



Cite this: *Phys. Chem. Chem. Phys.*,  
2021, 23, 2637

# Relaxometry models compared with Bayesian techniques: ganglioside micelle example†

Pär Håkansson 

In this work a methodology to perform Bayesian model-comparison is developed and exemplified in the analysis of nuclear magnetic relaxation dispersion (NMRD) experiments of water in a ganglioside micelle system. NMRD is a powerful tool to probe slow dynamics in complex liquids. There are many interesting systems that can be studied with NMRD, such as ionic and lyotropic liquids or electrolytes. However, to progress in the understanding of the studied systems, relatively detailed theoretical NMRD-models are required. To improve the models, they need to be carefully compared, in addition to physico-chemical considerations of molecular and spin dynamics. The comparison is performed by computing the Bayesian evidence in terms of a thermodynamic integral solved with Markov chain Monte Carlo. The result leads to a conclusion of two micelle water-pools, and rules out both less and more parameters, *i.e.*, one and three pools. On the other hand, if only the quality of the fits is considered (*i.e.*, mean square deviation or  $\chi^2$ ) a three water-pool model is the best. The latter can be expected since with more adjustable parameters a better fit is more likely. Bayesian evidence is needed to rank the uncertainty of the models, and in order to explain the outcome a notation of Ockham-entropy is defined. The two approximate selection tools, Akaike and Bayesian information criterions, may lead to wrong conclusions compared to the full integration. This methodology is expected to be one of several important tools in NMRD model development; however, it is completely general and should find awakened use in many research areas.

Received 8th September 2020,  
Accepted 2nd January 2021

DOI: 10.1039/d0cp04750c

rsc.li/pccp

## 1 Introduction

The measurement of relaxation rates of nuclear spins *versus* resonance frequency is typically denoted as relaxometry and the recorded data form a relaxation dispersion (NMRD) profile. The experiment typically covers kHz to MHz resonance frequency and adds additional information to NMR spectroscopy (denoted as NMR relaxation). In particular, relaxometry can characterise the dynamics in a wide range from the nano- to micro-second regime.<sup>1</sup> These properties open the way to study complex systems such as cement, petroleum fluids, biological tissue and collective motion of whole proteins,<sup>2</sup> and dynamics in liquid crystals.<sup>3</sup>

To obtain more meaningful information from relaxation experiments, relaxation models are utilized to provide details of the mechanisms underlying spin relaxation.<sup>4</sup> Starting from the spin-system under study, the relevant mechanism may involve dipole-dipole, chemical exchange, paramagnetic and quadrupole contributions,<sup>5–12</sup> as well as chemical shift anisotropy.<sup>13</sup>

The challenges in building models involve the incorporation of consistent molecular dynamics processes to explain experimental observations and account for spin-dynamics with

sufficient accuracy.<sup>14–16</sup> Hence, these challenges may come in the form of (i) multiple physical interpretations and (ii) more than one parametrization within a specific physical picture. By addressing (i) and (ii), the relaxometry-research community can strengthen the communication of results and make progress in model development for challenging systems.

Bayesian model selection is expected to provide a powerful method to rank conclusions drawn from different models and thus navigate past the challenges. Hence, in light of the data, NMRD-profile and additional experiments such as NMR-relaxation, diffusion and X-ray, we can determine the “best” model out of several competing proposals. The definition of “best” needs to include more than minima of cost functions like  $\chi^2$ .<sup>17</sup> Rather, it should focus on the reproducibility and a balance quality of fit *versus* uncertainty due to the multitude of parameters to be determined.<sup>18</sup>

Model selection is expected to help in navigating the physics and chemistry underpinning the relaxation models. This goes beyond the qualitative guide for model selection *via* Ockham’s razor, which may be formulated as “it is futile to do with more what can be done with fewer”.<sup>19</sup>

For the water study in this work, the proton dipole-dipole (DD)-relaxation is the most important mechanism to model. Even for one mechanism, multiple relaxation pathways can prove important, such as intramolecular contribution,<sup>20</sup>

NMR Research Unit, University of Oulu, P.O. Box 3000, 90014 Oulu, Finland.

E-mail: par.hakansson@oulu.fi

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0cp04750c



intermolecular DD modulated by molecular diffusion,<sup>21,22</sup> and/or proton exchange.<sup>23</sup> The vast majority of relaxometry studies merge relevant physics in an analytical model describing the dynamics of the system and lead to a relatively large set of adjustable parameters.<sup>5</sup> Ideally a signature NMRD feature can be used and strengthen the model interpretation with fewer parameters such as DD-relaxation enhancement.<sup>11</sup> Such pronounced features are explored for paramagnetic<sup>10,16</sup> super-paramagnetic<sup>24</sup> and quadrupole spin-systems.<sup>25,26</sup>

In this work the methodology is provided for Bayesian model selection and tested on micelle NMRD data.<sup>5</sup> In particular the parametrization for a class of models (noted as (ii) above), addresses the question: why does the two water-pool model give the best explanation of data and not a simpler or more complex model? Hence, does the data support this more complex model?

The communication can thus be strengthened, especially towards researchers who are not experts in this field. Other research-fields such as astrophysics and cosmology have a long tradition of using and benefiting from these methods.<sup>27,28</sup>

A deeper understanding of approximations in analytical models as well as a partial or complete reduction of adjustable parameters has been obtained with molecular dynamics simulations. In relaxometry studies, this is achieved for paramagnetic relaxation enhancement,<sup>29</sup> rotation and translation of small molecules,<sup>30</sup> water at the bilayer interface,<sup>8</sup> water in cement,<sup>31</sup> and ionic liquids,<sup>13</sup> lyotropic phases,<sup>14,32</sup> porous organic cages<sup>7</sup> and proteins.<sup>33</sup> The simulation studies opens a way to explore the validity of analytical models and for some systems serves as the method of choice. However, model comparisons are not removed by using molecular dynamics simulations; instead, questions such as what is the sufficient size and timespan as well as what forcefield or quantum-chemistry methods are sufficiently accurate arise. Hence, based on different types of simulations, what outcome is the best in describing NMR-relaxation/relaxometry?

The estimation of model parameters and model comparison is centred around a cost function such as  $\chi^2$ :

$$\chi^2(\mathbf{x}) = \sum_{i=1}^N \frac{(R_i^T(\mathbf{x}) - R_i^{\text{exp}})^2}{\sigma_i^2}, \quad (1)$$

where  $R_i^T(\mathbf{x})$  is the theoretical model, dependent on adjustable parameters  $\mathbf{x}$ ,  $R_i^{\text{exp}}$  are the experimental observations and  $\sigma_i$  is the standard deviation in experimental observation  $i$ . The adjustable parameters of the model ( $\mathbf{x}$ ) are of size  $k$ , *i.e.*, correlation times, order parameter diffusion constant, *etc.*, dependent on the physics and chemistry of the system under study. In this format we may merge multiple relaxometry experiments at various temperatures, diffusion measurements and high-field relaxation data.<sup>5,34</sup>

There are large sets of tools proposed around cost functions to compare models; the reduced  $\chi^2$  can be computed by dividing by  $N - k$  and for instance over-fitting can be proposed. However, in particular for nonlinear models we don't know how many degrees of freedom are present and the guess " $N - k$ ",

where  $k$  is the number of parameters, is not robust.<sup>35</sup> Hence, the  $\chi^2$ -value does not allow for model comparison in general.

Two approximate routes: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), both attempt to balance quality of fit with complexity of the model. These methods have as one assumption an asymptotic limit of  $N$  for a reduced error.<sup>36</sup> In recent work these have been tested and found to have limited value in NMRD micelle studies.

The cross-validation (analysing part of the data) or more generally the Bootstrap-sampling have the drawbacks that the validity is in the asymptotic limit<sup>17,37</sup> and for model comparison, only approximate forms are provided.<sup>36,38</sup> Given the common situation of a small data set ( $N$ ) in NMR-relaxation and NMRD the bootstrap method will not be explored in the current work.

In this work we follow instead Bayesian theory and compute the model evidence, which combines uncertainty due to a multitude of adjustable parameters and the quality of fit. Thus, a quantitative tool corresponding to Ockham's razor is obtained.<sup>36,39,40</sup> A path to understand Bayesian evidence is assisted by identifying and defining Ockham-entropy in this work, since suitable terms to discuss the results are lacking.<sup>18</sup>

Bayesian evidence is a challenging high-dimensional integral to compute even with Monte Carlo methods. However, by introducing a thermodynamic integration the sought property is computed with multiple but vastly less demanding Markov chain Monte Carlo (MCMC) methods.<sup>41</sup> In this work we implement thermodynamic integration and show that this is a computationally feasible route for relaxometry model comparison. The result is quantitative discrimination between models combining quality of fit and penalising over parametrization with Ockham-entropy.

In this paper, the relaxation model derived in previous work<sup>5</sup> is presented in Section 2. In Section 3, the relevant Bayesian theory is provided with discussion of the introduced Ockham-entropy. This is followed by aspects to consider in MCMC simulations and the code is discussed in Section 4. Finally the NMRD example is analysed and presented in Section 5, followed by a summarising conclusion in Section 6.

## 2 Relaxation model

In this work we reexamine and follow closely the notation for the model presented in ref. 5, and refer to this work for further model discussions. The experimental observation of longitudinal relaxation rate  $R_1^{\text{exp}}(\omega_0)$  is the rate-of-change in the spin ensemble due to the influence of the environment at resonance frequency  $\omega_0$ . At relatively low fields the dominating relaxation pathway is the DD-mechanism where the relaxation matrix for pairwise DD-coupling is well documented in the literature.<sup>4,23</sup>

The key to the relaxation study<sup>5</sup> is the formulation of a relevant physical picture, *i.e.* the two water-pool model depicted in Fig. 1. The two types of waters are denoted ( $\beta_i$ ,  $i = 1, 2$ ) both located at the micelle. Reasons for more than one pool may be due to the large ganglioside head-group allowing for several water locations, leading to different exchange times ( $\tau_{w,i}$ ) and



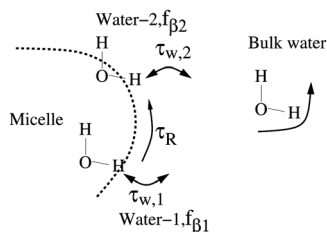


Fig. 1 Scheme of the two-pool H<sub>2</sub>O relaxation model. Hence, two water-pools at the micelle with populations  $f_{\beta 1}$  and  $f_{\beta 2}$ , in addition to the bulk water. Micelle overall rotation has a characteristic time  $\tau_R$  and the waters have residence times  $\tau_{w,i}$  ( $i = 1, 2$ ).

populations ( $f_{\beta i}$ ), but modulation by the same overall micelle rotation  $\tau_R$ . The non-bound waters are denoted as bulk water in Fig. 1. In addition there are 2–3% ganglioside protons that are not addressed with the model.<sup>5</sup>

The theoretical prediction ( $R_1^T(\omega)$ ) takes the form

$$R_1^T(\omega) = \alpha + f_{\beta 1}R_{1,\beta 1}(\omega) + f_{\beta 2}R_{1,\beta 2}(\omega) \quad (2)$$

where  $\alpha$  are the sum of field independent contributions.<sup>5</sup> The order parameters originate from expressing spectral densities within a two step model<sup>42</sup>

$$J_i(n\omega) = (1 - S_i^2)\tau_f + S_i^2\tau_{c,i} \frac{1}{1 + (n\tau_{c,i}\omega)^2}, \quad (3)$$

where  $\tau_f$  denotes fast local dynamics not explicitly accounted for in the final model, but enters the parameters  $\alpha$ . The field dependent parts for the two waters are given by

$$R_{1,\beta i} = S_i^2\tau_{c,i}\beta_{\text{intra}} \left\{ \frac{0.2}{1 + (\omega_0\tau_{c,i})^2} + \frac{0.8}{1 + (2\omega_0\tau_{c,i})^2} \right. \\ \left. + W_D \left[ 0.1 + \frac{0.3}{1 + (\omega_0\tau_{c,i})^2} + \frac{0.6}{1 + (2\omega_0\tau_{c,i})^2} \right] \right\}, \quad (4)$$

where  $W_D$  denotes weight factor for inter-molecular proton interaction (second part of eqn (4)),  $\tau_{c,i}$  represent an effective correlation time and  $\beta_{\text{intra}}$  is the amplitude of intra-water proton interaction. The effective correlation time accounts for the combined processes of micelle reorientation and water exchange:

$$\tau_{c,i} = \frac{\tau_R \cdot \tau_{w,i}}{\tau_R + \tau_{w,i}}. \quad (5)$$

## 2.1 Proposed models

The intra-molecular DD-amplitude may be determined from

$$\beta_{\text{intra}} \equiv \frac{3}{2} \left( \frac{\mu_0}{4\pi} \hbar \gamma_H^2 r_{\text{HH}}^{-3} \right)^2, \quad (6)$$

with proton distance in water  $r_{\text{HH}} = 1.58 \text{ \AA}$  and vacuum permeability  $\mu_0$ , Planck's constant per radians ( $\hbar$ ), and the gyromagnetic ratio of a proton ( $\gamma_H$ ). However, the amplitude for inter-molecular proton interactions is not known and takes the effective form  $\beta_{\text{inter}} = \beta_{\text{intra}}W_D$ . In this format we can not resolve if intra or inter are the dominating relaxation pathways and this is left for future work.

Relaxometry experiments are less sensitive to some model aspects and it makes sense to reduce the number of determined parameters. Following ref. 5, the product  $S_i^2f_{\beta i}$  is estimated, together with  $\tau_{c,i}$  and  $\alpha$ . Thus, the two water-pool model contains five adjustable parameters per temperature<sup>5</sup>  $\mathbf{x} = \{\alpha, S_i^2f_{\beta i}, \tau_{c,i}\}$ , ( $i = 1, 2$ ) and 15 in total.

We can not tell in advance how many water-pools there actually are, and more importantly, how many the NMRD experiments are sensitive to. This opens up a qualitative discussion along the line of Occam's razor to motivate a suitable number of water-pools, two in the original work.<sup>5</sup> The number of water-pools, however, forms a transparent example where quantitative model comparison may be followed. Thus we consider the models  $M_a$ ,  $M_b$  and  $M_c$  with two, one and three water-pools, respectively, leading to the number of adjustable parameters being:  $k_a = 15$ ,  $k_b = 9$  and  $k_c = 21$ .

The results are discussed in terms of Stokes–Einstein–Debye (SED) relation and furthermore, a challenge is observed in extracting temperature dependence in ( $S_i^2f_{\beta i}$ ). These aspects motivate a two-pool model  $M_d$  with a reduced set of parameters ( $k_d = 10$ ) explicitly including SED and with a constant parameter ( $S_i^2f_{\beta i}$ ).

As pointed out,<sup>5</sup> modelling can be refined if we have complementary heavy water experiments.<sup>43</sup> Furthermore, classical molecular dynamics may propose individual values to the merged ( $S_i^2f_{\beta i}$ ) parameter,<sup>44</sup> and quantum chemical molecular dynamics may suggest constraints on proton exchange times. High field  $T_1$  measurement increases the spanned frequency range with one order of magnitude. Hence, the model comparison is evolving and needs to be open to new information.

## 3 Theory: Bayesian approach to model selection

To define the model selection, consider the probability of model  $M_j$  in the light of available data  $R^{\text{exp}}$  (in this work the NMRD profile) that is formulated with Bayesian theory. With the given (additional) information  $I$  (for instance known concentration of surfactants, temperature of experiments, etc.)<sup>41</sup> the probability takes the form

$$P(M_j|R^{\text{exp}}, I) \\ = \frac{P(M_j|I)p(R^{\text{exp}}|M_j, I)}{Z} \propto P(M_j|I)p(R^{\text{exp}}|M_j, I), \quad (7)$$

where the left hand side reads “probability of model  $M_j$  given experimental data ( $R^{\text{exp}}$ )” and  $Z$  is a normalisation constant. Thus, two models can be compared ( $P(M_a|R^{\text{exp}}, I)/P(M_b|R^{\text{exp}}, I)$ ) without knowing  $Z$ . The right hand side contains the prior probability  $P(M_j|I)$ , i.e., the probability given to the model before we have the data and the so called evidence  $p(R^{\text{exp}}|M_j, I)$ . The evidence may be expanded as<sup>41</sup>

$$p(R^{\text{exp}}|M_j, I) = \int_{V(\mathbf{x}_j)} L(R^{\text{exp}}|M_j, \mathbf{x}_j)p(\mathbf{x}_j|M_j)d\mu(\mathbf{x}_j), \quad (8)$$



where the likelihood function in this work is computed from eqn (1) as:

$$L(R^{\text{exp}}|M_j, \mathbf{x}_j) = \frac{1}{\prod_i \sqrt{2\pi\sigma_i^2}} \exp[-\chi^2(\mathbf{x}_j)/2]. \quad (9)$$

Thus an integral of likelihood function over the whole multi-dimensional parameter volume ( $V(\mathbf{x}_j)$ ) where parameters are considered random variables with probability measure  $\mu(\mathbf{x}_j)$ , hence, integral fulfils normalization criteria and parameters are well defined over the domain. To get the probability of model  $M_j$  we need to integrate the likelihood over the whole weighted parameter space given experimental data  $R^{\text{exp}}$  and information  $I$ .

### 3.1 Thermodynamic integration

For many problems, eqn (8) is a high-dimensional integral and is difficult to estimate with numerical methods.<sup>36,40</sup> A flavour of the numerical challenge is given in Fig. 2 where the details of the structure of the likelihood function needs to be integrated.<sup>40</sup>

By introducing a one-dimensional thermodynamic integral instead of attempting direct computation of eqn (8), a much less demanding problem is obtained. In fact, as exemplified,<sup>41</sup> this may reduce the required number of sampling configurations from  $10^{138}$  to  $10^6$ ! To compute eqn (8) as a thermodynamic integral (without introducing approximations) an auxiliary function is introduced:<sup>40,41</sup>

$$F(\beta) = \int_{V(\mathbf{x}_j)} L(R^{\text{exp}}|M_j, \mathbf{x}_j)^\beta p(\mathbf{x}_j|M_j) d\mu(\mathbf{x}_j), \quad (10)$$

and the derivate of  $\ln[F]$  ( $\partial \ln[F]/\partial \beta = F^{-1} \partial F/\partial \beta$ ) is used. This gives:

$$\begin{aligned} \frac{\partial}{\partial \beta} \ln[F(\beta)] &= -\frac{1}{F(\beta)} \int_{V(\mathbf{x}_j)} \frac{\chi^2(\mathbf{x}_j)}{2} L(R^{\text{exp}}|M_j, \mathbf{x}_j)^\beta p(\mathbf{x}_j|M_j, I) d\mu(\mathbf{x}_j) \quad (11) \\ &= -\frac{1}{2} \langle \chi^2(\mathbf{x}) \rangle_\beta \approx -\frac{1}{2N_{\text{MCMC}}} \sum_i^{N_{\text{MCMC}}} \chi^2(\mathbf{x}_{I(\beta)})^2, \end{aligned}$$

where the specific likelihood function (see eqn (9)) to power  $\beta$  is used and the mean value  $\langle \cdot \rangle_\beta$  is computed for the  $\beta$ -probability density  $L(R^{\text{exp}}|M_j, \mathbf{x}_j)^\beta p(\mathbf{x}_j|M_j) d\mu(\mathbf{x}_j)/F(\beta)$ . The approximation

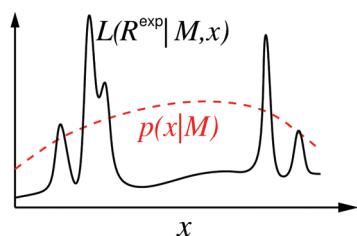


Fig. 2 One-dimensional sketch illustrating the challenge with eqn (8), the likelihood function ( $L$ ) in solid-black line and prior probability distribution of parameter ( $\mathbf{x}$ ) in dashed red line. The detailed structure of the likelihood function makes a high-dimensional integral challenging to compute by direct Monte Carlo sampling.

involves the statistical uncertainty from a finite number of MCMC configurations, described in Section 4.

Integrating eqn (11) over  $[0,1]$  gives

$$\ln[F(1)] = \ln[F(0)] - \int_0^1 \frac{1}{2} \langle \chi^2(\mathbf{x}_j) \rangle_\beta d\beta = \ln[p(R^{\text{exp}}|M_j, I)], \quad (12)$$

where  $\ln[F(0)]$  is zero (by normalisation) and the natural logarithm of the desired probability is obtained.

We note that in the limit  $\beta \rightarrow 1$  the integral in eqn (12) has a contribution from the quality of fit, hence, a smaller value of  $\langle \chi^2 \rangle_{\beta=1}$  reduces the negative log or increases the model probability. The meaning of the complete integral in eqn (12) is addressed in what follows.

### 3.2 Entropy and model complexity

Although the integral of eqn (12) is a mathematic path to compute Bayesian evidence, it also provides an insight on its meaning, as seen from the reformulation:

$$\begin{aligned} -\frac{1}{2} \langle \chi(\mathbf{x})^2 \rangle_\beta &= -C \int_{V(\mathbf{x})} \frac{\chi^2(\mathbf{x})}{2} L(\mathbf{x})^\beta d\tilde{\mu}(\mathbf{x}) \\ &= \frac{C}{\beta} \int_{V(\mathbf{x})} (\ln[L(\mathbf{x})^\beta]) L(\mathbf{x})^\beta d\tilde{\mu}(\mathbf{x}), \end{aligned} \quad (13)$$

where a compact notation for the probabilities in eqn (11) is introduced and the uniform prior-probability considered in this work  $[p(\mathbf{x}|M)]$  is contained in the integration boundaries and normalization constant  $C$  instead of explicitly included. The format of eqn (13) allows a comparison with the statistical mechanics formulation of entropy:<sup>45</sup>

$$S = -k_B \sum_n (\ln[P_n]) P_n \rightarrow -k_B \int_{V(\mathbf{x})} (\ln[P(\mathbf{x})]) P(\mathbf{x}) d\tilde{\mu}(\mathbf{x}), \quad (14)$$

taken in the continuous limit, where  $k_B$  is Boltzmann's constant and  $P$  is the probability density for physical states.

By comparing eqn (13) and (14) we see that the integral in eqn (12) collects the entropy, or equivalently, the uncertainty in going from only prior to final parameter information, from the optimisation (or fit) of model to the experimental data. The ( $\beta = 1$ ) contribution is lower since here the best information content is provided. Thus, a model is found that minimizes the  $\beta$ -integral but also maximizes the propability of the model [cf. eqn (7), simply considering log and non-log form]. With the entropy connection, just like a volume of ideal-gas compared with corresponding real-gas influences the entropy of a physical system, we can expect that the number of parameters, total volume, and extent parameters depend on each-other and also influence the entropy (cf. eqn (13)).

The literature proposes that Bayesian evidence consists of a best fit contribution and "model complexity",<sup>18,19,28,36,39,40</sup> however the correspondence and direct connection to entropy is not done. We find that for the physical chemistry community





**Table 1** Jeffrey's scale rating evidence against  $b$  via Bayes factor, assuming equal prior model probabilities for model  $a$  and  $b$ 

$2 \ln[B_{ab}]$	Evidence against $M_b$
0 to 2	Just mention
2 to 6	Positive
6 to 10	Strong
> 10	Very strong

it is useful to define Ockham's entropy:

$$\Delta S_{j,\text{Ockham}} \equiv \int_0^1 \langle \chi^2(\mathbf{x}_j) \rangle_\beta d\beta - \chi^2(\mathbf{x}_{j,\text{opt}}), \quad (15)$$

$$\Delta S_{j,\text{Ockham}} + \chi^2(\mathbf{x}_{j,\text{opt}}) = -2 \ln[p(R^{\text{exp}}|M_j, I)],$$

where the optimal fit ( $\chi^2(\mathbf{x}_{j,\text{opt}})$ ) is discussed in Section 4.1. With Ockham's entropy we can begin to follow the consequences of "too many parameters" when building models.

### 3.3 Bayes factor

With two models proposed ( $M_a, M_b$ ), eqn (12) enables the calculation of Bayes factor:<sup>36</sup>

$$\underbrace{\frac{P(M_a|R^{\text{exp}}, I)}{P(M_b|R^{\text{exp}}, I)}}_{\text{posterior odds}} = \underbrace{\frac{p(R^{\text{exp}}|M_a, I)}{p(R^{\text{exp}}|M_b, I)}}_{=B_{ab}(\text{Bayes factor})} \times \underbrace{\frac{P(M_a|I)}{P(M_b|I)}}_{\text{prior odds}}, \quad (16)$$

hence, for models that can be assumed equally probable before data is analysed the  $B_{ab}$  provides the posterior odds. Commonly Bayes-factor is discussed in terms of the empirical Jeffreys scale for the strength of evidence against model  $b$ , given in Table 1:

### 3.4 Approximate model selection criteria

There are a number of approximate model selection criteria and we will explore the Akaike and Bayesian information criterion, (AIC) and (BIC), respectively:

$$\text{AIC} = -2 \ln(L_{\text{opt}}) + 2k,$$

$$\text{BIC} = -2 \ln(L_{\text{opt}}) + k \ln(N),$$

where  $L_{\text{opt}}$  is the maximum likelihood.<sup>40</sup> For the case of Gaussian likelihood used in this work (see eqn (11)) further simplification is  $2 \ln(L_{\text{opt}}) = \chi^2_{\text{opt}} + \text{"constant independent of model"}$ . As for the full calculation provided in this work (see eqn (12)) BIC and AIC penalise a model with poor fit (large  $\chi^2$ ) and excessive adjustable parameters, with expected  $\mathcal{O}(N^{-1})$  error, of experimental data  $N$ .<sup>36</sup>

## 4 Methods

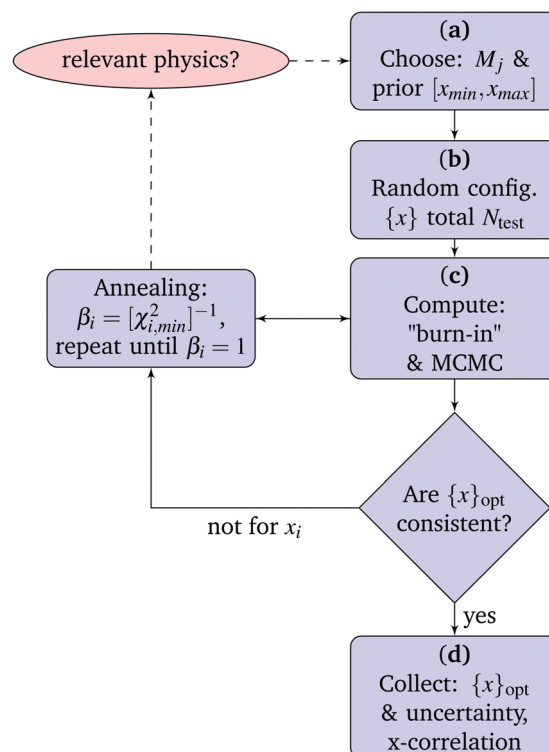
We present the parameter estimation with MCMC followed by the procedure for the thermodynamic integral (eqn (12)); for details we refer to the Matlab code and "readme" file presented in the GitHub example.<sup>46</sup> It is noted that central to the study is the details of  $\chi^2$  and thus the standard deviation ( $\sigma_i$ ) of experimental relaxation rates in eqn (1). The experimental relaxation rates in the work of ref. 5 where obtained by fitting

a function to  $R^{\text{exp}}$  data points, leading to a much too small standard deviation that was then calibrated with additional experiments. It is noted that the calibration factor of 10 for  $\sigma_i$  is used,<sup>5</sup> corresponding in this work to a factor  $10/\sqrt{2}$  due to the normalization of 2 in this work (see eqn (1)).

### 4.1 MCMC parameter estimation

Before any work with model comparison, parametrization is needed. The ideas are implemented in Matlab/octave code<sup>46</sup> and the steps are given in the flow-chart of parameter estimation (see Fig. 3). The start (a) for one of the models  $M_a, \dots, M_d$ , the parameter prior ( $p(\mathbf{x}_j|M_j, I)$ ) is chosen, that for this work is the uniform (uninformed) prior. For the user this involves setting the parameter-boundaries  $\mathbf{x}_{\min}$  and  $\mathbf{x}_{\max}$  (see Section 5). The choice of these ensures that parametrisation is reproducible within boundaries and no model is given extra advantage due to small interval. The parameter prior is used in (b) to generate  $N_{\text{test}}$  starting configurations (in the range 4–10  $\{x\}_{N_{\text{test}}}$ ), where these parameter sets are random variables from the sought prior distribution.

The core of MCMC is efficient sampling of configurations (c). For ease of introducing alternative prior distributions the current version, function MCMCstep.m,<sup>46</sup> follows the Metropolis algorithm,<sup>7,47,48</sup> hence, the proposed new parameter configurations are sampled in a random direction (not for instance guided by the likelihood function). Each parameter is given a move individually, following the Gibbs sampling algorithm.<sup>49</sup> The parameter moves are reflected at the boundaries ( $\mathbf{x}_{\min}$  and  $\mathbf{x}_{\max}$ ).

**Fig. 3** Flowchart MCMC parameter optimization.

The acceptance ratio of steps should as a rule of thumb be  $\sim 0.5$  to efficiently probe lower probability regions but where higher value (shorter steps) leads to greater correlation between configurations. It is noted that more efficient sampling may be found, but these require model specific implementation.<sup>47</sup> The step-length can be optimised *versus* the measured step correlation to further improve sampling. The step correlations are computed from the saved configurations, *i.e.*, normalised auto and cross-correlation functions<sup>50</sup> (see ESI† and implementation in the code<sup>46</sup>).

The step-lengths were determined by computing a sub-set of steps where the mean acceptance ratio is computed followed by calibration until within the interval  $[0.4, 0.6]$ . This was followed by  $\sim 10^5$  MCMC iterations towards equilibrium, the region of smallest  $\chi^2$  (denoted “burn-in”). The burn-in is followed directly by attempt of productive MCMC-steps ( $\sim 10^6$ ). The above sequence of steps is typically not intervened in by the user. At this point comes the main question of MCMC, have the configurations reached equilibrium? This question is addressed with a set of  $N_{\text{test}}$  trajectories and looking for consistency in  $\chi^2_{\text{opt}}$  *versus* the standard deviation of  $\chi^2$  (see Fig. S1 in ESI†). If some configurations remain in a local minimum, the annealing procedure is performed where simulations corresponding to higher “temperatures” (low  $\beta$ ) are made and any local minima are more easily surpassed.

Obviously there is no global minimisation routine (MCMC or otherwise) capable of arbitrary complex models; thus if some very challenging problems occur, the user should also consider if the model represents relevant physics worth proceeding with.

With all the  $\{\mathbf{x}\}_{N_{\text{test}}}$  configuration successfully completed (d) additional sets are computed up to and including the burn-in step, to verify that we have a reproducible global minimum in  $\chi^2$ . The results are summarised with  $x_j$ -histograms, *i.e.* total density projected on each variable  $x_j$ . Histograms are integrated to provide 95%-confidence intervals. The optimized parameters are in this work given by

$$\mathbf{x}_{\text{opt}} = \arg \min_{\mathbf{x}} (\chi^2(\mathbf{x})), \quad (17)$$

reported together with the 95%-confidence intervals. For the considered models one calculation of MCMC takes 5–20 minutes per separate MCMC-run on a MacBook Pro 3.1 GHz intel processor. The implementation<sup>46</sup> in Matlab R2019b has one option to submit multiple runs to a supercomputer. The same code also runs with the freely available Octave.<sup>51</sup>

What is found valuable in the provided implementation<sup>46</sup> is the ease by which calibration of the MCMC step-length is made automatic with the univariate moves (Gibbs sampling<sup>49</sup>). In addition, how random initial configurations converge by exploring a set of fixed annealing temperatures for the challenging cases. Challenges may appear for models with larger number of parameters that are still interesting to give a fair comparison. The annealing enables completely random initial configurations to query if the proposed model-result is reproducible within the given parameter boundaries.

There are several MCMC programs and it is found that in ref. 52 (used in NMRD<sup>53</sup>) the configuration move used is not transparent to the implementation including annealing. Furthermore, initial configurations are proposed to be from a “suitable starting point”.<sup>52</sup> Such an approach does not work here where the goal is to probe the whole parameter volume as potential starting points.

## 4.2 Numerical thermodynamic integral

The numerical computation of the thermodynamic integral in eqn (12) is performed with the set of optimised configurations  $\{\mathbf{x}_{\text{opt}}\}_{N_{\text{test}}}$  (see Fig. 3) computing integrand at discrete  $\beta$  values. The discretization is found by starting from five  $\beta$  values with logarithmic discretisation on interval  $(0,1)$ , where it is noted we have the  $\beta = 1$  value from the parametrization and the  $\beta = 0$  case is most efficient to generate with independently sampled configurations from the prior parameter distribution.

Based on the integral computed with the trapezoidal rule,<sup>17</sup> the interval with the largest contribution is located and split in half and the integral is recomputed. This procedure continues until the difference between two subsequent integrals is less than the desired tolerance. The same discretization was then used for all the models with  $\beta$ -values provided in the GitHub example.<sup>46</sup>

## 5 Results and discussion

The main aim of this work is to provide numerical (MCMC) Bayesian model selection to the relaxometry community and exemplify how this can be used in the interpretation of NMRD profiles. The example considered is relaxometry study of water in ganglioside Micelles,<sup>5</sup> where previous work considers a model with two water-pools; this number (and not one, three or more pools) is deemed the feasible amount of information to extract from NMRD-profiles. The main question is if quantitative model selection can conclude how many pools is best with the given data.

### 5.1 Model parametrization

From the parametrization of the models ( $M_a$ ,  $M_b$ ,  $M_c$  and  $M_d$ ) following Fig. 3 it is identified if any of the models provide exceptional challenges before continuing work with model comparison. The first stage, the parametrization, with optimal fits is provided in Fig. 5 and the estimated parameters in Tables 2–5, listing the parameter boundaries in the bottom rows.

Common for models  $M_a$  to  $M_d$  are that optimisation is performed globally over the 90 data points and the exploration of weight of inter-DD contribution case  $W_D = 0.4$ . It was found<sup>5</sup> that the model dependence of  $W_D$  is weak and it is not the focus of this work to explore this further.

For model  $M_a$  it is found that MCMC estimation provide within 95%-confidence interval,  $\chi^2_{\text{opt}} = 68$  and largely, within the confidence interval, the same parameters as the original work,<sup>5</sup> with data in Table 2.



**Table 2**  $M_a$ , following the two-pool assumption of ref. 5. Global MCMC parameter estimation with three temperature data (sample concentration 15 mM). The fit is displayed in Fig. 5

$T$ (°C)	$\tau_{c,1}$ (ns)	$\tau_{c,2}$ (ns)	$\alpha$ (s <sup>-1</sup> )	$S_1 f_{\beta 1}$ (10 <sup>-3</sup> )	$S_2 f_{\beta 2}$ (10 <sup>-3</sup> )
10	<b>157</b> [144 168]	<b>11.6</b> [6.6 18.6]	<b>0.64</b> [0.58 0.69]	<b>0.25</b> [0.23 0.27]	<b>1.6</b> [1.1 2.6]
15	<b>132</b> [125 147]	<b>9.0</b> [5.9 17]	<b>0.58</b> [0.54 0.62]	<b>0.25</b> [0.22 0.27]	<b>1.8</b> [1.0 2.6]
20	<b>123</b> [115 135]	<b>9.6</b> [6.4 17]	<b>0.53</b> [0.49 0.57]	<b>0.23</b> [0.20 0.25]	<b>1.6</b> [0.9 2.3]
$x_{\min} \leftrightarrow x_{\max}$	1.0 $\leftrightarrow$ 1000	0.1 $\leftrightarrow$ 100	0.001 $\leftrightarrow$ 0.8	0.01 $\leftrightarrow$ 100	0.01 $\leftrightarrow$ 100

**Table 3**  $M_b$ , following the one-pool assumption of ref. 5. Global MCMC parameter estimation with three temperature data (sample concentration 15 mM). Total  $\chi^2$  is 1346. The fit is displayed in Fig. 5

$T$ (°C)	$\tau_{c,1}$ (ns)	$\alpha$ (s <sup>-1</sup> )	$S_1 f_{\beta 1}$ (10 <sup>-3</sup> )
10	<b>119</b> [115 124]	<b>0.77</b> [0.75 0.80]	<b>0.38</b> [0.36 0.39]
15	<b>102</b> [99 107]	<b>0.70</b> [0.69 0.72]	<b>0.37</b> [0.36 0.39]
20	<b>94</b> [91 98]	<b>0.64</b> [0.63 0.66]	<b>0.35</b> [0.34 0.36]
$x_{\min} \leftrightarrow x_{\max}$	1 $\leftrightarrow$ 1000	0.001 $\leftrightarrow$ 0.8	0.01 $\leftrightarrow$ 100

Thus, the same physical interpretation<sup>5</sup> is summarised, that is, a slow correlation time  $\tau_{c,1}$  dominated by overall rotational diffusion of ganglioside micelles; that from the Stoke–Einstein–Debye (SED) relation is consistent with the radius of 54 Å, concluded from measurements by other techniques.<sup>5</sup> The shorter correlation time, originating from  $\tau_{c,2}$  (see Fig. 1) is ascribed to the proton exchange-modulated reorientation where water protons exchange with labile surfactant protons.<sup>5,54</sup>

The difference in the size of the confidence intervals is striking; for MCMC they are always larger where the difference for ( $S_2^2 f_{\beta 2}$ ) is a factor 50.<sup>5</sup> Considering steps to check the MCMC

parameter estimation, there are 7 simulations with random starting configurations within parameter boundaries (see range in Table 2). By estimating the MCMC-parameter autocorrelation ( $\tau_{\text{steps}}$ ) a handle is obtained on how well the parameter space is sampled. It is found that  $\chi^2$ , ( $S_2^2 f_{\beta 2}$ ) and ( $S_1^2 f_{\beta 1}$ ) are sampled with  $4 \times 10^4$ ,  $2 \times 10^2$  and  $1 \times 10^3$  correlation times, respectively. This points to a sufficiently sampled parameter-space given that additional “burn-in” is separately performed (see ESI,† Fig. S2).

From the computed cross-correlations (see ESI†), a substantial amount is seen between ( $S_2^2 f_{\beta 2}$ )–( $\tau_{c,2}$ ) and ( $S_2^2 f_{\beta 2}$ )– $\alpha$  for the 20 °C NMRD-profile and they are illustrated with histograms in the upper panels of Fig. 4. The parameter correlation explains the large error boundary found in MCMC estimation. Note that this in itself is not a problem for the MCMC-validity, except it leads to a large parameter uncertainty. This is since the whole parameter space is still sampled provided sufficiently long MCMC simulation is performed. On the other hand, for conventional fit-routine, parameter correlations introduce an error in confidence intervals since the parameter distribution is not known in that case.<sup>39</sup>

The parametrization of model  $M_b$  is given in Table 3 and Fig. 5 ( $\chi_{\text{opt}}^2 = 1346$ ) assumes one water-pool. The model provides

**Table 4**  $M_c$ , following the three-pool assumption of ref. 5. Global MCMC parameter estimation with three temperature data (sample concentration 15 mM). The fit is displayed in Fig. 5

$T$ (°C)	$\tau_{c,1}$ (ns)	$\tau_{c,2}$ (ns)	$\tau_{c,3}$ (ns)	$\alpha$ (s <sup>-1</sup> )	$S_1 f_{\beta 1}$ (10 <sup>-3</sup> )	$S_2 f_{\beta 2}$ (10 <sup>-3</sup> )	$S_3 f_{\beta 3}$ (10 <sup>-3</sup> )
10	<b>183</b> [153 215]	<b>49</b> [17 70]	<b>4.5</b> [1.2 40]	<b>0.58</b> [0.1 0.65]	<b>0.18</b> [0.13 0.24]	<b>0.3</b> [0.23 0.85]	<b>3.5</b> [1.8 12]
15	<b>178</b> [132 204]	<b>63</b> [16 69]	<b>5.0</b> [1.1 11.1]	<b>0.54</b> [0.2 0.6]	<b>0.15</b> [0.12 0.25]	<b>0.3</b> [0.22 0.85]	<b>2.9</b> [1.4 11]
20	<b>140</b> [122 192]	<b>40</b> [15 69]	<b>2.5</b> [0.8 7.9]	<b>0.39</b> [0.1 0.54]	<b>0.16</b> [0.08 0.22]	<b>0.31</b> [0.17 0.85]	<b>5.0</b> [1.7 13]
$x_{\min} \leftrightarrow x_{\max}$	70 $\leftrightarrow$ 1000	15 $\leftrightarrow$ 70	0.1 $\leftrightarrow$ 15	0.001 $\leftrightarrow$ 0.8	0.01 $\leftrightarrow$ 100	0.01 $\leftrightarrow$ 100	0.01 $\leftrightarrow$ 100

**Table 5**  $M_d$ , following the two-pool assumption of ref. 5, modified form of model  $M_a$ . Global MCMC parameter estimation with three temperature data (sample concentration 15 mM). The optimal fit is displayed in Fig. 5

$T$ (°C)	$R$ (Å)	$\tau_{w,1}$ (μs)	$\tau_{c,2}$ (ns)	$\alpha$ (s <sup>-1</sup> )	$S_1 f_{\beta 1}$ (10 <sup>-3</sup> )	$S_2 f_{\beta 2}$ (10 <sup>-3</sup> )
10	<b>49</b> [48 50]	<b>8</b> [2.1 10]	<b>11</b> [9 16]	<b>0.64</b> [0.61 0.67]	<b>0.24</b> [0.23 0.26]	<b>1.7</b> [1.3 2.0]
15	—	—	<b>10</b> [8 14]	<b>0.58</b> [0.55 0.60]	—	—
20	—	—	<b>9</b> [7 12]	<b>0.54</b> [0.51 0.56]	—	—
$x_{\min} \leftrightarrow x_{\max}$	10 $\leftrightarrow$ 100	0.001 $\leftrightarrow$ 10	0.1 $\leftrightarrow$ 100	0.001 $\leftrightarrow$ 0.8	0.0 $\leftrightarrow$ 0.1	0.0 $\leftrightarrow$ 0.1



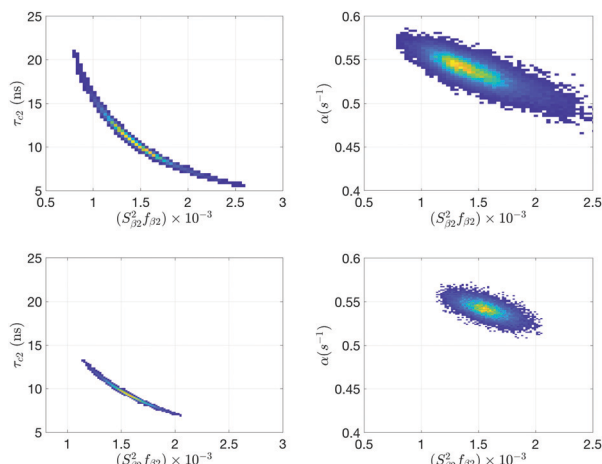


Fig. 4 Normalized histograms illustrating the main correlations for model  $M_a$  (upper row) and model  $M_d$  (lower row),  $(\tau_{c,2}-S_2f_{\beta 2})$  and  $(\alpha-S_2f_{\beta 2})$  in the left and right panel respectively. These are parameters for  $T = 20^\circ\text{C}$  NMRD-profile.

a poor description of high field data (see Fig. 5), however, the model is simpler with a total of 9 adjustable parameters. The least sampled parameter is  $(S_1^2f_{\beta 1})$  with 6700 correlation times within the MCMC run and auto-correlation (for each parameter in MCMC run) is in the same regime as cross-correlations. The correlation time  $\tau_{c,1}$  is about 20% shorter than for model  $M_a$  and corresponds to a micelle radius about 3 Å smaller with the correlation time interpreted in terms of SED.

For model  $M_c$ , the parametrization is provided in Table 4 and the best fit in Fig. 5 ( $\chi^2_{\text{opt}} = 18$ ) where three water-pools are assumed, is also the best of all models tested. However, it is noted that this  $\chi^2_{\text{opt}}$  is more than a factor of three lower than the number of experimental data-points, indicating an overfit of data. The parameter boundaries ( $\tau_{c,2}$  and  $\tau_{c,3}$ ) are chosen to be relatively narrow (based on the  $M_a$  result), in order for the study to be reproducible from random initial parameter configurations. The model  $M_c$  has a similar outcome to  $M_a$ , where  $\tau_{c,1}$  is longer, however, considering the 95%-confidence intervals the two models provide essentially the same outcome. The confidence intervals are much larger for  $M_c$ , especially for the high-field contribution  $\alpha$  and amplitude  $(S_3^2f_{\beta 3})$  as well as correlation times  $\tau_{c,2}$  and  $\tau_{c,3}$ . The correlation between several parameter pairs is substantial, in particular  $\tau_{c,i}$  and  $\alpha$ , and this is exemplified with histograms in the ESI.†

Based on what is learned from MCMC study of model  $M_a$  we propose an alternative in  $M_d$ , to explore a simpler model. First, from Table 2 the confidence intervals for  $(S_2^2f_{\beta 2})$  are overlapping and the parameter correlations suggest that one temperature independent parameter can be more useful for  $(S_i^2f_{\beta i})$ . Secondly, discussion of  $\tau_{c,1}$  is in terms of SED relation and can be incorporated explicitly and with a temperature independent micelle radius ( $R$ ) and water residence time ( $\tau_{w,1}$ ). In total 10 instead of 15 adjustable parameters in  $M_d$ .

Clearly this alters to some extent the physics of the model and must be kept in mind, for instance, a model that explains additional experimental data well is desired. The model  $M_d$  is

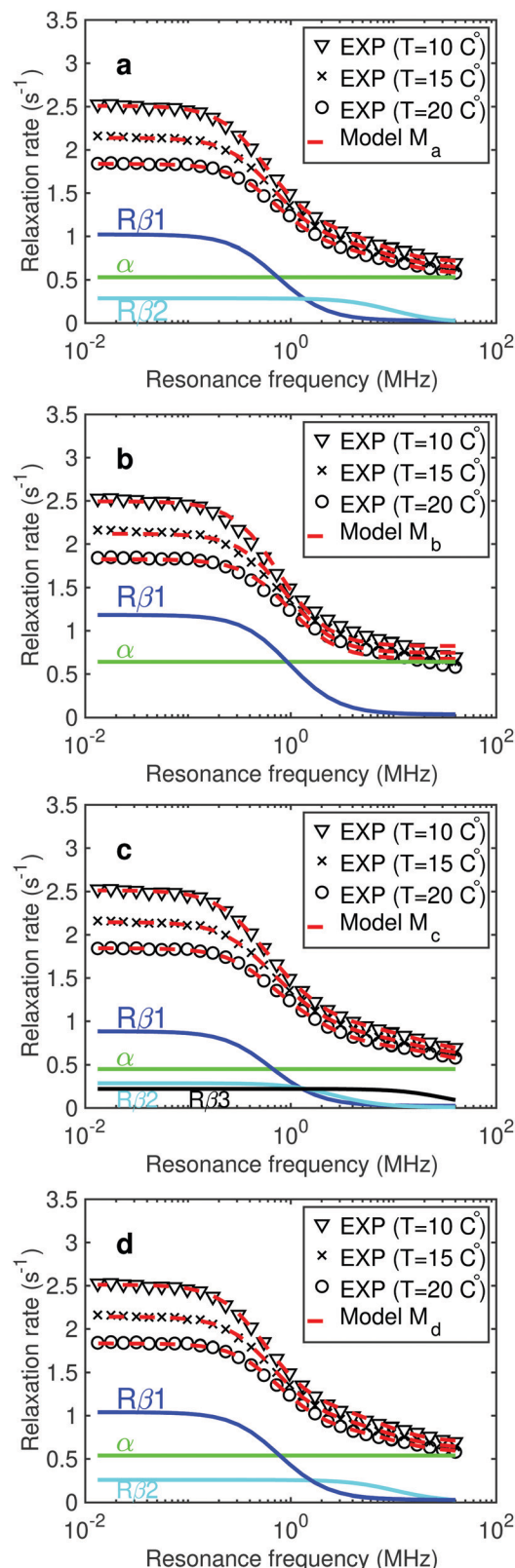


Fig. 5 The global fit with different temperatures (10, 15 and  $20^\circ\text{C}$ ) displayed with symbols  $\nabla$ ,  $\times$  and  $\circ$  respectively; the surfactant concentration is 15 mM and panels a–d display optimal fit of  $M_a \cdots M_d$ . The model contributions are exemplified at  $20^\circ\text{C}$  with  $\alpha$  as green and  $R_{i\beta 1}$ ,  $i = 1, 2, 3$  as blue, cyan and black respectively.



**Table 6** Model selection indicators for  $M_j$ ,  $j = a, b, c, d$ . Separate MCMC runs to compute error estimates. From left: number of water-pools, adjustable parameters ( $k_j$ ), optimal and mean  $\chi^2$ , Ockhams entropy (cf. eqn (15)) the Bayesian evidence, Bayesian factor, and the criteria AIC and BIC

$M_j$	#water	$k_j$	$\chi^2_{\text{opt}}$	$\Delta S_{j,\text{Ockham}}$	$-2 \ln[p(R^{\text{exp}} M_j, I)]$	$2 \ln[B_{aj}]$	AIC	BIC
$M_a$	2	15	$68.0 \pm 0.06$	$462 \pm 1.3$	$530 \pm 1.3$	—	$98.0 \pm 0.06$	$203.0 \pm 0.06$
$M_b$	1	9	$1346 \pm 0.04$	$248 \pm 1.0$	$1594 \pm 1.0$	$1064 \pm 2$	$1264 \pm 0.04$	$1427 \pm 0.04$
$M_c$	3	21	$18.0 \pm 0.4$	$692 \pm 40$	$710 \pm 40$	$180 \pm 40$	$60 \pm 0.4$	$207 \pm 0.4$
$M_d$	2	10	$75.3 \pm 0.04$	$246 \pm 5$	$321 \pm 5$	$-209 \pm 5$	$97.3 \pm 0.04$	$174.3 \pm 0.04$

first used to illustrate the computed Bayesian-evidence and it is left for future work to derive a less complex model with the relevant temperature dependence of  $(S_i^2 f_{\beta i})$ .<sup>5</sup>

The parametrization of model  $M_d$  is given in Table 5, with best fit in Fig. 5 ( $\chi^2_{\text{opt}} = 75$ ). Estimated micelle radius from the SED relation explicitly included in the model is 49 Å and is similar to what is obtained from  $M_a$  assuming  $\tau_c = \tau_R$ . The water residence time for pool-1 ( $\beta_1$ ) is larger than  $\tau_R$  ( $\tau_{w,1} \sim 10 \mu\text{s}$ ) with a large confidence interval. The amplitudes ( $S_i^2 f_{\beta i}$ ) are similar to model  $M_a$ . However, in  $M_d$  parameter cross-correlations (see histogram lower panels Fig. 4) are, together with parameter confidence intervals, reduced compared to model  $M_a$ .

## 5.2 Bayes evidence

Based on the fits in Fig. 5 and the  $\chi^2_{\text{opt}}$  values it is difficult to decide how to proceed. Clearly model  $M_b$  has a poor fit at higher frequency but still the parameters are the fewest and make sense, hence, according to Ockham's razor<sup>19</sup> this could be the best choice. Or it could be convincing with a good fit (model  $M_c$ ), especially if the additional parameters are useful in a discussion of the physical chemistry of micelles.

At this stage the Bayes-evidence and corresponding factor (see eqn (12) and (16) respectively), provide valuable information. The ranking properties in relation to the original model  $M_a$  with  $\Delta S_{j,\text{Ockham}}$ ,  $\chi^2_{\text{opt}}$ ,  $\ln[B_{aj}]$  and  $-2 \ln[p(R^{\text{exp}}|M_j, I)]$  are provided in Table 6.

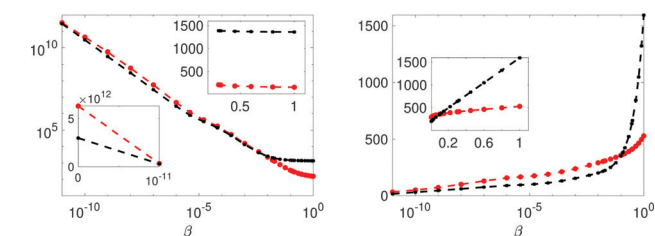
In Fig. 6 the Bayes-evidence for model  $M_a$  and  $M_b$ , the two and one water-pool models respectively, are shown with the left panel displaying the integrand of eqn (12) and the right the cumulative integral where  $\beta = 1$  is the sought value. For low- $\beta$  values it is noted that the  $M_b$  performs better, whereas at  $\beta$  close to 1 the model  $M_a$  is performing much better (see linear scale inserts in Fig. 6). At low  $\beta$  the likelihood function has almost no

contribution in eqn (11) and thus the integrand  $\langle \chi^2(\mathbf{x}) \rangle$  is expected to be small only in a few small volume segments, thus giving a smaller value due to the lower parameter volume of  $M_b$ . Considering the whole  $\beta$ -interval the large  $\chi^2$  for  $M_b$  (conventional poor fit) leads to a larger evidence integral for  $M_b$ . Thus, comparing the values  $(-2 \ln[p(R^{\text{exp}}|M_j, I)])$  in Table 6) the model  $M_a$  is significantly better and the evidence against  $M_b$  (single water-pool) is very strong on the Jeffery's scale. The latter is stated assuming that the prior probabilities for the models are equal (see eqn (16)).

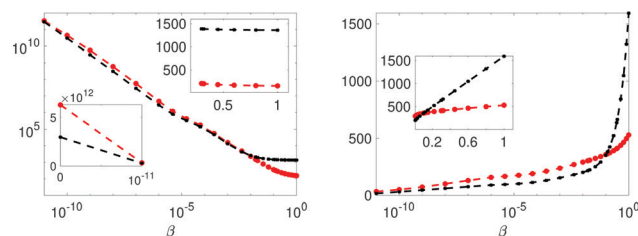
With three water-pools of model  $M_c$  the lowest  $\chi^2_{\text{opt}}$  (at 18) of the four models is reached. The computed evidence is provided in Fig. 7, where although a much better fit is achieved in terms of  $\chi^2_{\text{opt}}$ , the Bayes-factor dictates very strong evidence against  $M_c$  and the three water-pool model (following Jeffrey's scale, see Table 1). This has its origin in the increased Ockham-entropy of model  $M_c$ .

Finally, the evidence for alternative two-pool model  $M_d$  is provided in Fig. 8. Here an interesting observation is that a model with reduced Ockham-entropy ( $\Delta S_{j,\text{Ockham}}$ ) is favoured, in spite of its higher  $\chi^2(\mathbf{x}_{\text{opt}})$ . The boundaries of micelle radius are chosen such that the SED computed correlation time has similar range as in model  $M_a \dots M_c$ . However, noting that the amplitudes are assumed temperature independent makes the physics of the two models slightly different and care is needed not to over interpret the Bayes-factor, noting that a ranking assumes equal probability of the two models compared (see eqn (16)).

The approximate model selection criteria listed in Table 6 are the AIC and BIC. By comparing the magnitudes of  $-2 \ln[p(R^{\text{exp}}|M_j, I)]$  with AIC and BIC it is noted that the smallest AIC value is for  $M_c$  thus the wrong order in comparison with the complete integration. The BIC provides qualitatively the same order. Discrepancy is large for BIC compared to



**Fig. 6** The integrand and integral of eqn (12) in the left and right panels respectively, for models  $M_a$  (red) and  $M_b$  (black). Insets display the linear scale. The error bars are 3 standard errors. The  $\beta = 1$  value in the left panel is the sought evidence:  $-2 \ln[p(R^{\text{exp}}|M_j, I)]$ .



**Fig. 7** The integrand and integral of eqn (12) in the left and right panels respectively, for models  $M_a$  (red) and  $M_c$  (black). Insets display the linear scale. The error bars are 3 standard errors. The  $\beta = 1$  value in the left panel is the sought evidence:  $-2 \ln[p(R^{\text{exp}}|M_j, I)]$ .



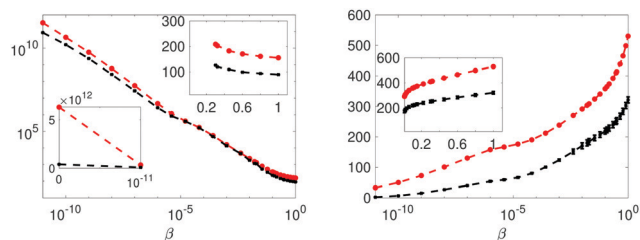


Fig. 8 The integrand and integral of eqn (12) in the left and right panels respectively, for models  $M_a$  (red) and  $M_d$  (black). Insets display the linear scale. The error bars are 3 standard errors. The  $\beta = 1$  value in the left panel is the sought evidence:  $-2 \ln[p(R^{\text{exp}}|M_j, I)]$ .

$-2 \ln[p(R^{\text{exp}}|M_j, I)]$ , larger than a factor two (on log-scale) for  $M_a$  and  $M_c$ .

The Ockham-entropy (see  $\Delta S_{j, \text{Ockham}}$  in Table 6) is lowest for  $M_d$  and  $M_b$ , where the number of parameters is also low. This property reflects the uncertainty in going from prior assumptions of parameters to the final parametrization and can be expected to increase for larger parameter volume.

Clearly we do not know *a priori* how large MCMC simulations are required, making researchers sometimes stay away from this technique. However, provided is a set of tools to monitor convergence and reproducibility and from a moderate computational investment it is clear how challenging (or easy) problems lie ahead. In particular with the option to follow annealing the experience so far is that MCMC can do well in parametrization of NMR-relaxation, relaxometry and line shape studies,<sup>6,7,53,55</sup> and should play an important role in Bayesian evidence calculation.

### 5.3 Model comparison summary

The components to pay attention to for useful Bayesian model comparison are summarized:

- Accurate estimation of experimental variances (see eqn (16)) plays a central role for the outcome model comparison. It is noted that the variance obtained by conventional fit to determine  $1/T_1$  (like Levenberg–Marquardt algorithm), systematically underestimate the variance. To circumvent this, the route<sup>5</sup> of calibrating or obtaining estimated variance with multiple experiments is used. An alternative route may be to estimate signal to noise in the measured intensities and apply estimation of  $1/T_1$  and its variance *via* a separate MCMC run following Fig. 3.

- Parameter prior is here assumed to be uniform, meaning the user sets the min and max of model parameters. The choice of uniform is suitable since parametrisation should be close to that achieved with conventional maximum likelihood methods, to ease the comparison.<sup>5,34</sup> Care is needed such that boundaries are set with equivalent strategy, noting that narrow interval around optimal minima may give unfair advantage with  $\Delta S_{\text{Ockham}}$  for that model. Also a verification that  $\chi^2$  is reproducible from multiple starting configurations (Fig. 3) is good to know before going further.

- The Bayesian evidence is accessible *via* MCMC (see Section 4.2), provided models have relatively low computational cost.

- Note that from a mathematical point of view the Bayesian evidence can be computed for arbitrary models since whole parameter space is integrated. This may allow for instance Lorentzian or non-Lorentzian character of spectral densities to be compared. It is noted, however, that for model comparison we have to assume equal *a priori* probabilities of the models that may not be true (see eqn (16)). Furthermore, extra attention is required for the motivation of parameter boundaries, not to favour one model more than the other.

## 6 Conclusions

In this work the methodology of Bayesian model comparison is provided and exemplified with nuclear magnetic relaxation dispersion (NMRD) data. The NMRD profile provides the otherwise difficult to measure slow dynamics in complex materials and biological systems. However, to reach conclusive results requires careful model development to capture the physics and chemistry of the studied system. Thus, Bayesian model comparison fills a gap with quantitative input on how many parameters and mechanisms are reasonable to include in a model based on the data we have. This is a practical question that should always occur when faced with data from a new system.

The computational approach follows a thermodynamic integration to compute the Bayesian model evidence from a set of MCMC simulations. The technique followed,<sup>41</sup> makes this challenging task feasible. Thus, in addition to quality of fit it is identified that the introduced definition of Ockham-entropy helps us understand the result. To work with entropy is suitable for the physical-chemistry community and valuable in the long-debated explanation of Bayesian evidence.<sup>18</sup> Ockham-entropy quantifies the uncertainty in going from prior assumptions, from the limited knowledge before study of the system to the final model parametrization. It is the Ockham-entropy that provides the penalty for over-use of parameters.

It was found that the evidence can be computed with only 10–30 relatively low cost MCMC simulations, in this work each MCMC was completed in 5–20 minutes, meaning that extensive model comparison can be performed with access to a workstation or single cluster-node.

The computed Bayesian evidence rules out the simplicity of one- and the uncertainty of three water-pools in favour of a two-pool model ( $M_a$ ) based on the current NMRD data. Such a conclusion can not be quantitatively drawn based on a conventional least squares fit (or MCMC parametrization). Furthermore, a two-pool model ( $M_d$ ) involving less uncertainty than ( $M_a$ ) is provided with improved Bayesian evidence. The proposed model  $M_d$  suggests a direction for future work on a suitable model that more accurately captures the relevant physics and chemistry. These models can incorporate more experimental data at higher field and another direction is to incorporate information from MD simulations and for instance resolve the order parameter and population.<sup>8</sup>



It is found that approximate tools to compare models can deviate from the complete integral (AIC), or qualitatively follow (BIC), however, the latter exhibit large discrepancy at logarithmic scale. These findings suggest that MCMC simulation of Bayesian evidence will play an important role in the development of NMRD theory; however, the method is completely general and can find ample use in other areas of physical chemistry as well.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

Prof. Per-Olof Westlund is acknowledged for valuable discussions initiating this work and for sharing NMRD experimental data. The COST Action CA15209 European Network on NMR Relaxometry is acknowledged and prof Pedro Sebastião for pointing out unclear aspects of the first draft. Financial support from the Kvantum institute (University of Oulu, Finland). The computational resources were provided by the CSC – IT Center for Science, Finland and the Finnish Grid and Cloud Infrastructure (FGCI) (urn:nbn:fi:research-infras-2016072533).

## Notes and references

- R. Kimmich and E. Anzardo, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2004, **44**, 257–320.
- J. P. Korb, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2018, **104**, 12–55.
- P. J. Sebastião, A. C. Ribeiro, H. T. Nguyen and F. Noack, *J. Phys. II*, 1995, **5**, 1707–1724.
- J. Kowalewski and L. Mäler, *Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications*, Taylor & Francis Group, New York, USA, 2006.
- P.-O. Westlund, *Univers. J. Chem.*, 2016, **4**, 69–73.
- M. Asadullah Javed, S. Ahola, P. Håkansson, O. Mankinen, M. Kamran Aslam, A. Filippov, F. Ullah Shah, S. Glavatskih, O. N. Antzutkin and V.-V. Telkki, *Chem. Commun.*, 2017, **53**, 11056–11059.
- P. Håkansson, M. A. Javed, S. Komulainen, L. Chen, D. Holden, T. Hasell, A. Cooper, P. Lantto and V. V. Telkki, *Phys. Chem. Chem. Phys.*, 2019, **21**, 24373–24382.
- P. Håkansson, T. Boirin and J. Vaara, *Langmuir*, 2018, **34**, 3755–3766.
- P. Håkansson and P.-O. Westlund, *Phys. Chem. Chem. Phys.*, 2004, **6**, 4321–4329.
- D. Kruk, J. Kowalewski and P.-O. Westlund, *J. Chem. Phys.*, 2004, **121**, 2215–2227.
- D. Kruk, P. Rochowski, M. Florek-Wojciechowska, P. José Sebastião, D. J. Lurie and L. M. Broche, *J. Magn. Reson.*, 2020, 106783.
- D. Kruk, A. Herrmann and E. A. Rössler, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2012, **63**, 33–64.
- P.-O. Westlund, G. Driver, Y. Wang, T. Sparrman, A. Laaksonen and Y. Huang, *Phys. Chem. Chem. Phys.*, 2017, **19**, 4975–4988.
- K. Åman, P. Håkansson and P.-O. Westlund, *Phys. Chem. Chem. Phys.*, 2005, **7**, 1394–1401.
- P. Håkansson and P. B. Nair, *Phys. Chem. Chem. Phys.*, 2011, **13**, 9578–9589.
- E. Belorizky, P. H. Fries, L. Helm, J. Kowalewski, D. Kruk, R. R. Sharp and P.-O. Westlund, *J. Chem. Phys.*, 2008, **128**, 052315.
- W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, 2nd edn, 1992.
- A. van der Linde, *Stat. Neerl.*, 2012, **66**, 253–271.
- R. Hoffmann, V. I. Minkin and B. K. Carpenter, *Hyle*, 1997, **3**, 3–28.
- I. Solomon, *Phys. Rev.*, 1955, **99**, 559–565.
- Y. Ayant, E. Belorizky, J. Alizon and J. Gallice, *J. Phys.*, 1975, **36**, 991–1004.
- L.-P. Hwang and J. H. Freed, *J. Chem. Phys.*, 1975, **63**, 4017–4025.
- B. Halle, V. P. Denisov and V. Kandadai, *Biological Magnetic Resonance*, Kluwer Academic/Plenum Publishers, New York, 1999, ch. Structure, vol. 17, pp. 419–484.
- A. Roch, R. N. Muller and P. Gillis, *J. Chem. Phys.*, 1999, **110**, 5403–5411.
- G. Voigt and R. Kimmich, *J. Magn. Reson.*, 1976, **24**, 149–154.
- D. Kruk, A. Kubica, W. Masierak, A. F. Privalov, M. Wojciechowski and W. Medycki, *Solid State Nucl. Magn. Reson.*, 2011, **40**, 114–120.
- T. J. Lored, in *Maximum Entropy and Bayesian Methods*, ed. P. F. Fougere, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990, pp. 81–142.
- R. Trotta, *Contemp. Phys.*, 2008, **49**, 71–104.
- J. Rantaharju and J. Vaara, *Phys. Rev. A*, 2016, **94**, 043413.
- P. Honegger, V. Overbeck, A. Strate, A. Appelhaugen, M. Sappl, E. Heid, C. Schröder, R. Ludwig and O. Steinhauser, *J. Chem. Phys. Lett.*, 2020, **11**, 2165–2170.
- J. P. Korb, P. J. McDonald, L. Monteilhet, A. G. Kalinichev and R. J. Kirkpatrick, *Cem. Concr. Res.*, 2007, **37**, 348–350.
- P. Håkansson, L. Persson and P.-O. Westlund, *J. Chem. Phys.*, 2002, **117**, 8634–8643.
- V. Wong and D. A. Case, *Annual Reports in Computational Chemistry*, Elsevier Masson SAS, 2008, vol. 4, pp. 139–154.
- P. J. Sebastião, *Eur. J. Phys.*, 2014, **35**, 1–15.
- R. Andrae, T. Schulze-Hartung and P. Melchior, 2010, arXiv, 1012.3754v, pp. 1–12.
- R. E. Kass and A. E. Raftery, *J. Am. Stat. Assoc.*, 1995, **90**, 773–795.
- N. Schenker, *J. Am. Stat. Assoc.*, 1985, **80**, 360–361.
- M. A. Newton and A. E. Raftery, *J. R. Stat. Soc. B*, 1994, **56**, 3–48.
- D. J. C. MacKay, *Neural Comput.*, 1992, **4**, 415–447.
- U. von Toussaint, *Rev. Mod. Phys.*, 2011, **83**, 943–999.
- W. Von Der Linden, R. Preuss and V. Dose, in *Maximum Entropy and Bayesian Methods*, ed. R. Fischer, Kluwer Academic Publishers, Dordrecht, 1999, pp. 285–292.



- 42 B. Halle and H. Wennerström, *J. Chem. Phys.*, 1981, **75**, 1928–1943.
- 43 T. Sparrman and P.-O. Westlund, *J. Phys. Chem. B*, 2001, **105**, 12524–12528.
- 44 K. Åman, E. Lindahl, O. Edholm, P. Håkansson and P.-O. Westlund, *Biophys. J.*, 2003, **84**, 102–115.
- 45 R. C. Tolman, *The Principles of Statistical Mechanics*, Dover Publications, New York, 1979.
- 46 P. Håkansson, BAYES-SELECT (GitHub v1.1), 2020, DOI: 10.5281/zenodo.4012179.
- 47 P. Håkansson and M. Mella, *J. Chem. Phys.*, 2008, **129**, 124101.
- 48 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1092.
- 49 M. A. Tanner, *Tools for statistical inference*, Springer-Verlag, New York, 3rd edn, 1996.
- 50 N. van Kampen, in *Stochastic Processes in Physics and Chemistry*, ed. N. van Kampen, Elsevier, Amsterdam, The Netherlands, 3rd edn, 2007, ch. 1, pp. 1–29.
- 51 J. W. Eaton, D. Bateman, S. Hauberg and R. Wehbring, GNU Octave version 5.1.0 manual: a high-level interactive language for numerical computations, 2019, <http://www.gnu.org/software/octave/>.
- 52 D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, *Publ. Astron. Soc. Pac.*, 2013, **125**, 306–312.
- 53 S. F. Cousin, P. Kadeřávek, N. Bolik-Coulon, Y. Gu, C. Charlier, L. Carlier, L. Bruschweiler-Li, T. Marquardsen, J. M. Tyburn, R. Bruschweiler and F. Ferrage, *J. Am. Chem. Soc.*, 2018, **140**, 13456–13465.
- 54 B. Halle, *Magn. Reson. Med.*, 2006, **56**, 60–72.
- 55 N. N. Kuzma, P. Håkansson, P. Pourfathi, R. K. Ghosh, H. Kara, S. J. Kadlecek, G. Pileio, M. H. Levitt and R. R. Rizi, *J. Magn. Reson.*, 2013, **234C**, 90–94.

