M&C 2017 - International Conference on Mathematics & Computational Methods Applied to Nuclear Science & Engineering, Jeju, Korea, April 16-20, 2017, on USB (2017)

Hierarchical Bayesian Modeling of Atomic Structural Disorder

Karl Pazdernik,* Brian J. Reich,* Katharine Page,† Alyson G. Wilson*

*Department of Statistics, North Carolina State University, 2311 Stinson Dr., Raleigh, North Carolina 27695
†Chemical and Engineering Materials Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831
ktpazder@ncsu.edu*

Abstract - *Many functional properties of different materials can be determined by an understanding of the local structure and atomic-level disorder. This insight is critical to the nuclear science community as it directly impacts the quality of materials and forensic science. Reverse Monte Carlo is a general method of structural modeling that has made it possible to generate three-dimensional structural models whose atomic configurations are consistent with experimental data. Although data and models are consistent, the specific structural model acquired from the Reverse Monte Carlo algorithm is a single non-unique solution. Therefore, multiple runs of the RMC algorithm are required to quantify the variability in model solutions and the material structures that they are meant to represent. We propose a Bayesian approach to structural modeling that provides improved estimates of the variances of all atomic coordinates without requiring multiple runs of the estimation algorithm.*

I. INTRODUCTION

Understanding the composition and spatial configuration of atoms within a material is critical to advance many areas of science and technology [1]. Knowledge of the atomic disorder can directly lead to the discovery of new materials with improved properties [2]. In particular, materials that are directly related to nuclear science could be improved, such as semiconductor detectors and scintillators [3, 4]. Understanding crystalline properties can also be useful to nuclear forensic science for inferring sample provenance [5, 6].

We can infer the atom location by the Reverse Monte Carlo (RMC) algorithm [7] which produces a set of plausible atomic locations consistent with the experimental data. However, the resulting solution set is not unique. Further, as a single realization of reasonable locations, it does not include any formal error quantification. Therefore, we consider the use of a hierarchical Bayesian framework that quantifies joint uncertainties of all of the atomic location estimates and provides a natural way to assess modeling assumptions.

II. ATOMIC STRUCTURAL MODELLING

1. Data Collection

Materials scientists describe the structure of materials in terms of the patterns that are combined to fill space, known as unit cells [8]. This substructure is then repeated to create a supercell (see example of barium titanate ($BaTiO_3$) in Figure 1). In $BaTiO_3$, a single barium atom, neighboring titanium atom, and three adjacent oxygen atoms describe the unit cell. The complete lattice, i.e. supercell, is then easily constructed by repeating the pattern of these six atoms throughout the predetermined domain. To improve materials, such as those used in semiconductor detectors and scintillators, scientists must understand their electrical and mechanical properties. That understanding comes from measuring the relative distance between all atoms in the supercell.

Unfortunately, the science of measuring relative atom distance is challenging due to the sub-microscopic scale of the measurements. Since direct measurements of distances are not attainable, diffraction methods are used to determine pairwise distances between all atoms in a supercell. Diffraction methods, including X-ray [9], electron [10], and neutron diffraction [11], involve a sample of material being bombarded with beams or subatomic particles, resulting in angles and intensity of diffraction. This pattern is then used to quantify the average number of atoms that occur as a function of relative distance.

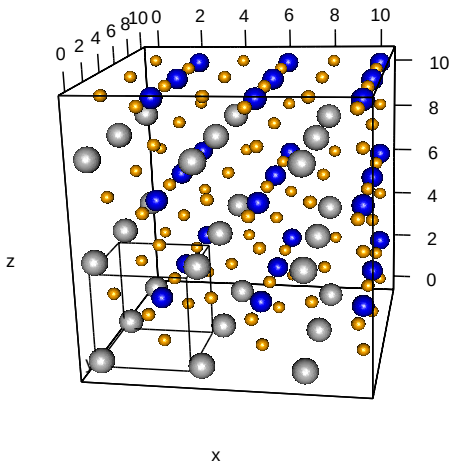

Fig. 1. Example supercell of barium titanate ($BaTiO_3$). Barium is grey, titanium is blue, and oxygen is orange. The smaller cube represents the unit cell structure.

One of the necessary steps in order to translate patterns of intensity into a function of relative distance involves describing the global pattern, i.e. the unit cell structure, in terms of

the angles of diffraction. The peaks of high intensity as a function of angle identify where common angles of diffraction exist. By fitting this curve while simultaneously estimating parameters that control atom positions within a unit cell, the global structure of the material is estimated. For X-ray and neutron powder diffraction, a commonly used method of global structure analysis is Rietveld refinement [12].

Unfortunately, this only provides information about the average unit cell structure and a material's properties can only be fully understood once patterns across and within different unit cells are illuminated. The RMC algorithm was developed to close this gap. Although this method is applicable to different forms of diffraction data (see [13] for details), to simplify notation, the following section assumes that the experimental data was obtained through neutron diffraction.

2. Reverse Monte Carlo

Reverse Monte Carlo (RMC) is an algorithm used for structural modeling where the result is a finite set of atomic coordinates ($\boldsymbol{s}$), that is consistent with experimental data. In order to estimate $\boldsymbol{s}$, the algorithm begins with an initial set of hypothesized atoms locations, which is typically modeled after *a priori* information about the general structure of the material. As discussed, for powder experiments, this information can be obtained from Rietveld refinement of powder diffraction data, or by other means of global crystal structure determination. With the properly constructed unit cell, the initial supercell is created by replicating the pattern across a user defined domain.

Given this initial configuration, the frequency of distances between atom locations is averaged and scaled to create the partial pair distribution functions,

$$g_{ij}(r) = \frac{\overline{n_{ij}(r)}}{(4\pi r^2 dr)(c_i \rho_0)}, \tag{1}$$

where $\overline{n_{ij}(r)}$ is the average number of atoms of type j at a distance of between rÅ and $(r + dr)$Å from a central atom of type i, c_i is the proportion of atoms of type i in the supercell, and ρ_0 is the average number density of the material, i.e. the total number of atoms divided by the total volume of the supercell (N/V). This information is then summarized into the total pair distribution function, defined as

$$G(r) = \sum_{i,j} c_i c_j b_i b_j [g_{ij}(r) - 1], \tag{2}$$

where b_i is the coherent neutron scattering length of atom type i (or the x-ray atomic form factor in the case of x-ray scattering data). This function is often normalized, thus representing the data in different units [14]. A few common transformations are the following:

$$G^{PDF}(r) = 4\pi r \rho_0 G(r) \left(\sum_i c_i b_i \right)^{-2}, \tag{3}$$

$$D(r) = 4\pi r \rho_0 G(r). \tag{4}$$

The RMC algorithm is an updating scheme that uses a normalized total pair distribution function to measure discrepancy between potential atomic configurations and the data. Specifically, denote the hypothesized and scaled total pair distribution function at iteration t as $D^t(\cdot)$ and the measured or "experimental" scaled total pair distribution function as $D^E(\cdot)$.

Although the partial pair distribution functions, and by extension the total pair distribution function, are continuous functions of r, experimental data is discretized to facilitate computation. Therefore, the hypothesized total pair distribution function must also be similarly discretized to evaluate pairwise differences. Let $\{(r_l, r_l + dr) : l = \{1, \ldots, L\}\}$ be the set of intervals used to bin the atomic distances. The loss function which is minimized by RMC is then

$$\chi_t^2 = \sum_{l=1}^{L} \frac{\left(D^t(r_l \mid \boldsymbol{s}) - D^E(r_l) \right)^2}{\sigma_D^2(r_l)} \tag{5}$$

where $\sigma_D^2(\cdot)$ is the variance of the data, defined as a function of r_l.

The updating algorithm is then as follows. Starting with an initial set of atomic locations ($t = 0$), a single, randomly selected atom is moved a random distance. An updated χ_{t+1}^2 is calculated and the move is accepted with probability: $\rho = \min\left(1, \exp\left(-(\chi_{t+1}^2 - \chi_t^2)/2\right)\right)$. This update is repeated until convergence and the final atomic configuration is the set of estimated locations.

The sequence of binning intervals is somewhat arbitrary, but guidelines exist to avoid introducing error into the result. For example, the shortest atomic bond known is 0.6Å [15], so including intervals below that threshold may not be necessary. The resolution of the most sophisticated instrumentation is 0.002Å [16], so smaller increments will not capture any additional information.

3. Prior and Related Work

There are several key features of the RMC algorithm that make it a robust method of atomic modeling, as pointed out by [17]. In contrast to previous work [18, 19, 20], edge and finite-size effects are avoided by using periodic boundary conditions and large supercells, respectively. The result is, at least theoretically, independent of the initial configuration since moves that increase χ^2 are also accepted with some probability. Although our description of the RMC algorithm is in terms of the total pair distribution function which exists in real space, the original experimental data observed in reciprocal space (angles and intensities), can be inserted into the algorithm and used equivalently, or in combination with real-space data.

To some, a disadvantage of the RMC algorithm is that it produces a non-unique solution of atomic disorder. There is no proof that the final χ^2 value will be a global minimum nor that the true atom distribution has been obtained. [17] argue that since the true structure of any material can never be known, an algorithm that produces a non-unique solution is actually an advantage. While the possibility of multiple solutions can be seen as an advantage, formal modeling of that variability remains an open topic.

Current research surrounding atomic structural modeling focuses on aspects such as simultaneous estimation of complex physical properties [21], grouping patterns in the resulting model [22], or improved computational efficiency

[23]. In order to minimize the effect uncertainty has on the final estimates, many practitioners, including those mentioned above, average results obtained through multiple independent runs of the RMC algorithm. This strategy can also be used to gain insight on the uncertainty of the RMC result, albeit in a fairly informal manner.

Another approach to uncertainty quantification is achieved by condensing the supercell into a unit cell, creating a distribution of atom locations within the unit cell (see [24, 25, 26, 27, 22] for examples). This clearly requires less computing power than the previous approach and is helpful for visualization, however variance is only defined globally instead of locally. Still, many practitioners make no attempt to address model uniqueness or quantify uncertainties.

Thus, we propose the use of hierarchical Bayesian modeling to achieve local uncertainty quantification through a single run of the estimation algorithm. A Bayesian framework provides the additional benefit of immediate quantification of the uncertainty of any function of atom locations as well. [28] and [29] outlined the use of Bayesian inference in place of Rietveld refinement to quantify the uncertainty in global structure with diffraction data in reciprocal space. However, to our knowledge, no use of Bayesian methodology has been incorporated when describing the local structure with total scattering data in real space.

III. BAYESIAN METHODOLOGY

The Bayesian approach to statistical inference combines prior information about the parameters of interest with the likelihood for the data. This is achieved by treating the parameters as random variables and making use of Bayes' theorem to compute the posterior distribution. Formally, let y be the observed random variable ($D^E(r)$ in equation (4)) and $\boldsymbol{\theta}$ be the set of parameters of interest (all atom locations $\boldsymbol{s}$ and all other unknowns such as variance parameters). The likelihood function and joint prior distribution are combined to create the posterior distribution, $p(\boldsymbol{\theta} \mid y) \propto p(y \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$, with which statistical inference is performed.

1. Overview of MCMC

We approximate the posterior distribution using Markov Chain Monte Carlo (MCMC) methods. Specifically, we use a combination of Gibbs [30] and Metropolis-Hastings [31, 32] sampling. As with RMC, the algorithm begins by setting initial values for all parameters and proceeds by cycling through the parameters and performing stochastic updates. After burn-in, the algorithm produces many samples from the joint posterior distribution which are used to approximate the posterior distribution and quantify parametric uncertainty.

Convergence of the algorithm is (theoretically) invariant to the initial values, but we use RMC results as initial values to encourage convergence. Parameters are updated from their full conditional posterior distributions, i.e., the distribution of the parameter conditioned on both the data and all other parameters in the model. If possible, the sample is simply drawn from the full conditional distribution in a Gibbs update; if the posterior is too complicated then it is sampled using Metropolis-Hastings rejection sampling. For the Metropolis-Hastings step at iteration t, a candidate value $\boldsymbol{\theta}^*$ is drawn from a proposal density $p(\boldsymbol{\theta}^t \mid \boldsymbol{\theta}^*)$. This candidate is then accepted by setting $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^*$ with probability

$$\rho = \min\left(1, \frac{p(\boldsymbol{\theta}^* \mid y)p(\boldsymbol{\theta}^t \mid \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^t \mid y)p(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^t)}\right), \tag{6}$$

otherwise, $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$. In our examples below, we run the algorithm for 2,000 iterations and discard the first 500 as burn-in, leaving 1,500 samples to approximate the posterior distribution.

2. Perovskite Model

To illustrate our method, we assume a perovskite structure, specifically tetragonal $BaTiO_3$ (Figure 1), however the Bayesian framework could be applied to more complex structures. Perovskite $BaTiO_3$ has become one of the most extensively studied functional materials since the discovery of its polar behavior in the 1940s [33, 34] due to its high dielectric constant and room temperature ferroelectric behavior. $BaTiO_3$ has a well-known global structure, described with a tetragonal *P4mm* space group, with Ba atoms at (0,0,0), Ti atoms at (1/2,1/2,z), and O atoms at (1/2,1/2,z) and (1/2,0,z). Its polarization at room temperature is a consequence of an average relative displacement of Ti and O atoms along the c-axis from centrosymmetric positions in the unit cell [33, 35]. There have been a series of investigations into the small local deviations hallmark to its ordering behavior, particularly in regards to the correlation of atomic displacements in the polarization direction [35, 36, 22, 37].

Let $i = \{Ba, O, Ti\}$ be the type of atom, $j \in \{1, \ldots, n_i\}$ index an atom of type i within the supercell, and $k \in \{1, 2, 3\}$ represent the dimension. Define $\boldsymbol{s}_{ij} = (s_{ij1}, s_{ij2}, s_{ij3})$ as latent coordinates of atom j of type i and $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \mu_{ij2}, \mu_{ij3})$ as the atom's coordinates if the atoms adhered to a perfect lattice. We assume that each atom location follows a normal distribution with known mean and unknown variance. The variance of the data measured in real space, σ_D^2, is assumed constant and modeled with an inverse gamma distribution. The hierarchical model is then

$$\begin{aligned}
&\text{Observational Model:} && D^E(r_l) \mid \boldsymbol{s} \sim N\left(D_{\boldsymbol{\theta}}(r_l \mid \boldsymbol{s}), \sigma_D^2\right)\\
&\text{Structural Model:} && s_{ijk} \sim N\left(\mu_{ijk}, \tau_i^2\right)\\
&\text{Hyperparameter Model:} && \tau_i^2 \sim Uniform\,(0, \delta)\\
& && \sigma_D^2 \sim IG\,(\alpha, \beta)
\end{aligned}$$

With priors specified as above, the full conditional distributions for parameters σ_D^2 and τ_i^2 are available in closed-form and are as follows,

$$\begin{aligned}
p(\sigma_D^2 \mid\ & D^E(\boldsymbol{r}), \boldsymbol{s}, \alpha, \beta) \sim\\
& IG\left(\alpha + \frac{L}{2}, \beta + \frac{1}{2}\sum_{l=1}^{L}[D^E(r_l) - D_{\boldsymbol{\theta}}(r_l \mid \boldsymbol{s})]^2\right)\\
p(\tau_i^2 \mid\ & D^E(\boldsymbol{r}), \boldsymbol{s}, \boldsymbol{\tau}^2 \setminus \{\tau_i^2\}, \delta) \sim\\
& tIG_{(0,\delta)}\left(\frac{3n_i}{2} - 1, \sum_{j=1}^{n_i}\sum_{k=1}^{3}\frac{[s_{ijk} - \mu_{ijk}]^2}{2}\right).
\end{aligned}$$

Here $tIG_{(0,\delta)}$ represents a truncated inverse gamma distribution, restricted to the interval $(0, \delta)$. These parameters are updated using Gibbs sampling by drawing new values at each iteration from their full conditional distributions.

The full conditional distribution for s_{ijk} is not available in closed form. Its posterior distribution has the complicated form,

$$p(s_{ijk} \mid D^E(\boldsymbol{r}), \boldsymbol{s} \setminus \{s_{ijk}\}, \boldsymbol{\tau}^2, \sigma^2) \propto \prod_{l=1}^{L} \phi\left(\frac{D^E(r_l) - D_{\boldsymbol{\theta}}(r_l \mid \boldsymbol{s})}{\sigma}\right) \phi\left(\frac{s_{ijk} - \mu_{ijk}}{\tau_i}\right),$$

where $\phi(\cdot)$ represents the probability density function of a standard normal distribution. We update the atom locations using a Metropolis-Hastings step as described above with an adaptive candidate distribution $N(s_{ijk}, \psi_i)$, where ψ_i is periodically adjusted to maintain an acceptance rate between 0.25 and 0.5.

3. Comparison of RMC and Bayesian analysis via MCMC

We see three main advantages to the proposed Bayesian methodology over RMC. The first advantage is a more complete assessment of uncertainty. RMC and MCMC are clearly similar algorithms, with both relying on Metropolis-Hastings steps. However, in a Bayesian analysis implemented via MCMC, the stationary portion of the Markov chain created through simulation is equal to the desired posterior distribution. Unlike RMC, where the sequence of updates are discarded once convergence is met, with MCMC all post-burn-in samples are used to approximate the posterior distribution. Functions of the posterior distribution (e.g., the average deviation from the lattice for each atom type) are proper distributions themselves, which allows for testing of complex hypotheses. And, unlike classical statistical methods, the credible intervals created through Bayesian inference are also easily interpretable as having a set probability of containing the true parameter value.

The second advantage is that we are also able to compare models with different mean or variances structures. For example, if we are interested in comparing models with and without equal variance components across all atom types, this could be achieved by fitting two models (one with assumed equal variance and one without), and comparing their Deviance Information Criterion (DIC) [38] values. Since separate models can be fit independently, there is essentially no added computational cost as it is easily parallelized.

Finally, we allow incorporation of prior knowledge in a statistically appropriate manner. Substantial prior information about atomic structure generally exists, and the Bayesian methodology provides a seamless way to formally incorporate this information. Structural modeling is extremely challenging and often plagued by non-unique solutions, and so incorporation of prior knowledge can stabilize estimation. In fact, if we ignore all prior information (i.e., structural and hyperparameter layers of the hierarchical model), fix the error variance σ_D^2, and use only one sample from the posterior distribution, we have the RMC algorithm as a special case of the Bayesian analysis.

IV. RESULTS AND ANALYSIS

1. Simulation

In order to test our methodology, we simulated an ideal $BaTiO_3$ structure and jittered the locations of atoms away from the lattice with variances 0.07^2, 0.09^2, and 0.05^2 for Ba, Ti, and O, respectively. The simulated data was constructed by following the equations (1)–(4) and then adding Gaussian noise, where $\sigma_D^2 = Var(G^{PDF}(r)) \left(\sum_i c_i b_i\right)^4 = 0.0027 \left(\sum_i c_i b_i\right)^4$.

The crystal chemistry of the perovskite oxide at ambient conditions precludes atom location swapping. Global estimates (through Reitveld refinement of atomic displacement parameters) indicates limited atom movements associated with thermal vibrations and disorder. Thus, atom movements were constrained to prevent swapping relative to their original lattice location. Specifically, atom moves beyond 0.4Å from the original lattice were automatically rejected. The corresponding hyperparameter ($\delta = 0.04$) was also chosen to model this behavior. Prior knowledge on the variance of data was purposely not incorporated to allow the data to drive the posterior distribution. Therefore, a diffuse prior was selected for $\sigma_D^2(r)$ by setting $\alpha = 0.5$ and determining β based on the solution to $\text{argmin}_\beta[Var(D^E(\boldsymbol{r})) - Q_\beta(0.9)]^2$, where $Q_\beta(p)$ represents the p-th quantile from an inverse gamma distribution with $\alpha = 0.5$. For our simulated data, this resulted in a hyperparameter, $\beta = 200$.

We assumed that the data originated from neutron diffraction, but the methodology can be applied to any diffraction data where a pair distribution function can be defined. For $BaTiO_3$, the neutron scattering lengths are $\boldsymbol{b} = \{5.07, -3.438, 5.803\} \times$ 1e−5Å and the concentrations are $\boldsymbol{c} = \{0.2, 0.2, 0.6\}$ for Ba, Ti, and O, respectively. The supercell was limited to a cube of dimension ≈ 10Å, which implies atom counts of Ba, Ti, and O equal to 27, 27, and 81, respectively. With a barrier added around the cube to facilitate using periodic boundary conditions, the supercell volume was $\approx 1741 \text{Å}^3$ and so $\rho_0 \approx 0.078$ atoms per Å^3.

We used 2000 MCMC iterations with a burn-in of 500, satisfying the diagnostic defined by [39]. Where the RMC algorithm is aided by an initial configuration determined by Rietveld refinement, we found the convergence of the MCMC chains can be greatly accelerated by a few iterations of the RMC algorithm. We fit both a full model, including different variance estimates for each atom type, and a reduced model, assuming that the variances for all atom types were equivalent.

The acceptance rate and convergence in χ^2 value does not have a theoretical basis. However, guidelines from the `RMCProfile` software [40] suggest that the acceptance rate should plateau at roughly between 0.25 and 0.6 and that the χ^2 value should no longer exhibit consistent reduction, but rather should resemble white noise. Practitioners often iterate well past these guidelines and we do the same, performing $200 \times N$ iterations. Although the RMC algorithm is a rejection sampling algorithm and, thus, is theoretically protected against initial configurations, we run the RMC algorithm ten times. This will allow for direct comparison of estimation performance and uncertainty quantification given equivalent computational cost. Assessment of convergence and accep-

tance rates of the sampling algorithms are provided in the Appendix.

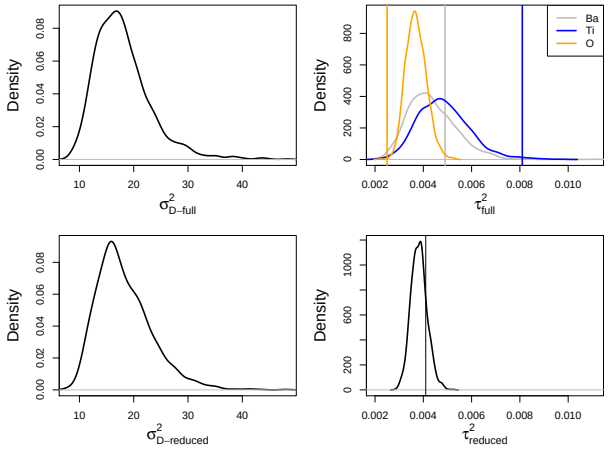

Fig. 2. Posterior distributions for variance parameters estimated from simulated data compared to their true value for both the full (top) and reduced (bottom) models.

The resulting posterior distributions for the variance estimates for both models are provided in Figure 2, along with the true values used to generate the data represented by vertical lines. As expected, the variance for each atom type in the full model has high posterior density near the true values. Using a single variance estimate for all atom types, $\tau^2_{reduced}$, the highest posterior probability is quite close to the average atom variance weighted by atom type frequency, which is 0.0041. However, this model, by design, cannot capture the different variance for each atom type. The posterior median for both distributions of σ^2_D are approximately $0.08\,(\sum_i c_i b_i)^4 \approx 17$, while the true value for σ^2_D is $0.0027\,(\sum_i c_i b_i)^4 \approx 0.568$. Our hypothesis is that σ^2_D cannot be effectively modeled in real space; we mention extensions to reciprocal space in Section V.

As a more formal assessment of model fit, we compared models with the DIC metric. Given posterior distributions for both the full (unequal variance) and reduced (equal variance) models, the DIC metric quantifies whether or not separate variance parameters for each atom type are truly necessary. Resulting DIC values were 285.14 and 286.39 for the full and reduced models, respectively. A lower DIC implies a better fit, so this confirms the truth that the variances of atom type are not equal and should be modeled separately.

The resulting atom locations for both the best RMC fit (designated by lowest χ^2 value) and the Bayesian analysis assuming a full model are provided in Figure 3, along with the true locations as determined by the simulation. Of note, the center of the posterior distributions are typically closer to the true values. To summarize, we compared the best fitting RMC estimate, the mean RMC estimate, and the mean posterior estimate of every atom location to their true values by Euclidean distance. This information is summarized in Figure 4.

It is clear that the Bayesian approach consistently reduces the error in location estimation as compared to a single RMC run, even the best fitting RMC output. However, if the Bayesian approach were to increase the error as compared to

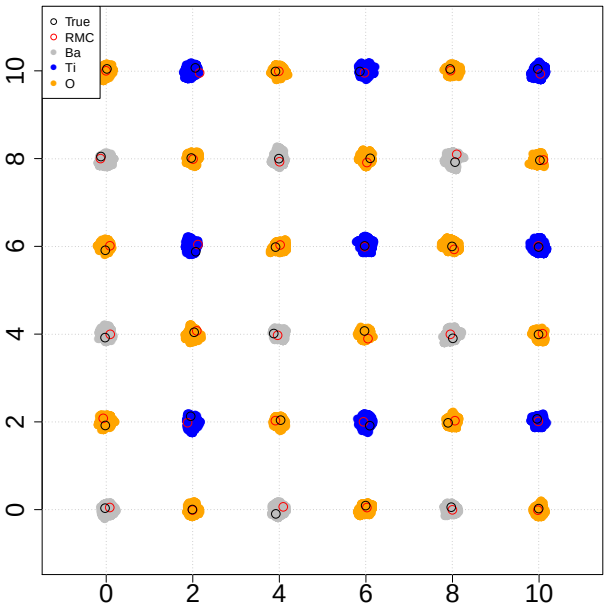

Fig. 3. Posterior distributions of atom locations for the full model and point estimates provided by a single run of RMC compared to the true simulated locations.

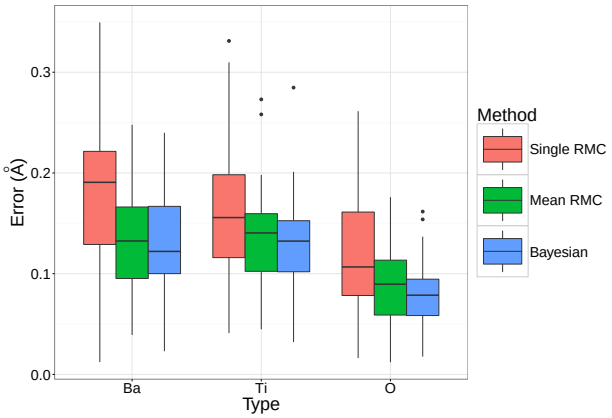

Fig. 4. Error in atom location for all atoms as measured by Euclidean distance, grouped by method and atom type.

an average of multiple RMC runs, then one would need to balance improved accuracy against the versatility of the Bayesian methodology. For this example, the Bayesian approach was more accurate than the mean of multiple RMC runs for 60% of all locations, was lower in median error for all atom types.

As mentioned, functions of posterior distributions are available at an additional computational cost that is negligible. The posterior mean and 95% credible region for the total pair distribution function is provided as an example and is shown in Figure 5, along with the data and the best RMC fit. Both the RMC and Bayesian approaches provide good fits to the data, however the posterior mean obtained from the Bayesian approach shows no discernible difference from the truth. Also, the Bayesian result includes credible regions, so it also quantifies the uncertainty in the simulated data.

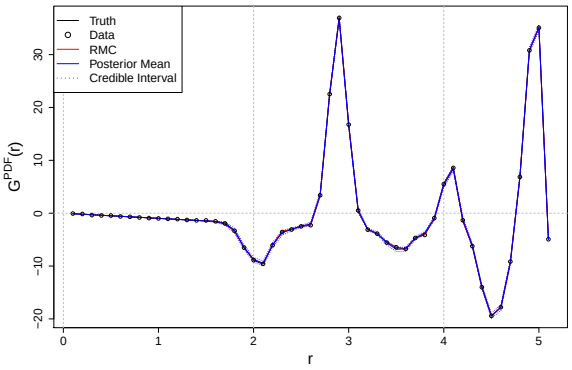

Fig. 5. Total pair distribution function for the simulated data, a single RMC run, and the posterior mean with a Bayesian 95% credible region.

2. Application

A structural analysis on real $BaTiO_3$ neutron data was also performed. Given the same hierarchical model as in the simulation, posterior distributions were obtained and results are provided in Figures 6–8.

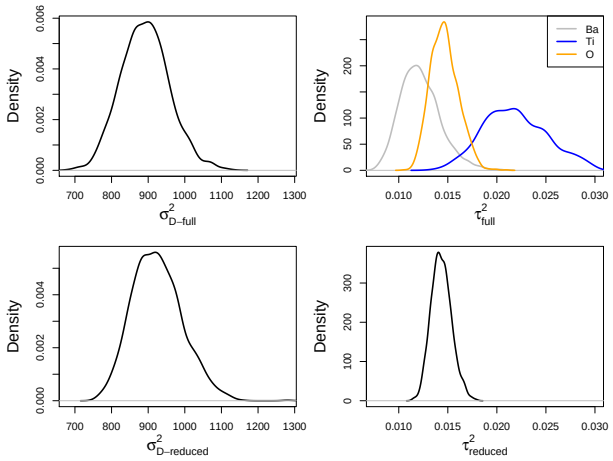

Fig. 6. Posterior distributions for variance parameters estimated from the observed data for both the full and reduced models.

In contrast to the simulation, the variance in the data is substantially larger. This pattern continues to the total pair distribution function, as the confidence bands in Figure 8 are much wider. The disparity between variances of atom type again indicates inequality and the DIC values, 3711.25 and 3723.20 for the full and reduced models, respectively, provides credence to this hypothesis. However, the magnitude and relative ranking of variability across different atom types is slightly different in the observed data. Specifically, the variability in atom location is larger in the observed data and the variability in titanium atoms is substantially larger than both that of barium and oxygen atoms. The pattern is clearly visible in Figure 7, where the barium and oxygen atoms are fairly stationary, while the titanium atoms have an expanded space of high posterior probability.

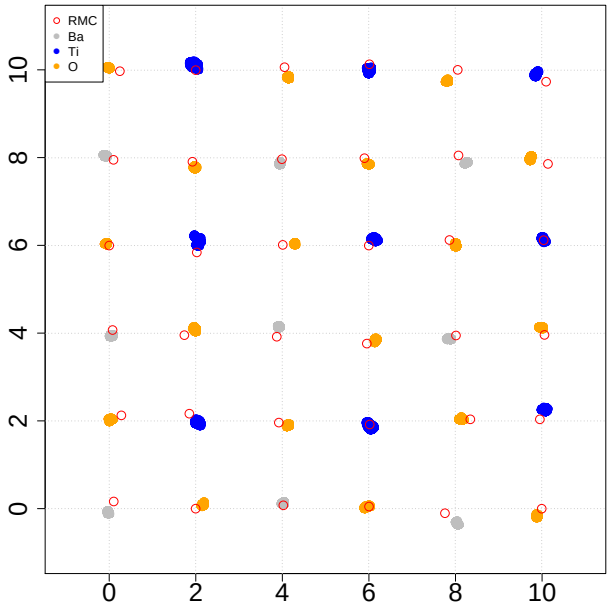

Fig. 7. Posterior distributions of atom locations for the full model compared to point estimates provided by a single run of RMC on the experimental data.

V. CONCLUSIONS AND FUTURE WORK

In conclusion, the use of Bayesian statistics provides a natural framework to quantify the uncertainty in every atom location as well as any function of atom location. Model comparison regarding mean or covariance parameters is also available, allowing users to determine how best to describe atom movement within a material. Quantifying the uncertainty in the data can also be an essential tool for practitioners attempting to determine when their data is of sufficient quality. An additional benefit of our method is that the error associated with estimated atomic location is reduced.

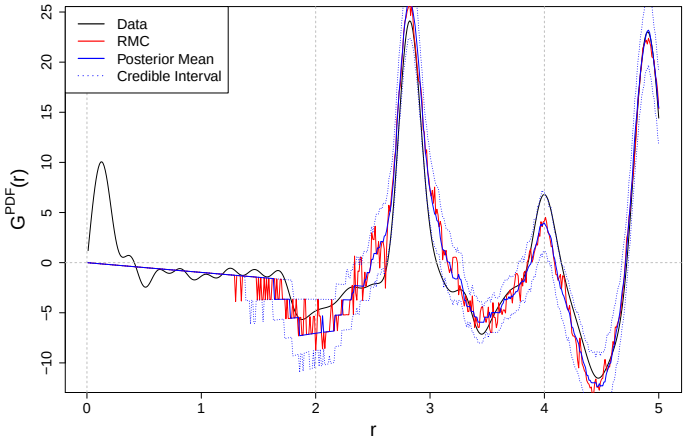

Fig. 8. Total pair distribution function for the observed data, a single RMC run, and the posterior mean with a Bayesian 95% credible region.

Although our proposed methodology quantifies uncer-

tainty in real space, the measured data originates in reciprocal space and is then transformed into real space. Thus, possible improvements include modeling the measurement error variance in reciprocal space and then transforming to real space or assuming the observational model itself occurs in reciprocal space. The data in reciprocal space, being count data, is often assumed Poisson in nature. We look to exploit that behavior in our hierarchical model, while assessing the assumption that the variance is proportional to the square root of the measured intensities [41].

Modeling the variance in both real and reciprocal space will also allow for comparison of data collected by different metrologies and under different settings, which would be of great value to the materials science community.

APPENDIX

The posterior draws and chains of χ^2 values for the MCMC and RMC algorithms, respectively, for the simulated data are provided below in Figures 9–10. We also include the acceptance rates for both procedures in Figures 11–12. Likewise, chains and acceptance rates for both algorithms applied to the observed data are provided in Figures 13–16.

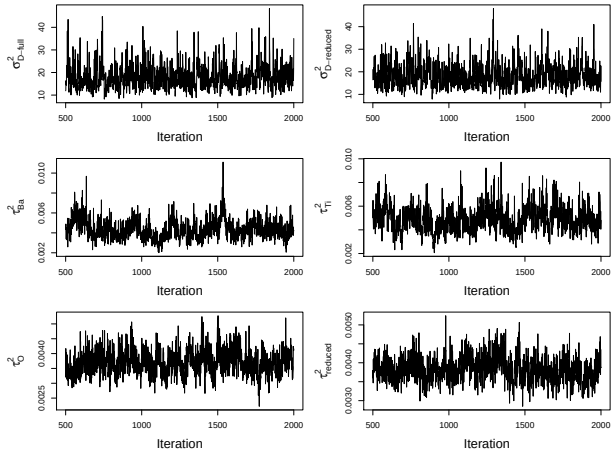

Fig. 9. MCMC chains for all variances and both models for the simulated data.

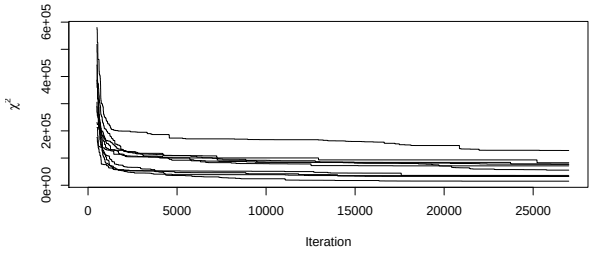

Fig. 10. χ^2 values for the RMC algorithm applied to the simulated data for all ten chains.

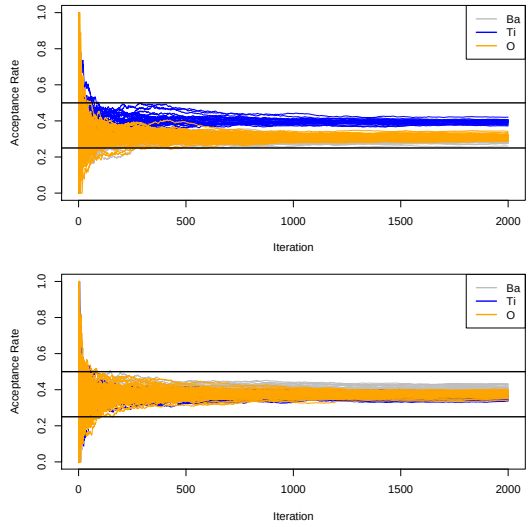

Fig. 11. Acceptance rates for all atom locations of the MCMC algorithm, applied to the simulated data, assuming the full model (top) and the reduced model (bottom).

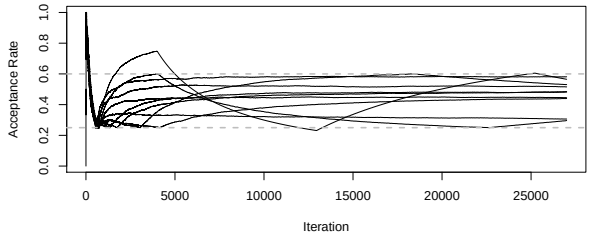

Fig. 12. Acceptance rates for the RMC algorithm applied to the simulated data for all ten chains.

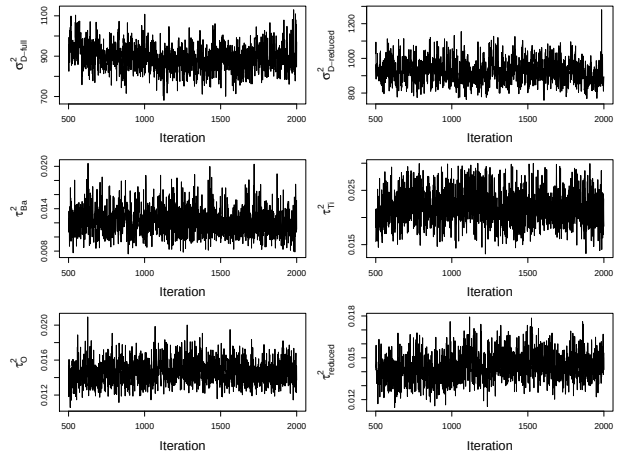

Fig. 13. MCMC chains for all variances and both models for the observed data.

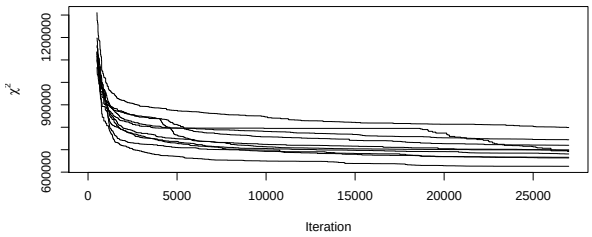

Fig. 14. χ^2 values for the RMC algorithm applied to the observed data for all ten chains.

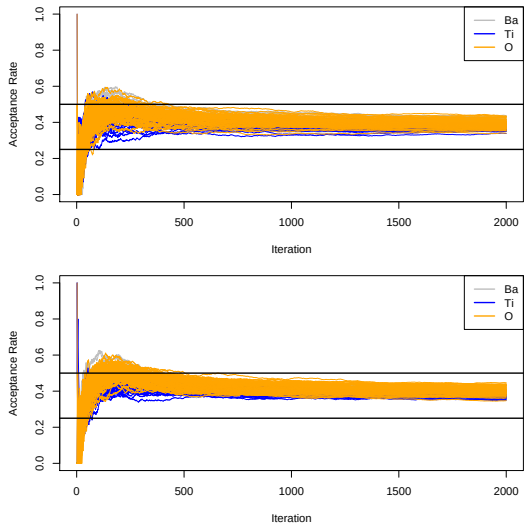

Fig. 15. Acceptance rates for all atom locations of the MCMC algorithm, applied to the observed data, assuming the full model (top) and the reduced model (bottom).

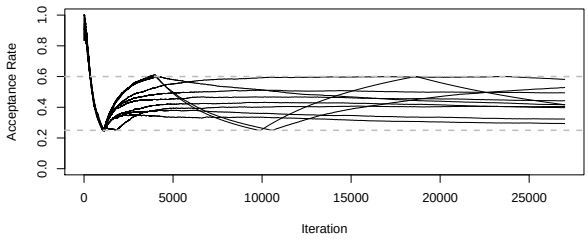

Fig. 16. Acceptance rates for the RMC algorithm applied to the observed data for all ten chains.

ACKNOWLEDGMENTS

The research at NC State was funded by the Consortium for Nonproliferation Enabling Capabilities (CNEC), which is sponsored by the Department of Energy, NNSA Office of Defense Nuclear Nonproliferation, and also National Science Foundation Grant DGE-1633587. Experimental data was collected at the POWGEN instrument at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. We also thank Christopher Fancher, Daniel Olds, Elizabeth Dickey, and Matt Tucker for their valuable input and suggestions.

REFERENCES

1. C. A. YOUNG and A. L. GOODWIN, "Applications of pair distribution function methods to contemporary problems in materials chemistry," *Journal of Materials Chemistry*, **21**, *18*, 6464–6476 (2011).
2. S. J. BILLINGE and I. LEVIN, "The problem with determining atomic structure at the nanoscale," *Science (New York, N.Y.)*, **316**, *5824*, 561–565 (Apr 27 2007), lR: 20070427; JID: 0404511; ppublish.
3. M. NIKL, "Wide band gap scintillation materials: progress in the technology and material understanding," *physica status solidi (a)*, **178**, *2*, 595–620 (2000).
4. Y. ARIKAWA, K. YAMANOI, T. NAKAZATO, E. S. ESTACIO, T. SHIMIZU, N. SARUKURA, M. NAKAI, T. NORIMATSU, H. AZECHI, and T. MURATA, "Pr 3 -doped fluoro-oxide lithium glass as scintillator for nuclear fusion diagnostics," *Review of Scientific Instruments*, **80**, *11*, 113504 (2009).
5. K. MAYER, M. WALLENIUS, and I. RAY, "Nuclear forensics–a methodology providing clues on the origin of illicitly trafficked nuclear materials," *Analyst*, **130**, *4*, 433–441 (2005).
6. M. J. KRISTO and S. J. TUMEY, "The state of nuclear forensics," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, **294**, 656–661 (2013).
7. R. MCGREEVY and L. PUSZTAI, "Reverse Monte Carlo simulation: a new technique for the determination of disordered structures," *Molecular Simulation*, **1**, *6*, 359–367 (1988).
8. M. DE GRAEF and M. E. MCHENRY, *Structure of materials: an introduction to crystallography, diffraction and symmetry*, Cambridge University Press (2007).
9. W. FRIEDRICH, P. KNIPPING, and M. LAUE, "Interferenzerscheinungen bei Röntgenstrahlen," *Annalen der Physik*, **346**, *10*, 971–988 (1913).
10. G. THOMSON and A. REID, "Diffraction of cathode rays by a thin film," *Nature*, **119**, 890 (1927).
11. C. SHULL, "Early development of neutron scattering," *Reviews of Modern Physics*, **67**, *4*, 753 (1995).
12. H. RIETVELD, "A profile refinement method for nuclear and magnetic structures," *Journal of applied Crystallography*, **2**, *2*, 65–71 (1969).
13. C. R. A. CATLOW, *Computer modeling in inorganic crystallography*, Academic Press (1997).

14. D. A. KEEN, "A comparison of various commonly used correlation functions for describing total scattering," *Journal of Applied Crystallography*, **34**, *2*, 172–177 (2001).
15. R. T. SHANNON and C. T. PREWITT, "Effective ionic radii in oxides and fluorides," *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, **25**, *5*, 925–946 (1969).
16. J. WANG, B. H. TOBY, P. L. LEE, L. RIBAUD, S. M. ANTAO, C. KURTZ, M. RAMANATHAN, R. B. VON DREELE, and M. A. BENO, "A dedicated powder diffraction beamline at the Advanced Photon Source: Commissioning and early operational results," *Review of Scientific Instruments*, **79**, *8*, 085105 (2008).
17. R. L. MCGREEVY, "Reverse monte carlo modelling," *Journal of Physics: Condensed Matter*, **13**, *46*, R877 (2001).
18. R. KAPLOW, T. ROWE, and B. AVERBACH, "Atomic arrangement in vitreous selenium," *Physical Review*, **168**, *3*, 1068 (1968).
19. A. RENNINGER, M. RECHTIN, and B. AVERBACH, "Monte Carlo models of atomic arrangements in arsenic-selenium glasses," *Journal of Non-Crystalline Solids*, **16**, *1*, 1–14 (1974).
20. W. SCHWEIKA and H.-G. HAUBOLD, "Neutron-scattering and Monte Carlo study of short-range order and atomic interaction in Ni 0.89 Cr 0.11," *Physical Review B*, **37**, *16*, 9240 (1988).
21. J. A. PADDISON, S. AGRESTINI, M. R. LEES, C. L. FLECK, P. P. DEEN, A. L. GOODWIN, J. R. STEWART, and O. A. PETRENKO, "Spin correlations in Ca 3 Co 2 O 6: Polarized-neutron diffraction and Monte Carlo study," *Physical Review B*, **90**, *1*, 014411 (2014).
22. J. NEILSON and T. MCQUEEN, "Representational analysis of extended disorder in atomistic ensembles derived from total scattering data," *Journal of applied crystallography*, **48**, *5*, 1560–1572 (2015).
23. B. AOUN, "Fullrmc, a rigid body reverse monte carlo modeling package enabled with machine learning and artificial intelligence," *Journal of Computational Chemistry* (2016).
24. D. NANU, M. TUCKER, W. HAIJE, J. VENTE, and A. BÖTTGER, "Atom configurations in Pd–Au and Pd–Au–D alloys: A neutron total scattering and Reverse Monte Carlo study," *Acta Materialia*, **58**, *16*, 5502–5510 (2010).
25. D. P. SHOEMAKER, R. SESHADRI, A. L. HECTOR, A. LLOBET, T. PROFFEN, and C. J. FENNIE, "Atomic displacements in the charge ice pyrochlore Bi 2 Ti 2 O 6 O' studied by neutron total scattering," *Physical Review B*, **81**, *14*, 144113 (2010).
26. C. YOUNG, E. DIXON, M. TUCKER, D. KEEN, M. HAYWARD, and A. GOODWIN, "Reverse Monte Carlo study of Cu–O bond distortions in $YBa_2Cu3O6.9$," *Zeitschrift für Kristallographie Crystalline Materials*, **227**, *5*, 280–287 (2012).
27. E. AKSEL, J. S. FORRESTER, J. C. NINO, K. PAGE, D. P. SHOEMAKER, and J. L. JONES, "Local atomic structure deviation from average structure of Na 0.5 Bi 0.5 TiO 3: Combined x-ray and neutron total scattering study," *Physical Review B*, **87**, *10*, 104113 (2013).
28. C. M. FANCHER, Z. HAN, I. LEVIN, K. PAGE, B. J. REICH, R. C. SMITH, A. G. WILSON, and J. L. JONES, "Use of Bayesian Inference in Crystallographic Structure Refinement via Full Diffraction Profile Analysis," *Scientific reports*, **6**, 31625 (Aug 23 2016), lR: 20160831; JID: 101563288; OID: NLM: PMC4994022; 2016/04/28 [received]; 2016/07/22 [accepted]; epublish.
29. J. E. LESNIEWSKI, S. M. DISSELER, D. J. QUINTANA, P. KIENZLE, and W. RATCLIFF, "Bayesian method for the analysis of diffraction patterns using BLAND," *Journal of Applied Crystallography*, **49**, *6*, 2201–2209 (2016).
30. S. GEMAN and D. GEMAN, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , *6*, 721–741 (1984).
31. N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, and E. TELLER, "Equation of state calculations by fast computing machines," *The Journal of chemical physics*, **21**, *6*, 1087–1092 (1953).
32. W. K. HASTINGS, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, **57**, *1*, 97–109 (1970).
33. H. D. MEGAW, "Ferroelectricity in crystals," (1957).
34. M. E. LINES and A. M. GLASS, *Principles and applications of ferroelectrics and related materials*, Oxford university press (1977).
35. G. KWEI, A. LAWSON, S. BILLINGE, and S. CHEONG, "Structures of the ferroelectric phases of barium titanate," *The Journal of physical chemistry*, **97**, *10*, 2368–2377 (1993).
36. K. PAGE, T. PROFFEN, M. NIEDERBERGER, and R. SESHADRI, "Probing local dipoles and ligand structure in BaTiO3 nanoparticles," *Chemistry of Materials*, **22**, *15*, 4386–4391 (2010).
37. M. SENN, D. KEEN, T. LUCAS, J. HRILJAC, and A. GOODWIN, "Emergence of Long-Range Order in BaTiO 3 from Local Symmetry-Breaking Distortions," *Physical Review Letters*, **116**, *20*, 207602 (2016).
38. D. J. SPIEGELHALTER, N. G. BEST, B. P. CARLIN, and A. VAN DER LINDE, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, *4*, 583–639 (2002).
39. A. E. RAFTERY and S. LEWIS, "How many iterations in the Gibbs sampler," *Bayesian statistics*, **4**, *2*, 763–773 (1992).
40. M. G. TUCKER, D. A. KEEN, M. T. DOVE, A. L. GOODWIN, and Q. HUI, "RMCProfile: reverse Monte Carlo for polycrystalline materials," *Journal of Physics: Condensed Matter*, **19**, *33*, 335218 (2007).
41. P. F. PETERSON, S. I. CAMPBELL, M. A. REUTER, R. J. TAYLOR, and J. ZIKOVSKY, "Event-based processing of neutron scattering data," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **803**, 24–28 (2015).