

Approximate Bayesian computation for estimating number concentrations of monodisperse nanoparticles in suspension by optical microscopy

Magnus Röding*

*SP Food and Bioscience, Soft Materials Science, Göteborg, Sweden and
School of Energy and Resources, UCL Australia, University College London, Adelaide, Australia*

Elisa Zagato, Katrien Remaut, and Kevin Braeckmans

*Laboratory of General Biochemistry and Physical Pharmacy, Ghent University, Ghent, Belgium and
Center for Nano- and Biophotonics, Ghent University, Ghent, Belgium*

(Dated: June 3, 2016)

We present an approximate Bayesian computation (ABC) scheme for estimating number concentrations of monodisperse diffusing nanoparticles in suspension by optical particle tracking microscopy. The method is based on the probability distribution of the time spent by a particle inside a detection region. We validate the method on suspensions of well-controlled reference particles. We illustrate its usefulness with an application in gene therapy, applying the method to estimate number concentrations of plasmid DNA molecules and the average number of DNA molecules complexed with liposomal drug delivery particles.

PACS numbers: 05.40.Jc, 02.50.Tt, 02.70.Uu, 83.10.Rs

I. INTRODUCTION

Understanding interactions of functional nanomaterials and nanoparticles with biological matter are at the heart of interdisciplinary research in biological physics, medicine, and nanotechnology [1, 2]. In particular, the design of man-made functional nanoparticles has received more and more attention in recent years. Applications are vast, including targeted delivery of therapeutic agents [3] and the use of tailor-made tracer particles for biomedical imaging [4]. As a consequence, there is an interest in the detection and characterization of particles of sub-micron size, both artificial and natural, in a range of situations including particles suspended in complex fluids like undiluted blood [5, 6]. Examples of applications are the detection of blood-borne specific cell-derived extracellular vesicles that serve as biomarkers for metabolic and systemic diseases [7, 8], of amyloid β aggregates in cerebrospinal fluid or plasma for diagnosis of e.g. Alzheimer's [9], and of subvisible protein aggregates that are important when developing protein-based drug formulations [10, 11].

With several substantial scientific advances, e.g. highly sensitive camera sensors with high spatial and temporal resolution, high-quality optics and lasers, and increased computer power for image analysis, optical microscopy and particle tracking techniques have emerged as cornerstone technology in this growing interdisciplinary field. While nanoparticles are often too small to be characterized by optical microscopy, still they can be visualized as diffraction limited spots of light in dark field or fluorescence microscopy. By imaging their Brownian motion in suspension, particle parameters including size and concentration can be derived. Indeed, a number of analysis

methods have emerged in recent years for characterization of such suspended nanoparticles [6, 12–15].

In previous work of Röding *et al.* [13], we demonstrated that using fluorescence microscopy and tracking freely diffusing particle species, an absolute concentration can be estimated with good accuracy and precision. While counting of the particles is relatively straightforward, a particular challenge in order to arrive at a concentration estimate is to determine precisely the size of the detection region in which the particles are detected. One approach is to indirectly estimate this size by a prior calibration using reference particles of known concentration [16]. However, the size of the detection region is affected by several factors, including the microscope's field of view and depth of field, sample illumination, particle fluorescence intensity, and particle detection and tracking algorithms. Consequently, a prior calibration using reference model particles (e.g. polymer nanospheres) is inherently flawed, because all these factors cannot be held constant between experiments on different particles. Therefore, we developed an approximate statistical model relating the probability distribution of the time spent by a particle inside the detection region, the size of the detection region, the diffusion coefficient estimated from particle tracking videos, and the number concentration.

While this previous method has proven quite useful in a number of applications [17–20], several approximations had to be introduced. No tractable, exact likelihood function can be formulated for the problem of concentration estimation in this setting: As is discussed in the Appendix of Röding *et al.* [13], formulating a likelihood function involves alternating Gaussian convolutions and multiplication with interval indicator functions, a procedure which becomes analytically intractable rather quickly. Moreover, for computational reasons, we had to approximate the three-dimensional geometrical model with a one-dimensional one. Nevertheless, the stochastic

* magnus.rodin@sp.se

data generating process (DGP) is simple and thus the problem naturally lends itself to simulation-based inference. In this work, we present an approximate Bayesian computation (ABC) scheme for estimating number concentrations of nanoparticles in suspension by fluorescence microscopy and particle tracking. First, we discuss aspects of the general concentration estimation problem and the tailoring of the ABC framework to it. Then, we validate the method experimentally on reference particles with well-controlled concentration. Finally, we illustrate its usefulness with an application in gene therapy, applying the method to estimating concentrations of plasmid DNA molecules and the average number of DNA molecules with liposomal drug delivery particles.

II. APPROXIMATE BAYESIAN COMPUTATION

Approximate Bayesian computation (ABC) is a widely used Bayesian inference framework for complex statistical models where the likelihood is intractable or computationally expensive. Although some underlying ideas for simulation-based inference date back to Rubin [21] and Diggle and Gratton [22], the original paper on the modern ABC framework is arguably Tavaré *et al.* [23], with other notable early contributions being Pritchard *et al.* [24] and Beaumont *et al.* [25], the latter in which the term 'ABC' was coined. Consider a data set $\mathcal{D} \in \mathcal{S}$ and a model class indexed by a general parameter vector θ . In Bayesian statistics, inference is drawn from the full posterior distribution

$$f(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)\pi(\theta)}{P(\mathcal{D})}, \quad (1)$$

where $P(\mathcal{D}|\theta)$ is the likelihood, $\pi(\theta)$ is the prior distribution, and $P(\mathcal{D})$ is the evidence. Conventionally, posterior inference relies on some method of generating random observations from $f(\theta|\mathcal{D})$ by means of actual computation of the likelihood, such as a simple rejection method. The rejection method involves generating a random θ from $\pi(\theta)$, and accepting θ as an observation from $f(\theta|\mathcal{D})$ with probability $h = P(\mathcal{D}|\theta)$. Now, if the likelihood cannot be explicitly computed but the stochastic data generating process (DGP) is known and it is simple to simulate realizations, one can

1. Generate a random θ from $\pi(\theta)$.
2. Simulate a realization \mathcal{D}' from the DGP for parameter θ .
3. Accept θ if and only if $\mathcal{D}' = \mathcal{D}$.

The obtained result after iterating this procedure many times is still a sample from the exact posterior. However, since $P(\mathcal{D}' = \mathcal{D}|\theta)$ is likely to be very small in most realistic settings this is an impractical approach. A way forward is to introduce a 'distance metric' $\rho : \mathcal{S} \times \mathcal{S} \rightarrow$

\mathbb{R}_+ , such that $\rho = 0$ if and only if $\mathcal{D}' = \mathcal{D}$, and a tolerance level $\epsilon \in \mathbb{R}_+$. Then, one can

1. Generate a random θ from $\pi(\theta)$.
2. Simulate a realization \mathcal{D}' from the DGP for parameter θ .
3. Compute the 'distance' $\rho(\mathcal{D}, \mathcal{D}')$ between the real data and the simulated realization.
4. Accept θ if and only if $\rho(\mathcal{D}, \mathcal{D}') \leq \epsilon$.

Whereas $f(\theta|\rho(\mathcal{D}, \mathcal{D}') = 0)$ is the exact posterior, $f(\theta|\rho(\mathcal{D}, \mathcal{D}') \leq \epsilon)$ for a non-zero ϵ constitutes a relaxation, the nature of which is determined by ρ and ϵ . The acceptance probability now depends on ϵ that indexes a continuous set of distributions from the posterior ($\epsilon = 0$) to the prior ($\epsilon \rightarrow \infty$). There are recent developments in the selection of ϵ such as cross-validation techniques [26], but we choose a more simple and pragmatic approach of carefully choosing ϵ manually such that a good trade-off between accuracy and computability can be found.

Since its introduction, different flavours of the ABC framework have emerged such as MCMC ABC [27] and sequential Monte Carlo ABC [28] (with correction, [29]). We shall focus on the rejection case herein while proceeding to the field of application.

III. ESTIMATING PARTICLE CONCENTRATIONS USING ABC

Consider a particle tracking experiment with freely diffusing, Brownian particles in a liquid suspension $\Omega = [-A/2, A/2]^3 \subset \mathbb{R}^3$. Assuming diffusion equilibrium and a concentration (density) c , a (fixed) number of particles with distribution $\text{Poi}(c|\Omega)$ are distributed (marginally) uniformly in Ω at any time. A fluorescence microscope setup is used to image and track particles in two dimensions inside a cuboid mid-section of the liquid suspension which we denote the detection region, $\omega = [-a_x/2, a_x/2] \times [-a_y/2, a_y/2] \times [-a_z/2, a_z/2] \subset \Omega$. Figure 1 shows the experimental setup. The lateral sizes (perpendicular to the optical axis z) a_x and a_y of the detection region are determined by the field of view of the microscope. Hence, they are known prior to the experiment with near-perfect accuracy. The axial size (parallel to the optical axis z) a_z , on the other hand, depends on experimental factors such as the depth of field of the microscope optics, sample illumination, particle fluorescence intensity, and particle detection and tracking algorithms. Hence, it is not known prior to the experiment because it may vary depending not only on the experimental setup but also on the image analysis settings. It is perhaps worth noting at this point that we do not know the particular shape of the liquid suspension, only do we know that $|\Omega| \gg |\omega|$ and that ω is far from the boundary of Ω so that any edge effects are negligible; the cuboid

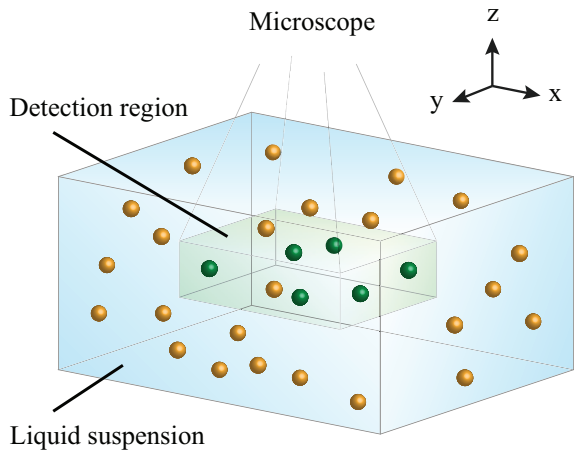


FIG. 1. (Color online) The experimental setup. Particles inside the cuboid-shaped detection region (green) are detected and tracked in two dimensions (x and y) using a fluorescence microscope. The x and y dimensions of the detection region, a_x and a_y , are known. The z dimension a_z needs to be estimated from the particle tracking data.

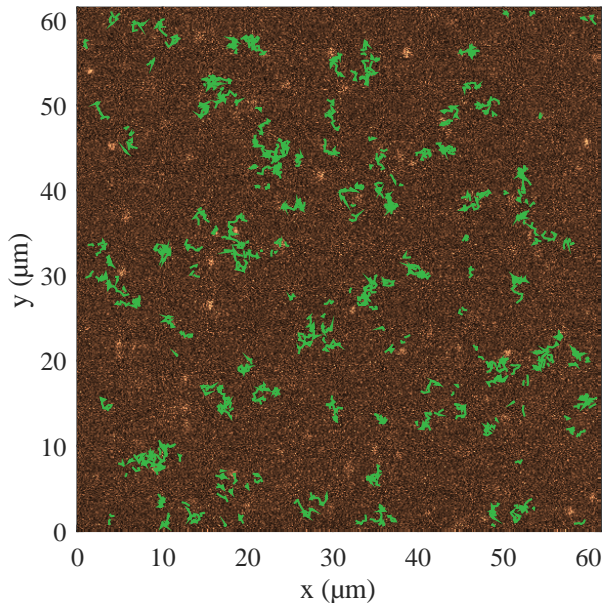


FIG. 2. (Color online) Example of a set of experimentally acquired trajectories superimposed on a still video frame.

shape is only a convenient assumption and we choose arbitrarily to let $A = 100 \mu\text{m}$.

Recording a number of time-lapse video sequences with sampling interval Δt and performing subsequent particle detection and tracking (in two dimensions) yields a set of particle trajectories. Figure 2 shows an example of particle trajectories taken from the data sets analyzed herein.

Because the particles move randomly in and out of the detection region, the trajectory durations are random,

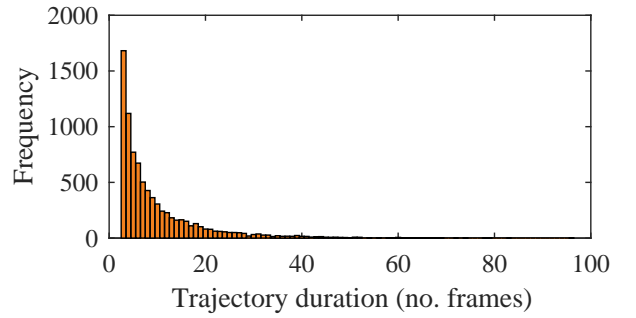


FIG. 3. (Color online) Example of an experimentally acquired trajectory duration distribution.

the number of positions being between 1 and k_{\max} (the number of frames in the longest video recorded). The real length of a trajectory can be longer than the duration of a video in principle; however, the experiments and the simulations share the property that trajectories longer than the duration of a video is recorded as being equally long as the video. Hence, the ABC framework elegantly takes care of this truncation. Figure 3 shows an example of a trajectory duration distribution taken from the data sets analyzed herein.

In practice, because very short trajectories often in fact are false positives due to noise in the images [30], a lower threshold k_{\min} (equal to, say, 2 or 3) is also introduced and shorter trajectories with $k < k_{\min}$ are discarded (this step reduces ambiguity in particle tracking). By definition of free diffusion and the mean squared displacement relation from statistical physics, between any two consecutive frames, each observed particle performs a Gaussian random walk with incremental (and mutually independent) displacements $\Delta x, \Delta y \sim \mathcal{N}(\mu = 0, \sigma^2 = 2D\Delta t)$ (and a third, unobserved displacement Δz with the same distribution) [31]. The data thus constitutes (i) a list of trajectory durations k_1, k_2, \dots , represented by a vector of non-negative integers (a histogram) $\mathbf{k} \in \mathbb{Z}_+^{k_{\max}}$, where \mathbf{k}_i is the number of trajectories consisting of i positions, and (ii) a list of two-dimensional particle displacement vectors $(\Delta x_1, \Delta y_1), (\Delta x_2, \Delta y_2), \dots$ of (random) length $m = (k_1 - 1) + (k_2 - 1) + \dots$ (since the number of displacements in each trajectory is the number of positions minus 1) summarized by their mean squared displacement

$$r^2 = \frac{1}{m} \sum_{i=1}^m (\Delta x_i^2 + \Delta y_i^2). \quad (2)$$

We proceed to the setting of the ABC rejection framework. The ultimate goal is to estimate the concentration (density) of particles c , but that also requires knowledge of their (common) diffusion coefficient D and the axial size of the detection region a_z . Hence, we wish to sample from the posterior for the entire parameter set (D, a_z, c) ,

$$f(D, a_z, c | \mathbf{k}, r^2) = \frac{P(\mathbf{k}, r^2 | D, a_z, c) \pi(D, a_z, c)}{P(\mathbf{k}, r^2)}. \quad (3)$$

We note that inasmuch as a_z and c influence the number of observed incremental displacements of particles considerably, their impact on the value of r^2 as such is very minor. This is related to the fact that the lateral dimensions of the detection region a_x and a_y are much larger than a_z (typically, 20-100 times larger). Therefore, particles mainly enter and exit the detection region along the axial dimension (in parallel to the optical axis), as a result of the unobserved displacements in the axial dimension rather than as a result of the observed displacements in the lateral dimensions. As a consequence, \mathbf{k} and r^2 are approximately conditionally independent (the exception being negligible 'edge effects' close to the lateral boundaries of the detection region, causing a very slight cut-off effect in the normal distribution of displacements). We utilize this by factoring the likelihood,

$$P(\mathbf{k}, r^2 | D, a_z, c) \approx P(\mathbf{k} | D, a_z, c) P(r^2 | D, a_z, c) \approx \quad (4) \\ P(\mathbf{k} | D, a_z, c) P(r^2 | D).$$

Now, $P(r^2 | D)$ is tractable and based on the normal distribution likelihood for the two-dimensional particle displacement vectors, for which r^2 is a sufficient statistic (the right hand side is a product of m normal densities),

$$P(r^2 | D, m) = (4\pi D \Delta t)^{-m} \exp\left(-\frac{mr^2}{4D\Delta t}\right). \quad (5)$$

Note that this part of the likelihood is also conditional on a fixed number of trajectories. Normalizing so that this forms a pdf for D , we obtain that $P(D | r^2, m)$ is an inverse gamma distribution,

$$D | r^2, m \sim \Gamma^{-1}\left(\alpha = m - 1, \beta = \frac{mr^2}{4\Delta t}\right), \quad (6)$$

and in effect, sampling values of D (marginally) is easy.

The other factor in the likelihood, $P(\mathbf{k} | D, a_z, c)$, is handled by straightforward simulation, and we can therefore easily cope with the fact that the number of trajectories is random. For a given candidate parameter vector (D, a_z, c) , $\text{Poi}(c|\Omega)$ noninteracting point particles are placed uniformly in the virtual liquid suspension Ω . For each time step Δt , particle positions are displaced in normally distributed increments with $\mu = 0$ and $\sigma^2 = 2D\Delta t$ in each direction. Periodic boundary conditions are assumed for Ω . When particles enter the virtual detection region ω , they are detected and tracked. The lateral sizes of the detection region is known prior to the experiment. The displacements from the simulation are not used, only the trajectory durations represented by \mathbf{k}' . As the 'distance metric' ρ , we choose simply

$$\rho(\mathbf{k}, \mathbf{k}') = \sum_{i=k_{\min}}^{k_{\max}} (\mathbf{k}_i - \mathbf{k}'_i)^2, \quad (7)$$

i.e. the Euclidean norm (it was found that other 'distance metrics' considered, also based on norms of 'histogram' differences, such as the total variation or the

Euclidean norm of the 'cumulative histograms' produced near-identical results). We mention in this context that there are recent investigations on optimal ways of summarizing the differences between data and simulations, including dimensionality reductions [32, 33]. However, our 'distance metric' does not constitute a dimensionality reduction in this sense, because the data is not reduced to e.g. moments. We choose a flat (improper) prior over $(D, a_z, c) \in (0, \infty) \times [a_{z,\min}, a_{z,\max}] \times [c_{\min}, c_{\max}]$. Restricting the choice of candidate values in this fashion is necessary because computational workload for each simulation is approximately linear in c , so extremely large candidate values of c must be avoided. It is also sensible from a physical point of view: the assumption of independence between particles requires c to be upper bounded, and to make meaningful inference possible by observing a sufficient number of trajectories, c must be lower bounded. Also, a_z is upper bounded by optical and other limitations, in practice this is in the order of a few micrometers. This prior is essentially non-informative from the point of inference, but still avoids the computation of extremely heavy cases (high c). The algorithm now becomes

1. Generate a random $D \sim \Gamma^{-1}(\alpha = m - 1, \beta = mr^2/(4\Delta t))$, using the estimated mean squared displacement r^2 and the number of observed increments m .
2. Generate random $a_z \sim U[a_{z,\min}, a_{z,\max}]$ and $c \sim U[c_{\min}, c_{\max}]$.
3. Simulate a realization of a tracking experiment using the parameter vector (D, a_z, c) , yielding simulated trajectory durations represented by \mathbf{k}' .
4. Compute the 'distance' $\rho(\mathbf{k}, \mathbf{k}')$ between the trajectory duration distributions of the real data and of the simulated realization.
5. Accept (D, a_z, c) as a sample from the (approximate) posterior if and only if $\rho(\mathbf{k}, \mathbf{k}') \leq \epsilon$.

In a sense, from the point of view of the ABC framework, $P(r^2 | D, m)\pi(D, a_z, c)$ now constitutes the prior (one could introduce the notation $\pi(D, a_z, c | r^2, m)$), and $P(\mathbf{k} | D, a_z, c)$ takes on the role of the full likelihood. There are several benefits to this factorization of the likelihood. First, the prior (incorporating the 'partial' posterior) comes 'closer' to the full posterior, so that less rejections are necessary to retain a certain accuracy of approximation of the posterior. Second, the choice of ρ is easier when r^2 is accounted for separately, and ρ does not incorporate a 'trade-off' between r^2 and \mathbf{k} .

It is hard to provide precise guidelines or rules-of-thumb for the number of draws in the sample that are necessary in order to obtain a certain accuracy, because this depends on the problem, the prior knowledge about the problem, and the parameter bounds set in the prior. In the studies performed herein, we use approximately 2×10^7 draws for each estimation.

The ABC algorithm is implemented in the high-performance scientific computing language Julia [34–36]. Matlab (Mathworks, Natick, MA, US) is used for graphics.

IV. EXPERIMENTS

A. Validation on polymer nanospheres

We perform a series of validation experiments on water dispersions of fluorescent polymer nanospheres with very narrow distribution in diameter ($d \approx 497 \pm 9$ nm). From a batch of known concentration, five different dilutions are made. Based on knowledge of the diameter, mass density, and weight percentage of particles in the water solution as well as the dilution factor, theoretical concentrations can be obtained including 95 % confidence intervals (see Rödning *et al.* [13]). For more details on the experiment, refer to Section IV C.

The proposed ABC framework is then used to generate a sample from the approximate posterior for (D, a_z, c) . After carefully assessing the impact of the choice of the tolerance level ϵ by manual inspection, it is chosen so that $P(\rho \leq \epsilon) = 10^{-3}$, keeping approximately 2×10^4 draws from each approximate posterior (from a total of approximately 2×10^7). Further decreasing the value of ϵ provides no substantial change in mean and variance of the approximate posterior (and some increase provides very little change, so the choice is not absolutely crucial). Figure 4 shows estimated concentrations (posterior mean) with credible intervals (2.5 % and 97.5 % percentiles of the marginal posterior) compared to the expected (theoretical) concentrations.

As is clear, good agreement is found between the estimated concentrations and the expected reference concentration (and with the ones estimated before using another method, as can be seen when comparing with Figure 4 in Rödning *et al.* [13]), confirming the validity of the approach in a highly controlled setting.

B. An application in gene therapy

We illustrate the application of the method with an example related to nanomedicine mediated gene therapy. The genes come in the form of plasmid DNA, small ring-shaped DNA molecules that can be manipulated for therapeutical purposes. These therapeutic macromolecules are typically incorporated into nanocarriers such as liposomes to enable their delivery into target cells. Complexes consisting of liposomes and DNA are termed lipoplexes (LPX). To allow for accurate dosing, a crucial question is to determine how many DNA particles are on average encapsulated in a single lipoplex. It is useful to design particles with a high number of DNA per LPX, because this means that less material is needed to bring the DNA to the cell nucleus. Hence, this reduces

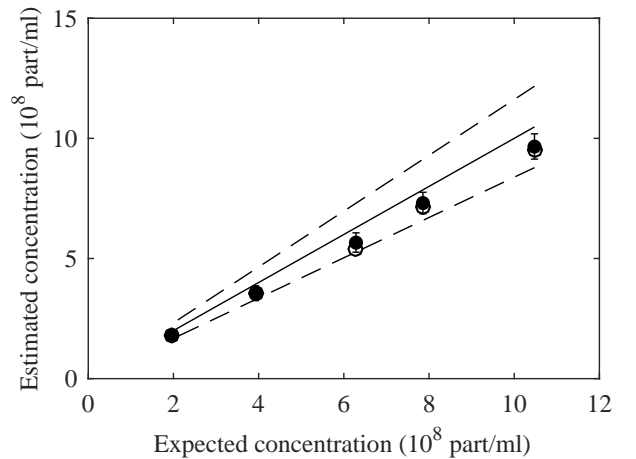


FIG. 4. (Color online) Estimated vs expected concentrations of a dilution series of water dispersions of fluorescent polymer nanospheres. Concentrations estimated with the proposed ABC framework (posterior mean) with credible intervals (2.5 % and 97.5 % percentiles) are compared to the expected (theoretical) concentrations (solid line) with error bounds (dashed lines). As is clear, good agreement is found between the estimated concentrations and the expected reference concentration. Furthermore, the results are slightly improved compared to the concentrations estimated before using the method of Rödning *et al.* [13] (circles, partly occluded).

the toxicity of the treatment (even biological materials are toxic in high concentration) and the risk of an immunoresponse from the organism.

We demonstrate how this number can be estimated using particle tracking in fluorescence microscopy and the proposed ABC framework for estimating number concentrations for both DNA and LPX. To the best of our knowledge, currently there is no alternative method for accurately estimating the concentration of LPX other than that proposed herein and in Rödning *et al.* [13]. Hence, this small example is an effective demonstration of the usefulness of the method in pharmaceutical applications.

First, the number concentration c_{DNA} of a suspension of only plasmid DNA molecules is estimated. Second, the plasmid DNA suspension is mixed in equal volumes (so that c_{DNA} is in effect halved) with a liposome suspension. LPX are formed and their number concentration c_{LPX} is estimated as well. The average number of DNA particles encapsulated in a single LPX can thus be estimated by the ratio

$$Q = \frac{1}{2} \frac{c_{\text{DNA}}}{c_{\text{LPX}}}. \quad (8)$$

The whole experiment is performed in triplicate to account for variations in e.g. dilution factors. The ABC framework is used to generate samples from the approximate posterior for (D, a_z, c) for the whole set of DNA and LPX suspensions. Once again, after carefully assessing the impact of the choice of the tolerance level ϵ by manual

inspection, it is chosen so that $P(\rho \leq \epsilon) = 10^{-3}$, keeping approximately 2×10^4 draws from each approximate posterior (from a total of approximately 2×10^7). Further decreasing the value of ϵ provided no substantial change in mean and variance of the approximate posterior.

We proceed with reporting posterior mean and 2.5% and 97.5% percentiles of the parameter values for diffusion coefficient, axial detection region size and particle number concentration for DNA and LPX as well as the ratio Q , see Table I. Observations from the posterior for the ratio Q in Eq. (8) can be formed simply by random pairing of observations from the posteriors for c_{DNA} and c_{LPX} .

We also illustrate the entire posterior distribution over the parameter space (D, a_z, c) for the second LPX replicate, see Figure 5. That a_z and c are clearly negatively correlated in the posterior does not come as a surprise because for a fixed number of trajectories, a larger detection region suggests a smaller concentration, and vice versa.

The average number of DNA per LPX is 1.59 ± 0.06 . It would be expected that the numbers are between 1 and 2, because the diameters of both DNA and LPX are roughly 100-150 nm. Although plasmid DNA can be compressed during formation of LPX, it would be unlikely that 3 or more would fit into one liposomal carrier.

C. Materials and experimental methods

Acquisition of video data and particle tracking is described elsewhere in detail [6]. The experimental methods and the nanoparticles involved are briefly described below.

Validation experiments are conducted using a custom-built widefield epi-fluorescence microscope setup with laser illumination and an EMCCD camera, based on a Nikon TE2000E microscope. Image acquisition is performed using the Nikon Elements R imaging software. For each experiment, 40 videos with 15 second duration of individual particles performing free diffusion in the liquid suspension are acquired. The frame rate is approximately 17 frames per second and the resolution is 512 by 512 pixels. The physical pixel size is 0.1329 μm ; hence, the lateral size of the detection region is $a_x = a_y \approx 68 \mu\text{m}$. Particle detection and tracking is performed using in-house software developed in Matlab (Mathworks, Natick, MA, US). Fluorescent polymer nanospheres (Bangs Laboratories, Fishers, US) with very narrow distribution in diameter ($d \approx 497 \pm 9 \text{ nm}$) and of well-controlled concentration are used for validation of the method. Trajectories shorter than $k_{\text{min}} = 3$ are excluded from further analysis. The tracking was performed at least 20 μm away from the cover glass to avoid deviations from free diffusion due to wall interactions.

For the gene therapy experiments, suspensions of DNA

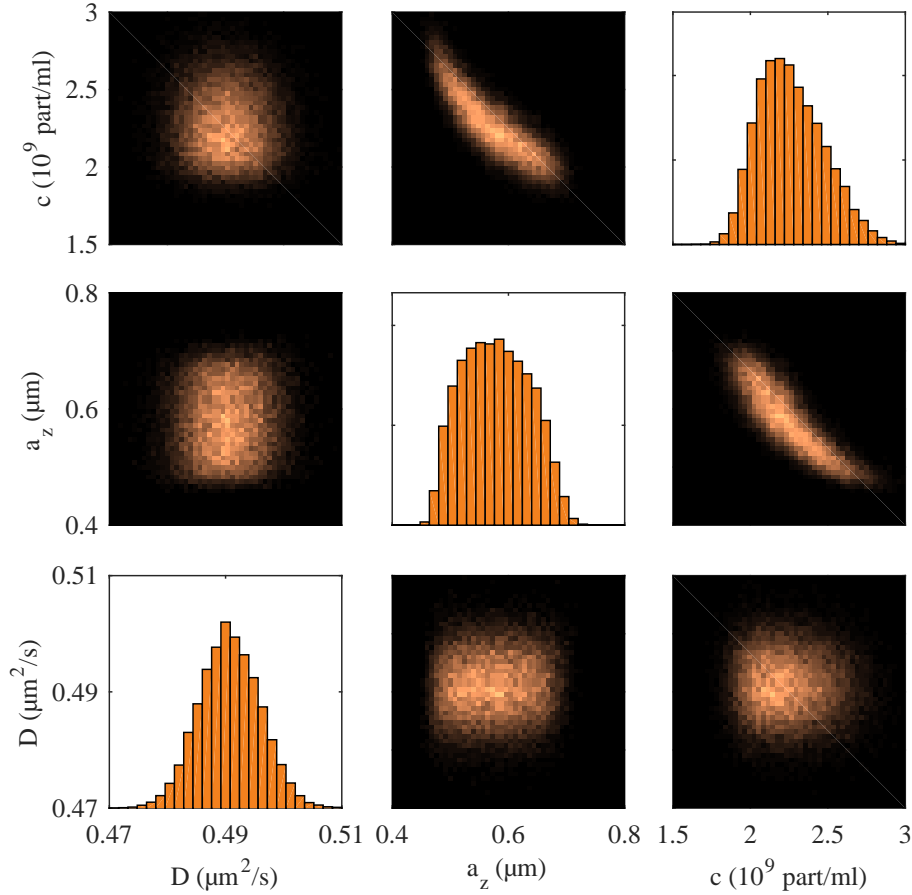
and LPX are used in triplicate. The experimental setup is similar as for the validation experiments, except a Nikon LiveScan Swept Field Confocal Microscope is now used. For each replicate, 25 videos with 5 second duration of individual particles performing free diffusion in the liquid suspension are acquired. The frame rate is approximately 20 frames per second and the resolution is 512 by 512 pixels. The physical pixel size is 0.12 by 0.12 μm ; hence, the lateral size of the detection region is $a_x = a_y \approx 61 \mu\text{m}$. The DNA is labeled with YOYO-1 Iodide (Life Technologies, Thermo Fisher Scientific, Waltham, MA, US), an intercalating dye whose quantum yield become 3200 times higher when complexed with DNA. This ensure brightly fluorescing particles so that purification steps are not necessary; moreover, the dye is quite photostable so that photobleaching is limited. Once again, trajectories shorter than $k_{\text{min}} = 3$ are excluded from further analysis. The tracking was performed at least 10 μm away from the cover glass to avoid deviations from free diffusion due to wall interactions. Gel electrophoresis experiments demonstrate no DNA band after complexation, suggesting complete complexation and hence that no free DNA is left. This suggests that the proposed procedure to quantify the average number of DNA particles encapsulated in a single LPX is appropriate.

V. CONCLUSION

We have considered the problem of estimating number concentrations of freely diffusing monodisperse nanoparticles using optical fluorescence microscopy and particle tracking. No tractable, exact likelihood function can be formulated for the considered case, and in a previously published method approximations had to be introduced. Nevertheless, the stochastic data generating process is simple and thus paves the way for an approximate Bayesian computation framework for joint estimation of the diffusion coefficient, the size of the detection region in which particles are tracked, and the particle number concentration. In an experiment performed on water dispersions of fluorescent polymer nanospheres with well-known concentration, we have demonstrated the validity of the proposed method in a controlled setting. Moreover, we have demonstrated the usefulness of this approach in an application related to nanomedicine mediated gene therapy, applying the method to estimate concentrations of plasmid DNA molecules and the average number of DNA molecules complexed with liposomal drug delivery particles. In conclusion, approximate Bayesian computation, although computationally heavy, provides a very suitable statistical estimation framework for this problem, and we consider it the preferred method of inference over the previously published method because no simplifying approximations are needed for tractability.

TABLE I. Estimated parameter values

Data sets and variables	Replicate no.		
	1	2	3
DNA suspensions			
D ($\mu\text{m}^2/\text{s}$)	0.420 [0.417, 0.423]	0.451 [0.447, 0.455]	0.453 [0.449, 0.456]
a_z (μm)	1.033 [0.955, 1.117]	0.894 [0.816, 0.977]	0.922 [0.848, 0.999]
c (10^9 part/ml)	8.900 [8.513, 9.294]	7.448 [7.074, 7.840]	8.700 [8.300, 9.113]
LPX suspensions			
D ($\mu\text{m}^2/\text{s}$)	0.495 [0.485, 0.505]	0.491 [0.481, 0.501]	0.410 [0.404, 0.416]
a_z (μm)	0.519 [0.433, 0.611]	0.579 [0.483, 0.683]	0.710 [0.621, 0.808]
c (10^9 part/ml)	2.844 [2.346, 3.489]	2.264 [1.916, 2.717]	2.841 [2.581, 3.127]
DNA per LPX ratios			
Q (dimensionless)	1.583 [1.268, 1.909]	1.659 [1.363, 1.962]	1.536 [1.376, 1.705]

FIG. 5. The posterior distribution of (D, a_z, c) for the second LPX replicate.

-
- [1] A. Deniz, S. Mukhopadhyay, and E. Lemke, *Journal of the Royal Society Interface* **5**, 15 (2008).
- [2] A. Nel, L. Mädler, D. Velegol, T. Xia, E. Hoek, P. Somasundaran, F. Klaessig, V. Castranova, and M. Thompson, *Nature materials* **8**, 543 (2009).
- [3] K. Remaut, N. Sanders, B. D. Geest, K. Braeckmans, J. Demeester, and S. D. Smedt, *Materials Science and Engineering: R: Reports* **58**, 117 (2007).
- [4] S. Nune, P. Gunda, P. Thallapally, Y.-Y. Lin, F. Laird, and C. Berkland, *Expert opinion on drug delivery* **6**, 1175 (2009).
- [5] I. Montes-Burgos, D. Walczyk, P. Hole, J. Smith, I. Lynch, and K. Dawson, *Journal of Nanoparticle Research* **12**, 47 (2010).
- [6] K. Braeckmans, K. Buyens, W. Bouquet, C. Vervaet, P. Joye, F. D. Vos, L. Plawinski, L. Doeuve, E. Angles-Cano, N. Sanders, J. Demeester, and S. D. Smedt, *Nano letters* **10**, 4435 (2010).
- [7] G. Chironi, C. Boulanger, A. Simon, F. Dignat-George, J.-M. Freyssinet, and A. Tedgui, *Cell and tissue research*

- 335**, 143 (2009).
- [8] L. Doeuvre, L. Plawinski, F. Toti, and E. Anglés-Cano, *Journal of neurochemistry* **110**, 457 (2009).
- [9] S. Khare and N. Dokholyan, *Current Protein and Peptide Science* **8**, 573 (2007).
- [10] V. Filipe, R. Poole, M. Kutscher, K. Forier, K. Braeckmans, and W. Jiskoot, *Pharmaceutical research* **28**, 1112 (2011).
- [11] V. Filipe, R. Poole, O. Oladunjoye, K. Braeckmans, and W. Jiskoot, *Pharmaceutical research* **29**, 2202 (2012).
- [12] H. Saveyn, B. D. Baets, O. Thas, P. Hole, J. Smith, and P. V. D. Meerem, *Journal of colloid and interface science* **352**, 593 (2010).
- [13] M. Röding, H. Deschout, K. Braeckmans, and M. Rudemo, *Physical Review E* **84**, 031920 (2011).
- [14] M. Röding, H. Deschout, K. Braeckmans, and M. Rudemo, *Journal of microscopy* **251**, 19 (2013).
- [15] M. Röding, H. Deschout, K. Braeckmans, A. Särkkä, and M. Rudemo, *Journal of microscopy* **252**, 79 (2013).
- [16] S. Du, K. Kendall, S. Morris, and C. Sweet, *Journal of chemical technology and biotechnology* **85**, 1223 (2010).
- [17] B. Naeye, H. Deschout, V. Caveliers, B. Descamps, K. Braeckmans, C. Vanhove, J. Demeester, T. Lahoutte, S. D. Smedt, and K. Raemdonck, *Biomaterials* **34**, 2350 (2013).
- [18] E. Zagato, K. Forier, T. Martens, K. Neyts, J. Demeester, S. D. Smedt, K. Remaut, and K. Braeckmans, *Nanomedicine* **9**, 913 (2014).
- [19] H. Deschout, K. Raemdonck, S. Stremersch, P. Maoddi, G. Mernier, P. Renaud, S. Jiguet, A. Hendrix, M. Bracke, R. van den Broecke, M. Röding, M. Rudemo, J. Demeester, S. C. de Smedt, F. Strubbe, K. Neyts, and K. Braeckmans, *Nanoscale* **6**, 1741 (2014).
- [20] K. Forier, A.-S. Messiaen, K. Raemdonck, H. Nelis, S. D. Smedt, J. Demeester, T. Coenye, and K. Braeckmans, *Journal of Controlled Release* **195**, 21 (2014).
- [21] D. Rubin, *The Annals of Statistics* **12**, 1151 (1984).
- [22] P. Diggle and R. Gratton, *Journal of the Royal Statistical Society. Series B (Methodological)* **46**, 193 (1984).
- [23] S. Tavaré, D. Balding, R. Griffiths, and P. Donnelly, *Genetics* **145**, 505 (1997).
- [24] J. Pritchard, M. Seielstad, A. Perez-Lezaun, and M. Feldman, *Molecular Biology and Evolution* **16**, 1791 (1999).
- [25] M. Beaumont, W. Zhang, and D. Balding, *Genetics* **162**, 2025 (2002).
- [26] K. Csilléry, O. François, and M. Blum, *Methods in ecology and evolution* **3**, 475 (2012).
- [27] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, *Proceedings of the National Academy of Sciences* **100**, 15324 (2003).
- [28] S. Sisson, Y. Fan, and M. Tanaka, *Proceedings of the National Academy of Sciences* **104**, 1760 (2007).
- [29] S. Sisson, Y. Fan, and M. Tanaka, *Proceedings of the National Academy of Sciences* **106**, 16889 (2009).
- [30] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstead, S. Schmid, and G. Danuser, *Nature methods* **5**, 695 (2008).
- [31] H. Berg, *Random walks in biology* (Princeton University Press, 1993).
- [32] P. Fearnhead and D. Prangle, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 419 (2012).
- [33] M. Blum, M. Nunes, D. Prangle, and S. Sisson, *Statistical Science* **28**, 189 (2013).
- [34] J. Bezanson, S. Karpinski, V. Shah, and A. Edelman, *CoRR* **abs/1209.5145** (2012).
- [35] J. Bezanson, A. Edelman, S. Karpinski, and V. Shah, *CoRR* **abs/1411.1607** (2014).
- [36] See Supplemental Material at [URL will be inserted by publisher] for code and data sets.