In this work a methodology to perform Bayesian model-comparison is developed and exemplified in the analysis of magnetic relaxation dispersion (NMRD) experiments of water in Ganglioside micelle system. The NMRD powerful probe of slow dynamics in complex liquids is obtained. There are many interesting systems to study with NMRD, such as ionic and Lyotropic liquids or electrolytes. However, to progress in the understanding of the physical chemistry of 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 sites, and rules out lower and higher complexity level, i.e., one and three sites. In contrast, and provided only with the quality of best fit, suggest a three site model. The two approximate selection tools, Akaike and Baysian information criterions, may lead to wrong conclusions compared to the the full integration. The methodology is expected to be one of several important tools in NMRD model development, however, is completely general and should find awakened use in many research areas.
Relaxometry models compared with Bayesian technique: Ganglioside micelle example†

Pår Håkansson,*a

In this work a methodology to perform Bayesian model-comparison is developed and exemplified in the analysis of magnetic relaxation dispersion (NMRD) experiments of water in Ganglioside micelle system. The NMRD powerful probe of slow dynamics in complex liquids is obtained. There are many interesting systems to study with NMRD, such as ionic and Lyotropic liquids or electrolytes. However, to progress in the understanding of the physical chemistry of 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 sites, and rules out lower and higher complexity level, i.e., one and three sites. In contrast, and provided only with the quality of best fit, suggest a three site model. The two approximate selection tools, Akaike and Baysian information criterions, may lead to wrong conclusions compared to the full integration. The methodology is expected to be one of several important tools in NMRD model development, however, is completely general and should find awakened use in many research areas.

1 Introduction

The measurement of relaxation rates of nuclear spins versus resonance frequency is typically denoted relaxometry and the recorded data forms a relaxation dispersion (NMRD) profile. The experiment typically cover kHz to MHz resonance frequency adds additional information to NMR spectroscopy (denotes NMR relaxation). In particular relaxometry can characterise the dynamics in a wide range from nano- to micro-second regime.[1] These properties opens for study of complex systems such as cement, petroleum fluids, biological tissue and collective motion of whole protein,[2] dynamics in liquid crystals.[3]

To obtain more meaningful information from relaxation experiments, relaxation models are utilized to provide details of the mechanisms underlying spin relaxation.[4] Starting from the spin-system under study relevant mechanism to include may involve dipole-dipole, chemical exchange, paramagnetic and quadrupole contributions.[5][6]

The challenges in building models involves incorporation of consistent molecular dynamics processes to explain experimental observations and account for spin-dynamics at sufficient accuracy.[7][11] Hence, these challenges may come in the form of (i) multiple physical interpretations, (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 have progress in model development for challenging systems.

The 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" need to include more than minima of cost function like $\chi^2$.[16] Rather, it should focus on reproducibility and a balance quality of fit versus complexity of the model.[17]

Model selection is expected to help navigate in the physics and chemistry underpinning the relaxation models. This goes beyond qualitative guide in model selection via the Ockham’s razor, that may be formulated as “it is futile to do with more what can be done with fewer”.[18]

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,[19] intermolecular DD modulated by molecular diffusion,[20][21] and/or proton ex-
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. Ideally a signature NMRD feature can be used and strengthen the model interpretation with fewer parameters such as DD-relaxation enhancement. Such pronounced features are explored for paramagnetic super-paramagnetic and quadrupole spin-system.

In this work the methodology is provided for Bayesian model selection and tested on micelle NMRD data. In particular the example (ii) above, i.e., the communication of result, why does the two water-site model give the best explanation of data and not a simpler or more complex model. An alternative hypothetical example, why did we model anisotropic rotational diffusion in spite this not being the dominating contribution? Hence, does the data support this more complex model?

The communication can thus be strengthened, especially towards researchers that are not experts in the field. Other research-fields such as astrophysics and cosmology has a long tradition of using and benefiting from these methods.

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, rotation and translation of small molecules water at bilayer interface, water in cement and ionic liquid lyotropic phases porous organic cages and proteins. The simulation studies open for exploration of the validity of analytical models and for some systems serves as the method of choice. However, model comparison 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 method are sufficiently accurate. 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 the $\chi^2$:

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

where $R_i^T(x)$ is the theoretical model, dependent on adjustable parameters $x$, $R_i^{exp}$ are the experimental observations and $\sigma_i$ the standard deviation in experimental observation $i$. The adjustable parameters of the model ($x$) is of size $k$, i.e., correlation times, order parameters 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 relaxometry data.

There are large set of tools proposed around cost functions to compare models, the reduced $\chi^2$, can be computed by dividing by $N$ minus the model degrees of freedom 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 $k$ are the number of parameters is not robust. Hence, $\chi^2$-value does not allow for model comparison in general.

Two approximate routes: Akaike information criterion (AIC) and the Bayesian information criterion (BIC), both attempting 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. In current work these are tested and to have limited value in NMRD micelle study.

The cross-validation (analysing part of the data) or more generally the Bootstrap-sampling have the drawbacks with validity in the asymptotic limit and for model comparison, only approximate form are provided. Given the common situation of small data set ($N$) in NMR-relaxation and NMRD the bootstrap method will not be explored in current work.

In this work we follow instead Bayesian theory and compute the model evidence, that balance model-complexity, that will be discussed further, and the quality of fit. Thus, a quantitative tool corresponding to Ockham’s razor is obtained.

The 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 method (MCMC). In this work we implement the thermodynamic integration and show that this is a computationally feasible route for relaxometry model comparison. The result is quantitative discrimination between models, finding suitable complexity of micelle water-sites, a property that is qualitatively discussed previously.

The structure of the paper present the relaxation model derived in previous work in section 2. In section 3, the relevant Bayesian theory is provided with discussion of model complexity. This is followed by aspects to consider in MCMC simulations and the code provided 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 Re and refer to this work for further model discussions. The experimental observation of longitudinal relaxation rate $R_1^{exp}(\omega_0)$ 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.

The key to the relaxation study is the formulation of a relevant physical picture, i.e. the two site water model depicted in figure[1]. The two types of waters are denoted ($\beta_i$, $i=1,2$) both located at the micelle. Reasons for more than one site may be due to the large Ganglioside head-group allowing for several water locations, leading to different exchange times ($\tau_{w,i}$) and populations ($f_{w,i}$), but modulation by the same overall micelle rotation $\tau_R$. The non-bounded waters are denoted bulk water in figure[1].

In addition there are 2-3% Ganglioside protons that are not addressed with the model.
The theoretical prediction \( R_T^2(\omega) \) takes the form
\[
R_T^2(\omega) = \alpha + f_{\beta_1} R_{1,\beta_1}(\omega) + f_{\beta_2} R_{1,\beta_2}(\omega)
\]
where \( \alpha \) are the sum of field independent contributions. The order parameters originates from expressing spectral densities within a two step model
\[
J_i(n\omega) = (1 - S_i^2)\tau_i + S_i^2\tau_{ci,i} + \frac{1}{1 + (\tau_i\omega)^2},
\]
with \( \tau_i \) denotes fast local dynamics not explicitly accounted for in final model, but enters the parameters \( \alpha \). The field dependent parts for the two waters are given by
\[
R_{1,\beta_i} = S_i^2\tau_{ci,i}\beta_{\text{intra}} \left\{ \frac{0.2}{1 + (\omega\tau_{ci,i})^2} + \frac{0.8}{1 + (\omega\tau_{ci,i})^2} \right\} + W_D(0.1 + \frac{0.3}{1 + (\omega\tau_{ci,i})^2} + \frac{0.6}{1 + (\omega\tau_{ci,i})^2}),
\]
where \( W_D \) denotes weight factor for inter-molecular proton interaction (second part of eq. 4). \( \tau_{ci,i} \) 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:
\[
\tau_{ci,i} = \frac{\tau_R + \tau_{wi,i}}{\tau_R \tau_{wi,i}}.
\]

### 2.1 PROPOSED MODELS
The intra-molecular DD-amplitude may be determined from
\[
\beta_{\text{intra}} = \left( \frac{3}{2} \frac{\mu_0}{\pi \hbar} \frac{\tau_{HH}^2}{r_{HH}^{-3}} \right)^2,
\]
with proton distance in water \( r_{HH} = 1.58 \) Å and vacuum permeability \( \mu_0 \). Planks constant per radians (h), gyromagnetic ratio of a proton (\( \gamma_H \)). However, the amplitude for inter-molecular proton interactions is not known and takes the effective form \( \beta_{\text{water}} = \beta_{\text{intra}} W_D \). In this format we can not resolve if intra or inter are the dominating relaxation pathways and 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 two combined parameters \( S_i^2 f_{\beta_i} \) are estimated, together with \( \tau_{ci,i} \) and \( \alpha \). Thus, two water site model contain five adjustable parameters per temperature \( x = \{\alpha, S_i^2 f_{\beta_i}, \tau_{ci,i}\}, (i=1, 2) \) and 15 in total.

We can not tell in advance how many water sites there actually are, and more importantly, how many the NMRD experiments are sensitive to. This opens for a qualitative discussion a long the line of Occham’s razor to motivate a suitable number of water sites, two in the original work. The number of water sites, however, forms a transparent example where quantitative model comparison may be followed. Thus we consider the models \( M_0, M_5 \) and \( M_6 \) with two, one and three number of water sites respectively, leading to number of adjustable parameters as: \( \dim(x_0) = 15, \dim(x_5) = 9 \) and \( \dim(x_6) = 21 \). In addition the model \( M_7 \), a two-site model but exploring a reduced set of parameters \( (\dim(x_7) = 10) \).

As pointed out[5] modelling can be refined if we have complementary heavy water experiments. Furthermore, classical molecular dynamics may propose individual values to the merged \( S_i^2 f_{\beta_i} \) parameter and quantum chemical molecular dynamics may suggest constraints on proton exchange times. High field T1 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.)[22] the probability takes the form
\[
P(M_j| R^{\text{exp}}, I) = \frac{P(M_j| I)p(R^{\text{exp}}| M_j, I)}{Z},
\]
where 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 contain 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[20]
\[
p(R^{\text{exp}}| M_j, I) = \int_{V(x_j)} L(R^{\text{exp}}| M_j, x_j)p(x_j|M_j)d\mu(x_j),
\]
and is thus an integral of likelihood function over the whole multidimensional parameter volume \( V(x_j) \) where parameters are considered random variables with probability measure \( \mu(x_j) \). Hence, to get the probability of model \( M_j \) we need to integrate the likelihood over the whole weighted parameter space.

#### 3.1 Model complexity
It is worthwhile to look at one of the asymptotic approximations[20] of eq. 5 \( p(R^{\text{exp}}| M_j) \approx L(R^{\text{exp}}| M_j, x_j|_{\text{opt}})(4\pi\hbar^2/|x_{\text{max}} - x_{\text{min}}|) \), containing a single likelihood maxima \( L \), with the ratio of the two parameter volumes, "the (small) likelihood" and "parameter boundary" volumes of dimension \( k \). Hence, the Bayesian
evidence balance quality of fit $L$ penalised by model complexity, here given by the number of parameters $k$ and the volume-ratio they span. It is important to note that the boundaries defined by the researcher (priors) have influence on the evidence. 

If we introduce more parameter in model, the fit will improve but at some point the model complexity dominates. Hence, with additional information, that in this work could be obtained from surfactant relaxation study or a molecular dynamics simulation, a reduction of some correlation times interval $[x_{\text{max}} - x_{\text{min}}]$ is motivated and there is an increase in the evidence, or a lowering of the negative logarithm, provided the fit is still good. 

In eq. $\ref{eq:8}$ the complexity involves also dependence between parameters and transition between multiple maxima, and thus includes more than the volume ratio. For instance, rotational correlation times may be expected to have inverse temperature dependence and by incorporating this assumption in the model the complexity may go down. An aspect explored with the proposed models.

### 3.2 Thermodynamic integration

For many problems, the eq. $\ref{eq:8}$ is a high-dimensional integral and is difficult to estimate with numerical methods.$^{17,38}$ A flavour of the numerical challenge is given in Figure 2, where the details of the structure of the likelihood function needs to be integrated.$^{17}$

By introducing a one-dimensional thermodynamic integral instead of attempting direct computation of eq. $\ref{eq:8}$ a much less demanding problem is obtained. In fact, as exemplified this may reduce the required number of sampling configurations from $10^3$ to $10^0$! To compute eq. $\ref{eq:8}$ as a thermodynamic integral (without introducing approximations) an auxiliary function is introduced function$^{17,49}$

$$F(\beta) = \int_{\gamma(x_j)} L(R^{\exp}|M_j, x_j)^\beta p(x_j|M_j) d\mu(x_j), \quad (9)$$

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

$$\frac{\partial}{\partial \beta} \ln[F(\beta)] = -\frac{1}{F(\beta)} \int_{\gamma(x_j)} \frac{\chi(x_j)^2}{2} L(R^{\exp}|M_j, x_j)^\beta p(x_j|M_j, I) d\mu(x_j),$$

$$= -\frac{1}{2} \langle \chi(x_j)^2 \rangle_\beta \approx -\frac{1}{2N_{\text{MCMC}}} \sum_i \chi(x_i(\beta))^2, \quad (10)$$

where the specific likelihood function in this work ($L(x_j)^\beta \propto \exp(-\beta \chi(x_j)^2/2)$) is used and the mean value $\langle \cdot \rangle_\beta$ is computed for the $\beta$-probability density $L(R^{\exp}|M_j, x_j)^\beta p(x_j|M_j, I)d\mu(x_j)/F(\beta)$. The approximation involve the statistical uncertainty from finite number of MCMC configurations, described in section 4.

Integrating eq. $\ref{eq:10}$ over $[0, 1]$ gives

$$\ln[F(1)] = \ln[F(0)] - \int_0^1 \frac{1}{2} \chi(x_j)^2 \beta d\beta = \ln[p(R^{\exp}|M_j, I)], \quad (11)$$

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

### 3.3 Bayes factor

With two models proposed ($M_a, M_b$) the eq.$\ref{eq:11}$ enables calculation of Bayes factor$^{50}$

$$\frac{P(M_a|R^{\exp}, I)}{P(M_b|R^{\exp}, I)} = \frac{P(R^{\exp}|M_a, I)}{P(R^{\exp}|M_b, I)} \times \frac{P(M_a|I)}{P(M_b|I)}, \quad (12)$$

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 empirical Jeffreys scale for the strength of evidence against model $b$, given in Table 1. This is an empirically calibrated scale.

| Evidence against $M_b$ | Bayes factor $B_{ab}$ |
|------------------------|-----------------------|
| 2 to 6                 | Positive              |
| 6 to 10                | Strong                |
| >10                    | Very strong           |

### 3.4 Approximate model selection criteria

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

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

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

where $L_{\text{opt}}$ is the maximum likelihood$^{17}$ for the case of gaussian likelihood used in this work (see eq. $\ref{eq:10}$) 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 eq. $\ref{eq:11}$) BIC and AIC penalise a model with poor fit (large $\chi^2$) and excessive adjustable parameters, with expected $O(N^{-1})$ error, of experimental data $N$.

### 4 Methods

We present the parameter estimation with MCMC followed by the procedure for thermodynamic integral (eq. $\ref{eq:11}$), for details we refer to the Matlab code and “readme” file presented in the GitHub example.$^{23}$ It is noted that central to the study is the details of $\chi^2$.
4.1 MCMC parameter estimation

Before any work with model comparison the parametrisation is needed. The ideas are implemented in Matlab/octave code\textsuperscript{42} and the steps given in flow-chart of parameter estimation (see figure 3). The start \textbf{(a)} for one of the models $M_1, \ldots, M_d$, the parameter prior ($p(x_j | M_j, I)$) is chosen, e.g., in this work is the uniform (uninformed) prior. For the user this involve setting the $x_{\min}$ and $x_{\max}$ boundaries (given under “range” in tables 2-4). The parameter prior is used in \textbf{(b)} to generating $N_{\text{test}}$ starting configurations (in the range 4-10 $\{x\}_{N_{\text{test}}}$), where these parameter sets are random variables from the sought prior distribution.

Core of the MCMC is efficient sampling of configurations \textbf{(c)}. For ease of introducing alternative prior distributions the current version, function MCMCstep.m\textsuperscript{42} follow Metropolis algorithm\textsuperscript{44} hence, the proposed new parameter configurations are sampled at random direction (not for instance guided by likelihood function). Each parameter is given a move individually (hence as a component of Gibbs sampling\textsuperscript{42}). The parameter moves are reflected at the boundaries ($x_{\min}$ and $x_{\max}$).

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 are require model specific implementation\textsuperscript{42}. 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\textsuperscript{48} (see SEI and implementation in the code\textsuperscript{42}).

The calibration of step length followed by performing $\sim 10^5$ MCMC iterations towards equilibrium (region of smallest $\chi^2$) is denoted “burn in”. The burn in is followed directly by attempt of productive MCMC-steps ($\sim 10^6$) and this procedure are typically not intervened by the user. At this point comes the main question of MCMC, have configurations reached equilibrium? This question is hard to address from a single configuration, but for a set of $N_{\text{test}}$ looking for consistency in $\chi^2_{\text{opt}}$ versus the spread (standard deviation of $\chi^2$) are examined, see SEI. If some configurations remain in a local minima, proceed with annealing MCMC, thus corresponding to simulation corresponding to higher “temperature” (low $\beta$) where longer steps are made and local minima more easily surpassed.

Obviously there are no global minimisation routine (MCMC or otherwise) capable of arbitrary complex model, thus if some very challenging problem occur, the user should also consider if actually relevant physics is sought.

With all the $\{x\}_{N_{\text{test}}}$ configuration successfully completed \textbf{(d)} an additional set is typically computed up to and including the burn-in step, to verify that we have a reproducible global minima in $\chi^2$. This follows by $x_j$-histogram, i.e. total density projected on variable $x_j$ that is integrated to give 95%-confidence intervals. The optimal parameters are in this work given by

$$x_{\text{opt}} = \arg\min_x (\chi^2),$$

reported in combination with 95%-confidence interval numerically integrated from MCMC histogram data. For the considered models one calculation of MCMC take 1-3 hours on MacBook Pro 3.1 GHz intel processor. The implementation\textsuperscript{42} in Matlab R2019b has one option to submit multiple runs to supercomputer (see next section). The same code also runs with freely available Octave\textsuperscript{49}.

What is found valuable in the provided implementation\textsuperscript{42} is the ease by which calibration of MCMC step-length is made automatic with the univariate moves (Gibbs sampling\textsuperscript{42}). In addition, how random initial configurations converge by exploring a set of fixed annealing temperatures for the challenging cases. These cases may in particular appear for complex models that still are interesting to give a fair trial. The annealing enables completely random initial configurations to query if proposed model result is reproducible within the given parameter boundaries.

There are several MCMC programs and it is found that in Ref\textsuperscript{50} (used in NMRD\textsuperscript{51}) a type of configuration move is used that is not transparent to implementation of annealing. Furthermore, initial configurations are proposed to be from a suitable staring point\textsuperscript{50} thus, the correlated set of configurations does unlikely explore the whole prior parameter volume as is done with annealing and random initial configurations.
4.2 Numerical thermodynamic integral

The numerical computation of thermodynamic integral in eq. 11 follow by taking the set of optimised configurations \( \{x_{\text{opt}}\} \) following from MCMC-parameter optimisation in Figure 3 and collecting data at discrete \( \beta \) values. The discretisation 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 trapezoidal rule, the interval with largest contribution is located and split in half and integral is recomputed. This procedure continue until 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.

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 Micelle, where previous work consider a model with two water sites, this number (and not one, three or more pools) are 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 he models \( (M_\alpha, M_b, M_c, \text{ and } M_d) \) following Figure 3 we can identify if any of the models provide exceptional challenges before continued work with model comparison. The first stage, the parametrization, with optimal fits (parameter inference) are provided in figure 5 and estimated parameters in table 3 providing the parameter boundaries in bottom row (i.e. the prior assumptions).

Common for models \( M_b \) to \( M_d \) are that optimisation is performed globally over the 90 data points and the exploration of weight of inter-DD contribution case \( D_W = 0.4 \). It was found that model dependence of \( D_W \) is weak and is not the focus of this work to explore further.

For model \( M_b \) it is found that MCMC estimation provide within 95%-confidence interval, \( X_{\text{conf}} = 35 \) and largely the same parameters as the original work with data in table 2. The exception is for \( (S_2/f_{b2}) \) that is slightly larger from present MCMC-analysis.

Thus, the same physical interpretation is summarised, that is, a slow correlation time \( \tau_{c \rightarrow 1} \) dominated by overall rotational diffusion of Ganglioside micelle, that from Stoke-Einstein-Debye (SED) relation is consistent with the radius of 54 Å, concluded from measurements by other techniques. The shorter correlation time, originating from \( \tau_{c \rightarrow 2} \) (see figure 4) is ascribed to the proton exchange-modulated reorientation where water protons exchange with labile surfactant protons.

The difference in the size of confidence intervals are striking, for MCMC they are always larger with the difference for \( (S_2^2/f_{b2}) \) is a factor 50. Considering steps to check the MCMC parameter estimation, there are 7 simulations starting from independent starting configurations within parameter boundaries (see range in table 2). By estimating the MCMC-parameter auto-correlation \( (\tau_{\text{auto}}) \) provides a handle how well the parameter space is sampled and it is found that \( \chi^2 \), \( (S_2^2/f_{b2}) \) and \( (S_1^2/f_{b1}) \) are sampled with 4·104, 2·105 and 1·103 correlation times respectively. This points to a sufficiently sampled parameter-space given that additional "burn-in" is separately performed (see SEI, figure S2).

From the computed cross-correlations (see SEI), a substantial amount is seen between \( (S_1^2/f_{b1})-(\tau_{c \rightarrow 2}) \) and \( (S_2^2/f_{b2})-\alpha \) for 20°C NMRD-profile and are illustrated with histograms in Figure 4. The parameter correlation explains the large error boundary found in MCMC estimation. Note that this in it self 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, parameter correlations introduce an error in uncertainty estimation from \( \chi^2 \)-values where parameter distribution is unknown and the parameters are assumed to be independent degrees of freedom of the model.

The parametrization of model \( M_b \) is given in table 3 and figure 5 \( (\chi^2_{\text{opt}} = 673) \) assumes one water pool. The model provide a poor description of high field data (see figure 5), however, complexity of model is less with total of 9 adjustable parameters. The least sampled parameter is \( (S_1^2/f_{b1}) \) 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 \( \tau_{c \rightarrow 1} \) is about 20% shorter than for model \( M_b \) 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 figure 5 \( (\chi^2_{\text{opt}} = 12) \) where three water pools are assumed, is also the best of all models tested. The parameter boundaries \( (\tau_{c \rightarrow 2} \text{ and } \tau_{c \rightarrow 3}) \) are chosen relatively narrow (based on the \( M_b \) result), in order for the study to be reproducible from random initial parameter configurations. The model \( M_c \) has similar
outcome as $M_d$, where $\tau_{w1}$ is longer (considering optimal fit values), however, considering the 95%-confidence intervals the two models provide essentially the same outcome. The confidence intervals are large for $M_3$, exceptionally for the high-field contribution $\alpha$ and amplitude ($S_2^2 f_{\beta 2}$) as well as correlation times $\tau_{c2}$ and $\tau_{c3}$. Correlation between several parameter pairs are substantial, in particular $\tau_{c1}$, $\alpha$ and ($S_2^2 f_{\beta 3}$) and this is exemplified with histograms in SEI.

Based on what is learned from MCMC study of model $M_d$ we propose an alternative in $M_{5d}$, to explore a simpler model. First, from table 2, the confidence intervals for ($S_2^2 f_{\beta 2}$) are overlapping and the parameter correlations, suggest that one temperature independent parameter can be more useful for ($S_2^2 f_{\beta 2}$). Secondly, discussion of $\tau_{w1}$ is in terms of SED relation and can be incorporated explicitly and with a temperature independent micelle radius ($R$) and water residence time ($\tau_{w1}$). In total 10 instead of 15 adjustable parameters in $M_d$.

Clearly this alter 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 first aimed 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 ($S_2^2 f_{\beta 1}$).

5.2 Bayes evidence

Based on the fits in figure 5 and the $\chi^2_{opt}$ values it is difficult do decide how to proceed. Clearly model $M_b$ has a poor fit at higher frequency but still parameters are the fewest and make sense, hence, according to Occam’s razor [18] this could be the choice? Or it could be convincing with a good fit (model $M_c$), especial 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 eqs 11 and 12 respectively), provide valuable information. The ranking properties in relation to the original model $M_c$ with $\chi^2_{opt}$, $\ln[Bayes]$ and $-\ln[p(R^{exp} | M_j, I)]$ are provided in table 5.

In figure 6 the Bayes-evidence for model $M_c$ and $M_b$ (two and one water pools respectively) are shown with the left panel displaying the integrand of eq. 11 and the right the cumulative integral where $\beta = 1$ is the sought value. From left panel it noted that for low-$\beta$ the $M_b$ performs better, whereas at $\beta$ close to 1 the model $M_c$ is performing better (see linear scale inserts in figure 5). It is noted that at low $\beta$ the Likelihood function has almost no contribution in eq 10 and thus the integral of $\chi^2(x)$ is expected to be small only in a few small volume elements obtain a smaller value due to the lower parameter volume of $M_c$. 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 ($-\ln[p(R^{exp} | M_j, I)]$ in table 5) the model $M_b$ is significantly better, evidence against $M_b$ (single water site) is very strong (with Jeffery’s scale). The later can be stated provided the prior probabilities for the models are equal (see eq. 12).

With three water pools of model $M_c$ the lowest $\chi^2_{opt}$ (at 12) of the four models is reached. The computed evidence is provided in figure 7 where although a much better fit is achieved in terms of $\chi^2_{opt}$, the Bayes-factor dictates very strong evidence against $M_c$ and the three water-pool model (following Jeffery’s scale, see table 1).
This has it’s origin in the increased complexity of model $M_d$.

Finally the evidence for alternative two-pool model $M_d$ is provided in figure 6 were although larger $\chi^2$ compared with $M_a$, the evidence supports model $M_d$. Hence, an interesting observation where a model with reduced complexity is favoured, in spite of the worse quality fit. However, noting that the amplitudes are assumed temperature independent make 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 eq. 11).

The approximate model selection criterions listed in table 6 are the AIC and BIC. By comparing the magnitudes of $-\ln[p(R^{opt}|M_f, I)]$ with AIC and BIC it is noted that smallest AIC value is for $M_d$, thus wrong order in comparison with the complete integration. The BIC provide qualitatively the same order and relatively close for $2\ln[b_{opt}]$ (533±1 for Bayesian and 585 for BIC). Discrepancy is large for $2\ln[b_{nc}]$, (234±9 for Bayesian and 30 for BIC).

If Bayesian evidence are relatively close (not the case in this study) the property $\langle \chi^2(x) \rangle - \chi^2_{opt}$ is proposed to be combined with Bayesian evidence for ranking. The property eliminates quality of fit and summarise model complexity as the effective number of parameters that in this work is found larger than the actual number (from table 5). This reflect how much the data can constrain versus model prediction. Hence, with additional parameter not constrained by the data, for example proposing multiple parameters instead of the single field independent $\alpha$, this property should increase.

The components for the presented Bayesian model comparison are, (i) accurately finding the experimental variance in $\chi^2$ (see eq. 1) here obtained from calibrating NMRD output with multiple experiments, (ii) parametrization of each model with MCMC (see figure 3) where the choice of likelihood function is Gaussian and parameter-priors are set to uniform type (in practise setting a max and min value of each adjustable parameter in the model, correlation time etc), (iii) computing 10-30 MCMC to obtain the integrand of thermodynamic integral. The stage (i) is important to address carefully, without correctly weighted experimental data model comparison is not useful. The choices in (ii) of likelihood and priors are useful to enable comparison with conventional maximum likelihood methods.

Clearly we do not know a priori how large MCMC simulations are required, making researchers sometimes stay away from this technique. However, provided are a set of tools to monitor convergence and reproducibly and from a moderate computational investment it is clear if it is worthwhile to complete the study. 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 and should in the future play an important role in Bayesian evidence calculation.

6 Conclusions

In this work the methodology of Bayesian model comparison are provided and exemplified with nuclear magnetic relaxation dispersion (NMRD) data. The NMRD profile provides otherwise difficult to probe slow dynamics in complex materials and biological systems. However, to reach conclusive results requires a 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. A practical question that should always occur when faced with data from a new system.

The computational approach follow a thermodynamic integration to compute the Bayesian model evidence from a set of MCMC simulations. The technique followed makes this challenging task feasible. It was found that this integral can be computed with only 10-30 relatively low cost MCMC simulations, in this work each completed at 1-2 hours, meaning that extensive model comparison can be performed with access to a workstation or cluster-node.

The computed Bayesian evidence rules out the simplicity of one- and the complexity of three-water cites in favour of a two-cite model ($M_a$) based on the current NMRD data. Such conclusion can not be quantitatively drawn based on a conventional least square fit (or MCMC parametrization). Furthermore, a two-site model ($M_d$) of less complexity than proposed is provided, interesting given the improved Bayesian evidence.

However, the represented physics in this model ($M_d$) with temperature independent amplitudes are different and illustrate a case where care must be held not to over-interpret the Bayesian evidence. The proposed model $M_d$ suggest a direction for future work on a suitable model that more accurately capture the relevant physics and chemistry. These models can can incorporate more experimental data at higher field and an other direction is to incorporate information from MD simulations and for-instance resolve the order parameter and population.

It is found that approximate tools to compare models can de-
violate from the complete integral (AIC), or qualitatively follow (BIC), however the later at large discrepancy at logarithmic scale. Suggesting that MCMC simulation of Bayesian evidence will play important role in development of NMRED theory, however, completely general and can find extensive use in other areas of physical chemistry as well.

**Conflicts of interest**

“There are no conflicts to declare”.

**Acknowledgements**

Please acknowledge prof Per-Olof Westlund for valuable discussions initiating this work and for sharing NMRED experimental data. The COST Action CA15209 European Network on NMR Relaxometry is acknowledged and prof Pedro Sebastião for pointing out unclear aspects of first draft. Financial support from Relaxometry is acknowledged and prof Pedro Sebastião for pointing out unclear aspects of 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**

1. R. Kimmich and E. Anoardo, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2004, **44**, 257–320.
2. J. P. Korb, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2018, **104**, 12–55.
3. P. J. Sebastião, A. C. Ribeiro, H. T. Nguyen and F. Noack, *J. Phys. II France*, 1995, **5**, 1707–1724.
4. J. Kowalewski and L. Mäler, *Nuclear Spin Relaxation in Liquids: Theory, Experiments, and Applications*, Taylor & Francis Group, New York, U.S.A., 2006.
5. P.-O. Westlund, *Univ. J. Chem.*, 2016, **4**, 69–73.
6. 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, *ChemComm*, 2017, **53**, 11056–11059.
7. 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.
8. P. Håkansson, T. Boirin and J. Vaara, *Langmuir*, 2018, **34**, 3755–3766.
9. P. Håkansson and P.-O. Westlund, *Phys. Chem. Chem. Phys.*, 2004, **6**, 4321–4329.
10. D. Kruk, J. Kowalewski and P.-O. Westlund, *J. Chem. Phys.*, 2004, **121**, 2215–27.
11. D. Kruk, P. Rochowski, M. Florek – Wojciechowska, P. José Sebastião, D. J. Lurie and L. M. Broche, *J. Magn. Reson.*, 2020, 106783.
12. D. Kruk, A. Herrmann and E. A. Rössler, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2012, **63**, 33–64.
13. K. Åman, P. Håkansson and P.-O. Westlund, *Phys. Chem. Chem. Phys.*, 2005, **7**, 1394–1401.
14. P. Håkansson and P. B. Nair, *Phys. Chem. Chem. Phys.*, 2011, **13**, 9578–89.
15. E. Belorizky, P. H. Fries, L. Helm, J. Kowalewski, D. Kruk, R. R. Sharp and P.-O. Westlund, *J Chem Phys*, 2008, **128**, 052315.
16. 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.
17. U. von Toussaint, *Rev. Mod. Phys.*, 2011, **83**, 943–999.
18. R. Hoffmann, V. I. Minkin and B. K. Carpenter, *Hyle*, 1997, **3**, 23–28.
19. I. Solomon, *Phys. Rev.*, 1955, **99**, 559–465.
20. Y. Ayant, E. Belorizky, J. Alizon and J. Gallicle, *J. Phys. France*, 1975, **36**, 991–1004.
21. L.-P. Hwang and J. H. Freed, *J. Chem. Phys.*, 1975, **63**, 4017–4025.
22. B. Halle, V. P. Denisov and V. Kandadai, in *Biological Magnetic Resonance*, Vol. 17, Kluwer Academic / Plenum Publishers, New York, 1999, ch. Structure, pp. 419–498.
23. A. Roch, R. N. Muller and P. Gillis, *Journal of Chemical Physics*, 1999, **110**, 5403–5411.
24. G. Voigt and R. Kimmich, *J. Magn. Reson.*, 1976, **24**, 149–154.
25. D. Kruk, A. Kubica, W. Masierak, A. F. Privalov, M. Wojciechowski and W. Medycki, *Solid State Nucl. Magn. Reson.*, 2011, **40**, 114–120.
26. T. J. Loredo, in *Maximum Entropy and Bayesian Methods*, ed. P.F. Fougeres, Kluwer Academic c Publishers, Dordrecht, The Netherlands, 1990, pp. 81–142.
27. R. Trotta, *Contemp. Phys.*, 2008, **49**, 71–104.
28. J. Rantaharju and J. Vaara, *Phys. Rev. A*, 2016, **94**, 044312.
29. P. Honegger, V. Overbeck, A. Strate, A. Appelhagen, M. Sappi, E. Heid, C. Schröder, R. Ludwig and O. Steinhauser, *J Chem Phys Lett.*, 2020, **11**, 2165–2170.
30. J. P. Korb, P. J. McDonald, L. Monteilhet, A. G. Kalinicchev and R. J. Kirkpatrick, *Cem Concr Res*, 2007, **37**, 348–350.
31. P.-O. Westlund, G. Driver, Y. Wang, T. Sparrman, A. Laaksonen and Y. Huang, *Phys. Chem. Chem. Phys.*, 2017, **19**, 4975–4988.
32. P. Håkansson, L. Persson and P.-O. Westlund, *J. Chem. Phys.*, 2002, **117**, 8634–8643.
33. V. Wong and D. A. Case, in *Annual Reports in Computational Chemistry*, Elsevier Masson SAS, 2008, vol. 4, pp. 139–154.
34. P. José Sebastião,* Eur. J. Phys.*, 2014, **35**, 1–15.
35. R. Andrae, T. Schulze-Hartung and P. Melchior, *arXiv*, 2010, **1012.3754v1**, 1–12.
36. R. E. Kass and A. E. Raftery, *J Am Stat Assoc.*, 1995, **90**, 773–795.
37. N. Schenker, *J Am Stat Assoc.*, 1985, **80**, 360–361.
38. M. A. Newton and A. E. Raftery, *J R STAT SOC B*, 1994, **56**, 3–48.
39. D. J. C. MacKay, *Neural Comput.*, 1992, **4**, 415–447.
40. 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.
41. B. Hallé and H. Wennerström, *J. Chem. Phys.*, 1981, **75**, 1928–1943.
42. T. Sparrman and P.-O. Westlund, *J. Phys. Chem. B*, 2001, **105**, 9578–89.
12524–12528.

43 K. Åman, E. Lindahl, O. Edholm, P. Håkansson and P.-O. Westlund, *Biophys. J.*, 2003, 84, 102–115.

44 P. Håkansson, *BAYES-SELECT* (github v1.1), 2020, https://github.com/Parcode/BAYES-SELECT/tree/v1.1.

45 P. Håkansson and M. Mella, *J. Chem. Phys.*, 2008, 129, 124101.

46 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, 21, 1087–1092.

47 M. A. Tanner, *Tools for statistical inference*, Springer-Verlag, New York, 3rd edn., 1996.

48 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.

49 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/.

50 D. Foreman-Mackey, D. W. Hogg, D. Lang and J. Goodman, *Publ. Astron. Soc. Pac.*, 2013, 125, 306–312.

51 S. F. Cousin, P. Kaderávek, N. Bolik-Coulon, Y. Gu, C. Charlier, L. Carlier, L. Bruschweiler-Li, T. Marquardsen, J. M. Tyburn, R. Brüschweiler and F. Ferrage, *J. Am. Chem. Soc.*, 2018, 140, 13456–13465.

52 B. Halle, *magn. reson. med.*, 2006, 56, 60–72.

53 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.
### Table 2
Model selection indicators for Table 6. Global MCMC parameter estimation with three temperature data (sample concentration 15mM). Total $\chi^2$ is 67.3 and fit displayed in figure 5.

| T (°C) | $\tau_{c1}$ (ns) | $\tau_{c2}$ (ns) | $\chi^2$ | $S_{1/f_1}$ (10$^{-1}$) | $S_{2/f_2}$ (10$^{-1}$) |
|--------|-----------------|-----------------|----------|-------------------------|-------------------------|
| 10     | 156 (140 171)   | 9.2 (5.1 18)    | 0.62     | 0.24                    | 2.0                     |
| 15     | 130 (122 154)   | 8.2 (5.0 20)    | 0.56     | 0.26                    | 1.9                     |
| 20     | 118 (112 138)   | 11 (4.6 18)     | 0.55     | 0.25                    | 1.4                     |
| (range)| 1.00±1000       | 0.1±100         | 0.001±0.8| 0.01±100                | 0.01±100                |

### Table 3
Model selection indicators for Table 6. Global MCMC parameter estimation with three temperature data (sample concentration 15mM). Total $\chi^2$ is 67.3 and fit displayed in figure 5.

| T (°C) | $\tau_{c1}$ (ns) | $\alpha$ (s$^{-1}$) | $S_{1/f_1}$ (10$^{-1}$) |
|--------|-----------------|---------------------|-------------------------|
| 10     | 120 (114 126)   | 0.77                | 0.38                    |
| 15     | 103 (97 109)    | 0.70                | 0.37                    |
| 20     | 94 (90 99)      | 0.64                | 0.35                    |
| (range)| 1.0±1000        | 0.001±0.8           | 0.001±100               |

### Table 4
Model selection indicators for Table 6. Global MCMC parameter estimation with three temperature data (sample concentration 15mM). Total $\chi^2$ is 12.0 and fit displayed in figure 5.

| T (°C) | $\tau_{c1}$ (ns) | $\tau_{c2}$ (ns) | $\alpha$ (s$^{-1}$) | $S_{1/f_1}$ (10$^{-1}$) | $S_{2/f_2}$ (10$^{-1}$) | $S_{3/f_3}$ (10$^{-1}$) |
|--------|-----------------|-----------------|---------------------|-------------------------|-------------------------|-------------------------|
| 10     | 174 (146 222)   | 3 (15 69)       | 0.54                | 0.20                    | 0.45                    | 4.1                     |
| 15     | 168 (127 211)   | 5.7 (15 69)     | 0.55                | 0.16                    | 0.3                    | 2.5                     |
| 20     | 145 (119 192)   | 4.2 (15 69)     | 0.49                | 0.17                    | 0.35                    | 3.0                     |
| (range)| 1.00±1000       | 0.1±15          | 0.001±0.8           | 0.01±100                | 0.01±100                | 0.01±100                |

### Table 5
Model selection indicators for Table 6. Modified form of model $M_w$. Global MCMC parameter estimation with three temperature data (sample concentration 15mM). Optimized $\chi^2_{opt}$ is 37.98±0.02 and fit displayed in figure 5.

| T (°C) | $R$ (Å) | $\tau_{c1}$ (µs) | $\tau_{c2}$ (ns) | $\alpha$ (s$^{-1}$) | $S_{1/f_1}$ (10$^{-1}$) | $S_{2/f_2}$ (10$^{-1}$) |
|--------|---------|-----------------|-----------------|---------------------|-------------------------|-------------------------|
| 10     | 48 (49.9 50.1) | 9.7 (1.8 10) | 12 (9 15)       | 0.64                | 0.24                    | 1.6                     |
| 15     | —       | —              | —               | 10                  | 0.58                    | —                       |
| 20     | —       | —              | —               | 9                   | 0.54                    | —                       |
| (range)| 1.00±100| 0.001±10       | 0.1±100         | 0.01±0.5            | 0.0±0.1                 | 0.0±0.1                 |

### 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-sites, adjustable parameters ($n$), optimal and mean $\chi^2$, the Bayesian evidence and factor, the criteria AIC and BIC. The multiplication by two in Bayesian factor provides same scale as direct subtraction of BIC and AIC estimates.

| $M_j$ | #water | $n$ | $\chi^2_{opt}$ | $(\chi^2(x) - \chi^2_{opt})$ | $-\ln[p(R^{exp}(M_j, f))]$ | $2\ln[|B_{opt}^j|]$ | AIC   | BIC   |
|-------|--------|----|----------------|------------------------------|----------------------------|---------------------------|-------|-------|
| $M_a$ | 2      | 15 | 35±0.1         | 20.9±0.16                    | 186.7±0.7                 | 0                         | 64.8±0.1 | 169.8±0.1 |
| $M_b$ | 1      | 9  | 67.3±0.04      | 12.6±0.1                    | 453.8±0.8                 | 533                       | 691.4±0.04 | 754.4±0.04 |
| $M_c$ | 3      | 21 | 12.0±0.5       | 29±1                        | 303.9±0.5                 | 234                       | 540.5±0.5 | 200.6±0.5 |
| $M_d$ | 2      | 10 | 37.89±0.02     | 13.0±0.1                    | 144±3                     | 86                        | 59.9±0.2 | 136.9±0.02 |
Supplementary electronic information (SEI) to the article: **Relaxometry models compared with Bayesian technique: Ganglioside micelle example**

Pär Håkansson.
NMR Research Unit, University of Oulu, P.O.Box 3000, 90014 Oulu (Finland); E-mail: par.hakanson@oulu.fi

1. \( \chi^2 \)-Plots

A first guidance on consistency of Markov chain Monte Carlo simulation (MCMC) is to consider the step sequences of \( \chi^2 \), to make the decision after stage (c) in figure 3 (section 4.1 in main article). In Figure S1, the stacked plots of \( \chi^2 \) and \( \chi_{opt}^2 \) for a few simulations are provided. If one or more of the independent simulations show \( \chi^2 \) significantly larger than the others, it is a case for additional burn-in preparation with elevated “temperature” (\( \beta \ll 1 \)) for that particular simulation (see figure 3).

![Figure S1: Stacked plots of \( \chi^2 \) and \( \chi_{opt}^2 \) (solid and dashed lines respectively) for models M_a, M_b, M_c, and M_d.](image)

2. Step-correlation functions

As noted in section 4.1, convergence of MCMC is explored with step-correlation functions for MCMC-configurations. These have the form:\(^1\)

\[
C_{ij}(n) = \frac{(X_i(0)X_j(n))-(X_i)(X_j)}{\sqrt{[(X_i)^2-(X_i)^2][X_j^2-(X_j)^2]}},
\] (S1)

for auto- (\( i=j \)) and cross-correlations (\( i \neq j \)). To obtain efficient numerical calculation of eq. (S1) a implementation inspired by an example-code (F.27) in CPPS\(^2\) was followed. In the procedure (Figure 3 of main paper), a set of \( N_{test} \) MCMC-trajectories are generated. Each of these trajectories are expected to be equilibrated, hence, are stationary random processes,\(^3\) an assumption that is exploited to make the statistical averaging. These ideas are used in Matlab functions\(^4\) scalarcorr.m and CORR_MCMC.m.

The \( (S_{\chi}^2f\beta^2) \) step-correlation functions are provided in Figure S2 for models M_a-M_c. The characteristic step length of the step-correlations are obtained by fitting an exponential function (blue line) to simulated results (red line) in figure S2. The number of steps characterizing the decay rate (the correlation “time”) are provided in variable “corrEstimates” in the example code\(^4\) (see file PLOT_v0c.m).
Figure S2: Step-correlation function for autocorrelation of $(S^2_{f_1})$ and model $M_a$, $M_b$ and $M_c$ respectively. The thick blue line is fitted exponential function and thin red MCMC-simulation.

3. Variable histograms

The variable histograms are provided in Figure S3-S6 for models $M_a$, $M_b$, $M_c$ and $M_d$ respectively.

Figure S3. Fitted parameter histogram for model $M_a$ (data from 20 °C parameterization).

Figure S4. Fitted parameter histogram for model $M_b$ (data from 20 °C parameterization).
Figure S5. Fitted parameter histogram for model $M_1$ (data from 20 °C parameterization).
Figure S6. Fitted parameter histogram for model $M_4$ (data from 20 °C parameterization).

References

1. van Kampen, N. G. STOCHASTIC VARIABLES. in Stochastic Processes in Physics and Chemistry (ed. van Kampen, N. G.) 1–29 (Elsevier, 2007). doi:10.1016/B978-044452965-7/50004-

2. Allen, M. P. CCP5, Collaborative Computational Project. Available at: http://www.ccp5.ac.uk/software/allen_tildersley.

3. McQuarrie, D. A. Statistical Mechanics. (Harper & Row, 1976).

4. Håkansson, P. BAYES-SELECT (github v1.1). (2020). doi:10.5281/zenodo.4012179
