Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 17 January 2011
Accepted 16 June 2011

Iteratively reweighted least squares in crystal structure refinements

Marcello Merli* and Luciana Sciascia

Dipartimento di Scienze della Terra e del Mare (DiSTeM), Università degli Studi di Palermo, Via Archirafi 36, Palermo, I-90123, Italy. Correspondence e-mail: merli@unipa.it

The use of robust techniques in crystal structure multipole refinements of small molecules as an alternative to the commonly adopted weighted least squares is presented and discussed. As is well known, the main disadvantage of least-squares fitting is its sensitivity to outliers. The elimination from the data set of the most aberrant reflections (due to both experimental errors and incompleteness of the model) is an effective practice that could yield satisfactory results, but it is often complicated in the presence of a great number of bad data points, whose one-by-one elimination could become unattainable. This problem can be circumvented by means of a robust least-squares regression that minimizes the influence of outliers. This work is aimed at showing the capability of a robust regression to achieve an higher reliability of the least-squares estimates with respect to the traditional weighted least-squares crystal structure refinement in terms of both accuracy and precision. The results can be considered encouraging and represent a starting point for future developments.

1. Introduction

The current requirement for a high level of reliability of the results of a crystal structure refinement in all investigations involving crystallographic models obliges us to try to improve any protocol usually adopted in any common crystallographic practice.

Over the years, the efforts of a number of investigators have been turned into fruitful suggestions proposed by the International Union of Crystallography aimed at improving each step of the crystallographic analysis of a structure. These efforts have been in many directions, from data-collection techniques to data treatment, from the choice of the minimization function to the algorithm of optimization, and so on.

Within this framework we have undertaken the present work in order to adopt some common robust techniques of refinement, and to improve the least-squares estimates in crystal structure refinement with respect to the commonly used weighted least-squares procedures. This need is even more pressing when dealing with multipole refinements, when subsequent calculation for reconstructing reliable electron densities as a function of the refined multipole parameters is involved.

Robust techniques hinge on some robust statistical estimators based on knowledge of the leverage of each data point, so the present study can be considered as a natural evolution of our previous research (Merli, 2005) involving regression diagnostics based on leverage analysis aimed at achieving a higher accuracy of the estimated variables of the crystal structure refinement.

2. Theoretical background

2.1. The outliers of a structure refinement

It is well known that if a data point has an observed value markedly different from its calculated value, it means that the fitting algorithm is unable to resolve this aberrant discrepancy, and the data point becomes an outlier. The effect of trying to fit an outlier is to make the fits of all other data points a little bit worse, with the consequent introduction of bias into the parameter estimates.

For this reason, a suitable identification of the influential points (*i.e.* the points that remarkably affect the model parameters) is therefore critically important if a highly accurate estimation of the model parameters is required. A reliable tool for detecting the influence of each data entry on the regression is represented by a number of regression diagnostics (Belsey *et al.*, 1980). Such an approach allows a reliable identification and elimination of the actually dangerous outliers, *i.e.* the data points with a large discrepancy between observed and calculated values whose fitting may actually affect some estimates.

We successfully applied this procedure to the crystallographic least-squares refinement (Merli, 2005), and have recently extended this approach to chemical kinetic calculations (Merli *et al.*, 2010), resulting in a significant improvement of any fitted model.

It must be noted that the simultaneous elimination of the outliers detected by means of some suitable diagnostics could be a dangerous practice. This procedure would involve a multi-collinearity analysis, which can be a non-trivial task from several points of view. The progressive one-by-one elimination of the bad observations should be recommended, but it becomes very time consuming, especially when a large number of outliers are present. Besides, this procedure would sometimes be complicated by the appearance of new outliers during the cycles: such situations would make the process too laborious or even impossible.

These considerations led us to test a robust regression procedure aimed at reducing the negative influence on the estimates introduced by those observations with a relatively large discrepancy with respect to the calculated ones, allowing us to simultaneously downweight the bad reflections and avoid their elimination from the data set.

2.2. Robust weights in a crystal structure refinement

The heteroskedasticity of the residuals in a least-squares crystal structure refinement obviously requires a weighted least-squares (WLS) regression using some suitable weighting scheme to be introduced in the loss function. A number of effective weighting schemes are described in the crystallographic literature. For the sake of brevity we address the reader to the work by Spagna & Camalli (1999) and references therein, as well as to the extensive review by Watkin (2008).

Particularly interesting is the robust weighting scheme adopted by Carruthers & Watkin (1979) implemented in the crystallographic least-squares code *CRYSTALS* (Betteridge *et al.*, 2003). This weight is a function of the discrepancy between the observed reflection and the calculated one. If this difference is too large compared with those estimated from the Chebychev fitting of the residuals, the reflection is downweighted. This weighting scheme actually reduces the bad influence of the large-discrepancy outliers on the refinement.

Robust statistics have been previously discussed (for instance, Prince, 1982) in the field of crystallography but, in general, their implementation has been largely heuristic. These statistics are used when it is known that there are rogue values or outliers in the data, since standard least-squares analysis associates a particularly significant penalty with these points. Heuristic robust techniques generally involve a reasonable modification to the least-squares procedure. Prince (1982) and Prince & Nicholson (1985) mentioned two modifications for the least-squares algorithm. Spagna & Camalli (1999) included robust statistics in their analysis of weighting schemes. Box & Tiao (1968) and Sivia (1996), however, have shown that the outlier problem may be developed within a Bayesian approach to produce probability distribution functions that have a well reasoned basis.

2.3. Robust regression techniques

Robust regression techniques are, in principle, both less sensitive to the presence of the outliers and to some departures from general idealized assumptions introduced in the optimization (for example, the normality of the residuals). Obviously, the prediction and the estimation of the model may become biased when these axiomatic assumptions are not met. The advantage in using this kind of approach hinges on the robustness of the statistic estimators involved in this context, that have a lower dependency on the mere discrepancy between observed and calculated data points and allow an effective downweighting of the dangerous outlier.

It should be noticed that robust methods for regression are still not widely used, even if they often yield better results with respect to the least-squares estimation (see Hampel *et al.*, 1986). It is our opinion that the main (historical) reason is that the robust estimation is a very resource-demanding computation. Because of the great increase in computer performance in recent years, however, robust regression should not be considered as an insurmountable obstacle. We hope that these methods will come into wider use in crystallographic practice.

There are a number of robust regression techniques which replace the least-squares loss function with one less influenced by the presence of outliers in the data set and which can be insensitive to departures from the model assumptions. We can summarize the most common estimators used in robust regression as follows:

(i) L-estimators, based on linear combinations of order statistics;

(ii) R-estimators, based on the ranks of the residuals;

(iii) M-estimators, based on maximum-likelihood arguments;

(iv) S-estimators, that minimize a robust M-estimate of the residual scale;

(v) MM-estimators, that build on both M-estimation and S-estimation to achieve a high breakdown point with high asymptotic efficiency.

Let us briefly consider some peculiar features of the regression estimators listed above. The reader may refer to Rousseeuw & Leroy (1987) for an extended review of these arguments.

Among the L-estimators class we recall the least absolute value (LAV) regression, in which the model estimates are found by minimizing the absolute value of the residuals instead of the weighted sum of squares as in the WLS regression. LAV is less affected by the presence of outliers but is not robust in the presence of gross outliers in the data set. Least median of squares (LMS) regression, first introduced by Rousseeuw (1984), in which the loss function is represented by the median of the squared residuals, is another estimator belonging to the L-estimators class, as well as the least trimmed squares (LTS) regression (Rousseeuw, 1985). Both LMS and LTS show some limitations but play a significant role in the calculation of other estimators.

R-estimators (first introduced by Jaeckel, 1972) involve dispersion measures based on linear combinations of ordered residuals (*i.e.* on the rank of the residuals). R-estimators often show an undesirable 'breakdown point', *i.e.* the least number of outliers that affect the estimation, which actually 'breaks down'. M-estimation for regression was introduced by Huber (1964, 1973). This estimator combines the efficiency of the least-squares estimators and the resistance of the LAV estimators.

An M-estimator minimizes a less rapidly increasing function of the residuals, which requires the use of an iterative procedure, since the residuals cannot be found until the model is fitted. The so-called iteratively reweighted least-squares (IRLS) is thus employed, as in this work.

Because of the breakdown points of M-estimators, Hampel (1975) introduced the important concept of the scaling of the residuals. Such an approach was followed by Rousseeuw & Leroy (1987), who proposed the so-called S-estimator. This technique seeks a solution that finds the smallest possible dispersion of the residuals, namely, a robust estimate of the scale (from which the method gets the 'S' in its name) of the residuals.

MM-estimators attempt to combine the robustness and resistance of S-estimation with the efficiency of M-estimation. The method finds a highly robust and resistant S-estimate that minimizes an M-estimate of the scale of the residuals (*i.e.* the reason for the first 'M' in the name of this method). The scale is then kept constant until a close-by M-estimate of the parameters is located (*i.e.* the second 'M' in the name).

3. The robust procedure used in this work

We have implemented an MM robust regression algorithm by:

(i) Performing a usual WLS regression until convergence is reached, using one of the weighting schemes commonly used in crystallographic practice. The preliminary p vector of the regression coefficients and the n vector of the residuals are thus obtained, together with the $(n \times p)$ design matrix of the system, which will be used to calculate robust weights in the following steps.

(ii) Choosing one of the estimators of the dispersion of the residuals among MSE, MAD or Rousseeuw's estimator as reported in §*A*2 (the choice is empirical: in this work MAD has been adopted).

(iii) Calculating the scaled (robust) residual as reported in §*A*3.

(iv) Choosing one of the M-estimator functions as reported in §*A*4.

(v) Performing one cycle of least-squares and restarting from step (ii) until convergence is reached.

(vi) Using some regression diagnostics to detect outliers.

For the sake of clarity, we would like to give more details for steps (iv) to (vi). As for step (iv), note that – depending on the results of the IRLS procedure – an adjustment of the tuning constant c is required (see §*A*4 for further details). Unfortunately, a proper value of the tuning constant can be obtained only through a trial-and-error procedure. In general, a smaller tuning constant tends to downweight large residuals more severely, while a larger tuning constant downweights large residuals less severely. The default tuning constants, as has been proposed by the authors, yield coefficient estimates that are approximately 95% as efficient as least-squares estimates when the response has a normal distribution without outliers. A decrease of c involves a lowering of the asymptotic Gaussian efficiency of the refinement, while an augmentation of c yields an increase of the efficiency, approaching the WLS regression. Thus, if the outlier detection diagnostics reveal the presence of some aberrant reflections, the tuning constant should be lowered until the disappearance of the outliers is achieved. Besides, if the diagnostics do not reveal the presence of outliers, it could be worthwhile increasing the value of c and then comparing some related figure of merit with that obtained using a different tuning constant. The choice of the 'best' setup (*i.e.* the best choice of both weighting function and tuning constant) should involve the use of some statistical criteria for model selection. The problem is that the most widely used model choice criteria, such as Akaike's information criteria (AIC, Akaike, 1973; Burnham & Anderson, 2002), Bayesian information criteria (BIC, Schwarz, 1978; Kass & Raftery, 1995) and the PRESS statistics analysis (Neter *et al.*, 1990), always depend on the scaling of the residuals. Consequently, it would be difficult to compare models with the same weighting function but different tuning constants. The analysis of any scaled residuals (StdRes, StudRes or StuDel, see §*A*3) can be helpful in this kind of problem, as shown in the next section.

Relative to step (v), it should be noticed that in IRLS regressions the convergence is always slower and sometimes oscillating with respect to WLS regression. Thus, it is worth adopting, together with a convergence criterion commonly adopted in crystallographic works (for instance, maximum shift/s.u. < 0.01), some other criterion: in this work, we stopped IRLS when the relative difference in *w*SSE between two consecutive cycles was less than 0.001% for two or three cycles.

The outlier detection [step (vi)] can be performed in a number of ways (Belsey *et al.*, 1980). In this work, the Cook's distance and the COVRATIO estimators have been considered. Cook's distance is a measure of how much all the other residuals would change if the ith observation is deleted from the analysis. Cook's distance is greater than 0, and may be arbitrarily large. COVRATIO examines how the precision of the parameter estimates changes with the removal of the ith observation. A small COVRATIO is bad, since the variance is smaller without the ith observation, whereas a big COVRATIO involves larger variance without the ith observation. In other words, a big COVRATIO just indicates an extremely influential observation, not necessarily one that is dangerously aberrant. However, if the observation also has a high leverage, the precision of the estimates may be worse.

The thresholds adopted in this work for leverage and Cook's distance were $3p/n$ and $4/(n - p - 1)$, respectively, whereas the lower and upper bounds for COVRATIO were $1 - 3p/n$ and $1 + 3p/n$, respectively. Note that for the leverage cutoff we have adopted that introduced by Velleman & Welsch (1981), who suggested that, when $p > 6$ and $n - p > 12$, $3p/n$ is more appropriate than the usual $2p/n$.

The so-called William's graph, in which Cook's distance or alternatively one of the scaled residuals is plotted against leverage and the related thresholds are superimposed, is an effective tool to detect the dangerous outliers. Data points with leverage and Cook's distance greater than the corresponding thresholds are to be considered potentially dangerous outliers. If the correspondent COVRATIO is less than the lower bound, the outlier recognized in the William's graph can be considered as an actually aberrant point, whereas reflections with a COVRATIO value higher than the upper thresholds are not to be considered bad observations, but just highly influential data on the least-squares estimates. This diagnostic, based on the combination of leverage value, Cook's distance and COVRATIO analysis, has been found to be quite effective when dealing with this kind of data.

4. The experiments: results and discussion

Several different crystal structures have been used to test the reliability of the IRLS regression described above, taken from the inorganic, organic and metal–organic databases. For the sake of brevity, we present just three selected examples from the whole set of theoretical cases considered, together with an example of the application of the robust procedure to experimental data. In particular, we analyse here the multipole refinement results for (i) an organic compound, the non-standard amino acid sarcosine (Dittrich & Spackman, 2007), (ii) the natural borate datolite Ca[BOH(SiO_4)] (Ivanov & Belokoneva, 2007) and (iii) an experimental formamidine (Giumanini *et al.*, 1999). These theoretical and experimental case studies can be considered as representative of the overall behaviour of the whole set of structures considered. For cases (i) and (ii), synthetic data sets have been produced on the basis of the model obtained from the structure refinement of the experimental data up to the experimental resolution. Random errors taken by normal populations with mean zero and variances comparable to the experimental ones have been added to the synthetic data points. In particular, a mean variance value has been evaluated for ten different resolution shells, in order to model a noise pertaining to the reality. It can be noticed that we observed that the higher the bias of the data, the higher is the effectiveness of the IRLS procedures with respect to the WLS refinements. IRLS becomes useless if the residuals are homoskedastic.

The multipole refinements have been carried out using the *XD* program (Koritsanszky *et al.*, 1995). The WLS regressions have been performed using the weighting scheme implemented in *XD* (with the coefficient *a* adjusted to correct the goodness-of-fit value and the coefficient *f* set to 1/3; see the *XD* manual for further details), and the IRLS method has been implemented in the *XD* code.

Once convergence has been reached, calculation of the leverage and some related diagnostics as described above has been performed.

Reflections with both Cook's distance and COVRATIO values outside the suggested thresholds and, simultaneously, a high leverage have been considered 'dangerous outliers' and progressively eliminated from the data sets.

For cases (i) and (ii), the results of each refinement have been evaluated by averaging the absolute values of the discrepancies between the calculated parameters and those of the reference model, both for the whole set of the parameters and for each class of variables (*i.e.* atom coordinates, atomic displacement parameters, κ shrinking factors, multipole population coefficients and overall scale factor). In all cases the precision of the estimates has been evaluated by calculating the average value of both the standard deviations associated with all the variables and with each class of variables as described above.

Table 1
Selected figures of merit for WLS and IRLS refinements of sarcosine.

Weighting function	WLS	Huber (c = 1.345)	Huber (c = 0.100)	Logistic (c = 1.205)	Logistic (c = 0.200)
Gaussian efficiency	100%	95%	67%	95%	73%
$R(F_o)$	0.0160	0.0145	0.0143	0.0145	0.0143
$R(F_o^2)$	0.0256	0.0232	0.0232	0.0232	0.0233
wSSE	2371.22	12.40	1.28	10.69	2.49
$\mathrm{MSE}^{1/2}$	0.9410	0.0680	0.0219	0.0632	0.0305
R	0.9994	0.9998	1.0000	0.9998	0.9999
R^2	0.9989	0.9996	0.9999	0.9997	0.9998
adjR^2	0.9989	0.9996	0.9999	0.9996	0.9998
Max. shift/s.u.	0.0000	0.0058	0.0103	0.0061	0.0093
r.m.s. (shift/s.u.)	0.0000	0.0270	0.0327	0.0278	0.0374
⟨shift/e.s.d.⟩	0.0000	0.0014	0.0020	0.0015	0.0024
ME	−0.0077	−0.0020	−0.0022	−0.0021	−0.0021
σ^2	0.0108	0.0102	0.0107	0.0101	0.0107
MAE	0.0622	0.0560	0.0552	0.0559	0.0553
MARE	0.0210	0.0184	0.0180	0.0184	0.0181
MEDE	−0.0077	−0.0023	−0.0007	−0.0023	−0.0011
MAD	0.0617	0.0560	0.0553	0.0559	0.0553
SAE	175.95	158.64	156.40	158.50	156.60
AIC	19176.4	7350.5	5183.3	7022.6	5442.1
PRESS	3931.4	15.6	1.4	13.3	2.8
predR^2	0.9982	0.9995	0.9999	0.9996	0.9998

Moreover, for cases (i) and (ii) the relative absolute difference between the discrepancy measures related to each IRLS run and those for the WLS has been evaluated, in order to have an idea about the 'gain' in accuracy with respect to the traditional WLS when a robust technique is adopted. Similarly, the relative absolute difference between the ⟨s.u.⟩ evaluated in WLS and each IRLS run has been considered for all cases.

4.1. Synthetic sarcosine

A synthetic set of 2831 structure factors up to a reciprocal resolution of $\sin(\theta)/\lambda = 1.18$ Å^{-1} was generated on the basis of the model given by Dittrich & Spackman (2007). The errors added to each reflection (taken from a normal population with zero mean and the variance observed in the experimental data) involved a mean discrepancy between the theoretical $|F_o|$ and the synthetic noise-induced $|F_c| \simeq 1.6\%$.

In order to facilitate a comparison between WLS and all of the IRLS procedures considered, the results of both methods have been summarized in Table 1, where some selected figures of merit related to the refinements are shown. Table 2 gives the overall discrepancies between the theoretical and the refined parameters, as well as the departures from the theoretical values of each class of variables, as explained above. Fig. 1 shows the comparison of the William's graph and the COVRATIO *versus* leverage plot between the WLS run (Figs. 1*a* and 1*b*, respectively) and the best robust run (Huber weights with c = 0.1; Figs. 1*c* and 1*d*, respectively). Fig. 2 compares the StudRes observed in the WLS case (Fig. 2*a*) with that of the best robust run (Fig. 2*b*).

Table 2
Absolute discrepancy measurements between each class of the theoretical and refined parameters for sarcosine.

Percentages in parentheses represent the relative improvement with respect to the WLS results. all = summation over the discrepancies related to the whole set of refined variables; coord = summation over the discrepancies related to the atom coordinates; a.d.p. = summation over the discrepancies between the a.d.p.'s; κ = summation over discrepancies between the κ values; mult = summation over discrepancies between the multipole terms; scale = discrepancy between theoretical and refined scale factor; $\max(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm var}$ = maximum absolute discrepancy between the true and calculated values for the class of variables denoted by 'var'; $\min(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm var}$ = minimum absolute discrepancy between the true and calculated values for the class of variables denoted by 'var'.

Weighting function	WLS	Huber (c = 1.345)	Huber (c = 0.100)	Logistic (c = 1.205)	Logistic (c = 0.200)
$(\Sigma\vert x_i^{\rm true} - x_i^{\rm calc}\vert/p)_{\rm all}$	0.0175	0.0158 (10%)	0.0144 (18%)	0.0159 (10%)	0.0146 (17%)
$\langle {\rm s.u.}\rangle_{\rm all}$	0.0086	0.0065 (25%)	0.0041 (52%)	0.0064 (26%)	0.0048 (45%)
$(\Sigma\vert x_i^{\rm true} x_i^{\rm calc}\vert/p)_{\rm coord}$	5.08E-05	2.9E-5 (43%)	2.1E-05 (59%)	2.8E-05 (45%)	2.0E-05 (61%)
$\langle {\rm s.u.}\rangle_{\rm coord}$	5.66E-05	3.4E-05 (40%)	1.7E-05 (69%)	3.2E-05 (43%)	2.1E-05 (63%)
$\max(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm coord}$	1.7E-04	8.3E-05	5.6E-05	7.6E-05	5.2E-05
$\min(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm coord}$	8.0E-06	1.0E-06	2.0E-06	0.0	0.0
$(\Sigma\vert x_i^{\rm true} - x_i^{\rm calc}\vert/p)_{\rm a.d.p.}$	0.0013	0.00135 (9%)	0.00090 (28%)	0.00138 (11%)	0.00098 (22%)
$\langle {\rm s.u.}\rangle_{\rm a.d.p.}$	0.0011	0.00082 (25%)	0.00052 (52%)	0.00082 (26%)	0.00060 (45%)
$\max(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm a.d.p.}$	2.7E-02	1.3E-02	1.3E-02	1.3E-02	1.4E-02
$\min(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm a.d.p.}$	1.2E-05	0.0	0.0	0.0	3.0E-06
$(\Sigma\vert x_i^{\rm true} - x_i^{\rm calc}\vert/p)_{\kappa}$	0.0076	0.0013 (83%)	0.0010 (87%)	0.0008 (90%)	0.0009 (88%)
$\langle {\rm s.u.}\rangle_{\kappa}$	0.0029	0.0017 (44%)	0.0010 (66%)	0.0016 (45%)	0.0012 (60%)
$\max(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\kappa}$	1.1E-02	2.5E-03	2.0E-03	1.5E-03	1.7E-03
$\min(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\kappa}$	4.2E-03	3.0E-06	3.0E-06	5.0E-06	5.0E-06
$(\Sigma\vert x_i^{\rm true} - x_i^{\rm calc}\vert/p)_{\rm mult}$	0.0296	0.0266 (10%)	0.0244 (18%)	0.0267 (10%)	0.0247 (16%)
$\langle {\rm s.u.}\rangle_{\rm mult}$	0.0143	0.0107 (25%)	0.0068 (53%)	0.0106 (26%)	0.0079 (44%)
$\max(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm mult}$	1.2E-01	1.7E-01	1.5E-01	1.8E-01	1.5E-01
$\min(\vert x_i^{\rm true} x_i^{\rm calc}\vert)_{\rm mult}$	5.0E-05	6.2E-04	1.3E-04	1.8E-05	9.7E-05
$(\vert x_i^{\rm true} - x_i^{\rm calc}\vert)_{\rm scale}$	4.16E-03	3.0E-06 (100%)	3.0E-06 (100%)	5.0E-06 (100%)	5.0E-06 (100%)
${\rm s.u.}_{\rm scale}$	2.13E-03	1.2E-03 (45%)	6.3E-04 (70%)	1.1E-03 (46%)	7.6E-04 (65%)

4.1.1. WLS regression. The initial WLS crystal structure refinement on the synthetic set of structure factors was carried out using the weighting scheme implemented in the *XD* code, with the coefficient a set to 0.0 and the coefficient f set to 1/3. The refinement of 153 model parameters gave final $R(F_o)$ = 0.0160 and $R(F_o^2)$ = 0.0256 (Table 1). The overall measure of the discrepancy between the refined parameters and the theoretical ones was 0.0175, corresponding to a relative absolute error on the estimates ~5%, while the mean s.u. was 0.0086 (Table 2).

Once the convergence criterion was satisfied, the regression diagnostics were calculated.

For this structure the thresholds for the estimators listed above were 0.16 for the leverage, 0.0014 for the Cook's distance, 0.84 for the lower bound of the COVRATIO estimator and 1.19 for the upper bound of COVRATIO.

The regression diagnostics revealed the presence of 63 influential outliers. This is a typical case in which the one-by-one elimination of the outliers is unsuitable because of the large number of outliers to be eliminated and because of the usual appearance of new outliers during the stepwise elimination of the aberrant points. Figs. 1(*a*), 1(*b*) and Fig. 2(*a*) allow a visual analysis of the outliers. In particular, the points possibly lying in the lower-right quadrant in Figs. 1(*b*) and 1(*d*) represent the actual outliers of the refinement.

4.1.2. IRLS regression. In this work all of the weighting functions listed in §*A*3 have been tested for a number of structures. In this paper we report just the results for the Huber and the logistic weight, since they have been observed to be the most effective in terms of robustness, rate of convergence and precision.

The figures of merit reported in Table 1 (columns 2–6) clearly show a general improvement with respect to the WLS procedures when both of the robust regression techniques are adopted.

In all of the IRLS cases a decrease of 9–11% of the value of the crystallographic $R(F_o)$ and $R(F_o^2)$ factors has been observed. The slight improvement of the values of the R, R^2 and adjR^2 factors also indicates either a greater capability of the model to explain the variance of the dependent variables involved, or, equivalently, a reduction of the errors associated with each dependent variable.

Some figures of merit based on the unweighted residuals such as ME, σ^2, MAE, MARE, MAD and SAE listed in Table 1 are also useful in model choice, being independent of the weights. These unweighted estimators have been taken into account because the weighted estimators harm the intuitive appeal of a measure of the actual error. As can be seen, the comparison between IRLS and WLS always indicates a significant improvement if a robust regression is performed.

The improvement of the figures of merit is in agreement with the overall 'gain' in accuracy of the estimates with respect to the WLS case, which ranges from ~10% up to ~18% (Table 2). The most important contribution to the overall gain in accuracy is due to the improvement of the multipole parameters that involve the greater number of variables (88 parameters on 153). A very high level of improvement is observed for the atom coordinates (up to 60%), atomic displacement parameters (~28%), κ factors (~87%) and overall scale factor (~99%). The last two classes of variables are typically affected by the strongest error in the structure refinements.

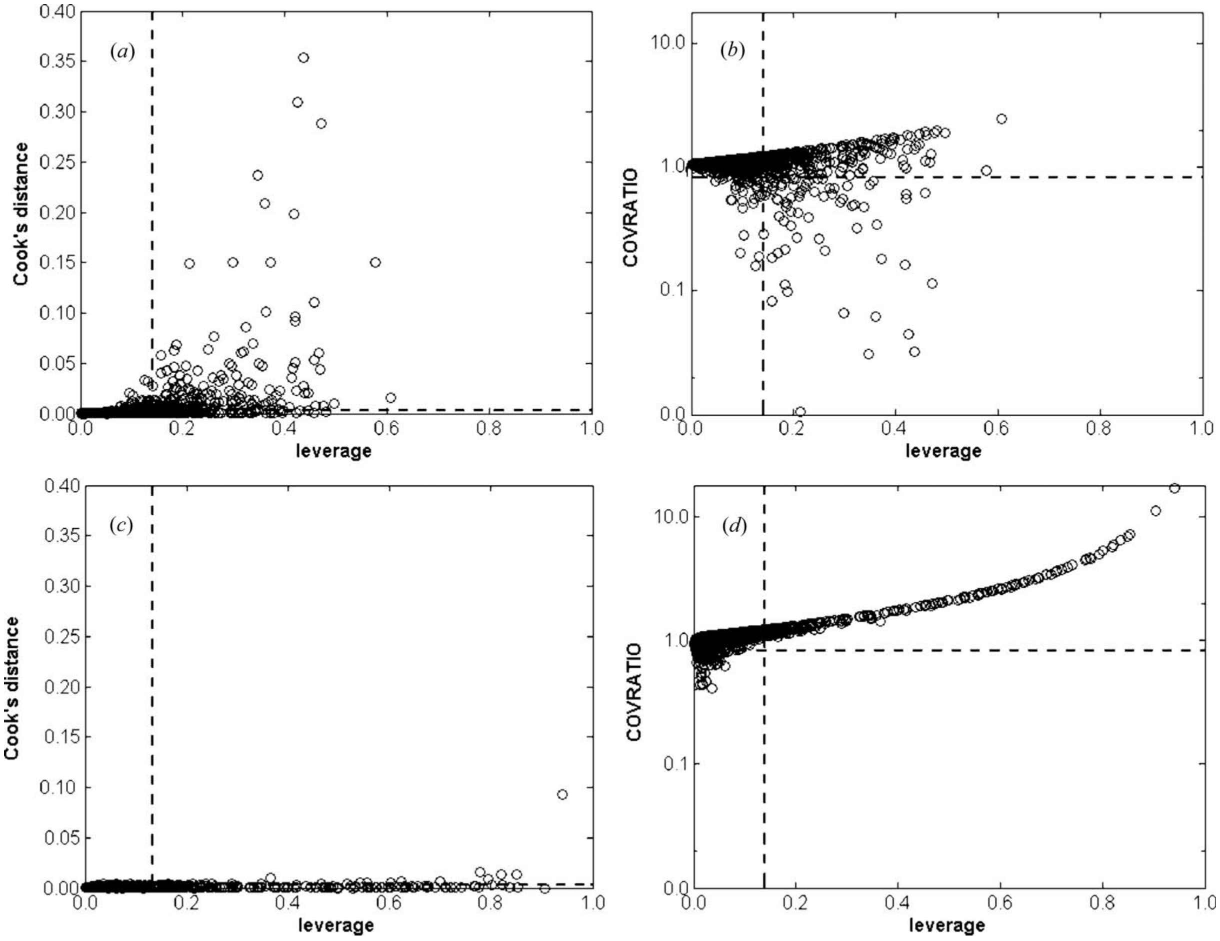

Figure 1
(*a*), (*c*) Cook's distance *versus* leverage and (*b*), (*d*) COVRATIO *versus* leverage for the refinements of sarcosine. (*a*), (*b*) WLS refinement; (*c*), (*d*) IRLS refinement with Huber function and $c = 0.100$. Vertical dashed lines = leverage cutoff value; horizontal dashed lines = Cook's distance cutoff value (*a*), (*c*) and low COVRATIO cutoff value (*b*), (*d*).

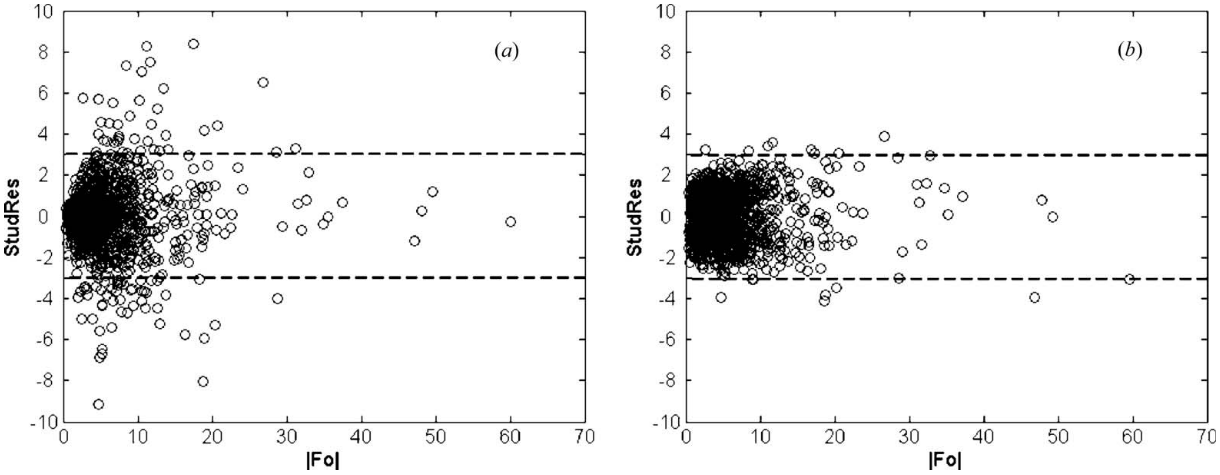

Figure 2
StudRes *versus* $|F_o|$ for the sarcosine WLS refinement (*a*) and IRLS refinement with Huber function and $c = 0.100$ (*b*). Dashed lines = StudRes cutoff (±3).

The effect of reducing the bad influence of the outliers on the refinement has an impact on the values of the scaled residuals (Table 3), which lie in ranges almost halved with respect to the WLS case (Table 3 and Fig. 2). For either StdRes or StudRes (when n is large), one should expect no more than 5% of the absolute residuals to exceed the value of 1.96, no more than 1% to exceed the value of 3 (in the WLS procedure the percentages of reflections with an absolute

Table 3
Minimum and maximum values for scaled residuals, Cook's distance and COVRATIO for different tuning constant values in the Huber weighting scheme.

	Tuning constant								
	1.400	1.345	1.100	1.000	0.900	0.800	0.500	0.100	0.010
Gaussian efficiency	96%	95%	92%	90%	89%	87%	79%	69%	64%
max(StdRes)	4.58	4.54	4.36	4.29	4.22	4.16	3.98	3.86	4.12
min(StdRes)	−4.89	−4.85	−4.64	−4.56	−4.47	−4.39	−4.14	−4.05	−4.73
max(StudRes)	4.71	4.67	4.47	4.39	4.32	4.25	4.05	3.90	4.14
min(StudRes)	−5.03	−4.99	−4.79	−4.70	−4.61	−4.53	−4.31	−4.13	−4.89
max(StuDel)	4.73	4.68	4.49	4.41	4.33	4.26	4.06	3.91	4.16
min(StuDel)	−5.05	−5.01	−4.81	−4.72	−4.63	−4.54	−4.32	−4.64	−4.91
max(Cook)	0.041	0.040	0.042	0.045	0.049	0.055	0.051	0.094	0.012
max(COVRATIO)	3.48	3.54	3.88	4.03	4.21	4.42	6.08	17.18	93.24
min(COVRATIO)	0.27	0.27	0.30	0.32	0.33	0.35	0.39	0.41	0.29
No. of outliers	6	4	2	1	0	0	0	0	0

Table 4
Selected figures of merit for WLS and IRLS refinements of datolite.

Symbols as in Table 1.

Weighting function	WLS	Huber ($c = 1.345$)	Huber ($c = 1.250$)	Logistic ($c = 1.205$)	Logistic ($c = 1.100$)
Gaussian efficiency	100%	95%	94%	95%	94%
$R(F_o)$	0.0122	0.0094	0.0094	0.0093	0.0093
$R(F_o^2)$	0.0260	0.0159	0.0159	0.0158	0.0158
wSSE	2553.68	66.10	63.07	56.76	53.44
MSE$^{1/2}$	1.0480	0.1688	0.1648	0.1564	0.1517
R	0.9992	0.9999	0.9999	0.9999	0.9999
R^2	0.9985	0.9997	0.9997	0.9997	0.9998
adjR^2	0.9983	0.9997	0.9997	0.9997	0.9997
Max. (shift/e.s.d.)	0.0000	0.0085	0.0099	0.0084	0.0050
r.m.s. (shift/e.s.d.)	0.0000	0.0269	0.0303	0.0297	0.0282
⟨shift/e.s.d.⟩	0.0000	0.0012	0.0015	0.0014	0.0013
ME	0.0025	0.0004	0.0003	-0.0001	-0.0001
σ^2	0.1255	0.0706	0.0708	0.0699	0.0701
MAE	0.1764	0.1355	0.1353	0.1351	0.1348
MARE	0.0134	0.0136	0.0135	0.0135	0.0135
MEDE	−0.0020	−0.0014	−0.0010	−0.0010	−0.0010
MAD	0.1765	0.1355	0.1353	0.1351	0.1348
SAE	455.72	349.96	349.38	348.91	348.30
AIC	17245.0	12308.7	12274.4	12078.2	12030.2
PRESS	14872.1	97.8	92.1	81.8	76.0
predR^2	0.9912	0.9996	0.9996	0.9996	0.9997

scaled residual >1.96 and >3 were 7.3% and 3.2%, respectively).

Robust weighting also increases the minimum values of COVRATIO, indicating a reduction of the bad influence of some reflections on the estimates. The precision of the estimates, evaluated by means of the overall ⟨s.u.⟩, is greatly increased, as expected, yielding a gain with respect to WLS ranging from ~25% up to ~52%. The latter feature is a physiological consequence of the scaling of the residuals.

Note that in this experiment the tuning constants of the IRLS weighting functions have been adjusted, since the 'default' values ($c = 1.345$ and $c = 1.205$ for Huber and logistic weights, respectively) yielded an ineffective downweighting of the outliers [four outliers still detected in both Huber and logistic runs with this setup of the tuning constants (Table 3 and Figs. 1*c*, 1*d*).

Perusal of Table 3 gives an idea of the behaviour of the regression at different values of the tuning constant. As can be seen, the effect of increasing c from 1.345 to 1.4 is to punish the residuals less severely, with a consequent increase in the number of outliers detected (from 4 to 6). The progressive lowering of c turns into a complete disappearance of the outliers when $c = 0.9$, at a Gaussian efficiency level ~89%. The results in terms of accuracy of the model improve until $c = 0.1$ (Gaussian efficiency ~69%). Lower values of the tuning constant bias the refinement, worsening the accuracy (the overall gain in accuracy with respect to the WLS decreases from 18 to 12%): the scaled residuals become larger, indicating that there is an overfiltering of the mismeasured reflections which turns into a bias in the estimates. This is a case in which the adjustment of the tuning constant provides a significant improvement of the results, especially for some classes of variables such as atomic displacement parameters and multipole parameters. Note that, in general, it is wise not to set a tuning constant value corresponding to a Gaussian efficiency less than ~69%.

In our experience, it has been observed that elimination of the possible outliers in an IRLS procedure can lead to a slight worsening of the refinement in terms of difficulty in reaching convergence, as well as with respect to the precision and accuracy of the estimates.

4.2. Synthetic datolite

We present here another case study that turned out to be quite interesting and whose IRLS procedure proved to be very effective. The data collection for this kind of compound is sometimes affected by Renninger effects and/or large undesired extinction corrections, making a subsequent multipole refinement difficult to carry out. As in the previous case study, the results of both methods have been summarized (Table 4) and the William's graph/COVRATIO/ leverage comparison is depicted in Fig. 3 (symbols as in Fig. 1). Even in this case the robust weight functions adopted were Huber's and logistic schemes.

4.2.1. WLS regression. The noise introduced in the structure factors with the same scheme adopted for sarcosine

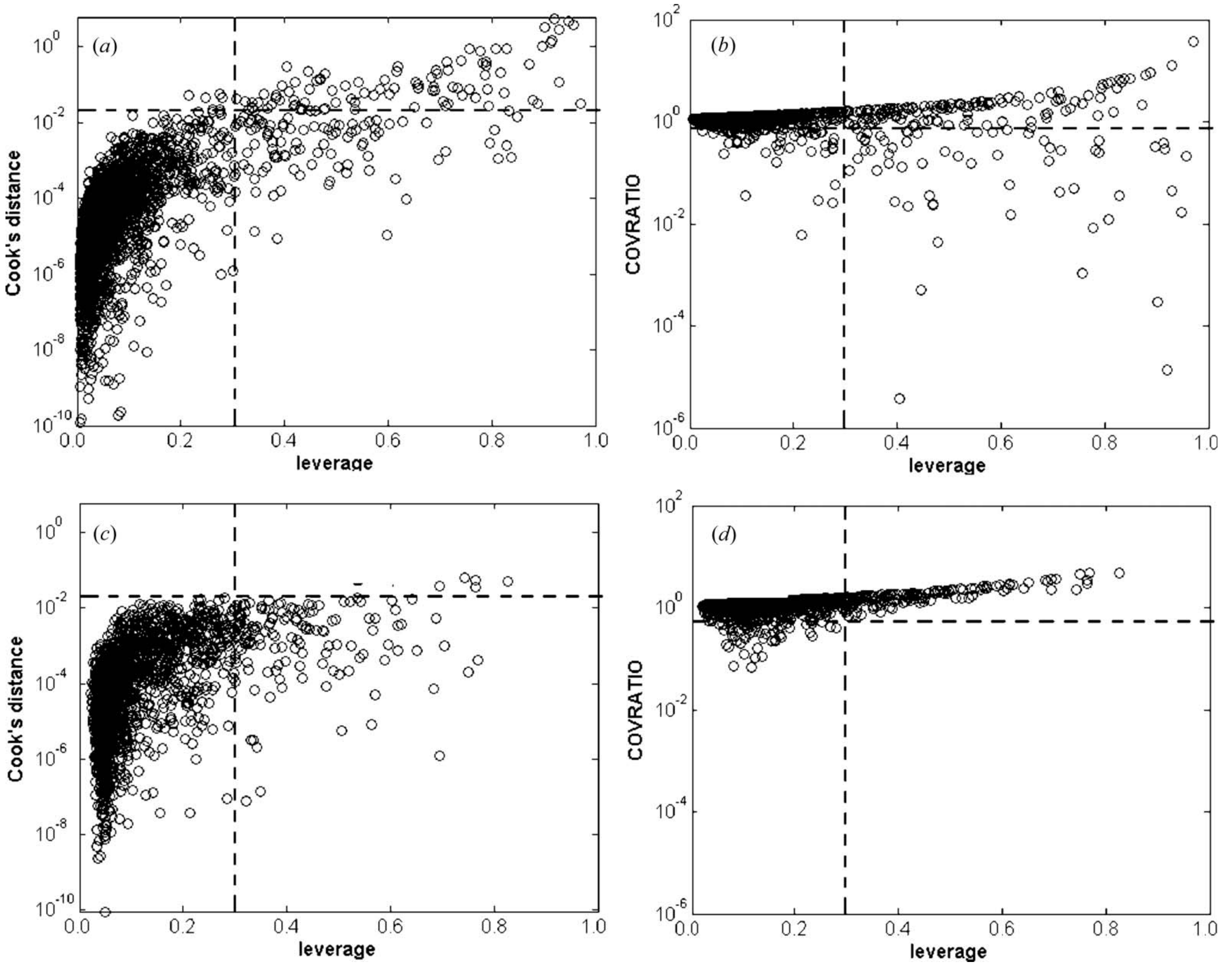

Figure 3
(*a*), (*c*) Cook's distance *versus* leverage and (*b*), (*d*) COVRATIO *versus* leverage for the refinements of datolite. (*a*), (*b*) WLS refinement; (*c*), (*d*) IRLS refinement with logistic function and c = 1.100. Dashed lines as in Fig. 1.

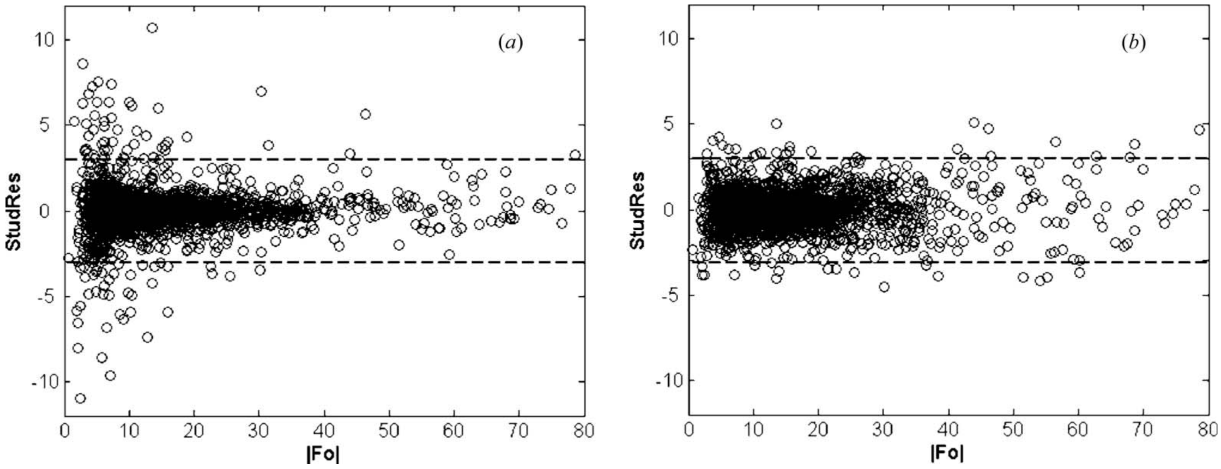

Figure 4
StudRes *versus* $|F_o|$ for the datolite WLS refinement (*a*) and IRLS refinement with Huber function and c = 0.100 (*b*). Dashed lines as in Fig. 2.

involved a mean discrepancy between theoretical and synthetic $|F_o|$ ~1.1%. This quite large noise turned into a certain difficulty in refining the κ values for all of the atoms involved (maximum shift/s.u. > 0.01). They were kept fixed.

WLS refinement gave final $R(F_o)$ = 0.0122 and $R(F_o^2)$ = 0.0260 (Table 3) using the weighting scheme implemented in *XD* by setting the coefficients a and f to 0.04 and 1/3, respectively. The overall measure of the discrepancy between

Table 5
Absolute discrepancy measurements between each class of the theoretical and refined parameters for datolite.

Symbols as in Table 2.

Weighting function	WLS	Huber (c = 1.345)	Huber (c = 1.250)	Logistic (c = 1.205)	Logistic (c = 1.100)
$(\Sigma\vert x_i^{\rm true}- x_i^{\rm calc}\vert/p)_{\rm all}$	0.0578	0.0348 (40%)	0.0346 (40%)	0.0347 (40%)	0.0346 (40%)
$\langle \text{s.u.}\rangle_{\rm all}$	0.0188	0.0165 (23%)	0.0144 (24%)	0.0142 (24%)	0.0140 (26%)
$(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert/p)_{\rm coord}$	5.7E-05	3.9E-05 (31%)	3.8E-05 (32%)	3.7E-05 (34%)	3.6E-05 (35%)
$\langle \text{s.u.}\rangle_{\rm coord}$	4.9E-05	4.3E-05 (14%)	4.2E-05 (15%)	4.1E-05 (18%)	4.0E-05 (20%)
$\max(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\rm coord}$	2.3E-04	1.4E-04	1.4E-04	1.4E-04	1.4E-04
$\min(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\rm coord}$	3.0E-06	0.0	0.0	0.0	0.0
$(\Sigma\vert x_i^{\rm true}- x_i^{\rm calc}\vert/p)_{\rm a.d.p.}$	9.5E-05	6.0E-05 (37%)	5.9E-05 (38%)	5.8E-05 (39%)	5.7E-05 (40%)
$\langle \text{s.u.}\rangle_{\rm a.d.p.}$	8.9E-05	7.7E-05 (14%)	7.5E-05 (16%)	7.3E-05 (18%)	7.1E-05 (20%)
$\max(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\rm a.d.p.}$	3.1E-04	1.9E-04	1.9E-04	1.8E-04	1.8E-04
$\min(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\rm a.d.p.}$	0.0	0.0	1.0E-06	1.0E-06	0.0
$(\Sigma\vert x_i^{\rm true}- x_i^{\rm calc}\vert/p)_{\kappa}$	0.0376	0.000 (100%)	0.0030 (92%)	0.0006 (98%)	0.0024 (94%)
$\langle \text{s.u.}\rangle_{\kappa}$		0.0235	0.0234	0.0233	0.0232
$\max(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\kappa}$	1.0E-01	1.7E-05	7.7E-03	1.0E-03	7.8E-03
$\min(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\kappa}$	6.4E-03	3.0E-06	2.6E-04	1.6E-04	2.2E-04
$(\Sigma\vert x_i^{\rm true}- x_i^{\rm calc}\vert/p)_{\rm mult}$	0.0826	0.0487 (41%)	0.0485 (41%)	0.0486 (41%)	0.0484 (41%)
$\langle \text{s.u.}\rangle_{\rm mult}$	0.0261	0.0199 (24%)	0.0196 (25%)	0.0195 (25%)	0.0193 (26%)
$\max(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\rm mult}$	2.4E+00	1.4E+00	1.4E+00	1.4E+00	1.4E+00
$\min(\Sigma\vert x_i^{\rm true}x_i^{\rm calc}\vert)_{\rm mult}$	6.8E-04	3.4E-04	1.0E-04	1.6E-04	4.6E-05
$(\vert x_i^{\rm true}- x_i^{\rm calc}\vert)_{\rm scale}$	0.0630	0.0007 (99%)	0.0007 (99%)	0.0007 (99%)	0.0004 (99%)
$\text{s.u.}_{\rm scale}$	0.0011	0.0005 (58%)	0.0004 (58%)	0.0004 (59%)	0.0004 (59%)

the refined parameters and the theoretical ones was 0.0578, corresponding to a relative absolute error on the estimates of ~16%. The mean value of the s.u. was 0.0188.

For this structure the thresholds for the estimators listed above were 0.30 for the leverage, 0.0017 for the Cook's distance, 0.72 for the lower bound of the COVRATIO estimator and 1.40 for the upper bound of COVRATIO. Very large values of |StdRes| have been observed (~8.3), as well as of |StudRes| and |StuDel| (~11.0), largely exceeding the typical threshold values of 3. The 2% of the reflections showed a |StdRes| > 3, *i.e.* twice the expected value. The regression diagnostics revealed the presence of 60 influential outliers, due both to the errors introduced (*i.e.* to the 'mismeasurement' of a number of data) and the incompleteness of the model, since the κ factors have been kept fixed.

4.2.2. IRLS regression. Even in this case the results of the IRLS runs are much improved (Table 4) with respect to the WLS procedures, providing a gain in accuracy with respect to the WLS case of ~40%. The overall measure of the discrepancy between the refined parameters and the theoretical ones was ~0.0345, corresponding to a relative absolute error on the estimates of ~9% (Table 5). The mean value of the s.u. ranged from 0.0165 to 0.0140, corresponding to an improvement of 25% with respect to the WLS procedure.

In this case study, the most important advantage of using robust regression has been the capability of refining the κ factors ($\langle$shift/s.u.$\rangle$ ~0.001), in addition to the general improvement usually observed in IRLS regressions.

Even for datolite, the analysis of the scaled residuals showed ranges almost halved with respect to the WLS case (Fig. 4), together with a significant reduction of the percentage of reflections with |StdRes| > 3 (from ~2% to ~1%, the latter being the expected value).

In this experiment no dramatic adjustment of tuning constants of the IRLS weighting was needed: the 'default' values mentioned above were good enough to ensure a complete disappearance of the outliers. Fine tuning of the weight constants (c = 1.250 and c = 1.100 for Huber's and logistic weights, respectively, both with Gaussian efficiency ~94%) yielded just a slight improvement of the results.

4.3. Experimental formamidine

The comparison between the WLS and the IRLS refinement of a noise-induced experimental data set presented in this section is aimed at emphasizing the impact on the practical use of the robust approach.

X-ray data for *N*-(4-methoxyphenyl)-*N*-phenyl-*N*-oxyformamidinium species 5 were previously collected and processed by MM (more details on the crystal structure and the data-collection settings can be found in Giumanini *et al.*, 1999).

Because of the noise, this is a typical case in which the multipole model refined with a traditional WLS procedure can be limited at most to the evaluation of the κ factors and the monopole terms for the non-hydrogen atoms. It will be shown that, using a robust procedure, not only can the results be improved with respect to the WLS refinement, but the multipole terms for the non-hydrogen atoms can be expanded up to the octupole terms, and the evaluation of the monopole terms for the hydrogen atoms can also be done. Since a reference model is not known, the comparison between each WLS and IRLS run can be made through the evaluation of the figures of merit, the AIC statistics, the analysis of the s.u. associated with each variable and the analysis of the scaled residuals. In addition, in order to check the reliability of the regression coefficients, a Student's *t*-test has been performed for all of the models.

4.3.1. WLS regressions. The WLS refinements have been performed adopting the following models:

Table 6
Selected figures of merit for WLS and IRLS refinements of experimental formamidine.

WLS0 = WLS refinement with κ and monopole terms for non-hydrogen atoms; WLS1 = WLS refinement as WLS0 + dipole terms for non-hydrogen atoms; WLS2 = as WLS1 but cutting reflections with $I/\sigma(I) < 3$; Huber0 = IRLS as WLS0 with Huber function and $c = 0.100$; Huber1 = IRLS as WLS1 with Huber function and $c = 0.100$; Huber2 = IRLS as WLS2 with Huber function and $c = 0.100$; Huber3 = IRLS as WLS1 with Huber function and $c = 0.100$, + the monopole terms for hydrogen atoms and the octupole terms for non-hydrogen atoms. Other symbols as in Table 1.

Weighting function	WLS0	WLS1	WLS2	Huber0 ($c = 0.100$)	Huber1 ($c = 0.100$)	Huber2 ($c = 0.100$)	Huber3 ($c = 0.100$)
Gaussian efficiency	100%	100%	100%	67%	67%	67%	67%
n	2181	2181	1472	2181	2181	1472	2181
p	225	279	279	225	279	279	495
$R(F_o)$	0.0767	0.0755	0.0460	0.0728	0.0710	0.0417	0.0599
$R(F_o^2)$	0.0529	0.0539	0.0491	0.0435	0.0411	0.0353	0.0294
wSSE	8490.23	8142.60	5868.54	197.84	190.89	96.22	135.29
MSE$^{1/2}$	2.0834	2.0691	2.2179	0.3180	0.3168	0.2840	0.2833
R	0.9896	0.9902	0.9925	0.9995	0.9995	0.9997	0.9997
R^2	0.9794	0.9806	0.9851	0.9990	0.9990	0.9995	0.9993
adjR^2	0.9778	0.9777	0.9817	0.9989	0.9989	0.9993	0.9991
Max. shift/s.u.	0.0000	0.0641	0.4058	0.0076	0.0084	0.0081	0.0095
r.m.s. (shift/s.u.)	0.0000	0.0688	0.0633	0.0142	0.0359	0.0404	0.0350
⟨shift/e.s.d.⟩	0.0000	0.0073	0.0073	0.0006	0.0018	0.0021	0.0016
ME	0.2477	0.2466	0.1419	0.2917	0.2884	0.2884	0.2415
σ^2	1.1619	1.1596	0.9943	1.1329	1.1083	0.9477	0.8817
MAE	0.8171	0.8110	0.7155	0.7792	0.7663	0.6533	0.6357
MARE	0.2206	0.2198	0.0889	0.2114	0.2083	0.0868	0.1804
MEDE	0.1349	0.1200	0.0525	0.0940	0.0924	0.0301	0.0036
MAD	0.8148	0.8094	0.7182	0.7914	0.7794	0.6813	0.6680
SAE	1782.10	1768.55	1053.27	1699.56	1671.34	961.69	1386.56
AIC	2964.47	2873.34	2036.13	−5234.36	−5312.33	−4014.91	−6062.92
PRESS	10798.1	11144.8	9310.4	214.64	213.52	114.07	164.67
predR^2	0.9738	0.9734	0.9764	0.9989	0.9989	0.9994	0.9992

Table 7
Mean s.u. for all the variables and for each class of the refined parameters for experimental formamidine.

Percentages in parentheses for Huber0,1,2 represent the relative improvement with respect to the WLS0,1,2 results, respectively. ext = isotropic extinction parameter; other symbols as in Table 2 and Table 6.

Weighting function	WLS0	WLS1	WLS2	Huber0 ($c = 0.100$)	Huber1 ($c = 0.100$)	Huber2 ($c = 0.100$)	Huber3 ($c = 0.100$)
$\langle \text{s.u.} \rangle_{\text{all}}$	1.03E-02	2.81E-02	4.18E-02	4.02E-03 (61%)	1.42E-02 (49%)	2.05E-02 (45%)	8.32E-02
$\langle \text{s.u.} \rangle_{\text{coord}}$	1.25E-03	1.51E-03	1.72E-03	6.32E-04 (49%)	7.96E-04 (36%)	7.70E-04 (55%)	2.01E-03
$\langle \text{s.u.} \rangle_{\text{a.d.p.}}$	3.01E-03	3.67E-03	5.21E-03	1.55E-03 (49%)	1.88E-03 (37%)	2.51E-03 (48%)	3.69E-03
$\langle \text{s.u.} \rangle_{\kappa}$	3.21E-02	5.37E-02	1.14E-01	1.29E-02 (60%)	2.01E-02 (37%)	5.28E-02 (54%)	2.84E-02
$\langle \text{s.u.} \rangle_{\text{mult}}$	9.73E-02	1.03E-01	1.73E-01	3.95E-02 (58%)	5.30E-02 (47%)	7.72E-02 (55%)	1.42E-01
$\langle \text{s.u.} \rangle_{\text{ext}}$	1.16E-01	1.16E-01	1.09E-01	1.05E-02 (91%)	1.18E-02 (90%)	1.47E-02 (86%)	1.73E-02
$\langle \text{s.u.} \rangle_{\text{scale}}$	7.36E-03	7.69E-03	2.37E-02	3.41E-03 (54%)	4.14E-03 (56%)	1.91E-02 (50%)	4.85E-03

(*a*) a refinement with the evaluation of the κ shrinking factors for all the atoms and the monopole terms for the non-hydrogen atoms only (hereafter WLS0);

(*b*) a refinement with the evaluation of the κ shrinking factors for all the atoms, the monopole terms for the non-hydrogen atoms and the dipole terms for the non-hydrogen atoms (hereafter WLS1);

(*c*) a refinement as described in (*b*) but cutting the reflections with $I/\sigma(I) < 3$ (hereafter WLS2).

For all the runs the a and f coefficients of the weighting scheme implemented in *XD* were 0.0 and 1/3, respectively. Figures of merit and some statistics for WLS0,1,2 are presented in Table 6. The ⟨s.u.⟩ values for the whole set of variables and for each class of the refined parameters are presented in Table 7, while an analysis of the scaled residuals is summarized in Table 8. As can be argued, only the WLS0 run reached convergence (maximum shift/s.u. < 0.01). The quality of the X-ray data collected cannot allow a reliable estimation of the multipole parameters, with the exception of the κ factors and the monopole terms for the non-hydrogen atoms. The patterns of the scaled residuals also confirm this fact, showing a strong skewness to the highest values [for instance, max(StdRes) = 4.65, min(StdRes) = −2.67 in the WLS0 run]. Indeed, the elimination of the worst reflections [709 data points with $I/\sigma(I) < 3$] in the WLS2 run did not improve the results, which are presented here just for comparison with the IRLS refinements. Note that WLS1 and WLS2 show the presence of one outlier (namely the 413 and the 321 reflections for the WLS1 and WLS2 cases, respectively). Student's t-tests performed on the variables suggested rejecting the null hypothesis for all the variables in all the runs but WLS2, for which the evaluation of the isotropic atomic displacement

Table 8
Minimum and maximum values for scaled residuals, Cook's distance and COVRATIO for WLS0,1,2 and Huber0,1,2,3 IRLS of experimental formamidine.

Weighting function	WLS0	WLS1	WLS2	Huber0 (c = 0.100)	Huber1 (c = 0.100)	Huber2 (c = 0.100)	Huber3 (c = 0.100)
max(StdRes)	4.65	4.62	4.54	2.45	2.47	2.49	2.55
min(StdRes)	−2.67	−2.74	−2.42	−2.12	−2.21	−2.18	−2.16
max(StudRes)	4.77	4.79	4.76	2.46	2.60	2.56	2.56
min(StudRes)	−2.90	−3.06	−3.60	−2.15	−2.25	−2.24	−2.19
max(StuDel)	4.80	4.82	4.81	2.46	2.61	2.60	2.57
min(StuDel)	−2.91	−3.07	−3.61	−2.15	−2.25	−2.24	−2.20
max(Cook)	0.029	0.083	0.132	0.003	0.008	0.011	0.005
max(COVRATIO)	3.49	5.97	6.99	16.18	15.21	11.16	15.73
min(COVRATIO)	0.085	0.042	0.007	0.562	0.473	0.296	0.196
No. of outliers	0	1	1	0	0	0	0

parameters of the hydrogen atoms suffers from the cutting of the weakest reflections.

4.3.2. IRLS regressions. The IRLS refinements have been performed using the Huber weighting function with c = 0.100 for the models listed in the previous section (Huber0,1,2), and a refinement involving the quadrupole and octupole terms for the non-hydrogen atoms in addition to the multipole terms introduced in WLS1 and IRLS1 (Huber3). Perusal of Table 8 shows that in all of the IRLS procedures convergence has been reached (maximum shift/s.u. < 0.01). Moreover, Student's t-tests performed on the regression variables suggested rejecting the null hypothesis for all the cases investigated.

A direct comparison between the WLS and IRLS procedures can be made for Huber0,1,2: for these cases, the selected figures of merit presented in Table 6 show a significant improvement of the refinements. For instance, by looking at the values of $R(F_o)$, it can be seen that the Huber0 case shows an improvement with respect to the WLS0 result of ∼5%, while the improvement of Huber1 with respect to the WLS1 result is ∼6%, and the Huber2 run yielded an improvement with respect to WLS2 of ∼9%. The 'gain' in precision of the estimates for each IRLS run is even more significant (on average ∼50%), as shown in Table 7. Note that the highest improvement is related to the isotropic extinction coefficient (∼90%). In all the IRLS runs, the scaled residuals are lower than those recorded for the corresponding WLS refinements and, overall, the residuals are no longer skewed. Cook's distances are ∼90% lower than the WLS values, and the min(COVRATIO) values are much greater than those recorded for all the WLS cases. The severe value of the tuning constant (c = 0.100) allowed for all the IRLS runs an effective downweighting of the outliers (Table 8).

Starting from the WLS0 model, the simultaneous introduction of all the multipole terms up to the octupoles for the non-hydrogen atoms as well as the monopole terms for the hydrogen atoms in the IRLS regression yielded a model with $R(F_o)$ = 0.0599, corresponding to an improvement of ∼22% with respect to the stable WLS0 refinement. The values of $\langle$s.u.$\rangle$ for each class of variables listed in Table 7 are actually comparable with those of the WLS and IRLS refinements performed with a very low number of parameters (Table 6). AIC statistics strongly suggest that the Huber3 model is much more accurate than the other refinements, as well as the values of R, R^2 and adjR^2. This example, which can be considered as an 'extreme' application of the robust approach, shows how powerful the robust procedure is in crystal structure (multipole) refinements.

5. Conclusions

This explorative attempt to use a robust regression technique in a least-squares structure refinement of small molecules yielded encouraging results both in terms of accuracy and precision of the estimates, showing an overall improvement of the regression results with respect to the traditional WLS refinement. In particular, both the tests on the synthetic data and the application to experimental cases presented here showed that the fitting quality is always much better after using the robust algorithm (as demonstrated by the figures of merit discussed) and the precision of the estimates is much higher. Besides, the synthetic runs showed a better fit of the estimates to the theoretical ones that turns into a greater reliability of the structures and, consequently, into a greater reliability of any subsequent calculation (first of all, the electron-density reconstruction). These features can be reasonably extended even to the application of the robust procedure to experimental practice.

Moreover, these techniques allow one to downweight outliers simultaneously, save time and avoid any naive eliminations of reflections only apparently aberrant. The latter feature is even more true in the presence of a large number of outliers.

It must be noticed that in this round-robin experiment the IRLS tests suggested that the definition of the robust residuals should involve MAD and StudRes, and that the use of the Huber or logistic function could be considered as a proper choice. This is just one of the possible recipes to adopt.

It is our opinion that further investigations aimed at testing other robust techniques as well as other robust weighting functions could shed light on unknown behaviours for these algorithms, enhancing the control on this kind of optimization process. We hope that this research area in crystallography will be thoroughly investigated, from both a theoretical and practical point of view.

APPENDIX *A*

A1. List of symbols and abbreviations

n: number of observations.

p: number of variables in the least-squares procedure.

F_o, F_c: observed structure factor, calculated structure factor.

w: statistical weight of the least-squares refinement.

e_i: residual error $F_{o_i} - F_{c_i}$ associated with the ith reflection.

SSE: error sum of squares, defined as $\sum_{i=1}^{n} e_i^2$.

wSSE: weighted error sum of squares, defined as $\sum_{i=1}^{n} w_i e_i^2$.

ME: mean error, defined as $(1/n)\sum_{i=1}^{n} e_i$.

σ^2: error variance, defined as $(1/n)\sum_{i=1}^{n}(e_i - \mathrm{ME})^2$.
MAE: mean absolute error, defined as $(1/n)\sum_{i=1}^{n}|e_i|$.
MARE: mean absolute relative error, defined as $(1/n)\sum_{i=1}^{n}|e_i|/|f_{o_i}|$.
MEDE: errors median, *i.e.* the 50th percentile of the errors.
MAD: median absolute deviation of the errors, *i.e.* the median of the absolute deviation of the errors from MEDE.
MSE: mean squared error, defined as $[1/(n-p)]\sum_{i=1}^{n}e_i^2$.
SAE: sum of absolute errors, *i.e.* $\sum_{i=1}^{n}|e_i|$.
h_i: ith diagonal element of the $(n \times p)$ projection matrix, *i.e.* leverage of the ith data point.
PRESS: value of the PRESS statistics, defined as $\sum_{i=1}^{n}[e_i/(1-h_i)]^2$.
R, R^2: correlation coefficient, coefficient of determination.
adjR^2: adjusted R^2, defined as $1-(1-R^2)(n-1)/(n-p-1)$.
predR^2: predicted R^2, defined as $1-\mathrm{PRESS}/\mathrm{TSS}$, where TSS is the total sum of squares, *i.e.* $\sum_{i=1}^{n}(f_{o_i} - \langle f_o\rangle)^2$.
AIC: value of the Akaike's information criterion, defined as $n \times \ln(\mathrm{SSE}/n) + 2p/n$.
$R(F_o)$, $R(F_o^2)$: crystallographic discrepancy factor on $|F_o|$, crystallographic discrepancy factor on F_o^2.
s.u.: standard uncertainty associated with each least-squares parameter.
shift/s.u.: ratio of the final least-squares parameter shift to the final s.u.
a.d.p.'s: atomic displacement parameters.

A2. Robust estimators formulae

(*a*) MSE σ estimator, calculated as: $s = \mathrm{MSE}^{1/2}$.

(*b*) MAD σ estimator, calculated as: $s = \mathrm{MAD}/\mathrm{K}$, where the constant K is set to 0.6745, which makes the estimate unbiased for the normal distribution.

Alternatively, s can be calculated as (Rousseeuw, 1985)

$$s = 1.4826\left[1 + \frac{5}{(n-p)}\right]s.$$

The scaled robust residual rr_i involved in the weight function is given by

$$rr_i = \frac{e_i}{(sc)},$$

where c is the tuning constant.

A3. Scaled residuals formulae

(*a*) Standardized residual (StdRes), computed as

$$\mathrm{StdRes}_i = \frac{e_i(w_i)^{1/2}}{s}.$$

(*b*) Studentized residuals (StudRes), computed as

$$\mathrm{StudRes}_i = \frac{\mathrm{StdRes}_i}{(1-h_i)^{1/2}}.$$

(*c*) Studentized deleted residuals (StudDel), computed as

$$\mathrm{StudDel}_i = \frac{e_i(w_i)^{1/2}}{[\mathrm{StudVar}_i(1-h_i)]^{1/2}},$$

where StudVar$_i$ is computed as

$$\mathrm{StudVar}_i = \frac{(s^2\mathrm{SSE} - w_i e_i)^2}{(1-h_i)^{1/2}}.$$

A4. Robust weight functions

The reader may refer to Holland & Welsch (1977) for a summary of the most commonly used weighting functions and for further details. We can categorize them into three groups:

(I) 'Hard redescenders';

(I*a*) Andrews weighting scheme (Andrews *et al.*, 1972);

(I*b*) Tukey's bisquare function (Beaton & Tukey, 1974);

(I*c*) Talwar's weighting function (Hinich & Talwar, 1975).

All of them involve $w \rightarrow 0$ for $|e|$ sufficiently large. Talwar's scheme assigns unit/zero weights depending on $|e|$.

(II) 'Soft redescenders';

(II*a*) Cauchy's weights (or *t*-likelihood);

(II*b*) Welsch's function (Dennis & Welsch, 1976).

(III) 'Monotone redescenders';

(III*a*) Huber's function (Huber, 1964);

(III*b*) logistic weight;

(III*c*) Fair's function (Fair, 1974).

The schemes adopted in this work are (III*a*) and (III*b*). In particular:

(*a*) Huber's function is defined as

$$w_i = \frac{1}{\max(1, |rr(e_i|c)|)},$$

where c is set to 1.345 to have a 95% asymptotic Gaussian efficiency.

(*b*) The logistic function is defined as

$$w_i = \tanh(|rr_i(e_i|c)|^{-1}),$$

where c is set to 1.205 to have a 95% asymptotic Gaussian efficiency.

Financial support has been provided by MIUR.

References

Akaike, H. (1973). *2nd International Symposium on Information Theory*, edited by B. N. Petrov & F. Csaki, pp. 267–281. Budapest: Akademiai Kiado.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. & Tukey, J. W. (1972). *Robust Estimates of Location.* Princeton: Princeton University Press.

Beaton, A. E. & Tukey, J. W. (1974). *Techometrics*, **16**, 147–185.

Belsey, D. A., Kuh, E. & Welsh, R. E. (1980). *Regression Diagnostics: Identify Influential Data and Sources of Collinearity.* New York: John Wiley and Sons.

Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.

Box, G. E. & Tiao, G. C. (1968). *Biometrika*, **55**, 119–129.

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: a Practical Information–Theoretic Approach*, 2nd ed. New York: Springer-Verlag.

Carruthers, J. R. & Watkin, D. J. (1979). *Acta Cryst.* A**35**, 698–699.

Dennis, J. E. & Welsch, R. E. (1976). *Techniques for Nonlinear Least Squares and Robust Regression.* Proceedings of the American Statistical Association, Washington DC, USA.

Dittrich, B. & Spackman, M. A. (2007). *Acta Cryst.* A**63**, 426–436.

Fair, R. C. (1974). *Ann. Econ. Soc. Meas.* **3**, 667–677.

Giumanini, A. G., Toniutti, N., Verardo, G. & Merli, M. (1999). *Eur. J. Org. Chem.* pp. 141–143.

Hampel, F. R. (1975). *Bull. Int. Stat. Inst.* **46**, 375–391.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust Statistics – the Approach Based on Influence Functions.* New York: John Wiley.

Hinich, M. T. & Talwar, P. P. (1975). *J. Am. Stat. Assoc.* **70**, 113–119.

Holland, P. W. & Welsch, R. E. (1977). *Commun. Stat. Theory Methods*, **A6**, 813–828.

Huber, P. J. (1964). *Ann. Math. Stat.* **101**, 35–73.

Huber, P. J. (1973). *Ann. Stat.* **1**, 799–821.

Ivanov, Y. V. & Belokoneva, E. L. (2007). *Acta Cryst.* B**63**, 49–55.

Jaeckel, L. A. (1972). *Ann. Math. Stat.* **43**, 1449–1458.

Kass, R. E. & Raftery, A. E. (1995). *J. Am. Stat. Assoc.* **90**, 773–795.

Koritsanszky, T., Howard, S. T., Richter, T., Mallinson, P. R., Su, Z. & Hansen, N. K. (1995). *XD – a Computer Program Package for Multipole Refinement and Analysis of Charge Density from Diffraction Data.* Free University of Berlin, Germany.

Merli, M. (2005). *Acta Cryst.* A**61**, 471–477.

Merli, M., Sciascia, L. & Turco Liveri, M. L. (2010). *Int. J. Chem. Kinet.* **42**, 587–607.

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wassermann, W. (1990). *Applied Linear Regression Models.* Chicago: IRWIN.

Prince, E. (1982). *Mathematical Techniques in Crystallography*. New York: Springer-Verlag.

Prince, E. & Nicholson, W. L. (1985). *Structure and Statistics in Crystallography*, edited by A. J. C. Wilson, pp. 183–195. New York: Adenine Press.

Rousseeuw, P. J. (1984). *J. Am. Stat. Assoc.* **79**, 871–880.

Rousseeuw, P. J. (1985). In *Mathematical Statistics and Applications*, edited by W. Grossmann, G. Pflug, I. Vincze & W. Wertz. Dordrecht: Reidel.

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York: John Wiley and Sons.

Schwarz, G. (1978). *Ann. Stat.* **6**, 461–464.

Sivia, D. S. (1996). *Data Analysis – a Bayesian Tutorial*. Oxford: Clarendon Press.

Spagna, R. & Camalli, M. (1999). *J. Appl. Cryst.* **32**, 934–942.

Velleman, P. F. & Welsch, R. E. (1981). *Am. Stat.* **35**, 234–242.

Watkin, D. (2008). *J. Appl. Cryst.* **41**, 491–522.