American Mineralogist, Volume 85, pages 532–542, 2000

Leverage analysis and structure refinement of minerals

MARCELLO MERLI,[1,2,*] LUCIANO UNGARETTI,[1,2] AND ROBERTA OBERTI[1]

[1]CNR-Centro di Studio per la Cristallochimica e la Cristallografia (CSCC), via Ferrata 1, I-27100 Pavia, Italy
[2]Dipartimento di Scienze della Terra, via Ferrata 1, I-27100 Pavia, Italy

ABSTRACT

Leverage analysis allows detection of the reflections that have the greatest influence on the estimate of a refined variable. We have applied leverage analysis in the structure refinement of some rock-forming minerals, in the attempt of settling the best procedures to obtain accurate and reliable results. Attention has been focused on those variables (e.g., the refined site-scatterings) which are fundamental to the determination of site populations, which in turn are the basis for the study of order-disorder processes and petrogenetic modeling. A garnet (pyrope) and an amphibole (tremolite) were used to test the procedure, being illustrative of different symmetries and structural complexity. As the omission of reflections with high leverage on a particular variable may strongly decrease the accuracy in its estimate, the results indicate that systematic procedures for data truncation [either on the basis of (low or high) $\sin\theta/\lambda$, high $\Delta F/\sigma F_o$ or of low I/σ_I] are potentially dangerous.

The choice of a correctly ionized model is shown to be critical to obtain accurate estimates of the refined site-scatterings. The set of reflections with the highest leverage value with respect to a peculiar group of variables was found to be nearly invariant within isomorphous solid-solutions; this implies that the results of this work are valid throughout the amphibole and garnet compositional spaces. The warnings proposed to the treatment and refinement of X-ray diffracted data may be of general utility in the structure refinement of minerals.

INTRODUCTION

Crystal structure refinement is used to obtain accurate and precise values of atom coordinates, atom displacement parameters, and scattering powers at the structural sites (hereafter abbreviated as site scattering). The last item is particularly important when dealing with minerals, owing to the widespread occurrence of isomorphous substitution at several structural sites. Refined site-scattering values depend on the model used (e.g., use of the scattering factors of the correct atom species and appropriate ionization state) and on the refinement procedure (e.g., the weighting scheme, selection of the "observed" reflections, adequate corrections for absorption and secondary extinction, use of constraints, and/or restraints, control of correlations between atomic-displacement parameters and site scattering, etc.). Additionally, correct determination of site scattering is necessary to obtain reliable site populations.

Studies of cation ordering in rock-forming minerals combine chemical analysis with structure refinement and/or spectroscopic characterization of cation ordering at the short- and long-range level as a function of bulk composition and intensive parameters of crystallization. In particular, variations in the site populations as a function of temperature (T) are currently used as input for thermodynamic and kinetic modeling of ordering processes in minerals.

Convergence in the modeling procedure is particularly sensitive to the starting point, (i.e., to the estimate of the site populations). To give an idea of this dependence, a precision of ±0.0007 (which is the highest precision claimed so far) in the (Mg vs. Fe^{2+}) occupancy of an octahedral site of an orthopyroxene may give calculated cooling rates ranging from 0.06 to 0.6 °C/y (Kroll et al. 1997). Different precisions and systematic errors are inherent in the experimental procedures used. As a consequence, different (but highly precise) strategies used in two laboratories on two orthopyroxene crystals separated from the same portion of a meteorite may result in equilibration temperatures of 388 and 467 °C, respectively, corresponding to cooling rates of 0.2 °C/y and 18 °C/y (Zema et al. 1996; Kroll et al. 1997 respectively). Taking into account all the sources of random and systematic error, Kroll et al. (1997) concluded that these values are to be considered in satisfactory agreement.

Precision and accuracy of estimated site populations are strongly dependent on: (1) the ability to apply the best experimental conditions and procedures for each analytical technique, and (2) systematic biases between the different techniques which have been used for characterization of the crystal composition (most often single-crystal structure-refinement, electron-microprobe analysis, and Mössbauer spectroscopy). In this paper, we focus only on the optimization of the procedures for single-crystal structure-refinement.

*E-mail: merli@crystal.unipv.it

0003-004X/00/0304–532$05.00

THE STATE OF THE ART

Kirfel (1996) reported the results of a Round-Robin on least-squares structure refinement of the same sets of diffracted intensities obtained from an olivine crystal and an orthopyroxene crystal. He concluded that (1) different models and strategies used for the refinement of poor-to-moderate quality data (as those obtained by sealed X-ray tube) give occupancy values accurate within the calculated standard deviation, the latter being too high to allow further thermodynamic calculations; (2) high quality data (in terms of internal consistency and distribution of σ_{F^2}/F^2 vs. F^2) are more sensitive to refinement strategy, and may give highly precise results which are actually significantly biased; (3) as a consequence, several refinements have to be done and compared to obtain results sufficiently accurate to be used to model cation-ordering processes. Recently, Stimpfl et al. (1999) gave a more optimistic diagnosis, showing that the same set of data collected on an orthopyroxene gives site populations in agreement within the estimated standard deviation (esd) when refined with the programs STRUCSY, SHELXL-93, and RFINE90 (STOE-AEDII diffractometer software; Sheldrick 1993; Finger and Prince 1975) according to different procedures. However, they found that the choice of the model and of the refinement strategies is crucial for lowering the esd, and, in particular, they suggested adoption of ionized scattering-factors and of adjustable weighting schemes. These conclusions are in agreement with our experience at CSCC on systematic structure refinement of rock-forming minerals (e.g., Ungaretti 1980; Hawthorne et al. 1995, and references therein).

Kroll et al. (1997) examined the effects of choosing different refinement strategies (weighting schemes, bonding models, selection of "observed" reflections) on the estimate of site occupancies in orthopyroxene. They proposed a novel bivariate analysis of the set of collected intensities; it consists of stepwise truncation of low-order (LOT) or high-order data (HOT) and rejection of the outliers under different $\Delta F/\sigma$ cut-off thresholds. In particular, they suggested that the LOT procedure reduces effects due to chemical bonding and poor extinction correction. The proposed strategy also eliminates correlation between site occupancies and atom-displacement parameters by fixing the B(M2)/B(M1) ratio to a value that provides invariance of the occupancies with respect to the choice of the data subset.

The empirical procedure proposed by Kroll et al. (1997) may yield highly precise results. However, systematic cut-off of categories of reflections (low- or high-order, high $\Delta F/\sigma$) may be in conflict with basic statistical axioms of the least-squares method. Statistical sciences have provided various techniques which can lower the negative influence of outliers, which may result from the non-correctness of the model or from the non-normality of the errors associated to each observation [cf. Prince and Collins (1991) for an exhaustive review].

Maximum-likelihood methods tend to maximize the likelihood function where $\Phi(\mathbf{x})$ is the

$$S = \sum_i \ln\left[\Phi\left(x_i\right)\right]$$

probability density function (pdf) of the error distribution associated with the observations, and $\mathbf{x}$ is the set of the "true" values of the parameters. Maximization of S is a more powerful alternative to the least-squares method.

Robust-resistant methods are most useful when the error distribution is not well known. They are called "robust" if they work well for a great number of error distributions, and "resistant" if the results are not affected by significant fluctuation in the estimation of the parameters. Therefore, least-squares refinement based on a weighting scheme that takes into account the agreement between observed and calculated data [e.g., that proposed by Prince (1982), which is available in the least-squares structure refinement program CRYSTALS (Watkin et al.1996)] can be considered as a robust-resistant method of refinement.

When the correctness of the model is doubtful, any prejudice that can potentially generate outliers can be avoided by means of Maximum-Entropy Methods [MEM; cf. Shannon (1948) and Jaynes (1968) for the theory, and Collins (1982) and Bricogne (1984) for important application in crystallography]. When dealing with the estimation of structural parameters, the MEM approach involves maximization of the function

$$H = -\sum_i p_i \ln\left(\frac{p_i}{m_i}\right)$$

in which a set of experimental observations is given as (normalized) prior proportions m and the calculated values are given as (normalized) p. An advantage of this method with respect to standard least-squares structure refinement is that there is no need of overall scale determination; this is particularly important when the overall scale-factor and the atomic- displacement parameters are strongly correlated.

To sum up, there are several statistical approaches that can be chosen according to the peculiarities of the data set. Therefore, to solve the general problem of detection and correction of outliers, there is no need to apply any empirical strategy to a method that may result in an inadequate answer.

We assume that errors associated with the observations (i.e., the diffracted intensities) are normally distributed in a data set collected from a single-crystal, and that the model of the structure is substantially correct; therefore, there is no need to use a method alternative to least-squares refinement. Nevertheless, some improvements to the strategies in treating and refining diffracted intensities can be obtained by use of leverage analysis (Prince and Nicholson 1985). This procedure should allow us to monitor the influence of each data point (i.e., single reflection) on the results of the structure refinement. In our opinion, this is the correct way to identify potential outliers and to assess the reliability of refinement results.

THE THEORY OF LEVERAGE ANALYSIS AND ITS APPLICATION TO CRYSTAL-STRUCTURE REFINEMENT OF MINERALS

We summarize here the most important features of leverage analysis (Prince and Nicholson 1985). If $\mathbf{y}$ is the experimental-observation vector and $\mathbf{M}(\mathbf{x})$ is the set of equations which represent the model (i.e., $y_i = M_i(\mathbf{x}) + e_i$, where e_i is the experimental error), a matrix $\mathbf{A}$ can be defined such that $A_{ij} = \partial M_i/\partial x_j$. If $\mathbf{W}$ is the positive-definite matrix of the weight, a triangular matrix

U can also be defined such that $\mathbf{W} = \mathbf{U}^T\mathbf{U}$. Under the condition that the errors are not correlated, **U** is a diagonal matrix and $U_{ii} = 1/\sigma_i$. If we consider a linear model $\mathbf{M}(\mathbf{x}) = \mathbf{A}(\mathbf{x})$, which has a minimum at $\mathbf{x}_M = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{y}$, and let **Z** be the matrix **UA**, which has dimensions $n \times p$ (n being the number of observations and p the number of parameters in the model), the relation between the estimate $\mathbf{x}_M$ and the matrix **Z** is $\mathbf{x}_M = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$. The $n \times n$ matrix $\mathbf{P} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ is called the *projection matrix* or the *hat matrix*. It is symmetric, and it is possible to show that are trace (**P**) = p and $0 \le P_{ii} \le 1$; obviously, the average value of P_{ii} is p/n.

Each diagonal element P_{ii} of the matrix **P** is the leverage of the ith observation, and represents the rate of variation of the calculated value of a data point as a function of a change in the observed value. If x_j is the refined variable, the ith reflection introduced in the refinement does not strongly affect the estimate of x_j when P_{ii} is close to zero (or is significantly lower than the average P_{ii} value). Vice versa, when P_{ii} is close to 1, the ith reflection strongly affects the least-squares process.

We can now apply this approach to situations that may be encountered during structure refinement of minerals. (1) The ith reflection has a high leverage with respect to variation of a structural parameter (e.g., an atomic coordinate or a site-scattering value), and the refined model is not able to reproduce this reflection satisfactorily (high ΔF). In this case it is likely that the parameter under consideration has not been correctly estimated. (2) The ith reflection cannot be properly reproduced but has a low leverage with respect to the refined structural parameter. In this case, its presence can be tolerated, as it cannot strongly affect the estimate of the variable. (3) Most of the high-leverage reflections can be satisfactory reproduced; the estimate of the variable of interest is highly reliable.

When the outliers in the refinement have high leverage values, two hypotheses must be investigated: (1) the model has not been correctly defined; (2) there are significant experimental errors which could derive from the Renninger effect, thermal diffuse scattering, low quality of the crystal, etc. [cf. Abrahams (1969) for an extensive review of experimental errors which may occur during X-ray data-collection]. In the second case, the outlier reflections should be re-measured, although they will not improve the agreement if the aberrant F_o values were due to non-transient reasons (e.g., inhomogeneity of the crystal, poor absorption correction). In any case, the reason for the bad agreement should be found by careful consideration of both the data and the model (Prince and Nicholson 1985), and its effect on the model should be evaluated.

Leverage analysis also allows evaluation of the decrease of the variance of the estimate of a given parameter any time a data point (a reflection) is repeated. For the ith data point, if **z** is the ith row of the matrix **Z** defined above, $\mathbf{z}\mathbf{V}_n\mathbf{z}^T = P_{ii}$ (where $\mathbf{V}_n$ is the variance-covariance matrix for an n-reflection model), the updated variance-covariance matrix is $\mathbf{V}_{n+1} = \mathbf{V}_n - \mathbf{V}_n\mathbf{z}^T\mathbf{z}\mathbf{V}_n/(1+P_{ii})$, and $\mathbf{t} = \mathbf{z}\mathbf{V}_n$, $t_j^2/(1+P_{ii})$ is the amount by which the variance of the estimate of a parameter would be reduced by repetition of the ith data point. If we apply this reasoning to the collection of X-ray diffracted intensities, it is evident that we can avoid repeated measurement of reflections with low leverage values; conversely, it is wise to find the reason(s) for high $F_o - F_c$ values for reflections with high leverage values.

To our knowledge, only Hazen and Finger (1989) have used leverage analysis in the structural study of minerals. They made a leverage analysis of the intensities collected for garnets to optimize the data collection by checking that the high-leverage reflections were present in the reduced data sets.

Experimental procedures and results

We applied a leverage analysis to the reflections of high-resolution intensity-data sets collected from two crystals with rather simple compositions: garnet (space group $Ia\bar{3}d$), and amphibole (space group $C2/m$). The crystals were nearly stoichiometric pyrope ($Mg_3Al_2Si_3O_{12}$, a = 11.4573 (9) Å, V = 1504.00 Å^3) and tremolite [$\square Ca_2Mg_5Si_8O_{22}(OH)_2$, a = 9.8359 (3) Å, b = 18.0450 (6) Å, c = 5.2752 (2) Å, β = 104.750 (3)°, V = 905.43 Å^3]. These minerals were used because (1) they are widespread rock-forming minerals that record changes in petrogenetic conditions; (2) they have significantly different symmetry and structural complexity; (3) the expertise on garnet and amphibole crystal-chemistry developed in the last twenty years at CSCC during the refinement of approximately 500 garnet and 1000 amphibole crystals allows straightforward and reliable interpretation of the results. Both the selected crystals were gem-quality, and were ground to spheres of ≈0.4 mm diameter to reduce errors due to inaccurate absorption correction. They were mounted on a Philips PW-1100 four-circle diffractometer and analyzed with graphite-monochromatized Mo$K\alpha$ X-radiation up to θ = 67° (six equivalent reflections) for pyrope and to θ = 62° (two equivalent reflections) for tremolite. Intensities were corrected for Lorentz-polarization and for spherical absorption, and then merged by applying a weight equal to $1/\sigma$ to each equivalent reflection. R_{sym} values were 2.0% for pyrope (1052 unique reflections) and 1.5% for tremolite (6942 unique reflections), respectively. The structure refinements were done with the program CRYSTALS (Watkin et al. 1996), by using the robust-resistant weighting scheme provided by the program. The results are reported in Table 1. After convergence, the refined models were used as input for a locally written full-matrix least-squares program which allows us to store the **W** and **A** matrices defined above and to apply leverage analysis to the results.

Strictly speaking, leverage analysis should be done with respect to changes of a single variable; when a group of variables is allowed to vary, the true effect of each reflection on the estimate of each variable is partly obscured. In fact, the total leverage of the data points (reflections) in each analysis is normalized to one, and a single data point can have a leverage near to 1 for a single variable in the group, and near to 0 for all the other variables, giving an average leverage for the group of variables still higher than the mean value <L> but far lower than if only the first variable were taken into account. Therefore, a correct test should have been done by doing separate leverage analysis for all the reflections (1052 and 6852, respectively) with respect to each of the 22 variables in pyrope and the 120 variables in tremolite. These laborious calculations would probably be also incompatible with a concise and effective discussion of the results. Therefore, we decided to group analogous variables and to do leverage analysis with respect to variation of (1) the overall scale-

TABLE 1. Selected refinement parameters of pyrope* and tremolite†

Site	*x/a*	*y/b*	*z/c*	ion/ss‡	U_{eq}	U_{11}	U_{22}	U_{33}	U_{23}	U_{13}	U_{12}
					pyrope*§						
O	0.0330(1)	0.0503(1)	0.6533(1)	0.72(1)	0.0053	551(4)	674(5)	443(4)	–17(3)	–118(3)	78(3)
X	1/8	0	1/4	12.26(2)	0.0080	490(8)	1063(7)	1063(7)	281(6)	0	0
Y	0	0	0	0.57(1)	0.0038	377(3)	377(3)	377(3)	1(2)	1(2)	1(2)
Z	3/8	0	1/4	0.40(1)	0.0034	304(4)	354(3)	354(3)	0	0	0
					tremolite †‖						
O1	0.1119(1)	0.0857(1)	0.2180(1)	0.83(1)	0.0056	447(4)	607(4)	649(4)	–19(3)	117(3)	–30(3)
O2	0.1187(1)	0.1709(1)	0.7244(1)	0.92(1)	0.0060	459(4)	722(4)	646(4)	–45(3)	140(3)	–37(3)
O3	0.1085(1)	0	0.7155(1)	0.69(1)	0.0073	834(7)	656(6)	733(7)	0	218(5)	0
O4	0.1351(1)	0.2519(1)	0.2069(1)	0.87(1)	0.0070	872(5)	579(4)	860(5)	–83(4)	313(4)	–276(4)
O5	0.3466(1)	0.1344(1)	0.1005(1)	0.79(1)	0.0069	664(5)	921(5)	646(5)	271(4)	165(4)	–65(4)
O6	0.3440(1)	0.1188(1)	0.5891(1)	0.82(1)	0.0068	680(4)	878(5)	592(5)	–256(4)	139(4)	16(4)
O7	0.3377(1)	0	0.2928(1)	0.75(1)	0.0075	787(7)	407(6)	1318(9)	0	294(6)	0
T1	0.2806(1)	0.0839(1)	0.2972(1)	0.30(1)	0.0043	421(2)	403(2)	460(2)	–17(1)	105(1)	–37(1)
T2	0.2884(1)	0.1711(1)	–0.1953(1)	0.35(1)	0.0044	423(2)	464(2)	443(2)	–18(1)	124(1)	–78(1)
M1	0	0.0878(1)	1/2	12.12(6)	0.0058	656(3)	535(3)	577(3)	0	191(3)	0
M2	0	0.1765(1)	0	12.17(6)	0.0059	608(3)	557(3)	628(3)	0	190(3)	0
M3	0	0	0	12.17(7)	0.0058	656(3)	491(3)	576(3)	0	125(3)	0
M4	0	0.2780(1)	1/2	20.10(2)	0.0073	970(2)	605(2)	1011(3)	0	610(2)	0
H	0.196(3)	0	0.764(7)	1.00	0.15(2)						

Note: The anisotropic-displacement parameter is of the form exp $(-2\pi^{2*}(h \cdot h \cdot a^* \cdot a^* \cdot U_{11} + k \cdot k \cdot b^* \cdot b^* \cdot U_{22} + l \cdot l \cdot c^* \cdot c^* \cdot U_{33} + 2\ k \cdot l \cdot b^* \cdot c^* \cdot U_{23} + 2\ h \cdot l \cdot a^* \cdot c^* \cdot U_{13} + 2\ h \cdot k \cdot a^* \cdot b^* \cdot U_{12})$; U_{ij} are multiplied by 10^5.

* DM 3 from Dora Maira (Italy)(no. 44 in the CSCC garnet database).

† 9527 from Val Tremola (Italy) (no. 850 in the CSCC amphibole database).

‡ Refined occupancies of the neutral vs. fully ionized scattering-curve or refined site-scatterings.

§ R = 2.75%; R_w = 1.66%; GoF = 0 .98; overall scale factor = 0.34(1); secondary extinction (Larson 1969) = 52.9(34).

‖ R = 3.16%; R_w = 1.86%; GoF = 1.15; overall scale factor = 1.12(6); secondary extinction = 164.0(40).

factor; (2) the secondary extinction parameter; (3) the atom coordinates which are not fixed by symmetry (3 for pyrope, 31 for tremolite); (4) the anisotropic components of the atom-displacements parameters (13 for pyrope, 74 for tremolite); (5) the refined site-scattering values (4 cation and anion sites in pyrope, 13 in tremolite). This choice allowed us to monitor the effects of the reflections on each type of variable during least-squares refinement and to draw some general conclusions. Separate runs were also done for the refined scattering values at the cation site in pyrope and are discussed below.

RESULTS

For each group of variables, the calculated leverage values have been plotted as a function of the amplitude of the observed structure factor ($|F_o|$) and the resolution ($\sin\theta/\lambda$) of the single reflections (i.e., of the parameters often used for cutting the data off). For clarity, the behavior of the same group of variables against the same parameter in the two minerals (a: pyrope; b: tremolite) are compared in each figure. It must be stressed that the absolute values of the leverage depend both on the structure and on the types and number of variables that have been simultaneously allowed to vary in each calculation. The influence of each reflection on the refinement results can be evaluated only with respect to the average value of the leverage ($\langle L \rangle = p/n$) (which is reported in the captions).

Scale factor and extinction coefficient

In both minerals, the leverage on the scale factor increases parabolically as a function of $|F_o|$ (Figs. 1a and 1b). Conversely, no systematic trend is observed as a function of $\sin\theta/\lambda$, the leverage being high (i.e., far higher than $\langle L \rangle$) for only a fraction of (intense) reflections in the $\sin\theta/\lambda$ ranges 0.2–0.8 in pyrope (Fig. 2a) and 0.05–0.4 in tremolite (Fig. 2b). Similar, although less regular, trends are observed for the extinction coefficient; in this case, a dozen reflections with high $|F_o|$ values and low $\sin\theta/\lambda$ are dominant (Figs. 3a, 3b, 4a, and 4b). This evidence explains why truncation of low-order reflections gives results less sensitive to incomplete correction for extinction (Kroll et al. 1997); however, it also indicates that such truncation may prevent a correct estimate of that variable.

Atom coordinates

Only the oxygen atom occupies a general position in the garnet structure, and thus its three coordinates are the only positional parameter free to vary during structure refinement. There are fewer symmetry constraints present in the amphibole structure, in which 30 to 34 atom coordinates are free to vary, depending on the presence or absence of A-site cations and of splitting at the A and M4 sites.

For pyrope, reflections with the highest leverage on the determination of the oxygen-atom coordinates are irregularly distributed over the whole range of both $|F_o|$ (Fig. 5a) and $\sin\theta/\lambda$ (Fig. 6a). Conversely, for tremolite, the leverage decreases regularly as a function of $\sin\theta/\lambda$ (Fig. 6b), and no interpretable trend is observed as a function of $|F_o|$ (Fig. 5b). The trend in Figure 6a suggests the importance of high-resolution data ($\sin\theta/\lambda_{max} \geq 1.2$ Å^{-1}) when refining the geometrical features of the garnet structure, a peculiarity that had been already observed in a systematic work on garnet crystal-chemistry (Merli et al. 1995). Conversely, a data set of medium resolution ($0.70 < \sin\theta/\lambda_{max} < 0.80$ Å^{-1}) includes all the high-leverage reflections in the case of amphibole (Fig. 6b). Again, this is in agreement with the results obtained during systematic work on amphibole crystal-chemistry at the CSCC: the geometrical model is not significantly modified by introduction of reflections with $\sin\theta/\lambda > 0.7$–0.8 Å^{-1}.

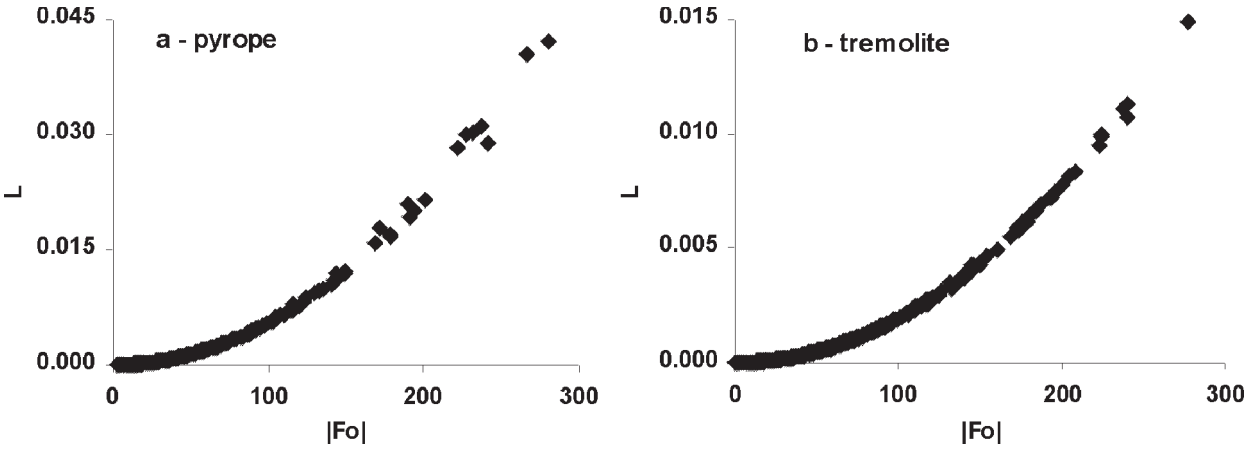

FIGURE 1. The patterns of the leverage values obtained during the refinement of the scale factor plotted as a function of F_o. <L> values are 0.001 and 0.0001, respectively.

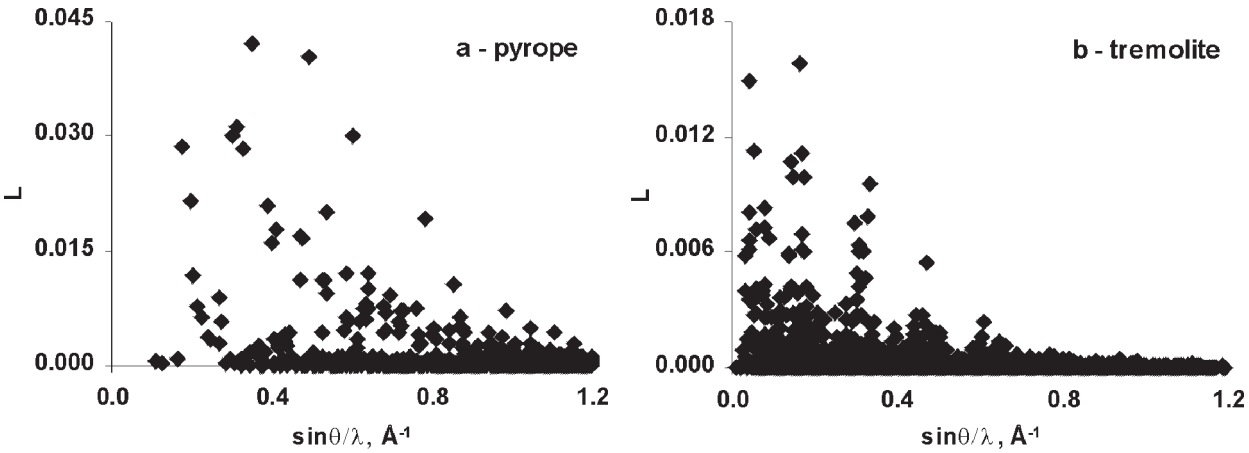

FIGURE 2. The patterns of the leverage values obtained during the refinement of the scale factor plotted as a function of sin θ/λ. <L> values as in Figure 1.

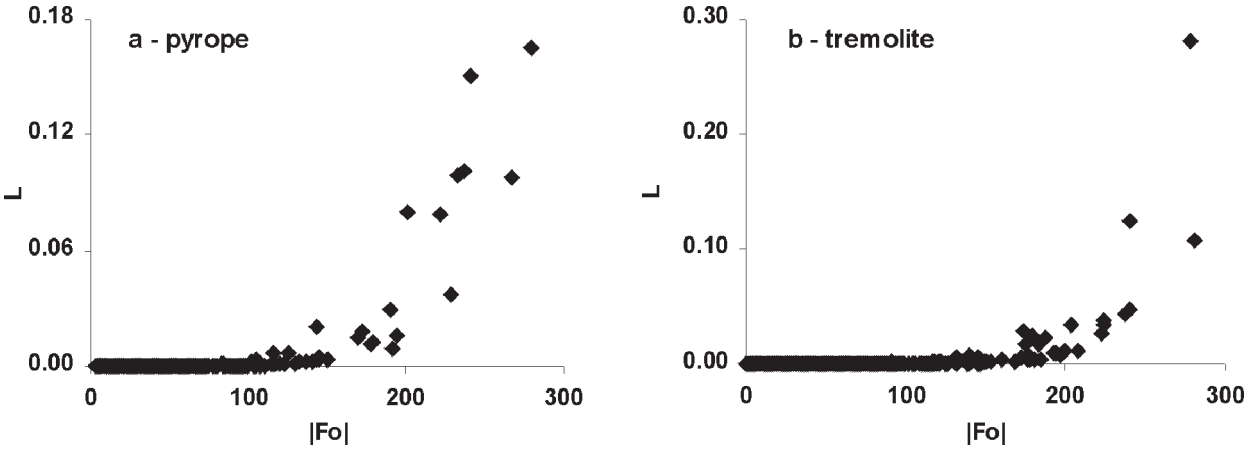

FIGURE 3. The patterns of the leverage values obtained during the refinement of the extinction coefficient plotted as a function of F_o. <L> values are 0.001 and 0.0001, respectively.

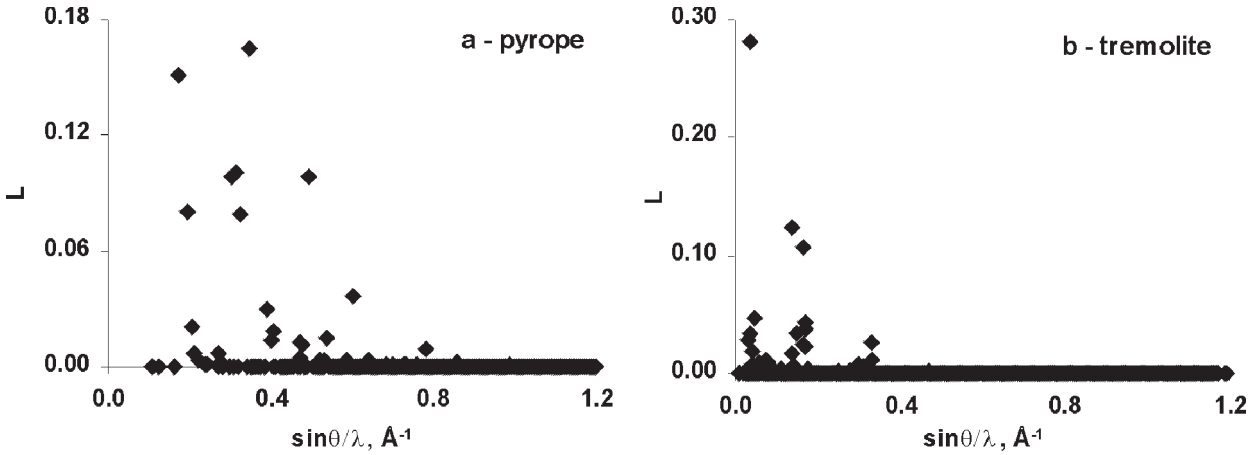

FIGURE 4. The patterns of the leverage values obtained during the refinement of the extinction coefficient plotted as a function of sin θ/λ.

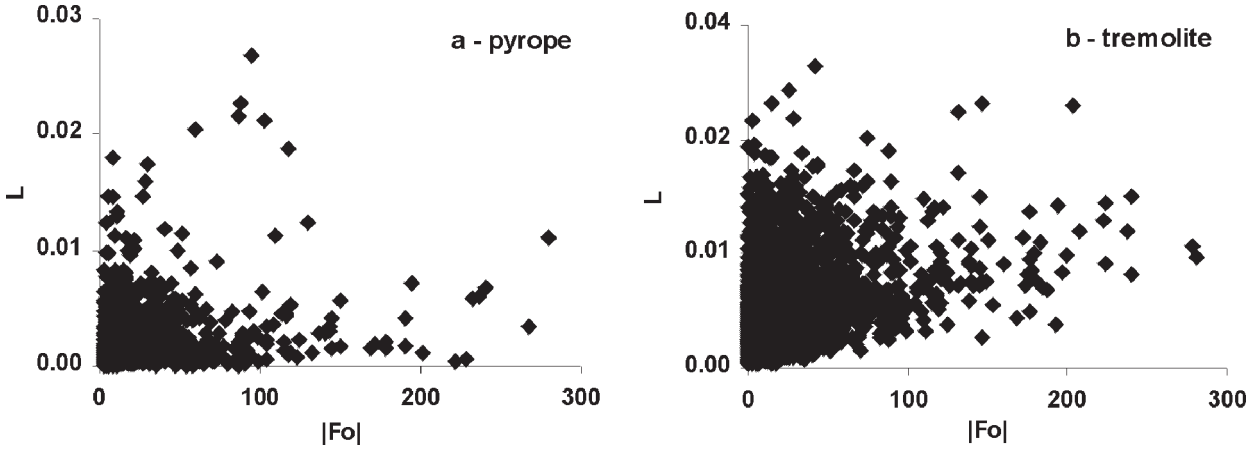

FIGURE 5. The patterns of the leverage values obtained during the refinement of the atomic coordinates plotted as a function of F_o. <L> values are 0.003 and 0.004, respectively.

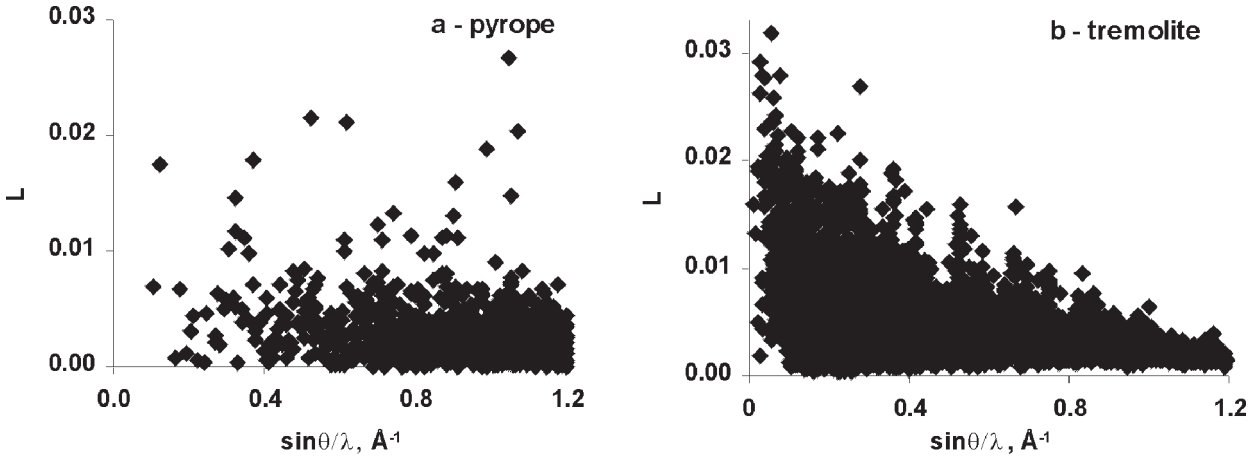

FIGURE 6. The patterns of the leverage values obtained during the refinement of the atomic coordinates plotted as a function of sin θ/λ.

Atom displacement parameters

No clear indication for refinement strategies can be obtained by plotting the leverage as a function of $|F_o|$ (Figs. 7a and 7b); however, it is evident that high-resolution data are critical for accurate refinement of the garnet structure (Fig. 8a) and are less critical for the amphibole structure (Fig. 8b).

Site scattering and site occupancies

The x and y occupancies of two atom species, represented by the appropriate scattering curves, are usually refined under the constraint $x + y = 1$ at each structural site. Sometimes, least-squares programs allow variation of the occupancies of a larger number of scattering curves (or combinations of two scattering curves) during the structure refinement. The selected combination of scattering curves cannot often represent the very complex site populations of minerals; therefore, it is wiser to refer to the physically meaningful quantity which is actually refined, i.e. the scattering power at the structural site (abbreviated as site scattering; Hawthorne et al. 1995). This is actually the quantity that is converted into site populations in the final crystal-chemical formulae, and which must be determined with high precision and accuracy, especially when small variations in cation ordering need to be detected.

In minerals, most of the sites occupied by O and several of those occupied by Si do not show mixed occupancy. For such cases, the occupancies of pairs of scattering curves corresponding to the extreme ionization states (O vs. O^{2-}, Si vs. Si^{4+}) can be refined with the aim of an approximate estimate of the ionization state. This procedure is particularly fruitful with very accurate data, and provides lower disagreement indices and

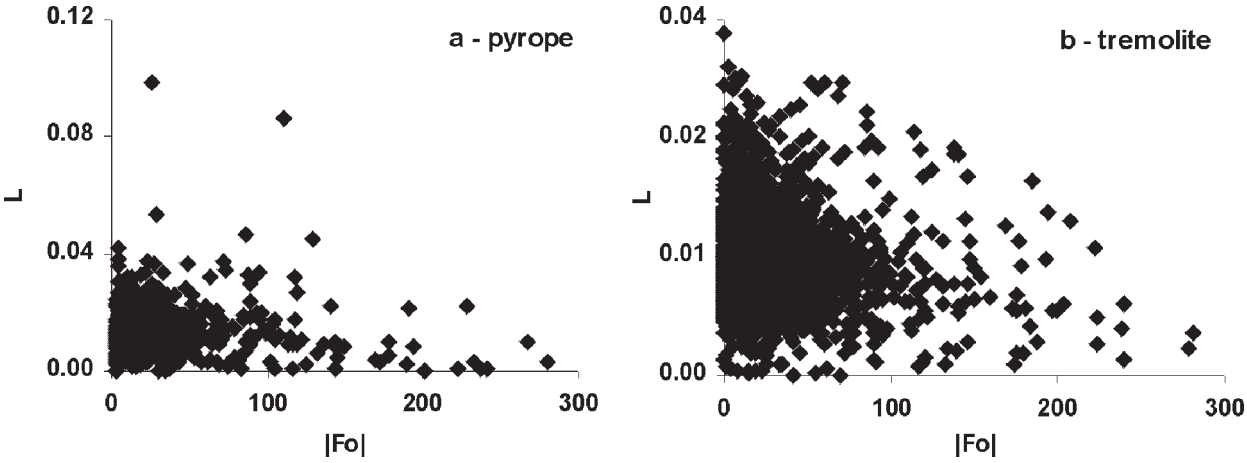

FIGURE 7. The patterns of the leverage values obtained during the refinement of the atom displacement parameters plotted as a function of F_o. <L> values are 0.012 and 0.011, respectively.

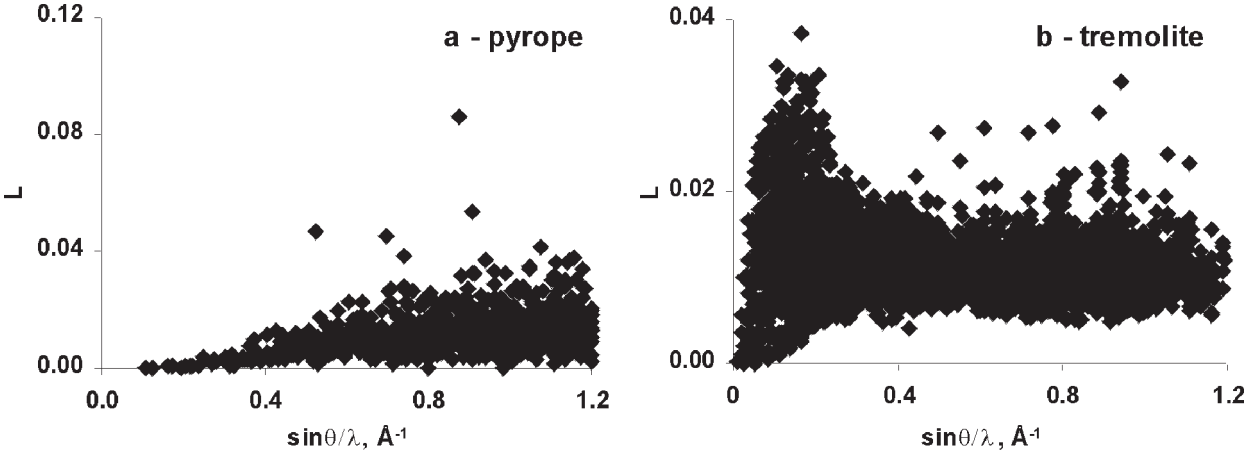

FIGURE 8. The patterns of the leverage values obtained during the refinement of the atom displacement parameters plotted as a function of sin θ/λ.

physically satisfactory models; however, it also increases the degrees of freedom of the least-squares system and, consequently, lowers the standard deviations of other variables (Ungaretti et al. 1983; Hawthorne et al. 1995 and references therein).

The scattering curves for neutral and fully ionized states differ significantly only at very low sin θ/λ values (<0.3 Å^{-1}), and the effects of ionization state are detectable only in that region. Low-θ reflections (several of which are also weak reflections) have the highest leverage in the estimate of site scattering and ionization state (Figs. 9 and 10). The leverage values obtained in this calculation are by far the highest reported in this work, which implies that the estimate of the site scattering may be significantly biased if an incorrect model for ionization is adopted at the O and Si sites. This further confirms that neutral scattering curves must not be used for the assessment of accurate site populations.

Trends in leverage values as a function of changes in a single variable (ionization state of species at the O and Z sites, site scattering at the X and Y sites) were monitored for the pyrope structure (Fig. 11). Distinct sets of reflections at distinct intervals of (low) sin θ/λ contribute to the estimate of ionization state at the O and Z sites. However, the same low (weak-to-medium) reflections contribute to the estimate of site scattering at the X and (especially) the Y sites. The distinct populations in Figures 11c and 11d can by explained by noting that distinct categories of reflections contribute in different ways to the estimate of site scattering, some of them being totally insensitive in this regard.

DISCUSSION

The results of the present work show that the highest leverage values (those with $L > 100 \langle L \rangle$) are obtained when the extinction coefficient (Figs. 3 and 4) and the ionization state

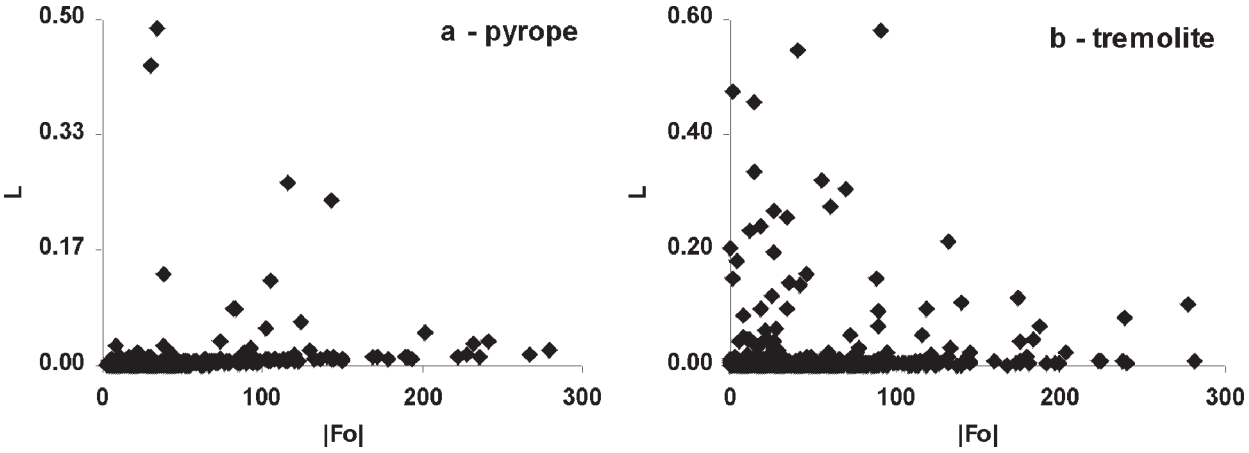

FIGURE 9. The patterns of the leverage values obtained during the refinement of the site scattering plotted as a function of F_o. <L> values are 0.004 and 0.0019, respectively.

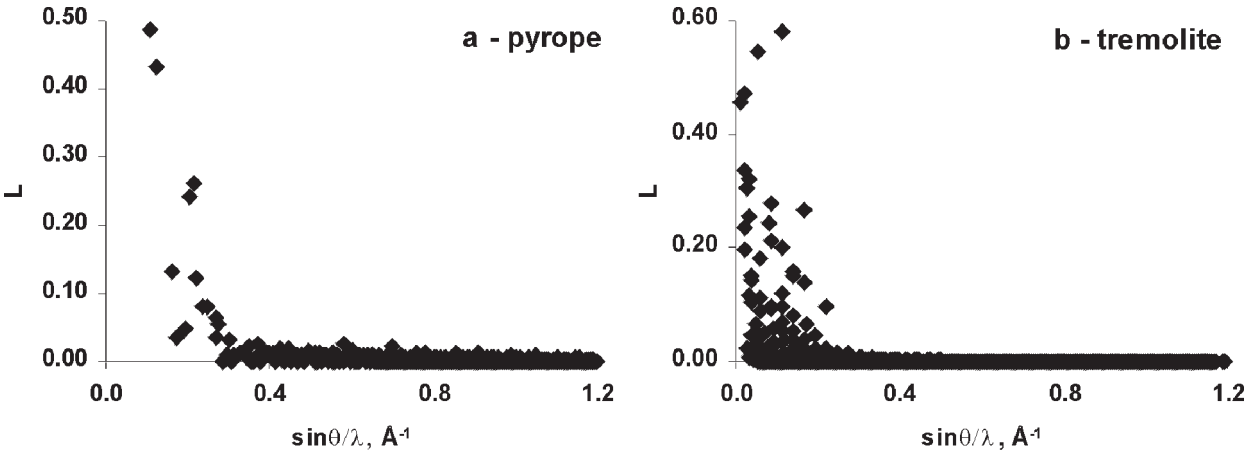

FIGURE 10. The patterns of the leverage values obtained during the refinement of the site scattering plotted as a function of sin θ/λ.

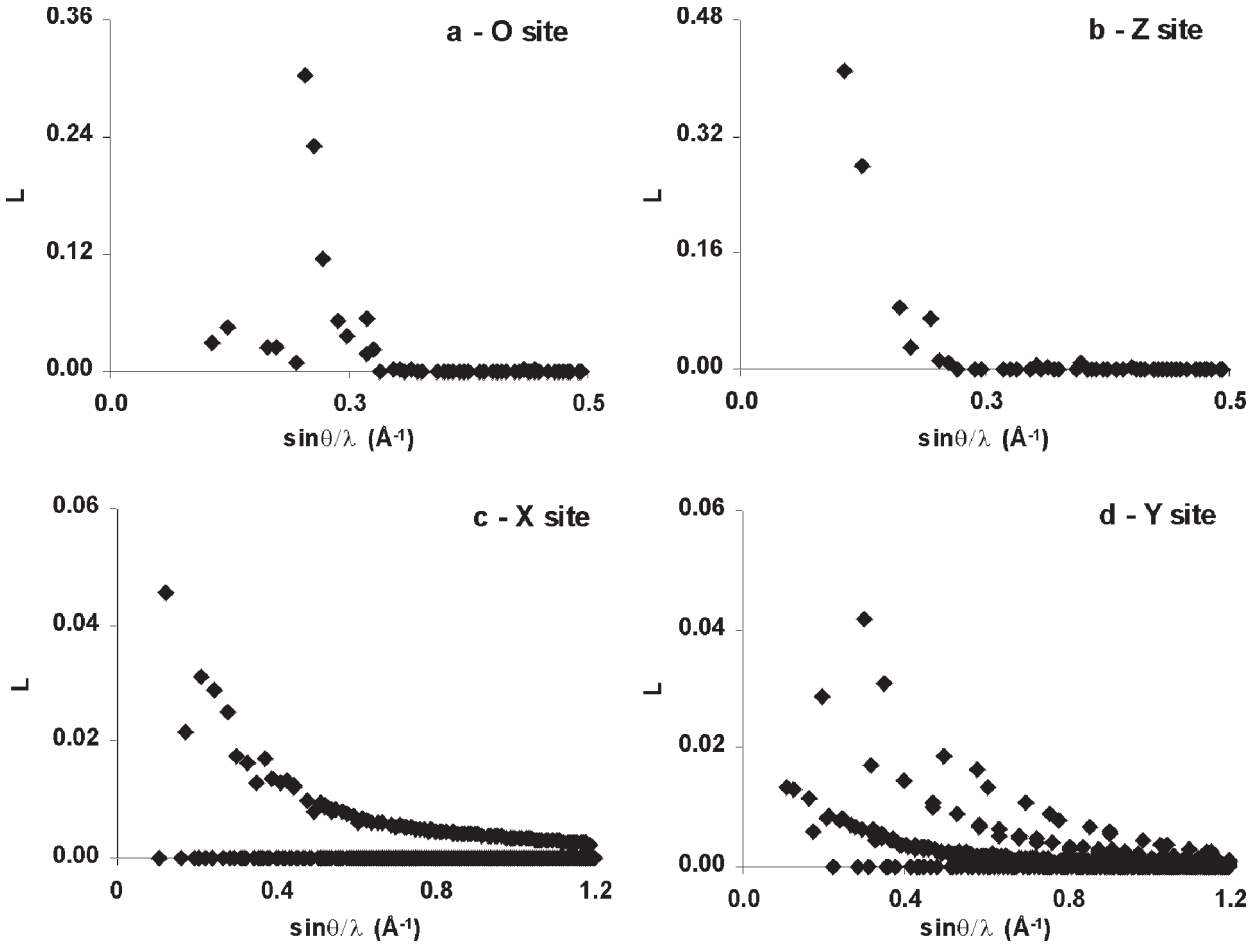

FIGURE 11. The patterns of the leverage values obtained during the refinement of the site scattering in pyrope, plotted as a function of sin θ/λ; (**a**) O site; (**b**) Z site; (**c**) Y site; (**d**) X site. <L> value is 0.001.

(Figs. 9, 10, and 11) are allowed to vary, and that low sin θ/λ reflections are particularly critical in this regard. Some reflections have high leverage values with regard to the estimation of more than one variable (e.g., extinction coefficient and ionization state at O and Z, ionization state at Z and site scattering at Y in pyrope); an incorrect value for one of the two variables could significantly affect the estimation of the other.

We observed a remarkable difference in the trends of the calculated leverage of reflections in pyrope and tremolite, suggesting that different crystal structures are most likely characterized by different leverage patterns. Table 2 provides lists (derived for the pyrope) of the ten reflections with the highest leverage values with respect to the derivation of the scale factor, the extinction coefficient, the oxygen-atom coordinates, the anisotropic-displacement parameters, and the ionization state and site-scattering value in the garnet structure.

To check whether the results obtained with pyrope and tremolite could be of more general use, the leverage analysis was done for garnet crystals within the pyralspite, ugrandite, and schörlomite compositional ranges, obtaining remarkable invariance; only the relative order of importance changes. The same behavior has been noticed for two other amphibole compositions (taramite and richterite). When working on a given mineral, a preliminary mapping of the leverage of the reflections allows us to know a priori whether possible outliers in the structure refinement are likely to affect the estimate of a variable, and thus cannot be discarded but need further and more accurate measurement or correction of the diffracted intensity. It is evident from the above discussion that leverage analysis should be done before planning any refinement strategy which involves truncation of the data, either as a function of $|F_o|$ or/and of sin θ/λ. $\langle L \rangle$ being defined as p/n, at constant number of parameters (p) a decrease in n (the number of reflections used in the refinement) increases $\langle L \rangle$ and thus the leverage of each observed reflection. Leverage analysis is thus even more important when dealing with low-resolution data.

Systematic truncation of the data can eliminate irreplaceable information and artificially enhance the importance of some reflections. If reflections at low sin θ/λ have the highest leverage values (as it is the case for most of the variables ex-

TABLE 2. The reflections which have the highest leverage values with respect to the refinement of the most important parameters in the garnet structure

h	*k*	*l*	F_o	σ_{F_o}	F_c	*R*	I/σ_I	sinθ/λ	Leverage
scale factor (<L> = 1/1052 = 0.00095)									
0	8	0	279.83	2.74	277.59	0.01	51	0.35	0.0423
8	0	8	266.92	5.07	271.54	0.02	26	0.49	0.0404
4	6	0	236.55	5.18	238.68	0.01	22	0.31	0.0312
4	4	4	232.07	0.91	234.50	0.01	127	0.30	0.0302
8	8	8	228.10	1.42	234.06	0.03	80	0.60	0.0300
0	4	0	240.77	2.73	228.89	0.05	44	0.17	0.0287
4	6	2	222.05	4.94	227.34	0.02	22	0.33	0.0283
2	4	0	201.22	3.77	197.41	0.02	26	0.20	0.0214
2	12	2	194.16	1.43	191.30	0.01	67	0.54	0.0201
4	8	0	189.92	2.95	195.31	0.03	32	0.39	0.0209
secondary extinction coefficient (<L> = 1/1052 = 0.00095)									
0	8	0	279.83	2.74	277.59	0.01	51	0.35	0.1654
0	4	0	240.77	2.73	228.89	0.05	44	0.17	0.1508
4	6	0	236.55	5.18	238.68	0.01	22	0.31	0.1014
4	4	4	232.07	0.91	234.50	0.01	127	0.30	0.0989
8	0	8	266.92	5.07	271.54	0.02	26	0.49	0.0984
2	4	0	201.22	3.77	197.41	0.02	26	0.20	0.0800
4	6	2	222.05	4.94	227.34	0.02	22	0.33	0.0787
8	8	8	228.10	1.42	234.06	0.03	80	0.60	0.0370
4	8	0	189.92	2.95	195.31	0.03	32	0.39	0.0296
2	3	3	143.21	1.79	147.25	0.03	40	0.20	0.0208
oxygen coordinates (<L> = 3/1052 = 0.00285)									
0	24	0	94.94	1.77	94.47	0.00	26	1.05	0.0268
16	16	16	87.52	0.44	85.82	0.02	98	1.21	0.0228
0	12	0	86.13	0.52	87.24	0.01	82	0.52	0.0215
10	0	10	102.82	1.79	101.28	0.01	28	0.62	0.0212
10	20	10	58.87	0.56	58.40	0.01	52	1.07	0.0204
16	0	16	117.25	0.34	115.09	0.02	171	0.99	0.0188
6	0	6	7.94	1.55	3.57	0.55	2	0.37	0.0180
2	0	2	29.46	0.29	25.87	0.12	51	0.12	0.0175
12	12	12	28.41	0.53	28.60	0.01	26	0.91	0.0160
10	22	0	7.89	0.99	7.89	0.00	3	1.05	0.0147
thermal parameters (<L> = 13/1052 = 0.0124)									
0	28	0	26.38	0.68	27.10	0.03	19	1.22	0.0984
0	20	0	110.02	1.05	108.31	0.02	52	0.87	0.0863
12	12	12	28.41	0.53	28.60	0.01	26	0.91	0.0538
0	12	0	86.13	0.52	87.24	0.01	82	0.52	0.0469
0	16	0	129.87	0.96	129.75	0.00	67	0.70	0.0453
4	24	4	3.65	1.24	0.98	0.73	1	1.08	0.0418
4	16	4	4.59	1.04	2.22	0.52	2	0.74	0.0383
8	24	8	72.04	0.63	71.51	0.01	57	1.16	0.0375
12	20	12	23.13	1.15	23.13	0.00	10	1.14	0.0373
8	20	0	48.24	1.28	48.84	0.01	18	0.94	0.0367
occupancies (<L> = 4/1052 = 0.00380)									
2	1	1	34.49	0.23	34.89	0.01	75	0.11	0.4871
2	0	2	29.46	0.29	25.87	0.12	51	0.12	0.4329
2	4	2	116.19	1.88	119.31	0.03	30	0.21	0.2629
2	3	3	143.21	1.79	147.25	0.03	40	0.20	0.2404
3	2	1	37.96	0.60	36.90	0.03	31	0.16	0.1331
4	3	1	104.48	1.71	107.45	0.03	30	0.22	0.1235
4	0	4	82.12	0.43	80.53	0.02	95	0.25	0.0816
5	2	1	83.28	0.68	82.90	0.00	61	0.24	0.0807
6	1	1	124.59	1.00	127.48	0.02	62	0.27	0.0640
2	6	0	101.93	1.40	101.65	0.00	36	0.28	0.0541

amined), their truncation would make the remaining reflections less sensitive to variation in the model; this would imply that the consequent small variations in the F_c would not provide clear indication of the best solution.

Truncation of high-order reflections seems to be less critical, especially in complex structures with a large number of variables, such as the amphiboles. Conversely, highly symmetric structures (e.g., garnet) and less complex structures (e.g., olivine, orthopyroxene, as we are presently verifying) do need high-resolution data.

A list of outliers is often obtained at the end of the structure refinement. When the overall disagreement indices are very low (1–2%), the reasons for their incorrect reproduction are experimental errors and/or incorrect or incomplete data-reduction. Thus a possible (and frequent) solution to this problem is their omission from the data set. It is now clear that this procedure can be followed only when the bad reflections do not have high leverage values; however, in this case, they would not have the possibility to affect the results of the refinement. If the outliers have high leverage values, their truncation can be

very deleterious, and the reason for their disagreement must be checked by further measurement and/or improved corrections. Conversely, the satisfactory reproduction of (at least most of the) high-leverage reflections validates the estimate of each variable.

Another possible explanation for deviance of some reflections from their calculated values is obviously related to problems in the refined model; two examples are discussed below to show the importance of an incorrect choice of the model in the case of the pyrope structure.

The refinement was done varying the relative occupancies of the O and O^{2-} scattering curves as described before, and converged to (F_o-F_c) differences for the reflections 211 and 202 (those with the highest leverage with respect to the derivation of the occupancies, Table 2) of 4.5% and –0.1%, respectively. The refinement was also done with only the scattering curve of neutral oxygen, and converged to the same figures of merit (R and goodness of fit, GoF), but with (F_o-F_c) differences equal to 10.6% and –8.1%, respectively, for the same reflections (which therefore became outliers); furthermore, the site scattering at the X site increased from 12.23 to 12.31 electrons. This is a good example of the effect of an incorrect model for ionization state on site-scattering values.

If we remove from the model at convergence the six reflections with the highest leverage, either with respect to the extinction or the occupancies, then further refinements of these two truncated data sets do not diverge significantly. Conversely, if the same reflections are removed before convergence of the least-squares refinement, the truncation of such important data strongly affects the refinement, which ends up with unacceptable ionization states and negative anisotropic-displacement parameters.

This latter example is to be considered as an extreme case, as any (stepwise) truncation of the data is rarely so drastic; however, it gives an idea of the bias that may derive from inappropriate and inopportune truncation of data.

Caveats

We show that the investigation of the different leverage patterns in different mineral structures can provide a powerful tool to improve the strategy for structure refinement and to validate the reliability of its results. Some general caveats are: (1) The highest the symmetry of the structure, the most important is the use of high-resolution data. (2) The scale factor and the anisotropic extinction coefficient are the variables most sensitive to accurate measurement of intense reflections. The exclusion of a single high-leverage intense reflection may lead to their incorrect estimation, which in turn may lead to unpredictable errors in the estimation of other variables. (3) Refining the ionization state at all the sites occupied by a single chemical constituent is critical to obtain reliable results at all the other sites. (4) Site occupancies (and thus site scatterings and ionization states) are particularly sensitive to low-θ reflections. The latter must accurately measured and kept in the refined data set to obtain reliable results especially when thermodynamic and kinetic modeling is provided. (5) Atom-displacement parameters are sensitive to high-θ reflections; the availability of good-quality high-resolution data is particularly critical in high-symmetry mineral structures.

Work in progress focuses on detailed analysis of the effect of truncation of single reflections on each refined variable in olivine and orthopyroxene, the two mineral groups most germane to kinetic and thermodynamic modeling of order-disorder processes at present.

Acknowledgments

This work benefited from funding of the CNR to the CSCC. Marcello Merli also acknowledges funding from the MURST project "Relation between structure and properties in minerals: analysis and applications." Frank C. Hawthorne and Lee A. Groat are gratefully acknowledged for constructive criticisms and editorial assistance.

References cited

Abrahams, S.C. (1969) Indicators of accuracy in structure factor measurement. Acta Crystallographica, A25, 165–173.

Bricogne, G. (1984) Maximum entropy and the foundation of direct methods. Acta Crystallographica, A40, 410–415.

Collins, D. (1982) Electron density images from imperfect data by iterative entropy maximization. Nature, 298, 49–51

Finger, L.W. and Prince, E. (1975) A system of FORTRAN IV computer programs for crystal-structural computations. NBS Tech Note 854.

Hawthorne, F.C., Ungaretti, L., and Oberti, R. (1995) Site populations in minerals: terminology and presentation of results of crystal-structure refinement. Canadian Mineralogist, 33, 907–911.

Hazen, R.M. and Finger, L.W. (1989) High-pressure crystal-chemistry of andradite and pyrope: Revised procedures for high-pressure diffraction experiments. American Mineralogist, 74, 352–359.

Kirfel, A. (1996) Cation distribution in olivines and orthopyroxenes. An interlaboratory study. Physics and Chemistry of Minerals, 23, 503–519.

Kroll, H., Lueder, T., Schlenz, H., Kirfel, A., and Vad, T. (1997) The Fe^{2+}-Mg distribution in orthopyroxene: a critical assessment of its potential as a geospeedometer. European Journal of Mineralogy, 9, 705–733.

Jaynes, E.T. (1968) Prior probabilities. IEEE Transactions on Systems Science and Cybernetics, SSC-4, 227–241.

Larson, A.C. (1969) The inclusion of secondary extinction in least-squares refinement of crystal strcutures. In F.R. Ahmed, Ed., Crystallographic Computing, 303 p., Munksgaard, Copenhagen, Denmark.

Merli, M., Callegari, A., Cannillo, E., Caucia, F., Leona, M., Oberti, R., and Ungaretti, L. (1995) Crystal-chemical complexity in natural garnets: structural constraints on chemical variability. European Journal of Mineralogy, 7, 1239–1249.

Prince, E. (1982) Mathematical techniques in crystallography and materials science, 427 p., Springer Verlag, New York, USA.

Prince, E. and Collins, D.M. (1992) Other refinement methods. In A.J.C. Wilson, Ed., International Tables of X-Ray Crystallography vol. C, 605–608, Kluwer Academic Publisher.

Prince, E. and Nicholson, W.L. (1985) The influence of individual reflections on the precision of parameter estimates in least-squares refinement. In Structure and statistics in crystallography, 183–195, Adenine Press, Guilderland NY, USA.

Shannon, C.E. (1948) The mathematical theory of communication. Bell Laboratories Technical Journal, 27, 399.

Sheldrick, G.M. (1993) SHELX93. Program for crystal-structure refinement. University of Göttingen, Germany.

Stimpfl, M., Ganguly, J., and Molin, G. (1999) Fe^{2+}-Mg order-disorder in orthopyroxene: equilibrium fractionation between octahedral sites and thermodynamic analysis. Contribution to Mineralogy and Petrology, 136, 297–309.

Ungaretti, L. (1980) Recent developments in X-ray single crystal diffractometry applied to the crystal-chemical study of amphiboles. Godis˘njak Jugoslavenskog centra za kristalografiju, 15, 29–65.

Ungaretti, L., Lombardo, B., Domeneghetti, M.C., and Rossi G. (1983) Crystal-chemical evolution of amphiboles from eclogitised rocks of the Sesia-Lanzo Zone, Italian Western Alps. Bulletin de Minéralogie, 106, 645–672.

Watkin, D.J., Prout, C.K. Carruthers, J.R., and Betteridge, P.W. (1996) CRYSTALS, Issue 10. Chemical Crystallography Laboratory, Oxford, UK.

Zema, M., Domeneghetti, M.C., and Molin, G. (1996) Thermal history of Acapulco and ALHA81261 acapulcoites constrained by Fe^{2+}-Mg ordering in orthopyroxene. Earth and Planetary Sciences Letters, 144, 359–367.

Manuscript received February 17, 1999
Manuscript accepted October 19, 1999
Paper handled by Lee A. Groat