Estimation of bivariate probability distributions of nanoparticle characteristics, based on univariate measurements

Orkun Furat*†[1], Uwe Frank†[2], Matthias Weber[1], Simon Wawra[2], Wolfgang Peukert[2], and Volker Schmidt[1]

[1]Institute of Stochastics, Ulm University, Ulm, Germany
[2]Institute of Particle Technology, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

Abstract

The properties of complex particle systems typically depend on multivariate distributions of particle properties, like size and shape characteristics. Multidimensional particle property distributions can be a very powerful tool to describe these systems. However, only few techniques exist which are able to simultaneously measure more than one property of individual particles in a fast and efficient way. In the present paper it is shown how a two-dimensional property space can be constructed by the combination of two univariate measurements to obtain a bivariate particle size distribution. The proposed reconstruction method is a general approach, which can be applied to a wide spectrum of particle systems and measurement devices. In this paper, the results of a case study are presented, which allow the estimation of the bivariate distribution of length and diameter of gold nanorods, solely using univariate distributions of their particle mass and extinction-weighted sedimentation coefficient distributions. Each of these two quantities contains joint information about the particle lengths and diameters, which is used for the reconstruction. The method is validated in a simulation study in which both the bivariate distribution to be reconstructed and the reconstruction parameters, such as the number of measured univariate distributions used for reconstruction, are varied. In addition, regularization techniques are introduced to reduce methodical errors. This approach can be transferred to other particle systems and measurement techniques, for which functional relationships between vectors of particle properties and measured quantities are well described.

Keywords: multidimensional characterization; particle size distributions; reconstruction; inverse problem; copulas

*Corresponding author. E-mail address: orkun.furat@uni-ulm.de, Phone: +49731/5023555, Fax: +49731/50 23649

†OF and UF contributed equally to this paper.

1 Introduction

Anisotropic plasmonic, surface functionalized, hierarchically structured or other multi-dimensionally characterized particles offer vast possibilities of product design (Amendola et al., 2017; Mehringer et al., 2014; Voigt et al., 2011; Wang et al., 2018). Elucidating the formation mechanisms allows tailoring properties of particulate products. Therefore, suitable characterization in terms of a multi-dimensional property space is mandatory. A particle ensemble can be described by the multivariate distribution of its properties which, in many applications, can be represented by its multivariate probability density. However, despite the importance of accessing multi-dimensional features in contemporary research, techniques to access more than a single particle attribute are rare. While techniques like transmission electron microscopy (TEM) or tomography offer huge capabilities in terms of single particle characterization, they are often tedious and require complex sample preparation (Elazzouzi-Hafraoui et al., 2007). To avoid this, in situ measurement techniques are preferable. In the gas phase, the combination of a differential mobility analyzer and an aerosol particle mass analyzer (DMA-APM) allows obtaining bivariate probability densities of particle size and mass via an inversion routine, which allows investigating the state of nanoparticle aggregation in non-equilibrium plasma synthesis (Rawat et al., 2016; Chen et al., 2018). Another example is the multi-parameter measurement of aerosols through wide-angle light scattering, laser scattering and differential aerodynamic particle sizing (Babick et al., 2018).

Recently, the number-weighted probability density of length and diameter of rod-shaped gold nanoparticles could be determined in the liquid phase via the optical back-coupling (OBC) method (Wawra et al., 2018). This technique relies on coupling of optical and hydrodynamic properties of plasmonic particles within an analytical ultracentrifugation (AUC). For other particle systems, the investigation of multi-dimensional particle characteristics is challenging from an experimental point of view, as different applied measurement techniques obey varying physical principles. Thus, the measured distributions are differently weighted which can lead to non-comparability of particle size distributions (PSDs). This problem was recently overcome by a scheme which enables the conversion of differently weighted multivariate PSDs into each other, additionally the connection between bivariate PSDs and univariate measurements was elaborated in Frank et al. (2019). This development allows predicting the resulting distributions of particle characteristics for different independent measurement techniques. Obviously, the inversion of this procedure is important for applications. More precisely, many laboratories are equipped with standard measurement techniques, like dynamic light scattering (DLS), sedimentation analysis or UV/VIS spectroscopy which measure univariate distributions of particle properties. Thus, the possibility to reconstruct the multivariate distribution from various univariate distributions obtained by measurements could open vast new possibilities. As different as these measurement techniques might be, their respective results all depend on the characterizing multivariate distribution. Therefore, the latter can be

reconstructed by combining measured univariate distributions, as long as the different measurement techniques deliver sufficient information.

In the present paper, the idea of reconstructing bivariate PSDs of gold nanorods, which are characterized by their length ℓ and diameter d, from univariate measurements is elaborated with simulated data. However, the presented reconstruction technique is not necessarily limited to gold nanorods—other particle systems could be investigated as well provided that the functional relationships between vectors of particle properties and measured quantities are well described. For the particle system considered in the present paper, we assume that the bivariate PSD, which will be reconstructed, can be represented by a bivariate piecewise constant probability density. Thus, we have a flexible family of bivariate probability densities which are parameterized by a finite set of function values, denoted by the vector p. Then, we represent these values as a solution of a system of linear equations $Tp = b$, where the entries of the matrix T and the right-hand side b are derived from measured univariate distributions. This inverse problem is then solved in a least-squares sense, i.e., by minimizing the term $\|Tp - b\|_2^2$, where $\|\cdot\|_2$ denotes the Euclidean norm (Kirsch, 2011; Idier, 2013). Note that inverse problems have applications in various fields, like, for example, in geology (Tarantola, 2005), computed tomography (Trampert and Leveque, 1990) and image analysis (Chalmond, 2012). In order to specify the entries of the coefficient matrix T of the linear equations we consider transformation functions γ, given either by analytical formulas or numerical computation schemes, which assign to each vector (ℓ, d) of particle characteristics the corresponding measured quantity $\gamma(\ell, d)$. Therefore, in Section 2 we introduce the transformation functions γ considered in the present paper. Furthermore, some measurement techniques do not provide number-weighted distributions — thus, we describe measurement-specific weight functions in Section 2.2. Then, in Section 2.3 we state the inverse problem which is solved for estimating the bivariate probability density of nanorod length and diameter, and apply the estimation procedure to measured data in Section 2.4. Furthermore, we show how regularization techniques can be utilized to obtain more reliable results (Engl et al., 1996). In Section 2.5, an error analysis is performed via simulation studies for investigating the robustness of the estimation procedure. Section 3 deals with further robustness investigations by applying the presented estimation procedure to various parametric bivariate probability densities. Therefore, we introduce the notion of copulas (Nelsen, 2007) in Section 3.1 which has been used to model bivariate probability densities in various applications like, for example, finance and insurance (McNeil et al., 2005), weather forecast models (von Loeper et al., 2020), climate research (Schölzel and Friederichs, 2008) and the characterization of particles and porous media (Furat et al., 2019; Neumann et al., 2020; Spettl et al., 2017). In Section 3.2 we describe a procedure for fitting a bivariate probability density from parametric families of densities, where the latter are modeled by copulas. The robustness analysis mentioned above for the estimation procedure presented in Section 2.3 is explained in Section 3.3. Finally, Section 4 concludes.

2 Theoretical considerations & results

In the present paper, we assume that each probability distribution considered in the following can be represented by a probability density. Thus, from this point on, we use the terms "probability distribution" and "probability density" interchangeably.

2.1 Number-weighted univariate distributions

From a known bivariate, number-weighted, probability density q_0 of particle characteristics it is possible to obtain number-weighted univariate probability densities in various ways. A simple two-dimensional description of a system of nanorods with hemispherical end caps is given by the joint probability density q_0 of particle length ℓ and diameter d. From such a representation, univariate probability densities of length ℓ and diameter d can be obtained via marginalization, which is equivalent to an integration with respect to the complementary variable. With this, one obtains e.g. the number-weighted probability density of particle lengths

$$\varrho_0(\ell) = \int_0^\infty q_0(\ell, d)\,\mathrm{d}d. \tag{1}$$

If more complex distributions, like the surface or volume/mass distributions shall be computed from q_0, marginalization cannot be directly used as e.g. the volume is a linked variable of length and diameter. For the general case of such a linked variable $y = \gamma(\ell, d)$, described by some transformation function $\gamma\colon \mathbb{R}^2 \to \mathbb{R}$, one obtains for the corresponding (univariate) probability density ϱ_0 by marginalizing (Frank et al., 2019), i.e., by

$$\varrho_0(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_{\{(\ell,d)\in\mathbb{R}^2_+ \,:\, \gamma(\ell,d)\le y\}} q_0(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell. \tag{2}$$

The number-weighted probability density of the particle mass distribution can be, for example, obtained if the function γ in Equation (2) is a suitably selected transformation function for the mass. For nanorods with hemispherical endcaps the volume is obtained as sum of the volumes of a cylinder with diameter d and length $\ell - d$ and a sphere with diameter d. Multiplication with the particle bulk density ρ_{part} yields the mass m_{part} of each rod:

$$\gamma_m(\ell, d) = m_{\mathrm{part}}(\ell, d) = \rho_{\mathrm{part}}\pi d^2 \left(\frac{1}{4}\ell - \frac{1}{12}d\right). \tag{3}$$

The distributions of the aspect ratio ν and surface S can be obtained in a similar manner if γ is chosen as $\gamma_\nu(\ell, d) = \nu(\ell, d) = \frac{\ell}{d}$ and $\gamma_S(\ell, d) = S(\ell, d) = \pi d(\ell - d) + \pi d^2 = \pi d\ell$, respectively.

In the more general case, the measured univariate distributions are weighted as a result of the underlying measurement principle. To obtain the transformed and weighted univariate measured distribution ϱ_κ from q_0, the latter has to

be weighted with the corresponding weight function κ first, resulting in the weighted bivariate PSD q_κ given by

$$q_\kappa(\ell, d) = \frac{q_0(\ell, d)\kappa(\ell, d)}{\int_{\mathbb{R}_+^2} q_0(\ell, d)\kappa(\ell, d)\,\mathrm{d}(d, \ell)} = \frac{q_0(\ell, d)\kappa(\ell, d)}{M_{\kappa,0}} \tag{4}$$

where κ is the weight function and $M_{\kappa,0}$ the involved moment (Frank et al., 2019). From the weighted PSD q_κ the univariate distribution ϱ_κ weighted with κ can be obtained by means of (2):

$$\varrho_\kappa(y) = \frac{\mathrm{d}}{\mathrm{d}y} \int_{\{(\ell,d)\in\mathbb{R}_+^2\,:\,\gamma(\ell,d)\leq y\}} q_\kappa(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell. \tag{5}$$

2.2 Extinction-weighted distribution of the sedimentation coefficient

An important case for the previously discussed weighted univariate distributions are extinction weighted sedimentation coefficients, which will be used in the present paper for the reconstruction of bivariate PSDs. The hydrodynamical properties of particles can be observed with analytical ultracentrifugation (AUC). Here a centrifugal force is applied to a nanoparticle water dispersion within a centrifuge. This induces the sedimentation of nanoparticles to higher radial positions. Wawra et al. (2018) The sedimentation speed u at radial position r and angular velocity ω can be expressed in terms of the sedimentation coefficient s, which solely depends on particle and solvent properties:

$$s = \frac{u}{\omega^2 r} = \frac{m\left(1 - \frac{\varrho_{\text{solvent}}}{\varrho_{\text{part,eff}}}\right)}{3\pi\eta\frac{f}{f_0}x_V}. \tag{6}$$

Here, ϱ_{solvent} corresponds to the solvent mass density, η is the solvent viscosity, $\varrho_{\text{part,eff}}$ the effective mass density of the particle, x_V the volume-equivalent diameter and $\frac{f}{f_0}$ the frictional ratio. The latter is defined as the ratio of the hydrodynamic diameter x_H to the volume-equivalent diameter x_V, which corresponds to the diameter of a sphere experiencing the same drag respectively the same volume as the particle under consideration. Specifically for cylinders the frictional ratio $\frac{f}{f_0}$ is approximately given by

$$\begin{aligned}\frac{f}{f_0} =& 1.0304 + 0.0193\ln(\nu_s) + 0.06229\left(\ln(\nu_s)\right)^2 + 0.00476\left(\ln(\nu_s)\right)^3 \\ &+ 0.00166\left(\ln(\nu_s)\right)^4 + 2.66\cdot 10^{-6}\left(\ln(\nu_s)\right)^7,\end{aligned} \tag{7}$$

where ν_s is the particle's aspect ratio (Hansen, 2004). Equation (7) also provides a good approximation for the frictional ratio $\frac{f}{f_0}$ of gold nanorods with hemispherical end caps (Wawra et al., 2018). Alternatively the expression obtained by Walter et al. (2015) or the Perrin formula for ellipsoids of revolution (Perrin, 1936) could be used. It should be noted that the hydrodynamic

properties measured by AUC correspond to complex particles, which consist of a core particle, a shell including attached ligand and hydration. For shape-anisotropic analytes, like gold nanorods, the sedimentation coefficient in Equation (6) does not only depend on the length, diameter and mass density of the core nanorod, but also on the corresponding parameters of the shell, as shown schematically in Figure SI1 of the Supporting Information (SI). Therefore, x_V, m, $\varrho_{\mathrm{part,eff}}$ as well as $\frac{f}{f_0}$ have to be calculated based on the core-shell particle. Details regarding the determination of these parameters for hemispherical end capped nanorods can be found in the SI, where also the usage of Hansen's formula for hemispherical end capped cylinders is discussed. With Equations (2), (6) and (7) being available, the sedimentation coefficients of all nanorods within a bivariate number-weighted particle size distribution q_0 can be calculated. For plasmonic particles the distribution q_0 can be determined via the optical back-coupling method (OBC) for multiwavelength analytical ultracentrifugation (MWL-AUC). This technique relies on coupling the measured sedimentation coefficient with the optical extinction spectra of the corresponding species. The embedded geometrical information within the unique optical signature of anisotropic plasmonic particles together with s allows to determine the bivariate size distribution within one experiment (Wawra et al., 2018). However, in general, particles do not exhibit this explicit geometrical information within their extinction spectra, which induces that additional information from the spectra is not available and q_0 cannot be determined. Still, the optical properties of the particle ensemble play a decisive role in order to determine the distribution of the sedimentation coefficient. For MWL-AUC, measuring the radial and temporal evolution of the particles' extinction during centrifugation allows to quantify the distribution of the measured sedimentation coefficients. Extinction of light is a combined process of scattering and absorption induced by the presence of an obstacle in the path of light (Bohren and Huffman, 1998). MWL-AUC uses an extinction detector, leading to sedimentation coefficient distributions which are extinction-weighted and not number-weighted, as the particle ensemble is fractionated and quantified according to the optical signal of the individual species. From this follows that the measured distributions depend on the wavelength λ as well as on the optical properties of the analyte under consideration. Its probability density $\varrho_{\mathrm{ext},\lambda}$ can be obtained from the density $q_{\mathrm{ext},\lambda}$ of the extinction-weighted bivariate size distribution:

$$\varrho_{\mathrm{ext},\lambda}(s) = \frac{\mathrm{d}}{\mathrm{d}s} \int_{\left\{(\ell,d)\in\mathbb{R}_+^2 \,:\, \gamma_s(\ell,d)\leq s\right\}} q_{\mathrm{ext},\lambda}(\ell,d)\,\mathrm{d}d\,\mathrm{d}\ell, \tag{8}$$

where γ_s is the transformation function which assigns, to a particle with length ℓ and diameter d, the sedimentation coefficient s given by (6), see Frank et al. (2019). The density of the extinction-weighted bivariate size distribution $q_{\mathrm{ext},\lambda}$ can be obtained via the extinction weighting function $\sigma_{\mathrm{ext},\lambda}$ and the extinction

moment $M_{\mathrm{ext},\lambda,0}$ by means of (4):

$$q_{\mathrm{ext},\lambda}(\ell,d) = \frac{q_0(\ell,d)\sigma_{\mathrm{ext},\lambda}(\ell,d)}{\int_{\mathbb{R}^2_+} q_0(\ell,d)\sigma_{\mathrm{ext},\lambda}(\ell,d)\,\mathrm{d}(d,\ell)} = \frac{q_0(\ell,d)\sigma_{\mathrm{ext},\lambda}(\ell,d)}{M_{\mathrm{ext},\lambda,0}}. \tag{9}$$

Note that the weight function κ was replaced with the length- and diameter-dependent value of the extinction cross section $\sigma_{\mathrm{ext},\lambda}$, representing the fraction of the individual species of the integral extinction signal. Thus, it can be calculated from the scattering and absorption cross sections $\sigma_{\mathrm{sca},\lambda}$ and $\sigma_{\mathrm{abs},\lambda}$ by

$$\sigma_{\mathrm{ext},\lambda}(\ell,d) = \sigma_{\mathrm{sca},\lambda}(\ell,d) + \sigma_{\mathrm{abs},\lambda}(\ell,d), \tag{10}$$

see Bohren and Huffman (1998). The formulas for the calculation of the scattering and absorption cross sections $\sigma_{\mathrm{sca},\lambda}$ and $\sigma_{\mathrm{abs},\lambda}$ are given in the SI. In general, these cross sections can be computed in different ways, which depend on the investigated system. Spheres are described well by the Mie theory, spheroids are calculated via the Mie-Gans theory, but for most particle systems the cross sections have to be computed numerically, e.g., via boundary element methods or finite element methods, based on the Maxwell equations for classical electrical fields (Bohren and Huffman, 1998; Hohenester and Trügler, 2012). Alternatively, semi empirical models can be used (Yu et al., 2017). The extinction-weighted univariate distribution of the sedimentation coefficient can thus be regarded as a conversion of the bivariate number-weighted size distribution via the weighting function in Equation (10) and the sedimentation coefficient in Equation (6).

2.3 Reconstruction of bivariate densities

Now, we present a method for estimating the bivariate PSD of nanorod ensembles from univariate measurements, as they were introduced in Sections 2.1 and 2.2. Therefore, we will assume that the bivariate probability density $q_0\colon \mathbb{R}^2 \to [0,\infty)$ of the length ℓ and diameter d of gold nanorods with hemispherical end caps is piecewise constant. This simple model assumption will make the reconstruction from univariate measurements easier, yet, the assumption is not too restrictive, providing a flexible class of distributions. More precisely, q_0 will be uniquely described by just a few function values which are in a linear relationship with measured quantities. Thus, the reconstruction of q_0 will entail (approximately) solving a system of linear equations. First, we show how these linear equations are obtained from univariate number-weighted measurements. Then, the more general case of non-constantly weighted measurements is considered. In Section 2.4 the method is applied to reconstruct the probability density q_0 of gold nanorods. Note that, even though we only consider bivariate densities, the methods presented in this section can be easily transfered to higher dimensions. Furthermore, although the present paper focuses on the bivariate probability density q_0 of length ℓ and diameter d of gold nanorods, the method could also be utilized when considering other particle systems. More precisely, the reconstruction scheme could be applied to systems of particles which are described by another set of particle characteristics and

for which experimental measurement techniques measure different quantities. To do so, the transformation functions γ matching particle characteristics with measured quantities have to be well known. Since some measurement techniques provide measurements which are weighted differently in dependence of particle characteristics their corresponding weight functions have to be known (e.g. MWL-AUC measurements of gold nanorods have the weight function $\sigma_{\mathrm{ext},\lambda}$).

In this section, we use the mathematical concept of a random vector (L, D) of length and diameter corresponding to the distribution q_0. More precisely, for any rectangle $R \subset \mathbb{R}^2$ the probability that the values of (L, D) belong to R is given by

$$\mathbb{P}\left((L, D) \in R\right) = \int_R q_0(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell. \tag{11}$$

For the estimation procedure of q_0 based on measured *univariate distributions* presented in Sections 2.1 and 2.2, we assume that the density function q_0 has a bounded support $\operatorname{supp} q_0$ and is piecewise constant, i.e., there are n bounded, pairwise disjoint and closed polygons $A_1, \ldots, A_n \subset \mathbb{R}^2$ and some non-negative scalar values $p_1, \ldots, p_n$, such that

$$q_0(\ell, d) = \sum_{k=1}^{n} p_k \mathbb{1}_{A_k}(\ell, d) \tag{12}$$

for each $(\ell, d) \in \mathbb{R}^2$ and $\mathbb{1}_{A_k}$ denotes the indicator function.[1] Note that, since q_0 is a probability density, the values $p_1, \ldots, p_n$ have to fulfill the following linear equation:

$$\int_{\mathbb{R}_+^2} q_0(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell = \sum_{k=1}^{n} p_k |A_k| = 1, \tag{13}$$

where $|A_k|$ denotes the area of the set A_k. Using this approach, we can estimate q_0 by determining the coefficients $p_1, \ldots, p_n$ based on measured univariate distributions. In the following, we will show that $p = (p_1, \ldots, p_n)$ is in a linear relationship with information gained from measured univariate distributions ϱ_κ, i.e., there is some matrix T and a vector b with $Tp = b$, where the entries of T and b depend on measurements and functional relationships between (ℓ, d) and the measured quantities. Figure 1 visualizes how linear equations are obtained from number-weighted measured distributions ϱ_κ, i.e., for $\kappa \equiv 1$. More precisely, by using the piecewise constant representation of q_0 given in (12) the linear relationship between p and information obtained from measurements is given

[1] For some set A the indicator function $\mathbb{1}_A$ is defined by

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

A common convention is to denote $\mathbb{1}_A(x)$ by $\mathbb{1}_{x \in A}$. Also, for intervals $A = (-\infty, y]$ with $y \in \mathbb{R}$ we use the notation $\mathbb{1}_{x \leq y} = \mathbb{1}_{(-\infty, y]}(x)$.

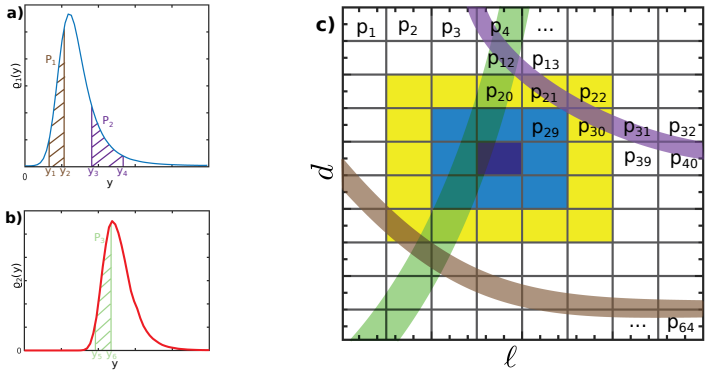


Figure 1: Scheme for obtaining linear equations from univariate number-weighted measurements, which are used to reconstruct the bivariate PSD. Left: two univariate number-weighted distributions ϱ_1 (a) and ϱ_2 (b), which can be described by transforming a bivariate PSD with some transformation functions γ_1 and γ_2, respectively. The brown and purple areas represent the probabilities of the measured value belonging to the corresponding intervals, i.e., $P_1 = \int_{y_1}^{y_2} \varrho_1(y)\mathrm{d}y$ and $P_2 = \int_{y_3}^{y_4} \varrho_1(y)\mathrm{d}y$. Analogously, the green area corresponds to $P_3 = \int_{y_5}^{y_6} \varrho_2(y)\mathrm{d}y$. Right: A piecewise constant bivariate PSD (c) which is uniquely described by its function values $p_1, \dots, p_{64}$ on a 8×8 grid. The "purple stripe" corresponds to the (ℓ, d) vectors with $y_3 \le \gamma_1(\ell, d) \le y_4$. The integral of the bivariate PSD over this stripe has the value P_2 resulting in the linear equation $P_2 = a_3p_3 + a_4p_4 + a_{12}p_{12} + a_{13}p_{13} + \dots + a_{32}p_{32} + a_{39}p_{39} + a_{40}p_{40}$. The coefficient a_k is the area of the intersection of the stripe and the rectangle corresponding to p_k. Analogously, the "brown and green stripes" provide linear equations. By partitioning the supports of ϱ_1 and ϱ_2 into further intervals, or by considering additional univariate measurements we obtain even more linear equations, which are used to determine a solution p.

by

$$\begin{aligned}\int_{y_1}^{y_2} \varrho_\kappa(y)\,\mathrm{d}y &= \int_{\{(\ell,d)\in\mathbb{R}_+^2\colon\, y_1\le\gamma(\ell,d)\le y_2\}} q_0(\ell,d)\,\mathrm{d}d\,\mathrm{d}\ell \\ &= \sum_{k=1}^{n} p_k |\{(\ell,d)\in\mathbb{R}_+^2 \colon y_1 \le \gamma(\ell,d) \le y_2\} \cap A_k|, \end{aligned} \tag{14}$$

where $y_1 \le y_2$ and γ is the transformation function corresponding to the measurement. By varying y_1 and y_2, or by considering other measurements, we obtain additional linear equations. Thus, the estimation of q_0 can be performed by solving the system of linear equations $Tp = b$ for p. In the following, we show how these linear equations are derived in the general case of non-constant weight functions κ.

Note that each univariate cumulative distribution function $\Xi_\kappa\colon \mathbb{R} \to [0,\infty)$ which is measured in the setting of the present paper can be expressed in terms of

the bivariate probability density q_0 of the random vector (L, D). More precisely, using a weight function $\kappa\colon \mathbb{R}^2 \to [0, \infty)$ and a transformation $\gamma\colon \mathbb{R}^2 \to \mathbb{R}$, the measured distribution function Ξ_κ is given by

$$\Xi_\kappa(y) = \mathbb{P}\left(\gamma(\widetilde{L}, \widetilde{D}) \leq y\right) \quad \text{for each } y \in \mathbb{R}, \tag{15}$$

where $(\widetilde{L}, \widetilde{D})$ is a random vector distributed according to a weighted version q_κ of the probability density q_0 which is given by

$$q_\kappa(\ell, d) = \frac{q_0(\ell, d)\kappa(\ell, d)}{M_{\kappa,0}} = \frac{q_0(\ell, d)\kappa(\ell, d)}{\int_{\mathbb{R}^2_+} q_0(\ell, d)\kappa(\ell, d)\,\mathrm{d}(d, \ell)}, \tag{16}$$

provided that $M_{\kappa,0} < \infty$. For example, for the number-weighted mass measurements described in Section 2.1 the weight function is given by $\kappa \equiv 1$ and the value $\gamma(\ell, d)$ of the transformation corresponds to the mass of a nanorod with length ℓ and diameter d given by Equation (3).

By combining Equations (15) and (16) we get

$$\Xi_\kappa(y) M_{\kappa,0} = \int_{\{(\ell,d)\in\mathbb{R}^2_+ \colon \gamma(\ell,d)\leq y\}} q_0(\ell, d)\kappa(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell. \tag{17}$$

Then, the piecewise constant representation of q_0 given in Equation (12) leads to

$$\Xi_\kappa(y) \sum_{k=1}^{n} p_k \int_{A_k} \kappa(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell = \sum_{k=1}^{n} p_k \int_{A_k} \mathbb{1}_{\gamma(\ell,d)\leq y}\,\kappa(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell \tag{18}$$

or, equivalently,

$$\sum_{k=1}^{n} p_k \int_{A_k} \kappa(\ell, d)\left(\mathbb{1}_{\gamma(\ell,d)\leq y} - \Xi_\kappa(y)\right)\,\mathrm{d}d\,\mathrm{d}\ell = 0. \tag{19}$$

Note that for absolutely continuous distribution functions Ξ_κ of measurements, i.e., functions Ξ_κ for which there is a probability density $\varrho_\kappa\colon \mathbb{R} \to [0, \infty)$ with $\Xi_\kappa(y) = \int_{-\infty}^{y} \varrho_\kappa(y')\,\mathrm{d}y'$ for each $y \in \mathbb{R}$, one can show that

$$\sum_{k=1}^{n} p_k \int_{A_k} \kappa(\ell, d)\left(\mathbb{1}_{\gamma(\ell,d)\in B} - \int_B \varrho_\kappa(y)\,\mathrm{d}y\right)\,\mathrm{d}d\,\mathrm{d}\ell = 0 \tag{20}$$

for each interval $B \subset \mathbb{R}$. Equation (20) is the basis for the procedure of estimating q_0 with M measured univariate probability densities $\varrho_{\kappa_1}, \ldots, \varrho_{\kappa_M}$, where the probability density ϱ_{κ_j} is obtained by weighting q_0 with some weight function $\kappa_j\colon \mathbb{R}^2 \to [0, \infty)$ followed by a transformation with the transformation function $\gamma_j\colon \mathbb{R}^2 \to \mathbb{R}$ for each $j = 1, \ldots, M$, see Equations (15) and (16).

Therefore, let Ξ_{κ_j} denote the cumulative distribution function corresponding to the probability density ϱ_{κ_j} for each $j = 1, \ldots, M$, respectively. Furthermore, we assume that ϱ_{κ_j} has a bounded support, i.e., $\operatorname{supp} \varrho_{\kappa_j} \subset \gamma_j(\operatorname{supp} q_0)$ is bounded for each $j = 1, \ldots, M$. For example, due to the bounded support of

q_0, see Equation (12), this condition is fulfilled for continuous transformation functions γ_j which include the transformation functions considered in the present paper, see Equations (3) and (6). Beyond that, this assumption is reasonable since distributions of measured quantities usually have a bounded support. Then, for each measured probability density ϱ_{κ_j}, its support $\operatorname{supp} \varrho_{\kappa_j} \subset [\Xi_{\kappa_j}^{-1}(0), \Xi_{\kappa_j}^{-1}(1))$ is partitioned into $N > 0$ intervals which, based on Equation (20), lead to N linear equations for each measured probability density $\varrho_{\kappa_1}, \ldots, \varrho_{\kappa_M}$. For the partitioning of $\operatorname{supp} \varrho_{\kappa_j}$ we construct disjoint equiprobable intervals. More precisely, for each $j = 1, \ldots, M$ we define a monotone sequence $-\infty < w_0^{(j)} \leq w_1^{(j)} \leq \cdots \leq w_N^{(j)} < \infty$, where

$$w_g^{(j)} = \Xi_{\kappa_j}^{-1}\left(\frac{g}{N}\right) \tag{21}$$

for each $g = 0, \ldots, N$. These values define the endpoints of the intervals $B_1^{(j)}, \ldots, B_N^{(j)}$ which are given by

$$B_g^{(j)} = [w_{g-1}^{(j)}, w_g^{(j)}) \quad \text{for each } g = 1, \ldots, N. \tag{22}$$

By Equation (21), we have

$$\int_{B_g^{(j)}} \varrho_{\kappa_j}(y') \, \mathrm{d}y' = \mathbb{P}\left(\gamma_j(\widetilde{L}, \widetilde{D}) \in B_g^{(j)}\right) = \Xi_{\kappa_j}(w_g^{(j)}) - \Xi_{\kappa_j}(w_{g-1}^{(j)}) = \frac{1}{N}. \tag{23}$$

Thus, $B_1^{(j)}, \ldots, B_N^{(j)}$ partition the support $\operatorname{supp} \varrho_{\kappa_j}$ into equiprobable intervals. Then, by inserting Equation (23) into (20) we obtain the linear equation

$$\sum_{k=1}^{n} p_k \int_{A_k} \kappa_j(\ell, d) \left(\mathbb{1}_{\gamma_j(\ell,d) \in B_g^{(j)}} - \frac{1}{N}\right) \mathrm{d}d \, \mathrm{d}\ell = 0 \tag{24}$$

for each $j = 1, \ldots, M$ and $g = 1, \ldots, N$. Actually, Equation (24) describes a system of $N \cdot M$ linear equations with unknowns $p_1, \ldots, p_n$, i.e., there is a matrix $T \in \mathbb{R}^{(N \cdot M) \times n}$, whose entries are given by (24) such that the system of linear equations is given by

$$Tp = 0. \tag{25}$$

Furthermore, we have the constraints $p_1, \ldots, p_n \geq 0$ and the constraint given in Equation (13) which ensure that the function $\sum_{k=1}^{n} p_k \mathbb{1}_{A_k}$ induced by the solution p is a probability density. If the underlying bivariate probability density q_0 is piecewise constant, this system of linear equations is solvable. If the number of linear independent equations in (25) is smaller than $n - 1$, the solution is not unique. However, this can be remedied by deriving additional linear equations obtained by considering additional univariate probability densities or finer partitions of their supports. Due to numerical errors or the fact that the bivariate probability density q_0, from which measured univariate probability densities $\varrho_{\kappa_1}, \ldots, \varrho_{\kappa_M}$ are obtained, is not piecewise constant, the system of linear equations is not necessarily solvable. Therefore, we compute values for $p_1, \ldots, p_n$, which fulfill the constraints and minimize the mean squared error.

More precisely, instead of the system of linear equations given by (25) we consider the optimization problem

$$\begin{aligned} \min_{p=(p_i)_{i=1}^n} \quad & \|Tp\|_2^2 \\ \text{such that} \quad & \sum_{k=1}^{n} p_k |A_k| = 1 \\ & p_k \geq 0 \text{ for } k = 1, \ldots, n, \end{aligned} \tag{26}$$

where $\|\cdot\|_2$ denotes the Euclidean norm and the entries of the matrix $T \in \mathbb{R}^{(N \cdot M) \times n}$ are given by Equation (24). Then, approximate solutions p for minimization problem given in (26) can be computed iteratively using interior point methods Altman and Gondzio (1998).

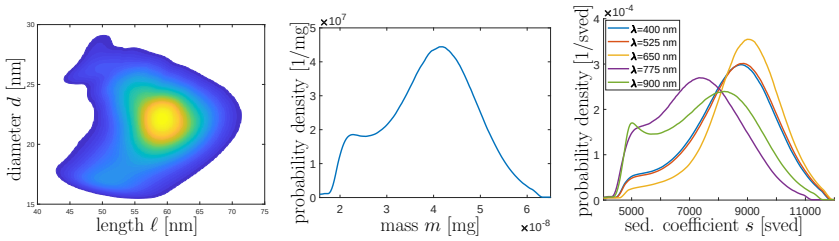
2.4 Application to length and diameter of nanorods

The outlined method is applied to a particle system of gold nanorods, which are distributed in length ℓ and diameter d according to the bivariate PSD q_0. The used univariate measurements are the previously discussed measurements of particle masses m and extinction weighted sediment coefficients s, that are characterized via their univariate probability densities. In order to increase the readability the notation is slightly modified. From hereon, we denote the probability density of the particle mass and sedimentation coefficient by ϱ_m and $\varrho_{s,\lambda}$, respectively, where the latter depends on the wavelength λ.

In order to demonstrate the performance of the reconstruction method a known bivariate PSD of gold nanorods (Wawra et al., 2018) is used as initial data set. From this, univariate distributions of s and m are computed and used as an input for the described method. After reconstruction one receives an estimate $\widehat{q_0}$ for the bivariate PSD, which itself is compared to the true density q_0 for the error investigation of this procedure.

First, the cumulative distribution function Ξ_m of the mass m of particles is derived from the bivariate probability density q_0 by means of Equation (15), with the constant weight function $\kappa \equiv 1$ and the transformation $\gamma = m_{\text{part}}$ given in Equation (3). Then, derivation of Ξ_m leads to the corresponding probability density ϱ_m, see Figure 2b. Analogously, the cumulative distribution function $\Xi_{s,\lambda}$ of the extinction-weighted sedimentation coefficients measured at wavelength λ can be obtained. The corresponding weight function, which depends on the wavelength λ, and the transformation function are given in Equations (10) and (6), respectively. After derivation of $\Xi_{s,\lambda}$, the resulting probability densities $\varrho_{s,\lambda}$ are visualized in Figure 2c for various wavelengths λ. The corresponding bivariate probability density q_0 of particle length and diameter is depicted in Figure 2a.

For estimating the bivariate probability density q_0 based on the probability densities ϱ_m and $\varrho_{s,\lambda_1}, \ldots, \varrho_{s,\lambda_M}$, we assume that diameters and lengths within the system are bounded by minimal and maximal lengths $\ell_{\min}, \ell_{\max}$ and diameters $d_{\min}, d_{\max}$, which induces that the support $\operatorname{supp} q_0$ of q_0 is bounded



(a) (b) (c)

Figure 2: Bivariate probability density q_0 of length and diameter of gold nanorods, adapted with permission from Wawra et al. (2018) under the Creative Commons Attribution 4.0 International License (2020). Dark blue color corresponds to small values and yellow color to large values of the probability density. Negligibly small values are white (a). The corresponding distributions of mass (b) and extinction-weighted sedimentation coefficients of the particles (c) which were computed from (a) via the outlined integration and weighting techniques described in Sections 2.1 and 2.2.

by some rectangle $R = [\ell_{\min}, \ell_{\max}) \times [d_{\min}, d_{\max})$. In real life applications these boundaries can be obtained in two different ways which can also be combined. The synthesis of nanoparticles usually follows strict protocols, which dictate a rough size estimation of the nanoparticles a priori. This information can be used to approximate the lower and upper limits of the rectangle R. Additionally, the univariate measurements yield important information. Figure 1 illustrates that each point on the x-axis of a univariate measurement corresponds to a curve in the 2D length-diameter space. Thus, the lower and upper bounds of the support of the univariate distribution generate two curves between which the support of the bivariate distribution is located. The usage of additional univariate measurements restricts the set in which the support of the bivariate distribution is located even further. Then, the resulting set can be approximated by a rectangle which, a posteriori, defines the bounded support.

The estimated density $\widehat{q}_0$ will be of the form

$$\widehat{q}_0(\ell, d) = \sum_{i,j=1}^{n} p_{i,j} \mathbb{1}_{A_{i,j}}(\ell, d) \quad \text{for each } (\ell, d) \in \mathbb{R}^2 . \tag{27}$$

Thus, $\widehat{q}_0$ is piecewise constant on n^2 pairwise disjoint rectangular sets $A_{i,j}$ (which we call bins from now on) given by

$$A_{i,j} = \left[\ell_{\min} + \frac{i-1}{n}\Delta_\ell, \ell_{\min} + \frac{i}{n}\Delta_\ell\right) \times \left[d_{\min} + \frac{j-1}{n}\Delta_d, d_{\min} + \frac{j}{n}\Delta_d\right) \tag{28}$$

for all $i, j = 1, \ldots, n$, where $\Delta_\ell = \ell_{\max} - \ell_{\min}$ and $\Delta_d = d_{\max} - d_{\min}$. Then, the bins $(A_{i,j})_{i,j=1}^n$ are partitioning the rectangle $R \supset \operatorname{supp} q_0$, i.e., $R = \bigcup_{i,j=1}^{n} A_{i,j}$.

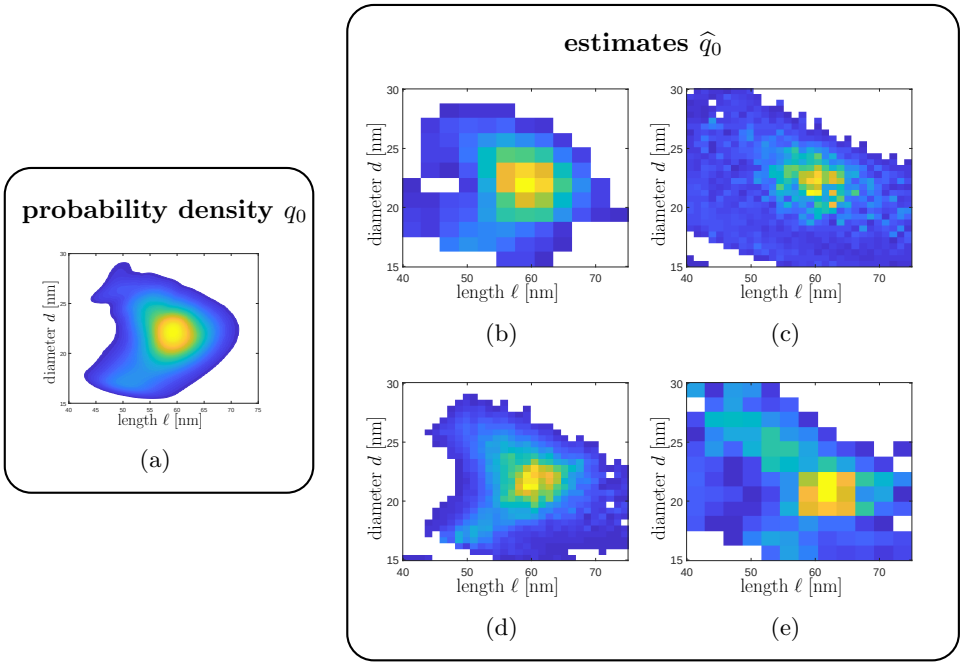


Figure 3: Bivariate probability density q_0 to be reconstructed (a) and the corresponding reconstructions (b-e). The estimated probability density $\widehat{q}_0$ in (b) utilized the mass distribution and the sedimentation coefficient distribution for the wavelength $\lambda = 480\,\text{nm}$ with $n^2 = 12^2$ bins. The same univariate distributions were used to estimate the probability density in (c) with $n^2 = 30^2$ bins. The estimate $\widehat{q}_0$ in (d) was computed with $n^2 = 30^2$ bins using twelve additional sedimentation coefficient distributions. For the estimate in (e) the mass distribution and the sedimentation coefficient distribution for the wavelength $\lambda = 673\,\text{nm}$ was used with $n^2 = 12^2$ bins.

Note that, in general, one could partition the ℓ- and d-axis into n and m intervals with $n \neq m$, respectively, leading to a partition with nm bins. By means of Equation (24), from the univariate probability densities $\varrho_m, \varrho_{s,\lambda_1}, \ldots, \varrho_{s,\lambda_M}$ we obtain a system of $(M+1)N$ linear equations with n^2 unknowns, namely $(p_{i,j})_{i,j=1}^n$, where N is the number of partition components into which the support of the $M+1$ univariate measurements is divided into, see Equation (22). Note that the coefficients of $(p_{i,j})_{i,j=1}^n$ given by (24) are given by integrals which we computed numerically. This can introduce errors in the reconstructed probability densities which are not investigated in the present paper. The values $(p_{i,j})_{i,j=1}^n$ which describe the estimate $\widehat{q}_0$ can then be obtained by finding an optimal solution for the system of linear equations under the constraints $p_{i,j} \geq 0$ and $\sum_{i,j=1}^n p_{i,j}|A_{i,j}| = 1$ with the interior point algorithm, see Equation (26).

Figure 3b depicts the estimated probability density $\widehat{q}_0$, where we used the density ϱ_m of the mass and one density $\varrho_{s,480\mathrm{nm}}$ of sedimentation coefficients with $n = 12$ and $N = 300$. Even though the resolution given by the number of bins $n^2 = 12^2$ is relatively low, the estimated density $\widehat{q}_0$ reassembles q_0 with good accuracy. Quantitatively, the mismatch between $\widehat{q}_0$ and q_0 can be described by the error term $\|\widehat{q}_0 - q_0\| = \left(\int_{\mathbb{R}^2} |\widehat{q}_0(\ell, d) - q_0(\ell, d)|^2 \,\mathrm{d}d\,\mathrm{d}\ell\right)^{\frac{1}{2}}$. Note that $\|\cdot\|$ denotes the L^2 norm.[2] For the estimate $\widehat{q}_0$ depicted in Figure 3b this error $\|\widehat{q}_0 - q_0\|$ is 0.016, see also Table 1. If we utilize the same amount of information for the estimation of $\widehat{q}_0$, namely the univariate probability densities ϱ_m and $\varrho_{s,480\mathrm{nm}}$, but increase the number of bins used in the estimation to $n^2 = 40^2$, the error increases significantly to $\|\widehat{q}_0 - q_0\| = 0.035$, see also Figure 3c. Then, if we include further information in form of twelve additional densities of sedimentation coefficients (for the wavelengths $\lambda \in \{495\,\mathrm{nm}, 510\,\mathrm{nm}, 547\,\mathrm{nm}, 561\,\mathrm{nm}, 575\,\mathrm{nm}, 590\,\mathrm{nm}, 604\,\mathrm{nm}, 618\,\mathrm{nm}, 632\,\mathrm{nm}, 646\,\mathrm{nm}, 660\,\mathrm{nm}, 673\,\mathrm{nm}\}$) we can reduce the error down to $\|\widehat{q}_0 - q_0\| = 0.018$ while remaining at the finer resolution of $n^2 = 40^2$, see Figure 3d.

However, when estimating q_0 with linear equations derived from certain univariate probability densities $\varrho_{s,\lambda}$ by means of (24), a relatively large error $\|\widehat{q}_0 - q_0\|$ can occur. For example, when estimating the bivariate probability density q_0 depicted in Figure 3a for $n^2 = 12^2$ utilizing the univariate probability densities ϱ_m and $\varrho_{s,673\mathrm{nm}}$ we obtain the error $\|\widehat{q}_0 - q_0\| = 0.045$, see also Figure 3e. A possible reason for this effect is that the matrix T whose entries are given by Equation (24) is ill-posed when using certain univariate probability densities $\varrho_{s,\lambda}$ for estimating q_0. The reasons for this effect are currently not fully understood and will be studied in future. The ill-posedness of a matrix T can be characterized by the condition number k which is given by

$$k(T) = \frac{\tau_{\max}}{\tau_{\min}}, \tag{30}$$

[2] The L^2 norm of a bivariate function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\|f\| = \left(\int_{\mathbb{R}^2} f(x,y)^2 \,\mathrm{d}x\mathrm{d}y\right)^{\frac{1}{2}}. \tag{29}$$

Table 1: Reconstruction parameters and errors of the estimates $\widehat{q}_0$ depicted in Figures 3b-4b.

Figure	number of bins n^2	number of measurements $M+1$	regularization	error $\|q_0 - \widehat{q}_0\|$
3b	12^2	2	no	0.0164
3c	30^2	2	no	0.0346
3d	30^2	14	no	0.0177
3e	12^2	2	no	0.0450
4a	12^2	2	yes	0.0467
4b	30^2	2	yes	0.0248

where $\tau_{\min}$ and $\tau_{\max}$ are the smallest and largest singular values of T, respectively, where the singular values of T are given by the square roots of the eigenvalues of the matrix $T^\top T$ (Trefethen and Bau III, 1997). For example, the matrix T used for the estimate depicted in Figure 3e has a condition number of $1.2 \cdot 10^{17}$, whereas the matrix used for estimating the probability density depicted in Figure 3b has a condition number of $6 \cdot 10^{16}$. Note that this effect on the estimate $\widehat{q}_0$ can be decreased by considering additional univariate probability densities $\varrho_{s,\lambda}$, see Figure 3d.

Another approach to handle possibly ill-posed matrices T—causing oscillations in the estimate $\widehat{q}_0$, e.g., see Figures 3c and 3e—are so-called regularization techniques which are used for increasing the stability of solutions of inverse problems with respect to both numerical and measurement errors (Engl et al., 1996). More precisely, it is possible to introduce a regularization term to the function to be minimized in (26), i.e.,

$$\min_{p=(p_i)_{i=1}^n} \quad \|Tp\|_2^2 + \mu \mathcal{R}\left(\sum_{k=1}^n p_k \mathbb{1}_{A_k}\right), \tag{31}$$

where $\mu > 0$ and $\mathcal{R}$ is a functional which, for example, characterizes the amount of oscillation in a bivariate function. One possible way to do so, is considering the squared L^2 norm of the absolute gradient of the estimate, i.e.,

$$\mathcal{R}(q_0) = \left\| \sqrt{(\partial_\ell q_0)^2 + (\partial_d q_0)^2} \right\|^2 = \|\partial_\ell q_0\|^2 + \|\partial_d q_0\|^2, \tag{32}$$

where $\partial_\ell q_0$ and $\partial_d q_0$ denote the partial derivatives of q_0. Since the regularization term suggested in (32) requires, among other things, differentiability which is not given for $q_0 = \sum_{k=1}^n p_k \mathbb{1}_{A_k}$, we propose a modified regularization term which is given by

$$\mathcal{R}(f) = \|q_0 * \partial_\ell g\|^2 + \|q_0 * \partial_d g\|^2, \tag{33}$$

where $g\colon \mathbb{R}^2 \to \mathbb{R}$ is a centered Gaussian kernel and $*$ denotes the convolution operation. For $q_0 = \sum_{k=1}^n p_k \mathbb{1}_{A_k}$, the values of the function $q_0 * \partial_\ell g$ can be

approximated by using discrete convolution kernels or, in other words, by linear combinations of the entries of p. Furthermore, the value of $\|q_0 * \partial_\ell g\|^2 = \int_{\mathbb{R}^2_+} (q_0 * \partial_\ell g)^2(\ell, d)\,\mathrm{d}d\,\mathrm{d}\ell$ can be numerically computed by a weighted sum of values of $(q_0 * \partial_x g)^2$. Therefore, we can determine a matrix L_ℓ, such that $\|q_0 * \partial_\ell g\|^2 \approx \|L_\ell p\|_2^2$ and analogously a matrix L_d with $\|q_0 * \partial_d g\|^2 \approx \|L_d p\|_2^2$. From here on, we therefore use the regularization term

$$\mathcal{R}(p) = \|L_\ell p\|_2^2 + \|L_d p\|_2^2 = \|Lp\|_2^2\,, \tag{34}$$

where $L = \begin{pmatrix} L_\ell \\ L_d \end{pmatrix}$. The representation of the regularization term $\mathcal{R}$ given in (34) has the advantage that the regularized minimization problem is again of the form given in (26). More precisely, the regularized minimization problem is given by

$$\begin{aligned} \min_{p=(p_i)_{i=1}^n} \quad & \|Tp\|_2^2 + \mu\,\|Lp\|_2^2 = \min_{p=(p_i)_{i=1}^n} \quad \left\| \begin{pmatrix} C \\ \sqrt{\mu} L \end{pmatrix} p \right\|_2^2 \\ \text{such that} \quad & \sum_{k=1}^n p_k |A_k| = 1 \\ & p_k \geq 0 \text{ for } k = 1, \ldots, n, \end{aligned} \tag{35}$$

where we determine a good choice for the regularization parameter μ by following the procedure described in Cultrera and Callegaro (2016). Note that there are various procedures for choosing an appropriate regularization parameter μ with different a priori assumptions on the inverse problem. For further references we refer the reader to Sparacino et al. (2001). By utilizing this regularization scheme for computing estimates $\widehat{q}_0$, the latter exhibits less oscillation, which is depicted in Figure 4a. Furthermore, the use of the regularization term can even reduce estimation errors. For example, the estimates depicted in Figures 4b were computed from the same system of linear equations given by T, while for the latter the regularization scheme was employed—reducing the error $\|q_0 - \widehat{q}_0\|$ from 0.0346 to 0.0248.

This first analysis of the reconstruction and regularization scheme made in this section indicates that for larger numbers of bins n^2 used for the estimated density $\widehat{q}_0$ more information in form of univariate probability densities obtained by weighting and transformation of q_0 is required, i.e., more univariate measurements are needed. Furthermore, the utilization of regularization techniques can lead to more realistic estimates $\widehat{q}_0$ and even smaller estimation errors. Therefore, in the next section, we will give a more detailed error analysis.

2.5 Error analysis

The estimation method presented in Sections 2.3 and 2.4 can reconstruct a piecewise constant bivariate with varying accuracy, depending on how well-determined the resulting system of linear equations is. However, several error sources need to be taken into account, e.g.

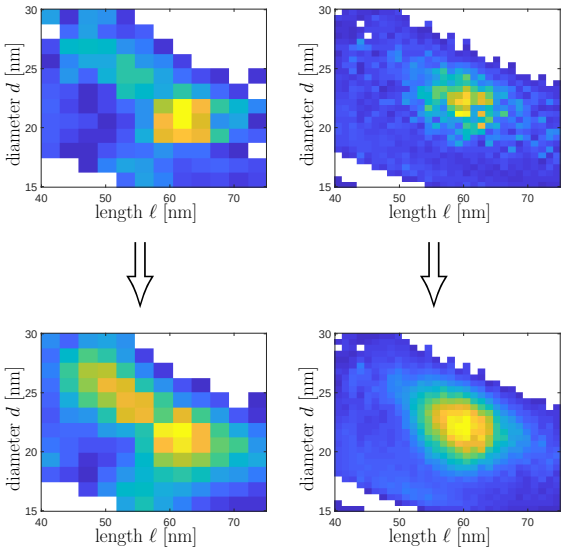


Figure 4: Estimates $\widehat{q_0}$ of the bivariate probability density q_0 which were computed using regularization techniques (bottom row). The corresponding estimates which were computed without regularization are depicted in the top row. For the estimate in (a) the mass distribution and sedimentation coefficient distribution for the wavelength $\lambda = 673\,\text{nm}$ was used with $n^2 = 12^2$ bins. The estimate in (b) was computed with $n^2 = 30^2$ bins using the mass distribution and sedimentation coefficient distribution for the wavelength $\lambda = 480\,\text{nm}$.

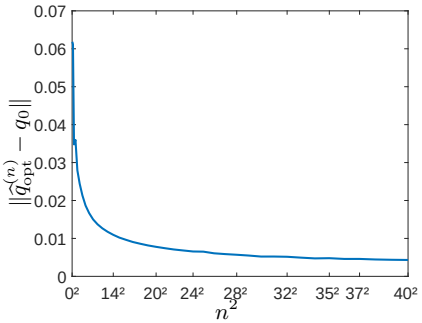


Figure 5: The smallest possible approximation error $\|\widehat{q}_{\text{opt}}^{(n)} - q_0\|$ for the probability density q_0 depicted in Figure 3a versus the number of bins n^2.

(i) numerical errors while applying the method,

(ii) numerical or measurement errors in the univariate densities,

(iii) the real density is *not* piecewise constant, resulting in non-avoidable approximation errors.

On one hand, the numerical (and measurement) errors tend to grow when the number of bins n^2 or the number of partition components N increase. On the other hand, the smallest possible approximation error decreases when using a finer grid size (larger number of bins n^2). More precisely, the best possible piecewise constant approximation $\widehat{q}_{\text{opt}}^{(n)}$ of q_0 with n^2 bins is characterized by

(i) $\widehat{q}_{\text{opt}}^{(n)}$ is a piecewise constant density with n^2 bins, i.e.,

$$\widehat{q}_{\text{opt}}^{(n)} = \sum_{i,j=1}^{n} p_{i,j} \mathbb{1}_{A_{i,j}} \text{ and } \|\widehat{q}_{\text{opt}}^{(n)}\| = 1 \tag{36}$$

(ii) For any bivariate piecewise constant probability density $\widetilde{q}_0 = \sum_{i,j=1}^{n} \widetilde{p}_{i,j} \mathbb{1}_{A_{i,j}}$ with n^2 bins, the discrepancy between $\widetilde{q}_0$ and q_0 has to be larger than between $\widehat{q}_{\text{opt}}^{(n)}$ and q_0, i.e., the following inequality holds:

$$\|\widehat{q}_{\text{opt}}^{(n)} - q_0\| \leq \|\widetilde{q}_0 - q_0\|. \tag{37}$$

Then the value $\|\widehat{q}_{\text{opt}}^{(n)} - q_0\|$ is the smallest possible approximation error as illustrated in Figure 5.

However, a larger number of bins n^2 leads to more unknowns which consequently implies the necessity for more equations, i.e., a larger number of partition components N and/or more univariate distributions. These complex relationships between the different types of errors make it hard to perform

an analysis of the underlying error contributions. Thus, we will perform a simulation study to heuristically determine best practices for applying the reconstruction method proposed in the present paper and reducing the overall error (approximation and numerical error). In order to do so, we computed the estimate $\widehat{q}_0$ of the bivariate probability density q_0, depicted in Figure 3a, for various constellations of the number of bins n^2, the number of considered univariate probability densities M and the number of partition components N. More precisely, in this simulation study, the number of bins n^2 was taken from $\{8^2, 10^2, \ldots, 26^2, 28^2, 30^2\}$, while the number of partition components N was taken from the set $\{100, 300, 500\}$. The number M of considered univariate probability densities was varied from 2 to 32, where the mass distribution ϱ_m was considered for each estimate $\widehat{q}_0$ of the bivariate density q_0. The remaining $M-1$ univariate probability densities were distributions of extinction-weighted sedimentation coefficients at different wavelengths λ between 480 nm and 908 nm. Figure 6 depicts the computed errors $\|\widehat{q}_0 - q_0\|$, which indicate that for small values of the number of bins n^2, the errors stay relatively large. This is expected, a bivariate continuous distribution cannot be properly represented by a few bins. Increasing the number M of considered univariate probability densities up to a number of 20 distributions, results in the considered scenario to relatively large errors. This effect is not understood so far, but it can be reduced by using additional univariate probability densities.

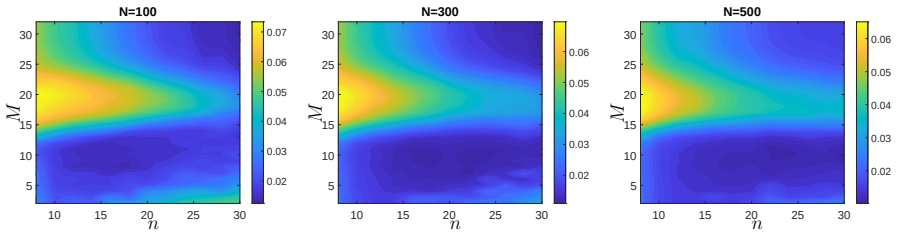


Figure 6: Error $\|\widehat{q}_0 - q_0\|$ for the estimation of the bivariate probability density q_0 depicted in Figure 3a in dependency of the number of bins n^2 and the number of considered univariate distributions considered for computing $\widehat{q}_0$. From each univariate distribution 100 (a), 300 (b) and 500 (c) linear equations were derived.

Since the estimation procedure described in Sections 2.3 and 2.4 is based on the solution of a system of linear equations with $N \cdot M$ equations and n^2 unknowns (or $n^2 - 1$ unknowns, due to the normalization of probability densities, see Equation 13), we now analyze the estimation error with respect to the number $\alpha = NM$ of equations provided. The scatter plot depicted in Figure 7a indicates that the estimation error $\|\widehat{q}_0 - q_0\|$ converges with increasing α. However, there seem to be subsets of points which converge to different lower limits. This effect can be explained by the smallest possible approximation error $\|\widehat{q}^{(n)}_{\text{opt}} - q_0\|$, see Figure 5, which is a lower limit for the estimation error

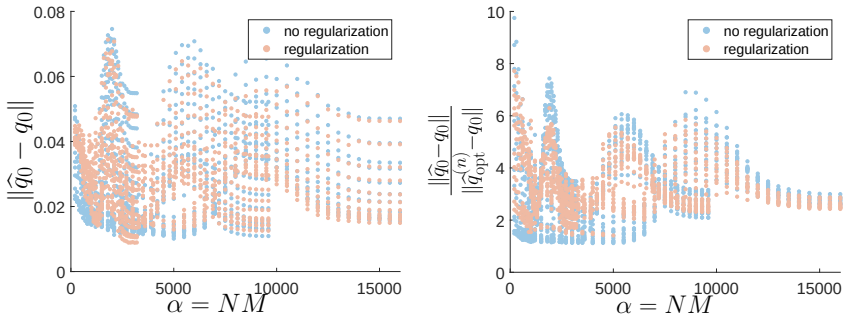


Figure 7: (a): Scatter plot of estimation errors in dependency of the number $\alpha(M, N) = NM$ of linear equations. (b): Scatter plot of normalized estimation errors in dependency of $\alpha(M, N) = NM$.

$\|\widehat{q}_0 - q_0\|$ in dependency of the number of bins n^2. Therefore, in Figure 7b, we consider the normalized estimation error $\frac{\|\widehat{q}_0 - q_0\|}{\|\widehat{q}_{\text{opt}}^{(n)} - q_0\|}$, where we can see that for increasing $\alpha = MN$ the errors seem to converge against a relatively small approximation error of about 2.5. Note, that the estimate $\widehat{q}_0$ depicted in Figure 3d, which visually reassembles the true density q_0 quite nicely, has a normalized estimation error of 3.39. Furthermore, Figure 7 indicates, that the utilization of regularization techniques during the estimation of q_0 can reduce the error even further. Also, we observed that for large values of M the error cannot be decreased by further increasing M. A possible explanation for this is the fact that for larger number M of partitions the integration area ("stripes" in Figure 1) for computing entries of the matrix T become smaller, which might cause redundant linear equations.

3 Reconstruction of parametric distributions

Until now, we used the estimation technique presented in Sections 2.3 and 2.4 to obtain estimates $\widehat{q}_0$ of the bivariate probability density q_0 depicted in Figure 3a. However, the method is not limited to this particular case. Therefore, in this section, we investigate a parametric family $\{q_\theta : \theta \in \Theta\}$ of bivariate probability densities, where Θ is some parameter space. Then, by considering various parameters $\theta_1, \ldots, \theta_z \in \Theta$, we obtain z bivariate probability densities $q_{\theta_1}, \ldots, q_{\theta_z}$ on which we can test our estimation method. This workflow is visualized in Figure 8. First, we derive the univariate probability distributions of measured quantities for each bivariate probability density q_{θ_j} where $j = 1, \ldots, z$. Based on these univariate distributions, the method presented in Sections 2.3 and 2.4, provides estimates $\widehat{q}_{\theta_1}, \ldots, \widehat{q}_{\theta_z}$ of $q_{\theta_1}, \ldots, q_{\theta_z}$, respectively. Thus, we can analyze the performance of the method for a wide range of bivariate distributions. Moreover, it is possible to fit, once again, parametric bivariate

probability densities from the family $\{q_\theta \colon \theta \in \Theta\}$ to the piecewise constant estimates $\widehat{q}_{\theta_j}$ by determining a parameter $\widehat{\theta}_j$ with

$$q_{\widehat{\theta}_j} \approx \widehat{q}_{\theta_j} \tag{38}$$

for each $j = 1, \ldots, z$. This provides the means to investigate the robustness of the estimation method, by comparing the initial parameter θ_j with the parameter $\widehat{\theta}_j$ of the fit.

In order to obtain a wide variety of bivariate probability densities $q_{\theta_1}, \ldots, q_{\theta_z}$ to test the estimation method, we consider so-called copulas which are a useful tool for describing non-Gaussian multivariate distributions with correlated marginals (Nelsen, 2007; Joe, 2014). We will specifically focus on Archimedean copulas which provide a wide variety of parametric bivariate probability densities while still having analytical representations.

3.1 Representation of bivariate distributions by copulas

To demonstrate the importance of copulas for modeling bivariate probability densities q, we have to go back to the notion of the cumulative distribution function $Q \colon \mathbb{R}^2 \to [0,1]$ corresponding to q which is given by

$$Q(x_1, x_2) = \int_{(-\infty, x_1] \times (-\infty, x_2]} q(y_1, y_2) \, \mathrm{d}y_1 \mathrm{d}y_2. \tag{39}$$

Then, the marginal cumulative distribution functions Ξ_1 and Ξ_2 of the bivariate distribution function Q are given by

$$\Xi_1(x) = \lim_{x_2 \to \infty} Q(x, x_2) \tag{40}$$

and

$$\Xi_2(x) = \lim_{x_1 \to \infty} Q(x_1, x). \tag{41}$$

In particular, a bivariate copula $C \colon \mathbb{R}^2 \to [0,1]$ is a bivariate distribution function whose marginal distributions are uniform distributions on the interval $[0,1]$, i.e., a copula C is a bivariate distribution function for which

$$\lim_{x_1 \to \infty} C(x_1, x) = \lim_{x_2 \to \infty} C(x, x_2) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } x \in [0,1], \\ 1, & \text{if } x > 1, \end{cases} \tag{42}$$

see Nelsen (2007). Copulas are of great interest for modeling distributions of dependent random variables since for any copula C and for any univariate distribution functions $\Xi_1, \Xi_2 \colon \mathbb{R} \to [0,1]$ the function $Q \colon \mathbb{R}^2 \to [0,1]$ given by

$$Q(x_1, x_2) = C(\Xi_1(x_1), \Xi_2(x_2)) \quad \text{for each } (x_1, x_2) \in \mathbb{R}^2, \tag{43}$$

is a bivariate distribution function whose marginals are Ξ_1 and Ξ_2 (Nelsen, 2007). Note that there is still another interesting aspect of copulas

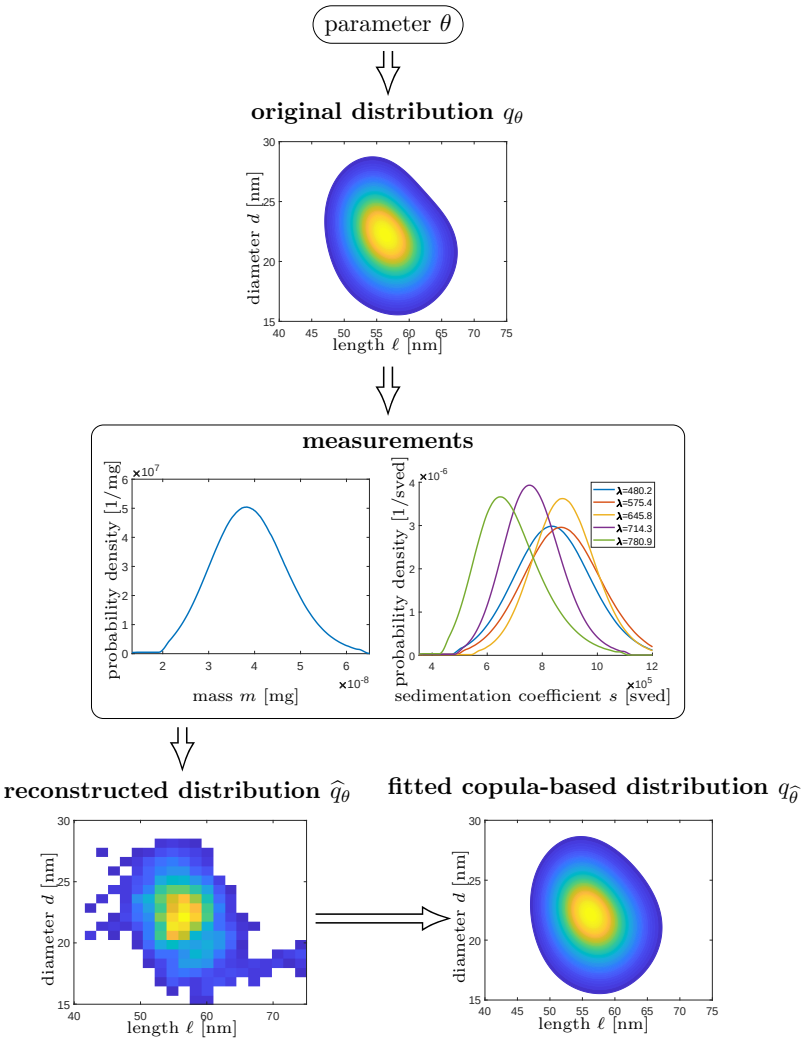


Figure 8: Workflow of the simulation study performed in Section 3. Any parameter vector $\theta \in \Theta$ defines a copula-based bivariate probability density q_θ which are to be reconstructed using the method presented in Sections 2.3 and 2.4. Therefore, univariate distributions of mass and extinction-weighted sedimentation coefficient are derived from q_θ. Using the method presented in Section 2.4 we obtain a piecewise constant estimate $\widehat{q}_\theta$ of q_θ. An additional estimate $q_{\widehat{\theta}}$ of q_θ is obtained by fitting a copula-based probability density to $\widehat{q}_\theta$. For both estimates $\widehat{q}_\theta$ and $q_{\widehat{\theta}}$ the discrepancy to the original probability density q_θ are investigated. Additionally, the discrepancy between the parameter vectors θ and $\widehat{\theta}$ is analyzed, where the latter is given by the estimate $q_{\widehat{\theta}}$. By varying the initial parameter vector $\theta \in \Theta$ we consider a wide range of initial probability densities q_θ to which this workflow can be applied on.

which motivates their use for modeling multivariate distributions. Namely, any bivariate distribution function $Q\colon \mathbb{R}^2 \to [0,1]$ can be represented as in Equation (43) by its marginal distribution functions Ξ_1, Ξ_2 and some bivariate copula C which describes the dependence structure between the two components. The characterization of bivariate distribution functions by their marginals and a copula function, given in Equation (43), holds for any bivariate distribution (Sklar, 1959; Nelsen, 2007). However, the method presented in Sections 2.3 and 2.4 deals with the estimation of bivariate probability densities. Therefore, we consider a differential version of Equation (43). More precisely, let Q be a bivariate distribution function of the form (43), such that the marginal distribution functions Ξ_1, Ξ_2 have probability densities $\varrho_1, \varrho_2\colon \mathbb{R} \to [0,\infty)$, respectively. Moreover, let the copula C be differentiable such that it has the bivariate probability density $c\colon \mathbb{R}^2 \to [0,\infty)$ (also called copula density) given by

$$c(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} C(x_1, x_2). \tag{44}$$

Then, differentiating both sides of Equation (43) twice, implies that Q has the bivariate probability density q which is given by

$$q(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} Q(x_1, x_2) = \varrho_1(x_1)\varrho_2(x_2) c(\Xi_1(x_1), \Xi_2(x_2)). \tag{45}$$

3.2 Fitting of model parameters

Now, we briefly explain how to parametrically model a given bivariate probability density q with copulas. In a first step we parametrically model the univariate probability densities of its marginals ϱ_1, ϱ_2 which are given by

$$\varrho_1(x) = \int_{-\infty}^{\infty} q(x, y)\,\mathrm{d}y \tag{46}$$

and

$$\varrho_2(x) = \int_{-\infty}^{\infty} q(y, x)\,\mathrm{d}y, \tag{47}$$

see also Equation (1). Therefore, we chose suitable parametric families $\{\varrho_{\theta_1}^{(1)}\colon \theta_1 \in \Theta_1\}, \{\varrho_{\theta_2}^{(2)}\colon \theta_2 \in \Theta_2\}$ of univariate probability densities (for example, the log-normal, normal or exponential distribution, etc.) from which we fit densities to ϱ_1 and ϱ_2, respectively. More precisely, we determine optimal parameters $\widehat{\theta}_i \in \Theta_i$ such that

$$\widehat{\theta}_i = \arg\min_{\theta_i \in \Theta_i} \|\varrho_i - \varrho_{\theta_i}^{(i)}\| \tag{48}$$

holds for $i = 1, 2$. In a second step, we chose a suitable parametric family $\{c_{\theta_3}\colon \theta_3 \in \Theta_3\}$ of copula densities for modeling q. Note that, similarly to the univariate case, there are numerous parametric families of copula densities which

can be used for modeling bivariate distributions (Joe, 2014; Nelsen, 2007). Using Equation (45) we then obtain a parametric family

$$\Big\{ q_{\theta_1,\theta_2,\theta_3} \colon (\theta_1,\theta_2,\theta_3) \in \Theta_1 \times \Theta_2 \times \Theta_3, \\ q_{\theta_1,\theta_2,\theta_3}(x_1,x_2) = \varrho^{(1)}_{\theta_1}(x_1)\, \varrho^{(2)}_{\theta_2}(x_2)\, c_{\theta_3}\left(\Xi^{(1)}_{\theta_1}(x_1), \Xi^{(2)}_{\theta_2}(x_2)\right) \Big\} \quad (49)$$

of bivariate probability densities, where $\Xi^{(1)}_{\theta_1}, \Xi^{(2)}_{\theta_2}$ denote the distribution functions of $\varrho^{(1)}_{\theta_1}$ and $\varrho^{(2)}_{\theta_2}$, respectively. Then, the marginal densities of the bivariate density $q_{\theta_1,\theta_2,\theta_3}$ belong to the parametric families $\{\varrho^{(1)}_{\theta_1} \colon \theta_1 \in \Theta_1\}, \{\varrho^{(2)}_{\theta_2} \colon \theta_2 \in \Theta_2\}$. Especially, when choosing $\theta_1 = \widehat{\theta}_1$ and $\theta_2 = \widehat{\theta}_2$ according to (48), the bivariate density $q_{\widehat{\theta}_1,\widehat{\theta}_2,\theta_3}$ has, for each $\theta_3 \in \Theta_3$, the marginal probability densities $\varrho^{(1)}_{\widehat{\theta}_1}, \varrho^{(2)}_{\widehat{\theta}_2}$ which were fitted in the first step. Then, we obtain from the parametric family of bivariate probability densities given in (49) a parametric fit $q_{\widehat{\theta}_1,\widehat{\theta}_2,\widehat{\theta}_3}$ of the initial bivariate density q by solving

$$\widehat{\theta}_3 = \arg\min_{\theta_3 \in \Theta_3} \|q - q_{\widehat{\theta}_1,\widehat{\theta}_2,\theta_3}\|. \quad (50)$$

The main advantage of copulas for modeling bivariate distributions is that it is possible to model the marginal distributions first. Then, in a second step the correlation structure is modeled with copulas. Due to this hierarchical modeling approach, the resulting bivariate density has marginals which are identical to the univariate fits of the first step.

3.3 Simulation study: Fitting Archimedean copulas

In this section we investigate the robustness of the method proposed in Sections 2.3 and 2.4 for the reconstruction of parametric bivariate densities. Assume that the true bivariate probability densities q_{θ_3} which are to be estimated are given by

$$q_{\theta_3}(\ell, d) = \varrho_1(\ell)\varrho_2(d)c_{\theta_3}(\Xi_1(\ell), \Xi_2(d)), \quad (51)$$

where Ξ_1 is the cumulative distribution function of nanorod lengths which is assumed to be log-normally distributed with fixed parameters $\mu = 4$ and $\sigma = 0.075$ which is truncated on the interval $[40, 75]$. Furthermore, we assume that the cumulative distribution function Ξ_2 of diameters follows a normal distribution with parameters $\mu = 22$ and $\sigma = 2.7$ which was truncated on the interval $[15, 30]$, where ϱ_1, ϱ_2 are the probability densities of Ξ_1 and Ξ_2 respectively. Note that the marginal distributions Ξ_1 and Ξ_2 were chosen in this manner, such that they fit the marginal distributions of the probability density depicted in Figure 3a. Finally we assume that the bivariate function

c_{θ_3} in Equation (51) is the density of an Ali-Mikhail-Haq copula which is a parametric copula given by

$$c_{\theta_3}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} \frac{x_1 x_2}{1 - \theta_3(1 - x_1)(1 - x_2)}, \tag{52}$$

with copula parameter $\theta_3 \in (-1, 1)$ (Ali et al., 1978; Nelsen, 2007; Joe, 2014). In the simulation study presented in this section, we have chosen the values for the copula parameter θ_3 from the set $\{\theta_{3,k} = -1 + \frac{2(k-1)}{15} : k = 2, \ldots, 15\}$. Thus, we have 14 bivariate densities $q_{\theta_{3,k}}$ from which we can compute univariate densities which represent measurements, such that the estimation scheme presented in Sections 2.3 and 2.4 can be applied to obtain piecewise constant estimates. In the simulation study the number M of considered bivariate probability densities was varied from 2 to 32, while the number of bins and partition components were fixed to $n = 20^2$ and $N = 300$, respectively.

After computing (either regularized or non-regularized) estimates $\widehat{q}_{\theta_{3,k}}$ for $q_{\theta_{3,k}}$ for various values of M, see Figure 9a (blue, yellow) for errors in dependency of $\alpha = NM = 300M$, we use the approach described in Section 3.2 to fit parametric bivariate densities $q_{\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_{3,k}}$ to the estimates. As already mentioned above, for the parametric family of bivariate probability densities, see (49), we chose bivariate densities which have log-normally and normally distributed marginals with Ali-Mikhail-Haq copulas describing their dependence structure. Figure 9a indicates a significant decrease of errors when fitting a density from a parametric family to the piecewise constant estimates. A reason for this might be the fact that the true bivariate densities $q_{\theta_{3,k}}$ are smooth. Thus, by fitting bivariate densities from a smooth parametric family to the piecewise constant estimates $\widehat{q}_{\theta_{3,k}}$, we obtain more realistic smooth versions of $\widehat{q}_{\theta_{3,k}}$.

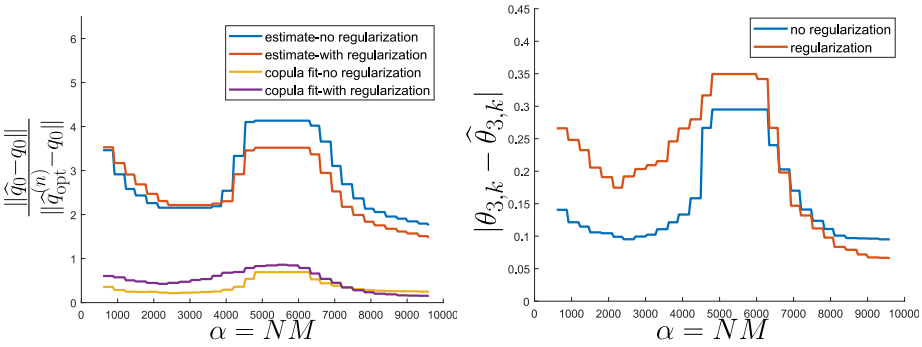


Figure 9: (a): Moving median of estimation errors $\|\widehat{q} - q_{\theta_{3,k}}\|$ and $\|q_{\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_{3,k}} - q_{\theta_{3,k}}\|$ in dependency of the number $\alpha(M, N) = NM$ of linear equations for piecewise-constant estimates (blue, yellow) and for parametric fits (red, purple), respectively. (b): Error of the copula parameter in dependency of $\alpha(M, N) = NM$.

Since the bivariate densities q_{θ_k}, which are to be estimated, and the parametric fits belong to the same parametric family, we are able to investigate how well we estimated model parameters. More precisely, we obtain for each configuration of model parameters (n, N, M) and for each $k = 1, \ldots, 14$ an estimate $\widehat{\theta}_{3,k}$ of the real copula parameter $\theta_{3,k}$. The estimation errors $|\theta_{3,k} - \widehat{\theta}_{3,k}|$ are visualized in Figure 9b, where we observe that for some values of $\alpha = NM$ the copula parameter is estimated rather poorly. However, when using additional information, i.e., when increasing the values of α, the errors are relatively small.

Therefore, the error analysis performed in this section indicates that the method described in Sections 2.3 and 2.4 for estimating the joint distribution of length and diameter of gold nanorods based on univariate measurements can be utilized successfully for different manifestations of the underlying bivariate PSD q_0.

4 Conclusions

Particle technology faces the development of particle systems of ever increasing complexity in terms of disperse properties like their size, shape and composition. Multidimensional particle size distributions can be a powerful tool to describe these particle ensembles. However, apart from a few examples, measurement techniques providing this multidimensional information are missing (Wawra et al., 2018; Li et al., 2020). Therefore, we introduced a general approach to reconstruct multidimensional property distributions. Exemplarily, the combination of independently measured univariate mass and sedimentation coefficient distributions allows the reconstruction of the bivariate size distribution of gold nanorods. This is possible as the different univariate measurements are results of the same underlying multi-dimensional probability distribution of the particle system. The performance of the reconstruction is evaluated via comparison to the original bivariate size distribution. A forthcoming study will focus on the mathematical details, e.g., uniqueness and stability of the solution, and the experimental realization of the technique for different particle systems and sets of measurement techniques.

Supporting information

Supporting information is available in the online version or from the author.

Acknowledgments

The authors are grateful to the Deutsche Forschungsgemeinschaft (DFG) for financial support through the priority program SPP 2045: “Highly specific and multidimensional fractionation of fine particle systems with technical relevance.”

Conflict of interest

The authors declare no conflict of interest.

Data availability

The measurement data used in this study has been part of a prior study (Wawra et al., 2018) and is available upon request at `https://doi.org/10.1038/s41467-018-07366-9`.

References

Ali, M. M., Mikhail, N., and Haq, M. (1978). A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8(3):405 – 412.

Altman, A. and Gondzio, J. (1998). Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. *Optimization Methods and Software*, 11-12:275–302.

Amendola, V., Pilot, R., Frasconi, M., Maragò, O. M., and Iatì, M. A. (2017). Surface plasmon resonance in gold nanoparticles: A review. *Journal of Physics: Condensed Matter*, 29(20):203002.

Babick, F., Hillemann, L., Stintz, M., Dillenburger, T., Pitz, M., Hellmann, A., Antonyuk, S., Ripperger, S., Huber, F. J., Will, S., Wernet, R., Seipenbusch, M., Gensch, M., Weber, A., Kiesler, D., Kruis, E., Friehmelt, R., and Sachweh, B. (2018). Multiparameter characterization of aerosols. *Chemie Ingenieur Technik*, 90(7):923–936.

Bohren, C. F. and Huffman, D. R. (1998). *Absorption and Scattering of Light by Small Particles*. J. Wiley & Sons.

Chalmond, B. (2012). *Modeling and Inverse Problems in Imaging Analysis*. Springer.

Chen, X., Ghosh, S., Buckley, D. T., Sankaran, R. M., and Hogan Jr, C. J. (2018). Characterization of the state of nanoparticle aggregation in non-equilibrium plasma synthesis systems. *Journal of Physics D: Applied Physics*, 51(33):335203.

Cultrera, A., and Callegaro, L. (2018). A simple algorithm to find the L-curve corner in the regularization of inverse problems. *arXiv preprint*, arXiv:1608.04571.

Elazzouzi-Hafraoui, S., Nishiyama, Y., Putaux, J.-L., Heux, L., Dubreuil, F., and Rochas, C. (2007). The shape and size distribution of crystalline nanoparticles prepared by acid hydrolysis of native cellulose. *Biomacromolecules*, 9(1):57–65.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*. Springer.

Frank, U., Wawra, S. E., Pflug, L., and Peukert, W. (2019). Multidimensional particle size distributions and their application to nonspherical particle systems in two dimensions. *Particle & Particle Systems Characterization*, 36(7):1800554.

Furat, O., Leißner, T., Bachmann, K., Gutzmer, J., Peuker, U., and Schmidt, V. (2019). Stochastic modeling of multidimensional particle properties using parametric copulas. *Microscopy and Microanalysis*, 25(3):720–734.

Hansen, S. (2004). Translational friction coefficients for cylinders of arbitrary axial ratios estimated by Monte Carlo simulation. *The Journal of Chemical Physics*, 121(18):9111–9115.

Hohenester, U. and Trügler, A. (2012). MNPBEM–A Matlab toolbox for the simulation of plasmonic nanoparticles. *Computer Physics Communications*, 183(2):370–381.

Idier, J. (2013). *Bayesian Approach to Inverse Problems*. J. Wiley & Sons.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman and Hall/CRC.

Kirsch, A. (2011). *An Introduction to the Mathematical Theory of Inverse Problems*. Springer.

Li, C., Lee, A. L., Chen, X., Pomerantz, W. C. K., Haynes, C. L., and Hogan, C. J. (2020). Multidimensional nanoparticle characterization through ion mobility-mass spectrometry. *Analytical Chemistry*, 92:2503–2510.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Mehringer, C., Wagner, R., Jakuttis, T., Butz, B., Spiecker, E., and Peukert, W. (2014). Gas phase synthesis of anisotropic silicon germanium hybrid nanoparticles. *Journal of Aerosol Science*, 67:119 – 130.

Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer.

Neumann, M., Charry Machado, E., Zojer, K., and Schmidt, V. (2020). On variability and interdependence of local porosity and local tortuosity in porous materials: A case study for sack paper. *Methodology and Computing in Applied Probability*. (in print).

Perrin, F. (1936). Mouvement Brownien d'un ellipsoide (II). Rotation libre et dépolarisation des fluorescences. Translation et diffusion de molécules ellipsoidales. *Journal de Physique et le Radium*, 7(1):1–11.

Rawat, V. K., Buckley, D. T., Kimoto, S., Lee, M.-H., Fukushima, N., and Hogan Jr, C. J. (2016). Two dimensional size–mass distribution function inversion from differential mobility analyzer–aerosol particle mass analyzer (DMA–APM) measurements. *Journal of Aerosol Science*, 92:70–82.

Schölzel, C. and Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research — introduction to the copula approach. *Nonlinear Processes in Geophysics*, 15(5):761–772.

Sklar, M. (1959). Fonctions de répartition àn dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231.

Sparacino, G., De Nicolao, G. and Cobelli, C. (2001). Deconvolution. In: Carson, E. and Cobelli, C. (eds.), *Modeling Methodology for Physiology and Medicine*, Academic Press, Biomedical Engineering Series. 45–75.

Spettl, A., Dosta, M., Klingner, F., Heinrich, S., and Schmidt, V. (2017). Copula-based approximation of particle breakage as link between DEM and PBM. *Computers & Chemical Engineering*, 99:158–170.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM.

Trampert, J. and Leveque, J.-J. (1990). Simultaneous iterative reconstruction technique: Physical interpretation based on the generalized least squares solution. *Journal of Geophysical Research: Solid Earth*, 95(B8):12553–12559.

Trefethen, L. N. and Bau III, D. (1997). *Numerical Linear Algebra*. SIAM.

Voigt, M., Klaumünzer, M., Ebel, A., Werner, F., Yang, G., Marczak, R., Spiecker, E., Guldi, D. M., Hirsch, A., and Peukert, W. (2011). Surface functionalization of ZnO nanorods with C60 derivatives carrying phosphonic acid functionalities. *Journal of Physical Chemistry C*, 115(13):5561–5565.

von Loeper, F., Schaumann, P., de Langlard, M., Hess, R., Bäsmann, R., and Schmidt, V. (2020). Probabilistic prediction of solar power supply to distribution networks, using forecasts of global horizontal irradiation. *Solar Energy*. (submitted).

Walter, J., Nacken, T. J., Damm, C., Thajudeen, T., Eigler, S., Peukert, W. (2015). Determination of the lateral dimension of graphene oxide nanosheets using analytical ultracentrifugation. *Small*, 11(7):814–825.

Wang, J., Mbah, C. F., Przybilla, T., Apeleo Zubiri, B., Spiecker, E., Engel, M., and Vogel, N. (2018). Magic number colloidal clusters as minimum free energy structures. *Nature Communications*, 9(1):5259.

Wawra, S. E., Pflug, L., Thajudeen, T., Kryschi, C., Stingl, M., and Peukert, W. (2018). Determination of the two-dimensional distributions of gold nanorods by multiwavelength analytical ultracentrifugation. *Nature Communications*, 9(1):4898.

Yu, R., Liz-Marzán, L. M., and García de Abajo, F. J. (2017). Universal analytical modeling of plasmonic nanoparticles. *Chemical Society Reviews*, 46:6710–6724.