Application Of Robust/Resistant Techniques
To Crystal Structure Refinement

Will be part of the National Bureau of Standards FY76
Yearly Summary of Activities Report
Washington D.C. January , 1977

by
W.L. Nicholson
E. Prince

August, 1977

BATTELLE
Pacific Northwest Laboratories
Richland, Washington 99352

MASTER

This work was done for the Energy Research and Development
Administration under contract E(45-1): 1830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

APPLICATION OF ROBUST/RESISTANT TECHNIQUES TO
CRYSTAL STRUCTURE REFINEMENT

E. Prince

and

W. L. Nicholson*

(Battelle Pacific Northwest Laboratories, Richland, WA)

A technique for fitting a theoretical model to a set of experimental data points and estimating the best values of adjustable parameters in the model is said to be "robust," or, more precisely, "robust of efficiency," if the parameter estimates have near minimum variance for a broad class of distributions for the errors in experimental data. A technique is "resistant" if the estimates are not highly dependent on any small subset of the experimental data. To date, all techniques that are robust of efficiency are also resistant. Many data analysts feel that this is inevitable, hence, the splice word, robust/resistant.

The technique of least squares, the one most commonly used for refining crystal structures, is neither robust nor resistant. Least squares was designed specifically by Gauss for normally distributed errors. With broad tailed error distributions and/or the presence of a few bad data points in a large data set, least squares can give anomalous results. In recent years statisticians have been developing methods of fitting and parameter estimation which work nearly as well as least squares for ideal data sets but which are robust/resistant. The purpose of this research is to apply these methods to crystal structure refinement, both as a non-trivial test of the robust/resistant approach and as a development in calculational crystallography.

*Funded by the U.S. Energy Research and Development
Administration under contract E(45-1):1830.

The technique of least squares is based on finding the minimum value of the quantity $\sum_i w_i (y_{io} - y_{ic})^2$, where y_{io} is the value of the observed data, y_{ic} is a calculated value of that point based on some theoretical function containing adjustable parameters, and w_i is a weight. For the method to work well the weight, w_i , must be proportional to $1/\sigma_i^2$, where σ_i^2 is the variance of the observation y_{io} . This variance is, however, usually not known a priori, but is approximated by a formula containing a term arising from an assumed Poisson distribution of random counts and one or more other terms which estimate additional experimental error associated with the observed value, y_{io} .

Tukey⁽¹⁾ describes a robust/resistant alternative to least squares estimation which is particularly attractive for crystal structure refinement. In the least squares program the response of a refinement is modified by adjusting the weight according to the formula $w'_i = w_i W(x)$, where $x = (y_{io} - y_{ic}) / K \hat{\sigma}_i$. $\hat{\sigma}_i$ is the standard deviation approximation described above and K is a robust/resistant estimate of scale. The function $W(x)$ is called a "W-function". Characteristics that a W-function should have in order to create a robust/resistant refinement are discussed elsewhere by the authors.⁽²⁾ To perform the refinement an initial set of parameters is used to form normalized residuals $r = (y_{io} - y_{ic}) / \hat{\sigma}_i$. These are used to construct K . Independent work of Andrews⁽³⁾ and Beaton and Tukey⁽⁴⁾ suggest a K of 6 times the median absolute residual. With K specified, $W(x)$ is calculated and the weighted least squares iteration performed. After each iteration K is recalculated as above.

The crystallographic least squares program RFINE4⁽⁵⁾ was modified to include, as options, several different W-functions. One of these W-functions was used to investigate the operational characteristics of the robust/resistant approach using a lead deuterium orthophosphate data set. The details of the investigation are described elsewhere.^(2,6) A second option is the biweight W-function $W(x) = (1-x^2)^2$ ($|x| \leq 1$); = 0 (otherwise).

RFINE4 with the Tukey biweight W-function is being used to repeat the refinements of the structure of D(+) tartaric acid originally carried out by Hamilton and Abrahams⁽⁷⁾ as part of the single-crystal intensity project of the International Union of Crystallography. Three of the 17 experiments considered by Hamilton and Abrahams, selected on the basis of reasonably large numbers of reflections, have been analyzed so far. The refinement for each data set has three stages: 1) attempt to "recreate" the results of Hamilton and Abrahams by repeating as closely as possible the condition of their refinements; 2) add secondary extinction to the classical weighted least squares refinement; and 3) incorporate the biweight W-function and do a robust/resistant refinement. The results of Hamilton and Abrahams are not reproduced exactly, possibly because of small differences in the handling of atomic scattering factors in the original refinement. Inclusion of the Zachariasen model for secondary extinction significantly improves the refinements of all three experiments. Inclusion of the biweight function again significantly improves the refinement with weighted r indices decreasing by up to 40%.

Based on the refinement of these three experiments the robust/resistant approach looks very promising. Individual atom positions and thermal vibration parameters agree more closely than in the original analysis. For each experiment a small fraction of the data fail to fit the robust/resistant estimated model and, hence, are identified as extreme outliers. These outliers can be identified on theoretical grounds as very imprecise reflections. For example, for one experiment the extreme outliers were mostly weak reflections. Such residual patterns suggest that the biweight approach succeeds in minimizing the influence of poor data on the refinement. The above work was reported at the 1976 winter meeting of the American Crystallographic Association, Clemson, SC. A written version is being submitted for open literature publication.

Work is now in progress on refinement of the remaining experimental data sets reported by Hamilton and Abrahams. Several of their refinements give very discrepant individual atom parameter estimates. Elimination or a least reduction of these discrepancies would establish robust/resistant techniques as a viable approach to crystal structure refinement.

REFERENCES

1. J. W. Tukey in Critical Evaluation of Chemical and Physical Structural Information, National Academy of Sciences, Washington DC, 1974.
2. E. Prince and W. L. Nicholson, NBS Tech. Note 896 (1976).
3. D. F. Andrews, Technometrics 16, p 523 (1974).
4. A. E. Beaton and J. W. Tukey, Technometrics 16, p. 147 (1974).
5. L.W. Finger and E. Prince, NBS Tech. Note 854 (1975).
6. C. S. Brickenkamp and E. Prince, NBS Tech. Note 860, p 55 (1975).
7. W. C. Hamilton and S. C. Abrahams, Acta. Cryst., A26, 18, (1970).