Optimal Tikhonov regularization for DEER spectroscopy

Thomas H. Edwards, Stefan Stoll *

Department of Chemistry, University of Washington, Seattle, WA 98103, United States

ARTICLE INFO

ABSTRACT

Tikhonov regularization is the most commonly used method for extracting distance distributions from experimental double electron-electron resonance (DEER) spectroscopy data. This method requires the selection of a regularization parameter, α, and a regularization operator, L. We analyze the performance of a large set of α selection methods and several regularization operators, using a test set of over half a million synthetic noisy DEER traces. These are generated from distance distributions obtained from *in silico* double labeling of a protein crystal structure of T4 lysozyme with the spin label MTSSL. We compare the methods and operators based on their ability to recover the model distance distributions from the noisy time traces. The results indicate that several α selection methods perform quite well, among them the Akaike information criterion and the generalized cross validation method with either the first- or second-derivative operator. They perform significantly better than currently utilized L-curve methods.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Double electron–electron resonance (DEER) spectroscopy, also called pulsed electron–electron double resonance (PELDOR) spectroscopy, measures the magnetic dipolar coupling between two or more paramagnetic centers, such as spin labels attached to proteins [1–3]. DEER data analysis usually involves the removal of a background signal followed by a transformation of the oscillatory time-domain signal into a distance-domain distribution function describing the distances between nearby paramagnetic centers (1.5–10 nm).

There exist several different approaches for extracting distance distributions from DEER data: Tikhonov regularization [4–7], Gaussian mixture models [8–10], Tikhonov regularization postprocessed with Gaussians [11,12], Tikhonov regularization combined with maximum entropy [13], Bayesian inference (based upon Tikhonov regularization) [14], regularization by limiting the number of points in the distance domain [15], wavelet denoising [16], truncated singular-value decomposition [6,17], and neural networks [18]. Among them, Tikhonov regularization is the most widely employed method.

In this paper, we are concerned with the determination of optimal settings for Tikhonov regularization. This involves the choice of a regularization operator L and of a value for the regularization parameter α. An optimal choice of L and α ensures good distance distribution recovery and prevents overfitting the data; a bad choice causes poor recovery and either under- or overfitting to the data. There are several operators to choose from, and many methods are available for selecting α, each based on a defensible rationale. However, they vary greatly both in terms of theoretical justification and empirical track record. Therefore, the selection of the method/operator combination ought to be based on a thorough comparison of their performance for a practically relevant benchmark set of data analysis problems.

Tikhonov regularization was introduced to NMR for de-Pake-ing [19], for the extraction of internuclear distances from dipolar time-domain signals such as those from REDOR [20], for the determination of orientational distributions from ^{2}H NMR data [21,22], and for the calculation of relaxation rate distributions [23,24]. These approaches used the self-consistent method for selecting α, as introduced and implemented in the program FTIKREG [25,26]. In the context of extracting distance distributions from DEER data, Tikhonov regularization was initally mentioned in 2002 [27,28], and first applications appeared in 2004 [4,5]. In these papers, the regularization parameter was selected manually or using FTIKREG. A thorough paper examining Tikhonov regularization and introducing the use of the L-curve maximum-curvature criterion appeared in 2005 [6]. A different L-curve method, the minimum-radius criterion, was introduced in 2006 in the program DeerAnalysis [7] and is used in its current release (2016).

Despite the long history and the widespread use of Tikhonov regularization for DEER data analysis, there has been no systematic assessment and efficiency comparison of regularization operators and regularization parameter selection methods. This is what we present here. We evaluate a large number of regularization param-

* Corresponding author.

E-mail addresses: edwardst@uw.edu (T.H. Edwards), stst@uw.edu (S. Stoll).

eter selection methods and three regularization operators with respect to their performance against a large set of synthetic DEER time-domain data derived from model distributions obtained by *in silico* spin labeling of a PDB crystal structure of T4 lysozyme. We examine and compare all methods by their efficiency in recovering the underlying model distributions. The results indicate that the performance of the methods varies significantly, and that there are methods that perform better than the currently employed L-curve α selection methods.

The paper is structured as follows. Section 2 summarizes the principles of Tikhonov regularization for DEER and introduces the necessary notation. Section 3 explains the construction and characteristics of the large test set of distance distributions and associated synthetic DEER time-domain traces. Section 4 describes the assessment methodology. Section 5 presents the results and discusses benefits and drawbacks of various methods in light of these results. The paper ends with recommendations regarding the preferred methods. Mathematical details of the various regularization parameter selection methods are given in the Appendix.

2. DEER and Tikhonov regularization

DEER measures the electron spin echo intensity V as a function of the position in time of one or more pump pulses, t. For dilute samples of doubly-labelled proteins or complexes, V is a product of an overall amplitude, V_0, a modulation function due to intra-complex coupling, F, and a background modulation function due to the interaction between spins on different complexes, B:

$$V(t) = V_0 \cdot F(t) \cdot B(t) = V_0 \cdot ((1 - \lambda) + \lambda S(t)) \cdot B(t) \quad (1)$$

λ is the modulation depth parameter related to the excitation efficiency of the DEER experiment. When utilizing Tikhonov regularization, a background model is typically fitted to $V(t)/V_0$ and divided out, yielding the isolated intra-complex modulation function $S(t)$ after scaling by λ. $S(t)$ is normalized such that $S(0) = 1$ in the noise-free limit.

In the absence of orientation selection, exchange coupling, differential relaxation, and multi-spin effects, $S(t)$ is related to the distance distribution between the two spins on the complex, $P(r)$, by

$$S(t) = \int_0^\infty dr\, K(t, r) P(r) \quad (2)$$

where

$$K(t, r) = \int_0^{\pi/2} d\theta \sin\theta \cos\left[(1 - 3\cos^2\theta)\omega_\perp t\right] \quad (3)$$

and $\omega_\perp = \mu_0 \mu_B^2 g_A g_B / (4\pi\hbar r^3)$. $P(r)$ integrates to one: $\int_0^\infty dr\, P(r) = 1$.

When S is discretized as an n_t-element vector with elements $S_i = S(t_i)$, and P as an n_r-element vector with elements $P_j = P(r_j)$, the integral transformation from P to S is represented by the matrix–vector multiplication

$$S = KP \quad (4)$$

where K is the $n_t \times n_r$ kernel matrix. Its elements are $K_{ij} = K(t_i, r_j) \cdot \Delta r$, with the distance increment Δr. Usually, n_r is set equal to n_t.

The matrix K is close to singular, and the calculation of the inverse of $K^T K$—needed to obtain P from S in an ordinary least-squares fitting procedure—is very inaccurate. To enable the solution, Tikhonov regularization is used:

$$P_{\alpha L} = \operatorname*{argmin}_{P \geq 0} \left(\|S - KP\|^2 + \alpha^2 \|LP\|^2 \right) \quad (5)$$

This is a form of penalized least-squares fitting. The first term is the least-squares term capturing the misfit between the model KP

and the data S. The second term penalizes for unwanted properties of the solution P and depends on a specific form for the regularization operator L and a specific value for the regularization parameter α. The fit is constrained by the requirement that all elements of P be non-negative ($P \geq 0$). The subscripts in $P_{\alpha L}$ indicate that the solution depends on the particular choices of α and L.

L defines the criterion by which P should be penalized. Physically reasonable distance distributions between spin labels on proteins are smooth on a tenths-of-nanometer scale. Three L choices are possible that all encourage smoothness and penalize roughness of P, in one sense or another. The second derivative, represented by the $(n_r - 2) \times n_r$ second-order difference matrix

$$L_2 = \begin{pmatrix} 1 & -2 & 1 & & & 0 \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ 0 & & & 1 & -2 & 1 \end{pmatrix} \quad (6)$$

penalizes sharp turns in the distribution, which arise from sharp peaks. The first derivative, represented by the $(n_r - 1) \times n_r$ first-order difference matrix

$$L_1 = \begin{pmatrix} -1 & 1 & & & 0 \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ 0 & & & -1 & 1 \end{pmatrix} \quad (7)$$

penalizes steep slopes, which are also associated with sharp peaks. Note that the matrices defined in Eqs. (6) and (7) do not contain Δr. Lastly, the $n_r \times n_r$ identity matrix $L_0 = I$ can be used. It penalizes tall peaks, which tend to be narrow due to the overall normalization of P.

The value of α determines the balance between the two terms in Eq. (5). Large α values lead to oversmoothing of P and a poor fit to the time-domain data S, while small α values lead to unrealistically spiky distributions (undersmoothing) and overfitting of the data. Therefore, a proper choice of α is essential. A large number of α selection methods are described in the literature, and we include many of them in our performance analysis. They are described in the Appendix.

The optimal values of both α and L, α_{opt} and L_{opt}, are the ones which together best recover the underlying true distribution P_0, i.e. the ones for which the model recovery error (mre), defined as

$$\text{mre}(\alpha, L) = \|P_{\alpha L} - P_0\| \quad (8)$$

is minimal. For synthetic data, as used in this study, this error and thereby α_{opt} and L_{opt} can be determined, since P_0 is given. However, in experimental practice P_0 is not known, and one has to resort to other methods to choose a form for L and select a value for α.

3. The test set

In order to test α selection methods and regularization operators on DEER data from practically relevant distance distributions, we numerically generated a large set of synthetic noisy DEER time-domain signals based on a crystal structure of T4 lysozyme (PDB ID 2LZM, 1.7 Å resolution); its structure is shown in Fig. 1. We use this protein since it is currently one of the most thoroughly investigated proteins by DEER and other EPR techniques [29–37] and is of a size commensurate with the typical DEER distance range of 1.5–6 nm. We *in silico* labeled the 2LZM protein structure with the nitroxide spin label MTSSL which is to date, by a wide margin, the most commonly used label for DEER. In this section, we detail the construction of the test set.
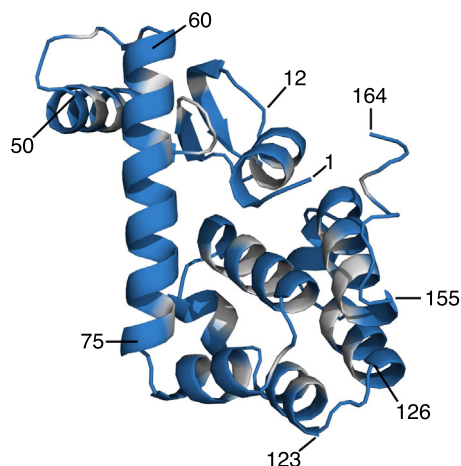
Fig. 1. Crystal structure of T4 lysozyme (164 residues, PDB 2LZM, 1.7 Å resolution). Residues with low predicted labeling probabilities that were excluded from the test set are shown in grey. Several residues are labeled as guides to the eye.

3.1. Construction of model distributions

Using scripts based on MMM 2017.1 [38–40] with the default 216-member rotamer library R1A_298K_UFF_216_r1_CASD for MTSSL at ambient temperature, we calculated rotamer distributions for all 164 sites on the protein. For each site, this yielded coordinates of the mid-points of the N—O bonds of all 216 rotamers and their associated populations, as well as the site partition function. In order to keep the test set realistic, we eliminated all sites with low predicted labelling probability (partition function smaller than 0.05), leaving 129 sites. These sites are indicated in Fig. 1.

For each of the resulting 8256 site pairs, we calculated the associated model distance distribution, consisting of a sum of Gaussian lineshapes, with centers corresponding to the inter-rotamer distances, intensities corresponding to the products of the respective populations, and with a uniform full width at half maximum

(FWHM) of 0.15 nm to account for structural heterogeneity around the energy minimum of each rotamer due to librational motion [38]. Each distribution was discretized with a high resolution of $\Delta r = 0.005$ nm, corresponding to 1/30 of the FWHM. To keep the test set experimentally realistic, we discarded all distributions with more than 5% integrated population below 1.7 nm. This guarantees that there is no population below 1.5 nm. Site pairs with population at such short distances are always avoided in experimental studies in order to avoid potential distortions due to exchange coupling and incomplete excitation.

This procedure resulted in a set of 5622 model distance distributions. Their properties are summarized in Fig. 2. The distribution modes (location of the distribution maxima) range from 1.7 to 6.1 nm, and the inter-quartile ranges (iqr; range of central 50% integrated population) are between 0.09 and 1.1 nm wide. Long, short, narrow, wide, unimodal, multimodal, symmetric, and skewed distributions are all well-represented in the test set. Skewness is relatively evenly distributed across both the mode and iqr. The number of significant peaks is evenly distributed across modal distance, but there is a positive correlation between iqr and number of peaks. Fig. 3 shows several representative example distributions.

An alternative strategy of constructing a test set would be to generate completely artificial distributions without reference to a protein where the center, width, multi-modality, and skewness is varied either systematically or randomly. However, such a test set will be less experimentally relevant. Fig. 2 shows that the structure-based test set used here covers a sufficiently diverse range of distribution centers, widths, and shapes.

3.2. Generation of time-domain traces

Each model distribution was used to generate a set of time-domain traces S with varying maximum evolution times t_{max}, time increments Δt, noise levels σ, and noise realizations, subject to the constraints that: (1) each S must be long enough to permit accurate recovery of the longest distances in the corresponding P, and (2) the sampling rate must be high enough to recover the shortest distances. The starting time for all time traces was set to $t_{min} = 0$. We
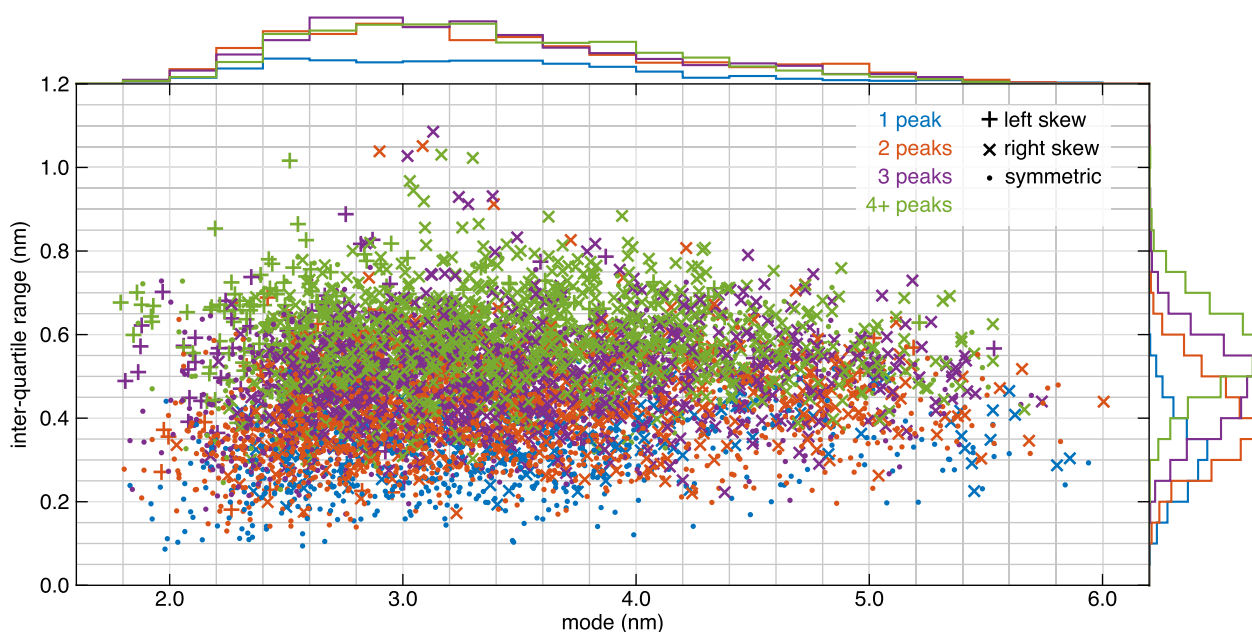
Fig. 2. Summary statistics of the 5622 model DEER distance distributions derived from the crystal structure of T4 lysozyme (PDB 2LZM). The inter-quartile range (iqr) of r is plotted vs the mode of r for each distribution. Color indicates the number of peaks (significant maxima), and the symbol indicates the type and degree of skew. The histogram colors correspond to the number of peaks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
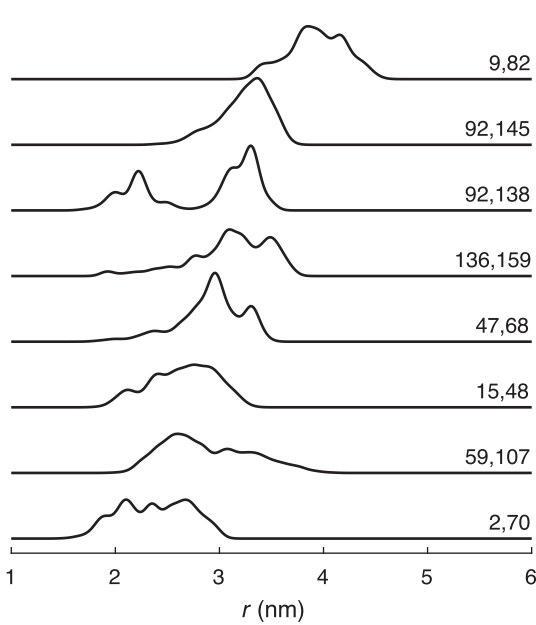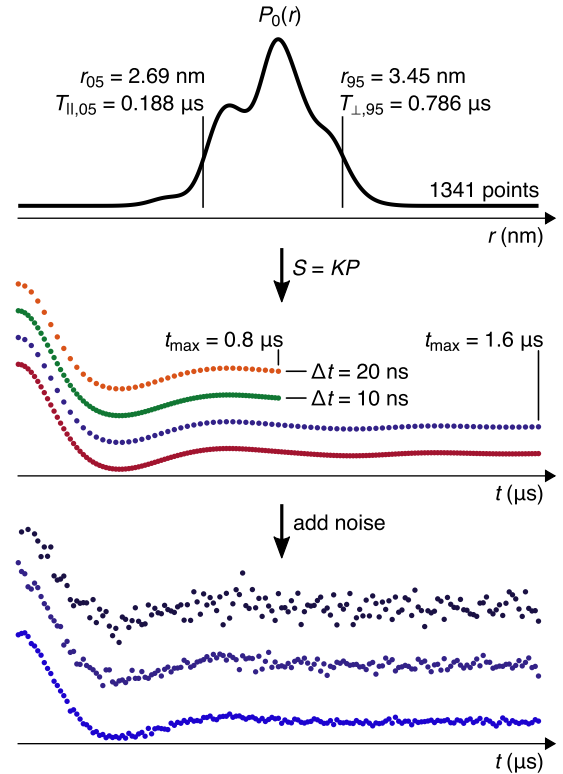
Fig. 3. Examples of distance distributions calculated via *in silico* double labeling of the 2LZM structure of T4 lysozyme with MTSSL. The site pairs corresponding to each distribution are listed in the figure.

Fig. 4. Generation of time traces ($S(t)$) from a distance distribution. Top panel: the model distribution (for labeled residues 112 and 143). In this case, $3T_{\perp,95} = 2.36\,\mu s$, so t_{max} was set to 1.6 and 0.8 μs in accordance with the rules given in Section 3.2. $T_{\parallel,05}/6 = 31.3$ ns, so Δt was set to 10 and 20 ns. Middle panel: the four resulting time-domain traces generated using Eq. (4). Bottom panel: three noisy realizations with σ = 0.1, 0.05, and 0.02 for the 3rd (dark blue) trace in the middle panel. A total of 120 noisy time-domain traces were generated from this particular distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

used multiple values of t_{max} and Δt for each model distribution in order to be able to assess whether or not any α selection methods display uneven performance with respect to these parameters.

To choose a set of appropriate t_{max} for a given distribution, we take as an upper limit $t_{mm} = 3T_{\perp,95}$, where $T_{\perp,95}$ is the period corresponding to the perpendicular-orientation dipolar frequency of the 95-percentile distance, i.e., $T_{\perp,95} = (r_{95}/nm)^3/(52\,MHz)$. The factor 3 prolongs the time in order to completely capture the long-distance tail of the distribution, which has the lowest frequencies. With this, we choose the set of t_{max} according to the following procedure: if $t_{mm} > 6.4\,\mu s$, then t_{max} is 6.4 and 3.2 μs; if $3.2\,\mu s < t_{mm} < 6.4\,\mu s$, then t_{max} is 3.2 and 1.6 μs; if $1.6\,\mu s < t_{mm} < 3.2\,\mu s$, then t_{max} is 1.6 and 0.8 μs; if $0.8\,\mu s < t_{mm} < 1.6\,\mu s$, then t_{max} is 0.8 μs; and if $t_{mm} < 0.8\,\mu s$, then t_{max} is 0.4 μs. Fig. 4 shows an example distribution for which $t_{mm} = 2.36\,\mu s$, hence t_{max} is 1.6 and 0.8 μs. Note that even the shorter t_{max} value of 0.8 μs is long enough to accommodate a full oscillation of period $T_{\perp,95} = 0.786\,\mu s$.

To find a reasonable set of time increments Δt for each distribution, we first determined the longest allowable increment, Δt_{max}, that correctly samples all frequencies in the signal. This is dictated by the highest dipolar frequency in the time-domain signal, which corresponds to the shortest distance in the distribution. We use the period $T_{\parallel,05}$ of the parallel-orientation dipolar frequency corresponding to the 5-percentile distance. According to the Nyquist theorem, Δt_{max} must be less than $T_{\parallel,05}/2$ to avoid frequency aliasing. To capture the short-distance tail of the distribution and to obtain enough time-domain points to achieve reasonable distance resolution in the Tikhonov fits in all cases, we chose $\Delta t_{max} = T_{\parallel,05}/6$. This resulted in values from 8 ns to 280 ns. Δt_{max} now guides the choice of Δt: If $\Delta t_{max} > 200$ ns, then Δt is 50, 100, and 200 ns; if $100\,ns < \Delta t_{max} < 200\,ns$, then Δt is 20, 50, and 100 ns; if $50\,ns < \Delta t_{max} < 100\,ns$, then Δt is 10, 20, and 50 ns; if $20\,ns < \Delta t_{max} < 50\,ns$, then Δt is 10 and 20 ns; if $10\,ns < \Delta t_{max} < 20\,ns$, then Δt is 10 ns; and if $\Delta t_{max} < 10\,ns$, then Δt is 5 ns. For the example distribution in Fig. 4, $\Delta t_{max} = T_{\parallel,05}/6 = 31.3$ ns, so Δt was set to 10 and 20 ns.

To generate the noise-free time trace for a particular combination of t_{min}, t_{max}, and Δt (with n_t = 33, 41, 65, 81, 129, 161, 321, or 641 points), we used in Eq. (4) the high-resolution model distribu-

tion ($n_0 = 1341$ points) and generated the n_t-element time trace using an $n_t \times n_0$ kernel matrix. Therefore, the model distributions used for this forward modeling have a higher resolution than the ones obtained via Tikhonov regularization (Eq. (5)) of the simulated noisy data, with $n_r = n_t$. This avoids circular reasoning and committing the "inverse crime" [41,42]. Fig. 4 shows the synthetic time-domain traces resulting from the above rules for an example distribution.

The combinations of $\Delta t, t_{min}$, and t_{max} resulted in a total of 20701 noise-free time-domain traces with between 33 and 641 points. A few traces with fewer than 33 points were thrown out, since such short traces are not acquired in practice.

Finally, we generated 30 noisy traces from each noise-free trace. We utilized uncorrelated Gaussian noise [14] with standard deviations $\sigma = 0.02, 0.05$, and 0.1. These noise levels span the range typically obtained experimentally for proteins. In experimental settings, the background removal step typically increases the magnitude of the noise at the end of the $S(t)$ trace relative to the start. This effect is more pronounced for steeper background functions and can be negligible for shallower ones. We did not attempt to simulate this characteristic of the data. For each noise level, $m = 10$ noise realizations were generated using a random-number generator. For reproducibility, seeds for the random-number generator are stored as part of the test set.

Altogether, with these selections for $t_{max}, \Delta t, \sigma$, and m, the 5622 model distributions P_0 resulted in 621030 noisy time-domain traces S. These constitute the final test data set, which is available online (see Appendix B).

4. Performance evaluation

For each of the traces in the test set, we use 60 different variants of α selection methods to determine an appropriate regularization parameter value α_{sel} and the associated Tikhonov solution. Mathematical details of the methods are given in the Appendix. Each method is referred to via a short acronym, which is listed in Table 1. We evaluate each α selection method for each of the three regularization operators L (identity, first derivative, and second derivative). Thus, we examine Tikhonov regularization along both its degrees of freedom (L and α), with a total of 180 approaches.

Each method determines α by minimizing or thresholding some cost function over α. To search for these points, we use an α range from 10^{-3} to 10^{3} with 61 points on a logarithmic scale, i.e. 10 points per decade (1 dB increments). All methods require the data S as input and need to calculate $P_{\alpha L}$ over the entire α range for each L. For this, $P_{\alpha L}$ can be determined with or without the constraint $P \geq 0$ in Eq. (5), and we test both variants in all cases. Several methods require knowledge of the time-domain noise standard deviation, σ. The cost functions of a few methods are equipped with a tuning or scaling parameter that can be adjusted to increase stability.

The analysis procedure also requires the choice of a distance range and resolution for $P_{\alpha L}$. We use a range of 1.0–7.0 nm for all cases. These limits encompass the full ranges of all model distributions in the test set. The number of points in the distance domain, n_r, is determined separately for each case and is set equal to the number of points in the time-domain trace, n_t.

Since the goal of analyzing the time-domain data is to reveal the underlying model distribution, we base our performance evaluation of the various selection methods not on α (which is a nuisance parameter whose numerical value is physically irrelevant), but on the model recovery error defined in Eq. (8). It quantifies how close the calculated Tikhonov solution, $P_{\alpha L}$, for a given α and L is to the model distribution P_0. This error is generally non-zero, since noise in the data and the fundamental ill-posedness of the problem prevent full recovery of the model from the data, no matter which α and L are used. As the actual performance measure, we use the inefficiency [43]

Table 1
List of acronyms for regularization parameter selection methods. To indicate the use of unconstrained $P(r)$ in a method, the suffix 'u' is appended to the acronym. If a tuning parameter is used, its value is appended as well.

Acronym	Full name
AIC	Akaike information criterion
AICC	corrected Akaike information criterion
BIC	Bayesian information criterion
BP	balancing principle
CV	leave-one-out cross validation
DP	discrepancy principle
EE	extrapolated error
GCV	generalized cross validation
GML	generalized maximum-likelihood
hBP	hardened balancing principle
ICOMP	information complexity criterion
LC	L-curve, maximum curvature
LR	L-curve, minimum radius
LR2	L-curve, minimum radius 2
MCL	Mallows' C_L
mGCV	modified generalized cross validation
NCP	normalized cumulative periodogram
QO	quasi-optimality criterion
rGCV	robust generalized cross validation
RM	residual method
SC	self-consistency method
srGCV	strong robust generalized cross validation
tDP	transformed discrepancy principle

$$\text{inefficiency} = \frac{\text{rmsd}(\alpha, L)}{\text{rmsd}(\alpha_{opt}, L_{opt})} \qquad (9)$$

with $\text{rmsd}(\alpha, L) = \text{mre}(\alpha, L)/\sqrt{n_t} = \|P_{\alpha L} - P_0\|/\sqrt{n_r}$. This measure relates the model recovery error for a given L and the particular α value chosen by a selection method to the smallest possible model recovery error, obtainable with α_{opt} and L_{opt}. This error, and the associated optimal solution P_{opt} are found by minimizing the mre as a function of α and L. The inefficiency equals 1 when $P_{\alpha L} = P_{opt}$ and is greater otherwise. The closer the inefficiency is to 1, the better the method.

5. Results and discussion

5.1. Histograms

Fig. 5 shows an overview of the calculated inefficiencies for all 621030 time traces, 60 α selection methods, and three regularization operators. Each histogram shows the distribution of inefficiencies for a particular combination of method and operator. The methods are sorted by increasing 75th-percentile inefficiency from top left to bottom right. There is a wide range of performances. Most methods in the top row achieve low inefficiencies that are sharply peaked at 1. Methods shown lower down have a propensity towards worse inefficiencies. The high-inefficiency tails of the histograms are not shown, but are protracted in many cases, indicating a not insignificant failure rate for those methods. It is apparent from the histograms that the L_0 operator, shown in blue, tends to perform worse than the others for the same method, except for some with overall bad performance. For most methods, L_1 and L_2 appear to give similar results.

5.2. Performance comparison

To compare the method/operator combinations more quantitatively, we use the 99th percentile from each inefficiency histogram as the metric. This stringent choice is motivated by the consideration that a method/operator combination can be deemed good and safe if it delivers overall low inefficiencies with negligible risk of large inefficiencies (i.e. severe under- or oversmoothing). Fig. 6 shows the results. There is a group of methods that perform similarly well, leading to small 99th-percentile inefficiencies just above 2. Mallows' C_L (MCL), the Akaike information criterion (AIC) and the generalized cross validation (GCV) perform equally well with L_1 and L_2, whereas the generalized maximum likelihood method and its unconstrained variant (GML and GMLu) perform well only with L_1. L_0 is the worst-performing operator for these top methods. The methods ranking below this top group are modifications of GCV and AIC. Since they are inferior to their parent methods, they can be disregarded. Currently, the most commonly applied α selection methods are based on the L-curve. The results show that the L-curve methods LC, LCu, LR, LRu, LR2, LR2u are not competitive with the top group. Their overall tendency is to oversmooth. Similarly, the self-consistent methods (SC and SCu) display elevated inefficiency for DEER. The computational cost of all methods is essentially identical, since it is dominated by the solution of the Tikhonov minimization problem for each α.
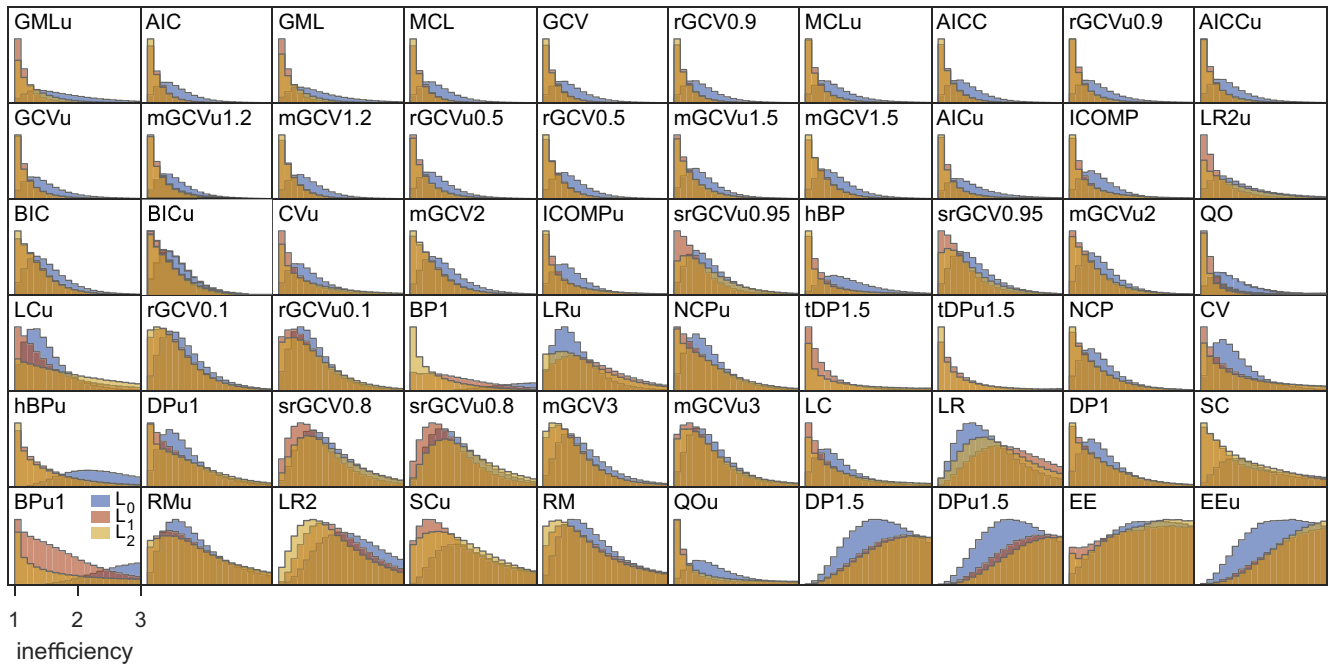
5.3. Sensitivity to metric

The performance comparison shown in Fig. 6 is based on our particular choice of metric, the 99th percentile of the ratio of rmsds, defined in Eq. (9). Varying the percentile to 90, 75, or 50 does not significantly affect the composition of the top group, although it affects the detailed rankings (see SI). Altering the definition of inefficiency in Eq. (9) by using the difference instead of the ratio,

Fig. 5. Inefficiency distributions for each α selection method (indicated by their acronyms) and regularization operator (color-coded). The panels are arranged from top left to bottom right in increasing order of lowest 75th-percentile inefficiency among L_0, L_1, and L_2. For each histogram, only the part with inefficiency ≤ 3 is shown, and the often prolonged tails are not visible. The histograms are scaled such that the maxima in all panels are identical. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
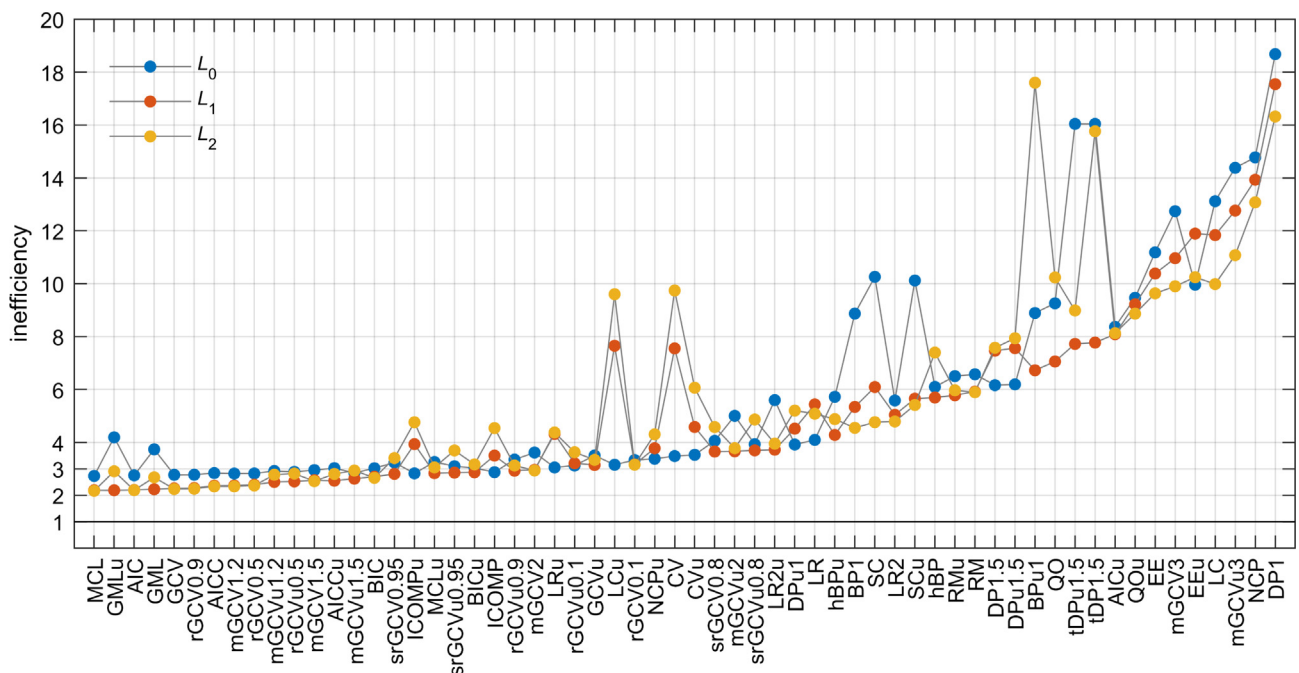
Fig. 6. Performance comparison of all combinations of α selection methods and regularization operators, based on their 99th-percentile inefficiency. The methods are sorted based on best 99th-percentile inefficiency among L_0, L_1, and L_2. The gray lines serves as guides to the eye. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

or by using the maximum absolute deviation (mad) instead of the rmsd, yields similar results (see SI). Therefore, we conclude that the identity of the top methods is insensitive towards the particular choice of ranking metric, and that our assessment procedure is overall robust.

5.4. Sensitivity to data characteristics

To check for uneven performance for certain subsets of the test set, we examined the performance across noise level, the ratio of iqr to mode (a measure of the damping rate of oscillations in S),

number of points, and number of modal periods, using the 99th percentile rmsd ratio inefficiency metric. The breakdowns are available in the SI. For all values of these variables, MCL, AIC, GCV, GML, and GMLu are usually within 5% of the optimal method/operator combination and never deviate by more than 11%. The specific rankings vary between subsets, but actual performance changes to such a small degree that this is not significant. The conclusions reached above regarding the three operators generalize to the subsets considered in the SI. In addition to poor performance relative to L_1 and L_2, L_0 also displays much higher variation in performance across test subsets. Several variations on GCV (mGCV, mGCVu, rGCV, and rGCVu) with varied tuning parameter values, as well as AICC, come within 10% of the best case for many of the test subsets. However, none display sufficient consistency or quality of performance to surpass MCL, AIC, GCV, GML, or GMLu.

Since there is no correlation between the performance of any top method and these subset characteristics, the results are therefore likely to be applicable to situations with different relative representations of distance distribution characteristics, such as other spin labels or proteins with more β sheets than T4 lysozyme.

5.5. Method choice

The performance differences among the top method/operator combinations are so small that we cannot identify one as an evident best choice. However, additional considerations can provide some guidance. MCL requires the time-domain noise variance σ^2 as an input. An under- or overestimation of σ^2, which is likely in experimental settings, will affect the performance and likely degrade it. GML and GMLu depend on a thresholding value to remove eigenvalues close to zero, and the choice of this value can affect the performance. In contrast, AIC and GCV do not require *a priori* knowledge of σ^2 nor do they depend on thresholding or tuning parameters. Therefore, the parameter-free AIC and GCV methods with either L_1 or L_2 appear to be the simplest, best, and safest choices for practical applications.

Due to the diversity of distribution shapes and time traces, and due to the wide range of inefficiencies, it is impossible to visualize method performances with a few sample datasets. Nevertheless, for the sake of illustration, Fig. 7 shows the relative performance of LR2u, GCV, and AIC for a typical test dataset. This example is typ-
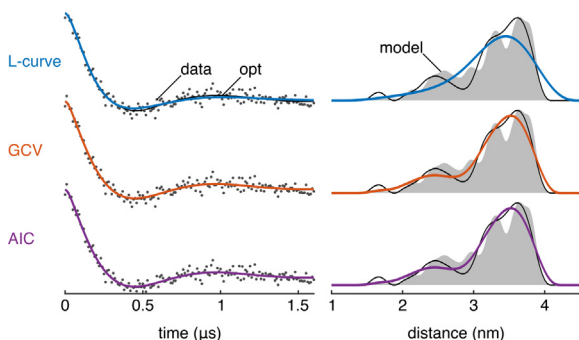
Fig. 7. Example case comparing the LR2u L-curve, GCV, and AIC methods (dataset 115333). For this particular test case, each method features an inefficiency very close to that method's overall median inefficiency (LR2u 1.38, GCV 1.09, AIC 1.09). Left column: the simulated data are shown as grey dots, the Tikhonov fit with optimum α is shown as a black line, and the Tikhonov fits for LR2u, GCV, and AIC are shown as blue, orange, and purple lines, respectively. Right column: the model P is shown in grey, the Tikhonov solution for P with optimum α is shown as a black line, and the Tikhonov Ps for LR2u, GCV, and AIC are shown as blue, orange, and purple lines, respectively. In all cases, the L_2 operator was used. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ical in these sense that the inefficiency of each method is very close to the method's median inefficiency. The figure shows that the L-curve has a tendency to oversmooth, and that AIC and GCV perform similarly well. Fig. S1 in the Supplementary Material illustrates the uncertainty in the extracted $P(r)$ based on these α selection methods, quantified using Bayesian inference [14].

6. Conclusions

Our performance analysis of a large number of regularization parameter selection methods over a large set of protein-based synthetic DEER data indicates that there are several method/operator combinations that perform equally well. Among these, the Akaike information criterion (AIC) and the generalized cross validation (GCV) are preferable, as they are parameter-free and do not require accurate knowledge of the noise level. They work equally well with the first- and second-derivative operator, but not with the identity. L-curve methods, some of which are currently widely employed, perform distinctly worse.

The structure-based test set developed in this work is useful beyond Tikhonov regularization, as it can be used to assess the performance of other existing analysis methods (truncated singular-value decomposition, Gaussian mixture models) and of any new solution approaches.

Acknowledgements

Appendix A. Selection methods

Here, we summarize all α selection methods included in this study and give the expressions necessary to implement them. Further details about the methods can be found in the cited statistics literature.

Most methods are derived assuming an unconstrained Tikhonov regularization, i.e. without the $P \geq 0$ constraint in Eq. (5). In this case, the solution $P_{\alpha L}$ can be expressed in closed form as

$$P_{\alpha L} = \overline{K}_{\alpha L} S \qquad \text{with} \qquad \overline{K}_{\alpha L} = \left(K^T K + \alpha^2 L^T L \right)^{-1} K^T \qquad (10)$$

For the physically relevant constrained problem (with $P \geq 0$), a closed-form solution is not possible, and we obtain $P_{\alpha L}$ via an iterative optimization algorithm [44]. In either case, once $P_{\alpha L}$ is obtained, the time-domain fit is

$$S_{\alpha L} = K P_{\alpha L} \qquad (11)$$

We apply each method described below in two ways, one using the unconstrained solutions and one using the constrained solutions. The methods are referred to by short acronyms (see Table 1). A suffix 'u' is appended to the method acronym when unconstrained $P_{\alpha L}$ are used. In addition, if the method contains a tuning or scaling parameter, then its value is appended to the acronym as well.

L-Curve methods (LC, LR, LR2)

Several methods select α based on heuristic considerations about a parametric log–log plot of the Tikhonov penalty term,

$\eta = \|LP_{\alpha L}\|$, against the residual norm, $\rho = \|S - S_{\alpha L}\|$, as a function of α [45,46]. Such a plot shows a monotonically decaying $\eta(\rho)$ and is called the L-curve since it tends to form a characteristic "L" shape with a visual corner. The optimal value for α is considered to correspond to this corner, since it intuitively represents a reasonable balance between the fitting error ρ and the regularization error η. The corner is not a mathematically defined quantity, therefore different operational definitions of locating such a corner exist.

One of them is the maximum-curvature method (LC) [46]. It selects the α corresponding to the point of maximum (positive) curvature of the L-curve, given by

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmax}} \frac{\hat{\rho}'\hat{\eta}'' - \hat{\rho}''\hat{\eta}'}{(\hat{\rho}'^2 + \hat{\eta}'^2)^{3/2}} \qquad (12)$$

where $\hat{\eta} = \lg\|LP_{\alpha L}\|$, $\hat{\rho} = \lg\|S - S_{\alpha L}\|$, and the primes indicate derivatives with respect to $\lg\alpha$ or α. This method picks the global maximum of the curvature, even though there can be several local maxima.

Another possible definition of corner is the point closest to the lower left corner of the L-curve plot [45]. DeerAnalysis uses one implementation of this idea [7]. The two coordinates $\hat{\rho}$ and $\hat{\eta}$ are evaluated over a range of α values and then rescaled to the interval [0,1]. The corner is determined as the location on the L-curve that is closest to the origin in these rescaled coordinates:

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \left(\left(\frac{\hat{\rho} - \hat{\rho}_{min}}{\hat{\rho}_{max} - \hat{\rho}_{min}} \right)^2 + \left(\frac{\hat{\eta} - \hat{\eta}_{min}}{\hat{\eta}_{max} - \hat{\eta}_{min}} \right)^2 \right) \qquad (13)$$

The results from this method depend on the α range. We refer to this method as the minimum-radius method and use two variants of it. In the first one (LR), we use the same α range as for all other methods (10^{-3} to 10^3), and in the other one (LR2), we employ the same α range as DeerAnalysis2016. The latter extends from the largest generalized singular eigenvalue of K and L, s_{max}, to a value cs_{max} with $c = 16\epsilon \cdot 10^6 \cdot 2^{\sigma/0.0025}$, where $\epsilon \approx 2.2 \cdot 10^{-16}$ and σ is the noise level. In order to keep the range large enough, we limit c to $\leq 10^{-6}$.

Cross validation (CV, GCV, mGCVc, rGCVγ, srGCVγ)

We use several methods based on the idea of cross validation. Leave-one-out cross validation (CV) [47,48] is conceptually the simplest of them. For a given α, it minimizes the total prediction error, which is obtained as the sum of the prediction errors for each individual data point. For that, a single data point is excluded and a fit to the remaining data is calculated. That fit is judged based upon its ability to reproduce the excluded data point. By repeating the method for each data point, the total prediction error is obtained, and the α value is selected that minimizes it. This procedure can be condensed into the following expression:

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \left(\sum_{i=1}^n \left| \frac{S(i) - S_{\alpha L}(i)}{1 - H_{\alpha L}(i,i)} \right|^2 \right) \qquad (14)$$

with the α-dependent influence matrix $H_{\alpha L} = K\overline{K}_{\alpha L}$.

Generalized cross validation (GCV) [49,50] is very similar to leave-one-out cross validation, but the diagonal matrix element $H_{\alpha L}(i,i)$ in the denominator is replaced by the average of all diagonal elements, rendering the method more stable. The expression simplifies to

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \frac{\|S - S_{\alpha L}\|^2}{[1 - \mathrm{tr}(H_{\alpha L})/n_t]^2} \qquad (15)$$

The modified GCV (mGCVc) [51,52] method is a simple tunable modification to GCV intended to stabilize the method further.

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \frac{\|S - S_{\alpha L}\|^2}{[1 - c\,\mathrm{tr}(H_{\alpha L})/n_t]^2} \qquad (16)$$

with the tuning parameter $c \geq 1$. We use this with $c = 1.2$, 1.5, 2, and 3, to test a range of stabilizations.

The robust GCV (rGCVγ) method [53,54] is also designed to exhibit greater stability than the GCV method. It is given by

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \left(\frac{\|S - S_{\alpha L}\|^2}{[1 - \mathrm{tr}(H_{\alpha L})/n_t]^2} \left(\gamma + (1 - \gamma)\mathrm{tr}\left(H_{\alpha L}^2\right)/n_t \right) \right) \qquad (17)$$

with the tuning parameter $\gamma \leq 1$. With $\gamma = 1$, the method reduces to GCV. As γ gets smaller, the method becomes increasingly stable and less likely to undersmooth. We examine γ values of 0.1, 0.5, and 0.9, with increasing contributions from the second term.

Strong robust GCV (srGCVγ) [55] is another tunable modification to GCV that is based on stronger statistical arguments than rGCV. It selects α via

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \left(\frac{\|S - S_{\alpha L}\|^2}{[1 - \mathrm{tr}(H_{\alpha L})/n_t]^2} \left(\gamma + (1 - \gamma)\mathrm{tr}(\overline{K}_{\alpha L}^{\mathrm{T}}\overline{K}_{\alpha L})/n_t \right) \right) \qquad (18)$$

with $\gamma \geq 1$. Like rGCV, with $\gamma = 1$, the method reduces to GCV. We use γ values of 0.8 and 0.95.

Quasi-optimality (QO)

This criterion, from Tikhonov and coworkers [56–58], selects α such that a small change in α from that selected value has minimal effects on the resulting $P_{\alpha L}$:

$$\alpha_{sel} = \underset{\alpha}{\mathrm{argmin}} \left\| \frac{dP_{\alpha L}}{d \lg \alpha} \right\| \qquad (19)$$

The underlying rationale is that the model recovery error is flat at its minimum.

Discrepancy principle (DPτ)

This principle [59–61] is predicated upon the idea that the root-mean-square residual, $\|S - S_{\alpha L}\|/\sqrt{n}$, should be on the order of the noise standard deviation in the data, σ. It requires *a priori* knowledge of σ. The value for α is selected as the largest value such that

$$\|S - S_{\alpha L}\|/\sqrt{n_t} \leqslant \tau\sigma \qquad (20)$$

where τ is a safety factor $\geqslant 1$ to guard against undersmoothing in the case σ is underestimated. We use this principle with $\tau = 1$ and 1.5.

The transformed discrepancy principle (tDPτ) [62–64] performs the comparison in the distance domain, choosing the largest α that satisfies

$$\|\overline{K}_{\alpha L}(S - S_{\alpha L})\|/\sqrt{n_r} \leqslant \tau \frac{b_0\sigma}{\alpha} \qquad (21)$$

where $b_0 = 3\sqrt{3}/16 \approx 0.325$. We use $\tau = 1.5$.

Balancing principle (BP, hBP)

The balancing principle (BP) [65,66] balances the propagated noise error with the unknown model recovery error. Using

$$B(\alpha) = \underset{\alpha' < \alpha}{\max} \left(\underset{\alpha'' < \alpha'}{\max} \frac{\|P_{\alpha'} - P_{\alpha''}\|}{4\sigma\phi(\alpha'')} \right) \qquad (22)$$

and $\phi(\alpha) = \sqrt{\mathrm{tr}(\overline{K}_{\alpha L}\overline{K}_{\alpha L}^{\mathrm{T}})}$, it selects α_{sel} as the largest α value that satisfies $B(\alpha) \leqslant 1$. The hardened balancing principle (hBP) [66] selects α using

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(B(\alpha) \sqrt{\phi(\alpha)} \right) \qquad (23)$$

Residual method (RM)

This method [43,67] determines α by minimizing a scaled norm of the residual vector

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \frac{\|S - S_{\alpha L}\|^2}{\sqrt{\operatorname{tr}(B^{\mathrm{T}} B)}} \qquad (24)$$

where $B = K^{\mathrm{T}}(I - H_{\alpha L})$. The scaling penalizes undersmoothing.

Self-consistency method (SC)

This method [25,26], utilized in FTIKREG, seeks to minimize the sum of the estimated model recovery error and the propagated data noise variance.

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(\left\| (I - \overline{K}_{\alpha L} K) P_{\alpha L} \right\|^2 + \sigma^2 \operatorname{tr}(\overline{K}_{\alpha L} \overline{K}_{\alpha L}^{\mathrm{T}}) \right) \qquad (25)$$

This expression is valid for the unconstrained case. In order to include the non-negativity constraint, further steps are necessary. First, α_{sel} for the unconstrained case is determined. Next, the constrained solution is determined using this α value and the indices q of the active non-negativity constraints are stored. Finally, the columns of K and L with indices q are removed and Eq. (25) is re-evaluated.

Generalized maximum likelihood (GML)

This method [68] selects the α that maximizes the likelihood (or, equivalently, minimizes the negative log-likelihood) and is given by

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \frac{S^{\mathrm{T}}(S - S_{\alpha L})}{\sqrt[m]{\det_{\text{nz}}(I - H_{\alpha L})}} \qquad (26)$$

where $\det_{\text{nz}}(\cdot)$ indicates the product of the non-zero eigenvalues, and m is the their number. To account for numerical errors, we treat all eigenvalues with magnitude below 10^{-8} as zero.

Extrapolated error (EE)

This method [69–71] minimizes an estimate of the regularization error via

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \frac{\|S - S_{\alpha L}\|^2}{\left\| K^{\mathrm{T}}(S - S_{\alpha L}) \right\|} \qquad (27)$$

Normalized cumulative periodogram (NCP)

This method [72–74] is based on the idea that the power spectrum of the residuals should match the power spectrum of the noise. The unscaled power spectrum (periodogram) for a given α is a vector p with elements

$$p(k) = \left| \operatorname{dft}(S - S_{\alpha L})_k \right|^2 \qquad (28)$$

where $k = 1, \ldots, n_t$, dft refers to the discrete Fourier transform. The normalized cumulative periodogram is an $(n_t - 1)$-element vector c with elements

$$c(i) = \frac{\|p(2 : i + 1)\|_1}{\|p(2 : n_t)\|_1} \qquad (29)$$

where $\|\ldots\|_1$ is the ℓ_1 norm (sum of absolute values). The zero-frequency component, p_1, is omitted. c represents the integrated power spectrum of the residuals. The α value is selected that minimizes the deviation between c and the integrated power spectrum c_{noise} expected for the noise:

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \|c(\alpha) - c_{\text{noise}}\|_1 \qquad (30)$$

where c_{noise} is a $(n_t - 1)$-element vector, with elements $c_{\text{noise},i} = i/(n_t - 1)$ for white noise.

Mallows' C_L (MCL)

This method [75] minimizes an approximation to the model recovery error, derived under the assumption of unconstrained regularization and uncorrelated Gaussian noise.

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(\|S - S_{\alpha L}\|^2 + 2\sigma^2 \operatorname{tr}(H_{\alpha L}) - 2n_t \sigma^2 \right) \qquad (31)$$

This requires the knowledge of the noise level σ.

Information-theoretical criteria (AIC, AICC, BIC, ICOMP)

In the context of information theory [76,77], the set of $P_{\alpha L}$ is regarded as a set of candidate models, and criteria have been developed that select a parsimonious model that balances a small fitting error with a low model complexity. The general expression is

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(n_t \ln \frac{\|S - S_{\alpha L}\|^2}{n_t} + c \operatorname{tr}(H_{\alpha L}) \right) \qquad (32)$$

where the second term is a measure of model complexity, which decreases with increasing α. The term $\operatorname{tr}(H_{\alpha L})$ can be regarded as the effective number of free parameters in the model. The constant c depends on the particular criterion: $c = 2$ for the Akaike information criterion (AIC) [78], $c = 2n_t/(n_t - \operatorname{tr}(H_{\alpha L}) - 1)$ for the corrected AIC (AICC) [79,80], and $c = \ln(n_t)$ for the Bayesian information criterion (BIC) [81]. The information complexity criterion (ICOMP) [82,83] is yet another information-theoretical procedure and is given by

$$\alpha_{\text{sel}} = \underset{\alpha}{\operatorname{argmin}} \left(\frac{1}{\sigma^2} \|S - S_{\alpha L}\|^2 + 2 \operatorname{tr}(H_{\alpha L}) + n_t \ln \frac{\bar{s}_a}{\bar{s}_g} \right) \qquad (33)$$

where $\bar{s}_a$ and $\bar{s}_g$ are the arithmetic and geometric means of the singular values of $(K^{\mathrm{T}} K + \alpha^2 L^{\mathrm{T}} L)^{-1}$. The last term penalizes for interdependence among model parameters. In contrast to the other information-theoretical criteria, this one requires knowledge of the noise level σ.

Appendix B. Supplementary material

The online supplementary material associated with this article can be found at https://doi.org/10.1016/j.jmr.2018.01.021. It contains (a) tables of relative performance based on different metrics and percentiles, including breakdowns into subsets based on data characteristics, (b) Bayesian uncertainty analysis for the test case in Fig. 7. The complete set of synthetic model distributions and test time-domain traces is freely available for download from the University of Washington ResearchWorks public repository at https://doi.org/10.6069/H5S75DCG.

References

[1] A.D. Milov, K.M. Salikhov, M.D. Shchirov, Application of the double resonance method to electron spin echo in a study of the spatial distribution of paramagnetic centers in solids, Fiz. Tverd. Tela 23 (1981) 975–982.
[2] A.D. Milov, A.B. Ponomarev, Yu.D. Tsvetkov, Electron-electron double resonance in electron spin echo: Model biradical systems and the sensitized photolysis of decalin, Chem. Phys. Lett. 110 (1984) 67–72.

[3] R.G. Larsen, D.J. Singel, Double electron-electron resonance spin-echo modulation: Spectroscopic measurement of electron spin pair separations in orientationally disordered solids, J. Chem. Phys. 98 (1993) 5134–5146.

[4] M.K. Bowman, A.G. Maryasov, N. Kim, V.J. DeRose, Visualization of distance distributions from pulse double electron-electron resonance data, Appl. Magn. Reson. 26 (2004) 23–39.

[5] G. Panek, A. Godt, A. Bender, H. Paulsen, Data analysis procedures for pulse ELDOR measurements of broad distance distributions, Appl. Magn. Reson. 26 (2004) 223–244.

[6] Y.-W. Chiang, P.B. Borbat, J.H. Freed, The determination of pair distance distributions by pulsed ESR using Tikhonov regularization, J. Magn. Reson. 172 (2005) 279–295.

[7] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banjam, C.R. Timmel, D. Hilger, H. Jung, DeerAnalysis2006: a comprehensive software package for analyzing pulsed ELDOR data, Appl. Magn. Reson. 30 (2006) 473–498.

[8] K.I. Sen, T.M. Logan, P.G. Fajer, Protein dyanmics and monomer-monomer interactions in AntR activation by electron paramagnetic resonance and double electron-electron resonance, Biochemistry 46 (2007) 11639–11649.

[9] S. Brandon, A.H. Beth, E.J. Hustedt, The global analysis of DEER data, J. Magn. Reson. 218 (2012) 93–104.

[10] R.A. Stein, A.H. Beth, E.J. Hustedt, A straightforward approach to the analysis of Double Electron-Electron Resonance data, Methods Enzymol. 563 (2015) 531–567.

[11] M.E. Blackburn, A.M. Veloro, G.E. Fanucci, Monitoring inhibitor-induced conformational population shifts in HIV-1 protease by pulsed EPR spectroscopy, Biochemistry 48 (2009) 8765–8767.

[12] T.M. Casey, G.E. Fanucci, Spin labeling and Double Electron-Electron Resonance (DEER) to Deconstruct Conformational Ensembles of HIV Protease, Method Enzymol. 564 (2015) 153–187.

[13] Y.-W. Chiang, P.P. Borbat, J.H. Freed, Maximum entropy: A complement to Tikhonov regularization for determination of pair distance distributions by pulsed ESR, J. Magn. Reson. 177 (2005) 184–196.

[14] T.H. Edwards, S. Stoll, A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy, J. Magn. Reson. 270 (2016) 87–97.

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

[16] M. Srivastava, E.R. Georgieva, J.H. Freed, A new wavelet denoising method for experimental time-domain signals: pulse dipolar electron spin resonance, J. Phys. Chem. A 121 (2017) 2452–2465.

[17] M. Srivastava, J.H. Freed, Singular value decomposition method to determine distance distributions in pulsed dipolar electron spin resonance, J. Phys. Chem. Lett. 8 (2017) 5648–5655.

[18] I. Kuprov, Deep neural network processing of DEER data, The 50th Annual International Meeting of the Electron Spin Resonance Group of the Royal Society of Chemistry, 2017.

[19] H. Schäfer, B. Mäder, F. Volke, De-Pake-ing of NMR powder spectra by nonnegative least-squares analysis with Tikhonov regularization, J. Magn. Reson. A 116 (1995) 145–149.

[20] F.G. Vogt, D.J. Aurentz, K.T. Mueller, Determination of internuclear distances from solid-state nuclear magnetic resonance: Dipolar transforms and regularization methods, Mol. Phys. 95 (1998) 907–919.

[21] H. Schäfer, B. Mädler, E. Sternin, Determination of orientational order parameters from ^{2}H NMR spectra of magnetically partially oriented lipid bilayers, Biophys. J. 74 (1998) 1007–1014.

[22] J. Winterhalter, D. Maier, D.A. Grabowski, J. Honerkamp, Determination of orientational distributions from ^{2}H NMR data by a regularization method, J. Chem. Phys. 110 (1999) 4035–4046.

[23] H. Schäfer, E. Sternin, Inverse ill-posed problems in experimental data analysis in physics, Phys. Can. 53 (1997) 77–85.

[24] E. Sternin, Use of inverse theory algorithms in the analysis of biomembrane NMR data, Method Mol. Bio. 400 (2007) 103–125.

[25] J. Honerkamp, J. Weese, Tikhonovs regularization method for ill-posed problems, Continuum Mech. Thermodyn. 2 (1990) 17–30.

[26] J. Weese, A reliable and fast method for the solution of Fredholm integral equations of the first kind based on Tikhonov regularization, Comput. Phys. Commun. 69 (1992) 99–111.

[27] G. Jeschke, A. Koch, U. Jonas, A. Godt, Direct conversion of EPR dipolar time evolution data to distance distributions, J. Magn. Reson. 155 (2002) 72–82.

[28] G. Jeschke, Distance measurements in the nanometer range by Pulse EPR, Chem. Phys. Chem. 3 (2002) 927–932.

[29] H.S. Mchaourab, T. Kálai, K. Hideg, W.L. Hubbell, Motion of spin-labeled side chains in T4 lysozyme: Effect of side chain structure, Biochemistry 38 (1999) 2947–2955.

[30] R. Langen, K.J. Oh, D. Cascio, W.L. Hubbell, Crystal structures of spin labeled T4 lysozyme mutants: Implications for the interpretation of EPR spectra in terms of structure, Biochemistry 39 (2000) 8396–8405.

[31] L. Columbus, T. Kálai, J. Jekö, K. Hideg, W.L. Hubbell, Molecular motion of spin labeled side chains in α-helices: Analysis by variation of side chain structure, Biochemistry 40 (2001) 3828–3846.

[32] Z. Liang, Y. Lou, J.H. Freed, L. Columbus, W.L. Hubbell, A multifrequency electron spin resonance study of T4 lysozyme dynamics using the slowly relaxing local structure model, J. Phys. Chem. B 108 (2004) 17649–17659.

[33] K. Jacobsen, S. Oga, W.L. Hubbell, T. Risse, Determination of the orientation of T4 lysozyme vectorially bound to a planar-supported lipid bilayer using site-directed spin labeling, Biophys. J. 88 (2005) 4351–4365.

[34] C. Altenbach, W. Froncisz, R. Hemker, H. Mchaourab, W.L. Hubbell, Accessibility of nitroxide side chains: Absolute Heisenberg exchange rates from power saturation EPR, Biophys. J. 89 (2005) 2103–2112.

[35] J. Pyka, J. Ilnicki, C. Altenbach, W.L. Hubbell, W. Froncisz, Accessibility and dynamics of nitroxide side chains in T4 lysozyme measured by saturation recovery EPR, Biophys. J. 89 (2005) 2059–2068.

[36] C.J. Lopez, Z. Yang, C. Altenbach, W.L. Hubbell, Conformational selection and adaptation to ligand binding in T4 lysozyme cavity mutants, Proc. Natl. Acad. Sci. USA 110 (2013) E4306–E4315.

[37] M.T. Lerch, C.J. López, Z. Yang, M.J. Kreitman, J. Horwitz, W.L. Hubbell, Structure-relaxation mechanism for the response of T4 lysozyme cavity mutants to hydrostatic pressure, Proc. Natl. Acad. Sci. USA 112 (2015) E2437–E2446.

[38] Y. Polyhach, E. Bordignon, G. Jeschke, Rotamer libraries of spin labelled cysteines for protein studies, Phys. Chem. Chem. Phys. 13 (2011) 2356–2366.

[39] Y. Polyhach, G. Jeschke, Prediction of favourable sites for spin labelling of proteins, Spectrosc. Int. J. 24 (2010) 651–659.

[40] G. Jeschke, MMM: A toolbox for integrative structure modeling, Prot. Sci. 27 (2018) 76–85.

[41] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer, 2005.

[42] J. Kaipio, E. Somersalo, Statistical inverse problems: Discretization, model reduction and inverse crimes, J. Comput. Appl. Math. 198 (2007) 493–504.

[43] F. Bauer, M.A. Lukas, Comparing parameter choice methods for regularization of ill-posed problems, Math. Comput. Simulat. 81 (2011) 1795–1841.

[44] R. Bro, S. de Jong, A fast non-negativity-constrained least squares algorithm, J. Chemometr. 11 (1997) 393–401.

[45] P.C. Hansen, D.P. O'Leary, The use of the L-curve in the regularization of discrete ill-posed problems, SIAM J. Sci. Comp. 14 (1993) 1487–1503.

[46] P.C. Hansen, The L-Curve and its use in the numerical treatment of inverse problems, in: P. Johnston (Ed.), Advances in Computational Bioengineering: Computational Inverse Problems in Electrocardiology, WIT Press, 2000, pp. 119–142.

[47] G. Wahba, Spline Models for Observational Data, SIAM, 2007.

[48] P.C. Hansen, Discrete Inverse Problems: Insight and Algorithms, SIAM, 2010.

[49] G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numer. Anal. 14 (1977) 651–667.

[50] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215–223.

[51] D.J. Cummins, T.G. Filloon, D. Nychka, Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage, J. Am. Stat. Assoc. 96 (2001) 233–246.

[52] R. Vio, P. Ma, W. Zhong, J. Nagy, L. Tenorio, W. Wamsteker, Estimation of regularization parameters in multiple-image deblurring, Astron. Astrophys. 423 (2004) 1179–1186.

[53] M.A. Lukas, Robust generalized cross-validation for choosing the regularization parameter, Inverse Probl. 22 (2006) 1883–1902.

[54] T. Robinson, R. Moyeed, Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression, Stat. Theory Methods 18 (1989) 523–539.

[55] M.A. Lukas, Strong robust generalized cross-validation for choosing the regularization parameter, Inverse Probl. 24 (2008).

[56] A. Tikhonov, V. Glasko, Use of the regularization method in non-linear problems, U.S.S.R. Comput. Math. Math. Phys. 5 (1965) 93–107.

[57] A. Tikhonov, V. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.

[58] B. Hofmann, Regularization of Applied Inverse and Ill-Posed Problems, Teubner, Leipzig, 1986.

[59] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. Assoc. Comput. Mach. 9 (1962) 84–97.

[60] V.A. Morozov, On the solution of functional equations by the method of regularization, Soviet Math. Dokl. 7 (1966) 414–417.

[61] V.A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer, New York, 1984.

[62] T. Raus, An a posteriori choice of the regularization parameter in case of approximately given error bound of data, Acta et Comment. Univ. Tartuensis 913 (1990) 73–87.

[63] T. Raus, About regularization parameter choice in case of approximately given error bounds of data, Acta et Comment. Univ. Tartuensis 937 (1992) 77–89.

[64] U. Hämarik, T. Raus, On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data, J. Inverse Ill-Posed Probl. 14 (2006) 251–266.

[65] O. Lepskij, On a problem of adaptive estimation in gaussian white noise, Theor. Probab. Appl. 35 (1990) 454–466.

[66] F. Bauer, Some considerations concerning regularization and parameter choice algorithms, Inverse Probl. 23 (2007) 837–858.

[67] F. Bauer, P. Mathé, Parameter choice methods using minimization schemes, J. Complexity 27 (2011) 68–85.

[68] G. Wahba, A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem, Ann. Stat. 13 (1985) 1378–1402.

[69] G. Auchmuty, A posteriori error estimates for linear equations, Numer. Math. 61 (1992) 1–6.

[70] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for linear systems with applications to regularization, Numer. Algorithms 49 (2008) 85–104.

[71] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for the regularization of least squares problems, Numer. Algorithms 51 (2009) 61–76.

[72] P.C. Hansen, M.E. Kilmer, R.H. Kjeldsen, Exploiting residual information in the parameter choice for discrete ill-posed problems, BIT 46 (2006) 41–59.

[73] B.W. Rust, Parameter selection for constrained solutions to ill-posed problems, Comput. Sci. Stat. 32 (2000) 333–347.

[74] B.W. Rust, D.P. O'Leary, Residual periodograms for choosing regularization parameters for ill-posed problems, Inverse Probl. 24 (2008) 034005.

[75] C.L. Mallows, Some comments on C_P, Technometrics 15 (1973) 1–6.

[76] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference, Springer, 2002.

[77] Y. Zhang, R. Li, C.-L. Tsai, Regularization parameter selections via generalized information criterion, J. Am. Stat. Assoc. 105 (2010) 312–323.

[78] H. Akaike, A new look at the statistical model identification, IEEE T. Automat. Contr. 19 (1974) 716–723.

[79] N. Sugiura, Further analysis of the data by Akaike's information criterion and the finite corrections, Commun. Stat. Theor. M. A7 (1978) 13–26.

[80] C.M. Hurvich, C.-L. Tsai, Regression and time series model selection in small samples, Biometrika 76 (1989) 297–307.

[81] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.

[82] H. Bozgodan, Akaike's information criterion and recent developments in information complexity, J. Math. Psychol. 44 (2000) 62–91.

[83] A.M. Urmanov, A.V. Gribok, H. Bozdogan, J.W. Hines, R.E. Uhrig, Information complexity-based regularization parameter selection for solution of ill conditioned inverse problems, Inverse Probl. 18 (2002) L1–L9.