Bonded-particle extraction and stochastic modeling of internal agglomerate structures

Aaron Spettl[a,*], Simon Bachstein[a], Maksym Dosta[b], Monika Goslinska[b], Stefan Heinrich[b], Volker Schmidt[a]

[a] *Institute of Stochastics, Ulm University, Germany*
[b] *Institute of Solids Process Engineering and Particle Technology, Hamburg University of Technology, Germany*

Abstract

The discrete element method (DEM) is an effective computational technique that is used to investigate the mechanical behavior of various particle systems like, for example, agglomerates. However, for systems of perfectly spherical and non-overlapping particles, the structural input is almost always based only qualitatively on experimentally observed structures. In this paper, we consider the case of agglomerates where particles are nearly spherical and connected by bonds. A novel bonded-particle extraction (BPE) method is proposed for the automated approximation of such agglomerate structures from tomographic data sets. This method can be effectively used in conjunction with various commercial or open-source DEM simulation systems. By BPE, sphere-like primary particles are represented each by exactly one (perfect) sphere, and the set of spheres is non-overlapping. Furthermore, the solid bridge bonds between primary particles are retained. Having derived such a simple description of complex tomographic data sets, one can perform DEM simulations with well-established models like the bonded-particle model. Moreover, it is shown that a larger data base of statistically equivalent microstructures can be generated by a stochastic modeling approach. This approach reduces the need for (time-consuming) experimental agglomerate production and characterization.

Keywords: agglomerate structure, bonded-particle model, bonded-particle extraction, segmentation, stochastic microstructure model

1. Introduction

The discrete element method (DEM) [1] is a modern and effective computational technique to simulate the mechanical behavior of granular systems on microscale. A common approach to the simulation-based investigation of agglomerates is the bonded-particle model [2]. The particles are specified as a dense packing of spheres that are bonded together. Usually, particles are spherical and bonds cylindrical. More complex geometries are possible, but they lead to higher computational effort. This is due to the need for more complex contact models or the description of complex objects as a cluster of spheres [3, 4, 5, 6, 7]. So far, very often agglomerate microstructures are obtained by generating packings of (bonded) particles, such that they have similar properties as observed experimentally in real agglomerates (see, *e.g.*, [8, 9, 10]). However, more realistic morphologies are desirable [7]. In the last few years, the microstructural characterization in 3D using μCT has become feasible and, therefore, more popular. For example, agglomerate microstructures have been studied [11, 12] and, *e.g.*, their relationship to process variables has been investigated [13, 14]. As a link to DEM, a direct approach is to approximate 3D structures of real agglomerates by (idealized) objects. Recently, experimentally obtained structures have been represented by different kinds of objects. For example, individual particles may be approximated by ellipsoids [15, 16], clusters of (non-overlapping or overlapping) spheres [15, 17, 18], polyhedra [19] or splines [20]. In that case, DEM simulations and real experiments can be compared directly – the structures are not only statistically equivalent with respect to some characteristics, they are identical (under the restriction of having idealized objects).

In this paper, we consider the case of highly spherical primary particles. In contrast to the literature mentioned above, this allows us to represent each particle by exactly one sphere. However, this simplicity comes at a cost. Even for a packing of particles with high sphericities of about 0.9, it is hard to find a non-overlapping set of spheres without changing the structure too much. To the authors knowledge, volume-equivalent spheres with a subsequent DEM relaxation step have been used to obtain particle configurations from tomographic data; see, *e.g.*, [21]. However, the presence of solid bridges in agglomerates is a problem because mechanical equilibrium of the system is not a sufficient criterion for relaxation, and, of course, solid bridges have a volume as well. With these problems in mind, we propose a new automated method to extract bonded particle systems from tomographic data. We state an optimization problem to find a configuration of bonded particles such that (i) the spherical particles are non-overlapping, (ii) the cylindrical bonds match the thickness of solid bridges, (iii) the agglomerate mass does not change, and (iv) the shape and size of the agglomerate is captured well. A stochastic optimization method is applied to find a solution to this high-dimensional and multi-extremal optimization problem. We call this method the bonded-particle extraction (BPE) method.

*Corresponding author. Tel.: +49 731 50 23555; fax: +49 731 50 23649.
Email address: aaron.spettl@uni-ulm.de (Aaron Spettl)

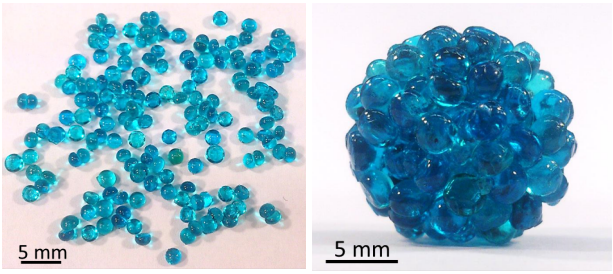
In the second part of the paper, we show how statistically equivalent microstructures can be obtained by a parametric stochastic model for agglomerate microstructures. The model proposed in [22] is fitted to real agglomerate structures. There are several advantages of such an approach. First, one can reduce the number of real agglomerates that have to be produced and characterized experimentally. Second, realizations generated by a fitted stochastic model are more closely related to the real microstructures than packings generated with less information (*e.g.*, only based on aggregated characteristics like the mean coordination number). Third, parametric models allow the systematic variation of individual structural parameters in a realistic setting. As mentioned in [22], this is an important step to obtain reliable results by means of DEM, which leads to a better understanding of the relationship between microstructure and mechanical properties.

2. Experimental data

In this contribution, agglomerates consisting of maltodextrin (DE47) primary particles have been analyzed. Maltodextrin is a material which is widely used as a model substance for amorphous food powders and employed often as carrier or filler in the food industry. The production of agglomerates was performed in three subsequent steps: (1) creation of spherical primary particles; (2) storage of primary particles under specific conditions (temperature, humidity); (3) agglomeration of primary particles under high temperature.

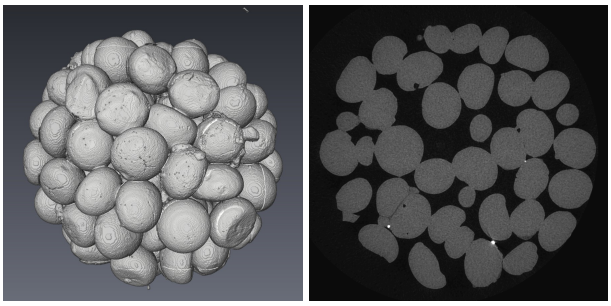
To produce spherical maltodextrin particles, a solution consisting of 70 wt% maltodextrin and 30 wt% water was prepared. In order to decrease the water content, the solution was preheated in the microwave oven. Afterwards, it was dropped via separate droplets into an oil bath and primary particles were formed. For further decrease of the water content, the particles were placed in the oven at a temperature of 85°C. This allows to get primary particles with a mean sphericity of 0.864 (standard deviation 0.026; sphericity as defined by Wadell [23]) and water content of 3 wt% (Figure 1(a)). In the second step, in order to reproduce different storage conditions, the water content of primary particles was increased by placing them into a chamber with high air humidity and temperature. Finally, agglomerates were assembled by putting particles together into spherical (Figure 1(b)) or cylindrical form and placing them in the oven at a temperature of 65°C.

In total, 34 maltodextrin agglomerates were produced and characterized using a µCT 35 of SCANCO Medical AG. The tomographic reconstruction was performed based on 100 rotations, where each 2D cross-section was captured at a resolution of 2048 × 2048 pixels. A visualization is given in Figure 2. There are 10 spherical agglomerates (diameter about 16 mm) and 24 cylindrical agglomerates (diameter about 16 mm, height about 11 mm). A dataset label is assigned to each agglomerate. The spherical agglomerates are labeled by elements of a certain set $\mathcal{L}_S$, the cylindrical agglomerates by $\mathcal{L}_C$, and $\mathcal{L} = \mathcal{L}_S \cup \mathcal{L}_C$ corresponds to all datasets. Furthermore, the mass of each agglomerate was measured. It is in the range of 1.8 to 2.25 g, where each agglomerate consists of about 130 to 170 primary particles.



(a) Primary particles. (b) Spherical agglomerate.

Figure 1: Maltodextrin primary particles and agglomerates.



(a) 3D visualization. (b) Planar 2D section.

Figure 2: Tomographic data of a spherical agglomerate.

3. Image processing

The tomographic data sets are denoted by $I_\ell = \{I_\ell(x,y,z) \in \{0,\dots,255\} : (x,y,z) \in W\}$, where $\ell \in \mathcal{L}$ denotes the considered dataset and $W \subset \mathbb{N}^3$ is a grid of voxel coordinates. Because of the high resolution, the original image data were (down)scaled by factor 0.5 in all three directions. The final grid of voxel coordinates has a size of $1024 \times 1024 \times 1024$ and the edge length of a voxel in the processed data corresponds to 0.02 mm.

In a first step, all images are preprocessed. As the sample holder is visible, it is removed from all data sets. This is easily possible by setting the affected voxels to black, *i.e.*, grayscale value zero, because the sample holder is located at the same position in all datasets. Furthermore, a median-filter [24] with a box size of $3 \times 3 \times 3$ voxels is applied, which reduces noise without losing much structural specifics. The resulting images are denoted by I'_ℓ.

Global thresholding is used to binarize all images I'_ℓ. For a threshold τ_ℓ, the resulting binary image $I_\ell^{\tau_\ell}$ is given by

$$I_\ell^{\tau_\ell}(x,y,z) = \begin{cases} 255 & \text{if } I'_\ell(x,y,z) \geq \tau_\ell, \\ 0 & \text{if } I'_\ell(x,y,z) < \tau_\ell, \end{cases}$$

where the value 255 corresponds to the solid (maltodextrin) phase, also called foreground phase. Because the mass of each

agglomerate and the density $\rho = 1500$ kg/m^3 of maltodextrin are known, the threshold τ_ℓ can be chosen such that the "mass" of the foreground phase

$$\#\left\{(x,y,z) \in W : I_\ell^{\tau_\ell}(x,y,z) = 255\right\} \cdot (0.02\text{ mm})^3 \cdot \rho$$

corresponds to the experimentally measured mass as closely as possible (#A denotes the number of elements in a set A).

Having obtained the thresholds τ_ℓ and the corresponding thresholded images $I_\ell^{\tau_\ell}$, a further processing step is necessary to obtain the final binary image that will be used in all subsequent steps. One can observe that there are very small foreground or background clusters of voxels present in the thresholded images, which are obviously not relevant. The Hoshen–Kopelman algorithm [25] is applied to detect clusters in the foreground as well as in the background. Small clusters with a volume of at most 5^3 voxels are removed. The resulting binary images are denoted by I_ℓ^{bin}.

3.1. Agglomerate shape parameters

Later on, the position and exact shape of agglomerates is required. Therefore, it is necessary to detect the spheres or cylinders that match the agglomerates best. It is already known which agglomerate has a spherical or cylindrical shape. However, the exact coordinates and sizes of these spheres and cylinders have to be determined.

The Hough transform (HT) [26, 27, 28] is used to estimate the "best-fit" spheres or cylinders, respectively. The idea of the HT is to maximize the agreement between the boundary of an object observed in (binary) image data, and a simple geometric object that has one or several parameters. Essentially, the (discretized) parameter space is scanned and a so-called accumulator space is constructed, where a score is assigned to every parameter. The score is the number of voxels that belong to the object boundary in both the image and the parametric object. Then, the global maximum in this accumulator space identifies the object providing the best match. For example, for a sphere there are four parameters: the three coordinates of the center and the radius. For a cylinder, there are 7 parameters: the coordinates of two extreme points on the cylinder axis, and the radius.

Given the binary images of spherical agglomerates, the best-fit spheres are detected as follows. First, the convex hull [29, 30] of all foreground voxels is computed. The primary particles do not fill out the complete space of the agglomerate – by computing the convex hull, pore phase voxels lying "between" pairs of solid phase voxels are added, yielding a single (non-parametric) object corresponding to the agglomerate shape. The boundary of the convex hull is discretized to an image with a certain thickness. (In our case, we dilated the convex hull surface with a sphere as structuring element [24, 31], where the sphere centered in the origin has a radius of three voxels.) Based on this image, the HT is applied to detect a sphere whose surface matches the convex hull boundary best. Note that the thickness used for the convex hull boundary discretization is helpful because the convex hull is, of course, not a perfect sphere, and some tolerance is sensible.

For the binary images of the cylindrical agglomerates, the procedure is exactly the same. The only difference is that one has to take even more care when determining a suitable discretization of the 7-dimensional parameter space, to avoid issues with very long computing times and high memory requirements.

The detected objects describing the agglomerate positions and shapes are denoted by S_ℓ, $\ell \in \mathcal{L}$. Note that the best-fit shape may not cover all voxels of the foreground. This can be observed in Figure 3. The inner circle is the best-fit; however, there are some particles (partially) outside. The object S_ℓ is dilated with a sphere as structuring element such that the foreground is just completely contained; this result is denoted by S'_ℓ.

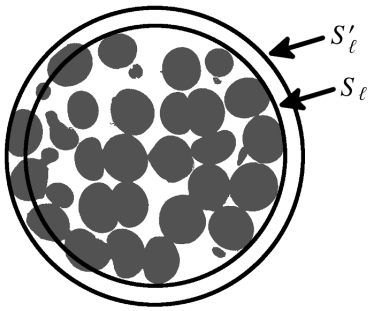


Figure 3: Planar 2D section with best-fit object (inner circle: cross-section of 3D sphere) and best-fit object dilated such that the complete solid volume is contained (outer circle: cross-section of dilated 3D sphere).

3.2. Regions of primary particles

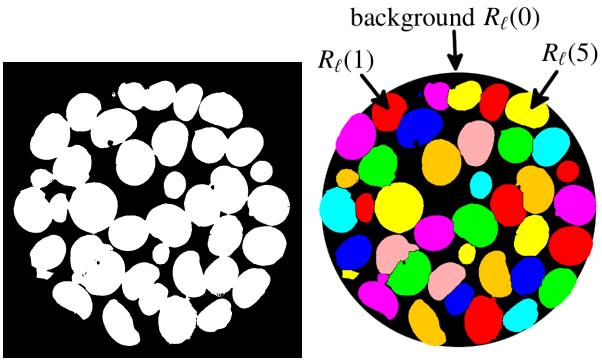
The individual particles are identified in the binarized tomographic data by means of the watershed algorithm [32, 33, 34, 35]. The idea and standard technique is as follows. One considers shortest distances from voxels of the foreground to the background phase. An image containing such shortest distances is called (Euclidean) distance transform [24, 36]. Taking the negative image, "valleys" with respect to the grayscale values (*i.e.*, high distances) are starting points for the objects that one wants to detect. Then, starting from these points, the valleys are flooded until the basins touch each other. The process is called watershed transformation; the voxels where two basins meet are called watersheds.

In the standard watershed transformation, all local minima are used as starting points for basins. With this method, over-segmentation is a typical problem [31, 33]. The reason is that the distance transform is very sensitive to small problems in the original binary image, therefore many local minima are generated. A solution to this problem are marked-based watershed transformations [31, 33]. In that case, instead of using the local minima, some predefined markers are employed. Often, these markers are chosen based on the regional minima, but some minima are left out or joined. The agglomerates investigated in this study consist of primary particles with a high sphericity. Therefore, the distance to the pore phase that is associated with every local minimum can be thought of as the radius of a sphere around that local minimum. Every other

local minimum contained in that sphere and having a smaller distance to the background is removed.

Using the thinning procedure of local minima as described above yields very good results. However, small unnecessary regions can exist near the boundary of a particle or between two particles. Therefore, in a last step, watershed regions with a small volume are assigned to an adjacent (large) region. This is useful as it is known that primary particles have a certain minimum size. Note that it is not necessary to distinguish between particles and bonds at this point. The identified particle regions consist each of one primary particle and, potentially, one or several (partial) bonds. The optimization problem stated below will assign volume to particles and bonds as necessary.

The final result is given by labeled images, where each particle region is identified by an (integer) label. For $N_\ell > 1$ particle regions in data set $\ell \in \mathcal{L}$, the labeled image is given by $\{I_\ell^{\text{regions}}(x, y, z) \in \{0, \dots, N_\ell\} : (x, y, z) \in W\}$. In particular, the voxels of the ith particle region are given by $R_\ell(i)$. Note that $R_\ell(0)$ does not correspond to a particle region, it denotes the background, *i.e.*, the pore phase plus watersheds that separate particle regions by a layer with one voxel thickness. A visualization of the result is given in Figure 4.



(a) Binary image data. (b) Identified particle regions.

Figure 4: Identification of particle regions with watershed transformation, based on binary data (planar 2D section). Colors of particles are put at random.

3.3. Extraction of region data

As an input to the BPE method described below, not only the agglomerate shape is needed, but also information describing the particle regions. These are the volumes of particle regions and the contact areas between particle regions (which describe the thickness of solid bridges). It is sensible to assume that these two characteristics have a large influence on the mechanical behavior. In the following, the volume of a particle region is called a *local volume*. It contains the volume of one primary particle and (partially) the volume of solid bridges. The local volume $V_\ell(i)$ for particle i can be obtained by counting the voxels of the particle region, *i.e.*, $V_\ell(i) = \#R_\ell(i)$.

The contact areas between particle regions are estimated based on ideas used for surface area estimation in voxelized data. A simple method is to use a local weighting of voxel configurations as presented in [37]. First, particles in contact are determined by detecting for every background voxel whether it is in contact to two particles (*i.e.*, it is 26-adjacent[1] to two regions with different particle labels). All voxels in contact to two particles determine the approximate region of interest. The local weights introduced in [37] are evaluated for all voxel configurations that touch the region of interest. With this technique, the surface area from both particles is summed up. Therefore, the contact area is given by half the obtained sum; it is denoted by $A_\ell(i, j)$ for two particle regions with labels i and j. The variable $A_\ell(i, j)$ is set to zero if the two particles are not adjacent.

4. Bonded-particle extraction

The primary goal is a parametric description of the agglomerates such that the data may be used as input for DEM simulations. We restrict ourselves to the case of approximately spherical primary particles and want to represent one primary particle by exactly one (solid) sphere. The system of spheres must be non-overlapping because we want the system to be in an equilibrium state (when put into DEM simulations as an initial configuration, without external forces applied). The solid bridges between primary particles are approximated by cylindrical bonds. Natural conditions are the preservation of agglomerate shape as well as its mass (or its volume, which is equivalent).

4.1. Optimization problem

Recall that the following information was extracted from the tomographic data sets. For every analyzed agglomerate, there is

- the "best-fit" object describing the agglomerate shape S_ℓ, including S'_ℓ,
- the local volume $V_\ell(i) > 0$ for every particle region $i = 1, \dots, N_\ell$, and
- the contact area $A_\ell(i, j) \geq 0$ between two particle regions.

Based on this information, the aim is to find a bonded-sphere description of the original data. We assume that the local volumes as well as the thickness of contacts dominate the mechanical behavior. There is flexibility in the positions and radii of the spheres; under the condition that the local volume is matched quite well and the contact area (or bond radius) is fixed. The radius of the bond between particles i and j is computed as $r_{i,j} = \sqrt{A_\ell(i, j)/\pi}$. The fact that primary particles are not perfectly spherical and the condition that the spheres should be non-overlapping can lead to bonded-sphere configurations where the spheres drift apart when only minimizing overlap. To lessen this effect, a good fit to the agglomerate shape is required as well. An intuitive further restriction is that the radius of a particle may not be smaller than the radius of a bond emanating from this particle.

[1] The 26-neighborhood of a voxel (x, y, z) is given by all voxels that share a face, edge or corner with (x, y, z).

We state the problem of finding a suitable bonded-sphere configuration as an optimization problem, where all undesired configurations are assigned a cost. As we only allow flexibility in the sphere centers and radii, the optimization problem can be formulated as follows. A set of spheres $\{B(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{r}_i),\ i = 1, \ldots, N_\ell\}$ should be detected, where $B(x, y, z, r)$ denotes the sphere with center (x, y, z) and radius $r > 0$. The optimal parametrization $\{(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{r}_i),\ i = 1, \ldots, N_\ell\}$ of the spheres is obtained by minimizing a cost function, *i.e.*,

$$\{(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{r}_i)\} = \underset{\{(x_i, y_i, z_i, r_i) \in P_i\}}{\arg\min}\ C(\{(x_i, y_i, z_i, r_i)\}), \tag{1}$$

where $P_i = \{(x_i, y_i, z_i, r_i) \in \mathbb{R}^4 : r_i \geq \max_{j \neq i} r_{i,j}\}$ denotes the admissible parameters for each sphere $i = 1, \ldots, N_\ell$; the total cost of a configuration $\{p_i\} = \{(x_i, y_i, z_i, r_i) \in P_i, i = 1, \ldots, N_\ell\}$ of spheres is given by

$$\begin{aligned} C(\{p_i\}) =& \frac{1}{N_\ell} \sum_{j=1,\ldots,N_\ell} C^{\{p_i\}}_{\text{overlap}}(j) + \frac{1}{2} \max_{j=1,\ldots,N_\ell} C^{\{p_i\}}_{\text{overlap}}(j) \\ &+ \frac{1}{N_\ell} \sum_{j=1,\ldots,N_\ell} C^{\{p_i\}}_{\text{volume}}(j) + \frac{1}{2} \max_{j=1,\ldots,N_\ell} C^{\{p_i\}}_{\text{volume}}(j) \\ &+ \frac{1}{N_\ell} \sum_{j=1,\ldots,N_\ell} C^{\{p_i\}}_{\text{shape}}(j) + \frac{1}{2} \max_{j=1,\ldots,N_\ell} C^{\{p_i\}}_{\text{shape}}(j) \end{aligned} \tag{2}$$

meaning that the mean value and maximum value of three different cost functions is minimized, which each map the particle label to a cost value. The cost values themselves are volumes, which makes interpretation and weighting easier. The three cost functions are given by

- the *overlap volume* of particle j with all other particles, *i.e.*,

  $$C^{\{p_i\}}_{\text{overlap}}(j) = \sum_{k \neq j} V_{\text{intersection}}(p_j, p_k)$$

  where $V_{\text{intersection}}(p_j, p_k)$ denotes the intersecting volume of the two spheres p_j and p_k,

- the *local volume mismatch* in comparison to the original particle region j, *i.e.*,

  $$\begin{aligned} C^{\{p_i\}}_{\text{volume}}(j) =& \Bigg| V_\ell(j) - V_{\text{sphere}}(p_j) + \frac{1}{2} C^{\{p_i\}}_{\text{overlap}}(j) - \\ & \frac{1}{2} \sum_{k \neq j \text{ with } r_{j,k} > 0} V_{\text{bond}}(p_j, p_k, r_{j,k}) \Bigg| \end{aligned}$$

  where $V_{\text{sphere}}(p_j)$ denotes the volume of the sphere p_j, and $V_{\text{bond}}(p_j, p_k, r_{j,k})$ stands for the volume of the cylindrical bond with radius $r_{j,k}$ between particles j and k, and

- the *volume outside of the agglomerate shape* S_ℓ of particle j, *i.e.*,

  $$C^{\{p_i\}}_{\text{shape}}(j) = V_{\text{sphere}}(p_j) - V_{\text{intersection}}(p_j, S_\ell)$$

  where $V_{\text{intersection}}(p_j, S_\ell)$ denotes the intersecting volume of the sphere p_j and the agglomerate shape S_ℓ.

Note that the appendix contains some comments on how to compute the required volumes, *i.e.*, the bond volumes, sphere–sphere intersections, and sphere-cylinder intersections.

4.2. The cross-entropy method

The optimization problem given in (1) is high-dimensional and multi-extremal, which makes it hard to solve. To compute an approximative solution to this problem, we apply a stochastic optimization technique, namely the cross-entropy (CE) method [38, 39, 40, 41]. The basic idea is to interpret the solution as a (degenerate) random variable, whose distribution is learned starting from a plausible initial configuration. A standard choice is to use normal densities whose variances shrink until they are below a predefined threshold. In every iteration, samples are drawn using the current normal densities, the best-performing samples are selected, then they are used to update the normal densities. This procedure has proven to be suitable for high-dimensional and multi-extremal optimization problems [39, 41, 42].

For a given dataset $\ell \in \mathcal{L}$, the natural initial configuration for the spheres describing the particles is given by their centers of mass $(x_i^{(0)}, y_i^{(0)}, z_i^{(0)})$ and their volume-equivalent radii $r_i^{(0)} = (3V_\ell(i)/(4\pi))^{1/3}$. However, one must take care that a particle has at least the same radius as each bond that starts or ends in this particle. Therefore, the condition $r_i^{(0)} \geq \max_{j \neq i} r_{i,j}$ is enforced.

The application of the CE method to our problem is given by the following algorithm.

1) **Initialization.** Set $t = 0$. Obtain the initial parametrization of the normal parameter densities from the initial configuration given above, *i.e.*, set the mean values

   $$(\mu_{X_i}^{(0)}, \mu_{Y_i}^{(0)}, \mu_{Z_i}^{(0)}) = (x_i^{(0)}, y_i^{(0)}, z_i^{(0)}), \quad \mu_{R_i}^{(0)} = r_i^{(0)},$$

   and standard deviations

   $$(\sigma_{X_i}^{(0)}, \sigma_{Y_i}^{(0)}, \sigma_{Z_i}^{(0)}) = (d_c, d_c, d_c), \quad \sigma_{R_i}^{(0)} = d_r,$$

   for all $i = 1, \ldots, N_\ell$, where $d_c,\ d_r > 0$ denote the initial standard deviations of coordinates and radii, respectively. For example, they can be chosen as a fraction of the mean diameter of all particle regions.

2) **Set up next iteration.** Increment t and define independent normally-distributed random variables, *i.e.*,

   $$\begin{aligned} X_i^{(t)} &\sim \mathcal{N}\left(\mu_{X_i}^{(t-1)}, (\sigma_{X_i}^{(t-1)})^2\right), \\ Y_i^{(t)} &\sim \mathcal{N}\left(\mu_{Y_i}^{(t-1)}, (\sigma_{Y_i}^{(t-1)})^2\right), \\ Z_i^{(t)} &\sim \mathcal{N}\left(\mu_{Z_i}^{(t-1)}, (\sigma_{Z_i}^{(t-1)})^2\right), \\ R_i^{(t)} &\sim \mathcal{N}_{\geq \max_{k \neq i} r_{i,k}}\left(\mu_{R_i}^{(t-1)}, (\sigma_{R_i}^{(t-1)})^2\right), \end{aligned}$$

   for all $i = 1, \ldots, N_\ell$. Note that the index "$\geq \max_{k \neq i} r_{i,k}$" to $\mathcal{N}$ emphasizes the natural constraint that the radius of a particle must be equal to or larger than all radii of its emanating bonds.

3) **Sample generation.** Generate M samples

   $$\left\{((x_i^{(t,s)}, y_i^{(t,s)}, z_i^{(t,s)}), r_i^{(t,s)}),\ s = 1, \ldots, M\right\}$$

   of $((X_i^{(t)}, Y_i^{(t)}, Z_i^{(t)}), R_i^{(t)})$, $i = 1, \ldots, N_\ell$.

4) **Sample evaluation.** Compute the cost function given in Eq. (2) for all samples, denote the cost values by $c^{(t,s)}$, *i.e.*, set

$$c^{(t,s)} = C\left(\{((x_i^{(t,s)}, y_i^{(t,s)}, z_i^{(t,s)}), r_i^{(t,s)})\}\right)$$

for all $s = 1, \ldots, M$.

5) **Update parameters of densities.** Update the mean values and standard deviations based on the best-performing samples, which is called the *elite set*. More precisely, for rarity parameter $\varrho \in (0, 1)$, use the $\lceil \varrho M \rceil$ samples with the lowest cost values $c^{(t,s)}$ and use their coordinates and radii to estimate the new mean values $\mu_{X_i}^{(t)}, \mu_{Y_i}^{(t)}, \mu_{Z_i}^{(t)}, \mu_{R_i}^{(t)}$ and standard deviations $\sigma_{X_i}^{(t)}$, $\sigma_{Y_i}^{(t)}$, $\sigma_{Z_i}^{(t)}$, $\sigma_{R_i}^{(t)}$. This is done by using the sample mean and sample standard deviation estimators of the elite set.

6) **Termination condition.** If all standard deviations are sufficiently small, *i.e.*,

$$\max\{\sigma_{X_i}^{(t)}, \sigma_{Y_i}^{(t)}, \sigma_{Z_i}^{(t)}, \sigma_{R_i}^{(t)}, i = 1, \ldots, N_\ell\} < \varepsilon$$

for some $\varepsilon > 0$, then return with the result

$$\{(\hat{x}_i, \hat{y}_i, \hat{z}_i, \hat{r}_i)\} = \left\{(\mu_{X_i}^{(t)}, \mu_{Y_i}^{(t)}, \mu_{Z_i}^{(t)}, \mu_{R_i}^{(t)})\right\} .$$

Otherwise, go back to step 2.

A good choice for the parameters of this algorithm is required. Our investigations showed that $M = 4000$, $\varrho = 0.05$, and $\varepsilon = 0.1$ are suitable. This means that the elite set consists of $\lceil \varrho M \rceil = 200$ samples, which is a good data base for estimating sample means and sample variances. If the standard deviations of coordinates and radii are below 0.1 voxels for all particles, the algorithm is terminated. The initial standard deviations were chosen as $d_c = \bar{r}_\ell/6$ and $d_r = \bar{r}_\ell/20$, where $\bar{r}_\ell = N_\ell^{-1} \sum_{i=1}^{N_\ell} r_i^{(0)}$ denotes the mean radius of volume-equivalent spheres of the particle regions. The exact choice of these parameters is not very important for convergence, but it may affect the computational cost to perform one iteration and the number of required iterations. For example, for larger M one could choose a smaller ϱ – then one iteration would take more time, but typically less iterations are required in total. Note that the CE method has a natural aptitude to escape local minima. This ability can be improved even more using fixed smoothing, dynamic smoothing or variance injection [39]. However, for our problem, these techniques seem not to be necessary – the results did not improve substantially.

4.3. Postprocessing

In optimization problem (1), it is not clear whether there exists a configuration of spheres $\{(x_i, y_i, z_i, r_i) \in P_i\}$ such that $C(\{(x_i, y_i, z_i, r_i)\}) = 0$. Particles observed in real materials are not perfectly spherical and they are often densely packed. As a consequence, it is to be expected that a "zero cost" is often not possible. This implies that all three cost functions will have non-zero cost in the typical case. Therefore, there may be differences in the local volume, spheres may overlap or the shape condition may be slightly violated. This is not critical for local volumes or the shape – but the overlapping of spheres should be minimized or eliminated to avoid unphysical starting configurations for DEM simulations, and to have an equilibrium configuration without external forces applied (because overlapping is often interpreted as elastic deformation of particles).

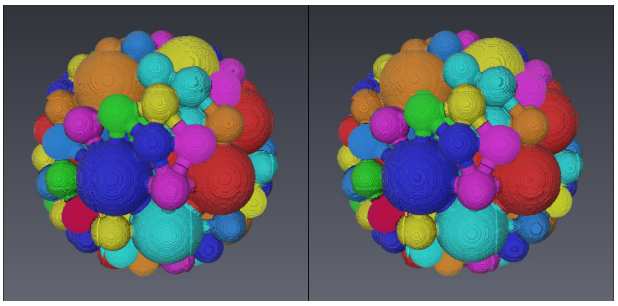
For this reason, we propose a postprocessing step. Overlapping is removed with the force-biased algorithm [43, 44], a collective rearrangement algorithm for the packing of spheres. The algorithm is modified such that each sphere is moved at most a distance of 1/20 voxel per iteration. By applying this restriction, it is ensured that the structure itself does not change fundamentally. Typically, the CE method already detects a solution very close to a non-overlapping configuration. However, in many cases, it did not completely coincide with such a configuration because the other (local volume and shape) cost components would have increased. In principle, it is possible to modify the CE method to obtain an admissible solution directly. However, this would require a dynamic weighting of the three cost functions, *i.e.*, the penalty of overlapping configurations would have to increase the closer one gets to terminating the algorithm. But, note that even for very high penalties, a configuration without any overlapping is not guaranteed.

4.4. Validation with artificial data

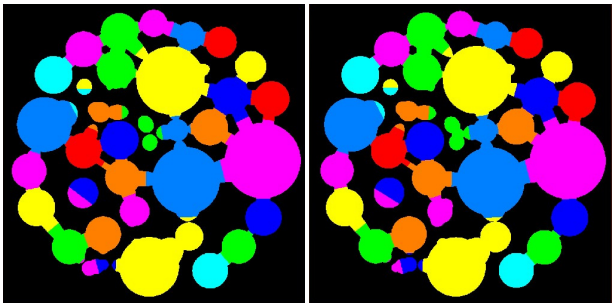
To check the suitability of the proposed BPE method, artificial data sets were generated, where the exact locations and radii of particles are known. We used a stochastic agglomerate microstructure model that has been presented recently [22], where a random mixing ratio controls the particle size distribution. For the artificial data set considered in the following, a mixing ratio of $a = 0.6$ was chosen. A realization of the model is a bonded-sphere configuration, where only two sizes of spherical primary particles are present. In expectation, 60% of the solid volume consists of large spheres, 40% of small spheres, which have exactly half the radius of the large spheres. Bonds are generated such that all particles are connected and a certain total bond volume is obtained [22].

In a first step, the local volumes (of primary particles including partial bonds) and contact areas are computed directly, *i.e.*, without discretization to a voxel grid. We chose this technique to avoid the influence of image data preprocessing, where regions and contact areas are not perfectly estimated. Then, the CE method is applied to detect the original spheres. A visual inspection shows that the centers of all extracted particles are very close to their original positions and their radii are almost perfect. The centers of particles are slightly shifted: the absolute center displacement divided by the mean particle radius is 7%. The mean absolute difference in particle radii is 0.3% of the mean particle radius. Figure 5 shows a visualization of both the original as well as the extracted bonded-particle system.

Furthermore, we discretized the simulated agglomerate and stored it as a binary image. All steps from detecting agglomerate shape, extracting particle regions, estimating particle region volumes and particle contact areas as well as

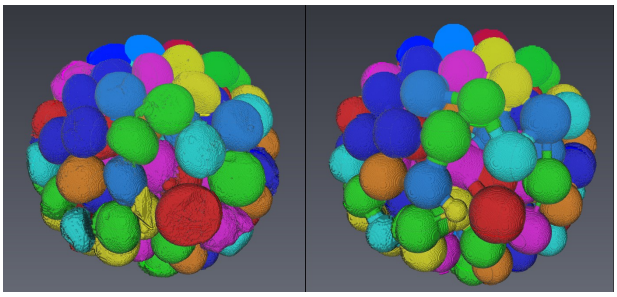


(a) 3D visualization.

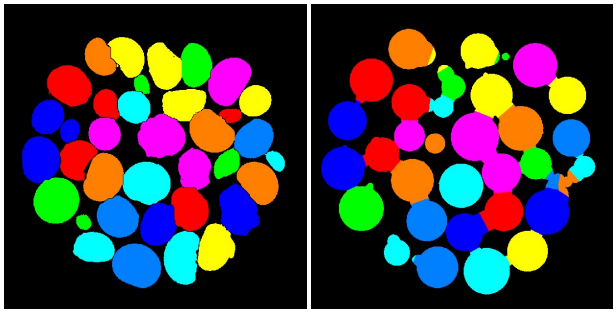


(b) Planar 2D section.

Figure 5: Original discretized microstructure of an agglomerate realized by the model presented in [22] (left), and result of BPE method (right).

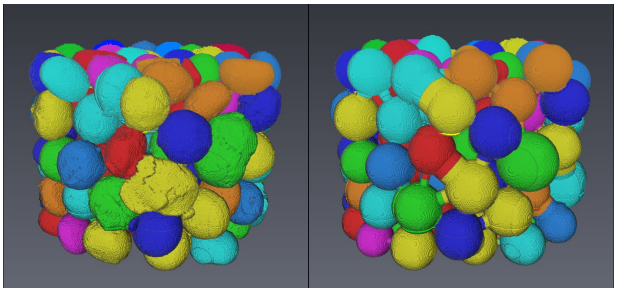


(a) 3D visualization.

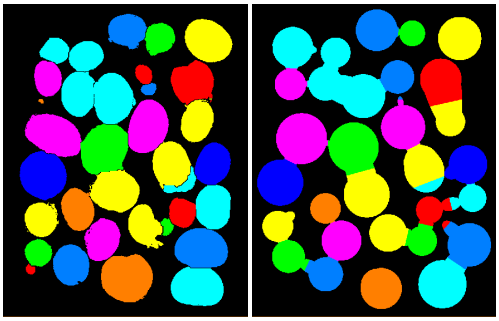


(b) Planar 2D section.

Figure 6: Original microstructure of spherical agglomerate (left), and result of BPE method (right).



(a) 3D visualization.



(b) Planar 2D section.

Figure 7: Original microstructure of cylindrical agglomerate (left), and result of BPE method (right).

the CE method were applied. The relative displacement of particle centers is 10%, the mean absolute difference in particle radii is 1.2% (again, both percentages are proportions of the mean particle radius). The mean absolute difference in bond radii is 7.8% of the mean bond radius. This deviation is caused by the estimation of contact areas from image data. Typically, the estimated contact areas are too high. There are two reasons for this. First, there is no absolute guarantee that the watersheds, which separate particle regions, have optimal properties. Second, surface area estimation techniques for voxelized data are not perfect, either. Considering these facts, we think that the approximation is quite good, although this could be improved certainly.

4.5. Results for experimental data

The BPE method has been applied to all experimental data sets; we call the extracted systems of particles and bonds bonded-particle approximations (BPAs). The main part of the computation time is spent on the Hough transform that is used to detect the agglomerate shape parameters. The CE method itself takes only 4 to 8 minutes for one agglomerate (using multi-threading on a standard quad-core processor). The CE method terminates after 250 to 450 iterations, depending on the data set. A visualization of the results is given in Figure 6 (for a spherical agglomerate) and in Figure 7 (for a cylindrical agglomerate).

The quality of the BPAs is investigated in the following. First, we look at the total volume of the extracted agglomerates. The volume of the configuration obtained directly from the CE method should correspond to experimental solid volume almost exactly, as the local volumes are optimized. Table 1 confirms that the volume before postprocessing is almost perfect. There are only three (of 34) agglomerates where the ratio of volumes from BPAs and original data exceeds 1.02. These problematic cases will be discussed below. It

is clear that the postprocessing, which removes all remaining overlapping, increases the volume. By shifting overlapping particles apart, the volume of the bonds gets larger. This effect is in an acceptable range for most agglomerates. 30 of 34 agglomerates have volume ratios between 1.004 and 1.031. However, the analysis of 4 agglomerates is problematic. An investigation of the tomographic data shows that primary particles in these agglomerates are (partially) melted together, effectively violating the assumption of a sphere-like shape. This makes it impossible to represent the same volume with non-overlapping spheres. Furthermore, this manifests itself already in the result of the CE method (where the cost value after termination is still very high), but becomes really distinct in the structure after enforcing non-overlapping particles.

Table 1: Total volume of extracted bonded-particle system divided by original volume obtained from image data; once before and once after postprocessing (removal of remaining overlapping).

volume ratio	CE result	after postprocessing
minimum	0.993	1.004
median	0.999	1.016
mean value	1.002	1.025
maximum	1.066	1.138

Another important characteristic of agglomerates is their porosity. As discussed in Section 3.1, not all particles are completely contained in the detected shape object S_ℓ. Therefore, the container S'_ℓ has been introduced, which covers the complete solid volume observed in tomographic data. Table 2 shows that the fraction of the solid volume with respect to S_ℓ or S'_ℓ changes in the approximations: On the one hand, in S_ℓ, the solid volume fraction of the approximation decreases by one percent point in the median. On the other hand, the same difference evaluated in S'_ℓ shows an absolute increase of +0.8 %. This increase can be attributed to the slight volume increase after postprocessing, which we discussed above. The large minimum and maximum absolute differences observed for our 34 agglomerates correspond again to agglomerates where particles had to be shifted significantly outside of the agglomerate shape to obtain a non-overlapping packing of particles. Summarizing, slightly larger agglomerates with correspondingly higher porosity are obtained owing to the non-perfect sphericity of experimental primary particles. However, this plays no role once one considers S'_ℓ. This means that primary particles in the BPAs may be pushed outwards, but this effect is comparable to some of the experimental primary particles, which are also not perfectly contained in S_ℓ.

Altogether, we can state that the method works fine as long as the spherical shape condition of primary particles is not violated too badly. This is an inherent problem that cannot be solved directly. If the assumption of sphere-like primary particles is not reasonable, then one should think about, *e.g.*, using ellipsoids or several spheres to represent one particle – either by bonding several non-overlapping spheres together to form the primary particle, or by using the multi-sphere DEM approach [4, 5, 6]. Another very flexible alternative is voxel-based DEM simulation [45].

Table 2: Percent points in the difference of solid volume fractions: volume fraction in extracted bonded-particle systems minus the respective original volume fractions, which are obtained directly from tomographic data.

solid volume fraction	difference in S_ℓ	difference in S'_ℓ
minimum	−4.7 %	+0.2 %
median	−1.0 %	+0.6 %
mean value	−1.2 %	+0.8 %
maximum	+3.4 %	+4.5 %

5. Stochastic modeling

The stochastic model for agglomerate microstructures presented in [22] has been used to investigate the effect of structural changes on breakage behavior of agglomerates. The model is not based on experimental data – this has the advantage that one can tailor the microstructures exactly as desired. However, as mentioned in [22], fitting the stochastic model to experimental agglomerate structures is important to ensure that the simulation results are in good agreement with experiments. The validity of results from DEM calculations can be verified only if exactly the same agglomerates or at least the same type of agglomerates (*i.e.*, statistically equivalent agglomerates) can be characterized experimentally. The BPE method proposed in the present paper is a first step to facilitate the validation of simulated mechanical behavior by comparing it to the measurements – based on exactly the same samples. Nonetheless, it remains useful to be able to generate a larger number of agglomerate structures that are statistically equivalent. Furthermore, slightly modified structures allow the investigation of effects caused by structural changes in a realistic setting. Note that such approaches have been used previously for other types of materials. For example, in [46], a stochastic model is proposed for coral-like microstructures in Al–Si alloys, which has been fitted to real data. Then, this model has been used to determine the mechanical behavior of real and virtually generated alloy microstructures by help of the finite element method [47].

In this section, we show how the stochastic agglomerate microstructure model [22] can be fitted to experimental data. First, the idea of the modeling approach is briefly repeated. Then, model parameters are estimated. Finally, the quality of fit is discussed by comparing characteristics obtained directly from experimental data with simulated microstructures.

5.1. Model description

The stochastic model introduced in [22] is based on the idea of a network of primary particles, which are connected to each other directly or indirectly through a bond system. We summarize the main requirements and the model construction.

Requirements are isotropy of the structure, non-overlapping spheres interpreted as primary particles, connectivity of the primary particles through bonds, and a defined total volume of the bonds. The model construction follows directly from these requirements.

- **Primary particles.** The primary particles are obtained from a sphere-packing algorithm that generates isotropic structures. A commonly used technique is the *force-biased algorithm* [43, 44]. The initial configuration is generated such that all sphere centers are placed randomly in the agglomerate shape W (uniform distribution on W, independent placement). The radii of spheres $r_1, r_2, \ldots$ are drawn independently from a random radius R_p. The packing density η of the spheres is a parameter of the model. The number of spheres is chosen such that the packing density η is matched as well as possible.

- **Bond system.** The main parameters for bond generation are the radius distribution for the cylindrical bonds and the total volume of bonds. The total volume of bonds is defined relative to the total volume of primary particles; the 'bond/particle volume ratio' is called b. Bond radii are defined indirectly. For every possible bond connecting particle i and j, a factor is applied to the minimum radius of its primary particles, *i.e.*, $r_{i,j} = k_b \min\{r_i, r_j\}$ for some factor $k_b \in (0, 1]$. The value k_b is deterministic in the model of [22]. For real data, a random factor K_b is used to draw the factor k_b for every bond independently.

  The generation of the bond system itself is a bit more involved. The bonds between particles are put as follows. Intuitively, primary particles with small distances are likely to be connected by a solid bridge. For this reason, the existence of a bond is controlled by a thresholding value on the particle–particle distances. After having generated a radius for every possible bond, the threshold is selected such that the correct total volume (specified by b) is obtained. However, one cannot be sure that all primary particles of an agglomerate are connected to each other. Therefore, tools from graph theory like the minimum-spanning tree are used to guarantee connectivity by adding additional bonds if necessary (for details, see [22]).

- **A comment on the bond network model.** In this paper, the bond radius factor k_b is obtained from some random variable K_b. Therefore, a clarification of the minimization problem [22, Section 2.1.3] is necessary. The threshold $l^\star$ for bond lengths is determined such that the total volume of bonds is as close as possible to the desired bond volume. The bond volumes depend on the bond radii. Therefore, bond radii have to be realized for every possible bond (*i.e.*, every pair of primary particles) before the minimization problem. This means that bond radii are already fixed previously to the selection of the threshold $l^\star$ – only the existence of a bond for a particular pair of particles depends on the threshold.

In [22], all parameters except a mixing ratio (which determined R_p) were constant. For real agglomerates, it is clear that this is not reasonable. As mentioned above, it makes sense to allow random bond radii by introducing a random factor K_b. Furthermore, we take η to be a realization of a random packing density H and b to be a realization of a random bond/particle volume ratio B. Therefore, the parameters are the four random variables R_p, K_b, H and B. It is assumed that these four random variables are independent. (Note that particle radii and bond radii are obtained from independent copies of R_p and K_b.) All four random variables are modeled using parametric distributions.

5.2. Model fitting

In this section, we describe how we selected the parametric families of distributions and their parameters for the random variables H, K_b, R_p, and B. All characteristics are fitted based on the data from both spherical and cylindrical agglomerates because we assume that they are statistically equivalent in their internal microstructure. Note that three agglomerates were excluded manually (in order to not use them for fitting): these are exactly the three agglomerates from experiments that were very conspicuous regarding the bonded-particle extraction (compare Section 4.5). The agglomerate shape is not fitted to experimental data – later on, if we want to generate spherical or cylindrical agglomerates, we define the desired (non-random) shape object directly.

First, the packing density of primary particles is considered. The results shown in Figure 8 were obtained from the experimental data sets by computing the primary particle volume inside the detected agglomerate shape, *i.e.*, the packing density is estimated by computing the ratio of the volume of all extracted primary particles intersected with S_ℓ and the volume of S_ℓ. Figure 8 shows a histogram and a (non-parametric) estimation of the density function (obtained by kernel density estimation, see [48, 49]), which is both useful to get a feeling for the shape of a distribution. In this case, however, it does not make much sense to try to match the exact shape because the number of data sets is only 31. Therefore, we chose the normal distribution and fitted it to the samples using maximum-likelihood [50]. We obtained $H \sim \mathcal{N}_{[\eta_{\min}, \eta_{\max}]}(\mu_H, \sigma_H^2)$ with $\mu_H = 0.549$ and $\sigma_H = 0.025$. The additional condition $H \in [\eta_{\min}, \eta_{\max}]$ is natural: very small packing densities below approximately $\eta_{\min} = 0.45$ do not make much sense, and packing densities above $\eta_{\max} = 0.65$ are hard to obtain with perfect hard spheres.

Essentially, the same procedure is used for fitting the distribution of the the bond/particle volume ratio. The BPAs allow to compute the ratios of the total bond volume to the total primary particle volume. A histogram of the resulting samples is drawn in Figure 9. Again, a non-parametric estimate of the density is shown. We chose a (shifted) gamma distribution in order to have a positive lower bound and a skew distribution. The parameters of the gamma distribution were fitted to the shifted samples $\tilde{b}_\ell$, *i.e.*, $\tilde{b}_\ell = b_\ell - 0.02$, where the shift parameter 0.02 was selected manually. Therefore, the bond volume has always at least 2% of the primary particle

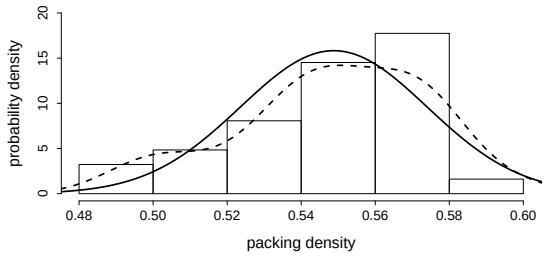


Figure 8: Estimated packing densities of primary particles observed in experimental data (histogram), non-parametric density function (dashed curve) and fitted normal distribution (solid curve).

volume. The random bond/particle ratio volume is given by $B \sim 0.02 + \text{Gamma}_{[0,0.18]}(\alpha_B, \beta_B)$, where the shape parameter was determined as $\alpha_B = 6.214$ and the rate parameter as $\beta_B = 141.951$. The condition $B \in [0.02, 0.2]$ is applied such that no unreasonable large values are obtained (which has a low probability, but would be possible otherwise). It is assumed that B is independent of H – we could not observe an obvious relationship when looking at the scatter plot of the experimental bond/particle volume ratios against the particle packing densities.

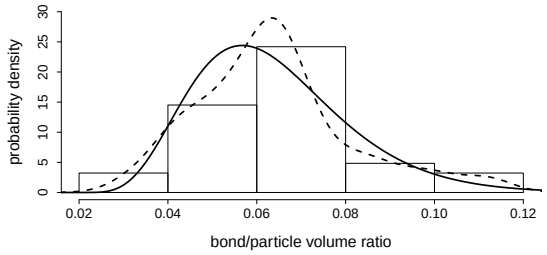
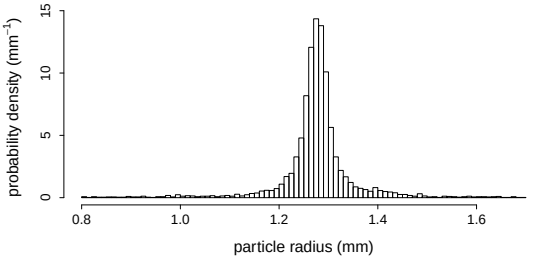
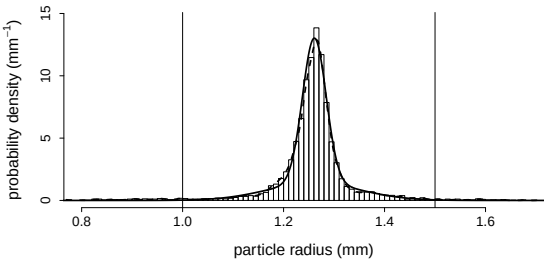


Figure 9: Estimated bond/particle volume ratios in bonded-particle approximations of experimental data (histogram), non-parametric density function (dashed curve) and fitted (shifted) gamma distribution (solid curve).

The two remaining distributions are those of the particle radii and the (relative) bond radii. The samples consist of all particles or bonds, respectively, of all agglomerates. Therefore, the sample sizes are much larger. Figure 10 shows histograms of the sample of particle radii: once for the volume-equivalent radii of the original particle regions, and once for particle radii of bonded-particle approximations (which are the basis for fitting). It is clear that the distribution is very narrow. Nonetheless, there are some outliers, which are ignored by requiring all radii to be in the range of 50 to 75 voxels (corresponding to 1 to 1.5 mm). A very good fit can be obtained by a mixture of two normal distributions, *i.e.*, $R_p = R_{p,1}$ with probability $w \in (0, 1)$ and $R_p = R_{p,2}$ with probability $1 - w$, where $R_{p,1} \sim \mathcal{N}_{[50,75]}(\mu_{R_{p,1}}, \sigma^2_{R_{p,1}})$ and $R_{p,2} \sim \mathcal{N}_{[50,75]}(\mu_{R_{p,2}}, \sigma^2_{R_{p,2}})$. Therefore, the mixture distribution has 5 parameters. Maximum-likelihood fitting yields $\mu_{R_{p,1}} = 63.047$, $\sigma_{R_{p,1}} = 1.169$, $\mu_{R_{p,2}} = 62.997$, $\sigma_{R_{p,2}} = 4.599$ (all in voxels) and $w = 0.682$.



(a) Volume-equivalent radii of particle regions.



(b) Particle radii in BPAs of experimental data.

Figure 10: Histograms of particle radii obtained directly from experimental data and from BPAs. For the latter, a particle radius distribution was fitted and its density is shown as a solid curve. The dashed curve is a non-parametric estimation of the density. The two vertical lines indicate a manual truncation of the samples, because very small and very large radii are considered as artifacts from image processing.

Finally, the relative bond radii are considered. Again, a histogram is shown in Figure 11, and only samples between 0.05 and 0.9 are used. We do not want to consider the slight clustering effect visible near zero, as we can suspect that very thin bonds may originate from particle contacts (or just particles that are very close together; *i.e.*, without solid bridges). On the other hand, bonds whose radii are almost as large as their connected primary particles are likely to correspond to segmentation problems or primary particles that melted together. The remaining data follows a skew distribution. Therefore, we fitted a gamma distribution (truncated to the interval $(0, 1]$). The obtained result is $K_b \sim \text{Gamma}_{(0,1]}(\alpha_{K_b}, \beta_{K_b})$ with $\alpha_{K_b} = 5.105$ and $\beta_{K_b} = 15.349$. The fitted density is shown in Figure 11.

5.3. Model validation

In the previous section, four distributions were fitted to the experimental data. Based on these fits and the stochastic model for agglomerate microstructures, 100 internal structures were simulated each for the spherical and cylindrical agglomerates. In this section, the quality of fit regarding further agglomerate characteristics is evaluated. Namely, a comparison is done for the number of contacts between primary particles, for the structure of the pore phase, for the arrangement of primary particles and for bond characteristics.

First, we consider the coordination number, *i.e.*, the number of bonds per primary particle. Figure 12 shows the distribution of coordination numbers obtained from experimental data as

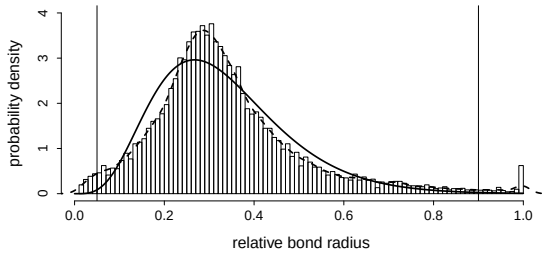
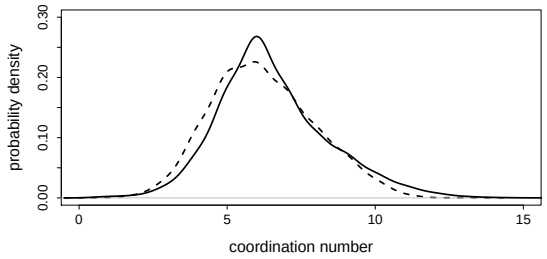
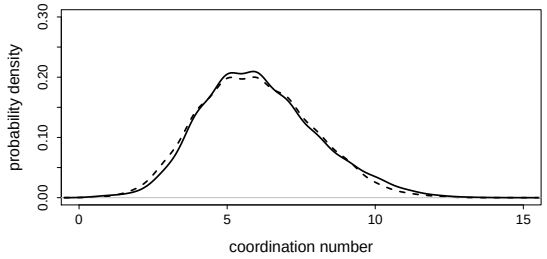


Figure 11: Samples of relative bond radius K_b in BPAs of experimental data (histogram), non-parametric density function (dashed curve) and fitted gamma distribution (solid curve). The two vertical lines indicate a manual truncation of the samples, because very thin and very thick bonds are considered as artifacts from image processing.

well as from simulated data. Recall that the coordination numbers of the particle regions from tomographic data are identical to the coordination numbers in the BPAs. Estimated density functions are used instead of histograms, which makes it easier to compare the two distributions in a single figure. We can state that the agreement is quite good, even though the number of contacts is not directly modeled in the stochastic modeling approach. Note that extremely high coordination numbers (> 12) are very rare; however, they are possible due to the non-equally sized primary particles, their medium packing density of about 0.55, and the existence of bonds (*i.e.*, primary particles do not need to be in direct contact).



(a) Spherical agglomerates.



(b) Cylindrical agglomerates.

Figure 12: Coordination numbers in BPAs of experimental data (dashed curve) and from simulated agglomerates (solid curve).

An important characteristic describing the pore space is the spherical contact distribution. A point of the pore space is selected randomly according to the uniform distribution on the pore space, and the resulting distribution of shortest distances to the solid phase is the quantity of interest. An illustration of this concept is given in Figure 13. The smallest distances to the solid phase evaluated for a large number of randomly placed points describe the pore sizes and pore shapes in an aggregated way, without the need to give an (object-based) definition of individual pores. Figure 14 shows the estimated density functions for experimental data. For simulated data, a 96% point-wise confidence band is computed. The density functions obtained from experimental data should in almost all cases be between the upper and lower bound of the confidence band. This is the case for spherical agglomerates. Only for cylindrical agglomerates, there are a few experimental data sets that seem to have a bit more larger pores. Looking at the microstructure of cylindrical agglomerates, we can confirm that there are some agglomerates where the cylindrical shape is not as perfect as desired, which leads to large "pores" at the boundary of the detected shape object.

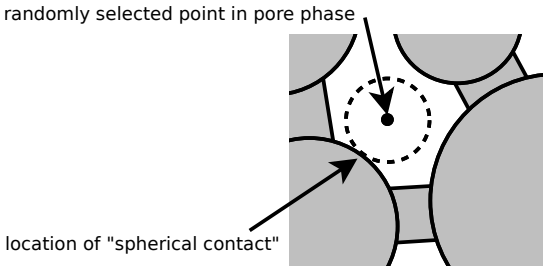
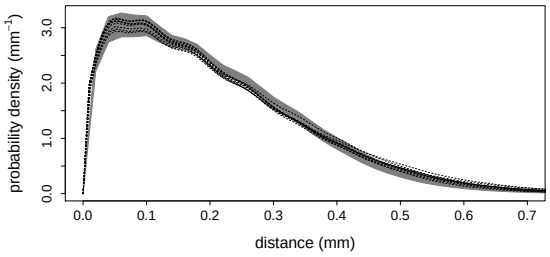


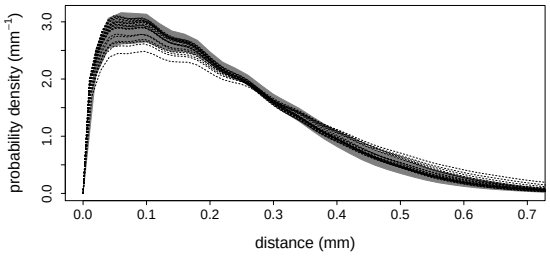
Figure 13: 2D illustration of the spherical contact distance for a point in the pore phase to the solid phase. The spherical contact distance is given by the distance of the point to the location of "spherical contact", *i.e.*, it corresponds to the radius of the dashed circle. The evaluation of all spherical contact distances for all possible points yields a distribution, which can be understood as the "spherical contact distribution" of a randomly chosen point.

The pair correlation function is a tool that is often used to describe attraction and repulsion in point patterns, in dependence of pairwise point distances. A reference scenario is the case of "complete spatial randomness" – *i.e.*, there is no interaction between points. In that case, the pair correlation equals one. As shown in Figure 15 and as expected for packings of spheres, there is a clear attraction of points with certain distances (distances that correspond to about two times the mean particle radius). This effect is visible for the same distances in both experimental and simulated data and its strength is very similar for the first two peaks.

Finally, two characteristics of the bonds are considered: their lengths and their radii. The experimental data has two main features: an increasing bond radius for increasing bond length, and a proportion of bonds where the bond radius seems to be independent of the bond length. This can be observed in Figure 16(a) and (c). Note that the bond length is not the distance between particle centers, but the length of the cylinder surface that is outside the particles. The scattering of length–radius pairs is not identical for experimental and simulated data, *cf.* Figure 16(b) and (d). The first main feature is also observed in the simulated data, which leads us to believe

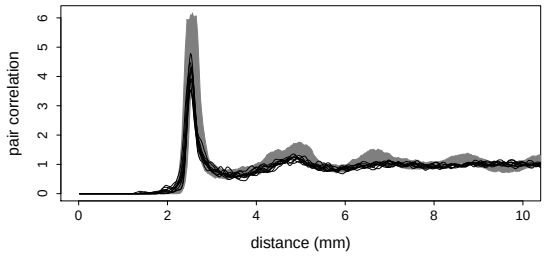


(a) Spherical agglomerates.

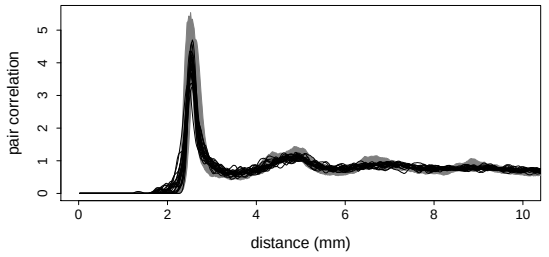


(b) Cylindrical agglomerates.

Figure 14: Spherical contact distribution function: distribution of shortest distances when considering a random point in the pore phase and evaluating its shortest distance to the solid phase. The density functions are shown for experimental data (black dashed lines), and 96% point-wise confidence bands were estimated from simulated data (gray area). Note that the black lines from experimental data are hard to distinguish. However, it is important to see that almost no black lines are outside the confidence bands.



(a) Spherical agglomerates.



(b) Cylindrical agglomerates.

Figure 15: Pair correlation function of primary particle centers: the relative frequency of particle centers in a given distance to each other (relative to point patterns with 'complete spatial randomness'). The pair correlation functions are shown for BPAs of experimental data (black lines), and 96% point-wise confidence bands were estimated from simulated data (gray area).

that the bond network modeling approach is realistic. However, the second feature is missing and could certainly be integrated as a modification to the stochastic model. We refrain from making such a change for two reasons. First, at least two parameters would be needed to integrate these effects into the model (width of bonds of the second type and their proportion) – this would complicate the model, its implementation, and its fitting significantly. Second, we do not yet know why these bonds with nearly constant thickness are present, and whether they can be found in other types of agglomerates as well.

Altogether, we can state that we obtained a very realistic stochastic model that is able to imitate the structural features and randomness of real agglomerates quite well.

6. Conclusions and outlook

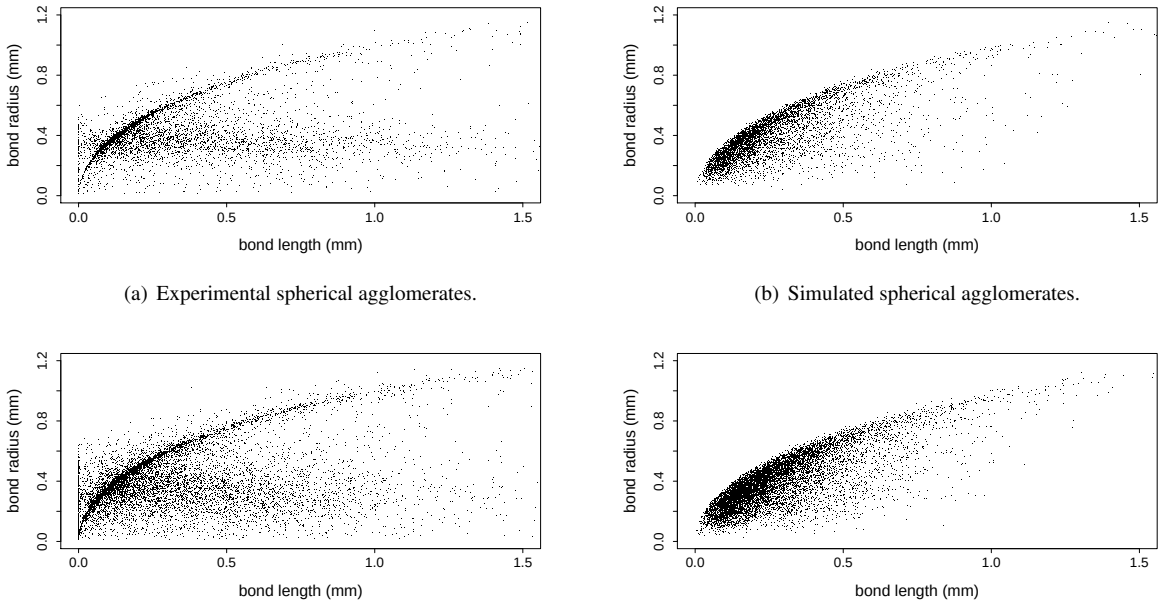
In this paper, we consider the problem of extracting optimal and generating realistic particle structures suitable for DEM simulations, where particles are bonded together. For example, this is the case for agglomerates, which are produced in various apparatuses by addition of binder or after sintering of stored material. Maltodextrin agglomerates have been produced and characterized, they consist of primary particles with a high sphericity. The tomographic data is processed to detect the agglomerate shape parameters as well as individual primary particle regions. We present a novel algorithm to transform this data into spherical particles that are bonded by cylindrical bonds. We call this method the bonded-particle extraction (BPE) method. This method can be effectively integrated in various DEM simulation systems, where the bonded-particle model is employed. The method is validated with artificial and experimental data. Using the obtained agglomerate structures, an extended version of the stochastic model proposed in [22] has been fitted to experimental data. A validation shows a high agreement with respect to various structural features. Advantages of our approach are:

1) Parametric representation of real agglomerate structures with bonded (spherical) particles allows a more direct calibration of DEM model parameters and/or validation of DEM simulation results.

2) The stochastic modeling technique provides the possibility of generating as many agglomerate structures as desired, which are statistically equivalent to those of experimental agglomerates.

3) The stochastic model is parametric. Therefore, individual parameters can be modified to investigate structural effects on the mechanical behavior of agglomerates in a realistic setting.

However, it should be mentioned that the BPE method can be effectively applied only to systems with relative high sphericity of primary particles. Application of this method for strongly deformed primary particles can lead to a large discrepancy between the generated DEM model and reality.



(a) Experimental spherical agglomerates.

(b) Simulated spherical agglomerates.

(c) Experimental cylindrical agglomerates.

(d) Simulated cylindrical agglomerates.

Figure 16: Scatter plots of bond radii versus bond lengths: obtained from BPAs of experimental data and simulated data for both spherical and cylindrical agglomerates.

An obvious next step is to perform DEM simulations for bonded-particle approximations (BPAs) of real agglomerates. The results obtained by DEM have to be compared to those of real experiments. A problem is that some parameters required in DEM simulations (like bond strengths) are hard to measure or calibrate. Furthermore, the effect of the approximation by spherical particles and cylindrical bonds has to be investigated. These are non-trivial problems and subject to further research.

Appendix A. Formulas for bond volumes and intersections

Some comments on analytical formulas for bond and intersection volumes might be helpful regarding the (fast) evaluation of the cost function defined in Eq. (2).

Appendix A.1. Volume of cylindrical bonds

It is required to compute the volume of a bond $V_{\text{bond}}(p_i, p_j, r_{i,j})$ connecting two particles p_i and p_j. The distance between the two particle centers is denoted by d. The bond radius must not be larger than the particle radii, *i.e.*, $r_{i,j} \leq \max\{r_i, r_j\}$. If the particles are non-overlapping ($d \geq r_i + r_j$), the bond volume is given by the formula

$$V_{\text{bond}}(p_i, p_j, r_{i,j}) = \pi r_{i,j}^2 d - \frac{2}{3}\pi(r_i^3 - (r_i^2 - r_{i,j}^2)^{3/2}) - \frac{2}{3}\pi(r_j^3 - (r_j^2 - r_{i,j}^2)^{3/2})\,.$$

However, the case $d < r_i + r_j$ must be considered as well: the particles may overlap during CE optimization. In the trivial case $d = 0$, the volume is zero. Otherwise, if $d > 0$ and $d < r_i + r_j$, the bond volume is given by

$$\begin{aligned} V_{\text{bond}}(p_i, p_j, r_{i,j}) &= \pi r_{i,j}^2 d \\ &\quad - \begin{cases} \frac{2}{3}\pi(\frac{3}{2}r_i^2 d_1 - \frac{1}{2}d_1^3 - (r_i^2 - r_{i,j}^2)^{3/2}) & \text{if } r_i^2 - r_{i,j}^2 \leq d_1^2 \\ \pi r_{i,j}^2 d_1 & \text{if } r_i^2 - r_{i,j}^2 > d_1^2 \end{cases} \\ &\quad - \begin{cases} \frac{2}{3}\pi(\frac{3}{2}r_j^2 d_2 - \frac{1}{2}d_2^3 - (r_j^2 - r_{i,j}^2)^{3/2}) & \text{if } r_j^2 - r_{i,j}^2 \leq d_2^2 \\ \pi r_{i,j}^2 d_2 & \text{if } r_j^2 - r_{i,j}^2 > d_2^2 \end{cases} \end{aligned}$$

with $d_1 = \max\{0, \min\{d, (r_i^2 - r_j^2 + d^2)/2d\}\}$, and $d_2 = d - d_1$.

Appendix A.2. Sphere–sphere intersections

Suppose again there are two spheres p_i and p_j with radii r_i and r_j, and the distance between their centers is d. Their intersection should be non-empty, *i.e.*, $d \leq r_i + r_j$. If $d \leq |r_i - r_j|$, then the smaller sphere is completely contained in the larger sphere, and the intersection volume is given by $\frac{4}{3}\pi \min\{r_i, r_j\}^3$. Otherwise, the volume of the intersection is given by

$$V_{\text{intersection}}(p_i, p_j) = \pi(r_i + r_j - d)^2(d^2 + 2dr_j - 3r_j^2 + 2dr_i + 6r_i r_j - 3r_i^2)/12d\,.$$

Note that this formula applies as well for the computation of $V_{\text{intersection}}(p_i, S_\ell)$ if the agglomerate shape S_ℓ is spherical.

Appendix A.3. Sphere–cylinder intersections

To compute the volume $V_{\text{intersection}}(p_i, S_\ell)$ of the intersection of a sphere p_i with a cylindrical agglomerate, a general method to compute sphere–cylinder intersection volumes is required.

One possibility is to use numerical integration. In particular, by integrating along the axis of the cylinder, the problem reduces to computing the one-dimensional integral over circle–circle intersection areas.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft under grant number SCHM 997/14-2 in the priority program 1679 “Dynamische Simulation vernetzter Feststoffprozesse”.

References

[1] P. A. Cundall, O. D. L. Strack, A discrete numerical model for granular assemblies, Géotechnique 29 (1979) 47–65.
[2] D. O. Potyondy, P. A. Cundall, A bonded-particle model for rock, Int. J. Rock Mech. Min. 41 (2004) 1329–1364.
[3] P. Thomas, J. Bray, Capturing nonspherical shape of granular media with disk clusters, J. Geotech. Geoenviron. 125 (3) (1999) 169–178.
[4] J. F. Favier, M. H. Abbaspour-Fard, M. Kremmer, A. O. Raji, Shape representation of axisymmetrical, non-spherical particles in discrete element simulation using multi-element model particles, Eng. Computations 16 (4) (1999) 467–480.
[5] R. P. Jensen, P. J. Bosscher, M. E. Plesha, T. B. Edil, DEM simulation of granular media–structure interface: effects of surface roughness and particle shape, Int. J. Numer. Anal. Met. 23 (6) (1999) 531–547.
[6] H. Kruggel-Emden, S. Rickelt, S. Wirtz, V. Scherer, A study on the validity of the multi-sphere Discrete Element Method, Powder Technol. 188 (2) (2008) 153–165.
[7] C. O’Sullivan, Particulate Discrete Element Modelling: A Geomechanics Perspective, Spon Press, Abingdon, 2011.
[8] S. Antonyuk, S. Palis, S. Heinrich, Breakage behaviour of agglomerates and crystals by static loading and impact, Powder Technol. 206 (2011) 88–98.
[9] X. Fu, M. Dutt, A. C. Bentham, B. C. Hancock, R. E. Cameron, J. A. Elliott, Investigation of particle packing in model pharmaceutical powders using X-ray microtomography and discrete element method, Powder Technol. 167 (3) (2006) 134–140.
[10] S. Kozhar, M. Dosta, S. Antonyuk, S. Heinrich, U. Bröckel, DEM simulations of amorphous irregular shaped micrometer-sized titania agglomerates at compression, Adv. Powder Technol. 26 (3) (2015) 767–777.
[11] L. Farber, G. Tardos, J. N. Michaels, Use of X-ray tomography to study the porosity and morphology of granules, Powder Technol. 132 (2003) 57–63.
[12] M. Dadkhah, M. Peglow, E. Tsotsas, Characterization of the internal morphology of agglomerates produced in a spray fluidized bed by X-ray tomography, Powder Technol. 228 (2012) 349–358.
[13] M. Dadkhah, E. Tsotsas, Influence of process variables on internal particle structure in spray fluidized bed agglomeration, Powder Technol. 258 (2014) 165–173.
[14] M. Dadkhah, E. Tsotsas, Study of the morphology of solidified binder in spray fluidized bed agglomerates by X-ray tomography, Powder Technol. 264 (2014) 256–264.
[15] L. Wang, J.-Y. Park, Y. Fu, Representation of real particles for DEM simulation using X-ray tomography, Constr. Build. Mater. 21 (2007) 338–346.
[16] P.-Y. F. Robin, C. R. J. Charles, Quantifying the three-dimensional shapes of spheroidal objects in rocks imaged by tomography, J. Struct. Geol. 77 (2015) 1–10.
[17] X. Yang, Z. You, C. Jin, H. Wang, Aggregate representation for mesostructure of stone based materials using a sphere growth model based on realistic aggregate shapes, Mater. Struct. 49 (6) (2016) 2493–2508.
[18] C. Jin, X. Yang, Z. You, Automated real aggregate modelling approach in discrete element method based on X-ray computed tomography images, Int. J. Pavement Eng. (published online), DOI: 10.1080/10298436.2015.1066006.
[19] Y. Lee, C. Fang, Y.-R. Tsou, L.-S. Lu, C.-T. Yang, A packing algorithm for three-dimensional convex particles, Granular Matter 11 (5) (2009) 307–315.
[20] J. E. Andrade, K.-W. Lim, C. F. Avila, I. Vlahinić, Granular element method for computational particle mechanics, Computer Methods in Applied Mechanics and Engineering 241–244 (2012) 262–274.
[21] G. W. Delaney, T. Di Matteo, T. Aste, Combining tomographic imaging and DEM simulations to investigate the structure of experimental sphere packings, Soft Matter 6 (2010) 2992–3006.
[22] A. Spettl, M. Dosta, S. Antonyuk, S. Heinrich, V. Schmidt, Statistical investigation of agglomerate breakage based on combined stochastic microstructure modeling and DEM simulations, Adv. Powder Technol. 26 (3) (2015) 1021–1030.
[23] H. Wadell, Volume, shape and roundness of quartz particles, J. Geol. 43 (3) (1935) 250–280.
[24] B. Jähne, Digital Image Processing, 6th Edition, Springer, Berlin, 2005.
[25] J. Hoshen, R. Kopelman, Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm, Phys. Rev. B 14 (1976) 3438–3445.
[26] W. Burger, M. J. Burge, Digital Image Processing - An Algorithmic Introduction using Java, Springer, Berlin, 2008.
[27] R. O. Duda, P. E. Hart, Use of the Hough transformation to detect lines and curves in pictures, Commun. ACM 15 (1) (1972) 11–15.
[28] D. H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recogn. 13 (2) (1981) 111–122.
[29] F. P. Preparata, M. I. Shamos, Computational Geometry: An Introduction, Springer, New York, 1985.
[30] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls, ACM T. Math. Software 22 (4) (1996) 469–483.
[31] P. Soille, Morphological Image Analysis: Principles and Applications, 2nd Edition, Springer, Berlin, 2004.
[32] S. Beucher, C. Lantuéjoul, Use of watersheds in contour detection, in: Proceedings of the International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation, Rennes, France, 1979, pp. 2.1–2.12.
[33] S. Beucher, F. Meyer, The morphological approach to segmentation: the watershed transformation, in: E. R. Dougherty (Ed.), Mathematical Morphology in Image Processing, Marcel Dekker, New York, 1993, pp. 433–481.
[34] J. B. T. M. Roerdink, A. Meijster, The watershed transform: definitions, algorithms, and parallellization strategies, Fundam. Inform. 41 (2001) 187–228.
[35] R. Beare, G. Lehmann, The watershed transform in ITK - discussion and new developments, The Insight Journal, Paper ID: 92.
[36] J. C. Russ, The Image Processing Handbook, 5th Edition, CRC Press, Boca Raton, 2007.
[37] J. Lindblad, Surface area estimation of digitized 3D objects using weighted local configurations, Image Vision Comput. 23 (2005) 111–122.
[38] R. Y. Rubinstein, D. P. Kroese, The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning, Springer, New York, 2004.
[39] D. P. Kroese, S. Porotsky, R. Y. Rubinstein, The cross-entropy method for continuous multi-extremal optimization, Methodol. Comput. Appl. Probab. 8 (3) (2006) 383–407.
[40] D. P. Kroese, T. Taimre, Z. I. Botev, Handbook of Monte Carlo Methods, J. Wiley & Sons, Hoboken, 2011.
[41] Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, P. L’Ecuyer, The cross-entropy method for optimization, in: V. Govindaraju, C. R. Rao (Eds.), Machine Learning: Theory and Applications, North-Holland, Oxford, 2013, pp. 35–59.
[42] A. Spettl, T. Brereton, Q. Duan, T. Werz, C. E. Krill III, D. P. Kroese, V. Schmidt, Fitting Laguerre tessellation approximations to tomographic image data, Philos. Mag. 96 (2) (2016) 166–189.
[43] A. Bezrukov, M. Bargieł, D. Stoyan, Statistical analysis of simulated random packings of spheres, Part. Part. Syst. Char. 19 (2) (2002) 111–118.
[44] J. Mościński, M. Bargieł, Z. A. Rycerz, P. W. M. Jacobs, The force-biased algorithm for the irregular close packing of equal hard spheres, Mol. Simulat. 3 (4) (1989) 201–212.
[45] X. Jia, M. Gan, R. A. Williams, D. Rhodes, Validation of a digital packing algorithm in predicting powder packing densities, Powder Technol. 174 (1–2) (2007) 10–13.

[46] G. Gaiselmann, O. Stenzel, A. Kruglova, F. Mücklich, V. Schmidt, Competitive stochastic growth model for the 3D morphology of eutectic Si in Al-Si alloys, Comp. Mater. Sci. 69 (2013) 289–298.
[47] M. Roland, A. Kruglova, G. Gaiselmann, T. Brereton, V. Schmidt, F. Mücklich, S. Diebels, Numerical simulation and comparison of a real Al-Si alloy with virtually generated alloys, Arch. Appl. Mech. 85 (8) (2015) 1161–1171.
[48] B. W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall/CRC, London, 1986.
[49] M. P. Wand, M. C. Jones, Kernel smoothing, Chapman & Hall/CRC, New York, 1995.
[50] G. Casella, R.-L. Berger, Statistical Inference, 2nd Edition, Duxbury, Pacific Grove, 2002.