2-10-2016

What Is the Best Method to Fit Time-Resolved Data? A Comparison of the Residual Minimization and the Maximum Likelihood Techniques As Applied to Experimental Time-Correlated, Single-Photon Counting Data

Kalyan Santra
Iowa State University and Ames Laboratory, kalyans@iastate.edu

Jinchun Zhan
Iowa State University

Xueyu Song
Iowa State University and Ames Laboratory, xsong@iastate.edu

Emily A. Smith
Iowa State University and Ames Laboratory, esmith1@iastate.edu

Namrata Vaswani
Iowa State University, namrata@iastate.edu

See next page for additional authors

What Is the Best Method to Fit Time-Resolved Data? A Comparison of the Residual Minimization and the Maximum Likelihood Techniques As Applied to Experimental Time-Correlated, Single-Photon Counting Data

Abstract

The need for measuring fluorescence lifetimes of species in subdiffraction-limited volumes in, for example, stimulated emission depletion (STED) microscopy, entails the dual challenge of probing a small number of fluorophores and fitting the concomitant sparse data set to the appropriate excited-state decay function. This need has stimulated a further investigation into the relative merits of two fitting techniques commonly referred to as "residual minimization" (RM) and "maximum likelihood" (ML). Fluorescence decays of the well-characterized standard, rose bengal in methanol at room temperature (530 ± 10 ps), were acquired in a set of five experiments in which the total number of "photon counts" was approximately 20, 200, 1000, 3000, and 6000 and there were about 2–200 counts at the maxima of the respective decays. Each set of experiments was repeated 50 times to generate the appropriate statistics. Each of the 250 data sets was analyzed by ML and two different RM methods (differing in the weighting of residuals) using in-house routines and compared with a frequently used commercial RM routine. Convolution with a real instrument response function was always included in the fitting. While RM using Pearson's weighting of residuals can recover the correct mean result with a total number of counts of 1000 or more, ML distinguishes itself by yielding, in all cases, the same mean lifetime within 2% of the accepted value. For 200 total counts and greater, ML always provides a standard deviation of <10% of the mean lifetime, and even at 20 total counts there is only 20% error in the mean lifetime. The robustness of ML advocates its use for sparse data sets such as those acquired in some subdiffraction-limited microscopies, such as STED, and, more importantly, provides greater motivation for exploiting the time-resolved capacities of this technique to acquire and analyze fluorescence lifetime data.

Disciplines

Atomic, Molecular and Optical Physics | Chemistry | Physical Chemistry

Comments

This document is the Accepted Manuscript version of a Published Work that appeared in final form in Santra, Kalyan, Jinchun Zhan, Xueyu Song, Emily A. Smith, Namrata Vaswani, and Jacob W. Petrich. "What is the best method to fit time-resolved data? A comparison of the residual minimization and the maximum likelihood techniques as applied to experimental time-correlated, single-photon counting data." *The Journal of Physical Chemistry B* 120, no. 9 (2016): 2484-2490.

Authors

Kalyan Santra, Jinchun Zhan, Xueyu Song, Emily A. Smith, Namrata Vaswani, and Jacob W. Petrich

What Is the Best Method to Fit Time-resolved Data? A Comparison of the Residual Minimization and the Maximum Likelihood Techniques as Applied to Experimental Time-correlated, Single-photon Counting Data.

Kalyan Santra,[1,2] Jinchun Zhan,[3] Xueyu Song,[1,2] Emily A. Smith,[1,2] Namrata Vaswani,[3] and Jacob W. Petrich[1,1,2]

[1]Department of Chemistry, Iowa State University, Ames, Iowa 50011, USA

[2]U. S. Department of Energy, Ames Laboratory, Ames, Iowa 50011, USA

[3]Department of Electrical and Computer Engineering, Iowa State University, Ames, Iowa 50011, USA

Email: jwp@iastate.edu

Phone: +1 515 294 9422. Fax: +1 515 294 0105

ABSTRACT

1 To whom correspondence should be addressed.

The need for measuring fluorescence lifetimes of species in subdiffraction-limited volumes in, for example, stimulated emission depletion (STED) microscopy, entails the dual challenge of probing a small number of fluorophores and fitting the concomitant sparse data set to the appropriate excited-state decay function. This need has stimulated a further investigation into the relative merits of two fitting techniques commonly referred to as "residual minimization," RM, and "maximum likelihood," ML. Fluorescence decays of the well-characterized standard, rose bengal in methanol at room temperature (530 ± 10 ps), were acquired in a set of five experiments in which the total number of "photon counts" was approximately 20, 200, 1000, 3000, and 6000; and there were from about 2 to 200 counts at the maxima of the respective decays. Each set of experiments was repeated 50 times in order to generate the appropriate statistics. Each of the 250 data sets was analyzed by ML and two different RM methods (differing in the weighting of residuals) using in-house routines and compared with a frequently-used commercial RM routine. Convolution with a real instrument response function was always included in the fitting. While RM using Pearson's weighting of residuals can recover the correct mean result with a total number of counts of 1000 or more, ML distinguishes itself by yielding, in all cases, the same mean lifetime within 2% of the accepted value. For 200 total counts and greater, ML always provides a standard deviation of less than 10% of the mean lifetime; and even at 20 total counts there is only 20% error in the mean lifetime. The robustness of ML advocates its use for sparse data sets such as those acquired in some subdiffraction-limited microscopies, such as STED, and, more importantly, provides greater motivation for exploiting the time-resolved capacities of this technique to acquire and analyze fluorescence lifetime data.

INTRODUCTION

Time-resolved spectroscopic techniques provide an important portfolio of tools for investigating fundamental processes in chemistry, physics, and biology as well as for evaluating the properties of a wide range of materials.[1,2] One of the most powerful time-resolved techniques is that of time-correlated, single-photon counting (TCSPC), which is explained in detail in the texts by Fleming[1] and O'Conner and Phillips.[2] Traditionally, this method requires constructing a histogram of arrival time differences between an excitation pulse and pulse resulting from an emitted photon and fitting this histogram to an exponential decay (or perhaps, a sum of exponential decays in more complicated systems). We shall refer to this method of analysis as the Residual Minimization technique (RM). Phase fluorometry is an exception.[3-5] The quality of the histogram directly determines the quality of the fit, and hence, the accuracy of the extracted decay time. Thus, if the sample does not have a high fluorescence quantum yield (number of photons emitted per number of photons absorbed), one must collect data for a longer period of time in order to obtain a histogram of commensurate quality. This, however, is not always practical. For example, the sample may not have a high fluorescence quantum yield, or it may degrade after prolonged exposure to light. Figure 1 provides examples of such histograms.

The difficulties cited above are illustrated by a certain class of fluorescence microscopy experiments, in particular, those involving subdiffraction-limited spatial resolution, which usually require rapid data acquisition times and the use of fluorescent probes that may not be stable at the high laser powers that these techniques often require.[6,7] The experimental technique also limits the probe volume, thus reducing the concentration of excited-state fluorophores, and thereby contributing to the reduction of the fluorescence signal. One of the ways to overcome this is to bin the adjacent pixels of the image to increase the number of photons in the time

channels. This, however, compromises the spatial resolution, which is clearly undesirable in an experiment whose objective is super resolution imaging. We have recently discussed these difficulties as they pertain to stimulated emission depletion (STED) microscopy.[7] In particular, a major challenge in STED fluorescence lifetime imaging has been, as we have indicated above, collecting a sufficient number of photons with which to construct a histogram of photon arrival times from which a fluorescence lifetime may be extracted. We discussed[7] the utility of binning time channels in order to convert a sparse data set, whose histogram may bear a faint resemblance to an exponential decay, into a histogram that may be fit with sufficient accuracy to an exponential decay with a well-resolved time constant. An example of binning is given in the inset of the 200-count data set of Figure 1. One difficulty presented by binning time channels, however, is that it reduces the dynamic range over which the data are fit and thus renders the accurate determination of a time constant--or several time constants in a heterogeneous system--problematic.

An alternative to RM exists, however, in recognizing that given a certain model for the fluorescence decay, there is a well-defined probability of detecting a certain number of photons in a given bin (or channel) of the histogram. The time constant for fluorescence decay can thus be extracted by comparing this probability distribution function with the number of photons in the set of bins. In this technique, it is advantageous to maximize the number of bins used to construct the histogram. This method of analysis is referred to as the Maximum Likelihood technique (ML).[8]

Here we present a detailed and systematic comparison of RM with ML using the very-well characterized dye, rose bengal in methanol, as our standard (Figure 1). The excited state lifetime, τ, at 20°C in methanol is 530 ± 10 ps.[1] A more recent study gives 516 ps (with no

error estimate).[9] The fluorescence decay of rose bengal is collected over a total of 1024 bins in a set of five experiments in which the total number of arrival times (counts) in all the bins is approximately 20, 200, 1000, 3000, and 6000, respectively. Each set of experiments was repeated 50 times in order to obtain appropriate statistics. Each of the 250 fluorescence decays was analyzed using both RM and ML.

Analyzing data via RM and ML methods has, of course, been previously discussed.[8,10-29] With a few exceptions,[19,20,22,26] these analyses were limited to simulated data. Our work has been stimulated by the efforts of Maus *et al.*[20], who provided a careful and detailed comparison of the RM (to which they refer as LS, "least squares") and ML methods using experimental data. Maus *et al.* used Neyman[12,30,31] weighting in their RM analysis. They find that such weighting underestimates the mean lifetime. In addition, they find that ML effectively generates the correct lifetime down to about 1000 total counts, the lowest number of total counts that they considered. We have extended their analysis in two significant ways. In order to push the comparison between RM and ML as far as possible, our data sets were designed to be considerably sparser than those considered before, ranging from about 2 to 200 counts at the maximum of the respective fluorescence decays, whereas those of Maus *et al*. range from about 60 to 1300. We note that from 200 total counts and below, the data bear little or no resemblance to an exponential decay (Figure 1); and this is precisely where one might expect the distinction between RM and ML to be most marked. We also employ two different methods of weighting residuals in RM, that of Neyman and that of Pearson.[12,30,31] Our results are consistent with those of Maus et al. in that we also observe that Neyman weighting, except in one instance, underestimates the target answer. We find, however, that at 1000 total counts and greater, Pearson weighting affords an acceptable answer. Furthermore, and most importantly, we too find

that ML can be an effective analysis tool, but that its utility can be extended down to 200 total counts and even fewer. For example, at 20 total counts, the correct target lifetime is recovered with 20% error, which in some cases may be sufficiently accurate. Finally, we explicitly point out that the ML method (estimating the parameters that maximize the data likelihood under the assumed model) as it is traditionally and originally formulated[32] yields the exact same maximizers as the modified method introduced by Baker and Cousins[12] and employed by others,[19,20,22,25] which invokes a "likelihood ratio." Finally, we note for completeness that there are other methods of analysis[2,33-37], such as, for example, Bayesian[33,34], Laguerre expansion[35], and Laplace transform[2] analyses.

MATERIALS AND METHODS

Experimental Procedure

Rose bengal (Sigma) was purified by thin-layer chromatography using silica-gel plates and a solvent system of ethanol, chloroform, and ethyl acetate in a ratio of 25:15:30 by volume. Solvents were used without further purification. The R_f (retardation factor) value of the pure dye in this mixture was approximately 0.51. The purified dye was stored in methanol. Rose bengal absorbs in the region of 460-590 nm. Time-resolved data were collected using a home-made time-correlated, single-photon counting (TCSPC) instrument that employs a SPC-830 TCSPC module from Becker & Hickl GmbH. A Fianium pulsed laser (Fianium Ltd, Southampton, UK) operating at 570 nm and 2 MHz was used for the excitation of the sample. Emission was collected using a 590 nm long-pass filter. The instrumental response function was measured by collecting scattered light at 570 nm from the pure methanol solvent. The full-width at half-maximum of the instrument function was typically ~120 ps. Sparser data sets were obtained by attenuating the excitation laser beam with neutral density filters. The TCSPC data were

collected in 1024 channels, providing a time resolution of 19.51 ps/channel, and a full-scale time window of 19.98 ns. Experiments were performed at 19.7 ± 0.2°C. Five different data sets consisting of 50 fluorescence decays were collected with total counts of approximately 20, 200, 1000, 3000, and 6000, respectively. The photon arrival times are used to build histograms comprised of 1024 bins (channels).

Data Analysis

Modeling the Time-Correlated, Single-Photon Counting Data

Let t_j, j=1, 2,…, 1024 represent the center of the jth bin (or channel); and ϵ=19.51 ps, the time width of each bin in the histogram. Then, $t_1 = \epsilon/2$, $t_2 = t_1 + \epsilon$,... $t_j = t_1 + (j\text{-}1)\ \epsilon$,…, $t_{max} = t_{1024} = t_1 + 1023\epsilon$. Let $C(t)=\{c_1, c_2,\ldots, c_{1024}\}$ represent the set of counts obtained experimentally in all 1024 bins. Similarly, we can have $I(t)=\{ I_1, I_2,\ldots, I_{1024}\}$ as the set of counts for the experimentally measured IRF. We thus assume that the IRF consists of a series of 1024 delta pluses (δ-IRFs) having intensity $I_1, I_2,\ldots, I_{1024}$, respectively.

The probability that a photon is detected in the jth bin, p_j, is proportional to the convolution of the IRF and the model for the fluorescence decay.

$$p_j \mu \sum_{i=1}^{j-j_0} I_i e^{-\frac{t_{j-j_0}-t_i}{\tau}} \qquad (1)$$

where, j_0 is given by $b = j_0\epsilon$. b is a linear shift between the instrument response function and the fluorescence decay. This shift parameter is necessary because the lower energy ("redder") fluorescence photons travel at a different speed through dispersive optics than the higher energy ("bluer") excitation photons that are used to generate the IRF in a scattering experiment.[1,38,39]

The probability that a photon is detected in the range $t_1 \le t \le t_{max} = t_{1024}$ must be $\sum_j p_j = 1$. We have, therefore:

$$p_j = \frac{\sum_{i=1}^{j-j_0} I_i e^{-\frac{t_{j-j_0} - t_i}{\tau}}}{\sum_{j=1}^{1024} \left(\sum_{i=1}^{j-j_0} I_i e^{-\frac{t_{j-j_0} - t_i}{\tau}} \right)} = \frac{\sum_{i=1}^{j-j_0} I_i e^{-\frac{t_{j-j_0} - t_i}{\tau}}}{\sum_{k=1}^{1024} \left(\sum_{i=1}^{k-j_0} I_i e^{-\frac{t_{k-j_0} - t_i}{\tau}} \right)} \qquad (2)$$

The denominator acts as the normalization factor for the probability and it is independent of the index j. We can, therefore, change the dummy index, j, to another dummy index, k, for clarity, while retaining j_0, as this is a constant unknown shift applied for all bins. The denominator is proportional to the total convoluted counts generated from the IRF.

Let $\hat{c}_j$ represent the number of predicted counts from the single-exponential model in the jth bin, taking into account convolution. The number of predicted counts in a given bin is directly proportional to the probability that a photon is detected in that bin: $\hat{c}_j \propto p_j$. Thus, the sequence $\{\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_{1024}\}$ is the predicted data for a decay. The area under the decay curves obtained from the observed counts $C(t)$ and from the predicted counts $\hat{C}(t)$ must be conserved during optimization of the fitting parameters. In other words, the total number of predicted counts must be equal to the total number of observed photon counts. Therefore, the number of predicted counts in the jth bin is given by:

$$\hat{c}_j = C_T \frac{\sum_{i=1}^{j-j_0} I_i e^{-\frac{t_{j-j_0}-t_i}{\tau}}}{\sum_{k=1}^{1024}\left(\sum_{i=1}^{k-j_0} I_i e^{-\frac{t_{k-j_0}-t_i}{\tau}}\right)} \qquad (3)$$

where $C_T = \sum_j c_j$.

Finally, we note that the shift parameter, b, need not be an integral multiple of ϵ. If we assume that b can take continuous values, then we can always find an integer, j_0, such that $b = j_0\epsilon + \zeta$, where ζ lies between 0 and ϵ, the time width of the bin. The probability, p_j, and predicted number of counts, $\hat{c}_j$, are thus given by:

$$p_j = \frac{\sum_{i=1}^{j-j_0-1} I_i e^{-\frac{t_j-t_i-b}{\tau}}}{\sum_{k=1}^{1024}\left(\sum_{i=1}^{k-j_0-1} I_i e^{-\frac{t_k-t_i-b}{\tau}}\right)} \quad ; \quad \hat{c}_j = C_T \frac{\sum_{i=1}^{j-j_0-1} I_i e^{-\frac{t_j-t_i-b}{\tau}}}{\sum_{k=1}^{1024}\left(\sum_{i=1}^{k-j_0-1} I_i e^{-\frac{t_k-t_i-b}{\tau}}\right)} \qquad (4)$$

Residual Minimization Method (RM)

In this method, the sum of the squares of the residuals, as given in Eq. 5, is minimized over the parameters, τ and b, to obtain the optimal values.

$$S = \sum_{j=1}^{1024} (c_j - \hat{c}_j)^2 \qquad (5)$$

It is well established that minimization of the weighted square of the residuals provides a better fit than minimization of the unweighted square of the residuals.[12,19,40] We, therefore, construct a weighted square of the residuals:

$$S_w = \sum_j w_j (c_j - \hat{c}_j)^2 \tag{6}$$

where w_j is the weighting factor. Depending on the choice of w_j, Eq. 6 often takes the form of the classical chi squared, for example :[12,16,19,20,25,30,31,40]

$$\chi_P^2 = \text{Pearson's}\chi^2 = \sum_{j=1}^{1024} (c_j - \hat{c}_j)^2 / \hat{c}_j \tag{7}$$

or,

$$\chi_N^2 = \text{Neyman's}\chi^2 = \sum_{j=1}^{1024} (c_j - \hat{c}_j)^2 / c_j \tag{8}$$

The reduced χ^2 is obtained by dividing by the number of degrees of freedom:

$$\chi_{red}^2 = \frac{1}{n-p}\chi^2 \tag{9}$$

where n is the number of data points; and p, the number of parameters and constraints in the model. For example, in our case we have 1024 data points, two parameters (τ and b), and one constraint, $C_T = \hat{C}_T$. This gives $n - p$ =1021. For an ideal case, $\chi_{¿}^2$ will be unity; and $\chi_{¿}^2 < 1$ signifies overfitting the data. Therefore, the closer $\chi_{¿}^2$ is to unity (without being less than

unity), the better the fit. The program is run so as to vary τ and *b* in such a manner as to minimize χ^2_{ι} .

Maximum Likelihood Method (ML)

The total probability of having a sequence $\{c_1, c_2,\ldots, c_{1024}\}$ subject to the condition, $C_T = \sum_j c_j$, follows the multinomial distribution:

$$Pr(c_1, c_2, \ldots c_{1024}) = \frac{C_T!}{c_1! c_2! \ldots c_{1024}!} \prod_{j=1}^{1024} (p_j)^{c_j} = C_T! \prod_{j=1}^{1024} \frac{(p_j)^{c_j}}{c_j!} \qquad (10)$$

We can define a likelihood function as the joint probability density function above: $L(\hat{c}, c) = Pr(c_1, c_2, \cdots, c_{1024})$.

Substituting the expression for the probability using Eq. 4, we have:

$$\mathsf{L}(\hat{c}, c) = C_T! \prod_{j=1}^{1024} \frac{(\hat{c}_j / C_T)^{c_j}}{c_j!} \qquad (11)$$

Following the treatment of Baker and Cousins,[12] we let $\backslash\{c'\backslash\}$ represent the true value of $\{c\}$ given by the model. A likelihood ratio, λ, can be defined as:

$$\lambda = \mathsf{L}(\hat{c}, c) / \mathsf{L}(c', c) \qquad (12)$$

According to the likelihood ratio test theorem,[20,25,41,42] the "likelihood χ^2" is defined by

$$\chi^2_\lambda = -2\ln\lambda \tag{13}$$

which obeys a chi-squared distribution as the sample size (or number of total counts) increases.

For the multinomial distribution, we may replace the unknown $\{c'\}$ by the experimentally observed {c}.[12] This gives:

$$\lambda = \left[C_T ! \prod_{j=1}^{1024} \frac{(\hat{c}_j / C_T)^{c_j}}{c_j !} \right] / \left[C_T ! \prod_{j=1}^{1024} \frac{(c_j / C_T)^{c_j}}{c_j !} \right] = \prod_{j=1}^{1024} \left(\frac{\hat{c}_j}{c_j} \right)^{c_j} \tag{14}$$

And the "likelihood χ^2" becomes:

$$\chi^2_\lambda = -2\ln\lambda = -2\ln \prod_{j=1}^{1024} \left(\frac{\hat{c}_j}{c_j} \right)^{c_j} = 2\sum_{j=1}^{1024} c_j \ln\left(\frac{c_j}{\hat{c}_j} \right) \tag{15}$$

The minimization of the "likelihood χ^2," described in Eq. 15 , is thus performed to obtain the optimum values of τ and *b*.

It is important to stress that the form of the maximum likelihood method given in Eq. 10 is used widely by statisticians [32] and that Eq. 15, popularized by Baker and Cousins[12] and used in several instances to fit photon-counting data [19,20,22,25] is formally identical to it, as Baker and Cousins themselves point out. Namely, maximizing Eq. 10 is equivalent to minimizing Eq. 15. Specifically, from Eq. 10:

$$Pr(c_1, c_2, \ldots c_{1024}) = C_T ! \prod_{j=1}^{1024} \frac{(p_j)^{c_j}}{c_j !}$$

$$\ln Pr(c_1, c_2, \ldots c_{1024}) = const. + \sum_{j=1}^{1024} c_j \ln p_j = const. + \sum_{j=1}^{1024} c_j \ln \hat{c}_j$$

since, $p_j = \hat{c}_j / C_T$. The *const.* includes the terms involving only C_T or c_j, as they are experimentally observed numbers and independent of the parameters τ and *b*. From Eq. 15:

$$\chi^2_\lambda = 2\sum_{j=1}^{1024} c_j \ln\left(\frac{c_j}{\hat{c}_j}\right) = 2\sum_{j=1}^{1024} c_j \ln c_j - 2\sum_{j=1}^{1024} c_j \ln \hat{c}_j = const. - 2\sum_{j=1}^{1024} c_j \ln \hat{c}_j$$

Again the *const.* includes the terms which are independent of the parameters τ and *b*. Equation 10 may be considered to be simpler in form than Eq. 15 and, for some models, may prove to be less computationally expensive as well.

For completeness, we mention the Bayesian analysis, which offers another approach in terms of a likelihood function. The Bayesian analysis starts with a prior distribution of the parameters in the appropriate range. The "posterior distribution" is calculated using the likelihood of the observed distribution for a given "prior distribution."[33,34] In the case of our model system, let $P(\tau,b)$ represent the prior distribution of the parameters. We can write the likelihood of having an observed distribution, $\{c\}=\{c_1, c_2,\ldots, c_{1024}\}$, subject to the prior distribution as $Pr(\{c\}|\,\tau,b)$. Therefore, the posterior distribution is given by:

$$P(\tau, b \mid \{c\}) = \frac{P(\tau,b) Pr(\{c\} \mid \tau, b)}{\int d\tau' db' P(\tau', b') Pr(\{c\} \mid \tau', b')} \qquad (16)$$

where the denominator is acting as the normalization factor. Maximization of the posterior distribution will furnish the desired value of the parameters. The results are often greatly

affected by the choice of the prior distribution. Usually the prior distribution is chosen in such a way that the entropy of the distribution is maximized.

Computational Tools

The RM and ML analyses described above are performed using codes written in MATLAB. We employ the GlobalSearch toolbox, which uses the "fmincon" solver. In each calculation, a global minimum was found. Finally, for comparison, the data were also analyzed with the proprietary SPCImage software v. 4.9.7 (SPCI), provided by Becker & Hickl GmbH. As this program is based upon a method of RM, it should, in principle, perform identically to our in-house code. In all the fitting comparisons to be discussed, there are only two variable parameters, the lifetime (τ), and the shift parameter (b), see below. With our in-house routines, we experimented with different initial values of the lifetime and shift parameters, ranging from 0.3 to 0.7 ns and from -0.02 to 0.02 ns, respectively. In all cases, we retrieved the same fit results through the third decimal place.

RESULTS AND DISCUSSION

Each of the 250 fluorescence decays for the five sets of data (taken with approximately 20, 200, 1000, 3000, and 6000 total counts) is analyzed by the four methods described above: ML; RM-Neyman; RM-Pearson; and the commercial SPCI. As noted, the ML results obtained from Eq. 10 and Eq. 15 are formally identical; and the fits obtained using the two equations yield the same results. Figure 1 presents a sample decay from each of the five data sets. Figure 2(a) provides a scatter plot of each lifetime obtained for each method of fitting. The horizontal red dashed line represents the value of a recently acquired lifetime of rose bengal in methanol at room temperature of 516 ps,[9] which we use as reference. Histograms of lifetimes obtained for the different fitting methods are presented in Figures 2(b)-(f). The mean (average) lifetime plus

or minus one standard deviation, $\langle \tau \rangle \pm \sigma$, obtained from the results are computed and summarized in Table I.

The salient results are the following. Concerning the RM methods, we note that because the SPCI source code is not available, the details of the differences arising between it and our code cannot be determined. One noticeable and important difference between SPCI and our RM (Table I) is that SPCI does not converge for the 20-total-counts data set. On the other hand, our RM-Neyman and RM-Pearson methods fit the data in all cases, but with varying degrees of success. Except for the case of 200 counts, RM-Neyman consistently underestimates the target value. For 200 counts, all RM methods overestimate the target value, and SPCI yields an aberrant result of 600 ± 700 ps. From 1000 counts onward, RM-Pearson provides results close to those of the target value and similar to those of SPCI. RM-Pearson appears to be more robust and reliable than either RM-Neyman or SPCI.

In contrast, at 20 counts, ML yields 500 ± 100 ps, which brackets the target result and which is to be compared with 320 ± 30 ps for RM-Neyman and with 460 ± 70 ps for RM-Pearson. For 200 total counts and greater, ML always provides an acceptable result with a standard deviation of less than 10% of the mean lifetime. The RM techniques achieve this level of precision only as of 1000 counts; and, as mentioned above, RM-Neyman generally underestimates the target value. Perhaps the most significant difference among the ML and the RM methods is that ML, within 2%, always produces the same mean lifetime, whereas this is not the case for RM, especially for total counts of 1000 and less.

In the Introduction, we commented on the careful comparison of the RM and ML methods by Maus *et al.*[20] and noted that our results presented here are not only consistent with theirs but also suggest that the ML method can be extended to considerably fewer counts than

they explored in their study. We summarize some of the more important differences between our work and that of Maus *et al*.

1. Our data sets were designed to be considerably sparser than those considered before, ranging from about 2 to 200 counts *at the maximum* of the respective fluorescence decays, whereas those of Maus *et al*. range from about 60 to 1300. From 200 total counts and below, the data bear little or no resemblance to an exponential decay (Figure 1); and this is precisely where one might expect the distinction between RM and ML to be the greatest—and the most useful.

2. Maus *et al.* use only 180 time channels (140 ps/channel) to study a molecule (hexaphenylbenzene-perylenemonoimide) whose lifetime is ~4500 ps, whereas we have used 1024 time channels (19.51 ps/channel) to study rose bengal, whose lifetime is ~530 ps. In other words, our experimental conditions (both the time window and the excited-state lifetime under consideration) are determined to distribute the data over as many time channels as possible in order to minimize the effects of time-binning, which we have discussed elsewhere[7] and to highlight instances where the differences between ML and RM might be the most pronounced.

3. There are some subtle but significant differences in the details of the fitting procedures. For example, we argue that it is necessary to conserve the total number of counts (which is proportional to the area under the fitted curve) during the optimization process. Maus *et al.*, however, permit the amplitude (our total counts) to vary for RM but keep it fixed for ML. Also, all of our fitting comparisons involve two variable parameters, the lifetime and the shift, τ and b. Maus *et al*. only have one variable parameter for ML, τ; but they employ two for RM, τ and the amplitude. We suggest that a close comparison between the methods should maintain as many similarities as possible.

In addition, we note that Köllner and Wolfrum[8] have discussed the use of ML. They suggested, based on simulations (some including 20% of a constant background), that one needs to have at least 185 photon counts in a time window of 8 ns with 256 time channels to measure a 2.5-ns lifetime with 10% variance without background.

CONCLUSIONS

We have performed a comparison of the maximum likelihood (ML) and residual minimization (RM) fitting methods by applying them to experimental data incorporating a convoluted instrument function. While RM using Pearson's weighting of residuals can recover the correct mean result with a total number of counts of 1000 or more, ML distinguishes itself by yielding, in all cases, the same mean lifetime within 2% of the accepted value. For total counts of 200 and higher, ML always provides a standard deviation of less than 10% of the mean lifetime. Even at 20 total counts, ML provides a 20% error. The robustness of ML advocates its use for sparse data sets such as those acquired in some subdiffraction-limited microscopies, such as STED, and, more importantly, provides greater motivation for exploiting the time-resolved capacities of this technique to acquire and analyze fluorescence lifetime data.

ACKNOWLEDGMENTS

We thank Mr. Ujjal Bhattacharjee for assistance in the early stages of this work. The work performed by K. Santra, E. A. Smith, and J. W. Petrich was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences through the Ames Laboratory. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. DE-AC02-07CH11358. X. Song was supported by The Division of Material Sciences and Engineering, Office of Basic Energy Sciences, U.S. Department of Energy, under Contact No. W-7405-430 ENG-82 with

Iowa State University. The work of J. Zhan and N. Vaswani was partly supported by grant CCF-1117125 from the National Science Foundation.

REFERENCES

1. Fleming, G. R. *Chemical Application of Ultrafast Spectroscopy* Oxford University Press: New York, 1986.

2. O'Connor, D. V.; Phillips, D. *Time Correlated Single Photon Counting* Academic Press Inc.: London, 1984.

3. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer: 2011.

4. Stringari, C.; Cinquin, A.; Cinquin, O.; Digman, M. A.; Donovan, P. J.; Gratton, E. Phasor Approach to Fluorescence Lifetime Microscopy Distinguishes Different Metabolic States of Germ Cells in a Live Tissue. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (33), 13582-13587.

5. Colyer, R. A.; Lee, C.; Gratton, E. A Novel Fluorescence Lifetime Imaging System That Optimizes Photon Efficiency. *Microsc. Res. Techniq.* **2008**, *71* (3), 201-213.

6. Lesoine, M. D.; Bose, S.; Petrich, J. W.; Smith, E. A. Supercontinuum Stimulated Emission Depletion Fluorescence Lifetime Imaging. *J. Phys. Chem. B*. **2012**, *116* (27), 7821-7826.

7. Syed, A.; Lesoine, M. D.; Bhattacharjee, U.; Petrich, J. W.; Smith, E. A. The Number of Accumulated Photons and the Quality of Stimulated Emission Depletion Lifetime Images. *Photochem. Photobiol.* **2014**, *90* (4), 767-772.

8. Köllner, M.; Wolfrum, J. How Many Photons Are Necessary for Fluorescence-Lifetime Measurements? *Chem. Phys. Lett.* **1992**, *200* (1), 199-204.

9. Luchowski, R.; Szabelski, M.; Sarkar, P.; Apicella, E.; Midde, K.; Raut, S.; Borejdo, J.; Gryczynski, Z.; Gryczynski, I. Fluorescence Instrument Response Standards in Two-Photon Time-Resolved Spectroscopy. *Appl. Spectrosc.* **2010**, *64* (8), 918-922.

10. Ankjærgaard, C.; Jain, M.; Hansen, P. C.; Nielsen, H. B. Towards Multi-Exponential Analysis in Optically Stimulated Luminescence. *J. Phys. D: Appl. Phys.* **2010**, *43* (19), 195501.

11. Bajzer, Ž.; Therneau, T. M.; Sharp, J. C.; Prendergast, F. G. Maximum Likelihood Method for the Analysis of Time-Resolved Fluorescence Decay Curves. *Eur. Biophys. J.* **1991**, *20* (5), 247-262.

12. Baker, S.; Cousins, R. D. Clarification of the Use of Chi-Square and Likelihood Functions in Fits to Histograms. *Nucl. Instr. Meth. Phys. Res.* **1984**, *221* (2), 437-442.

13. Bevington, P.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences*, 3rd ed.; McGraw-Hill: New York, 2002.

14. Grinvald, A.; Steinberg, I. Z. On the Analysis of Fluorescence Decay Kinetics by the Method of Least-Squares. *Anal. Biochem.* **1974**, *59* (2), 583-598.

15. Hall, P.; Selinger, B. Better Estimates of Exponential Decay Parameters. *J. Phys. Chem.* **1981**, *85* (20), 2941-2946.

16. Hauschild, T.; Jentschel, M. Comparison of Maximum Likelihood Estimation and Chi-Square Statistics Applied to Counting Experiments. *Nucl. Instr. Meth. Phys. Res. A*. **2001**, *457* (1), 384-401.

17. Hinde, A. L.; Selinger, B.; Nott, P. On the Reliability of Fluorescence Decay Data. *Aust. J. Chem.* **1977**, *30* (11), 2383-2394.

18. Kim, G.-H.; Legresley, S. E.; Snyder, N.; Aubry, P. D.; Antonik, M. Single-Molecule Analysis and Lifetime Estimates of Heterogeneous Low-Count-Rate Time-Correlated Fluorescence Data. *Appl. Spectrosc.* **2011**, *65* (9), 981-990.

19. Laurence, T. A.; Chromy, B. A. Efficient Maximum Likelihood Estimator Fitting of Histograms. *Nat. Methods*. **2010**, *7* (5), 338-339.

20. Maus, M.; Cotlet, M.; Hofkens, J.; Gensch, T.; De Schryver, F. C.; Schaffer, J.; Seidel, C. An Experimental Comparison of the Maximum Likelihood Estimation and Nonlinear Least-Squares Fluorescence Lifetime Analysis of Single Molecules. *Anal. Chem.* **2001**, *73* (9), 2078-2086.

21. Moore, C.; Chan, S. P.; Demas, J.; DeGraff, B. Comparison of Methods for Rapid Evaluation of Lifetimes of Exponential Decays. *Appl. Spectrosc.* **2004**, *58* (5), 603-607.

22. Nishimura, G.; Tamura, M. Artefacts in the Analysis of Temporal Response Functions Measured by Photon Counting. *Phys. Med. Biol.* **2005**, *50* (6), 1327.

23. Periasamy, N. Analysis of Fluorescence Decay by the Nonlinear Least Squares Method. *Biophys. J.* **1988**, *54* (5), 961-967.

24. Sharman, K. K.; Periasamy, A.; Ashworth, H.; Demas, J. Error Analysis of the Rapid Lifetime Determination Method for Double-Exponential Decays and New Windowing Schemes. *Anal. Chem.* **1999**, *71* (5), 947-952.

25. Turton, D. A.; Reid, G. D.; Beddard, G. S. Accurate Analysis of Fluorescence Decays from Single Molecules in Photon Counting Experiments. *Anal. Chem.* **2003**, *75* (16), 4182-4187.

26. Tellinghuisen, J.; Goodwin, P. M.; Ambrose, W. P.; Martin, J. C.; Keller, R. A. Analysis of Fluorescence Lifetime Data for Single Rhodamine Molecules in Flowing Sample Streams. *Anal. Chem.* **1994**, *66* (1), 64-72.

27. Enderlein, J.; Köllner, M. Comparison between Time-Correlated Single Photon Counting and Fluorescence Correlation Spectroscopy in Single Molecule Identification. *Bioimaging*. **1998**, *6* (1), 3-13.

28. Bialkowski, S. E. Data Analysis in the Shot Noise Limit. 1. Single Parameter Estimation with Poisson and Normal Probability Density Functions. *Anal. Chem.* **1989**, *61* (22), 2479-2483.

29. Bialkowski, S. E. Data Analysis in the Shot Noise Limit. 2. Methods for Data Regression. *Anal. Chem.* **1989**, *61* (22), 2483-2489.

30. Neyman, J.; Pearson, E. S. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I. *Biometrika*. **1928**, *20A* (1/2), 175-240.

31. Neyman, J.; Pearson, E. S. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part Ii. *Biometrika*. **1928**, *20A* (3/4), 263-294.

32. Poor, H. V. An Introduction to Signal Detection and Estimation, 2nd ed.; Springer: 1994.

33. Barber, P.; Ameer-Beg, S.; Pathmananthan, S.; Rowley, M.; Coolen, A. A Bayesian Method for Single Molecule, Fluorescence Burst Analysis. *Biomed. Opt. Express.* **2010**, *1* (4), 1148-1158.

34. Rowley, M. I.; Barber, P. R.; Coolen, A. C.; Vojnovic, B. Bayesian Analysis of Fluorescence Lifetime Imaging Data. *SPIE BiOS* International Society for Optics and Photonics: 2011, 790325-790325-790312.

35. Jo, J. A.; Fang, Q.; Marcu, L. Ultrafast Method for the Analysis of Fluorescence Lifetime Imaging Microscopy Data Based on the Laguerre Expansion Technique. *IEEE J. Sel. Top. Quantum Electron.* **2005**, *11* (4), 835-845.

36. Večeř, J.; Kowalczyk, A.; Davenport, L.; Dale, R. Reconvolution Analysis in Time-Resolved Fluorescence Experiments--an Alternative Approach: Reference-to-Excitation-to-Fluorescence Reconvolution. *Rev. Sci. Instrum.* **1993**, *64* 3413-3424.

37. Istratov, A. A.; Vyvenko, O. F. Exponential Analysis in Physical Phenomena. *Rev. Sci. Instrum.* **1999**, *70* (2), 1233-1257.

38. Calligaris, F.; Ciuti, P.; Gabrielli, I.; Giamcomich, R.; Mosetti, R. Wavelength Dependence of Timing Properties of the Xp 2020 Photomultiplier. *Nucl. Instr. Meth.* **1978**, *157* (3), 611-613.

39. Sipp, B.; Miehe, J.; Lopez-Delgado, R. Wavelength Dependence of the Time Resolution of High-Speed Photomultipliers Used in Single-Photon Timing Experiments. *Opt. Commun.* **1976**, *16* (1), 202-204.

40. Jading, Y.; Riisager, K. Systematic Errors in X 2-Fitting of Poisson Distributions. *Nucl. Instr. Meth. Phys. Res. A*. **1996**, *372* (1), 289-292.

41. Wilks, S. The Likelihood Test of Independence in Contingency Tables. *Ann. Math. Stat.* **1935**, *6* (4), 190-196.

42. Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann. Math. Stat.* **1938**, *9* (1), 60-62.

Table I

Total counts	ML	RM		SPCI
		Neyman	Pearson	
20	500 ± 100	320 ± 30	460 ± 70	
200	510 ± 40	690 ± 20	600 ± 50	600 ± 700
1000	510 ± 20	490 ± 30	560 ± 20	520 ± 30
3000	510 ± 10	480 ± 20	540 ± 10	520 ± 20
6000	501 ± 8	480 ± 10	520 ± 20	520 ± 10

Table I. Mean lifetime ± One Standard Deviation (ps) Associated with Each Method of Analysis ML, maximum likelihood method; RM-Neyman, residual minimization method weighting the residuals by $1/c_j$, where c_j is the number of counts in a channel (Eq. 8); RM- Pearson, residual minimization method weighting the residuals by $1/\hat{c}_j$, where $\hat{c}_j$ is the predicted number of counts in a channel (Eq. 7); SPCI, commercially supplied residual minimization software.

Figure 1

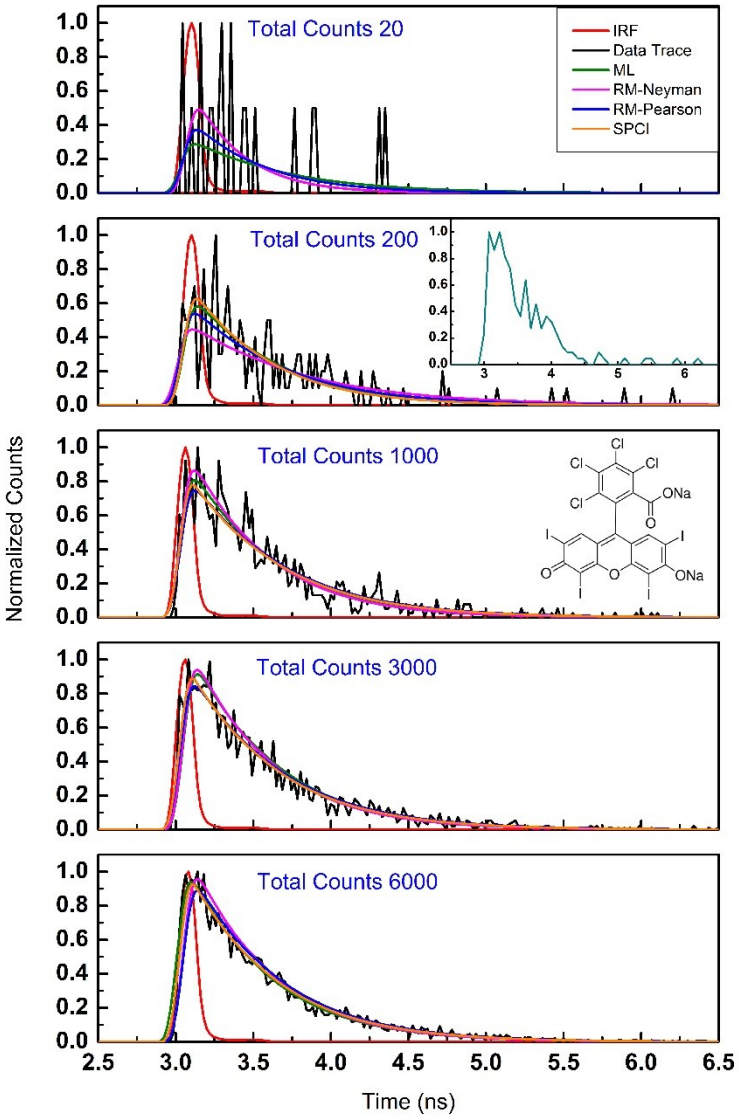

Figure 1. A representative histogram for a given number of total counts is presented. Each panel gives the raw data (black), the instrument response function (IRF, red), the ML fit (green), the RM-Neyman fit (magenta), the RM-Pearson fit (blue), and the SPCI fit (orange). The inset in the 200-count panel gives the result of binning four contiguous time channels, reducing the number from 1024 to 256. The inset in the 1000-count panel presents the structure of the sodium salt of rose bengal.

Figure 2

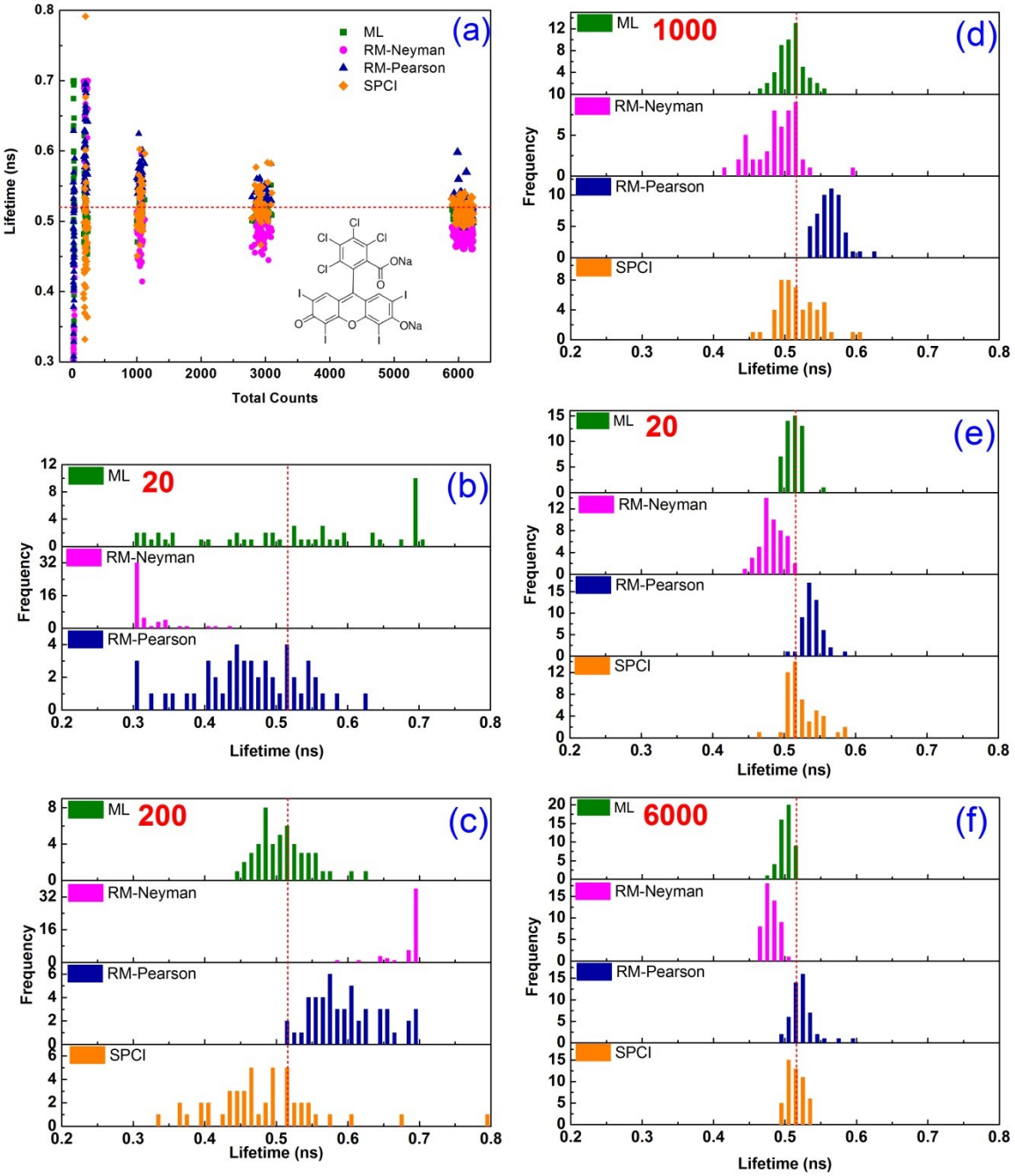

Figure 2. Estimated lifetime of rose bengal by ML (green), RM-Neyman (magenta), RM-Pearson (blue) and SPCI (orange). (a) The scatter plot of the lifetime with respect to the total

counts in a decay. (b)-(f) Histograms of the lifetimes obtained by the above four methods for total counts of 20, 200, 1000, 3000, and 6000 respectively. The bins for all of the histograms are 10 ps wide. The red dashed lines give, as a benchmark, a recent value of τ = 516 ps.[9]

TOC Graphic