y)]_t = \exp (y_t^2/2) \left[R_1(y) - \sqrt{\frac{\pi}{2}} \int_0^{y_t} \int_0^{\infty} |y_N - y_P^c| \right.$$

$$\times P(y_N, y_P^c) dy_N dy_P^c \left. \right]. \quad (23)$$

It is convenient to change the variables of integration to u and v where

$$u = y_N/(1 + y_N) \quad \text{and} \quad v = y_P^c/(1 + y_P^c). \quad (24)$$

Thus, we obtain

$$[R_1(y)]_t = \exp (y_t^2/2) \left[R_1(y) \right.$$

$$- \sqrt{\frac{\pi}{2}} \int_0^{y_t/(1+y_t)} \int_0^1 \left| \frac{u}{1-u} - \frac{v}{1-v} \right|$$

$$\times P \left(\frac{u}{1-u}, \frac{v}{1-v} \right)$$

$$\times \frac{du}{(1-u)^2} \frac{dv}{(1-v)^2} \left. \right]. \quad (25)$$

Here, $P[u/(1-u), v/(1-v)]$ means that in expression (15) for $P(y_N, y_P^c)$, y_N is to be replaced by $u/(1-u)$ and y_P^c by $v/(1-v)$.

Index $[R_1(z)]_t$. From the definition of $[R_1(z)]_t$ in Table 1 we have

$$[R_1(z)]_t = \frac{1}{\langle y_N^2 \rangle_t} \int_{y_t}^{\infty} \int_0^{\infty} |y_N^2 - (y_P^c)^2| P_t(y_N, y_P^c) dy_N dy_P^c. \quad (26)$$

Following the method used for $[R_1(y)]_t$ and using (A-13) and the definition (see SP, 1976)

$$R_1(z) = \langle |z_N - z_P^c| \rangle = \langle |y_N^2 - (y_P^c)^2| \rangle,$$

we can rewrite (26) as

$$[R_1(z)]_t = \left[R_1(z) - \int_0^{y_t} \int_0^\infty |y_N^2 - (y_P^c)^2| \right. \\ \left. \times P(y_N, y_P^c) dy_N dy_P^c \right] \\ \times [\beta_C + \sqrt{(2/\pi)} y_t \exp(-y_t^2/2)]^{-1}. \quad (27)$$

It is known that (SP, 1976)

$$R_1(z) = \frac{4}{\pi} \sigma_B. \quad (28)$$

Making use of the substitution (24), we can rewrite (27) as

$$[R_1(z)]_t = \left[R_1(z) - \int_0^{y_t/(1+y_t)} \int_0^1 \left| \left(\frac{u}{1-u} \right)^2 \right. \right. \\ \left. \left. - \left(\frac{v}{1-v} \right)^2 \right| P\left(\frac{u}{1-u}, \frac{v}{1-v} \right) \right. \\ \left. \times \frac{du}{(1-u)^2} \frac{dv}{(1-v)^2} \right] \\ \times [\beta_C + \sqrt{(2/\pi)} y_t \exp(-y_t^2/2)]^{-1}. \quad (29)$$

Index $[{}_B R_1(y)]_t$. From the definition of $[{}_B R_1(y)]_t$ given in Table 1 we can obtain

$$[{}_B R_1(y)]_t = \frac{\langle y_N^2 \rangle_t + \langle (y_P^c)^2 \rangle_t - 2 \langle y_N y_P^c \rangle_t}{\langle y_N^2 \rangle_t}. \quad (30)$$

Substituting for $\langle y_N^2 \rangle_t$, $\langle (y_P^c)^2 \rangle_t$ and $\langle y_N y_P^c \rangle_t$ from (A-13), (A-18) and (A-23) respectively and simplifying the result with (16), we obtain

$$[{}_B R_1(y)]_t = \left[{}_B R_1(y) + 2(\beta_C - 1) \right. \\ \left. + \sqrt{(2/\pi)} y_t (1 + \sigma_A^2) \exp(-y_t^2/2) \right. \\ \left. - \frac{4}{\pi} \left\{ \sigma_B^3 [\exp(-y_t^2/2\sigma_B^2) - 1] \right. \right. \\ \left. \left. - \sigma_A \sqrt{\pi} \int_0^{y_t^2/2} \sqrt{u} \exp(-u) \right. \right. \\ \left. \left. \times \operatorname{erf}\left(\frac{\sigma_A}{\sigma_B} \sqrt{u} \right) du \right\} \right] \\ \times [\beta_C + \sqrt{(2/\pi)} y_t \exp(-y_t^2/2)]^{-1}. \quad (31)$$

${}_B R_1(y)$ in (31) is given by (SP, 1976)

$${}_B R_1(y) = \langle (y_N - y_P^c)^2 \rangle / \langle y_N^2 \rangle = \langle (y_N - y_P^c)^2 \rangle \\ = 2 - \frac{4}{\pi} [\sigma_B + \sigma_A \sin^{-1}(\sigma_A)]. \quad (32)$$

Index $[{}_B R_1(z)]_t$. From the definition of $[{}_B R_1(z)]_t$ given in Table 1 we can write

$$[{}_B R_1(z)]_t = \frac{\langle y_N^4 \rangle_t + \langle (y_P^c)^4 \rangle_t - 2 \langle y_N^2 (y_P^c)^2 \rangle_t}{\langle y_N^4 \rangle_t}. \quad (33)$$

Substituting for $\langle y_N^4 \rangle_t$, $\langle (y_P^c)^4 \rangle_t$ and $\langle y_N^2 (y_P^c)^2 \rangle_t$ from (A-15), (A-22) and (A-24) respectively and simplifying, we obtain

$$[{}_B R_1(z)]_t = \{ {}_B R_1(z) + \sqrt{(2/\pi)} (1 - \sigma_A^2) (y_t/3\beta_C) \\ \times \exp(-y_t^2/2) [(1 + y_t^2) \\ + \sigma_A^2 (3 - y_t^2)] \} [1 + \sqrt{(2/\pi)} (y_t/3\beta_C) \\ \times (3 + y_t^2) \exp(-y_t^2/2)]^{-1}. \quad (34)$$

In obtaining (34) we have used (16) as well as the result (SP, 1976)

$${}_B R_1(z) = \frac{4}{3} \sigma_B^2. \quad (35)$$

Index $[R_1^f(y)]_t$. From the definition of $[R_1^f(y)]_t$ given in Table 1, we can write

$$[R_1^f(y)]_t = 2 \int_0^{y_t} \int_0^\infty \left| \frac{y_N - y_P^c}{y_N + y_P^c} \right| P(y_N, y_P^c) dy_N dy_P^c. \quad (36)$$

With (13) we can rewrite (36) as

$$[R_1^f(y)]_t = \frac{1}{\beta_C} \left[2 \int_0^{y_t} \int_0^\infty \left| \frac{y_N - y_P^c}{y_N + y_P^c} \right| P(y_N, y_P^c) dy_N dy_P^c \right. \\ \left. - 2 \int_0^{y_t} \int_0^\infty \left| \frac{y_N - y_P^c}{y_N + y_P^c} \right| \right. \\ \left. \times P(y_N, y_P^c) dy_N dy_P^c \right]. \quad (37)$$

The first term within the square brackets represents $R_1^f(y)$ (see PP, 1977). Making use of the transformation (24) in the integrals occurring as the second term on the right-hand side of (37), we obtain

$$[R_1^f(y)]_t = \frac{1}{\beta_C} \left[R_1^f(y) - 2 \int_0^{y_t/(1+y_t)} \int_0^1 \left| \left(\frac{u}{1-u} \right. \right. \right. \\ \left. \left. - \frac{v}{1-v} \right) \right| \left(\frac{u}{1-u} + \frac{v}{1-v} \right) \right. \\ \left. \times P\left(\frac{u}{1-u}, \frac{v}{1-v} \right) \right. \\ \left. \times \frac{du}{(1-u)^2} \frac{dv}{(1-v)^2} \right]. \quad (38)$$

where $R_1^f(z)$ is given by (PP, 1977)

$$R_1^f(z) = \frac{2\sigma_B}{\pi\sigma_A} \ln \left(\frac{1 + \sigma_A}{1 - \sigma_A} \right). \quad (42)$$

3. Discussion of the theoretical results

The theoretical expressions for the six normalized R indices valid for truncated data have been derived in (25), (29), (31), (34), (38) and (41). A study of these equations shows that we can, in general, write the R indices applicable to the truncated data in the general form as

$$[R]_t = A(y_t) [R + B(\sigma_A, y_t)], \quad (43)$$

where R stands for any one of the normalized R indices, $A(y_t)$ is a function of y_t and $B(\sigma_A, y_t)$ is a function of σ_A and y_t . If $y_t \rightarrow 0$, $A(y_t) \rightarrow 1$ and $B(\sigma_A, y_t) \rightarrow 0$ so that $[R]_t \rightarrow R$ as required.

A study of (25), (29), (31), (34), (38) and (41) shows that these R indices depend on the parameters y_t (which is a characteristic of the data) and σ_A . σ_A is in turn a function of σ_1^2 (i.e. the fractional contribution to the local mean intensity from the known atoms), $\langle |\Delta r| \rangle$ (i.e. the mean positional error in the atoms of the model) and $\sin \theta/\lambda$. Putting $\sigma_A = \sigma_1$ (i.e. $\langle |\Delta r| \rangle = 0$) in these equations, we can obtain the results for the related case. Similarly, putting $\sigma_A = 0$ (i.e. $\langle |\Delta r| \rangle$ large), we can obtain the results for the unrelated case. The local values of these R indices (in percent) are given in Tables 2-4 as functions of σ_A for different fixed values of y_t ,

Table 3. Values of $[{}_B R_1(y)]_t$ and $[{}_B R_1(z)]_t$ as functions of σ_A and y_t for the centrosymmetric case

σ_A ↓	$[{}_B R_1(y)]_t$					$[{}_B R_1(z)]_t$				
	$y_t \rightarrow 0.00$	0.15	0.30	0.45	0.60	$y_t \rightarrow 0.00$	0.15	0.30	0.45	0.60
0.000	72.676	62.143	54.377	49.045	45.697	133.333	121.469	110.218	100.125	91.595
0.040	72.574	62.059	54.307	48.985	45.643	133.120	121.294	110.077	100.004	91.523
0.080	72.268	61.806	54.097	48.804	45.479	132.480	120.767	109.654	99.677	91.233
0.120	71.758	61.385	53.746	48.502	45.205	131.413	119.887	108.946	99.113	90.777
0.160	71.043	60.794	53.253	48.078	44.820	129.920	118.653	107.951	98.318	90.131
0.200	70.121	60.031	52.617	47.530	44.322	128.000	117.062	106.662	97.283	89.287
0.240	68.991	59.096	51.837	46.857	43.709	122.853	112.498	103.075	96.002	88.022
0.280	67.652	57.986	50.909	46.055	42.978	122.680	112.498	103.075	95.464	86.966
0.320	66.100	56.700	49.831	45.123	42.126	119.680	110.110	100.972	92.656	85.462
0.360	64.333	55.233	48.601	44.055	41.149	116.053	107.052	98.438	89.176	83.705
0.400	62.347	53.583	47.214	42.849	40.042	112.000	103.613	95.567	88.166	81.688
0.440	60.140	51.688	45.660	41.499	38.800	109.813	101.749	94.001	86.863	80.569
0.460	59.951	50.754	44.830	40.768	38.000	107.520	99.787	92.346	85.467	79.376
0.480	57.705	49.715	43.951	39.968	37.416	105.120	97.726	90.600	83.991	78.155
0.500	56.401	48.626	43.029	39.190	36.668	102.613	95.565	88.760	82.427	76.754
0.520	55.038	47.487	42.063	38.341	35.882	100.000	93.304	86.865	80.773	75.318
0.540	53.616	46.297	41.053	37.451	35.056	97.280	90.941	84.775	78.959	73.715
0.560	52.132	45.054	39.996	36.518	34.189	94.453	88.475	82.664	77.155	72.181
0.580	50.588	43.759	38.892	35.542	33.280	91.520	85.905	80.432	75.245	70.473
0.600	48.981	42.410	37.740	34.520	32.328	88.480	83.229	78.097	73.204	68.666
0.610	48.854	42.314	37.615	34.422	32.238	88.333	83.120	78.015	73.104	68.565
0.620	48.731	42.218	37.490	34.328	32.149	88.180	82.980	77.870	72.987	68.464
0.630	48.608	42.122	37.365	34.232	32.060	88.027	82.833	77.720	72.850	68.363
0.640	48.485	42.026	37.240	34.136	31.971	87.873	82.680	77.567	72.704	68.262
0.650	48.362	41.930	37.115	34.040	31.882	87.719	82.523	77.410	72.558	68.161
0.660	48.239	41.834	37.000	33.944	31.793	87.565	82.367	77.253	72.411	68.060
0.670	48.116	41.738	36.875	33.848	31.704	87.410	82.210	77.096	72.264	67.959
0.680	47.993	41.642	36.750	33.752	31.615	87.256	82.053	76.939	72.117	67.858
0.690	47.870	41.546	36.625	33.656	31.526	87.101	81.896	76.782	71.970	67.757
0.700	47.747	41.450	36.500	33.560	31.437	86.947	81.739	76.625	71.823	67.656
0.710	47.624	41.354	36.375	33.464	31.348	86.792	81.582	76.468	71.676	67.555
0.720	47.501	41.258	36.250	33.368	31.259	86.638	81.425	76.311	71.529	67.454
0.730	47.378	41.162	36.125	33.272	31.170	86.483	81.268	76.154	71.382	67.353
0.740	47.255	41.066	36.000	33.176	31.081	86.328	81.111	76.000	71.235	67.252
0.750	47.132	40.970	35.875	33.080	31.000	86.173	80.954	75.843	71.088	67.151
0.760	47.009	40.874	35.750	32.984	30.919	86.018	80.797	75.686	70.941	67.050
0.770	46.886	40.778	35.625	32.888	30.838	85.863	80.640	75.529	70.794	66.949
0.780	46.763	40.682	35.500	32.792	30.757	85.708	80.483	75.372	70.647	66.848
0.790	46.640	40.586	35.375	32.696	30.676	85.553	80.326	75.215	70.500	66.747
0.800	46.517	40.490	35.250	32.600	30.595	85.398	80.169	75.058	70.353	66.646
0.810	46.394	40.394	35.125	32.504	30.514	85.243	79.999	74.901	70.206	66.545
0.820	46.271	40.298	35.000	32.408	30.433	85.088	79.842	74.744	70.059	66.444
0.830	46.148	40.202	34.875	32.312	30.352	84.933	79.685	74.587	69.912	66.343
0.840	46.025	40.106	34.750	32.216	30.271	84.778	79.528	74.430	69.765	66.242
0.850	45.902	40.010	34.625	32.120	30.190	84.623	79.371	74.273	69.618	66.141
0.860	45.779	39.914	34.500	32.024	30.109	84.468	79.214	74.116	69.471	66.040
0.870	45.656	39.818	34.375	31.928	30.028	84.313	79.057	73.959	69.324	65.939
0.880	45.533	39.722	34.250	31.832	29.947	84.158	78.900	73.802	69.177	65.838
0.890	45.410	39.626	34.125	31.736	29.866	84.003	78.743	73.645	69.030	65.737
0.900	45.287	39.530	34.000	31.640	29.785	83.848	78.586	73.488	68.883	65.636
0.905	45.242	39.500	33.975	31.625	29.770	83.822	78.571	73.473	68.878	65.631
0.910	45.197	39.470	33.950	31.610	29.755	83.797	78.556	73.458	68.873	65.626
0.915	45.152	39.440	33.925	31.595	29.740	83.772	78.541	73.443	68.868	65.621
0.920	45.107	39.410	33.900	31.580	29.725	83.747	78.526	73.428	68.863	65.616
0.925	45.062	39.380	33.875	31.565	29.710	83.722	78.511	73.413	68.858	65.611
0.930	45.017	39.350	33.850	31.550	29.695	83.697	78.496	73.398	68.853	65.606
0.935	44.972	39.320	33.825	31.535	29.680	83.672	78.481	73.383	68.848	65.601
0.940	44.927	39.290	33.800	31.520	29.665	83.647	78.466	73.368	68.843	65.596
0.945	44.882	39.260	33.775	31.505	29.650	83.622	78.451	73.353	68.838	65.591
0.950	44.837	39.230	33.750	31.490	29.635	83.597	78.436	73.338	68.833	65.586
0.955	44.792	39.200	33.725	31.475	29.620	83.572	78.421	73.323	68.828	65.581
0.960	44.747	39.170	33.700	31.460	29.605	83.547	78.406	73.308	68.823	65.576
0.965	44.702	39.140	33.675	31.445	29.590	83.522	78.391	73.293	68.818	65.571
0.970	44.657	39.110	33.650	31.430	29.575	83.497	78.376	73.278	68.813	65.566
0.975	44.612	39.080	33.625	31.415	29.560	83.472	78.361	73.263	68.808	65.561
0.980	44.567	39.050	33.600	31.400	29.545	83.447	78.346	73.248	68.803	65.556
0.985	44.522	39.020	33.575	31.385	29.530	83.422	78.331	73.233	68.798	65.551
0.990	44.477	38.990	33.550	31.370	29.515	83.397	78.316	73.218	68.793	65.546
0.995	44.432	38.960	33.525	31.355	29.500	83.372	78.301	73.203	68.788	65.541
1.000	44.387	38.930	33.500	31.340	29.485	83.347	78.286	73.188	68.783	65.536

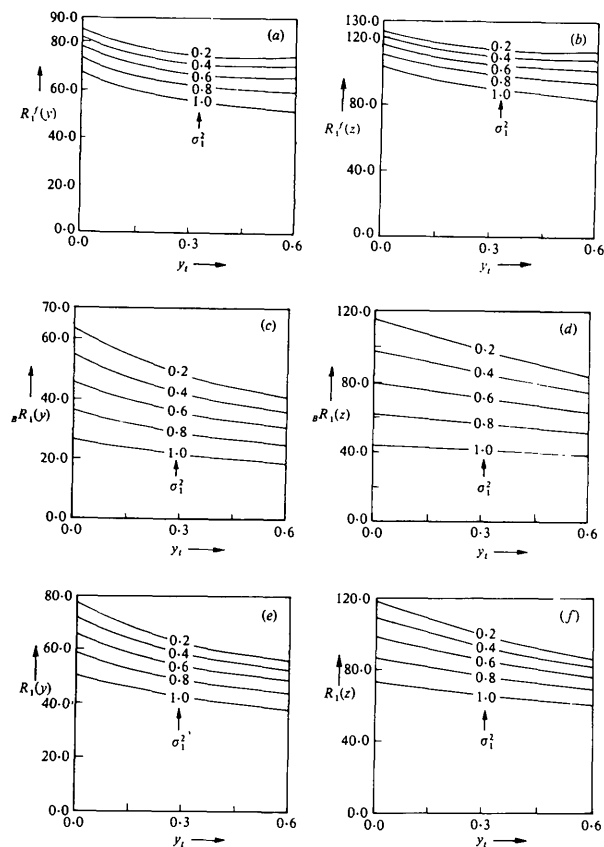


Fig. 1. Variation of the local values of normalized R indices as a function of y_i for different fixed values of σ_i^2 . For the definition of $R_1(y)$, $R_1(z)$, etc., see column 6 of Table 1.

P/N . (iii) Divide the range $0 \leq s \leq S_{\max}$ into a number of narrow subintervals. Determine from the data the number (n_s) of observed reflections in each of these sub-intervals. (iv) From the mean value of $\sin \theta/\lambda$ corresponding to each sub-interval determine the values of f . (v) Calculate σ_A (by taking $\langle |\Delta r| \rangle = 0.2 \text{ \AA}$) corresponding to each of these sub-intervals. (vi) Make use of the results in Tables 2–4 and determine the values of $[R]_i$ for these sub-intervals by bilinear interpolation. (vii) Make use of the results thus obtained in the appropriate expressions for $[\bar{R}]_i$ and compute the overall values $[\bar{R}]_i$. The values thus obtained represent the theoretical overall values of the normalized R indices corresponding to a model for which $\langle |\Delta r| \rangle = 0.2 \text{ \AA}$ and for data in which $y_N \geq y_i$ and $0 \leq s \leq S_{\max}$.

References

- PARTHASARATHI, V. & PARTHASARATHY, S. (1977). *Pramana*, **9**, 61–69.
 PARTHASARATHY, S. (1975). *J. Indian Inst. Sci.* **57**, 127–159.
 PARTHASARATHY, S. & PARTHASARATHI, V. (1975). *Acta Cryst.* **A31**, 178–185.
 PONNUSWAMY, M. N. & PARTHASARATHY, S. (1977). *Acta Cryst.* **A33**, 838–844.
 SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
 STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination*, p. 182. New York: Macmillan.
 WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–320.

Acta Cryst. (1981). **A37**, 480–485

Theoretical Evaluation of the Overall Values of Discrepancy Indices for Truncated Data. II. Normalized R indices for a Non-centrosymmetric Crystal with Similar Atoms

BY D. VELMURUGAN AND S. PARTHASARATHY

Department of Physics (Crystallography and Biophysics), University of Madras, Guindy Campus, Madras-600025, India*

(Received 2 June 1980; accepted 13 January 1981)

Abstract

Theoretical expressions and numerical tables for the local values of six types of normalized R indices are obtained for an imperfectly related incomplete model of

a non-centrosymmetric crystal with truncated data. Under similar conditions, the curve of the local value of an R index *versus* the truncation limit y_i is relatively more flat for the non-centrosymmetric case than for the centrosymmetric case particularly in the region where y_i is small.

* Contribution No. 546.