This is the accepted version of a paper published in *Journal of Microscopy*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Single particle raster image analysis of diffusion for particle mixtures
*Journal of Microscopy*, 269(3): 269-281
https://doi.org/10.1111/jmi.12625

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

This is the peer reviewed version of the following article: LONGFILS, M., RÖDING, M., ALTSKÄR, A., SCHUSTER, E., LORÉN, N., SÄRKKÄ, A. and RUDEMO, M. (2018), Single particle raster image analysis of diffusion for particle mixtures. *Journal of Microscopy*, 269: 269-281, which has been published in final form at https://doi.org/10.1111/jmi.12625. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-33367
Single particle raster image analysis of diffusion for particle mixtures

Marco Longfils*  Magnus Röding†  Annika Altskär†
Erich Schuster†  Niklas Lorén†  Aila Särkkä*
Mats Rudemo*
July 17, 2017

*Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg,
Gothenburg, Sweden
†RISE Bioscience and Materials, Gothenburg, Sweden

Key words: bootstrap, confocal laser scanning microscopy, diffusion, fluorescent beads, maximum likelihood, particle mixtures, single particle tracking.

Summary

Recently we complemented the Raster Image Correlation Spectroscopy (RICS) method of analyzing raster images via estimation of the image correlation function with the method Single Particle Raster Image Analysis (SPRIA). In SPRIA, individual particles are identified and the diffusion coefficient of each particle is estimated by a maximum likelihood method. In this paper, we extend the SPRIA method to analyse mixtures of particles with a finite set of diffusion coefficients in a homogeneous medium. In examples with simulated and experimental data with two and three different diffusion coefficients, we show that SPRIA gives accurate estimates of the diffusion coefficients and their proportions. A simple technique for finding the number of different diffusion coefficients is also suggested. Further, we study the use of RICS for mixtures with two different diffusion coefficients and investigate by plotting level curves of the correlation function how large the quotient between diffusion coefficients needs to be in order to allow discrimination between models with one and two diffusion coefficients. We also describe a minor correction (compared to published papers) of the RICS autocorrelation function.

*Correspondence (including request for programs) to Marco Longfils. E-mail:longfils@chalmers.se
Introduction

A well established method to study mass transport in biomaterials using a confocal laser scanning microscope is Raster Image Correlation Spectroscopy (RICS) (Digman et al., 2005). RICS is a powerful method based on correlation function estimation. It has also been applied to models of diffusion and binding (Digman & Gratton, 2009) and to mapping spatial heterogeneity (Schuster et al., 2016)-(Hendrix et al., 2016). In various areas such as polymer physics and sizing of biomolecules, it is useful to determine diffusion properties of a sample of particles consisting of a mixture of components with different diffusion coefficients. Hence, methods to simultaneously estimate all the diffusion coefficients are needed. For example, fluctuation correlation spectroscopy has been used to measure the diffusion coefficients of gold nanoparticles in semidilute polyethylene glycol as a function of their radius (not measured simultaneously) (Kohli & Mukhopadhyay, 2012). Many systems are characterized by the presence of polydisperse mixtures. In particular, the majority of proteins fulfill their biological roles not as monomeric species but as components of larger functional complexes. Microfluidics has been used as a platform for sizing of biomolecules by measuring diffusion, thus enabling determination of their hydrodynamic radius (Arosio et al., 2016). In the present paper, we present an extension of the Single Particle Raster Image Analysis (SPRIA) method (Longfils et al., 2017) to study systems of particles having a finite number of different diffusion coefficients. A simple method to select the number of components in the mixture model is also proposed, while more formal techniques will be a subject of a future study.

In the case of applying RICS to particle mixtures, we show by plotting the level curves of different correlation functions that the quotient between different diffusion coefficients needs to be rather large to be able to distinguish between models with two components and only one component. By using simulated data, we also investigate the performance of RICS when the number of components is known and no model selection is needed. In the following section, we briefly describe the raster scanning procedure of images with a confocal laser scanning microscope, and extend the SPRIA method to study systems of particles with different diffusion coefficients. A minor correction (compared to published papers) of the RICS autocorrelation function is presented, and we delineate a RICS modelling of particle mixtures. Then, we present the results of the SPRIA method applied to both simulated and experimental data and validate a simple model selection criterion. Finally, the applicability of RICS for such systems is discussed by use of plots of correlation function level curves and by aid of simulated data. Experimental results for RICS are provided in the Supplementary material.
Image analysis, statistical methods and experimental setup

Raster scanning of images

Raster images used in both SPRIA and RICS are collected by sampling the region of interest according to a raster pattern. The observation volume is initially centered at the first (from left to right) pixel of the image. Then, after the pixel dwell time $\tau_p$, the second pixel in the first line is scanned. Scanning pixel-by-pixel, the first line of the image is collected. In the next step, after the line time $\tau_l$, the observation volume is retraced from the beginning of the second line of pixels. Thus the second line is recorded, and by iterating this process the whole image is sampled.

The SPRIA method with one diffusion coefficient

The principles of SPRIA are briefly recalled here and the reader is referred to (Longfils et al. 2017) for more details. In the recommended experimental setup for SPRIA, it should be possible to identify the individual particles and the time between two consecutive lines should be such that a particle moves non-negligibly. SPRIA consists of two parts: first, the particles are extracted through simple image segmentation, and second, the trajectory of each particle in a single image is estimated providing an estimate of the diffusion coefficient. Here, we introduce a slight modification of the estimation procedure of the trajectory of single particles. Let us define, as in (Longfils et al. 2017), a particle $P$ as an axis-parallel rectangle

$$P = \{ (x,y) : a < x < a + L, \ b < y < b + K \}$$

around a local maximum of photon counts. Here $(a,b)$ is the position of the top-left pixel in the corner of the rectangle $P$ in the image, and $L$ and $K$ are the lengths of its sides. The trajectory of a particle can then be estimated based on the extracted image, see Figure 1. Let $N(x,y,t_k)$ denote the measured number of photons for a given particle at the pixel with center $(x,y)$ at time $t_k$, where $t_k = t_k(y)$, $k = 0, \ldots, K$, is the time at which we observe the horizontal line at $y$. Furthermore, let

$$S = (S(t_0), \ldots, S(t_K))$$

denote the vector of particle positions at times $t_0, \ldots, t_K$, where $S(t_k) = (X_k, Y_k, Z_k)$. (Longfils et al. 2017) introduced a maximum likelihood method to estimate the $x$ position of the particle on each line. Here, we introduce an alternative and more direct way to estimate the $x$ position, namely

$$\psi_k = \frac{\sum_{\{x:(x,y)\in P\}} N(x,y,t_k) \cdot x}{\sum_{\{x:(x,y)\in P\}} N(x,y,t_k)}.$$
Let $\Delta t = \tau$ denote the time needed to scan a line. Then,

$$
\tilde{D} = \frac{1}{2\Delta t K} \sum_{k=1}^{K} (X_k - X_{k-1})^2
$$

(4)

is an unbiased estimator of the diffusion coefficient $D$. However, we cannot observe the $X_k$ variables and can thus not compute $\tilde{D}$. By noting that $\psi_k$ in Equation (3) should be close to $X_k$, we find that

$$
\hat{D} = \frac{1}{2\Delta t K} \sum_{k=1}^{K} (\psi_k - \psi_{k-1})^2
$$

(5)

is a suitable observable estimate of the diffusion coefficient $D$. A comparison of the method presented here to estimate the position and the one used in [Longfils et al. 2017] is illustrated in Figure 1. More details can be found in Supplementary Table 8.

Figure 1: A simulated raster scan image of a particle with the true trajectory (red), the corresponding estimated trajectory computed using the maximum likelihood method from [Longfils et al. 2017] (green), and the centroid based method (black) presented in the present paper.
Consider a diffusion trajectory in the $x$ direction with $K+1$ locations $X_0, X_1, \ldots, X_K$, i.e. $K$ steps, corresponding to consecutive observations with a time interval $\Delta t = \tau$ and diffusion coefficient $D$. We will first consider the estimate $\hat{D}$ of $D$. The probability density of $\hat{D}$ can be obtained in the following way. From the assumption that the increments $X_k - X_{k-1}$ have independent normal distributions with mean zero and variance $2D\Delta t$, it follows that

$$\chi^2 = \frac{1}{2D\Delta t} \sum_{k=1}^{K} (X_k - X_{k-1})^2$$

is chi-square distributed with $K$ degrees of freedom and probability density

$$f_{\chi^2}(y) = \frac{2^{-\frac{K}{2}}}{\Gamma(\frac{K}{2})} y^{\frac{K}{2}-1} e^{-\frac{y}{2}},$$

where $\Gamma(\cdot)$ is the gamma function. Further, $\hat{D} = \frac{D}{K}\chi^2$ and with the transformation $y = \frac{D}{K}x$ we find that

$$f_D(y) = f_{\chi^2}(x) \frac{dx}{dy} = f_{\chi^2}\left(\frac{Ky}{D}\right) \frac{K}{D},$$

which gives

$$f_D(y) = \left(\frac{K}{\Gamma(K)}\right)^{\frac{K}{2}} y^{\frac{K}{2}-1} e^{-\frac{yK}{2D}}.$$

Letting $f_\gamma$ denote the gamma density,

$$f_\gamma(y; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y},$$

we can write

$$f_D(y) = f_\gamma\left(y; \frac{K}{2}, \frac{K}{2D}\right).$$

The likelihood function for the model with one diffusion coefficient

Suppose that we have observed $J$ independent particles with $K_j$ steps for particle $j$, $j = 1, \ldots, J$, and let $\tilde{D}_j$, $j = 1, \ldots, J$, denote the corresponding values of $\hat{D}$. We will first assume that all particles have the same diffusion coefficient $D$. Considering $\tilde{D}_j$, $j = 1, \ldots, J$, as observations we find that the likelihood function for the diffusion coefficient $D$ is

$$L(D) = \prod_{j=1}^{J} f_\gamma(\tilde{D}_j; \frac{K_j}{2}, \frac{K_j}{2D}) = \prod_{j=1}^{J} \left(\frac{K_j}{\Gamma(K_j/2)}\right) \frac{K_j}{D_j^{K_j/2}} \frac{1}{D_j} \frac{1}{\Gamma(K_j/2)} \frac{1}{e^{\frac{D_j K_j}{2}}}$$

and the corresponding log-likelihood $\ell(D) = \log(L(D))$ is

$$\ell(D) = C - \log(D) \sum_{j=1}^{J} \frac{K_j}{2} - \frac{1}{D} \sum_{j=1}^{J} \frac{K_j \tilde{D}_j}{2},$$
where
\[ C = \sum_{j=1}^{J} \left( \frac{K_j}{2} - 1 \right) \log(\hat{D}_j) - \sum_{j=1}^{J} \log \left( \Gamma \left( \frac{K_j}{2} \right) \right) + \sum_{j=1}^{J} \frac{K_j}{2} \log \frac{K_j}{2}. \] (14)

The log-likelihood \( \ell(D) \) is maximized for \( D = \hat{D} \), where
\[ \hat{D} = \frac{\sum_{j=1}^{J} K_j \hat{D}_j}{\sum_{j=1}^{J} K_j}. \] (15)
We see that in the estimation of \( D \), the diffusion coefficient estimates \( \hat{D}_j \) are quite naturally weighted according to the corresponding number \( K_j \) of steps.

The second derivative of \( \ell(D) \) is
\[ \ell''(D) = \frac{1}{D^2} \sum_{j=1}^{J} \frac{K_j}{2} - \frac{2}{D^3} \sum_{j=1}^{J} \frac{K_j \hat{D}_j}{2} \] (16)
and the corresponding Fisher information is
\[ J_F(\hat{D}) = -\mathbb{E}[\ell''(\hat{D})] = \frac{1}{\hat{D}^2} \sum_{j=1}^{J} \frac{K_j}{2}. \] (17)

The asymptotic variance of \( \hat{D} \) is then
\[ \frac{1}{J_F(\hat{D})} = \frac{2 \hat{D}^2}{\sum K_j}. \] (18)
It can be noted that in the present case we can compute the variance of \( \hat{D} \) exactly. We have that \( \text{Var}(\hat{D}_j) = 2D^2/K_j \), which gives
\[ \text{Var}(\hat{D}) = \frac{2D^2}{\sum K_j}. \] (19)

The \( \hat{D}_j \)'s are not observable and are replaced by the estimates \( \hat{D}_j \) in Equation (12) and subsequent formulae. Then \( \ell(D) \) in Equation (12) becomes an approximate log-likelihood and relations such as Equation (18) should hold approximately.

**SPRIA applied to a mixture with different diffusion coefficients**

The likelihood function for a mixture of diffusion coefficients

Suppose now that we have diffusion coefficients drawn independently from a mixture distribution with distribution function \( F_D \). If the corresponding density exists it is denoted \( f_D(x) \). The density in Equation (11) of the diffusion coefficient estimator \( \hat{D} \) is replaced by
\[ f_{\hat{D}}(y) = \int f_\gamma \left( y; \frac{K}{2}, \frac{K}{2x} \right) F_D(dx), \] (20)
which for a continuous mixing distribution can be written as

\[ f_D(y) = \int f_f(y; K \frac{2}{2x}) f_D(x) dx. \]  \hfill (21)

Here, we restrict ourselves to models consisting of a finite mixture of \( M \) unknown diffusion coefficients \( D_1, \ldots, D_M \) with unknown proportions \( \pi_1, \ldots, \pi_M \) satisfying \( \pi_i \geq 0 \) and \( \sum_i \pi_i = 1 \). Thus the number of free parameters is

\[ n_{par} = 2M - 1. \]  \hfill (22)

With \( \pi = (\pi_1, \ldots, \pi_M) \) and \( D = (D_1, \ldots, D_M) \) the log-likelihood takes the form

\[ \ell(\pi, D) = \sum_{j=1}^{J} \log \left( \sum_{i=1}^{M} \frac{\pi_i (K_j^2)}{\Gamma(\frac{K_j^2}{2})} D_j^{-\frac{K_j^2}{2}} e^{-\frac{\pi_i}{2D_j}} \right) . \]  \hfill (23)

**Standard errors, model choice and bootstrap for SPRIA mixture models**

Standard errors can be computed using the asymptotic variance obtained from the derivatives of the log-likelihood function. However, from the simulations and experiments with fluorescent beads it turns out, see the results below, that the standard errors computed in this way are too small. Furthermore, the models corresponding to different numbers of components in the mixture are nested. Therefore, choosing between a model and the same model with one component added could be based on the fact that, according to Wilks’ theorem (Wilks 1938), twice the log-likelihood differences between the models are approximately chi-square distributed with two degrees of freedom, compare Equation (22). However, this procedure often leads to models with too many components.

Here, we compute standard errors using a bootstrap method (Efron & Tibshirani 1993), where we resample the identified particles. As we will see from the results with simulations and experiments, this method of estimating standard errors performs well.

The choice of mixture model order, i.e. the number of components with different diffusion coefficients, turns out to be more complicated and will be discussed in more detail in a future work. In this paper, we suggest a simple model choice rule, namely to accept the model with the largest number of components such that the estimated proportions are all greater than a certain threshold value. As shown in the results section below, a limit of at least 15% performs reasonably well in cases with two or three types of particles. The 15% rule should be reconsidered in case of samples with more than three types of particles.

To illustrate the 15% rule we show in Figure 2 the histogram of the distribution of \((\hat{D}_1, \ldots, \hat{D}_J)\) from an experiment with \( J = 3388 \) particles in a mixture of 175nm and 1000nm particles together with the SPRIA diffusion coefficient estimates for the models with 1, 2, 3 or 4 components. The histogram is clearly bimodal, indicating that a mixture
of two components would be preferable compared to a model with a single component. Here, the log-likelihood difference testing would incorrectly select 4 components. This is not surprising, because, as indicated in Figure 2, adding components will always give an equal or slightly better fit to the histogram. On the other hand, we observe that the extra components have very small proportions (in this case 3.0% for 3 components and 2.4% and 2.2% for 4 components). The 15% rule performs well here, and gives two components.

Figure 2: Histogram of the SPRIA estimates ($\hat{D}_1, \ldots, \hat{D}_J$) and the SPRIA estimated diffusion coefficients for the mixture models with 1 (green), 2 (red), 3 (black), and 4 (blue) components from an experiment with two true components. The expected diffusion coefficients according to the Stokes-Einstein equation are 0.48 and 2.5 with proportions 0.476 and 0.524.
The RICS autocorrelation function

The correlation function for RICS corresponding to two points \((x, y)\) and \((x + \xi, y + \psi)\) is of the form

\[
G(\xi, \psi) = \frac{1}{\langle N \rangle} e^{\left[\frac{(S\xi)^2 + (S\psi)^2}{\omega_0^2 + 4D\tau(\xi, \psi)}\right]} \left(1 + \frac{4D\tau(\xi, \psi)}{\omega_0^2}\right)^{-1} \left(1 + \frac{4D\tau(\xi, \psi)}{\omega_0^2}\right)^{-1/2},
\]

where \(\langle N \rangle\) is the average number of particles in the observation volume, \(S\) is the pixel size and the function \(\tau(\xi, \psi)\), corresponding to the time between the two points, takes slightly different forms in different papers. Thus in \(\text{Digman et al. 2005}\) it is given as

\[
\tau(\xi, \psi) = \tau_p\xi + \tau_l\psi,
\]

where \(\tau_p\) and \(\tau_l\) are, respectively, the pixel dwell time and the line time, while in \(\text{Brown et al. 2008}\) it is given as

\[
\tau(\xi, \psi) = \tau_p|\xi| + \tau_l|\psi|.
\]

We find that the correct form is

\[
\tau(\xi, \psi) = |\tau_p\xi + \tau_l\psi|.
\]

corresponding to the time it takes to move between the points \((x, y)\) and \((x + \xi, y + \psi)\).

We note that if \(\psi < 0\), or if \(\psi = 0\) and \(\xi < 0\), then we reach the point \((x + \xi, y + \psi)\) before \((x, y)\), otherwise we reach the point \((x, y)\) before \((x + \xi, y + \psi)\).

Note that for \(G(\xi, \psi)\) to be a valid autocorrelation function it, and also the function \(\tau(\xi, \psi)\), need to be invariant under the transformation \((\xi, \psi) \rightarrow (-\xi, -\psi)\). Note also that for positive \(\xi\) and \(\psi\) the formulas \((25)\) and \((27)\) coincide but otherwise \((25)\) needs modification and we thus find that the correct modification is \((27)\) rather than \((26)\).

By plotting the contour levels of the correlation functions we compare in Figure 3 the autocorrelation functions corresponding to \(\tau(\xi, \psi)\) given by \((26)\) and \((27)\) for typical parameter values. We see that for these values the difference between the autocorrelation functions is small. But if the quotient between \(\tau_p\) and \(\tau_l\) gets larger the difference could increase, and under all circumstances it seems preferable to use the correct formula \((27)\).
RICS for a mixture of $M$ components

RICS has previously been applied to study mixtures of diffusing fluorophores in (Sanabria et al. 2008). There, a model with only one diffusing component produced a poor fit to the data and the addition of a slower component was required. The faster component, corresponding to enhanced green fluorescent protein, was found to have diffusion coefficient $D_1 = 20.1 \mu m^2 s^{-1}$ and the slower, possibly caused by interaction or corresponding to large molecular complexes, had diffusion coefficient $D_2 \sim 0.4 \mu m^2 s^{-1}$.

Consider the following theoretical correlation function for a mixture of $M$ diffusion coefficients

$$G(\xi, \psi, \theta) = \sum_{m=1}^{M} \alpha_m e^{-\left(\frac{(S\xi)^2+(S\psi)^2}{w_0^2+4D_m(|\tau_p\xi+\gamma\psi|)}\right)} \left(1 + \frac{4D_m(|\tau_p\xi+\gamma\psi|)}{w_0^2}\right)^{-\frac{1}{2}} + \gamma,$$

where $\theta = (\alpha, \gamma, D)$, $\alpha = (\alpha_1, \ldots, \alpha_M)$ is the vector of amplitudes for each component, $\gamma$ is an offset, $D = (D_1, \ldots, D_M)$ is the vector of diffusion coefficients, and $\tau_p, \tau_l$ and $S$ are, respectively, the pixel dwell time, the line time and the pixel size. In particular, the amplitude $\alpha_m$ depends on the particle brightness $\eta_m$ and the average concentration of
such particles $\langle C_m \rangle$ as

$$
\alpha_m = \frac{1}{V_{\text{eff}}} \frac{\eta_m^2 \langle C_m \rangle}{(\sum_{l=1}^{K} \eta_l \langle C_l \rangle)^2},
$$

(29)

where $V_{\text{eff}} = \pi^2 w_0^2 w_z$ is the effective volume of the point spread function. Consider

$$
\beta_m = \frac{\alpha_m}{\sum_{j=1}^{K} \alpha_j} = \frac{\eta_m^2 \langle C_m \rangle}{\sum_{l=1}^{K} \eta_l^2 \langle C_l \rangle},
$$

(30)

which can be interpreted as the contribution of the $m$-th species to the overall correlation function of the sample. These contributions sum up to 1 and they are proportional to the concentration and brightness of each species. Let $\hat{\theta} = (\hat{D}, \hat{\alpha}, \hat{\gamma})$ be the weighted least squares estimate of $\theta$, i.e.

$$
Q_M(\theta) = \sum_{(\xi, \psi) \in A} w(\xi, \psi, I) \left[ \hat{C}(\xi, \psi, I) - G(\xi, \psi, \theta) \right]^2,
$$

(31)

where $\hat{C}(\xi, \psi, I)$ is the empirical correlation function of the stack of images $I$, $w(\xi, \psi, I) = \left[ \text{Var}(\hat{C}(\xi, \psi)) \right]^{-1}$, and

$$
\hat{\theta} = \arg \min_{\theta} Q_M(\theta).
$$

(32)

The residual sum of squares is then $Q_M(\hat{\theta})$.

Concerning standard errors for the RICS method we found in (Longfils et al. 2017) that standard errors computed by bootstrapping images worked well for systems of particles with identical diffusion coefficients. However, for systems of particles with mixed diffusion coefficients studied in the present paper it turns out that this method in most cases gives unrealistically small standard errors, and therefore, they are not included here in the tables with RICS results for mixtures.

**Computational details**

All the programs were developed in Matlab (Matlab R2015b, MathWorks, Natick, Massachusetts) and run on a desktop PC. The computations for the RICS mixture models took a few minutes to run. The SPRIA method is computationally more demanding than RICS, but it never took more than one hour on a 3.2 GHz dual core computer to obtain the results for the analyses of the simulated and experimental data described below. The centroid based method introduced in this paper is faster than the maximum likelihood method used in (Longfils et al. 2017). In particular, the maximum likelihood method would have taken about a day to analyse the largest dataset presented here. The large number of detected particles, together with the cost of maximizing a complex function, lead to low computational efficiency. The centroid in Equation 3 can be quickly computed by taking advantage of matrix multiplication.
Simulation study setup

In the main simulation study, particles trajectories were simulated using Gaussian random walk (discrete-time Brownian motion) generated in a box with periodic boundary conditions. The mixtures had two or three components with bead diameters chosen among the values of 50nm, 100nm, 175nm, 500nm, and 1000nm and the diffusion coefficients were obtained by using the Stokes-Einstein equation assuming the particles were in water. The side length of the box was two times the side length of the image to avoid edge effects. Image data were collected by simulating a raster scan pattern and Poisson distributed photon counts were generated. An image size of 256×256 was chosen. Furthermore, 100 images were simulated, as suggested for good RICS performance (Digman et al. 2005, Brown et al. 2008). To facilitate comparison, parameters in the simulation study were kept close to the experimental setup settings. Lateral and axial waists of the point spread function were 248nm and 1270nm, respectively, as obtained through a Gaussian fit to the averaged z-scan of immobile 175nm fluorescent beads, as described in the Experimental setup section below. The pixel dwell time was \( \tau_p = 1.71 \times 10^{-7} \text{s} \), the line time \( \tau_l = 1.44 \times 10^{-3} \text{s} \), and the pixel size \( S_x = 0.03 \mu\text{m} \).

To choose the values for \( \tau_p \) and \( \tau_l \) for the SPRIA analysis, let \( S_s \) denote the size (diameter) of a bead and \( S_x \) the pixel size. Then it takes the time \( \Delta t = \tau_p S_s / S_x \) to scan a bead. The root mean square distance traveled in the \( x \)-direction during the time \( \Delta t \) is \( \text{RMSD} = \sqrt{2 \Delta t D} \). To call the motion of a bead during the line scan of the bead small we require that the quotient

\[
Q_{\text{val}} = \frac{\text{RMSD}}{S_s} = \frac{\sqrt{2 \tau_p D}}{(S_x S_s)} < 0.025.
\]

One can choose \( \tau_p \) so that Equation 33 is satisfied and let values of \( \tau_l \) be \( 10^3 - 10^4 \) times longer than the pixel dwell times to allow particles to move significantly between the lines.

Experimental setup with fluorescent beads

A Leica SP5 AOB (Heidelberg, Germany) confocal laser scanning microscope (CLSM) with a 63x, 1.2 NA water objective and a GaAsP hybrid detector was used for the raster image measurements. The experiments were carried out with the 488nm line of an argon laser having 10 \( \mu\text{W} \) laser power on stage. The motion of very dilute solutions of three different types of fluorescent beads were recorded: FluoSpheres carboxylate-modified (ex/em at 505/515 nm) of 175nm (actual size 175 \( \pm \) 5 nm, determined by the supplier) diameter (P7220, Lifetechnologies, Eugene, Oregon), TetraSpeck microspheres of 1000nm (actual size 1062 \( \pm \) 35 nm) diameter (T7282, Lifetechnologies, Eugene, Oregon), and molecular probes of 500nm (actual size 490nm) diameter (T7281, Lifetechnologies, Eugene, Oregon) were investigated. All measurements were carried out at ambient conditions (\( \approx \)298K). The fluorescent beads were measured either as a sole solution or as a mixture of two or three bead sizes. The beads were first sonicated in order to separate
aggregates. Further, the mixture solutions were also vortexed to ensure mixing. Furthermore, 7µl of microsphere solution was loaded onto a cover-glass slide and locked in a sandwich manner by a second cover-glass slide by using Secure-Seal spacers (MolecularProbes, Eugene, Oregon). The apparent axial and lateral waists of the point spread function were determined by fitting a 3-dimensional Gaussian to a z-stack series of 175nm immobilized beads (Cole et al. 2011). The beads were immobilized in a 5 % w/w gelatin gel (G-1890, Sigma-Aldrich, Stockholm, Sweden) and the z-stack was imaged with a step size of 42nm on a Leica Super-z galvanometer stage (Heidelberg, Germany).

Results

Simulation study using SPRIA

In Table 1 we report results for the different simulation settings with mixtures of particles. The estimated diffusion coefficients for the components are generally close to the correct values, and, except for the mixtures of 100/1000nm and 50/175nm beads, also the proportions are well estimated. The two cases with 50nm particles stand out as they show a negative bias for the diffusion coefficient estimates. The column Comp15 gives the number of components selected if we use the criterion described above with a minimum proportion for each component of 15%. The two cases where we select three components had a third component with an estimated proportion of 18% and 22%, respectively. The standard error estimates are also reasonable although they are too low in some cases. Thus, the corresponding 95% confidence intervals cover the true $D_1$ in 6 out of 8 cases, the true $D_2$ value in 2 out of 8 cases, and the true proportion value 0.5 in 5 out of 8 cases. The standard errors in Table 1 are computed by bootstrapping. Alternatively one could use the asymptotic variance computed from the derivatives of the log-likelihood function to estimate standard errors. However, as mentioned before, standard errors computed from the likelihood were generally too small. For the simulations described in Table 1 it gave standard errors that were about half the size of the bootstrap standard errors (not shown) and consequently gave too short confidence intervals.
Table 1: Results from SPRIA simulations with two component bead mixture models with equal proportions 0.5 for each component. Sizes gives the bead diameters, \( N \) is the observed number of particles, \( D_1 \) and \( D_2 \), where \( D_1 < D_2 \), are the corresponding true diffusion coefficients, and \( D_{SPRIA1} \) and \( D_{SPRIA2} \) are the corresponding estimated diffusion coefficients for the mixture model with two components. \( C_{SPRIA1} \) is the estimated proportion for particles with diffusion coefficient \( D_{SPRIA1} \) and \( D_{SPRIA} \) is the diffusion coefficient estimated for the single component model. The values after the ± signs are standard errors computed by bootstrapping. Comp15 shows the number of components selected if we use the criterion described above with a minimum proportion for each component of 15%.

<table>
<thead>
<tr>
<th>Sizes</th>
<th>( N )</th>
<th>( D_1(\mu m^2s^{-1}) )</th>
<th>( D_2(\mu m^2s^{-1}) )</th>
<th>( D_{SPRIA1}(\mu m^2s^{-1}) )</th>
<th>( D_{SPRIA2}(\mu m^2s^{-1}) )</th>
<th>( C_{SPRIA1} )</th>
<th>( D_{SPRIA} )</th>
<th>Comp15</th>
</tr>
</thead>
<tbody>
<tr>
<td>100/175</td>
<td>1364</td>
<td>2.5</td>
<td>4.8</td>
<td>2.48 ± 0.11</td>
<td>4.83 ± 0.20</td>
<td>0.53 ± 0.05</td>
<td>3.52 ± 0.05</td>
<td>2</td>
</tr>
<tr>
<td>100/500</td>
<td>940</td>
<td>0.98</td>
<td>4.8</td>
<td>1.01 ± 0.02</td>
<td>5.20 ± 0.12</td>
<td>0.46 ± 0.02</td>
<td>2.01 ± 0.02</td>
<td>2</td>
</tr>
<tr>
<td>100/1000</td>
<td>739</td>
<td>0.48</td>
<td>4.8</td>
<td>0.52 ± 0.02</td>
<td>4.84 ± 0.19</td>
<td>0.38 ± 0.02</td>
<td>1.43 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td>175/500</td>
<td>942</td>
<td>0.98</td>
<td>2.5</td>
<td>0.99 ± 0.02</td>
<td>2.82 ± 0.11</td>
<td>0.53 ± 0.02</td>
<td>1.48 ± 0.01</td>
<td>3</td>
</tr>
<tr>
<td>175/1000</td>
<td>1415</td>
<td>0.48</td>
<td>2.5</td>
<td>0.44 ± 0.02</td>
<td>2.05 ± 0.15</td>
<td>0.47 ± 0.03</td>
<td>0.87 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td>500/1000</td>
<td>436</td>
<td>0.48</td>
<td>0.98</td>
<td>0.46 ± 0.01</td>
<td>1.27 ± 0.05</td>
<td>0.50 ± 0.03</td>
<td>0.80 ± 0.01</td>
<td>3</td>
</tr>
<tr>
<td>50/175</td>
<td>2343</td>
<td>2.5</td>
<td>9.6</td>
<td>2.18 ± 0.05</td>
<td>8.86 ± 0.28</td>
<td>0.69 ± 0.01</td>
<td>4.98 ± 0.11</td>
<td>2</td>
</tr>
<tr>
<td>50/100</td>
<td>2224</td>
<td>4.8</td>
<td>9.6</td>
<td>3.20 ± 0.24</td>
<td>7.63 ± 0.37</td>
<td>0.54 ± 0.06</td>
<td>5.37 ± 0.08</td>
<td>2</td>
</tr>
</tbody>
</table>

In Table 2 we report the results for simulations with three subpopulations of beads. Overall, the precision of the estimates decreases when the number of components increases. However, the point estimates for the diffusion coefficients are still close to the expected values except for the case corresponding to the second to last line in Table 2. The standard error computed from bootstrap are sometimes rather large and the corresponding 95% confidence intervals cover the true \( D_1 \), \( D_2 \), and \( D_3 \) in, respectively, 7, 5, and 7 cases out of 8. The estimated proportions are often far from the expected 0.33 value and usually one of the subpopulations is overrepresented by our model with one proportion as high as 50 – 60%. Finally, in 5 out of 8 cases, the correct number of components is selected by the simple principle proposed.
We studied the performance of the SPRIA method for experimental data consisting of mixtures with various combinations of two (see Table 3) and three (see Table 4) components of different sized beads. Here, the expected diffusion coefficient for each type of bead is calculated via the Stokes-Einstein equation, knowing the viscosity of water at 298K and the actual size of the beads, as given by the manufacturers.

In Table 3 with the 175/1000nm mixtures the two component SPRIA model provides useful point estimates for both the diffusion coefficients and the proportions. For the 175/500nm mixtures SPRIA provides good point estimates for the proportion of the components, but the two diffusion coefficient estimates $D_{SPRIA1}$ and $D_{SPRIA2}$ are both underestimating the two true corresponding quantities, with a large bias for the lower one. The bias of the diffusion coefficient estimates for the 500nm beads is of the order of 30% and consistent when the proportion of the 500nm was changed, compare the three experiments corresponding to lines 4-6 in Table 3. However, we did not find this bias when mixing the same 500nm beads with the 1000nm beads. To check the 500nm beads we performed a separate study, see Table 10 in the supplementary material. For the 500-1000nm beads SPRIA works well, providing precise estimates for both the diffusion coefficients and the proportions. In the last line of Table 3 the standard deviations of the estimates are large, which may partly be due to the relatively small number of particles available. When we look at the estimated diffusion coefficient $D_{SPRIA}$, disregarding that we have a mixture, we obtain a value which is between the two expected diffusion coefficients. It may be noted that when the proportion of the slower particles increases, $D_{SPRIA}$ decreases as expected. As in Table 1, standard errors generally seem slightly
Table 3: Results with SPRIA for experimental two component mixtures of 175nm, 500nm or 1000nm fluorescent beads with varying proportions. Sizes indicates which beads have been mixed in each row and in parentheses we report the expected diffusion coefficients according to the Stokes-Einstein. $N$ is the observed number of particles, $C_{\text{low}}$ is the expected proportion of the larger beads (with lower diffusion coefficient) in the mixture. $D_{\text{SPRIA1}}$ and $D_{\text{SPRIA2}}$ are the two estimated diffusion coefficients for the mixture model and $D_{\text{SPRIA}}$ is the diffusion estimated for the single diffusion coefficient model. $C_{\text{SPRIA1}}$ is the estimated proportion of the component with diffusion coefficient $D_{\text{SPRIA1}}$. The values after the $\pm$ signs are standard errors computed by bootstrapping. Comp15 indicates the number of component selected if we use the criterion described above with a minimum proportion for each component of 15%.

<table>
<thead>
<tr>
<th>Sizes (D)</th>
<th>$N$</th>
<th>$C_{\text{low}}$</th>
<th>$D_{\text{SPRIA1}}$ ($\mu$m$^2$s$^{-1}$)</th>
<th>$D_{\text{SPRIA2}}$ ($\mu$m$^2$s$^{-1}$)</th>
<th>$C_{\text{SPRIA1}}$</th>
<th>$D_{\text{SPRIA}}$</th>
<th>Comp15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 / 175nm (0.48 / 2.5 $\mu$m$^2$s$^{-1}$)</td>
<td>5936</td>
<td>0.23</td>
<td>0.45 ± 0.01</td>
<td>2.35 ± 0.02</td>
<td>0.21 ± 0.01</td>
<td>1.89 ± 0.03</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3388</td>
<td>0.48</td>
<td>0.47 ± 0.01</td>
<td>2.48 ± 0.04</td>
<td>0.41 ± 0.01</td>
<td>1.25 ± 0.02</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1883</td>
<td>0.73</td>
<td>0.42 ± 0.00</td>
<td>2.19 ± 0.05</td>
<td>0.72 ± 0.01</td>
<td>0.65 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td>500 / 175nm (0.98 / 2.5 $\mu$m$^2$s$^{-1}$)</td>
<td>5310</td>
<td>0.33</td>
<td>0.68 ± 0.02</td>
<td>2.17 ± 0.03</td>
<td>0.38 ± 0.01</td>
<td>1.70 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4035</td>
<td>0.60</td>
<td>0.64 ± 0.02</td>
<td>2.03 ± 0.08</td>
<td>0.60 ± 0.02</td>
<td>1.23 ± 0.02</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2922</td>
<td>0.82</td>
<td>0.65 ± 0.01</td>
<td>2.10 ± 0.09</td>
<td>0.79 ± 0.02</td>
<td>0.95 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td>1000 / 500nm (0.48 / 0.98 $\mu$m$^2$s$^{-1}$)</td>
<td>1553</td>
<td>0.38</td>
<td>0.38 ± 0.01</td>
<td>0.84 ± 0.03</td>
<td>0.59 ± 0.02</td>
<td>0.55 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1088</td>
<td>0.65</td>
<td>0.40 ± 0.01</td>
<td>0.90 ± 0.05</td>
<td>0.71 ± 0.03</td>
<td>0.51 ± 0.01</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>811</td>
<td>0.85</td>
<td>0.41 ± 0.18</td>
<td>1.02 ± 0.28</td>
<td>0.86 ± 0.39</td>
<td>0.46 ± 0.01</td>
<td>1</td>
</tr>
</tbody>
</table>

In Table 4 we report the result in the case of a mixture of 175, 500, and 1000nm beads with approximately the same concentrations. The SPRIA diffusion coefficient and proportion estimates seem generally satisfactory although all three diffusion coefficient estimates are slightly lower than the corresponding theoretical values. The standard errors in Table 4 are rather large.
Table 4: Results with SPRIA for a mixture of 175nm, 500nm and 1000nm beads with equal concentrations. \( N \) is the observed number of particles, and \( C_{1000}, C_{500}, \) and \( C_{175} \) are, respectively, the expected proportions of 1000nm, 500nm, and 175nm beads. \( D_{SPRIA1}, D_{SPRIA2} \) and \( D_{SPRIA3} \) are the estimated diffusion coefficients for the mixture model with the expected values in parentheses. \( C_{SPRIA2} \) is the estimated proportion of the species with diffusion coefficient \( D_{SPRIA2} \), and similarly the others. The values after the ± signs are standard errors computed by bootstrap. Comp15 indicates the number of components selected if we use the criterion described above with a minimum proportion for each component of 15%.

<table>
<thead>
<tr>
<th>( N )</th>
<th>( C_{1000} )</th>
<th>( C_{500} )</th>
<th>( C_{175} )</th>
<th>( D_{SPRIA1}(0.48 \mu m^2 s^{-1}) )</th>
<th>( D_{SPRIA2}(0.98 \mu m^2 s^{-1}) )</th>
<th>( D_{SPRIA3}(2.5 \mu m^2 s^{-1}) )</th>
<th>( C_{SPRIA2} )</th>
<th>( C_{SPRIA3} )</th>
<th>Comp15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2622</td>
<td>0.34</td>
<td>0.34</td>
<td>0.32</td>
<td>0.32 ± 0.15</td>
<td>0.84 ± 0.44</td>
<td>2.33 ± 0.70</td>
<td>0.31 ± 0.12</td>
<td>0.24 ± 0.15</td>
<td>3</td>
</tr>
</tbody>
</table>

The correlation between the estimated parameters in SPRIA (not reported here) has been investigated for both experimental and simulated data. Overall, all variables were correlated and the correlation increased as difference between the diffusion coefficients was smaller. The estimate of the different diffusion coefficients were positively correlated and this was consistent in the different mixtures and when changing concentrations.

**Validation of the simple SPRIA model selection criterion**

Table 5 shows the confusion matrix for the model selection criterion and summarises the number of cases (both in simulated and experimental data) where we correctly select the right number of components for mixtures of one, two, or three components with a 15% threshold for the proportions. Overall, we make a mistake in about 28% of the cases, even though some cases might be borderline for the small difference between the diffusion coefficients and variability in the estimate of the proportions.

<table>
<thead>
<tr>
<th># Selected</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td># True</td>
<td>2</td>
<td>1</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

For experimental data, the model selection criterion selects the correct number of components in 11 out of 17 cases. Moreover, in the case of the mixture of 10/90%
volume fraction of 500nm and 1000nm beads the expected proportion of 500nm beads is 15\% which is equal to the 15\% threshold and could explain the selection of only one component.

We have computed correlation coefficients between estimated parameters both for simulated and experimental data. Some results on such correlations are given in the Supplementary material. Thus we find for data with two diffusion coefficients that the correlation between the two diffusion coefficient estimates are typically positive and of the order 0.5 both for simulated and experimental data.

**Discussion about estimating mixtures with RICS**

Although the theory for RICS applied to mixtures is straightforward to obtain from the RICS correlation function, in practice one has to be cautious. Correlation functions of two different models can be very similar, which can lead to wrong estimates for the parameters. For example, let us consider a situation where we do not know whether we have a mixture of two components or just one. Then, our approach could be to fit the correlation function of each model to the empirical correlation function of the data. We would choose the model with the smallest residual sum of squares $Q_M$ as in Equation (31). However, the empirical correlation function is affected by noise and the correlation functions of the two models can be quite similar, and therefore fit the empirical correlation function almost equally well. Thus, model selection and parameter estimation would become noise-sensitive. In Figure 4 we have plotted the theoretical correlation function for RICS in the case of a mixture of two components with diffusion coefficients ($\mu m^2 s^{-1}$) respectively $D_1 = 1$ and $D_2 = 2$ (A), 4 (B), 8 (C), 16 (D), 32 (E), 64 (F) in red and the theoretical correlation function for RICS with only one diffusing component in blue. In an experimental setting, it would be quite hard to distinguish the two models in Figure 4 (A) and (B) even with a good signal to noise ratio, while the cases in (C)-(F) the models would be distinguishable. Thus, on the basis of these observations, we recommend to have at least a factor 8 between the diffusion coefficients of the components in order to use RICS for mixtures. This condition could be relaxed if one uses a global analysis methods as multiple scan speed image correlation spectroscopy (Gröner et al. 2010), where different scan rates are employed to cover a broad range of dynamics. Further investigation of such methods will be the study of future work.
Figure 4: Contour plots of the correlation function of different mixtures of two components with $D_1 = 1 \mu m^2 s^{-1}$ and $D_2 = 2, 4, 8, 16, 32, 64 \mu m^2 s^{-1}$ together with the "closest" correlation function from a model with a single diffusion component. In all cases the proportions $\beta$ in Equation 30 of the two components are equal to 0.5.

We have also tested RICS on 20 simulated datasets with a mixture of particles with the same proportions $\beta$ and diffusion coefficients corresponding to the six cases in Figure 4. The results are reported in Figure 5. In the first two cases, where the factor between the two diffusion coefficients is smaller than 8, the estimates of the parameters are far from the expected values but become increasingly more accurate when the factor increases.
Figure 5: (A) Box plots of the estimated diffusion coefficient (in $\mu m^2 s^{-1}$) $D_1$ divided by the true value of $D_1$ in different simulations with $D_1=1$ and $D_2=2, 4, 8, 16$ and 32. (B) Box plots of the estimated diffusion coefficient $D_2$ divided by the true value of $D_2$ in different simulations with $D_1=1$ and $D_2=2, 4, 8, 16$ and 32. A logarithmic scale has been used on both axes in both plots.

In Table 6 we provide the results with RICS applied to some additional simulated datasets. RICS provides good estimates of $D_1$ and $D_2$ in all cases except in the first one, where the quotient between the two diffusion coefficients is two. The estimated proportion is not always close to the expected one. However, based on these results, one would say that RICS works quite well when the number of components is known and no model selection is needed. To see whether this is the right conclusion, we made some additional simulations with diffusion coefficients $D_1 = 1 \mu m^2 s^{-1}$ and $D_2 = 4 \mu m^2 s^{-1}$ and equal proportions. The first line in Table 7 shows the results for this case when applying RICS without any further constrains. The estimated larger diffusion coefficient is far from the expected value and so is the proportion. In the second line, we observe the results of RICS if we constrain both proportions to be higher than 40%. With the added restriction, the estimated diffusion coefficients are precise, but the estimated proportion lies on the boundary of the region of possible values for the proportions. However, the comparison of the two residual sums of squares $Q_M$ indicates that the best fit is given by the results in the first line.
Table 6: Results with RICS for a mixture of two diffusing species with different diffusion coefficients and equal proportions. $D_1$ and $D_2$ are the two diffusion coefficients used to simulate the mixture of beads, and $\beta_1$ corresponds to the proportion of the first component in the correlation function. $D_{RICS1}$ and $D_{RICS2}$ are the two estimated diffusion coefficients using of RICS, and $\beta_{RICS1}$ is the estimated proportion to be compared with $\beta_1$.

<table>
<thead>
<tr>
<th>$D_1(\mu m^2 s^{-1})$</th>
<th>$D_2(\mu m^2 s^{-1})$</th>
<th>$\beta_1$</th>
<th>$D_{RICS1}(\mu m^2 s^{-1})$</th>
<th>$D_{RICS2}(\mu m^2 s^{-1})$</th>
<th>$\beta_{RICS1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0.50</td>
<td>0.84</td>
<td>3.98</td>
<td>0.62</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.50</td>
<td>0.76</td>
<td>4.74</td>
<td>0.53</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>0.50</td>
<td>0.79</td>
<td>7.01</td>
<td>0.60</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
<td>0.50</td>
<td>0.99</td>
<td>19.15</td>
<td>0.77</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>0.50</td>
<td>0.92</td>
<td>35.28</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 7: Results for a mixture of two diffusing components. The first line corresponds to the application of RICS, while the second corresponds to the application of RICS if we add the constraints $\beta_i \geq 0.4$, $i = 1, 2$. $D_1$ and $D_2$ are the two diffusion coefficients used to simulate the mixture of beads, and $\beta_1$ corresponds to the proportion of the first component in the correlation function. $D_{RICS1}$ and $D_{RICS2}$ are the two estimated diffusion coefficients. Finally, $\beta_{RICS1}$ is the estimated proportion to be compared with $\beta_1$, and $Q_M$ the weighted residual sum of squares as in Equation (31).

<table>
<thead>
<tr>
<th>$D_1(\mu m^2 s^{-1})$</th>
<th>$D_2(\mu m^2 s^{-1})$</th>
<th>$\beta_1$</th>
<th>$D_{RICS1}(\mu m^2 s^{-1})$</th>
<th>$D_{RICS2}(\mu m^2 s^{-1})$</th>
<th>$\beta_{RICS1}$</th>
<th>$Q_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>0.50</td>
<td>1.60</td>
<td>29.95</td>
<td>0.88</td>
<td>1.08-10^5</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.50</td>
<td>1.00</td>
<td>3.98</td>
<td>0.60</td>
<td>1.31-10^5</td>
</tr>
</tbody>
</table>

Finally, in Figure 6 we plot the true correlation function for two different mixture models with two diffusive components. The models differ both in terms of the diffusion coefficients and the proportion of the components but the correlation functions are almost the same. In this case, RICS could be applied if more information about the proportions of the single components was available, and attention was paid to ensure convergence to the global minimum of the residual sum of squares during the fitting of the RICS correlation function.
Figure 6: Contour plots of the correlation function of one mixture of two species with $D_1=1\mu m^2 s^{-1}$ and $D_2=2\mu m^2 s^{-1}$ with the same proportions together with the correlation function of a model with $D_1 = 1.25\mu m^2 s^{-1}$ and $D_2 = 4\mu m^2 s^{-1}$ with proportions $\beta_1 = 0.85$ and $\beta_2 = 0.15$. $\xi, \psi$ are the spatial lags expressed in pixels.

**Conclusion**

The SPRIA method has been extended to analyse systems with mixtures of particles with varying diffusion coefficients. We have also introduced an alternative way of estimating the trajectory of each particle, which compared to the method introduced in (Longhls et al. 2017) works better for particles of varying size and which also allows for faster analysis of large datasets. First, particles are identified and their trajectories estimated via a centroid based technique. Second, the parameters of a model with a mixture of particles are determined with the maximum likelihood method. Third, standard errors of the parameter estimates are obtained by bootstrapping particles. Finally, the number of components that best represents the sample is chosen by using a simple criterion requiring that proportions of all the included components must be above a certain threshold.

The SPRIA mixture model has been applied to simulated and experimental data of fluorescent beads and has been shown to perform satisfactorily in most cases.

We have also studied RICS applied to mixtures. The main result obtained by plotting level curves of mixture correlation functions is that quotients between different diffusion...
coefficients need to be rather large, in the order of 8 or larger, to be identifiable. In such cases we obtain satisfactory parameter estimates. A minor correction is described for the RICS autocorrelation function as given in the literature. It is also shown by plots of level curves that for parameters typically used the effect of the correction is small.

Acknowledgements

Financial support from the Swedish Foundation for Strategic Research, SSF, and Knut and Alice Wallenberg Foundation, KAW, is highly appreciated. We are also indebted to Chris Glasbey for pointing out that the usual formula for the RICS autocorrelation function in the literature needs to be corrected.

References


**Supplementary material**

**Old and new trajectory estimation methods**

We give in Table 8 a comparison on experimental data of the old and the new trajectory estimation methods, compare Figure 1.

Table 8: \( N \) is the observed number of particles, \( D \) is the expected diffusion coefficient, \( D_{\text{old}} \) is the estimated diffusion coefficient as presented in [Longfils et al., 2017], and \( D_{\text{new}} \) is the estimated diffusion coefficient with the method suggested herein. The values after the ± signs are standard errors computed by bootstrapping.

<table>
<thead>
<tr>
<th>( N )</th>
<th>( D (\mu m^2 s^{-1}) )</th>
<th>( D_{\text{old}} (\mu m^2 s^{-1}) )</th>
<th>( D_{\text{new}} (\mu m^2 s^{-1}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>491</td>
<td>2.5</td>
<td>2.0 ± 0.20</td>
<td>1.87 ± 0.09</td>
</tr>
<tr>
<td>160</td>
<td>2.5</td>
<td>2.20 ± 0.32</td>
<td>2.16 ± 0.04</td>
</tr>
<tr>
<td>139</td>
<td>2.5</td>
<td>2.4 ± 0.33</td>
<td>2.25 ± 0.10</td>
</tr>
<tr>
<td>140</td>
<td>2.5</td>
<td>2.75 ± 0.44</td>
<td>2.75 ± 0.13</td>
</tr>
<tr>
<td>134</td>
<td>0.48</td>
<td>0.62 ± 0.07</td>
<td>0.64 ± 0.02</td>
</tr>
<tr>
<td>118</td>
<td>0.48</td>
<td>0.63 ± 0.09</td>
<td>0.63 ± 0.02</td>
</tr>
<tr>
<td>101</td>
<td>0.48</td>
<td>0.58 ± 0.07</td>
<td>0.66 ± 0.03</td>
</tr>
</tbody>
</table>

**Correlations between SPRIA parameter estimates**

We have computed estimates of the correlations between the parameter estimates in Tables 1–4 in the paper. We find that, although some correlation estimates are rather
close to +1 or -1, most of them are of moderate size, and typically of the same sign column-wise. For example, we find that the (column-wise) average correlation between the estimates $D_{SPRIA1}$ and $D_{SPRIA2}$ in Tables 1 and 3 are 0.52 and 0.51, respectively, the average correlation between the estimates $D_{SPRIA1}$ and $C_{SPRIA1}$ in Tables 1 and 3 are 0.52 and 0.52, respectively, and the average correlation between the estimates $D_{SPRIA2}$ and $C_{SPRIA2}$ in Tables 1 and 3 are -0.45 and -0.58, respectively.

**RICS on experimental data**

In Table 9 we report results for RICS on the experimental data previously used to study SPRIA. Generally, the RICS results here are less useful. For instance, in the case of the mixture of 175 and 1000nm beads, RICS consistently gives one estimate where one of the two species is almost non-existent. One can note that the quotient between different diffusion coefficients in each mixture varies from about 2 to 5, i.e. considerably less than our recommendation based on Figure 4 of at least a factor 8.

Table 9: Results for a mixtures of 175nm, 500nm, and 1000nm beads with different proportions. Sizes indicates which beads have been mixed in each row and in parentheses we report the expected diffusion coefficients according to Stokes-Einstein relation. $C_{low}$ is the expected proportion of the larger bead species in the mixture ($=C_1/(C_1+C_2)$). $D_{RICS1}$ and $D_{RICS2}$ are the two estimated diffusion coefficients for the mixture model, and $D_{RICS}$ is the diffusion estimated for the single diffusion coefficient model. $\beta_{RICS1}$ is the estimated proportion $\beta_1$.

<table>
<thead>
<tr>
<th>Sizes (D)</th>
<th>$C_{low}$</th>
<th>$D_{RICS1}(\mu m^2 s^{-1})$</th>
<th>$D_{RICS2}(\mu m^2 s^{-1})$</th>
<th>$\beta_{RICS1}$</th>
<th>$D_{RICS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000 / 175nm</td>
<td>0.23</td>
<td>0.58</td>
<td>2.29</td>
<td>0.04</td>
<td>2.29</td>
</tr>
<tr>
<td>(0.48 / 2.5 $\mu m^2 s^{-1}$)</td>
<td></td>
<td></td>
<td></td>
<td>0.48</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td>0.47</td>
<td>3.01</td>
<td>1.00</td>
<td>1.93</td>
</tr>
<tr>
<td>500 / 175nm</td>
<td>0.33</td>
<td>0.39</td>
<td>1.83</td>
<td>0.80</td>
<td>1.47</td>
</tr>
<tr>
<td>(0.98 / 2.5 $\mu m^2 s^{-1}$)</td>
<td></td>
<td></td>
<td></td>
<td>0.60</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.19</td>
<td>2.18</td>
<td>0.12</td>
<td>0.90</td>
</tr>
<tr>
<td>1000 / 500nm</td>
<td>0.38</td>
<td>0.00</td>
<td>1.56</td>
<td>0.31</td>
<td>1.97</td>
</tr>
<tr>
<td>(0.48 / 0.98 $\mu m^2 s^{-1}$)</td>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>0.22</td>
<td>1.22</td>
<td>0.88</td>
<td>2.13</td>
</tr>
</tbody>
</table>
Non-mixed particles experimental data

Here, we analyse separately the same three sizes of beads as above. In Table 10, we report the results for the 175nm, 500nm, and 1000nm beads. For the 175nm beads, the estimated diffusion coefficient in a model with a single component ($D_{SPRIA}$) in all three cases is underestimating the true one. Moreover in two cases out of three we overfit the data by selecting a model with two components. On the other hand, RICS provides good point estimates. Regarding the results for 500nm particles, both SPRIA and RICS provide estimates that are lower than the expected diffusion coefficient. Again, SPRIA selects the incorrect number of components. Finally, in the case of 1000nm beads both RICS and SPRIA provide good point estimates and SPRIA selects the correct number of components in all cases.

Table 10: Results for single populations of beads. Sizes indicates which beads have been considered in each row and in parentheses we report the expected diffusion coefficients according to Stokes-Einstein. $D_{SPRIA}$ is the estimated diffusion coefficient by SPRIA with one component, while $D_{SPRIA1}$ and $D_{SPRIA2}$ are the two estimates for a mixture with two diffusing species. $C_{SPRIA1}$ is the estimated proportion of particles with diffusion coefficient $D_{SPRIA1}$. Comp15 is the selected number of components with a 15% threshold on the minimum proportion and finally, $D_{RICS}$ is the estimated diffusion coefficient by RICS. The values after the ± signs are standard errors computed by bootstrap.

<table>
<thead>
<tr>
<th>Sizes (D)</th>
<th>$D_{SPRIA}$($\mu$m$^2$s$^{-1}$)</th>
<th>$D_{SPRIA1}$($\mu$m$^2$s$^{-1}$)</th>
<th>$D_{SPRIA2}$($\mu$m$^2$s$^{-1}$)</th>
<th>$C_{SPRIA1}$</th>
<th>Comp15</th>
<th>$D_{RICS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>175nm (2.5µm$^2$s$^{-1}$)</td>
<td>2.03 ± 0.03</td>
<td>1.09 ± 0.45</td>
<td>2.53 ± 5.78</td>
<td>0.44 ± 0.29</td>
<td>2</td>
<td>2.60 ± 0.03</td>
</tr>
<tr>
<td></td>
<td>2.04 ± 0.02</td>
<td>0.82 ± 0.29</td>
<td>2.31 ± 1.66</td>
<td>0.27 ± 0.18</td>
<td>2</td>
<td>2.41 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>2.07 ± 0.03</td>
<td>1.92 ± 0.31</td>
<td>7.63 ± 2.46</td>
<td>0.99 ± 0.21</td>
<td>1</td>
<td>2.41 ± 0.03</td>
</tr>
<tr>
<td>500nm (0.98µm$^2$s$^{-1}$)</td>
<td>0.79 ± 0.01</td>
<td>0.87 ± 0.03</td>
<td>0.32 ± 0.01</td>
<td>0.80 ± 0.03</td>
<td>2</td>
<td>0.66 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>0.79 ± 0.01</td>
<td>0.87 ± 0.03</td>
<td>0.30 ± 0.01</td>
<td>0.81 ± 0.03</td>
<td>2</td>
<td>0.63 ± 0.02</td>
</tr>
<tr>
<td>1000nm (0.48µm$^2$s$^{-1}$)</td>
<td>0.45 ± 0.01</td>
<td>0.43 ± 0.06</td>
<td>1.51 ± 0.72</td>
<td>0.97 ± 0.15</td>
<td>1</td>
<td>0.46 ± 0.04</td>
</tr>
<tr>
<td></td>
<td>0.44 ± 0.01</td>
<td>0.20 ± 0.12</td>
<td>0.46 ± 0.49</td>
<td>0.10 ± 0.36</td>
<td>1</td>
<td>0.50 ± 0.05</td>
</tr>
</tbody>
</table>

Simulation of non-mixed particles

To check the results shown in Table 10, we also performed simulations for each of the three sizes of the beads separately. The results are given in Table 11.
Table 11: Results for simulated non-mixed beads. $D_{\text{True}}$ is the true diffusion coefficient used for the simulation, $D_{\text{SPRIA}}$ is the estimated diffusion coefficient by SPRIA with one component, while $D_{\text{SPRIA1}}$ and $D_{\text{SPRIA2}}$ are the two estimates for a mixture with two diffusing species, and $C_{\text{SPRIA1}}$ is the estimated proportion of particles with diffusion coefficient $D_{\text{SPRIA1}}$. Comp15 is the selected number of components with a 15% threshold on the minimum proportion and finally $D_{\text{RICS}}$ is the estimated diffusion coefficient by RICS. The values after the ± signs are standard errors computed by bootstrapping.

<table>
<thead>
<tr>
<th>$D_{\text{True}}$ ($\mu m^2 s^{-1}$)</th>
<th>$D_{\text{SPRIA}}$ ($\mu m^2 s^{-1}$)</th>
<th>$D_{\text{SPRIA1}}$ ($\mu m^2 s^{-1}$)</th>
<th>$D_{\text{SPRIA2}}$ ($\mu m^2 s^{-1}$)</th>
<th>$C_{\text{SPRIA1}}$</th>
<th>Comp15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>2.18 ± 0.03</td>
<td>2.30 ± 2.10</td>
<td>0.54 ± 0.70</td>
<td>0.91 ± 0.40</td>
<td>1</td>
</tr>
<tr>
<td>2.5</td>
<td>2.24 ± 0.03</td>
<td>2.35 ± 0.16</td>
<td>0.54 ± 0.23</td>
<td>0.92 ± 0.10</td>
<td>1</td>
</tr>
<tr>
<td>0.98</td>
<td>1.05 ± 0.01</td>
<td>1.12 ± 0.20</td>
<td>0.35 ± 0.11</td>
<td>0.90 ± 0.13</td>
<td>1</td>
</tr>
<tr>
<td>0.98</td>
<td>1.03 ± 0.01</td>
<td>1.96 ± 1.05</td>
<td>0.92 ± 0.24</td>
<td>0.10 ± 0.33</td>
<td>1</td>
</tr>
<tr>
<td>0.48</td>
<td>0.66 ± 0.01</td>
<td>1.30 ± 0.15</td>
<td>0.57 ± 0.02</td>
<td>0.12 ± 0.05</td>
<td>1</td>
</tr>
<tr>
<td>0.48</td>
<td>0.67 ± 0.01</td>
<td>1.34 ± 0.12</td>
<td>0.60 ± 0.01</td>
<td>0.08 ± 0.03</td>
<td>1</td>
</tr>
</tbody>
</table>