Acta Cryst. (1995). A**51**, 112–121

A Systematic Study of Coordinate Precision in X-ray Structure Analyses. II. Predictive Estimates of E.S.D.'s for the General-Atom Case

BY FRANK H. ALLEN*

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England

AND JASON C. COLE AND JUDITH A. K. HOWARD*

Department of Chemistry, University of Durham, South Road, Durham DH1 3LE, England

(*Received* 10 *March* 1994; *accepted* 29 *June* 1994)

Abstract

The relationship between the mean isotropic e.s.d. $\overline{\sigma}(A)_o$ of any element type A in a crystal structure and the R factor and atomic constitution of that structure is explored for 124905 element-type occurrences calculated from 33955 entries in the Cambridge Structural Database. On the basis of the work of Cruickshank [*Acta Cryst.* (1960), **13**, 774–777], it is shown that $\overline{\sigma}(A)_p$ values can be estimated by equations of the form $\overline{\sigma}(A)_p = KRN_c^{1/2}/Z_A$ where N_c is taken as $\sum Z_i^2/Z_C^2$, the Z_i are atomic numbers and the summation is over all atoms in the asymmetric unit. Values of K were obtained by regression techniques using the $\overline{\sigma}(A)_o$ as basis. The constant K_{nc} for noncentrosymmetric structures is found to be larger than K_c for centrosymmetric structures by a factor of $\sim 2^{1/2}$, as predicted by Cruickshank (1960). Two predictive equations are generated, one for first-row elements and the second for elements with $Z_A > 10$. The relationship between the different constants K that arise in these two situations is linked to shape differentials in scattering-factor (f_i) curves for light and heavy atoms. It is found that predictive equations in which the Z_i are selectively replaced by f_i at a constant $\sin\theta/\lambda$ of 0.30 Å^{-1} generate closely similar values of K for the light-atom and heavy-atom subsets. The overall analysis indicates that atomic e.s.d.'s may be seriously underestimated in the more precise structure determinations, that e.s.d.'s for the heaviest atoms may be less reliable than those for lighter atoms and that e.s.d.'s in noncentrosymmetric structures may be less accurate than those in centrosymmetric structures.

Introduction

In paper I (Allen, Cole & Howard, 1995), we presented descriptive statistics concerning the precision of structures stored in the Cambridge Structural Database (CSD; Allen *et al.*, 1991). We also used correlation and linear-regression techniques to analyse the relationship between the mean isotropic e.s.d.'s of C atoms, $\overline{\sigma}$(C), for more than 20000 X-ray structures and the R factors and atomic constitutions of those structures. It was found that an equation of the form

$$\overline{\sigma}(\mathrm{C})_p = kRN_c^{1/2}/e \qquad (1)$$

($e = 1$ for noncentrosymmetric structures and 1.399 for centrosymmetric structures), where

$$N_c = \sum_{i=1}^{N_{nh}} Z_i^2/Z_C^2, \qquad (2)$$

in which the Z_i are atomic numbers, $Z_C = 6$ and N_{nh} is the number of non-H atoms in the asymmetric unit, was able to generate reasonable estimates of $\overline{\sigma}$(C) across a very broad range of R factors and atomic constitutions.

The rationale for this work lies in the relative lack of indicators of structural precision stored in the CSD for ~40000 structures published prior to 1984. Since that time, individual e.s.d.'s for atomic coordinates are incorporated in the database, together with the R factor and a flag, denoted AS (*Cambridge Structural Database User's Manual*, 1992), which records the mean e.s.d. of a C–C bond [$\overline{\sigma}$(C–C)] in banded ranges: AS = 1, 2, 3 and 4 for $\overline{\sigma}$(C–C) ≤ 0.005, 0.005–0.010, 0.010–0.030 and > 0.030 Å. For pre-1984 entries, only R and AS are recorded in the CSD and, regrettably, AS is unavailable for ~16% of those entries. Hence, the work reported here and in paper I represents an attempt to provide estimates of coordinate e.s.d.'s that are based on information that *is* present in all CSD entries. The earlier study concentrated almost entirely on the precision of C-atom coordinates as indicated by (1). In this paper, we extend the analysis to provide estimates of $\overline{\sigma}(A)$, the mean isotropic e.s.d. of a general atom A.

* Authors for correspondence.

ISSN 0108-7673

Factors affecting structural precision

The theoretical background to this work is provided by Cruickshank (1960), who analysed the precision of X-ray intensity data that is required to yield a mean isotropic coordinate e.s.d., $\overline{\sigma}(A)$, for an element A in a structure that may contain other elements. He derived the simple approximation

$$\overline{\sigma}(A) \simeq RN_A^{1/2}/\overline{s}(mp)^{1/2}, \quad (3)$$

where $\overline{s}$ is the r.m.s. reciprocal radius for the reflections observed, p is $(N_r - N_p)$, where N_r is the number of independent reflections and N_p is the number of refined parameters, and m is 4 for noncentrosymmetric space groups or 8 for centrosymmetric space groups. The parameter N_A is the number of atoms similar to A that are required to give a scattering power at $\overline{s}$ that is equivalent to the N (non-H) atoms of the asymmetric unit, *i.e.*

$$N_A = \sum_{i=1}^{N_{nh}} f_i^2/f_A^2, \quad (4)$$

where the f_i are scattering factors at $\overline{s}$.

However, $\overline{s}$, N_r and N_p are not stored for CSD entries, nor are they calculable from information stored in the database. Hence, in our earlier work (paper I), we were forced to assume constancy of $\overline{s}$ and p [(3)] across all database entries and to approximate the calculation of N_A by use of atomic numbers (Z_i) as shown for the carbon case in (2). Despite these approximations, it was possible to predict values of $\overline{\sigma}(C)_p$ *via* (1) for which more than 75% of the $\overline{\sigma}(C)_p$ were within 0.0025 Å of the 'observed' values, $\overline{\sigma}(C)_o$, obtained from the published coordinate e.s.d.'s. The effectiveness of a four-parameter equation (3) was investigated in paper I using a small (~550 entry) data set for which values of N_r and N_p were manually abstracted from the literature and values of $\overline{s}$ were then estimated by a method due to Cruickshank (1993) reported in paper I. Although estimates of $\overline{\sigma}(C)$ derived from a more complete version of (3) were improved slightly, the results were somewhat disappointing. Possible reasons for these results are discussed in detail in paper I and are not repeated here.

In paper I, we also noted that, since R, $\overline{s}$, m and p in (3) are constants for a given structure, we may manipulate (3) and our approximation of N_A embodied in (2) so that, for two atom types A and B, we obtain

$$\overline{\sigma}(A)_p/\overline{\sigma}(B)_p = Z_B/Z_A. \quad (5)$$

If B is carbon, then we obtain

$$\overline{\sigma}(A)_p = 6\overline{\sigma}(C)_p/Z_A \quad (6)$$

and, by substitution from (1),

$$\overline{\sigma}(A)_p = 6kRN_c^{1/2}/eZ_A, \quad (7)$$

i.e. an equation that relates $\overline{\sigma}(A)$ for any atom to quantities that are readily available in CSD entries. Equation (1) is, of course, a special case of (7) for which $Z_A = 6$.

A preliminary experiment (paper I) used regressions based on (7) to estimate $K = 6k$ and obtain a predictive equation for $\overline{\sigma}(A)$. However, this preliminary work was applied to an atom type A only when it was the *heaviest* element in a given structure. In the present work, we explore the applicability of (7) to *all* atom types in CSD structures for which published ('observed') mean isotropic e.s.d.'s $\overline{\sigma}(A)_o$ are available in the database.

Terminology

Throughout this paper, the mean values of parameters that can vary within a single structure are denoted, $\overline{\sigma}(A)$, $\overline{\sigma}(C)$ *etc.* Mean values of parameters taken over many structures are denoted $\langle\overline{\sigma}(A)\rangle$, $\langle R\rangle$, $\langle N_c^{1/2}\rangle$ *etc.*

Methodology

The April 1993 release of Version 5 of the CSD was processed using the search program *QUEST3D* (*Cambridge Structural Database User's Manual*, 1992) and by local code to generate a formatted work file that was used in subsequent data analyses. Initial data retrieval (*QUEST3D*) was restricted to entries for which atomic coordinate e.s.d.'s were available in the CSD and which: (*a*) were determined from X-ray data collected on a diffractometer; (*b*) had $R \le 0.100$; (*c*) were error free after CSD check procedures; and (*d*) contained no disorder or polymeric (catena) bonding.

Local code was then written to convert the binary CSD records for retrieved entries to a simple formatted work file containing the information items defined in Table 1. Methods used to calculate the mean isotropic e.s.d.'s $\overline{\sigma}(A)$ for each (non-H) element type in each structure were identical to those employed in paper I. These calculations took account of the recent cautionary note of Muir & Mallinson (1993) concerning the hazards of calculating $\overline{\sigma}(A)$ in oblique coordinate systems using simple analytical equations. As before, we used the analysis of Templeton (1959) to approximate the coordinate correlations that exist in these systems.

As in paper I, a number of checks were included in the code for work-file generation in an attempt (*a*) to improve the validity of statistical analyses of the work file and (*b*) to guard against possible numerical errors in the coordinate e.s.d.'s stored in the CSD. The following checks were employed for this study:

(1) The distribution of the n_A individual $\sigma(A)$ values for each element A in each CSD entry was

Table 1. *Principal information fields in the work file*

CSD entry information record (one per entry)

REFCOD	CSD reference code
R	Crystallographic R factor
N	Number of non-H atoms
SPGN	Space-group number
CENT	Noncentrosymmetric = 1, centrosymmetric = 2
V	Unit-cell volume
Z_{max}	Atomic number of heaviest element
$N_o^{1/2}$	As defined by (2)
ITYPE	Number of element type records that follow for this entry

Element-type records (ITYPE per entry)

A	Element symbol
Z	Atomic number
$\overline{\sigma}(A)_o$	Mean isotropic e.s.d. (Å) for this element type
N_σ	Number of occurrences of element A contributing to the mean e.s.d. value

examined by calculating the sample standard deviation:

$$S = \left\{ \sum_{i=1}^{n_A} [\overline{\sigma}(A) - \sigma(A)]^2/(n_A - 1) \right\}^{1/2}. \quad (8)$$

Any $\sigma(A)$ value for which $|\overline{\sigma}(A) - \sigma(A)| > 4S$ was omitted from a final calculation of $\overline{\sigma}(A)_o$ for inclusion in the work file.

(2) The skewness of the $\sigma(A)$ distribution in each CSD entry was examined by calculating

$$M = [\sigma(A)_{max} + \sigma(A)_{min}]/2 \quad (9)$$

as an approximation of the median $\sigma(A)$ value. Element types were omitted from the work file if $|\overline{\sigma}(A) - M|/\overline{\sigma}(A) > 0.25$.

(3) Only those element types having $\overline{\sigma}(A) \leq 0.030$ Å after all checks were included in the work file.

The final work file contained 124905 $\overline{\sigma}(A)_o$ values derived from 990458 individual $\sigma(A)$ values obtained from 33955 CSD entries.

Results and discussion

Descriptive statistics

The work file contains $\overline{\sigma}(A)_o$ values for 80 element types ranging from lithium ($Z = 3$) to neptunium ($Z = 93$) and with occurrences ranging from 1 (xenon) to 32715 (carbon). 24 elements, spanning a Z range of 3–78, occur more than 500 times in the file. The $\langle\overline{\sigma}(A)_o\rangle$ values for these elements, given in Table 2, decrease in a reasonably systematic manner with increasing Z, as might be expected from values averaged over a wide variety of structure types and levels of crystallographic precision. We note, however, that certain elements, in particular the halogens (fluorine, chlorine, bromine and iodine), tend to disturb the downward trend in $\langle\overline{\sigma}(A)_o\rangle$. These monovalent elements exist on the molecular periphery, tend to exhibit higher displacement parameters than atoms that are constrained by two or more chemical bonds and, hence, are relatively less well located by X-ray analysis.

It is obviously inappropriate to depict $\overline{\sigma}(A)$ distributions even for the most commonly observed elements; however, we can rearrange (6) so as to convert or normalize each $\overline{\sigma}(A)_o$ value to its carbon equivalent, $\overline{\sigma}(C)_e$, for each structure *via*

$$\overline{\sigma}(C)_e = \overline{\sigma}(A)_o Z_A/6. \quad (10)$$

Further, if we can assume that the most frequently occurring elements are distributed across a wide variety of structure types and levels of crystallographic precision, then values of $\langle\overline{\sigma}(C)_e\rangle$ for these elements should be roughly constant. Table 2 shows that 17 of the 24 $\langle\overline{\sigma}(C)_e\rangle$ values calculated for the full data set lie in the narrow range 0.0058–0.0081 Å with the seven higher values (0.0092–0.0116 Å) arising from three halogens (fluorine, bromine and iodine) and four of the heaviest elements (tungsten, rhenium, osmium and platinum). This latter observation is in agreement with the study of Taylor & Kennard (1986), who found that the e.s.d.'s of heavy-atom positions may be relatively less reliable than those of light-atom positions. The overall behaviour of $\langle\overline{\sigma}(A)_o\rangle$ with increasing Z and the relative imprecision of heavier atoms are clearly revealed in Table 3. Here, $\overline{\sigma}(A)_o$ values for all elements are included in the statistics and averages are taken over five ranges of Z.

The results of Tables 2 and 3 would indicate that (10) provides a reasonable approximation by which we may normalize all $\overline{\sigma}(A)_o$ values to a common basis. In Fig. 1, we show the full distribution of the normalized $\overline{\sigma}(C)_e$ values to a maximum of 0.030 Å. We note, however, that 2190 $\overline{\sigma}(A)$ values (1.8% of the total) convert to $\overline{\sigma}(C)_e$ that exceed this limit. As with the distribution of $\overline{\sigma}$(C–C) values shown in

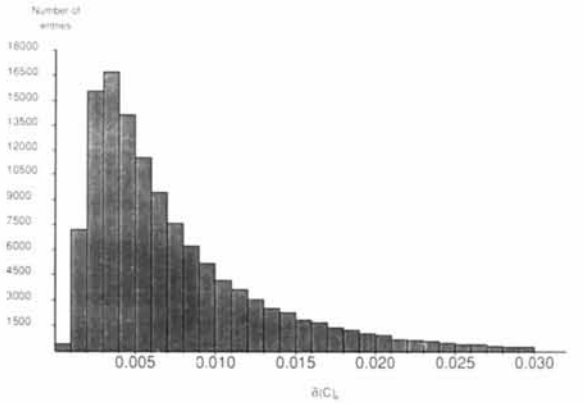

Fig. 1. Distribution of the 'mean carbon equivalent e.s.d.' $\overline{\sigma}(C)_e$ (see text) for all atoms in the work file.

Table 2. *Values of* $\langle\overline{\sigma}(A)_o\rangle$ *and* $\langle\overline{\sigma}(C)_e\rangle$ *(see text) in* Å *for the* 24 *elements with* ≥ 500 *occurrences in the full work file*

Values are given (*a*) for all data and (*b*) after removal of the upper and lower decile of the $\langle\overline{\sigma}(C)_e\rangle$ distribution shown in Fig. 1. El is the element symbol, Z is the atomic number and N_{OCC} is the number of occurrences of each element type.

		(*a*) For all data			(*b*) After decile removal		
El	Z	N_{OCC}	$\langle\overline{\sigma}(A)_o\rangle$	$\langle\overline{\sigma}(C)_e\rangle$	N_{OCC}	$\langle\overline{\sigma}(A)_o\rangle$	$\langle\overline{\sigma}(C)_e\rangle$
B	5	1583	0.00852	0.0071	1295	0.00828	0.0069
C	6	32715	0.00792	0.0079	26760	0.00678	0.0068
N	7	18564	0.00591	0.0069	15017	0.00531	0.0062
O	8	24574	0.00585	0.0078	19722	0.00488	0.0065
F	9	2192	0.00733	0.0110	1631	0.00520	0.0078
Si	14	1615	0.00249	0.0058	1316	0.00236	0.0055
P	15	6927	0.00276	0.0069	5693	0.00252	0.0063
S	16	6895	0.00236	0.0063	5518	0.00221	0.0059
Cl	17	6552	0.00286	0.0081	5374	0.00233	0.0066
Cr	24	582	0.00138	0.0055	471	0.00138	0.0055
Fe	26	1828	0.00136	0.0059	1493	0.00141	0.0061
Co	27	1100	0.00138	0.0062	872	0.00140	0.0063
Ni	28	759	0.00139	0.0065	626	0.00135	0.0063
Cu	29	1265	0.00136	0.0066	1034	0.00132	0.0064
Br	35	1600	0.00158	0.0092	1374	0.00130	0.0076
Mo	42	1299	0.00104	0.0073	1042	0.00106	0.0074
Ru	44	1041	0.00094	0.0069	837	0.00094	0.0069
Rh	45	811	0.00097	0.0073	644	0.00100	0.0075
Sn	50	532	0.00095	0.0079	433	0.00084	0.0070
I	53	946	0.00120	0.0106	743	0.00097	0.0086
W	74	1042	0.00084	0.0103	741	0.00059	0.0073
Re	75	564	0.00078	0.0098	398	0.00055	0.0069
Os	76	587	0.00092	0.0116	427	0.00065	0.0082
Pt	78	941	0.00084	0.0109	639	0.00055	0.0071

Table 3. *Values of* $\langle\overline{\sigma}(A)_o\rangle$ *and* $\langle\overline{\sigma}(C)_e\rangle$ *(see text) in* Å *for all* 80 *element types represented in the full work file averaged over ranges of atomic number* Z

Values are given (*a*) for all data and (*b*) after removal of upper and lower deciles of the $\langle\overline{\sigma}(C)_e\rangle$ distribution of Fig. 1. N_{OCC} is the number of element occurrences in each range.

			(*a*) For all data			(*b*) After decile removal			
Z_{min}	Z_{max}	N_{OCC}	$\langle Z\rangle$	$\langle\overline{\sigma}(A)_o\rangle$	$\langle\overline{\sigma}(C)_e\rangle$	N_{OCC}	$\langle Z\rangle$	$\langle\overline{\sigma}(A)_o\rangle$	$\langle\overline{\sigma}(C)_e\rangle$
3	10	79960	6.90	0.00670	0.0077	64707	6.88	0.00567	0.0065
11	18	22562	15.74	0.00267	0.0070	18354	15.74	0.00236	0.0062
19	36	9981	28.48	0.00145	0.0069	8214	28.57	0.00137	0.0065
37	57	6703	46.27	0.00102	0.0079	5330	46.21	0.00096	0.0074
58	92	5144	76.27	0.00083	0.0106	3639	76.21	0.00059	0.0075

paper I, the $\overline{\sigma}(C)_e$ distribution of Fig. 1 is highly skewed towards lower values, with a median value of *circa* 0.006 Å.

Finally, in this overview, we examine the variation of $\overline{\sigma}(C)_e$ with increasing R. The full $RN_c^{1/2}$ distribution was divided into 16 bins as shown in Table 4 and values of $\langle RN_c^{1/2}\rangle$ and $\langle\overline{\sigma}(C)_e\rangle$ computed for each bin. Centrosymmetric and noncentrosymmetric structures were treated together in this overall survey. As expected from (7), there is a steady increase in $\langle\overline{\sigma}(C)_e\rangle$ with increasing $\langle RN_c^{1/2}\rangle$ and the approximate linearity of this relationship is confirmed by the constant ($\sim$0.015) values of k [(7)] over a large part of the $\langle RN_c^{1/2}\rangle$ range. However, values of k at the lowest points of the joint distribution are considerably larger than the norm, while values of k for higher values of $RN_c^{1/2}$ begin to decrease quite rapidly from the norm. These observa-

Table 4. *Values of* $\langle RN_c^{1/2}\rangle$, $\langle\overline{\sigma}(C)_e\rangle$ *and the ratio* $k = \langle\overline{\sigma}(C)_e\rangle/\langle RN_c^{1/2}\rangle$ *for ranges of* $RN_c^{1/2}$

N_{OCC} is the number of element-type occurrences in each range. Binning procedures are described in the text.

$RN_c^{1/2}$ range	N_{OCC}	$\langle RN_c^{1/2}\rangle$	$\langle\overline{\sigma}(C)_e\rangle$	k
<0.1	390	0.0818	0.0034	0.0412
0.1–0.2	10620	0.1669	0.0032	0.0190
0.2–0.3	26015	0.2507	0.0041	0.0164
0.3–0.4	23647	0.3476	0.0056	0.0160
0.4–0.5	17437	0.4466	0.0071	0.0159
0.5–0.6	12721	0.5462	0.0086	0.0157
0.6–0.7	9447	0.6473	0.0102	0.0157
0.7–0.8	6961	0.7469	0.0116	0.0155
0.8–0.9	4824	0.8466	0.0129	0.0153
0.9–1.0	3542	0.9468	0.0140	0.0148
1.0–1.1	2472	1.0484	0.0153	0.0146
1.1–1.2	1725	1.1439	0.0164	0.0144
1.2–1.3	1445	1.2462	0.0172	0.0138
1.3–1.5	1366	1.3875	0.0187	0.0135
1.5–2.0	1162	1.6621	0.0203	0.0122
>2.0	576	2.2802	0.0233	0.0102

tions would imply that the lowest values of $\overline{\sigma}(C)_e$ and, hence, of $\overline{\sigma}(A)_p$ that are predicted by (6) and (7) are likely to be high relative to the values observed in crystal structures, whilst the opposite effect will occur for the highest $\overline{\sigma}(C)_e$ and $\overline{\sigma}(A)_p$ values. As in our previous study (paper I), it would appear that the most appropriate predictive equations will be obtained from regression experiments that exclude some proportion of the very low and very high observations from the $\sigma(C)_e$ distribution.

Regression analyses based on (7)

Simple linear regressions with an imposed zero intercept were carried out using local code that employed subroutine *G02CBF* of the NAG Library (Numerical Algorithms Group, 1990). In all regression experiments, the crystallographically observed $\overline{\sigma}(A)_o$ was the dependent variable and the quantity $RN_c^{1/2}/Z_A$ was the independent variable in the estimation of K [$=6k$ in (7)]. Predicted e.s.d.'s $\overline{\sigma}(A)_p$ were then obtained from (7) using the regression estimate of K. A variety of numerical criteria, defined in Table 5 and used previously in paper I, were employed to assess and compare the predictive abilities of the various regression experiments. We note that the quantities N_{50}, N_{25} and N_{10} are, perhaps, less valuable in this study since the overall numerical levels of $\overline{\sigma}(A)_o$ and $\overline{\sigma}(A)_p$ are much lower for general elements A than they were for the carbon case treated in paper I. Separate regressions for centrosymmetric (c) and noncentrosymmetric (nc) structures were carried out for each subset of data employed in the analysis. Data-selection criteria for the various data subsets are given in Table 6 and regression results are collected in Table 7.

The initial regression 7.1 made use of all available $\overline{\sigma}(A)$ values except for a very small number (554) for which $\overline{\sigma}(C)_e$ exceeded 0.040 Å. The R_σ values for the c and nc regressions are high (36.5 and 36.1%) compared with values for $\overline{\sigma}$(C–C) and $\overline{\sigma}$(C) regressions in paper I (typically 25–30%). Values of n_{50}, n_{25} and n_{10} obtained for regressions 1c and 1nc are also significantly lower than for the carbon case. Most importantly, the distributions of absolute errors, $\overline{\sigma}(A)_o - \overline{\sigma}(A)_p$, from regressions 7.1$c$ and 7.1nc are both skewed towards positive values [$\overline{\sigma}(A)_o > \overline{\sigma}(A)_p$] and these errors occur for the larger (>0.015 Å) values of $\overline{\sigma}(A)_o$. 6.9% of atoms have an absolute error that exceeds +0.005 Å. Conversely, the distribution of absolute percentage errors, $100[\overline{\sigma}(A)_o - \overline{\sigma}(A)_p]/\overline{\sigma}(A)_o$, is very severely skewed towards negative values and for the lowest values (<0.002 Å) of $\overline{\sigma}(A)_o$. Obviously, a given absolute error $\overline{\sigma}(A)_o - \overline{\sigma}(A)_p$ will generate a larger percentage error for a small $\overline{\sigma}(A)_o$ than it will for higher values. However, 13.6% of all atoms show a negative percentage error $>100\%$, a figure that seems over large from past experience.

Table 5. *Parameters used to assess and compare the results of the various regression experiments*

$\overline{\sigma}(A)_o$ are e.s.d.'s reported in crystal structure analyses; $\overline{\sigma}(A)_p$ are e.s.d.'s predicted by the regression equations.

Item	Description
R_σ	A pseudo-R-factor measuring the discrepancy between the distributions of $\overline{\sigma}(A)_o$ and $\overline{\sigma}(A)_p$. $R_\sigma = 100[\sum\lvert\overline{\sigma}(A)_o - \overline{\sigma}(A)_p\rvert/\sum\overline{\sigma}(A)_o]$
r.m.s.(σ)	The root-mean-square error, *i.e.* r.m.s.$(\sigma) = \{\sum[\overline{\sigma}(A)_o - \overline{\sigma}(A)_p]^2/n\}^{1/2}$ for the n observations in any data subset
pc(σ)	The mean percentage error, *i.e.* $\mathrm{pc}(\sigma) = 100(\sum\{[\overline{\sigma}(A)_o - \overline{\sigma}(A)_p]/\overline{\sigma}(A)_o\})/n$ for the n observations in any data subset
n_{50}	Percentage of atoms with $\lvert\overline{\sigma}(A)_o - \overline{\sigma}(A)_p\rvert \le 0.0050$ Å
n_{25}	Percentage of atoms with $\lvert\overline{\sigma}(A)_o - \overline{\sigma}(A)_p\rvert \le 0.0025$ Å
n_{10}	Percentage of atoms with $\lvert\overline{\sigma}(A)_o - \overline{\sigma}(A)_p\rvert \le 0.0010$ Å
n_{50}	Percentage of atoms for which $\overline{\sigma}(A)_p$ is within 50% of $\overline{\sigma}(A)_o$
n_{25}	Percentage of atoms for which $\overline{\sigma}(A)_p$ is within 25% of $\overline{\sigma}(A)_o$
n_{10}	Percentage of atoms for which $\overline{\sigma}(A)_p$ is within 10% of $\overline{\sigma}(A)_o$

For reasons similar to those noted above, we restricted the $\overline{\sigma}$(C–C) and $\overline{\sigma}$(C) regressions in paper I to the most typical and broad ranges of these quantities by eliminating the lowest and highest values. Here, the summary statistics of Tables 2–4 indicate that the highest and lowest $\overline{\sigma}(C)_e$ deviated significantly from general trends and we experimented with data sets that excluded lower and upper quantiles of the $\overline{\sigma}(C)_e$ distribution of Fig. 1. As a result, it was found that a data set that excludes the upper and lower deciles of this distribution generates $\langle\overline{\sigma}(C)_e\rangle$ values that are much more consistent over a range of individual elements and element groups. These $\langle\overline{\sigma}(C)_e\rangle$ values are included in Tables 2 and 3. All $\langle\overline{\sigma}(C)_e\rangle$ values now fall within the relatively narrow range of 0.0059–0.0086 Å for the 24 elements of Table 2, while the $\langle\overline{\sigma}(C)_e\rangle$ values for the five Z ranges of Table 3 now lie in the range 0.0062–0.0075 Å.

The data subsets for regressions 7.2–7.4 were all selected on the basis of $\overline{\sigma}(C)_e$ limits (see Table 6) derived from analysis of the complete $\overline{\sigma}(C)_e$ distribution, *i.e.* the distribution that included both centrosymmetric and noncentrosymmetric structures. Thus, there is no guarantee that the numbers of atoms (N_-, N_+; Table 6) that are below and above the $\overline{\sigma}(C)_e$ limits will be equal for the c and nc cases. Table 6 shows that, in practice, $N_- \simeq N_+$ for atoms from centrosymmetric structures but that $N_- << N_+$ for the noncentrosymmetric cases, *i.e.* there is a relative lack of noncentrosymmetric structures that exhibit e.s.d.'s of the lowest magnitudes. This observation would agree with the tentative evidence provided by Taylor & Kennard (1986) that e.s.d.'s in noncentrosymmetric structures may be relatively less accurate than those in centrosymmetric structures.

Table 6. *Summary of data subsets used in regression analyses based on* (7)

RN is the regression number; Z indicates the range of element types covered by each analysis; R_{min} and R_{max} define R-factor limits; σ_{min} and σ_{max} are $\overline{\sigma}(C)_e$ limits (see text); c/nc indicates centrosymmetric or noncentrosymmetric structures included; N_{ent} is the number of CSD entries used; N_{tot} is the total number of $\overline{\sigma}(A)_o$ values available; N_- and N_+ are numbers of atoms with $\overline{\sigma}(C)_e$ below and above the σ limits; N_{al} is the total number of $\overline{\sigma}(A)_o$ finally used in each regression.

RN	7.1		7.2		7.3		7.4		7.5	
Z	All		All		All		$Z \geq 10$		$3 \leq Z < 10$	
R_{min}	0.01		0.01		0.02		0.01		0.01	
R_{max}	0.10		0.10		0.07		0.10		0.10	
σ_{min}	0.0001		0.0023		0.0022		0.0023		0.0023	
σ_{max}	0.0400		0.0160		0.0140		0.0160		0.0160	
c/nc	c	nc	c	nc	c	nc	c	nc	c	nc
N_{ent}	25685	8066	25685	8066	22757	7087	25685	8066	25685	8066
N_{tot}	97874	27031	97874	27031	87068	23919	37345	7577	60529	19454
N_-	1	0	10465	1419	8979	1127	4260	468	6205	951
N_+	375	179	9470	3307	9159	3251	3540	1117	5930	2190
N_{at}	97480	26852	77939	22305	68930	19541	29545	5992	48394	16313

Table 7. *Results of regression analyses based on* (7)

The parameters used as assessment criteria are defined in Table 5. RN is the regression number and c/nc indicates that only centrosymmetric or noncentrosymmetric structures were included.

RN	7.1		7.2		7.3		7.4		7.5	
c/nc	c	nc	c	nc	c	nc	c	nc	c	nc
K	0.09286	0.12542	0.08143	0.10789	0.08093	0.10868	0.06160	0.08581	0.08354	0.10948
$\sigma(K)$	0.00014	0.00037	0.00011	0.00030	0.00012	0.00030	0.00014	0.00046	0.00015	0.00035
R_σ	36.5	36.1	30.9	31.0	30.0	30.1	31.8	31.2	28.8	30.2
r.m.s.(σ)	0.0029	0.0034	0.0019	0.0022	0.0017	0.0019	0.0008	0.0009	0.0024	0.0026
pc(σ)	53.1	48.0	37.5	35.0	37.2	34.8	34.4	34.0	29.4	30.9
N_{50}	93.1	90.1	96.5	95.1	97.8	97.1	99.9	99.9	94.7	93.5
N_{25}	80.5	73.8	85.3	81.1	88.0	84.8	98.4	97.4	78.2	75.5
N_{10}	49.3	41.4	57.8	51.3	61.2	54.5	85.6	81.4	45.3	42.2
n_{50}	63.6	67.7	76.0	78.8	76.2	79.0	77.3	78.9	84.3	83.3
n_{25}	37.0	39.4	45.6	46.6	45.9	47.2	44.5	46.4	52.0	49.7
n_{10}	15.6	16.5	19.1	19.6	19.4	19.9	18.4	19.4	22.1	21.0

Regression 7.2 was conducted exactly as for regression 7.1 but with lower and upper limits of $\overline{\sigma}(C)_e$ set at 0.0023 and 0.0160 Å. This decreased the total number of $\overline{\sigma}(A)_o$ values used in regression 7.2 by 24661 or 19.7%, with 11884 values below the lower $\overline{\sigma}(C)_e$ limit and 12777 values above the upper limit. Results (Table 7) for regressions 7.2c and 7.2nc are significantly improved by comparison with their regression 7.1 counterparts. The R_σ values are now close to 30%, pc(σ) decreases and there are major increases in n_{50}, n_{25} and n_{10} and in N_{50}, N_{25} and N_{10}. We note, in particular, that ~82% of $\overline{\sigma}(A)_p$ are within 0.0025 Å of their observed values and that nearly 80% of the predicted values are within ±50% of their observed values across a very wide numerical range of $\overline{\sigma}(A)_o$.

The composite (c+nc) absolute and percentage error distributions from regression 7.2 are illustrated in Figs. 2(*a*) and (*b*). The distribution of $\overline{\sigma}(A)_o$–$\overline{\sigma}(A)_p$ is now much less skewed and only 2.9% of atoms have an absolute error that exceeds 0.005 Å. The percentage error distribution still retains the long tail towards large negative values, for reasons already noted, but the percentage of atoms exhibiting a negative percentage error that exceeds 100% is significantly reduced at 6.5%.

The results of Fig. 2 and Table 7 for regression 7.2 indicate that the removal of the lower and upper deciles of the $\overline{\sigma}(C)_e$ distribution have removed a large proportion of those $\overline{\sigma}(A)_o$ values that yielded inconsistent results in regression 7.1. However, in regression 7.3 we have applied an additional restriction by selecting $\overline{\sigma}(A)_o$ values arising from structures having R in the range 2–7%. Although this does generate assessment criteria that are marginally improved over those from regression 7.2, we note that the regression constants K_c and K_{nc} are identical to three significant figures for experiments 7.2 and 7.3.

The ratios of the regression constants for noncentrosymmetric and centrosymmetric subsets, K_{nc}/K_c, are 1.351, 1.325 and 1.343 for regressions 7.1, 7.2 and 7.3, respectively, in fair agreement with the value of 1.414 expected from the theoretical analysis of Cruickshank (1960). However, these ratios from the $\overline{\sigma}(A)$ regressions are not as close to $2^{1/2}$ as the 1.399 and 1.417 obtained from the $\overline{\sigma}$(C–C) and $\overline{\sigma}$(C) regressions of paper I. A possible reason for this

effect is that, although regressions 7.1–7.3 of Table 7 involve $\overline{\sigma}(A)_o$ values from a very broad distribution of atoms A, this distribution is dominated by contributions from carbon, nitrogen and oxygen. These element types contribute 75853 (61%) of observations to regression 7.1 and 61499 (61%) of observations to regression 7.2. Further, we have already noted (Table 4) a steady decrease in k $[=\langle\overline{\sigma}(\mathrm{C})_e\rangle/\langle RN_c^{1/2}\rangle]$ with increasing $RN_c^{1/2}$ that implies a dependency of k on increasing size of atom A.

Further evidence supporting treating heavier atoms separately from light atoms such as carbon, nitrogen and oxygen can be obtained by comparing the results from the current regression 7.2 with the results of paper I from regressions based on (7) for $\overline{\sigma}(\mathrm{C})$ and $\overline{\sigma}(E)$, where E is the *heaviest* non-C atom in a structure. Thus, for equations cast in the form of (7), paper I reports K_c, $K_{nc} = 0.0792$, 0.11266 for a regression based entirely on $\overline{\sigma}(\mathrm{C})_o$ data: values that are remarkably close to K_c and K_{nc} for regression 7.2 of Table 7. For the regression based on $\overline{\sigma}(E)_o$ data, however, the K_c and K_{nc} values of paper I were significantly lower at 0.0678 and 0.1006, respectively.

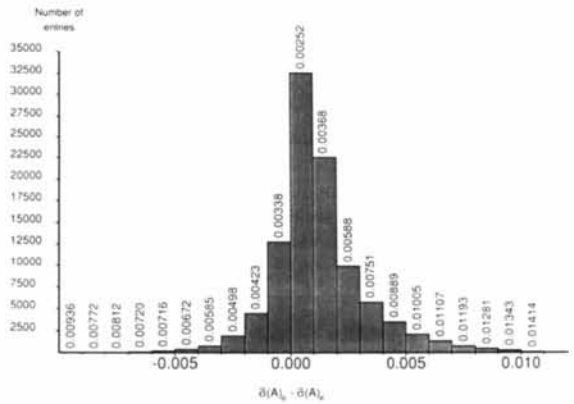

Fig. 2. (*a*) Distribution of numerical errors $\overline{\sigma}(A)_o - \overline{\sigma}(A)_p$ from regression 7.2. (*b*) Distribution of percentage errors $[\overline{\sigma}(A)_o - \overline{\sigma}(A)_p]/\overline{\sigma}(A)_o$ from regression 7.2.

For these reasons, we have performed regression 7.4 of Tables 6 and 7, in which all contributions from first-row elements have been removed from the $\overline{\sigma}(A)_o$ distribution. Other limitations on the data subset are exactly as for regression 7.2. Here, of course, we are dealing with the low end of the $\overline{\sigma}(A)_o$ distribution that has a mean $\langle\overline{\sigma}(A)_o\rangle$ of 0.00141 Å. Hence, it is not surprising that values of N_{50}, N_{25} and N_{10} are relatively high for regression 7.4. However, other assessment criteria in Table 7 are directly comparable with those from regression 7.2. In particular, only 3.5% of the very lowest $\overline{\sigma}(A)$ now have a negative percentage error that exceeds 100%.

The most important features of regression 7.4 are the constants K_c and K_{nc}, at 0.06160 (14) and 0.08581 (46), respectively, very close to the 0.0678 and 0.1006 obtained in the $\overline{\sigma}(E)$ regression of paper I. The ratio K_{nc}/K_c from regression 7.4 is 1.393, much closer to the expectation value of $2^{1/2}$ (Cruickshank, 1960) than ratios obtained from regressions 7.1–7.3. The reason for the lower K values from regression 7.4 by comparison with those from the $\overline{\sigma}(C)_o$ regression of paper I or the carbon-, nitrogen- and oxygen-dominated regression 7.2 of Table 7 lies in the approximations we have used in calculating N_c *via* (2) (Cruickshank, 1993). Properly, N_c should be calculated as described by (4) and, hence, (7) should be written as

$$\overline{\sigma}(A)_p = \overline{f}_C k R N_c^{1/2}/\overline{f}_A, \qquad (11)$$

where $\overline{f}_C$ and $\overline{f}_A$ are scattering factors for carbon and element A at $\overline{s}$, the r.m.s. reciprocal radius for the observed reflections. Since we do not know the $\overline{s}$ values from CSD information, we have substituted Z values in our approximation of (2). However, since scattering factors for carbon, nitrogen and oxygen fall off more rapidly than those for heavier elements (and these elements will tend to have lower atomic displacement parameters relative to C atoms in the same structure), then the ratio $Z_C(=6)/Z_A$ of (7) will be larger than the ratio $\overline{f}_C/\overline{f}_A$ of (11) for heavier elements A. Thus, when (7) is applied to heavier elements alone, as in regression 4, the constant K will compensate for the higher Z_C/Z_A ratio and be lower than for regressions based on, or dominated by, the lighter carbon, nitrogen and oxygen.

Results for regression 7.5, based on $\overline{\sigma}(A)_o$ for first-row elements alone, confirm this point quite conclusively. The K_c and K_{nc} values are considerably larger [at 0.08354 (15) and 0.10948 (35)] than for the heavier-atom case, although their ratio of 1.311 is low by comparison with that from regression 7.4. The assessment criteria from regression 7.5 (Table 7) are, however, an improvement on those obtained in

the general regression 7.2. In particular, R_σ, pc(σ) and n_{50}, n_{25} and n_{10}, which are comparable across all regressions, show their best values for regression 7.5.

Regression analyses incorporating scattering-factor data

We have briefly investigated the effect of the differential fall-off in scattering power between lighter and heavier elements by assuming a constant $\bar{s}$ that is consistent with perceived data-collection practice. We base our $\bar{s}$ on a θ_{max} value of 55° for Cu $K\alpha$ radiation, whence s_{max} ($=2\sin\theta_{max}/\lambda$) = 1.063 Å^{-1} and $\bar{s}$ [$=(3/5)^{1/2}s_{max}$; see Appendix to paper I] = 0.823 Å^{-1}, from which $\bar{\theta}$ (the r.m.s. θ value for the diffraction data) is 27.55° and $\sin\bar{\theta}/\lambda = 0.30$ Å^{-1}. The corresponding θ_{max} and $\bar{\theta}$ for Mo $K\alpha$ radiation are 22.19 and 12.31°. Values of neutral-atom scattering factors ($\bar{f}_i$) at $\sin\theta/\lambda = 0.30$ Å^{-1} (*International Tables for Crystallography*, 1992) were then used (*a*) to replace the Z_i in (7) so as to transform it into (11) and/or (*b*) to replace the Z_i in (2) so as to better mimic the original Cruickshank (1960) expression for N_c *via* (4). To avoid confusion in describing the remainder of our investigation, we denote this $\bar{f}_i$-based N_c value as $\bar{N}_c$ and retain N_c as the symbol for the Z_i-based calculation of (2). Values of $\bar{N}_c^{1/2}$ were computed for all structures and were included, alongside $N_c^{1/2}$, in a regenerated work file.

Three further regression experiments were performed, each based on variants of (7), in which the $\bar{f}_i$ are used in different combinations of (*a*), (*b*) of the previous paragraph, *viz*

$$\overline{\sigma}(A)_p = 6kR\bar{N}_c^{1/2}/Z_A = KR\bar{N}_c^{1/2}/Z_A \qquad (12)$$

$$\overline{\sigma}(A)_p = 2.494kN_c^{1/2}/\bar{f}_A = KRN_c^{1/2}/\bar{f}_A \qquad (13)$$

$$\overline{\sigma}(A)_p = 2.494k\bar{N}_c^{1/2}/\bar{f}_A = KR\bar{N}_c^{1/2}/\bar{f}_A, \qquad (14)$$

where 2.494 is $\bar{f}_c$ (f_c at $\sin\theta/\lambda = 0.30$ Å^{-1}).

Regressions 12.2 (all data), 12.4 (atoms with $Z \geq 10$) and 12.5 (atoms with $Z \geq 10$) were based on (12) with data-selection limits (Table 6) exactly as for regressions 7.2, 7.4 and 7.5 [σ limits are identical since (10) for $\overline{\sigma}(C)_e$ still applies here]. Full regression results from (12) have been deposited,* since they are almost identical to those for regressions 7.2, 7.4 and 7.5 of Table 7. Indeed, substitution of $\bar{N}_c^{1/2}$ for $N_c^{1/2}$ generates assessment criteria that are marginally worse than those of Table 7. Further, the constants K_c and K_{nc} for regression 12.4 (heavier atoms) at 0.0402 (1) and 0.00593 (5) are considerably lower than their counterparts from regression 12.5 (first-row elements) at 0.0570 (1) and 0.0838 (3), *i.e.* the pattern observed in regressions 7.4 and 7.5 is not corrected by the use of the f_i-based $\bar{N}_c^{1/2}$ values.

For regressions based on (13) and (14), the expression used to generate $\overline{\sigma}(C)_e$ involving Z_i [(10)] must be altered to one involving $\bar{f}_i$:

$$\overline{\sigma}(C)_f = \overline{\sigma}(A)_o\bar{f}_A/\bar{f}_C, \qquad (15)$$

so that appropriate limits can be established to effect removal of upper and lower deciles of the complete $\overline{\sigma}(C)_f$ distribution. The analysis yielded σ_{min} and σ_{max} of 0.0028 and 0.0200 Å, somewhat higher than the 0.0023 and 0.0160 Å for the $\overline{\sigma}(C)_e$ distribution of Fig. 1: a reflection of the fact that $\bar{f}_A/\bar{f}_C > Z_A/Z_C$. Descriptive statistics for $\overline{\sigma}(C)_f$ analogous to those given in Tables 2–4 for $\sigma(C)_e$ show similar trends, but with enhanced magnitudes for $\langle\overline{\sigma}(C)_f\rangle$ values. Thus, the $\langle\overline{\sigma}(C)_f\rangle$ that correspond to the final column of Table 3 are, respectively, 0.00779, 0.00802, 0.00913, 0.00990 and 0.00996 Å.

Full results for regressions 13.2, 13.4 and 13.5, shown in Table 8, are based on an equation that uses our original Z_i-based $N_c^{1/2}$ value, but uses the $\bar{f}_i$ (rather than Z_i) in (5). Error distributions have the same form as Figs. 2(*a*) and (*b*) but most of the assessment criteria in Table 8 are now improved over those for regressions 7.2, 7.4 and 7.5 (Table 7). In particular, the criteria associated with percentage differences [pc(σ), n_{50}, n_{25}, n_{10}] now show some very significant improvements over the corresponding values in Table 7. These improvements relate entirely to the heavier atoms, as can be seen by comparing n_{50}, n_{25} and n_{10} values from regression 13.4 with those from 7.4. Values for lighter elements (13.5) are almost identical to those from 7.5. The most important change in Table 8 is that the K values for the heavier-atom subset (13.4) are now almost identical to those for the first-row elements (13.5). The ratios of K_{heavy}/K_{light} are now 0.910 (centrosymmetric) and 0.965 (noncentrosymmetric) by comparison with the 0.737 and 0.781 obtained from regressions 7.4 and 7.5 (Table 7).

Finally, results from (14) in which all Z_i values in our original (7) are replaced by $\bar{f}_i$, have marginally improved assessment criteria by comparison with those from (7), but are worse than those from (13). Full details have been deposited* but we note that the significant improvements in n_{50}, n_{25} and n_{10} values for heavier atoms observed in 13.4 (Table 8) are not maintained in the analogous regression 14.4. Further, the ratios K_{heavy}/K_{light} are 0.890 (centrosymmetric) and 0.899 (noncentrosymmetric), an improvement over the results of (7) but not as close to our 'expectation' value of unity as those from (13) (Table 8).

* The results of regressions 12.2, 12.4 and 12.5 (Table *A*) and regressions 14.2, 14.4 and 14.5 (Table *B*) have been deposited with the IUCr (Reference: MU0317). Copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

* See deposition footnote.

Table 8. *Results of regression analyses based on* (13)

The parameters used as assessment criteria are defined in Table 5. RN is the regression number and *c*/nc indicates that only centrosymmetric or noncentrosymmetric structures were included. The selection criteria for regressions 13.2, 13.4 and 13.5 are analogous to those used for regressions 2, 4 and 5 of Table 6 and Table 7. Other notations are as in Table 6.

RN	13.2		13.4		13.5	
Z	All		$Z \geq 10$		$Z < 10$	
R_{min}	0.01		0.01		0.01	
R_{max}	0.10		0.10		0.10	
σ_{min}	0.0028		0.0028		0.0028	
σ_{max}	0.0200		0.0200		0.0200	
c/nc	*c*	nc	*c*	nc	*c*	nc
N_{ent}	25685	8066	25685	8066	25685	8066
N_{tot}	97874	27031	37345	7577	60529	19454
N_-	10877	1684	2535	280	8342	1404
N_+	9781	3208	5461	1593	4320	1615
N_{at}	77216	22139	29349	5704	47867	16435
K	0.03814	0.05041	0.03487	0.04869	0.03831	0.05046
$\sigma(K)$	0.00006	0.00014	0.00008	0.00026	0.00007	0.00016
R_σ	30.0	30.6	31.1	30.5	29.6	30.7
r.m.s.(σ)	0.0022	0.0025	0.0007	0.0008	0.0027	0.0029
pc(σ)	31.4	31.2	31.9	31.6	29.5	30.8
N_{50}	95.5	93.8	99.9	99.9	92.7	91.6
N_{25}	84.0	79.0	98.9	98.2	74.9	72.3
N_{10}	59.4	51.4	87.0	83.1	42.7	40.4
n_{50}	81.5	82.5	80.3	82.0	84.2	83.1
n_{25}	49.2	49.3	46.7	47.7	51.4	49.7
n_{10}	20.8	20.7	19.5	20.6	21.9	20.8

Concluding remarks

The work reported here represents an extension and generalization of our earlier carbon-based study (paper I). Again, the study has necessarily been restricted by the information content of the CSD. However, our earlier work did indicate that equations that incorporated details of the diffraction experiment, *via* inclusion of $p^{1/2}$ and estimates of $\bar{s}$ [(13)], generated only marginally better estimates of $\overline{\sigma}(A)_p$ than two-parameter equations involving only R and $N_c^{1/2}$. Possible reasons for this behaviour are discussed in paper I.

The present analysis provides additional support for some of the conclusions concerning atomic e.s.d.'s drawn by Taylor & Kennard (1986) from their study of 100 structures, each of which had been determined by two independent research groups. They showed that atomic e.s.d.'s were typically underestimated by a factor of 1.4–1.45 overall, and noted specifically that: (*a*) e.s.d.'s are underestimated by a greater factor in the more precise structure determinations; (*b*) e.s.d.'s of heavier atom positions are relatively less reliable than those for light atoms; and (*c*) there was tentative evidence that e.s.d.'s in noncentrosymmetric structures may be less accurate than those in centrosymmetric structures. Further work with our current work files is planned to address these and other systematic aspects of atomic precision in X-ray crystallographic analyses.

Despite the systematic effects that are inherent in our $\overline{\sigma}(A)_o$ data, we feel that the regression experiments summarized in Tables 6, 7 and 8 provide estimated e.s.d.'s that are suitable for their intended purpose: as a substitute measure of structural precision for those CSD entries that do not contain the published e.s.d.'s of atomic positions. Specifically, we recommend use of the following equations:

(*a*) for first-row elements ($Z_A \leq 10$),

$$\overline{\sigma}(A)_p = 0.0835RN_c^{1/2}/Z_A$$

for centrosymmetric structures, (16*a*)

$$\overline{\sigma}(A)_p = 0.1095RN_c^{1/2}/Z_A$$

for noncentrosymmetric structures; (16*b*)

(*b*) for heavier elements ($Z_A \leq 10$),

$$\overline{\sigma}(A)_p = 0.0616RN_c^{1/2}/Z_A$$

for centrosymmetric structures, (17*a*)

$$\overline{\sigma}(A)_p = 0.0858RN_c^{1/2}/Z_A$$

for noncentrosymmetric structures; (17*b*)

where N_c is given by (2) above. In the light of our experiments with scattering-factor data, we also recommend use of the results of regressions 13.2 as a composite equation to predict e.s.d.'s for any element A, irrespective of Z value, *viz*:

$$\overline{\sigma}(A)_p = 0.0381RN_c^{1/2}/\bar{f}_A$$

for centrosymmetric structures; (18*a*)

$$\overline{\sigma}(A)_p = 0.0504RN_c^{1/2}/\bar{f}_A$$

for noncentrosymmetric structures; (18*b*)

with N_c calculated by (2) and where $\bar{f}_A$ is the scattering factor for element A at $\sin\theta/\lambda = 0.30$ Å^{-1}. We acknowledge that (16)–(18) will tend to overestimate

the lowest e.s.d. values, in some cases by a factor of 2 or more, but in the light of the work of Taylor & Kennard (1986) this may not be unrealistic.

We are greatly indebted to Professor Durward Cruickshank FRS for valuable discussions and comments. We thank Drs Olga Kennard FRS, Robin Taylor and Michael Doyle for their interest in this project and for helpful suggestions. The Science and Engineering Research Council (UK) is thanked for financial support to JCC.

References

ALLEN, F. H., COLE, J. C. & HOWARD, J. A. K. (1995). *Acta Cryst.* A51, 95–111.

ALLEN, F. H., DAVIES, J. E., GALLOY, J. J., JOHNSON, O., KENNARD, O., MACRAE, C. F., MITCHELL, E. M., MITCHELL, G. F., SMITH, J. M. & WATSON, D. G. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 187–204.

Cambridge Structural Database User's Manual (1992). Version 5.1. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England.

CRUICKSHANK, D. W. J. (1960). *Acta Cryst.* **13**, 774–777.

CRUICKSHANK, D. W. J. (1993). Private communication.

International Tables for Crystallography (1992). Vol. C. *Mathematical, Physical and Chemical Tables*, edited by A. J. C. WILSON. Dordrecht: Kluwer Academic Publishers.

MUIR, K. W. & MALLINSON, P. R. (1993). *J. Appl. Cryst.* **26**, 142–143.

Numerical Algorithms Group (1990). Subroutine Library. Numerical Algorithms Group Ltd, Wilkinson House, Jordan Hill Road, Oxford OX2 5DR, England.

TAYLOR, R. & KENNARD, O. (1986). *Acta Cryst.* **B42**, 112–120.

TEMPLETON, D. H. (1959). *Acta Cryst.* **12**, 771–773.

Acta Cryst. (1995). A**51**, 121–123

On the Harker–Kasper Inequalities

BY SALVINO CICCARIELLO

Dipartimento di Fisica 'G. Galilei' and Sezione INFM, via Marzolo 8, *I*-35131 *Padova, Italy*

(*Received* 1 *June* 1994; *accepted* 17 *July* 1994)

Abstract

It is shown that two unitary structure factors generate a two-parameter family of Harker–Kasper inequalities and that the strongest of these coincides with the third-order determinant Karle–Hauptman inequality.

Inequality relationships among structure factors are called Harker–Kasper (HK) inequalities when they are obtained by the Cauchy–Schwartz inequality (Harker & Kasper, 1948). Their derivation requires algebraic manipulations of the unitary structure-factor (u.s.f.) expressions. Thus, HK inequalities in general have considerably different expressions, as is illustrated by, for instance, equations (6), (15) and (16) of Woolfson (1988). Moreover, Woolfson showed that some HK inequalities can be more effective that the lowest determinant Karle–Hauptman (Karle & Hauptman, 1950) inequalities. Since it is commonly believed that HK inequalities are contained within the *complete* set of determinant inequalities, the former result does not conflict with this idea. This paper reports explicit proof of this in the case of the lowest-order inequality. More definitely, it is shown that a *two*-parameter family of HK inequalities, involving three reflexions, can be constructed from *two* unitary structure factors and that the strongest of these inequalities, *viz* the one that holds true whatever the parameter values, is the third-order determinant Karle–Hauptman (Karle & Hauptman, 1950) inequality, in the form obtained by Goedkoop (1950). [For a recent review, see Hauptman (1991).]

Since the analysis is derived from that of Woolfson (1988; hereinafter referred to as I), his notation is adopted. In order to prove the above statement, the effect of translations on the derivation of the HK inequality [equation (6*a*) of I] is first analysed. Translation of the origin by **a** implies that the u.s.f. $U(\mathbf{h})$ [whose modulus $|U(\mathbf{h})|$ is denoted $\mu(\mathbf{h})$, so that $U(\mathbf{h}) = \exp i\varphi(\mathbf{h})\,\mu(\mathbf{h})$] becomes

$$\begin{aligned} U_{\mathbf{a}}(\mathbf{h}) &= \sum_{j=1}^{N} n_j \exp[i2\pi\mathbf{h}\cdot(\mathbf{r}_j + \mathbf{a})] \\ &= \exp[i2\pi\mathbf{a}\cdot\mathbf{h}]U(\mathbf{h}) \\ &\equiv \exp[i\phi_{\mathbf{a}}(\mathbf{h})]U(\mathbf{h}). \end{aligned} \tag{1}$$

The same calculations of *I*, yielding inequality (6*a*) of I, can now be repeated bt starting with $U_{\mathbf{a}}(\mathbf{h})$ and $U_{\mathbf{a}}(\mathbf{k})$. They give

$$\begin{aligned} &|\exp[i\phi_{\mathbf{a}}(\mathbf{h})U(\mathbf{h})] + \exp[i\phi_{\mathbf{a}}(\mathbf{k})]U(\mathbf{k})|^2 \\ &\quad \le 2\{1 + \mu(\mathbf{h}-\mathbf{k})\cos[\varphi(\mathbf{h}-\mathbf{k}) + 2\pi\mathbf{a}\cdot(\mathbf{h}-\mathbf{k})]\}. \end{aligned} \tag{2}$$

With the definition

$$\delta \equiv \varphi(\mathbf{h}-\mathbf{k}) + 2\pi\mathbf{a}\cdot(\mathbf{h}-\mathbf{k}), \tag{3}$$