Bayesian method for the analysis of diffraction patterns using *BLAND*

Joseph E. Lesniewski,[a,b,c] Steven M. Disseler,[c] Dylan J. Quintana,[d] Paul A. Kienzle[c] and William D. Ratcliff[c]*

[a]Mount St Mary's University, Maryland, USA, [b]Georgetown University, Washington, DC, USA, [c]NIST Center for Neutron Research, National Institute of Standards and Technology, Maryland, USA, and [d]Carnegie Mellon University, Pennsylvania, USA. *Correspondence e-mail: william.ratcliff@nist.gov
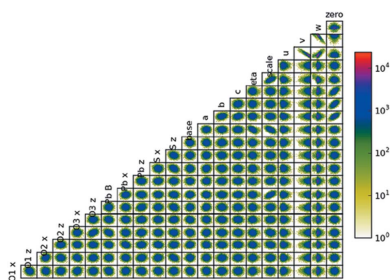
Rietveld refinement of X-ray and neutron diffraction patterns is routinely used to solve crystal and magnetic structures of organic and inorganic materials over many length scales. Despite its success over the past few decades, conventional Rietveld analysis suffers from tedious iterative methodologies, and the unfortunate consequence of many least-squares algorithms discovering local minima that are not the most accurate solutions. Bayesian methods which allow the explicit encoding of *a priori* knowledge pose an attractive alternative to this approach by enhancing the ability to determine the correlations between parameters and to provide a more robust method for model selection. Global approaches also avoid the divergences and local minima often encountered by practitioners of the traditional Rietveld technique. The goal of this work is to demonstrate the effectiveness of an automated Bayesian algorithm for Rietveld refinement of neutron diffraction patterns in the solution of crystallographic and magnetic structures. A new software package, *BLAND* (Bayesian library for analyzing neutron diffraction data), based on the Markov–Chain Monte Carlo minimization routine, is presented. The benefits of such an approach are demonstrated through several examples and compared with traditional refinement techniques.

1. Introduction

For many solid-state materials, only polycrystalline samples are initially available for analysis. Historically, the first approach to the analysis of diffraction data taken from polycrystalline materials was to start with the integrated intensities of the peaks. These intensities were then fitted using the standard techniques of single-crystal crystallography. However, because of peak overlap, a great deal of information was lost. Rietveld (1967, 1969) suggested that instead one should fit the entire pattern, point by point. While a simple proposition in theory, in practice the intensity at each point comes from a variety of factors, with tens or hundreds of variables which must all be fitted or fixed using *a priori* knowledge and experience to obtain quantitative information from the diffraction patterns. Within this framework, the scattered intensity as a function of 2θ is given by

$$I_i = B_i + S \sum_j L_j |F_j|^2 P(2\theta_i - 2\theta_j). \qquad (1)$$

Here, I_i is the intensity at point i, B_i is the background at point i, S is the scale factor, L_j is the Lorentz factor of the jth reflection, F_j is the structure factor (including form factors *etc.* for magnetic structures) and P is the instrument profile of the jth reflection measured at point i. For simplicity, we have neglected absorption, extinction and numerous other effects.

For most diffraction patterns of interest, experiments are performed such that individual peak intensities are hundreds to thousands of counts, meaning that the Poisson noise can be approximated using a Gaussian distribution. Any fit of equation (1) to the data thus results in a χ^2 goodness of fit as described by equation (2):

$$\chi^2 = \sum_i \frac{\left(I_{\mathrm{calc},i} - I_{\mathrm{obs},i}\right)^2}{\sigma_i^2}, \tag{2}$$

where the variance at a given point, σ_i, is simply given by $(I_{\mathrm{obs},i})^{1/2}$. In the traditional Rietveld approaches, including most of the widely available software packages (Rodríguez-Carvajal, 1993; Larson & Von Dreele, 1994; Toby & Von Dreele, 2013; Toby, 2001; Coelho, 2000), it is this goodness of fit or a similar weighted residual function which is minimized using damped nonlinear least-squares techniques such as the Levenberg–Marquardt algorithm. While this works quite well if the starting parameters are close to the ultimate global minimum, the overall multidimensional χ^2 surface is often far from concave and thus it is easy for the solution to become trapped in a local minimum which fails to describe the most accurate model. In other cases the Hessian, or matrix of second partial derivatives, used in Levenberg–Marquardt minimization becomes nearly singular and the refinement diverges before an adequate solution can be found. In practice, the experimenter must have a reasonably good guess for the phase space he or she wishes to explore, leading to significant 'art' involved in adding variables to a refinement.

Maximum likelihood and maximum entropy methods related to Bayesian statistical inference have been used in previous studies of various aspects of crystallographic refinement, but the computational cost has prevented such approaches from gaining widespread use (Gilmore, 1996). More recently, Bayesian analysis has resurfaced in a growing number of approaches as a viable method of pattern refinement, particularly in the case of the determination of microstructural/strain parameters and other systematic deviations which are heavily influenced by a detailed accounting of experimental and refinement error (Wiessner & Angerer, 2014; Gagin & Levin, 2015; Toby & Von Dreele, 2013; Fancher et al., 2016).

With efficient sampling and computing efficiency, however, the types of problems which may be addressed by a Bayesian or probabilistic approach may be greatly expanded. In the present work, we introduce a fully probabilistic method of refining neutron powder and single-crystal diffraction patterns, including magnetic and structural components, using a Bayesian approach. We implement this routine as part of the software package *BLAND* built on the *Bumps* (Kienzle et al., 2015) fitting package which utilizes, among other fitting engines, the differential evolution adaptive Metropolis (*DREAM*) algorithm to traverse parameter space efficiently (Vrugt et al., 2008). In order to demonstrate the power of this method, we present several refinements solved using the *BLAND* package, including examples contained within the *FullProf* suite, for comparison of this method with standard least-squares approaches.

2. Methodology

2.1. Bayes' theorem

We begin by briefly discussing our approach employing Bayes' theorem to model crystallographic parameters akin to Rietveld refinement from diffraction patterns. We note that a full derivation of the Bayesian approach to data analysis can be found in a number of excellent introductory texts on statistical analysis (Jruschke, 2011; Taylor, 1990). Formally stated, we seek to determine the posterior distribution for a set of parameters, φ, for a parametric model, μ, given the experimentally observed intensities, I_{obs}. Casting this in the language of Bayes' theorem this is defined as

$$p(\varphi \mid I_{\mathrm{obs}}, \mu) = \frac{p(I_{\mathrm{obs}} \mid \varphi, \mu)\, p(\varphi \mid \mu)}{p(I_{\mathrm{obs}} \mid \mu)}, \tag{3}$$

where $p(\varphi \mid I_{\mathrm{obs}}, \mu)$ is the probability of obtaining a vector of variables φ given the experimental observed intensities I_{obs} and model μ. Here, $p(\varphi \mid \mu)$ is the *a priori* distribution of φ, typically assumed to be uniformly distributed over parameter space unless there is prior information to restrict the value of a parameter.

In the case of a diffraction measurement, φ includes all unrestrained quantities such as lattice parameters, atomic positions and thermal displacement parameters, and any variables describing the magnetic structure if required. In contrast, model-defining quantities such as the space group, the profile shape function for powder diffraction and any other fixed parameters are contained in μ.

While this objective sounds reasonable, in practice equation (3) is determined by maximizing the likelihood function $p(I_{\mathrm{obs}} \mid \varphi, \mu)$. Even with the assumption of normally distributed parameters, with $p \propto \exp[-\chi^2(\varphi)/2]$ for χ^2 defined by equation (2), this entails sampling over a substantial range of the multidimensional space defined by the bounds of φ. If one is given detailed information about the starting values of φ then one may be able to confine the search to a narrow region and perform efficient searches, but this would then defeat our originally stated goal of minimizing the amount of *a priori* knowledge. An important example of why this is necessary is in the case of complex magnetic structures where there may be many independent components or propagation vectors which are not reasonably constrained by representation theory alone (Bertaut, 1968).

Therefore, to use a Bayesian approach in a reasonable computational time frame we must make efficient choices in the way φ space is sampled. Many such algorithms devoted solely to this global optimization problem have been proposed since the Metropolis algorithm (Metropolis et al., 1953), but a drawback of many of these is that, even if they can be proved to converge, they converge very slowly. Related methods such as simulated annealing and simple Markov–Chain Monte Carlo (MCMC) require careful tuning of the statistical

temperature profile by the user to achieve this criterion. Differential evolution (DE) is a rather efficient algorithm for problems with many minima or non-differentiable χ^2 surfaces, but its control parameters need to be tuned to the problem in order to obtain good performance, and even then it can become trapped in a local minimum. We have found that the *DREAM* algorithm is highly convergent and requires no additional tuning beyond identifying the parameter ranges (Vrugt *et al.*, 2008), so we use it as the primary optimization engine within *BLAND*.

2.2. *DREAM* algorithm

DREAM uses an adaptive MCMC algorithm that combines DE with the random-walk Metropolis algorithm in order to provide global optimization of highly nonlinear complex problems, while maintaining good efficiency (Vrugt *et al.*, 2008; ter Braak, 2006). Starting with a population of Markov chains, DE guides the chains through the search space towards the solution, while the Metropolis algorithm prevents them from being trapped by local minima.

2.3. *Bumps* and *BLAND*

Bumps is a Python package which implements the *DREAM* algorithm, among others (Kienzle *et al.*, 2015). It provides a generalized fitting framework for use with multi-parameter problems where the problem is described by the negative log-likelihood and it has been parallelized for both multi-core systems and MPI-based clusters. This backbone has been successfully implemented in neutron reflectometry (Kienzle *et al.*, 2011) and small-angle scattering (Butler *et al.*, 2013) software packages which also suffer from a similar problem of non-analytic and multi-modal χ^2 surfaces.

BLAND utilizes the general framework of *Bumps*, not only to determine the best fit of the model to the diffraction data, but also to illustrate clearly any correlations between various parameters over an extremely large parameter space. The underlying crystallographic calculations of structure factors, multiplicities *etc.* are provided by *CrysFML*, the Crystallographic Fortran Modules Library (Rodriguez-Carvajal, 2001). *CrysFML* is a collection of modules that provide numerous crystallographic calculations for use by other Fortran programs, including *FullProf* (Rodriguez-Carvajal, 2001; Rodríguez-Carvajal, 1993). By utilizing this well established collection of general routines, *BLAND* may be tuned to handle any number of crystallographic refinement problems, including magnetism, single crystals *etc.* (Lesniewski *et al.*, 2016).

3. Application of *BLAND* to powder neutron diffraction patterns

3.1. Nuclear crystal structures

In order to demonstrate the advantages of the Bayesian method implemented by *BLAND*, and to compare its accuracy with that of traditional Rietveld analysis, we have performed a series of model refinements on example data sets of simple

materials and those packaged with *Fullprof* as examples (Rodríguez-Carvajal, 1993). This includes the neutron diffraction pattern of the corundum phase of Al_2O_3 and orthorhombic $PbSO_4$, the latter of which was used in a round-robin study evaluating the systematic differences of various crystallographic analysis software packages (Hill, 1992). We also examine the low-temperature diffraction pattern of CuF_2 (Fischer *et al.*, 1974) to show how *BLAND* may be used to determine the simultaneous solution of nuclear and magnetic structures, noting that this approach has been used successfully to solve or verify the magnetic structures in several other materials thus far (Disseler *et al.*, 2015; Maruyama *et al.*, 2014).

For each example the parameters of interest were constrained only in that the results be physically meaningful; for example, atomic positions were confined only with displacements limited to the maximum dimension of the unit cell, and with occupancies fixed by the chemical formula units per unit cell or Wyckoff sites. The initial distribution of MCMC chains within the multidimensional parameter space was selected according to the Latin hypercube sampling (LHS) routine (McKay *et al.*, 1979). This ensures sufficient distribution of the initial parameters in order to remove artificial bias towards a known solution, and that all regions of parameter space are sampled. The experimental background was taken to be a simple linear interpolation of select points taken from the observed data, with an additional additive constant, or 'base' value, used as a fine-tuning parameter for each refinement.

We introduce the various outputs from the *BLAND* package by first demonstrating the refinement of an Al_2O_3 phase (corundum). A polycrystalline sample was measured on the BT-1 powder diffractomter at the NIST Center for Neutron Research using neutrons of wavelength λ = 1.5403 Å. The *BLAND* package was used to refine the nuclear structure, assuming pseudo-Voigt peak shapes with the standard empirical profile function $H^2 = U^2 \tan^2(\theta) + V \tan(\theta) + W$ and a Gaussian–Lorentzian interpolating parameter η ('eta' in the figures) to describe the peak width and shape as a function of diffraction angle. The lattice parameters, atomic positions and thermal displacement parameters, the four profile parameters, and the overall scale and background were all fitted simultaneously, with no constraints other than the fixed symmetries defined by the atomic Wyckoff site for each ion in the $R\bar{3}c$ space group. The resulting best-fit diffraction profile is shown in Fig. 1(*a*), with the individual Bragg peaks and difference shown below the best-fit line in the figure. A value of χ^2 = 7.39 was determined, with the error stemming mostly from a disagreement with the peak shape function over the entire range. In this case, χ^2 could be further reduced by implementing a more advanced peak shape or asymmetric profile functions.

In Fig. 1(*b*) we show the probability distribution of each of the parameters fitted using the LHS initialization method. The probability density of each continuous variable is given by the histogram in each respective panel, where the height of each bar is given by the number of cycles or steps taken in a given parameter bin value during the course of the fit, normalized to

the total number of cycles. The corresponding green line shows the largest log-likelihood value within the corresponding histogram bin. If the maximum of the likelihood (green) line is not coincident with the maximum of the histogram distribution, this may indicate that the fit has found a better value but stopped running before exploring the parameter space around that value, or it may indicate odd correlations between variables. In either case, more cycles would be needed to obtain the proper error distribution or values of the parameter. Other plots not shown here are displayed within *Bumps* to judge the quality of the fit, including the collective log-likelihoods at each step and parameter values to ensure the fit is not stuck in one region of space. The 68% confidence interval, equal to one standard
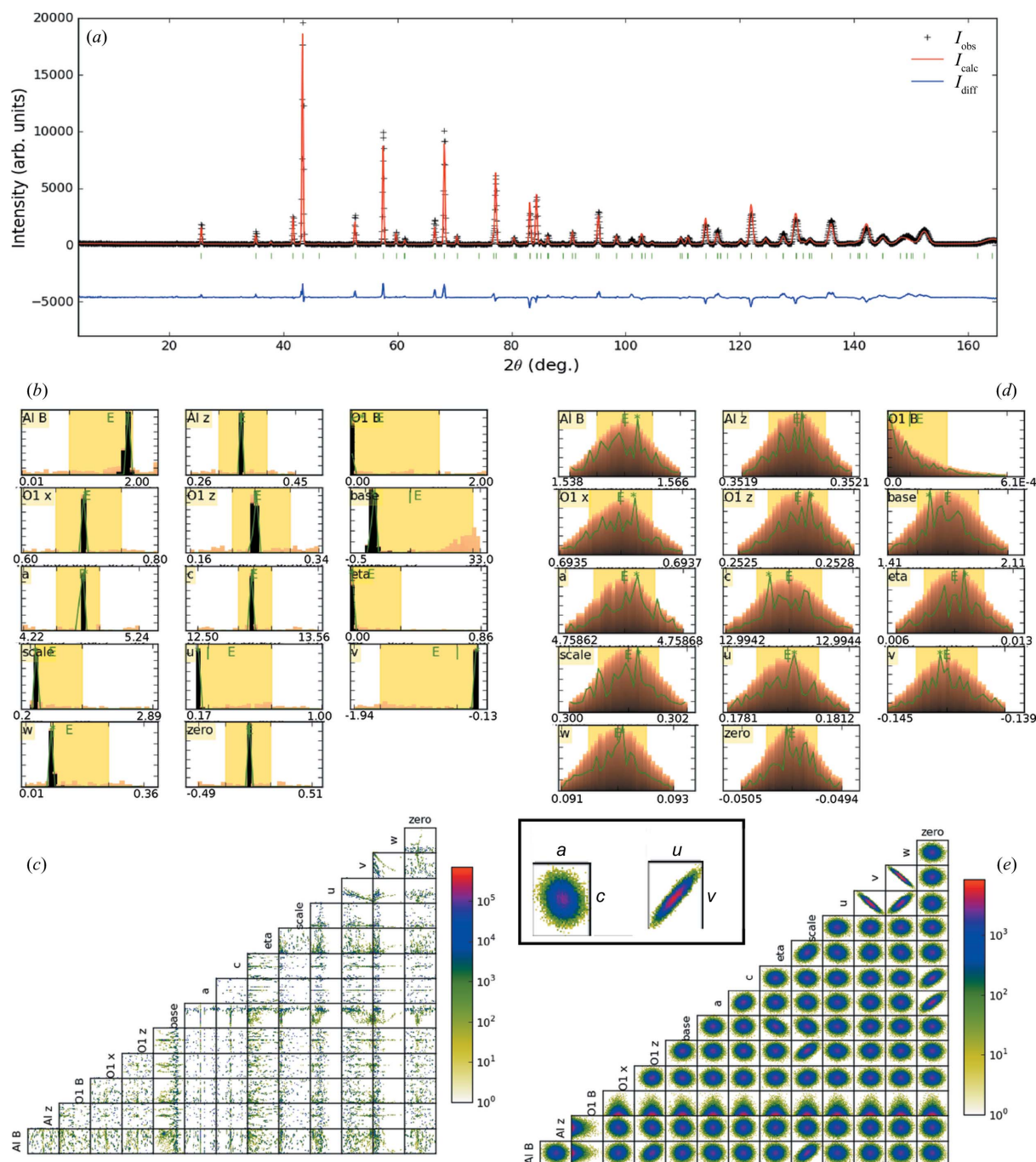

Figure 1
BLAND refinement of Al_2O_3. (*a*) Observed data (+), calculated pattern (red), difference (blue, offset) and locations of Bragg reflections (green bars). (*b*) Probability distributions for each refined parameter from the wide range LHS sampling method; the total area under each plot is normalized to unity. The *x*-axis values in each panel denote the bounding values of the 95% confidence interval. (*c*) Correlation matrix between parameters. (*d*), (*e*) Parameter distribution and correlation plots, respectively, both determined using the epsilon-ball approach near the optimized parameters. The inset shows examples of uncorrelated and correlated parameter distributions.

deviation for normal distribution, is given by the lightly shaded region centered at the maximum of the distribution, while the values corresponding to the 95% confidence interval are labeled on the x axis in each panel for scale. From this, we see the values peak very strongly over a narrow window, indicating a single well defined solution. Parameters such as the scale factor and thermal displacement have peaks close to their respective lower boundary, as the sampling range is not symmetric owing to the imposed physically meaningful boundaries.

A subsequent fit was performed beginning with these best-fit parameters using the so-called 'epsilon-ball' approach, in which the starting population is initialized near the provided parameters and the solver explores parameter space by expanding around this point. This tends to produce an excellent description of the parameter space near a known solution, but it does not quickly traverse a large area of parameter space if the initial conditions are not close to a correct solution. The histograms of the individual parameters from this second fit are shown in Fig. 1(d) and, with the exception of the O1 thermal displacement parameter, all are normally distributed over a very narrow window, indicating the solution found from the LHS initialization search is indeed a strong minimum on the χ^2 surface.

In Figs. 1(c) and 1(e) we show the two-dimensional correlation plots between pairs of parameters generated as a part of the *DREAM* algorithm utilized in the *BLAND* package for the LHS and subsequent epsilon-ball fits, respectively. The peaks in the histograms for each parameter found in Fig. 1(b) are quite sharp relative to the range, and are therefore represented by only a small number of red pixels in Fig. 1(c). The correlations near the minimum provide more information in this case, as shown in Fig. 1(e). We note that the information represented here is different from the uncertainties estimated from the covariance matrix determined after Levenberg–Marquardt least-squares refinement, which yields the parameter sensitivity rather than parameter uncertainty.

In each panel in Fig. 1(e), a tightly clustered circular pattern in the center of a box indicates that the fit was able to determine a value for the parameter and that there is no strong correlation between the values of the two parameters; they are essentially independent of each other. This can be seen in Fig. 1(e) where the a and c lattice parameters are uncorrelated. On the other hand, patterns indicating a strongly correlated relationship between two parameters are seen as elongated ellipsoids with a slope which depends on the sign of the correlation. An example of this type of pattern can be seen in the peak shape parameters, u, v and w, and to a lesser extent the scale and oxygen position. If the refinement depends strongly on one parameter but not the other, the distribution will appear as a horizontal or vertical band with no slope.

Other possible correlation patterns not observed here include a box filled completely with a random distribution of points, indicating that the fit was unable to confidently determine a value for either of the parameters, and therefore the observed diffraction intensity is not sensitive to either parameter. When sampling over a large parameter space one

Table 1
List of refined parameters for $PbSO_4$ determined using *BLAND* and *FullProf*.

For *BLAND*, $R_f = 5.06$, $\chi^2 = 2.07$; for *FullProf*, $R_f = 2.71$, $\chi^2 = 4.2$.

Parameter	*BLAND* (epsilon-ball)	*BLAND* (LHS)	*FullProf*
a (Å)	8.47818 (10)	8.478	8.47883 (15)
b (Å)	5.397039 (67)	5.397	5.3967195 (99)
c (Å)	6.958488 (89)	6.9585	6.9583 (34)
x_{Pb}	0.18739 (11)	0.188	0.18749 (9)
z_{Pb}	0.16704 (17)	0.167	0.16719 (15)
b_{Pb}	0.890 (21)	0.969	1.42 (2)
x_S	0.06424 (35)	0.064	0.0654 (3)
z_S	0.67844 (49)	0.068	0.6833 (4)
x_{O1}	0.90720 (20)	0.908	0.9076 (2)
z_{O1}	0.59579 (23)	0.595	0.5953 (2)
x_{O2}	0.19324 (20)	0.193	0.1937 (2)
z_{O2}	0.54217 (25)	0.542	0.5432 (3)
x_{O3}	0.08093 (13)	0.081	0.0810 (1)
z_{O3}	0.80940 (16)	0.809	0.80905 (15)

may also find a multimodal distribution in the correlation plots. This is indicative of an additional symmetry which was not explicitly limited by the original model; obvious examples include inversion-symmetry-breaking distortions corresponding to distinct ferroelectric domains, and the sign of a ferromagnetic moment due to the breaking of time-reversal symmetry. In both cases we have observed symmetric maxima in the probability distributions spaced evenly around the paraelectric or paramagnetic value of the respective parameter, indicating that *Bumps* is indeed returning appropriate probabilities. Multiple maxima can also occur if symmetry-related atomic positions for a given atom lie within the constrained region of parameter space, as will be described in detail in the following example.

Most materials of interest are much more complex than Al_2O_3, however, leading to a large increase in the number of variables and the dimensionality of the problem. In general, this also decreases the likelihood that a least-squares algorithm will find the correct minimum χ^2 solution without extensive guidance. To demonstrate how *BLAND* handles such problems, we have refined the room-temperature neutron diffraction pattern for $PbSO_4$, originally measured on the D1A diffractometer at the Institut Laue–Langevin (Grenoble, France) and used in a previous round-robin study of different refinement software packages (Hill, 1992). This compound is orthorhombic in the *Pnma* space group and has atomic positions at relatively low symmetry positions, such that there are a significant number of parameters which must be refined. Here, we have refined lattice constants, atomic positions and thermal displacement parameters for each species in the unit cell listed in Table 1, for a total of 17 free parameters.

The data have been fitted using the epsilon-ball approach beginning with crystallographic parameters determined previously (Hill, 1992), and well as with the full LHS initialization, again over a wide parameter range. For both initializations, the lattice parameters were limited to a range of ±0.5 Å about the known values. The atomic displacements

were allowed to vary by up to 60% along each unit-cell dimension, centered about the known positions.

BLAND produces an excellent fit of the reported data, with similar χ^2 values for both initialization approaches ($\chi^2 \simeq 2.0$). The refinement from the epsilon-ball initialization is shown in Fig. 1(*a*). Here again, the primary source of error between the reported and fitted profiles stems from the lack of higher-

order corrections, such as asymmetric instrumental broadening and other angle-dependent corrections to the peak shapes or widths, which are not currently applied within the *BLAND* routines.

While the final results and best-fit solutions are quite similar for both initialization conditions, the probability distributions of each parameter are quite different. In the case of the
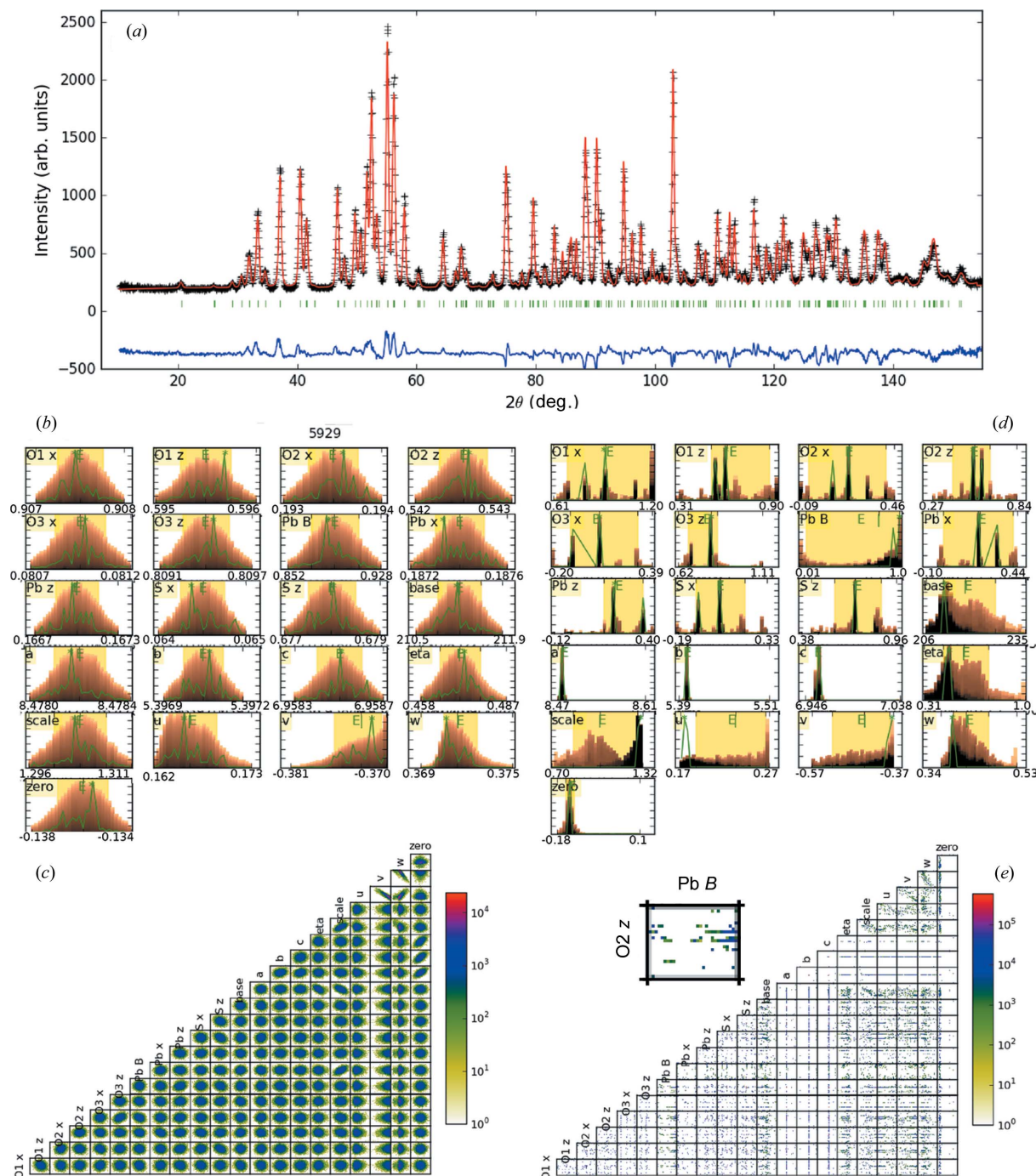

Figure 2
BLAND refinement of PbSO₄. (*a*) Best fit including observed data (+), calculated pattern (red), difference (blue, offset) and locations of Bragg reflections (green bars). (*b*), (*d*) Probability distributions for each parameter after runs with initial starting populations that were (*b*) tightly spaced near the expected values or (*d*) randomly distributed over space. It should be noted that the horizontal scale in each panel of part (*d*) is much larger than that presented in part (*b*) and thus the panels should not be compared directly. (*c*), (*e*) Correlation matrices between parameters for each of the two starting populations, respectively.

epsilon ball, the parameters shown in Fig. 2(b) are all normally distributed and centered near the initial conditions, with the best-fit values also in good agreement with this center position. On the other hand, when the initial parameters are widely distributed according to the LHS routine, as would be the case if the atomic positions were completely unknown, we find multiple peaks in the distribution of the atomic positions. Upon closer examination, each of these maxima corresponds to one of two features: (i) symmetrically equivalent positions generated by the symmetry operators of the space group, or (ii) switching of identical atom types sharing the same Wyckoff symmetry, such as O1 and O2 atomic x positions, both of which result in identical crystallographic structures.

The multimodal distribution of the atomic positions is also observed in the correlation plots shown in Fig. 2(e). The enlarged panel highlights the correlations between the O2 z and Pb B parameters. Here, the horizontal bands correspond to distinct high-likelihood values of the O2 z parameter, which are each largely independent of the value of Pb B. Each maximum in the distribution appears as a distinct set of bands and may be independently correlated from other parameters. In the case of the oxygen positions, the intersections of these bands results in 'patches' of parameter space corresponding to the various symmetry-generated or switched oxygen positions. Near the ideal position, we find that the parameters are mostly independent of one another from the correlation plot in Fig. 2(c). In fact, only the peak-shape refinement parameters exhibit any substantial dependence, as expected for a phenomenological function. It is this lack of correlation between parameters that allows different Rietveld refinement approaches based on least-squares routines to result in precise and accurate refinements, even with the exceedingly large number of parameters here (Hill, 1992).

The resulting atomic displacements are shown in Table 1, where we compare them with those obtained for the full refinement obtained by the *FullProf* program beginning with these same values. The error shown in parentheses is given as one standard deviation, or a 68% confidence interval for *BLAND*. One can see that the values obtained by *BLAND* for the epsilon ball are within the error of those obtained using *FullProf* and exhibit similar measures of error, demonstrating that *BLAND* has been implemented correctly and dependably. Unlike the epsilon-ball approach, the uncertainty calculated from the distributions in Fig. 2(d) encompasses several minima and thus the error bars are not representative of true parameter error. With the locations of the minima known, however, detailed errors and model comparisons could then be easily obtained by performing the fit over a more limited range or with the epsilon-ball approach.

We note that, while this method provides a powerful approach to solving unknown crystallographic structures, searching over such a wide range of parameter space leads to a dramatic loss in the speed of convergence. For example, to obtain the fits shown for Al_2O_3 and the epsilon-ball approach to $PbSO_4$ required fewer than a thousand steps, and even fewer to obtain the statistics necessary if one simply wanted an estimation. By comparison, the random initialization approach used for $PbSO_4$ required over an order of magnitude more steps to obtain sufficient statistics and to discern each of the independent minima shown Fig. 2(d). In practice, one would ideally utilize such a broad search if little *a priori* information was known, then greatly reduce the parameter range to isolate a single maximum for rapid fitting. The tools within *Bumps* provide valuable feedback for estimating performance and convergence criteria, particularly when parallelization is used in multi-core or cluster systems (Kienzle *et al.*, 2015).

3.2. Magnetic structures

In addition to refining the nuclear or crystalline structures as examined in previous implementations of Bayes' theorem (Wiessner & Angerer, 2014; Gagin & Levin, 2015), we have extended *BLAND* to make use of the magnetic structure calculations within the *CrysFML* library (Rodriguez-Carvajal, 2001). As a means of demonstrating the effectiveness of this approach, we use example data also found in the *FullProf* example libraries for CuF_2. This monoclinic compound (space group $P2_1/n$) orders antiferromagnetically, with a magnetic supercell described by a unit-cell doubling along the a and c directions, or equivalently by a propagation vector $\mathbf{k} = (\frac{1}{2}, 0, \frac{1}{2})$ (Fischer *et al.*, 1974). From a symmetry analysis following the theory of irreducible representations (Bertaut, 1968), one finds that the magnetic moment of the Cu atom is, in principle, described by three basis vectors corresponding to each of the crystallographic directions. However, from the extinction of specific peaks, only moments along the b axis are allowed.

In Fig. 3 we demonstrate a fit of the low-temperature data using *BLAND*. In addition to parameters such as the lattice constants and peak shapes, we consider an additional parameter corresponding to the coefficient of the irreducible representation basis vector defining the magnitude of the magnetic moments. A wide search is first performed using the LHS initialization, the results of which are shown in Figs. 3(b) and 3(c). The magnetic parameter C_0 clearly shows two maxima at symmetric points about zero, corresponding to a symmetry in the sign of the magnetic moment to point along either the positive or the negative b direction. One would therefore naturally expect domains of both types to be present in the material in bulk.

A finer fit is performed using the epsilon-ball initialization about these best-fit parameters, with the additional constraint that the magnetic moment be positive. The parameter distributions and correlation plots for this fit are shown in Figs. 3(d) and 3(e), respectively. The magnetic moment from this refinement is 0.793 ± 0.012 μ_B Cu. This is slightly larger than one standard deviation from that determined using *Fullprof* directly (0.757 ± 0.015 μ_B), quite close considering the small number and intensity of the magnetic peaks compared with the much brighter nuclear reflections. This fit resulted in $\chi^2 = 4.22$, where again the largest source of disagreement stems from subtleties in the peak shape not captured by the simple pseudo-Voigt peak shape used here. From the correlation plots in Figs. 3(c) and 3(e), most parameters are largely
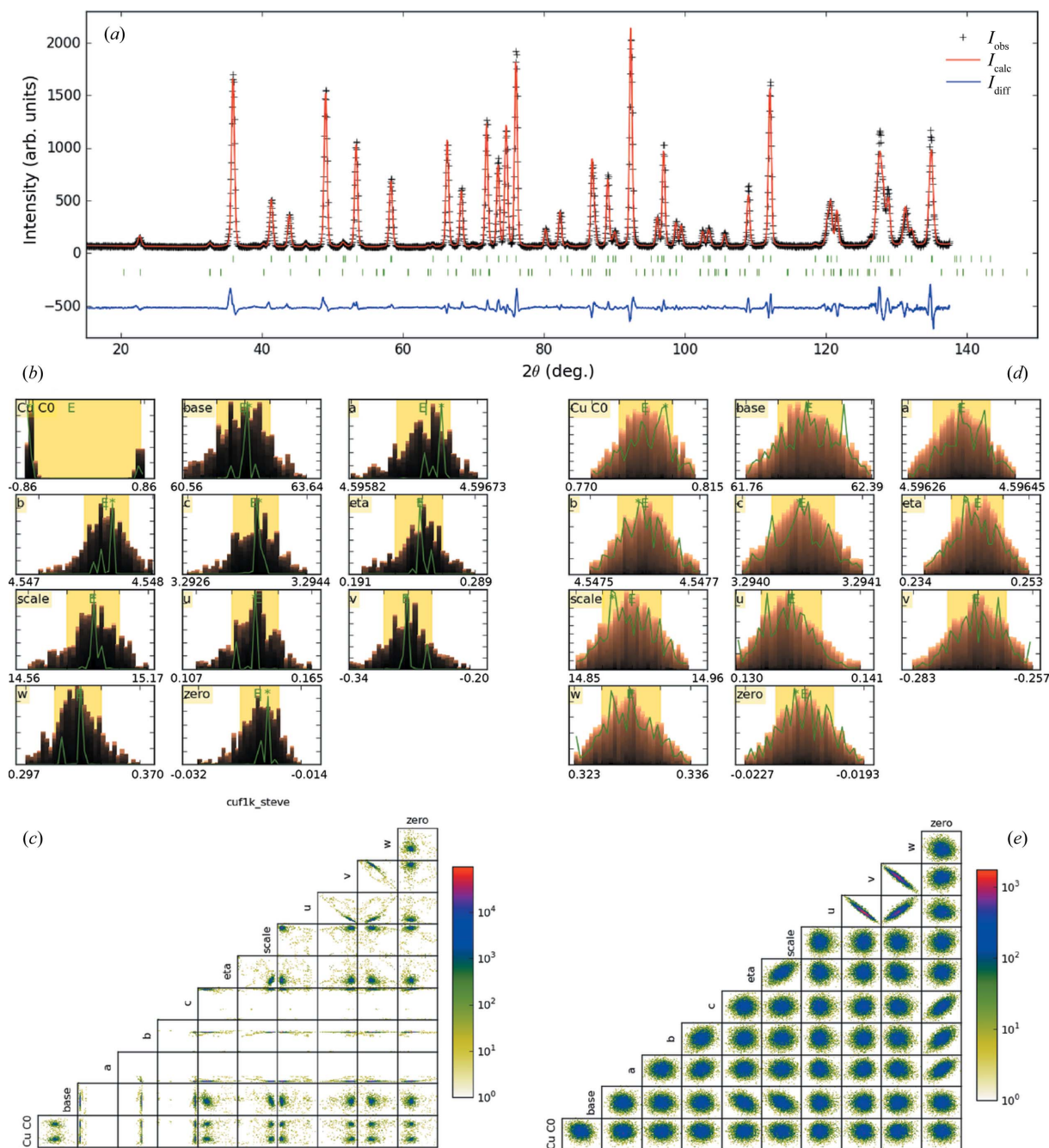
Figure 3
(a) *BLAND* model refinement for magnetically ordered CuF$_2$, showing observed data (+), calculated pattern (red), difference (blue, offset) and locations of Bragg reflections (green bars). The top are nuclear Bragg peaks and the bottom are the locations of magnetic Bragg peaks. (b) Probability distributions determined for each refined parameter as noted, determined from LHS initialization. (c) Correlation matrix between parameters for LHS. (d) and (e) are similar to parts (b) and (c), but using the epsilon-ball initialization about the LHS best-fit parameters.

independent of one another, with the exception of the peak-width profile function parameters.

4. Conclusions

Bayesian statistical inference methods are a powerful alternative to the standard Reitveld technique for the analysis of crystallographic and magnetic structures. As implemented in the *BLAND* algorithm, one gains substantial statistical infor-

mation on the correlation, errors and distributions of parameters compared with that possible from simple least-squares routines. By employing Markov–Chain and differential evolution-based parameter sampling routines, *BLAND* can more accurately fit high-dimensional parameter spaces with mulitmodal χ^2 surfaces. In this work we have demonstrated that this method accurately reproduces crystallographic and magnetic structures obtained by standard Rietveld refinement without *a priori* knowledge of the precise atomic positions,

magnetic moments or other instrumentation parameters. In doing so, we have shown that this package is amenable even when only very wide limits can be placed on various parameter ranges.

Importantly, we have demonstrated that this method can find adequate solutions when only the space group, composition and site symmetries are known, using a true random initialization of the starting values over a wide range of parameter space. We have also demonstrated that this method yields far greater statistical information about parameter distributions and correlations in the multidimensional space than current least-squares-based approaches. The Bayesian approach implemented in *BLAND* to determine the log-likelihood distribution also allows for detailed comparisons of different models which are not subsets of one another. This lends itself naturally to an extension based on information theory approaches such as Bayesian information criteria or Akaike information criteria (Jruschke, 2011) in exploring whole families of models. An obvious example of this would be in exploring various subsets of crystallographic structures about a lattice distortion, or when other experimental evidence suggests a number of different possible space groups.

References

Bertaut, E. F. (1968). *Acta Cryst.* A**24**, 217–231.
Braak, C. J. F. ter (2006). *Stat. Comput.* **16**, 239–249.
Butler, P. A., Doucet, M., Jackson, A. & King, S. (2013). *SasView for Small-Angle Scattering Analysis*, http://www.sasview.org/.
Coelho, A. A. (2000). *TOPAS*. Version 2.0. Bruker AXS, Karlsruhe, Germany.
Disseler, S. M., Luo, X., Gao, B., Oh, Y. S., Hu, R., Wang, Y., Quintana, D., Zhang, A., Huang, Q., Lau, J., Paul, R., Lynn, J. W., Cheong, S. & Ratcliff, W. (2015). *Phys. Rev. B*, **92**, 054435.
Fancher, C. M., Han, Z., Levin, I., Page, K., Reich, B. J., Smith, R. C., Wilson, A. G. & Jones, J. L. (2016). *Sci. Rep.* **6**, 31625.
Fischer, P., Hälg, W., Schwarzenbach, D. & Gamsjäger, H. (1974). *J. Phys. Chem. Solids*, **35**, 1683–1689.
Gagin, A. & Levin, I. (2015). *J. Appl. Cryst.* **48**, 1201–1211.
Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.
Hill, R. J. (1992). *J. Appl. Cryst.* **25**, 589–610.
Jruschke, J. (2011). *Doing Bayesian Data Analysis*. Burlington: Academic Press.
Kienzle, P. A., Krycka, J., Patel, N. & Sahin, I. (2011). *NCNR Reflectometry Software*, *Reflpak*, http://www.ncnr.nist.gov/reflpak.
Kienzle, P., Krycka, J., Patel, N. & Sahin, I. (2015). *Bumps*, Version 0.7.5.7, https://github.com/bumps/bumps and https://zenodo.org/badge/latestdoi/18489/bumps.bumps.
Larson, A. C. & Von Dreele, R. B. (1994). *GSAS*. Report LAUR 86-748. Los Alamos National Laboratory, New Mexico, USA.
Lesniewski, J., Kienzle, P. & Ratcliff, W. (2016). *BLAND Beta Release*, https://github.com/scattering/pycrysfml and https://zenodo.org/badge/latestdoi/10650860.
Maruyama, S., Anbusathaiah, V., Fennell, A., Enderle, M., Takeuchi, I. & Ratcliff, W. D. (2014). *APL Mater.* **2**, 116106.
McKay, M. D., Beckman, R. J. & Conover, W. J. (1979). *Technometrics*, **2**, 239–245.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087.
Rietveld, H. M. (1967). *Acta Cryst.* **22**, 151–152.
Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.
Rodríguez-Carvajal, J. (1993). *Phys. B Condens. Matter*, **192**, 55–69.
Rodríguez-Carvajal, J. (2001). *IUCr Commission on Powder Diffraction Newsletter*, No. 26, pp. 12–19.
Taylor, J. K. (1990). *Statistical Techniques for Data Analysis*. Chelsea: Lewis Publishers.
Toby, B. H. (2001). *J. Appl. Cryst.* **34**, 210–213.
Toby, B. H. & Von Dreele, R. B. (2013). *J. Appl. Cryst.* **46**, 544–549.
Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M. & Robinson, B. A. (2008). *Water Resour. Res.* **44**, W00B09.
Wiessner, M. & Angerer, P. (2014). *J. Appl. Cryst.* **47**, 1819–1825.