

**DRAFT DOCUMENT: This was the project of an article that we never finished ....It is not finished and there are still errors ...**

**Do not distribute this document outside the Short Rietveld Course in Atlanta (May 2001)**

## **SYSTEMATIC ERRORS AND STRUCTURAL COMPLEXITY IN THE RIETVELD METHOD**

Juan Rodríguez-Carvajal

Laboratoire Léon Brillouin (CEA-CNRS), CE Saclay, 91191 Gif sur Yvette Cedex, France

and

Maria Luisa Medarde

Laboratory for Neutron Scattering, ETH Zürich and Paul Scherrer Institute, CH-5232 Villigen PSI,  
Switzerland

### **ABSTRACT**

The results of Rietveld refinements of a series of simulated patterns is presented. We have studied the effect of systematic errors (in the profile function) and the resolution in the quality of the refined structural parameters. The behaviour of the common R-factors as a function of counting time (statistics) for refinement with the exact model and with the wrong profile function is analysed. The profile R-factors depend on both the quality of the data and the adequacy of the structural and profile model. A way for getting the “expected R-factors” for the best structural model is proposed. We have also studied the limits of the Rietveld method for complex structures as a function of resolution. Finally we introduce some concepts and prescriptions helping to asses the “accuracy” of the structural parameters obtained by the Rietveld method.

## Introduction

The Rietveld Method [1] (RM) is now widely accepted as the natural way of refining crystal structures from powder diffraction data. However, the common practice concerning the standard deviations ( $\sigma$ 's) of the obtained structural parameters is to ignore the problems related with their significance.

In 1979 arose a strong criticism concerning the nearly systematic underestimation of the standard deviations of the structural parameters [2]. Prince clarified the problem stating that, in presence of systematic errors, “uncertainties in parameters estimates cannot be evaluated by statistical method” [3]. An important work was published by Scott [4] who clearly distinguished between *accuracy* and *precision* and gave a method for giving a “corrected” set of standard deviations. Hill and co-workers [5-9] have provided new insight into the above and related problems of the Rietveld method by using different experimental conditions and studying the behaviour of the parameters of interest. Some authors [4, 10-12] have proposed different methods to modify the  $\sigma$ 's multiplying the obtained values by a factor greater than 1, in order to take into account the “correlation of residuals” [11, 12], or to distinguish the effect of the heterogeneous set of parameters in the RM [4, 10].

Another important point that will be addressed in this article is the significance of the R-factors used in the RM. Recently some proposal and prescriptions concerning R-factors have been given [13, 14], we will show that profile R-factors cannot be used as indicators of the goodness of structural parameters if the “expected” R-factors for the best structural model are not known.

We have undertaken a simulation work for studying the effect of “known systematic errors” in structural parameters, their corresponding standard deviations, and R-factors. The ultimate goal is to find a method for assessing the true accuracy of the refined structural parameters that is still lacking. In this article we will present some aspects our work and we will also discuss the behaviour of R-factors and other indicators against statistical accuracy (counting time). Finally we introduce some definitions, as the “effective number of reflections”, and new indicators helping in the assessment of the quality of a Rietveld refinement.

## The Rietveld Method, R-Factors And Standard Deviations

The model used for describing the powder pattern profile is based in the following expression:

$$y_{ical} = \sum_{\phi} S_{\phi} \sum_{\mathbf{h}} I_{\phi\mathbf{h}} [\mathbf{\beta}_I] \Omega(T_i - T_{\phi\mathbf{h}}) [\mathbf{\beta}_P] + B_i [\mathbf{\beta}_B] \quad (1)$$

where  $\mathbf{h}$  labels the Bragg reflections, the subscript  $\phi$  labels the “phase” and vary from 1 up to the number of “phases” existing in the model.  $\Omega(T_i - T_{\phi\mathbf{h}})$  is the value of the function  $\Omega(x)$ , normalised to unit area, selected to describe the peak shape, at the position  $T_i$  due to the reflection centred at  $T_{\phi\mathbf{h}}$ . Finally,  $S_{\phi}$  is a scale factor and  $B_i$  is the background at  $T_i$ .

In the RM, the functions  $I[\mathbf{\beta}_I]$ ,  $\Omega[\mathbf{\beta}_P]$  and  $B[\mathbf{\beta}_B]$ , are calculated using a model that depends on a series of parameters. The structural information is contained in the integrated intensities  $I_{\phi\mathbf{h}}$  through the structural parameters  $[\mathbf{\beta}_I]$ , while the instrumental and microstructural effects are included in the peak shape function which depends on the “profile parameters”  $[\mathbf{\beta}_P]$ . The background parameters  $[\mathbf{\beta}_B]$  contain information about disorder, short range order, amorphous phases, etc. If the set of model parameters is  $\mathbf{\beta} = [\mathbf{\beta}_I] \cup [\mathbf{\beta}_P] \cup [\mathbf{\beta}_B] = (\beta_1, \beta_2, \dots, \beta_P)$ , where  $P = N_I + N_P + N_B$  is the total number free parameters, the Rietveld method tries to optimise the chi-square function:

$$\chi_p^2 = \sum_i w_i \{y_{iobs} - y_{ical}(\mathbf{\beta})\}^2 \quad (2)$$

where  $w_i$  is the inverse of the variance associated to the observation “i” ( $\sigma^2(y_{iobs})$ ).

If there is no model for the structural problem (crystal and/or magnetic structure) the Rietveld method is not applicable. The variance of the refined parameters are calculated using the formula:

$$\sigma^2(\beta_j) = (A^{-1})_{jj} \cdot \chi_v^2 = (A^{-1})_{jj} \chi_p^2 / (N - P + C) \quad (3)$$

where A is the matrix of the normal equations and  $\chi^2_v$  is the reduced chi-square.

For assessing the global quality of the refinement several R-factors are used in the literature. Let us summarise the formula of the most usual indicators used in Rietveld refinements.

(i) R-pattern:

$$R_P = 100 \sum_i |y_{iobs} - y_{ical}| / \sum_i y_{iobs}$$

$$cR_P = 100 \sum_i |y_{iobs} - y_{ical}| / \sum_i |y_{iobs} - B_i| \quad (4)$$

(ii) R-weighted pattern:

$$R_{wP} = 100 [ \sum_i w_i \{y_{iobs} - y_{ical}\}^2 / \sum_i w_i \{y_{iobs}\}^2 ]^{1/2}$$

$$cR_{wP} = 100 [ \sum_i w_i \{y_{iobs} - y_{ical}\}^2 / \sum_i w_i \{y_{iobs} - B_i\}^2 ]^{1/2} \quad (5)$$

(iii) R-expected:

$$R_E = 100 [ (N-P+C) / \sum_i w_i \{y_{iobs}\}^2 ]^{1/2}$$

$$cR_E = 100 [ (N-P+C) / \sum_i w_i \{y_{iobs} - B_i\}^2 ]^{1/2} \quad (6)$$

$$(iv) \text{ Reduced Chi-square: } \chi^2_v = \sum_i w_i \{y_{iobs} - y_{ical}\}^2 / (N-P+C) = (R_{wP}/R_E)^2 \quad (7)$$

(v) Durbin-Watson statistics: DW (see manual of FullProf)

The sums can be extended to all non-excluded points of the pattern or only to those points where there is at least the contribution of a Bragg reflection:

(vi) R- Bragg factor (other notations  $R_{Nuclear}$ ,  $R_I$ ):

The sums below are extended to the Bragg reflections in non-excluded regions.

$$R_B = 100 \sum_h |I_h(\text{"obs"}) - I_h(\text{calc})| / \sum_h I_h(\text{"obs"}) \quad (8)$$

where:

$$I_h(\text{"obs"}) = I_h(\text{calc}) \sum_i \Omega(T_i - T_h)_i \{y_{i\text{obs}} - B_i\} / \{y_{i\text{cal}} - B_i\} \quad (9)$$

The symbols starting with “c” correspond to background-corrected indicators that are usually higher than the corresponding non corrected R-factors.

## The Simulation Method

The simulation method is preferred over real experiments because we can quantitatively control the influence of “known ” systematic errors in the “refined ” structural parameters, enabling the evaluation of the bias. The procedure we have used can be summarised as follows:

- 1- For a hypothetical (or real) compound characterised by a given crystal structure, scale factor (which is related to a “ counting time ”), resolution function, peak shape, angular range  $\{2\theta_i = 2\theta_s + (i-1) \cdot \Delta, i=1, N\}$  and background level, we calculate a theoretical (deterministic) diffraction pattern  $\{y_i(\text{theo}), i=1, N\}$ .
- 2-  $N_r$  versions of the “observed ” diffraction pattern are generated by corrupting the deterministic pattern by a noise calculated with a generator of pseudo-random numbers according to the Poisson distribution  $y_i(\text{obs}) = P\{y_i(\text{theo})\}$ .
- 3- The process is repeated for  $N_t$  different values of the “counting time ” (scale factor), thus producing  $N_r \times N_t$  observed diffraction patterns for the same compound.
- 4- Each pattern is refined by the RM by using either the “true ” model or a biased model (e.g. wrong peak shape).
- 5- The values of the refined parameters are then compared to the true valued (bias and dispersion) and the computed standard deviations are compared to the empirical ones obtained from the  $N_r$  versions of the same diffraction pattern, as a function of the counting time.
- 6- Finally we study the behaviour of the R-factors versus counting time for the different conditions of refinement.

- 7- We also refine the diffraction patterns without structural model to obtain the “expected ” R-factors for the best structural model by profile-matching (PM) iterating the Rietveld formula for the Bragg R-factor

### **The profile matching method (usually known as Le Bail fit)**

Pawley [15] proposed a method to obtain integrated intensities (for a single “phase”) by refinement of the whole profile using  $I_h$  in expression (1) as least squares (LS) parameters. The matrix of the normal LS-equations is usually singular, due to the reflection overlap. Pawley solved the problem by using “slack constraints” [15]. However, there is an intrinsic degeneracy causing an infinite number of solutions. When the reflections  $\mathbf{h}_1$  and  $\mathbf{h}_2$  are accidentally at the same position  $T_{\mathbf{h}_1}=T_{\mathbf{h}_2}$  the global intensity  $I(\mathbf{h}_1, \mathbf{h}_2)$  can be decomposed,  $I_{\mathbf{h}_1} + I_{\mathbf{h}_2}$ , in an infinite number of ways. Usually the *a priori* Bayesian guess,  $I_{\mathbf{h}_1} = I_{\mathbf{h}_2}$ , is chosen. This uncertainty is the fundamental point limiting the capability of getting a structural solution from powder data. The LS approach proposed by Pawley has the drawback of needing the inversion of a matrix that can be of considerable size even if it is sparse. An alternative procedure is to iterate the calculated profile up to “match” the observed pattern. For doing that a criterion to approach the observed pattern must be defined. The problem can be formulated, for a single phase, as follow. Let  $\mathbf{y}_{\text{obs}} = \{y_{\text{obs}}\}_{i=1, \dots, N}$  (N: number of points in the profile) be the observed diffraction pattern. This set of values can be calculated if the integrated intensities  $\mathbf{I}^0 = \{I^0_j\}_{j=1, \dots, n}$  (n: number of reflections) the cell parameters and the peak shape are known. In that case the expression (1) can be rewritten as:

$$y_{\text{obs}} = y_{\text{ical}} + e_i = \sum_j I^0_j \Omega_{i,j} + B_i + e_i \quad (10)$$

where  $e_i$  is a statistic fluctuation of the order of  $\sigma(y_{\text{obs}})$ , and (2) is a minimum by definition. In practice, the set  $\mathbf{I}^0$  is not known and the cell and profile parameters are known approximately. The set  $\mathbf{I}^0$  can be considered as a *fixed point* of the n-dimensional space,  $R_+^n$  (positive components), associated with the pattern  $\mathbf{y}_{\text{obs}}$ . If we start from an arbitrary point  $\mathbf{I}^1$ , of  $R_+^n$ , we have to find an

operator,  $\Gamma$ , that acting on that point transforms it into another closer to  $\mathbf{I}^0$ . The term “closer” means that the distance  $d(\mathbf{I}^0, \Gamma[\mathbf{I}^1])$  is smaller than  $d(\mathbf{I}^0, \mathbf{I}^1)$ . Whatever formal definition of distance is good. Actually, due to the degeneracy discussed above,  $\mathbf{I}^0$  is not a single point but belongs to an extended region:  $V(\mathbf{I}^0) \subset \mathbb{R}_+^n$ . A good operator must verify:

$$\mathbf{I}^k = \Gamma[\mathbf{I}^{k-1}] = \Gamma^{k-1}[\mathbf{I}^1] \quad (11)$$

$$\lim_{k \rightarrow \infty} \mathbf{I}^k = \lim_{k \rightarrow \infty} \Gamma^{k-1}[\mathbf{I}^1] = \mathbf{I}^0 \in V(\mathbf{I}^0) \quad (11')$$

The expression provided by Rietveld [1] to estimate the “observed” integrated intensity in order to mimic the classical crystallographic R-factor (now called R-Bragg), defined as expression Whatever Rietveld program can be easily modified to include the possibility of “fitting” the whole profile without structural model using the expression (9) for iterative calculation of the integrated intensities. Of course, the rest of profile parameters, including the propagation vector(s) of a magnetic structure, can be refined simultaneously with the usual LS procedure. The method of “profile matching” is extremely efficient and fast. This procedure has been used first by Taylor [16] and A. Le Bail [17] and it is implemented in the program FullProf [18] developed at the ILL/LLB. This program allows the mixture of the Rietveld and “profile matching” methods for the analysis of a diffraction pattern. This mixed procedure has to be used with caution: the use of both methods simultaneously is necessary when one tries to get the integrated intensities of an unknown structure in presence of another, already known, phase. Thus, no structural parameter of the known phase must be refined. This is the usual case of neutron diffraction patterns of magnetically ordered compounds, where the nuclear reflections coexist with the magnetic reflections.

## **Simulations**

**The text describing the different simulations is not yet written. Some of the figures are given in a separate document.**

## CONCLUSIONS.

- 1- When the refined model is unbiased (no systematic errors), least squares estimates of the standard deviations represent not only the precision but also the accuracy. For biased models (usual practical situation) the standard deviations do not represent accuracy. Simulation completely confirms the theses of E. Prince.
- 2- For biased models (a known systematic error in profile is introduced) in simple structures some fitted parameters (profile and thermal parameters) show a bias while others are barely affected (position parameters).
- 3- Concerning accuracy, nothing is gained repeating the measurement several times if the systematic error is expected to be the same. The standard deviation obtained from different measurement is not, in such a case, a measure of the accuracy.
- 4- For complex structures high statistical accuracy and high resolution is required for getting the true parameter values even if the refined model is unbiased. Suggestion: the solvability index ( $r = N_{\text{eff}}/N_f$  for  $p=1/2$ ) should be largely greater than 4-5 to be sure that the structural parameters are accurate enough. More experience is needed to establish precise rules.
- 5- The absolute value of the profile R-factors has little significance because their values depend on the quality of the data as well as on the goodness of the structural model. The R-factors obtained by a refinement of the whole pattern without structural model provide the “expected” values for the best structural model.
- 6- Well behaved peak shape could be more important than resolution in some cases.



## REFERENCES

1. H.M. Rietveld, J. Appl. Cryst. **2**, 65 (1969).
2. M. Sakata & M. J. Cooper, J. Appl. Cryst. **12**, 554 (1979).
3. E. Prince, J. Appl. Cryst. **14**, 157 (1981)
4. H.G. Scott, J. Appl. Cryst. **16**, 159 (1983)
5. R.J. Hill & I.C. Madsen, J. Appl. Cryst. **17**, 297 (1984),
6. R.J. Hill & C.J. Howard, J. Appl. Cryst **18**, 173 (1985),
7. R.J. Hill & I.C. Madsen, J. Appl. Cryst **19**, 10 (1986)
8. R.J. Hill & H.D. Flack, J. Appl. Cryst **20**, 356 (1987)
9. I.C. Madsen & R.J. Hill, J. Appl. Cryst **21**, 398 (1988)
10. G.S. Pawley, Acta Cryst **B34**, 523 (1978)
11. J.F. Bérar & P. Lelann, J. Appl. Cryst **24**, 1 (1991)
12. Yu.G. Andreev, J. Appl. Cryst **27**, 288 (1994)
13. R.J. Hill & R.X. Fischer, J. Appl. Cryst **23**, 462 (1990)
14. E. Jansen, W. Schäfer & G. Will, J. Appl. Cryst **27**, 492 (1994)
15. G.S. Pawley, J.Applied Crystallogr. **14**, 357 (1981)
16. J.C. Taylor, Zeitschrift für Kristallographie, **176**, 183 (1986) and, Zeitschrift für Kristallographie, **181**, 151 (1987).
17. A. Le Bail, H. Duroy and J.F. Fourquet, Material Research Bulletin, **23**, 447 (1988).
18. J. Rodríguez-Carvajal, "FullProf: A Program for Rietveld Refinement and Profile Matching Analysis of Complex Powder Diffraction Patterns". (ILL, unpublished)

**Few slides.**

## COMPLEXITY VERSUS RESOLUTION

The model used for describing the powder pattern profile is based in the expression

$$y_{\text{ical}} = \sum_{\phi} S_{\phi} \sum_{\mathbf{h}} I_{\phi\mathbf{h}}[\mathbf{B}_I] \Omega(T_i - T_{\phi\mathbf{h}})[\mathbf{B}_P] + B_i[\mathbf{B}_B]$$

$\mathbf{h}$  ( $\mathbf{H}$  or  $\mathbf{H}+\mathbf{k}$ ) labels the Bragg reflections

$\phi$  labels the phase

$I_{\phi\mathbf{h}}[\mathbf{B}_I]$  Integrated intensities depending on structural parameters: atom positions, magnetic moments, ...

$\Omega(T_i - T_{\phi\mathbf{h}})[\mathbf{B}_P]$  Peak shape depending on profile parameters, instrument, defects, domain size, etc.

$B_i[\mathbf{B}_B]$  Background function: diffuse scattering, short range order, parasitic, etc.

Rietveld Method: Minimise the chi-square function against an heterogeneous set of parameters

$$\chi_p^2 = \sum_i w_i \{y_{\text{ioobs}} - y_{\text{ical}}(\mathbf{B})\}^2$$

Reduced chi-square:

$$\chi_v^2 = \chi_p^2 / (N - P + C) \text{ with } P = N_I + N_P + N_B$$

- If one is interested in “structural parameters” the number of independent observations *is not* the number of points in the pattern  $N$ .
- What is the number of “independent” observations?

**There is no rigorous answer...**

## Complexity of a structural problem:

Points to be considered:

- Signal-to-noise ratio, statistics.
- Number of independent Bragg reflections:  $N_B$
- Number of intensity-dependent free parameters:  $N_I = N_f$
- Degree of reflection overlap: resolution versus separation between consecutive reflections.

The above considerations give rise to the concept of:

- Effective number of observations (resolution weighted):  **$N_{eff}$**
- “Solvability” index: ratio between the effective number of observations and the number of structural parameters

$$r = N_{eff}/N_I$$

### Guest

The solvability index must be higher than 5-8 (?) for powder diffraction

### Criteria for calculating $N_{\text{eff}}$ :

Two reflections separated by  $\Delta(Q)$  can be discriminated properly if the following relation holds:

$$\Delta(Q) = 2\pi^2 j / (Q^2 V_o) \geq p D_Q$$

The “separability factor”  $p$  is of the order of the unity

A single reflection at  $Q_o$  contributes to  $N_{\text{eff}}$  as  $x/(x+\text{nearest})$ , where “ $x$ ” is the fraction of the current phase and “nearest” is the number of adjacent reflections, weighted by the corresponding  $x$ , verifying:

$$Q_o - p D_Q \leq Q_{\text{adjacent}} \leq Q_o + p D_Q$$

The formula for calculating  $N_{\text{eff}}$  is:

$$N_{\text{eff}}(\phi) = \sum_{i(\phi)} \{x[\phi:i] / \sum_{j(\phi)} x[\phi:j]\}$$

$x[\phi:i]$ : fractional contribution to the diffraction pattern (area) of the phase “ $\phi$ ” to which belongs the reflection “ $i$ ”

$i(\phi)$ : index of reflections of the phase  $\phi$ .

$j$  index of reflections adjacent to reflection “ $i$ ”. The reflection “ $i$ ” always belong to the set of  $j$

For a single phase  $x[\phi:i]=1$

$$N_{\text{eff}} = \sum_i 1 / (1 + N_i)$$

$N_i$  is the number of reflections in the neighbourhood of reflection “ $i$ ”

## Simulation of “systematic errors” in the Rietveld Method

- The presence of systematic errors in the model prevents the possibility of a rigorous analysis of the “accuracy” of the refined parameters.
- The standard deviations are a measure of the accuracy only when statistical errors dominate the residuals

An empirical approach can help the assessment of the accuracy of the refined parameters:

- (1) Select a structural and profile model (real or artificial) and calculate its theoretical deterministic diffraction pattern. Use different scale factors to simulate different counting times.
- (2) Add a “Poissonian noise” to the deterministic pattern.
- (3) Generate  $N_R$  diffraction patterns corresponding to the same scale factor
- (4) Repeat for  $N_T$  different “counting times” ( $N_R \times N_T$  diffraction patterns)
- (5) Refine, using the Rietveld Method, the generated diffraction patterns with the correct model or with a partial model.
- (6) Analysis of the “fitted” parameters and comparison with the “true values”
- (7) Behaviour of the reliability indices
- (8) Plot results

## Standard deviations

$$\sigma^2(\beta_j) = (A^{-1})_{jj} \cdot \chi^2_v = (A^{-1})_{jj} \chi^2_p / (N-P+C)$$

- ⇒ The standard deviation  $\sigma(\beta_j)$  is a measure of “precision”. It is also a measure of the “accuracy” only if there is no “systematic error”.
- ⇒ In the Rietveld Method the standard deviations calculated with the above formula could be much smaller than the corresponding “accuracy” (case of cell parameters).
- ⇒ There is no general method of converting the standard deviations obtained in least squares to estimated “true errors” of the parameters.
- ⇒ ⇒ The repetition of the measurement in presence of systematic errors does not help. The bias in some parameters remains.
- ⇒ For structural parameters the two step method (extraction of integrated intensities and refinement based on peak clusters) could be applied for getting a new set of standard deviations.