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

Andrew R. McCluskey,[1,†] Andrew J. Caruana,[2] Christy J. Kinane,[2] Alexander J. Armstrong,[2] Tom Arnold,[1] Joshaniel F. K. Cooper,[2] David L. Cortie,[3] Arwel V. Hughes,[2] Jean-François Moulin,[4] Andrew R. J. Nelson,[3] Wojciech Potrzebowski,[1] and Vladimir Starostin[5]
(Open Reflectometry Standards Organisation)

[1]*European Spallation Source ERIC, P.O. Box 176, SE-221 00, Lund, SE*
[2]*ISIS-Neutron and Muon Source, Rutherford Appleton Laboratory, Didcot, Oxon OX11 0QX, GB*
[3]*Australian Nuclear Science and Technology Organisation, Lucas Heights, New South Wales, AU*
[4]*German Engineering Material Science at Heinz Maier-Leibnitz Zentrum, Helmholtz-Zentrum Hereon, Lichtenbergstraße 1, 85748 Garching, DE*
[5]*Institute of Applied Physics, University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, DE*

Driven by the availability of modern software and hardware, 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 provide 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, and 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 some prior understanding of the system is used to aid our understanding of some reflectivity profile [3–5]. Recently, developments in the availability of computer software for reflectometry analysis that include Bayesian functionality, such as Refl1d, refnx, anaklasis, and RasCAL [6–9], which implement sampling 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].

The discussion in this work will focus on the best practice for reporting the results from Bayesian and sampling-based analysis of neutron and X-ray reflectivity data. This work will not introduce Bayesian or sampling methods for neutron and X-ray reflectometry analysis. For those unfamiliar with these techniques, we suggest the work of Sivia and co-workers [5, 15] and more recent work focusing on reflectometry analysis [7, 14, 16, 17]. We hope that this work will inform best practices in data sharing from reflectometry analysis and inspire software developers to enable these to be accessed easily by the user.

Reflectometry analysis can be described, in the most simplistic terms, as a comparison and refinement of a model based on some parameters, $\mathbf{x}$, to reproduce some reflectivity data set, $\mathbf{D}$. This refinement process involves comparing the model to the data and calculating some goodness-of-fit parameter or likelihood, $p(\mathbf{D}|\mathbf{x})$, and modifying the model to either minimise or maximise this, as appropriate. A commonly used goodness-of-fit parameter is the χ^2 parameter which is found as [7],

$$\chi^2 = \sum_{q=q_{\min}}^{q_{\max}} \left[\frac{(R(q) - R(q)_{\mathrm{m}})}{\sigma_R(q)} \right]^2, \tag{1}$$

where, $R(q)$ and $R(q)_{\mathrm{m}}$ are the measured and modelled reflectivity at a given q, the measured momentum transfer vector, and $\sigma_R(q)$ is the uncertainty associated with the measured reflectivity at each q point. The likelihood is related to the χ^2 parameter, however, a larger likelihood is indicative of better agreement between the model and data,

$$\ln[p(\mathbf{D}|\mathbf{x})] = -\frac{1}{2}\left(\chi^2 + \sum_{q=q_{\min}}^{q_{\max}} \ln\left[2\pi\sigma_R(q)^2\right] \right). \tag{2}$$

The input for this refinement process is the model and some initial parameter values, which may be an absolute value or some parameter range, depending on the refinement algorithm. The output is a set of values for $\mathbf{x}$, potentially with associated error bars, where these are present they typically describe a standard deviation from the mean of a Gaussian probability distribution. This process implicitly assumes that the data is completely

* This work was developed as a part of the Open Reflectometry Standards Organisation Workshop in 2022 with contributions from all authors as part of a round-table discussion.

† andrew.mccluskey@ess.eu

reduced, accounting for all experimental parameters, uncertainties are accurately described and the model can accurately describe the data.

The input required depends on a minimisation algorithm being used, with some algorithms requiring a single starting guess (such as 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 shared as part of a publication describing the work. Furthermore, the minimisation is often performed with bounds in place, defining that the parameter values will lie within a given range. This range can be thought of as having a prior probability distribution, $p(\mathbf{x})$, where values of $\mathbf{x}$ outside of this range have a probability of 0. Even when a non-Bayesian approach is used in the analysis (i.e. Bayes theorem is not utilised), the result where bounds are set would be analogous to a Bayesian analysis with a uniform prior probability.

The optimised parameters from the minimisation algorithm, which depend on the particular algorithm, often include some statistical uncertainty. How these statistical uncertainties are found is beyond the scope of this work, but it is important to acknowledge that this uncertainty typically assumes that the probability distribution of the parameter is Gaussian in nature. This probability distribution is either the partial likelihood or posterior, the latter when some prior is included and Bayes theorem is applied and describes our understanding of the parameter values. When Bayesian modelling is used and the prior is included, the posterior probability is found as,

$$p(\mathbf{x}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{x})p(\mathbf{x}). \tag{3}$$

Therefore, when Bayesian modelling is performed, the priors and likelihood are of fundamental importance to the results that are obtained, the posterior, and therefore any scientific conclusions that are drawn.

The use of Bayesian inference can be valuable in the interpretation of reflectivity data, however, inconsistency in the description of the process will result in analysis that cannot be reproduced or easily understood. This can range from not giving the bounds of a uniform prior probability, to failing to share the complete sampling chain of a Markov chain Monte Carlo sampling or details of any autocorrelation analysis (the last of which, the authors of this work admit to being guilty of [13]). In this letter, we outline the best practice for reporting the results of Bayesian analysis for neutron and X-ray reflectometry, we hope that this work will engage others to carefully consider how they report this information. Furthermore, uptake of the approaches discussed herein will lead to greater clarity about the models and assumptions used in, and reproducibility of, our analyses.

TABLE I. An example of the presenting uniform priors in a tabular format. Reproduced from [14], where each parameter was either constrained to a given value or sampled within the prior range.

Parameter	Constrained Value	Prior Range
d_h/Å	10.0	$[8.0, 16.0)$
V_h/Å^3	339.5	$[300.0, 380.0)$
d_t/Å	21.0	$[10.0, 26.0)$
ϕ_t	1.0	$[0.5, 1.0)$
V_t/Å^3	850.4	$[800.0, 1000.0)$
σ/Å	2.9	$[2.9, 5.0)$

II. PRIOR

The most common prior probability that is used for a parameter is uniform between two values. The use of a bounded parameter along with some traditional χ^2-minimisation method and a parameter with a uniform prior and a Bayesian maximum *a-posteriori* approach will lead to the same result. For priors that are uniform is it important that the upper and lower bounds are reported, this can be achieved with a simple table (see Table I for an example) to be included in the article or supplementary information.

Currently, the use of non-uniform, informative priors is less common in the analysis of reflectometry. However, the increasing popularity of Bayesian methods and interest in using complementary methods for analysis means that these are likely to become more 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.

When a prior probability can be described with a mathematical function, this should be done by providing this function in the clearest possible language. For example, if the prior is taken from a single complementary measurement that is defined as a value with some uncertainty, which represents a normal distribution with a mean and standard deviation, this information should be provided. This is shown in Figure 1 for the density of silicon nitride (Si_3N_4) produced by atomic layer deposition [18]) which is used to inform the value for a scattering length density for some layer of the material. Such a prior probability distribution could be described graphically as in Figure 1, in prose as being "normally distributed with a mean of $2.9\,\mathrm{g\,cm^{-3}}$ and a standard deviation of $0.1\,\mathrm{g\,cm^{-3}}$", mathematically as,

$$p(\rho_m) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\rho_m - \mu}{\sigma}\right)^2\right] \tag{4}$$

where, $\mu = 2.9\,\mathrm{g\,cm^{-3}}$ and $\sigma = 0.1\,\mathrm{g\,cm^{-3}}$, or more concisely $p(\rho_m) \sim \mathcal{N}(\mu = 2.9\,\mathrm{g\,cm^{-3}}, \sigma = 0.1\,\mathrm{g\,cm^{-3}})$. The same descriptive approach could be taken for any

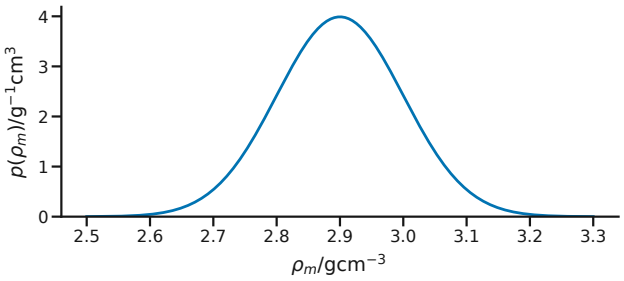

FIG. 1. A prior probability distribution for Si_3N_4 with a density of $\rho_m = (2.9 \pm 0.1)\,\mathrm{g\,cm^{-3}}$.

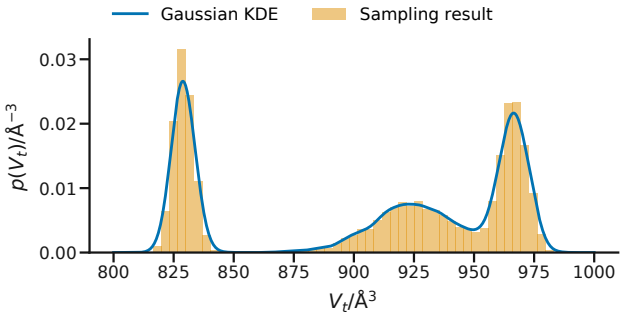

FIG. 2. A hypothetical prior probability distribution for a DPPC lipid that could arise from a molecular dynamics simulation (orange histogram) and a Gaussian kernel density estimation for the probability distribution using a bandwidth factor of 0.05 (blue line).

common statistical distribution, including log-normal or truncated normal distributions.

In the case that the prior distribution cannot be described easily with some mathematical function, for example, if it is a multimodal model result from some other sampling method, then the chain from the prior sampling should be given. The chain is all of the samples investigated in the sampling and should be shared, although in the case that this chain is very large a subsampled object may be shared, in which case, the autocorrelation analysis performed should be described (see Section IV for a more complete discussion of this). To use such a prior probability in Bayesian analysis, some functional description of the prior must be generated, most commonly this will be some kernel density estimation, when this is used it is also necessary to state the structure of the kernel being used. An example of this is shown in Figure 2, where the prior probability for the volume of a phospholipid tail could be found from molecular dynamics simulation where there are three common conformers that the lipid is likely to have.

III. LIKELIHOOD

Bayes theorem (given in Equation 3) consists of the product of the prior and the likelihood. The former describes our current understanding of the system before we conduct some experiments, while the latter describes the understanding that we gain from the data. Therefore, we must state how the likelihood for a given model and data is calculated, in particular, because although Equation 2 is a common approach, it assumes a normally distributed uncertainty. While a normally distributed uncertainty is the most common, it may not be accurate in all circumstances, for example, when low numbers of counts are present, a Poisson uncertainty may offer a more accurate description.

IV. POSTERIOR

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. Assuming there are m parameters under investigation, the posterior will be an m-dimensional probability distribution. The result of a Bayesian sampling process is a "chain" consisting of n samples for each parameter. Therefore, the full chain has a shape (m, n). Typically these are histogrammed to show the probability of different values of the parameters, however, to identify independent samples in the chain, autocorrelation analysis may be performed and the chain "thinned". We will not cover in detail autocorrelation analysis other than to say that it helps to identify the length of separation required for samples to be independent and that thinning means that we only included samples separated by this length in the final chain. Additionally, it is valuable to rreport the use of convergence diagnostics, such as the Gelman-Rubin statistic [19], which can assist in determining if a chian appropriately describes a posterior.

Either the full posterior chain or the thinned chain should be shared along with details of any autocorrelation analysis to accompany any Bayesian or sampling analysis. This will allow the best replication and verification of any results obtained from the data. Furthermore, large output files such as these chains can be easily shared using some versioned data repository, such as Zenodo or those available at specific institutions. Additionally, to allow the reader to quickly interpret the sampled posterior, a graphical description (such as that in Figure 3) should be included, at a minimum, in the supplementary information of the work. The importance of presenting the full posterior graphically lies in the ability to easily interpret correlations between parameters through this medium.

To report values for parameters and some form of statistical uncertainty, two clear approaches can be taken from the posterior chain. The first is to use some known statistical distribution that describes the samples well.

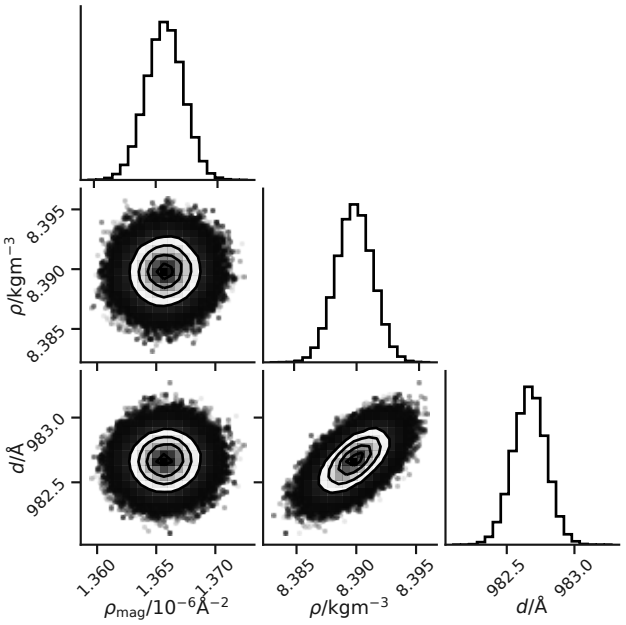

FIG. 3. An example of a graphical depiction of the unthinned posterior as a corner plot (produced using the `corner.py` package [20]), representing a three-dimensional probability distribution showing the posterior distribution for the parameters of nickel magnetic scattering length density, nickel mass density, and nickel layer thickness, from the analysis of a nickel layer on a silicon block.

This is best defined for a normal distribution, for which there are statistical tests to check normality, such as the D'Agostino and Pearson's test [21, 22] (which is available in the SciPy library as `scipy.stats.normaltest` [23]). As with 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 greater than 0.01 when 1000 random samples are used, therefore we can quote the parameters as normal distributions; $\rho_{\mathrm{mag}} = (1.366 \pm 0.001)\,10^{-6}\,\text{Å}^{-2}$, $\rho_m = (8.390 \pm 0.001)\,\mathrm{kgm}^{-3}$, and $d = (982.668 \pm 0.121)\,\text{Å}$.

If it is not possible to describe the m-dimensional distribution using some statistical test and a common distribution type, then confidence intervals can be given. Where these are used the percentage of the confidence interval must be defined alongside it. Alongside these confidence intervals, it is typically most accurate to give the maximum probability value for the parameter, rather than the numerical mean which may sit in a region of low probability.

Regardless of how the chain is shared, as a component of a fully reproducible analysis, the author should give details of the software packages, scripts, and data used to produce the analysis and any random number seeds that were defined. This means that if the chain is not available, the reader can rerun the sampling and replicate the results. Included in this is information regarding specific version numbers for different software packages, as these can create irreplicable results between version numbers.

V. CONCLUSION

The use of Bayesian analysis in neutron and X-ray reflectometry is increasing, and alongside this, there is a need for analytical clarity and reproducibility. We have outlined the best practice, based on experience, for reporting information from Bayesian analysis. 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 or maybe not be described mathematically. We mentioned the importance of including the specific likelihood function used in an analysis. 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 and confidence intervals in reporting. 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. Furthermore, we hope that developers of analysis software will take this work as a call to arms to include these best practices as easy-to-access methods in their software. Finally, if results of neutron and X-ray Bayesian analysis is shared as outlined in this work, then the analysis will be both reproducible and comprehensible.

DATA AVAILABILITY

Electronic Supplementary Information (ESI) available: All analysis/plotting scripts and data files allowing for a fully reproducible and automated analysis workflow, using show your work! [24], for this work is available at `https://github.com/arm61/reporting_sampling` (DOI: 10.5281/zenodo.xxxxxxx) under an MIT license, while the paper is shared under a CC BY-SA 4.0 license [25].

CReDiT AUTHOR STATEMENT

A.R.M: Conceptualization, Methodology, Resources, Writing - original draft, Writing - review & editing, Visualisation, Project administration. A.J.C & C.J.K.: Con-

ceptualization, Writing - review & editing. Other authors: Writing - review & editing.

ACKNOWLEDGMENTS

All authors gratefully acknowledge the contribution of Rev. T. Bayes (1701-1761) and D. S. Sivia, an early pioneer in using Bayesian methods in reflectometry analysis.

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

[16] Hughes, A. V. *et al.* Physical properties of bacterial outer membrane models: Neutron reflectometry & molecular simulation. *Biophys. J.* **116**, 1095–1104 (2019).

[17] Aboljadayel, R. O. M. *et al.* Determining the proximity effect induced magnetic moment in graphene by polarized neutron reflectivity and x-ray magnetic circular dichroism (2021). arXiv:2101.09946.

[18] Knoops, H. C. M. *et al.* Atomic Layer Deposition of Silicon Nitride from Bis(*tert*-butylamino)silane and N_2 Plasma. *ACS Appl. Mater. Interfaces* **7**, 19857–19862 (2015).

[19] Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).

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

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

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

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

[24] Luger, R. showyourwork. `https://github.com/rodluger/showyourwork` (2021).

[25] McCluskey, A. R. reporting_sampling (2022). URL `https://github.com/arm61/reporting_sampling`.