Confidence Intervals for Concentration and Brightness from Fluorescence Fluctuation Measurements

Kenneth M. Pryse,[†△*] Xi Rong,[‡△] Jordan A. Whisler,[‡] William B. McConnaughey,[†] Yan-Fei Jiang,[†] Artem V. Melnykov,[†] Elliot L. Elson,[†‖] and Guy M. Genin[‡‖*]

[†]Department of Biochemistry and Molecular Biophysics and [‡]Department of Mechanical Engineering and Materials Science, Washington University in St. Louis, St. Louis, Missouri

ABSTRACT The theory of photon count histogram (PCH) analysis describes the distribution of fluorescence fluctuation amplitudes due to populations of fluorophores diffusing through a focused laser beam and provides a rigorous framework through which the brightnesses and concentrations of the fluorophores can be determined. In practice, however, the brightnesses and concentrations of only a few components can be identified. Brightnesses and concentrations are determined by a nonlinear least-squares fit of a theoretical model to the experimental PCH derived from a record of fluorescence intensity fluctuations. The χ^2 hypersurface in the neighborhood of the optimum parameter set can have varying degrees of curvature, due to the intrinsic curvature of the model, the specific parameter values of the system under study, and the relative noise in the data. Because of this varying curvature, parameters estimated from the least-squares analysis have varying degrees of uncertainty associated with them. There are several methods for assigning confidence intervals to the parameters, but these methods have different efficacies for PCH data. Here, we evaluate several approaches to confidence interval estimation for PCH data, including asymptotic standard error, likelihood joint-confidence region, likelihood confidence intervals, skew-corrected and accelerated bootstrap (BCa), and Monte Carlo residual resampling methods. We study these with a model two-dimensional membrane system for simplicity, but the principles are applicable as well to fluorophores diffusing in three-dimensional solution. Using simulated fluorescence fluctuation data, we find the BCa method to be particularly well-suited for estimating confidence intervals in PCH analysis, and several other methods to be less so. Using the BCa method and additional simulated fluctuation data, we find that confidence intervals can be reduced dramatically for a specific non-Gaussian beam profile.

INTRODUCTION

As a fluorescent particle diffuses through a laser excitation beam, pulses of light are emitted that are proportional to the particle brightness, defined as the number of photons emitted per second when the molecule is at the position of maximum intensity of the laser beam. In the absence of electronic interactions among fluorophores, the brightness is proportional to the number of fluorophores in the diffusing particle. Hence, measurements of brightness provide an indicator of fluorophore aggregation. Indeed, fluorescence correlation spectroscopy (FCS) yields the average brightness and concentration of fluorescent particles in terms of the mean fluorescence and the zero-time amplitude of the fluorescence fluctuation autocorrelation function (1,2). However, it would also be very useful to have some indication of the distribution of concentrations and brightnesses of a mixture of fluorescent particles. Several methods based on moments of the fluorescence fluctuations (3), high-order fluorescence correlation functions (4), the photon count histogram (PCH) (5), and fluorescence intensity distribution analysis (6,7), which is based on the PCH, have been developed for this purpose. Our use of PCH analysis to study aggregation in model membranes has led us to investigate its limits of applicability in a two-dimensional (membrane) system. Our overall approach is applicable to three-dimensional systems as well, and we present a careful analysis of this.

The two-dimensional membrane system chosen as a model for this purpose is of interest in its own right as a framework for detecting nanoscopic clusters of lipid and protein molecules in lipid bilayer membranes and for characterizing their sizes. One application is the formation in the cell plasma membrane of nanodomains called "rafts" that are vital to several important biological functions (8–10). Properly chosen fluorescent lipid molecules will selectively partition in a nanodomain, and, if there are no substantial electronic interactions among these fluorescent lipid molecules, the brightness of a nanodomain containing n fluorescent lipid probes is n times that of a single probe. If the number of fluorescent lipid or lipid probe molecules scales with the size of the nanodomain, PCH analysis enables identification of relative sizes of nanodomains, and estimation of the distribution of their sizes, even below the resolution limit of light microscopy. Similarly, PCH has been used to detect clusters of epidermal growth factor receptors on the surfaces of living cells in culture (7).

The analysis involves fitting a model that specifies the numbers and brightnesses of the fluorescent species in the

Submitted January 23, 2012, and accepted for publication July 10, 2012.

[△]Kenneth M. Pryse and Xi Rong contributed equally to this work.

[‖]Elliott L. Elson, and Guy M. Genin contributed equally to this work.

*Correspondence: pryse@wustl.edu or genin@wustl.edu

Jordan A. Whisler's present address is Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Editor: Petra Schwille.

0006-3495/12/09/0898/9 $2.00

http://dx.doi.org/10.1016/j.bpj.2012.07.045

system to the experimental PCH through a nonlinear least-squares procedure by finding the set of parameter values (the maximum-likelihood estimates) that minimizes the χ^2 error between the model prediction and the data. We show that, due to the relatively shallow curvature of the χ^2 hypersurface in the neighborhood of the optimum, PCH analysis can yield optimal estimates of parameters with high degrees of uncertainty unless applied to impractically large datasets. Statistical theory affords different approaches to estimate the uncertainties of parameters, and these methods have different efficacies for PCH data. An earlier analysis of the statistical errors of estimated parameters, presumably by the asymptotic standard error method, focused on the dependence of the errors on time bin duration, concentration, and brightness (11).

Modern implementations of nonlinear least-squares regressions enable scientists to fit complex, multiparameter models to large data sets. High-level programming languages combined with packaged data analysis software make such complex analyses routine for most physical and biological scientists. The determination of the optimal parameter values is the primary objective, but an important secondary goal, that of estimating the uncertainty of the optimal parameter values, is often underappreciated and not well understood, despite efforts to remedy this (12,13). The uncertainty of the parameter values is quantified by the calculation of confidence regions or confidence intervals. It is assumed that a set of true parameter values exists that give rise to the experimental data (which also contains random noise). It is the size of the noise relative to the signal that makes estimation of the underlying true parameters uncertain. Probability theory shows that there is a range of values of the parameters within which one can be certain to a specified probability the true values of the parameters lie. The most common method to calculate confidence intervals of model parameters is the asymptotic standard error (14). This method is exact for linear models, but can be quite inaccurate for highly nonlinear models. An extreme example is given of a microwave absorption data set for which the nominal 95% asymptotic confidence intervals include the true values only 10.8% of the time (15). The asymptotic confidence intervals are symmetric about the optimal value, which is not the case for many nonlinear models.

An important distinction should be made between joint confidence regions for all parameters and confidence intervals for each parameter separately. A linear model with two parameters, p_1 and p_2, will have a joint confidence region in the p_1-p_2 plane that is a circle (with proper scaling of the axes) if there are no parameter interactions, where a change in one parameter can compensate for a change in the other, and an ellipse with major axis lying between the parameter axes if there are compensatory parameter interactions. Nonlinear models can have confidence regions that are highly curved, disjoint, or even unbounded along one parameter axis (14,16). With three or more parameters, multidimensional joint confidence regions can be difficult to visualize and understand. Confidence intervals calculated for each parameter are generally more useful for the practicing scientist seeking to understand the limits of data analyses.

In this article, we evaluate several approaches to confidence interval estimation for PCH analysis, including asymptotic standard error, likelihood joint-confidence region, likelihood confidence interval, skew-corrected and accelerated bootstrap (BCa), and a Monte Carlo residual resampling method. Using simulated fluorescence fluctuation data, we find the BCa method to be particularly well-suited for estimating confidence intervals in analysis of PCHs collected from fluorophores diffusing on a membrane, and several other methods to be ill-suited. We find both the BCa method and asymptotic standard error to be suitable for studying fluorophores diffusing in a three-dimensional liquid. Using the BCa method and additional simulated fluctuation data for fluorophores diffusing on a membrane, we confirm earlier predictions (e.g., (17)) that confidence intervals can be reduced dramatically for specific non-Gaussian beam profiles.

METHODS

Statement of problem

We consider a distribution of J types of fluorophores diffusing either on a flat membrane or within a three-dimensional solution. A single fluorescent lipid probe has one fluorophore, and may be 1) outside any fluorophore cluster, or 2), diffusing independently, or 3), within a cluster that contains additional fluorescent lipids. Because the fluorescent lipids in a cluster diffuse as a unit, we refer for convenience to single fluorescent lipids and clusters with more than one fluorescent lipid simply as fluorophores of different brightness. We define the brightness, q_j ($j = 1, 2, \ldots J$), of each of the J species of fluorophores as the number of photons per unit time emitted when a member of the species is at the position of maximum intensity of an illuminating laser. The brightness q_j is related to the fluorescence yield Q, absorbance ε, and the number of fluorophores, n_j, in the species by $q_j = n_j g \varepsilon Q B_o$, where g accounts for losses and geometric effects in the imaging system.

The beam excitation intensity is Gaussian,

$$B(r,z) = B_0 \exp\left(-\frac{2r^2}{\omega^2} - \frac{2z^2}{\omega_z^2}\right), \quad (1)$$

where z is the position perpendicular to the focus plane, r is the radial position parallel to the focus plane and measured from the point of peak excitation, B_0 is the peak intensity of the beam, and ω and ω_z are the characteristic length-scales of the Gaussian intensity profile, i.e., the nominal (exp(–2)) beam radius. $B(r)$ is determined by the combination of laser excitation and fluorescence detection efficiency, normalized so that $B(0) = B_0 = 1$. The average number of each type of fluorophore in the observational beam volume ($\pi^{3/2}\omega^2\omega_z$) is n_j (e.g., Rüttinger et al. (18)).

For the two-dimensional case, the excitation intensity is planar ($\omega_z \gg \omega$) and Eq. 1 simplifies to a planar Gaussian profile, $B(r) = B_o \exp(-2r^2/\omega^2)$. In two dimensions, n_j is the average number of fluorophores of type j that appear in the nominal beam area defined as $\pi\omega^2$ (1). The goal in both cases is to find the brightnesses q_j and the average numbers n_j of each type of fluorophore.

PCH analysis

During an experiment, the numbers of photons registered during time windows of duration T, called bins, are counted and the fraction of bins containing m photons is plotted versus m to create the experimental PCH. T is chosen to be less than one-fifth the smallest diffusion time of the system under study, so that the molecular motion during each bin time is small and yet the bin time is as long as possible to minimize shot noise. For a single fluorophore component the probability $P(m)$ of measuring m photon counts from an area illuminated with uniform laser intensity is

$$P(m) = \sum_{n=0}^{\infty} P(m|n)P(n) = \sum_{n=0}^{\infty}\left(\frac{(nqt)^m}{m!}e^{-nqT}\right)\left(\frac{\langle n\rangle^n}{n!}e^{-\langle n\rangle}\right), \quad (2)$$

where n and $\langle n\rangle$ denote the number and average number of fluorophores in the illuminated area, and $P(m|n)$ is the probability of detecting m photons if n fluorophores are in this region (6). Both $P(n)$ and $P(m|n)$ can be represented as Poisson distributions in which q is the emission rate (photons per s) of a fluorophore, and nqT is the mean number of photons emitted per bin interval when n fluorophores are in the sampling region.

To account both for systems with several species of fluorophores with different brightnesses (i.e., different numbers of individual fluorescent lipids) and also for the spatial variation of the excitation intensity (a two- or three-dimensional Gaussian in this case), it is useful to express the generating function, $G(\xi)$, of the PCH distribution (6). The generating function is

$$G(\xi) = \sum_{m=0}^{\infty} P(m)\xi^m. \quad (3)$$

When $\xi = \exp(i\mu)$, is chosen, with $i = \sqrt{-1}$, $P(m)$ becomes the Fourier transform of its generating function, with μ the Fourier space variable. For a single diffusing species of concentration n_j and brightness q_j, the series method of Meng and Ma (19) was used as

$$\ln\big(G_j(\xi)\big) = n_j\sum_{k=1}^{\infty}\frac{(\xi-1)^k}{k!}\gamma_k\big(q_jT\big)^k, \quad (4)$$

where

$$\gamma_k = \int_{-\infty}^{+\infty}\int_{0}^{+\infty}(B(r,z))^k r\,dr\,dz$$

for three dimensions and

$$\gamma_k = \int_{0}^{+\infty}(B(r))^k r\,dr$$

for two dimensions. For the Gaussian beam shapes of Eq. 1, the following recursion relations are convenient: $\gamma_k = \gamma_1/k^{3/2}$ for three dimensions and $\gamma_k = \gamma_1/k$ for two dimensions.

Conventional PCH (6) is adapted easily to measurements on membranes (7). Adopting a two-dimensional Gaussian shape for the area illuminated on the bilayer surface, we note that, because the system is quasi-two-dimensional, it is not necessary to characterize the shape of the laser-illuminated volume along the optical axis. For experiments on giant unilamellar vesicles (GUVs), we suppose that the curvature of the GUV is small over distances comparable to ω. For a GUV with radius ≥ 10 μm, this will be valid for typical values of ω in the range of 0.3 μm.

The generating function of a system with J species of fluorophores diffusing in two or three dimensions, each with a possibly unique concentration n_j and unique brightness q_j, can be written via the following superposition:

$$\ln(G(\xi)) = \sum_{j=1}^{J}\ln\big(G_j(\xi)\big) = \sum_{j=1}^{J}\sum_{k=1}^{\infty}\frac{(\xi-1)^k}{k!}n_j\gamma_k\big(q_jT\big)^k. \quad (5)$$

Simulations

Monte Carlo simulations were performed on idealized monomer/tetramer systems ($J = 2$, $q_2 = 4q_1$) using MATLAB (The MathWorks, Natick, MA). The monomer fluorophores and nanodomains containing four fluorophores were assumed to diffuse at the same speed.

For simulations of fluorophores on a planar membrane, populations of prescribed numbers of monomers and tetramers were randomly dispersed over a circular region of radius 15ω. At each time step, each fluorophore shifted a distance $\Delta = 0.15\omega$ in a random direction from the previous time step. Each species then emitted a random number p of photons according to a Poisson distribution $P(p)$, with the mean emission rate (photons per time bin of duration T) varying with radial position of the fluorophore and the peak emission rate q_j:

$$P(p) = \frac{\big(q_jTB(r)\big)^p}{p!}\exp\big(-q_jTB(r)\big). \quad (6)$$

The time- and distance scales of the steps are related to the diffusion constant D of the fluorophores, which can be measured independently using standard FCS techniques,

$$\frac{\Delta^2}{4T} = D = \frac{\omega^2}{4\tau_D}, \quad (7)$$

where τ_D is the FCS diffusion time. Thus, $T = 0.0225\ \tau_D$, fulfilling the requirement that the time bin be short relative to the diffusion time. The brightness values in the simulations are in photons/bin time, are in the upper range of brightness values that we have seen experimentally in GUVs, and correspond to several kHz/molecule.

For the coverage analysis of the three-dimensional system, 500 PCHs were generated by adding binomial noise (5) to an ideal PCH, using MATLAB's "binornd" function with 10^5 trials.

Simulated data took the form of time histories of photon counts $F(sT)$, $s = 1, 2, \ldots S$, during each of S time bins of duration T. For the simulations in this study, $S = 10^5$. From these data, a PCH was derived as

$$PCH(m) = \sum_{s=1}^{S}H(s,m), \quad (8)$$

where

$$H(s,m) = 1 \text{ if } F(sT) = m, \text{ and } H(s,m) = 0 \text{ otherwise.}$$

Nonlinear regression and confidence interval computations

Several statistical approaches are available to assess the degree of uncertainty of the fitted parameter estimates from the nonlinear regression. We present here a range of available statistical approaches, and assess their

efficacy for interpreting confidence intervals derived from PCH data. In each of these approaches, the first step is to fit a model function $f(x,\theta)$ with p parameters to n experimental data points $y(x_i)$, $i=\{1, 2, \ldots n\}$. The parameter values are contained in the p-dimensional vector θ. The fitting is accomplished by finding the set of optimal parameter values, $\widehat{\boldsymbol{\theta}}$, which minimizes the χ^2 error. All computations were done with MATLAB. We use the MATLAB nonlinear regression function with the Levenberg-Marquardt algorithm, and our PCH model function uses the generating function and the discrete Fourier transform to guarantee accuracy and decrease computation time. The termination tolerance on the change in the residual sum of squares was set to 10^{-8}. All regressions were performed with weighting by using the standard deviation of the 1000 simulated data sets at each point. In practice, such a large number of replicate data sets is rarely available, but other methods of calculating the weights, such as using the binomial distribution (5) or the diagonal covariance matrix elements (11), are easily implemented and are essentially equivalent to our method.

We assess the confidence interval methods by fitting 1000 separate simulated data sets, calculating the confidence intervals for all sets by each method at different probability levels, and seeing what percentage of the confidence intervals contain the true values of the parameters, which are known from the simulations. For a 95% confidence interval 950 of the 1000 confidence intervals should contain the true parameters.

Asymptotic standard error

The asymptotic approach is the most common method of evaluating confidence intervals in linear and nonlinear regressions because of its relative simplicity and computational speed (14,16). The fitting function $f(x,\boldsymbol{\theta})$ is assumed to be linear such that it can converge with only the terms before the second-order derivatives if expanded as a Taylor series. The sensitivity of the estimate to experimental error is then given fully by the degree to which $f(x,\boldsymbol{\theta})$ changes with infinitesimal perturbations to each of the parameters $\boldsymbol{\theta}$ at each experimental measurement condition x_i. The system of first-order derivatives from Taylor expansions representing these changes, is composed of the weighted partial derivatives of $f(x,\boldsymbol{\theta})$ with respect to each parameter, evaluated at each experimental measurement condition,

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sigma_1}\cdot\frac{\partial f(x_1,\,\boldsymbol{\theta})}{\partial\theta^{(1)}} & \cdots & \frac{1}{\sigma_1}\cdot\frac{\partial f(x_1,\,\boldsymbol{\theta})}{\partial\theta^{(p)}} \\ \vdots & \ddots & \vdots \\ \frac{1}{\sigma_n}\cdot\frac{\partial f(x_n,\,\boldsymbol{\theta})}{\partial\theta^{(1)}} & \cdots & \frac{1}{\sigma_n}\cdot\frac{\partial f(x_n,\,\boldsymbol{\theta})}{\partial\theta^{(p)}} \end{bmatrix}, \quad (9)$$

where σ_i is the statistical weighting function of the data point (x_i, y_i). The confidence error is derived from the diagonal of the following matrix that is based on the matrix $\mathbf{A}$:

$$\mathbf{C} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}. \quad (10)$$

The confidence interval of the r^{th} parameter, $\boldsymbol{\theta}^{(r)}$, or the range above or below the estimated parameter $\widehat{\theta}^{(r)}$, in which the true parameter $\boldsymbol{\theta}^{(r)}$ lies with a probability of $1-\alpha$, is then

$$\mathrm{CI}_{1-\alpha}\left(\theta^{(r)}\right) = \widehat{\theta}^{(r)} \pm t_{n-p}^{\alpha/2}\cdot\sqrt{\widehat{c}^{(rr)}}\cdot\sqrt{\frac{\sum_{i=1}^{n}\left[\frac{f\left(\widehat{\boldsymbol{\theta}},x_i\right)-y_i}{\sigma_i}\right]^2}{n-p}}, \quad (11)$$

where $\widehat{c}^{(rr)}$ is defined as the r^{th} diagonal element of $\mathbf{C}$ evaluated at the estimated parameters $\widehat{\boldsymbol{\theta}}$, the terms in the radical represent the square-root of the least-squares norm at the minimum, n is the size of the dataset, p is the number of parameters in $\boldsymbol{\theta}$, and $t_{n-p}^{\alpha/2}$ is the upper $\alpha/2$ percentage point of the t-distribution with the number of degrees of freedom specified at $(n-p)$. Because the $\mathbf{A}$ matrix is evaluated at each iteration during the optimal parameter search, there is very little additional computation in the confidence interval calculation by this method. This asymptotic approach is strictly valid only for linear regression and is unable to produce precise confidence intervals if $f(\bullet)$ is nonlinear, because the first-order derivatives of Taylor expansions of $f(\bullet)$ plus the $f(\bullet)$ value at the estimated parameters may diverge from the real value of $f(\bullet)$ (12).

Likelihood method

The likelihood method is an application of the classical hypothesis testing of a null hypothesis and an alternative hypothesis (i.e., can one reject, with a chosen level of certainty, that the value of a particular parameter, or parameter set, lies within a subset of parameter space (20)). The metric used for the test is the sum of squares of weighted errors as a function of $\boldsymbol{\theta}$,

$$S(\theta) = \sum_{i=1}^{n}\left[\frac{f(\boldsymbol{\theta},x_i)-y_i}{\sigma_i}\right]^2,$$

which is compared to the sum of squares of weighted errors at the optimum parameter set, $s(\widehat{\boldsymbol{\theta}})$.

The likelihood joint confidence region using the F-test was first proposed by Beale (21), and gives the confidence region of all parameters together. The confidence region is defined as all values of $\boldsymbol{\theta}$ such that

$$s(\theta) \le \left(1+\frac{p}{n-p}\cdot F_{p,n-p}^{\alpha}\right)\bullet s\left(\widehat{\boldsymbol{\theta}}\right), \quad (12)$$

where $F_{p,n-p}^{\alpha}$ is the upper-α percentage point of the F-distribution with its two degrees of freedom specified as p and $(n-p)$. Note that p and $(n-p)$ are also degrees of freedom for $s(\boldsymbol{\theta})$ and $s(\widehat{\boldsymbol{\theta}})$, respectively. For linear models, the joint confidence region will be a hyperellipsoid of dimension p, but for nonlinear models the region can be far from ellipsoidal, and even disjoint or unbounded (16). Evaluation of the joint confidence region is computationally intensive, as it requires a search in p-dimensional parameter space to find the contours of the region defined by Eq. 12.

There is also difficulty in visualizing the complex, high-dimensional shape of the confidence regions for most nonlinear problems. For these reasons, the full confidence region is less than satisfactory for practical purposes. Confidence regions for subsets of the parameters can be defined in an analogous way, which produce two-dimensional confidence regions for pairs of parameters, for instance, that are slices of the higher-dimensional region. There may be many pairwise combinations of parameters for a complex model, however, which again makes visualization and evaluation difficult. Confidence intervals for each parameter separately can also be defined by the likelihood method. The confidence interval for the r^{th} parameter, θ^r, is all values of θ^r such that

$$s\left(\theta^{\mathrm{p}-1},\theta'\right) \le \left(1+\frac{1}{n-p}\cdot F_{1,n-p}^{\alpha}\right)\bullet s\left(\widehat{\boldsymbol{\theta}}\right), \quad (13)$$

where $\boldsymbol{\theta}^{\boldsymbol{p-1}}$ is the parameter vector of all parameters except θ^r, and $s(\boldsymbol{\theta}^{\boldsymbol{p-1}}, \theta^r)$ is the sum of squares of weighted errors for θ^r, minimized over the other parameters. Practically, the upper and lower limits of the confidence interval are found by one-dimensional searches along the chosen parameter axis. Starting at the optimal value, θ^r is incremented, a new regression is performed holding θ^r constant and varying the others to find a new least-squares minimum, and the sum of squares is evaluated by Eq. 13. This procedure is continued until the upper limit for the value of θ^r satisfying Eq. 13 can be found to any desired precision. A similar

search is then performed along the parameter axis in the opposite direction to find the lower limit. An optimized search strategy can find each limit with relatively few steps.

Here, we generate two-dimensional contours of this volume as follows. The contributed PCH dataset for Fig. 1, *A* and *B*, was obtained from Monte Carlo simulation with supplied parameter values of $n_1 = 1.2$, $n_2 = 0.3$, $q_1 = 1.0$, and $q_2 = 4.0$, whereas the estimated values from nonlinear fit are 1.208, 0.236, 1.141, and 4.332, respectively. All plots are produced by starting with two fixed parameters, varying two other parameters around their estimated values, and getting each corresponding $s(\boldsymbol{\theta})$. Then it is possible to generate contour plots with $s(\boldsymbol{\theta})$ less than certain level values from Eq. 13. It is shown in Fig. 1 that both the two concentrations and the two brightnesses for the PCH system are highly correlated. This is indicated by the high degrees of flatness and inclination of the ellipses. In other words, the increase of one parameter is compensated by a decrease in the other parameter with the overall compensatory interaction indicated by the slope of the major axis of the ellipse. Clearly, as the degree of flatness of the ellipse increases and the absolute value of the slope of the major axis becomes closer to 1, the degree of correlation of the two parameters increases.

Bootstrap method

Bootstrapping was first proposed by Efron (22,23). The fundamental concept in bootstrapping is the construction of a sample distribution for a particular statistic by resampling from the datasets that are at hand. To estimate bootstrap confidence intervals of a parameter-vector $\boldsymbol{\theta}$ based on an original dataset $\boldsymbol{P}$, B replicate data sets $\{\widehat{\boldsymbol{P}}_1, \widehat{\boldsymbol{P}}_2, \ldots, \widehat{\boldsymbol{P}}_{\boldsymbol{B}}\}$ are generated by random sampling of data with replacement from the original dataset. Each resampled data set, $\widehat{\boldsymbol{P}}_{\boldsymbol{b}}$, thus has the same number of data points as the original. The number of replicates, B, is typically on the order of 1000. Then optimal parameter sets $\{\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \ldots, \widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}\}$ are computed by least-squares fitting for all bootstrap samples. Therefore, confidence intervals of $\boldsymbol{\theta}$ can be determined from the generated replicates distributions. This approach is schematically illustrated in Fig. 2.

Several techniques have been developed to build confidence intervals from replicate distributions (24). The normal-theory interval assumes that $\{\widehat{\boldsymbol{\theta}}_1, \widehat{\boldsymbol{\theta}}_2, \ldots, \widehat{\boldsymbol{\theta}}_{\boldsymbol{B}}\}$ is normally distributed, and uses normal distribution properties to determine confidence intervals of $\boldsymbol{\theta}$. The percentile-theory interval is based on the empirical quantiles such that the $(1 - \alpha)$ percentile confidence intervals of $\boldsymbol{\theta}$ are created by the $\alpha/2$ and the $(1 - \alpha/2)$ quantiles of replicates distributions. Because the normal-theory works well only for normal distributions, and the percentile-theory requires that the distributions be symmetric about the observed statistic, we recommend the more general bias-corrected and accelerated approach (BCa) to compute PCH confidence intervals.

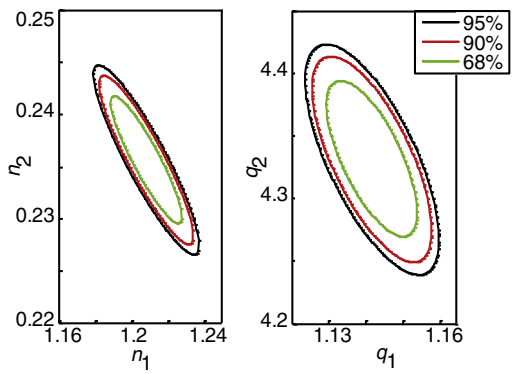

FIGURE 1 Two-dimensional slices of the nearly four-dimensional elliptical confidence volume of a noise-assigned monomer/tetramer system with $n_1 = 1.2$, $q_1 = 1$, $n_2 = 0.3$, and $q_2 = 4$. (*A*) Correlation between n_1 and n_2 at $q_1 = 1.1$ and $q_2 = 4.3$. (*B*) Correlation between q_1 and q_2 at $n_1 = 1.2$ and $n_2 = 0.2$.

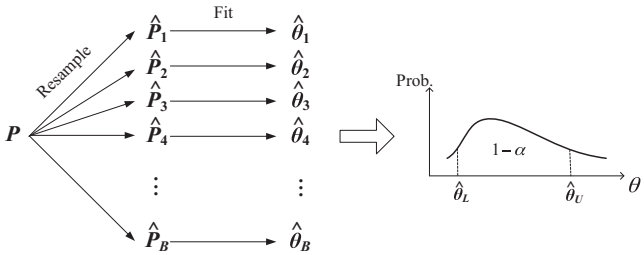

FIGURE 2 Schematic of procedures for the bootstrap method: P, original dataset; $\widehat{\boldsymbol{P}}_{\boldsymbol{b}}$, resampled data sets; and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{b}}$, optimal parameter value sets from regression of resampled data sets.

BCa adjusts for both bias and skewness in the replicates distributions. It is verified by Efron (24) to be of second-order accuracy. The bias $z_0^{(r)}$ and the skewness $a^{(r)}$ of the replicates distribution of the r^{th} parameter, $\theta^{(r)}$, are first computed from

$$z_o^{(r)} = \Phi^{-1}\left\{\frac{\sum_{b=1}^{B} I\left\{\widehat{\theta}_b^{(r)} < \widehat{\theta}^{(r)}\right\}}{B}\right\}, \tag{14}$$

$$a^{(r)} = \frac{\sum_{i=1}^{n}\left(\overline{\widehat{\theta}}_{(-i)}^{(r)} - \widehat{\theta}_{(-i)}^{(r)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\overline{\widehat{\theta}}_{(-i)}^{(r)} - \widehat{\theta}_{(-i)}^{(r)}\right)^2\right]^{\frac{3}{2}}}. \tag{15}$$

Then the $(1 - \alpha)$ percentile confidence interval of $\theta^{(r)}$ is generated from

$$\mathrm{CI}_{1-\alpha}\left(\theta^{(r)}\right) = \left[\widehat{G}^{-1}\left\{\Phi\left[z_0^{(r)} + \frac{z_0^{(r)} + z_{\alpha/2}}{1 - a^{(r)}\left(z_0^{(r)} + z_{\alpha/2}\right)}\right]\right\}, \widehat{G}^{-1}\left\{\Phi\left[z_0^{(r)} + \frac{z_0^{(r)} + z_{1-\alpha/2}}{1 - a^{(r)}\left(z_0^{(r)} + z_{1-\alpha/2}\right)}\right]\right\}\right] \tag{16}$$

where $\Phi^{-1}(\cdot)$ is the standard-normal quantile function, with $z_\alpha = \Phi^{-1}(\alpha)$,

$$\left(\frac{\sum_{b=1}^{B} I\left\{\widehat{\theta}_b^{(r)} < \widehat{\theta}^{(r)}\right\}}{B}\right)$$

is the proportion of bootstrap replicates below the original sample estimated parameter $\widehat{\theta}^{(r)}$, n is the size of original sample, $\widehat{\theta}_{(-i)}^{(r)}$ is the value of $\theta^{(r)}$ produced when the i^{th} observation is deleted from the original sample, $\overline{\widehat{\theta}}_{(-i)}^{(r)}$ is the average of $\widehat{\theta}_{(-i)}^{(r)}$, and $\widehat{G}^{-1}(\cdot)$ is the cumulative-replicates quantile function. Note that if the correction factors $a^{(r)}$ and $z_0^{(r)}$ are both 0, the replicates distribution is symmetric about the estimated parameter value. In this case, the BCa method is the same as the percentile method.

Monte Carlo with residuals resampling

Another nonparametric approach similar to the bootstrap method, Monte Carlo with residuals resampling (MCRR), resamples the residuals, rather than the data points, and uses them to construct synthetic data sets (25). It is applied by: Step 1. Least-squares fitting of the data to find the optimal parameter values; Step 2. Calculating a theoretical, or perfect, dataset from the model using the optimal parameter values; Step 3. Generation of many synthetic data sets by adding noise (obtained by resampling the residuals from Step 1 analysis) to the theoretical dataset; Step 4. Least-squares fitting of the noise-containing, synthetic datasets and tabulating the distribution of optimal parameter values from these fits; and finally Step 5. Generation of confidence intervals by the percentile-theory (25). This approach is schematically illustrated in Fig. 3. Note that the bias and the acceleration of the replicates distributions need further exploration, which is beyond the scope of this article.

RESULTS AND DISCUSSION

Comparison of coverages

To test the accuracy of each confidence interval computation method, we generated confidence intervals/regions for 1000 simulated experiments on a planar membrane. We analyzed PCH datasets by using each method, and calculated the coverage, i.e., the fraction of the confidence intervals that include the true parameter values, known from the simulation. Specifically, we assigned representative values of 1.2, 0.3, 1.0, and 4.0 to n_1, n_2, q_1, and q_2, respectively, and used Monte Carlo simulation to produce 1000 two-component PCH datasets and fit these 1000 data sets using nonlinear regression. Note that all datasets are fitted with statistical weights corresponding to different photon numbers for each measurement. Each data set and corresponding analysis yielded slightly different optimal parameter values and confidence intervals/regions, attributed to the limited number of bins (10^5) in the simulation and the random noise inherent in the experiment. Then we determined the fraction of the parameter sets having the true value within the derived confidence intervals. The comparison is summarized in Table 1.

It is obvious from Table 1 that no method is able to provide absolutely accurate confidence intervals/regions for PCH parameters. Bootstrap (BCa) is overall the most accurate and is recommended as an effective approach to predict PCH confidence intervals. The asymptotic method is a fast operational approach that does not produce bad coverage; thus, it can be recommended as a prompt and rough estimate of PCH confidence intervals. However, joint confidence region is not well applied to a PCH model. It is not clear why Monte Carlo with residuals resampling, which has similarities to BCa, does as poorly as shown in Table 1. It is possibly due to the lack of bias correction of the replicates distributions that is applied in the BCa method, and warrants further exploration along this path. Also, it is interesting that the coverage of the likelihood method is not better than the asymptotic standard method. This is probably due to the fact that both the likelihood and asymptotic methods are approximate methods (15) An earlier study has shown that the two methods produce confidence intervals of essentially the same length for a different model system (26).

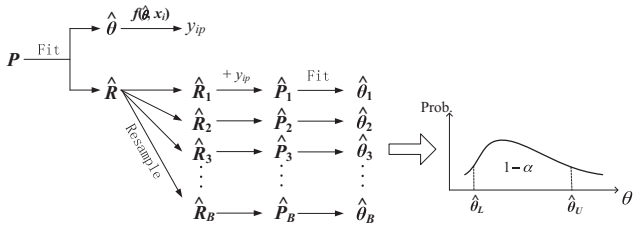

FIGURE 3 Schematic of procedures for the Monte Carlo with residuals resampling (MCRR) method: $\boldsymbol{P}$, the original dataset; $\widehat{\boldsymbol{R}}_{\boldsymbol{b}}$, a residual resample; $\widehat{\boldsymbol{P}}_{\boldsymbol{b}}$, a database resample; and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{b}}$, the regression result of a resample.

For fluorophores diffusing through a three-dimensional fluid, 500 two-component PCH datasets were studied using the two methods that performed best for analysis of PCHs acquired from flat membranes. The BCa and asymptotic standard error methods both produced reasonably good coverage with the BCa method slightly better than the asymptotic (Table 2).

Although we have rigorously explored only a subset of the four-dimensional parameter space for PCHs acquired from two-dimensional membranes, there is evidence that the conclusions apply to more of the full space. Fig. 4 shows a comparison of the 95% confidence level coverage by the asymptotic standard error and the BCa methods for a range of brightness values. As the brightness of the two species increases, the confidence interval becomes smaller, but the BCa method is substantially better than the asymptotic standard method for all brightness values examined.

In principle, PCH analysis promises to supply the brightness and concentrations of many fluorescent components. In practice, however, only a few components can ordinarily be identified, and there can be large uncertainties in the optimal values of concentration and brightness determined from least-squares regression analysis. Fig. 5 suggests why this is so for an analysis of a typical two-component system. The value of χ^2 is plotted against n_1 and n_2 for a system with $n_1 = 1.2$, $q_1 = 1$, $n_2 = 0.3$, and $q_2 = 4$. The plot shows that the χ^2 surface has the shape of a curving elongated valley with low curvature along the valley floor due to the compensation between n_1 and n_2 in the PCH model.

Longer data collection that provides larger data sets will, of course, reduce the relative noise level and decrease the size of the confidence intervals for all methods, thus more accurately determining the correct values of the concentrations and brightnesses. We can see this effect clearly by increasing the length of the simulations (data not shown.) The decrease in signal/noise and reduction of the confidence interval length depends only on the square root of the number of data points, which puts practical limits on how much one can reduce the uncertainty in parameters

TABLE 1 Comparison of confidence interval coverage for two-dimensional PCH analysis

Method	Theoretical confidence level	Simulated confidence interval coverage				Approximate computation time (seconds)
		n_1	n_2	q_1	q_2	
Asymptotic standard error	95%	88.9%	86.4%	85.9%	87.2%	0.3
	90%	82.8%	79.8%	79.4%	80.8%	
	68%	57.6%	58.2%	58.6%	60.8%	
Likelihood joint confidence region	95%		84.6%			
	90%		77.7%			
	68%		57.8%			
Likelihood confidence interval	95%	85.6%	88.2%	87.5%	88.2%	15
	90%	78.3%	82.0%	80.4%	82.7%	
	68%	54.9%	59.3%	59.2%	61.7%	
Bootstrap (BCa)	95%	98.2%	96.5%	96.9%	94.8%	220
	90%	92.5%	90.3%	89.2%	88.4%	
	68%	57.5%	62.8%	60.1%	64.6%	
Monte Carlo with residuals resampling (MCRR)	95%	81.6%	84.0%	83.1%	83.3%	130
	90%	72.7%	76.6%	74.1%	75.8%	
	68%	47.3%	54.5%	53.0%	55.6%	

this way. Even with larger data sets, however, the BCa method outperforms the asymptotic method.

An additional contribution to the uncertainty of parameter estimation is the presence of systematic noise that is not taken into account in this study. There can be many sources of such noise, e.g., photobleaching, vibrations, laser fluctuations, etc., which contribute differently in different experiments. We have focused on the stochastic elements inherent in the method that contribute in all experimental situations.

One factor that contributes to difficulty of determining adequately the brightnesses and concentrations in a multicomponent system is the Gaussian shape of the excitation laser beam. Ambiguity results from that fact that a dimmer fluorophore in the center of the beam will yield an emission intensity comparable to a brighter fluorophore in the periphery. Benefits from hard boundaries of the laser excitation area have been proposed for some time (e.g., (17)). This idea is corroborated by examining the results of simulations using a disk-shaped excitation beam profile in which the excitation intensity is constant over the disk and is zero outside the disk. With this excitation profile, the fluorescence signal from a fluorophore anywhere in the disk is the same (apart from shot noise) and is proportional to the brightness of the fluorophore.

As expected and as illustrated in Fig. 6, the confidence intervals are tighter for both the concentrations and the brightnesses for a system measured using a disk-shaped beam profile than for the same system measured with a Gaussian excitation profile. Evidently, using a disk-shaped excitation profile would be advantageous for PCH measurements, but implementing this approach is difficult. FCS measurements are commonly performed with a laser excitation profile of minimal size, i.e., a diffraction-limited profile. This minimizes the background fluorescence and also the diffusion correlation time. It does not appear to be possible to have both a diffraction-limited and also a disk-shaped beam profile. To approximate a disk it would be necessary to enlarge the beam so that the diffraction fringes at the beam periphery (27) would contribute relatively little to the overall excitation intensity. But this would, then, increase background and correlation time. It would be interesting in future work to explore the possibility of using different beam shapes to determine the concentrations and brightnesses by PCH measurements more precisely.

CONCLUSIONS

An important component of the analysis of experimental data by least-squares curve fitting is the calculation of how trustworthy, or alternatively, how uncertain, the results of the regression analysis really are. The most common approach for determining confidence intervals of model parameters is the asymptotic standard error, due to its speed and ease of computation. Although valid for linear models, the asymptotic approach applied to nonlinear models may severely underestimate the size of the confidence intervals,

TABLE 2 Comparison of confidence interval coverage for three-dimensional PCH analysis

Method	Theoretical confidence level	Simulated confidence interval coverage				Approximate computation time (seconds)
		n_1	n_2	q_1	q_2	
Asymptotic standard error	95%	93.8%	93.6%	95.6%	92.6%	0.3
	68%	76.6%	66.4%	65.8%	67.6%	
Bootstrap (BCa)	95%	95.9%	98.3%	95.1%	94.5%	220–440
	68%	63.7%	72.6%	68.6%	65.9%	

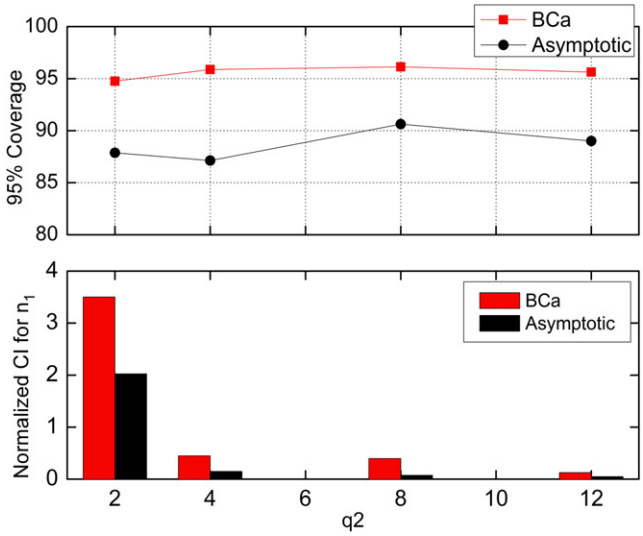

FIGURE 4 95% confidence interval coverage and the normalized confidence interval width of n_1 versus q_2 given that $n_1 = 1.2$, $n_2 = 0.3$, and $q_2/q_1 = 4$. The BCa method outperforms asymptotic standard method over a range of brightness values.

leading to erroneous conclusions. The theory describing PCH data, either directly in terms of the super-Poissonian distribution or in terms of the generating function approach, is highly nonlinear in the variables of interest, i.e., the number and brightness of the species in the system. A surprising result is that the asymptotic approach leads to very good and quite consistent coverage over a broad range of conditions. However, as shown in Fig. 4, the asymptotic approach underestimates the confidence intervals for all conditions for fluorophores diffusing on a membrane, and for many conditions does so significantly. We have shown in this article that the bias-corrected and accelerated bootstrap method is a reliable and accurate way to calculate

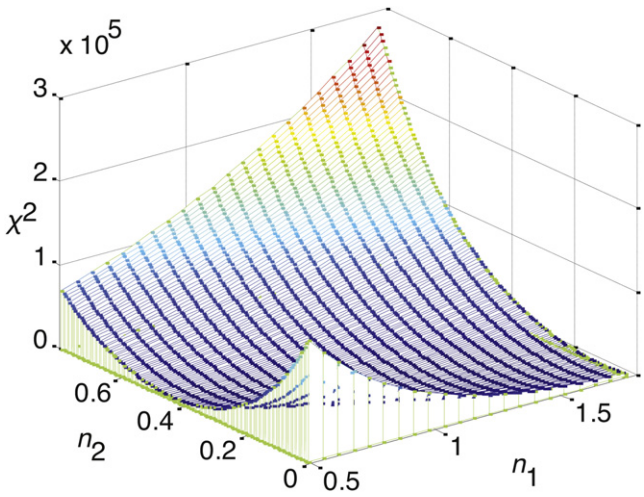

FIGURE 5 χ^2 surface of a simulated monomer/tetramer system with $n_1 = 1.2$, $q_1 = 1$, $n_2 = 0.3$, and $q_2 = 4$.

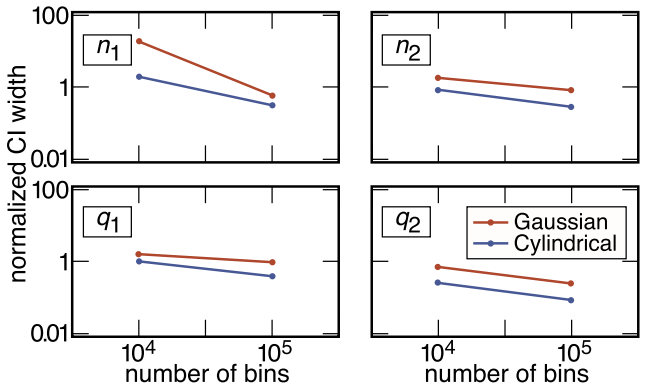

FIGURE 6 Comparisons of normalized confidence-interval widths (using BCa) of monomer/tetramer systems diffusing on a Gaussian beam (*Gaussian*) and a disk-shaped beam (*Cylindrical*). Each width above is the average of widths from 100 simulated monomer/tetramer systems with $n_1 = 1.2$, $q_1 = 1$, $n_2 = 0.3$, and $q_2 = 4$.

the confidence intervals for the four parameters in a two-component PCH in these cases. For fluorophores diffusing in a three-dimensional fluid, the BCa and asymptotic standard error approaches appear equally good.

The selection of a method for identifying confidence intervals is a trade-off between experimental and computational time. Improvement over confidence intervals calculated from a dataset by asymptotic standard error can be made in two ways. The first is to use the BCa method, and the second is to acquire a longer data record and use asymptotic standard error, recognizing that the confidence intervals would be better still if one were to use the BCa method on this longer data record. Deciding between these requires a decision about the relative importance of experimental accuracy versus computational speed. We note as well that one can never be certain of the accuracy of confidence intervals calculated using asymptotic standard error without performing lengthy computations like those presented in this article. Although the computation time of the BCa method is roughly three orders-of-magnitude greater than the asymptotic standard method, it nevertheless requires only several minutes for such an analysis, and the use of optimized software packages such as MATLAB can make its application routine. Confidence intervals calculated using BCa are significantly better in many cases and require little additional computation time.

The authors thank Yu Li for many valuable discussions.

This work was supported by National Institutes of Health grants R01 GM084200 and K25 HL079165, and National Science Foundation grant CMMI 0826518.

REFERENCES

1. Elson, E., and D. Magde. 1974. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. *Biopolymers.* 13:1–27.

2. Magde, D., E. L. Elson, and W. W. Webb. 1974. Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers.* 13:29–61.

3. Qian, H., and E. L. Elson. 1990. On the analysis of high order moments of fluorescence fluctuations. *Biophys. J.* 57:375–380.

4. Thompson, N. L., and J. L. Mitchell. 2001. High order autocorrelation in fluorescence correlation spectroscopy. *In* Fluorescence Correlation Spectroscopy, Theory and Applications. R. Rigler and E. L. Elson, editors. Springer-Verlag, Berlin, Germany. 438–458.

5. Chen, Y., J. D. Müller, …, E. Gratton. 1999. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys. J.* 77:553–567.

6. Kask, P., K. Palo, …, K. Gall. 1999. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proc. Natl. Acad. Sci. USA.* 96:13756–13761.

7. Saffarian, S., Y. Li, …, L. J. Pike. 2007. Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. *Biophys. J.* 93:1021–1031.

8. Pike, L. J. 2006. Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function. *J. Lipid Res.* 47:1597–1598.

9. Elson, E. L., E.Fried, …, G. M.Genin. Phase separation in biological membranes: integration of theory and experiment. *Annu. Rev. Biophys.* 39:207–226.

10. Waugh, M. G., D. Lawson, and J. J. Hsuan. 1999. Epidermal growth factor receptor activation is localized within low-buoyant density, non-caveolar membrane domains. *Biochem. J.* 337:591–597.

11. Kask, P., and K. Palo. 2001. Introduction to the theory of fluorescence intensity distribution analysis. *In* Fluorescence Correlation Spectroscopy, Theory and Applications. R. Rigler and E. L. Elson, editors. Springer-Verlag, Berlin, Germany. 396–409.

12. Johnson, M. L., and L. M. Faunt. 1992. Parameter estimation by least-squares methods. *Methods Enzymol.* 210:1–37.

13. Johnson, M. L. 2008. Nonlinear least-squares fitting methods. *Methods Cell Biol.* 84:781–805.

14. Seber, G. A. F., and C. J. Wild. 2003. Nonlinear Regression. John Wiley and Sons, Hoboken, NJ.

15. Donaldson, J. R., and R. B. Schnabel. 1987. Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics.* 29:67–82.

16. Bates, D. M., and D. G. Watts. 1988. Nonlinear Regression Analysis and its Applications. John Wiley & Sons, Hoboken, NJ.

17. Mertz, J., C. Xu, and W. W. Webb. 1995. Single-molecule detection by two-photon-excited fluorescence. *Opt. Lett.* 20:2532–2533.

18. Rüttinger, S., V. Buschmann, …, F. Koberling. 2008. Comparison and accuracy of methods to determine the confocal volume for quantitative fluorescence correlation spectroscopy. *J. Microsc.* 232:343–352.

19. Meng, F., and H. Ma. 2006. A comparison between photon counting histogram and fluorescence intensity distribution analysis. *J. Phys. Chem. B.* 110:25716–25720.

20. DeGroot, M. H., and M. J. Schervish. 2002. Probability and Statistics. Addison-Wesley, Boston, MA.

21. Beale, E. M. L. 1960. Confidence regions in non-linear estimation. *J. R. Stat. Soc. B.* 22:41–76.

22. Efron, B. 1979. Bootstrap methods: another look at the jack-knife. *Ann. Stat.* 7:1–26.

23. Efron, B. 1985. Bootstrap confidence intervals for a class of parametric problems. *Biometrika.* 72:45–58.

24. Efron, B. 1996. Bootstrap confidence intervals. *Stat. Sci.* 11:189–212.

25. Straume, M., and M. L. Johnson. 1992. Monte Carlo method for determining complete confidence probability distributions of estimated model parameters. *Methods Enzymol.* 210:117–129.

26. Gallant, A. R. 1975. Testing a subset of the parameters of a nonlinear regression model. *J. Am. Stat. Assoc.* 70:927–932.

27. Hess, S. T., and W. W. Webb. 2002. Focal volume optics and experimental artifacts in confocal fluorescence correlation spectroscopy. *Biophys. J.* 83:2300–2317.