Acta Cryst. (1990). B**46**, 717–723

On Metal–Oxygen Coordination. A Statistical Method to Determine Coordination Number. I. Calcium*

By Giacomo Chiari

Dipartimento di Scienze della Terra, Sezione di Mineralogia e Cristallografia, Università di Torino, Via Valperga Caluso 37, *I*-10125 *Torino, Italy*

(*Received* 17 *March* 1990; *accepted* 25 *June* 1990)

Abstract

The environment around the Ca cation for compounds involving bonded oxygen has been studied for 254 inorganic structures containing a total of 368 polyhedra. Selection was made on the basis of the accuracy of the structural data. The coordination number (CN) was assigned using the criteria of maximum gap in the Ca—O distances and balanced bond-valence sums for Ca, but 32 cases were still difficult to assign unambiguously. A series of variables related to the Ca—O distances were calculated and averaged for each value of CN. The Gaussian curves representing the distribution of these variables for each CN overlap strongly. By way of contrast, the volume of the coordination polyhedra (Vol) showed well separated curves. Statistical analysis was applied to the set of structures with known CN, with seven variables and then with Vol alone, which seems to discriminate between the various CN groups equally well. A strong linear dependence was found for CN *versus* Vol. A method is proposed to assign CN in uncertain cases based on the equation: CN = 0·197 (2)Vol + 2·83 (5). Application of this equation to the unassigned cases compares favourably with discriminant analysis using the larger set of variables.

Introduction

One of the most common ways of describing an inorganic crystal structure is to define coordination polyhedra around the cations. It is therefore important to have precise rules in order to establish which anions are coordinated by the cations. Pfeiffer (1915, 1916) introduced the concept of coordination number (CN hereafter), defined as 'the number of near neighbours linked by valence forces to a central atom'. In most cases one can distinguish between the nearest and second-nearest neighbours on the basis of the maximum gap in the cation–anion distances ranked in increasing order (Brunner & Schwarzenbach, 1971). Unfortunately, this gap does not always exist, especially for distorted polyhedra and high values of CN. Brown (1988) affirms that 'the assignment of CN in inorganic structures is usually a matter of individual judgment, although there have been a number of attempts to devise systematic definitions'.

Pauling (1929, 1960) calculates the predicted CN (PCN) on the basis of ionic radius ratio. Frank & Kasper (1958), for alloy structures, define CN by considering the Voronoi (1908) polyhedron surrounding each atom [the difficulties arising from this approach were discussed by O'Keeffe (1979)]. Brunner (1977) suggests taking the largest gap in the differences of the reciprocals of the interatomic distances as the cut-off criterion. This method was applied in order to determine CN for calcium coordinated to water (Einspahr & Bugg, 1980), except that no O atom with Ca—O less than 2·8 Å was excluded. Brunner (1977) and Bhandary & Girgis (1977) proposed giving each atom a weight which decreases with cation–anion distance. This leads to a non-integral CN. Carter (1978) and O'Keeffe (1979) regard coordinating atoms as contributing faces to the Voronoi polyhedron and their contributions are weighted in proportion to the solid angle subtended by the face at the centre, which again leads to non-integral CN's. The same is true for the effective coordination numbers proposed by Hoppe (1979). Gelato (1981) calculates the domain of an atom in a structure, making use of the concept of 'radical planes' introduced by Fischer, Koch & Hellner (1971). Using the cut-off criteria proposed by Brunner & Schwarzenbach (1971), Brunner (1977) and O'Keeffe (1979) three weighted CN's are calculated. Sandomirskiy & Baturin (1985) assign CN for potassium on the basis of the bond-valence balance for the anions as calculated by the Pyatenko (1973) method. They also made use of the maximum-gap criterion. Altermatt & Brown (1985) select CN by checking the agreement between the sum of the bond valence [calculated according to Brown & Altermatt (1985)] and the oxidation state of the cation. In this way anomalous situations in which the cation receives an excess or a deficiency of bond valence from the anions can be made evident. It should be emphasized

---

* Preliminary results on this work were presented at the XIth European Crystallographic Meeting, Vienna (Chiari, 1988; Chiari & Ferraris, 1988).

0108-7681/90/060717-07$03.00

though, that the spread of bond-valence sums for various polyhedra is such that this method does not always lead to certain results. Furthermore, the anions for which a doubt exists have the longest distances from the cation, and their contribution to the bond-valence sum is small. Therefore the case for considering them, bonded or not, is a weak one.

An indication of the expected value of bond length is given by the sum of the ionic radii, but this value is of very little use in establishing CN, since normally there is a large spread in the individual cation–anion distances in the polyhedra.

Another criterion used to establish CN is to define the largest cation–anion distance, $L_{\max}$, that will still be considered a bond. All anions which form a bond exceeding that value are excluded from the coordination sphere. Donnay & Allmann (1970) estimate the values of $L_{\max}$ by extrapolating the 'effective ionic radii' of Shannon & Prewitt (1969), which are themselves dependent on CN, to bond valence = 0. These values are approximate and often appear to be too large (3·25 Å for Ca—O) but they give an idea of how far the search for possibly coordinated atoms should be extended. Furthermore, every anion at a shorter distance from the cation than $L_{\max}$ cannot be considered coordinated, because it could be 'screened' by other closer atoms.

It seems therefore that the problem of establishing CN is still not completely solved. It is the author's intention to tackle this problem for a series of cations by analysing, in an empirical way using statistical methods, the geometry of their environments for the well-refined structures present in the literature. The first cation selected for this inquiry is calcium since it presents a large variety of CN's.

The goal of this work is to find a statistical test, based purely on structural information, which would help in evaluating CN for the uncertain cases concerning polyhedra where calcium is bonded to oxygen. For this purpose: (*a*) only very accurate structure determinations were selected; (*b*) CN was assigned, for all cases were there was no ambiguity, using a combination of the criteria of the maximum gap of the Ca—O distances and the bond-valence sums for Ca; (*c*) a set of geometrical variables were calculated for each polyhedron and averaged by CN group; (*d*) the assumption was made that, for ambiguous cases, an indication of the most plausible CN can be obtained by comparing the values of the variables calculated for the CN's in doubt with the average values of the groups for which a CN was certainly determined.

Selection of the data

In order to obtain the structural information in a form easily processable by computer and to cover the majority of the inorganic crystal structures available in the literature, the Inorganic Crystal Structure Database (Bergerhoff, Hundt, Sievers & Brown, 1983) was used. The number of structures containing calcium and oxygen, not necessarily bonded to one another, is quite large (1866). It was greatly reduced by careful inspection of the data. This was done in an automatic way at first, using the retrieval program *CRYSTIN*. The structures eliminated were: 417 defective; 69 disordered; 426 representing solid solutions; 91 refined by X-ray powder, neutron powder or electron diffraction methods; 43 not refined at room temperature, in order to have consistent data; and 410 having a final *R* value greater than 10%. Only 410 structures were retrieved.

It is worth noting that the number of entries rejected on the basis of the *R* value is quite large. A trial run was carried out which also included refinements with *R* values between 10 and 15%. However, the great majority of these 'poorly refined' structures showed some fault, mainly unreasonable distances, and are not considered in this work.

Only Ca—O polyhedra are taken into account in the present paper. By inspection of the cation–anion distances selected up to 3·3 Å, 50 structures were eliminated because they contained other anions besides oxygen. Multiple refinements of the same compound (46 cases), were also excluded to avoid sample bias. Another 60 entries did not contain the coordinates, or showed unreasonable distances, and therefore an error in the coordinates. At the end of the process only 254 structures (*i.e.* 13·6%) for a total of 368 independent polyhedra passed the screening.* All subsequent calculations were carried out on this set of data. Although this initial selection of the starting set was very laborious and time consuming, it is of paramount importance for a statistical work such as the present one, and the very sharp reduction in the number of 'good structures' shows this.

General comments on the data

The Ca—O distance expected on the basis of the ionic radius sum is 2·40 Å. The grand mean of the average Ca—O distances for the polyhedra with certainly assigned CN (see below) is 2·46 Å.

The PCN for Ca—O calculated on the basis of the radius ratio (Pauling, 1960) is 8·7. Brown (1988) found, by empirical examination of a large number of structures, an average observed CN of 7·31, while Nord & Kierkegaard (1984) using a different set of

---

* A list of references for the structures used has been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 53389 (6 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

data and accepting the CN assignments of the original authors found a value of 7·48. These values compare well with the average CN of 7·43 found in this study.

The CN for Ca ranges from 6 to 12, and the polyhedra are often very irregular. Nord & Kierkegaard (1984) quote two cases of CN = 5. The first concerns Ca_{11} in α-$Ca_3(PO_4)_2$ (Mathew, Schroeder, Dickens & Brown, 1977). The environment of the Ca atom was checked using the program *MOLDRAW* (Ugliengo, Borzani, Chiari & Viterbo, 1989) and two other non-screened O atoms were found at distances of 2·907 and 3·107 Å. This polyhedron was inserted in the uncertain cases (between CN = 6 and CN = 7, third case with COL = 923 in Table 3) and was assigned by discriminant analysis to CN = 7 (see below), therefore including the two atoms with larger distances as well.

As for the other example of CN = 5, namely $CaGa_4O_7$ (Deiseroth & Mueller-Buschbaum, 1971), the *R* value of the refinement is not given and there is a very short Ca—O distance (2·13 Å). A drawing of the polyhedron by *MOLDRAW* shows that it is quite open on the side opposite to the short distance, where there are two O atoms at a distance of 3·24 Å. The suspicion remains that there could be an error in the Ca coordinates. Although the presence of CN = 5 for Ca cannot be excluded, it is not taken into account in this paper.

Experimental method

The CN was assigned on the basis of the two most used criteria: the maximum gap in the Ca—O distances and the bond-valence sums for Ca. For 32 polyhedra it was not possible to assign the CN on this basis, and they were left out of the starting set to be used later as a check of the statistical test still to be found. For some polyhedra (see Table 3) a pronounced gap existed but the first excluded atom was less than 3 Å from the Ca atom. These cases were also considered ambiguous. The sums of bond valence on Ca (Brown & Wu, 1976) were checked in order to see that they did not differ from 2 by more than 0·5 v.u. Some structures containing errors were detected because the bond-valence sums showed abnormal values [a discussion concerning Ca—O bond valence has already been published (Chiari & Ferraris, 1990) and is omitted from this paper]. In this way 336 polyhedra were grouped by an unambiguously assigned CN, corresponding to 2504 individual Ca—O bonds.

Table 1. *Variables related to* Ca—O *distance* (Å), *listed according to CN*

For each variable: first row, the average value by CN; second row, the e.s.d. of the distribution (in parentheses), which gives and indication of the spread of values; third row, the minimum individual value; fourth row, the maximum individual value. Only the 336 polyhedra for which the CN was unambiguously assigned (see text) are considered.

	CN = 6	CN = 7	CN = 8	CN = 9	CN = 10	CN = 11	CN = 12
⟨*d*(Ca—O)⟩	2·371	2·445	2·493	2·546	2·586	2·685	2·647
	(0·03)	(0·03)	(0·04)	(0·03)	(0·04)	–	(0·08)
	2·310	2·391	2·442	2·508	2·538	–	2·549
	2·442	2·528	2·584	2·588	2·621	–	2·777
$\sigma_{\langle d(Ca—O)\rangle}$	0·061	0·114	0·118	0·146	0·186	0·231	0·046
	(0·05)	(0·06)	(0·06)	(0·05)	(0·04)	–	(0·08)
	0·	0·026	0·002	0·070	0·145	–	0·003
	0·180	0·251	0·269	0·250	0·239	–	0·252
$d_{\min}$	2·304	2·322	2·365	2·378	2·362	2·279	2·596
	(0·04)	(0·04)	(0·05)	(0·07)	(0·05)	–	(0·08)
	2·208	2·224	2·205	2·221	2·326	–	2·459
	2·392	2·414	2·490	2·479	2·441	–	2·708
$d_{\max}$	2·458	2·631	2·676	2·756	2·827	2·951	2·691
	(0·10)	(0·13)	(0·12)	0·10)	(0·08)	–	(0·16)
	2·310	2·429	2·446	2·590	2·734	–	2·549
	2·724	2·929	2·927	2·952	2·958		3·042
δ_d	0·155	0·308	0·311	0·378	0·465	0·672	0·096
	(0·12)	(0·15)	(0·15)	(0·15)	(0·11)	–	(0·20)
	0·	0·008	0·003	0·139	0·352	–	0·
	0·442	0·652	0·678	0·669	0·632	–	0·583
No. of cases	79	96	131	17	5	1	8

For each polyhedron several quantities, all related to the Ca—O distances, were calculated, that is: the average Ca—O for each polyhedron, $\langle d(Ca—O)\rangle$; the e.s.d. of the Ca—O distribution, $\sigma_{\langle d(Ca—O)\rangle}$; the minimum and maximum individual Ca—O distance, $d_{\min}$ and $d_{\max}$ respectively; the distance range $\delta_d = d_{\max} - d_{\min}$; and the bond-valence sums on Ca (Brown & Wu, 1976).

For every group of assigned CN a histogram was drawn for each variable to test its distribution, which was found to be close to normal for all cases. All the variables were averaged by group of assigned CN, and the e.s.d. of the distribution was calculated. Table 1 shows, grouped by CN, the values of these averages (e.s.d. of the distribution in parentheses), together with their minimum and maximum values (which obviously refer to different polyhedra), and the number of cases found for each CN. Thus, for example, reported in the third line of the variable $d_{\min}$ are the shortest individual Ca—O distances found in each group of CN (the fourth line shows the largest values of the minimum distances). The largest individual Ca—O distances for each CN are shown in the fourth line of the variable $d_{\max}$. These data can be useful to inorganic and structural chemists as well as to mineralogists, during the preliminary steps of crystal structure determination, or to assess the probability of correctness of a structure refinement.

From an inspection of Table 1, the following observations can be made: the most common CN for calcium is 8 (34% of the total), followed by CN = 7 with 24% and CN = 6 with 20%. Only one case with CN = 11 was found, and it was omitted from all subsequent statistics. This case refers to Ca_3 in a perovskite-type structure: $CsCa_2Nb_3O_{10}$ (Dion, Ganne & Tournoux, 1984). In the same structure there are two other Ca atoms with CN = 8 and one with CN = 9.

The shortest individual Ca—O distance (2·205 Å) found was for CN = 8, while the longest (3·042 Å) was for CN = 12. The $\sigma_{\langle d(\mathrm{Ca—O})\rangle}$ and the δ_d give an idea of the distortion of the polyhedra. The polyhedra with CN = 6 and 12 are the most regular. The Ca—O distances (average, minimum and maximum) increase with CN, as one would expect. Increasing the number of coordinated O atoms obviously increases the dimensions of the polyhedra, and decreases the bond valence of each individual distance. Therefore all the Ca—O distances should lengthen. Plots were drawn for each of these variables, and regression coefficients with respect to CN were calculated. The spread of points was large, and the dependence not linear.

Klebe & Lentz (1988) studied the bonding geometry of Ca in molecular structures. They found only six-, seven- and eightfold coordination in the organic complexes, and the ⟨*d*(Ca—O)⟩ reported by them are systematically smaller than those found in this work (2·322 Å for CN = 6, 2·394 Å for CN = 7 and 2·458 Å for CN = 8).

In contrast, the ⟨*d*(Ca—O)⟩ reported by Nord & Kierkegaard (1984) compare well with those shown in Table 1, namely 2·374 Å for CN = 6, 2·460 Å for CN = 7, 2·497 Å for CN = 8, 2·561 Å for CN = 9, 2·614 Å for CN = 10 and 2·686 Å for CN = 12.

The theoretical Gaussian curves calculated for each CN on the basis of the average value of each variable and the e.s.d. of the distribution, were renormalized taking into consideration the number of polyhedra *N*, which is different for each CN. The formula used is: $Y = [N\mathrm{step}/\sigma_x(2\pi)^{1/2}]\exp[-0{\cdot}5(X - X_{\mathrm{av}})/\sigma_x]^2$, where X_{av} is the mean value, σ_x the e.s.d. of the distribution and the value of step = $(X_{\max} - X_{\min})/N_{\mathrm{bars}}$ was adapted in order that $N_{\mathrm{bars}} = 6$ for all CN (N_{bars} is the number of bars in the histogram). Consequently, the area delimited by each Gaussian is equal to *N*, and the curves can be compared in the same plot. Fig. 1 shows these curves for ⟨*d*(Ca—O)⟩, which show a large degree of overlap (see also Table 1). The same kind of plots were drawn for all the above quoted variables, with similar results. One can conclude that the spread of the ⟨*d*(Ca—O)⟩ and the other related variables is such that these cannot, by themselves, be used in a discriminant analysis to assign CN's in ambiguous cases.

Other variables were then calculated, using a modified version of the computer program *POLYVOL* (Swanson & Peterson, 1980) which finds the triangular faces having anions as vertices. The program checks that no other anion lies on the faces, within a small tolerance; otherwise the face is considered to be a polygon of higher order. By connecting the vertices with the cation one obtains a series of pyramids, whose volume is calculated and added to obtain the total volume of the polyhedron. The variables calculated for each polyhedron were: the volume (Vol); the total surface area ($\mathrm{Area_{tot}}$); the minimum ($\mathrm{Area_{min}}$), maximum ($\mathrm{Area_{max}}$) and average value (⟨Area⟩) of the area of the lateral faces and the e.s.d. of their distribution ($\sigma_{\langle\mathrm{Area}\rangle}$).

It is evident that the actual position of the cation inside the polyhedron is irrelevant to the calculation of these variables which, on the contrary, include information on the position of the O atoms with respect to one another.

Table 2 lists these variables, averaged by group of CN. Vol and $\mathrm{Area_{tot}}$ increase with CN, since they depend on the dimension of the polyhedra. The ⟨Area⟩ decreases for increasing CN, as one would expect, since the larger number of O atoms forces them to be closer to each other. The values of $\sigma_{\langle\mathrm{Area}\rangle}$ are smaller for CN = 6 and 12, indicating that the polyhedra tend to be more regular for these CN, not only with respect to the distribution of the O atoms around Ca (as already pointed out) but also among themselves.

The theoretical Gaussian curves were also plotted for these variables. Fig. 2 shows the plot for the volume. The overlap between the curves is reduced with respect to Fig. 1, and it is almost non-existent for CN = 6, 7, 8, which are the most common (see also Table 2). For $\mathrm{Area_{tot}}$, a very similar plot was obtained. It seems, therefore, that including these variables in a discriminant analysis can improve the results.

Statistical analysis

Linear-correlation coefficients were calculated among all the variables, averaged by CN, including CN itself. The ⟨Vol⟩ is the variable that best

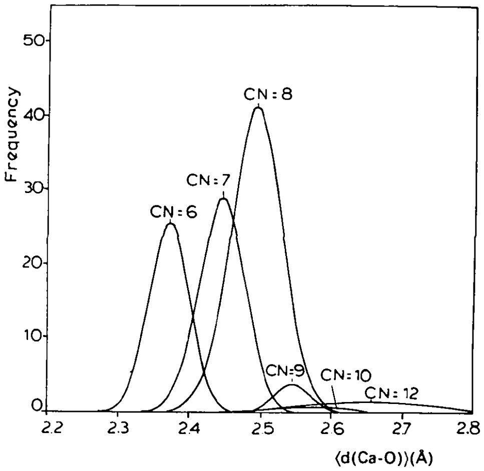

Fig. 1. Plot of the theoretical Gaussian distributions of the variable ⟨*d*(Ca—O)⟩ for the various CN groups. It can be seen that a large overlap exists between the curves.

Table 2. *Polyhedron volumes* (Å^3) *and areas* (Å^2) *according to CN (for the unambiguous cases)*

For each variable: first row, the average value by CN; second row, the e.s.d. of the distribution (in parentheses), which gives an indication of the spread of values; third row, the minimum individual value found in that CN group; fourth row, the maximum individual value.

	CN = 6	CN = 7	CN = 8	CN = 9	CN = 10	CN = 11	CN = 12
Vol	16·76	20·81	26·29	31·29	35·56	42·79	46·05
	(0·57)	(0·78)	(1·25)	(1·55)	(1·16)	-	(3·16)
	15·44	18·99	22·74	27·61	34·41	-	42·01
	18·10	22·81	29·55	33·86	36·92	-	50·90
Area$_{tot}$	37·89	43·91	49·42	54·46	58·82	65·89	66·80
	(0·71)	(0·91)	(1·38	(1·04)	(1·93)	-	(3·93)
	36·30	41·40	46·40	52·80	56·60	-	62·20
	39·50	46·40	53·10	56·30	61·30	-	73·00
Area$_{min}$	4·08	3·70	3·45	3·29	3·12	3·29	3·18
	(0·32)	(0·24)	(0·31)	(0·27)	(0·15)	-	(0·11)
	3·44	3·12	2·44	2·76	2·98	-	3·02
	4·71	4·17	4·10	3·69	3·36	-	3·39
Area$_{max}$	5·43	5·54	5·51	5·20	5·60	8·63	3·50
	(0·33)	(1·06)	(1·44)	(1·13)	(1·52)	-	(0·34)
	4·74	4·58	4·22	4·19	4·35	-	3·11
	6·24	10·60	9·49	7·53	7·46	-	4·05
⟨Area⟩	4·74	4·46	4·17	3·93	3·84	3·88	3·34
	(0·09)	(0·27)	(0·19)	(0·13)	(0·36)	-	(0·20)
	4·54	4·15	3·87	3·77	3·54	-	3·11
	4·94	5·70	4·83	4·29	4·39	-	3·65
$\sigma_{\langle Area \rangle}$	0·49	0·64	0·64	0·58	0·78	1·24	0·16
	(0·18)	(0·45)	(0·40)	(0·26)	(0·37)	-	(0·14)
	0·09	0·18	0·12	0·31	0·46	-	0·03
	0·77	2·70	1·96	1·11	1·38	-	0·43

correlates with CN ($R = 0{\cdot}998$). Fig. 3 shows this plot (the point corresponding to CN = 11 was inserted in the drawing, but not used to calculate the R value). This result means that each addition of an O atom to the polyhedra causes Vol to increase by an equal amount. As we have qualitatively seen, each new O atom entering the coordination sphere forces the other Ca—O distances to lengthen, and the O···O distances to decrease. However, none of these variations are linearly correlated with CN. The volume, which is influenced by their combined effect, on the contrary, shows the above mentioned marked correlation.

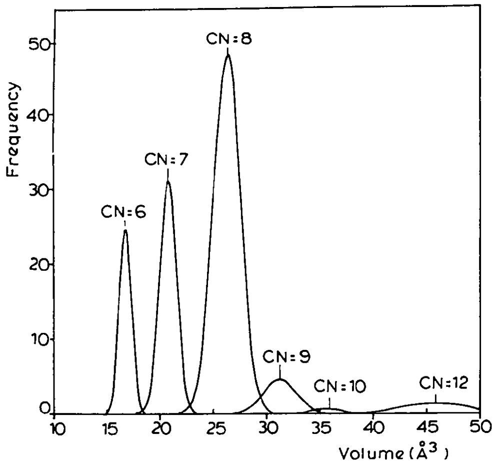

Fig. 2. Plot of the theoretical Gaussian distributions of the polyhedral volume for the various CN groups. The curves are almost completely resolved.

Statistical analyses were carried out using the *SPSS* package (Microsoft Corporation, 1984), on the basic set of 335 polyhedra (CN = 11 excluded) for which CN was unambiguously assigned. This was done to verify whether the selected variables were able to discriminate among the various groups.

A factor analysis was carried out, using all variables (with the obvious exclusion of the assigned CN), to see which were relevant to the analysis. Two factors were selected by the program: the first, grouped together in order of importance Vol, Area$_{tot}$, $\langle d(\text{Ca—O})\rangle$ and d_{min}, and can be considered dependent on the 'dimension' of the polyhedra. The second involved d_{max}, $\sigma_{\langle d(\text{Ca—O})\rangle}$ and δ_d, and is related to the 'distortion' of the polyhedra.

A cluster analysis was carried out using these seven variables. It should be remembered that the clusters found by the procedure were obtained without any information about the preassigned CN. The results were as follows: for CN = 6, 7, 8 and 10 correctly separated clusters were obtained; CN = 9 was not resolved (six cases grouped with CN = 8 and ten cases with CN = 10); CN = 12 gave rise to two clusters, one of five and one of three cases. A second cluster analysis was carried out using only Vol as the grouping variable. The results were the same as before for CN = 6, 7, 10 and 12, while three cases of CN = 8 were grouped with CN = 9; as for CN = 9, it constituted a cluster by itself with the exception of two cases.

A discriminant analysis was carried out using these seven independent variables, and the assigned CN as the group variable. All the polyhedra were assigned to the expected CN group. In a second run, with Vol alone as the independent variable, 98% of the

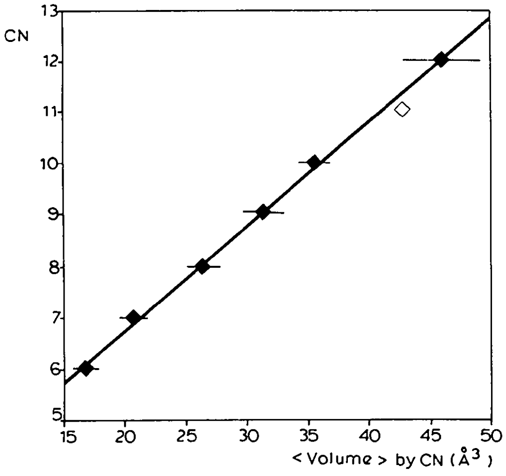

Fig. 3. Plot of CN *versus* the polyhedral volume averaged by CN. Bars represent one e.s.d. The point for CN = 11, referring to one polyhedron only, was not used to calculate the correlation coefficient ($R = 0{\cdot}998$).

polyhedra were properly assigned (three errors for CN = 8, two for CN = 9 and two for CN = 12).

In conclusion, for the cluster and discriminant analyses, the information contained in the volume alone allows the proper assignment in the majority of cases.

Control of the method for the ambiguous cases and conclusions

One can try to assign CN for the ambiguous cases by use of discriminant analyses in which the 'certain' cases characterize the groups by CN, and the 'uncertain' cases are entered as 'ungrouped', *i.e.* without CN.

The normal procedure of attributing the ungrouped cases consists of checking which of the preformed groups each uncertain case is closest to. In most applications the ungrouped cases have only one value of the variables selected for the analysis. In the present study the attribution is complicated by the fact that, for the uncertain cases, one set of variables can be calculated for each possible CN. Each polyhedron is therefore entered at least twice, and for each entry a separate attribution is made to the groups. For each entry the program calculates: $P(G/D)$, the posterior probability that the case belongs to one group rather than to any of the other groups; $P(D/G)$, the conditional probability that the discriminant score of the case belongs to the Gaussian distribution of the group to which the case is assigned.

By inspection of these values assignment of the most plausible CN was attempted. It should be stressed that if the values of the discriminant scores are lower (even though very different) than the average for CN = 6, the $P(G/D)$ is equal to one for CN = 6, while for very large values of the variables, $P(G/D)$ is equal to one for CN = 12. Furthermore, in many cases the $P(G/D)$'s for alternative CN's were either equal or very close to each other. In these instances the attribution of CN was based on the comparison of the $P(D/G)$'s.

A discriminant analysis was carried out, with the above mentioned seven variables and Rao's V method, also known as the Lawley–Hotelling trace (*SPSS*, Microsoft Corporation, 1984, p. B22), which includes the variables using a stepwise algorithm: each variable introduced is that which results in the largest difference among the groups means. This method gives an idea of the importance of each variable to the model. The variables were introduced in the following order: Vol, $\langle d(\mathrm{Ca—O})\rangle$, $\mathrm{Area}_{\mathrm{tot}}$, $d_{\max}$, $\sigma_{\langle d(\mathrm{Ca—O})\rangle}$, δ_d and $d_{\min}$. (The first three variables depend upon the dimensions of the polyhedra, which seem therefore more important than the distortion for the discriminant analysis). A second run was carried out using Vol alone. Table 3 shows the values of the variables for the uncertain cases. For each case there are at least two sets of variables, corresponding to the CN in doubt (in two instances – COL = 4345 and COL = 17034 – there are three lines corresponding to three possible CN). The assignments on the basis of the two discriminant analyses are marked in the third and fourth columns with the symbols: *R* for Rao's method using seven variables, and *V* for Vol alone. These attributions do not always agree with each other. The symbol * indicates that the suggested CN does not correspond to the number of O atoms in the polyhedron. These are the best assignments, since one of the alternatives is obviously wrong. For example, the first case in Table 3 indicates that both discriminant analyses assigned CN = 9 as the most probable (the two * on the first line, which refer to a polyhedron having ten O atoms, indicate that both discriminant analyses assigned it to a CN group different from ten).

For ten cases the results were in contrast. Using the program *MOLDRAW*, these polyhedra were drawn, to check the assignments on the basis of geometrical considerations. For instance, an atom may be screened by two other atoms at shorter distance or, on the contrary, an oxygen with a long Ca—O distance may face a large empty space in the polyhedron. Except for two of them (COL = 9276 and COL = 16039) the indication given by Vol alone was the most plausible.

In conclusion, the statistical method used seems capable of resolving most uncertain cases.

A useful test in assigning CN for uncertain cases found in new structure refinements should be as simple as possible. The discriminant analysis described above is not simple to apply, since it requires the full set of data for the 'certain' cases. On the other hand, it is evident that Vol is by far the most important variable and with a single independent variable the whole procedure is greatly simplified. The linear regression analysis with CN as dependent variable and Vol as independent variable gave a correlation coefficient $R = 0{\cdot}982$ and the following regression equation:

$$\mathrm{CN} = 0{\cdot}197\,(2)\mathrm{Vol} + 2{\cdot}83\,(5). \qquad (1)$$

Using (1), or the plot of Fig. 3, from the volume of a polyhedron containing *n* O atoms, one can derive the expected CN, which is in general non-integer. The most probable CN is the one that gives the smallest absolute value of the difference, $\delta_{\mathrm{CN(Vol)}}$, between the expected CN and the number of O atoms *n*. For the uncertain cases, $\delta_{\mathrm{CN(Vol)}}$ is reported in the last column of Table 3, for all the alternatives. Of course the attributions made using this method coincide with the ones obtained by the discriminant analysis using Vol alone.

Table 3. *List of variables used in the statistical analyses for the* 32 *cases with uncertain CN*

The second and third columns give the assignments of the discriminant analyses (*R* = Rao method with seven variables; *V* = Vol alone; * indicates that the assigned CN is not equal to the number of O atoms). COL is the collection code in the ICSD. In the last column $\delta_{CN(Vol)}$ is the difference between the estimated CN using equation (1) and the number of O atoms in the polyhedra (see text). Distances are given in Å, areas in Å^2 and volumes in Å^3.

CN			COL	$\langle d \rangle$	σ_d	d_{min}	d_{nax}	δ_d	Vol	A_{tot}	$\delta_{CN(Vol)}$
10	*	*	11	2·545	0·167	2·405	2·970	0·565	33·42	55·78	−0·60
9	*R*	*V*	11	2·497	0·078	2·405	2·623	0·218	29·97	52·48	−0·28
8	*R*	*V*	13	2·557	0·206	2·369	3·013	0·644	26·45	50·48	0·03
7			13	2·492	0·100	2·369	2·623	0·253	19·50	43·46	−0·34
10	*R*	*V*	131	2·603	0·173	2·479	3·043	0·564	37·13	59·61	0·13
9	*	*	131	2·554	0·082	2·479	2·699	0·220	28·51	53·04	−0·57
9		*	131	2·557	0·204	2·370	3·007	0·637	28·59	53·52	0·55
8	*R*	*V*	131	2·501	0·122	2·370	2·686	0·316	24·13	48·98	−0·43
8		*V*	923	2·519	0·252	2·313	3·009	0·696	25·79	48·75	−0·10
7	*R*		923	2·449	0·169	2·313	2·807	0·494	20·13	42·76	−0·22
8	*R*	*V*	923	2·581	0·253	2·233	3·012	0·780	26·70	51·75	0·08
7			923	2·533	0·224	2·233	2·850	0·617	20·27	44·30	−0·19
7	*R*	*V*	923	2·537	0·327	2·321	3·107	0·786	21·98	45·73	0·15
6			923	2·441	0·229	2·321	2·907	0·586	14·35	37·75	−0·35
8			1017	2·539	0·266	2·265	3·003	0·737	27·70	51·60	0·27
7	*R*	*V*	1017	2·473	0·204	2·265	2·901	0·636	21·04	43·97	−0·04
6			1017	2·400	0·084	2·265	2·507	0·242	14·73	37·49	−0·28
8		*V*	1391	2·492	0·190	2·341	2·939	0·598	27·21	49·92	0·18
7	*R*		1391	2·427	0·062	2·341	2·490	0·149	19·95	43·03	−0·25
7			1841	2·525	0·299	2·266	2·076	0·810	22·51	46·23	0·25
6	*R*	*V*	1841	2·434	0·192	2·266	2·781	0·515	16·80	38·80	0·13
9	*	*	1841	2·694	0·240	2·524	3·014	0·489	39·20	61·81	1·53
6	*R*	*V*	1841	2·534	0·011	2·524	2·544	0·020	16·71	38·87	0·11
8			4345	2·565	0·322	2·251	3·061	0·810	26·58	50·85	0·05
7			4345	2·494	0·273	2·251	3·017	0·766	20·80	45·04	−0·08
6	*R*	*V*	4345	2·407	0·160	2·251	3·706	0·455	16·34	37·61	0·04
7	*R*		9276	2·577	0·302	2·329	3·014	0·685	22·66	47·80	0·28
6		*V*	9276	2·504	0·255	2·329	2·964	0·635	15·77	40·47	−0·07
8		*V*	14296	2·539	0·226	2·314	2·998	0·684	26·55	50·06	0·05
7	*R*		14296	2·474	0·140	2·314	2·737	0·423	20·49	43·60	−0·14
8		*V*	15059	2·543	0·218	2·330	2·985	0·655	26·69	50·22	0·07
7	*R*		15059	2·479	0·135	2·330	2·726	0·396	20·59	43·76	−0·13
9			15519	2·607	0·243	2·316	3·043	0·727	32·85	56·66	0·28
8	*R*	*V*	15519	2·553	0·193	2·316	2·901	0·585	26·97	50·54	0·13
7			16039	2·463	0·221	2·273	2·947	0·674	21·78	44·38	0·11
6	*R*	*V*	16039	2·382	0·062	2·273	2·447	0·174	16·41	37·62	0·05
8	*R*		16039	2·546	0·182	2·278	2·919	0·641	28·79	51·98	0·49
7	*	*V*	16039	2·492	0·109	2·278	2·623	0·325	19·47	42·84	−0·35
9		*V*	16764	2·565	0·252	2·409	3·088	0·679	31·75	55·40	0·07
8	*R*		16764	2·499	0·169	2·409	2·903	0·494	25·82	49·18	−0·10
7			16975	2·457	0·209	2·293	2·894	0·601	20·64	43·67	−0·12
6	*R*	*V*	16975	2·384	0·087	2·293	2·519	0·226	16·20	37·30	0·01
8			16975	2·516	0·232	2·345	2·888	0·543	25·17	49·79	−0·22
6	*R*	*V*	16975	2·392	0·038	2·345	2·428	0·083	17·11	38·58	0·19
9			17034	2·595	0·290	2·333	3·159	0·826	32·98	57·06	0·31
8		*V*	17034	2·525	0·212	2·333	2·995	0·661	26·41	50·16	0·02
7	*R*		17034	2·457	0·102	2·333	2·658	0·325	20·86	44·38	−0·07
9			20196	2·551	0·203	2·359	2·977	0·619	29·63	53·07	−0·35
8	*R*	*V*	20196	2·498	0·134	2·359	2·752	0·393	24·91	48·39	−0·28
8	*R*	*V*	20255	2·579	0·183	2·438	2·864	0·427	28·77	53·04	0·48
6			20255	2·484	0·058	2·437	2·558	0·121	13·62	38·88	0·50
8		*V*	21033	2·495	0·218	2·304	2·981	0·677	26·12	49·05	−0·04
7	*R*		21033	2·426	0·102	2·304	2·599	0·295	20·91	43·19	−0·06
8	*R*	*V*	23641	2·546	0·173	2·416	2·907	0·491	26·84	50·98	0·10
7			23641	2·494	0·101	2·416	2·654	0·238	19·35	43·97	−0·37
9			28425	2·577	0·236	2·415	3·058	0·643	32·26	55·90	0·17
8	*R*	*V*	28425	2·517	0·163	2·415	2·906	0·492	26·47	49·86	0·03
7		*V*	31269	2·455	0·164	2·338	2·805	0·467	21·23	43·62	0·00
6	*R*		31269	2·397	0·060	2·338	2·485	0·147	17·15	38·75	0·20
9			31280	2·562	0·172	2·391	2·964	0·573	33·15	55·74	0·34
8	*R*	*V*	31280	2·512	0·088	2·391	2·633	0·242	27·09	50·00	0·15
8			35085	2·568	0·234	2·344	2·956	0·612	27·63	51·60	0·26
7	*R*	*V*	35085	2·513	0·187	2·344	2·776	0·432	20·88	45·14	−0·07
7		*V*	100074	2·473	0·212	2·290	2·917	0·626	21·49	45·22	0·05
6	*R*		100074	2·399	0·091	2·290	2·563	0·273	16·94	38·20	0·16
12			100082	2·722	0·250	2·448	3·060	0·612	49·23	70·34	0·50
9	*R*	*V*	100082	2·615	0·185	2·448	2·876	0·428	29·71	56·12	−0·33

In conclusion, the simplest suggested test consists of calculating the volume of the polyhedra, and applying either the plot of Fig. 3, or equation (1) to estimate the most probable CN.

This research was supported by grants from the Consiglio Nazionale delle Ricerche (Roma). I am grateful to the student Alessandra Amisano Canesi for helping me in this work, and to Professor Ferraris for reading the manuscript.

References

Altermatt, D. & Brown, I. D. (1985). *Acta Cryst.* **B41**, 240–244.

Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. (1983). *J. Chem. Inf. Comput. Sci.* **23**, 66–69.

Bhandary, K. K. & Girgis, K. (1977). *Monatsh. Chem.* **108**, 341–349.

Brown, I. D. (1988). *Acta Cryst.* **B44**, 545–553.

Brown, I. D. & Altermatt, D. (1985). *Acta Cryst.* **B41**, 244–247.

Brown, I. D. & Wu, K. K. (1976). *Acta Cryst.* **B32**, 1957–1959.

Brunner, G. O. (1977). *Acta Cryst.* **A33**, 226–227.

Brunner, G. O. & Schwarzenbach, D. (1971). *Z. Kristallogr.* **133**, 127–133.

Carter, F. L. (1978). *Acta Cryst.* **B34**, 2962–2966.

Chiari, G. (1988). *Z. Kristallogr.* **185**, 504.

Chiari, G. & Ferraris, G. (1988). *Z. Kristallogr.* **185**, 508.

Chiari, G. & Ferraris, G. (1990). *Z. Kristallogr.* **191**, 39–43.

Deiseroth, H. J. & Mueller-Buschbaum, H. (1971). *Z. Anorg. Chem.* **382**, 149–156.

Dion, M., Ganne, M. & Tournoux, M. (1984). *Rev. Chim. Miner.* **21**, 92–103.

Donnay, G. & Allmann, R. (1970). *Am. Mineral.* **55**, 1003–1015.

Einspahr, H. & Bugg, C. E. (1980). *Acta Cryst.* **B36**, 264–271.

Fischer, W., Koch, E. & Hellner, E. (1971). *Neues Jahrb. Mineral. Monatsh.* pp. 227–237.

Frank, F. C. & Kasper, J. S. (1958). *Acta Cryst.* **11**, 184–190.

Gelato, L. M. (1981). *J. Appl. Cryst.* **14**, 151–153.

Hoppe, R. (1979). *Z. Kristallogr.* **150**, 23–52.

Klebe, G. & Lentz, A. (1988). *Z. Kristallogr.* **185**, 227.

Mathew, M., Schroeder, L. W., Dickens, B. & Brown, W. E. (1977). *Acta Cryst.* **B33**, 1325–1333.

Microsoft Corporation (1984). *SPSS.* Microsoft Corporation, Chicago, USA.

Nord, A. G. & Kierkegaard, P. (1984). *Chem. Scr.* **24**, 151–158.

O'Keeffe, M. (1979). *Acta Cryst.* **A35**, 772–775.

Pauling, L. (1929). *J. Am. Chem. Soc.* **51**, 1010–1026.

Pauling, L. (1960). *The Nature of Chemical Bond.* Ithaca: Cornell Univ. Press.

Pfeiffer, P. (1915). *Z. Anorg. Chem.* **92**, 376–380.

Pfeiffer, P. (1916). *Z. Anorg. Chem.* **97**, 161–174.

Pyatenko, Y. A. (1973). *Sov. Phys. Crystallogr.* **17**, 677–682.

Sandomirskiy, P. A. & Baturin, S. V. (1985). *Geokhimiya*, **2**, 234–247.

Shannon, R. D. & Prewitt, C. T. (1969). *Acta Cryst.* **B25**, 925–946.

Swanson, D. K. & Peterson, R. C. (1980). *Can. Mineral.* **18**, 153–156.

Ugliengo, P., Borzani, G., Chiari, G. & Viterbo, D. (1989). XIIth *European Crystallographic Meeting, Moscow*, Vol. 3, p. 191. Moscow: USSR Academy of Sciences.

Voronoi, G. (1908). *J. Reine Angew. Math.* **134**, 198–287.