Advice on describing Bayesian analysis of neutron and X-ray reflectometry

Andrew R. McCluskey,[1,*] Andrew J. Caruana,[2] and Christy J. Kinane[2]
(Open Reflectometry Standards Organisation)
[1] *European Spallation Source ERIC, P.O. Box 176, SE-221 00, Lund, Sweden*
[2] *ISIS-Neutron and Muon Source, Rutherford Appleton Laboratory, Didcot, Oxon OX11 0QX, U.K.*

Driven by the availability of modern software, Bayesian analysis is becoming more popular in neutron and X-ray reflectometry analysis. The understandability and replicability of these analyses may be harmed by inconsistencies in how the probability distributions central to Bayesian methods are represented in the literature. Herein, we porvide advice on how to report the results of Bayesian analysis as applied to neutron and X-ray reflectometry. This includes the clear reporting of initial starting conditions, the prior probabilities, and results of any analysis, the posterior probabilities that are the Bayesian equivalent of the error bar, to enable replicability and improve understanding. We believe that this advice, grounded in our experience working in the field, will enable greater analytical reproducibility among the reflectometry community, as well as improve the quality and usability of results.

I. INTRODUCTION

Neutron and X-ray reflectometry are powerful tools to probe the interfacial structure of materials [1]. However, as a result of the "phase-problem", the analysis of these techniques is ill-posed in nature, as there are multiple possible solutions [2]. This has led to the use of Bayesian analysis, where our prior understanding of the system, can help in our understanding of the results [3–5]. Recently, developments in the availability of computer software for analysis that include Bayesian functionality, such as the Refl1d, refnx, anaklasis, and RasCAL [6–9] which implement methods from bumps, emcee, and dynesty [10–12], have led to an increase in the utilisation of Bayesian methods by the reflectometry community [13, 14].

Reflectometry analysis can be described in the most simplistic terms as the comparison and refinement of a model based on some parameters, $\mathbf{x}$, to reproduce some reflectivity data set, $\mathbf{D}$. This refinement process is involves comparing the model to the data via a statistical test, such as the χ^2-test, to produce a set of parameters that reproduce well the experimental data. The input for this refinement process is the model and some initial parameter values, which may be absolute value or parameter ranges, which the output is a set of value for $\mathbf{x}$ with associated error bars, which describe the mean and standard deviation of a Gaussian posterior probability distribution. Most analysis packages use some minimisation algorithm to vary the parameters and therefore obtain a set that are optimised with respect to the data, with the implicit assumption that the model can and does accurately describe the data.

The input required depends on the minimisation algorithm being used, with some algorithms requiring a single starting guess (such as a traditional Newtonian methods) and others taking a range of potential values (more common in stochastic approaches like differential evolution). The nature of these inputs define the results of the analysis, therefore it is of the utmost importance that these are share as part of a publication describing the work. Furthermore, often the minimisation is performed with bounds in place, defining parameter values to lie within a given range. This range can be thought of as a probability distribution, where values of x outside of this range have a probability of 0. The use of such a constraint leads to a Bayesian approach to our anlaysis, where we use some prior probability to inform our analysis.

The final values from the minimisation algorithm give the best guess (minimisation algorithm dependent) estimate for the parameters, often with a statistical uncertainty. How this statistical uncertainty is obtained from a minimisation algorithm is beyond it scope of this work, but it is important to acknowledgment that this uncertainty typically assumes that the posterior probability for the parameter is Gaussian in nature. The posterior probability is our understanding of the distribution of parameter values that can be used to describe the data and our prior knowledge. The posterior probability is found as,

$$p(\mathbf{D}|\mathbf{x}) \propto p(\mathbf{x}|\mathbf{D})p(\mathbf{x}), \tag{1}$$

where $p(\mathbf{x}|\mathbf{D})$ is our likelihood, a description of goodness-of-fit between our parameterised model and the data and $p(\mathbf{x})$ is the prior probability associated with our parameters. Similar to the importance of a given goodness-of-fit metric in traditional model-dependent analysis processes, the definition of the priors are extremely important in the use of Bayesian modelling. Additionally, the description of the posterior distribution that results from the Bayesian sampling processes is of paramount importance in any scientific conclusions drawn.

Although the use of Bayesian inference can be valuable in the interpretation of reflectometry data, inconsistency

* andrew.mccluskey@ess.eu

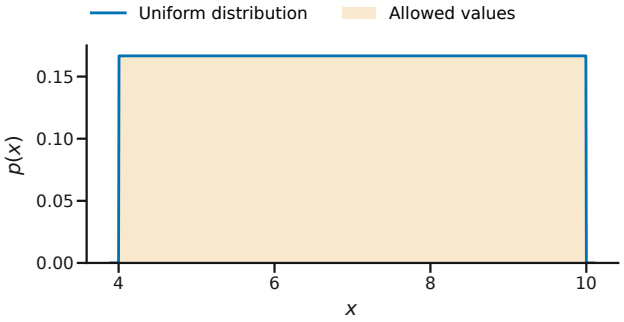

FIG. 1. The normalised prior probability for some parameter, say a scattering length density, that is bounded between 4 and 10, a uniform distribution that can also be written as $x \sim \mathcal{U}[4, 10]$ or $4 \leq x \leq 10$.

in the description of the process may result in analysis that cannot be reproduced or easily understood. This can range from information critical to the analysis replication being omitted, to failure to accurately describe the structure of a parameter's posterior probability. A common example of this (that the authors of this work admit to being guilty of [15]) is failure to recognise that the utilisation of a bounded fitting, where a parameter is limited to be within a given range of values, is a Bayesian analysis and subsequently the bounds are then not stated in the resulting publication, while this has no impact on the quality of the results, some clarity and reproducibility will be lost. The definition of a set of bounds on a parameter is fundamentally stating that the prior probability for the parameter is uniform and non-zero between these bounds and zero outside of them (Figure 1), a Bayesian statement. This is incorporated into most fitting packages by allowing maximum and minimum fitting bounds for fitting parameters.

In this letter, we outline some advice and define some best practices for those analysing neutron and X-ray reflectometry measurements looking to report parameter prior and posterior probability distributions. This letter is by no means an exhaustive guide to Bayesian analysis of reflectometry data but we hope that this will help to engage others in best practice. Furthermore, uptake of the approaches discussed herein will lead to greater clarity about the models and assumptions used in, and reproducibility of, our analyses.

II. PRIOR DEFINITION

The most common prior probability that is used for a parameter is uniform between two values. The fact that often it is overlooked by researchers that a bounded parameter is in essence represents a uniform prior probability has been discussed above. For priors that are uniform, the most important information to provide to the user are (1) that the parameter is truly uniform (i.e. all values are equally likely in the prior) and (2) the values of the upper and lower bounds.

There are a variety of different ways that a uniform prior probability may be written; including upper and lower bounds definition, or the range is given as a mathematical interval or expression. While the upper and lower bound description is potentially the most legible, it is also the least precise. The interval and expression approach are both capable of representing open and closed intervals, i.e. for ρ in Table I the upper bound is included in the probability, whereas for d the upper bound is not included. For this reason, we suggest either the interval or expression description to define a uniform probability, or bounded fitting, for a given parameter.

The use of non-uniform informative priors is less common in the analysis of reflectometry data currently. However, with the growth of Bayesian inference and interest in using complementary methods for analysis, these are likely to become very popular in the coming years. Here we will define two potential types of informative prior probabilities, those that can be described with some mathematical function and those that cannot, for example arising from the application of a sampling-based analysis of a complementary technique.

Where it is possible to describe the prior probability as a mathematical function, this should be done by providing the function in the clearest possible language. For example, if the prior probability is taken from a single complementary measurement that is defined as a value with some uncertainty, this represents a normal distribution with a mean and standard deviation. This is shown in Figure 2a for the volume of phospholipid head group that is has been found, from the analysis of a molecular dynamics simulation [16], to have a value of $(320.9 \pm 20.1)\,\text{Å}^3$. Such a prior probability should be described as being normally distributed with a mean of $320.9\,\text{Å}^3$ and a standard deviation of $20.1\,\text{Å}^3$ or more concisely $p(V_h) \sim \mathcal{N}(\mu = 320.9\,\text{Å}^3, \sigma = 20.1\,\text{Å}^3)$. This same approach and be taken for any common statistical distribution, including log-normal and truncated Gaussian distributions.

In the case that the prior distribution cannot be described easily with some mathematical function, for example, if it is a multimodal result from some other sampling method, then the probability distribution should be given in full. It is not straightforward to describe this distribution with some common statistical object, therefore the full probability distribution will be plotted, as shown in Figure 2b for a phospholipid tail volume. Futhermore, this should also be given as either a chain (discussed for a posterior distribution in more detail below) and as a set of histogrammed bin centres and counts in a data file (see `prior.txt` in the ESI for this work). This means that it is possible for the reader to easily visualise the prior distribution and to reproduce the analysis if desired.

TABLE I. Examples of defining different bounds for a uniform prior probability.

Parameter	Lower	Upper	Interval	Expression
$\rho/1\times10^{-6}$ Å^{-2}	4.0	10.0	$p(\rho)\sim\mathcal{U}[4.0,10.0]$	$4.0\leq\rho\leq10.0$
d/Å	100.0	500.0	$p(d)\sim\mathcal{U}[100.0,500.0)$	$100.0\leq d<500.0$

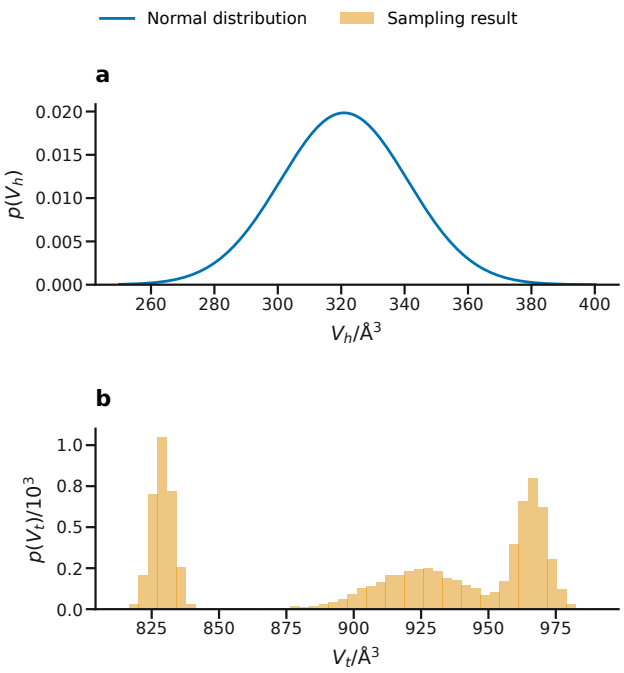

FIG. 2. Potential prior probability distributions; (a) a normal distribtution based on a literature value of (320.9 ± 20.1) Å^3, (b) a potential result, for a phospholipid tail volume, from some sampling approach to a complementary technique where there is not clear distribution that can be defined.

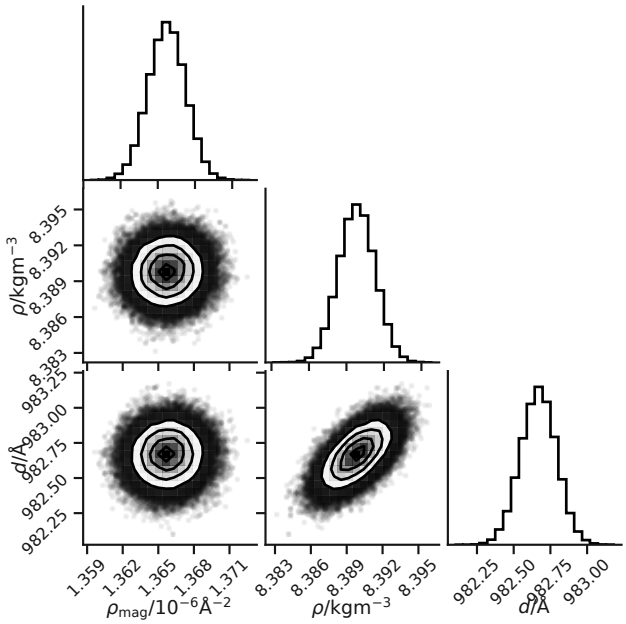

FIG. 3. An example of a corner plot (produced using the corner.py package [17]), representing a three-dimensional probability distribution showing the posterior distribution for the parameters of nickel magnetic scattering length density, nickel mass density, and nickel layer thickness, from the analysis of a nickel layer on a silicon block.

III. POSTERIOR DESCRIPTION

The Bayesian analysis methods typically involve using some sampling process, such as Markov chain Monte Carlo, to estimate the posterior probability distributions for each of the parameters. For most analyses, there will be N parameters under investigation, meaning that the sampling is that of an N-dimensional probability distribution. Therefore, it is important to provide information both about the individual parameter posteriors and the covariance between the parameters. An example of a 3-dimensional posterior probability distribution is given for a nickel layer on silicon in Figure 3, showing the distribution of the magnetic and nuclear scattering length densities for nickel and the thickness of the nickel layer.

While, ideally, the full posterior distribution plot and chain would be provided for a given analysis [18], we accept that this may be unfeasible. Therefore, we suggest that all of the posterior distributions that appear to be Gaussian in nature as subjected to a statistical test to check normality, such as the D'Agostino and Pearson's test [19, 20] (which is available in the SciPy library as `scipy.stats.normal_test` [21]). As will all statistical tests, this requires some threshold value to be defined to reject the null hypothesis, for this value we recommend 0.001 but accept that this is at the discretion of the user. If the parameter distribution passes a statistical test for a given distribution type, this can be quoted in the work, with information about the distribution type and the threshold value used, and the distribution can be described based on fitted parameters of the distribution as is discussed above for the Gaussian distribution. For example, the three parameters in Figure 3 pass this statistical tests, with p-values of less than 1×10^{-15}, therefore we can quote the parameters as normal distributions; $\rho_{\mathrm{mag}}\sim\mathcal{N}(\mu=1.4\times10^{-6}\,\text{Å}^{-2},\sigma=1.5\times10^{-9}\,\text{Å}^{-2})$, $\rho\sim\mathcal{N}(\mu=8.4\times10^{-6}\,\mathrm{kgm}^{-3},\sigma=1.4\times10^{-9}\,\mathrm{kgm}^{-3})$, and $d\sim\mathcal{N}(\mu=9.8\times10^{2}\,\text{Å},\sigma=1.2\times10^{-1}\,\text{Å})$.

In the case of an N-dimensional sampled distribution, where all of the individual dimensions pass a normal test, it is possible to describe the covariance between

the distributions. This can be achieved with a covariance/correlation matrix or correlation coefficients, both of which are accessible from common programming languages. These parameters describe the amount of correlation between different parameters, i.e. if parameter A increases, does this make an increase in parameter B more likely? Understanding these correlations are of fundamental importance to the sampling process, in particular for the ill-posed reflectometry technique. Figure 3 allows us to observe the correlation between the different parameters easily, with showning the join variability of the parameters as a heat map, where the more circular the heat map the less correlation present. We can also find the correlation matrix to be the following,

$$corr[p(\mathbf{D}|\mathbf{x})] = \begin{bmatrix} 1.000 & 0.047 & 0.040 \\ 0.047 & 1.000 & 0.620 \\ 0.040 & 0.620 & 1.000 \end{bmatrix}, \qquad (2)$$

where the off-diagonal elements describe the correlation. It can be seen that the larger correlation between the mass density and thickness of the nickel layer is quantified with the correlation matrix.

If it is not possible to describe the N-dimensional distribution using a series of statistical tests, then the probability distributions must be given in full, ideally as both a plot and a chain of values. The plot should be provided as a corner-type plot (Figure 3), and therefore make clear the correlations present between the different parameters. These corner-type plots, or similar, are available from many common reflectometry analysis packages. In addition to plots, the full chain (the values that are sampled in the sampling of the posterior distribution) should be given. These chains can be very large, therefore it is suggested that they are included in the electronic supplementary information of the work as a compressed file.

IV. CONCLUSIONS

The use of Bayesian analysis in neutron and X-ray reflectometry is increasing, and alongside this, these analyses must be clear to readers and replicable by others. We have outlined some advice, based on experience, on the best practice from reporting information about Bayesian analysis that is performed for reflectometry. Specifically, we have outlined how the prior probabilities used to inform our analyses should be stated, either as uniform or more informed probability distributions that may be described mathematically or as a series of histogrammed values or the complete chain of values. Additionally, we described how best to present the results from our Bayesian analysis in a clear and precise fashion, including the importance of statistical tests for describing our results and the inclusion of correlation between the different parameters. We hope that this advice will be taken on by the reflectometry community and in future, there will be greater consistency and clarity in the reporting of results from Bayesian methods.

CREDIT AUTHOR STATEMENT

A.R.M: Conceptualization, Methodology, Resources, Writing - Original Draft, Visualisation, Project administration.

ACKNOWLEDGMENTS

[1] Lovell, M. R. & Richardson, R. M. Analysis methods in neutron and x-ray reflectometry. *Curr. Opin. Colloid Interface Sci.* **4**, 197–204 (1999).

[2] Majkrzak, C. F. & Berk, N. F. Exact determination of the phase in neutron reflectometry. *Phys. Rev. B* **52**, 10827–10830 (1995).

[3] Sivia, D., Hamilton, W. & Smith, G. Analysis of neutron reflectivity data: maximum entropy, bayesian spectral analysis and speckle holography. *Physica B: Condens. Matter* **173**, 121–138 (1991).

[4] Geoghegan, M., Jones, R. A. L., Sivia, D. S., Penfold, J. & Clough, A. S. Experimental study of surface segregation and wetting in films of a partially miscible polymer blend. *Phys. Rev. E* **53**, 825–837 (1996).

[5] Sivia, D. & Webster, J. The bayesian approach to reflectivity data. *Physica B: Condens. Matter* **248**, 327–337 (1998).

[6] Kienzle, P. A., Krycka, J., Patel, N. & Sahin, I. Refl1d 0.8.15. `https://refl1d.readthedocs.io/` (2021).

[7] Nelson, A. R. J. & Prescott, S. W. *refnx*: neutron and x-ray reflectometry analysis in python. *J. Appl. Crystallogr.* **52**, 193–200 (2019).

[8] Koutsioubas, A. anaklasis: a compact software package for model-based analysis of specular neutron and x-ray reflectometry data sets. *J. Appl. Crystallogr.* **54**, 1857–1866 (2021).

[9] Hughes, A. V. Rascal v1.1.0. `https://github.com/arwelHughes/RasCAL_2019` (2021).

[10] Kienzle, P. A., Krycka, J., Patel, N. & Sahin, I. Bumps 0.8.1. `https://bumps.readthedocs.io/` (2021).

[11] Foreman-Mackey, D. *et al.* emcee v3: A python ensemble sampling toolkit for affine-invariant mcmc. *J. Open Source Softw.* **4**, 1864 (2019).

[12] Speagle, J. S. dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences. *Mon. Notices Royal Astron. Soc.* **493**, 3132–3158 (2020).

[13] McCluskey, A. R. *et al.* Bayesian determination of the effect of a deep eutectic solvent on the structure of lipid monolayers. *Phys. Chem. Chem. Phys.* **21**, 6133–6141 (2019).

[14] McCluskey, A. R., Cooper, J. F. K., Arnold, T. & Snow, T. A general approach to maximise information density

in neutron reflectometry analysis. *Mach. Learn.: Sci. Technol.* **1**, 035002 (2020).

[15] Griggs, W. *et al.* Polarized neutron reflectometry characterization of interfacial magnetism in an FePt/FeRh exchange spring. *Phys. Rev. Mater.* **6** (2022).

[16] Armen, R. S., Uitto, O. D. & Feller, S. E. Phospholipid component volumes: Determination and application to bilayer structure calculations. *Biophys. J.* **75**, 734–744 (1998).

[17] Foreman-Mackey, D. corner.py: Scatterplot matrices in python. *J. Open Source Softw.* **1**, 24 (2016).

[18] Sivia, D. S. *Data Analysis : a Bayesian tutorial* (Oxford University Press, Oxford, U.K., 2006).

[19] D'Agstino, R. B. An omnibus test of normality for moderate and large size samples. *Biometrika* **58**, 341–348 (1971).

[20] D'Agostino, R. & Pearson, E. S. Tests for departure from normality. *Biometrika* **60**, 613–622 (1973).

[21] Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).