Bayesian Inference Frameworks for Fluorescence Microscopy Data Analysis

by

Ross Wallgren

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts in Mathematics

Approved April 2019 by the
Graduate Supervisory Committee:

Hans Dieter Armbruster, Co-Chair
Steve Pressé, Co-Chair
Robert McCulloch

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

In this work, I present a Bayesian inference computational framework for the analysis
of widefield microscopy data that addresses three challenges: (1) counting and localiz-
ing stationary fluorescent molecules; (2) inferring a spatially-dependent effective fluo-
rescence profile that describes the spatially-varying rate at which fluorescent molecules
emit subsequently-detected photons (due to different illumination intensities or dif-
ferent local environments); and (3) inferring the camera gain. My general theoretical
framework utilizes the Bayesian nonparametric Gaussian and beta-Bernoulli processes
with a Markov chain Monte Carlo sampling scheme, which I further specify and im-
plement for Total Internal Reflection Fluorescence (TIRF) microscopy data, bench-
marking the method on synthetic data. These three frameworks are self-contained,
and can be used concurrently so that the fluorescence profile and emitter locations
are both considered unknown and, under some conditions, learned simultaneously.
The framework I present is flexible and may be adapted to accommodate the infer-
ence of other parameters, such as emission photophysical kinetics and the trajectories
of moving molecules. My TIRF-specific implementation may find use in the study
of structures on cell membranes, or in studying local sample properties that affect
fluorescent molecule photon emission rates.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

Recent advances in optics have allowed the scientific community unprecedented, beautiful access into biology down to the level of single molecules (Liu *et al.* (2015); Su and Ju (2018); Sako *et al.* (2000); Grimm *et al.* (2016); Kong *et al.* (2016); Aguet *et al.* (2016); Chen *et al.* (2014); Specht *et al.* (2017)). Often, the imaging techniques involve molecules fluorescently labeled to contrast with the background. These fluorescent molecules are then excited with light at a particular wavelength, and they subsequently emit photons at a differing wavelength, and the excitation light can be filtered out before reaching the detector, creating images that can resolve single molecules (Shashkova and Leake (2017); Lichtman and Conchello (2005)).

Because the point-source (the labeled molecule) emits light with a wavelength many times larger than itself (typically two orders of magnitude larger), the image produced by the emitting sources is diffraction-limited. That is, each point-source (∼1-5nm) appears as a large diffraction-limited spot (∼150-300nm) called the emission point-spread function (PSF). Two point-sources of light that are sufficiently close in space (closer than ∼200-300nm, depending on parameters such as emission wavelength and refractive index) will appear as one coinciding object (Born and Wolf (2013)). Strategies attempting to circumvent the diffraction limit to better count and localize individual molecules are known as super-resolution techniques, and many of them are applicable to biological systems (Schermelleh *et al.* (2010); Huang *et al.* (2009); Fernández-Suárez and Ting (2008); Bates *et al.* (2007); Shim *et al.* (2012); Zhi *et al.* (2018)). Many quantitative methods for counting and localization in widefield microscopy exist already, employing techniques such as calculating centroids, fitting

Gaussians, least-squares, deconvolution, and (frequentist) maximum likelihood; some are simple and quick to implement, while others are more sophisticated and precise, and many take into account other factors such as camera shot noise and background pixel noise (Chen *et al.* (2014); Lee *et al.* (2017); Deschout *et al.* (2014); Lee *et al.* (2012); Wöll and Flors (2017); Ovesnỳ *et al.* (2014)).

Despite the progress of quantitative counting and localization there are two critical issues that are often not considered, the first being how to deal with closely-neighboring (in xy) or stacked (in z) molecules, and the second being how to account for spatially-varying fluorescence (effective emission) properties within the region of interest (ROI). Fundamentally these are model selection problems (Fig. C.1): how many overlapping emission PSFs underlay the image we see? How do we select the fluorescence profile?

In this work, we present a framework to count and localize single stationary fluorescent molecules (fluorophore) at super-resolution in fluorescence microscopy and a separate, self-contained framework to infer a spatially-dependent function that describes the rate at which an extant fluorophore would emit detectable photons when located at a particular location in the region of interest, which we call the fluorescence profile. With (EM)CCD cameras in mind (Mortensen and Flyvbjerg (2016)), we also incorporate a relatively simple framework for the inference of the camera gain parameter, which is itself worth describing. At the basis of these three methods is a simple, general model of fluorescence microscopy. We specify this model further and implement our inference frameworks on synthetic data for Total Internal Reflection Fluorescence (TIRF) microscopy (Axelrod *et al.* (1983); Schneckenburger (2005); Fish (2009); Mattheyses *et al.* (2010)). The concurrent inference of all three or any two of the aforementioned parameters is possible, though the functionality of inferring both emitter count/localization and fluorescence profile together is limited by fun-

damental model nonidentifiability (App. C); that is, there exists some fundamental and intractable ambiguity between the number and/or location of the molecules, the fluorescence profile, and to some extent also the gain, so that the observations we make can have come about by qualitatively differing scenarios.

Our method largly relies on Bayesian nonparametrics (Hjort *et al.* (2010); Ferguson (1973); Tavakoli *et al.* (2019)), and in particular two separate nonparametric tools: the Gaussian process (Rasmussen (2004); Williams and Rasmussen (2006); Wilson and Adams (2013)) and beta-Bernoulli process (Jazani *et al.* (2019); MacEachern (2016); Broderick *et al.* (2012)), used respectively for the inference of the fluorescence profile and fluorophore counts/localizations (App. D). These two methods are said to be nonparametric in the sense that they allow for inference over an infinite (or in practice, an arbitrarily large) number of random variables. In the case of the Gaussian process, this is an undetermined and possibly infinite number of points in a continuous space over which we infer values for the profile, and in the case of the beta-Bernoulli process, an infinite number of possible fluorophore counts/localizations.

Bayesian nonparametrics are relatively new and sophisticated mathematical tools, which have been applied to other problems in fluorescence microscopy (Lee *et al.* (2017); Tavakoli *et al.* (2019); Sgouralis *et al.* (2018a); Sgouralis and Pressé (2017); Tavakoli *et al.* (2016); Kilic and Pressé (2019)), including tracking moving molecules (Sgouralis *et al.* (2018b)) and estimating diffusion coefficients (Jazani *et al.* (2019)) in confocal microscopy. The specific implementation we present is aimed toward static regimes in widefield TIRF microscopy. Our general framework, by virtue of the Bayesian inference paradigm, allows for easy adaption to a wide variety of imaging methods and problems.

Chapter 2

METHODS

2.1 Description of General Model

Here, we describe our general model for fluorescence microscopy, which we further specify and adapt to TIRF in the following section.

We have M stationary point-source emitters (fluorophore) in the ROI, with locations $X_{1:M} \triangleq [x_1, x_2, \ldots x_M]$. The fluorescence profile, which we denote $F(\cdot)$ and assume to be continuous, describes the rate at which an emitter at a particular location in space will emit "detectable" photons (photons that are subsequently detected, though they may be emitted outside the ROI; this accounts for emission and effects that prevent photons from being registered). The integral of the PSF over the area covered by a pixel describes the probability that an emitted photon will be registered to that particular pixel. The intensity of detection-events (photon arrivals) in pixel p for some fixed period of time, which we denote E_p, is the sum of the integrals of all M PSFs over pixel p, scaled by the value of the fluorescence profile at the locations of the emitters. We also include B_p, a background photon-arrival rate term for pixel p, and consider the camera-exposure time, Δt_e, to be fixed:

$$E_p \triangleq \Delta t_e \big(B_p + \sum_{m=1}^{M} F(x_m) \text{INTPSF}(x_m, p) \big). \tag{2.1}$$

Note that here, E_p is time invariant.

We choose to model the PSF as an isotropic Gaussian distribution, centered at the location of the emitter, with FWHM parameter σ_{PSF}. This is an approximation of the more-realistic Airy function, and allows for fast computation of the integral

(Zhang *et al.* (2007); Deschout *et al.* (2012)):

$$\text{INTPSF}(x_m, p) := \varphi^{-1/2} \iint_p \exp[-\pi ||X - x_m||^2/\varphi]dX \tag{2.2}$$

$$\varphi = 2\pi\sigma_{\text{PSF}}^2. \tag{2.3}$$

Photon-detection intensity is not measured directly by (EM)CCD cameras, but given in arbitrary digital units (ADUs) in a way that is stochastically related to E_p, and so we need a statistical model for the data, that relates the intensity-reading (in ADUs) to E_p. We denote the intensity-reading of pixel p for collection-interval t as w_p^t, and the collection of all data is represented by W. The following is a simple phenomenological model that has been found to effectively match the signal-to-shot-noise ratio provided by (EM)CCD camera manufacturers (Huang *et al.* (2013)):

$$w_p^t :\sim \text{Gamma}(E_p/f, fg). \tag{2.4}$$

The parameter f is the excess noise factor, and is a known, fixed quantity that is specific to the camera provided by the manufacturer. The gain, g, is a constant, physically-adjustable parameter of the camera that (not unlike a guitar amplifier) increases the "loudness" of the signal (expected value), but also its "noisiness" (variance). It is typically not known *precisely* (but should be known within some range), and so considering it unknown in an inference framework may be useful.

2.2 Description of TIRF Model

In TIRF microscopy, illuminating light is made incident to a coverslip such that the light is totally reflected off of the coverslip, creating an evanescent wave that travels into the sample, and decays exponentially (see Fig. 2.1). We let $Z_{1:M} \triangleq [z_1, z_2, \ldots, z_M]$ denote the depth-in-sample of the emitters, and let $X_{1:M}$ represent the xy-coordinates of the emitters. The fluorescence profile is decomposed into the

product of separate functions $F_{xy}(x, y)$ and $F_z(z)$. We capture the behavior of the evanescent wave excitation with $F_z(z)$:

$$F(x, y, z) := F_{xy}(x, y)F_z(z) \tag{2.5}$$

$$F_z(z) = \exp(-z/d) \tag{2.6}$$

$$d \triangleq \left(\lambda_0/(4\pi)\right)\left[n_{\mathrm{cs}}^2\sin^2(\theta_{\mathrm{inc}}) - n_{\mathrm{s}}^2\right]^{-\frac{1}{2}}. \tag{2.7}$$

The parameter λ_0 is the vacuum wavelength of the incidence light, θ_{inc} is the incidence angle (measured from the vector normal to the coverslip surface), n_{cs} and n_{s} are the coverslip and sample refractive indexes respectively.

Because the emitters take 3D coordinates, we must now take into account that the emitters may not be located on the focal plane ($z = 0$), and consider the width of the PSF to be a function of the emitter's distance from the focal plane at the coverslip (Zhang *et al.* (2007)):

$$\varphi(z) = (\lambda_{\mathrm{em}}/n_{\mathrm{s}})^2(Z/(2\pi)) + (1/(4Z))z^2 \tag{2.8}$$

$$Z = \frac{7\left[1 - \left[1 - (N/n_{\mathrm{s}})^2\right]^{\frac{3}{4}}\right]}{4 - 7\left[1 - (N/n_{\mathrm{s}})^2\right]^{\frac{3}{4}} + 3\left[1 - \left[1 - (N/n_{\mathrm{s}})^2\right]^{\frac{7}{4}}\right]}. \tag{2.9}$$

Here, N is the detection objective numerical aperture.

2.2.1 Assumptions of the TIRF Model

We have assumed that emitter behavior is uniform, except for differing, spatially-dependent effective fluorescence, and that there are no effective optical aberrations on the detection path (so that our model for the PSF in z is correct). We also assume that the focal plane is fixed perfectly at the coverslip. The coverslip and the sample are both assumed to have uniform refractive indexes.

Because we have formally stated the z-dependence of the fluorescence profile with $F_z(z)$, and will ultimately seek to infer $F_{xy}(x, y)$, we assume that all spatially-dependent fluorescence effects other than the decaying evanescent wave will depend on xy, but not z. The ROI in z may be constricted, due to the thin samples typically imaged in TIRF.

Emission kinetics, which is to say time-varying fluorophore chemical states during which the emitters may fluoresce at different intensities, or cease to fluoresce at all (dark state), either temporarily (blinking) or permanently (bleaching) are assumed to be captured by our stochastic camera model. In other words, we assume that these kinetics occur on time-scales that are much faster than the exposure time, and that none of the molecules photobleach during imaging, so that there is in some sense a photo-kinetic equilibrium assumption. Many molecules are indeed suitably stable (Giepmans $et\ al.$ (2006); Zheng $et\ al.$ (2014)).

2.3   Computational Framework

We will now detail our Bayesian inference framework to infer the parameters $X_{1:M}$, $F(\cdot)$, and g, when given the data W and other relevant information about the system that we assume to know, namely in our implementation the size and coverage of the pixels, camera exposure time, incidence light angle and wavelength, emission wavelength, refractive indexes of the sample and coverslip, objective numerical aperture, excess noise factor, and photon background parameter(s) (see Table A.2 for a glossary of variables). In essence, we seek to characterize the following posterior probability distribution:

$$P(X_{1:M}, F(\cdot), g|W) = \frac{P(W|X_{1:M}, F(\cdot), g)P(X_{1:M}, F(\cdot), g)}{P(W)}. \qquad (2.10)$$

The distribution $P(W|X_{1:M}, F(\cdot), g)$ is the likelihood, and is specified by our model (Eq. 2.11):

$$P(W|X_{1:M}, F(\cdot), g) = \prod_{t=1}^{T} \prod_{p=1}^{P} \text{Gamma}(w_p^t; E_p/f, fg). \tag{2.11}$$

The distribution $P(X_{1:M}, F(\cdot), g)$ is the prior, which is interpreted in a Bayesian context as the distribution encapsulating our prior beliefs about the random variables in the absence of the data, which we specify later. Because these random variables are independent:

$$P(X_{1:M}, F(\cdot), g) = P(X_{1:M})P(F(\cdot))P(g). \tag{2.12}$$

The distribution $P(W)$, sometimes referred to in the literature as the evidence, is a normalizing constant; typically it is unimportant, barring the case that it is zero (signifying, in some sense, invalid data).

Our general strategy is to use Markov chain Monte Carlo (MCMC, see App. D.1), specifically the Metropolis-Hastings algorithm, to draw samples of the variables from the posterior distribution of interest. With a sufficiently large number of samples, we can characterize the distributions of the individual random variables, given the data. Overall, we use a special case of the Metropolis-Hastings algorithm, called Gibbs sampling, to iteratively sample the individual variables conditioned on the others:

$$\vdots$$

$$X_{1:M}^{(i)} \sim P(X_{1:M}|F(\cdot)^{(i-1)}, g^{(i-1)}, W) \tag{2.13}$$

$$F(\cdot)^{(i)} \sim P(F(\cdot)|X_{1:M}^{(i)}, g^{(i-1)}, W) \tag{2.14}$$

$$g^{(i)} \sim P(g|X_{1:M}^{(i)}, F(\cdot)^{(i)}, W). \tag{2.15}$$

$$\vdots$$

We index this sampling scheme by (i), and refer to it as the main scheme or main loop. The method we use to sample each individual conditional distribution can be

thought of as a self-contained framework for inferring the variable in question, given all the others. It is trivial to add additional variables to this framework, though of course sampling the appropriate distribution may not be.

The gain (g) is sampled using a relatively simple Metropolis-Hastings step that is detailed in App. E.1. We choose the uniform distribution as its prior; reasons for this choice are discussed in App. C:

$$P(g) := \mathrm{Unif}(g_{\mathrm{min}}, g_{\mathrm{max}}). \tag{2.16}$$

Our strategies for sampling the conditional distributions of $X_{1:M}$ and $F(\cdot)$ are non-trivial nested MCMC schemes that we summarize in the following sections.

2.3.1 Sampling the Fluorescence Profile

Using Bayes' rule with Eq. 2.14:

$$P(F(\cdot)|X_{1:M}, g, W) \propto P(W|F(\cdot), X_{1:M}, g)P(F(\cdot)). \tag{2.17}$$

We specify the prior on $F(\cdot)$ to be a Gaussian process (Rasmussen (2004); Williams and Rasmussen (2006); Wilson and Adams (2013)):

$$P(F(\cdot)) := \mathrm{GP}(\mu(\cdot), K(\cdot, \cdot)). \tag{2.18}$$

In short, every arbitrary, finite collection of fluorescence profile values (which we denote $F_{1:N} \triangleq [F(x_1), \ldots, F(x_N)]$) will have a joint multivariate normal distribution, with the mean and covariance functions $\mu(\cdot)$ and $K(\cdot, \cdot)$ describing how the mean and covariance are to be constructed, given any arbitrary collection of input points $X_{1:N} \triangleq [x_1, x_2, \ldots, x_N]$. Our choice of $\mu(\cdot)$ and $K(\cdot, \cdot)$ will encapsulate our prior assumptions about the properties of the profile (e.g., periodicity and smoothness) because choosing a particular mean and covariance function will result in higher probabilities for such behavior.

Though our framework allows for flexibility in such choices, in our implementation we have chosen a constant mean function and the squared exponential kernel with a nugget regularizer (Mohammadi *et al.* (2016)) as the covariance function:

$$\mu(\cdot) \coloneqq H \tag{2.19}$$

$$K(\cdot, \cdot) \coloneqq \mathrm{SE}(\cdot, \cdot) + \mathrm{NR}(\cdot, \cdot) \tag{2.20}$$

$$\mathrm{SE}(x_1, x_2) \triangleq \sigma_{\mathrm{GP}}^2 \exp\left(\frac{\|x_1 - x_2\|^2}{2l_{\mathrm{GP}}^2}\right) \tag{2.21}$$

$$\mathrm{NR}(x_1, x_2) \triangleq \begin{cases} \tau^2 & x_1 = x_2 \\ 0 & x_1 \neq x_2 \end{cases}. \tag{2.22}$$

App. E.2 contains a more detailed discussion of our choice for the covariance function.

In our implementation, we treat H as a random variable and refer to it as the baseline fluorescence, choosing for it a normal distribution as the prior:

$$P(H) \coloneqq \mathrm{N}(H_0, \sigma_H^2). \tag{2.23}$$

To put the formulation in a slightly more convenient form, we introduce the fluorescence deviation profile: $\Delta(\cdot) \triangleq F(\cdot) - H$. We can think of $\Delta(\cdot)$ as the zero-mean, spatially-dependent component of the fluorescence, which has the same covariance properties as $F(\cdot)$:

$$P(\Delta(\cdot)) \coloneqq \mathrm{GP}(0, \mathrm{SE}(\cdot, \cdot) + \mathrm{NR}(\cdot, \cdot)). \tag{2.24}$$

Now that our prior is fully specified, we consider how to arrive at a sample for $F(\cdot)$. First, we sample the values of the profile at locations where emitters exist (on which our likelihood will depend), $F_{1:M} \triangleq F(X_{1:M})$, and use these values as input to subsequently sample values for the profile on a fine, fixed grid of points $G_{1:N} \triangleq [x_{g1}, \ldots, x_{gN}]$. We can approximate the full function by interpolating between

the values for the profile over the grid-points, $F_{1:N} \triangleq F(G_{1:N})$:

$$F_{1:N} \sim \mathrm{N}(F_{1:N}|F_{1:M}, X_{1:M}, G_{1:N}). \tag{2.25}$$

$$F(\cdot) \approx \mathrm{Interpolate}(F_{1:N}). \tag{2.26}$$

The form of the distribution in Eq. 2.25 is given in App. E.2. We have that $F_{1:M} \equiv H + \Delta_{1:M}$, where $\Delta_{1:M} \triangleq [\Delta(x_1), \dots, \Delta(x_M)]$, and so need a strategy to sample H and $\Delta_{1:M}$. For this, we use a Gibbs subsampling scheme, indexed by (k) (not to be confused with the main Gibbs scheme, indexed by (i)):

$$\vdots$$

$$H^{(k)} \sim P(H|\Delta_{1:M}^{(k-1)}, X_{1:M}, g, W) \tag{2.27}$$

$$\Delta_1^{(k)} \sim P(\Delta_1|H^{(k)}, \Delta_{2:M}^{(k-1)}, X_{1:M}, g, W)$$

$$\vdots$$

$$\Delta_m^{(k)} \sim P(\Delta_m|H^{(k)}, \Delta_{1:m-1}^{(k)}, \Delta_{m+1:M}^{(k-1)}, X_{1:M}, g, W) \tag{2.28}$$

$$\vdots$$

$$\Delta_M^{(k)} \sim P(\Delta_M|H^{(k)}, \Delta_{1:M-1}^{(k)}, X_{1:M}, g, W).$$

$$\vdots$$

Samples from the distributions in Eq. 2.27 and Eq. 2.28 are drawn using a Metropolis-Hastings step, with the relevant equations detailed in App. E.2.

If one does not expect the fluorescence profile to be approximately uniform, such as in the case of a structured illumination profile (Gustafsson *et al.* (2008); Gustafsson (2005)), it may be wise to use a fixed, spatially varying term $C(\cdot)$, in the shape of the expected fluorescence profile, and use the formulation $\Delta(\cdot) \triangleq F(\cdot) - C(\cdot)$. Here, $\Delta(\cdot)$ is zero-mean, and we do not need to use a baseline term H. In the case that one wishes to assuming the fluorescence profile is uniform for the entire ROI, the

problem reduces to inferring $F(\cdot) \equiv H$ (i.e., we need only to consider Eq. 2.27 and no longer need the Gaussian process or the interpolation approximation). Both are trivial changes to the above analysis.

2.3.2 Sampling Emitter Counts/Locations

In order to sample emitter counts/locations (Eq. 2.13), we utilize the beta-Bernoulli process (Jazani *et al.* (2019); MacEachern (2016); Broderick *et al.* (2012)) (App. D.3). That is, we introduce random vector $Y_{1:L}$ of L load locations that one can think of as "candidate locations" (may be 3D coordinates), along with a random vector $S_{1:L}$ of Bernoulli random variables associated with the load locations that indicate whether the load locations are active ($S_j = 1$ implying the existence of an emitter at load location Y_j) or inactive ($S_j = 0$, implying nonexistence). We treat the probability that a load is active as a random variable itself, and thus need a hyperprior on the probability that $S_j = 1$, which we denote Q_j (the random vector of probabilities corresponding to $S_{1:L}$ being $Q_{1:L}$):

$$P(Q_j) \coloneqq \text{beta}(\alpha_q, \beta_q) \tag{2.29}$$

$$P(S_j = 1|Q_j) \triangleq Q_j \tag{2.30}$$

$$P(Y_j) \equiv P(X_j) \coloneqq \text{Unif(ROI)} \tag{2.31}$$

$$X_{1:M} \equiv \{\bigcup_{i=1}^{L} Y_j | S_j = 1\}, \quad M \equiv \sum_{j=1}^{L} S_j. \tag{2.32}$$

Here, α_q and β_q are free, fixed hyperparameters.

Now that we have fully specified the variables and their priors, we concern ourselves with the inference of the random vectors $Q_{1:L}$, $S_{1:L}$, $Y_{1:L}$. We use a Gibbs

sampling scheme, which we index by (k):

$$\vdots$$

$$Q_{1:L}^{(k)} \sim P(Q_{1:L}|S_{1:L}^{(k-1)}, Y_{1:L}^{(k-1)}, F(\cdot), g, W) \tag{2.33}$$

$$S_{1:L}^{(k)} \sim P(S_{1:L}|Q_{1:L}^{(k)}, Y_{1:L}^{(k-1)}, F(\cdot), g, W) \tag{2.34}$$

$$Y_{1:L}^{(k)} \sim P(Y_{1:L}|Q_{1:L}^{(k)}, S_{1:L}^{(k)}, F(\cdot), g, W). \tag{2.35}$$

$$\vdots$$

We need a strategy to sample each of these three distributions. Eq. 2.33 is able to be sampled in closed form (see App. E.3). The vector $S_{1:L}$ we typically sample in blocks (which improves mixing, discussed further in App. E.3) using a Metropolis-Hastings scheme, and $Y_{1:L}$ is split into two separate subsets: active and inactive loads:

$$M^{(k)} \equiv \sum_{j=1}^{L} S_j^{(k)} \tag{2.36}$$

$$\hat{Y}_{1:M^{(k)}} \triangleq \{\bigcup_{i=1}^{L} Y_j | S_j = 1\} \tag{2.37}$$

$$\overline{Y}_{1:(L-M^{(k)})} \triangleq \{\bigcup_{i=1}^{L} Y_j | S_j = 0\} \tag{2.38}$$

$$Y_{1:L}^{(k)} \equiv \hat{Y}_{1:M^{(k)}} \cup \overline{Y}_{1:(L-M^{(k)})}. \tag{2.39}$$

Individual $\overline{Y}_j$ are sampled directly from the prior (as the likelihood does not depend on these locations), thereby scrambling inactive load locations. The elements $\hat{Y}_j$ are sampled individually using a Metropolis-Hastings step (App. E.3). An illustration of our method is given in Fig. E.2

2.4   Method Summary and Pseudocode

The pseudocode presented as Fig. 1 summarizes the structure of our full MCMC scheme. The scheme for sampling $X_{1:M}$ is encapsulated in lines 3-8, the scheme for

$F(\cdot)$ in lines 10-14, and g in line 15; note that in all of these cases, we will want many (i) indexed samples, so each (k) loop will be executed many times (in other words, we will always need lines 1 and 16, even if we are sampling only one of the random variables).

A graphical model and informal break-down of all variables are presented in Fig. 2.2, and Table E.4 contains a table that summarizes all of the random variables, their priors, and summarizes the sampling method.

2.5 Chapter Figures and Algorithms

Fig. 2.1

Illustration of TIRF, labeled with our notation for the parameters.

Alg. 1

Our full sampler algorithm. Here, $\star$ indicates that elements of the vector are sampled individually, and $\star\star$ indicates that elements of the vector are sampled in randomized blocks (both cases are nested Gibbs schemes). The sampler method is indicated in parenthesis. Variables on which the relevant distributions do not depend are omitted from the notation.

Fig. 2.2

Graphical representations of the random variables. (a) Formal graphical model. Blue circles indicate random variables, and the gray circle denotes observations. (b) Informal breakdown of all variables. Those with circles are model parameters, and variables that are not circled are intermediary and computational in nature.
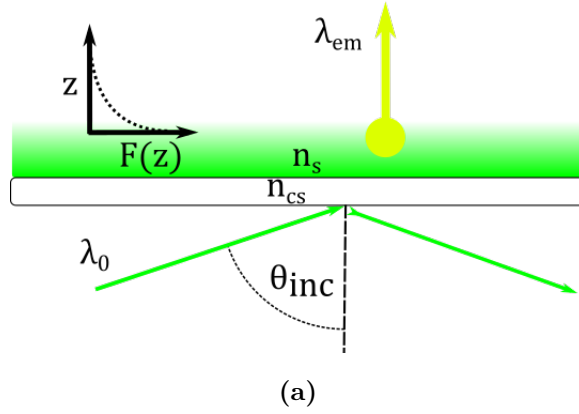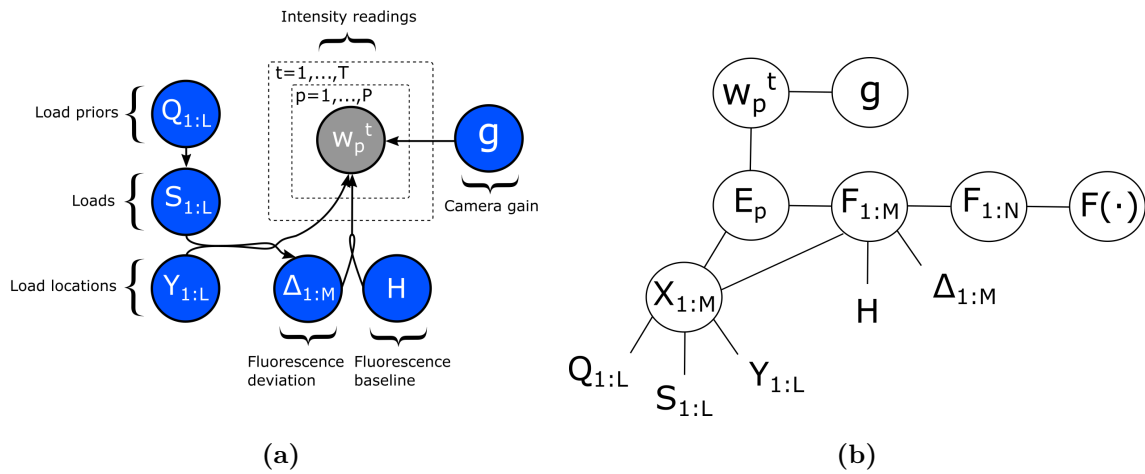
Figure 2.1

(a)

(b)

Figure 2.2

Algorithm 1 Full Sampler

1: for i:

2: for k:

3:   $Q_{1:L}^{(k)} \sim P(Q_{1:L}|S_{1:L}^{(k-1)})$ (Sampled directly)

4:   $S_{1:L}^{(k)} \sim P(S_{1:L}|Y_{1:L}^{(k-1)}, W, F(\cdot)^{(i-1)}, g^{(i-1)})$ ($\star\star$, Met.-Hast.)

5:   $\overline{Y}_{1:(L-M^{(k)})}^{(k)} \sim P(\overline{Y}_{1:(L-M^{(k)})})$ ($\star$, Sampled directly)

6:   $\hat{Y}_{1:M^{(k)}}^{(k)} \sim P(\hat{Y}_{1:M^{(k)}}|W, F(\cdot)^{(i-1)}, g^{(i-1)})$ ($\star$, Sampled directly)

7:   $Y_{1:L}^{(k)} = \overline{Y}_{1:(L-M^{(k)})}^{(k)} \cup \hat{Y}_{1:M^{(k)}}^{(k)}$

8: end

9: $X_{1:M^{(i)}}^{(i)} = \{\bigcup\limits_{i=1}^{L} Y_j^{(k_{\max})}|S_j^{(k_{\max})} = 1\}$

10: for k:

11:   $H^{(k)} \sim P\big(H|\Delta_{1:M^{(i)}}^{(k-1)}, W, X_{1:M^{(i)}}^{(i)}, g^{(i-1)}\big)$ (Met.-Hast.)

12:   $\Delta_{1:M^{(i)}}^{(k)} \sim P\big(\Delta_{1:M^{(i)}}|H^{(k)}, W, X_{1:M^{(i)}}^{(i)}, g^{(i-1)}\big)$ ($\star$, Met.Hast.)

13: end

14: $F(\cdot)^{(i)} \sim P\big(F(\cdot)|H^{(k_{\max})} + \Delta_{1:M^{(i)}}^{(k_{\max})}\big)$ (Sampled directly)

15: $g^{(i)} \sim P\big(g|W, F(\cdot)^{(i)}, X_{1:M^{(i)}}^{(i-1)}\big)$ (Met.-Hast.)

16: end

Chapter 3

RESULTS - TIRF

First, we generate synthetic data from our TIRF model given in Sec. 2.2. We assume in our analysis that emitters are located within 10nm of the focal plane ($z = 0$), suggestive of an imaging experiment for a thin sample, such as a cell membrane, or for objects stuck to the coverslip (Jaiswal and Simon (2007); Betzig *et al.* (2006); Yildiz *et al.* (2003)); the ROI in xy is 12×12 pixels, each pixel covering a 100nm×100nm area in image space. A table of parameters, motivated by cited literature, is provided in App. B, and these are the values used in the simulations/analysis unless otherwise noted. Four representative datasets are shown in Fig. 3.2, which are generated using the fluorescence profile in Fig. 3.1a. In Fig. 3.2a and 3.2b, emitters are placed on a diagonal line in xy with equal xy spacing, and z alternates between 0nm and 10nm. Figure 3.2c and 3.2d were generated with 8 and 16 emitters placed randomly in the ROI.

In the following subsections, we demonstrate the efficacy of our subsampling scheme for g, $X_{1:M}$, and $F_{xy}(x, y)$, given the others, and then show results for our method when inferring all three variables concurrently.

3.1   Inferring the Gain

Figure 3.3 shows results for our subsampling scheme on the gain g, given every other variable/parameter (line 15 in Alg. 1). These are obtained almost instantly, and performance (computational time and the accuracy/precision) generally does not depend on the number or concentration of emitters or the fluorescence profile. Varying the gain over a range of reasonable values (∼10-100) (Mortensen and Fly-

17

vbjerg (2016)), and even outside this range, also does not change the efficiency or accuracy/precision of the inference.

3.2   Inferring the Fluorescence Profile

Figure 3.4 shows results for our subsampling scheme on F_{xy}, given every other variable/parameter (lines 10-14 in Alg. 1). These results are obtained in about 5 seconds. Computational costs increase as emitters are added, but the matrices involved in sampling F_{xy} need only be calculated once, because they depend on the location of the stationary emitters, which is given in this context.

Though our likelihood depends only on the value of the profile at emitter locations, given the assumptions about the profile encapsulated by the choice of the Gaussian process covariance function and its hyperparameters, we are able to infer the function over the entire ROI. The error and standard deviation of the samples typically grow as one moves farther away from the emitters, because we have less information; more emitters provides more information and thus improves the inference of the profile.

3.3   Inferring Emitter Locations

Figure 3.5 shows results for our subsampling scheme on $X_{1:M}$ (lines 2-9 in Alg. 1), which are attained in about 15 seconds. By pooling emitter localizations from every step in the chain, we obtain a heat map of emitter locations. As a means to obtain point estimates for the locations of the emitters, we examine the $M^{(i)}$ samples and select a value for k by hand and perform k-means clustering (Hartigan and Wong (1979)) on the pooled samples where $M^{(i)} = k$.

When emitters make distinct PSFs, they are easy to count and localize. However, when the PSFs become muddled, localizations are more ambiguous, though typically the count remains fairly accurate (the sampler sometimes under-count or over-count

by 1 or 2 emitters) given that the sampler has had enough time to converge (more ambiguity requires more sampling). Heat maps of locations for clusters of fluorophore typically resemble the cluster, making localizations imprecise in this case. Localization in z is particularly imprecise when there are multiple co-local emitters, but in contrast, for unambiguously isolated emitters, the additional information we have through the intensity of fluorescence helps to localize in z (see Fig. B.2). For this reason, having a known and sharply-varying profile in xy may also aid in the xy localization precision.

3.4 Inferring All Three Variables

Figure 3.6 shows results when we sample $X_{1:M}$, F_{xy}, and g altogether (following the pseudocode presented in Alg. 1), with additional results using a higher molecule concentration in App. B. The computational time required to obtain good sampling ergodicity is much higher in this case than for sampling any individual variable, because the parameter space for all three variables, particularly in combining unknown $X_{1:M}$ with F_{xy}, is far larger than that of any one variable; our results are obtained in about 2 hours. The size of the grid $G_{1:N}$ becomes a significant factor in computation times, because most of the matrices involved in the computations for sampling F_{xy} need to be re-computed after the locations of the active emitters move (that is, at every iteration of the main loop).

We have some degree of nonidentifiability between z depth, emitter count, F_{xy} profile, and even the gain that creates difficulty for our sampler. Nonidentifiability issues are discussed at length in App. C. With low ambiguity between these variables (as is the case with the results in Fig. 3.6), our inference for the gain is generally good, and xy counting/localization is quite accurate, but we can see how z depth ambiguity affects our inference for the profile: the emitters are estimated, on average,

in the middle of the ROI in z. The profile is overestimated where the emitters are located at $z = 0$nm, and underestimated where the emitter is at $z = 10$nm. Shallow ROIs in z helps to mitigate this nonidentifiability. Typically, when inferring $X_{1:M}$ and F_{xy} together, the best results are obtained when PSFs are distinct, the fluorophore count is unambiguous, and there is a somewhat restrictive prior on g. Figure B.1 shows results for when these conditions begin to break down.

3.5 Chapter Figures, Tables, and Algorithms

Fig. 3.1

Two different F_{xy} fluorescence profiles and a cross-sections of the PSF at various z depths. The profiles are generated with one (a) or two (b) Gaussian depressions, units in kHz. (c) A cross-section of the PSF (Eq. 2.2), the height (Eq. 2.6) and width (Eq. 2.8) of which is scaled depending on various z values. One can think of this as a $y = 0$ cross-section, with the emitter centered at the origin, and $F_{xy}(x, y) \equiv 1$. (d) An alternate visualization (contour plot) of the same profile shown in (b).

Fig. 3.2

Maximum intensity projections of representative synthetic datasets. Black dots correspond to actual xy locations of the emitters. (a,b) Emitters are located equidistant on the line $x = y$; the distance in xy being 500nm and 150nm respectively. The z values alternate between 0nm and 10nm. (c,d) Emitters ($M = 8, 16$) are placed stochastically in the ROI.

Fig. 3.3

Histograms of $g^{(i)}$ sampling chains for data generated with various g values. Burn-in is considered the first 20% of the 10,000 samples. In each figure, the emitter locations correspond to those reflected in Fig. 3.2c, and the profile is shown in Fig. 3.1a (single Gaussian depression). We vary the ground truth gain, shown with a dashed line in each figure.

Fig. 3.4

Mean-of-chain and/or standard error results for the inference of varying F_{xy}. Burn-in is considered the first 20% of the 500 samples. The black dots are ground-truth xy locations of the emitters. (a) The mean of the F_{xy} samples, analyzing data with emitter locations represented in Fig. 3.2b, but using the xy profile shown in Fig. 3.1b; (b) shows the standard error (mean-of-chain minus ground truth over the sample standard deviation) of the chain represented in (a). (c) We change the emitter locations to those reflected in Fig. 3.2c, but not the profile, and show the standard error when analyzing the new data. (d) We change the profile to that shown in Fig. 3.1a, and show the standard error of the subsequent results.

Fig. 3.5

Pooled scatterplots of $X_{1:M}$ and $Z_{1:M}$ samples. Burn-in was considered the first 20% of the 10,000 samples. Small blue dots represent the pooled samples ($X_{1:M^{(i)}}^{(i)}$ and $Z_{1:M^{(i)}}^{(i)}$), which may be repeated, possibly a large number of times. With k set to the mode of $M^{(i)}$, k-means clustering was performed. Median separation is the median distance from the k-means clusters to the actual emitter location. The only set for which a vast majority of $M^{(i)}$ did *not* equal the set k value was in (d), where

there was ambiguity between counts of $M =$ 13, 14, 15, and 16. The locations of the emitters correspond to the respective maximum intensity projections in Fig. 3.2; the profile used in generating the data is shown in Fig. 3.1a.

Fig. 3.6

Results of sampling $X_{1:M}$**,** F_{xy}**, and** g **altogether.** The data analyzed is represented in Fig. 3.2a. Burn-in is considered the first 20% of the 250 main loop samples. (a) Plot of pooled xyz localizations; $M^{(i)} = 4$ for all samples after burn-in. Also plotted is the maximum *a posteriori* estimate of the localizations, and the resulting k-means clusters with $k = 8$. (b) Histogram of $g^{(i)}$ values. (c) Standard error of the $F_{xy}^{(i)}$ profile samples. (d) Standard deviation of the $F_{xy}^{(i)}$ profile samples.
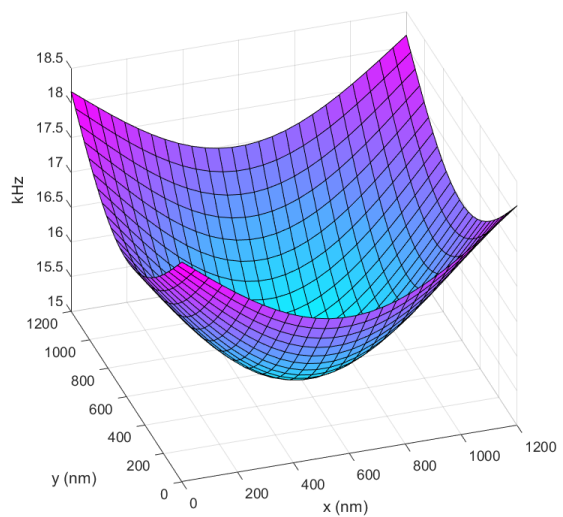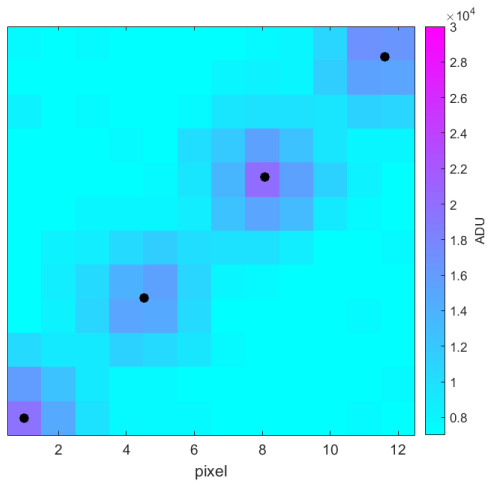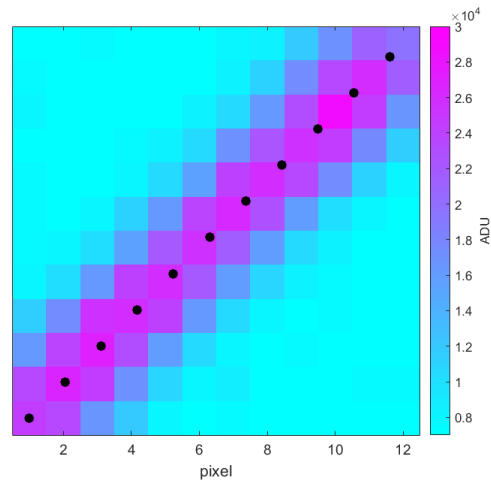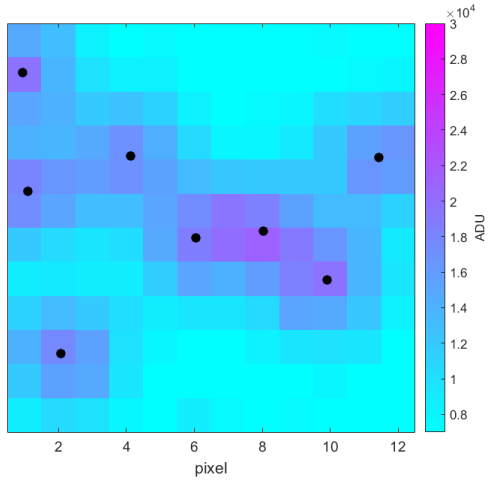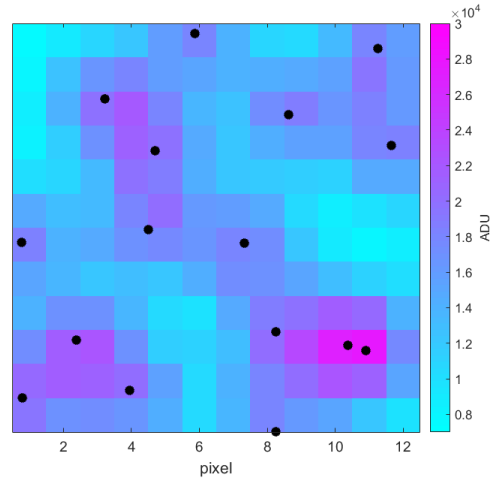
(a)
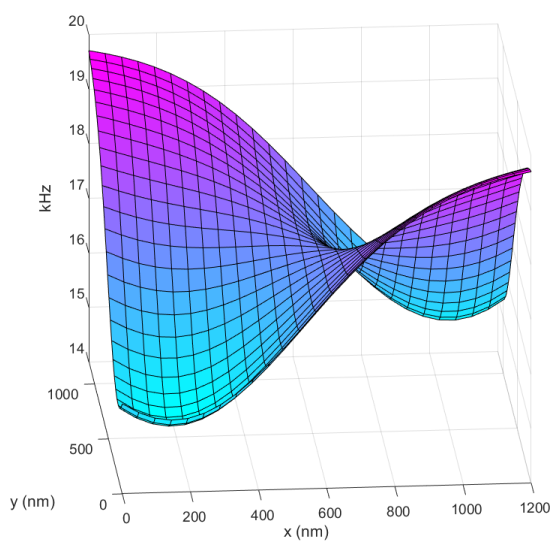
(b)
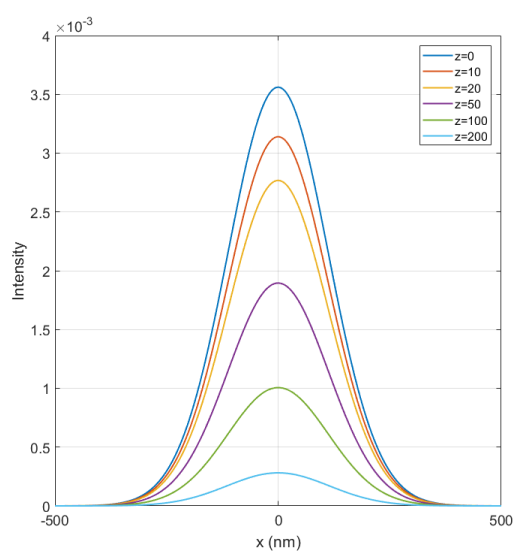
(c)
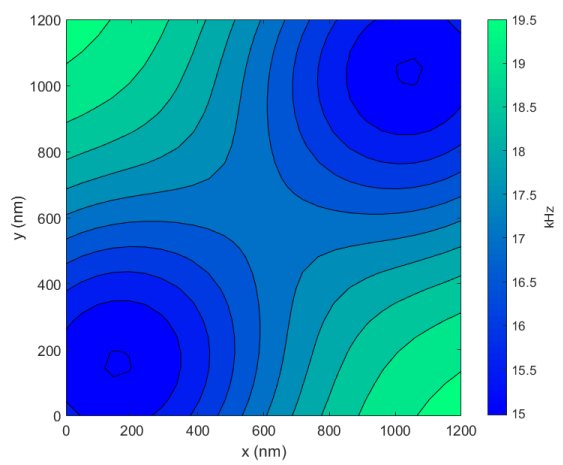
(d)

Figure 3.1

(a)

(b)

(c)

(d)

Figure 3.2
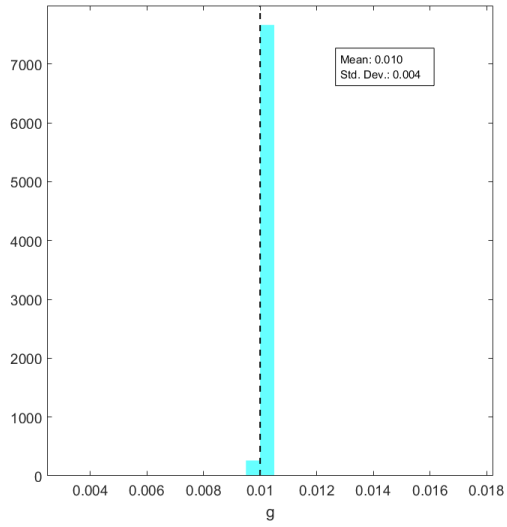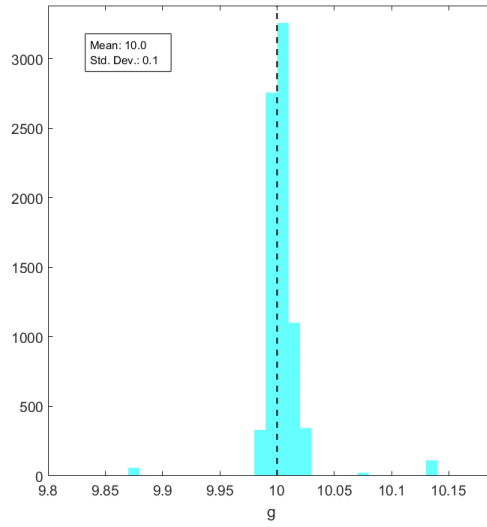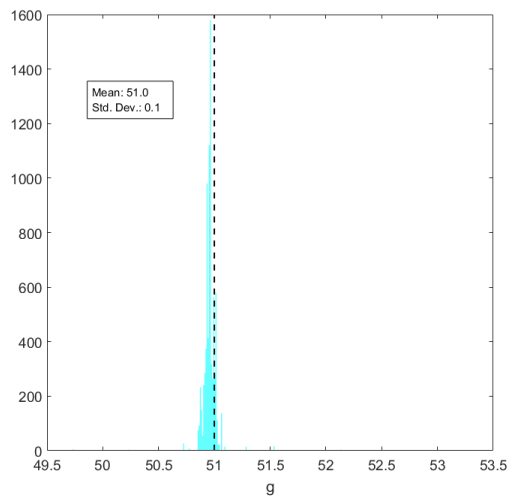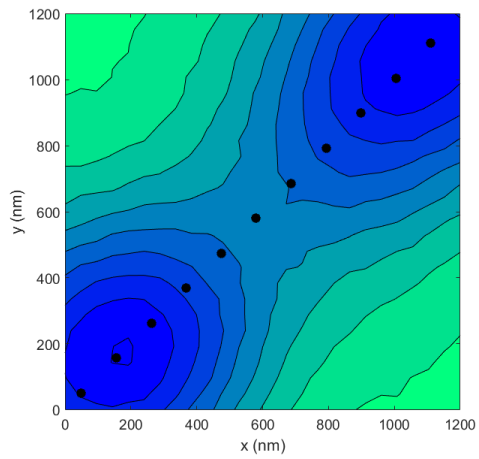
Figure 3.3

(a)
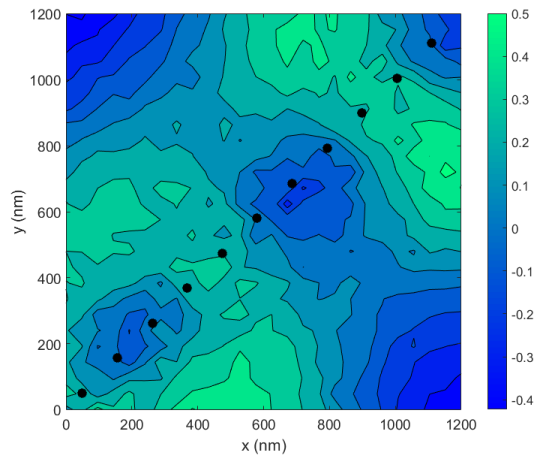
(b)
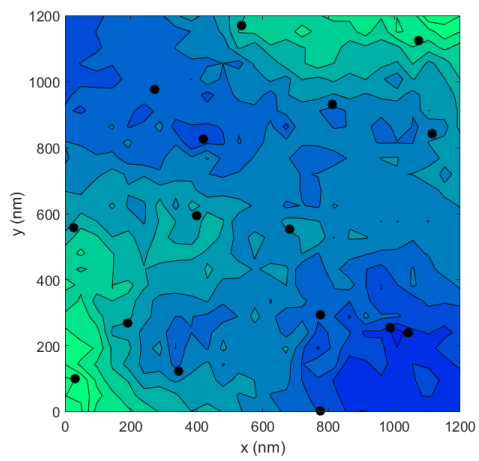
(c)
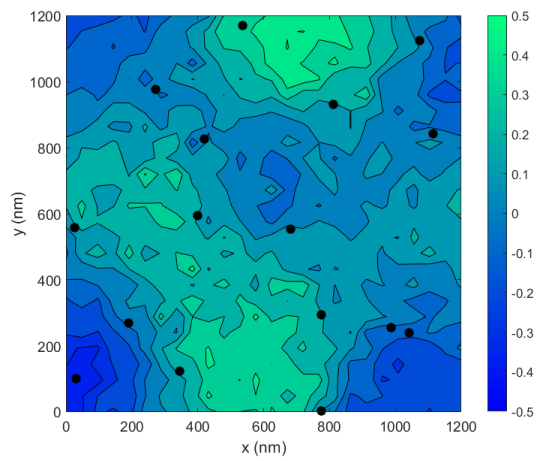
(d)

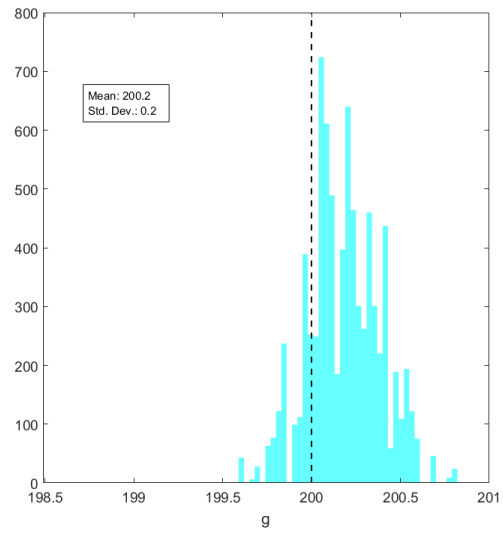**Figure 3.4**
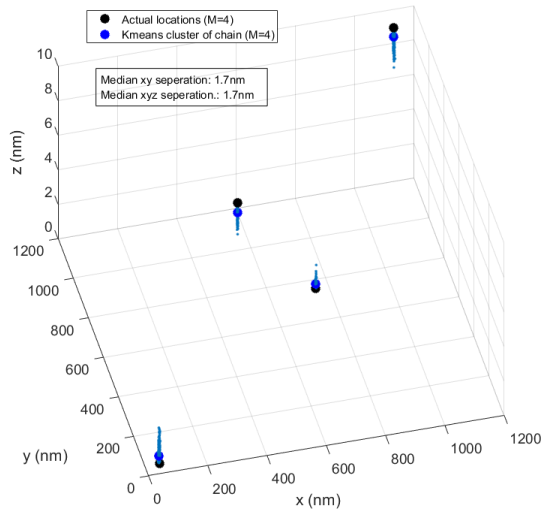
(a)
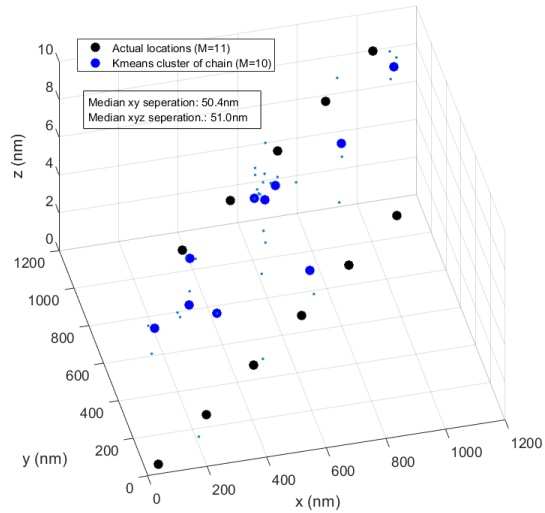
(b)
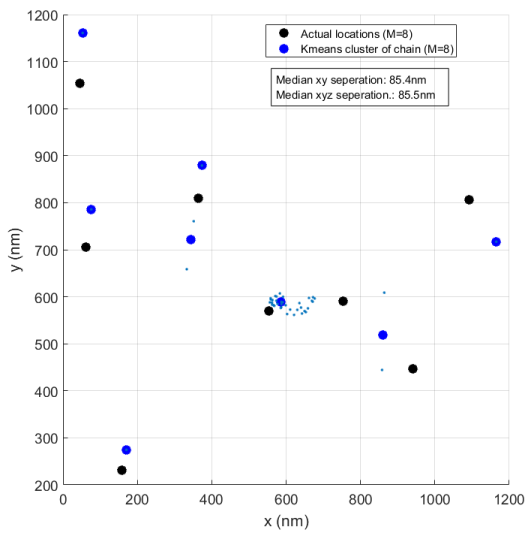
(c)
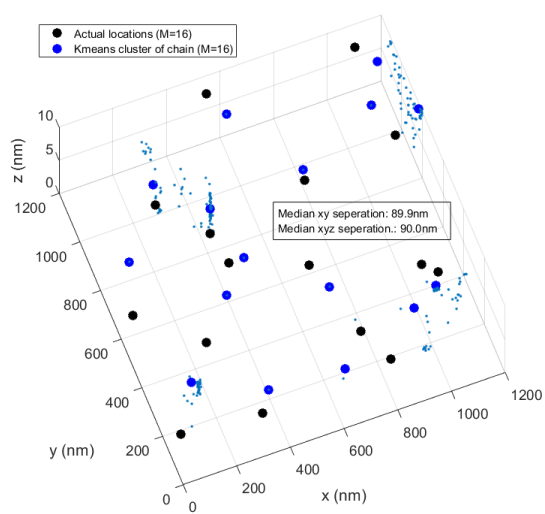
(d)

Figure 3.5
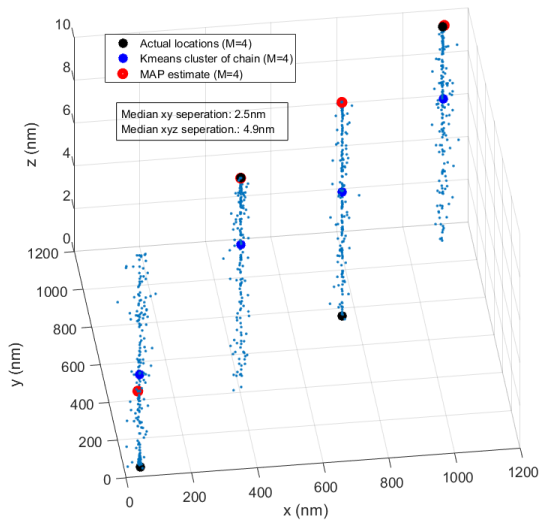
(a)
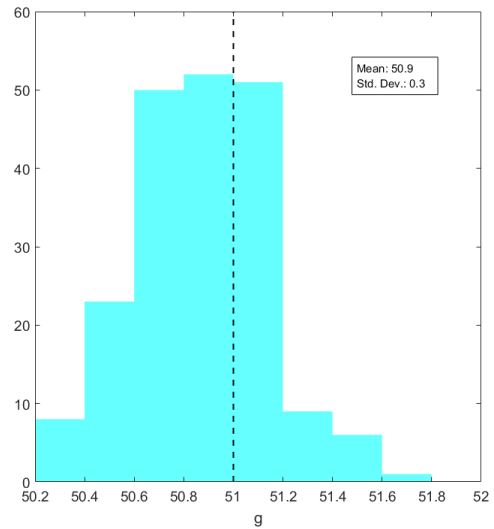
(b)
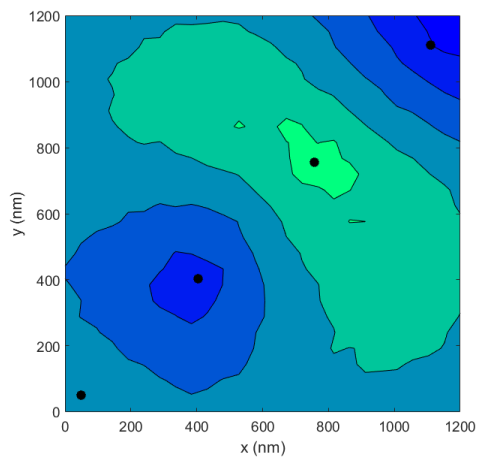
(c)
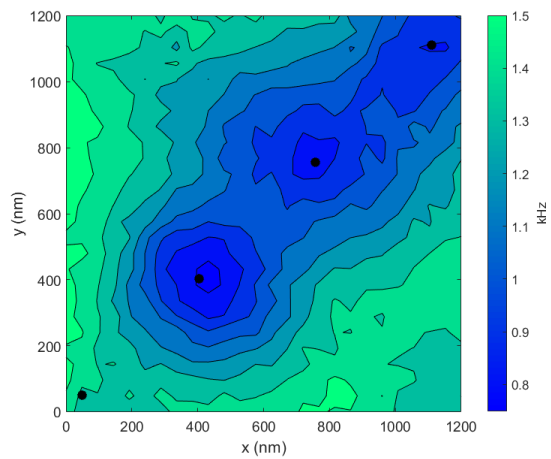
(d)

Figure 3.6

Chapter 4

DISCUSSION

Our TIRF implementation is able to accurately count the number of fluorophores in an ROI, and localize them with very high accuracy and precision (within 2nm) if PSFs are isolated and the fluorescence profile is known (Fig. 3.5 and Fig. B.2). Inference on the gain (Fig. 3.3) and fluorescence profile (Fig. 3.4), given emitter locations, are also quite pleasing. Although functionality is limited, and effective sampling orders of magnitude slower in the case that both emitter locations and fluorescence profile are unknown, concurrent analysis may in some cases be useful (Fig. 3.6).

This method is best applied to experiments in which either the emitter locations or fluorescence profile is in question, with the other known and fixed. One may perhaps be interested in performing a two-phase experiment, in which fluorophore are imaged with a known fluorescence profile, localized, and then imaged after some kind of change to the system that induces a new fluorescence profile that can then be inferred, given the locations of the emitters. One may also be interested in using the method to localize isolated fluorophores with very high precision, perhaps in techniques that feature stochastically-activated molecules such as PALM (Patterson *et al.* (2007); Lee *et al.* (2012); Chen *et al.* (2014)) and STORM (Rust *et al.* (2006); Hainsworth *et al.* (2018)). Researchers may perhaps find our framework for the gain (detailed in App. E.1), useful in calibrating their cameras or optical system, or perhaps want to adapt the method for other univariate calibration parameters.

The process is highly automated, requiring as inputs known information about the system, the data, and a few computational parameters that have to do with the

program, like the size of the grid $G_{1:N}$ and the number of component subsampler iterations. Though accurate and precise values for prior parameters will help, their effect on the results diminishes as the number of main loop iterations grows, because the posterior distribution tends to be characterized more by the likelihood (for this reason, the prior(s) will matter more when W contains little information, e.g., for small $T \cdot \Delta t_e$). Coming up with hyperparameters for the Gaussian process prior may be a bit tricky, and one possible future direction is to use MCMC to infer values for these hyperparameters (Titsias *et al.* (2008)), or even use model selection techniques to find a covariance function that works well for the data.

The flexibility of our general framework opens up the possibility of many similar implementations, perhaps for the inference of different parameters or applications to different microscope set-ups. For instance, we have only considered (EM)CCD cameras because they are the main workhorse, but an almost-identical scheme can be applied with a different stochastic camera model (e.g., for sCMOS cameras). We may also re-formulate the model for any PSF. Many exotic, engineered detection PSFs have been developed for methods that improve localization accuracy by inducing a PSF that provides additional information (especially in z) (Quirin *et al.* (2012); Berlich *et al.* (2016); Pavani *et al.* (2009)). Also, we have not taken particular interest in characterizing emission kinetics, but they can be modeled using Bayesian techniques and incorporated into the framework. One could perhaps calibrate or study illumination profiles in Structured Illumination Microscopy/Mode (SIM) (Gustafsson *et al.* (2008); Gustafsson (2005)) by providing some kind of fixed grid of fluorescent objects that form distinct PSFs, as these objects will be easy to localize and thus inference of the profile will be very accurate/precise.

Other types of spatially-varying functions may also be inferred using similar analysis, such as a refractive index map in TIRF, or a complex dielectric profile in 3D

imaging techniques. One may also want to consider a fluorescence profile that varies in time, and quantifying such time-varying fluorescence may aid in the study of dynamic local sample environments. Our general counting/localization scheme may find application in many other applications or fields.

For our general method, the high computational cost may be a drawback, and indeed preclude some potential applications. It is difficult, for instance, to imagine mapping large specimens in their entirety, like a whole zebrafish, so our method is perhaps best thought of as a precision tool for mapping smaller structures, or possibly a low-definition estimator of a smooth fluorescence profile over a large ROI. Computational cost is difficult to quantify because it is so dependent on the individual computer system and many program parameters, but it may be possible to generally speed-up computation time by using the GPU to calculate larger covariance matrices and their operations, allowing larger or finer $G_{1:N}$ grids, larger ROIs, denser fluorophore concentration, etc. Parallel MCMC chains may also be applied for uncorrelated components of the framework.

Ultimately, what we have presented is a general framework for inferring parameters of a widefield fluorescence optical system where, given a specific model, we have inferred a specific (1) univariate random variable; (2) random vector of undetermined size; and (3) continuous, smooth function over physical space. The appropriation of these methods may lead to a plethora of new analysis tools.

REFERENCES

Aguet, F., S. Upadhyayula, R. Gaudin, Y.-y. Chou, E. Cocucci, K. He, B.-C. Chen, K. Mosaliganti, M. Pasham, W. Skillern *et al.*, "Membrane dynamics of dividing cells imaged by lattice light-sheet microscopy", Molecular biology of the cell **27**, 22, 3418–3435 (2016).

Andrieu, C., N. De Freitas, A. Doucet and M. I. Jordan, "An introduction to mcmc for machine learning", Machine learning **50**, 1-2, 5–43 (2003).

Axelrod, D., N. L. Thompson and T. P. Burghardt, "Total internal reflection fluorescent microscopy", Journal of microscopy **129**, 1, 19–28 (1983).

Bates, M., B. Huang, G. T. Dempsey and X. Zhuang, "Multicolor super-resolution imaging with photo-switchable fluorescent probes", Science **317**, 5845, 1749–1753 (2007).

Berlich, R., A. Bräuer and S. Stallinga, "Single shot three-dimensional imaging using an engineered point spread function", Optics express **24**, 6, 5946–5960 (2016).

Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution", Science **313**, 5793, 1642–1645 (2006).

Bishop, C., *Pattern Recognition and Machine Learning* (Springer, 2006).

Born, M. and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).

Broderick, T., M. I. Jordan, J. Pitman *et al.*, "Beta processes, stick-breaking and power laws", Bayesian analysis **7**, 2, 439–476 (2012).

Chen, B.-C., W. R. Legant, K. Wang, L. Shao, D. E. Milkie, M. W. Davidson, C. Janetopoulos, X. S. Wu, J. A. Hammer, Z. Liu *et al.*, "Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution", Science **346**, 6208, 1257998 (2014).

Chib, S. and E. Greenberg, "Understanding the metropolis-hastings algorithm", The american statistician **49**, 4, 327–335 (1995).

Cowles, M. K. and B. P. Carlin, "Markov chain monte carlo convergence diagnostics: a comparative review", Journal of the American Statistical Association **91**, 434, 883–904 (1996).

Deschout, H., K. Neyts and K. Braeckmans, "The influence of movement on the localization precision of sub-resolution particles in fluorescence microscopy", Journal of biophotonics **5**, 1, 97–109 (2012).

Deschout, H., F. C. Zanacchi, M. Mlodzianoski, A. Diaspro, J. Bewersdorf, S. T. Hess and K. Braeckmans, "Precisely and accurately localizing single emitters in fluorescence microscopy", Nature methods **11**, 3, 253 (2014).

Ferguson, T. S., "A bayesian analysis of some nonparametric problems", The annals of statistics pp. 209–230 (1973).

Fernández-Suárez, M. and A. Y. Ting, "Fluorescent probes for super-resolution imaging in living cells", Nature reviews Molecular cell biology **9**, 12, 929 (2008).

Fish, K. N., "Total internal reflection fluorescence (tirf) microscopy", Current protocols in cytometry **50**, 1, 12–18 (2009).

Gelfand, A. E., "Gibbs sampling", Journal of the American statistical Association **95**, 452, 1300–1304 (2000).

Giepmans, B. N., S. R. Adams, M. H. Ellisman and R. Y. Tsien, "The fluorescent toolbox for assessing protein location and function", science **312**, 5771, 217–224 (2006).

Grimm, J. B., B. P. English, H. Choi, A. K. Muthusamy, B. P. Mehl, P. Dong, T. A. Brown, J. Lippincott-Schwartz, Z. Liu, T. Lionnet *et al.*, "Bright photoactivatable fluorophores for single-molecule imaging", Nature methods **13**, 12, 985 (2016).

Gustafsson, M. G., "Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution", Proceedings of the National Academy of Sciences **102**, 37, 13081–13086 (2005).

Gustafsson, M. G., L. Shao, P. M. Carlton, C. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard and J. W. Sedat, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination", Biophysical journal **94**, 12, 4957–4970 (2008).

Hainsworth, A., S. Lee, P. Foot, A. Patel, W. Poon and A. Knight, "Super-resolution imaging of subcortical white matter using stochastic optical reconstruction microscopy (storm) and super-resolution optical fluctuation imaging (sofi)", Neuropathology and applied neurobiology **44**, 4, 417–426 (2018).

Hartigan, J. A. and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm", Journal of the Royal Statistical Society. Series C (Applied Statistics) **28**, 1, 100–108 (1979).

Hjort, N. L., C. Holmes, P. Müller and S. G. Walker, *Bayesian nonparametrics*, vol. 28 (Cambridge University Press, 2010).

Huang, B., M. Bates and X. Zhuang, "Super-resolution fluorescence microscopy", Annual review of biochemistry **78**, 993–1016 (2009).

Huang, F., T. M. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes *et al.*, "Video-rate nanoscopy using scmos camera–specific single-molecule localization algorithms", Nature methods **10**, 7, 653 (2013).

Jaiswal, J. K. and S. M. Simon, "Imaging single events at the cell membrane", Nature chemical biology **3**, 2, 92 (2007).

Jazani, S., I. Sgouralis, O. M. Shafraz, S. Sivasankar and S. Presse, "An alternative framework for fluorescence correlation spectroscopy", Biophysical Journal **116**, 3, 282a (2019).

Kilic, Z. and S. Pressé, "Bayesian nonparametric analysis of transcriptional processes", Biophysical Journal **116**, 3, 299a (2019).

Kong, M., L. Liu, X. Chen, K. I. Driscoll, P. Mao, S. Böhm, N. M. Kad, S. C. Watkins, K. A. Bernstein, J. J. Wyrick *et al.*, "Single-molecule imaging reveals that rad4 employs a dynamic dna damage recognition process", Molecular cell **64**, 2, 376–387 (2016).

Lee, A., K. Tsekouras, C. Calderon, C. Bustamante and S. Pressé, "Unraveling the thousand word picture: an introduction to super-resolution data analysis", Chemical reviews **117**, 11, 7276–7330 (2017).

Lee, S.-H., J. Y. Shin, A. Lee and C. Bustamante, "Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (palm)", Proceedings of the National Academy of Sciences **109**, 43, 17436–17441 (2012).

Lichtman, J. W. and J.-A. Conchello, "Fluorescence microscopy", Nature methods **2**, 12, 910 (2005).

Link, W. A. and M. J. Eaton, "On thinning of chains in mcmc", Methods in ecology and evolution **3**, 1, 112–115 (2012).

Liu, Z., L. D. Lavis and E. Betzig, "Imaging live-cell dynamics and structure at the single-molecule level", Molecular cell **58**, 4, 644–659 (2015).

MacEachern, S. N., "Nonparametric bayesian methods: a gentle introduction and overview", Communications for Statistical Applications and Methods **23**, 6, 445–466 (2016).

Mattheyses, A. L., S. M. Simon and J. Z. Rappoport, "Imaging with total internal reflection fluorescence microscopy for the cell biologist", J Cell Sci **123**, 21, 3621–3628 (2010).

Mohammadi, H., R. L. Riche, N. Durrande, E. Touboul and X. Bay, "An analytic comparison of regularization methods for gaussian processes", arXiv preprint arXiv:1602.00853 (2016).

Mortensen, K. I. and H. Flyvbjerg, ""calibration-on-the-spot": How to calibrate an emccd camera from its images", Scientific reports **6**, 28680 (2016).

Ovesnỳ, M., P. Křížek, J. Borkovec, Z. Švindrych and G. M. Hagen, "Thunderstorm: a comprehensive imagej plug-in for palm and storm data analysis and super-resolution imaging", Bioinformatics **30**, 16, 2389–2390 (2014).

Patterson, G. H., E. Betzig, J. Lippincott-Schwartz and H. F. Hess, "Developing photoactivated localization microscopy (palm)", in "2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro", pp. 940–943 (IEEE, 2007).

Pavani, S. R. P., M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun and W. Moerner, "Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function", Proceedings of the National Academy of Sciences **106**, 9, 2995–2999 (2009).

Quirin, S., S. R. P. Pavani and R. Piestun, "Optimal 3d single-molecule localization for superresolution microscopy with aberrations and engineered point spread functions", Proceedings of the National Academy of Sciences **109**, 3, 675–679 (2012).

Rasmussen, C. E., "Gaussian processes in machine learning", pp. 63–71 (2004).

Rust, M. J., M. Bates and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)", Nature methods **3**, 10, 793 (2006).

Sako, Y., S. Minoghchi and T. Yanagida, "Single-molecule imaging of egfr signalling on the surface of living cells", Nature cell biology **2**, 3, 168 (2000).

Schermelleh, L., R. Heintzmann and H. Leonhardt, "A guide to super-resolution fluorescence microscopy", The Journal of cell biology **190**, 2, 165–175 (2010).

Schneckenburger, H., "Total internal reflection fluorescence microscopy: technical innovations and novel applications", Current opinion in biotechnology **16**, 1, 13–18 (2005).

Sgouralis, I., S. Madaan, F. Djutanta, R. Kha, R. F. Hariadi and S. Pressé, "A bayesian nonparametric approach to single molecule forster resonance energy transfer", The Journal of Physical Chemistry B **123**, 3, 675–688 (2018a).

Sgouralis, I. and S. Pressé, "An introduction to infinite hmms for single-molecule data analysis", Biophysical journal **112**, 10, 2021–2029 (2017).

Sgouralis, I., M. Whitmore, L. Lapidus, M. J. Comstock and S. Pressé, "Single molecule force spectroscopy at high data acquisition: A bayesian nonparametric analysis", The Journal of chemical physics **148**, 12, 123320 (2018b).

Shashkova, S. and M. C. Leake, "Single-molecule fluorescence microscopy review: shedding new light on old problems", Bioscience reports p. BSR20170031 (2017).

Shim, S.-H., C. Xia, G. Zhong, H. P. Babcock, J. C. Vaughan, B. Huang, X. Wang, C. Xu, G.-Q. Bi and X. Zhuang, "Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes", Proceedings of the National Academy of Sciences **109**, 35, 13978–13983 (2012).

Souza, C. R., "Kernel functions for machine learning applications", Creative Commons Attribution-Noncommercial-Share Alike **3**, 29 (2010).

Specht, E. A., E. Braselmann and A. E. Palmer, "A critical and comparative review of fluorescent tools for live-cell imaging", Annual review of physiology **79**, 93–117 (2017).

Su, Q. P. and L. A. Ju, "Biophysical nanotools for single-molecule dynamics", Biophysical reviews pp. 1–9 (2018).

Tavakoli, M., S. Jazani, I. Sgouralis and S. Presse, "Bayesian nonparametrics for fluorescence methods", Biophysical Journal **116**, 3, 39a (2019).

Tavakoli, M., J. N. Taylor, C.-B. Li, T. Komatsuzaki and S. Pressé, "Single molecule data analysis: an introduction", arXiv preprint arXiv:1606.00403 (2016).

Titsias, M. K., N. Lawrence and M. Rattray, "Markov chain monte carlo algorithms for gaussian processes", Inference and Estimation in Probabilistic Time-Series Models **9** (2008).

Williams, C. K. and C. E. Rasmussen, "Gaussian processes for machine learning", the MIT Press **2**, 3, 4 (2006).

Wilson, A. and R. Adams, "Gaussian process kernels for pattern discovery and extrapolation", in "International Conference on Machine Learning", pp. 1067–1075 (2013).

Wöll, D. and C. Flors, "Super-resolution fluorescence imaging for materials science", Small Methods **1**, 10, 1700191 (2017).

Yildiz, A., J. N. Forkey, S. A. McKinney, T. Ha, Y. E. Goldman and P. R. Selvin, "Myosin v walks hand-over-hand: single fluorophore imaging with 1.5-nm localization", science **300**, 5628, 2061–2065 (2003).

Zhang, B., J. Zerubia and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models", Applied optics **46**, 10, 1819–1829 (2007).

Zheng, Q., M. F. Juette, S. Jockusch, M. R. Wasserman, Z. Zhou, R. B. Altman and S. C. Blanchard, "Ultra-stable organic fluorophores for single-molecule research", Chemical Society Reviews **43**, 4, 1044–1056 (2014).

Zhi, B., Y. Cui, S. Wang, B. P. Frank, D. N. Williams, R. P. Brown, E. S. Melby, R. J. Hamers, Z. Rosenzweig, D. H. Fairbrother *et al.*, "Malic acid carbon dots: from super-resolution live-cell imaging to highly efficient separation", ACS nano **12**, 6, 5741–5752 (2018).

APPENDIX A

NOTATION TABLES

Table A.2

Summary of notation: important quantities and variables.

Symbol	Definition
$\triangleq$	(By) mathematical definition
$\equiv$	(By) mathematical equivalency
$:=$	"Set to" or "chosen to be"
$:\sim$	"Chosen to be sampled from"
$/$	Multiplication by matrix inverse (i.e., $A/B \equiv AB^{-1}$)
$A_{1:N}$	Vector or indexed set with elements 1 to N
$A^{(i)}$	Sample index i of an MCMC chain for random variable A
A^T	Transpose of matrix or vector A
$[A_1, A_2, \ldots]$	A row vector
$[A_1, A_2, \ldots]^T$	A column vector
$[A_1, A_2, \ldots] - x$	$\equiv [A_1 - x, A_2 - x, \ldots]$ (x is a number)
$A\|_{x=b}$	A evaluated with the value x set to b. May be shortened to $A\|_b$.
$\|A\|_{\text{size}}$	The size (length) of a matrix (vector)
I_N	The $N \times N$ identity matrix
$\ln a$	Natural logarithm of a (log base e)
$N(\mu, \sigma^2)$	Normal distribution, mean μ and variance σ^2
$\text{Unif}(a,b)$, $\text{Unif}(A)$	Uniform distribution from a to b or over A, respectively
$\text{beta}(\alpha, \beta)$	Beta distribution with two parameters
$\text{Gamma}(k, \theta)$	Gamma distribution with *shape* parameter k and *scale* parameter θ
ROI	Region of interest
MCMC	Markov chain Monte Carlo
MAP	Maximum *a posteriori*
TIRF	Total internal reflection fluorescence (microscopy)
PSF	Point-spread function
ADU	Arbitrary digital unit
(EM)CCD	(Electron multiplying) charge-coupled device
FWHM	Full width (at) half maximum
SIM	Structured Illumination Microscopy (or Mode)

Table A.1

Value	Definition
(i), (k)	Index of main and sub- sampling schemes, respectively
M	Number of emitters in the ROI
$X_{1:M}$	Locations of all M emitters in the ROI ($x_m \in \mathbb{R}^n$)
$G_{1:N}$	An arbitrarily-fine grid ($G_n \in$ ROI) with N elements.
$F(\cdot)$	Fluorescence profile
$F_{1:M}$	Value of fluorescence profile at locations where emitters exist
$F_{1:N}$	Value of fluorescence profile at grid-point locations
H, $\Delta(\cdot)$	Fluorescence baseline and fluorescence deviation profile ($F(\cdot) \triangleq H + \Delta(\cdot)$)
E_p	Intensity of photon detection events
L	Number of loads
$Y_{1:L}$	Load locations
$\hat{Y}$	Active load(s)
$\overline{Y}$	Inactive load(s)
$S_{1:L}$	Loads
$Q_{1:L}$	Load prior parameter
g	Camera gain
f	Camera excess noise factor
B_p	Background noise constant for pixel p
T	Number of (collection-intervals) exposures
P	Number of camera pixels
w_p^t	Intensity-reading in pixel p for collection-interval t
W	All readings for all pixels
$\varphi(\cdot)$	FWHM of (Gaussian) PSF
n_{cs}, n_{s}	Refractive indexes of coverslip and sample respectively
λ_0, λ_{em}	Vacuum wavelength of illumination and emission light respectively
$K(\cdot,\cdot)$	Gaussian process covariance function
$K_{A,B}$	$A \times B$ covariance matrix resulting in evaluating input set A against set B
Δt_e	Camera exposure time
σ_{PSF}	FWHM of the Gaussian PSF

Table A.2

APPENDIX B

SUPPLEMENTARY RESULTS

Table B.1 contains the default values of various modeling and computational parameters used in our implementation. Figure B.1 shows additional results obtained for inference of $X_{1:M}$, F_{xy}, and g together (discussed in Sec. 3.4). Here, some of the conditions under which the sampler performs well (unambiguous emitter counts) begin to break down, consequentially affecting the precision and accuracy of the results.

We show the precision of localization for a small ROI with one emitter for both known fluorescence and gain (Fig. B.2) and unknown fluorescence and gain (Fig. B.3) with data given in Fig. B.3a. When the other parameters are known, localization is extremely precise (within 2nm) in xy and in z, aided by the additional information we have from the exact fluorescence intensity. With an unknown fluorescence profile and gain, we lose our ability to localize the fluorophore precisely in z, and it is shown in Fig. B.3d how our sampler has converged to a case in which the emitter is located at the far end of the z boundary with a higher F_{xy} fluorescence profile. However, we do not seem to lose precision in xy.

B.1   Chapter Figures and Tables

Table B.1

Set (default) values of modeling and computational parameters.

Fig. B.1

Results of sampling $X_{1:M}$, F_{xy}, and g altogether. Emitters are placed on the line $x = y$ and spaced 250nm apart from each other in xy, and depths alternate between $z = 0$nm and $z = 10$nm. Burn-in is considered the first 20% of the 250 main loop samples. (a) Histogram for $M^{(i)}$. (b) Plot of pooled xyz localizations, for every $X_{1:M^{(i)}}^{(i)}$ and $Z_{1:M^{(i)}}^{(i)}$ such that $M^{(i)} = 7$ (the mode). Also plotted is the MAP estimate of the locations, and the resulting k-means clusters with $k = 7$. (c) Standard error of the $F_{xy}^{(i)}$ profile samples (mean of the chain, minus the ground truth profile, over the sample standard deviation). (d) Histogram of $g^{(i)}$.

Fig. B.2

Results for the sampling scheme with one emitter in a small ROI, with known profile and gain. The actual profile is given in Fig. 3.1a. The data is represented in Fig. B.3a (a) Scatter plot of pooled xyz localizations (all samples counted one emitter). (b) Histogram of (absolute) errors in xyz, with the mean and standard deviation of errors in xyz, xy, and z.

Fig. B.3

**Results for the sampling scheme with one emitter in a small ROI, with unknown profile and gain.** (a) A maximum intensity representation of the data used. (b) Error of the mean of the function samples, over the standard deviation profile for the samples. (c) Histogram of (absolute) errors in xyz, with the mean and standard deviation of errors in xyz, xy, and z. (d) Plot of pooled xyz localizations.

Also plotted is the MAP estimate of the locations, and the resulting kmeans clusters with $k = 1$.

| Parameter	(Default) value
Δt_e	100ms
g	51
T	100
λ_{em}	560nm
λ_0	520nm
n_{cs}	1.52
n_s	1.33
θ_{inc}	70°
f	2
N (numerical aperture)	1.1
B_p	1 (all p)
k_{max} for sampling g	100
k_{max} for sampling $X_{1:M}$	10,000
k_{max} for sampling $F(\cdot)$	200
l_{GP}	300
σ^2_{GP}	2
τ^2_{GP}	0.15
N (size of grid)	676
L	Varying (always at least double the number of actual emitters)
α_q	$1/L$
β_q	$1 - 1/L$
$\sqrt{\sigma^2_x}$	10nm
$P(H)$	$N(H_0 = 20, \sqrt{\sigma^2_H} = 0.5)$
$P(g)$	$\text{Unif}(g_{min} = 45, g_{max} = 55)$
Pixel size	100nm × 100nm

Table B.1
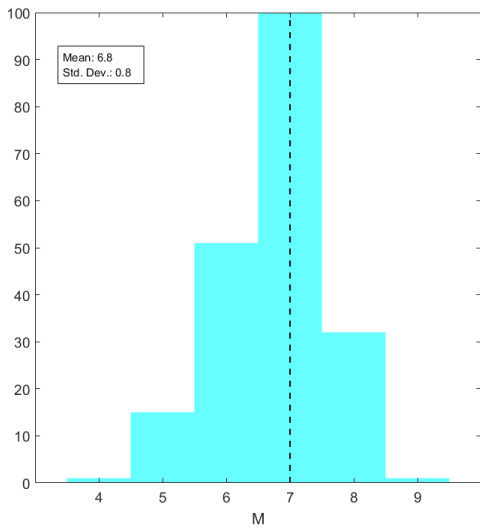
(a)
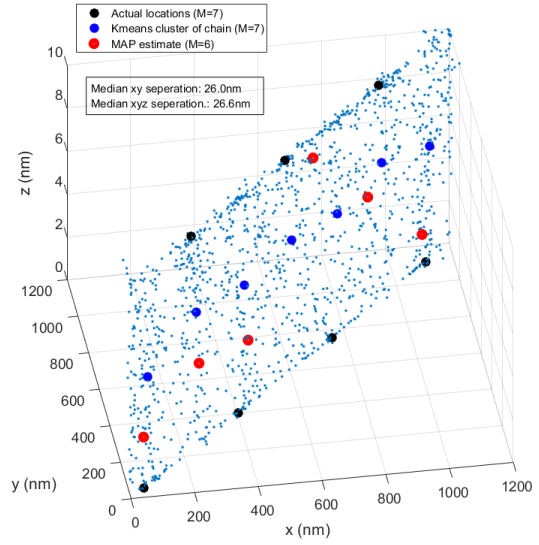
(b)
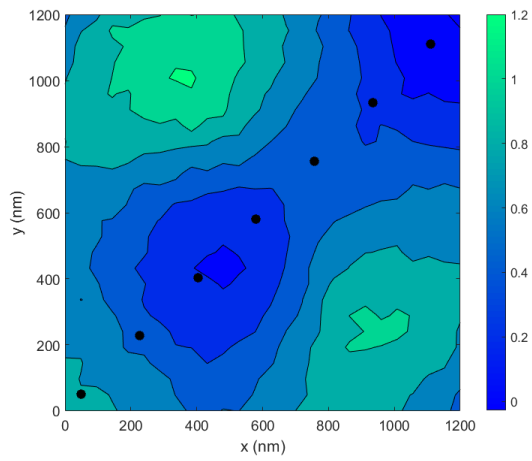
(c)
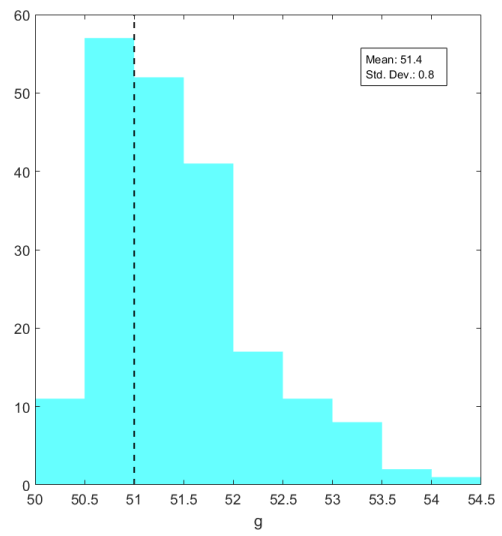
(d)

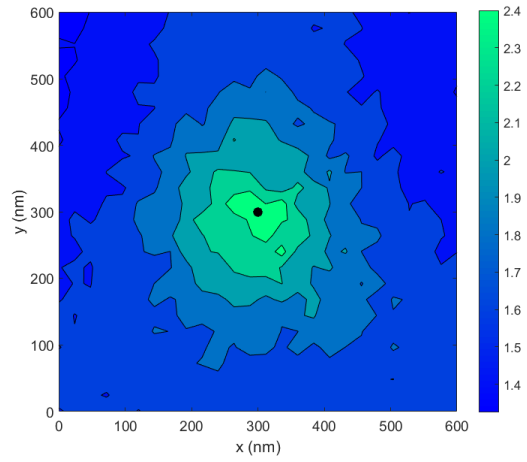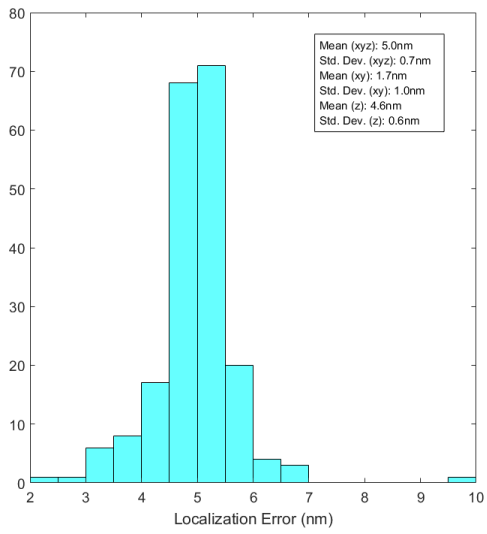**Figure B.1**
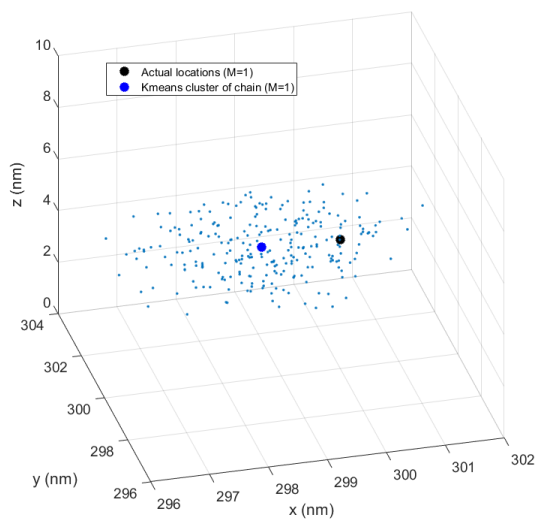
(a)
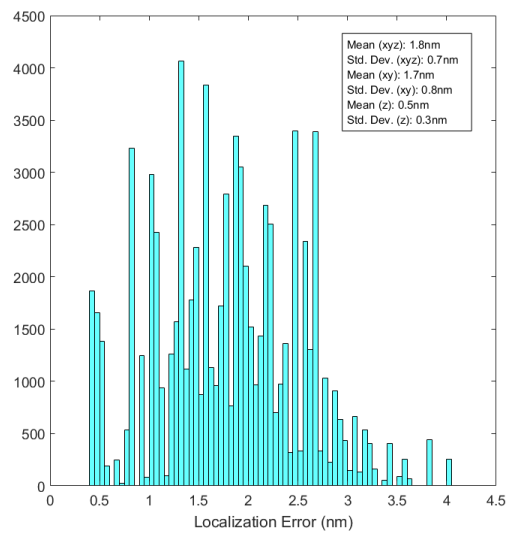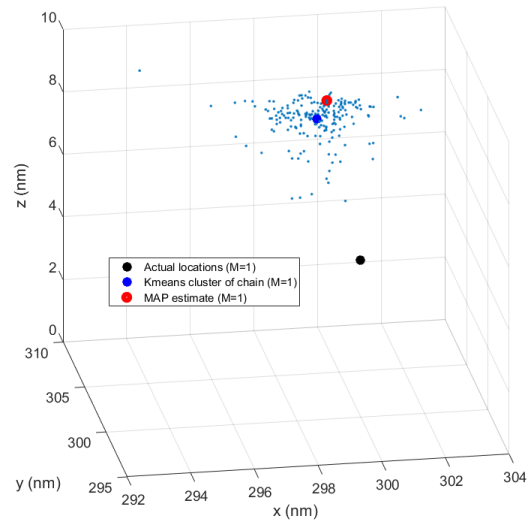

(b)

Figure B.2
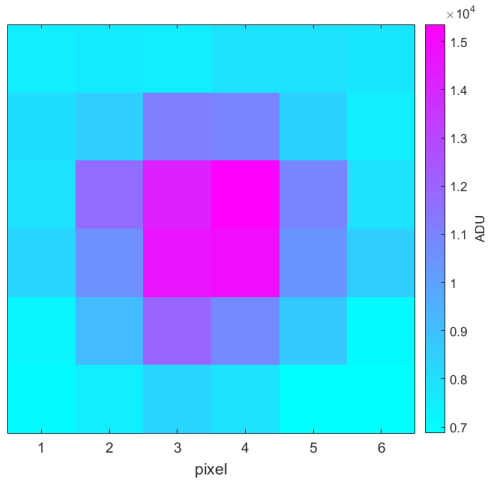
(a)

(b)

(c)

(d)

Figure B.3

APPENDIX C

NONIDENTIFYABILITY - TIRF

In this section, we discuss the possible nonidentifiability issues of our TIRF model in depth, and also briefly discuss our consequential choice in using the uniform prior on g. Figure C.1 shows four separate *mean* intensity projections that feature four distinct scenarios in which very similar results are produced. The xy coordinates for all emitters remains fixed, but we vary the F_{xy} value and refer to it in the relative sense as: *low* fluorescence ($F_{xy} \equiv 10\text{kHz}$), *medium* fluorescence ($F_{xy} \equiv 20\text{kHz}$), and *high* fluorescence ($F_{xy} \equiv 40\text{kHz}$). The evanescence profile $F_z(z)$ remains fixed, as we are not varying any of the parameters that constitute d (see Eq. 2.7). We vary the number of emitters and their z locations; the z location (depth) is referred to in relative terms as *surface* ($z = 0\text{nm}$) and *deep* ($z = 55.1\text{nm}$). A virtually infinite number of variations in emitter count, z, and F_{xy} may be constructed to produce the same or nearly same visual image.

In our implementation, the sort of nonidentifiability we see in Scenario A vs. Scenario C (count/profile nonidentifiability) is rare, and the sampler tends to become "stuck" in one of these local solutions, rather than tending to sample these distinct but visually-similar scenarios. This is mostly due to our priors: it is typically impossible for our fluorescence profile sampler to suddenly double or reduce to half of its previously sampled value, given our priors. This is also why we do not simply model F_{xy} as a Gaussian process with prior mean set to zero (i.e., let $P(F_{xy}(x, y)) = \text{GP}(0, K(\cdot, \cdot))$ and not bother with the decomposition), because the $F_{xy}(x, y)$ profile will tend to zero far away from the emitter locations, and a fluorophore under very low fluorescence may be difficult for the sampler to discern from background, leading to overfitting emitter counts.

However, the nonidentifiability we see in Scenario C vs. Scenario D, (depth/profile) is more of an issue in our implementation. The smoothness assumptions embedded in the choice for the Gaussian process covariance function and its hyperparameters are useful here because the prior prevents the sampler from creating sharp peaks/valleys in the profile to accommodate fewer/greater emitter counts (i.e., prevents profile overfitting).

There is also, in a sense, some nonidentifiability with the gain (g) and fluorescence profile. Consider the mean and variance in our camera model:

$$\mathbb{E}[\text{Gamma}(E_p/f, fg)] = E_p g \tag{C.1}$$

$$\mathbb{V}[\text{Gamma}(E_p/f, fg)] = E_p f g^2. \tag{C.2}$$

The gain will linearly scale the mean of the data, in a way that may appear similar to scaling the fluorescence profile in the same way, especially if there is low background. However, the variance of w_p^t will increase with the *square* of g, thus information about the gain is manifested in the observed variance in a way that is distinct from scaling $F(\cdot)$. Still, apparent small increases and decreases in the gain will be difficult to distinguish from small apparent changes in the fluorescence profile, and for this reason we have chosen to use our uniform prior on g, to put a "hard boundary" on the maximum/minimum values so that the sampler doesn't drift toward unreasonably high or low values, affecting our inference on the profile. If one were interested only in inferring the gain, it may be more prudent to use a different prior, such as a normal distribution. Changing this prior is trivial in our framework.

Fig. C.1

Four qualitatively separate scenarios that illustrate nonidentifiability in our model. (A) One shallow emitter under medium fluorescence. (B) One deep emitter under high fluorescence. (C) Two shallow emitters under low fluorescence. (D) Two deep emitters under medium fluorescence.
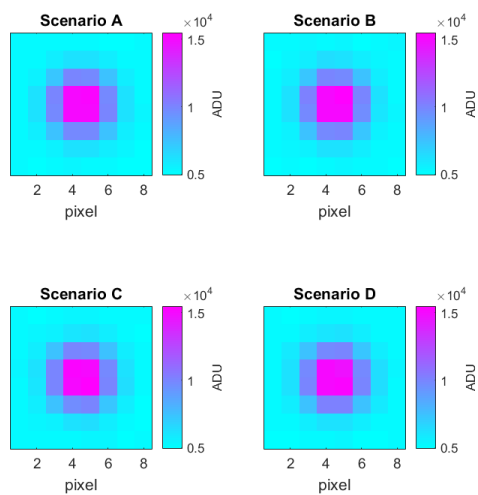
(a)

Figure C.1

APPENDIX D

MATHEMATICAL PRELIMINARIES

The following is an overview of some of the mathematical concepts used in our analysis.

D.1   MCMC, Metropolis-Hastings, and Gibbs Sampling

Markov Chain Monte Carlo (MCMC) methods are a class of algorithms used to sample from a probability distribution, and is most useful when the distribution in question is analytically intractable or too complicated to work with in simpler ways. In essence, these techniques are used to form a Markov chain of samples that randomly "walk" about a parameter space, in such a way that the chain will eventually converge to the stationary distribution of interest, i.e., that the samples will, after some time, be sampled from the stationary distribution. After the chain converges, we can analyze the samples to characterize the distribution. Samples taken during the so-called "burn-in" period, before the chain has converged, are simply discarded. It is generally not possible to know when the chain has converged, but heuristics can usually applied to help provide insight (Cowles and Carlin (1996)).

These methods tend to be, in practice, computationally expensive (Andrieu *et al.* (2003)). Large amounts of memory may also be consumed storing the samples, particularly if the chain is very long or the random variable is something significantly more complicated than a simple univariate, like large vectors, matrices or even in principal something like a high definition image or video, though thinning (Link and Eaton (2012)) can be applied and intermediary samples discarded (this also helps to mitigate potential issues with the inherently autocorrelated samples).

Implicit to the goal of sampling from the stationary distribution of a Markov process is that the stationary distribution exists and is unique. A sufficient (but not necessary) condition for the existence and uniqueness of an underlining stationary distribution is that the Markov chain preserves detailed balance. The two MCMC techniques used in our method are techniques that generate reversible Markov chains which, by definition, preserve detailed balance, namely the Metropolis-Hastings algorithm (Chib and Greenberg (1995)) and the Gibbs sampling scheme, which is actually a special (but very significant) case of Metropolis-Hastings (Gelfand (2000)).

In Metropolis-Hastings, we need to have a function that is proportional to the probability density function that we are interested in sampling. This is very convenient in a Bayesian context, because for random variable x (a posterior distribution for which we want to sample) and data D, we have that $P(x|D) \propto P(D|x)P(x)$, where $P(D|x)$ and $P(x)$ are the likelihood and prior. We propose values x (indexing these samples $x^{(i)}$) using a separate proposal distribution, which can be chosen freely, though it can only depend on the previous sample value ($x^{(i-1)}$) because of the memoryless property of the Markov process. Intelligent choices for the proposal can help reduce the inherent sample autocorrelation or help the chain converge faster.

The probability of accepting a proposed sample must only be such that detailed balance is preserved, but the most common choice (and the one used ubiquitously in

our analysis) is known as the Metropolis choice:

$$\alpha \triangleq P(x^i := x') := \min(1, p) \tag{D.1}$$

$$p = \frac{P(x'|D)G(x^{i-1}|x')}{P(x^{i-1}|D)G(x'|x^{i-1})}$$

$$\equiv \frac{P(D|x')P(x')G(x^{i-1}|x')}{P(D|x^{i-1})P(x^{i-1})G(x'|x^{i-1})}. \tag{D.2}$$

Here, x' is the proposed sample, and G is the proposal distribution; $G(x^{i-1}|x')$ should be read as "the probability of proposing x^{i-1} given x' as the current value," where $G(x'|x^{i-1})$ is the obvious analog. The statement $P(x^i := x')$ is read as "the probability of accepting x' as the new sample in the chain." Alg. 2 details the pseudocode of the Metropolis-Hastings sampler algorithm.

In Gibbs sampling, we consider sampling multiple random variables or a multivariate random variable; we propose changes to only a subset of the variables or variable indices, conditioned on the current value in the sampler (i.e., the most recently accepted sample) for the other variables not currently being sampled. Proposing from such a conditional distribution intrinsically preserves detailed balance, so samples are automatically accepted, a very powerful tool if the appropriate conditional distributions exist and can be sampled from, because it provides a theoretically guaranteed solution to virtually any Bayesian inference inverse problem (Bishop (2006)). The Gibbs sampling algorithm is summarized in the Alg. 3 pseudocode.

D.2   Gaussian Process

The Gaussian process is a stochastic process for which a finite and arbitrary collection of random variables, most typically indexed through time or space, is assumed to have a joint multivariate normal distribution. It is considered nonparametric in the sense that the collection of variables, though technically finite, can be arbitrarily large. Because variables can be indexed over continuous spaces, the Gaussian process is used as a convenient statistical model (e.g., as a prior) for continuous functions.

The Gaussian process is defined by a prior mean function and covariance function that instruct how the mean and covariance of the multivariate normal distribution are to be constructed for a set of specific input points:

$$F(X) \sim \text{GP}(m(X), K(X, X)) \tag{D.3}$$

$$X \triangleq [x_1, \ldots, x_N].$$

The prior mean function describes how the function $F(\cdot)$ will tend in the lack of information and is often, in practice, set to zero for computational simplicity. The covariance function is usually more important, describing how "similar" any two points in the input space will be (it is oftentimes referred to as the kernel function), determining the covariance of any two points based on this evaluation. Different covariance functions result in different model behaviors for the same input, and choices for covariance functions embed prior assumptions about the latent function smoothness, periodicity, and other qualities (Souza (2010)).

Often, covariance functions depend on the temporal or spatial (or spaciotemporal) distance between input points. Valid covariance functions create symmetric positive definite matrices, and sums and products of valid covariance functions are also valid covariance functions, allowing simpler, more generic and widely used covariance functions to be combined to achieve specific behavior. Covariance functions may also depend on some parameters, which we refer to as the Gaussian process hyperparameters; they may be chosen, inferred using Markov chain Monte Carlo techniques (Titsias *et al.* (2008)), or set using optimization techniques like gradient descent.

One can use the Gaussian process to perform nonlinear regression of continuous functions. Let us say that we have observed a function $F(\cdot)$ at a set of observation points in input space, X_{obs}, and let $F_{\mathrm{obs}} \triangleq F(X_{\mathrm{obs}})$. In order to infer the values of $F(\cdot)$ at an entirely different set of input points, which we call X_{grid} (letting $F_{\mathrm{grid}} \triangleq F(X_{\mathrm{grid}})$), we begin by stating the Gaussian process model, for some arbitrary $m(\cdot)$ and $K(\cdot, \cdot)$:

$$\begin{bmatrix} F_{\mathrm{obs}} \\ F_{\mathrm{grid}} \end{bmatrix} \sim \mathrm{N}\left(m\left(\begin{bmatrix} X_{\mathrm{obs}} \\ X_{\mathrm{grid}} \end{bmatrix} \right), \begin{bmatrix} K_{\mathrm{obs}} & K_{\mathrm{obs,grid}} \\ K_{\mathrm{obs,grid}}^T & K_{\mathrm{grid}} \end{bmatrix} \right).$$

Here, the $K_{A,B}$ matrices ($K_A \equiv K_{A,A}$) are constructed by evaluating input points A against B using the covariance function $K(\cdot, \cdot)$, i.e.,:

$$K_{A,B} \triangleq K(A, B) \equiv \begin{bmatrix} K(a_1, b_1) & K(a_1, b_2) & \ldots & K(a_1, b_N) \\ K(a_2, b_1) & K(a_2, b_2) & \ldots & K(a_2, b_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(a_M, b_1) & K(a_M, b_2) & \ldots & K(a_M, b_N) \end{bmatrix} \tag{D.4}$$

$$|A|_{\mathrm{size}} = M, \quad |B|_{\mathrm{size}} = N, \quad |K_{A,B}|_{\mathrm{size}} = M \times N.$$

Since we are interested in inferring F_{grid}, and all of the other variables are fixed and known in this context, the posterior probability distribution that we are interested in constructing is of the form F_{grid} conditioned on F_{obs}, X_{obs} and X_{grid}. The derivation is non-trivial, but the standard results is that this conditional normal distribution is itself normal:

$$P(F_{\mathrm{grid}}|F_{\mathrm{obs}}, X_{\mathrm{obs}}, X_{\mathrm{grid}}) =$$
$$\mathrm{N}(m(X_{\mathrm{grid}}) + (K_{\mathrm{obs,grid}}^T / K_{\mathrm{obs}})(F_{\mathrm{obs}} - m(X_{\mathrm{obs}})), F_{\mathrm{grid}} - (K_{\mathrm{obs,grid}}^T / K_{\mathrm{obs}})K_{\mathrm{obs,grid}}) \tag{D.5}$$

Letting $m(\cdot) \equiv 0$ significantly simplifies this expression.

The Gaussian process is a beautiful theoretical tool, but presents some difficulty in practice, as the construction of and operation on the covariance matrices may be computationally expensive and numerically unstable. However, there are quite a few tricks and approximations to mitigate these challenges, and if they can be overcome, then the Guassian process offers a robust framework for inference on functions in a Bayesian context. See (Rasmussen (2004); Williams and Rasmussen (2006); Wilson and Adams (2013)) for further details.

D.3 Beta-Bernoulli Process

The beta-Bernoulli process is a Bayesian nonparametric technique used to count a finite number of elements in a random vector (MacEachern (2016)). We introduce a random vector $S_{1:N}$ with N elements that we call "loads," which are Bernoulli random variables, and so take values of either 0 or 1, interpreted respectively as (by convention) *inactive* and *active*. We choose a prior on the loads: $P(S_i = 1) := Q_i$, and treat the random vector $Q_{1:N}$, called the "load priors," as beta-distributed random variables, such that prior $P(Q_i) := \text{beta}(\alpha_q, \beta_q)$. The parameters α_q and β_q are the load prior hyperparameters, the classic choices for which are $\alpha_q := 1/N$ and $\beta_q := 1 - 1/N$.

We can sample the values of $Q_{1:N}$ and $S_{1:N}$ in a Gibbs sampling scheme (see App. D.1), using the appropriate conditional distributions:

$$P(S_i = 1|Q_i) \equiv Q_i \tag{D.6}$$

$$P(Q_i|S_i) \propto \text{beta}\big(\alpha_q + S_i, \beta_q + (1 - S_i)\big). \tag{D.7}$$

In practice, it may be useful to introduce a third random vector $Y_{1:N}$, the elements of which are active or inactive, depending on the value of the corresponding loads (i.e., $S_{1:N}$ are loads over $Y_{1:N}$). For instance, we can count and localize the number of objects inside an ROI by defining $Y_{1:N}$ as the "load locations." Implicitly, the locations of all the objects in the ROI are the locations Y_i such that $S_i = 1$. The values for $Y_{1:N}$ may be sampled in the same Gibbs scheme in which we sample values for $S_{1:N}$ and $Q_{1:N}$, and one should note that in a Bayesian context, the likelihood will typically not depend on the values of Y_i where $S_i = 0$.

It is important to mind the fact that, due to memory and time constraints, one cannot use an infinite number of loads, and care must be taken to choose a number of loads that exceeds the likely number of on loads that we might expect to see. However, overshooting the number of loads may have its own associated cost in computational efficiency. Figure E.2 is a graphical representation showing all of the steps in the inference of $Q_{1:L}$, $S_{1:L}$, and $Y_{1:L}$. See also: (Jazani *et al.* (2019); Broderick *et al.* (2012)).

D.4 Chapter Algorithms

Alg. 2

Metropolis-Hastings sampler algorithm. For numerical stability, one may want to evaluate the natural logarithm of p, in which case we sample r from an exponential distribution such that $r \sim -\exp(1)$.

Alg. 3

Gibbs sampler algorithm.

Algorithm 2 Metropolis-Hastings Sampler

1: $x^0 = \text{InitialGuess}()$
2: for i:
3:     $x' \sim G(x; x^i | x^{i-1})$
4:     $p = \frac{P(D|x')P(x')G(x^{i-1}|x')}{P(D|x^{i-1})P(x^{i-1})G(x'|x^{i-1})}$
5:     if $(p \geq 1)$
6:         $x^i = x'$
7:     else
8:         $r \sim \text{Unif}(0, 1)$
9:         if $(p > r)$
10:             $x^i = x'$
11:         else
12:             $x^i = x^{i-1}$
13: end

Algorithm 3 Gibbs Sampler

1: $[x_1^0, x_2^0, \ldots, x_M^0] = \text{InitialGuess}()$
2: for i:
3:     for m:
4:         $x_1^i \sim P(x_1|x_2^{i-1}, \ldots, x_M^{i-1})$
5:         $\vdots$
6:         $x_m^i \sim P(x_m|x_1^i, \ldots, x_{m-1}^i, x_{m+1}^{i-1}, \ldots, x_M^{i-1})$
7:         $\vdots$
8:         $x_M^i \sim P(x_M|x_1^i, \ldots, x_{M-1}^i)$
9:     end
10: end

APPENDIX E

FURTHER MATHEMATICAL DETAILS OF THE METHOD

In this section, we provide further details on the individual parameter inference frameworks (i.e., to sample the distributions described in Eqs. 2.14, 2.13, and 2.15). The information in the main text is summarized and elaborated further. It will be noted in various sections what random variables are fixed to known values, and will likely be omitted from the notation in that context.

E.1 Sampling the Gain

We seek to sample a value for $g^{(i)}$ (Eq. 2.15). In this context, $X_{1:M}$ and $F(\cdot)$ are fixed, with known values and in the context of the full sampler, these are the samples obtained most recently in the steps described by Eq. 2.13 and Eq. 2.14.

Using Bayes' rule:

$$g^{(i)} \sim P(W|g)P(g). \tag{E.1}$$

We have chosen a uniform prior on g, with left and right parameters $g_{\min}$ and $g_{\max}$. Unreasonable large or small values are precluded entirely with this prior, rather than allowing improbable values to overfit for $X_{1:M}$ and $F(\cdot)$. The likelihood is given in Eq. 2.11, and it is not conjugate to the prior we have chosen, nor can it be sampled directly, even if it were to be. To draw the sample $g^{(i)}$, we use a Metropolis-Hastings algorithm, which we index by (k). We may draw many $g^{(k)}$, and ultimately set $g^{(i)}$ to the final sample in this chain. With the Metropolis-choice acceptance ratio (see App. D.1), we choose to use the prior as our proposal distribution:

$$g' :\sim \text{Unif}(g_{\min}, g_{\max}) \tag{E.2}$$

$$\alpha := \min(1, p) \tag{E.3}$$

$$p = \frac{P(W|g')P(g')}{P(W|g^{(k-1)})P(g^{(k-1)})} \cdot \frac{P(g^{(k-1)})}{P(g')}$$

$$= \prod_{t=1}^{T} \prod_{p=1}^{P} \left(g^{(k-1)}/g' \right)^{E_p/f} \exp\left[(w_p^t/f)(1/g^{(k-1)} - 1/g') \right] \tag{E.4}$$

$$\ln(p) = (1/f)\left[T\left[ln(g^{(k-1)}) - ln(g') \right] \left[\sum_{p=1}^{P} E_p \right] + (1/g^{(k-1)} - 1/g')\left[\sum_{t=1}^{T} \sum_{p=1}^{P} w_p^t \right] \right]. \tag{E.5}$$

Note that E_p does not depend on g.

This completes our sampling scheme for g. A summary of all variables and relevant sampling information is summarized in Table E.4.

E.2 Sampling the Fluorescence Profile

In seeking a method to sample $F(\cdot)^{(i)}$ from the distribution described in Eq. 2.14 (where $X_{1:M}$ and g are, in the context of the full sampler, the most recent samples from Eq. 2.13 and Eq. 2.15), we have decomposed the fluorescence profile into the form $F(\cdot) \equiv H + \Delta(\cdot)$, where H is the baseline fluorescence and $\Delta(\cdot)$ is the zero-mean fluorescence deviation profile. We have chosen a Gaussian process prior for $F(\cdot)$, which has the same covariance properties as $\Delta(\cdot)$, so that $P(F(\cdot)) \equiv P(\Delta(\cdot))$.

We cannot, in practice, sample a continuous function, because of the discrete nature of computers. So we will seek to sample values $F_{1:N} \triangleq F(G_{1:N})$, where $G_{1:N}$ is a fine grid over the ROI, and approximate a function by interpolating between these grid-points. A fine enough grid and interpolation will be nearly indistinguishable from a not-interpolated function. In our implementation, we use a linear interpolation, because it is computationally the most efficient, but or large ROIs or grids that are not so fine, one may want to consider a smoother interpolation that will work well with the smoothness assumptions that one has for the function (embedded mathematically by the choice in the Gaussian process prior mean and covariance functions). We are provided by the Gaussian process model a method to sample such values (see App. D.2):

$$P(F_{1:N}|F_{1:M}, X_{1:M}, G_{1:N}) = H + \mathrm{N}\big((K_{X,G}^T/K_X)\Delta_{1:M}, K_G - (K_{X,G}^T/K_X)K_{X,G}\big). \tag{E.6}$$

Here, $K_{A,B}$ (with $K_A \equiv K_{A,A}$) is a covariance matrices constructed by evaluating input points A against input B, using the chosen covariance function, and $F_{1:M} \triangleq F(X_{1:M}) \equiv H + \Delta_{1:M}$ is the value of the profile at locations where the M emitters exist. The likelihood will depend on these values (and we will elaborate on how to sample these values later).

We have defined our covariance function (kernel) as the sum of a squared exponential kernel and nugget regularizer (Eqs. 2.19-2.22). The squared exponential kernel $\mathrm{SE}(\cdot, \cdot)$ is translation invariant, depending on the (Euclidean) distance between input points, along with two hyperparamters: σ_{GP}^2 and l_{GP}. Parameter σ_{GP}^2 is a scaling prefactor that, roughly, describes the freedom that the posterior function will have to deviate from the prior mean function, and l_{GP}, known as the "length scale," describes the distance over which significant changes to the function may occur, i.e., the smoothness of its variations. This kernel is quite standard, and because only two parameters need to be specified, it is flexible and intuitive (see Fig. E.1 for an illustration of this kernel).

The nugget regularizer (Mohammadi *et al.* (2016)), involving the hyperparameter τ^2, adds a constant to the variance of the fluorescence profile value, and we have chosen to do this for two reasons: (1) to reduce column redundancy, preserving computational stability when working with nearly-singular, ill-conditioned covariance matrices (that are sadly common in practice) (2) to produce more varied samples of $F(\cdot)$.

Note that we have now introduced three new computational parameters to our framework. In practice, one may wish to sample values for these hyperparameters along with our other random variables, but in our implementation these are values that we specify.

There is now the matter of inferring values $F_{1:M}$, which our likelihood will depend on. We have chosen to use a Gibbs sampling scheme (App. D.1), where we iteratively sample H and individual $\Delta_m \in \Delta_{1:M}$, conditioned on the others (Eq. 2.27 and Eq. 2.28). We index these samples by (k), and may draw many $H^{(k)}$ and $\Delta_{1:M}^{(k)}$ before setting $H^{(i)}$ and $\Delta_{1:M}^{(k)}$ to the final sample in this chain.

Consider Eq. 2.27; in this context, $\Delta_{1:M}$ is fixed. Using Bayes' rule:

$$H^{(k)} \sim P(W|H, \Delta_{1:M}^{(k-1)})P(H). \tag{E.7}$$

With the Metropolis-choice acceptance ratio (see App. D.1), we choose to use the prior (Eq. 2.23) as our proposal distribution:

$$H' :\sim N(H_0, \sigma_H^2) \tag{E.8}$$

$$\alpha := \min(1, p) \tag{E.9}$$

$$p = \frac{P(W|H', \Delta_{1:M}^{(k-1)})P(H')}{P(W|H^{(k-1)}, \Delta_{1:M}^{(k-1)})P(H^{(k-1)})} \cdot \frac{P(H^{(k-1)})}{P(H')}$$
$$= \prod_{t=1}^{T}\prod_{p=1}^{P} \frac{\Gamma\big((E_p|_{H^{(k-1)}})/f\big)}{\Gamma\big((E_p|_{H'})/f\big)} \left[\frac{w_p^t}{fg}\right]^{(E_p|_{H'} - E_p|_{H^{(k-1)}})/f} \tag{E.10}$$

$$\ln(p) = \sum_{p=1}^{P} T\big[\ln\Gamma(E_p|_{H^{(k-1)}}/f) - \ln\Gamma(E_p|_{H'}/f)\big]$$
$$+ \big[(E_p|_{H'} - E_p|_{H^{(k-1)}})/f\big]\big[-T\ln(fg) + \sum_{t=1}^{T}\ln w_p^t\big]. \tag{E.11}$$

Now, consider Eq. 2.28. We again use the Metropolis-choice acceptance ratio, and for the proposal we use the conditional distribution of each Δ_m that is obtained by virtue of the Gaussian process prior:

$$\Delta_m' :\sim \mathrm{GP}(\Delta_m|\Delta_\star) \equiv P(\Delta_m|\Delta_\star, X_{1:M}) \tag{E.12}$$

$$\mathrm{GP}(\Delta_m|\Delta_\star) \triangleq N((K_{\star,m}^T/K_\star)\Delta_\star, K_m - (K_{\star,m}^T/K_\star)K_{\star,m}) \tag{E.13}$$

$$\alpha := \min(1, p) \tag{E.14}$$

$$p = \frac{P(W|\Delta_m', \Delta_\star, H^{(k)})\mathrm{GP}(\Delta_m'|\Delta_\star)}{P(W|\Delta_m^{(k-1)}, \Delta_\star, H^{(k)})\mathrm{GP}(\Delta_m^{(k-1)}|\Delta_\star)} \cdot \frac{\mathrm{GP}(\Delta_m^{(k-1)}|\Delta_\star)}{\mathrm{GP}(\Delta_m'|\Delta_\star)}$$
$$= \prod_{t=1}^{T}\prod_{p=1}^{P} \frac{\Gamma\big((E_p|_{\Delta_m^{(k-1)}})/f\big)}{\Gamma\big((E_p|_{\Delta_m'})/f\big)} \left[\frac{w_p^t}{fg}\right]^{(E_p|_{\Delta_m'} - E_p|_{\Delta_m^{(k-1)}})/f} \tag{E.15}$$

$$\ln(p) = \sum_{p=1}^{P} T\big[\ln\Gamma(E_p|_{\Delta_m^{(k-1)}}/f) - \ln\Gamma(E_p|_{\Delta_m'}/f)\big]$$
$$+ \big[(E_p|_{\Delta_m'} - E_p|_{\Delta_m^{(k-1)}})/f\big]\big[-T\ln(fg) + \sum_{t=1}^{T}\ln w_p^t\big]. \tag{E.16}$$

Here, $\Delta_\star$ refers to the most recent sample of each element in $\star \equiv 1 : M\backslash m$ (all indexes 1 through M except m) so that, strictly speaking, the index of each element may be either (k) or $(k-1)$. Note that $X_{1:M}$ is implicitly required to evaluate $\mathrm{GP}(\Delta_m|\Delta_\star)$. We have also introduced the notation $\ln\Gamma(\cdot)$, which denotes the natural logarithm of the gamma-function.

Note that $\lim_{x\to 0}\Gamma(x)$ is undefined, so we will have computational instability as $E_p \to 0$. In practice however, this should not be troublesome, because background noise is ubiquitous for virtually any fluorescence microscope system and it should

never be the case that we expect zero photon arrivals in a realistic collection-interval. It's also worth noting that the covariance matrices K depend only on the elements $X_{1:M}$ and associated hyperparameters, and because these values do not change at all during this nested sampling scheme, we only need to evaluate the $M \times M$ covariance matrix once and draw from it the appropriate rows and columns to carry out computations. We can even store every matrix $(K_{m,\star}/K_{\star})$ and $K_m - (K_{m,\star}/K_{\star})K_{m,\star}^T$ in memory, which makes most of the computational expense up-front, allowing a large number of samples to be taken rather quickly.

This completes our sampling scheme for $F(\cdot)$. A summary of all variables and relevant sampling information is summarized in Table E.4.

E.3   Sampling Emitter Locations

We seek a method for sampling $X_{1:M}^{(i)}$ from the distribution described in Eq. 2.13. In this context, $F(\cdot)$ and g are fixed with known values (in the full sampler, the most recent samples from Eq. 2.14 and Eq. 2.15). Previously, we have introduced three random vectors in a beta-Bernoulli scheme (see App. D.3): loads $S_{1:L}$, load locations $Y_{1:L}$, and load priors $Q_{1:L}$. An active load ($S_i = 1$) implies the existence of an emitter at the corresponding load location (Y_i). These variables are described mathematically in Eqs. 2.29-2.32. We choose to sample these random vectors altogether in a Gibbs sampling scheme, detailed in Eqs. 2.33-2.35. Figure E.2 features a graphic of the sampling method.

We need a strategy to sample each equation in the Gibbs scheme. First, consider Eq. 2.33. It is easy to show that the distribution has a closed form, and for $Q_j \in Q_{1:L}$:

$$Q_j^{(k)} \sim \text{beta}(\alpha_q + S_j^{(k-1)}, \beta_q + 1 - S_j^{(k-1)}). \tag{E.17}$$

Now, consider Eq. 2.34. With Bayes' rule:

$$S_{1:L}^{(k)} \sim P(W|S_{1:L}, Q_{1:L}^{(k)}, Y_{1:L}^{(k-1)})P(S_{1:L}). \tag{E.18}$$

We choose to sample values for the random vector in randomized blocks that we periodically mix-up, in order to achieve good mixing and computational efficiency.

For load block $S_{B_i}^{(k)}$:

$$S'_{B_i} :\sim \text{Bern}(Q_{B_i}^{(k)}) \equiv P(S_{B_i}|Q_{1:L}^{(k)}) \tag{E.19}$$

$$\alpha := \min(1, p) \tag{E.20}$$

$$p = \frac{P(W|Q_{1:L}^{(k)}, S'_{B_i}, S_\star, Y_{1:L}^{(k-1)})\text{Bern}(S'_{B_i}|Q_{1:L}^{(k)})}{P(W|Q_{1:L}^{(k)}, S_{B_i}^{(k-1)}, S_\star, Y_{1:L}^{(k-1)})\text{Bern}(S_{B_i}^{(k-1)}|Q_{1:L}^{(k)})} \cdot \frac{\text{Bern}(S_{B_i}^{(k-1)}|Q_{1:L}^{(k)})}{\text{Bern}(S'_{B_i}|Q_{1:L}^{(k)})}$$

$$= \prod_{t=1}^{T}\prod_{p=1}^{P} \frac{\Gamma\big((E_p|_{S_{B_i}^{(k-1)}})/f\big)}{\Gamma\big((E_p|_{S'_{B_i}})/f\big)} \left[\frac{w_p^t}{fg}\right]^{(E_p|_{S'_{B_i}} - E_p|_{S_{B_i}^{(k-1)}})/f} \tag{E.21}$$

$$\ln(p) = \sum_{p=1}^{P} T\big[\ln\Gamma(E_p|_{S_{B_i}^{(k-1)}}/f) - \ln\Gamma(E_p|_{S'_{B_i}}/f)\big]$$

$$+ \big[(E_p|_{S'_{B_i}} - E_p|_{S_{B_i}^{(k-1)}})/f\big]\big[-T\ln(fg) + \sum_{t=1}^{T}\ln w_p^t\big]. \tag{E.22}$$

Here, $S_\star$ refers to the current sample of each element in $\star \equiv 1 : L\backslash B_i$ (all indexes 1 through L except those in block B_i) so that, strictly speaking, the index of each element may be either (k) or $(k-1)$. We have abused notation slightly: Eq. E.19 features a random vector sampled from a univariate distribution with a random vector in its argument, which is meant to be interpreted as meaning "the individual elements of this random vector are sampled from a distribution for which the corresponding element in the argument vector is the sampling distribution parameter."

Finally, we consider Eq. 2.35. Using Bayes' rule:

$$Y_{1:L}^{(k)} \sim P(W|Y_{1:L}, Q_{1:L}^{(k)}, S_{1:L}^{(k)})P(Y_{1:L}). \tag{E.23}$$

Recall that we have separated $Y_{1:L}$ into two subsets: those with active loads (i.e., $S_j^{(k)} = 1$), which we denote $\hat{Y}_j \in \hat{Y}_{1:M^{(k)}}$, and those with inactive loads ($S_j^{(k)} = 0$), which we denote $\overline{Y}_j \in \overline{Y}_{1:(L-M^{(k)})}$, described respectively in Eq. 2.37 and Eq. 2.38. We will sample new locations for each $\hat{Y}_j$ and $\overline{Y}_j$ individually, but treat sampling active and inactive load locations differently, because the likelihood will depend on the active load locations (locations of emitters), but not on the inactive load locations (which are transient and computational in nature).

Consider now an active load location $\hat{Y}_j$. We will draw a new sample for the location in a Metropolis-Hastings step, which we will not formally index; it is valid to propose and accept/reject multiple samples, and set $\hat{Y}_j^{(k)}$ to the final sample in this scheme, but in our implementation we only make a single new proposal for each $\hat{Y}_j$. The proposal distribution that we choose is isotropic normal centered at the current location (the most recently accepted sample, variance σ_x^2). It is unwise to sample a new location from the prior (as we have done in most of our other Metropolis-Hastings sampling steps), because it is reasonable to suspect that an active load will be an actual emitter location or close to it, and also do not wish to bias any one

direction, so the isotropic normal seems to be a reasonable choice. What we have is otherwise much like our previous Metropolis-Hastings schemes:

$$\hat{Y}_j' :\sim \mathrm{N}(\hat{Y}_j^{(k-1)}, \sigma_x^2) \tag{E.24}$$

$$\alpha := \min(1, p) \tag{E.25}$$

$$p = \frac{P(W|\hat{Y}_j', Y_\star, Q_{1:L}^{(k)}, S_{1:L}^{(k)})\mathrm{Unif(ROI)}}{P(W|\hat{Y}_j^{(k-1)}, Y_\star, Q_{1:L}^{(k)}, S_{1:L}^{(k)})\mathrm{Unif(ROI)}} \cdot \frac{\mathrm{N}(\hat{Y}_j'|\hat{Y}_j^{(k-1)})}{\mathrm{N}(\hat{Y}_j^{(k-1)}|\hat{Y}_j')}$$

$$= \prod_{t=1}^{T}\prod_{p=1}^{P} \frac{\Gamma\big((E_p|_{\hat{Y}_j^{(k-1)}})/f\big)}{\Gamma\big((E_p|_{\hat{Y}_j'})/f\big)} \left[\frac{w_p^t}{fg}\right]^{(E_p|_{\hat{Y}_j'} - E_p|_{\hat{Y}_j^{(k-1)}})/f} \tag{E.26}$$

$$\ln(p) = \sum_{p=1}^{P} T\big[\ln\Gamma(E_p|_{\hat{Y}_j^{(k-1)}}/f) - \ln\Gamma(E_p|_{\hat{Y}_j'}/f)\big]$$

$$+ \big[(E_p|_{\hat{Y}_j'} - E_p|_{\hat{Y}_j^{(k-1)}})/f\big]\big[-T\ln(fg) + \sum_{t=1}^{T}\ln w_p^t\big]. \tag{E.27}$$

The symmetric proposal distribution cancels in the ratio for p. The prior cancels as well, assuming that samples for $\hat{Y}_j$ are restricted to the ROI (i.e., $P(Y_j) \neq 0$; a reflective boundary when sampling new locations works well in our implementation to achieve this restriction).

Now, we consider an inactive location $\overline{Y}_j$. The likelihood does not depend on inactive locations, so we are left with only the prior in Eq. E.23; we will sample inactive load locations at random from it, effectively scrambling them at this step in the sampling scheme, thereby achieving a high level of mixing:

$$\overline{Y}_j^{(k)} :\sim \mathrm{Unif(ROI)}. \tag{E.28}$$

Ultimately, after we have finished the overall Gibbs sampling scheme for $Q_{1:L}$, $S_{1:L}$, and $Y_{1:L}$, we set $X_{1:M}^{(i)}$ to the set of active load locations (Eq. 2.32). In our implementation, we then discard the values of $Q_{1:L}$, $S_{1:L}$, and $Y_{1:L}$, reinitializing these variables at function call for the subsampler (i.e., we start over with some initial values every time we sample from Eq. 2.13), preventing localization from becoming stuck in a local minimum and not moving.

This completes our sampling scheme for $X_{1:M}$. A summary of all variables and relevant sampling information is found in Table E.4.

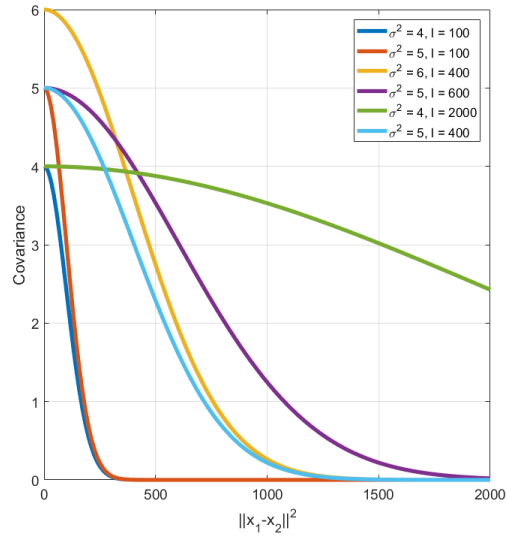### E.4 Chapter Figures and Tables

Fig. E.1

The squared exponential kernel, evaluated with various values of the hyperparameters $\sigma_{\mathbf{GP}}^2$ and $l_{\mathbf{GP}}$. $\mathrm{SE}(x_1, x_2) = \sigma^2 \exp\big(\frac{\|x_1 - x_2\|^2}{2l^2}\big)$.

Fig. E.2

A graphical representation of our sampling scheme for $X_{1:M}$.

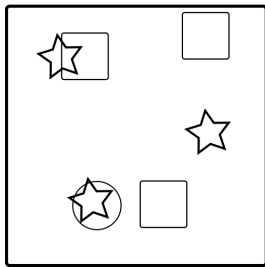A table of our method random variables, their priors, and the proposal or sampling distribution.
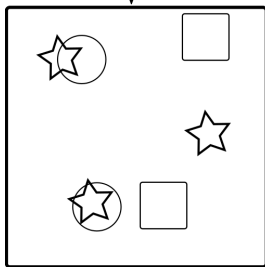

(a)

Figure E.1

One iteration of the beta-Bernoulli process sampling scheme for emitter locations. The bold box represents the ROI, and the shapes represent the positions of:
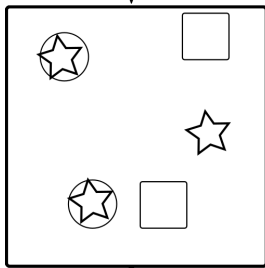
- Actual (latent) emitter locations

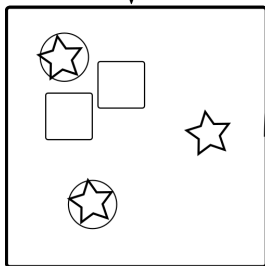- Active loads

- Inactive loads

A) The starting locations of actual emitters (stars), active loads (circles), and inactive loads (squares) in the ROI.

B) New load values (active/inactive) are sampled and accepted/rejected with a Metropolis-Hastings step. We typically accept/reject changes in blocks, to improve mixing. In our illustration, a previously inactive load in the upper left quadrant was changed to an active load.

C) New locations for active loads are sampled and accepted/rejected with a Metropolis-Hastings step. We typically accept/reject individual location changes. In our illustration, both of the active loads have moved closer to the actual (latent) location of the respective nearby emitters.

D) New locations for inactive loads are sampled from the prior, which is uniform over the ROI. In our illustration, both of the inactive loads have moved to two new, random locations.

Figure E.2

Variable	Definition	(Conditional) Prior	Proposal/Sampling distribution
Q_j	$P(S_j = 1)$	beta(α_q, β_q)	$Q_j^{(k)} \sim$ beta$(\alpha_q + S_j, \beta_q + 1 - S_j)$
S_j	Load on Y_j	Q_j	$S_j' \sim P(S_j^{(k)}|Q_j^{(k)})$
Y_j	Load location	Unif(ROI)	$Y_{1:L}^{(i)} \equiv \overline{Y}_{1:(L-M^{(i)})}^{(i)} \cup \hat{Y}_{1:M^{(i)}}^{(i)}$
$\hat{Y}_j$	Active load location, $S_j = 1$	...	$\hat{Y}_j' \sim N(\hat{Y}_j^{(k-1)}, \sigma_x^2)$
$\overline{Y}_j$	Inactive load location, $S_j = 0$	...	$\overline{Y}_j^{(k)} \sim P(\overline{Y}_j)$
$X_{1:M}$	Emitter locations	...	$X_{1:M^{(i)}}^{(i)} \equiv \{\bigcup_{j=1}^{L} Y_j^{(i)} \,|\, S_j^{(i)} = 1\}$
M	Emitter count	...	$M^{(i)} \equiv \sum_{j=1}^{L} S_j^{(i)}$
H	Fluorescence baseline	$N(H_0, \sigma_H^2)$	$H' \sim P(H)$
Δ_m	$F(x_m) - H$	GP$(\Delta_m|\Delta_{1:M\setminus m}, X_{1:M})$	$\Delta_m' \sim P(\Delta_m|\Delta_{1:M\setminus m}, X_{1:M^{(i)}})$
$F_{1:M}$	$F(X_{1:M})$	GP$(F_{1:M}|X_{1:M})$	$F_{1:M^{(i)}}^{(i)} \equiv H^{(i)} + \Delta_{1:M^{(i)}}^{(i)}$
$F_{1:N}$	$F(G_{1:N})$	GP$(F_{1:N}|F_{1:M^{(i)}}, X_{1:M^{(i)}}, G_{1:N})$	$F_{1:N}^{(i)} \sim P(F_{1:N}|F_{1:M^{(i)}}^{(i)}, X_{1:M^{(i)}}^{(i)}, G_{1:N}^{(i)})$
g	Camera gain	Unif$(g_{\min}, g_{\max})$	$g' \sim P(g)$