THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

# Bayesian Probabilistic Inference of Nonparametric Distance Distributions in DEER Spectroscopy

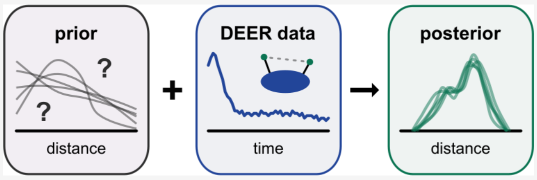
Sarah R. Sweger, Julian C. Cheung, Lukas Zha, Stephan Pribitzer, and Stefan Stoll*

**Cite This:** *J. Phys. Chem. A* 2024, 128, 9071−9081

Read Online

ACCESS | Metrics & More | Article Recommendations

**ABSTRACT:** Double electron−electron resonance (DEER) spectroscopy measures distance distributions between spin labels in proteins, yielding important structural and energetic information about conformational landscapes. Analysis of an experimental DEER signal in terms of a distance distribution is a nontrivial task due to the ill-posed nature of the underlying mathematical inversion problem. This work introduces a Bayesian probabilistic inference approach to analyze DEER data, assuming a nonparametric distance distribution with a Tikhonov smoothness prior. The method uses Markov Chain Monte Carlo sampling with a compositional Gibbs sampler to determine a posterior probability distribution over the entire parameter space, including the distance distribution, given an experimental data set. This posterior contains all of the information available from the data, including a full quantification of the uncertainty about the model parameters. The corresponding uncertainty about the distance distribution is visually captured via an ensemble of posterior predictive distributions. Several examples are presented to illustrate the method. Compared with bootstrapping, it performs faster and provides slightly larger uncertainty intervals.



## 1. INTRODUCTION

Double electron−electron resonance (DEER) spectroscopy is a pulse electron paramagnetic resonance (EPR) technique utilized for determining distances between electron spin centers on a nanometer scale,[1,2] predominantly on proteins. DEER resolves the full distribution of distances in an ensemble of proteins, making it possible to directly quantify conformational ensembles and the underlying conformational landscapes.[3−5] DEER measures an oscillatory time-domain signal that depends on the magnitude of the magnetic dipole−dipole interaction between the spin centers, which in turn depends on the inverse cube of the distance $r$. In the analysis, this signal is fitted with a model that includes a distance distribution $P(r)$. Mathematically, this constitutes an ill-posed inversion problem. Assessment of uncertainty in the fitted distance distribution is therefore challenging but is crucial for making sound conclusions about the conformational landscape of the protein.

Analysis approaches for obtaining a distance distribution from a measured DEER signal range from analytical solutions[6] to neural networks.[7,8] The least-squares fitting methods that have seen the widest practical application utilize one of two models for $P(r)$: either a Gaussian mixture model or a nonparametric representation combined with Tikhonov regularization.[9−16] Gaussian mixture models are parametric and represent the distribution as a linear combination of several Gaussian functions. Nonparametric models represent $P(r)$ as a histogram over a discretized distance range, combined with Tikhonov regularization that includes a roughness penalty for the distribution into the fitting objective function. Both Gaussian mixture models and nonparametric models with Tikhonov regularization can be fit directly to the raw data in a single step.[17] In both approaches, however, correctly quantifying and visualizing uncertainty is challenging.

For Gaussian mixture models, uncertainty analysis relies primarily on parameter confidence intervals, which are obtained from the covariance matrix or by explicitly exploring the sensitivity of the objective function on the parameter values.[12−14,18] The parameter confidence intervals are then propagated to the distance domain to yield error bands for the distribution. This method assumes that the error surface is quadratic and that the parameters are unbounded, neither of which is generally true. Our previous work extended the available uncertainty analysis methods for Gaussian mixture models by implementing a Bayesian inference approach.[16] The method directly models the raw DEER data and yields a full joint probability distribution over all model parameters, thereby fully quantifying their uncertainties. Similar Bayesian data analysis methods have been implemented previously for NMR and EPR.[16,19−23]

For nonparametric distribution models with Tikhonov regularization, partial uncertainty analysis is commonly conducted by manually varying some parameters in the

**Received:** July 26, 2024
**Revised:** September 18, 2024
**Accepted:** September 24, 2024
**Published:** October 4, 2024

analysis (intermolecular background parameters, modulation depth, and noise) and summarizing the sensitivity of the extracted distance distribution to these parameters into error bands around the fitted distribution.[11,24] Another method to obtain confidence intervals for both approaches is bootstrapping. As implemented in DeerLab,[17] it generates an ensemble of distributions by analyzing a large number of synthetically generated hypothetical signals based on the fitted model. In previous work, we introduced a partial approach based on Bayesian inference to quantify the uncertainty in the distribution due to the noise in the signal.[22] Unfortunately, this work required prior processing to remove the intermolecular background contribution and could not incorporate parameters beyond noise.

In this paper, we present a Bayesian inference approach for analyzing a DEER trace using a nonparametric distance distribution with Tikhonov smoothing. It extends our previous work[22] and models the raw DEER data directly without prior background correction. It yields a full probability distribution for all model parameters, providing complete quantitative information about uncertainty and correlations for all parameters without any implicit limiting assumptions. We also introduce distribution ensembles as a visual tool to effectively represent uncertainty about the distance distribution, including correlations that are neglected when using visualizations based on error bands.

The paper is structured as follows. Section 2 presents the model, and Section 3 outlines the inference methodology. Section 4 shows examples using synthetic and experimental data, including a comparison between parametric and nonparametric $P(r)$ models, an analysis of the dependence on the distance range, and a comparison of the quantified uncertainty with that obtained from bootstrapping. Finally, Section 5 discusses the merits of this method in comparison to others.

## 2. PROBABILISTIC DEER MODEL

To model the DEER data, we start from the general noise-free continuous-time physical model[17,25]

$$V_{\mathrm{M}}(t) = V_0 \int_0^\infty K(t, r)P(r)\mathrm{d}r \tag{1}$$

where $V_{\mathrm{M}}(t)$ represents the DEER signal as a function of dipolar evolution time $t$, $V_0$ is an overall amplitude factor, $P(r)$ is the distribution of intramolecular interspin distances $r$ normalized such that $\int_0^\infty P(r)\,\mathrm{d}r = 1$, and $K(t, r)$ is the kernel function that provides the DEER signal as a function of $t$ and $r$. In this work, we use

$$K(t, r) = [(1 - \lambda) + \lambda K_0(t, r)]\cdot B(t) \tag{2}$$

with the modulation depth $\lambda$, the elementary kernel function

$$K_0(t, r) = \int_0^1 \cos[(1 - 3\cos^2\theta)Dr^{-3}t]\mathrm{d}\cos\theta \tag{3}$$

and the constant $D = (\mu_0/4\pi)\, g^2\mu_{\mathrm{B}}^2/\hbar$, with the electron g factor $g$, the spin concentration $c$, the Bohr magneton $\mu_{\mathrm{B}}$, the reduced Planck constant $\hbar$, and the magnetic constant $\mu_0$. $\theta$ is the angle between the applied magnetic field and the interspin direction.

The first factor in eq 2 represents the intramolecular contribution. The second factor, $B(t)$, represents the intermolecular contribution, also called the background and sometimes denoted as $V_{\mathrm{inter}}(t)$. Here, we use an exponential decay corresponding to a homogeneous three-dimensional distribution of spins

$$B(t) = \exp(-k|t|) \tag{4}$$

with $k = (8\pi^2/9\sqrt{3})Dc\lambda$. More extended background models that incorporate fractal dimensions or volume exclusions are occasionally needed and can be incorporated easily,[26] although they will increase the number of model parameters. We will use the value of the background function at the end of the DEER time trace

$$B_{\mathrm{end}} = B(t_{\mathrm{end}}) \tag{5}$$

as an alternate way to specify the background decay rate, via $k = -\log(B_{\mathrm{end}})/t_{\mathrm{end}}$.

Experimentally, the DEER signal is measured at a set of linearly spaced discrete time points $t = t_i$ and is therefore represented as $n_t$-element vector $\boldsymbol{V}$ with elements $V_i$. The measured values typically include normally distributed noise $\epsilon_i$ with mean zero and $t$-independent variance $\sigma^2$.[22] This is represented as

$$V_i = V_{\mathrm{M}}(t_i) + \epsilon_i \text{ with } \epsilon_i \sim \text{normal}(0, \sigma^2) \tag{6}$$

or, equivalently,

$$V_i \sim \text{normal}(V_{\mathrm{M}}(t_i), \tau^{-1}) \tag{7}$$

The tilde ∼ indicates that the quantity on the left is distributed according to the probability distribution on the right. The precision $\tau$ is the inverse of the variance, $\tau = 1/\sigma^2$.

While it is possible to use a closed-form expression for the angular integral in eq 3, the integral over the distance distribution in eq 1 can be evaluated only numerically. We do this by discretizing $P(r)$ at a set of equidistant distances $r = r_j$, giving the $n_r$-element vector $\boldsymbol{P}$. This gives

$$\boldsymbol{V} = \boldsymbol{V}_{\mathrm{M}} + \boldsymbol{\epsilon} = \boldsymbol{KP} + \boldsymbol{\epsilon} \tag{8}$$

with the kernel matrix $\boldsymbol{K}$ with elements $K_{i,j} = V_0\, K(t_i, r_j)\Delta r$, where $\Delta r = r_{j+1} - r_j$ is the increment in the $r$ domain. $\boldsymbol{\epsilon}$ is the noise vector with elements $\epsilon_i$.

Extracting the distance distribution from eq 8 using Tikhonov regularization is done by minimizing an objective function that includes a misfit term and a Tikhonov regularization term,

$$\boldsymbol{P}_\alpha = \min_{\boldsymbol{P}\geq 0}(\|\boldsymbol{V} - \boldsymbol{KP}\|^2 + \alpha^2\|\boldsymbol{LP}\|^2) \tag{9}$$

Here, $\alpha$ is the Tikhonov regularization parameter, and $\boldsymbol{L}$ is an operator matrix, most commonly the $(n_r - 2) \times n_r$ second-order difference matrix

$$\boldsymbol{L} = \begin{pmatrix} 1 & -2 & 1 & & & 0 \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & 1 & -2 & 1 \end{pmatrix} \tag{10}$$

With $\boldsymbol{L}$ as defined, the end points of the distance range are neglected in this penalty term. They can be included by extending $\boldsymbol{L}$ with an additional first row with −2 and 1 as the first elements and with an additional last row with 1 and −2 as the last two elements.

The model specified above depends on a set of parameters: the distance distribution vector penalized for roughness $\boldsymbol{P}_\alpha$, the modulation depth $\lambda$, the end point $B_{\mathrm{end}}$ of the intermolecular

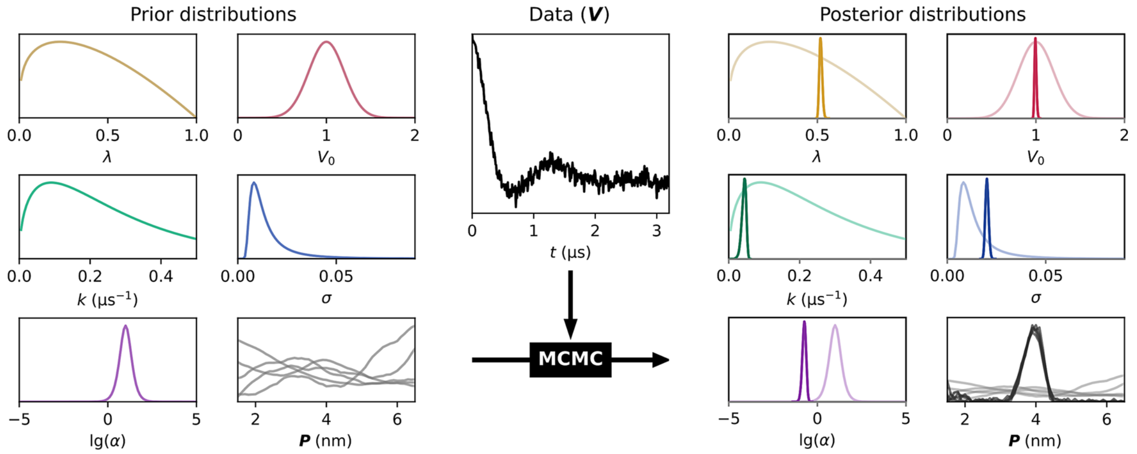


**Figure 1.** Visualization of Bayesian inference. Prior distributions for model parameters (modulation depth ($\lambda$), echo amplitude ($V_0$), background decay constant ($k = -\log(B_{\text{end}})/t_{\text{end}}$), noise ($\sigma = 1/\sqrt{\tau}$), Tikhonov smoothness parameter ($\alpha = \sqrt{\delta/\tau}$), and the distance distribution ($\boldsymbol{P}$) are combined with the data $\boldsymbol{V}$ to yield a posterior probability distribution, with the marginalized posterior distributions for individual parameters shown. An ensemble of five distance distributions is shown for the prior and posterior of $\boldsymbol{P}$.

background decay, the noise precision $\tau$, and the overall amplitude $V_0$. We indicate the parameter set as

$$\boldsymbol{\theta} = \{\boldsymbol{P}_\alpha, \lambda, B_{\text{end}}, \tau, V_0\} \tag{11}$$

Although the model is formulated here in terms of $B_{\text{end}}$ and $\tau$ for more efficient numerical sampling, we will show the results in terms of $k$ and $\sigma$ to remain consistent with the standard practices in DEER data analysis.

The goal of analyzing $\boldsymbol{V}$, given the chosen model, is to determine the range of values for the model parameters that are consistent with the data. Now, since $\boldsymbol{V}$ is an incomplete representation of $V(t)$ due to time truncation, time discretization, amplitude discretization, and amplitude noise, there will be uncertainty associated with these parameters. It is important to quantify this uncertainty as well. Therefore, the goal of the analysis is to determine the full joint probability distribution of all parameters, given the data $\boldsymbol{V}$, symbolically written as

$$p(\boldsymbol{\theta}|\boldsymbol{V}) \tag{12}$$

This distribution is called the posterior distribution or simply posterior. It is posterior in the sense that it represents the probability distribution of the parameters *after* the data are taken into account. Once calculated, the posterior can be visualized or used to obtain statistics on the parameters such as means and spreads.

Using Bayes' theorem, the posterior can be calculated via

$$p(\boldsymbol{\theta}|\boldsymbol{V}) \propto p(\boldsymbol{V}|\boldsymbol{\theta})\cdot p(\boldsymbol{\theta}) \tag{13}$$

The first term on the right provides the probability of the data given specific values for the model parameters. When seen as a function of the parameters, it is called the likelihood function. Based on eqs 8 and 9, it is

$$\begin{aligned} p(\boldsymbol{V}|\boldsymbol{\theta}) &= \text{normal}(\boldsymbol{V}; \boldsymbol{KP}, \tau^{-1}) \\ &\propto \tau^{n_t/2}\exp\left(-\frac{1}{2}\tau\left\|\boldsymbol{V} - \boldsymbol{KP}\right\|^2\right) \end{aligned} \tag{14}$$

The second term on the right in eq 13 is the prior distribution or simply prior. It represents the probability distribution of the parameters prior to taking the data into account, summarizing information about the parameters that is available independently of the given data set. For example, we know without any data that $P_i$ is non-negative, that $\lambda$ is between 0 and 1, that $V_0$ is around 1 (assuming the experimental trace is normalized to maximum 1), and that $B_{\text{end}}$ is between 0 and 1. We take the prior as a product of independent distributions over individual parameters:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{P}, \delta)p(\lambda)p(V_0)p(\tau)p(B_{\text{end}}) \tag{15}$$

with the smoothness hyperparameter $\delta$ that is related to the Tikhonov regularization parameter $\alpha$ by[22]

$$\delta = \alpha^2\tau \tag{16}$$

We include $\delta$ as an additional parameter in $\boldsymbol{\theta}$. Note that we treat $B_{\text{end}}$ (or $k$) as an independent parameter, even though it depends on $\lambda$ (see eq 4). Alternatively, the concentration $c$ could be used as a model parameter instead of $k$ or $B_{\text{end}}$.

As prior for $\boldsymbol{P}$, we encode our knowledge that $\boldsymbol{P}$ is element-wise non-negative, normalized, and expected to be smooth, i.e., $P_i$ and $P_j$ should be similar if the distances $r_i$ and $r_j$ are similar. For this, we write

$$p(\boldsymbol{P}, \delta) = p(\boldsymbol{P}|\delta)p(\delta)f(\boldsymbol{P}) \tag{17}$$

The function $f(\boldsymbol{P})$ in eq 17 is an indicator function that equals one if all elements of $\boldsymbol{P}$ are non-negative and $\boldsymbol{P}$ integrates to 1, and zero otherwise.

For the smoothness prior, we assume a normal distribution

$$\begin{aligned} p(\boldsymbol{P}|\delta) &= \text{normal}(\boldsymbol{P}; \boldsymbol{0}, (\delta\boldsymbol{L}^{\text{T}}\boldsymbol{L})^{-1}) \\ &\propto \delta^{n/2}\exp\left(-\frac{1}{2}\delta\left\|\boldsymbol{LP}\right\|^2\right) \end{aligned} \tag{18}$$

where $n$ is the number of nonzero elements in $\boldsymbol{P}$ (to ensure proper normalization[27,28]). This distribution assigns high prior probabilities to smooth distributions (where $\|\boldsymbol{LP}\|^2$ is small) and low prior probabilities to rough distributions (where $\|\boldsymbol{LP}\|^2$ is large). This is motivated physically by the flexibility of the spin labels and the biomacromolecule to which the pair of spin labels is attached. This corresponds to the penalty term in eq 9.

For the hyperprior for the regularization parameter $\delta$, we select a gamma distribution[22]

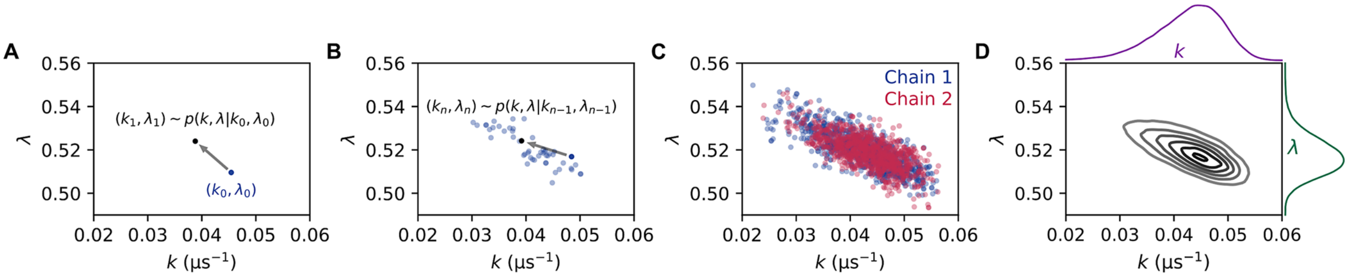


**Figure 2.** Principle of MCMC sampling of the posterior distribution. (A) A random starting value is chosen from the joint posterior of $k$ and $\lambda$ (blue). The next random sample is chosen from the probability distribution of $k$ and $\lambda$ given the starting value (black), and thus, every step is dependent upon the previous step. (B) This process continues, where each new step is a random draw from a probability distribution for $k$ and $\lambda$ given the position of the last step. (C) Regions of higher probability are sampled more frequently until the chain converges. Multiple chains are run, indicated by color, to obtain inter- and intrachain convergence. (D) The result of the sampling is a representation of the full posterior distribution of the parameters. Shown here are the 1D and 2D marginal posteriors for $k$ and $\lambda$.

$$p(\delta) = \mathrm{gamma}(\delta; a_\delta, b_\delta) \tag{19}$$

with $a_\delta = 1$ and $b_\delta = 10^{-6}$. This gives a very broad distribution function that decays exponentially with increasing $\delta$.

For the priors for the other parameters, we use the same broad distributions as used in our earlier work on parametric $\boldsymbol{P}$ models[16]

$$p(\lambda) = \mathrm{beta}(\lambda; 1.3, 2) \tag{20}$$

$$p(B_{\mathrm{end}}) = \mathrm{beta}(B_{\mathrm{end}}; 1.0, 1.5\ \mu\mathrm{s}) \tag{21}$$

$$p(V_0) = \mathrm{bnd}(\mathrm{normal}(V_0; 1, 0.\,2^2), 0) \tag{22}$$

$$p(\tau) = \mathrm{gamma}(\tau; a_\tau, b_\tau) \tag{23}$$

with $a_\tau = 1$ and $b_\tau = 10^{-4}$. However, the priors do not necessarily need to follow these distributions. Instead, they should be chosen based on the known information about the system and setup. The same applies for the hyperparameter $\delta$ which will be reported as $\lg(\alpha)$ throughout by using eq 16. The estimation of posterior probabilities from prior probabilities and observed data through Bayesian inference is visualized in Figure 1.

With the above expressions, the posterior $p(\boldsymbol{\theta}|\boldsymbol{V})$ is fully defined. It has some structure that is important to recognize. (a) It is a gamma distribution in $\tau$

$$p(\tau | \boldsymbol{V}, \boldsymbol{\theta}_{-\tau}) = \mathrm{gamma}(\tau; \tilde{a}_\tau, \tilde{b}_\tau) \tag{24}$$

where $\boldsymbol{\theta}_{-\tau}$ indicates the set of all parameters except $\tau$. The distribution parameters are $\tilde{a}_\tau = a_\tau + n_t/2$ and $\tilde{b}_\tau = b_\tau + \|\boldsymbol{V} - \boldsymbol{KP}\|^2/2$, where $n_t$ is the number of elements in $\boldsymbol{V}$. (b) It is a gamma distribution in $\delta$

$$p(\delta | \boldsymbol{V}, \boldsymbol{\theta}_{-\delta}) = \mathrm{gamma}(\delta; \tilde{a}_\delta, \tilde{b}_\delta) \tag{25}$$

with $\tilde{a}_\delta = a_\delta + n/2$ and $\tilde{b}_\delta = b_\delta + \|\boldsymbol{LP}\|^2/2$. (c) It is a truncated multivariate normal distribution in $\boldsymbol{P}$

$$p(\boldsymbol{P} | \boldsymbol{V}, \boldsymbol{\theta}_{-\boldsymbol{P}}) = \mathrm{normal}(\boldsymbol{P}; \overline{\boldsymbol{P}}, \boldsymbol{\Sigma}) f(\boldsymbol{P}) \tag{26}$$

with center $\overline{\boldsymbol{P}} = \tau \boldsymbol{\Sigma} \boldsymbol{K}^{\mathrm{T}} \boldsymbol{V}$ and covariance matrix $\boldsymbol{\Sigma} = (\tau \boldsymbol{K}^{\mathrm{T}} \boldsymbol{K} + \delta \boldsymbol{L}^{\mathrm{T}} \boldsymbol{L})^{-1}$. We will make use of these structures for the sampling methodology in the next section.

## 3. INFERENCE

The analytical form of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{V})$ is intractable. In particular, it is not possible to evaluate integrals required to determine the mean, the variance, or marginalized distributions of individual parameters. Therefore, we resort to representing the distribution by a finite set of samples generated numerically such that the density of samples in various regions in parameter space is proportional to the local probability density (see Figure 2). These samples are then used to evaluate (approximately) the aforementioned integrals and to construct visualizations.

We use Markov Chain Monte Carlo (MCMC) sampling to generate a Markov chain of samples from the posterior,[29] where each sample $i$, containing ($\boldsymbol{P}_i$, $\tau_i$, $\delta_i$, $V_{0,i}$, $B_{\mathrm{end},i}$, and $\lambda_i$), is generated from the previous sample $i - 1$, containing ($\boldsymbol{P}_{i-1}$, $\tau_{i-1}$, $\delta_{i-1}$, $V_{0,i-1}$, $B_{\mathrm{end},i-1}$, and $\lambda_{i-1}$). A simplified example of the MCMC sampling process is shown in Figure 2.

We investigated two separate MCMC sampling strategies. The first strategy involves a compositional Gibbs sampling approach that utilizes independent draws for the three parameters that have simple analytical conditional posterior distributions ($\tau$, $\delta$, and $\boldsymbol{P}$) and Hamiltonian Monte Carlo (HMC) draws for the remaining three ($V_0$, $B_{\mathrm{end}}$, and $\lambda$). The second one utilizes HMC sampling for all model parameters.

The compositional Gibbs sampling approach is based on our previous work.[22] After a starting point is chosen, it proceeds iteratively as follows:

(1) Generate the $i$th random sample of the precision $\tau$ from its full conditional posterior distribution eq 24 using the values of all other parameters from sample $i - 1$.

(2) Generate the $i$th random sample of the regularization parameter $\delta$ from its posterior distribution eq 25 using the new value for $\tau$ and the previous values of all other parameters.

(3) Generate the $i$th random sample of $\boldsymbol{P}$ from its full conditional posterior distribution eq 26 with the new values for $\tau$ and $\delta$ and the previous values for all other parameters from sample $i - 1$. To generate a random sample from this distribution, we use the fast non-negative least-squares (FNNLS) algorithm by Bro and De Jong[30] to enforce the non-negativity and normalization constraints imposed by $f(\boldsymbol{P})$. Other algorithms for generating samples from a truncated multivariate normal distribution are available in the literature.[31,32]

For the remaining three parameters, since their posteriors are not of a form for which independent sampling is possible, we use an HMC algorithm known as the no-U-turn sampler (NUTS)[33] to simultaneously generate the next samples of these parameters:

(4) Generate the $i$th random sample of the remaining parameters ($V_0$, $B_{\mathrm{end}}$, and $\lambda$) with the NUTS sampler, using the

new values of $\boldsymbol{P}$, $\tau$, and $\delta$ and the previous values of $V_0$, $B_{\text{end}}$, and $\lambda$.

HMC methods take the negative logarithm of the posterior distribution to draw from as a potential-energy landscape and the parameters as position variables of fictitious particles. Samples are then generated by simulating particle trajectories on this landscape with classical Hamiltonian dynamics using momenta that are drawn from a multivariate normal distribution.[34,35] NUTS autotunes the step size and the number of steps used in the integration of the Hamiltonian dynamics.

The second sampling approach we investigated samples all six parameters simultaneously with the NUTS sampler. However, since eq 18 would yield negative values for $\boldsymbol{P}$ when sampled with NUTS, we represent $\boldsymbol{P}$ as a uniform Dirichlet distribution $\text{Dir}(\boldsymbol{P};\,\boldsymbol{1})$ to encode the non-negativity and normalization constraints from $f(\boldsymbol{P})$ in eq 17. To accommodate the smoothness prior in this implementation, we add the $p(\boldsymbol{P}|\delta)$ term from eq 18 directly as an additional term to the potential-energy function of the NUTS sampler. Thus, eq 17 becomes

$$p(\boldsymbol{P},\,\delta) \propto p(\boldsymbol{P}|\delta)p(\delta)\text{Dir}(\boldsymbol{P};\,\boldsymbol{1}) \tag{27}$$

We continue to use eqs 19−23 as our priors for the other parameters.

We implemented both sampling algorithms in the Python package PyMC 5.10.4,[36] which uses autodifferentiation for the calculation of the gradients necessary for calculating the Hamiltonian trajectories in NUTS sampling. For each analysis, four chains containing several thousand to several tens of thousands of samples are run. These chains are then assessed for convergence using the rank-normalized split $\hat{R}$ statistic, which compares intra- and interchain variances.[37−39] Values of $\hat{R}$ very close to 1 indicate that the chains are stationary and similar, such as in the example in Figure 2C. Chains in this work are considered converged when $\hat{R} < 1.05$. Although convergence is essential to proper analysis, some runs take impracticably long to converge. When this occurs, we exclude chains one by one and observe the effect on $\hat{R}$, and then remove the chain that leads to the largest reduction in $\hat{R}$ when excluded. In practice, we did not need to remove more than one chain to attain convergence, but this process can be repeated. All code used for modeling and sampling in this paper was run within our Python package dive, which can be accessed at https://github.com/StollLab/dive.

After convergence, the pooled samples from all chains represent the full multidimensional posterior $p(\boldsymbol{\theta}|\boldsymbol{V})$. Due to its large dimensionality, it is not possible to visualize it directly. Instead, we examine each parameter individually using its marginalized posterior, which is obtained by integrating the full posterior over all other parameters. This integral is approximated by generating a histogram of the parameter values from all samples, smoothed with a Gaussian with a line width of 1/5 of the standard deviation of the parameter values. This results in a one-dimensional distribution that can be easily plotted. On the right of Figure 1, the marginalized posteriors are shown in color, together with the priors in less saturated color. In this case, the posteriors are much narrower than the corresponding priors. The spread of the posterior distribution is a quantitative measure of inferential uncertainty, and its narrowing compared to the prior distribution is a direct measure of information content of the data.

However, marginalization discards all information about correlations between parameters. It is also possible to display and examine 2D marginalized posteriors between pairs of parameters, which is particularly helpful for assessing issues of convergence often caused by highly correlated parameters. Examples of 1D and 2D marginalized posteriors are shown in Figure 2D.

We additionally visualize the results of Bayesian inference using posterior predictive samples for $V(t)$ and $P(r)$. For this, we randomly pick a small set of samples (about 50−100) of the parameters, including the distance distribution, from the pooled MCMC samples. An ensemble of noise-free time-domain signals and background decays is then generated from the sampled parameters. Plotting these posterior predictive samples of the time-domain trace and the distance distribution allows for a visual assessment of fit quality and of uncertainty in the inferred distance distribution.

When visualized, we found that the two MCMC methods we investigated had some differences. Specifically, when using the NUTS sampler for all parameters, the sampled distance distributions differ from those of the compositional Gibbs−NUTS sampler; whereas the compositional sampler generates $\boldsymbol{P}$ vectors with a significant number of points equal to exactly 0, the NUTS sampler never generates $\boldsymbol{P}$ with points equaling 0.

This difference arises from the use of the FNNLS algorithm to generate non-negative $\boldsymbol{P}$ draws in the compositional sampling approach. The FNNLS algorithm initially sets all points in $\boldsymbol{P}$ to zero in a non-negativity constrained "active set" and improves the fit by iteratively moving points to an unconstrained "passive set" until the fit can no longer be improved.[30] At this point, $\boldsymbol{P}$ consists of some positive points and some zero points, meaning that the probability of points in $\boldsymbol{P}$ being equal to 0 is significant. This is similar to a spike-and-slab prior, a common distribution in Bayesian inference involving a discontinuity at 0 in an otherwise smooth distribution to increase the probability of 0.[28] Spike-and-slab priors are examples of priors that encourage sparse distributions; the continuous Laplace, double Pareto, and horseshoe priors are other examples of sparsity-inducing priors.[40] Thus, the FNNLS algorithm effectively adds an implicit sparsity-inducing term to eq 17 that encourages points in $\boldsymbol{P}$ to be 0, representing our knowledge (or assumption) that there are many distances that the spin label pair does not populate. Although a similar effect could be achieved in the NUTS sampler by using an explicit sparsity-inducing prior for $\boldsymbol{P}$ in eq 27, few multidimensional, non-negative, constant-sum, sparsity-inducing priors are available in the literature.

Furthermore, in our implementation, the NUTS sampler runs more slowly than the Gibbs−NUTS sampler due to the larger number of parameters that are included in the calculation of the potential-energy landscape. We also found that the additional complexity of this landscape leads to a greater frequency of undesirable divergences, which occur when the NUTS sampler, which takes discrete steps, encounters regions that are too steep to sample accurately. For these reasons, we chose to use the compositional Gibbs−NUTS sampling method for the rest of this discussion, taking note of the implicit sparsity bias. However, the NUTS sampling approach remains as an alternative.

## 4. RESULTS AND DISCUSSION

**4.1. Basic Illustrations.** We first illustrate the probabilistic analysis method on synthetic data, using the large simulated T4 lysozyme (T4L) test data set published by Edwards et al.[15,41] The distributions in this test data set were generated computationally from an in silico spin-labeled crystal structure of T4L. Distribution 3992 from the test data set is taken as ground truth, and two DEER traces of differing quality were generated and then analyzed with the Bayesian inference method using the aforementioned nonparametric model. The traces and the analysis results are shown in Figure 3. The trace in panel A has a large modulation depth, a slow background decay rate, a long trace length, and a small noise level. Comparatively, the trace in panel B has less ideal values for all of these parameters, in particular, a shorter trace length and a higher noise level. This provides a challenging case with a higher inferential uncertainty.

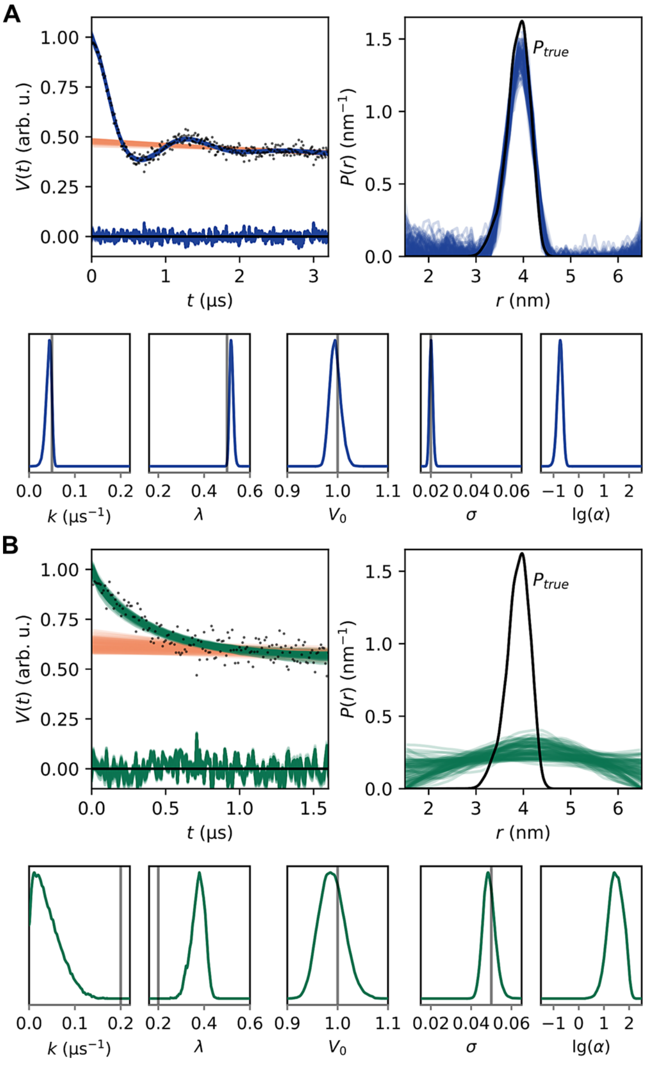


**Figure 3.** Validation of the Bayesian inference method using synthetic data generated from distribution 3992 from the Edwards test set.[15,41] The time trace in panel A was generated using $\lambda = 0.5$, $k = 0.05\ \mu s^{-1}$, $t_{\max} = 3.2\ \mu s$, and $\sigma = 0.02$, and the trace in panel B was generated using $\lambda = 0.2$, $k = 0.2\ \mu s^{-1}$, $t_{\max} = 1.6\ \mu s$, and $\sigma = 0.05$. The MCMC simulation of each was run with four chains and 20,000 samples per chain. In each panel, the top left plot shows the time-domain data (gray dots), ensembles of fitted signals $\boldsymbol{V}$ (blue/green) and backgrounds $(1 - \lambda)\boldsymbol{B}$ (orange), and the residuals. The top right plot shows an ensemble of distance distributions drawn from the posterior distribution (blue/green) and the ground-truth distance distribution (black). The bottom plots show the marginalized parameter posteriors. The gray lines indicate true parameter values.

For both cases, the marginalized posteriors for all scalar parameters are shown (Figure 3A,B, bottom). The gray lines indicate the ground-truth values used in generating the trace. There is no line shown for $\alpha$ as this is a nonphysical parameter introduced in the analysis. For the longer and less noisy trace, $V_0$ and $\sigma$ are recovered accurately and with little uncertainty, as indicated by the narrow posterior distributions with modes close to the ground-truth values. For $k$ and $\lambda$, the posterior modes align less with the ground-truth values, and the spread in the case of $k$ is a bit larger. The cause for this is discussed in more depth below.

The parameter posteriors most directly show the outcome of the Bayesian analysis and are useful for identifying the impact of individual parameters on the overall fit. However, the time-domain fit and the distance distribution are the results of primary interest. These quantities are shown via ensembles of posterior predictive samples (Figure 3A,B, top), as described in Section 3. As in our previous work, this visualization of uncertainty is preferred, as it does not emphasize any particular distribution and more completely encompasses the range of fits compatible with the data. The time-domain plot shows that the fit is excellent, and the distance-domain plots show that there is little scatter in center, width, and shape among the $P(r)$ distributions. The ensemble also overlaps well with the ground truth, indicating that Bayesian analysis recovers the distribution from the data with little uncertainty. There is somewhat elevated uncertainty at short distances and at the long-distance edge.

Analysis of the posterior modes for the shorter, noisier data set (Figure 3B) shows that for $V_0$ and $\sigma$, the modes are reasonably well-identified, but the spreads are wider than for the first data set. For $k$ and $\lambda$, the method is unable to recover the ground-truth values accurately. The difficulty of recovering these parameters is indicative of a problem inherent to the underlying physical model—parameter nonidentifiability.[42] The signals from the intermolecular background and from intramolecular long distances are very similar. Given the short and noisy data, there is insufficient information for identifying and separating the two contributions, resulting in skewed and broad posterior distributions for $k$ and $\lambda$. In addition, the inferred distance distribution is very broad and uncertain, preventing specific structural conclusions. This broadness indicates that it is dominated by the prior, and the data did not provide significant additional information about $P(r)$.

Both data sets in Figure 3 show posterior distributions of $\alpha$. In least-squares fitting approaches, a single value of $\alpha$ is selected ad hoc or based on one of a series of criteria (L-curve, Akaike information criterion, and so forth). In the inference approach presented here, the prior together with the data results in a distribution of likely $\alpha$ values, without the need to pick a particular value or criterion.

### 4.2. Model Comparison.

The rest of the examples presented utilize the DEER data recently published as part of a benchmark test and guidelines paper for DEER.[43] Four constructs of the Yersinia outer protein O (YopO) from *Yersinia enterocolitica* without its membrane anchor were measured by seven different laboratories. We use the data from lab B. YopO contains an $\alpha$-helix that is 43 amino acids long, allowing for three site pairs to be chosen that encompass the short- (S585R1/Q603R1), mid- (V599R1/N624R1), and long- (Y588R1/N624R1) range distances accessible by DEER.

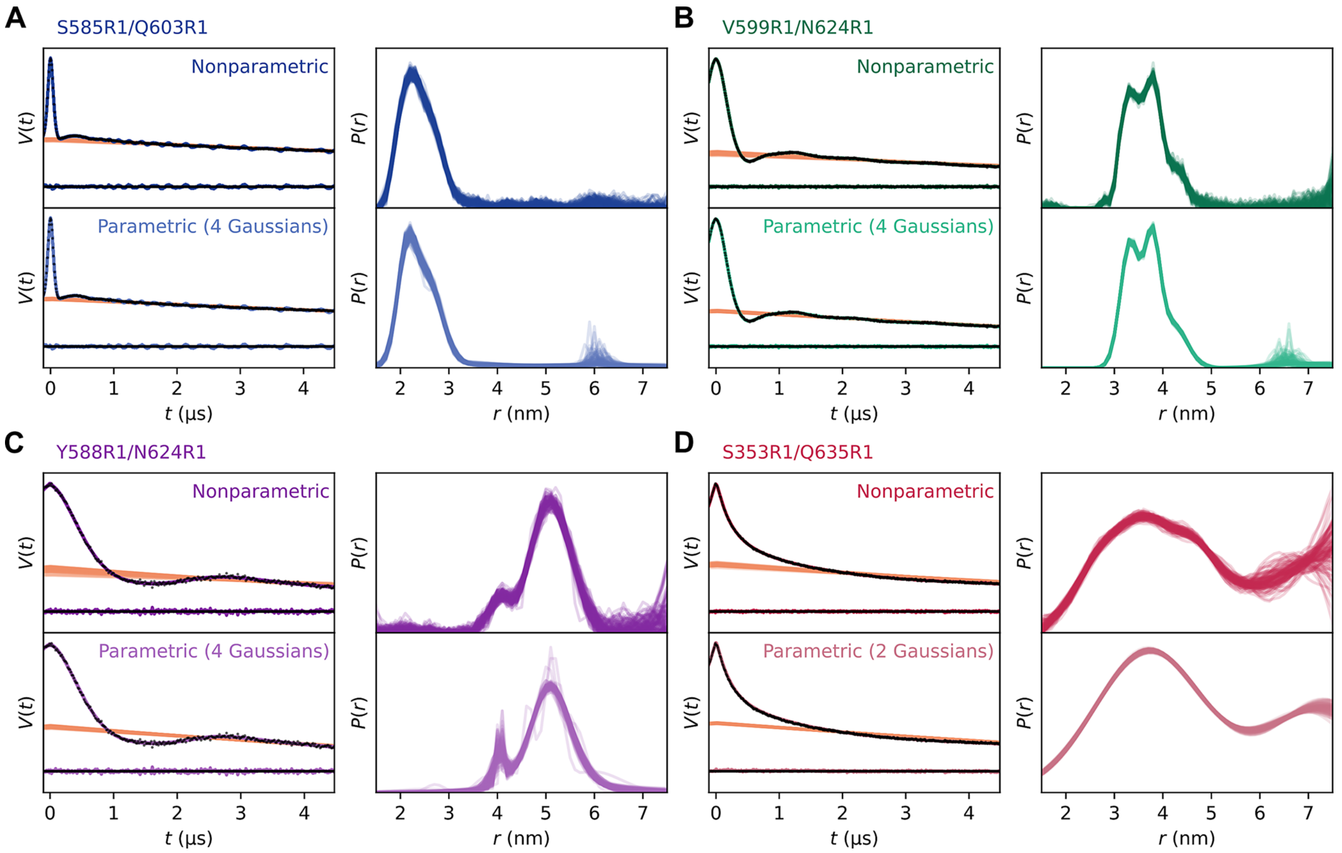


**Figure 4.** Bayesian inference using a nonparametric model with Tikhonov regularization and a multi-Gauss parametric model on DEER data obtained for four constructs of YopO.[43,44] Most MCMC data shown were run with four chains of 20,000 samples per chain from which 100 samples were randomly selected. The nonparametric models for panels B, C, and D were run with four chains of 100,000 samples to achieve convergence according to $\hat{R} < 1.05$. For the nonparametric model for panel B and the parametric models for panels A, B, and D, one chain was dropped due to nonconvergence. The raw experimental data are shown in gray overlaid with the time-domain full and intermolecular fits for the nonparametric model (top, dark) and the parametric multi-Gauss model (bottom, light). Beneath the time-domain data are the residuals from the posterior predictive ensemble. The distance distributions for each are shown according to the same color scheme.

A fourth site pair was chosen to include a spin label on a flexible loop, giving rise to a very broad distribution (S353R1/Q635R1).

The results of using the Bayesian inference approach with both the nonparametric model for $P(r)$ described in this paper and the parametric multi-Gauss model from our previous work[16] are shown in Figure 4. For each panel, the darker colored, top ensembles are the results of using the nonparametric model, and the lighter colored, bottom ensembles are the parametric model results with number of Gaussians indicated. For all four samples, convergence is achieved for both models, and the distance distributions show good agreement to those previously published.[43] The uncertainty is relatively low, given the small scatter of the posterior predictive distributions. Uncertainty increases at long distances but does not affect interpretation of the primary features. The parametric results for all site pairs show good agreement with the nonparametric results. The presence of peaks with high uncertainty in the parametric models in Figure 4A–C is indicative of limitations of the parametric approach, namely, imposing an underlying shape to the distance distribution. The multi-Gauss parametric model has more difficulty recovering distributions that have multiple overlapping peaks of similar width or intensity. Distributions of this nature show larger correlations between distribution parameters, making exploring the parameter space significantly less efficient and convergence more difficult to achieve. This was shown previously when analyzing synthetic data.[16]

### 4.3. Dependence on Distance Axis.

A nonparametric $P(r)$ is not entirely free of parameters—it depends on the fixed parameters that define the distance axis. These are the minimum distance $r_{\min}$, the maximum distance, $r_{\max}$, and, for a linear axis, the resolution $\Delta r$. This forces the distribution to be zero outside the distance range and imposes a fixed resolution within the range. Figure 5 shows a series of results for the midrange YopO data set with different $r_{\max}$ at a constant resolution $\Delta r$. Two effects of $r_{\max}$ can be discerned. First, as $r_{\max}$ is decreased, the uncertainty in the region near $r_{\max}$ shrinks. Second, it becomes less arduous to obtain converged chains (see Figure 5 caption), and the resulting posteriors indicate more efficient sampling and certainty around the parameters. When run for 20,000 samples per chain, the MCMC sampler was only able to converge for the model with an $r_{\max}$ of 6.5 nm. The model with an $r_{\max}$ of 8.5 nm (which is past heuristic values often used for $r_{\max}$, $\sqrt[3]{Dt_{\text{end}}/2\pi n_{\min}}$, where $n_{\min}$ is the number of required dipolar oscillation periods ($n_{\min} > 1/2$)),[43] did not attain convergence, as evidenced by $\hat{R}$ values much greater than 1.05 even when run for 100,000 samples per chain.

Both effects are a consequence of the fact that the models with long $r_{\max}$ are overspecified—there is not enough

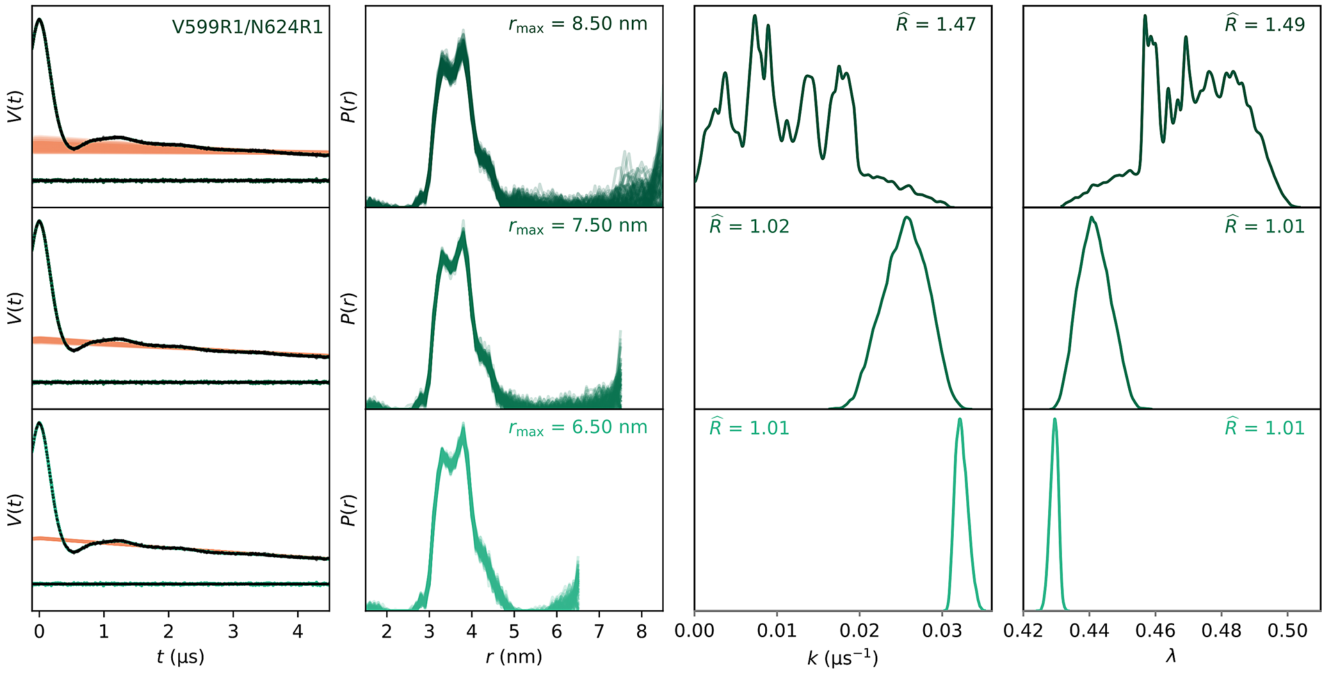


**Figure 5.** Effect of upper distance limit $r_{max}$ on quality of fit and $P(r)$ uncertainty. Data from the midrange site pair (YopO V599R1/N624R1) were evaluated for three values of $r_{max}$ and a constant resolution ($\Delta r$) of 0.1 nm, using MCMC with four chains and 20,000 samples (for $r_{max}$ = 6.5 nm) or 100,000 samples ($r_{max}$ = 7.5 and 8.5 nm). One chain was then dropped from the 7.5 nm model to attain convergence. The 8.5 nm model did not converge regardless of how many chains were dropped, so all four chains were kept. From these runs, 100 distance distributions are randomly sampled and plotted. The raw experimental data are shown in gray overlaid with the associated time-domain fits. The saturation of the color decreases with $r_{max}$. Left: time-domain fits and the resulting distance distributions. Right: the posterior distributions for the background decay rate constant $k$ and the modulation depth $\lambda$ with decreasing $r_{max}$.

information in the data to clearly distinguish between mass in this region of $r$ and the intermolecular background. The marginal distributions of the background parameter $k$ and the modulation depth $\lambda$ broaden substantially as $r_{max}$ is increased. These effects are also noticeable in the accompanying increased spread of background fits. Note that these effects are not a peculiarity of the MCMC method but rather an intrinsic parameter nonidentifiability issue of the model given the data. Any analysis method that uses this model will encounter similar difficulties in identifying unique parameter values. Based on the posterior distributions, the Bayesian approach provides a direct way of diagnosing these situations.

Several approaches can be considered to prevent model overspecification: (a) restrict $r_{max}$ to shorter values, (b) include additional information that $P(r)$ is close to zero at long distances, (c) include additional information that $P(r)$ is compact,[42] or (d) use a less flexible model with stronger assumptions about the $r$ distribution, such as a multi-Gauss model or a model with significantly fewer distance points.[45]

### 4.4. Comparison with Bootstrapping.

The Bayesian analysis presented here provides full quantification of the uncertainty for all model parameters, including correlations. Another, although conceptually different, approach is bootstrapping, a Monte Carlo resampling method. In the bootstrapping variant implemented in DeerLab,[17] synthetic data traces are generated by adding different noise realizations to a fitted signal obtained by least-squares fitting. These new traces are then analyzed according to the same procedure used for the original experimental data. This results in several fitted parameter sets and distance distributions. The distribution of fitted values compares in nature to the posterior output from Bayesian inference and can be randomly sampled from to produce ensemble plots representing the spread of uncertainty around a particular set of data.

In Figure 6, we show the Bayesian analysis using a nonparametric model of the YopO data from above and compare it to the data analyzed with bootstrapping. Compared to our previous figures, the distance axis resolution was doubled to 0.05 nm, and $r_{max}$ was set to 6.5 nm. For each panel, the darker colored, top ensembles are the results of using Bayesian inference and the lighter colored, bottom ensembles are the bootstrapping results. For bootstrapping, an initial fit was achieved using a regularization parameter selected by the Bayesian information criterion.[15] This value of $\alpha$ was then frozen for the bootstrap analysis wherein 1000 bootstrapped samples were taken, i.e., 1000 new signal traces were generated and fit. A set of 100 parameter vectors and distributions were randomly drawn from the 1000 samples and plotted alongside the Bayesian inference ensembles.

For all of the site pairs, the ensembles of distance distributions are very similar between Bayesian analysis and bootstrapping (see Figure 6). However, the bootstrap ensembles generally have less scatter, since all synthetic signals are generated from the same initial fit, leading to less exploration of the combined parameter space compared to the Bayesian approach.

A crucial difference between Bayesian inference and bootstrapping is that in the latter method, the Tikhonov smoothing parameter $\alpha$ is fixed, whereas it is a floating (hyper)parameter in the former. Therefore, bootstrapping does not incorporate the uncertainty due to $\alpha$. However, the value and uncertainty of $\alpha$ are crucial components for assessing whether over- or underfitting is occurring and provide insights into the overall shape of the resulting distance distribution.

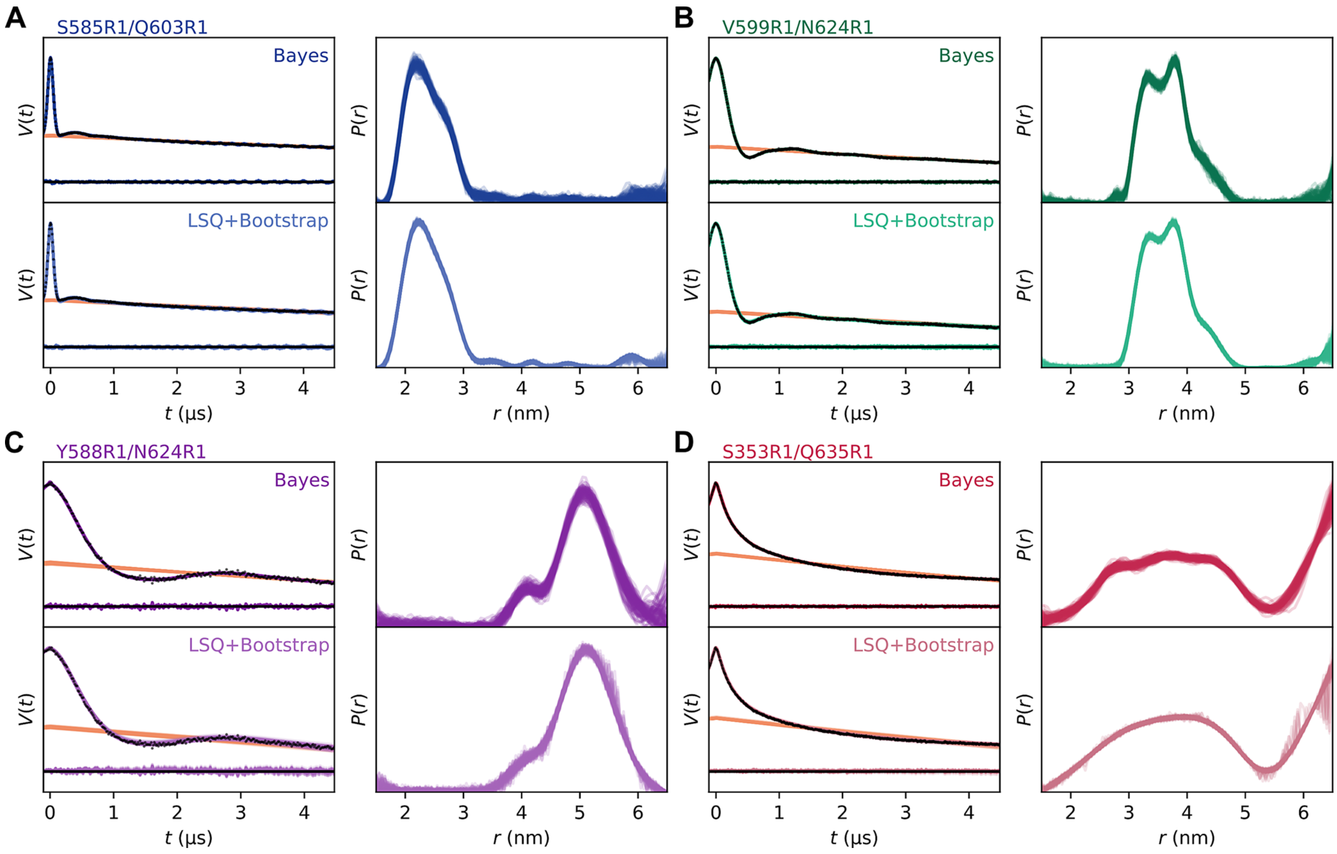


**Figure 6.** (A–D) Comparison of experimental fits and visualization of uncertainty between Bayesian inference and bootstrapping. The MCMC data shown were run with four chains of 100,000 samples per chain. For both the Bayesian MCMC data and the bootstrapping fits, 100 samples were randomly selected to plot. The raw experimental data are shown in gray overlaid with the time-domain full and intermolecular fits for the Bayesian analysis (top, dark) and bootstrapping (bottom, light). The distance distributions for each are shown according to the same color scheme.

With Bayesian inference, the uncertainty of $\alpha$, along with all other model parameters, is the direct output of the analysis. This is illustrated in Figure 7, which plots the marginalized posterior of $\lg(\alpha)$ in the Bayesian analysis against the fixed value of $\lg(\alpha)$ used in the bootstrap analysis (vertical line).

In terms of computational cost, the time taken to run 1,000 bootstrapped samples for the bootstrap analysis was similar to the time taken to run approximately 100,000 MCMC draws for the Bayesian analysis (a couple hours on a typical laptop computer). Thus, the Bayesian approach yields a more complete uncertainty analysis at roughly the same computational cost. Note, however, that computational performance depends on implementation, and there is a wide range of possible bootstrapping approaches beyond the one implemented in DeerLab.

## 5. CONCLUSIONS

The Bayesian method outlined in this work presents a rigorous, complete, and conceptually simple inference approach for analyzing DEER data. Given the experimental data and the choice of a particular physical model, the joint probability distribution of all model parameters is determined. This provides, completely, any information that can be gleaned from the data under the assumption of the chosen model and provided prior information. If the information content of the data is low (noisy and truncated trace), then this method captures the resulting significant uncertainty about the model parameters, particularly if a Tikhonov model is used. Introducing additional constraints to the $r$ distribution, such as by using a multi-Gaussian model, could be advantageous, at the expense of biasing the analysis.

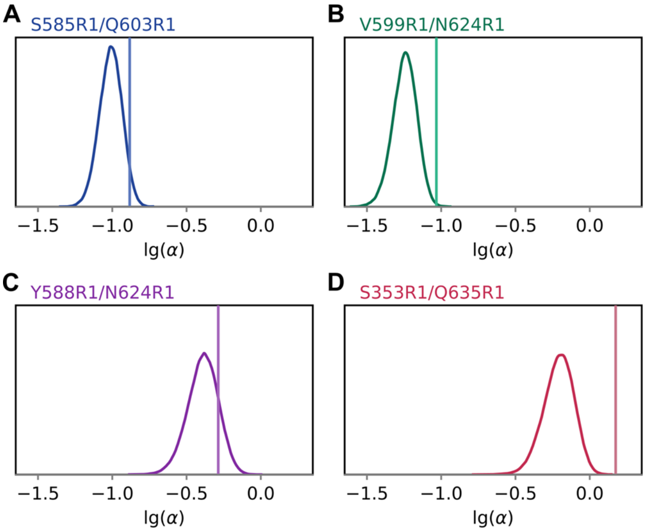


**Figure 7.** (A–D) Uncertainty assessment for Tikhonov smoothness parameter $\alpha$. The posterior distributions for $\lg(\alpha)$ from the Bayesian analysis are shown, maintaining the same panel order and coloring as those of previous figures for the site pairs. The single values for the $\lg(\alpha)$ parameter from the bootstrap fits are shown as vertical lines.

The complete uncertainty quantification provided by Bayesian inference has advantages over other forms of uncertainty quantification. Compared with confidence intervals based on maximum-likelihood estimation and the Laplace approximation (which assumes the likelihood function is Gaussian around its maximum), it is more complete and accurate and can capture asymmetric uncertainty as well as broad areas in the parameter space that leads to equally good model fits. Therefore, it can help spot identifiability problems with the model parameters. Compared to bootstrapping, which can be used to determine more comprehensive confidence intervals within the maximum-likelihood estimation approach, Bayesian inference does not synthetically generate new data sets. The only data set that is used in Bayesian inference is the given experimental data set. From a principled standpoint, this is indeed the only data set that should matter. Also, the Bayesian approach saves significant computational time on a per-sample basis. In general, the logically rigorous foundation of the Bayesian inference approach makes it possible to use it as a reference method for analyzing dipolar EPR data.

Although we presented the method using a Tikhonov model for the intramolecular distance distribution combined with a simple exponential decay for the intermolecular contribution, this method is very general and can handle more complex models. Extensions to use more sophisticated intermolecular models, to include multiple dipolar pathways, and to handle multiple data sets simultaneously are conceptually straightforward, although computationally more demanding. Therefore, the Bayesian framework provides a robust and flexible tool for DEER data analysis, providing complete quantitative information about the uncertainty.

## AUTHOR INFORMATION

### Corresponding Author

**Stefan Stoll** − *Department of Chemistry, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0000-0003-4255-9550; Email: stst@uw.edu

### Authors

**Sarah R. Sweger** − *Department of Chemistry, University of Washington, Seattle, Washington 98195, United States*

**Julian C. Cheung** − *Department of Chemistry, University of Washington, Seattle, Washington 98195, United States;* orcid.org/0009-0005-9534-7158

**Lukas Zha** − *Department of Chemistry, University of Washington, Seattle, Washington 98195, United States*

**Stephan Pribitzer** − *Department of Chemistry, University of Washington, Seattle, Washington 98195, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.4c05056

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation with grant CHE-1452967 (S.S.) and the National Institutes of Health with grants R01 GM125753 and R01 GM127325 (S.S.). We thank Jonathan Bardsley (University of Montana) for helpful discussions.

## REFERENCES

(1) Milov, A. D.; Salikhov, K. M.; Shchirov, M. D. Application of the Double Resonance Method to Electron Spin Echo in a Study of the Spatial Distribution of Paramagnetic Centers in Solids. *Sov. Phys. Solid State* **1981**, *23*, 565−569.

(2) Milov, A. D.; Tsvetkov, Y. D. Double Electron−Electron Resonance in Electron Spin Echo: Conformations of Spin-Labeled Poly-4-Vinylpyridine in Glassy Solutions. *Appl. Magn. Reson.* **1997**, *12*, 495−504.

(3) Pannier, M.; Veit, S.; Godt, A.; Jeschke, G.; Spiess, H. W. Dead-Time Free Measurement of Dipole−Dipole Interactions between Electron Spins. *J. Magn. Reson.* **2000**, *142*, 331−340.

(4) Jeschke, G.; Polyhach, Y. Distance Measurements on Spin-Labelled Biomacromolecules by Pulsed Electron Paramagnetic Resonance. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1895−1910.

(5) Jeschke, G. DEER Distance Measurements on Proteins. *Annu. Rev. Phys. Chem.* **2012**, *63*, 419−446.

(6) Matveeva, A. G.; Nekrasov, V. M.; Maryasov, A. G. Analytical Solution of the PELDOR Inverse Problem Using the Integral Mellin Transform. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32381−32388.

(7) Worswick, S. G.; Spencer, J. A.; Jeschke, G.; Kuprov, I. Deep Neural Network Processing of DEER Data. *Sci. Adv.* **2018**, *4*, No. eaat5218.

(8) Keeley, J.; Choudhury, T.; Galazzo, L.; Bordignon, E.; Feintuch, A.; Goldfarb, D.; Russell, H.; Taylor, M. J.; Lovett, J. E.; Eggeling, A.; et al. Neural networks in pulsed dipolar spectroscopy: A practical guide. *J. Magn. Reson.* **2022**, *338*, No. 107186.

(9) Jeschke, G.; Koch, A.; Jonas, U.; Godt, A. Direct Conversion of EPR Dipolar Time Evolution Data to Distance Distributions. *J. Magn. Reson.* **2002**, *155*, 75−82.

(10) Chiang, Y.-W.; Borbat, P. P.; Freed, J. H. Maximum Entropy: A Complement to Tikhonov Regularization for Determination of Pair Distance Distributions by Pulsed ESR. *J. Magn. Reson.* **2005**, *177*, 184−196.

(11) Jeschke, G.; Chechik, V.; Ionita, P.; Godt, A.; Zimmermann, H.; Banham, J.; Timmel, C. R.; Hilger, D.; Jung, H. DeerAnalysis2006—A Comprehensive Software Package for Analyzing Pulsed ELDOR Data. *Appl. Magn. Reson.* **2006**, *30*, 473−498.

(12) Sen, K. I.; Logan, T. M.; Fajer, P. G. Protein Dynamics and Monomer-Monomer Interactions in AntR Activation by Electron Paramagnetic Resonance and Double Electron-Electron Resonance. *Biochemistry* **2009**, *46*, 11639−11649.

(13) Brandon, S.; Beth, A. H.; Hustedt, E. J. The Global Analysis of DEER Data. *J. Magn. Reson.* **2012**, *218*, 93−104.

(14) Stein, R. A.; Beth, A. H.; Hustedt, E. J. A Straightforward Approach to the Analysis of Double Electron−Electron Resonance Data. *Methods Enzymol.* **2015**, *563*, 531−567.

(15) Edwards, T. H.; Stoll, S. Optimal Tikhonov Regularization for DEER Spectroscopy. *J. Magn. Reson.* **2018**, *288*, 58−68.

(16) Sweger, S. R.; Pribitzer, S.; Stoll, S. Bayesian Probabilistic Analysis of DEER Spectroscopy Data using Parametric Distance Distribution Models. *J. Phys. Chem. A* **2020**, *124*, 6193−6202.

(17) Fábregas Ibáñez, L.; Jeschke, G.; Stoll, S. DeerLab: A Comprehensive Toolbox for Analyzing Dipolar EPR Spectroscopy Data. *Magn. Reson.* **2020**, *1*, 209−224.

(18) Hustedt, E. J.; Marinelli, F.; Stein, R. A.; Faraldo-Gómez, J. D.; Mchaourab, H. S. Confidence Analysis of DEER Data and its Structural Interpretation with Ensemble-Biased Metadynamics. *Biophys. J.* **2018**, *115*, 1200−1216.

(19) Bretthorst, G. L. Bayesian analysis. III. Applications to NMR signal detection, model selection, and parameter estimation. *J. Magn. Reson.* **1990**, *88*, 571−595.

(20) Hao, J.; Astle, W.; De Iorio, M.; Ebbels, T. M. D. BATMAN−an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. *Bioinformatics* **2012**, *28*, 2088−2090.

(21) Ikeya, T.; Ikeda, S.; Kigawa, T.; Ito, Y.; Güntert, P. Protein NMR structure refinement based on Bayesian inference. *J. Phys.: Conf. Ser.* **2016**, *699*, No. 012005.

(22) Edwards, T. H.; Stoll, S. A Bayesian Approach to Quantifying Uncertainty from Experimental Noise in DEER Spectroscopy. *J. Magn. Reson.* **2016**, *270*, 87−97.

(23) De Lorenzi, F.; Weinmann, T.; Bruderer, S.; Heitmann, B.; Henrici, A.; Stingelin, S. Bayesian analysis of 1D $^1$H-NMR spectra. *J. Magn. Reson.* **2024**, *364*, No. 107723.

(24) Altenbach, C. LongDistances. https://sites.google.com/site/altenbach/labview-programs/epr-programs/long-distances.

(25) Fábregas-Ibáñez, L.; Tessmer, M. H.; Jeschke, G.; Stoll, S. Dipolar Pathways in Dipolar EPR Spectroscopy. *Phys. Chem. Chem. Phys.* **2022**, *24*, 2504−2520.

(26) Kattnig, D. R.; Reichenwallner, J.; Hinderberger, D. Modeling Excluded Volume Effects for the Faithful Description of the Background Signal in Double Electron−Electron Resonance. *J. Phys. Chem. B* **2013**, *117*, 16542−16557.

(27) Bardsley, J. M.; Fox, C. An MCMC Method for Uncertainty Quantification in Nonnegativity Constrained Inverse Problems. *Inverse Probl. Sci. Eng.* **2012**, *20*, 477−498.

(28) Bardsley, J. M.; Hansen, P. C. MCMC Algorithms for Computational UQ of Nonnegativity Constrained Linear Inverse Problems. *SIAM J. Sci. Comput.* **2020**, *42*, A1269−A1288.

(29) Robert, C. P.; Casella, G. *Monte Carlo Statistical Methods*; Springer: New York, 2010.

(30) Bro, R.; De Jong, S. A fast non-negativity-constrained least squares algorithm. *J. Chemom.* **1997**, *11*, 393−401.

(31) Moré, J. J.; Toraldo, G. On the Solution of Large Quadratic Programming Problems with Bound Constraints. *Siam J. Control Optim.* **1991**, *1*, 93−113.

(32) Van Benthem, M. H.; Keenan, M. R. Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. *J. Chemom.* **2004**, *18*, 441−450.

(33) Hoffman, M. D.; Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593−1623.

(34) Neal, R. M. *MCMC Using Hamiltonian Dynamics*; CRC Press: Boca Raton, 2011.

(35) Betancourt, M. *A Conceptual Introduction to Hamiltonian Monte Carlo*, 2018; https://arxiv.org/abs/1701.02434.

(36) Salvatier, J.; Wiecki, T. V.; Fonnesbeck, C. Probabilistic Programming in Python Using PyMC3. *Peer J. Comput. Sci.* **2016**, *2*, No. e55.

(37) Gelman, A.; Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **1992**, *7*, 457−472.

(38) Brooks, S. P.; Gelman, A. General Methods for Monitoring Convergence of Iterative Simulations. *J. Comput. Graph. Stat.* **1998**, *7*, 434−455.

(39) Vehtari, A.; Gelman, A.; Simpson, D.; Carpenter, B.; Bürkner, P.-C. Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC. *Bayesian Anal.* **2021**, *1*, 667−718.

(40) Engelhardt, B. E.; Adams, R. P. *Bayesian Structured Sparsity from Gaussian Fields*, 2014; https://arxiv.org/abs/1407.2235.

(41) Edwards, T. H.; Stoll, S. *Synthetic Test Data Set for DEER Spectroscopy Based on T4 Lysozyme*, 2018

(42) Fábregas-Ibáñez, L.; Jeschke, G.; Stoll, S. Compactness regularization in the analysis of dipolar EPR spectroscopy. *J. Magn. Reson.* **2022**, *339*, No. 107218.

(43) Schiemann, O.; Heubach, C. A.; Abdullin, D.; Ackermann, K.; Azarkh, M.; Bagryanskaya, E. G.; Drescher, M.; Endeward, B.; Freed, J. H.; Galazzo, L.; et al. Benchmark test and guidelines for DEER/PELDOR experiments on nitroxide-labeled biomolecules. *J. Am. Chem. Soc.* **2021**, *143*, 17875−17890.

(44) Peter, M. F.; Tuukkanen, A. T.; Heubach, C. A.; Selsam, A.; Duthie, F. G.; Svergun, D. I.; Schiemann, O.; Hagelueken, G. Studying conformational changes of the Yersinia type-III-secretion effector YopO in solution by integrative structural biology. *Structure* **2019**, *27*, 1416−1426.

(45) Dzuba, S. A. The determination of pair-distance distribution by electron−electron resonance: regularization by the length of distance discretization with Monte Carlo calculations. *J. Magn. Reson.* **2016**, *269*, 113−119.