

Copula-based approximation of particle breakage as link between DEM and PBM

Aaron Spettl^{a,*}, Maksym Dosta^b, Frederik Klingner^a, Stefan Heinrich^b, Volker Schmidt^a

^aInstitute of Stochastics, Ulm University, Germany

^bInstitute of Solids Process Engineering and Particle Technology, Hamburg University of Technology, Germany

Abstract

In process engineering, the breakage behavior of particles is needed for the modeling and optimization of comminution processes. A popular tool to describe (dynamic) processes is population balance modeling (PBM), which captures the statistical distribution of particle properties and their evolution over time. It has been suggested previously to split up the description of breakage into a machine function (modeling of loading conditions) and a material function (modeling of particle response to mechanical stress). Based on this idea, we present a mathematical formulation of machine and material functions and a general approach to compute them. Both functions are modeled using multivariate probability distributions, where in particular so-called copulas are helpful. These can be fitted to data obtained by the discrete element method (DEM). In this paper, we describe the proposed copula-based breakage model, and we construct such a model for an artificial dataset that allows to evaluate the prediction quality.

Keywords: discrete element method, population balance modeling, breakage probability, breakage function, copula

1. Introduction

Population balances are a widely used tool in engineering, especially in the field of particulate materials. They describe disperse properties of entities like particles; these properties are time- and possibly space-dependent (Ramkrishna, 2000). In population balance modeling (PBM) (Ramkrishna, 2000; Ramkrishna and Mahoney, 2002), the aim is to describe processes like, *e.g.*, crystallization or comminution by suitable population balance equations (see, *e.g.*, Briesen, 2006; Bilgili and Scarlett, 2005). These equations model the change in the number of particles with a given property; they are (partial integro-)differential equations. In such a setting, it is clear that a model-based description of particle breakage is required for all processes where breakage occurs. Breakage frequencies may be seen as a functional of the properties of the individual particle and the loading conditions. This separation goes back to Rumpf (1967) and it is stated more precisely in Peukert and Vogel (2001), where the process function is controlled by a machine function and a material function. In the case of comminution, this means that the machine function specifies the loading conditions (kind of stress, number of stress events and stress intensity), and the material function describes how a particle reacts to a given stress event. Combining these two functions leads to the description of breakage on the apparatus-scale.

A complementary but very different approach is the discrete element method (DEM) (Cundall and Strack, 1979; O'Sullivan, 2011). Individual particles are considered explicitly and contact models describe how they interact with each other. However,

this technique is computationally expensive, in particular for large-scale simulations. Yet, this method can be used to investigate the loading conditions in processes, and it is suitable to determine the breakage behavior of agglomerated particles. Quantitative information gained in this way can be used as input to PBM, see, *e.g.*, Freireich et al. (2011); Dosta et al. (2012, 2013); Barrasso and Ramachandran (2015).

In this paper, we present a new approach to the stochastic modeling of loading conditions and single-particle breakage behavior, *i.e.*, breakage probability and breakage function. The breakage function is essentially a conditional probability density function (Otwinowski, 2006) — however, this fact has not been exploited so far. We propose to construct multivariate (copula-based) probability distributions in order to use the resulting density functions to derive particle-dependent loading frequencies, breakage probabilities, and breakage functions. This information can then be combined to describe the apparatus-scale breakage behavior. As a consequence, this method provides a link between DEM and PBM, which is much more flexible than existing approaches. Note that the copula-based modeling of multivariate distributions is already applied in various areas. Typical applications are in finance and insurance (McNeil et al., 2005), but copulas are also used in, *e.g.*, climate research (Schölzel and Friederichs, 2008).

The present paper is structured as follows. First, in Section 2, the basic ideas of PBM and DEM are explained. Furthermore, copula-based modeling of multivariate distributions is introduced shortly. In Section 3, the copula-based modeling of loading conditions and particle breakage is explained. Then, we present an example in Section 4. We generate a simple data set, fit copula-based models to the data, and show that the copula-based models are able to predict breakage probabilities and fragment size

*Corresponding author. Tel.: +49 731 50 23555; fax: +49 731 50 23649.

Email address: aaron.spettl@uni-ulm.de (Aaron Spettl)

Nomenclature

$b_{\text{fragm}}(x; y)$	breakage function: number-scaled density function of fragment properties x for original particle y (e.g., x and y could be the particle volume, then [1/mm ³])
C	a copula, see Appendix A
f_{mach}	machine function, see Eq. (5)
f_{mat}	material function, see Eq. (6)
$f_{\text{fragm,int}}$	probability density function of internal properties of a fragment
f_{Δ}	probability density function of a random variable or random vector Δ
$f_{(\Delta \square)}$	(conditional) density function of Δ under condition \square
F_{Δ}	cumulative distribution function of a random variable or random vector Δ
ℓ	loading conditions vector
m_{int}	number of internal particle properties [-]
m_{ext}	number of external particle properties [-]
m_{load}	size of loading conditions vector [-]
$n(x, t)$	number-scaled density function for particle properties x in PBM at time t (e.g., x can be the particle volume, then [1/mm ³])
$n_{\text{DEM}}(x)$	probability density function of particle properties in an apparatus-scale DEM simulation
$N_{\text{DEM-particles}}$	total number of particles in DEM simulation [-]
$N_{\text{DEM-stress}}$	total number of stress events in DEM simulation [-]
N_{fragm}	expected number of fragments of a broken particle [-]
$N_{\text{particles}}(t)$	expected total number of particles at time t [-]
P_{break}	breakage probability [-]
$r_{\text{break}}(x)$	breakage rate of particle with properties x [1/s]
$r_{\text{load}}(x)$	loading frequency [1/s]
x	particle properties (e.g., particle volume, then [mm ³])
x_{int}	internal particle properties
x_{ext}	external particle properties
$(X_{\text{int}}, L_1, \dots, L_{m_{\text{load}}-1}, C_{\text{crit}})$	random vector of internal particle properties X_{int} , loading conditions L without last component and critical threshold C_{crit}
$(\tilde{X}_{\text{int}}, \tilde{L}, X_{\text{fragm,int}})$	random vector of internal fragment properties $X_{\text{fragm,int}}$ and their corresponding original internal particle properties \tilde{X}_{int} and loading conditions \tilde{L}
$(X_{\text{stress}}, L_{\text{stress}})$	random vector of stress events L_{stress} in DEM simulation and their corresponding particle properties X_{stress}

distributions quite well. In Section 5, the results are summarized and an outlook for future work is provided. Finally, Appendix A provides an overview on copulas and typical procedures for fitting such models to sample data.

2. Methods

In this section, we briefly describe the methods required to develop the copula-based breakage models in Section 3. This includes a short description of population balance modeling, the discrete element method, and copula-based multivariate probability distributions.

2.1. Population balance modeling (PBM)

The population balance equation is a differential equation that describes the change over time in the number of particles having certain properties. Particle properties are usually split up into internal and external properties, *i.e.*, $x = (x_{\text{int}}, x_{\text{ext}})$, where both $x_{\text{int}} \in \mathbb{R}^{m_{\text{int}}}$ and $x_{\text{ext}} \in \mathbb{R}^{m_{\text{ext}}}$ are vectors and $m_{\text{int}}, m_{\text{ext}}$ denote the numbers of internal and external properties. Internal properties describe the particle itself, *e.g.*, its size or shape. Particle coordinates are an example for external properties. The dynamic particle system is described by a time-dependent density function n . For fixed time t , $n(x, t)$ describes the distribution of particle properties, and the expected total number of particles $N_{\text{particles}}(t)$ can be obtained from the density by computing

$$N_{\text{particles}}(t) = \int_{\mathbb{R}^{m_{\text{int}}+m_{\text{ext}}}} n(x, t) dx.$$

In the simplest form of PBM, one assumes a well-mixed system of particles (no external properties) and considers only one internal property, *i.e.*, $m_{\text{int}} = 1, m_{\text{ext}} = 0$. Usually, the particle size in some sense (*e.g.*, its diameter, volume, or mass) is used as property coordinate. Then, the population balance equation that describes an aggregation or breakage process can be written as

$$\frac{\partial n(x, t)}{\partial t} = b(x, t) - d(x, t), \quad (1)$$

where $b(x, t)$ and $d(x, t)$ describe birth and death frequencies of particles with property x at time t . For example, inlet and outlet streams can be modeled by these birth and death terms, *i.e.*, $b(x, t) = n_{\text{inlet}}(x, t), d(x, t) = n_{\text{outlet}}(x, t)$. However, both the birth and death terms may depend on the entire current population $n(x, t)$, therefore, in general, Equation (1) is an integro-differential equation. Even in such a simple example, several important effects can be modeled that change the particle population over time: aggregation, breakage, and nucleation. All these phenomena can be described by the “death” of the original particle(s) and “birth” of one or more new particles (that have other properties).

In this paper, the focus lies on particle breakage. Therefore, we consider (1) with respect to particle breakage. Birth and death terms $b(x, t), d(x, t)$ can be obtained by the

standard approach using breakage rate and breakage functions (Kostoglou, 2007). A breakage process can be described by

$$\frac{\partial n(x, t)}{\partial t} = \int_x^\infty b_{\text{fragm}}(x; y) r_{\text{break}}(y) n(y, t) dy - r_{\text{break}}(x) n(x, t), \quad (2)$$

where $r_{\text{break}}(x)$ is the *breakage rate* of a single particle with property x , and b_{fragm} specifies the so-called *breakage functions*. In principle, both may depend on t and n . The object b_{fragm} can be understood as a family $\{b_{\text{fragm}}(x; y), y > 0\}$ of breakage functions, which means that there is an individual breakage function for every original particle’s size. A breakage function $b_{\text{fragm}}(x; y)$ for some original particle y is the probability density function of the fragment size distribution scaled with the expected number of fragments.

It is clear that the breakage behavior depends on the physical (communition) process. Therefore, the main problem in using PBM is to determine suitable breakage frequencies $r_{\text{break}}(x)$ and the breakage functions $\{b_{\text{fragm}}(x; y), y > 0\}$. There are many different approaches in literature. Very often, so-called algorithmic breakage functions are used (Kostoglou, 2007): A family of functions is described by a small number of parameters, and these parameters are estimated based on experimental data. The breakage rate is called homogeneous if $r_{\text{break}}(x) = Kx^a$ for some $K, a > 0$. This means that larger particles have a higher breakage rate, which is frequently observed in breakage processes (Ramkrishna, 2000). The breakage functions are called homogeneous if $b_{\text{fragm}}(x; y) = \theta(x/y)/y$ for some suitable function θ . Most algorithmic breakage functions are homogeneous (Kostoglou, 2007). However, in general, breakage frequencies may depend on internal and external properties of particles as well as on the entire particle population and possibly even on external (time-dependent) process parameters. With the concept of considering machine and material function separately (Peukert and Vogel, 2001), it is possible to model much more complex breakage processes. In this paper, we present a general approach to the modeling of breakage frequencies and breakage functions by construction of suitable machine and material functions.

2.2. Discrete element method

The discrete element method (DEM, see, *e.g.*, Cundall and Strack, 1979; O’Sullivan, 2011) is a computational technique to investigate the dynamics of particles on the microscale. Every particle is represented as a separate object and the forces acting on each particle are evaluated to predict particle movement, rotation, *etc.* based on Newton’s laws of motion. Contact models define the physical laws (or approximations thereof) that are used in the simulation. The simulation is performed in sufficiently short discrete time steps.

DEM is very powerful; however, a problem is that it is hard to investigate large particle systems as necessary for industrial-scale processes. For example, in Torbahn et al. (2016), micron-sized particles in a mm-sized shear-tester are used to compare experimental results directly to those of DEM simulations. For the modeling of larger particle systems,

multi-scale approaches are used frequently. One example is the combination of PBM and DEM (Freireich et al., 2011; Dosta et al., 2013, 2014; Barrasso and Ramachandran, 2015). On the one hand, DEM can be used to understand the loading of particles in processes (apparatus-scale DEM simulation of the process). On the other hand, DEM may be used to investigate the breakage behavior of individual particles (single-particle DEM simulations as performed, e.g., in Spettl et al. (2015) using the bonded-particle model (BPM) presented in Potyondy and Cundall (2004)). Both the single-particle breakage behavior and the loading conditions can then be combined using PBM to describe the process on the macro-scale.

2.3. Multivariate probability distributions and copulas

Recall that a (real-valued) random variable X is described by its cumulative distribution function $F_X(x) = \mathbb{P}(X \leq x)$ for $x \in \mathbb{R}$. Distribution functions are often specified by parametric models like, e.g., the normal distribution, the gamma distribution or the uniform distribution. Now, a more general case is considered. If X is an m -dimensional random vector, its (joint) distribution function is defined by setting

$$F_X(x_1, \dots, x_m) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m), \quad x_1, \dots, x_m \in \mathbb{R}.$$

Analogously to the univariate case, the distribution of the random vector X is called absolutely continuous if there exists a density function $f_X : \mathbb{R}^m \rightarrow [0, \infty)$ such that

$$F_X(x_1, \dots, x_m) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f_X(z_1, \dots, z_m) dz_m \cdots dz_1,$$

for all $x_1, \dots, x_m \in \mathbb{R}$. Note that a probability density f is always normalized to $\int_{\mathbb{R}^m} f(x) dx = 1$. In this paper, we also use unnormalized density functions, e.g., for some time t , $n(x, t)$ is a density, which integrates to the number of particles.

However, there are few parametric models for the description of multivariate distributions. The reason is simple: parametric models with a reasonable number of parameters are often not capable to describe real-world data — such models are not flexible enough. For example, think about a two-dimensional random vector $X = (X_1, X_2)$. Even if both components (also called marginals) X_1 and X_2 are independent, there are many possibilities for different combinations of the univariate distributions of X_1 and X_2 . The two marginal distributions may be chosen from different families of distributions (normal, log-normal, gamma, uniform, ...) or may have different parameters. Even in that simple example, it becomes obvious that higher-dimensional modeling approaches make a high level of flexibility necessary.

A way out of this problem is to model every marginal distribution separately. This can be done with classical methods from statistics (see, e.g., Casella and Berger, 2002). However, the question remains how these marginal distributions can be recombined to form the joint distribution. This is far from being trivial, because, in general, the X_i , $i = 1, \dots, m$, are not independent. The question is answered by the relationship

$$F_X(x_1, \dots, x_m) = C(F_{X_1}(x_1), \dots, F_{X_m}(x_m)), \quad x_1, \dots, x_m \in \mathbb{R}, \quad (3)$$

which holds for any random vector X , where, of course, the choice of the function $C : [0, 1]^m \rightarrow [0, 1]$ depends on the distribution of X . In this formula, F_{X_i} denotes the distribution function of the i th marginal X_i , and C is a so-called *copula*. In other words, Equation (3) states that for every random vector X , it is possible to split up the complexity of its (joint) distribution function into marginals and the dependencies between the marginals. The dependencies are described by the copula C .

The copula C itself is a multivariate distribution function with marginal distributions that are all uniform distributions on $[0, 1]$. If C is selected from a parametric family of copulas, the parameters specifying C describe the dependence structure of X . For example, the copula corresponding to a bivariate normal distribution is called (two-dimensional) Gaussian copula and it has one single parameter $\rho \in [-1, 1]$, which is the correlation coefficient. Figure 1(top-left) shows samples of the form (x_1, x_2) obtained using a bivariate normal distribution with correlation coefficient $\rho = 0.7$. A sample (x_1, x_2) can be transformed to a sample of the copula by setting $(u_1, u_2) = (F_{X_1}(x_1), F_{X_2}(x_2)) \in [0, 1]^2$. The corresponding transformed samples describing the copula are visualized in Figure 1(top-right), and histograms for the two marginal distributions are given in Figure 1(bottom). We can see that the data can be split up into data describing only the dependence and data describing only the marginals. Vice versa, fitting parametric models to both the data of the marginals and the copula, the joint distribution of X can be reconstructed. This is the basic idea of copula-based modeling of multivariate distributions.

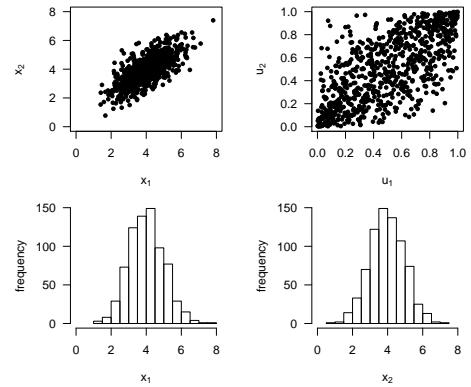


Figure 1: 200 samples of a bivariate normal distribution with expectation vector $(4, 4)$, variances $(1, 1)$ and correlation coefficient 0.7 : scatter plot of samples in the form (x_1, x_2) (top-left), scatter plot of pseudo-observations (vectors (x_1, x_2) transformed to (u_1, u_2) , cf. Appendix A.3, top-right), and histograms of marginals (bottom).

Summarizing, copulas are a tool for modeling the joint distribution function of random vectors. In particular, they provide an easy method to construct the multivariate distribution function by splitting up the complexity into several (less complex) sub-problems. The models of particle breakage presented in this paper are based on multivariate distributions, which are constructed using copulas. More details regarding the basics of copulas, some important parametric copula families and their fitting to data can be found in Appendix A and, e.g.,

Mai and Scherer (2012); Joe (2015).

3. Theory

A very general idea for modeling of particle breakage in processes is given by the approach of Peukert and Vogel (2001). The comminution process is described by a machine function and a material function. The machine function specifies the loading frequency and the loading conditions (kind of stress, stress intensity). These depend on the type of the mill and its operation parameters, which may change over time. Complementary, the material function describes how a particle will react to a given stress event.

It is useful to formalize this approach by defining the machine and material functions explicitly, *i.e.*, with the help of mathematical functions. This is done in Section 3.1. Their link to DEM simulations is described in Section 3.2. Then, in Sections 3.3 and 3.4, we explain how the machine and material functions can be specified based on techniques from probability calculus — in particular, by use of multivariate distributions. Finally, in Section 3.5, we state how machine and material functions can be recombined in order to describe the breakage behavior of the process.

3.1. Machine and material functions

We introduce a possible mathematical definition of **machine functions**. Recall that the particle properties are specified by internal and external properties as done in PBM, where the internal properties x_{int} are given as an m_{int} -dimensional numerical vector, the external properties x_{ext} as an m_{ext} -dimensional vector, and together they are denoted as $x = (x_{\text{int}}, x_{\text{ext}})$. For particles with properties vector x , a straightforward way would be to describe the loading frequency r_{load} and the (m_{load} -dimensional) loading conditions vector ℓ by a mapping

$$(r_{\text{load}}, \ell) = f_{\text{mach}}(x, p). \quad (4)$$

For particles with properties x , the function yields how often such particles are stressed individually. This means that r_{load} is the loading frequency of each particle with properties x , and ℓ is a vector containing information on the loading conditions like type of stress event (static loading, dynamic impact, *etc.*) and stress intensities (*e.g.*, acting forces, impact energy). The parameter vector p may be used to provide necessary information on the current operation parameters of the machine (*e.g.*, rotational velocity for ball mill, roll gap in roller mill).

However, in real processes all particles with the same properties x are not stressed equally. Therefore, we replace the loading conditions vector ℓ by a “distributed” vector. In particular, one can think of ℓ as being a random vector, which can be described by an m_{load} -dimensional probability density. This leads to our final definition of the machine function, which is as follows. In Equation (4), the vector ℓ is replaced by a probability density function $f_{\text{load}}(\ell)$, which essentially describes how often a certain ℓ is encountered. This leads to the mapping

$$(r_{\text{load}}, f_{\text{load}}) = f_{\text{mach}}(x, p), \quad (5)$$

where we write f_{load} without parameter to emphasize that the complete function is the returned “value” (*i.e.*, it is not only the function evaluated at some specific ℓ).

Material functions are constructed to return the expected behavior of a particle under a certain load (which we also call *single-particle breakage model*). In particular, the function takes the internal particle properties and the loading conditions as an argument and returns

$$(p_{\text{break}}, f_{\text{fragm,int}}, N_{\text{fragm}}) = f_{\text{mat}}(x_{\text{int}}, \ell), \quad (6)$$

where p_{break} is the breakage probability, $f_{\text{fragm,int}}$ is an (m_{int} -dimensional) probability density function that describes the internal properties $x_{\text{fragm,int}}$ of the fragments, and $N_{\text{fragm}} \geq 1$ is the expected number of fragments.

3.2. Multi-scale DEM simulations

Both the machine and material functions have to be determined using real experiments or computer simulations. Detailed simulations like DEM are ideal because all information can be accessed like, *e.g.*, the collisions in apparatus-scale simulations. Figure 2 illustrates the general approach. DEM simulations on the apparatus-scale are performed for realistic conditions and particle dynamics and their interactions are recorded. On the scale of single particles experimental or numerical investigations can be performed to obtain the material function.

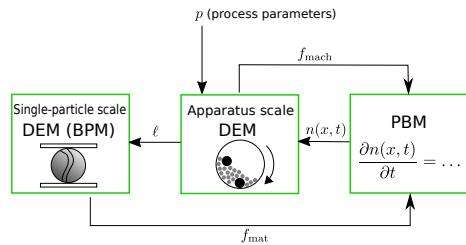


Figure 2: Schematic illustration of the modeling procedure for PBM.

The stress events from DEM calculations on the apparatus-scale can be employed as the main tool to predict the machine function. A stress event of a specific particle appears when this particle interacts with any other type of object. Therefore, a collision between particle and wall will result in one new stress event in the system. A collision between two particles will result in two events. Each stress event may be described by several parameters, *e.g.*, one being the stress intensity. These parameters are described by a *loading conditions vector* in the following. Naturally, different collision types (like particle–particle or particle–wall) have to be taken into account in a realistic model. There are two approaches to include these types of stress events. For one, it is possible to encode this information as a categorical variable in one entry of the loading conditions vector. This is the direct way. An alternative is to establish several machine functions — *i.e.*, split up the data in order to have one machine function as explained in Section 3.3 for every type of stress event. This leads to several independent birth and death terms in Equation (1).

In some cases, due to the high computational effort, the real distribution of particles cannot be directly simulated with DEM. For example, to reduce the number of simulated particles, their size can be increased with a scaling approach (Sutkar et al., 2013). Another problem is the modeling of wide particle size distributions, which cannot be effectively done. Therefore, the density distribution of particles in DEM, $n_{\text{DEM}}(x)$, can be different from $n(x, t)$. In that case additional extrapolation of data obtained from DEM is needed. However, in this paper, we consider the case when $n_{\text{DEM}}(x) \approx n(x, t)$ (up to a scaling constant; recall that $n(x, t)$ is not normalized to be a probability density). This means that the data of stress events from DEM can be transferred directly to the PBM approach. Note that it is hard to include breakage in the DEM simulations due to the resulting wide size distributions and large number of particles. This is exactly the reason why it is useful to only extract information on stress events from DEM and use these in a PBM approach.

3.3. Machine function: modeling of loading conditions

We can model the machine function f_{mach} by the following construction. For every particle with properties x , we model (a) the specific loading frequency r_{load} and (b) the distribution of loading conditions f_{load} . The construction is based on data that can be obtained by DEM simulations. A DEM simulation on the apparatus-scale yields information on stress events (*cf.* Section 3.2 and Figure 2). Recall that, in this paper, we interpret a collision between two particles as two stress events, one for each particle.

For convenience, we make some simplifying assumptions. We assume that we have a stationary regime, *i.e.*, loading frequencies r_{load} do not depend on time, and that the distribution of particle properties used in DEM is representative for the investigated process. From the apparatus-scale DEM simulations we learn which particle is stressed how and how often — in particular, we have a set of vectors $(x_{\text{stress}}^{(i)}, \ell_{\text{stress}}^{(i)})$ indexed by i . These are understood as a sample of a random vector $(X_{\text{stress}}, L_{\text{stress}})$. The meaning is as follows. The random vector X_{stress} denotes the $(m_{\text{int}} + m_{\text{ext}})$ -dimensional random vector of particle properties of a (random) stressed particle, and L_{stress} is the corresponding m_{load} -dimensional random vector of loading conditions. Therefore, the distribution of X_{stress} is not a standard, number-weighted distribution of particle properties in the system — rather, it is the distribution of particle properties weighted with their respective number of stress events. The model is based on describing the joint distribution of $(X_{\text{stress}}, L_{\text{stress}})$ by specifying a joint density function, which can be approximated based on the sample data from DEM simulations. Furthermore, we can easily determine the mean number of stress events per unit time interval occurring in whole apparatus, denoted by $N_{\text{DEM-stress}}$.

First, we consider the **loading frequency**. The aim is to determine the loading frequency $r_{\text{load}}(x)$ of an individual particle (which is different from $f_{X_{\text{stress}}}(x)$ because the latter must be interpreted with respect to all particles present in the DEM simulation). The probability density n_{DEM} describes the distribution of the particles that are present in the DEM

simulation, and $f_{X_{\text{stress}}}$ describes how often particles with properties vector x are stressed. The sought loading frequency $r_{\text{load}}(x)$ must fulfill

$$f_{X_{\text{stress}}}(x) \propto r_{\text{load}}(x) n_{\text{DEM}}(x), \quad (7)$$

i.e., the left- and right-hand sides must be proportional to each other. Recall that both n_{DEM} and $f_{X_{\text{stress}}}$ are probability density functions, which means that they integrate to unity. In order to compute r_{load} from (7), we need to determine the proportionality factor. Because the number of particles in the DEM system scaled with the respective loading frequencies should equal the total number of stress events (per unit time interval), we know that

$$\int_{\mathbb{R}^{m_{\text{int}}+m_{\text{ext}}}} r_{\text{load}}(x) n_{\text{DEM}}(x) N_{\text{DEM-particles}} dx = N_{\text{DEM-stress}}$$

where $N_{\text{DEM-particles}}$ is the total number of particles simulated in DEM. Therefore, with $c = N_{\text{DEM-stress}}/N_{\text{DEM-particles}}$, we have

$$f_{X_{\text{stress}}}(x) = \frac{1}{c} r_{\text{load}}(x) n_{\text{DEM}}(x).$$

A simple rearrangement of terms in the equations stated above leads to the formula

$$r_{\text{load}}(x) = \frac{N_{\text{DEM-stress}} \times f_{X_{\text{stress}}}(x)}{N_{\text{DEM-particles}} \times n_{\text{DEM}}(x)}.$$

Note that the ratio $f_{X_{\text{stress}}}(x)/n_{\text{DEM}}(x)$ will equal unity if all individual particles in the system are stressed equally often regardless of their properties because, in that case, both probability density functions would be identical. Then, only the ratio $N_{\text{DEM-stress}}/N_{\text{DEM-particles}}$ remains, where it can be easily seen that $r_{\text{load}}(x)$ is just the mean number of stress events per particle in a unit-length time interval.

Based on the random vector $(X_{\text{stress}}, L_{\text{stress}})$, the **distribution of loading conditions** can be predicted by evaluation of the (conditional) density of $(L_{\text{stress}} \mid X_{\text{stress}} = x)$. For x with $f_{X_{\text{stress}}}(x) > 0$, the conditional density $f_{(L_{\text{stress}}|X_{\text{stress}}=x)}$ is defined as

$$f_{(L_{\text{stress}}|X_{\text{stress}}=x)}(\ell) = f_{(X_{\text{stress}}, L_{\text{stress}})}(x, \ell) / f_{X_{\text{stress}}}(x), \quad \ell \in \mathbb{R}^{m_{\text{load}}}.$$

Putting everything together, by knowledge of r_{load} and $f_{(X_{\text{stress}}, L_{\text{stress}})}$, the machine function f_{mach} can be defined as

$$f_{\text{mach}}(x) = (r_{\text{load}}(x), f_{(L_{\text{stress}}|X_{\text{stress}}=x)}), \quad x \in \mathbb{R}^{m_{\text{int}}+m_{\text{ext}}}.$$

In this formula, we did not include the parameter p , which describes the process conditions. However, it should be mentioned that the results obtained from apparatus-scale DEM also depend on p . Note that, in a more sophisticated model, one should use a set of DEM apparatus-scale simulations to gain knowledge for different operation conditions.

3.4. Material function: single-particle breakage model

The material function f_{mat} , see Equation (6), maps the internal particle properties and loading conditions to the

breakage probability, fragment properties density function, and expected number of fragments.

The **breakage probability** p_{break} can be obtained as follows. Let $(x_{\text{int}}, \ell) \in \mathbb{R}^{m_{\text{int}}} \times \mathbb{R}^{m_{\text{load}}}$ be the vector that is given as input. The standard approach would be to predict p_{break} directly from (x_{int}, ℓ) — *e.g.*, by fitting a surface to data obtained from single-particle DEM simulations. However, depending on the data, a splitting approach can be applied more effectively. The idea is to consider one of the loading parameters (in this paper, it will be the last parameter $\ell_{m_{\text{load}}}$) separately from the vector $(x_{\text{int}}, \ell_1, \dots, \ell_{m_{\text{load}}-1})$. In this case, for every $(x_{\text{int}}, \ell_1, \dots, \ell_{m_{\text{load}}-1})$ there exists a one dimensional function that depends on $\ell_{m_{\text{load}}}$ and returns the breakage probability. Such a function can be constructed as follows. Suppose there is some random variable C_{crit} that describes the critical threshold for the last component, *i.e.*, the starting point for $\ell_{m_{\text{load}}}$ where the particle breaks. Then, the value $F_{C_{\text{crit}}}(\ell_{m_{\text{load}}})$ of the cumulative distribution function of C_{crit} returns the probability that the critical threshold is at most $\ell_{m_{\text{load}}}$. Therefore, we only need to predict the critical threshold C_{crit} (which is influenced by $(x_{\text{int}}, \ell_1, \dots, \ell_{m_{\text{load}}-1})$).

This can be implemented as follows. For every vector $(x_{\text{int}}, \ell_1, \dots, \ell_{m_{\text{load}}-1})$, there is some threshold c_{crit} that can be determined with DEM. There exist different types of characteristic types like octahedral shear stress, major principal stress, *etc.* which can be used as breakage criteria. An overview and a comparison between them can be found in De Bono and McDowell (2016). This threshold is essentially some critical stress intensity, *e.g.*, the energy required for breakage. We assume $(x_{\text{int}}, \ell_1, \dots, \ell_{m_{\text{load}}-1}, c_{\text{crit}})$ to be a realization of a random vector

$$(X_{\text{int}}, L_1, \dots, L_{m_{\text{load}}-1}, C_{\text{crit}}).$$

The distribution of $(X_{\text{int}}, L_1, \dots, L_{m_{\text{load}}-1}, C_{\text{crit}})$ can be obtained by fitting to data of single-particle DEM simulations (*cf.* Section 4.3), where stressing of particles is simulated, and the critical stress intensities required for breakage are recorded. Then, the breakage probability is directly given by

$$\begin{aligned} p_{\text{break}} &= \mathbb{P}(C_{\text{crit}} \leq \ell_{m_{\text{load}}} \mid X_{\text{int}} = x_{\text{int}}, L_1 = \ell_1, \dots, L_{m_{\text{load}}-1} = \ell_{m_{\text{load}}-1}) \\ &= F_{C_{\text{crit}}|X_{\text{int}}=x_{\text{int}}, L_1=\ell_1, \dots, L_{m_{\text{load}}-1}=\ell_{m_{\text{load}}-1}}(\ell_{m_{\text{load}}}). \end{aligned}$$

where $F_{C_{\text{crit}}|X_{\text{int}}=x_{\text{int}}, L_1=\ell_1, \dots, L_{m_{\text{load}}-1}=\ell_{m_{\text{load}}-1}}$ denotes the (conditional) cumulative distribution function of $(C_{\text{crit}} \mid X_{\text{int}} = x_{\text{int}}, L_1 = \ell_1, \dots, L_{m_{\text{load}}-1} = \ell_{m_{\text{load}}-1})$.

The same idea works for the **distribution of fragment properties**. Let $X_{\text{fragm,int}}$ be the random vector describing the properties of single fragments in the same manner as the internal properties $x_{\text{int}} \in \mathbb{R}^{m_{\text{int}}}$. The distribution of $X_{\text{fragm,int}}$ should be predicted from (x_{int}, ℓ) . Similar to the modeling of the critical stress intensity, the joint distribution of the random vector

$$(\tilde{X}_{\text{int}}, \tilde{L}, X_{\text{fragm,int}}),$$

should be modeled. Then, the distribution of the conditional random vector

$$(X_{\text{fragm,int}} \mid \tilde{X}_{\text{int}} = x_{\text{int}}, \tilde{L} = \ell) \quad (8)$$

provides all information on the fragments. A technical detail is that the distribution of $(\tilde{X}_{\text{int}}, \tilde{L})$ is not equal to that of (X_{int}, L) from above — the reason is simple: for the distribution of fragment properties, the original particle is weighted with the number of fragments it produces.

Ideally, the modeling of $(\tilde{X}_{\text{int}}, \tilde{L}, X_{\text{fragm,int}})$ already makes sure that a fragment may not be larger than the original particle. However, in practice, it is cumbersome to ensure that this is the case with probability 1. It is much more convenient to describe the distribution of $(\tilde{X}_{\text{int}}, \tilde{L}, X_{\text{fragm,int}})$ without this constraint. Then, by fitting the distribution to realistic data, a fragment may be too large, but this happens only with a small probability. This is a problem that is then solved by only considering $X_{\text{fragm,int}}$ conditioned on fragment sizes being small enough. Let the first component $X_{\text{fragm,int},1}$ of $X_{\text{fragm,int}}$ as well as the first component $X_{\text{int},1}$ of X_{int} denote the size of the random fragment and of the original (random) particle, respectively. We require that $X_{\text{fragm,int},1} < X_{\text{int},1}$ with probability 1. This is achieved by considering

$$X'_{\text{fragm,int}} = (X_{\text{fragm,int}} \mid X_{\text{fragm,int},1} < X_{\text{int},1})$$

instead of $X_{\text{fragm,int}}$. Therefore, the distribution of

$$(X'_{\text{fragm,int}} \mid \tilde{X}_{\text{int}} = x_{\text{int}}, \tilde{L} = \ell)$$

is used to predict fragment properties. Impossible outcomes are simply rejected. Note that, if sizes are specified using solid volumes or masses, by conservation of mass, it is clear that the expected number of fragments from a single breakage event is given by $N_{\text{fragm}}(x_{\text{int}}, \ell) = x_{\text{int},1}/\mathbb{E}(X'_{\text{fragm,int},1} \mid \tilde{X}_{\text{int}} = x_{\text{int}}, \tilde{L} = \ell)$, *i.e.*, the mean number of fragments can be determined by the original particle size and the mean fragment size.

Summarizing, the material function f_{mat} is given by

$$f_{\text{mat}}(x_{\text{int}}, \ell) = (p_{\text{break}}(x_{\text{int}}, \ell), f_{\text{fragm},(x_{\text{int}}, \ell)}, N_{\text{fragm}}(x_{\text{int}}, \ell)),$$

for $x_{\text{int}} \in \mathbb{R}^{m_{\text{int}}}$, $\ell \in \mathbb{R}^{m_{\text{load}}}$, with breakage probability

$$p_{\text{break}}(x_{\text{int}}, \ell) = F_{C_{\text{crit}}|X_{\text{int}}=x_{\text{int}}, L_1=\ell_1, \dots, L_{m_{\text{load}}-1}=\ell_{m_{\text{load}}-1}}(\ell_{m_{\text{load}}})$$

and fragment properties density function

$$f_{\text{fragm},(x_{\text{int}}, \ell)}(x_{\text{fragm,int}}) = f_{(X'_{\text{fragm,int}} \mid \tilde{X}_{\text{int}} = x_{\text{int}}, \tilde{L} = \ell)}(x_{\text{fragm,int}}),$$

for all $x_{\text{fragm,int}} \in \mathbb{R}^{m_{\text{int}}}$.

3.5. Linking material and machine functions together: apparatus-scale breakage model

The single-particle breakage behavior is described by the material function, and the loading frequencies and loading conditions are specified by the machine function. Both can be combined to obtain an apparatus-scale breakage model. On the apparatus scale, we need the distributions of breakage frequencies and fragment properties, where both must only depend on the particle properties. In particular, both do not depend on the loading frequencies or loading conditions because this information is already included by an appropriate averaging.

Let $x \in \mathbb{R}^{m_{\text{int}}+m_{\text{ext}}}$ denote some particle properties. By applying the machine function f_{mach} , we obtain the probability density function f_{load} , and the loading frequency r_{load} . Then, we can compute the breakage probability (averaged over all loading conditions) by evaluating

$$p_{\text{break}}(x) = \int_{\mathbb{R}^{m_{\text{load}}}} p_{\text{break}}(x_{\text{int}}, \ell) f_{\text{load}}(\ell) d\ell,$$

which leads directly to the breakage rate

$$r_{\text{break}}(x) = r_{\text{load}}(x) p_{\text{break}}(x).$$

The same averaging procedure is applied for the fragment properties. The distribution of internal fragment properties is given by the probability density function

$$f_{\text{fragm}, x_{\text{int}}}(x_{\text{fragm}, \text{int}}) = \int_{\mathbb{R}^{m_{\text{load}}}} f_{\text{fragm}, (x_{\text{int}}, \ell)}(x_{\text{fragm}, \text{int}}) f_{\text{load}}(\ell) d\ell,$$

for $x_{\text{fragm}, \text{int}} \in \mathbb{R}^{m_{\text{int}}}$. Furthermore, the expected number of fragments is given by

$$N_{\text{fragm}}(x_{\text{int}}) = \int_{\mathbb{R}^{m_{\text{load}}}} N_{\text{fragm}}(x_{\text{int}}, \ell) f_{\text{load}}(\ell) d\ell.$$

The breakage function $b_{\text{fragm}}(x_{\text{fragm}}; x)$ can be obtained directly from $f_{\text{fragm}, x_{\text{int}}}(x_{\text{fragm}, \text{int}})$ and $N_{\text{fragm}}(x_{\text{int}})$. Note that the procedure how the external properties of fragments are added depends on the meaning of the external properties. For example, spatial coordinates would just be transferred from the original particle.

4. Example

In this section, we explain the methodology with a simple example. First, we discuss the general procedure for applying the models of Section 3 in conjunction with DEM. However, in the present paper, for simplicity we generate a data set without using DEM. The incorporation of DEM simulations will be the subject of a forthcoming paper. The generated data is used as a basis to fit the breakage model, and the prediction quality of the fitted model is evaluated.

4.1. General procedure

The general procedure for obtaining the machine and material functions as described in Section 3 is as follows.

- 1) Determine the internal and external properties that shall be used in the modeling.
- 2a) Perform DEM simulations on the apparatus-scale to obtain information on stress events for given operation parameters. Note that the DEM simulations require suitable (calibrated and validated) contact models.
- 2b) Fit a multivariate distribution to the observations of the (unknown) random vector $(X_{\text{stress}}, L_{\text{stress}})$, which captures the particle-dependent loading conditions.

3a) Perform DEM simulations on the scale of a single particle to obtain information about influence of stress intensity on breakage probability and fragment properties. To obtain such information the bonded-particle model (BPM) can be effectively used. By BPM an investigated particle is represented as an agglomerate consisting of smaller primary particles connected with bonds (Dosta et al., 2013). During simulation bonds can be destroyed and, thus, breakage of the initial particle can be modeled.

- 3b) Fit a multivariate distribution to the observations of the (unknown) random vector $(X_{\text{int}}, L_1, \dots, L_{m_{\text{load}}-1}, C_{\text{crit}})$, which captures the critical stress intensities.
- 3c) Fit a multivariate distribution to the observations of the (unknown) random vector $(\tilde{X}_{\text{int}}, \tilde{L}, X_{\text{fragm}, \text{int}})$, which captures the properties of the fragments.
- 4) Compute the conditional density functions as necessary for evaluation of material and machine functions. Usually, discretization will be needed to solve the PBM.

As already mentioned above, in the following section we generate a sample data set without DEM. This has the advantage that the obtained data is simple to model. Otherwise, technical details would obfuscate the fitting procedure in steps 2b), 3b) and 3c). Furthermore, (realistic) DEM simulations would need further elaboration on model choice, model calibration and validation.

4.2. Sample data generation

We generate a simple data set to which we will fit a copula-based breakage model. Here the breakage behavior of particles is investigated where the particles are distributed only through one internal coordinate: their volume x . The initial distribution is described by the $\log N(\mu_X, \sigma_X)$ -distribution with location parameter $\mu_X = -1$ and scale parameter $\sigma_X = 0.2$. For modeling, 1000 representatives of the random volume X have been generated $(x^{(1)}, \dots, x^{(1000)})$. The particle volume density distribution is illustrated in Figure 3.

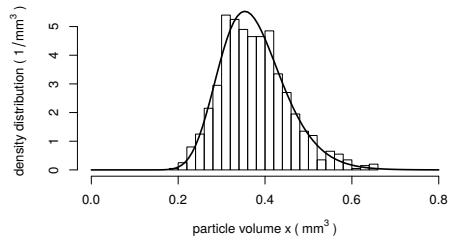


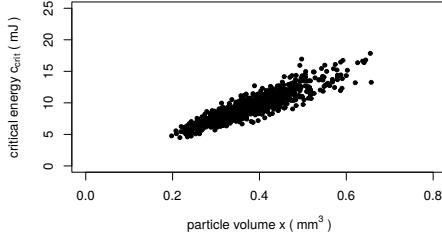
Figure 3: Particle volume distribution of sample data.

In our example, we assume that the vector ℓ of loading conditions is one-dimensional and contains the stress energy. The “critical” stress energy is the breakage energy and it is assumed to depend on the particle volume according to a linear

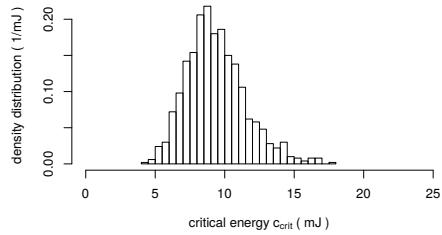
function. In particular, for the particle with volume X , the random critical energy C_{crit} (specified in mJ) is set to

$$C_{\text{crit}} = (25 + \varepsilon)X \quad (9)$$

where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is a normally-distributed noise with $\sigma_\varepsilon = 2.5$, and ε is independent of X . For every $x^{(i)}$, we simulate $c_{\text{crit}}^{(i)}$ according to Equation (9), and we obtain a scatter plot as shown in Figure 4(a). Looking only at the realizations $c_{\text{crit}}^{(1)}, \dots, c_{\text{crit}}^{(1000)}$, the histogram of critical energies in Figure 4(b) is obtained.



(a) Scatter plot of particle volumes and critical stress energies.



(b) Critical energy distribution of all particles.

Figure 4: Critical stress energies in sample data.

The data on fragments is generated as follows. We assume that the fragment sizes do not depend on the stress energy. Furthermore, we make the assumption that the relative fragment sizes do not depend on the initial particle volume. This scale invariance means that the distribution of fragments (described by the random fragment volume X_{fragm}) appearing after breakage of particle with random volume X is determined by

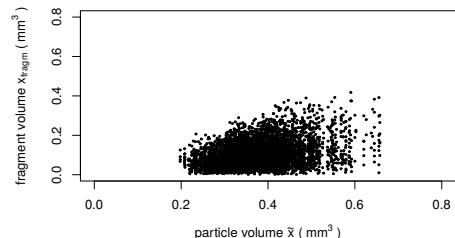
$$X_{\text{fragm}}/X \sim \text{Beta}(\alpha_{\text{fragm}}, \beta_{\text{fragm}}),$$

where $\text{Beta}(\alpha_{\text{fragm}}, \beta_{\text{fragm}})$ is the beta distribution with parameters $\alpha_{\text{fragm}} = 2$ and $\beta_{\text{fragm}} = 8$. Note that the beta distribution generates only values in the interval $[0, 1]$. Therefore, a fragment cannot be larger than the original particle. With this construction, the average number of fragments is given by $(\alpha_{\text{fragm}} + \beta_{\text{fragm}})/\alpha_{\text{fragm}}$. For our example, this leads to the average number of fragments of $(2 + 8)/2 = 5$. The sample data generation of the fragment volumes is organized as follows. We assume that the random number of fragments $N_{\text{fragm}} \sim \text{Poi}(5)$ has a Poisson distribution. This is a technical detail of sample data generation: the Poisson distribution is discrete and returns integer values, and the expected value is exactly the parameter. This is useful to obtain a statistically correct number of fragments (although in given simulation runs

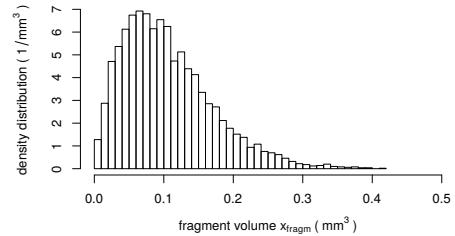
the relative fragment volumes do not necessarily sum up to unity). In DEM simulations on the single-particle scale this would not be a problem. For a given original particle volume $x^{(i)}$, the fragments are generated by

- 1) drawing a realization $n_{\text{fragm}}^{(i)}$ from N_{fragm} ,
- 2) sampling $n_{\text{fragm}}^{(i)}$ fragment volumes $\{x_{\text{fragm}}^{(i,j)}, j = 1, \dots, n_{\text{fragm}}^{(i)}\}$ from X_{fragm} .

The resulting data is shown in Figure 5, where both the scatter plot against original particle sizes and the histogram of the overall fragment volume distribution are shown.



(a) Scatter plot of (original) particle volumes and fragment volumes.



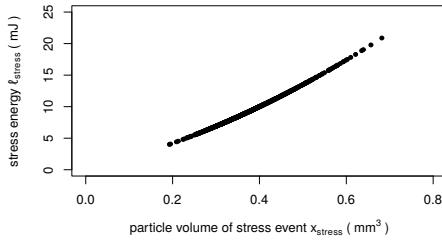
(b) Fragment volume distribution of all fragments.

Figure 5: Fragment volumes in sample data.

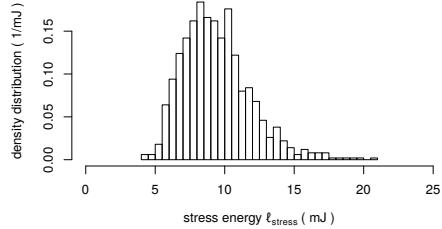
So far we have described the critical energies and the fragment volumes of the sample data, which is the data required for fitting the material function. Now, we consider the data required for the machine function, *i.e.*, we consider the loading conditions that are described by the stress event distribution. For simplicity, we assume that all particles are stressed equally often (*i.e.*, regardless of their size). Furthermore, we assume that there is a deterministic relationship between particle volume and stress energy. (In the reality, a particle with a specific volume is stressed with varying energies and the stressing frequency is influenced by its size.) The random stress energy is defined as $L_{\text{stress}} = 25X_{\text{stress}} + 20(X_{\text{stress}} - 0.4)X_{\text{stress}}$ for some random particle volume X_{stress} of stressed particles. Note that, here, X_{stress} has the same distribution as X because the particle volume does not influence the loading frequency. A scatter plot of realizations of $(X_{\text{stress}}, L_{\text{stress}})$ is shown in Figure 6(a), and a histogram of the marginal L_{stress} is given in Figure 6(b).

4.3. Fitting copula models to sample data

In this section, we assume that the there is no information about the distributions of (X, C_{crit}) , $(\bar{X}, X_{\text{fragm}})$ and



(a) Scatter plot of particle volumes of stress events and stress energies.



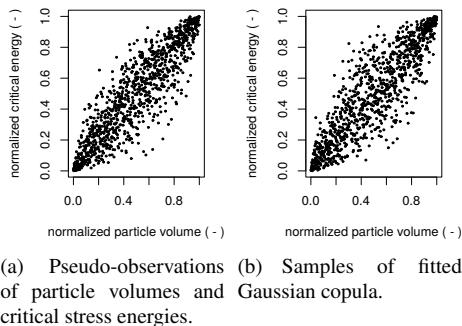
(b) Stress energy distribution of all stress events.

Figure 6: Stress events in sample data.

$(X_{\text{stress}}, L_{\text{stress}})$. Only the data generated above for the 1000 simulation runs is known. Based on this data and using copulas we reconstruct and estimate the initial distributions (previously defined in Section 4.2).

4.3.1. Fitting of (X, C_{crit})

We start with the distribution of (X, C_{crit}) , which will be estimated from the data shown in Figure 4(a). The pseudo-observations of the copula (see Appendix A.3) are shown in Figure 7(a). We fitted several families of copulas to the data (Clayton, Gumbel, Frank, Joe, Gaussian, t -copula, cf. Mai and Scherer, 2012; Nelsen, 2006) using maximum likelihood fitting, and we selected the best fit using the AIC (see Appendix A.4). The result is a Gaussian copula with correlation coefficient 0.906. Samples of this copula are shown in Figure 7(b), which confirms a good agreement.



(a) Pseudo-observations (b) Samples of fitted of particle volumes and Gaussian copula. critical stress energies.

Figure 7: Modeling of dependence structure for critical stress energies. The normalization to so-called pseudo-observations is made according to (A.3).

The copula describes the dependence structure of X and C_{crit} . However, we also need the marginal distributions of both X

and C_{crit} . This is done with maximum likelihood fitting and a manual suitable families of parametric distributions. Fitting a log-normal distribution to $\{x^{(i)}, i = 1, \dots, 1000\}$ yields an almost perfect fit with $\hat{\mu}_X = -0.998$, $\hat{\sigma}_X = 0.206$. The marginal distribution of C_{crit} is also chosen as log-normal, i.e., $C_{\text{crit}} \sim \log \mathcal{N}(\mu_{C_{\text{crit}}}, \sigma_{C_{\text{crit}}}^2)$. This leads to the estimate $\hat{\mu}_{C_{\text{crit}}} = 2.213$, $\hat{\sigma}_{C_{\text{crit}}} = 0.226$ based on the data $\{c_{\text{crit}}^{(i)}, i = 1, \dots, 1000\}$. The density function is shown in Figure 8.

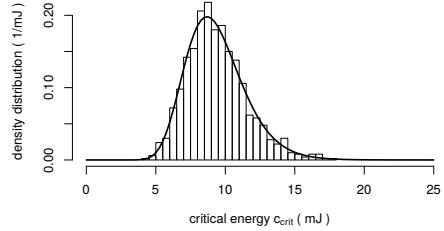
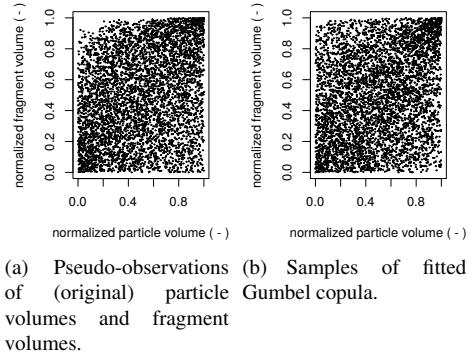


Figure 8: Modeling of critical energy distribution of all particles.

4.3.2. Fitting of $(\tilde{X}, X_{\text{fragm}})$

Similar to the modeling of particle volumes and critical energies, the original particle volumes and fragments have to be described by $(\tilde{X}, X_{\text{fragm}})$. The data is given as a set of vectors $\{\tilde{x}^{(i)}, x_{\text{fragm}}^{(i,j)}, i = 1, \dots, 1000, j = 1, \dots, n_{\text{fragm}}^{(i)}\}$. Fitting several copula families to the pseudo-observations and selection of the best fit with the AIC lead to a Gumbel copula with parameter 1.230. The pseudo-observations and samples drawn from the copula are shown in Figure 9. The marginal distribution of



(a) Pseudo-observations (b) Samples of fitted of (original) particle Gumbel copula. volumes and fragment volumes.

Figure 9: Modeling of dependence structure for fragment volumes.

\tilde{X} is log-normal with estimated parameters $\hat{\mu}_{\tilde{X}} = -1.002$, $\hat{\sigma}_{\tilde{X}} = 0.206$. Note that, in this case, all particles generate (in expectation) the same number of fragments, therefore the estimated distribution of \tilde{X} is essentially the same as the distribution of X . The marginal distribution of X_{fragm} is described by a gamma distribution with shape parameter 2.429 and rate parameter 22.957. The fit is illustrated in Figure 10.

4.3.3. Fitting of $(X_{\text{stress}}, L_{\text{stress}})$

The distribution of $(X_{\text{stress}}, L_{\text{stress}})$ is modeled a bit differently. Because we assumed a deterministic relationship for data generation in Section 4.2, we can use the so-called

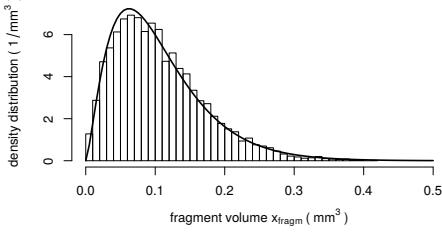


Figure 10: Modeling of fragment volume distribution of all fragments.

co-monotonicity copula. It ensures a perfect dependence without randomness, *i.e.*, it leads to a monotonically increasing stress energy in dependence on the particle volume. The co-monotonicity copula has no parameter — therefore, we need only to determine the marginal distributions. Note, however, that this is only the case for the sample data considered here. For more realistic data, other relationships can be captured by choosing another copula that fits the data well. Having selected the copula, the marginal distributions have to be determined. These include information on the distribution of particle volumes and loading frequencies. The marginal distribution of X_{stress} is again log-normal (with estimated parameters $\hat{\mu}_{X_{\text{stress}}} = -0.991$, $\hat{\sigma}_{X_{\text{stress}}} = 0.194$) because in this case, it was assumed for data generation that all particles are stressed equally often. For the marginal distribution of L_{stress} , we also use a log-normal distribution. The estimated parameters are $\hat{\mu}_{L_{\text{stress}}} = 2.209$, $\hat{\sigma}_{L_{\text{stress}}} = 0.253$. The fit is shown in Figure 11.

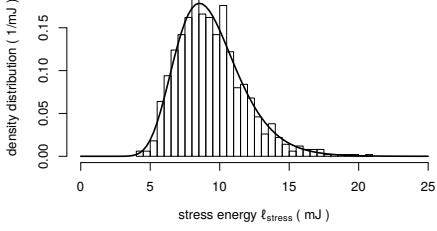
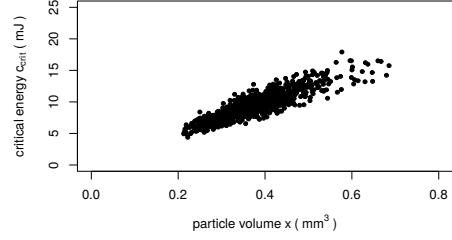
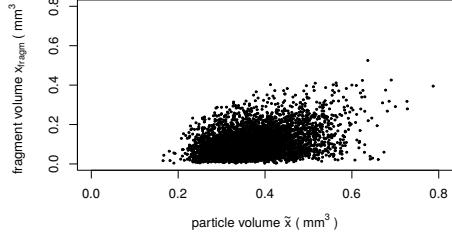


Figure 11: Modeling of stress energy distribution of all stress events.

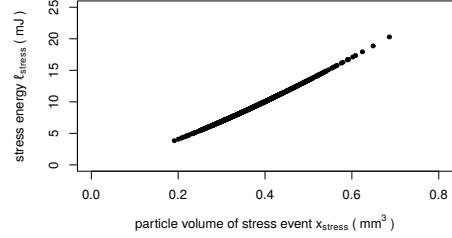
Finally, for all three modeled random vectors (X, C_{crit}) , $(\tilde{X}, X_{\text{fragm}})$ and $(X_{\text{stress}}, L_{\text{stress}})$, we can draw samples and compare them to the original data that was used for fitting. Figure 12 shows the samples drawn from these copula-based distributions. We can observe a quite nice agreement with the original data. In the next section, we evaluate the predictive abilities of the copula-based model.



(a) Samples of particle volumes and critical stress energies.



(b) Samples of (original) particle volumes and fragment volumes.

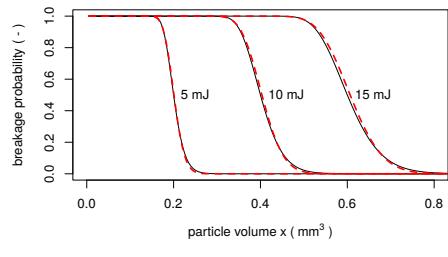


(c) Samples of particle volumes of stress events and stress energies.

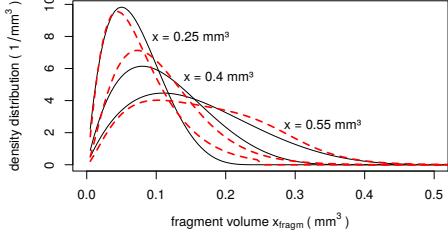
Figure 12: Samples of fitted distributions for (X, C_{crit}) , $(\tilde{X}, X'_{\text{fragm}})$ and $(X_{\text{stress}}, L_{\text{stress}})$.

4.4. Copula model validation

One quantity that is of great interest is the breakage probability under a given stress intensity. For that reason, we selected three different stress energies and evaluated the predicted breakage probability for all particle volumes. The results are shown in Figure 13(a). The prediction on the basis of the copula-based breakage model is shown using dashed red lines. Due to the data generation method of Section 4.2, we know the correct result — the desired outcome is shown using solid black lines. We see that there is a very good agreement. The interpretation is as follows. Consider $x = 0.4 \text{ mm}^3$. For a stress intensity of 5 mJ, the breakage probability is zero and therefore no particles with this volume are expected to break under this stress event. For 10 mJ on the other hand, the breakage probability is about 0.5, which means that every second particle with $x = 0.4 \text{ mm}^3$ is expected to break. For an even larger stress intensity of 15 mJ, the breakage probability is 1, which leads to the interpretation that all particles with this volume will break under this load.



(a) Breakage probabilities in dependence on particle volume for three different stress energies.



(b) Densities of fragment volume distributions for three different original particle volumes.

Figure 13: Comparison of data obtained from theoretical construction of Section 4.2 (solid black lines) and predictions from fitted distributions of Section 4.3 (dashed red lines).

Furthermore, we consider the predicted fragment volume distributions of the copula-based breakage model. We fix three different original particle volumes and compare the predicted fragment volume distributions to the theoretically expected distributions. In Figure 13(b) we can see the predicted density functions. As expected, large particles break up into large fragments, *i.e.*, we see a wider fragment volume distribution than the one for smaller original particles. The match of theoretical and obtained density functions is not perfect, but the data is described quite well, even on this limited data set consisting of 1000 particles. In this evaluation one has to keep in mind that the copula-based breakage model does not assume

scale invariance of the fragment volume distribution. (If this is desired, one could consider relative fragment volumes, thus avoiding the probabilistic dependence on the original volume and decreasing the dimensionality — however, our aim at this point is to make few assumptions and obtain good predictions anyway.)

5. Conclusions and outlook

In this paper, we presented a new approach to the modeling of breakage behavior of particles. This is important in process engineering, where comminution occurs and processes are modeled using PBM. The proposed approach is based on copulas, a well-known tool for the modeling of multivariate distributions. We showed that copula-based distributions are suitable to describe both the machine function as well as the material function. We generated a sample data set, fitted copula-based distributions and evaluated the quality of their predictions.

A big advantage of the proposed approach is its high flexibility. There are almost no restrictions on the effects that can be modeled. However, at the same time, this is also the largest disadvantage. Making only few assumptions in the model construction, there needs to be sufficiently many sample data in order to obtain realistic stochastic models. This makes extensive DEM simulations necessary.

In a follow-up paper, we will use data from DEM simulations to describe realistic scenarios, and use the developed models for a PBM-based description of a real process. This is beyond the scope of the present paper. In practice, there may be some technical problems — for example, the fragment size distribution is often not nearly as nice as in the example considered in the present paper. Then, mixed distributions or similar techniques can be used to obtain good stochastic models for the data. Note that this does not change anything in the methodology. From our point of view, the choice of (mixed) distribution families should be made according to the data because it is unlikely that one suggested distribution family for *e.g.* fragment sizes will work in all cases.

Appendix A. Copula-based modeling of multivariate distributions

Although copula-based modeling of multivariate distributions is already applied in different areas, it is not yet well-known on a broad basis. Typical applications are in finance and insurance (McNeil et al., 2005), but there are also some other applications like, *e.g.*, in climate research (Schölzel and Friederichs, 2008). In this appendix, we aim to give a short introduction to copulas, present typical copula families, and explain their practical application like model choice and fitting. More details can be found in books on copulas, *e.g.*, Mai and Scherer (2012); Nelsen (2006).

Appendix A.1. Definitions and basic properties

In this section, a short introduction to the modeling of random vectors and their multivariate distributions is given. The modeling idea is to consider marginal distributions and the dependence structure separately — and the dependence structure is described by a copula. More precisely, a copula is an m -dimensional distribution function with uniform marginals on the interval $[0, 1]$. The theoretical foundation for splitting marginals from the dependence structure is Sklar's theorem (Nelsen, 2006; Mai and Scherer, 2012; Joe, 2015; Ruppert and Matteson, 2015). Let $X = (X_1, \dots, X_m)$ denote an m -dimensional random vector with multivariate distribution function $F_X(x_1, \dots, x_m) = \mathbb{P}(X_1 \leq x_1, \dots, X_m \leq x_m)$. It has marginal distribution functions $F_{X_i}(x) = \mathbb{P}(X_i \leq x)$ for $X_i, i = 1, \dots, m$. Then, Sklar's theorem states that there exists a copula C such that

$$F_X(x_1, \dots, x_m) = C(F_{X_1}(x_1), \dots, F_{X_m}(x_m)) \quad (\text{A.1})$$

for all $x_1, \dots, x_m \in \mathbb{R}$ (this is exactly Equation (3)). Vice versa, given an m -dimensional copula C and m one-dimensional distribution functions F_{X_1}, \dots, F_{X_m} , then F_X defined as in (A.1) is a multivariate distribution function.

The most simple copula is the so-called independence copula (or product copula). It is given by

$$C^{\Pi}(u_1, \dots, u_m) = u_1 \cdots u_m,$$

which makes it obvious that a multivariate distribution with this copula must have independent components. On the other hand, a perfect positive linear dependence (without randomness) is given by the co-monotonicity copula

$$C^+(u_1, \dots, u_m) = \min\{u_1, \dots, u_m\}.$$

Appendix A.2. Parametric families of copulas

Most copulas used in practice belong to parametric copula families. For example, such a family is given by the Gaussian copulas. A Gaussian copula corresponds to the dependence structure of a multivariate normal distribution. Let $\Sigma \in \mathbb{R}^{m \times m}$ denote a correlation matrix. Then, the Gaussian copula with parameter matrix Σ is given by

$$C_{\Sigma}^{\text{Gauss}}(u_1, \dots, u_m) = \Phi_{\Sigma}\left(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m)\right),$$

where Φ_{Σ} is the distribution function of a multivariate normal distribution with expectation vector zero and covariance matrix Σ , and Φ^{-1} is the inverse distribution function (*i.e.*, quantile function) of a (univariate) standard normal distribution.

A similar approach works for the multivariate t -distribution, which is characterized by its degrees of freedom $\nu \in \mathbb{N}$, location parameter $\eta \in \mathbb{R}^m$ and positive definite scale matrix $\Sigma \in \mathbb{R}^{m \times m}$. The t -copula is given by

$$C_{\nu, \Sigma}^t(u_1, \dots, u_m) = F_{\nu, \Sigma}\left(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_m)\right)$$

where $F_{\nu, \Sigma}$ is the joint distribution function of a t -distributed random vector with ν degrees of freedom, scatter matrix $\Sigma =$

Table A.1: Some important Archimedean copula families and their generators. Note that “log” denotes the natural logarithm.

name of family	range of θ	generator $\varphi_{\theta}(t)$
Gumbel	$[1, \infty)$	$(-\log(t))^{\theta}$
Frank	$(-\infty, \infty) \setminus \{0\}$	$-\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$
Clayton	$[-1, \infty) \setminus \{0\}$	$\frac{1}{\theta}(t^{-\theta} - 1)$
Joe	$[1, \infty)$	$-\log\left(1 - (1-t)^{\theta}\right)$
Ali-Mikhail-Haq	$[-1, 1)$	$\log \frac{1-\theta(1-t)}{t}$

and location parameter $\eta = (0, \dots, 0)$, and t_{ν}^{-1} denotes the quantile function of a univariate standard t -distributed random variable with ν degrees of freedom. It is worth mentioning that both the Gaussian and t -copulas belong to the class of so-called elliptical copulas (Mai and Scherer, 2012).

Another important class of copulas is known as the Archimedean copulas. They are easily constructed, very different dependence structures can be modeled, and they have nice mathematical properties (Nelsen, 2006). Let $\varphi : [0, 1] \rightarrow [0, \infty]$ be a continuous, strictly decreasing function such that $\varphi(1) = 0$ and let

$$\varphi^{[-1]}(w) = \begin{cases} \varphi^{-1}(w) & \text{if } 0 \leq w \leq \varphi(0), \\ 0 & \text{if } \varphi(0) \leq w \leq \infty \end{cases}$$

denote its pseudo-inverse. If φ is convex, then

$$C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)), \quad u, v \in [0, 1],$$

is a two-dimensional copula (Nelsen, 2006). It is called an Archimedean copula with generator φ . Note that a generalization to higher dimensions is possible, but another condition is necessary for φ such that C is a copula (Nelsen, 2006).

By choosing different generator functions, it is possible to construct various families of Archimedean copulas. Very important families are: Gumbel, Frank, Clayton, Joe, Ali-Mikhail-Haq. These families are implemented in the R-package `copula` (Hofert et al., 2015) and their generators are given in Table A.1. (Note that some authors switch the meaning of the generator function and its pseudo-inverse, *i.e.*, notation is just the other way round. This may be confusing at first glance when the generator is “different”.)

Appendix A.3. Parametric pseudo-maximum likelihood

A common case is that there are $n \in \mathbb{N}$ observations $\{(x_{i,1}, \dots, x_{i,m}), i = 1, \dots, n\}$ of a random vector $X = (X_1, \dots, X_m)$, whose multivariate distribution function F_X is unknown. A standard approach in this case is to combine a (parametric) copula C and m (parametric) one-dimensional distribution functions F_{X_1}, \dots, F_{X_m} . This is possible by formula (A.1), yielding the multivariate distribution function F_X . In theory, it is possible to use standard maximum-likelihood fitting to estimate all parameters that describe the marginal distributions as well as the copula. However, this is often not

feasible because the optimization problem is high-dimensional, possibly multi-modal, it cannot be solved analytically and it is hard to choose a suitable initial guess for iterative numerical methods. Therefore, the univariate marginals are fitted separately using classical maximum-likelihood (Casella and Berger, 2002), *i.e.*, for every $j \in \{1, \dots, m\}$, the distribution function F_{X_j} is estimated from the sample $\{x_{i,j}, i = 1, \dots, n\}$. Finally, the copula is fitted to the transformed observations

$$\{(F_{X_1}(x_{i,1}), \dots, F_{X_m}(x_{i,m})), i = 1, \dots, n\} \quad (\text{A.2})$$

by maximum-likelihood. This methodology is called parametric pseudo-maximum likelihood (Ruppert and Matteson, 2015).

Note that, very often, the copula fitting is based on so-called pseudo-observations (Hofert et al., 2015), which are given by

$$\begin{aligned} & \{(u_{i,1}, \dots, u_{i,m}), i = 1, \dots, n\} = \\ & \left\{ \left(\frac{n}{n+1} \hat{F}_{X_1}(x_{i,1}), \dots, \frac{n}{n+1} \hat{F}_{X_m}(x_{i,m}) \right), i = 1, \dots, n \right\} \end{aligned} \quad (\text{A.3})$$

instead of (A.2), where $\hat{F}_{X_j}(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{x_{i,j} \leq z\}$ denotes the empirical distribution function of X_j . The scaling with $\frac{n}{n+1}$ is asymptotically negligible and only a technical detail (for practical reasons, it is an advantage for all values to lie in the open interval $(0, 1)$).

Appendix A.4. Selecting a parametric family using AIC

Sometimes, it is a problem to decide which family of distributions (or copulas) fits the data best. An optical impression may be misleading or different types of deviations cannot be compared easily. When using maximum-likelihood, a reasonable way is to choose the parametric family whose (log)likelihood value is the largest. However, in the case of a different number of parameters, the maximized likelihood value should not be compared directly — the additional flexibility that makes better fits possible comes at the cost of more parameters, which is a clear disadvantage. A standard technique to balance goodness-of-fit and complexity is Akaike's information criterion (AIC) (Akaike, 1974). For a given parametric family of distributions (or copulas), it is defined as

$$\text{AIC} = 2(p - \log L(\hat{\eta})),$$

where p is the number of parameters and $L(\hat{\eta})$ denotes the best likelihood value (which is adopted for parameters $\hat{\eta} \in \mathbb{R}^p$). Having computed the AIC for several different families, one selects the family that yields the smallest AIC value.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft [grant number SCHA 997/14-2] in the priority program 1679 “Dynamische Simulation vernetzter Feststoffprozesse”.

References

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE T. Automat. Contr.* 19(6), 716–723.
- Barrasso, D., Ramachandran, R., 2015. Multi-scale modeling of granulation processes: Bi-directional coupling of PBM with DEM via collision frequencies. *Chem. Eng. Res. Des.* 93, 304–317.
- Bilgili, E., Scarlett, B., 2005. Population balance modeling of non-linear effects in milling processes. *Powder Technol.* 153(1), 59–71.
- Briesen, H., 2006. Simulation of crystal size and shape by means of a reduced two-dimensional population balance model. *Chem. Eng. Sci.* 61(1), 104–112.
- Casella, G., Berger, R.-L., 2002. Statistical Inference. 2nd ed., Duxbury, Pacific Grove.
- Cundall, P.A., Strack, O.D.L., 1979. A discrete numerical model for granular assemblies. *Géotechnique* 29, 47–65.
- De Bono, J., McDowell, G., 2016. Particle breakage criteria in discrete-element modelling. *Géotechnique* 66(12), 1014–1027.
- Dosta, M., Antonyuk, S., Hartge, E.-U., Heinrich, S., 2014. Parameter estimation for the flowsheet simulation of solids processes. *Chem.-Ing.-Tech.* 86(7), 1073–1079.
- Dosta, M., Antonyuk, S., Heinrich, S., 2012. Multiscale simulation of fluidized bed granulation process. *Chem. Eng. Technol.* 35, 1373–1380.
- Dosta, M., Antonyuk, S., Heinrich, S., 2013. Multiscale simulation of agglomerate breakage in fluidized beds. *Ind. Eng. Chem. Res.* 52(33), 11275–11281.
- Freireich, B., Li, J., Litster, J., Wassgren, C., 2011. Incorporating particle flow information from discrete element simulations in population balance models of mixer-coaters. *Chem. Eng. Sci.* 66(16), 3592–3604.
- Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2015. copula: Multivariate Dependence with Copulas; R package version 0.999-13, <http://CRAN.R-project.org/package=copula>.
- Joe, H., 2015. Dependence Modeling with Copulas. CRC Press, Boca Raton.
- Kostoglou, M., 2007. The Linear Breakage Equation: From Fundamental Issues to Numerical Solution Techniques, in: Salman, A.D., Ghadiri, M., Hounslow, M.J. (Eds.), Particle Breakage. Elsevier, Amsterdam; pp. 793–835.
- Mai, J.-F., Scherer, M., 2012. Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications. Imperial College Press, Singapur.
- McNeil, A.J., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques, and Tools. Princeton University Press, Princeton.
- Nelsen R.B., 2006. An Introduction to Copulas. Springer, New York.
- O'Sullivan, C., 2011. Particulate Discrete Element Modelling: A Geomechanics Perspective. Spon Press, Abingdon.
- Otwowski, H., 2006. Energy and population balances in comminution process modelling based on the informational entropy. *Powder Technol.* 167(1), 33–44.
- Peukert, W., Vogel, L., 2001. Comminution of polymers – an example of product engineering. *Chem. Eng. Technol.* 24(9), 945–950.
- Potyondy, D.O., Cundall, P.A., 2004. A bonded-particle model for rock. *Int. J. Rock Mech. Min.* 41, 1329–1364.
- Ramkrishna, D., 2000. Population Balances: Theory and Applications to Particulate Systems in Engineering. Academic Press, San Diego.
- Ramkrishna, D., Mahoney, A.W., 2002. Population balance modeling. Promise for the future. *Chem. Eng. Sci.* 57, 595–606.
- Rumpf, H., 1967. Über die Eigenschaften von Nutzstäuben. *Staub* 1, 3–13.
- Ruppert, D., Matteson, D.S., 2015. Statistics and Data Analysis for Financial Engineering. 2nd ed., Springer, New York.
- Schölzel, C., Friederichs, P., 2008. Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlinear Proc. Geoph.* 15(5), 761–772.
- Spettl, A., Dosta, M., Antonyuk, S., Heinrich, S., Schmidt, V., 2015. Statistical investigation of agglomerate breakage based on combined stochastic microstructure modeling and DEM simulations. *Adv. Powder Technol.* 26(3), 1021–1030.
- Sutkar, V.S., Deen, N.G., Mohan, M., Salikov, V., Antonyuk, S., Heinrich, S., Kuipers, J.A.M., 2013. Numerical investigations of a pseudo-2D spout fluidized bed with draft plates using a scaled discrete particle model. *Chem. Eng. Sci.* 104, 790–807.
- Torbahn, L., Weuster, A., Handl, L., Schmidt, V., Kwade, A., Wolf, D.E., 2016. Mesostructural investigation of micron-sized glass particles during shear

deformation – An experimental approach vs. DEM simulation. Preprint
(submitted).