CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Jan-Otto Hooghoudt, Margarida Barroso and Rasmus Waagepetersen

Towards Bayesian Inference of the Spatial Distribution of Proteins from Three-Cube Förster Resonance Energy Transfer Data

No. 05, April 2017

Towards Bayesian Inference of the Spatial Distribution of Proteins from Three-Cube Förster Resonance Energy Transfer Data

Jan-Otto Hooghoudt[1], Margarida Barroso[2]
and Rasmus Waagepetersen[3]

[1]Department of Civil Engineering, Aalborg University, joh@civil.aau.dk
[2]Department of Molecular and Cellular Physiology, Albany Medical College, barrosm@mail.amc.edu
[3]Department of Mathematics, Aalborg University, rw@math.aau.dk

Abstract

Förster resonance energy transfer (FRET) is a quantum-physical phenomenon where energy may be transferred from one molecule to a neighbour molecule if the molecules are close enough. Using fluorophore molecule marking of proteins in a cell it is possible to measure in microscopic images to what extent FRET takes place between the fluorophores. This provides indirect information of the spatial distribution of the proteins. Questions of particular interest are whether (and if so to which extent) proteins of possibly different types interact or whether they appear independently of each other. In this paper we propose a new likelihood-based approach to statistical inference for FRET microscopic data. The likelihood function is obtained from a detailed modeling of the FRET data generating mechanism conditional on a protein configuration. We next follow a Bayesian approach and introduce a spatial point process prior model for the protein configurations depending on hyper parameters quantifying the intensity of the point process. Posterior distributions are evaluated using Markov chain Monte Carlo. We propose to infer microscope related parameters in an initial step from reference data without interaction between the proteins. The new methodology is applied to simulated and real data sets.

Keywords: Bayesian inference, Markov chain Monte Carlo, Förster resonance energy transfer, spatial point process, spatial distribution, proteins, fluorophores.

1 Introduction

In the biology community there is a vast interest in studying the biomolecular structure and dynamics of macromolecular assemblies in order to understand their functions (Alber et al., 2017; Polo and Jackson, 2011; Krissinel and Henrick, 2007; Puglisi, 2005). Because the interactions between proteins and the typical size of proteins (1 nm to 100 nm) is at the nanoscale level (Erickson, 2009), no information can be obtained from conventional optical microscopic techniques, which at best can resolve distances down to ∼200 nm (van Putten et al., 2011). Instead, Förster resonance energy transfer – also referred to as fluorescence resonance energy transfer – microscopy is widely used to obtain such information. Förster resonance energy transfer (FRET) provides information about distances of the order of 2 nm to 10 nm within or between molecular structures and is the preferred tool for investigating spatial relationships in biochemistry (Wu and Brand, 1994; Gryczynski et al., 2005; Clegg, 1995, 2006).

FRET is the nonradiative transfer of the surplus of energy from an excited *donor* fluorophore (fluorescent molecule) to a sufficiently nearby *acceptor* fluorophore by dipole-dipole interaction (Heitler, 1954; Rohatgi-Mukherjee, 1978). The widespread use of FRET in biological research is based on the possibility to label, in vivo or in vitro, proteins with fluorophores that are spectrally matched (Miyawaki et al., 2003; Bunt and Wouters, 2004). The energy transfer due to the FRET mechanism is a stochastic process and the probability that energy transfer occurs between a donor and an acceptor fluorophore is heavily dependent on the distance between them. The probability that energy transfer occurs is commonly referred to as the efficiency of the energy transfer. The usefulness of FRET lies in the fact that various techniques exist by which the fraction of donor excitations that result in energy transfer – i.e. the efficiency – can be quantified.

Two main methods for determining the FRET efficiency are: fluorescence lifetime measurements (Wallrabe and Periasamy, 2005; Lakowicz, 2009; Chen et al., 2013) and spectral methods (Sun et al., 2011; Zimmermann et al., 2003). We focus in this paper on the most commonly applied spectral method called three-cube FRET. Due to the FRET mechanism a certain fraction of the de-excitations of a donor result in energy transfer to an acceptor, instead of donor photon emission. Thereby, the rate by which photons are emitted from the donors decreases – a phenomenon referred to as quenching of the donor – while instead, photons are emitted by the acceptors. Spectral methods now rely on determining the decrease in the donor emission due to FRET. For three-cube FRET, intensity measurements are carried out using three different filter sets – often referred to as cubes – each comprising: an excitation filter, a dichroic mirror, and an emission filter. This results in three digital intensity images (Zal and Gascoigne, 2004; Wallrabe et al., 2006; Periasamy et al., 2008; Periasamy and Day, 2011). Two images are obtained by exposing the sample to light in the donor absorption spectrum and recording emitted intensities both in the donor and the acceptor excitation spectrum. The third image is obtained by exposing the sample to light in the acceptor spectrum and also recording light in the acceptor spectrum.

Given FRET image data, the task is to obtain information concerning the spatial configuration of the donors and acceptors in the sample. For example, Wallrabe et al. (2003) study the clustering of ligand-receptor complexes in endocytic membranes using confocal FRET microscopy. They differentiate between a clustered or a random distribution of proteins by considering the dependence of the FRET efficiency on donor and acceptor concentrations. In particular, independence of the efficiency on acceptor concentration, or a decrease in the efficiency for higher unquenched donor signal for a fixed acceptor concentration, are both indicators for clustering (Kenworthy and Edidin, 1998; Kenworthy, 2001). Goswami et al. (2008) instead compare observed distributions of fluorescence intensity and fluorescence anisotropy with values expected from a Poisson distribution of nanoclusters.

Other, computational, approaches mainly rely on the construction of a configuration of donors and acceptors and computing the FRET efficiency related to this configuration by numerical computation of the energy transfer probabilities for each of the donors (Wolber and Hudson, 1979; Corry et al., 2005). This simple approach has been extended by various authors by simulating FRET events explicitly using Monte-Carlo techniques. The extended approach gives the possibility to include additional physical complexity into the model to account for possible photobleaching of donors and acceptors during a FRET measurement or the effect that temporarily unavailable acceptors can have on the FRET efficiency (Frederix et al., 2002; Berney and Danuser, 2003; Corry et al., 2005). Corry et al. (2005) further carefully studied the FRET efficiency in relation to various fixed donor and acceptor configurations (e.g. pentamers) and gives a concise overview of the development of the numerical Monte Carlo approaches.

Loura and Prieto (2011), Loura et al. (2010) and Lakowicz (2009) give excellent reviews of methods to extract spatial information in membrane biophysics from FRET data. Methods determining the complex structures of a protein or the spatial distribution of protein complexes in living cells are given in e.g. Raicu et al. (2009) and Bonomi et al. (2014).

The previous mentioned contributions are based on a detailed understanding of the FRET data generating mechanism. This knowledge, however, so far has not been applied to obtain a complete statistical model of FRET data allowing for a principled statistical analysis. In this paper we present a first attempt to conduct a full likelihood-based Bayesian analysis of three-cube FRET image data. The potential advantages of such an approach is that the posterior distribution gives detailed quantitative information regarding model parameters and donor-acceptor interactions as well as measures of uncertainty regarding this information. To obtain the likelihood function we derive, based on physical considerations, an accurate statistical model for the distribution of the image intensities, conditional on a point pattern consisting of donors and acceptors. We further impose a spatial point process prior (Møller and Waagepetersen, 2003) for the unknown configuration of donors and acceptors. Since our resulting posterior distribution is of a complicated form we use Markov chain Monte Carlo (MCMC) to sample from the posterior distribution (Gamerman and Lopes, 2006; Gilks et al., 1995). It is difficult to infer simultane-

ously microscope related parameters and possible interactions between donors and acceptors. We therefore propose to infer microscope related parameters in an initial step based on reference data without interactions between donors and acceptors. We asses the Bayesian inference procedure by a simulation study and by applying it to an empirical in-vitro reference dataset.

2 Observation model for three-cube FRET image data

A three-cube FRET data set consists of three images each corresponding to a rectangular region W which is a union of rectangular pixels, $W = \bigcup_{i\in\mathcal{G}} C_i$, indexed by a grid $\mathcal{G}$. Each pixel C_i records a light intensity due to emission from donors or acceptors. The images are created by (1) excitation of donors and measurement of donor emission, (2) excitation of acceptors and measurement of acceptor emission or (3) excitation of donors and measurement of acceptor emission (due to FRET). We represent the images by vectors $Y_{DD} = (Y^i_{DD})_{i\in\mathcal{G}}$, $Y_{AA} = (Y^i_{AA})_{i\in\mathcal{G}}$ and $Y_{DA} = (Y^i_{DA})_{i\in\mathcal{G}}$. The first letter in the subscripts denotes whether donors (D) or acceptors (A) were excited and the second letter denotes in which channel emission was measured. We assume that a pixel value Y^i_k, $k = DD, AA, DA$, $i \in \mathcal{G}$, is subject to additive normal noise. That is,

$$Y^i_k = I^i_k + \varepsilon_i \tag{2.1}$$

where I^i_k denotes light intensity due to emission and the noise terms ε_i are independent and $N(0, \sigma^2)$ distributed.

We now specify models for the I^i_k given configurations of donor and acceptor proteins in W whose positions form point patterns respectively $\mathbf{X}_D$ and $\mathbf{X}_A$.

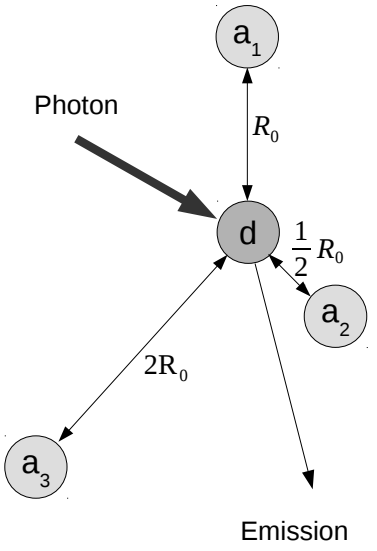
2.1 Some Fluorescence Resonance Energy Transfer Theory

An excited donor $d \in \mathbf{X}_D$ surrounded by a configuration $\mathbf{X}_A$ of acceptors can de-excite in three ways: either by emission with a rate k_{DE}, by non-radiative decay (e.g. internal heat conversion) with a rate k_{DN}, or by FRET to an acceptor $a \in \mathbf{X}_A$ with a rate $k_{F,da}$. We will refer to the sum of the first two mechanisms as the *intrinsic* de-excitation rate k_D, i.e. $k_D = k_{DE} + k_{DN}$. According to Förster (1948), $k_{F,da}$ is given by

$$k_{F,da} = k_D \left(\frac{R_0}{\|d - a\|} \right)^6$$

where R_0 is the so-called Förster distance, defined as the distance between the donor and acceptor at which the de-excitation rate due to FRET equals the intrinsic de-excitation rate. That is, $k_{F,da} = k_D$ if $\|d-a\| = R_0$. The probability that d de-excites due to FRET to a specific donor a in $\mathbf{X}_A$ thus becomes

$$P_{da} = \frac{k_{F,da}}{k_D + \sum_{\widetilde{a}\in\mathbf{X}_A} k_{F,d\widetilde{a}}} = \frac{(R_0/\|d-a\|)^6}{1 + \sum_{\widetilde{a}\in\mathbf{X}_A} (R_0/\|d-\widetilde{a}\|)^6}.$$



(a)

	$\left(\frac{R_0}{\|d-a\|}\right)^6$	P_{da}
a_1	1	$1/(1+S_d)$
a_2	64	$64/(1+S_d)$
a_3	$1/64$	$(1/64)/(1+S_d)$

$$S_d = \sum_{\tilde{a}\in\mathbf{X}_A}(R_0/\|d-\tilde{a}\|)^6 = 1 + 64 + 1/64$$

(b)

Figure 1: (a) Donor fluorophore surrounded by three acceptors at distances R_0, $\frac{1}{2}R_0$ and $2R_0$. (b) The table shows the de-excitation "path widths" $(R_0/||d-a||)^6$ for energy transfer from the donor to each of the acceptors and the corresponding energy transfer probabilities P_{da}.

Figure 1 shows an example of the computation of P_{da} for a specific configuration of acceptors a around a donor d. The total probability that d de-excitates due to FRET is $P_{dA} = \sum_{a\in\mathbf{X}_A} P_{da}$. The probability that d de-excites by emission or by non-radiative decay is $P_{dD} = 1 - P_{dA}$.

2.2 Model for intensities given protein configurations

Our model for the intensities given the configurations $\mathbf{X}_D$ and $\mathbf{X}_A$ is inspired by the model for simulation of FRET data in Corry et al. (2005). However, in contrast to Corry et al. (2005) we introduce the simplifying assumption that a donor or acceptor is always available for excitation (see also Wolber and Hudson, 1979; Berney and Danuser, 2003). This is a reasonable assumption if the intensity of the laser is moderate so that the inter arrival times of photons at a donor are large compared with the de-excitation times. We can then regard the times of excitations of donors and acceptors as Poisson processes and use standard results for Poisson processes to obtain closed form distributional results for the I_k^i.

In the Appendix (page 25) we show that $I_{DD}^i = G_D N_{DD}^i$ and $I_{DA}^i = G_A N_{DA}^i$ where N_{DD}^i and N_{DA}^i are the number of photons detected by the detector in, respectively, the DD-channel and the DA-channel, and N_{DD}^i and N_{DA}^i are both Poisson distributed. Further, G_D and G_A are unknown positive parameters related to the sensitivity of the detector in, respectively, the donor and acceptor emission spectrum. The means of I_{DD}^i and I_{DA}^i are

$$\mu_{DD}^i = M_D \sum_{d\in\mathbf{X}_D\cap C_i} (1 - P_{dA})$$

and

$$\mu^i_{DA} = GM_D \sum_{a \in \mathbf{X}_A \cap C_i} \sum_{d \in \mathbf{X}_D} P_{da},$$

where M_D and G are unknown positive parameters. We assume that the means μ^i_{DD}/G_D and μ^i_{DA}/G_A of, respectively, N^i_{DD} and N^i_{DA} are sufficiently large so the Poisson distributions of N^i_{DD} and N^i_{DA} can be well approximated by normal distributions. Then

$$I^i_{DD} \sim N(\mu^i_{DD}, G_D\mu^i_{DD}) \quad \text{and} \quad I^i_{DA} \sim N(\mu^i_{DA}, G_A\mu^i_{DA}). \tag{2.2}$$

By a similar line of arguments we also obtain

$$I^i_{AA} \sim N(\mu^i_{AA}, G_A\mu^i_{AA}) \tag{2.3}$$

where

$$\mu^i_{AA} = (M_D/K)\ n(\mathbf{X}_A \cap C_i),$$

K is an unknown positive parameter and $n(\mathbf{X}_A \cap C_i)$ denotes the number of acceptors within pixel C_i. In cases of intensity data with a large proportion of zeros we instead use truncated normal distributions with point masses at zero for I^i_{DD}, I^i_{DA} and I^i_{AA}, see Section B.2 of Supplement B.

The equations (2.1), (2.2) and (2.3) specify the distribution of the FRET data conditional on the protein configurations $\mathbf{X}_D$ and $\mathbf{X}_A$. The distribution is parameterized by $\psi = (M_D, G, K, G_D, G_A, \sigma^2)$. We refer to the components of ψ as microscope related parameters. The parameters G and K are known as the so-called G- and K-factors (Zal et al., 2002; Chen et al., 2007). The parameter M_D can be interpreted as the mean donor emission detector read-out intensity due to one donor excitation.

3 Bayesian inference of spatial characteristics of protein configurations

We adopt a Bayesian approach to infer the microscope related parameters ψ and spatial characteristics of the configurations $\mathbf{X}_D$ and $\mathbf{X}_A$ of proteins. A spatial point process prior (specified in Section 3.1) is used for $\mathbf{X} = (\mathbf{X}_D, \mathbf{X}_A)$ where this prior again depends on a parameter vector θ. We also assign a prior to θ thus including also this parameter in the posterior inference. As detailed later in Sections 4.4 and 5 we recommend to infer the microscope related parameters in an initial step using reference data without interactions between donors and acceptors. In a second step, investigating interactions in a data set of biological scientific interest, ψ can then be fixed at estimates obtained from the first step. Letting y denote an observation of $Y = (Y_{DD}, Y_{DA}, Y_{AA})$ and $(\mathbf{x}_D, \mathbf{x}_A)$ a realization of $\mathbf{X}$, the joint posterior distribution is

$$p(\mathbf{x}_D, \mathbf{x}_A, \theta | y, \psi) \propto p(y|\mathbf{x}_D, \mathbf{x}_A, \psi) p(\mathbf{x}_D, \mathbf{x}_A|\theta) p(\theta). \tag{3.1}$$

Here $p(z)$ and $p(z|u)$ is generic notation for a probability density of a random quantity Z and the conditional density of Z given another random quantity U.

3.1 Priors

We model a priori $\mathbf{X}_D$ and $\mathbf{X}_A$ as independent Poisson processes on W with intensities θ_D and θ_A. That is, the prior density of $(\mathbf{X}_A, \mathbf{X}_D)$ with respect to independent unit rate Poisson processes is

$$p(\mathbf{x}_D, \mathbf{x}_A|\theta) = \theta_A^{n(\mathbf{x}_A)}\theta_D^{n(\mathbf{x}_D)} \exp[-|W|(\theta_A + \theta_D - 2)] \tag{3.2}$$

where $|W|$ denotes the area of W and $n(\mathbf{x})$ denotes the number of points in a point configuration $\mathbf{x}$ (see for instance equation 6.2 in Møller and Waagepetersen, 2003). We further impose independent conjugate Gamma hyper priors for θ_D and θ_A. The Poisson prior can be viewed as a null model for the case of no interaction between donors and acceptors or within donors respectively acceptors. Compared with other more complex point process models like Markov point processes (like the Strauss hard core model considered in Section 4), the Poisson prior is advantageous in having a known density function. A potential problem is that the Poisson prior is in some sense a strong prior which assigns little probability to point configurations with strong clustering or regularity. This can lead to biased results as demonstrated in Section 4. Densities for more flexible Markov point process prior models on the other hand contain intractable normalizing constants that depend on the unknown parameters in the point process model. This then precludes the use of standard Markov chain Monte Carlo algorithms (Section 3.2) for evaluation of the posterior distribution.

The gamma distributions for θ_D and θ_A are defined through shape parameters α and rate parameters β. As the mean of the Gamma distribution is α/β and its variance α/β^2, the signal-to-noise ratio related to the distribution is defined by the square root of the shape parameter, i.e

$$\frac{S}{N} = \frac{\alpha/\beta}{\sqrt{\alpha/\beta^2}} = \sqrt{\alpha}.$$

In a typical FRET experiment there is quite some uncertainty concerning the true values of the numbers of proteins within the sample so we have defined not too confined priors for the intensities θ_D and θ_A. We have chosen to set the signal-to-noise ratio always equal to 2, resulting in the value of 4 for the shape parameter. In our applications we further specify the prior mean m of each of the parameters so that the rate parameter β follows from $\beta = \alpha/m = 4/m$. We also use gamma priors for the components of ψ, see the discussion of prior elicitation for ψ in Section 6.

3.2 Markov chain Monte Carlo

To evaluate the posterior distribution we use a Markov chain Monte Carlo algorithm (Gamerman and Lopes, 2006) where the components $(\mathbf{X}_D, \mathbf{X}_A)$, θ_D, θ_A and (if applicable) the components of ψ are updated in turn. Gibbs updates are used for the full conditional Gamma distributions of θ_D and θ_A while random walk Metropolis updates on the log scale are used for the components of ψ. For the point configurations

$(\mathbf{X}_A, \mathbf{X}_D)$ we first randomly choose to either update $\mathbf{X}_A$ or $\mathbf{X}_D$ (with probability $1/2$ for each choice). We then use birth-death updates as outlined in Sections 7.1.2–7.1.3 in Møller and Waagepetersen (2003). If e.g. $\mathbf{X}_A$ is chosen to be updated then with probability $1/2$ it is proposed to remove a point chosen from the uniform distribution on $\mathbf{X}_A$. Otherwise it is proposed to insert a new acceptor point at a location chosen from the uniform distribution on W. In case it is proposed to remove a point $u \in \mathbf{X}_A$, the Metropolis-Hastings ratio becomes

$$\frac{p(\mathbf{x}_D, \mathbf{x}_A \backslash \{u\}, \theta | y, \psi) n(\mathbf{x}_A)}{p(\mathbf{x}_D, \mathbf{x}_A, \theta | y, \psi) |W|} = \frac{p(y | \mathbf{x}_D, \mathbf{x}_A \backslash \{u\}, \psi) n(\mathbf{x}_A)}{p(y | \mathbf{x}_D, \mathbf{x}_A, \psi) \theta_A |W|}.$$

If it is proposed to insert a new acceptor point $v \in W$ the Metropolis-Hastings ratio is

$$\frac{p(y | \mathbf{x}_D, \mathbf{x}_A \cup \{v\}, \psi) \theta_A |W|}{p(y | \mathbf{x}_D, \mathbf{x}_A, \psi)(n(\mathbf{x}_A) + 1)}.$$

The expressions for updating $\mathbf{X}_D$ are similar. The described birth-death updates are repeated a large fixed number of times between the updates of the parameters θ_D, θ_A and ψ.

To keep the MCMC updates for donor and acceptor points numerically feasible, only those acceptors that reside within $4R_0$ of a donor are taken into account as a possible path for energy transfer for the donor. This important simplification will not lead to any significant difference in posterior results as the transfer probability P_{da} for a donor d and an acceptor a is very small when $\|d - a\| > 4R_0$. Thereby adding or removing a point in pixel i can only affect the values of likelihood factors $p(y^l_{DD}, y^l_{DA}, y^l_{AA} | \mathbf{x}_D, \mathbf{x}_A, \psi)$ for pixels l in a neighbourhood of i (note that the likelihood factors as $\prod_{i \in \mathcal{G}} p(y^i_{DD}, y^i_{DA}, y^i_{AA} | \mathbf{x}_D, \mathbf{x}_A, \psi)$). Exploiting this simplification, we have implemented an ingenious algorithm that recomputes the transfer probabilities P_{da} only for donors and acceptors which are influenced by the adding/removing of a point. A detailed description of the MCMC sampler is provided in Supplement B on page 37.

3.3 Inferring spatial characteristics

In statistics for spatial point processes, the K-function is a common tool for inferring interactions from a spatial point pattern. We adapt this approach and use the cross K function (e.g. Møller and Waagepetersen, 2003) to measure interactions between donors and acceptors given point configurations $\mathbf{x}_A$ and $\mathbf{x}_D$. In general, for point processes $\mathbf{X}_1$ and $\mathbf{X}_2$ of intensities ρ_1 and ρ_2, $\rho_2 K_{12}(t)$ is the expected number of $\mathbf{X}_2$ points within distance t from a typical point of $\mathbf{X}_1$. In case of no interaction between $\mathbf{X}_1$ and $\mathbf{X}_2$, $K_{12}(t) = \pi t^2$. Values of $K_{12}(t)$ greater (smaller) than πt^2 signifies positive (negative) interaction between $\mathbf{X}_1$ and $\mathbf{X}_2$. It is common to consider the cross L-function $L_{12}(t) = \sqrt{K_{12}(t)/\pi}$ which is equal to t in case of no cross interaction while $L_{12}(t) > t$ ($L_{12}(t) < t$) means positive (negative) cross interactions. For ease of presentation we refer to $L_{12}(t) - t$ as the 'centered' cross L-function.

Given configurations $\mathbf{x}_D$ and $\mathbf{x}_A$ of donors and acceptors we estimate the cross K function by

$$\widehat{K}_{\mathbf{x}}(t) = \sum_{u \in \mathbf{x}_A, v \in \mathbf{x}_D} \frac{1[\|u-v\| \leq t]}{n(\mathbf{x}_A) n(\mathbf{x}_D) |W \cap W_{u-v}| |W|^{-2}}$$

where W_{u-v} is W translated by $u-v$ (e.g. Section 4.4.3 in Møller and Waagepetersen, 2003). The cross L-function is estimated by $\widehat{L}_{\mathbf{x}}(t) = \sqrt{K_{\mathbf{x}}(t)/\pi}$. To infer cross spatial interactions between donors and acceptors given FRET data we consider the posterior distribution of

$$\widehat{L}_{\mathbf{X}}(t) = \sqrt{\widehat{K}_{\mathbf{X}}(t)/\pi}$$

or its centered version. We also considered so-called cross G- and J-functions (e.g. Møller and Waagepetersen, 2003) but in our simulation studies the cross L-function gave a more clear impression of the nature of donor-acceptor interactions.

4 Simulation studies

Our primary target of inference is the cross L-function, $\widehat{L}_{\mathbf{X}}$ for the configuration $\mathbf{X} = (\mathbf{X}_A, \mathbf{X}_D)$ of donors and acceptors which is unknown in practice. However, we also need to infer the microscope related parameters ψ. From a Bayesian perspective, if the right prior distribution is chosen, the posterior distribution by definition provides the correct inference given the data Y and prior information. However, in our case, the Poisson prior (3.2) is partly chosen for convenience in order to yield tractable MCMC computations and is not necessarily the best possible representation of prior information. Thus from a pragmatic point of view it makes sense to assess possible bias of our Bayesian inference procedure.

In particular we focus in Section 4.3.1 on the posterior mean $L_{|Y}$ of $\widehat{L}_{\mathbf{X}}$ as a predictor of $\widehat{L}_{\mathbf{X}}$. The posterior mean $L_{|Y}$ is further an estimate of $L^E = E[\widehat{L}_{\mathbf{X}}]$ which is the expected value of $\widehat{L}_{\mathbf{X}}$ over replicated data $\mathbf{X}$. Note in this connection that had we used the true distribution of $\mathbf{X}$ as the prior then $EL_{|Y}$ and L^E would be exactly equal – i.e. $L_{|Y}$ would be an unbiased predictor/estimate both of $\widehat{L}_{\mathbf{X}}$ and L^E. In Section 4.3.3 we assess the performance of the full posterior distribution of $\widehat{L}_{\mathbf{X}}$ given Y for inference regarding $\widehat{L}_{\mathbf{X}}$. In Section 4.3 we consider ψ to be a fixed known parameter. Section 4.4 is concerned with inference regarding ψ.

4.1 Simulation of synthetic data

To generate synthetic data for the simulation study, the point configuration $\mathbf{X} = (\mathbf{X}_A, \mathbf{X}_D)$ is generated on a 1000 nm by 1000 nm square region as a realization of a bivariate Strauss hard core process. This point process has density (with respect to a bivariate process of independent unit rate Poisson processes) of the form

$$f(\mathbf{x}_A, \mathbf{x}_D) \propto \beta_D^{n(\mathbf{x}_D)} \beta_A^{n(\mathbf{x}_A)} \gamma^{s_R(\mathbf{x}_A, \mathbf{x}_D)} \mathrm{HC}(\mathbf{x}_A, \mathbf{x}_D, \Re_A, \Re_D, \Re_{DA}) \tag{4.1}$$

where $s_R(\mathbf{x}_A, \mathbf{x}_D)$ is the number of unordered pairs of points $\{u, v\}$ with $u \in \mathbf{x}_A$, $v \in \mathbf{x}_D$, and interpoint distance $\|u - v\|$ less than R. Values of γ less than one lead to repulsion between donors and acceptors while values of γ greater than one lead to attraction. The term $\mathrm{HC}(\mathbf{x}_A, \mathbf{x}_D, \Re_A, \Re_D, \Re_{DA})$ is one if the following hard core condition is satisfied: all donors have an interpoint distance greater than $\Re_D$, all acceptors have an interpoint distance greater than $\Re_A$ and all pairs of points where one is a donor and the other an acceptor have an interpoint distance greater than $\Re_{DA}$. Otherwise the hard core term is zero whereby it serves to model that donors and acceptors have a physical extent that prevents them from getting arbitrarily close to each other. Different settings of the Strauss hard core process parameters are used to create different point pattern types described in Section 4.2 below.

Next, conditional on the configuration $\mathbf{X}$ and the various microscope related parameters ψ, the intensity data Y is generated from the model specified in Section 2. Regarding the observation model we fix the measurement variance σ^2 at 25, let each of G, K, G_D, G_A equal to 1, and consider values $1, 5, 20$ of M_D in order to generate data of varying signal to noise ratios defined by $\mathbf{E}[\mu^i_{DD}/(G_D\mu^i_{DD} + \sigma^2)^{1/2}]$ and $\mathbf{E}[\mu^i_k/(G_A\mu^i_k + \sigma^2)^{1/2}]$, $k = DA, AA$. For each point pattern type we generate 100 independent synthetic point patterns $\mathbf{X}^{\mathrm{synth},i}$ and associated synthetic image data $Y^{\mathrm{synth},i}$, $i = 1, \ldots, 100$.

4.2 Point pattern types

The basic point pattern types considered are *dimer*, *clustered*, *Poisson hard core* and *repulsive*. For all types, $\Re_D$, $\Re_A$ and $\Re_{DA}$ are at least 2 nm. The parameters β_D and β_A are further adjusted to have on average 1000 donors and 1000 acceptors.

In case of *dimer*, we specify large values of $\Re_D = \Re_A$ which essentially means that only proteins of different types can appear close to each other. Thus the only clusters possible are mini-clusters consisting of one donor and one acceptor, i.e dimer clusters. For the *clustered* case, $\Re_D = \Re_A$ are reduced which enables formation of a wider range of clusters containing several donors and acceptors. In case of *dimer* and *clustered*, values of $\gamma_{DA} = 2, 8$ correspond to respectively moderate and strong interaction. For *Poisson hard core*, $\gamma_{DA} = 1$ while all hard core distances are 2. In case of *repulsive*, varying values of $\Re_{DA}$, r and $\gamma_{DA} < 1$ generate different strengths of repulsive interaction between donors and acceptors. Table 1 gives an overview of the different parameter settings considered.

4.3 Inference regarding spatial characteristics

We estimate L^E by the empirical average of L-functions $\widehat{L}_{\mathbf{X}^{\mathrm{synth},i}}$ obtained from the $\mathbf{X}^{\mathrm{synth},i}$. From each synthetic data set $Y^{\mathrm{synth},i}$ we further obtain an MCMC estimate $\bar{L}_{|Y^{\mathrm{synth},i}}$ of the posterior mean $L_{|Y^{\mathrm{synth},i}}$ of $\widehat{L}_{\mathbf{X}}$ given $Y^{\mathrm{synth},i}$. The mean posterior L-function $EL_{|Y}$ is estimated by the mean of the $\bar{L}_{|Y^{\mathrm{synth},i}}$. The sampling variability of $L_{|Y}$ is further represented by the variation of the $\bar{L}_{|Y^{\mathrm{synth},i}}$. When considering inference for the cross L-function in the following Sections 4.3.1–4.3.3, ψ is fixed at the value used for generating the synthetic data sets.

Table 1: Parameter settings used in the Strauss hard core model (4.1) to create the various point pattern types. The values for the homo ($\Re_{DD}$, $\Re_{AA}$) and hetero ($\Re_{DA}$) hard core distances and the hetero interaction radius (R) presented in the table follow from the considerations stated in the text and by the choice of the Förster distance $R_0 = 6$.

Type number	$\Re_{DD} = \Re_{AA}$	$\Re_{DA}$	R	γ_{DA}	Short name
1	18	2	6	2	Dim.12
2	18	2	6	8	Dim.18
3	12	2	6	2	Dim.22
4	12	2	6	8	Dim.28
5	6	2	6	2	Clu.12
6	6	2	6	8	Clu.18
7	2	2	6	2	Clu.22
8	2	2	6	8	Clu.28
9	2	2	0	1.0	Poi.HC
10	2	6	0	1.0	Rep.h1
11	2	12	0	1.0	Rep.h2
12	2	2	10	0.5	Rep.s1
13	2	2	10	0.1	Rep.s2

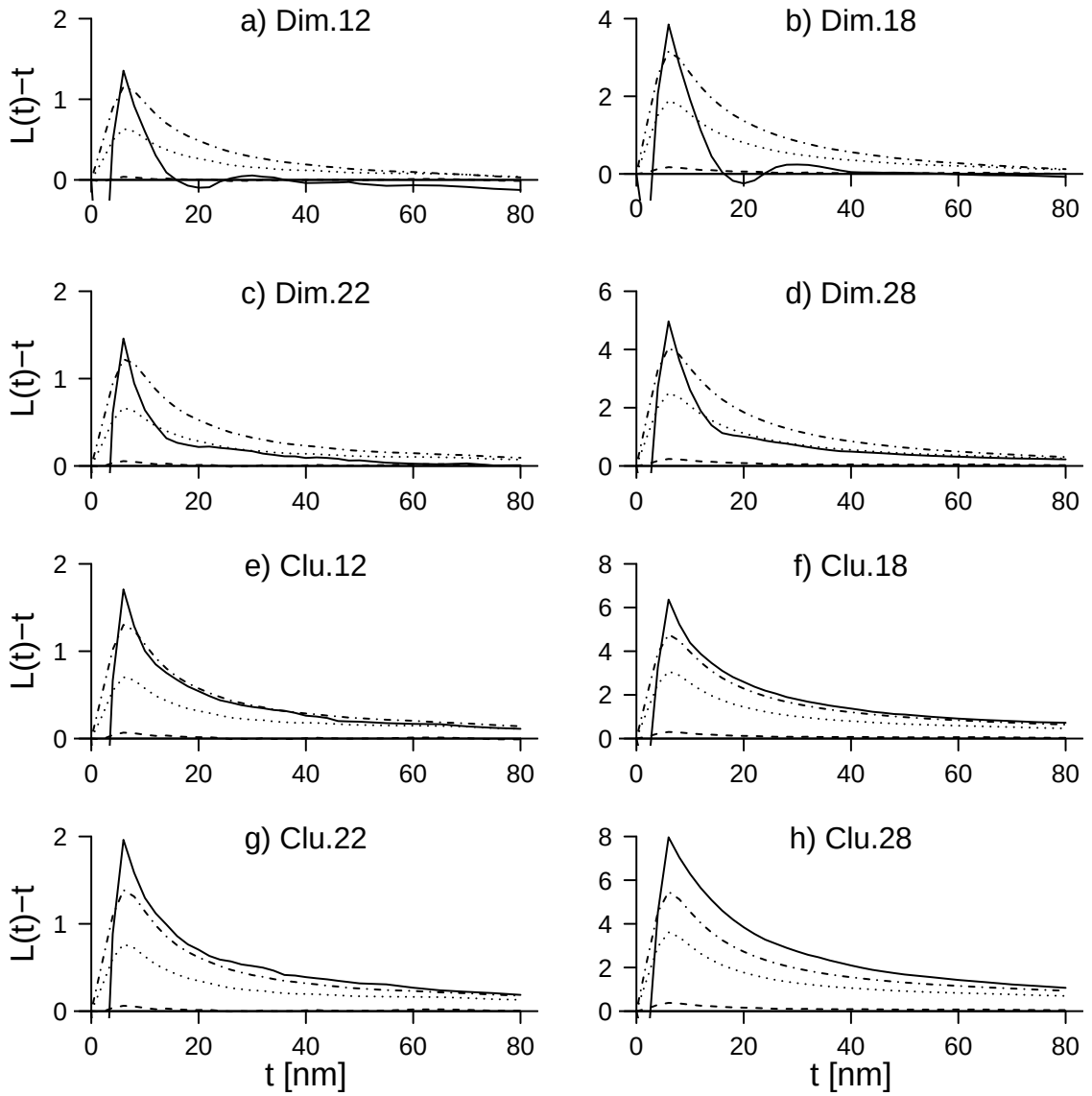


Figure 2: In each plot the solid line is the centered L^E function. The other lines show the centered $EL_{|Y}$ for varying M_D: dashed line: $M_D = 1$, dotted line: $M_D = 5$, dashed-dotted line: $M_D = 20$. The plots are for the dimer and clustered point pattern types.

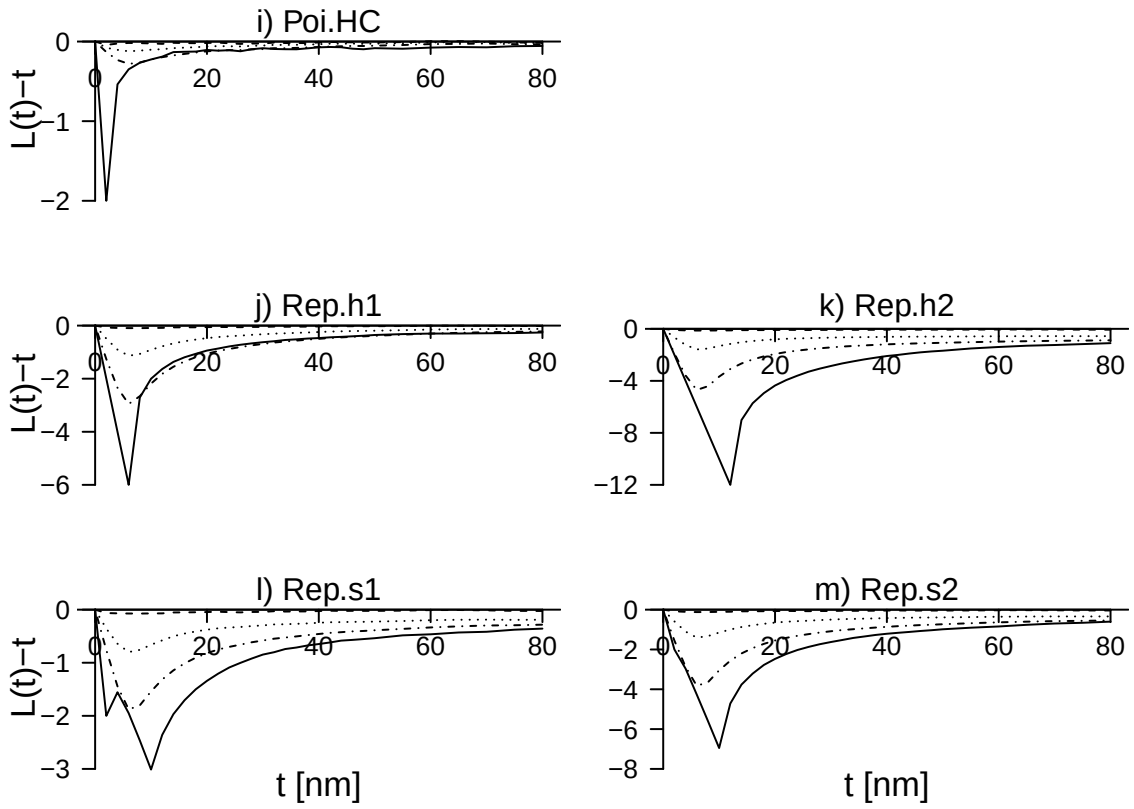


Figure 3: Continuation of Figure 2. Plots are for the Poisson hard core and repulsive types.

4.3.1 Bias of posterior mean

We assess the bias of $L_{|Y}$ by considering the mean (over replicated data $\mathbf{X}, Y$) of $\widehat{L}_{\mathbf{X}}$ and its prediction $L_{|Y}$. Thereby we also assess how $L_{|Y}$ performs as an estimate of L^E (the expected cross L-function for $\mathbf{X}$).

In Figures 2 and 3 the estimates of $L^E(t) - t$ and $EL_{|Y}(t) - t$ are shown for each of the point pattern types in Table 1 for the three values of M_D (1,5,20). In Figure 2 (a) for example, L^E for dimer-type 1.2 shows that the underlying point patterns are clustered for distances $r < 16$ (as $L^E > 0$) and are slightly repulsive for distances $16 < r < 24$. Further, for distances $24 < r < 40$ there seems to be some slight clustering again while for $r > 40$, the pattern displays complete spatial randomness. The negative values that occur for $r < 5$ is due to the minimum imposed hardcore distance of 2. The negative part for small r is also visible in the plots (b)–(h) where L^E otherwise indicates clustering among donors and acceptors and in plot (i) where $L^E(t) - t$ is close to zero otherwise.

The general impression from the plots is that $L_{|Y}$ is biased downwards when the true point patterns are of dimer or clustered types (a)–(h) and biased upwards in the cases of the repulsive types (i)–(m) (including the Poisson hard core case). For $M_D = 1$, where the signal to noise ratio is very low, the mean of $L_{|Y}$ is very close to zero, and it does not seem possible to infer in this case cross interactions between donors and acceptors. However, for $M_D = 5, 20$, there is always a pronounced peak (positive or negative) of the mean $L_{|Y}$ function where the peak is of the right sign and located in the right place of the peak of the L^E function. Moreover the bias consistently decreases when M_D and hence the signal to noise ratio increases. Despite of the bias the results suggest that qualitatively correct statements can be made regarding independence, clustering or repulsion between donors and acceptors.

4.3.2 Variability of posterior mean L-function

In addition to bias the extent to which valid qualitative conclusions can be made from the posterior $L_{|Y}$ function of course also depend on its variability. Figure 4 shows for each basic point pattern type and M_D either 5 or 20, 98% envelopes for $L_{|Y}(t)$ based for each t on the minimal and maximal values of $L_{|Y}(t)$ over the 100 replications. These envelopes are fairly narrow for distances up to 100 and shows that, in the setting of the simulation study, qualitative conclusions regarding the nature of interaction between donors and acceptors will be consistently correct over replicates.

4.3.3 Inference based on full posterior distribution

The results in Section 4.3.1 showed that the posterior mean L-function $L_{|Y}$ can exhibit substantial bias as an estimate of L^E and hence also as a predictor of $\widehat{L}_{\mathbf{X}}$. This can invalidate the use of the full posterior distribution for inferring the uncertainty regarding the estimation of L^E or the prediction of $\widehat{L}_{\mathbf{X}}$. As an example Figure 5 shows $L_{|Y^{\text{synth},1}}(t) - t$ and the 98% central posterior interval for $\widehat{L}_{\mathbf{X}}(t) - t$ given

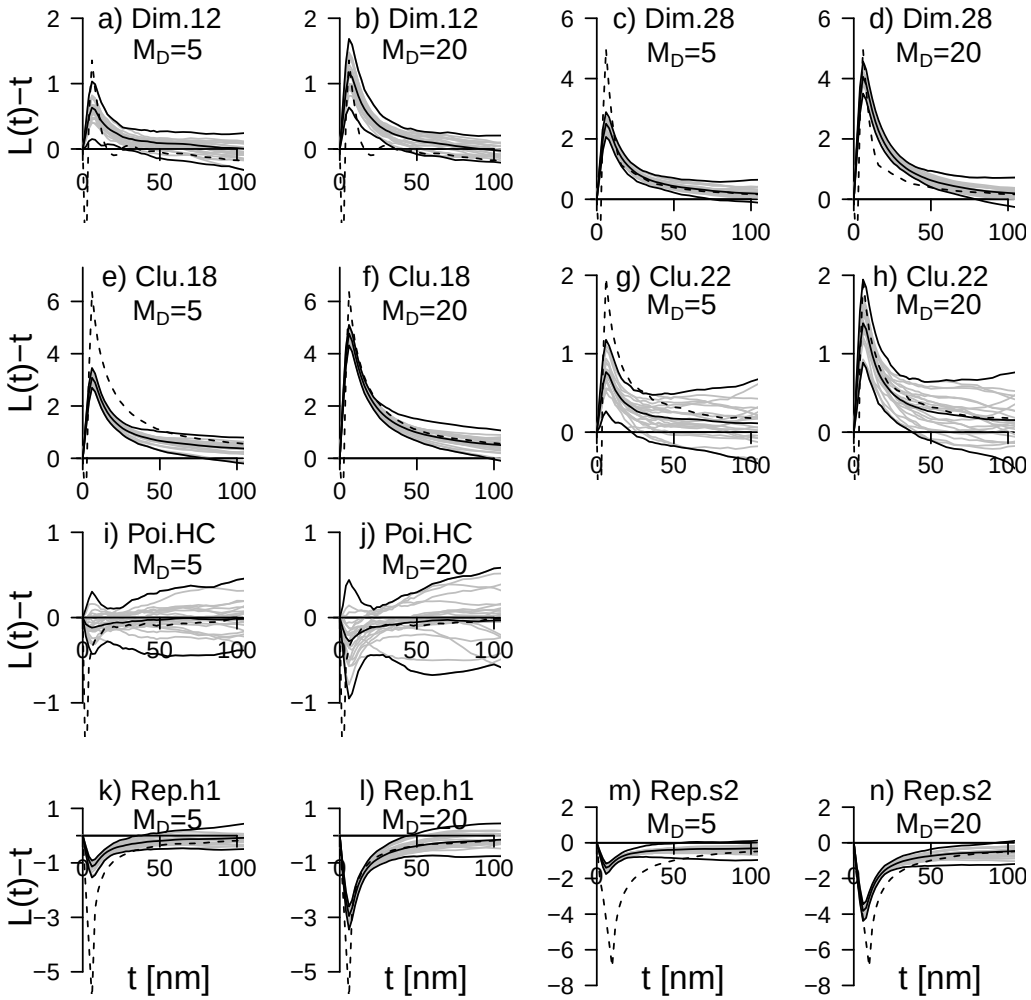


Figure 4: Distribution of the centered $\bar{L}_{|Y}$ function for the various point pattern types, summarized by 98% envelopes based on $\bar{L}_{|Y^{\text{synth},i}}$, $i = 1, \ldots, 100$, together with the mean value $\widehat{E}L_{|Y}$ (middle solid line). The value of M_D is 5 or 20. In each plot twenty of the $\bar{L}_{|Y^{\text{synth},i}}$'s are shown with solid gray curves, the dashed line is the centered $\widehat{L}^E$ function.

$Y = Y_{\text{synth},1}$ for the same point pattern types as in Figure 4 and $M_D = 5, 20$. Also the true $\widehat{L}_{\mathbf{X}^{\text{synth},1}}(t) - t$ are shown in each plot.

For the lower signal-to-noise ratio with $M_D = 5$, $\widehat{L}_{\mathbf{X}^{\text{synth},1}}(t)$ falls outside the 98% posterior interval for several point pattern types. Thus, the posterior intervals do not always give a useful quantification of the uncertainty regarding the knowledge of $L_{\mathbf{X}}$. However, for the higher signal-to-noise ratio with $M_D = 20$, the envelopes do include or almost include the $\widehat{L}_{\mathbf{X}^{\text{synth},1}}$ function.

4.4 Simulation studies for microscope parameters

So far the vector ψ of microscope related parameters has been assumed to be known which is rarely the case. We have investigated Bayesian inference for ψ in a simulation study for which the full details are given in Supplement C on page 48. We here just comment on results obtained for simulations with $G_D = G_A = K = G = 1$,

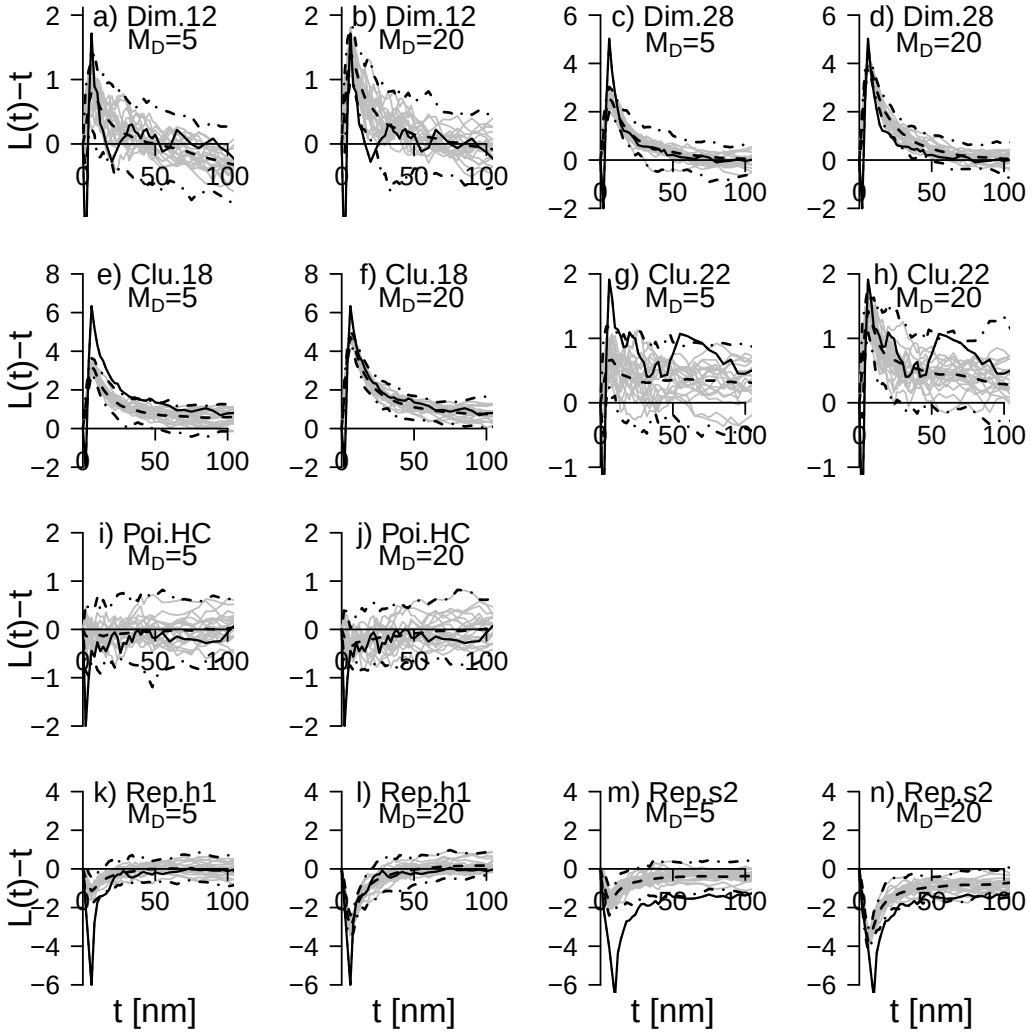


Figure 5: Posterior distribution of $\widehat{L}_{\mathbf{X}}$ given the first synthetic data set $Y^{\text{synth},1}$ for each point pattern type. Dashed: posterior mean $L_{|Y^{\text{synth},1}}$, dashed-dotted: 98% envelopes and solid gray: twenty posterior realizations of $\widehat{L}_{\mathbf{X}}$. Solid black shows the true $\widehat{L}_{\mathbf{X}^{\text{synth},1}}$.

$M_D = 20$ and $\sigma^2 = 25$. Figure 6 shows boxplots of the posterior mean of each microscope parameter over 40 replicated data sets for each point pattern type. The main features are as follows.

1. for the Poisson hard core patterns the posterior means of all the parameters coincide with or are very close to their respective synthetic values (see plots (a)–(f) for type number 9).
2. for all the clustered patterns (type number 1–8) inference for M_D, G and K is biased. The posterior means $\bar{M}_D$ are significantly below their synthetic value, while $\bar{G}$ and $\bar{K}$ are above their target values. Further the bias increases for the patterns generated with $\gamma_{DA} = 8$ (type number 2,4,6,8) compared to the corresponding patterns generated with $\gamma_{DA} = 2$ (type number 1,3,5,7).
3. for repulsive patterns (type number 11–13) $\bar{M}_D$ and $\bar{K}$ are on or close to target, while $\bar{G}$ is negatively biased.

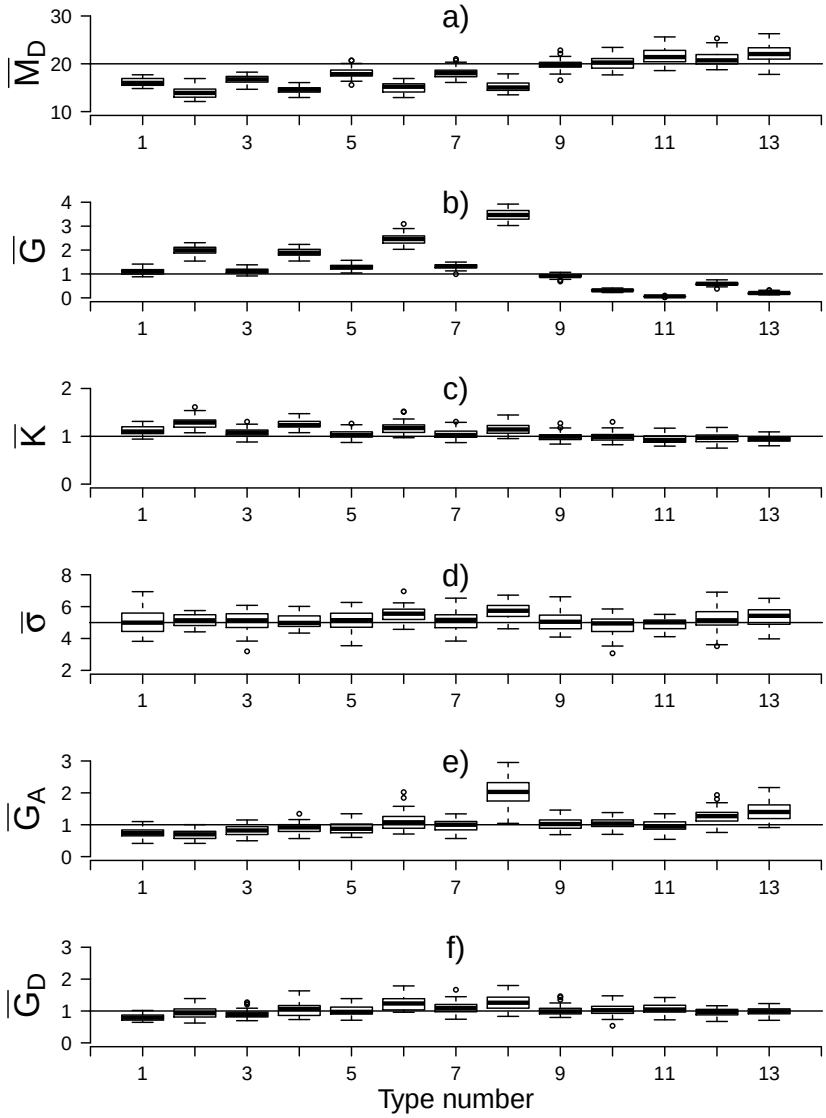


Figure 6: Boxplots of the posterior mean of all the six microscope parameters for the forty replicated runs for each of the point pattern types (referred to by their type number, see Table 1) for $M_D^s = 20$. The horizontal lines are drawn at the corresponding synthetic values.

4. the posterior means of σ, G_A and G_D are on target for almost all point pattern types (plots (d)–(f)).

As explained in detail in Supplement C, the biased results are due to the mismatch between the Poisson point process prior and the actual point processes used for the simulations. The microscope related parameters thus, a posteriori take on values to "soothe" this mismatch resulting in biased results. Further, in this setting with joint inference of ψ and spatial characteristics, the posterior mean L-functions are strongly biased as well, being close to zero for all distances for all point patterns types (not shown).

5 Two-step approach to likelihood-based inference

In the previous Section 4.4 we observed that applying the Bayesian inference methodology using the Poisson process prior on Poisson hard core patterns gave reliable estimates for all microscope parameters. This suggests an approach where microscope related parameters are inferred from reference data sets constructed with absence of donor-acceptor interactions. Therefore we propose a two-step approach where the microscope related parameters are inferred in a first step using reference data. In the second step the values of the microscope parameters are fixed at the posterior estimates from the first step in order to make inference on the spatial configuration of donors and acceptors of a three-cube FRET sample of biological scientific interest. To illustrate the approach we have carried out the first step on empirical reference three-cube FRET data, as dicussed in the following section.

6 Data example

In this section we apply our Bayesian methodology to empirical in vitro three-cube FRET data obtained from donor or acceptor fluorophore labeled transferrin proteins (Welch, 1992) attached to polylysine slides (Shima and Sakai, 1977). Transferrin bound to polylysine is known to be approximately randomly (i.e. Poisson hardcore) distributed (Wallrabe et al., 2007). The objective is to infer the microscope parameters related to the experimental set-up.

We initially conducted an exploratory analysis (described in Supplement A on page 27) where we quantified the amount of photobleaching and compared empirical mean-variance relationships of the image data with the ones implied by our model. From these mean-variance relationships, as well as other non-Bayesian methods discussed in Supplement A, we obtained rough estimates for the microscope parameters that were used to set the prior means in the Bayesian analysis. We thus use a pragmatic Bayesian approach where the rough non-Bayesian estimates entering in the priors are refined by introducing information obtained through the likelihood derived from our observation model.

Due to certain computational issues discussed in Section 6.2 and Section 6.3.1 we are at this stage only able to use a small subset of the full data in the Bayesian

inference. Improving the computational methodology is an important topic of further research.

6.1 The image data set

Three cube FRET measurements have been carried out on three samples, to which we refer as samples 1, 2 and 3. Sample 1 is prepared to consist of twice as many donors (D) as acceptors (A), that is D:A $\approx 1 : \frac{1}{2}$, while samples 2 and 3 are prepared such that, respectively, D:A $\approx 1 : 1$ and D:A $\approx \frac{1}{2} : 1$. Three-cube FRET data is obtained on each sample on a square grid containing 512×512 square pixels. The pixel side length is 0.279 µm and the focal volume depth is approximately 5 pixels (1.4 µm) (Wallrabe et al., 2007). The image data are shown in Figure 7.

The emission in the DD-channel (Y_{DD}) and AA-channel (Y_{AA}) are corrected for background emission, while the DA-channel data (Y_{DA}) is also corrected for spectral bleedthrough, by the methods described in Elangovan et al. (2003).

We noticed that around the edges of the 512×512 images, often very low or zero intensity regions occurred due to improper sample preparation. Therefore, the exploratory statistical analysis has been based solely on the central rectangular section of the images consisting of 100×100 pixels (see also Supplement A, Section A.1).

In order to obtain sufficient photon count statistics – that is sufficiently high signal-to-noise ratio for each pixel – each sample has been remeasured ten times. We then create an aggregated dataset by summing pixel wise over the ten measured intensities for each channel. We note that by remeasuring the sample instead of increasing the measurement time, we obtain information concerning: the amount of photobleaching occurring for remeasurements (Supplement A, Section A.3), and the pixel intensity variance in the three channels. The latter information gives the possibility to deduce the empirical mean-variance relationship of the image data and to obtain estimates for G_D and G_A (Supplement A, Section A.6).

6.2 Inference procedure set-up

The prior distributions for the microscope and Poisson point process parameters were specified as gamma distributions. For each parameter the shape parameter is set to 4 based on the reasoning in Section 3.1 and we use a pragmatic Bayesian approach where the prior means of five of the microscope parameters are set by aid of the rough estimates $M_D \approx 2.6, G \approx 0.7, K \approx 0.7$, $G_D \approx 7.4$ and $G_A \approx 5.5$ obtained from the preliminary statistical analysis in Supplement A, Sections A.3–A.6 and A.8. The prior mean of the measurement noise σ^2 we have set, rather ad-hoc, to 50.

By the statistical analysis in Section A.8 of Supplement A, it was further found that the point process intensities θ_A and θ_D of the samples can be roughly related to the donor and acceptor solution concentrations applied for the sample preparation. Thereby, we found that for sample 1, $\theta_D \approx 2 \times 10^3/\mu m^2$ and $\theta_A \approx 10^3/\mu m^2$; while for sample 2, $\theta_D \approx 2 \times 10^3/\mu m^2$ and $\theta_A \approx 2 \times 10^3/\mu m^2$; and for sample 3, $\theta_D \approx 10^3/\mu m^2$

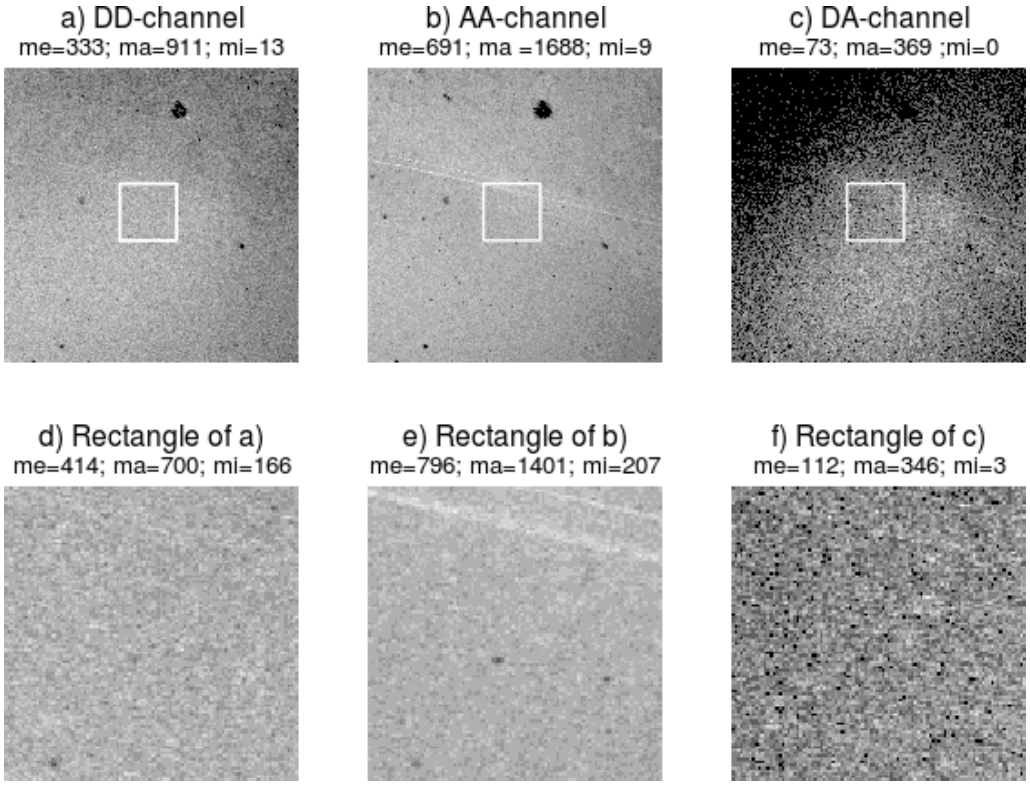


Figure 7: Channel intensity images of the aggregated channel dataset of sample 2. (a) DD-channel, (b) AA-channel, (c) DA-channel. Plots (a)–(c) consist each of 512×512 pixels. Plots (d)–(f) show enlargements of the square subregions of the plots (a)–(c), each consisting of 100×100 pixels. Above each plot is stated the mean (me), maximum (ma) and minimum (mi) pixel intensity value in the image. In each image the gray levels are constructed by using ten equally spaced intervals between zero and the maximum value of the image. Black/white refers to the lowest/highest intensity interval.

Table 2: Prior means for the microscope parameters and the Poisson point process intensities, as well as the values of the tuning parameter τ applied to generate proposals for the microscope parameters.

	Samples 1,2 and 3						Sample 1		Sample 2		Sample 3	
	M_D	G	K	G_D	G_A	σ^2	θ_D	θ_A	θ_D	θ_A	θ_D	θ_A
Prior mean	2.5	1.0	1.0	5.0	5.0	50.0	2e3	1e3	2e3	2e3	1e3	2e3
τ	0.05	0.1	0.1	0.5	0.5	1.0	—	—	—	—	—	—

and $\theta_A \approx 2 \times 10^3/\mu m^2$. We use these values as prior means for θ_A and θ_D for each of the samples. The applied prior means as used in the Bayesian analyses are summarized in Table 2.

In the MCMC computations we used random walk Metropolis-Hastings updates for the log microscope related parameters. The values of the random walk update standard deviations τ are also shown in Table 2. We tuned the τ values to get an approximately 30% proposal acceptance rate for each of the microscope parameters.

The total number of MCMC updates is 5×10^9 for each run. Metropolis-Hastings updates for the microscope parameters and Gibbs updates for the point process parameters are made after every 10^4 birth/death updates of donor or acceptor points.

The MCMC chains converge slowly due to bad mixing as discussed in Section 6.3. This means that we need many rounds of birth-death updates for the donor/acceptor points followed by updates of microscope and point process parameters. In each round we need to update a large fraction of the donors and acceptors. Thus, for a fixed fraction, each round takes more computing time the higher posterior expected number of donors and acceptors. The a posteriori expected number of donors and acceptors in each pixel is fairly high (of the order 300). To keep the computation time at an acceptable level, we therefore perform the inference on a small 10×10 subset of pixels which contain a posteriori of the order of 3×10^4 donors and acceptors.

6.3 Results of the inference

6.3.1 Assessment of MCMC samples

In Figure 8 the traceplots of the microscope parameters and the Poisson point process intensities for sample 1 are shown. Posterior mean values are displayed in the upper left corner of each of the plots. The traceplots indicate poor mixing of the MCMC samples except for G_A, G_D, and σ.

The poor mixing is due to high posterior correlation between certain parameters as visualized by the scatterplots in Figure 9 in which the posterior realizations of $M_D, G, K, \theta_D, \theta_A$ are plotted against each other. Especially M_D and θ_D and G and θ_A are highly correlated but fairly strong correlations are also evident between M_D and K and between K and θ_D. Similar scatterplots of G_D, G_A and σ^2 versus each of the other parameters (not shown), do not show any clear correlation with any of the parameters.

6.3.2 Posterior results

In Table 3 the 95% posterior intervals and posterior means for each of the microscope parameters and Poisson intensities are stated for each of the three samples. The microscope parameters M_D, G, G_A, G_D, and σ should be equal for all samples and this may seem contradicted by their posterior means that vary across samples. There is on the other hand considerable overlap between almost all 95% posterior intervals so the Bayesian inference does not contradict that the microscope parameters are equal across samples.

Figure 10 shows the posterior distributions of the centered L-function for the three samples. As expected, there is no indication of clustering nor repulsion since the centered posterior L-functions are close to zero and the posterior means are approximately zero.

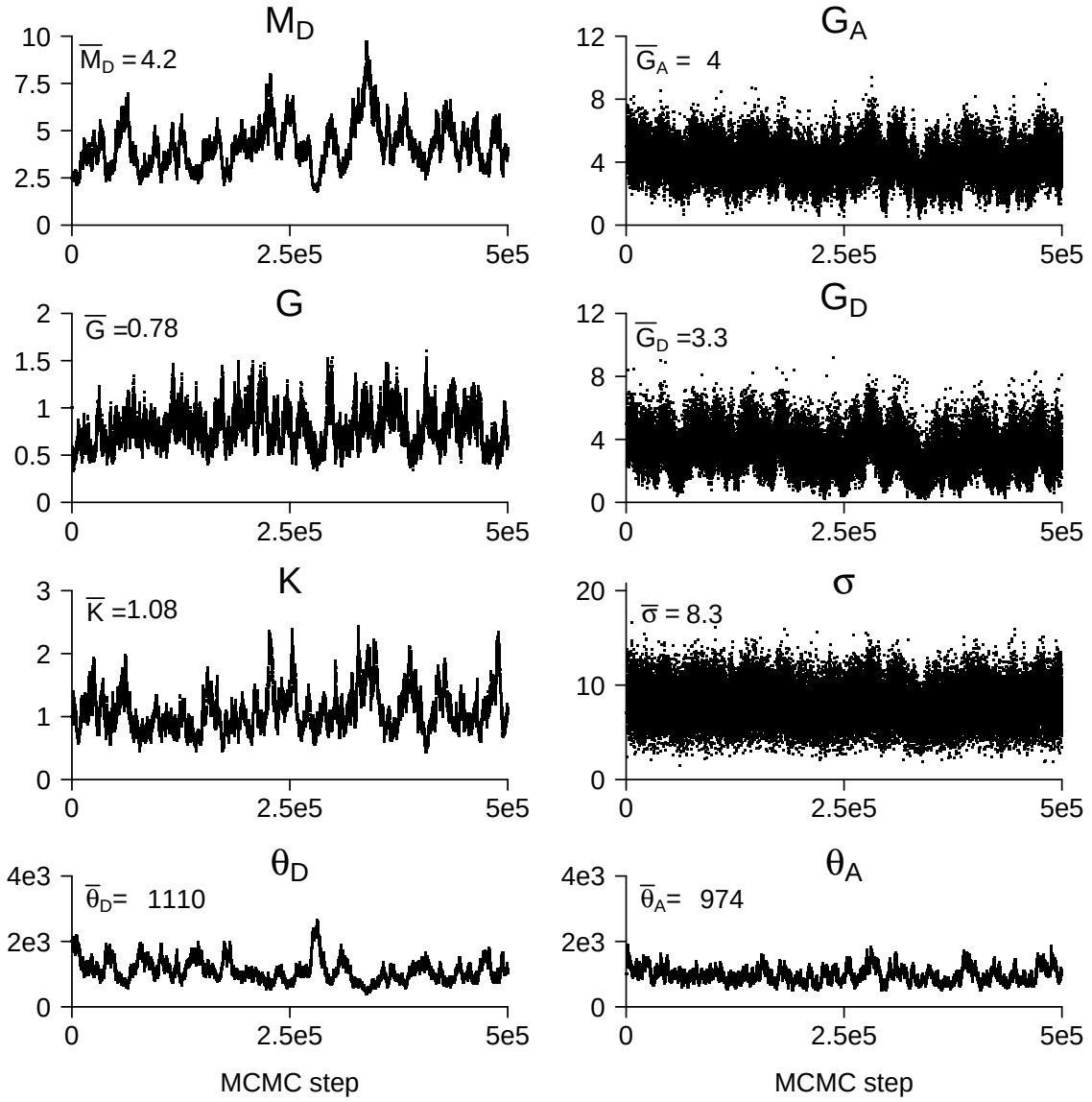


Figure 8: Traceplots of the microscope and the Poisson point process parameters for sample 1. Posterior mean values are displayed in the upper left corner of each of the plots. For plotting a subsampling of 10 has been applied.

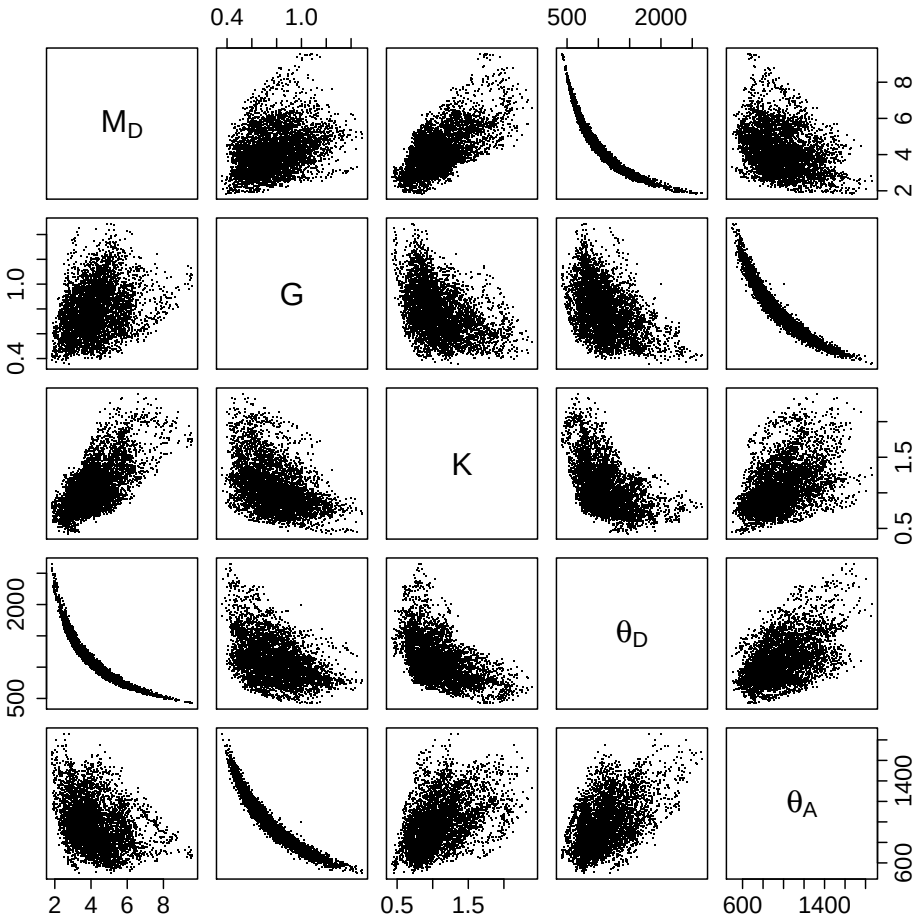


Figure 9: Scatterplots of the posteriors realizations of $M_D, G, K, \theta_D, \theta_A$ versus each other, for sample 1. For plotting a subsampling of 100 has been applied.

Table 3: Posterior results for each of the three samples: 95% posterior intervals and posterior means (in brackets) for each of the parameters. Prior means of the parameters are given in Table 2.

	Sample					
	1		2		3	
M_D	2.3–7.0	(4.2)	3.1–8.6	(5.5)	2.8–6.8	(4.7)
G	0.46–1.21	(0.78)	0.50–1.36	(0.89)	0.19–0.52	(0.32)
K	0.61–1.93	(1.08)	0.31–0.98	(0.60)	0.51–1.49	(0.92)
G_A	1.9–6.2	(4.0)	5.3–12.7	(8.9)	8.0–13.4	(10.6)
G_D	1.2–5.7	(3.3)	3.0–10.0	(6.3)	1.7–6.5	(3.9)
σ	4.4–12.0	(8.3)	3.8–10.7	(7.2)	4.2–11.6	(7.9)
θ_D	6.1e2–19.3e2	(11.1e2)	7.8e2–2282	(13.4e2)	9.0e2–23.5e2	(14.7e2)
θ_A	6.1e2–15.0e2	(9.7e2)	8.3e2–20.4e2	(12.8e2)	15.0e2–34.1e2	(23.5e2)

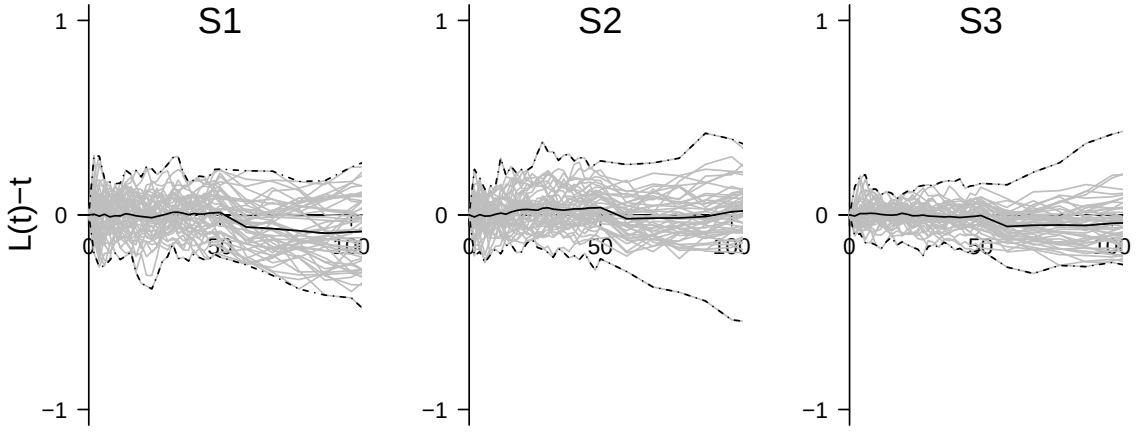


Figure 10: Posterior distribution of the L-function for samples 1–3. For each sample the posterior distribution is summarized by the posterior mean (solid line) of the L-function and 95% envelopes based on minimal (lower dashed-dotted line) and maximal (upper dashed-dotted line) values of 39 posterior realizations of the L-function.

7 Discussion

This paper presents a first attempt to implement likelihood based inference for FRET data. We thus, based on physical considerations, developed a realistic observational model for FRET data given the underlying configurations of donors and acceptors. Based on this model we proposed to implement Bayesian inference using MCMC.

We quantify spatial dependence by considering the posterior mean of the cross L-function for the donors and acceptors. Our simulation results show that the posterior mean of the L-function can be used to distinguish between clustering, absence of interaction and repulsion between donors and acceptors. Due to bias one needs to be careful when making quantitative statements regarding strength of interaction based on the posterior means of the L-functions. However, we believe that it is meaningful to make relative comparisons of strength of interactions between samples observed under the same experimental conditions and thus with same signal to noise ratios.

Partly due to poor mixing of the proposed MCMC procedure we were forced to consider only a small subset of the full data. A key objective for further research is therefore to obtain a more efficient MCMC scheme so that efficient use of the full data becomes feasible. Haario et al. (2001) suggest to use joint updates but they consider posterior distributions of fixed dimensional random vectors. However, preliminary experiments with this approach indicate that we need joint updates involving both the microscope parameters and the donor-acceptor point patterns. It is not clear how to do this. Our data example illustrated the use of reference data with no donor-acceptor interactions to infer the microscope related parameters. In future work it would interesting to apply an improved MCMC algorithm to conduct Bayesian inference for an experimental sample with possible interactions.

The Poisson point process prior for protein configurations was chosen partly for computational reasons. To implement Bayesian inference with more flexible Markov

point process priors allowing for both repulsive and attractive interactions requires more advanced Markov chain Monte Carlo methods developed in Møller et al. (2006) and Murray et al. (2006). However, these methods are highly computationally demanding since they involve so-called perfect simulation from the point process prior which can lead to unacceptable computing times in case of protein configurations of high cardinality which are frequently encountered for FRET data.

Acknowledgements

The project was supported by the Center for Stochastic Geometry and Advanced Bioimaging which is funded by the Villum Foundation. The research has been further supported by two grants of the Danish Council for Independent Research – Natural Sciences, grant 12-124675 "Mathematical and Statistical Analysis of Spatial Data" and grant 09-072331 "Point process modeling and statistical inference".

Appendix: Derivation of observation model

In this section we refer to notation introduced in Section 2. Let L_d denote the set of times in the observation time span $[0,T]$ where a donor d in $\mathbf{X}_D$ is excited by a photon from the laser. We assume that L_d is a homogeneous Poisson process on $[0,T]$ with intensity $\lambda_D > 0$. The process L_d can be decomposed as

$$L_d = L_{dE} \cup L_{dN} \bigcup_{a \in \mathbf{X}_A} L_{da}$$

where L_{dE} denotes the times of excitations of d which resulted in emission in the D channel, L_{dN} is the times of excitations resulting in non-radiative de-excitation and L_{da} denotes the times of excitations that resulted in FRET to acceptor a and subsequent emission in the A channel. The so-called quantum yield $0 < q_D < 1$ is the probability of emission for donors conditional on that de-excitation is by emission or non-radiatively, i.e. $q_D = k_{DE}/k_D$. Similarly, q_A denotes the quantum yield for acceptors. Invoking the random labelling theorem for Poisson processes (e.g. Proposition 3.7 in Møller and Waagepetersen, 2003), L_{dE}, L_{da}, and L_{dN}, are independent Poisson processes with intensities $\lambda_D q_D P_{dD}$, $\lambda_D q_A P_{da}$, and $\lambda_D[(1-q_D)P_{dD} + \sum_{a \in \mathbf{X}_A}(1-q_A)P_{da}]$.

For each $i \in \mathcal{G}$ we let

$$L^i_{DD} = \bigcup_{d \in \mathbf{X}_D \cap C_i} L_{dE}$$

and

$$L^i_{DA} = \bigcup_{d \in \mathbf{X}_D, a \in \mathbf{X}_A \cap C_i} L_{da}$$

be the Poisson processes of donor excitation times which result in photon emissions for respectively donors and acceptors in the pixel C_i.

The emitted photons fall on the detector independently of each other with a probability $0 < h < 1$ of detection. In the point process literature the detected

photons is called an independent thinning with retention probability h. Further, of the detected photons only independent thinnings with retention probabilities $0 < Q_D < 1$ ($0 < Q_A < 1$) are registrated in the donor (acceptor) channel of the detector. The probabilities Q_D and Q_A are respectively the detector quantum yields in the donor and acceptor channel (Pawley, 2006a). The detected photon counts N^i_{DD} and N^i_{DA} of emissions in the pixel C_i are thus Poisson distributed with means $Q_D h q_D \lambda_D T \sum_{d \in \mathbf{X}_D \cap C_i} P_{dD}$ and $Q_D h q_A \lambda_D T \sum_{d \in \mathbf{X}_D, a \in \mathbf{X}_A \cap C_i} P_{da}$. Finally, $I^i_{DD} = G_D N^i_{DD}$ and $I^i_{DA} = G_A N^i_{DA}$ where G_D and G_A are amplification factors depending on the detector and channel. Defining $M_D = G_D Q_D h q_D \lambda_D T$ and $G = (G_A Q_A q_A)/(G_D Q_D q_D)$ we arrive at the specified means of I^i_{DD} and I^i_{DA}.

The mean of I^i_{AA} is found in a similar fashion. In the AA-channel acceptors are directly excited by the laser which is now broadcasting in the acceptor excitation spectrum with an intensity $\lambda_A > 0$. An excited acceptor can only de-excite due to emission or non-radiatively and the detected photon counts N^i_{AA} of emissions in the pixel C_i are Poisson distributed with means $Q_A h q_A \lambda_A T \sum_{a \in \mathbf{X}_A \cap C_i} 1$ and $I^i_{AA} = G_A N^i_{AA}$. Defining $M_A = G_A Q_A h q_A \lambda_A T$ and $K = M_D/M_A$ we arrive at the specified mean of I^i_{AA}.

Supplement A

In this supplementary material we present a preliminary statistical analysis of the in vitro transferrin attached to polylysine slides three cube FRET data set. We start by describing the experimental set-up and sample preparation in Section A.1 and discuss the channel data set extracted from these samples in Section A.2. Then we study the influence that photobleaching of the donors and acceptors has on the intensity signal for the remeasurements in the three channels in Section A.3, and we present simple non-Bayesian methods for estimating the K and G factors in Section A.4 and A.5, respectively. Methods for estimating the G_A and G_D factors are presented in Section A.6 and an estimate of the ratio of the quantum efficiencies in the detector D- and A-channel is given in Section A.7. We conclude by obtaining an estimate of M_D in Section A.8.

A.1 The experimental set-up

Fourteen glass cover slips coated with polylysine have been prepared to contain various abundances of donor and acceptor fluorophores attached to it. Briefly, the procedure is as follows. Transferrin is labeled solely with donor fluorophores or solely with acceptor fluorophores, leading to Tfn-D and Tfn-A molecules. As donor the Alexa-488 fluorophore ($q_D = 0.92$) is used and as acceptor the Alexa-555 fluorophore ($q_A = 0.10$) is used. The Förster distance of the Alexa-488 and Alexa-555 fluorophore pair is 7 nm (Johnson, 2010). Solutions containing concentrations of approximately 2 µg/ml, 4 µg/ml or 8 µg/ml of solely Tfn-D or solely Tfn-A are prepared, and glass cover slips coated with polylysine are incubated for a certain period with Tfn-D and/or Tfn-A solution. The procedure followed to bind transferrin to polylysine plates is described in more detail in Wallrabe et al. (2006).

In Table A.1 the sample preparation set-up is shown. The samples 1–3 contain only donor fluorophores, while samples 4–6 consist of only acceptor fluorophores. The samples 7–14 consist of mixtures of donor and acceptor fluorophores. We note that the samples denoted as sample 1, 2 and 3 in the article correspond to, respectively, the samples 7, 9 and 11 in Table A.1.

Table A.1: Sample preparation set-up for transferrin attached to polylysine slides. The solution concentrations of Tfn-A ($[A]_{sol}$) and Tfn-D ($[D]_{sol}$) applied are stated for each sample in units of µg/ml.

	Sample number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$[A]_{sol}$	–	–	–	2	4	8	2	4	4	8	4	8	6	8
$[D]_{sol}$	2	4	8	–	–	–	4	8	4	8	2	4	2	2

A.2 The channel data set

Each sample (see Table A.1) has been measured at three sites. At each site three-cube FRET channel data is obtained on a square grid containing 512×512 square pixels. The pixel side length is 0.279 µm and the focal volume depth is approximately 5 pixels (1.4 µm) (Wallrabe et al., 2007).

In order to obtain sufficient photon count statistics – that is sufficiently high signal-to-noise ratio to apply our Bayesian inference method – each sample has been remeasured ten times.[1] We then create an aggregated dataset – the *aggregated channel dataset* – by summing pixel wise over the ten measured intensities for each channel. By remeasuring the sample instead of increasing the measurement time, we obtain information concerning: the amount of photobleaching occurring for remeasurements (see Section A.3) and the pixel intensity variance in the three channels. The latter information gives the possibility to compare the empirical pixel intensity mean-variance relationship with the one implied by our statistical model, as well as to obtain estimates for G_D and G_A (see Section A.6).

All three channels are corrected for background emission and the DA-channel is also corrected for spectral bleedthrough by the methods described in Elangovan et al. (2003). In the following, DA-channel data (Y_{DA}) refers to the intensities corrected for bleedthrough and background emission and the DD- and AA-channel data (Y_{DD} and Y_{AA}, respectively) to the intensities corrected for background emission. We will refer to a sample by its sample number (see Table A.1).

In Figure A.1 the channel intensity images of the aggregated channel dataset of sample 9 are shown in terms of gray levels. In plot (a) and (b) there are some very low intensity regions (black spots) and in the upper left corner of plot (c) most of the DA-channel intensities are zero. These artefacts are due to improper sample preparation. Therefore the statistical analysis as presented in the following sections are based on the channel data of the central square section consisting of 100×100 pixels (the white squares in plots (a)–(c).

In the following we will occasionally for any site and sample make use of the following generic notation. We denote the pixel intensity of pixel i measured in channel $k = DD, DA, AA$ and for measurement number $m = 1, \ldots, 10$ by $Y_k^{i,m}$. The *sample mean intensity* – taken over the 100×100 pixels – for channel k and measurement number m is denoted by $\bar{Y}_k^{\cdot m}$, that is $\bar{Y}_k^{\cdot m} = \frac{1}{10^4} \sum_{i=1}^{10^4} Y_k^{i,m}$. The *pixel mean intensity* for pixel i for the 10 remeasurements m for channel k will be denoted by $\bar{Y}_k^{i\cdot}$, that is $\bar{Y}^{i\cdot} = \frac{1}{10} \sum_{m=1}^{10} Y_k^{i,m}$. The *sample grand mean intensity* taken over the 100×100 pixels and the 10 remeasurements will be denoted by $\bar{\bar{Y}}_k = \frac{1}{10} \sum_{m=1}^{10} \bar{Y}^{\cdot m}$.

[1] For a typical three-cube FRET experiment less than 10 photons counts per pixel are registered by the detector in the DA-channel (Pawley, 2006b, Chapter 2; Clegg, 1996, Chapter 1).

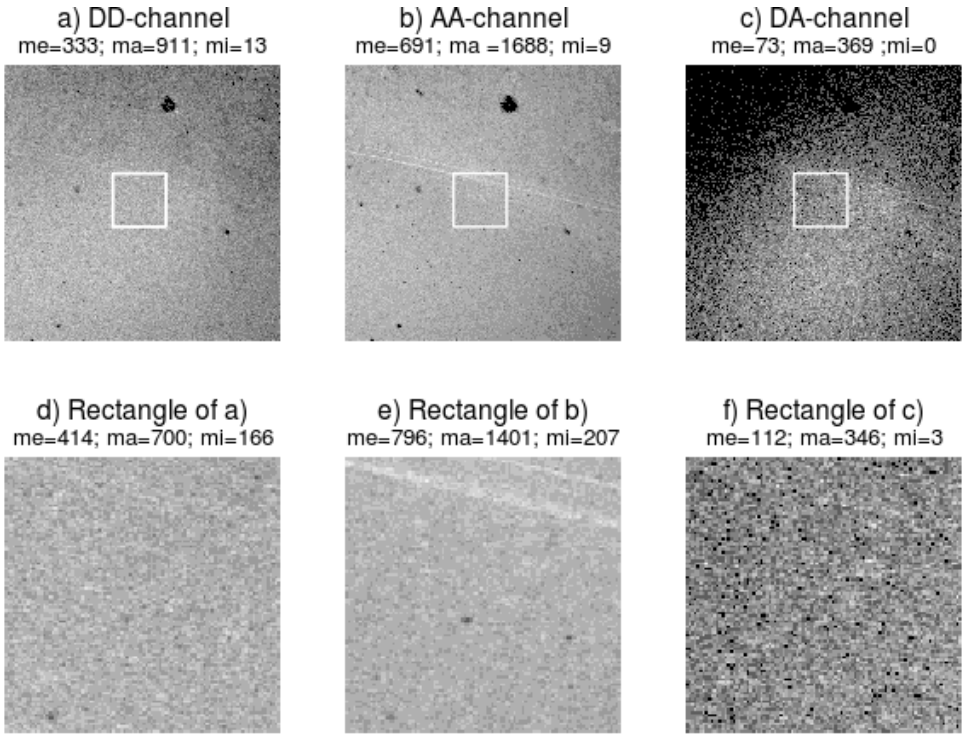


Figure A.1: Channel intensity images of the aggregated channel dataset of sample 9, site 2. (a) DD-channel, (b) AA-channel, (c) DA-channel. Plots (a)–(c) consist each of 512×512 pixels: Figures (d)–(f) show enlargements of the square subregions of the figures (a)–(c), each consisting of 100×100 pixels. Above each plot is stated the mean (me), maximum (ma) and minimum (mi) pixel intensity value in the image. In each image the gray levels are constructed by using ten equally spaced intervals between zero and the maximum value of the image. Black/white refers to the lowest/highest intensity interval.

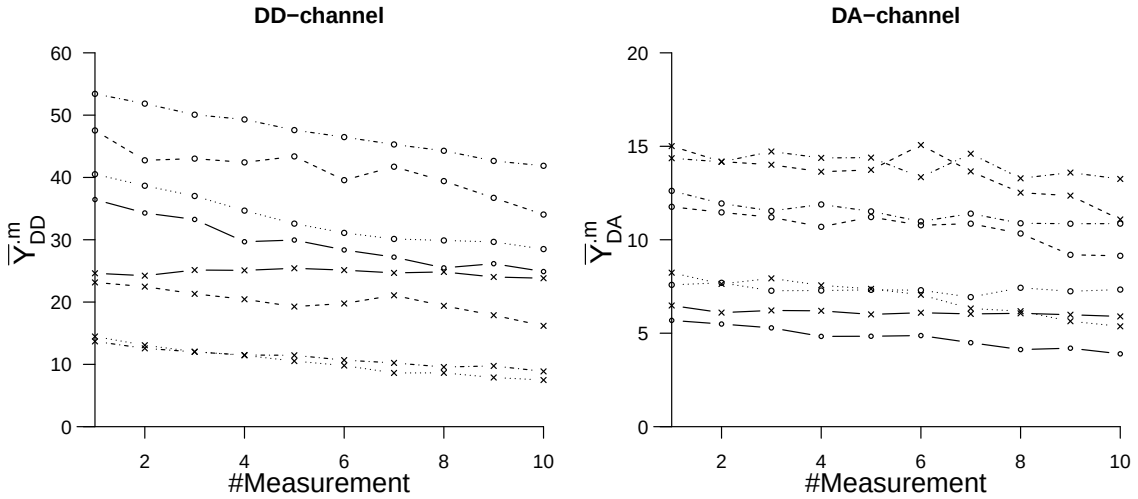


Figure A.2: Sample mean intensity versus measurement number shown for site 2 for the D+A samples: (left) *DD*-channel, (right) *DA*-channel. Solid-circle: sample 7 ; dashed-circle: sample 8; dotted-circle: sample 9; dashed-dotted-circle: sample 10; solid-triangle: sample 11; dashed-triangle: sample 12; dotted-triangle: sample 13; dashed-dotted-triangle: sample 14.

A.3 Photobleaching due to remeasuring

In Figure A.2 the sample mean intensity is shown as a function of the remeasurement number for the DD- and DA-channel, for site 2 of the D+A samples. Clearly, the sample mean pixel intensity has a slightly decreasing trend as a function of the measurement number. The decrease in intensity between measurement one and ten is roughly between 10 % to 30 % for all samples and for both channels. The same amount of decrease in intensity occurs for the AA-channel (not shown). The same analysis applied to the sites 1 and 3 of the D+A samples showed similar results as just discussed for site 2.

We note that the intensity remeasurements could be incorporated in our inference procedure by using the product of likelihoods for each measurement i.e. in equation (3.1) in the article we could replace each of the likelihood terms $p(y_k|\psi, \mathbf{x}), k = DD, DA, AA$ by $\prod_{m=1}^{M} p(y_k^m|\psi, \mathbf{x})$ with $M = 10$ the total number of measurements and y_k^m the observed intensity for channel k and measurement m. Thereby, a possible way to account for the decreasing linear trend in the pixel intensities due to bleaching is the following. Define $\mu_k^{i,1}$ – as in (2.2) and (2.3) in the article – to be the channel mean pixel intensity for measurement number 1 for pixel i and channel $k = DD, DA, AA$, and relate the mean pixel intensity $\mu_k^{i,m}$ for the remeasurements $m = 2, \ldots, 10$ to $\mu_k^{i,1}$ by

$$\mu_k^{i,m} = \mu_k^{i,1} - a(m-1)$$

and include a as a parameter in the model. For simplicity, however, we have chosen instead to create the aggregated photon count data set, to view it as resulting from one measurement and to apply our Bayesian inference procedure to this aggregated data set.

A.4 Estimate of the K-factor

The K-factor can be obtained experimentally by the preparation of a sample which contains equimolar concentrations of donor and acceptor fluorophores (Chen et al., 2006). Then by measuring the sample mean intensities in the three channels ($\bar{\bar{Y}}_{DD}$, $\bar{\bar{Y}}_{DA}, \bar{\bar{Y}}_{AA}$) an estimate for K is obtained by

$$K = \frac{\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G}{\bar{\bar{Y}}_{AA}}.$$

For a sample with unequimolar donor and acceptor concentrations [A] and [D] this relations becomes

$$K = \frac{[\mathrm{A}]}{[\mathrm{D}]} \frac{\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G}{\bar{\bar{Y}}_{AA}},$$

(Chen et al., 2006). The ratio $b_A = \bar{\bar{Y}}_{AA}/[\mathrm{A}]$ can be found as the slope of a regression line for pairs $([\mathrm{A}], \bar{\bar{Y}}_{AA})$ for samples with varying concentration $[A]$ and arbitrary donor concentration. Similarly, $b_D = (\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G)/[\mathrm{D}]$ can in principle be found as the slope of a regression line for pairs $([\mathrm{D}], \bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G)$. This, however, would

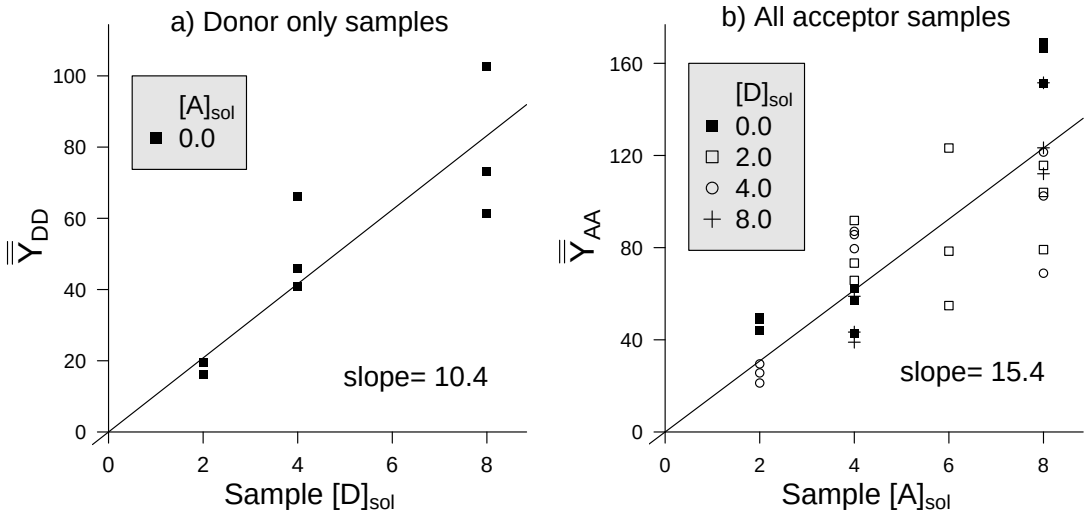


Figure A.3: Sample grand mean pixel intensity versus applied solution concentration. (a) DD-intensity vs donor solution concentration for each of the three sites of the donor only samples (samples 1–3), (b) AA-intensity vs acceptor sample concentration for for each of the three sites of the acceptor only (samples 4–6) and the D+A samples (samples 7–14).

require knowledge of G. In our data set, on the other hand, we have access to donor only samples and we can thus replace $\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G$ for a sample with both donors and acceptors of concentrations [D] and [A] with $\bar{\bar{Y}}_{DD}$ for a sample only containing donors of concentration [D].

In Figure A.3 (a) $\bar{\bar{Y}}_{DD}$ is plotted against $[\mathrm{D}]_{\mathrm{sol}}$ for the donor only samples 1–3.[2] The resulting slope of the regression line (of intercept 0) is $b_D = 10.6$.

Further, in Figure A.3 (b), $\bar{\bar{Y}}_{AA}$ is plotted versus the sample acceptor concentration $[\mathrm{A}]_{\mathrm{sol}}$ for acceptor only as well as D+A samples, and the slope b_A of the least squares line – found for a fixed intercept of zero – is 15.4. Thereby, we find as an estimate for the K-factor rounded to one decimal: $\hat{K} = b_D/b_A \approx 0.7$.

A.5 Estimate of the G-factor

The sample mean unquenched donor intensities for the D+A samples (samples 7–14) should scale with the applied sample donor solution concentration [D] by the same slope b_D as determined in the previous section. This observation provides a method to obtain a rough estimate for the G-factor.

Figure A.4 shows a similar plot as in Figure A.3 (a) where $\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}$ is plotted versus the applied donor solution concentration $[D]$ for the D+A samples. For this plot the slope of the least square estimate is 9.3, which is close to but slightly lower than the previous found estimate of $b_D = 10.4$. Since we have $\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G = b_D[D]$, this suggests a value of G less than one. We now simply tune G so that the regression for $\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G$ versus $[D]$ has the slope $b_D = 10.4$. This happens for $G = 0.66$,

[2]Because the true donor [D] and acceptor [A] concentrations in the samples are unknown, we use, respectively, the donor and acceptor concentrations $[\mathrm{D}]_{\mathrm{sol}}$ and $[\mathrm{A}]_{\mathrm{sol}}$ applied to prepare the samples, to find estimates for b_D and b_A.

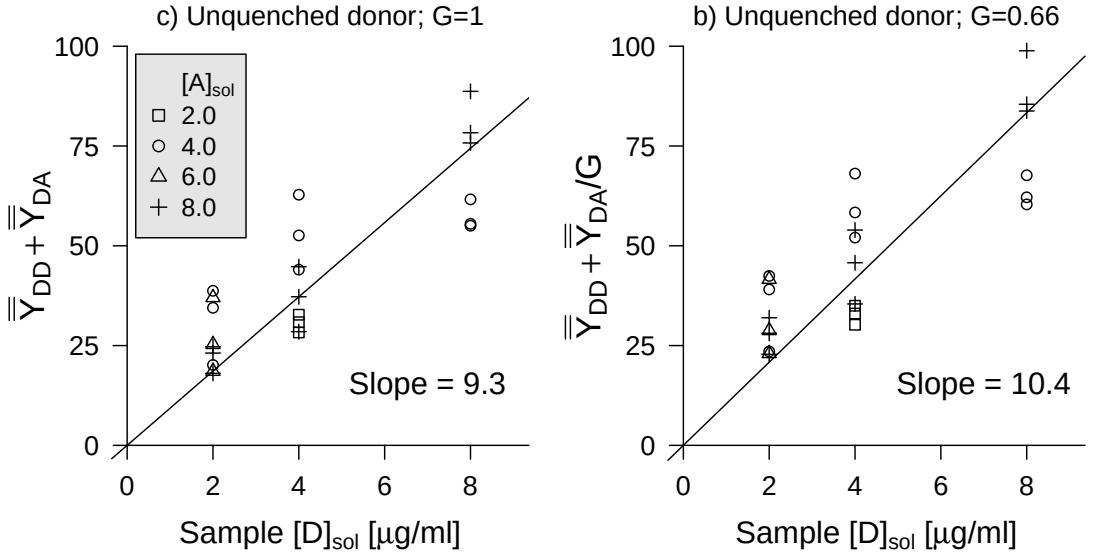


Figure A.4: Sample grand mean intensity of unquenched donor $\bar{\bar{Y}}_{DD} + \bar{\bar{Y}}_{DA}/G$ versus sample donor concentration for each of the three sites of the D+A samples (samples 7–14): (a) $G = 1$, (b) $G = 0.66$. In plot (b) the value of G has been tuned in order to obtain a least square line with a slope of 10.4. The least square lines are fitted in both plots with a fixed intercept of zero.

see Figure A.4 (b). We round this value to one decimal and find as an estimate for the G-factor, $\hat{G} = 0.7$.

A.6 Assessment of mean-variance relation for polylysine data

As each sample has been remeasured ten times, for pixel i we have observations $Y_k^{i,m}$ where $m = 1, \ldots, 10$ is the index of the replicates. Apart from a slight decrease in intensity due to bleaching (see Section A.3), we can view the $Y^{i,m}$ as independent and identically distributed. If we further ignore for the moment the additive noise, then the statistical model – equations (2.2)–(2.3) in the article – predicts a log-log linear relationship between the pixel mean μ_k^i and the pixel variance $\sigma_k^{2,i}$ of $Y_k^{i,m}$. E.g. for the DD-channel,

$$\log \sigma_{DD}^{2,i} = \log G_D + \log \mu_{DD}^i,$$

with G_D the amplification factor of the detector in the D-channel. Now, for each pixel we compute empirical means $\bar{Y}_k^{i\cdot} = \sum_{m=1}^{M} Y_k^{i,m}$ and empirical variances

$$s_k^{2,i} = \frac{1}{M-1} \sum_{m=1}^{M} \left(Y_k^{i,m} - \bar{Y}_k^{i\cdot} \right)^2, \qquad k = DD, AA$$

with $M = 10$ the total number of measurements. Figure A.5 shows for the two channels $k = DD, AA$ the empirical log variances $\log s_k^{2,i}$ against the log empirical

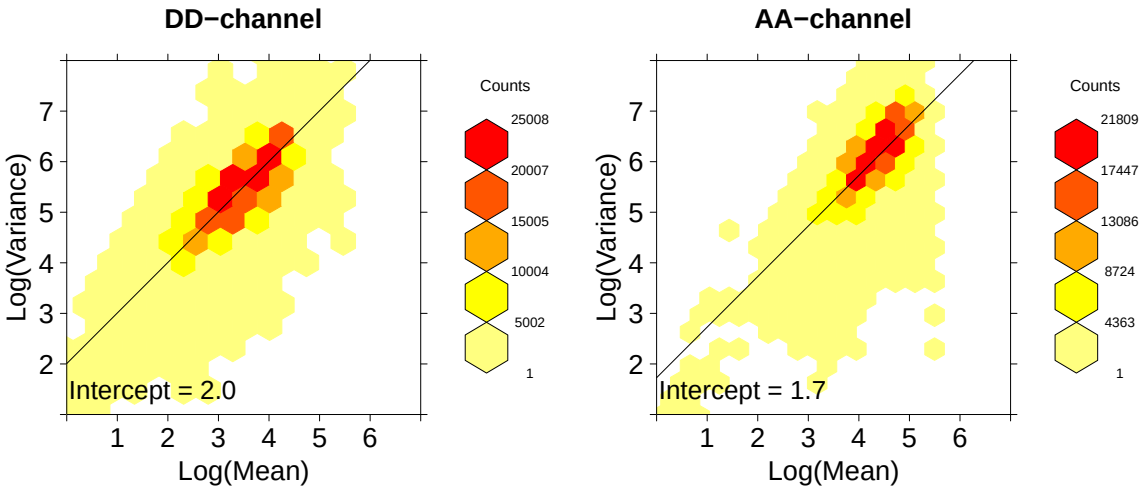


Figure A.5: Smoothed log-log scatterplot of the empirical pixel intensity variances $s_k^{2,i}$ versus the corresponding empirical pixel intensity means $\bar{Y}_k^{i\cdot}$. In both plots the least squares line is fitted for a fixed slope of one. For plot (a) the DD-channel intensity data is used for each of the three sites of the donor only samples (samples 1–3) and D+A samples (samples 7–14). For plot (b) the AA-channel intensity data is used for for each of the three sites of the acceptor only (samples 4–6) and D+A samples (samples 7–14). For further details we refer to the text.

means $\log \bar{Y}_k^{i\cdot}$. By fitting a regression line with a slope of 1 through the points, the intercept provides a rough estimate of respectively $\log G_D$ and $\log G_A$. The intercept is 2.0 for the DD-channel (plot (a)) and 1.7 for the AA-channel (plot (b)). Thereby we find, rounded to one decimal, the estimates $\hat{G}_D = \exp(2.0) \approx 7.4$ and $\hat{G}_A = \exp(1.7) \approx 5.5$.[3]

A.7 Estimate of the ratio of the detector quantum efficiencies

By the definition of the G-factor, as stated in the Appendix of the article (page 25),

$$\frac{Q_D}{Q_A} = \frac{G_A q_A}{G G_D q_D}.$$

Inserting the estimates for the parameters on the right-hand side ($\hat{G} = 0.7$, $\hat{G}_A = 5.5$, $\hat{G}_D = 7.4$) as well as the values for the quantum yield of the Alexa-488 donor fluorophore ($q_D = 0.92$) and of the Alexa-555 acceptor fluorophore ($q_A = 0.10$) gives that $Q_D/Q_A \approx 0.1$. So, the quantum efficiency of the applied detector in the A-channel is approximately 10 times larger than in the D-channel.

[3] We note that least square estimates for the slopes of the regression lines in Figure A.5 are 1.1 ± 0.2 for the DD- as well as the AA-channel, which is indeed close to one as predicted by the statistical model.

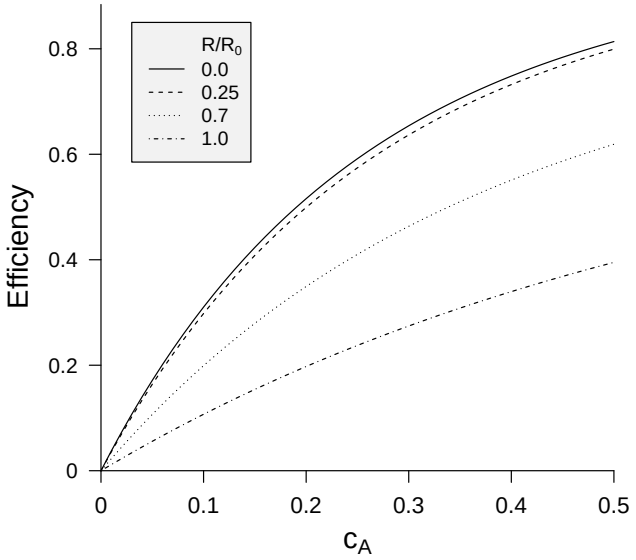


Figure A.6: Sample efficiency (A.1) for randomly distributed donors and acceptors in the plane as a function of c_A – i.e. the number of acceptors per area R_0^2 – for various ratios of R/R_0, with R_0 the Förster distance of the donor-acceptor pair and R the distance of closest approach between a donor and an acceptor. Adapted from Wolber and Hudson (1979).

A.8 Estimate of M_D

For randomly distributed donors and acceptors in the plane Wolber and Hudson (1979) provide an analytical expression for the sample mean FRET efficiency E as a function of: (1) the Förster distance R_0, (2) the acceptor concentration, and (3) the distance R of closest approach between a donor and an acceptor. Some of the limiting assumptions made in the derivation are: (a) donors do not compete with each other for transfer to an acceptor, (b) all donor-acceptor pairs have the same Förster distance and (c) the area contributed by the donors and acceptors is negligible i.e. there are no excluded area effects. The general solution is stated as an integral expression which should be solved numerically, but Wolber and Hudson (1979) also provide the following convenient and accurate approximation:

$$E = 1 - (A_1 e^{-k_1 c_A} + A_2 e^{-k_2 c_A}), \tag{A.1}$$

with A_1, A_2, k_1, k_2 constants depending on the ratio of R/R_0 (see Table I in Wolber and Hudson, 1979) and c_A is the acceptor concentration in units of the number of acceptors per area R_0^2. In Figure A.6 the efficiency as a function of the acceptor concentration is plotted for various R/R_0 ratios.

In the following we obtain an estimate for M_D by:

(i) applying (A.1) to find an estimate of c_A – and thereby of the mean number of acceptor points within a pixel – for the samples prepared with the highest acceptor solution concentration of $[\mathrm{A}]_{\mathrm{sol}} = 8$,

(ii) applying the equation for μ^i_{AA} in Section 2.2 of the article, to find an estimate of M_D

Regarding (i): The samples prepared with a solution concentration of $[\mathrm{A}]_{\mathrm{sol}} = 8$ are the samples 10, 12 and 14 (Table A.1). The sample mean efficiencies of sample 10, 12 and 14, averaged over the three sites are 0.32, 0.45 and 0.59. We believe that due to the high concentrations of donors in sample 10 (applied donor solution concentration is 8), the donors in this sample compete with each other for energy transfer to surrounding acceptors, which leads to a much lower value of the efficiency than for samples 10 and 12. We therefore here exclude sample 10 from the analysis.[4]

The average sample mean efficiency of sample 12 and 14 for the three sites is 0.52. Assuming the value of closest approach R to be zero then by aid of Figure A.6 we find that to $E = 0.55$ corresponds $c_A \approx 0.2$ [acceptors$/R_0^2$]. As for the Alexa-488 and Alexa-555 donor-acceptor pair $R_0 = 7\,\mathrm{nm}$, the average number of acceptors residing within a pixel of area $279\,\mathrm{nm} \times 279\,\mathrm{nm}$ for the samples 10 and 12 is approximately: $(0.2/(7^2)) \cdot 279^2 = 317.7 \approx 320$.

Regarding (ii): Summing the equation for μ_{AA}^i as stated in Section 2.2 of the article on the left and right hand-side over all pixels $i = 1, \ldots N$ on which our analysis is performed, i.e. the square subregion as displayed in Figure A.1 to which will we refer here as W, results in

$$\sum_{i=1}^{N} \mu_{AA}^i = \frac{M_D}{K} \sum_{a \in \mathbf{x}_A \cap W} 1,$$

where $N = 10^4$. Rewriting yields

$$M_D = \frac{K}{n(\mathbf{x}_A \cap W)} \sum_{i=1}^{N} \mu_{AA}^i, \tag{A.2}$$

with $n(\mathbf{x}_A \cap W)$ denoting the total number of acceptor points in W. We find an estimate of M_D by finding estimates for $\sum_{i=1}^{N} \mu_{AA}^i$ and $n(\mathbf{x}_A \cap W)$ in (A.2) for the samples 12 and 14.

In Section A.4 it was determined that the sample grand mean acceptor intensity $\bar{\bar{Y}}_{AA}$ scales with the sample acceptor concentration $[\mathrm{A}]_{\mathrm{sol}}$ by the slope $b_A = 15.4$. Thus for samples 12 and 14 we find that $\hat{\bar{\bar{Y}}}_{AA} \approx 15.4 \cdot 8 = 123.2$.

Because the Bayesian inference method applied in the article makes use of the aggregated data set summed over the ten measurements $m = 1, \ldots, 10$, the sample mean of the aggregated AA-intensity for the samples 12 and 14 is – ignoring the photobleaching effect – approximately $10 \cdot 123.2 = 1232$, which provides an estimate of $\frac{1}{N} \sum_{i=1}^{N} \mu_{AA}^i$ for the aggregated data set. Further, the in this section found estimate for the the mean number of acceptors per pixel of 320 for the samples 12 and 14 is an estimate of $\frac{1}{N} n(\mathbf{x}_A \cap W)$. Applying the not rounded estimate of $\hat{K} \approx 0.675$ previously found in Section A.4, we find as an estimate for M_D, rounded to one decimal,

$$\hat{M}_D \approx 0.675 \cdot \frac{1232}{320} = 2.6.$$

[4]For a random distribution of donors and acceptors and under the assumption that donors do not compete for energy transfer to an acceptor the sample efficiency is independent of the donor concentration for a fixed acceptor concentration (Kenworthy and Edidin, 1998)

We note that the estimate of a mean number of 320 acceptors per pixel for the samples prepared with an acceptor solution concentration of $8\,\mu g/ml$ is equally valid for donor fluorophores, i.e samples prepared with a donor solution concentration of $[D]_{sol} = 8\,\mu g/ml$ will contain approximately 320 donors per pixel. Further, the two results can be extrapolated, i.e. the average numbers of acceptors and donors within a pixel are, respectively $40 \cdot [A]_{sol}$ and $40 \cdot [D]_{sol}$. We use the latter relations to specify the prior means of the Poisson point process intensities in Section 6.2 in the article.

Supplement B

In this supplementary material we give a detailed description of the Markov chain Monte Carlo sampler used to draw samples from the posterior distribution (equation (3.1) in the article) in the first section. In the second section we briefly discuss the likelihood expressions we apply in the case the three-cube FRET channel intensities can not consist of negative values (i.e. empirical data).

B.1 Steps in the MCMC sampler

The steps involved in the sampler are illustrated in Figure B.1 and are described in detail below.

Step 1: Generate initial configuration for the posterior point pattern. An initial point pattern $\mathbf{X}_0 = (\mathbf{X}_D^0, \mathbf{X}_A^0)$ from which the chain starts is generated from the prior distribution by using the R-software (R Core Team, 2014) and the package `spatstat` (Baddeley and Turner, 2005). The initial point pattern is simulated on a square W corresponding to the area covered by the three-cube FRET channel data. The generated point pattern is stored by writing the coordinates x, y of the points as well as the marks of each point – 1 for a donor, 2 for an acceptor – to disk.

Step 2: Generate channel data from the initial posterior point pattern. Channel data conditional on the point pattern $\mathbf{X}_0$ is computed based on equations (2.1)–(2.3) in the article. To employ these equations the square area W is defined as a union of square pixels C_i, i.e. $W = \bigcup_{i\in\mathcal{G}} C_i$, with the C_i equal to the pixels in the three-cube FRET channel dataset indexed by $\mathcal{G}$.

To compute the μ_{DD}^i and μ_{DA}^i (defined in Section 2.2 of the article) the key component is to specify P_{da} for donor and acceptor pairs in $\mathbf{X}$. To keep the computation of transfer probabilities feasible, only those acceptors that reside within $4R_0$ of a donor are taking into account as a possible path for energy transfer for the donor. This important simplification will not lead to any significant difference in posterior results. By replacing P_{dA} in the equation for μ_{DD}^i by its definition $P_{dA} = \sum_{a\in\mathbf{X}_A} P_{da}$, and taking into account the cut-off radius of $4R_0$, the equations for μ_{DD}^i and μ_{DA}^i become

$$\mu_{DD}^i = M_D \sum_{d\in\mathbf{X}_D\cap C_i} \big(1 - \sum_{\substack{a\in\mathbf{X}_A \\ r_{da}<4R_0}} P_{da}\big), \tag{B.1}$$

$$\mu_{DA}^i = M_D G \sum_{a\in\mathbf{X}_A\cap C_i} \sum_{\substack{d\in\mathbf{X}_D \\ r_{da}<4R_0}} P_{da}, \tag{B.2}$$

with r_{da} the distance between a donor d and an acceptor a and (as previously defined

in Section 2.1 of the article)

$$P_{da} = \frac{(R_0/r_{da})^6}{1+S_d} \quad \text{with} \quad S_d = \sum_{\substack{\tilde{a}\in\mathbf{X}_A \\ r_{d\tilde{a}}<4R_0}} (R_0/r_{d\tilde{a}})^6.$$

For clarity and later use we further restate here the equation for μ^i_{AA} as previously defined in Section 2.2 of the article

$$\mu^i_{AA} = M_D/K \sum_{a\in\mathbf{X}_A\cap C_i} 1. \tag{B.3}$$

To benefit from the approach of excluding transfer probabilities P_{da} for donor and acceptors pairs which are further than $4R_0$ from each other, it is necessary to store donors and acceptors pixel wise. We have used so called *linked lists* to implement this. A linked list can be viewed as a list containing boxes and each box stores the values of some variables as well as a *pointer* to the next box. In our program for each pixel there are two linked-lists available, one storing the information concerning the donor points within the pixel and the other storing the information concerning the acceptor points within the pixel. To each point (donor or acceptor) corresponds precisely one box in the corresponding linked list. For an acceptor point a the box contains the x, y coordinates of a. For a donor point d also the value of S_d is stored. Storing of S_d gives the possibility to compute the channel data very efficiently when a proposal update for the acceptor point pattern is made. This is further discussed under step 6.

In our program the sums on the right of the factors M_D, M_DG and M_D/K in, respectively, equations (B.1), (B.2) and (B.3) are available and stored at every step in the three matrices XNDD, XNDA and XNA containing the elements

$$\begin{aligned}
\mathrm{XNDD}^i &= \sum_{d\in\mathbf{X}_D\cap C_i} (1 - \sum_{\substack{a\in\mathbf{X}_A \\ r_{da}<4R_0}} P_{da}) \\
&= \sum_{d\in\mathbf{X}_D\cap C_i} (1+S_d)^{-1}, \\
\mathrm{XNDA}^i &= \sum_{a\in\mathbf{X}_A\cap C_i} \sum_{\substack{d\in\mathbf{X}_D \\ r_{da}<4R_0}} \frac{(R_0/r_{da})^6}{1+S_d}, \\
\mathrm{XNA}^i &= \sum_{a\in\mathbf{X}_A\cap C_i} 1,
\end{aligned}$$

for $i = 1, \ldots, n$ and n the total number of pixels. The X in these names refers to the posterior pattern $\mathbf{X}$ and the N in XNDD and XNDA stands for *normalized* as the values of XNDD^i and XNDA^i correspond to the situation where every donor within pixel i is excited exactly one time. For XNA it is appropriate to think of the N in its name to refer to *number* – and we have used only a single A at the end of the name – as XNA is the matrix that stores the number of acceptors within each pixel of the pattern $\mathbf{X}_A$. Equivalent to the latter matrix, also a matrix XND is available within

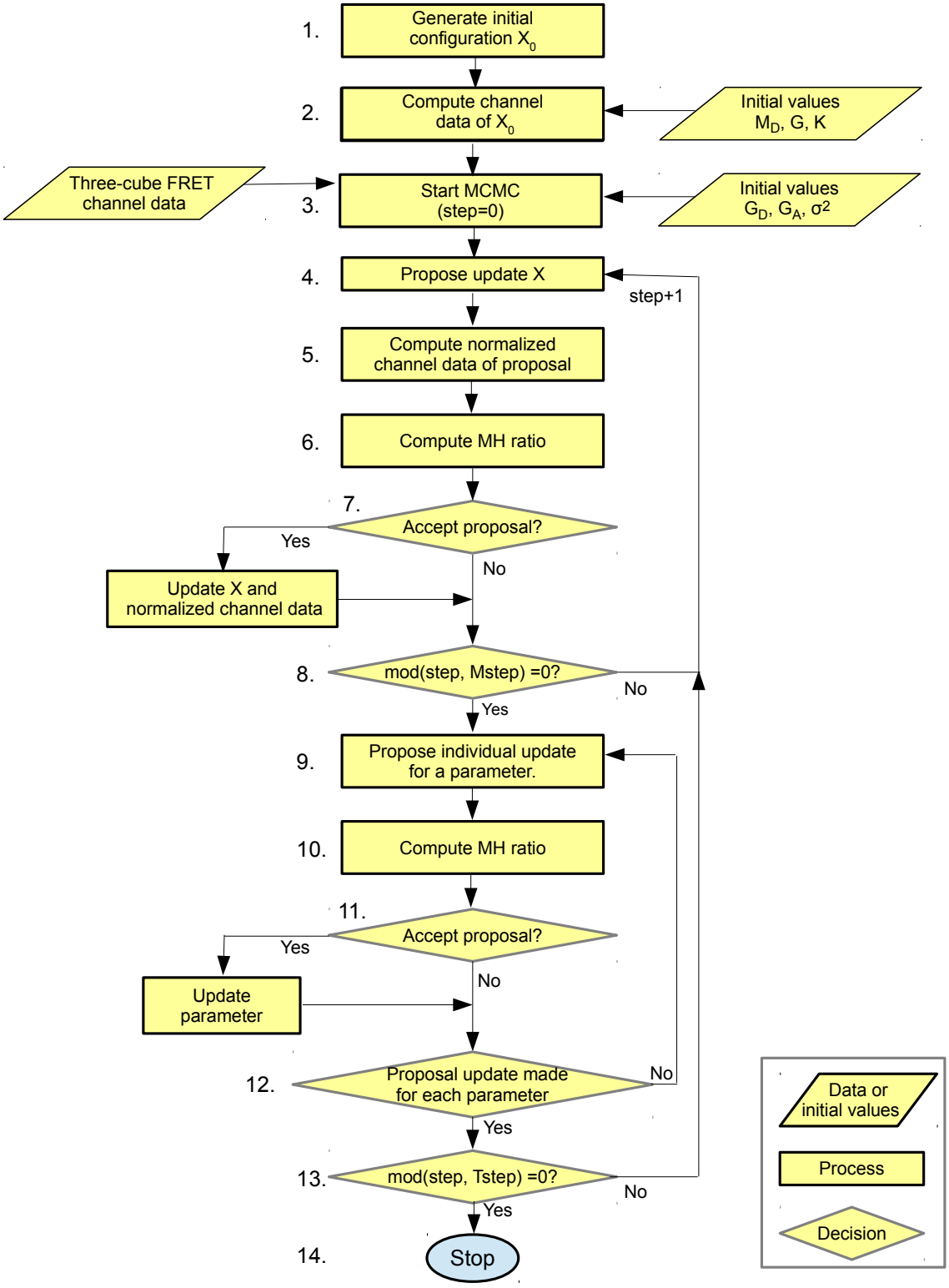


Figure B.1: Flow chart of the steps involved in the MCMC sampler.

the program which stores the number of donors within each pixel in the pattern $\mathbf{X}_D$. The values of μ^i_{DD}, μ^i_{DA} and μ^i_{AA} are now available at each step by multiplication of XNDD^i, XNDA^i and XNA^i by, respectively, the factors: M_D, $M_D G$ and M_D/K, as specified in (B.1), (B.2) and (B.3).

For the initial posterior point pattern $\mathbf{X}_0$ the value of each XNDD^i, XNDA^i and XNA^i is now computed as follows:

1. The initial point pattern is read from disk and depending on the mark (1 or 2) and the coordinates x, y a point is added to the corresponding donor or acceptor linked lists. For each donor point the value of S_d is set to zero. Within the process the total number of acceptors as well as donors that reside within a pixel i is counted and these values are stored, respectively, in the matrix elements XNA^i and XND^i.
2. An element XNDD^i (initially set to zero) is computed by looping over all donors d within pixel i. For each donors d the distance r_{da} to acceptors a which reside in the same pixel or directly neighboring pixels is calculated.[5] If the distance r_{da} is within $4R_0$, then the corresponding value of $(R_0/r_{da})^6$ is added to S_d (which initially is set to zero). After looping over all acceptors residing in the same or directly neighboring pixels the value S_d is stored and the value of $(1 + S_d)^{-1}$ is added to XNDD^i.
3. An element XNDA^i (initially set to zero) is computed by looping over all acceptors a within pixel i and for each acceptor a the distance r_{da} to donors d which reside in the same pixel or directly neighboring pixels is calculated. If the distance r_{da} is within $4R_0$, then the corresponding value of $P_{da} = (R_0/r_{da})^6/(1 + S_d)$ is added to XNDA^i.

Step 3: Start MCMC. By the initialization procedure, steps 1–2, and specifying initial values for the microscope parameters the following information is available within the program:

1. the point pattern $\mathbf{X}_0$ is stored as two sets of linked lists containing the coordinates of the donor and acceptor points.
2. the mean channel intensities $\mu^i_{DD}, \mu^i_{DA}, \mu^i_{AA}$ related to the initial point pattern $\mathbf{X}_0$ are available in the form of the three matrices XNDD, XNDA and XNA and the initial values of M_D, G and K.

Further, by specifying the prior distributions of the microscope and point process parameters and making a three-cube FRET data set available in the form of three matrices YDD, YDA and YAA containing the intensity values of respectively the DD-, DA- and AA-channel, the sampling procedure can start.

Step 4: Propose update pattern. A proposal is made to update the point pattern. With probability 1/2 the donor pattern is updated, otherwise the acceptor pattern is updated. Then it is proposed to add or remove a point, each with probability 1/2.

[5]For square pixels the approach of including only points residing within nearest neighbor pixels in the computation is valid for pixels with a side length $\geq 4R_0$. A condition always satisfied for empirical FRET datasets.

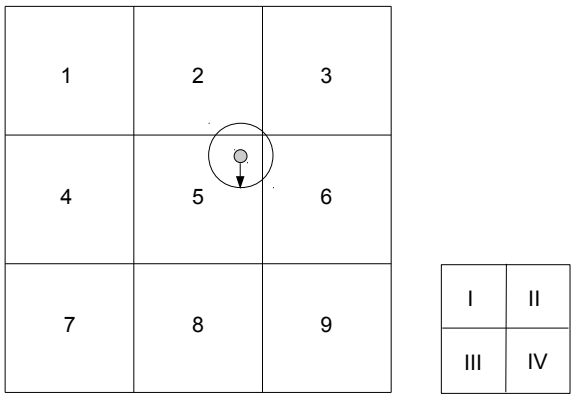


Figure B.2: Adding a donor d_0 (small gray circle) to the point pattern. The larger circle represents the radius of $4R_0$ around the donor. The figure is drawn on a scale with $R_0 = 6$ and the side length of the square pixels equal to 100. The small figure on the right shows how the quadrants within pixel 5 are labeled as used in the text.

- In the case a point is added, random coordinates for x and y are generated from the uniform distribution on W.
- In the case a donor (acceptor) point is removed from the pattern, a random integer is drawn between 1 and the total number of donor (acceptor) points in the current pattern. It is then proposed to remove the donor (acceptor) point that is labeled by this number. The latter method is implemented by implicitly labeling each donor and acceptor point. As an example, consider for instance a donor point which is stored in linked list number 4 box 8. This point is implicitely labelled by the integer value resulting from adding 8 (of box 8) to the total number of donors in linked lists 1 to 3.

Step 5: Calculate channel data for a proposal update of the point pattern. In order to gain computational speed, we have implemented a rather sophisticated procedure to compute the channel data for a proposal update of the point pattern. The implemented procedure distinguishes between computation of the channel data when an update for the donor pattern $\mathbf{X}_D$ or the acceptor pattern $\mathbf{X}_A$ is proposed. First, we discuss the algorithm for the donor case in terms of proposing to add a donor (proposing to remove a donor is similar). Secondly, we discuss the algorithm for the acceptor case in terms of proposing to add an acceptor (proposing to remove an acceptor is similar).

Compute proposal channel data: Adding a donor. In Figure B.2 a schematic representation of adding a donor d_0 (small gray circle) to the current pattern is shown. From the figure it is clear that if the new donor is placed in the second quadrant of a pixel only acceptors that reside within this same pixel, or in the pixels $2, 3$ and 6 can be affected by the newly placed donor.[6] Thereby, the only values of

[6]For square pixels the approach of only including acceptors within pixels which are nearest neighbors of the quadrant where the donor is added is valid for pixels with a side length $\geq 8R_0$. A condition always satisfied for empirical FRET datasets.

XNDA^i that need to be recomputed are for $i \in \{2,3,5,6\}$. The new value XNDA_p^i for each of the possibly affected XNDA^i is derived by computing the change $\Delta\mathrm{XNDA}^i$ that arises due to adding d_0, and adding it to the current value XNDA_c^i, i.e.

$$\mathrm{XNDA}_p^i = \mathrm{XNDA}_c^i + \Delta\mathrm{XNDA}^i.$$

The value of each $\Delta\mathrm{XNDA}^i$ for $i \in \{2,3,5,6\}$ is defined by

$$\Delta\mathrm{XNDA}^i = \sum_{\substack{a \in \mathbf{X}_A \cap C_i \\ r_{d_0 a} < 4R_0}} P_{d_0 a},$$

and easily obtained by: (i) computing S_{d_0} for donor d_0 by looping over the acceptors in the pixels $i \in \{2,3,5,6\}$, while storing $r_{d_0 a}$ for those a's for which $r_{d_0 a} < 4R_0$ and (ii) adding $P_{d_0 a}$ for each stored a to $\Delta\mathrm{XNDA}^i$ (which initially was set to zero) when a is located in pixel i.

The sole element of XNDD that is affected by the new donor is the element corresponding to the pixel where the donor is added, that is $i = 5$. The new proposed value for XNDD^5 is available as

$$\mathrm{XNDD}_p^5 = \mathrm{XNDD}^5 + \Delta\mathrm{XNDD}^5,$$

with

$$\Delta\mathrm{XNDD}^5 = 1 - \sum_{i \in \{2,3,5,6\}} \Delta\mathrm{XNDA}^i.$$

Compute proposal channel data: Adding an acceptor. Computing the channel data when an acceptor is added to the current pattern is rather more involved than for the donor case. We explain the procedure by example and in direct relation to the situation depicted in Figure B.3 in which an acceptor a_0 is added to a current pattern. The new acceptor point is placed in the second quadrant of pixel 5 and only donors that reside within the pixels $2, 3, 5$ and 6 can be affected by the new acceptor. Donors that reside within the solid circle are affected by the presence of the acceptor as they get an extra path for de-excitation by energy transfer to the newly placed acceptor. Now assume that the donor d_1 (small solid circle in pixel 3) which resides within the radius of $4R_0$ to a_0, has currently the possibility of energy transfer to a number of acceptors a ($r_{da} < 4R_0$) and denote one of these acceptors by a_1. Then in the current situation (a_0 not added) the probability of energy transfer from donor d_1 to acceptor a_1 is defined by

$$P_{d_1 a_1}^{\text{current}} = \frac{(R_0/r_{d_1 a_1})^6}{1 + S_{d_1}},$$

while in the proposed situation (acceptor a_0 added) this becomes

$$P_{d_1 a_1}^{\text{proposed}} = \frac{(R_0/r_{d_1 a_1})^6}{1 + S_{d_1} + (R_0/r_{d_1 a_0})^6}.$$

So, by placing a_0 within $4R_0$ of d_1, the probability $P_{d_1 a_1}$ for energy transfer from donor d_1 to acceptor a_1 decreases. Clearly, also for other acceptors residing within

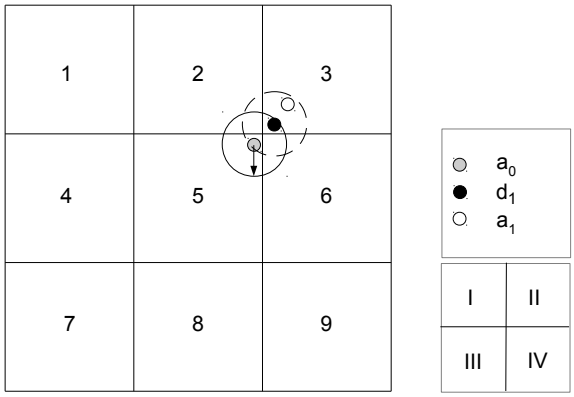


Figure B.3: Adding an acceptor a_0(small gray circle) to the point pattern. The larger circles around a_0 and d_1 (small solid circle) have a radius of $4R_0$. The figure is drawn on a scale with $R_0 = 6$ and the side length of the square pixels equal to 100. The small figure on the right shows how the quadrants within pixel 5 are labeled as used in the text

$4R_0$ of d_1 (other than a_0 and a_1), their respective probabilities to receive energy transfer from donor d_1 will decrease when acceptor a_0 is added. The algorithm for updating the possibly affected elements of XNDA is now as follows:

1. Compute for donors residing within possibly affected pixels, i.e. $d \in \mathbf{X}_D \cap C_i$, $i \in \{2, 3, 5, 6\}$, the distance r_{da_0} to a_0. If $r_{da_0} < 4R_0$, then store for each of these donors its coordinates as well as the value of $(R_0/r_{da_0})^6$ in a list.

2. Loop pixel wise over the acceptors residing in the possibly affected pixels, i.e. $a \in \mathbf{X}_A \cap C_i$, $i \in \{2, 3, 5, 6\}$.[7] Determine for each acceptor a the distance r_{da} to any of the donors in the list made in 1. If the distance $r_{da} < 4R_0$, then the value of XNDA_c^i will decrease due to a and d by

$$\frac{(R_0/r_{da})^6}{1 + S_d + (R_0/r_{da_0})^6} - \frac{(R_0/r_{da})^6}{1 + S_d}.$$

   Taking into account all the donors and acceptors of which the transfer probabilities change due to adding a_0, the algorithm computes ΔXNDA^i as

$$\Delta\text{XNDA}^i = \sum_{a \in \mathbf{X}_A \cap C_i} \sum_{\substack{j = \\ 2,3,5,6}} \sum_{\substack{d \in \mathbf{X}_D \cap C_j \\ r_{da_0} < 4R_0 \\ r_{da} < 4R_0}} \Big(\frac{(R_0/r_{da})^6}{1 + S_d + (R_0/r_{da_0})^6} - \frac{(R_0/r_{da})^6}{1 + S_d} \Big),$$

   for $i \in \{2, 3, 5, 6\}$. The values for ΔXNDA^i obtained by this equation are negative or zero as they compute the total decrease in the probability of energy transfer from donors d to acceptors $a \in \mathbf{X}_A \cap C_i$, due to adding a_0 to the current pattern. However, in the pixel where a_0 is added (pixel 5), there will

[7]For square pixels the approach of only including acceptors within pixels which are nearest neighbors of the quadrant where the proposed acceptor is placed, is valid for pixels with a side length $\geq 16R_0$. A condition in general satisfied for empirical three-cube FRET datasets.

be a possible increase in ΔXNDA, as a_0 can receive transfers from donors d possibly residing within $4R_0$ of it, which has to be added to ΔXNDA^5, i.e.

$$\Delta\text{XNDA}^5 = \Delta\text{XNDA}^5(\text{previous equation}) + \sum_{\substack{j= \\ 2,3,5,6}} \sum_{\substack{d\in\mathbf{X}_D\cap C_j \\ r_{da_0}<4R_0}} \left(\frac{R_0}{r_{da_0}}\right)^6.$$

Computation of the difference ΔXNDD^i between the current values XNDD^i and proposed values XNDD^i_p for the possibly affected elements $i \in \{2,3,5,6\}$ is much simpler than for ΔXNDA^i. Because the current value of S_d is stored for each donor d, no looping over the acceptors within the current pattern has to be carried out and

$$\Delta\text{XNDD}^i = \sum_{\substack{d\in\mathbf{X}_D\cap C_i \\ r_{da_0}<R_0}} \frac{1}{1+S_d+(r_{da_0}/R_0^6)} - \frac{1}{1+S_d}; \qquad i\in\{2,3,5,6\}.$$

We note that when an acceptor is added (removed) to (from) the current pattern this leads to to a redistribution of normalized pixel intensities between the DD- and DA-channel. Therefore the following equality holds

$$\sum_{i\in\{2,3,5,6\}} \Delta\text{XNDD}^i + \sum_{i\in\{2,3,5,6\}} \Delta\text{XNDA}^i = 0. \tag{B.4}$$

This observation provides a convenient way to check the proper implementation of the algorithms used to compute ΔXNDA^i and ΔXNDD^i. Computing both sums on the left-hand side of (B.4) and adding them should give the value of zero within numerical precision.

Step 6: Calculate Metropolis-Hastings ratio. We present here the steps involved to calculate (the logarithm of) the MH-ratio (Section 3.2 of the article) in terms of removing an accepter. The computation when adding an acceptor or adding/removing a donor is similar.

In case it is proposed to remove a point $u \in \mathbf{X}_A$, the Metropolis-Hastings ratio becomes

$$\frac{p(y|\mathbf{x}_D,\mathbf{x}_A\backslash\{u\},\psi)}{p(y|\mathbf{x}_D,\mathbf{x}_A,\psi)}\,\frac{p(\mathbf{x}_A\backslash\{u\}|\theta_A)}{p(\mathbf{x}_A|\theta_A)}\,\frac{n(\mathbf{x}_A)}{\theta_A|W|},$$

with the logarithm of first two ratios specified below:

1. the likelihood ratio term

$$\begin{aligned}
&\ln\left[\frac{p(y|\mathbf{x}_D,\mathbf{x}_A\backslash\{u\},\psi)}{p(y|\mathbf{x}_D,\mathbf{x}_A,\psi)}\right] \\
&= \frac{1}{2}\sum_{i\in N_c}\Big(\ln\left[\frac{G_D\mu^{i\mathbf{c}}_{DD}+\sigma^2}{G_D\mu^{i\mathbf{p}}_{DD}+\sigma^2}\right] + \frac{(y^i_{DD}-\mu^{i\mathbf{c}}_{DD})^2}{G_D\mu^{i\mathbf{c}}_{DD}+\sigma^2} - \frac{(y^i_{DD}-\mu^{i\mathbf{p}}_{DD})^2}{G_D\mu^{i\mathbf{p}}_{DD}+\sigma^2} \\
&\qquad + \ln\left[\frac{G_A\mu^{i\mathbf{c}}_{DA}+\sigma^2}{G_A\mu^{i\mathbf{p}}_{DA}+\sigma^2}\right] + \frac{(y^i_{DA}-\mu^{i\mathbf{c}}_{DA})^2}{G_A\mu^{i\mathbf{c}}_{DA}+\sigma^2} - \frac{(y^i_{DA}-\mu^{i\mathbf{p}}_{DA})^2}{G_A\mu^{i\mathbf{p}}_{DA}+\sigma^2} \\
&\qquad + \ln\big[\frac{G_A\mu^{i\mathbf{c}}_{AA}+\sigma^2}{G_A\mu^{i\mathbf{p}}_{AA}+\sigma^2}\big] + \frac{(y^i_{AA}-\mu^{i\mathbf{c}}_{AA})^2}{G_A\mu^{i\mathbf{c}}_{AA}+\sigma^2} - \frac{(y^i_{AA}-\mu^{i\mathbf{p}}_{AA})^2}{G_A\mu^{i\mathbf{p}}_{AA}+\sigma^2}\Big).
\end{aligned} \tag{B.5}$$

Here the summation is over the pixels that are direct neighbors of the pixel c where a point was added of removed, and defined as the neighborhood N_c.[8]

2. The Poisson process ratio term

$$\ln[\frac{p(\mathbf{x}_A \setminus \{u\}|\theta_A)}{p(\mathbf{x}_A|\theta_A)}] = -\ln[\theta_A].$$

Step 7: Accept proposal? With the logarithm of the MH-ratio available as MHR, we draw a uniform random number u between 0 and 1. If $u < \exp(\text{MHR})$ then the proposal is accepted and otherwise it is declined. If the proposal is accepted:

1. the current point pattern $\mathbf{X}$ is updated such as proposed, i.e. adding or removing a donor or acceptor point to or from the appropriate linked list.
2. XNA^i or XND^i are updated according to the accepted proposal.
3. The channel data matrices are updated by adding ΔXNDD^i and ΔXNDA^i to the values currently stored in XNDD^i and XNDA^i.

Step 8: mod(step, Mstep = 0?). In order to draw approximately independent realizations for the point process and microscope parameters, a large number Mstep of sequential updates of the point pattern are made between every parameter proposal update step.

Step 9: Propose update microscope parameter(s). We discuss this step assumming that all six microscope parameters and both the Poisson point process parameters are included in the Bayesian inference. At every Mstep'th step a proposal update is made for each of the parameters in a random order. As the support of the microscope parameters is on $\mathbb{R}^+$ we use as proposal distribution for a microscope parameter ω_j the lognormal distribution, effectively ensuring that a proposal value $\omega_j^{\mathbf{p}}$ is always strictly positive. Within the MCMC procedure a proposal $\omega_j^{\mathbf{p}}$ is now generated by drawing a random normal number $\varepsilon \sim N(0, \tau_j^2)$ and setting

$$\omega_j^{\mathbf{p}} = \omega_j^{\mathbf{c}} \exp(\varepsilon),$$

with τ_j^2 a tuning parameter, tuned such that for each of the microscope parameters the acceptance probability is around 30%.

For the Poisson point process parameters θ_D and θ_A independent updates are generated by a Gibbs step. If we wish to draw samples for θ_D conditional on the current state we need to specify $p(\theta_D|\mathbf{x}_D)$ where

$$\begin{aligned} p(\theta_D|\mathbf{x}_D) &\propto p(\mathbf{x}_D|\theta_D)p(\theta_D) \\ &\propto \theta_D^{n(\mathbf{x}_D)} \exp(-|W|\theta_D)\, \theta_D^{(\alpha_D-1)} \exp(-\beta_D\theta_D) \\ &\propto \theta_D^{(n(\mathbf{x}_D)+\alpha_D-1)} \exp(-(|W|+\beta_D)\theta_D) \\ &= \Gamma(n(\mathbf{x}_D)+\alpha_D, \beta_D+|W|), \end{aligned}$$

[8] In the case that the channel data set can not consist of negative values (i.e. empirical data), and $y_k^i = 0$ is observed, then the three terms in (B.5) related to channel k should be replaced by a so called truncated likelihood expression as discussed in Section B.2.

where in the second step for $p(\mathbf{x}_D|\theta_D)$ the density of an independent Poisson process on W with intensity θ_D (equation 6.2 in Møller and Waagepetersen, 2003) was inserted, as well as the density of the gamma distribution for $p(\theta_D)$. So, samples for θ_D are drawn by sampling from the gamma distribution $\Gamma(n(\mathbf{x}_D) + \alpha_D, \beta_D + |W|)$ with α_D, β_D the shape and rate (hyper) parameters defining the prior gamma distribution of θ_D. Equivalent, samples for θ_A are drawn by sampling from $\Gamma(n(\mathbf{x}_A) + \alpha_A$, $\beta_A + |W|)$ with α_A, β_A the shape and rate (hyper) parameters defining the prior gamma distribution of θ_A. When updating θ_D or θ_A the steps 10 and 11 are skipped and the program continues to make an update for the next parameter in line.

Step 10: Calculate MH ratio. The MH-ratio related to a proposal $\omega_j^{\mathbf{p}} \sim q_j(\cdot|\omega_j^{\mathbf{c}})$ for the j-th microscopic parameter is

$$\left(\frac{p(y|\psi^{\mathbf{p}}, \mathbf{x}^{\mathbf{c}})}{p(y|\psi^{\mathbf{c}}, \mathbf{x}^{\mathbf{c}})}\right)\left(\frac{p(\psi_j^{\mathbf{p}})}{p(\psi_j^{\mathbf{c}})}\right)\left(\frac{q(\psi_j^{\mathbf{c}}|\psi_j^{\mathbf{p}})}{q(\psi_j^{\mathbf{p}}|\psi_j^{\mathbf{c}})}\right),$$

where $\psi^{\mathbf{c}} = (\omega_1^{\mathbf{c}}, \ldots, \omega_6^{\mathbf{c}})$ contains the current values for the microscope parameters and $\psi^{\mathbf{p}}$ contains the elements

$$\psi_k^{\mathbf{p}} = \begin{cases} \omega_k^{\mathbf{c}} & k \neq j, \\ \omega_j^{\mathbf{p}} & k = j. \end{cases}$$

The logarithm of each of the terms in the MH-ratio are now specified by the

1. Likelihood ratio term[9]

$$\begin{aligned} &\ln\left[\frac{p(y|\psi^{\mathbf{p}}, \mathbf{x}^{\mathbf{c}})}{p(y|\psi^{\mathbf{c}}, \mathbf{x}^{\mathbf{c}})}\right] \\ &= \frac{1}{2}\sum_{i=1}^{n}\Big(\ln\left[\frac{G_D^{\mathbf{c}}\mu_{DD}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}}}{G_D^{\mathbf{p}}\mu_{DD}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}}}\right] + \frac{(y_{DD}^i - \mu_{DD}^{i\mathbf{c}})^2}{G_D^{\mathbf{c}}\mu_{DD}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}}} - \frac{(y_{DD}^i - \mu_{DD}^{i\mathbf{p}})^2}{G_D^{\mathbf{p}}\mu_{DD}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}}} \\ &\qquad + \ln\left[\frac{G_A^{\mathbf{c}}\mu_{DA}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}}}{G_A^{\mathbf{p}}\mu_{DA}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}}}\right] + \frac{(y_{DA}^i - \mu_{DA}^{i\mathbf{c}})^2}{G_A^{\mathbf{c}}\mu_{DA}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}}} - \frac{(y_{DA}^i - \mu_{DA}^{i\mathbf{p}})^2}{G_A^{\mathbf{p}}\mu_{DA}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}}} \\ &\qquad + \ln\left[\frac{G_A^{\mathbf{c}}\mu_{AA}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}}}{G_A^{\mathbf{p}}\mu_{AA}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}}}\right] + \frac{(y_{AA}^i - \mu_{AA}^{i\mathbf{c}})^2}{G_A^{\mathbf{c}}\mu_{AA}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}}} - \frac{(y_{AA}^i - \mu_{AA}^{i\mathbf{p}})^2}{G_A^{\mathbf{p}}\mu_{AA}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}}}\Big). \end{aligned} \tag{B.6}$$

2. Prior ratio term

$$\ln\left[\frac{p(\psi^{\mathbf{p}})}{p(\psi^{\mathbf{c}})}\right] = (\alpha - 1)\ln\left[\frac{\omega_j^{\mathbf{p}}}{\omega_j^{\mathbf{c}}}\right] + \beta(\omega_j^{\mathbf{c}} - \omega_j^{\mathbf{p}}),$$

   with α and β the hyper parameters specifying the (prior) gamma distribution of parameter ω_j.

3. Proposal ratio term

$$\ln\left[\frac{q(\psi_j^{\mathbf{c}}|\psi_j^{\mathbf{p}})}{q(\psi_j^{\mathbf{p}}|\psi_j^{\mathbf{c}})}\right] = \ln\left[\frac{\omega_j^{\mathbf{p}}}{\omega_j^{\mathbf{c}}}\right].$$

[9]The previous footnote on page 10 also applies to this likelihood expression.

Step 11: Accept proposal? With the value of the logarithm of the appropriate MH-ratio available as MHR, we draw a uniform random number u between 0 and 1. If $u < \exp(\text{MHR})$ then the proposal is accepted and otherwise it is declined. If the proposal is accepted, the current value of the microscope parameter is changed to the proposed value.

Step 12: mod(step, Tstep = 0?). If the current step number is equal to the total number of MCMC steps to be made (Tstep) the program stops, otherwise it continues.

B.2 Truncated likelihood expressions

For the typical case that the channel data does not allow for negative intensity values (i.e. non-synthetic data) we have defined the probability that a zero channel intensity value occurs as the probability mass in the left tail of the normal distribution over the negative intensity values. Therefore, in the case empirical three-cube FRET pixel intensities $y_k^i = 0$ are observed for pixel i and channel $k = DD, DA$ or AA, a truncated likelihood expression has to be used. In that case the three terms related to an observation $y_k^i = 0, k = DD$ in, for instance, (B.6) have to be replaced by

$$\ln\left[\frac{p(y_{DD}^i = 0|\psi^{\mathbf{p}}, \mathbf{x}^{\mathbf{c}})}{p(y_{DD}^i = 0|\psi^{\mathbf{c}}, \mathbf{x}^{\mathbf{c}})}\right] = \ln\left[\frac{\int_{-\infty}^{0} \varphi(z|\mu_{DD}^{i\mathbf{p}}, G_D^{\mathbf{p}}\mu_{DD}^{i\mathbf{p}} + \sigma^{2,\mathbf{p}})\mathrm{d}z}{\int_{-\infty}^{0} \varphi(z|\mu_{DD}^{i\mathbf{c}}, G_D^{\mathbf{c}}\mu_{DD}^{i\mathbf{c}} + \sigma^{2,\mathbf{c}})\mathrm{d}z}\right],$$

and when $k = DA$ or $k = AA$ by

$$\ln\left[\frac{p(y_k^i = 0|\psi^{\mathbf{p}}, \mathbf{x}^{\mathbf{c}})}{p(y_k^i = 0|\psi^{\mathbf{c}}, \mathbf{x}^{\mathbf{c}})}\right] = \ln\left[\frac{\int_{-\infty}^{0} \varphi(z|\mu_k^{i\mathbf{p}}, G_A^{\mathbf{p}}\mu_k^{i\mathbf{p}} + \sigma^{2,\mathbf{p}})\mathrm{d}z}{\int_{-\infty}^{0} \varphi(z|\mu_k^{i\mathbf{p}}, G_A^{\mathbf{c}}\mu_k^{i\mathbf{c}} + \sigma^{2,\mathbf{c}})\mathrm{d}z}\right],$$

where $\varphi(\cdot|\mu, \sigma^2)$ is the density of the normal distribution with mean μ and variance σ^2. The likelihood equations (B.5) are also updated accordingly to the strategy as described here above, in case the channel can not consist of negative intensity values.

Supplement C

In this supplementary material we present a detailed account of the results we have obtained concerning inference of the microscope parameters, when each parameter is singly introduced as a free parameter in the model. In Section C.1 we state the approach and applied settings and subsequently we discuss inference of σ^2 in Section C.2, G_D and G_A in Section C.3, M_D in Section C.4 and of G and K in Section C.5.

In order to explain many of the results, we often will refer to the equations for μ^i_{DD}, μ^i_{DA} and μ^i_{AA} as defined previously in the article. Therefore these relations are restated here with an equation number for easy referencing throughout this supplementary material:

$$\mu^i_{DD} = M_D \sum_{d \in \mathbf{X}_D \cap C_i} (1 - P_{dA}), \tag{C.1}$$

$$\mu^i_{DA} = G M_D \sum_{a \in \mathbf{X}_A \cap C_i} \sum_{d \in \mathbf{X}_D} P_{da}, \tag{C.2}$$

$$\mu^i_{AA} = M_D / K \; n(\mathbf{X}_A \cap C_i). \tag{C.3}$$

Throughout this supplementary material we refer to a point pattern type by its *point pattern name* or *number* as defined in Table 1 in the article.

C.1 Approach and settings

By including the microscope parameters in the Bayesian inference the joint posterior distribution – equation (3.1) in the article – reads as

$$p(\mathbf{x}_D, \mathbf{x}_A, \theta, \psi | y) \propto p(y | \mathbf{x}_D, \mathbf{x}_A, \psi) p(\mathbf{x}_D, \mathbf{x}_A | \theta) p(\theta) p(\psi), \tag{C.4}$$

with $p(\psi)$ the prior density of the miscroscope parameters. The support of the microscope parameters is on $\mathbb{R}^+$ and therefore a natural choice for the proposal distribution for a microscope parameter ω_j is the lognormal distribution, which ensures that a proposal value $\omega_j^{\mathbf{p}}$ is strictly positive. Denoting by $\varphi(z, \mu, \sigma^2)$ the density function of a normally distributed variable z with mean μ and standard deviation σ, the proposal density function for $\omega_j^{\mathbf{p}}$ conditionally under its current value $\omega_j^{\mathbf{c}}$ is

$$q(\omega_j^{\mathbf{p}} | \omega_j^{\mathbf{c}}) = \frac{1}{\omega_j^{\mathbf{p}}} \varphi(\ln[\omega_j^{\mathbf{p}}] \mid \ln[\omega_j^{\mathbf{c}}], \tau_j^2),$$

with τ_j^2 a tuning parameter, controlling the percentage of accepted proposals within a MCMC run. A proposal $\omega_j^{\mathbf{p}}$ is now generated by drawing a random normal number $\varepsilon \sim N(0, \tau_j^2)$ and setting $\omega_j^{\mathbf{p}} = \omega_j^{\mathbf{c}} \exp(\varepsilon)$. We note that as $\varphi(\cdot)$ is symmetric around its mean it follows that

$$\frac{q(\omega_j^{\mathbf{c}} | \omega_j^{\mathbf{p}})}{q(\omega_j^{\mathbf{p}} | \omega_j^{\mathbf{c}})} = \frac{\omega_j^{\mathbf{p}}}{\omega_j^{\mathbf{c}}}.$$

Applying (C.4), the Metropolis-Hastings ratio related to a proposal $u \sim q_j(\cdot|\omega_j^{\mathbf{c}})$ for the j-th microscopic parameter is

$$\left(\frac{p(y|\psi^{\mathbf{p}}, \mathbf{x}^{\mathbf{c}})}{p(y|\psi^{\mathbf{c}}, \mathbf{x}^{\mathbf{c}})}\right) \left(\frac{p(\psi_j^{\mathbf{p}})}{p(\psi_j^{\mathbf{c}})}\right) \left(\frac{q(\psi_j^{\mathbf{c}}|\psi_j^{\mathbf{p}})}{q(\psi_j^{\mathbf{p}}|\psi_j^{\mathbf{c}})}\right), \tag{C.5}$$

with $\psi^{\mathbf{c}} = (\omega_1^{\mathbf{c}}, \ldots, \omega_6^{\mathbf{c}})$ containing the current values for the microscope parameters and $\psi^{\mathbf{p}}$ containing the elements

$$\psi_k^{\mathbf{p}} = \begin{cases} \omega_j^{\mathbf{c}} & k \neq j, \\ u & k = j. \end{cases}$$

The expressions for the first two terms in (C.5) are specified under step 10 in Section B.1 of Supplement B while the third term equals $\omega_j^{\mathbf{p}}/\omega_j^{\mathbf{c}}$. Supplement B provides a detailed description of the MCMC sampler used for sampling from the posterior distribution (C.4).

C.1.1 Set up of the various simulation experiments

Below we give a short description of each of the simulations that have been carried out and state the settings that have been used. In order to study the effect a relatively low or high signal-to-noise ratio has on the inference, each simulation has been performed for three synthetic values of M_D $(1, 5, 20)$. In the next we denote the prior mean of a parameter by adding the superscript "pm" to the parameter.

A All microscope parameters are fixed in the inference procedure (the MCMC run) and set to the values of their synthetic counterparts applied to create the synthetic channel data. The synthetic value of the parameters G, K, G_D, G_A is set to 1 and the synthetic value of σ^2 is set to 25. Synthetic channel data is generated for a grid $\mathbf{G}$ which divides the window $W = 1\,\mu\text{m} \times 1\,\mu\text{m}$ in 10×10 equally sized square pixels. Accordingly the dimension of a pixel is $0.1\,\mu\text{m} \times 0.1\,\mu\text{m}$. The Poisson point process intensities θ_D and θ_A are free parameters and their prior means are set approximately equal to the intensities of the synthetic patterns, i.e. $\theta_D^{\text{pm}} = \theta_A^{\text{pm}} = 1000/\mu\text{m}^2$. In this simulation, inference is made on the spatial configuration of donors and acceptors and on the Poisson point process intensities θ_D and θ_A. Interest is in assessing how the accuracy of the posterior L-function depends on the underlying point pattern type. The outcomes of Simulation A are disussed in the article in Section 4.3 and its subsections.

B Settings as in Simulation A, but σ^2 is a free parameter. The prior mean value of σ^2 is set equal to the synthetic value of 25. Interest is in studying the inference for the parameter σ^2. Results are discussed in Section C.2.

C–G Settings as in Simulation A, but G_A, G_D, M_D, G, K are respectively (each separately) a free parameter in the model. Prior mean of the free parameter is always set equal to the synthetic value of 1. Results for G_A and G_D are discussed in Section C.3, for M_D in Section C.4 and for G and K in Section C.5.

H Settings as in Simulation A, but $M_D, G, K, G_D, G_A, \sigma^2$ are free parameters with their prior means set equal to the corresponding synthetic values. Interest is in studying the inference of all the microscope parameters in the setting that they are all free. Results are discussed in Section 4.4 in the article.

In Table C.1 the settings of each of the simulations are summarized.

Table C.1: Parameter settings as defined for the various simulations. All simulations are performed on a window $W = 1\,\mu\text{m} \times 1\,\mu\text{m}$ and a grid dividing W in 10×10 square pixels with a width of $0.1\,\mu\text{m}$. In each simulation the Poisson intensities are free parameters with a prior mean value of $1000/\mu\text{m}^2$. For Simulations C–G we have set the synthetic value of σ^2 (almost) equal to zero, in order to get the most accurate inference regarding G_A, G_D, K, G and M_D as possible.

	Within the MCMC run						Synthetic value					
Simulation	M_D	G	K	G_D	G_A	σ^2	M_D	G	K	G_D	G_A	σ^2
A	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$	1,5,20	1	1	1	1	25
B	$\star$	$\star$	$\star$	$\star$	$\star$	F	1,5,20	1	1	1	1	25
C	$\star$	$\star$	$\star$	$\star$	F	$\star$	1,5,20	1	1	1	1	0.1
D	$\star$	$\star$	$\star$	F	$\star$	$\star$	1,5,20	1	1	1	1	0.1
E	$\star$	$\star$	F	$\star$	$\star$	$\star$	1,5,20	1	1	1	1	0.1
F	$\star$	F	$\star$	$\star$	$\star$	$\star$	1,5,20	1	1	1	1	0.1
G	F	$\star$	$\star$	$\star$	$\star$	$\star$	1,5,20	1	1	1	1	0.1
H	F	F	F	F	F	F	1,5,20	1	1	1	1	25

$\star$: fixed parameter set equal to the synthetic value.
F: free parameter with its prior mean equal to the synthetic value.

C.1.2 Hyper parameters of the priors

The Gamma distribution is used as the prior density for the microscope parameters. Following the reasoning as outlined for setting the Poisson priors in Section 3.1 in the article, we also set the shape parameter of the microscope parameters equal to 4. In our simulations we specify the prior mean of each of the microscope parameters (see Table C.1). Then, if the value of the mean is denoted m the scale parameter β follows from $\beta = m/\alpha = m/4$.

C.1.3 Markov chain Monte Carlo settings

The total number of steps in each MCMC run is set to 10^7. A proposal update is made in each step to remove/add a donor or acceptor point to/from the current point pattern. A posterior point pattern is written to disk after every 10^5 steps. A proposal update for the microscope parameters (in case free) and for the Poisson point process intensities is made every 2.5×10^3 steps. We store the parameter and intensities values to disk every 10^4 steps. So, for each complete MCMC run,

one thousand values of the microscope parameters and Poisson intensities, and one hundred point patterns are stored to disk. The initial values of the point process intensities and the free microscope parameters are always set equal to their respective prior means.

C.1.4 Tuning parameters of the proposal distributions.

To generate proposals for the microscope parameters we have used the following values for the tuning parameter τ in the lognormal distribution, To generate proposals: for M_D we have used $\tau = 0.04$; for G and K we used $\tau = 0.05$; and for σ^2, G_A and G_D we used $\tau = 0.1$. In all cases the applied setting result in acceptance probabilities between 0.1–0.6 depending on the value of M_D. Higher values of M_D – corresponding to a higher signal-to-noise ratio – result in lower acceptance rates.

C.2 Inference of the measurement noise

In Figure C.1 the posterior means of the measurement noise standard deviation ($\bar{\sigma}$) for the replicated runs (from Simulation B) are summarized for each of the point pattern types by a boxplot. Results are shown for M_D equal to 1, 5 and 20. Clearly, the dimer and clustered point pattern types generated with $\gamma_{DA} = 8$ (type number: 2, 4, 6 and 8) show values for $\bar{\sigma}$ larger than the synthetic value of 5 for all three values of M_D. Further, saliently, the bias grows significantly for larger values of M_D. For all other point pattern types, $\bar{\sigma}$ is consistently close to five for all three values of M_D. We conclude that for modestly clustered point patterns (dimer and clustered types generated with $\gamma_{DA} = 2$) as well as for the Poisson and repulsive types, the inference of σ performs well. The results found for the dimer and clustered point patterns generated with $\gamma_{DA} = 8$ are surprising. We investigate this issue in the next section.

C.2.1 Persistent bias in posterior DA-channel intensity

To understand the cause of the offset between the posterior mean measurement noise standard deviation and its synthetic value we will study the posterior mean *pixel deviance* – as defined below – over replicated runs for each point pattern type, for the case that all microscope parameters are fixed to their synthetic value (that is settings as in Simulation A). We start by defining the *deviance* as used in this supplementary material. For a random variable Z, distributed with mean μ and variance σ^2 and for which n observations have been made, we refer to the quantity $\mathrm{DEV}(Z) = (1/n)\sum_{i=1}^{n}((Z_i - \mu)/\sigma)^2$ as the deviance. By the definition of the variance: $\sigma^2 = \mathrm{E}[(Z - \mathrm{E}[Z])^2]$, the expected value of the deviance is one, that is $\mathrm{E}[\mathrm{DEV}(Z)] = 1$. Applying the deviance definition to the pixel intensities in the three channels (equations (2.1)–(2.3) in the article), and defining n to be the number of pixels in a channel image, we define the *deviance* for each of the three channels

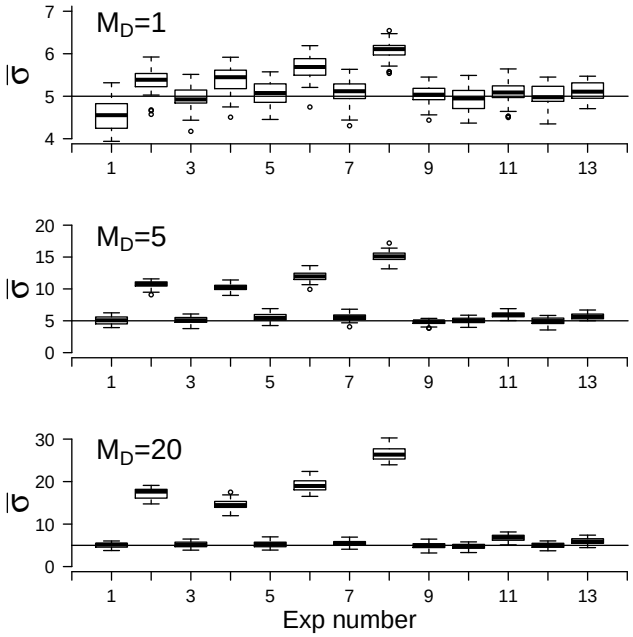


Figure C.1: Boxplot of the posterior mean measurement noise standard deviation of the forty replicated runs, for each of the point pattern types (referred to by their type number as denoted in Table 1 in the rticle), for $M_D = 1$ (upper), $M_D = 5$ (middle) and $M_D = 20$ (lower). Results are from Simulation B; σ^2 is a free parameter. The horizontal lines are drawn at the synthetic value of σ which is 5.

as

$$\mathrm{DEV}_{DD} \equiv \mathrm{DEV}(Y_{DD}|\mu_{DD}) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y^i_{DD} - \mu^i_{DD})^2}{G_D\mu^i_{DD} + \sigma^2}, \tag{C.6}$$

$$\mathrm{DEV}_k \equiv \mathrm{DEV}(Y_k|\mu_k) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y^i_k - \mu^i_k)^2}{G_A\mu^i_k + \sigma^2}; \qquad k = DA, AA. \tag{C.7}$$

In Figure C.2 the posterior mean of the channel deviance for the replicated runs (results from Simulation A) are summarized by aid of a boxplot for each of the point pattern types and each of the three channels. For $M_D = 1$ the posterior mean deviance in the DA-channel, plot (a), is clearly above one for the point patterns with type number: 2, 4, 6, 8 (that is the patterns generated with a $\gamma_{DA} = 8$), while it is on the target value of one for all other types. Also for $M_D = 20$ the posterior mean deviances in the DA-channel, plot (b), are persistently above one for the strongly clustered types (type number: 2, 4, 6, 8), while now also the deviance related to the strongly repulsive patterns (Rep.h2 and Rep.s2, type number 11 and 13) are slightly above target. In contrast, in the DD-channel, plot (c) and (d), and AA-channel, plot (e) and (f), the posterior mean deviance for all point pattern types are close to one for $M_D = 1$ as well as for $M_D = 20$. In short, our main observation from the various deviance plots is that for the strongly clustered and strongly repulsive point patterns, the inference procedure has – also for a high signal-to-noise ratio (high M_D) – significant problems to get on target in the DA-channel.

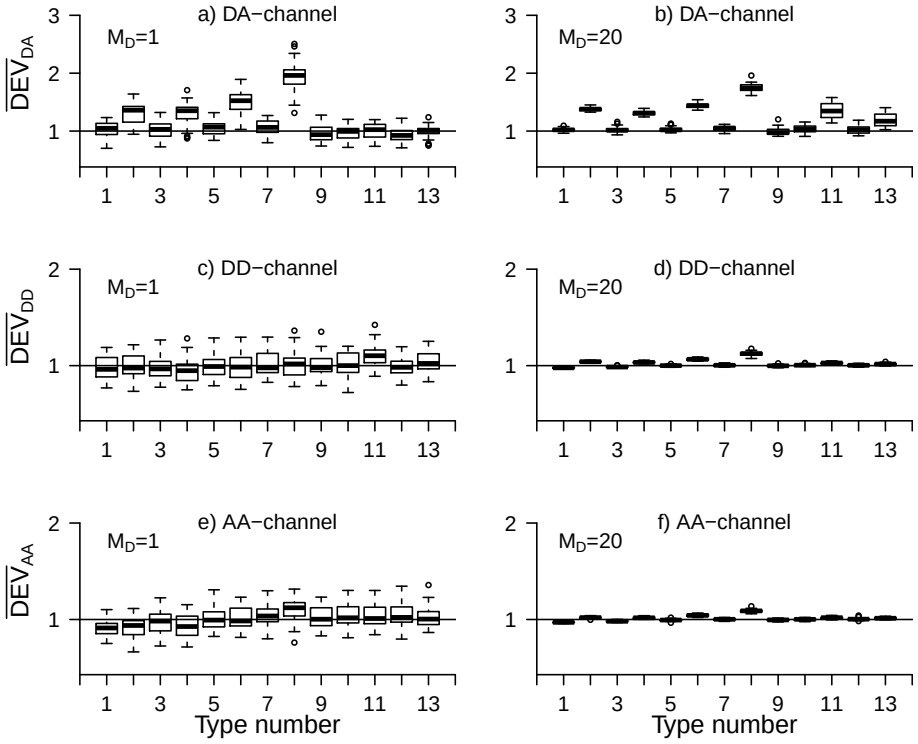


Figure C.2: Boxplot of the posterior mean deviance of the forty replicated runs for each of the point pattern types (referred to by their type number as defined in Table 1 in the article) and for each of the three channels. Upper plots: DA-channel, middle plots: DD-channel and lower plots: AA-channel, for $M_D = 1$ (left figures) and $M_D = 20$ (right figures). Results are from Simulation A.

To study this offset in the DA-channel further Figure C.3 shows the posterior mean pixel intensities $\bar{\mu}_k^i, k = DD, DA, AA$, versus the corresponding synthetic pixel intensities $Y_{k,\text{synth}}^i$ for the three values of M_D, and where $Y_{k,\text{synth}}$ is generated from a Clu.28 point pattern type. The results are from Simulation A; σ^2 is fixed and set equal to the synthetic value of 25. Clearly, the $\bar{\mu}_{DD}^i$'s and $\bar{\mu}_{AA}^i$'s get properly on target for higher values of M_D (from left to right in respectively the upper and lower plots). Also the $\bar{\mu}_{DA}^i$'s are closer to target for higher values of M_D, although for $M_D = 20$ still a persistent negative bias compared to the synthetic DA-channel pixel intensity is observed (right middle plot). Similar scatter plots for point pattern types Dim.18, Dim.28 and Clu.28, display a similar bias in the DA-channel (not shown). Scatter plots for point pattern types Rep.h2 and Rep.s2, show a positive bias in the DA-channel, that is $\bar{\mu}_{DA}^i$ is mostly larger than Y_{DA}^i (not shown). We conclude that a persistent negative (positive) bias in the posterior DA-channel pixel intensities is present when making inference on synthetic channel data constructed from strongly clustered (repulsive) point patterns.

By the previous analyses we can now explain the large offset observed in Figure C.1 between the posterior means of the measurement noise standard deviation and the synthetic value. Due to the bias in the DA-intensity channel – values of DEV_{DA} larger than one – the inference procedure with σ^2 a free parameter in the model (Simulation B), will increase σ^2 in order to get DEV_{DA} on target. However,

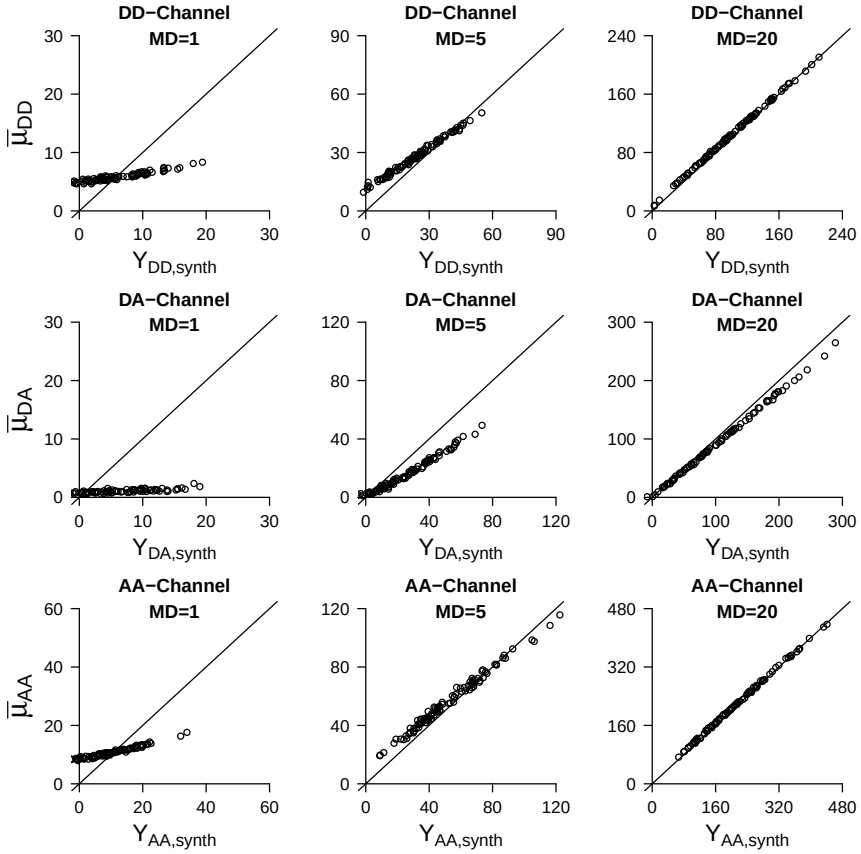


Figure C.3: Scatter plots of the posterior mean pixel intensities: $\bar{\mu}_k^i$, versus the corresponding synthetic pixel channel intensities $Y_{k,\text{synth}}^i$ $k = DD, DA, AA$, $i = 1, \ldots, n$; $n = 100$. With $Y_{k,\text{synth}}$ generated from a point pattern $\mathbf{X}_{\text{synth}}$ of Clu.28 type. From left to right: $M_D = 1, 5, 20$; from above to below: DD-, DA- and AA-channel. For higher values of M_D, the posterior mean DD- and AA-channel pixel intensities get on target, while the posterior mean DA-channel pixel intensities remain below target. Black line is reference line with slope 1.

the increase in σ^2 results in a further increase in the bias of the posterior DA-pixel intensities, resulting again in a further increase of σ^2. This explanation is supported by Figure C.4 in which the posterior mean values of σ are plotted versus the corresponding synthetic pixel intensities for the same Clu.28 point pattern as the results in Figure C.3 are based on, but now σ^2 is a free parameter in the model (posterior means are from Simulation B). Clearly, with σ^2 a free parameter in the model, the offset between posterior mean DA-intensity and synthetic DA-channel intensity increases (compare Figure C.4 middle plots, with corresponding plots in Figure C.3). Further, due to the increase in σ^2 now also the posterior mean intensities in the DD- and AA-channel have difficulty to get on target (compare Figure C.4, upper and lower plots, with corresponding plots in Figure C.3).

In summary, due to a relative large offset between the Poisson point process prior model and the strongly clustered and repulsive patterns, a persistent bias exist between the posterior pixel mean intensities and the synthetic pixel intensities in the DA-channel. This results in biased inference for σ.

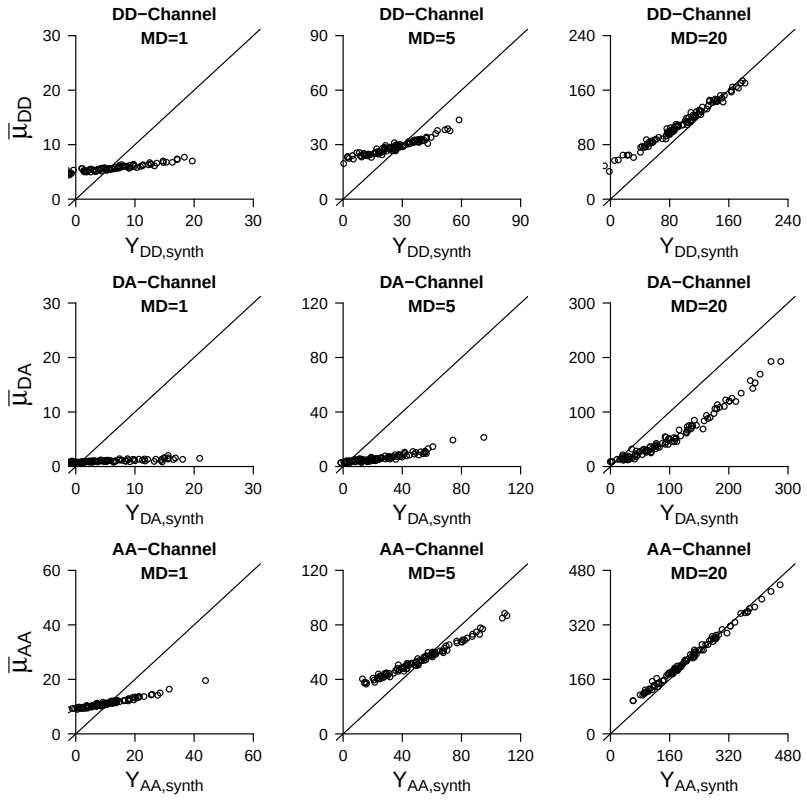


Figure C.4: As Figure C.3 but results are from Simulation B; σ^2 is a free parameter. Due to the bias of the posterior mean intensity in the DA-channel – see in Figure C.3 – with σ^2 a free parameter the inference procedure will start to increase σ^2, which leads to a further increase in the bias in the DA-channel. Because σ^2 defines the measurement noise in all the three channels also in the DD- and AA-channel the posterior pixel intensities now have difficulties to get on target (upper and lower plots).

C.3 Inference of the amplification factors

The results concerning inference of G_A are in accordance with the results as discussed in the previous section for the measurement noise. Due to the persistent bias in the posterior DA-channel intensities for strongly clustered patterns, for such patterns G_A is excessively increased in the inference procedure in order to get DEV_{DA} (C.7) on its target value of one. This is clearly seen in Figure C.5 (left plots), as for the pattern types 2, 4, 6 and 8 the values of $\bar{G}_A$ are clearly above the synthetic value of one for all three values of M_D, while on target or close to target for all other point pattern types.

The parameter G_D only effects DEV_{DD} (C.6) and thereby G_D can not be tuned in the inference procedure to adjust DEV_{DA} (C.7). Therefore, we would expect that proper inference for this parameter should be possible for all point pattern types. However, from Figure C.5 (right plots) we notice that for the strongly clustered patterns (type 2, 4, 6, 8) also the value of $\bar{G}_D$ is above the synthetic value of one, which is most clearly seen for Clu.28 (type 8). The probable explanation is that for the strongly clustered patterns also a small but significant persistent bias exists in

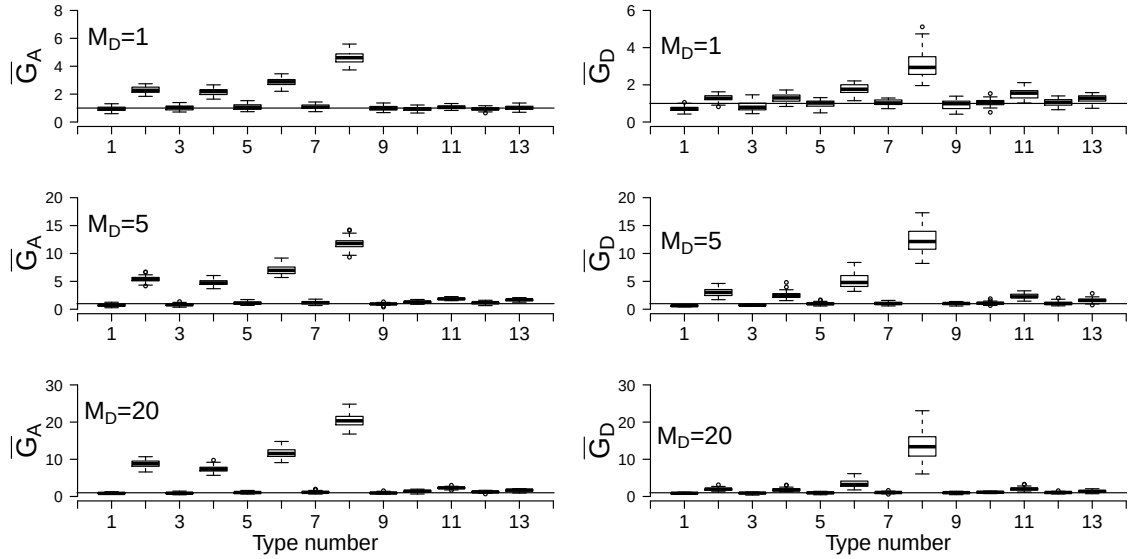


Figure C.5: Boxplots of the posterior mean of: (left) G_A and (right) G_D, for the forty replicated runs for each of the point pattern types (referred to by their type number as denoted in Table 1 in the article), for $M_D = 1$ (upper), $M_D = 5$ (middle) and $M_D = 20$ (lower). The horizontal lines are drawn at the synthetic values $G_A = 1$ (left plots) and $G_D = 1$ (right plots).

the DD-channel intensities, leading to values of DEV_{DD} slightly above the target value of one. This is supported by Figure C.2, where especially for Clu.28 (type 8) it is rather clearly seen that DEV_{DD} is above one for $M_D = 20$.

We conclude that, except for the strongly clustered point patterns, the inference procedure provides good estimates for G_A and G_D.

C.4 Inference of M_D

In Figure C.6 the posterior mean of M_D for the forty replicated runs is summarized by aid of a boxplot for each of the point pattern types and for $M_D = 1, 5, 20$. Results are of Simulation E. Inspection of the plots shows that there is a clear trend in the observed posterior means.

1. for underlying clustered patterns (type 1–8), the values of $\bar{M}_D$ are smaller than the corresponding synthetic values. Further, the bias is larger for the strongly clustered point patterns generated with $\gamma_{DA} = 8$ (type 2, 4, 6 and 8).[10]
2. for underlying repulsive patterns (type 10–13), the values of $\bar{M}_D$ are larger than the corresponding synthetic values and the bias increases for the strongly repulsive point patterns (type 11 and 13, that is Rep.h2 and Rep.s2).
3. for underlying Poisson hard core patterns (type 9), the value of $\bar{M}_D$ coincides with the corresponding synthetic value.

[10]We ignore here for the moment the values of $\bar{M}_D$ larger than their synthetic counterparts occurring for type 8 and M_D is 5 or 20. We comment on it at the end of this section.

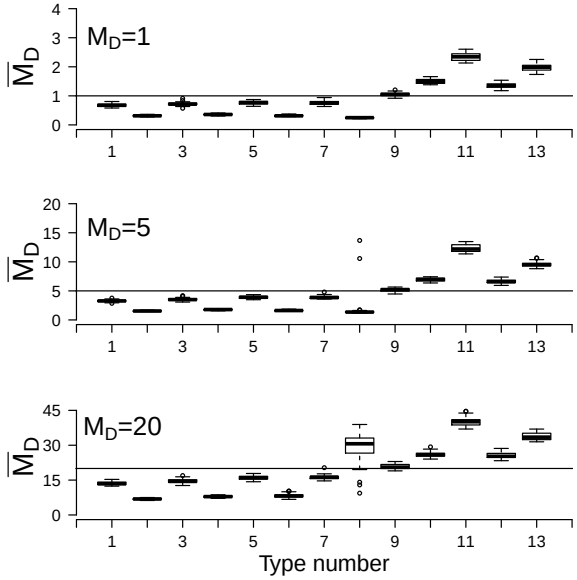


Figure C.6: Boxplot of posterior mean of M_D for the forty replicated runs for each of the point pattern types (referred to by their type number as denoted in Table 1 in the article), for $M_D = 1$ (upper), $M_D = 5$ (middle) and $M_D = 20$ (lower). In each plot the horizontal line is drawn at the synthetic values of M_D. Results are from Simulation E.

Further, the observed bias for the dimer, clustered and repulsive patterns is persistent in the sense that increasing the signal-to-noise ratio (by increasing M_D) does not result in a decrease of the bias.

The bias is an artifact of the use of the Poisson point process prior. The explanation is as follows. When inference is made with all the microscope parameters fixed to their synthetic values (Simulation A) on a *strongly hetero pair clustered* point pattern, the inference procedure has difficulties to create such clusters, because the Poisson point process prior will favor a more random distribution of the hetero points. This results, as previously discussed in Section C.2, in a negative bias of the posterior pixel mean DA-channel intensities, μ^i_{DA}, and of values of DEV_{DA} above the target value of one (Figures C.2 and C.3). To reduce the bias, the inference procedure – with M_D a free parameter (Simulation E) – favors to add more acceptors to the posterior pattern than the corresponding synthetic pattern contains. By adding extra acceptors, the average hetero pair inter distances decrease, resulting in an increase of the double summation term in (C.2) – and so of $\bar{\mu}^i_{DA}$ – thereby effectively removing the bias. Increasing the number of acceptors will lead to a positive bias of μ^i_{AA} (C.3). However, this bias is removed by the inference procedure by decreasing the value of M_D and so getting μ^i_{AA} on target again. As M_D is also present in the expression for μ^i_{DD} (C.1), this leads to a bias in the DD-channel, which, however, is removed by the inference procedure by adding more donors to the posterior patterns than the underlying synthetic pattern contains, effectively getting μ^i_{DD} back on target. Further, by (C.2), μ^i_{DA} is also proportional to M_D and by lowering M_D also μ^i_{DA} will be lower. However, the relative increase of the double summation term in (C.2) due to the higher concentrations of acceptors is larger than the relative decrease of M_D, resulting in an increase of the DA-channel pixel intensities and so bringing μ^i_{DA} onto target.

The same argument holds for underlying repulsive patterns but the other way around. In this case the posterior μ^i_{DA}'s – with all microscope parameters fixed to their synthetic values – show a positive bias with respect to the synthetic channel data. And the inference procedure – with M_D a free parameter – favors posterior patterns to contain less donors and acceptors than the synthetic pattern, thereby increasing the average hetero pair inter distances and bringing the posterior pixel intensities in the DA-channel onto target. This leads to a higher value of M_D with respect to the corresponding synthetic value.

We now will discuss the in the footnote on the previous page mentioned inference results concerning the Clu.28 patterns (type 8), for which for $M_D = 5$ some, and for $M_D = 20$ most of the posterior means $\bar{M}_D$ are higher than these synthetic values. Clearly these results are not captured by the explanation stated above. They can, however, be explained in relation to the large offset that exists between the *homogeneous* Poisson point process prior and the very *inhomogeneous* way donors and acceptors are distributed over the pixels in Clu.28 patterns. We will give a qualitative description. For the clustered point patterns type 1–7, hetero clustering occurs in such a way that the donors and acceptors are more or less homogeneously distributed over the pixels. This is schematically depicted in Figure C.7 (a), as a dimer pair residing in each of the pixels. For the point pattern types 1–7, to get the DA-channel intensities on target the posterior patterns will contain more donors and acceptors – as described in detail above – than the corresponding synthetic patterns (see Figure C.7 (b)), resulting in posterior mean values of M_D lower than the corresponding synthetic value. For the Clu.28 type point patterns, however, the hetero clustering is so strong that this results in large super clusters of donors and acceptors, leading to a very *inhomogeneous* distribution of donors and acceptors over the pixels. This is schematically depicted in Figure C.7 (c), in which all points are concentrated in only one pixel. As such a distribution is extremely unlikely to occur under the *homogeneous* Poisson point process prior, the inference procedure favors to: place only a few points in an inhomogeneous way over the pixels – which under the Poisson process prior and conditional under a fixed number of points, has a higher probability than to place many point inhomogeneously (see Figure C.7 (d)), while using a higher value of M_D than the synthetic value, in order to get the the posterior channel intensities onto target.

We conclude that the results of the inference of M_D are highly dependent on the synthetic point pattern type.

C.5 Inference of the G and K factor

In Figure C.8 (left) the results concerning inference of G (Simulation F) are shown. Clearly, $\bar{G}$ is larger than the synthetic value of one for clustered point patterns (type numbers 1–8) while smaller than one for the repulsive point patterns (type numbers 10–13). Further, the offset is larger for the more strongly clustered (types 2, 4, 6, 8)

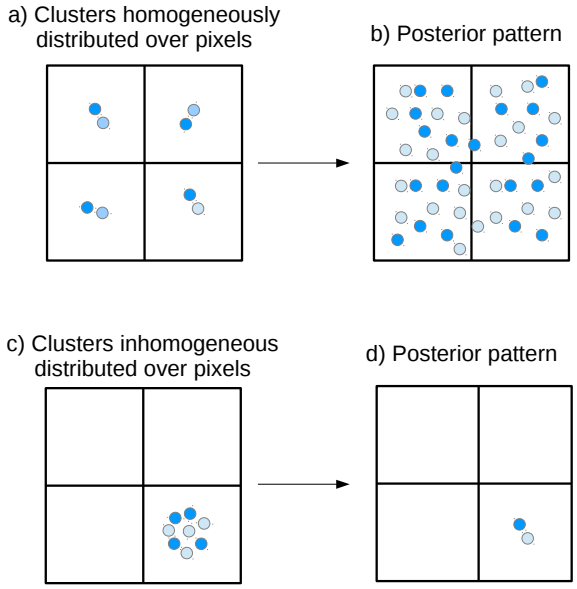


Figure C.7: Point patterns for which hetero cluster of points are distributed: (a) homogeneously, and (c) inhomogeneously, over the pixels. For hetero clusters distributed homogeneously over the pixels, the inference procedure – with M_D a free parameter – favors the posterior patterns to contain more donors and acceptors than the synthetic one (plot (b)) and to decrease the value of M_D below the synthetic value. For hetero clusters distributed inhomogeneously over the pixels, the inference procedure favors to decrease the number of donors and acceptors (plot (d)) and to increase M_D. See also the text.

and more strongly repulsive (types 11 and 13) point patterns. For underlying Poisson hard core patterns (type 9), the value of $\bar{G}$ coincides with, or is close to, the synthetic value of one. These results can again be explained by the relative offset between the prior Poison process model and the pattern types. With G a free parameter in the model, any existing consistent positive or negative bias of the posterior DA-channel pixel mean intensities – μ^i_{DA} (C.2) – and the corresponding synthetic data Y^i_{DA}'s can effectively be removed by the tuning of G in (C.2). For underlying clustered patterns, the inference procedure favors to increase G in (C.2) – in order to match the μ^i_{DA}'s with the Y^i_{DA}'s – instead of increasing the value of the double summation term in (C.2) by placing donors and acceptors sufficiently close to each other. Similar, for underlying repulsive patterns, the inference procedure favors to decrease G in (C.2) – in order to match the μ^i_{DA}'s with the Y^i_{DA}'s – instead of decreasing the value of the double summation term in (C.2) by placing donors and acceptors sufficiently far from each other.

In Figure C.8 (right) the results concerning inference of K (Simulation G) are shown. Clearly, $\bar{K}$ is larger/smaller than the synthetic value of one for clustered/ repulsive underlying point patterns (type number 1–8 and 10–13, respectively). Further, the offset is larger for the strongly clustered (type number 2,4,6,8) and strongly repulsive (type number 11 and 13) point patterns. For underlying Poisson hard core patterns (type 9), the value of $\bar{K}$ coincides with, or is close to, the correct value of one. With K a free parameter in the model, any existing consistent positive or negative bias of the posterior DA-channel pixel intensities can be removed by tun-

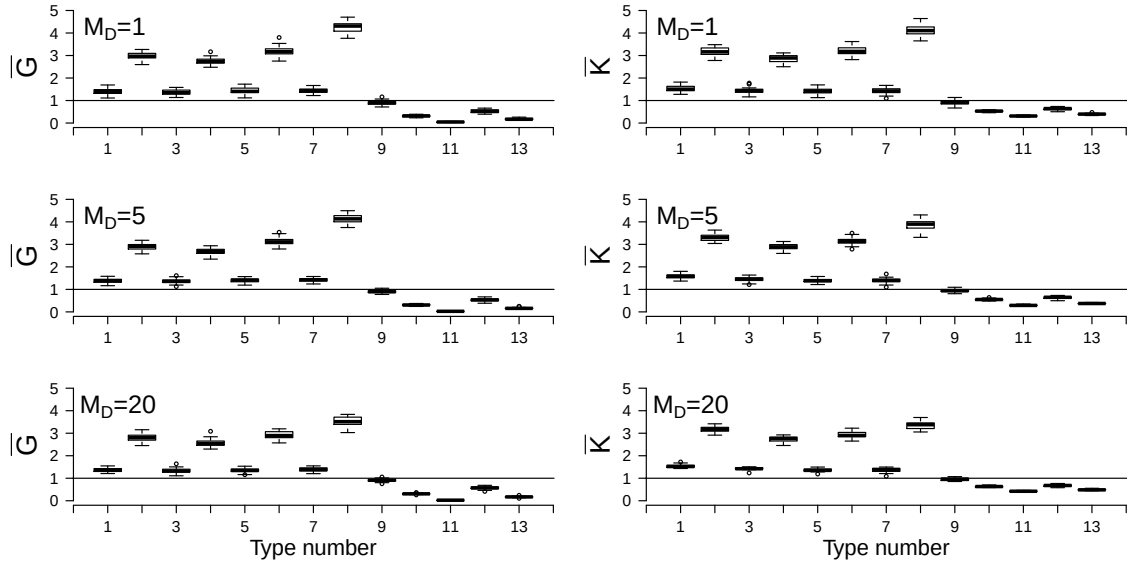


Figure C.8: Boxplot of posterior mean of: (left) G, and (right) K, for the forty replicated runs for each of the point pattern types (referred to by their type number as denoted in Table 1 in the article), for $M_D = 1$ (upper plots), $M_D = 5$ (middle plots) and $M_D = 20$ (lower plots). In each plot the horizontal line is drawn at the synthetic value $G = 1$ (left plots) and the synthetic value $K = 1$ (right plots). Results of $\bar{G}$ and $\bar{K}$ are from Simulations F and G, respectively.

ing the number of acceptors within the posterior patterns. For underlying clustered patterns, the inference procedure favors to increase the number of acceptors with respect to the underlying synthetic pattern. Hereby, the value of the double summation term in (C.2) increases, effectively removing the negative bias between the μ^i_{DA}'s and Y^i_{DA}'s. The increase in acceptors leads to an offset in the AA-channel intensities (C.2), but this offset is removed by increasing the value of K in (C.2). Similar, for underlying repulsive patterns, the inference procedure favors to remove acceptors, resulting in a smaller value of K compared to the synthetic value of one.

We conclude that the results of the inference of G and K are highly dependent on the point pattern type.

References

Frank Alber, Svetlana Dokudovskaya, Liesbeth M. Veenhoff, Wenzhu Zhang, Julia Kipper, Damien Devos, Adisetyantari Suprapto, Orit Karni-Schmidt, Rosemary Williams, Brian T. Chait, Michael P. Rout, and Andrej Sali. Determining the architectures of macromolecular assemblies. *Nature*, 450:683–694, 2017. `http://dx.doi.org/10.1038/nature06404`.

Adrian Baddeley and Rolf Turner. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005. `http://www.jstatsoft.org/v12/i06/`.

Claude Berney and Gaudenz Danuser. Fret or no fret: a quantitative comparison. *Biophysical Journal*, 84(6):3992–4010, 2003.

Massimiliano Bonomi, Riccardo Pellarin, Seung Joong Kim, Daniel Russel, Bryan A. Sundin, Michael Riffle, Daniel Jaschob, Richard Ramsden, Trisha N. Davis, Eric G. D. Muller, and Andrej Sali. Determining protein complex structures based on a Bayesian model of in vivo förster resonance energy transfer data. *Molecular and Cellular Proteomics*, 13(11):2812–2823, 2014. `http://www.mcponline.org/content/13/11/2812.abstract`.

Gertrude Bunt and Fred S. Wouters. Visualization of molecular activities inside living cells with fluorescent labels. In *International Review of Cytology*, volume 237, pages 205–277. Academic Press, 2004. `http://www.sciencedirect.com/science/article/pii/S0074769604370051`.

Huanmian Chen, Henry L Puhl, Srinagesh V Koushik, Steven S Vogel, and Stephen R Ikeda. Measurement of fret efficiency and ratio of donor to acceptor concentration in living cells. *Biophysical journal*, 91(5):L39–L41, 2006.

Huanmian Chen, Henry L Puhl, and Stephen R Ikeda. Estimating protein-protein interaction affinity in living cells using quantitative förster resonance energy transfer measurements. *Journal of biomedical optics*, 12(5):054011–054011, 2007.

Leng-Chun Chen, William R. Lloyd III, Ching-Wei Chang, Dhruv Sud, and Mary-Ann Mycek. Chapter 20 - fluorescence lifetime imaging microscopy for quantitative biological imaging. In Greenfield Sluder and David E. Wolf, editors, *Digital Microscopy*, volume 114 of *Methods in Cell Biology*, pages 457–488. Academic Press, 2013. `http://www.sciencedirect.com/science/article/pii/B9780124077614000208`.

R.M. Clegg. Fluorescence resonance energy transfer. In X.F. Wang and B. Herman, editors, *Fluorescence Imaging Spectroscopy and Microscopy*, pages 179–252. Wiley and Sons, NY, 1996.

Robert M Clegg. Fluorescence resonance energy transfer. *Current opinion in biotechnology*, 6(1):103–110, 1995.

RobertM. Clegg. The history of FRET. In ChrisD. Geddes and JosephR. Lakowicz, editors, *Reviews in Fluorescence 2006*, volume 2006 of *Reviews in Fluorescence*, pages 1–45. Springer US, 2006. ISBN 978-0-387-29342-4. `http://dx.doi.org/10.1007/0-387-33016-X_1`.

Ben Corry, Dylan Jayatilaka, and Paul Rigby. A flexible approach to the calculation of resonance energy transfer efficiency between multiple donors and acceptors in complex geometries. *Biophysical journal*, 89(6):3822–3836, 2005.

Masilamani Elangovan, Horst Wallrabe, Ye Chen, Richard N Day, Margarida Barroso, and Ammasi Periasamy. Characterization of one-and two-photon excitation fluorescence resonance energy transfer microscopy. *Methods*, 29(1):58–73, 2003.

Harold P Erickson. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration and electron microscopy. *Biological Procedures Online*, 11:32–51, apr 2009. ISSN 1480-9222. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3055910/.

Th. Förster. Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen der Physik*, 437(1-2):55–75, 1948. ISSN 1521-3889. http://dx.doi.org/10.1002/andp.19484370105.

Patrick LTM Frederix, Evert L de Beer, Wendie Hamelink, and Hans C Gerritsen. Dynamic monte carlo simulations to model fret and photobleaching in systems with multiple donor-acceptor interactions. *The Journal of Physical Chemistry B*, 106(26):6793–6801, 2002.

D. Gamerman and H.F. Lopes. *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2006. ISBN 9781584885870. https://books.google.dk/books?id=yPvECi_L3bwC.

W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995. ISBN 9780412055515. https://books.google.dk/books?id=TRXrMWY_i2IC.

Debanjan Goswami, Kripa Gowrishankar, Sameera Bilgrami, Subhasri Ghosh, Riya Raghupathy, Rahul Chadda, Ram Vishwakarma, Madan Rao, and Satyajit Mayor. Nanoclusters of gpi-anchored proteins are formed by cortical actin-driven activity. *Cell*, 135(6):1085–1097, 2008. ISSN 0092-8674. http://www.sciencedirect.com/science/article/pii/S0092867408015079.

Zygmunt Gryczynski, Ignacy Gryczynski, and joseph R Lakowicz. Basics of fluorescence and fret. *Molecular imaging: FRET microscopy and spectroscopy*, pages 21–56, 2005.

Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001. ISSN 13507265. http://www.jstor.org/stable/3318737.

W. Heitler. *The Quantum Theory of Radiation*. Dover Books on Physics and Chemistry. Dover Publications, 1954. ISBN 9780486645582. http://books.google.dk/books?id=L7w7UpecbKYC.

I. Johnson. *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies, 11th Edition*. Life Technologies Corporation, 2010. ISBN 9780982927915. http://books.google.dk/books?id=djuacQAACAAJ.

AK Kenworthy and M Edidin. Distribution of a glycosylphosphatidylinositol-anchored protein at the apical surface of mdck cells examined at a resolution of <100 Å using imaging fluorescence resonance energy transfer. *The Journal of cell biology*, 142(1):69–84, 1998.

Anne K. Kenworthy. Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods*, 24(3):289–296, 2001. ISSN 1046-2023. http://www.sciencedirect.com/science/article/pii/S1046202301911892.

Evgeny Krissinel and Kim Henrick. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372(3):774–797, 2007. ISSN 0022-2836. http//www.sciencedirect.com/science/article/pii/S0022283607006420.

Joseph R Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2009.

Luís M. S. Loura, Fábio Fernandes, and Manuel Prieto. Membrane microheterogeneity: Förster resonance energy transfer characterization of lateral membrane domains. *European Biophysics Journal*, 39(4):589–607, 2010. ISSN 1432-1017. http://dx.doi.org/10.1007/s00249-009-0547-5.

Luís MS Loura and Manuel Prieto. Fret in membrane biophysics: An overview. *Frontiers in physiology*, 2, 2011.

A. Miyawaki, A. Sawano, T. Kogure, and Others. Lighting up cells: labelling proteins with fluorophores. *Nature Cell Biology*, 2003.

Jesper Møller and Rasmus Plenge Waagepetersen. *Statistical inference and simulation for spatial point processes*. CRC Press, 2003.

Jesper Møller, Anthony N Pettitt, R Reeves, and Kasper Klitgaard Berthelsen. An efficient markov chain monte carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458, 2006.

I. Murray, Z. Ghahramani, and D. J. C. MacKay. MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pages 359–366. AUAI Press, 2006.

James Pawley. *Handbook of biological confocal microscopy*. Springer, 2006a.

James Pawley. Fundamental limits in confocal microscopy. In James Pawley, editor, *Handbook of biological confocal microscopy*, pages 20–42. Springer, 2006b.

Ammasi Periasamy and Richard Day. *Molecular imaging: FRET microscopy and spectroscopy*. Access Online via Elsevier, 2011.

Ammasi Periasamy, Horst Wallrabe, Ye Chen, and Margarida Barroso. Chapter 22 quantitation of protein-protein interactions: Confocal fret microscopy. In John J. Correia and H. William III Detrich, editors, *Biophysical Tools for Biologists, Volume Two: In Vivo Techniques*, volume 89 of *Methods in Cell Biology*, pages 569–598. Academic Press, 2008. http://www.sciencedirect.com/science/article/pii/S0091679X08006225.

Sophie E Polo and Stephen P Jackson. Dynamics of dna damage response proteins at dna breaks: a focus on protein modifications. *Genes & development*, 25(5):409–433, 2011.

J.D. Puglisi. *Structure, Dynamics and Function of Biological Macromolecules and Assemblies*. NATO Science. IOS Press, 2005. ISBN 9781586034757. https://books.google.dk/books?id=wK0oAAAAYAAJ.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. http://www.R-project.org/.

Michael R. Raicu, Valerica Stoneman, Russell Fung, Mike Melnichuk, David B. Jansma, Luca F. Pisterzi, Sasmita Rath, Michael Fox, James W. Wells, and Dilano K. Saldin. Determination of supramolecular structure and spatial distribution of protein complexes in living cells. *Nature Photonics*, 3:107–113, 2009.

K.K. Rohatgi-Mukherjee. *Fundamentals of Photochemistry*. A Halsted Press book. Wiley, 1978. ISBN 9780852267844. http://books.google.dk/books?id=TgOtDapLvEEC.

Shoji Shima and Heiichi Sakai. Polylysine produced by streptomyces. *Agricultural and Biological Chemistry*, 41(9):1807–1809, 1977.

Yuansheng Sun, Horst Wallrabe, Soo-Ah Seo, and Ammasi Periasamy. Fret microscopy in 2010: The legacy of theodor förster on the 100th anniversary of his birth. *European Journal of Chemical Physics and Physical Chemistry*, 12(3):462–474, 2011. ISSN 1439-4235.

E. G. van Putten, D. Akbulut, J. Bertolotti, W. L. Vos, A. Lagendijk, and A. P. Mosk. Scattering lens resolves sub-100 nm structures with visible light. *Phys. Rev. Lett.*, 106: 193905, May 2011. http://link.aps.org/doi/10.1103/PhysRevLett.106.193905.

H Wallrabe, G Bonamy, A Periasamy, and M Barroso. Receptor complexes cotransported via polarized endocytic pathways form clusters with distinct organizations. *Molecular biology of the cell*, 18(6):2226–2243, 2007.

Horst Wallrabe and Ammasi Periasamy. Imaging protein molecules using fret and flim microscopy. *Current opinion in biotechnology*, 16(1):19–27, 2005.

Horst Wallrabe, Masilamani Elangovan, Almut Burchard, Ammasi Periasamy, and Margarida Barroso. Confocal fret microscopy to measure clustering of ligand-receptor complexes in endocytic membranes. *Biophysical journal*, 85(1):559–571, 2003.

Horst Wallrabe, Ye Chen, Ammasi Periasamy, and Margarida Barroso. Issues in confocal microscopy for quantitative fret analysis. *Microscopy research and technique*, 69(3): 196–206, 2006.

S. Welch. *Transferrin: The Iron Carrier*. Taylor & Francis, 1992. ISBN 9780849367939. http://books.google.dk/books?id=BeyO7-wQVe8C.

PK Wolber and BS Hudson. An analytic solution to the förster energy transfer problem in two dimensions. *Biophysical journal*, 28(2):197–210, 1979.

P.G. Wu and L. Brand. Resonance energy transfer: Methods and applications. *Analytical Biochemistry*, 218(1):1–13, 1994. ISSN 0003-2697. http://www.sciencedirect.com/science/article/pii/S0003269784711341.

Tomasz Zal and Nicholas RJ Gascoigne. Photobleaching-corrected Fret efficiency imaging of live cells. *Biophysical journal*, 86(6):3923–3939, 2004.

Tomasz Zal, M Anna Zal, and Nicholas RJ Gascoigne. Inhibition of T cell receptor-coreceptor interactions by antagonist ligands visualized by live fret imaging of the T-hybridoma immunological synapse. *Immunity*, 16(4):521–534, 2002.

Timo Zimmermann, Jens Rietdorf, and Rainer Pepperkok. Spectral imaging and its applications in live cell microscopy. *FEBS Letters*, 546(1):87–92, 2003. ISSN 0014-5793. http://www.sciencedirect.com/science/article/pii/S0014579303005210.