

# Computational workflow for the characterization of size, shape and composition of particles and their separation behavior during processing

Sabrina Weber<sup>a,1,\*</sup>, Orkun Furat<sup>a,\*</sup>, Tom Kirstein<sup>a</sup>, Thomas Leißner<sup>b</sup>, Urs A. Peuker<sup>b</sup>, Volker Schmidt<sup>a</sup>

<sup>a</sup> *Institute of Stochastics, Ulm University, D-89069 Ulm, Germany*

<sup>b</sup> *Institute of Mechanical Process Engineering and Mineral Processing, Technische Universität Bergakademie Freiberg, D-09599 Freiberg, Germany*

## Abstract

Separation functions, so-called Tromp functions, are often used to quantitatively analyze the separation behavior in particle processing with respect to individual particle descriptors. However, since the separation behavior of particles is typically influenced by multiple particle descriptors, multivariate Tromp functions are required. This study focuses on methods that allow for the computation of multivariate parametric Tromp functions by means of statistical image analysis and copula based modeling. The computations are exemplarily performed for magnetic separation of Li-bearing minerals, including quartz, topaz, zinnwaldite, and muscovite based on micro-computed tomography images and scanning electron microscopy with energy-dispersive X-ray spectroscopy analysis. In particular, the volume equivalent diameter, zinnwaldite fraction, flatness and sphericity are examined as possible influencing particle descriptors. Moreover, to compute the Tromp functions, the probability distributions of these descriptors for concentrate and tailing should be used. In this study 3D image data depicting particles in feed, concentrate and tailings is available for the computation of Tromp functions. However, concentrate particles tend to be elongated, plate-like and densely packed, making segmentation for extracting individual particles from image data extremely difficult. Thus, information on the concentrate could not be obtained from the available database. To remedy this, an indirect optimization approach is used to estimate the distribution of particle descriptors of the concentrate. It turned out that this approach can be successfully applied to analyze the influence of size, shape and composition of particles on their separation behavior.

**Keywords**— Computed tomography, multivariate Tromp function, particle descriptor, probability density, separation process, statistical image analysis, stochastic modeling, copula

## 1 Introduction

Physical separation processes are often used to concentrate valuable components, allowing for the recovery of significant materials from intermediate streams. These processes usually work with a certain primary separation descriptor. To mention some examples, in the case of density separation this descriptor is the material density of particles, in magnetic separation it is the magnetizability of particles, and in flotation it is the wettability of the particle surface. However, the separation processes often do not only depend on a single particle descriptor, but are often influenced by further particle descriptors such as particle size or shape. The relationships between particle descriptors and separation behavior are therefore not univariate, but inevitably multivariate [1–4]. Thus, if a separation process is to be comprehensively characterized or modeled, all relevant particle descriptors must be taken into account and their interaction must be analyzed.

In many cases, particle systems are characterized by probability distributions of particle descriptors. These include the distributions of particle size, mass and composition (components) [5]. In general, the particle descriptors are correlated and the univariate distributions of individual particle descriptors do not capture such correlations. Thus, to characterize separation processes, the descriptor distributions must therefore be determined before and after separation and combined with each other. This can either be done directly using suitable multidimensional

---

<sup>1</sup>Corresponding author, *Email addresses: sabrina.weber@uni-ulm.de*

\*SW and OF contributed equally.

measurement methods (correlative methods) [6, 7], or the univariate descriptor distributions can be used to obtain a multivariate distribution by means of copulas [8, 9].

Once particle-based data is available for the analysis of separation processes, a statistical problem often needs to be solved first. Even if data of a large number of individual particles is available in the feed and tailings, it is often the case that there are only a few particles for which certain combinations of descriptors can be determined, thus impairing the reliability of the results. Therefore, various approaches exist, such as virtual sampling through bootstrap resampling, to determine the coefficient of variation of particle descriptors [10]. In this way, the representativeness of the separation function can be evaluated, allowing for a quantitative analysis of the separation behavior for each combination of descriptors. In the present paper, a threshold value is introduced, similar to [11], which defines a required minimum probability for observing certain combinations of descriptors in the feed. This ensures adequate information on the separation probability of the considered descriptor combinations.

In recent years, various methods have been developed that enable a multidimensional view on separation processes. This goal was pursued in particular in the framework of priority program (PP) 2045 of the German Research Foundation (DFG) dealing with “Highly specific and multidimensional fractionation of particle systems with technical relevance”. As already mentioned above, conventional methods characterize separation processes by means of univariate Tromp curves, where the influence of particle sizes is taken into account, in addition to the primary separation descriptor [12]. Furthermore, in [13], the flotation constant has been plotted as a function of particle size and liberation. To achieve this, the particles are sorted into different size categories. For each category, points are plotted on a graph, with axes representing the flotation constant and liberation classes. Another approach is to illustrate the recovery of a target component as a function of two descriptors on a two-dimensional grid. In this grid, each box is colored based on the number of occurrences of the target component, depending on the choice of descriptors [1]. If one wants to describe the selectivity in several dimensions, one possible approach is to compute entropy values for separation processes [5]. This enables the complex interactions that influence the separation result to be reduced to a single characteristic value. However, the multidimensional relationships are neither apparent nor quantified, and further detailed analysis of relevant data is required to identify them.

In the project “Stochastic modeling of multidimensional particle properties with parametric copulas for the investigation of microstructure effects on the fractionation of fine particle system” within PP 2045, we have developed numerous computational methods for statistical image processing (e.g., segmentation of image data into individual particles), stochastic modeling of particle systems using multivariate probability distributions, and quantitative characterization of separation processes by means of multivariate Tromp functions, see [8, 9, 11, 14–24]. The present paper serves as final report of the results obtained in the project of PP 2045 mentioned above. In particular, we describe a computational workflow for the characterization of particles/separation processes which can be deduced from the methods developed in this project. Note that this workflow does not depend on the length scale of the particle system under consideration, i.e., it solely requires image data that adequately resolves particles.

Moreover, in the present paper, new results are presented, by applying this workflow to magnetic separation processes performed on a Li-containing ore. The mineral ore is a Li-bearing mica (zinnwaldite), which differs in its magnetic descriptors and particle shape from the other minerals. In order to investigate the influence of the shape, size and composition of such particles on their magnetic separation behavior, multivariate Tromp functions are computed. Therefore, a parametric approach is presented to determine Tromp functions based on particle descriptor vectors, which are computed from 3D micro-computed tomography ( $\mu$ -CT) image data and from 2D scanning electron microscopy (SEM) with energy-dispersive X-ray spectroscopy (EDS) data of feed and tailings. For this, first, univariate parametric distributions for individual particle descriptors are fitted and then, in a subsequent step, the interdependencies between these descriptors are modeled with copulas to obtain a joint multivariate distribution of all relevant properties of the particle system [11]. Since for the distribution of descriptor vectors with dimension larger than two, so-called vine copulas have proven to be a flexible modeling tool [9, 25], they are used to fit the joint multivariate distributions of particle descriptor vectors in the presented paper. The resulting parametric distributions are used to compute multivariate Tromp functions in order to investigate the influence of particle descriptor vectors on the separation behavior of the particle system under consideration. However, to avoid numerical instabilities when computing the Tromp function, the distributions of particle descriptor vectors for concentrate and tailings are required. In the case of missing information regarding the separated fraction, an optimization approach has been introduced in [11] to nevertheless investigate the separation behavior. This approach is used in the present paper to compute multivariate Tromp functions for magnetic separation processes.

This paper contains a description of the mineral ore and the magnetic separation process. Additionally, it briefly covers image acquisition and data post-processing, followed by an explanation of the mathematical methods to

determine the Tromp functions based on size, shape and particle composition.

## 2 Materials and Methods

This section outlines the data and magnetic separation process, along with a detailed explanation of the imaging, segmentation, analysis and modeling used to determine Tromp functions.

### 2.1 Test Material

The material examined in this study is a Li-bearing ore from the Zinnwald deposit (Erzgebirge, Saxony). The main components are quartz, zinnwaldite, topaz and muscovite (in decreasing order with respect to their mass/volume fractions). Smaller amounts of iron oxides and other minerals can also be found in the material. The valuable mineral zinnwaldite (Li-containing mica, around 1.6% Li) as well as small amounts of muscovite (mica) are paramagnetic and can be extracted into a magnetic concentrate using high gradient separators. The modal mineralogy of the 3 samples (feed, concentrate and tailings) from magnetic separation is given in Table 1.

Table 1: Simplified mineral composition of the sample based on MLA.

Sample	Feed	Concentrate	Tailings
Muscovite	4.57	12.91	1.47
Quartz	61.59	6.85	89.85
Topaz	4.70	1.10	5.34
Zinnwaldite	26.07	74.63	0.67
Others	3.07	4.51	2.66

Due to their different crystal structure and mechanical properties, the mica has a different particle shape (plate-like) than quartz and topaz (isometric) after comminution. Thus, the components to be separated differ significantly with respect to these two properties. Furthermore, the minerals considered here also differ in other properties, such as density or conductivity. However, as these properties are not relevant for the separation process under consideration, they will not be taken into account any further.

### 2.2 Magnetic Separation Process

A ring-type magnetic separator was used to separate the crushed ore. This is a dry high gradient magnetic separator that works according to the lifting principle, see Figure 1. The particle mixture is fed onto a conveyor belt by a dosing unit and moved into the separation zone of the magnetic system. Magnetizable particles are lifted out in the separation zone of the magnetic field in the direction of the ring-shaped iron core and adhere to it. The rotation of the ring-shaped iron core moves the extracted particles laterally out of the separation zone of the magnetic field and discharges them into a concentrate container. The non-magnetized or only slightly magnetized particles remain on the conveyor belt and are discharged into a collecting container for the non-magnetic product.

By varying process parameters such as the gap width of the magnet system, the excitation current of the coils, as well as the speed and loading of the conveyor belt, it is possible to influence the separation process (throughput, separation efficiency, recovery). These are typically optimized in practical operation with regard to maximize product quality without knowing the interrelationships of the particle properties and taking these into account.

The ring-type magnetic separator used in this study is a laboratory separator manufactured in the 1980s by SKET (former GDR). The belt width of the separator is equal to 15 cm, whereas the distance from the flat pole to the surface of the conveyor belt is equal to 2 mm. It was operated using the following parameters: gap width 2 mm, magnetic flux density at belt surface 1.2 T and throughput  $200 \text{ g min}^{-1}$ . A size fraction of crushed ore from  $315 \mu\text{m}$  to  $500 \mu\text{m}$  was used as sample material. It was chosen in order to achieve a sufficient voxel number for scanned particles as well as a sufficient number of scanned particles when scanning a MLA sample.

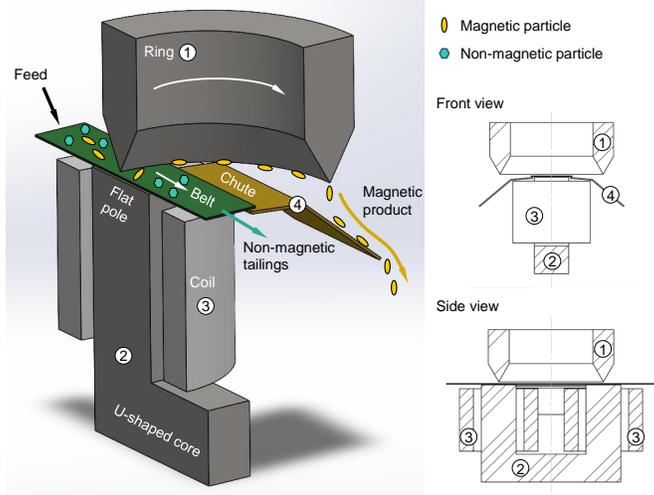


Figure 1: Sketch of the ring-type magnetic separator. 1 - ring-shaped wedge pole; 2 - conveyor belt; 3 - coils of the magnetic system.

### 2.3 Image Acquisition

Since the extraction of target particles is of great interest in the mining industry, the goal of the presented paper is a 3D characterization of particles with respect to their size, shape and fraction of valuable material and the subsequent investigation of the influence of these particle descriptors on the separation behavior. Therefore, we consider data that was previously published in [9] and whose post-processing has already been carried out there. For this reason, the data and post-processing are only briefly summarized in the present paper.

The sample material was prepared such that micro CT-imaging and SEM-EDS analyzes can be performed on the same sample. In order to achieve a better dispersion of the particles, they are blended with micron-sized graphite and then embedded in epoxy blocks. [26]. To prevent segregation effects from affecting the results of the 2D SEM-EDS analysis, the grain mount was cut in the direction of sedimentation, rotated by 90° and then re-embedded [26]. For SEM-EDS analysis a FEI Quanta 650F (Thermo Fisher Scientific, Waltham, MA, USA) SEM, equipped with two Bruker Quantax X-Flash 5030 EDS (Bruker Corporation, Billerica, MA, USA) was used. XCT was done using a Zeiss Xradia 510 Versa X-ray microscope (Carl Zeiss Microscopy GmbH, Jena, Germany). Further details of micro CT-imaging and SEM-EDS analysis are given in [15], measurement parameters can be found in the Appendix A and the related data publication (<https://dx.doi.org/10.25532/OPARA-684>).

### 2.4 Data post-processing

The imaging methods were applied to samples before separation as well as to samples after separation that are assigned to the tailings. To compute particle-wise descriptors, the image data must first be segmented and the results strongly depend on the quality of the segmentation. However, it turned out that conventional methods do not work sufficiently well, since many particles in the CT-image are elongated and plate-like. Thus, a convolutional neural network (CNN) with encoder-decoder architecture based on the 3D U-net of [27] is used. This architecture has the advantage that only 2D cutouts are required for training and thus, reduces the effort of manually labeling the data. Moreover, to obtain better results, the origin architecture of [27] is slightly adjusted, see [9].

To train the neural network, a ground-truth, i.e., labeled data, is required. For that purpose, 2D slices of the CT-image are thresholded and afterwards post-processed to avoid over segmentation. For more details on the labeling we refer to [9]. Then, for network training, the Adam algorithm [28] with step size  $\alpha = 10^{-4}$  is used to minimize the binary cross-entropy loss.

The output of the neural network is an image of the same size as the input with values in  $[0, 1]$ . The foreground-background segmentation is obtained by binarization with threshold 0.5. Then, connected components in the binarized image with more than 50 voxels are specified as particles  $P \subset W$ , where the CT-image is given by  $I : W \rightarrow \mathbb{R}$  for some sampling window  $W \subset \mathbb{Z}^3$ , with  $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$  denoting the set of integers. Moreover, a marker-based

watershed algorithm is used afterwards, to improve the shape recognition of the particles, see [9] for further details. Figure 2 shows a 2D slice extracted from a 3D CT image with the CNN-based particle-wise segmentation and the corresponding SEM-EDS slice, which is used to determine the mineralogical composition of each particle. Note that similar segmentation methods, which combine machine learning with conventional watershed algorithms, have been deployed in further papers within PP 2045, see [9, 14, 16, 18, 21].

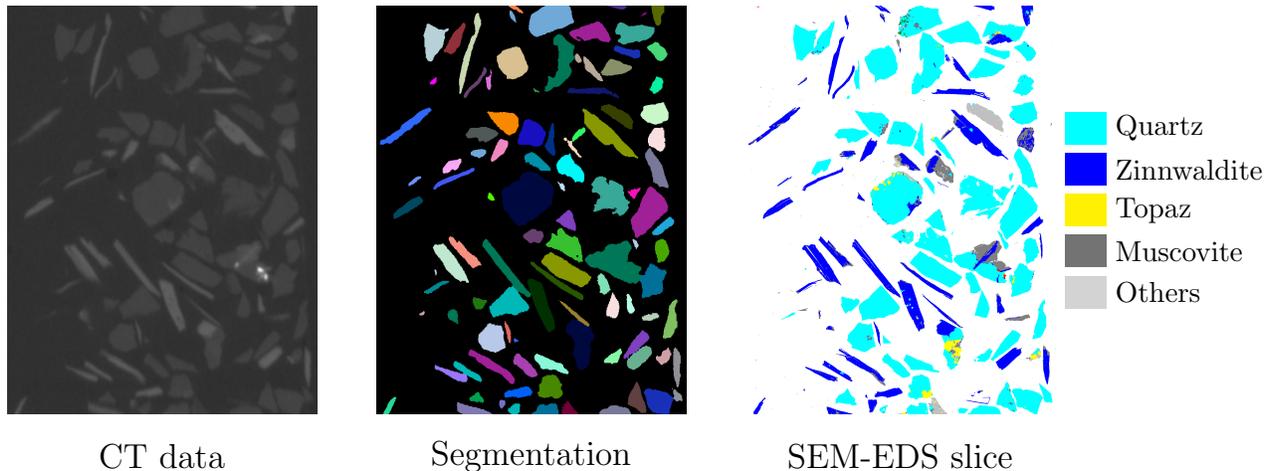


Figure 2: 2D slice of a 3D CT-image with corresponding CNN-based particle-wise segmentation and the corresponding SEM-EDS slice with color legend.

## 2.5 Computation of Particle Descriptors

In order to investigate the separation behavior with respect to shape, size and composition of particles, different particle descriptors are considered. To compute particle descriptors for shape and size, the particle-wise segmentation of CT-image data, as explained in Section 2.4, is used. However, CT-image data does not provide direct information about the volume fraction of zinnwaldite. Therefore, we additionally use segmented SEM-EDS data which is available for three slices of the CT data.

As size descriptor of a particle  $P \subset W$ , the volume equivalent diameter  $M_{\text{vol}}(P)$  is used. It is determined by computing the radius of a sphere with the same volume as the particle, where the volume is defined as the number of voxels belonging to  $P$ . More precisely, the volume equivalent diameter is given by

$$M_{\text{vol}}(P) = \sqrt[3]{\frac{6V(P)}{\pi}}, \quad (1)$$

where  $V(P)$  denotes the volume of  $P$ . Regarding the shape of a particle  $P$ , the flatness  $M_{\text{flat}}(P)$  and the sphericity  $M_{\text{sph}}(P)$  are used. These descriptors are given by

$$M_{\text{flat}}(P) = \frac{a_3(P)}{a_2(P)} \quad \text{and} \quad M_{\text{sph}}(P) = \frac{((36\pi M_{\text{vol}}(P))^2)^{\frac{1}{3}}}{M_{\text{area}}(P)}. \quad (2)$$

Here,  $a_2(P)$  and  $a_3(P)$  denote the length of the second and third longest axis of the minimum-volume bounding box of  $P$  [29]. Furthermore,  $M_{\text{area}}(P)$  denotes the surface area of  $P$ , which is computed by means of the algorithm described in [30]. Note that a more plate-like particle has a smaller value of  $M_{\text{flat}}$  than a more spherical one and a perfect sphere has a sphericity value of 1.

To correlate the 3D morphological and textural 3D characterization of particles with their volume fraction of zinnwaldite, SEM-EDS data for some slices  $W_z \subset W$  with  $W_z = \{(x_1, x_2, x_3) \in W : x_3 = z\}$  and  $z \in \mathbb{Z}$  are required. These slices do not necessarily have to be orthogonal to an axis of the coordinate system, but we assume this for simplicity of notation. The SEM-EDS data within a slice  $W_z \subset W$  for some  $z \in \mathbb{Z}$  is given by a map

$L^{\text{SEM}} : W_z \rightarrow \{0, 1, 2\}$ , where the presence of no particle at voxel  $x \in W_z$  is indicated by  $L^{\text{SEM}}(x) = 0$ . On the other hand,  $L^{\text{SEM}}(x) = 1$  indicates that zinnwaldite is identified at  $x \in W_z$ , whereas  $L^{\text{SEM}}(x) = 2$  means that non-valuable material is detected at  $x \in W_z$ . To determine the volume fraction of zinnwaldite three slices  $W_{z_1}, W_{z_2}, W_{z_3} \subset W$  with the corresponding maps  $L_j^{\text{SEM}} : W_{z_j} \mapsto \{0, 1, 2\}$  for  $j = 1, 2, 3$  are considered. Then, assuming that the observed particle composition within a slice is representative for the entire 3D particle, for each particle  $P \subset W$  with  $P \cap (W_{z_1} \cup W_{z_2} \cup W_{z_3}) \neq \emptyset$  the volume fraction  $M_{\text{rat}}(P)$  of zinnwaldite can be determined by

$$M_{\text{rat}}(P) = \frac{\#\left(\bigcup_{j=1}^3 \{x \in P \cap W_{z_j} : L_j^{\text{SEM}}(x) = 1\}\right)}{\#\left(\bigcup_{j=1}^3 \{x \in P \cap W_{z_j} : L_j^{\text{SEM}}(x) > 0\}\right)}, \quad (3)$$

where  $\#$  denotes cardinality [9].

## 2.6 Multivariate Probabilistic Modeling of Particle Descriptor Vectors

Suppose that for some  $n > 1$  we extracted the particles  $P_1, \dots, P_n \subset W$  from CT-image data, together with the corresponding SEM-EDS information, as explained in Section 2.4. For each of these particles, we compute the vector  $(M_{\text{vol}}(P_i), M_{\text{flat}}(P_i), M_{\text{sph}}(P_i), M_{\text{rat}}(P_i))$  of particle descriptors stated in Section 2.5, where  $i = 1, \dots, n$ .

Then, to determine various multivariate Tromp functions, we probabilistically model three different datasets of (sub-) vectors. Namely,

$$x_{\text{vr}}^{(i)} = (M_{\text{vol}}(P_i), M_{\text{rat}}(P_i)), \quad x_{\text{vsr}}^{(i)} = (M_{\text{vol}}(P_i), M_{\text{sph}}(P_i), M_{\text{rat}}(P_i)), \quad x_{\text{vfr}}^{(i)} = (M_{\text{vol}}(P_i), M_{\text{flat}}(P_i), M_{\text{rat}}(P_i)) \quad (4)$$

for  $i = 1, \dots, n$ . In the following, instead of writing  $x_{\text{vr}}^{(i)}, x_{\text{vsr}}^{(i)}$  and  $x_{\text{vfr}}^{(i)}$ , respectively, we refer to these sub-vectors just as  $x^{(i)}$  in order to simplify the notation. Furthermore, for each of the three datasets  $\{x^{(i)}, i = 1, \dots, n\}$  we consider two different cases, one dataset for particles before separation and one for particles in the tailings after separation, denoting them by  $D_{\text{f}} = \{x_{\text{f}}^{(i)}, i = 1, \dots, n_{\text{f}}\}$  and  $D_{\text{t}} = \{x_{\text{t}}^{(l)}, l = 1, \dots, n_{\text{t}}\}$ , respectively. Here,  $n_{\text{f}}$  is the number of particles extracted from tomographic image data for the feed, and  $n_{\text{t}}$  is the corresponding number of particles associated with tailings.

For modeling purposes, we interpret the particle descriptor vectors stated in Eq. (4) as realizations of random vectors, which will be denoted by  $X = (X_1, \dots, X_d)$ , where  $d = 2, 3$ . Note that the distribution of such random vectors is uniquely determined by its cumulative distribution function  $F_{1, \dots, d} : \mathbb{R}^d \rightarrow [0, 1]$  with  $F_{1, \dots, d}(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$  for each  $(x_1, \dots, x_d) \in \mathbb{R}^d$ , and the corresponding probability density  $f_{1, \dots, d} : \mathbb{R}^d \rightarrow [0, \infty)$ . Since the segmentation of the concentrate data is difficult due the elongated and plate-like particles, we use an optimization approach to estimate the density of the concentrate, see Section 2.6.5. Therefore, parametric Tromp functions are required. Thus, we parametrically model the density  $f_{1, \dots, d}$ , with a copula-based approach [31]. Recently, in [8, 16, 19, 20], we have deployed Archimedean copulas for modeling bivariate probability densities (i.e.,  $d = 2$ ). Furthermore, in [9] we showed that so-called R-vine copulas are a particularly suitable modeling tool if  $d > 2$ . To make the paper more self-contained, we provide some mathematical background for this approach in the following.

### 2.6.1 Sklar's Representation Formula

A function  $C : [0, 1]^d \rightarrow [0, 1]$  with  $d \geq 2$  is called a copula if it is the cumulative distribution function of a  $d$ -dimensional random vector  $(U_1, \dots, U_d)$  with standard uniform marginal distributions, i.e., for each  $j = 1, \dots, d$  it holds that  $C(x) = \mathbb{P}(U_j \leq x_j) = x_j$  if  $x = (1, \dots, 1, x_j, 1, \dots, 1) \in [0, 1]^d$ . An important tool for modeling the joint (multivariate) distribution of an arbitrary random vector  $X = (X_1, \dots, X_d)$  with correlated non-Gaussian components is Sklar's representation formula of its cumulative distribution function  $F_{1, \dots, d} : \mathbb{R}^d \rightarrow [0, 1]$  based on some copula  $C : [0, 1]^d \rightarrow [0, 1]$  and the (marginal) distribution functions  $F_1, \dots, F_d : \mathbb{R} \rightarrow [0, 1]$ , see e.g. [31]. Moreover, if the functions  $C$  and  $F_{1, \dots, d}$  are differentiable with densities  $c : [0, 1]^d \rightarrow [0, \infty)$  and  $f_{1, \dots, d} : \mathbb{R}^d \rightarrow [0, \infty)$ , respectively, Sklar's representation formula implies that

$$f_{1, \dots, d}(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{i=1}^d f_i(x_i), \quad (5)$$

for each  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ , where  $f_i : \mathbb{R} \rightarrow [0, \infty)$  is the density of  $F_i$  for each  $i \in \{1, \dots, d\}$ .

In the following, we use Eq. (5) in order to fit a parametric multivariate probability density  $f_{1, \dots, d}$  to data for particle descriptor vectors, i.e., to realizations of a random vector  $X = (X_1, \dots, X_d)$ , where we first fit univariate densities  $f_1, \dots, f_d$  to realizations of the individual components  $X_1, \dots, X_d$ . Then, in a second step, a  $d$ -dimensional copula density  $c : [0, 1]^d \rightarrow [0, \infty)$  is searched to model the dependencies between the random variables  $X_1, \dots, X_d$  as good as possible. The construction of multivariate copula densities by bivariate ones and their sequential estimation is explained in [9].

## 2.6.2 Adaptation of the Copula-Based Modeling Approach for Descriptor Vectors of Composite Particles

The copula-based modeling approach stated in Section 2.6.1, cannot directly be used for descriptor vectors of composite particles. To take into account the mineralogical composition of the particles, the descriptor  $M_{\text{rat}}$  is considered. But, the issue is that we observe a non-neglectable number of particles that are "pure", i.e., with  $M_{\text{rat}}$ -values being equal to 0 or 1. This indicates that the distribution of  $M_{\text{rat}}$  has atoms at 0 and 1, which means that it cannot be represented by a probability density as this has been done in Eq. (5). Therefore, we split the data in three disjoint sets, similarly to the procedure considered in [9]. More precisely, for the feed and tailings, i.e., for  $j \in \{f, t\}$ , the dataset  $D_j$  is further divided into three subsets  $D_v^j, D_{\text{nv}}^j$  and  $D_{\text{co}}^j$ , where

$$\begin{aligned} D_v^j &= \{x_{1, \dots, d-1}^{(i)} \in \mathbb{R}^{d-1} : i = 1, \dots, n_j, \text{ where } (x_{1, \dots, d-1}^{(i)}, x_d^{(i)}) \in D_j \text{ with } x_d^{(i)} \geq 1 - p\}, \\ D_{\text{nv}}^j &= \{x_{1, \dots, d-1}^{(i)} \in \mathbb{R}^{d-1} : i = 1, \dots, n_j, \text{ where } (x_{1, \dots, d-1}^{(i)}, x_d^{(i)}) \in D_j \text{ with } x_d^{(i)} \leq p\}, \\ D_{\text{co}}^j &= \{x_{1, \dots, d}^{(i)} \in D_j : i = 1, \dots, n_j, \text{ where } p < x_d^{(i)} < 1 - p\}, \end{aligned}$$

for some threshold  $p \in (0, 1)$ , which is used to subdivide the data based on the volume fraction of zinnwaldite. In the following, the cardinalities of the datasets  $D_v^f, D_{\text{nv}}^f$  and  $D_{\text{co}}^f$ , will be denoted by  $n_v^f, n_{\text{nv}}^f$  and  $n_{\text{co}}^f$ , respectively, whereas the cardinalities of  $D_v^t, D_{\text{nv}}^t$  and  $D_{\text{co}}^t$  will be denoted by  $n_v^t, n_{\text{nv}}^t$  and  $n_{\text{co}}^t$ .

Recall that  $x_d^{(i)}$  equals  $M_{\text{rat}}(P_i)$  for all particle descriptor vectors considered in this paper. Thus, putting  $p = 0.01$ , the subsets  $D_v^j, D_{\text{nv}}^j \subset \mathbb{R}^{d-1}$  contain the particles with almost exclusively valuable material, i.e., particles with a zinnwaldite fraction of at least  $1 - p = 0.99$ , and almost exclusively non-valuable material, i.e., particles with a zinnwaldite fraction of at most  $p$ , respectively. The subset  $D_{\text{co}}^j \subset \mathbb{R}^d$  contains the composite particles, i.e., particles with significant volume fractions of both valuable and non-valuable material.

Now, using the algorithm described in [9], we can fit  $(d-1)$ -variate probability densities  $f_v^f, f_{\text{nv}}^f, f_{\text{n}}^t : \mathbb{R}^{d-1} \rightarrow [0, \infty)$  to the datasets  $D_v^f, D_{\text{nv}}^f$  and  $D_{\text{nv}}^t$ , respectively. Similarly,  $d$ -variate densities  $f_{\text{co}}^f, f_{\text{co}}^t : \mathbb{R}^d \rightarrow [0, \infty)$  can be fitted to the datasets  $D_{\text{co}}^f$  and  $D_{\text{co}}^t$ . Note that the  $d$ -variate densities  $f_{\text{co}}^f, f_{\text{co}}^t$  are fitted such that the marginal densities  $f_{\text{co}, d}^f, f_{\text{co}, d}^t : \mathbb{R} \rightarrow [0, \infty)$  vanish outside of the interval  $[p, 1 - p]$ , which can be achieved by means of a truncated mixed beta distribution, as shown in [9].

To describe the multivariate probability density  $f^f : \mathbb{R}^d \rightarrow [0, \infty)$  of the  $d$ -dimensional particle descriptor vectors considered in Eq. (4) for the entire dataset  $D_f$ , we assume that the random variable  $M_{\text{rat}}$  is conditionally independent of the remaining particle descriptors, given that  $M_{\text{rat}} \in [0, p] \cup [1 - p, 1]$ . Moreover, we assume that  $M_{\text{rat}}$  is uniformly distributed on  $[0, p] \cup [1 - p, 1]$ . Then, the density  $f^f : \mathbb{R}^d \rightarrow [0, \infty)$  is given by

$$f^f(x) = \begin{cases} \frac{n_{\text{nv}}^f}{n_f} \frac{1}{0.01} f_{\text{nv}}^f(x_1, \dots, x_{d-1}), & \text{if } 0 \leq x_d \leq p, \\ \frac{n_{\text{co}}^f}{n_f} f_{\text{co}}^f(x), & \text{if } p < x_d \leq 1 - p, \\ \frac{n_v^f}{n_f} \frac{1}{0.01} f_v^f(x_1, \dots, x_{d-1}), & \text{if } 1 - p < x_d \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where  $x = (x_1, \dots, x_{d-1}, x_d) \in \mathbb{R}^d$ , see [9] for details. Similarly to Eq. (6), the density  $f^t : \mathbb{R}^d \rightarrow [0, \infty)$  of the

$d$ -dimensional particle descriptor vectors considered in Eq. (4) for the dataset  $D_t$  is given by

$$f^t(x) = \begin{cases} \frac{n_{nv}^t}{n_{nv}^t + n_{co}^t} \frac{1}{0.01} f_{nv}^t(x_1, \dots, x_{d-1}), & \text{if } 0 \leq x_d \leq p, \\ \frac{n_{nv}^t}{n_{nv}^t + n_{co}^t} f_{co}^t(x), & \text{if } p < x_d \leq 1 - p, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

Notice that the formulas given in Eqs. (6) and (7) are slightly different from each other. The reason for this is that, in our case, the set  $D_v^t$  contains only two particle descriptor vectors and therefore it is impossible to determine a density  $f_v^t : \mathbb{R}^{(d-1)} \rightarrow [0, \infty)$ , which would be necessary if we proceeded in the same way as in Eq. (6). As already mentioned above, the densities  $f_v^f, f_{co}^f, f_n^f, f_{co}^t, f_n^t$  will be constructed via the copula-based approach stated in Eq. (5). Specifically, bivariate Archimedean copulas will be used, with the univariate distributions of individual particle descriptors fitted by gamma or beta distributions, see Section 3.1 for further details.

Since, in our case, the concentrate almost exclusively consists of elongated and plate-like particles, it is difficult to obtain a sufficiently accurate segmentation of the particles. As this type of particles only occurs in small quantities in the tailings, it significantly reduces the effort to reliably segment the CT data of the tailings. That is why, instead of fitting the density function  $f^c : \mathbb{R}^d \rightarrow [0, \infty)$  for concentrate directly to CT data, in Section 2.6.5 we introduce an optimization routine to determine the density  $f^c$  for the concentrate. However, before doing so, we must first introduce a representation formula for the mass-weighted version  $f_m^f$  of  $f^f$ .

### 2.6.3 Mass-Weighted Probability Densities of Particle Descriptor Vectors

In addition to the number-weighted densities  $f^f, f^c, f^t$  considered so far, we consider the mass-weighted probability densities  $f_m^f, f_m^c, f_m^t : \mathbb{R}^d \rightarrow [0, \infty)$  for feed, concentrate and tailings, which are given by

$$f_m^f(x) = \frac{f^f(x)m(x)}{\int_{\mathbb{R}^d} f^f(y)m(y)dy}, \quad f_m^c(x) = \frac{f^c(x)m(x)}{\int_{\mathbb{R}^d} f^c(y)m(y)dy}, \quad f_m^t(x) = \frac{f^t(x)m(x)}{\int_{\mathbb{R}^d} f^t(y)m(y)dy} \quad (8)$$

for each  $x \in \mathbb{R}^d$ . The function  $m : \mathbb{R}^d \rightarrow [0, \infty)$  in Eq. (8) maps each particle descriptor vector  $x \in \mathbb{R}^d$  to the corresponding material density  $m(x)$ , see [11]. Since the particles consist mainly of quartz and zinnwaldite, we assume for the sake of simplicity that all particles are composed of these two components only. Consequently, the mass function  $m$  can be expressed as follows:

$$m(x) = \frac{4\pi}{3} \left(\frac{x_1}{2}\right)^3 (x_d \cdot \delta_Z + (1 - x_d) \cdot \delta_Q) \quad (9)$$

for each particle descriptor vector  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ , which contains the volume equivalent diameter  $x_1$  and the volume fraction of zinnwaldite  $x_d$ , where  $\delta_Z = 2.96$  and  $\delta_Q = 2.65$  are the mass densities of zinnwaldite and quartz, respectively.

### 2.6.4 Representation Formula for the Probability Density $f_m^f$ of Particle Descriptor Vectors of Feed

In separation processes, the sum of the numbers of particles in concentrate and tailings should theoretically be equal to the number of particles in the feed. In practice, however, the measurements of feed, concentrate and tailings are statistically representative samples for the corresponding particle systems and thus, can violate this equality. Nevertheless, the mass-weighted probability density  $f_m^f : \mathbb{R}^d \rightarrow [0, \infty)$  of particle descriptor vectors of feed can be represented as a convex combination of the corresponding densities  $f_m^t : \mathbb{R}^d \rightarrow [0, \infty)$  and  $f_m^c : \mathbb{R}^d \rightarrow [0, \infty)$  of tailings and concentrate, i.e.,

$$f_m^f(x) = \lambda f_m^c(x) + (1 - \lambda) f_m^t(x), \quad (10)$$

for each  $x \in \mathbb{R}^d$  and some mixing parameter  $\lambda \in [0, 1]$  given by

$$\lambda = \frac{m_c}{m_f},$$

where  $m_c$  and  $m_f$  denote the total mass of particles in concentrate and feed, see [1].

### 2.6.5 Determining the Probability Density $f^c$ of Particle Descriptor Vectors of Concentrate

As already mentioned in Section 2.6.2, the segmentation of particles in CT image data is difficult for the concentrate. Therefore, we aim to indirectly determine the probability density  $f^c$  of particle descriptor vectors of concentrate using the representation formula given in Eq. (10). For this, we assume that  $f^c$  belongs to a parametric family  $\{f_\theta : \theta \in \Theta\}$  of  $d$ -dimensional probability densities, where  $\Theta \subset \mathbb{R}^q$  denotes some set of  $q$ -dimensional parameter vectors for some integer  $q \geq 1$ . Then, similar to Eq. (6), we can write  $f^c$  as follows:

$$f^c(x) = \begin{cases} \theta_1 \frac{1}{0.01} f_{\text{nv}}^{(\theta_4)}(x_{1,\dots,d-1}), & \text{if } 0 \leq x_d \leq p, \\ \theta_2 f_{\text{co}}^{(\theta_5)}(x), & \text{if } p < x_d \leq 1-p, \\ \theta_3 \frac{1}{0.01} f_{\text{v}}^{(\theta_6)}(x_{1,\dots,d-1}), & \text{if } 1-p < x_d \leq 1, \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

for each  $x = (x_{1,\dots,d-1}, x_d)$ , where  $\theta_1, \theta_2, \theta_3 \in [0, 1]$  with  $\theta_1 + \theta_2 + \theta_3 = 1$ . Additionally, let  $q_1, q_2, q_3 \geq 1$  with  $q_1 + q_2 + q_3 + 3 = q$  and  $\theta_4 \in \mathbb{R}^{q_1}, \theta_5 \in \mathbb{R}^{q_2}, \theta_6 \in \mathbb{R}^{q_3}$  be the parameter vectors of the probability densities  $f_{\text{nv}}^{(\theta_4)}, f_{\text{co}}^{(\theta_5)}, f_{\text{v}}^{(\theta_6)}$  appearing on the right-hand side of Eq. (11). Therefore, the parameter vector  $\theta \in \Theta$  that we aim to optimize is given by  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$ . Notice that the probability density  $f^c$  is not yet mass-weighted, so this weighting must be carried out during the optimization process. Thus, the optimization problem is given by

$$\theta_c = \arg \min_{\theta \in \Theta} \int_{\mathbb{R}^d} \left| f_{\text{m}}^f(x) - \left( \lambda \frac{f_\theta(x)m(x)}{\int_{\mathbb{R}^d} f_\theta(y)m(y)dy} + (1-\lambda)f_{\text{m}}^t(x) \right) \right| dx, \quad (12)$$

where  $\theta_c$  is the optimal parameter vector for  $f^c$ , see also [11]. According to the rule given in Eq. (8), the resulting density  $f^c$  can then be transformed into the mass-weighted density  $f_{\text{m}}^c$ , which is used to determine Tromp functions as explained in the next section.

## 2.7 Multivariate Tromp Functions and Their Interpretation

To investigate the separation behavior of particles during separation processes, in [11, 22–24] we utilized copula-based probability densities of the considered particle descriptor vectors to compute multivariate Tromp functions  $T : \mathbb{R}^d \rightarrow [0, 1]$ , where  $T(x)$  can be interpreted as separation probability of a particle with descriptor vector  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ . In other words, a feed particle with descriptor vector  $x$  will be separated into the concentrate with probability  $T(x)$ . Formally, the Tromp function  $T : \mathbb{R}^d \rightarrow [0, \infty)$  of a separation process is given by the mass-weighted densities  $f_{\text{m}}^f : \mathbb{R}^d \rightarrow [0, \infty)$  and  $f_{\text{m}}^c : \mathbb{R}^d \rightarrow [0, \infty)$  as follows:

$$T(x) = \begin{cases} \frac{m_c}{m_f} \frac{f_{\text{m}}^c(x)}{f_{\text{m}}^f(x)}, & \text{if } f_{\text{m}}^f(x) > 0 \\ 0, & \text{if } f_{\text{m}}^f(x) = 0, \end{cases} \quad (13)$$

for each  $x \in \mathbb{R}^d$ , see [11]. However, to interpret the value  $T(x)$  as separation probability, it must hold that  $T(x)$  only takes values within the interval  $[0, 1]$ , so that particle descriptor vectors with a value of  $T(x)$  close to 1 have a high probability of being separated into the concentrate, and those with a value of  $T(x)$  close to 0 have a low probability of being separated to the concentrate. Due to numerical instabilities arising from the use of quotients, this can not always be guaranteed. In particular, the Tromp function is sensitive regarding the denominator in Eq. (13), when the feed contains only a small number of particles with a particular descriptor vector, i.e.,  $f_{\text{m}}^f(x)$  is close to zero, but such particles are enriched in the concentrate. For this reason, in [11], the reconstruction of the feed density in Eq. (10) has been used to rewrite Eq. (13) as follows:

$$T(x) = \begin{cases} \frac{m_c}{m_f} \frac{f_{\text{m}}^c(x)}{\lambda f_{\text{m}}^c(x) + (1-\lambda)f_{\text{m}}^t(x)}, & \text{if } \lambda f_{\text{m}}^c(x) + (1-\lambda)f_{\text{m}}^t(x) > 0 \\ 0, & \text{if } \lambda f_{\text{m}}^c(x) + (1-\lambda)f_{\text{m}}^t(x) = 0, \end{cases} \quad (14)$$

for each  $x \in \mathbb{R}^d$ , where  $\lambda = \frac{m_c}{m_f}$ . Thus, the multivariate densities of the concentrate and the tailing are used to compute a numerically stable Tromp function.

Formally, the values  $T(x)$  of a Tromp function can be defined for all  $x \in \mathbb{R}^d$ , but for descriptor vectors  $x \in \mathbb{R}^d$  such that  $f_m^f(x)$  is close to zero, this is not meaningful. Thus, in [11] the set

$$A = \{x \in \mathbb{R}^d : f_m^f(x) > \varepsilon\} \quad (15)$$

has been introduced, where  $\varepsilon = \inf\{s \in [0, \infty) : \int_{x \in \mathbb{R}^d : f_m^f(x) \leq s} f_m^f(x) dx \geq q\}$  for some  $q \in [0, 1]$ . The threshold  $q$  is used to indicate how likely it must be that particles with certain descriptor vectors are observed in the feed to provide sufficient information about the separation probability of such particles. Thus, the value  $T(x)$  of a Tromp function is meaningful only if  $x \in A$ .

### 3 Results

As already mentioned above, the separation behavior of particles in the magnetic separation process considered in this paper will be analyzed with respect to shape, size and composition of the particles. Thus, we analyze the separation process by means of the three particle descriptor vectors given in Eq. (4). In this section, the corresponding densities of the particle descriptor vectors are fitted and then used to compute Tromp functions.

#### 3.1 Fitted Univariate and Multivariate Probability Densities

To compute Tromp functions according to Eq. (14), we need to know the multivariate probability densities  $f_m^c$  and  $f_m^t$ . But, since the segmentation of particles in CT image data is difficult for the concentrate, we use the optimization approach stated in Section 2.6.5 in order to determine  $f_m^c$ . To achieve this, we first apply the algorithm proposed in [9] to fit the multivariate probability densities  $f_v^f, f_n^f, f_{co}^f, f_n^t$  and  $f_{co}^t$  appearing on the right-hand sides of Eqs. (6) and (7) to image data of feed and tailings for the three particle descriptor vectors considered in Eq. (4), where the marginal densities are fitted by mixed gamma and beta distributions for the shape and size descriptors, respectively, as they provide a suitable support, i.e., a support of  $[0, 1]$  for shape and composition descriptors and  $[0, \infty)$  for the size. Moreover, they are multimodal, which is required since the considered particles are often composites. Note that, formally, the probability density  $f_{\text{mixed}} : \mathbb{R} \rightarrow [0, \infty)$  of a mixture with two component densities  $f_1, f_2 : \mathbb{R} \rightarrow [0, \infty)$  is given by  $f_{\text{mixed}}(x) = wf_1(x) + (1 - w)f_2(x)$  for each  $x \in \mathbb{R}$ , where  $w \in [0, 1]$  denotes the mixing ratio [32].

##### 3.1.1 Univariate Probability Densities of Single Particle Descriptors for Feed and Tailings

To make the paper more self-contained, the formulas for the densities of beta and gamma distributions, which will be used in the following, are given in Table 2.

Table 2: Parametric families of univariate distributions with corresponding density, support and parameter space, where  $B, \Gamma$  denote the beta [33] and gamma function [34], respectively.

parametric family	probability density	support	parameters
beta distribution [33]	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$	$[0, 1]$	$\alpha, \beta > 0$
gamma distribution [34]	$\frac{x^{\alpha-1} \exp(-\frac{x}{\beta})}{\beta^\alpha \Gamma(\alpha)}$	$[0, \infty)$	$\alpha, \beta > 0$

To fit mixed gamma distributions, the expectation maximization algorithm [32] is used, whereas mixed beta distributions are fitted by means of the algorithm introduced in [35]. To ensure that the distribution corresponding to  $M_{\text{rat}}$  vanishes outside the interval  $[0.01, 0.99]$ , we used a truncated mixed beta distribution. The type and parameter values of the fitted univariate distributions are given in Table 3, see also Figures 3 and 4.

Table 3: Parametric families of fitted univariate distributions for the data sets  $D_v^f, D_{co}^f, D_{nv}^f, D_{co}^t, D_{nv}^t$  and the particle descriptors  $M_{vol}, M_{flat}, M_{sphe}, M_{rat}$ , together with their parameter values (trunc = truncated).

data set	descriptor	distribution	fitted parameter values
$D_v^f$	$M_{vol}$	mixed gamma	$w = 0.8989, \alpha_1 = 15.8647, \beta_1 = 12.4278, \alpha_2 = 6.4938, \beta_2 = 13.8301$
	$M_{flat}$	mixed beta	$w = 0.9265, \alpha_1 = 6.1468, \beta_1 = 36.1857, \alpha_2 = 310.4800, \beta_2 = 602.7691$
	$M_{sphe}$	beta	$\alpha_1 = 7.9504, \beta_1 = 10.4121$
$D_{co}^f$	$M_{vol}$	mixed gamma	$w = 0.81028, \alpha_1 = 15.9421, \beta_1 = 13.1879, \alpha_2 = 4.3932, \beta_2 = 58.3942$
	$M_{flat}$	mixed beta	$w = 0.6041, \alpha_1 = 5.4094, \beta_1 = 25.9257, \alpha_2 = 2.4341, \beta_2 = 2.6553$
	$M_{sphe}$	mixed gamma	$w = 0.3328, \alpha_1 = 20.3022, \beta_1 = 31.0702, \alpha_2 = 7.6441, \beta_2 = 5.7862$
	$M_{rat}$	trunc. mixed beta	$w = 0.2684, \alpha_1 = 0.3648, \beta_1 = 1.7342, \alpha_2 = 6.7577, \beta_2 = 0.7318$
$D_{nv}^f$	$M_{vol}$	mixed gamma	$w = 0.0539, \alpha_1 = 4.7095, \beta_1 = 14.7099, \alpha_2 = 11.7386, \beta_2 = 21.5318$
	$M_{flat}$	mixed beta	$w = 0.1335, \alpha_1 = 9.5198, \beta_1 = 11.2623, \alpha_2 = 3.131, \beta_2 = 2.1409$
	$M_{sphe}$	mixed beta	$w = 0.1097, \alpha_1 = 9.8598, \beta_1 = 9.4948, \alpha_2 = 19.6430, \beta_2 = 7.8057$
$D_{co}^t$	$M_{vol}$	mixed gamma	$w = 0.9599, \alpha_1 = 13.7417, \beta_1 = 19.3901, \alpha_2 = 29.1067, \beta_2 = 3.1535$
	$M_{flat}$	mixed beta	$w = 0.2916, \alpha_1 = 12.3731, \beta_1 = 11.5920, \alpha_2 = 12.6004, \beta_2 = 4.2472$
	$M_{sphe}$	gamma	$\alpha_1 = 15.01349, \beta_1 = 5.4848$
	$M_{rat}$	trunc. mixed beta	$w = 0.6736, \alpha_1 = 1.0630, \beta_1 = 16.4735, \alpha_2 = 4.0676, \beta_2 = 0.7839$
$D_{nv}^t$	$M_{vol}$	mixed gamma	$w = 0.9887, \alpha_1 = 11.3663, \beta_1 = 24.5910, \alpha_2 = 1.1749, \beta_2 = 78.2910$
	$M_{flat}$	mixed beta	$w = 0.4250, \alpha_1 = 8.2138, \beta_1 = 7.2888, \alpha_2 = 3.5589, \beta_2 = 1.9493$
	$M_{sphe}$	beta	$\alpha_1 = 20.0204, \beta_1 = 8.0359$

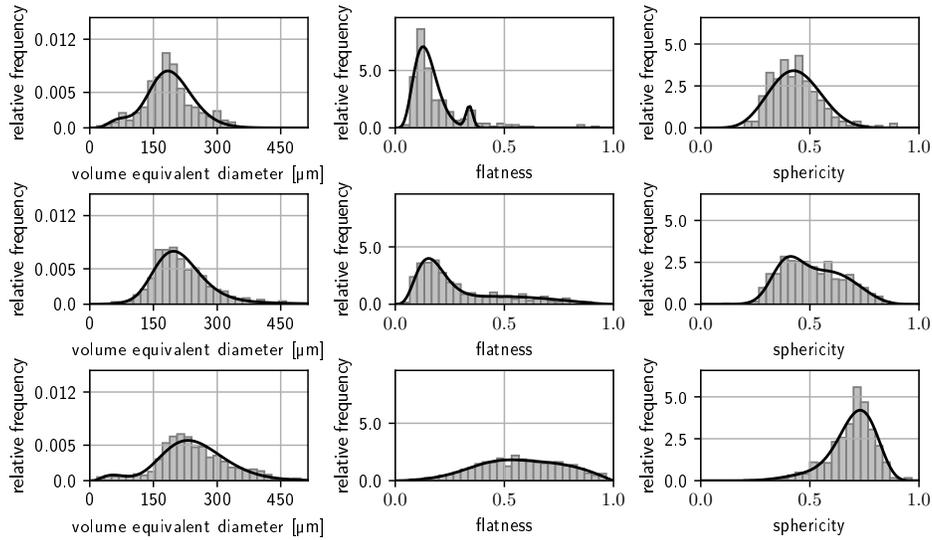


Figure 3: Histograms and fitted (univariate) densities of volume equivalent diameter, flatness and sphericity of particles in the feed, which consist almost exclusively of zinnwaldite (upper row), are a composition with significant fraction of both zinnwaldite and non-valuable material (middle row), and consist almost exclusively of non-valuable material (lower row).

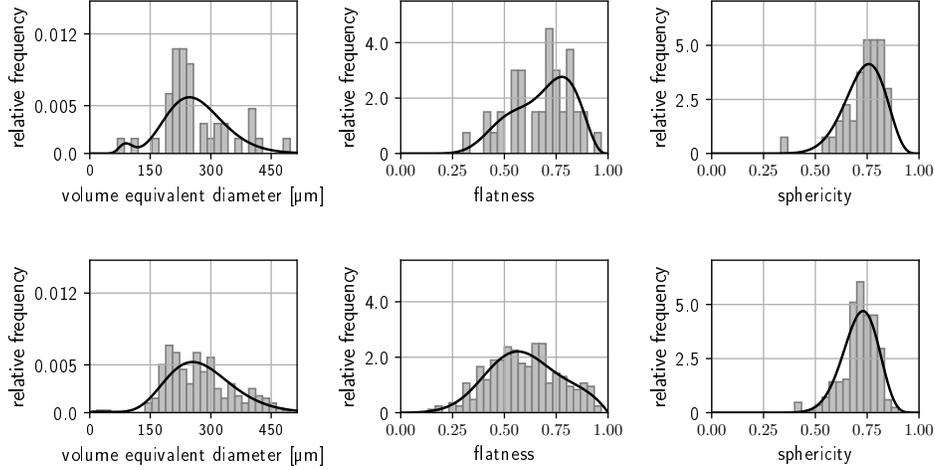


Figure 4: Histograms and fitted (univariate) densities of volume, equivalent diameter, flatness and sphericity of particles in the tailings which are a composition with significant fraction of both zinnwaldite and non-valuable material (upper row) and consist almost exclusively of non-valuable material (lower row).

### 3.1.2 Multivariate Probability Densities of Particle Descriptor Vectors for Feed and Tailings

Using the univariate distributions stated above, we determine the multivariate densities  $f_v^f, f_n^f, f_{co}^f, f_n^t$  and  $f_{co}^t$  appearing on the right-hand sides of Eqs. (6) and (7) by means of the algorithm given in [9]. For this, the independence copula and the following (bivariate) Archimedean copulas are taken into account as candidates. Namely, we consider the Frank, Joe, Clayton and Gumbel copula given in Table 4 as well as their rotations by 90, 180 and 270 degrees, see [36].

Table 4: Parametric families  $\{\phi_\theta : \theta \in \Theta\}$  of Archimedean generators, together with their set of parameters  $\Theta \subset \mathbb{R}$ .

copula	Frank	Joe	Clayton	Gumbel
$\phi_\theta(u)$	$-\ln \frac{\exp(-\theta u)-1}{\exp(-\theta)-1}$	$-\ln(1-(1-u)^\theta)$	$\frac{1}{\theta}(u^{-\theta}-1)$	$(-\ln u)^\theta$
$\Theta$	$\mathbb{R} \setminus \{0\}$	$[1, \infty)$	$(0, \infty)$	$[1, \infty)$

The type and parameter values of the fitted copulas for the particle descriptor vectors of Eq. (4) are given in Table 5. These copulas are then used to determine the multivariate densities  $f^f$  and  $f^t$  by means of Eqs. (6) and (7), which are transformed according to Eq. (8) to get the mass-weighted densities  $f_m^f$  and  $f_m^t$ .

### 3.1.3 Probability Densities of Particle Descriptors for Concentrate

In the next step, for each  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ , the value of  $f_m^f(x)$  is expressed by the right-hand side of Eq. (10), where the mixing parameter  $\lambda = 0.308$  is experimentally measured. Since we do not have full information about the particles in the concentrate, the minimization problem formulated in Eq. (12) is used to determine the mass-weighted density  $f_m^c$  for the three particle descriptor vectors given in Eq. (4), where the densities  $f_m^f$  and  $f_m^t$  fitted to segmented image data are inserted into Eq. (12).

However, due to the large number of parameters in the minimization problem stated in Eq. (12), it turns out that we have to divide it into several (smaller) minimization problems. Thus, we first fit the parameters of the univariate distributions of single particle descriptors for three subsets of particles in the concentrate, which consist of almost exclusively valuable material, almost exclusively non-valuable material, and a significant fraction of both, respectively. As candidates for families of univariate distributions, the same parametric families as for the feed are considered. Then, for all  $i \in \{1, \dots, d\}$  and  $k \in \{v, co, n\}$ , the parameter vector  $\theta^{(i,k)} \in \Theta^{(i,k)}$  of the univariate density

Table 5: Types and parameter values of copulas for the fitted multivariate densities of the particle descriptor vectors  $x_{\text{vr}}, x_{\text{vfr}}, x_{\text{vsr}}$  and the data sets  $D_{\text{v}}^{\text{f}}, D_{\text{co}}^{\text{f}}, D_{\text{nv}}^{\text{f}}, D_{\text{co}}^{\text{t}}, D_{\text{nv}}^{\text{t}}$ .

data set	descriptor vector	copula type	fitted copula parameter
$x_{\text{vr}}$			
$D_{\text{co}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{rat}})$	Clayton 90°	0.1855
$D_{\text{co}}^{\text{t}}$	$(M_{\text{vol}}, M_{\text{rat}})$	Gumbel 90°	1
$x_{\text{vfr}}$			
$D_{\text{v}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{flat}})$	independent	
$D_{\text{co}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{flat}})$	Clayton 180°	0.3779
	$(M_{\text{flat}}, M_{\text{rat}})$	Clayton 90°	0.5627
$D_{\text{nv}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{flat}})$	independent	
	$(M_{\text{vol}}, M_{\text{flat}})$	Frank	1.7726
$D_{\text{co}}^{\text{t}}$	$(M_{\text{flat}}, M_{\text{vol}})$	Frank	1.2837
	$(M_{\text{vol}}, M_{\text{rat}})$	Gumbel 90°	1
$D_{\text{nv}}^{\text{t}}$	$(M_{\text{flat}}, M_{\text{rat}} M_{\text{vol}})$	independent	
	$(M_{\text{vol}}, M_{\text{flat}})$	Gumbel 180°	1.2181
$x_{\text{vsr}}$			
$D_{\text{v}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{sphe}})$	Clayton 270°	0.2446
$D_{\text{co}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{rat}})$	Clayton 90°	0.1855
	$(M_{\text{sphe}}, M_{\text{rat}})$	Clayton 90°	0.5471
$D_{\text{nv}}^{\text{f}}$	$(M_{\text{vol}}, M_{\text{sphe}} M_{\text{rat}})$	independent	
	$(M_{\text{vol}}, M_{\text{sphe}})$	Joe 180°	1.1680
$D_{\text{co}}^{\text{t}}$	$(M_{\text{vol}}, M_{\text{rat}})$	Gumbel 90°	1
	$(M_{\text{sphe}}, M_{\text{rat}})$	Gumbel 180°	1.0470
$D_{\text{nv}}^{\text{t}}$	$(M_{\text{vol}}, M_{\text{sphe}} M_{\text{rat}})$	independent	
$D_{\text{nv}}^{\text{t}}$	$(M_{\text{vol}}, M_{\text{sphe}})$	independent	

$f_{\theta^{(i,k)}} : \mathbb{R} \rightarrow [0, \infty)$  of the  $i$ -th particle descriptor is fitted by solving a similar optimization problem as in Eq. (12). The type and parameter values of the fitted univariate distributions are given in Table 6.

Then, we optimize the remaining parameters to obtain a probability density as given in Eq. (11), where we assume that the multivariate densities appearing on the right-hand side of Eq. (11) have the same R-vine structure and the same bivariate copula types as the corresponding probability densities  $f_{\text{v}}^{\text{f}}, f_{\text{n}}^{\text{f}}, f_{\text{co}}^{\text{f}}$  and  $f_{\text{n}}^{\text{t}}, f_{\text{co}}^{\text{t}}$ . Moreover, the parameters  $\theta_1, \theta_2, \theta_3$  appearing in Eq. (11) are optimized. The obtained results are given in Table 7.

Table 6: Parametric families of fitted univariate distributions for the particle descriptors  $M_{\text{vol}}, M_{\text{flat}}, M_{\text{sphe}}, M_{\text{rat}}$  of the concentrate, together with their parameter values (trunc = truncated).

material	descriptor	distribution	fitted parameter values
valuable	$M_{\text{vol}}$	mixed gamma	$w = 0.9, \alpha_1 = 15.86, \beta_1 = 12.4120, \alpha_2 = 6.4900, \beta_2 = 13.8244$
	$M_{\text{flat}}$	beta	$\alpha_1 = 6.1400, \beta_1 = 36.1900$
	$M_{\text{sphe}}$	beta	$\alpha_1 = 7.9704, \beta_1 = 10.4121$
composite	$M_{\text{vol}}$	mixed gamma	$w = 0.7879, \alpha_1 = 15.8794, \beta_1 = 12.1410, \alpha_2 = 4.4159, \beta_2 = 58.4404$
	$M_{\text{flat}}$	gamma	$\alpha_1 = 5.1400, \beta_1 = 25.9300$
	$M_{\text{sphe}}$	mixed beta	$w = 0.6126, \alpha_1 = 19.6100, \beta_1 = 29.2378, \alpha_2 = 22.7139, \beta_2 = 17.6017$
	$M_{\text{rat}}$	mixed trunc. beta	$w = 0.1056, \alpha_1 = 3.1270, \beta_1 = 7.6151, \alpha_2 = 5.3638, \beta_2 = 0.5996$
non-valuable	$M_{\text{vol}}$	gamma	$\alpha_1 = 19.1034, \beta_1 = 7.0838$
	$M_{\text{flat}}$	mixed gamma	$w = 0.0027, \alpha_1 = 9.2924, \beta_1 = 11.4265, \alpha_2 = 1.0, \beta_2 = 1.4092$
	$M_{\text{sphe}}$	gamma	$\alpha_1 = 44.7555, \beta_1 = 84.8405$

Table 7: Types and parameter values of copulas of fitted multivariate densities for the particle descriptor vectors  $x_{\text{vr}}, x_{\text{vfr}}, x_{\text{vsr}}$  of the concentrate, together with the values of the parameters  $\theta_1, \theta_2, \theta_3$  appearing in Eq. (11).

material	$\theta_1, \theta_2, \theta_3$	descriptor vector	copula type	fitted copula parameter
$x_{\text{vr}}$ (fraction valuable: $\theta_1 = 0.1656$ )				
composite	$\theta_2 = 0.6299$	$(M_{\text{vol}}, M_{\text{rat}})$	Clayton $90^\circ$	0.0599
$x_{\text{vfr}}$				
valuable	$\theta_1 = 0.0930$	$(M_{\text{vol}}, M_{\text{flat}})$	independent	
composite	$\theta_2 = 0.8847$	$(M_{\text{vol}}, M_{\text{flat}})$	Clayton $180^\circ$	0.2614
		$(M_{\text{flat}}, M_{\text{ratio}})$	Clayton $90^\circ$	0.1359
non-valuable	$\theta_3 = 0.0223$	$(M_{\text{vol}}, M_{\text{ratio}} M_{\text{flat}})$	independent	
		$(M_{\text{vol}}, M_{\text{flat}})$	Frank	1.0923
$x_{\text{vsr}}$				
valuable	$\theta_1 = 0.0817$	$(M_{\text{vol}}, M_{\text{sph}})$	Clayton $270^\circ$	0.2505
composite	$\theta_2 = 0.9183$	$(M_{\text{vol}}, M_{\text{rat}})$	Clayton $90^\circ$	0.0651
		$(M_{\text{sph}}, M_{\text{ratio}})$	Clayton $90^\circ$	0.2396
non-valuable	$\theta_3 = 0$	$(M_{\text{vol}}, M_{\text{sph}})$	Joe $180^\circ$	1.1817

In Figure 5, bivariate probability densities of the particle descriptor vector  $x_{\text{vr}}$  are shown, which have been obtained for the feed and tailings, as well as for the concentrate by solving Eq. (12). Furthermore, in Figure 5d, the density of  $x_{\text{vr}}$  is shown, which has been obtained for the feed via reconstruction, i.e., by computing the convex combination of the densities for concentrate and tailings as given in Eq.(10). It is clearly visible that the densities shown in Figures 5a and 5d nicely coincide, although no segmented image data for the concentrate was available and therefore, the density of  $x_{\text{vr}}$  for the concentrate had to be determined indirectly, by solving the optimization problem stated in Eq. (12). Note also that Figure 5b shows that the tailings consist to a large extent of particles with a low zinnwaldite fraction, i.e., particles with a high quartz fraction.

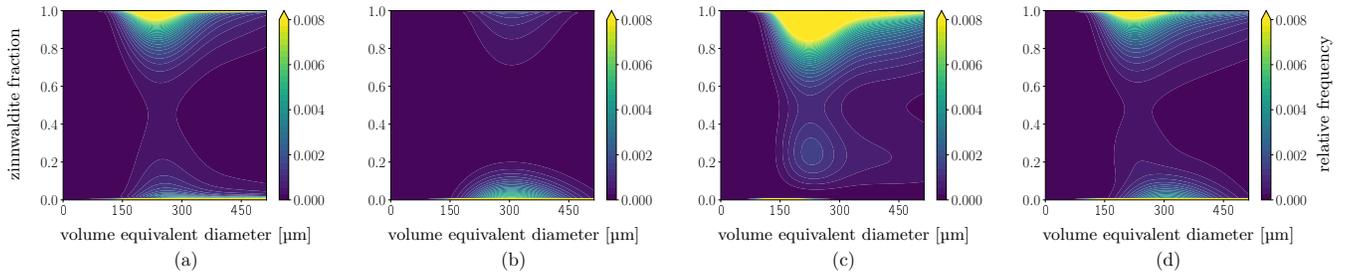


Figure 5: Bivariate probability densities of the particle descriptor vector  $x_{\text{vr}}$  for feed (a) and tailings (b), as well as for the concentrate (c) obtained by solving Eq. (12), and for the reconstructed feed (d) by means of Eq. (10).

To visualize the results, which have been achieved by solving the optimization problem stated in Eq. (12) for the three-dimensional particle descriptor vectors  $x_{\text{vfr}}$  and  $x_{\text{vsr}}$  given in Eq. (4), bivariate (marginal) probability densities are considered. In particular, in Figure 6, bivariate marginal densities are shown, which have been obtained for the volume equivalent diameter and flatness of particles in the feed, by (partially) integrating the corresponding trivariate densities of  $x_{\text{vfr}}$ .

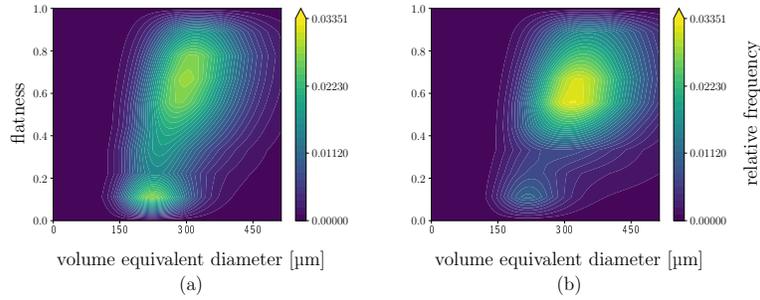


Figure 6: Bivariate densities of the volume equivalent diameter and flatness of particles in the feed, obtained by integrating the trivariate densities of  $x_{\text{vfr}}$ , which have been fitted to segmented image data (a) and reconstructed by means of Eq. (10) (b), respectively.

Moreover, in Figure 7, bivariate marginal densities are shown, which have been obtained for the volume equivalent diameter and sphericity of particles in the feed, by integrating the corresponding trivariate densities of the particle descriptor vector  $x_{\text{vsr}}$ .

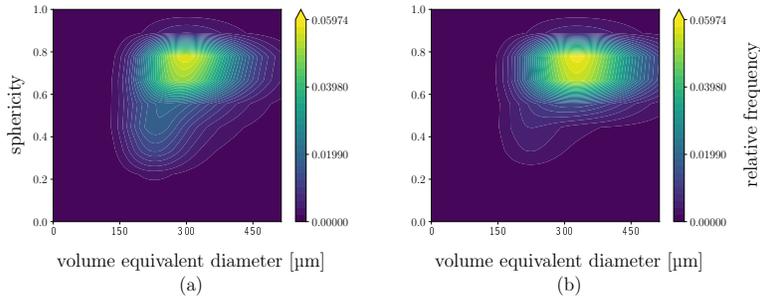


Figure 7: Bivariate densities of the volume equivalent diameter and sphericity of particles in the feed, obtained by integrating the trivariate densities of  $x_{\text{vsr}}$ , which have been fitted to segmented image data (a), and reconstructed by means of Eq. (10) (b), respectively.

It can be clearly seen that the pairs of bivariate probability densities shown in Figures 6 and 7 agree well with each other. This suggests that the method described in Section 2.6 works quite well.

### 3.2 Computed Multivariate Tromp Functions

Using the multivariate probability densities stated in Section 3.1, multivariate Tromp functions  $T : \mathbb{R}^d \rightarrow [0, \infty)$  are computed by means of Eq. (14). However, to obtain a meaningful interpretation of Tromp functions, the values  $T(x)$  are only computed for  $x \in A$ , where the set  $A \subset \mathbb{R}^d$  is given in Eq. (15) with  $q = 0.01$ , i.e., the values of Tromp functions are only computed for particles that are likely to be found in the feed.

For example, the bivariate Tromp function  $T : \mathbb{R}^2 \rightarrow [0, \infty)$  shown in Figure 8 is computed by means of the mass-weighted densities  $f_m^f : \mathbb{R}^2 \rightarrow [0, \infty)$  and  $f_m^c : \mathbb{R}^2 \rightarrow [0, \infty)$  of the descriptor vector  $x_{\text{vr}}$  for particles in the feed and concentrate. Here, as well as in Figure 9 and 10, white areas consist of  $x \notin A$ , i.e., particles that are not sufficiently often observed in the feed, whereas yellow areas indicate that particles with corresponding volume equivalent diameter and zinnwaldite fraction have a high probability of being separated into the concentrate. On the other hand, particles with values of  $x_{\text{vr}}$  in dark blue areas, i.e., particles with a low zinnwaldite fraction, are separated into the tailings with high probability.

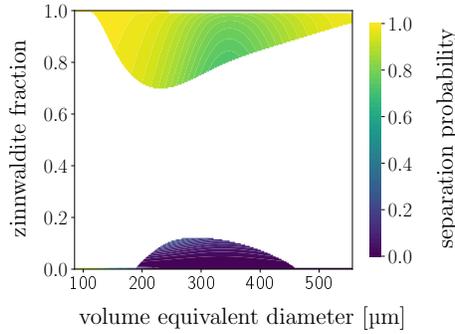


Figure 8: Bivariate Tromp function for the particle descriptor vector  $x_{vr}$ .

To investigate the impact of flatness and sphericity on the separation behavior of particles, the bivariate (marginal) densities of the respective shape descriptors and the zinnwaldite fraction are computed by integrating the trivariate densities of  $x_{vfr}$  and  $x_{vsr}$  for feed and concentrate, respectively. The corresponding bivariate Tromp functions are visualized in Figure 9.

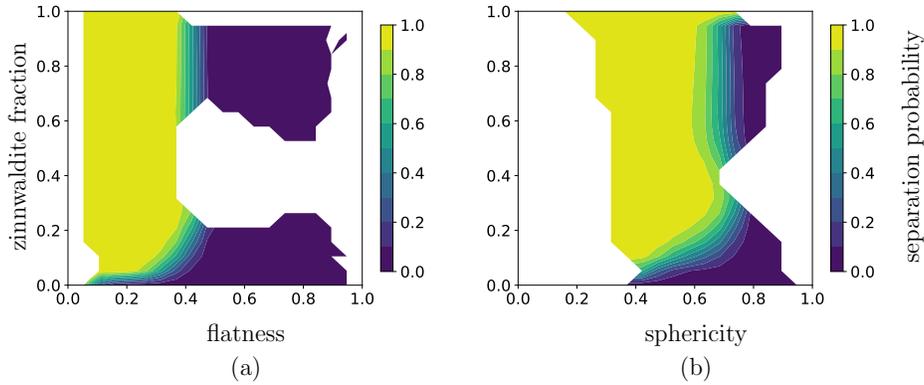


Figure 9: Bivariate Tromp functions for flatness and zinnwaldite fraction (a), as well as for sphericity and zinnwaldite fraction (b)

Moreover, using the trivariate probability densities fitted in Section 3.1 for the three-dimensional particle descriptor vectors  $x_{vfr}$  and  $x_{vsr}$ , trivariate Tromp functions can be computed. In particular, using the bivariate densities of  $(M_{\text{flat}}, M_{\text{rat}})$  and  $(M_{\text{sph}}, M_{\text{rat}})$  given that  $M_{\text{vol}} = v$  for some  $v > 0$ , the conditional bivariate Tromp functions  $T_{(M_{\text{flat}}, M_{\text{rat}} | M_{\text{vol}}=v)} : \mathbb{R}^2 \rightarrow [0, \infty)$  and  $T_{(M_{\text{sph}}, M_{\text{rat}} | M_{\text{vol}}=v)} : \mathbb{R}^2 \rightarrow [0, \infty)$  can be computed. This allows us to obtain 2D Tromp functions conditional on different sizes. Exemplary, conditional Tromp functions for two different particle sizes, i.e., for a volume equivalent diameters of  $v = 166 \mu\text{m}$  and  $v = 332 \mu\text{m}$ , are visualized in Figure 10.

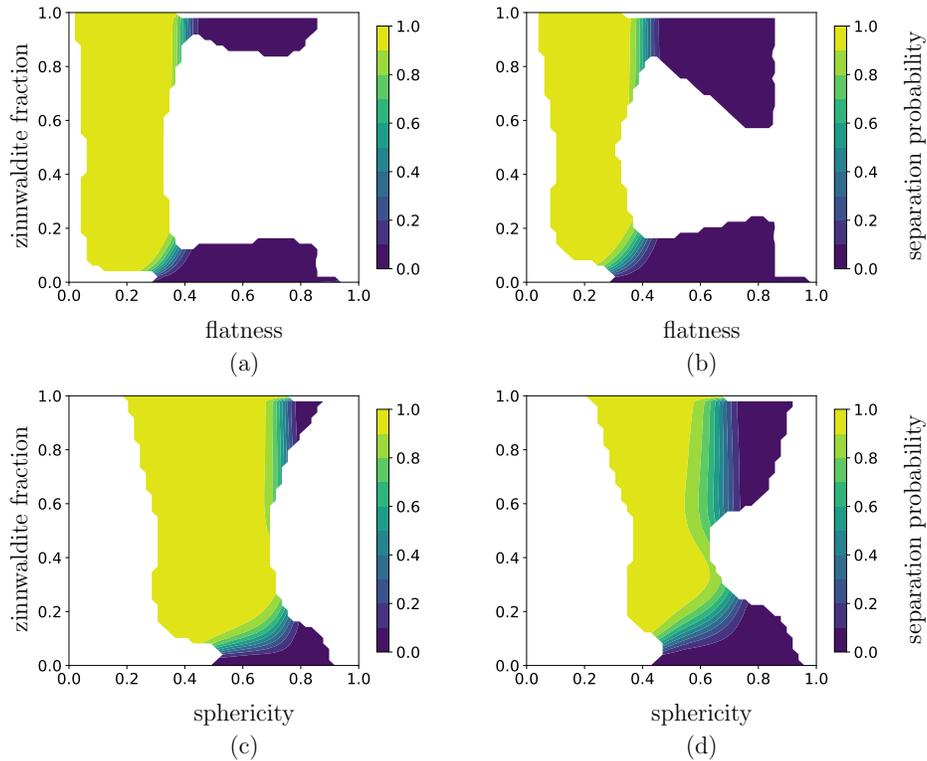


Figure 10: Conditional bivariate Tromp functions  $T_{(M_{\text{flat}}, M_{\text{rat}} | M_{\text{vol}}=v)}$  ((a), (b)) and  $T_{(M_{\text{sph}}, M_{\text{rat}} | M_{\text{vol}}=v)}$  ((c), (d)) for  $v = 166 \mu\text{m}$  ((a), (c)) and  $v = 332 \mu\text{m}$  ((b), (d)).

## 4 Discussion

As shown in Section 3, multivariate probability densities of particle descriptor vectors for the concentrate can be determined by means of corresponding densities for feed and tailings, where the latter densities are estimated based on segmented image data. Then, using the probability densities of particle descriptor vectors for feed and concentrate, bivariate Tromp functions have been established. In previous works [2, 37] bivariate Tromp functions have been computed based on kernel density estimators and 2D data. The advantage of a kernel density estimator is that no model assumptions on the distribution of descriptor vectors need to be made. However, in comparison to our parametric approach, kernel density estimation typically requires more data. Besides, the computational approach presented in Section 2.6.5 for reconstructing the concentrate distribution via an optimization scheme is not possible when deploying kernel density estimates. Moreover, the deployment of parametric Tromp functions can allow for the virtual optimization of process parameters, i.e., to identify parameters of the separation process that result in concentrates with desirable properties [8]. Recently, in [11] bivariate Tromp functions have been computed by means of low-parametric models for particle descriptor distributions. However, in [11] particle descriptor vectors have been computed solely from 2D image data acquired by means of SEM-EDS, which can lead to a stereological problems. In the present paper, we have combined both 3D  $\mu$ -CT and 2D SEM-EDS data for our analyses to counteract such problems.

In the following, it will be discussed in more detail how the Tromp functions demonstrate that the shape, size and composition of the particles have an influence on the separation behavior. In addition, the results of the concentrate reconstruction and the limitations of characterizing multivariate Tromp functions from 3D image data are discussed and some possible solutions to overcome these limitations.

## 4.1 Influence of Particle Descriptors on Magnetic Separation Results

It is well known that there is a strong connection between the results of magnetic separation and particle composition, since magnetizability and composition of particles are strongly correlated. Thus, particles rich in magnetizable zinnwaldite are more likely to be lifted into the concentrate, as illustrated in Figure 8. Still, the results presented in Section 3.2 highlight a complex interplay between separation results and the zinnwaldite fraction with further particle descriptors for their size and shape. For example, Figures 9 and 10 show that for plate-like particles, a lower zinnwaldite fraction is sufficient to be separated into the concentrate. Thus, plate-like particles are easier to be separated into the concentrate.

Figure 8 indicates a small decrease in separation probabilities for larger particles with the same zinnwaldite fraction. However, in previous studies on the same material as that one considered in the present paper, it has been observed that with increasing size the cut susceptibility decreases (i.e., the susceptibility for which 50 % of the particles are recovered in the concentrate) and thus the separation probabilities of particles increase, see e.g. [12]. This seems to be in contradiction with the results in the presented paper. Initially, this might seem attributable to the low relative frequency of particles larger than 300  $\mu\text{m}$ , as shown in Figure 3, potentially resulting in a less robust computation of Tromp functions in this region. However, in absolute terms there are 249 particles larger than 300  $\mu\text{m}$ , and in a comparable study [9], a similar number of particles have been effectively used to fit even higher-dimensional distributions—indicating that the low frequency of particles larger than 300  $\mu\text{m}$  does not have an adverse effect on the bivariate Tromp functions computed in the present paper.

An alternative explanation could be that in the present paper we quantify particle sizes using a different approach than that of [12]. More precisely, in [12] different size classes of the greisen materials were acquired by dry sieving. For size classes associated with larger particles, an overall lower likelihood of separation has been observed, indicating an influence of size on the separation probability. However, this could be attributed to the shape of the particles instead, which strongly correlates with their mineralogical composition. Intuitively speaking, a non-magnetizable quartz particle and a magnetizable zinnwaldite particle of the same volume-equivalent diameter usually have different spatial dimensions, as zinnwaldite tends to form plate-like shapes. As a result, the zinnwaldite particle would be classified into a larger size category during sieving than the quartz particle. This could explain the discrepancy between the results stated in Figure 8 of the present paper and in [12]. Moreover, these observations highlight the complexity of separation processes, i.e., the multivariate interplay between size, composition, shape and separation probabilities. In particular, a direct comparison of the results in the present paper and in [12] is not possible, since the mesh size of a sieve and the volume equivalent diameter cannot be compared.

Nonetheless, the computed Tromp functions provide separation probabilities for different particle descriptor vectors and can be used to predict the separation outcome for new feed materials. More precisely, the computed parametric Tromp functions can be used to predict the distribution of particle descriptors in the concentrate for any given descriptor distribution associated with the feed particles.

## 4.2 Reconstruction of Feed Distributions from Those of Concentration and Tailings

As outlined in Section 2, the relationship between particle descriptor vectors and separation probabilities can be analyzed by means of multivariate probability densities of descriptor vectors which in turn can be calibrated to particle-discrete descriptor vectors. To our knowledge the acquisition of microscopic image data is the most viable method to gain sufficient information on particle systems to compute multivariate probability densities of descriptor vectors, although, for some relatively simple geometries, there are computational methods that can determine such densities from other types of data, instead of using image data [19, 38]. Therefore, we believe that the workflow described in Section 2, for imaging, image processing, parametric multivariate stochastic modeling, and the computation of multivariate Tromp functions, is a valuable addition to the characterization toolbox for complex separation processes. Compared to related non-parametric approaches [37, 39], the parametric copula-based modeling approach, which is considered in the present paper, can address potential complications connected with data acquisition and data processing.

For example, as already mentioned in Section 2.6.4, the formula given in Eq. (10) for the reconstruction of feed distributions can be violated in practice, since the measurements of feed and tailings are only representative in a statistical sense. This effect can be seen in Figures 5, 6 and 7, where the bivariate probability densities of feed particle descriptors, which have been fitted to segmented image data, slightly differ from those obtained by means of Eq. (10). Therefore, fitting the probability densities of concentrate particle descriptors from image data would further improve the accuracy of Tromp functions. But, since in our case the concentrate almost exclusively consists

of strongly elongated and plate-like particles, it is significantly more difficult to correctly segment the CT data of the concentrate than it is for feed and tailings. Analyzing the influence of size, shape and composition on the separation behavior using probability densities of poorly segmented data would not yield sufficiently accurate results. However, the optimization routine, which heavily relies on our parametric modeling approach, allows us to make a statement when more precise conclusions cannot be drawn from the image data.

Note that in [12] and [39] similar difficulties have been reported regarding the accuracy of statements based on bad statistics due to the low number of particles, which are available to determine separation curves from two-dimensional SEM-EDS data.

### 4.3 Limitations of $\mu$ -CT to Characterize the Composition of Particles

Even though we have demonstrated that 3D imaging by means of  $\mu$ -CT is exceptionally informative for the characterization of separation processes, it still poses some limitations in comparison to 2D imaging techniques. For example,  $\mu$ -CT is not well-suited for detecting different fine-scale phases within the material, a task more effectively performed by modern 2D SEM-EDS measurement systems. On the other hand, analyses that are based solely on 2D image data are typically subject to a stereological error of unknown extent [12]. Thus, a correlative approach was presented utilizing the fine-scale phases of the 2D SEM-EDS data as well as the shape and size information observable in CT-data. This approach, leveraging the advantages of both 3D and 2D imaging, is sufficient for simply structured ore particles. However, for more complex particles, such as composite particles, the mineralogical composition in some 2D slices of the particle may not be representative. Therefore, for more complex particle systems, where such complex particles occur more frequently, obtaining a three-dimensional particle-wise segmentation might be necessary. Moreover, the resolution of 3D image data could be increased by deploying super-resolution techniques, while decorating it with features observable in 2D image data [40].

## 5 Conclusions

In this paper, particle systems consisting mainly of zinnwaldite and quartz are considered and both bivariate and trivariate mass-weighted Tromp functions are calculated, which take into account the zinnwaldite fraction of particles as well as their size and shape. Therefore, a parametric modeling approach is used based on 3D  $\mu$ -CT image data and 2D SEM-EDS analyses of feed and tailings. The corresponding multivariate probability distributions of particle descriptor vectors are fitted by means of copula techniques. To determine these distributions for the concentrate, an optimization routine is used, since it was not possible to determine the particle descriptors of the concentrate on the basis of image data. Using multivariate Tromp functions, the impact of shape, size and composition of particles can be investigated regarding their separation behavior in the magnetic separation process considered in the present paper, where it can be seen that all three kinds of particle properties have a significant influence on the separation result. More precisely, a higher zinnwaldite fraction and a flat shape lead to a higher separation probability.

In summary we can state that the considered workflow allows for the computation of multivariate probability distributions of descriptor vectors from image data in order to quantitatively characterize particle systems, independently of their length scale. Furthermore, our workflow allows for the characterization of separation processes using multivariate Tromp functions. Note that Tromp functions can not only be used to characterize the separation behavior of a process, but they also allow for predictive simulations. More precisely, Tromp functions can be used to predict the distribution of particle descriptors in the concentrate for any given descriptor distribution associated with feed particles. This kind of predictive simulation can be the basis for process optimization. More precisely, in [15] we parameterized Tromp functions with respect to process parameters. This allowed us to virtually optimize process parameters such that the distribution of descriptor vectors associated with the concentrate had desirable properties (e.g., a high content of valuable materials). Thus, the presented workflow for deriving a mathematical characterization of separation processes allows for computer-based process optimization, which can reduce costs in time and resources for the calibration of processes in particle technology. In forthcoming works, we intend to apply the proposed method to less ideal systems, such as slags processed with alternative separation techniques, as the separation quality of the considered data was quite good in the presented paper.

## Acknowledgements

The authors gratefully acknowledge funding by the German Research Foundation (DFG) within SPP 2045 (Highly specific and multidimensional fractionation of fine particle systems with technical relevance) under grants 313858392 and 381447825, and within SPP 2315 (Engineered Artificial Minerals /EnAM/ – a geo-metallurgical tool to recycle critical elements from waste streams) under grants 470202518, 470322626 and 470551727.

## A Measurement parameters imaging

Table 8: Measurement parameters of the MLA.

<b>Measurement Info</b>	
Measurement mode	GXMAP
Acceleration voltage	20 kV
Resolution	1 $\mu\text{m}$ /Pixel
Probe current	10 nA
Acquisition time	6 ms
Step size	6 Pixel

Table 9: Measurement and reconstruction parameters of the XCT scans.

<b>Measurement Info</b>	
Voltage	80kV
Power	7 W
Lens	0.4X
Source Position	-40 mm
Detector Position	120 mm
Filter	LE5
Binning	1
Pixel Size	8.56 $\mu\text{m}$
Exposure	15 s
<b>Reconstruction Info</b>	
Software	Zeiss Reconstructor
Center shift	automatic
Beam hardening	0.05
Byte scalinig	-0.05 to 0.3

## B Overview of important mathematical symbols:

Symbol	Description
$W \subset \mathbb{Z}^3$	sampling window
$W_z \subset W$	2D slice of $W$
$I : W \rightarrow \mathbb{R}$	CT-image
$P \subset W$	particle
$M_{\text{vol}}(P)$	volume equivalent diameter of particle $P$
$M_{\text{flat}}(P)$	flatness of particle $P$
$M_{\text{sph}}(P)$	sphericity of particle $P$
$M_{\text{area}}(P)$	surface of particle $P$
$M_{\text{rat}}(P)$	volume fraction of zinnwaldite in particle $P$
$x_{\text{vr}}$	particle descriptor vector $(M_{\text{vol}}(P), M_{\text{rat}}(P))$
$x_{\text{vsr}}$	particle descriptor vector $(M_{\text{vol}}(P), M_{\text{sph}}(P), M_{\text{rat}}(P))$
$x_{\text{vfr}}$	particle descriptor vector $(M_{\text{vol}}(P), M_{\text{flat}}(P), M_{\text{rat}}(P))$
$x = (x_1, \dots, x_d)$	some particle descriptor vector with $d = 2$ or $d = 3$ descriptors
$D_f$	dataset of particles before separation
$D_v^f$	dataset of particles containing almost exclusively valuable material in the feed
$D_{\text{nv}}^f$	dataset of particles containing almost exclusively non-valuable material in the feed
$D_{\text{co}}^f$	dataset of composite particles in the feed
$D_t$	dataset of particles in the tailing after separation
$D_v^t$	dataset of particles containing almost exclusively valuable material in the tailing
$D_{\text{nv}}^t$	dataset of particles containing almost exclusively non-valuable material in the tailing
$D_{\text{co}}^t$	dataset of composite particles in the tailing
$n_j$	number of particles in $D_j$ for $j \in \{f, t\}$
$n_j^i$	number of particles in $D_j^i$ for $i \in \{f, t\}$ and $j \in \{n, \text{nv}, \text{co}\}$
$p \in (0, 1)$	threshold, which is used to subdivide the data based on the volume fraction of zinnwaldite
$F_{1, \dots, d} : \mathbb{R}^d \rightarrow [0, 1]$	cumulative distribution function of a random vector with $d > 0$ entries
$F_j : \mathbb{R} \rightarrow [0, 1]$	marginal distribution function of the $j$ -th component
$f_{1, \dots, d} : \mathbb{R}^d \rightarrow [0, \infty)$	density function of a random vector with $d > 0$ entries
$f_j : \mathbb{R} \rightarrow [0, \infty)$	marginal density function of the $j$ -th component
$C : [0, 1]^d \rightarrow [0, 1]$	copula
$c : [0, 1]^d \rightarrow [0, \infty)$	copula density
$f^f : \mathbb{R}^d \rightarrow [0, \infty)$	multivariate density for particle descriptor vectors of the dataset $D_f$
$f^t : \mathbb{R}^d \rightarrow [0, \infty)$	multivariate density for particle descriptor vectors of the dataset $D_t$
$f_j^i : \mathbb{R}^{d-1} \rightarrow [0, \infty)$	multivariate density of particle descriptor vectors in $D_j^i$ for $i \in \{f, t\}$ and $j \in \{n, \text{nv}\}$
$f_{\text{co}}^i : \mathbb{R}^d \rightarrow [0, \infty)$	multivariate density of particle descriptor vectors in $D_{\text{co}}^i$ for $i \in \{f, t\}$
$f^c : \mathbb{R}^d \rightarrow [0, \infty)$	multivariate density for particle descriptors in the concentrate
$f_m^i : \mathbb{R}^d \rightarrow [0, \infty)$	mass-weighted density version of $f^i$ for $i \in \{f, t, c\}$
$m : \mathbb{R}^d \rightarrow [0, \infty)$	mass function, which maps a particle descriptor vector to the corresponding material density
$m_c$	total mass of particles in the concentrate
$m_f$	total mass of particles in the feed
$\lambda \in [0, 1]$	mass ratio of concentrate and feed
$T : \mathbb{R}^d \rightarrow [0, 1]$	Tromp function
$A \subset \mathbb{R}^d$	set of particle descriptor vector for which the Tromp function is meaningful
$q \in [0, 1]$	threshold indicating the required probability that particles with certain descriptor vectors are observed in the feed to sufficiently inform their separation probability.

## References

- [1] M. Buchmann, E. Schach, R. Tolosana-Delgado, T. Leißner, J. Astoveza, M. Kern, R. Möckel, D. Ebert, M. Rudolph, K. G. Van den Boogaart, U. A. Peuker, Evaluation of magnetic separation efficiency on a cassiterite-bearing skarn ore by means of integrative sem-based image and XRF–XRD data analysis, *Minerals* 8 (2018) 390.
- [2] T. Buchwald, E. Schach, U. A. Peuker, A framework for the description of multidimensional particle separation processes, *Powder Technology* 433 (2024) 119165.
- [3] F. Rhein, H. Ji, H. Nirschl, Multidimensional separation by magnetic seeded filtration: Theoretical study, *Powders* 3 (2) (2024) 217–232.
- [4] J. Sygusch, M. Rudolph, Multidimensional characterization and separation of ultrafine particles: Insights and advances by means of froth flotation, *Powders* 3 (3) (2024) 460–481.
- [5] E. Schach, T. Buchwald, T. Leißner, U. A. Peuker, R. T. Delgado, Concepts of entropy for raw materials, *Powder Technology* 435 (2024) 119398.
- [6] F. Reyes, Q. Lin, J. Cilliers, S. Neethling, Quantifying mineral liberation by particle grade and surface exposure using X-ray micro-CT, *Minerals Engineering* 125 (2018) 75 – 82.
- [7] J. R. A. Godinho, B. L. D. Grilo, F. Hellmuth, A. Siddique, Mounted single particle characterization for 3D mineralogical analysis—MSPaCMA, *Minerals* 11 (2021) 947.
- [8] O. Furat, T. Leißner, K. Bachmann, J. Gutzmer, U. A. Peuker, V. Schmidt, Stochastic modeling of multidimensional particle properties using parametric copulas, *Microscopy and Microanalysis* 25 (2019) 720–734.
- [9] O. Furat, T. Kirstein, T. Leißner, K. Bachmann, J. Gutzmer, U. A. Peuker, V. Schmidt, Multidimensional characterization of particle morphology and mineralogical composition using CT data and R-vine copulas, *Minerals Engineering* 206 (2024) 108520.
- [10] M. Buchmann, E. Schach, M. Rudolph, U. A. Peuker, K. G. van den Boogaart, R. Tolosana-Delgado, Multidimensional particle-based process characterization, in: *Proceedings of the 30th International Mineral Processing Congress*, Cape Town, South Africa, 2020, pp. 1–10.
- [11] T. Wilhelm, J. Sygusch, O. Furat, K. Bachmann, M. Rudolph, V. Schmidt, Parametric stochastic modeling of particle descriptor vectors for studying the influence of ultrafine particle wettability and morphology on flotation-based separation behavior, *Powders* 2 (2023) 353–371.
- [12] T. Leißner, K. Bachmann, J. Gutzmer, U. A. Peuker, MLA-based partition curves for magnetic separation, *Minerals Engineering* 94 (2016) 94–103.
- [13] S. D. D. Welsby, S. M. S. M. Vianna, J. P. Franzidis, Assigning physical significance to floatability components, *International Journal of Mineral Processing* 97 (2010) 59–67.
- [14] O. Furat, T. Leißner, R. Ditscherlein, O. Šedivý, M. Weber, K. Bachmann, J. Gutzmer, U. A. Peuker, V. Schmidt, Description of ore particles from X-ray microtomography (XMT) images, supported by scanning electron microscope (SEM)-based image analysis, *Microscopy and Microanalysis* 24 (2018) 461–470.
- [15] O. Furat, M. Masuhr, F. E. Kruis, V. Schmidt, Stochastic modeling of classifying aerodynamic lenses for separation of airborne particles by material and size, *Advanced Powder Technology* 31 (2020) 2215–2226.
- [16] R. Ditscherlein, O. Furat, M. de Langlard, J. Martins de Souza e Silva, J. Sygusch, M. Rudolph, T. Leißner, V. Schmidt, U. A. Peuker, Multiscale tomographic analysis for micron-sized particulate samples, *Microscopy and Microanalysis* 26 (2020) 676–688.
- [17] J. Schreier, O. Furat, M. Cankaya, V. Schmidt, U. Bröckel, Automated evaluation of contact angles in a three-phase system of selective agglomeration in liquids, *Image Analysis and Stereology* 39 (2020) 187–196.
- [18] O. Furat, D. P. Finegan, D. Diercks, F. Usseglio-Viretta, K. Smith, V. Schmidt, Mapping the architecture of single lithium ion electrode particles in 3D, using electron backscatter diffraction and machine learning segmentation, *Journal of Power Sources* 483 (2021) 229148.
- [19] O. Furat, U. Frank, M. Weber, S. Wawra, W. Peukert, V. Schmidt, Estimation of bivariate probability distributions of nanoparticle characteristics, based on univariate measurements, *Inverse Problems in Science and Engineering* 29 (2021) 1343–1368.
- [20] R. Ditscherlein, O. Furat, E. Löwer, R. Mehnert, R. Trunk, T. Leißner, M. J. Krause, V. Schmidt, U. A. Peuker, PARROT: A Pilot study on the open access provision of particle-discrete tomographic datasets, *Microscopy and Microanalysis* 28 (2022) 350–360.

- [21] S. Englisch, R. Ditscherlein, T. Kirstein, L. Hansen, O. Furat, D. Drobek, T. Leißner, B. Apeleo Zubiri, A. P. Weber, V. Schmidt, U. A. Peuker, E. Spiecker, 3D analysis of equally X-ray attenuating mineralogical phases utilizing a correlative tomographic workflow across multiple length scales, *Powder Technology* 419 (2023) 118343.
- [22] E. Schach, T. Buchwald, O. Furat, F. Tischer, A. Kaas, L. Kuger, M. Masuhr, J. Sygusch, T. Wilhelm, U. A. Peuker, Progress in the application of multidimensional particle property distributions: The separation function, *KONA Powder and Particle Journal* (in print).
- [23] J. Sygusch, T. Wilhelm, O. Furat, K. Bachmann, V. Schmidt, M. Rudolph, Application of multivariate Tromp functions for evaluating the joint impact of particle size, shape and wettability on the separation of ultrafine particles via flotation, *Powders* 3 (2024) 338–366.
- [24] J. Sygusch, T. Wilhelm, O. Furat, K. Bachmann, V. Schmidt, M. Rudolph, Investigating the multidimensional separation behavior of particles in a cyclosizer setting - A case study on calcite, fluorite and magnesite, in: *Proceedings of the International Mineral Processing Congress (2024)*, in print.
- [25] J. Dissmann, E. C. Brechmann, C. Czado, D. Kurowicka, Selecting and estimating regular vine copulae and application to financial returns, *Computational Statistics & Data Analysis* 59 (2013) 52–69.
- [26] T. Heinig, K. Bachmann, R. Tolosana-Delgado, G. Van Den Boogaart, J. Gutzmer, Monitoring gravitational and particle shape settling effects on MLA sampling preparation, in: *Proceedings of International Association for Mathematical Geosciences, 2015*, pp. 200–206.
- [27] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: Learning dense volumetric segmentation from sparse annotation, in: *Proceedings of the Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, Springer, 2016*, pp. 424–432.
- [28] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2017).
- [29] G. Barequet, S. Har-Peled, Efficiently approximating the minimum-volume bounding box of a point set in three dimensions, *Journal of Algorithms* 38 (2001) 91–109.
- [30] K. Schladitz, J. Ohser, W. Nagel, Measuring intrinsic volumes in digital 3D images, in: *Proceedings of the International Conference on Discrete Geometry for Computer Imagery, Springer, 2006*, pp. 247–258.
- [31] R. B. Nelsen, *An Introduction to Copulas*, Springer, 2006.
- [32] D. P. Kroese, Z. I. Botev, T. Taimre, R. Vaisman, *Data Science and Machine Learning: Mathematical and Statistical Methods*, CRC Press, 2019.
- [33] N. L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions, Volume 2*, J. Wiley & Sons, 1995.
- [34] N. L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*, J. Wiley & Sons, 1994.
- [35] C. Schröder, S. Rahmann, A hybrid parameter estimation algorithm for beta mixtures and applications to methylation state classification, *Algorithms for Molecular Biology* 12 (2017) 21.
- [36] M. Killiches, D. Kraus, C. Czado, Examination and visualisation of the simplifying assumption for vine copulas in three dimensions, *Australian & New Zealand Journal of Statistics* 59 (2017) 95–117.
- [37] M. Buchmann, E. Schach, T. Leißner, M. Kern, T. Mütze, M. Rudolph, U. A. Peuker, R. Tolosana-Delgado, Multidimensional characterization of separation processes – Part 2: Comparability of separation efficiency, *Minerals Engineering* 150 (2020) 106284.
- [38] U. Frank, D. Drobek, A. Sánchez-Iglesias, S. E. Wawra, N. Nees, J. Walter, L. Pflug, B. Apeleo Zubiri, E. Spiecker, L. M. Liz-Marzán, et al., Determination of 2D particle size distributions in plasmonic nanoparticle colloids via analytical ultracentrifugation: Application to gold bipyramids, *ACS nano* 17 (6) (2023) 5785–5798.
- [39] E. Schach, M. Buchmann, R. Tolosana-Delgado, T. Leißner, M. Kern, K. G. van den Boogaart, M. Rudolph, U. A. Peuker, Multidimensional characterization of separation processes – Part 1: Introducing kernel methods and entropy in the context of mineral processing using SEM-based image analysis, *Minerals Engineering* 137 (2019) 78–86.
- [40] A. Dahari, S. Kench, I. Squires, S. J. Cooper, Fusion of complementary 2D and 3D mesostructural datasets using generative adversarial networks, *Advanced Energy Materials* 13 (2) (2023) 2202407.