

CHARACTERIZATION OF THE 3D MICROSTRUCTURE OF IBUPROFEN TABLETS BY MEANS OF SYNCHROTRON TOMOGRAPHY

MATTHIAS NEUMANN¹, RAMON CABISCOL^{2,3}, MARKUS OSENBERG⁴,
HENNING MARKÖTTER⁵, INGO MANKE⁵, JAN-HENRIK FINKE^{2,3},
VOLKER SCHMIDT¹

¹*Institute of Stochastics, Ulm University, Helmholtzstr. 18, 89069 Ulm, Germany*

²*Institute for Particle Technology (iPAT), TU Braunschweig, Volkmaroder Str. 5,
38104 Braunschweig, Germany*

³*Center of Pharmaceutical Engineering (PVZ), Franz-Liszt-Str. 35-A, 38106 Braunschweig,
Germany*

⁴*Department of Materials Science and Technology, Technische Universität Berlin,
Hardenbergstr. 36, 10623 Berlin, Germany*

⁵*Institute of Applied Materials, Helmholtz-Zentrum Berlin, Hahn-Meitner-Platz 1,
14109 Berlin, Germany*

ABSTRACT. A new methodology to segment the 3D internal structure of Ibuprofen tablets from synchrotron tomography is presented, introducing a physically coherent trinarization for grey-scale images of Ibuprofen tablets consisting of three phases: microcrystalline cellulose, Ibuprofen and pores. For this purpose, a hybrid approach is developed combining a trinarization by means of statistical learning with a trinarization based on a watershed algorithm. This hybrid approach allows us to compute microstructure characteristics of tablets using methods of statistical image analysis. A comparison with experimental results shows that there is a significant amount of pores which is below the resolution limit. At the same time, results from image analysis let us conjecture that these pores constitute the great majority of the surface between pores and solid. Furthermore, we compute microstructure characteristics, which are experimentally not accessible such as local percolation probabilities and chord length distribution functions. Both characteristics are meaningful in order to quantify the influence of tablet compaction on its microstructure. The presented approach can be used to get better insight into the relationship between production parameters and microstructure characteristics based on 3D image data of Ibuprofen tablets manufactured under different conditions and elucidate key effects on the strength and solubility kinetics of the final formulation.

1. INTRODUCTION

In the past years, the so-called “Quality by design” principle has been targeted as the key strategy for the development and manufacturing of novel pharmaceutical products. This new method disrupts with the traditional “trial-and-error” and “quality-by-end-product” paths and seeks to foresee the performance of the final product from the understanding of process and component attributes. Direct compaction, a critical stage in the production of pharmaceutical tablets, is a main example of “trial-and-error” operation. The absence of defects throughout

Key words and phrases. 3D imaging, Ibuprofen tablet, microstructure characterization, statistical learning, image trinarization, watershed algorithm.

the operation (delamination, capping or chipping), a good powder flowability and an adequate breakage resistance constitute the product conformity criteria (Zhou and Qiu, 2010).

When it comes to a commercial multi-component tablet, the interplay of each constituent of the powder blend (excipients, lubricants, active ingredients, among others) affects drastically the strength and solubility kinetics of the final formulation. Many factors of the blends influence the properties of the final tablet such as porosity, shape, surface area and size (Sebhatu and Alderborn, 1999; Olsson and Nyström, 2001; Poquillon et al., 2002). Several empirical approaches have been reported in order to model the tensile strength from blend properties. Chan et al. (1983) provided a model considering the effects of particle size and composition of binary mixtures. Based on percolation theory, Kuentz and Leuenberger (1998) developed a model for the tensile strength of binary formulations, assuming that a tablet can only be produced with a solid fraction larger than some critical threshold, which is required to build a percolating system in the tablet.

However, these empirical relationships between properties of single component constituents of binary tablets and tensile strength do not capture microstructure characteristics like connectivity and surface area of each of the constituents, which also influence the strength of the tablet. The spread of visualization techniques resulting from computer aided reconstructions has opened a whole range of possibilities for a detailed microstructure analysis of tablets. Several imaging techniques have been applied towards this goal, including synchrotron tomography, terahertz pulse, Raman and NIR spectroscopy. For an overview, the reader is referred to (Müllertz et al., 2016) and the references therein. In the present study, a methodological approach to investigate binary tablets based on synchrotron tomography with a voxel size of $0.44\ \mu\text{m}$ is presented and demonstrated exemplarily for bicomponent tablets consisting of microcrystalline cellulose (MCC) and Ibuprofen (API) and working as follows. Water accesses the inner parts of the tablet via the pores and dissolves the MCC. Then, API is step by step given to the body. Since the underlying microstructure of the tablet influences the flow of water through the pores as well as the solubility kinetics, it influences the speed and amount of API given to the body. While the microstructure of individual MCC particles has recently been investigated based on synchrotron tomography with a voxel size of $0.65\ \mu\text{m}$ (Fang et al., 2017), our focus goes beyond and provides a characterization of all three phases in the tablet, i.e. the constituents MCC and API as well as pore space.

For this purpose, it is necessary to suitably trinarize the 3D images, i.e. each voxel has to be labeled as MCC, API or pore. Classical algorithms from morphological image analysis as reviewed in Schlüter et al. (2014) and algorithms based on statistical learning (James et al., 2013) do not lead to a sufficiently good trinarization. More precisely, they are not able to classify the pore space adequately. In order to overcome this limitation a new method to trinarize the greyscale images obtained by synchrotron tomography is developed. We propose a hybrid approach combining a watershed-based trinarization going back to ideas presented in Meyer and Beucher (1990) with a trinarization using a random forest algorithm, a tool from statistical learning. Another combination of tools from mathematical morphology with statistical learning has recently been used to improve the quality of particle-wise segmentation from 3D image data (Furat et al., 2018). In the present paper, the hybrid approach leads to a trinarization allowing for the computation of structural characteristics, which are only accessible via image analysis such as, e.g., local percolation probabilities or chord length distribution functions. These characteristics can be used for the quantification of anisotropy effects in a microstructure resulting from a uni-axial compaction of the tablet. The values of porosity and specific surface area obtained from image analysis are compared with the corresponding experimentally obtained values. The methodology considered in the present paper permits us to extend the characterization of the 3D microstructure of Ibuprofen tablets by image analysis, which is a first step towards the goal to relate microstructure characteristics with product conformity criteria like powder flow and breakage resistance.

The paper is organized as follows. Materials and experimental methods are described in Section 2, whereas the results of experimental measurements are given in Section 3. After the description of 3D imaging in Section 4, we present a hybrid approach for the algorithmic trinarization of 3D images in Section 5. The algorithm is tested for different cut-outs of one large 3D image of the microstructure of Ibuprofen tablets which allows for a statistical characterization of the microstructure by image analysis given in Section 6. Section 7 concludes the work.

2. MATERIALS AND EXPERIMENTAL METHODS

2.1. Materials. Binary tablets of microcrystalline cellulose (MCC) (Vivapur[®]12, JRS Pharma, Germany) and Ibuprofen Gracel are used to produce the three-phase tablets considered in this paper. In order to narrow the particle size distribution (PSD) of Ibuprofen, the fraction of Ibuprofen above 180 μm remaining after manual sieving is removed before admixing.

Due to the poor processability of this mixture, each batch is internally lubricated by blending of Magnesium stearate (MgSt) (Magnesia 4264; Magnesia GmbH, Germany). Blends are homogenized in a powder mixer ERWEKA AR-403/-S (ERWEKA GmbH, Germany) with a 3.5 l cubic container for 3 minutes at 16 rpm before tableting. The following formulation is considered for this analysis: 79.20 % (w/w) MCC; 19.80 % (w/w) API and 1.00 % (w/w) MgSt.

General powder characteristics, such as particle size distribution (PSD) and true density are determined. Powder particle size distribution is determined via dry dispersion by an air-flow injector (Mastersizer 300, Malvern Instruments Ltd, United Kingdom) using a differential pressure of 2 bar.

Density of both powders is required in order to determine the evolution of internal tablet porosity with the compaction pressure. Skeletal or true density of primary particles is determined by means of helium pycnometry (ULTRAPYC 1200e, Quantachrome GmbH, Germany). Samples were stored at 20°C and 45 % RH during the 24 h prior to the analysis. Averaged density values are extracted after 10 runs on the pycnometer assuring a typical relative standard deviation of less than 0.5 %.

Blends are compacted with the compaction simulator STYL'One Evolution (Medel Pharm S.A.S., France). Standard EURO B die and punches were set up in order to produce cylindrical tablets of 11.28 mm in diameter. The compaction sequence comprises the filling of the die by gravity up to a height of 10 mm with the blend of interest (previously stored at 20°C and 45 % RH for 24 h) and the symmetrical movement of the punches at a constant speed of 20.6 mm/s until the target pressure was achieved. In the current study compaction pressures running from 45 MPa up to 188 MPa were analysed.

2.2. Experimental methods. The internal surface area of tablets was determined by nitrogen sorption at the ASAP 2460 (Micromeritics Instrument Corp., USA). Sample conditioning proceeds as follows: after compaction, cylindrical tablets are cut up in quarters and inserted into the degassing units where they were treated for 24 h at room temperature at vacuum conditions in order to remove physisorbed compounds. Then, a conditioning interval of 500 s precedes the sorption of nitrogen at relative nitrogen pressure range P/P_0 from 0.10 to 0.30 and an absolute temperature of -196°C . Samples were measured in triplicate. Finally, the best linear fit was obtained for the Brunauer-Emmett-Teller (BET) model (Brunauer et al., 1938). The BET model is a standardized method based on the Langmuir isotherm (Fagerlund, 1973), which assumes a kinetic behaviour of the adsorption process, a rate of adsorption equal to the rate of desorption, a constant heat of adsorption and the formation of an adsorbed monolayer.

Pore size analysis experiments were executed using a mercury intrusion porosimeter (MIP) PoreMaster 60 (Quantachrome GmbH, Germany). Pressures ranging from 1 to 60,000 PSI were applied in a high pressure station. The pressure was exerted onto the sample using a penetrometer made of glass as specimen container. A penetrometer with a stem volume of 0.5 cm^3 and a sample container of 3.8 cm length is used.

One of the main assumptions of this technique is cylindrical shape pore configuration. Based on that, the MIP pore size distribution can be determined using a modified Young-Laplace equation, referred mainly as Washburn equation (Washburn, 1921),

$$\Delta P = \gamma \left(\frac{1}{r_1} + \frac{1}{r_2} \right) = \frac{2\gamma \cos \theta}{r_{\text{pore}}}, \quad (1)$$

where ΔP is the pressure difference, r_1 and r_2 describe the curvature of the interface, r_{pore} the pore size using the surface tension of mercury γ and the contact angle θ between the solid and mercury. A basic assumption for this relation is a constant surface tension $\gamma = 0.485 \text{ Nm}^{-1}$ and contact angle $\theta = 140^\circ$ of the intruded mercury and the substrate. According to the instrumental pressure range, a pore size span from 1.80 nm to 108 μm should be accessible with this set-up. The MIP pore size distribution is denoted by Q_3 in the following. As pointed out in Diamond (2000), the MIP pore size distribution measures the size of those pores, which are accessible by mercury. Thus, this characteristic can not be considered as a size distribution of all pores of a microstructure.

3. EXPERIMENTAL RESULTS

The density of the considered tablets experimentally determined by helium pycnometry is 1279 kg/m^3 and the specific surface area determined by BET measurements lies between $0.78 \text{ }\mu\text{m}^{-1}$ and $0.81 \text{ }\mu\text{m}^{-1}$. In Section 6 these results are compared with the results of image analysis. The MIP pore size distribution is presented in Figure 1. The results show that only about 25 % of the pore space can be reached by mercury through pore bottlenecks with a diameter larger than $0.438 \text{ }\mu\text{m}$, which is the voxel size of the 3D images. In particular, this means that there exist pores having a diameter below the resolution threshold (about $2 \text{ }\mu\text{m}$ after filtering, see Section 4.2). Note that the pores below the resolution of image data can not be taken into account for image analysis.

The 10%-,50%- and 90%-quantiles of the volumetric/mass particle size distribution (PSD), denoted by X_{10} , X_{50} and X_{90} , and the true density of MCC and Ibuprofen are summarized in Table 1. In contrast to the pore size, the values of X_{10} of both powders are some orders of magnitude higher than the resolution of image data. Therefore primary particles should be satisfactorily detected by synchrotron tomography.

4. IMAGING

4.1. Sample preparation. First measurements have shown that small empty pores lead to imaging artifacts arising due to refraction at the gas/material interface. This prevents successful classification/trinarization of the tablets material composition. Therefore, the pores are filled by a contrast medium, which reduces the refraction at the pore/material interface. This is achieved by cutting a piece of the Ibuprofen tablet and fixing it into a polyimide tube with an inner diameter of 1.6 mm, which is then filled up with a hydrocarbon based contrast medium and sealed on both tube ends. The tube with the sample therein is then imaged via synchrotron tomography. Note that no unfilled regions of the pore space are detected at least within the analyzed region of interest. Therefore, this procedure allows an artifact-free analysis.

4.2. Synchrotron tomography. The synchrotron tomography measurement is conducted at the BAMline at the electron storage ring Bessy II in Berlin, Germany (Görner et al., 2001). For optimized contrast of the light elements in the sample a beam energy of 9.8 keV is chosen with a double multilayer monochromator. After transmitting the sample, the X-ray beam is converted with a $60 \text{ }\mu\text{m}$ thick CdWO_4 scintillator into visible light, which is then imaged onto the CCD chip of a PCO.4000 camera (PCO AG). The used optical set-up yields a resolution of 4008×2672 pixels covering an area of $1.75 \times 1.17 \text{ mm}^2$, which corresponds to a pixel size of $0.438 \text{ }\mu\text{m}$. For the tomographic scan 2200 projections covering an angular range of 180°

are collected and used for the reconstruction via filtered back projection (Jähne, 2013). Each projection is exposed for 3 s. Together with flat field images for image normalization this results in a total scan time of approximately 2 h. A phase retrieval algorithm, according to Paganin et al. (2002), is applied on the projections in order to emphasize the contrast evoked by the phase shift in the material. After the application of the phase retrieval algorithm, the resolution limit is at $\approx 2 \mu\text{m}$.

5. TRINARIZATION OF GREY-SCALE IMAGES

In order to investigate the microstructure of an Ibuprofen tablet by statistical image analysis, the grey-scale images have to be trinarized. Figure 2a shows that 3D imaging described in Section 4 leads to a good contrast between pores, MCC and API. The darkest greyscale values correspond to the pore space, the medium values to API and the brighter values to MCC. Despite of the good contrast, algorithmic trinarization encounters two major challenges. On the one hand, there are voxels within MCC, the greyscale values of which are in the same range as the ones of voxels belonging clearly to API. Moreover, the greyscale values of thin pores, which are located at the boundary between different particles of MCC or which are located within a particle of MCC, are similar to the greyscale values of API. Thus, from a physical point of view, it is not reasonable to rely only on thresholding of greyscale values. In order to visualize what we understand by a realistic trinarization, a manual trinarization of a 2D cross-section of the greyscale image is shown in Figure 2b. The algorithmic trinarization proposed in this paper is based on a random forest algorithm, which meets the above mentioned challenges (Figure 2). In a second step, the results of this trinarization are improved by a combination with a trinarization based on the watershed algorithm (Figure 2d). This leads to a hybrid approach used for the final trinarization (Figure 2e), which takes benefit from the advantages of both trinarization algorithms.

5.1. Trinarization by a random forest algorithm. A random forest is an algorithm for classification from statistical learning based on decision trees. In contrast to classification by a single decision tree, a random forest is built by a large number of randomized decision trees. For a detailed description of random forests we refer to James et al. (2013). In the present paper, a random forest algorithm is used for the trinarization of greyscale images, which can be considered as a classification problem. Each voxel has either to be classified as pore, API or MCC.

The random forest has to be trained, which means that, roughly speaking, the algorithm has to learn how to trinarize a given greyscale image. For this purpose, a 2D slice of the 3D image is considered, in which N voxels are manually trinarized by visual inspection. Thereby, a trinarization mask for the selected 2D slice is obtained. The same 2D slice is then filtered in M different ways, such that we result in an $N \times (M + 2)$ -dimensional matrix. The rows of this matrix indicate the pixels, which are considered during the manual trinarization. In each row the original greyscale value of the pixel, its greyscale values after the application of each of the M filters as well as its manually determined class label are stored. The matrix represents the training data for the random forest.

In order to perform training and feature evaluation, we use Ilastik (Sommer et al., 2011) with the parallelised random forest implemented in the computer vision library VIGRA. Finally, the M filters are applied to all slices of the 3D images, which allows for a trinarization of all voxels in the 3D image based on the trained random forest. The random forest algorithm, the result of which is visualized in Figure 2c, works satisfactorily for trinarizing the considered image data. In particular, thin pores (e.g. green circles in Figure 2c) and darker greyscale values within MCC (e.g. blue circle in Figure 2c) are properly resolved. For example, the long and thin pore in the upper left part of the 2D slice in Figure 2c is recognized correctly. Furthermore, the random forest algorithm does not lead to misclassification within particles of the MCC, even if

smaller greyscale values may suggest an occurrence of API. However, comparing the result of the random forest algorithm with the original greyscale image one can observe that pores are detected between API and MCC, although there is no indication for pores, neither by greyscale values nor by physical reasons (e.g. red circles in Figure 2c). This effect occurs because the algorithm is trained to detect pores at the boundary of MCC. Moreover, unrealistically tiny connected components of all three phases are created using the random forest algorithm. Both effects can be corrected by using a hybrid approach for trinarization, i.e. by a combination of the random forest algorithm with a trinarization based on the watershed algorithm, see Section 5.3.

5.2. Watershed-based trinarization. The watershed algorithm is a method to partition an image into different regions, so-called watershed basins, see e.g. Roerdink and Meijster (2000). Consider any 3D image as a landscape, where its greyscale values are the local altitudes of the landscape. Then, the regions generated by the watershed algorithm can be interpreted as valleys, where their boundaries are given by the watershed lines of the landscape. Typically used for the extraction of individual particles from image data, the watershed algorithm has also been applied to classify different phases in multi-phase materials (Schlüter et al., 2014). In the present paper, we propose a new method how to use the watershed algorithm to trinarize greyscale images of three-phase materials, which only relies on the greyscale values of image data and does not require any type of training data.

The principal idea of the trinarization is the following. As a starting point, the pore space is classified by global thresholding, before the solid phase, i.e. the union set of MCC and API, is partitioned by a special type of watershed algorithm. According to its average greyscale value, each watershed basin is finally assigned to one of the three phases. Note that averaging over all greyscale values within a watershed basin has been used by Frucci et al. (2013) for filtering images.

For a more detailed description of the watershed-based trinarization, we denote the set of voxels by $W \subset \mathbb{Z}^3$. The greyscale image is denoted by I and for each voxel $w \in W$, we denote its corresponding greyscale value by $I(w)$. To distinguish the pore space from the solid phases we use global thresholding, where the threshold value t_1 is determined by visual inspection. This leads to a numerical value of $t_1 = 30739$. Note that we deal with 16-bit images, i.e., all greyscale values are between 0 and 65535. The determined pore space for a cut-out of a 2D slice is visualized in Figure 3b.

In the next step, a further image J is defined on the basis of which the watershed basins are computed. The greyscale values of J are low at voxels, for which it is clear whether they belong to MCC or API. Considering the image J as a landscape, as the watershed algorithm does, the uncertainty of classification is low at voxels with a low altitude. In turn, the uncertainty of the classification is high for voxels located close to the watershed lines. Thus, the algorithm is constructed such that phase transition only occurs at watershed lines.

Formally, J is defined as follows. Applying the iterative algorithm introduced by Ridler and Calvard (1978) to all voxels, not classified as pores before, we obtain a threshold between MCC and API, denoted by t_2 . The idea of the algorithm of Ridler and Calvard (1978) is to determine the threshold t_2 such that the between-class variance is maximized (Lin, 2003). This means that the absolute difference of mean greyscale values between the two classes, which are separated from each other by the threshold t_2 , is maximized. The voxels of greyscale values differing much from t_2 are considered as voxels which can be easily classified. Thus, greyscale values of J are defined by

$$J(x) = \begin{cases} -(I(w) - t_2)^2, & \text{if } I(w) \geq t_1, \\ 0, & \text{else.} \end{cases} \quad (2)$$

After smoothing J by a minimum filter with a radius of 2 voxels ($\approx 0.876 \mu\text{m}$), the watershed algorithm is applied on J , see Figure 3c. For this purpose, we use the algorithm which was

introduced by Meyer (1994). For an overview on watershed algorithms, we refer to Beare and Lehmann (2006).

The watershed basins are assigned to MCC or API by global thresholding with respect to the average greyscale values of the basins. Here t_2 is used as global threshold. The voxels, located at the watershed boundaries are finally classified by an application of a maximum filter to the trinarized image. In particular, voxels at the boundary between MCC and API are assigned to MCC. The resulting trinarization contains unrealistically many small connected components of API within the MCC. Therefore, in a post-processing step, all API clusters with less than 300000 voxels ($\approx 25200 \mu\text{m}^3$) are removed and assigned to MCC. Note that whenever cluster analysis is performed in the present paper, the algorithm of Hoshen and Kopelman (1976) is used.

The result of the watershed-based trinarization described above is visualized in Figure 2d. This approach does not require any manually trinarized training data and is strongly leaned on the values of the greyscale image. Considering only the greyscale values without any additional information about the material, the watershed algorithm leads to an appropriate trinarization. But, as already mentioned above, relying only on greyscale values is not reasonable here from a physical point of view. In particular, the random forest algorithm leads to a strongly improved detection of small pores at the boundary of MCC and the representation of MCC itself (e.g. red circles in Figure 2d). Nevertheless, the watershed algorithm does not detect unrealistic pores between MCC and API as the random forest algorithm does (e.g. green circles in Figure 2d). Thus, we can use the results obtained by the watershed-based trinarization to improve the trinarization by the random forest algorithm.

5.3. Hybrid approach. The trinarization obtained by the random forest algorithm, denoted by T_1 , is the basis of the hybrid approach. It is just modified using information from the watershed-based trinarization, denoted by T_2 . In a first step, tiny connected components of all three phases are removed from T_1 except of tiny connected components of the pore space which are completely surrounded by MCC. The existence of the latter parts of the pore space is not in contradiction with the prior knowledge about the material from a physical perspective. By connected components, we mean clusters of voxels all belonging to the same phase, such that for each pair of voxels w_1 and w_2 there exists a voxel path (with respect to the 26-neighbourhood, see Ohser and Schladitz (2009)) from w_1 to w_2 within this cluster. All connected components of the pore space with less than 1000 voxels ($\approx 84 \mu\text{m}^3$) are assigned to API. After that, all connected components of API with less than 8000 voxels ($\approx 670 \mu\text{m}^3$) are assigned to the pore phase. Finally all connected components of MCC with less than 8000 voxels ($\approx 670 \mu\text{m}^3$) are assigned to API.

Finally, all unrealistic pores between MCC and API detected in T_1 are removed. For this purpose, we use the pore space obtained by the watershed-based segmentation. To be precise, we assign each voxel $w \in W$, which belongs to the pore phase in T_1 , to API if the following two conditions are fulfilled: 1) The Euclidean distance from w to the closest pore voxel in T_2 is larger than 20 voxel ($\approx 8.76 \mu\text{m}$). 2) The Euclidean distance to the closest API voxel in T_1 is smaller than 20 voxel ($\approx 8.76 \mu\text{m}$). While Condition 1 uses T_2 to check if the pore voxel is wrongly detected in T_1 , Condition 2 ensures that the considered pore voxel is close to API. Condition 2 is necessary because pores within MCC, which are not detected by the watershed-based trinarization, should not be removed.

The results obtained by the hybrid approach, see Figure 2e, show that the required challenges can be met by combining the random forest algorithm with the watershed-based trinarization. Note that the tiny connected components, which can be observed in Figure 2e are connected in 3D. A 3D visualization of MCC as well as API determined by the hybrid approach is given in Figure 4. Moreover, a quantitative validation is performed based on the manually trinarized 2D slice (Figure 2b), which is not used for training the random forest algorithm. The quantitative validation in terms of misclassifications is visualized by a so-called cobweb graph (Patel and

Markey, 2005) in Figure 2f. The results show that the watershed-based trinarization classifies 55 % of the manually determined pore voxels as MCC and 18 % as API. These errors occur due to higher greyscale values of thin pores at the boundary of MCC. We conclude that an appropriate trinarization of the considered image data requires methods, which allow to take prior information about the material into account as, for instance, the existence of tiny pores at the boundary of MCC particles. For this purpose, we use the random forest algorithm described in Section 5.1 leading to a strongly improved trinarization. The fraction of manually determined pore voxels, which are wrongly classified is reduced to 13 % (3 % MCC, 10 % API). Moreover, the hybrid approach combining both algorithms leads to a further improvement. In particular, having removed the unrealistic pores in the segmentation based on random forests, only 1 % of the manually classified MCC voxels are determined as pores (compared to 3 % in the random forest based segmentation).

6. STATISTICAL MICROSTRUCTURE CHARACTERIZATION

Trinarization is performed on cutouts of the entire 3D image. The hybrid approach introduced in Section 5.3 is used for this segmentation. The trinarized cut-outs are statistically analysed using methods of spatial statistics, see e.g. Chiu et al. (2013). In Section 6.1 the results of the statistical analysis of trinarized images are compared with the results obtained by the experimental characterization of tablets described in Section 3. For this purpose, eight non-overlapping cubic cutouts with a side length of 306.6 μm are considered. The side length is chosen such that we obtain several non-overlapping cut-outs, where each of them contains sufficiently rich structural information for statistical analysis.

6.1. Volume fraction and surface area. In this section, the volume fractions and the specific surface areas of the three phases are considered. The volume fractions of MCC and API and their specific surface areas estimated from image data are visualized in Figure 5. Note that for a given porosity/density, the volume fraction of MCC is a function of the volume fraction of API. The graphs of these function are visualized for certain fixed values of density and porosity in Figure 5a. One can observe that volume fractions of MCC are between 0.7 and 0.77, while the volume fractions of API are between 0.21 and 0.28. Porosity varies between 0.025 and 0.075. These variations in terms of volume fractions indicate that one single cubic cutout with side length 306.6 μm is not sufficiently representative. This means that for an analysis of these particular microstructures, either a larger cut-out or several cut-outs (as in the present paper) have to be considered. Furthermore, a density between 1350 kg/m^3 and 1450 kg/m^3 is computed using the volume fractions of phases. These values are larger than the experimentally determined density of 1279 kg/m^3 . This overestimation is attributed to the pores with a size below the voxel size (0.44 μm), which are thus not visible in the 3D images. Moreover, note that the resolution of 3D images is about 2 μm , which does not allow for an exact determination of porosity via image analysis. From the mercury intrusion porosimetry (Figure 1), it can be concluded that barely 5% of the entire tablet volume is accessible through pathways with bottleneck diameters not smaller than 2 μm in intrusion direction. This result clearly demonstrates that the pore structure is not uniformly and not steadily tapering along the intrusion direction, but that larger cavities follow narrow constrictions, so-called bottleneck pores, in intrusion direction.

Moreover, the specific surface area, which is defined here as the ratio of surface area and the volume of the solid phase, i.e. the union of MCC and API, is estimated from image data using the method given by Ohser and Schladitz (2009). The values are higher for MCC than for API and pores for two reasons. First, MCC has the highest volume fraction of the three phases, and second, small pores are located at the boundary between MCC particles. The specific surface area of pores obtained from image analysis is much lower than the experimental values, which are between 0.78 μm^{-1} and 0.81 μm^{-1} . We attribute this to the different determination of the

specific surface area by image analysis and experiments and thus, the corresponding values are difficult to compare. More precisely the main difference is that the value obtained by image analysis also depends on the resolution of the image. Note that Figure 1 shows that there is a certain fraction of pores, which are below the resolution threshold. Such a restriction does not occur in the case of the BET measurement, which allows us – due to the adsorption of nitrogen molecules – to measure the roughness of particles on a scale which is not accessible by synchrotron tomography. Consequently, it is not surprising that the experimentally determined values of specific surface area are higher than the ones computed by image analysis.

6.2. Further microstructure characteristics. Besides estimating the surface area of each phase, the surface area of the interfaces, between pores and MCC, between pores and API, and between API and MCC can be computed. The surface areas of interfaces per unit cube are denoted by $I(\text{pore, MCC})$, $I(\text{pore, API})$, and $I(\text{API, MCC})$, respectively. Note that

$$I(\text{pore, MCC}) = \frac{1}{2} (S_{\text{pore}} + S_{\text{MCC}} - S_{\text{API}}), \quad (3)$$

where S_i denotes the surface area of phase i per unit cube. The surface areas of the other interfaces can be computed analogously. This allows us to determine the proportion of pairwise interfaces in the complete surface area, see Figure 6a. Depending on the cutout, the proportion of $I(\text{pore, MCC})$ and $I(\text{API, MCC})$ varies strongly for the eight samples in the range of $0.2 - 0.45$ and $0.45 - 0.7$, respectively. This variation can be explained by the variation of porosity, see Section 6.1. Further underpinning of this explanation is given by the fact that both, $I(\text{pore, MCC})$ and $I(\text{pore, API})$ strongly correlate with porosity, see Figure 6b.

Next, local percolation probabilities (Hilfer, 1991) are computed to quantify connectivity properties of the three phases API, MCC and pores, see Figure 7. For this purpose, we divide each cubic cut-out in sub-cubes with an edge length of $22 \mu\text{m}$. For each sub-cube and each phase, we compute the volume fraction of the phase and check if it is percolating in x -, y -, and z -direction. We say that a phase is percolating within a sub-cube $[0, 22 \mu\text{m}]^3$ in, e.g., x -direction if there exists a path (with respect to the 26-neighbourhood, i.e. each voxel $w \in W$ is connected to all other voxels which share at least one vertex with w (Ohser and Schladitz, 2009)) from the bottom $\{0\} \times [0, 22 \mu\text{m}]^2$ to the top $\{22 \mu\text{m}\} \times [0, 22 \mu\text{m}]^2$ within the considered phase. Then, the percolation probability $P(v)$ for a given volume fraction v is estimated by a Nadaraya-Watson estimator (Nadaraya, 1964) with a Gaussian kernel and a manually chosen bandwidth of $h = 0.3$. In Figure 7 histograms representing the frequency of volume fractions of the sub-cubes are given for each of the three phases. Since we do not observe sub-cubes with a porosity of more than 0.5, the percolation probabilities for higher porosities cannot be considered as accurate estimates. In Figure 7a, the black dotted line shows the upper volume fraction limit, where the estimation of percolation probabilities is no longer meaningful. Formally, we consider an estimate of $P(v)$ as meaningful if the denominator in the Nadaraya-Watson estimator is larger than 100. For API and MCC, we observe sufficiently many sub-cubes for the whole range of volume fractions in order to estimate the percolation probabilities. One can observe for all three phases that the percolation probabilities increase stronger with increasing volume fractions in x - and y -direction than in z -direction. This result is reasonable as the tablet is compacted in z -direction. Thus, the pores, as well as API and MCC are more elongated in the xy -plane leading, in turn, to better connectivity of all three phases in x - and y -direction than in z -direction. The computation of local percolation probabilities can be considered as a quantification of the influence of compaction on the tablet microstructure.

Another relevant descriptor of the microstructure is the so-called chord length distribution function. Note that the chord length distribution in a certain direction ϕ is the distribution of the lengths of subsequent intersections between the considered phase and a randomly chosen line in direction ϕ . For the estimation of chord length distributions from image data we refer to Ohser and Schladitz (2009). Figure 8 shows the mean chord length distribution functions in

x -, y - and z -direction for the eight samples to measure anisotropy effects. For all three phases, it can be seen that the chord length distribution functions in x - and y - directions are nearly identical and chord lengths in x - and y -direction are significantly larger than the ones in z -direction. The chord length distribution quantifies the elongation of phases in certain directions and is thus – besides local percolation probabilities – a further measure for the influence of uniaxial compression onto the microstructure.

6.3. Results and discussion. The presented statistical analysis characterizes the microstructure of tablets based on 3D images with a voxel size of 0.438 μm . Thus, pores with a diameter below the resolution threshold are not taken into account in the analysis, which leads to slightly different values of porosity and discrepancies of specific surface area compared to experimental results. Being aware that the analysis is performed on a certain length scale, which is determined by the voxel resolution, statistical image analysis nevertheless allows for the computation of microstructure characteristics, which are not accessible by experiments. The performed analysis shows that the interface area between pores and MCC, as well as between pores and API grows linearly in porosity for the considered material. The computation of local percolation probabilities and chord length distribution functions allows for a quantitative analysis of the impact of compaction on the microstructure, which can complement empirical models such as on the prediction of compactibility, compressibility and tabletability of compacts (Imbert et al., 1997).

A promising and powerful application of the presented trinarization, which identifies the different components, consists of a further discretisation which allows for numerical modeling of multicomponent formulations by means of the discrete element method (DEM). In particular, breakage and propagation of cracks can be numerically modeled. In future work, the statistically representative volume element size with respect to mechanical properties will be determined based on DEM simulations in the spirit of Dirrenberger et al. (2014).

In order to reinforce the current results, it might be necessary to include further microstructure characteristics in order to describe the microstructure influence on mechanical properties. Beyond local percolation probabilities, which are considered in the present paper as connectivity property, the quality of connecting pathways, important for effective conductivity in two-phase microstructures and reflected in the notions of mean geodesic tortuosity and constrictivity (Stenzel et al., 2017), could also affect the mechanical properties significantly.

7. CONCLUSIONS

A novel methodology to investigate the microstructure of Ibuprofen tablets based on 3D image data from synchrotron tomography has been presented. To be more precise, a new trinarization algorithm has been developed, which allowed for the identification and labelling of the three constituting phases, i.e., pores, API and MCC. Here an algorithm using methods from statistical learning, namely random forests, has been combined with a watershed-based algorithm. In contrast to using just one of these two algorithms, the presented hybrid approach matches the required challenges in the sense that it leads to a physically realistic trinarization. This means that the trinarization algorithm has not detected tiny constitutive parts of API within MCC particles and, on the other hand, long pores at the boundary of MCC particles are correctly detected. The trinarization allowed for an investigation of the microstructure by means of statistical image analysis, which has been illustrated by the example of a given tablet. The results obtained from statistical image analysis showed differences with experimental measurements, due to the pores with diameter below the resolution threshold. Furthermore, image analysis showed that the absolute change in the volume fractions of the three phases is up to 5 % when considering 8 different cubic cut-outs of the tablet with an edge length of 0.3 mm. Moreover, the computation of local percolation probabilities and chord length distribution functions enabled us to quantify the influence of compaction of the tablet on its

microstructure. The presented methodology can be used to investigate the relationship between production parameters of the tablet, as the ratio of volume fractions of API and MCC or the pressure of compaction, and the microstructure characteristics, which influence, in turn, effective properties of the tablet as its strength and solubility kinetics. As a possible subject of future research, a further powerful outcome of the trinarization might be a numerical modeling of multicomponent formulations by DEM, in order to elucidate the microstructure influence on mechanical properties of the tablet.

REFERENCES

- Beare, R. and Lehmann, G. (2006) The watershed transform in ITK-discussion and new developments. *The Insight Journal*, **92**, 1–24.
- Brunauer, S., Emmett, P. H. and Teller, E. (1938) Adsorption of gases in multimolecular layers. *Journal of the American Chemical Society*, **60**, 309–319.
- Chan, S. Y., Pilpel, N. and Cheng, D. C. H. (1983) The tensile strengths of single powders and binary mixtures. *Powder Technology*, **34**, 173–189.
- Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013) *Stochastic Geometry and its Applications*. Chichester: J. Wiley & Sons, 3rd edn.
- Diamond, S. (2000) Mercury porosimetry: an inappropriate method for the measurement of pore size distributions in cement-based materials. *Cement and Concrete Research*, **30**, 1517–1525.
- Dirrenberger, J., Forest, S. and Jeulin, D. (2014) Towards gigantic RVE sizes for 3D stochastic fibrous networks. *International Journal of Solids and Structures*, **51**, 359–376.
- Fagerlund, G. (1973) Determination of specific surface by the BET method. *Matériaux et Construction*, **6**, 239–245.
- Fang, L., Yin, X., Wu, L., He, Y., He, Y., Qin, W., Meng, F., York, P., Xu, X. and Zhang, J. (2017) Classification of microcrystalline celluloses via structures of individual particles measured by synchrotron radiation X-ray micro-computed tomography. *International Journal of Pharmaceutics*, **531**, 658–667.
- Frucchi, M., Nappi, M. and Riccio, D. (2013) Watershed based iris segmentation. In *Pattern Recognition: Proceedings of the 5th Mexican Conference, MCPR 2013, Querétaro, Mexico* (eds. J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, J. S. Rodríguez and G. S. di Baja), 204–212. Berlin: Springer.
- Furat, O., Leißner, T., Ditscherlein, R., Šedivý, O., Weber, M., Bachmann, K., Gutzmer, J., Peuker, U. and Schmidt, V. (2018) Description of ore particles from XMT images, supported by SEM-based image analysis. *Microscopy and Microanalysis*, **24**, 461–470.
- Görner, W., Hentschel, M. P., Müller, B. R., Riesemeier, H., Krumrey, M., Ulm, G., Diete, W., Klein, U. and Frahm, R. (2001) BAMline: the first hard X-ray beamline at BESSY II. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **467–468**, 703–706.
- Hilfer, R. (1991) Geometric and dielectric characterization of porous media. *Physical Review B*, **44**, 60–75.
- Hoshen, J. and Kopelman, R. (1976) Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. *Physical Review B*, **14**, 3438–3445.
- Imbert, C., Tchoreloff, P., Leclerc, B. and Couarraze, G. (1997) Indices of tableting performance and application of percolation theory to powder compaction. *European Journal of Pharmaceutics and Biopharmaceutics*, **44**, 273–282.
- Jähne, B. (2013) *Digital Image Processing*. Berlin: Springer, 6th edn.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An Introduction to Statistical Learning*. New York: Springer.
- Kuentz, M. and Leuenberger, H. (1998) Modified Young’s modulus of microcrystalline cellulose tablets and the directed continuum percolation model. *Pharmaceutical Development and Technology*, **3**, 13–19.

- Lin, K. C. (2003) Fast image thresholding by finding the zero(s) of the first derivative of between-class variance. *Machine Vision and Applications*, **13**, 254–262.
- Meyer, F. (1994) Topographic distance and watershed lines. *Signal Processing*, **38**, 113–125.
- Meyer, F. and Beucher, S. (1990) Morphological segmentation. *Journal of Visual Communication and Image Representation*, **1**, 21–46.
- Müllertz, A., Perrie, Y. and Rades, T. (2016) *Analytical Techniques in the Pharmaceutical Sciences*. New York: Springer.
- Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability & its Applications*, **9**, 141–142.
- Ohser, J. and Schladitz, K. (2009) *3D Images of Materials Structures: Processing and Analysis*. Weinheim: J. Wiley & Sons.
- Olsson, H. and Nyström, C. (2001) Assessing tablet bond types from structural features that affect tablet tensile strength. *Pharmaceutical Research*, **18**, 203–210.
- Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R. and Wilkins, S. W. (2002) Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *Journal of Microscopy*, **206**, 33–40.
- Patel, A. C. and Markey, M. K. (2005) Comparison of three-class classification performance metrics: a case study in breast cancer CAD. In *Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment* (eds. M. P. Eckstein and Y. Jiang), vol. 5749, 581–590. International Society for Optics and Photonics, SPIE.
- Poquillon, D., Lemaitre, J., Baco-Carles, V., Tailhades, P. and Lacaze, J. (2002) Cold compaction of iron powders—relations between powder morphology and mechanical properties: Part I: Powder preparation and compaction. *Powder Technology*, **126**, 65–74.
- Ridler, T. W. and Calvard, S. (1978) Picture thresholding using an iterative selection method. *IEEE Transactions on Systems*, **8**, 630–632.
- Roerdink, J. B. T. M. and Meijster, A. (2000) The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, **41**, 187–228.
- Schlüter, S., Sheppard, A., Brown, K. and Wildenschild, D. (2014) Image processing of multiphase images obtained via X-ray microtomography: A review. *Water Resources Research*, **50**, 3615–3639.
- Sebhatu, T. and Alderborn, G. (1999) Relationships between the effective interparticulate contact area and the tensile strength of tablets of amorphous and crystalline lactose of varying particle size. *European Journal of Pharmaceutical Sciences*, **8**, 235–242.
- Sommer, C., Straehle, C., Koethe, U. and Hamprecht, F. A. (2011) Ilastik: Interactive learning and segmentation toolkit. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 230–233.
- Stenzel, O., Neumann, M., Pecho, O. M., Holzer, L. and Schmidt, V. (2017) Big data for microstructure-property relationships: A case study of predicting effective conductivities. *AIChE Journal*, **63**, 4224–4232.
- Washburn, E. W. (1921) Note on a method of determining the distribution of pore sizes in a porous material. *Proceedings of the National Academy of Sciences*, **7**, 115–116.
- Zhou, D. and Qiu, Y. (2010) Understanding material properties in pharmaceutical product development and manufacturing: Powder flow and mechanical properties. *Journal of Validation Technology*, **16**, 65–77.

FIGURES

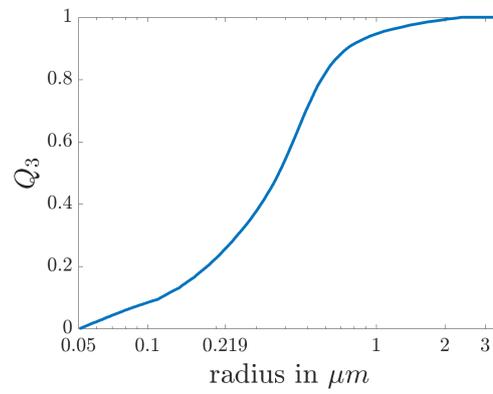


FIGURE 1. Experimentally determined MIP pore size distribution function Q_3 on a logarithmic scale.

TABLES

TABLE 1. Physical properties of the powders of interest: quantiles of volumetric/mass cumulative PSD and true/skeletal density.

Material	$X_{10}[\mu\text{m}]$	$X_{50}[\mu\text{m}]$	$X_{90}[\mu\text{m}]$	Span [-]	True Density [$\text{kg} \cdot \text{m}^{-3}$]
MCC	43.9	161.5	320.1	1.71	1573.0
Ibuprofen	7.7	25.4	69.2	2.42	1115.3

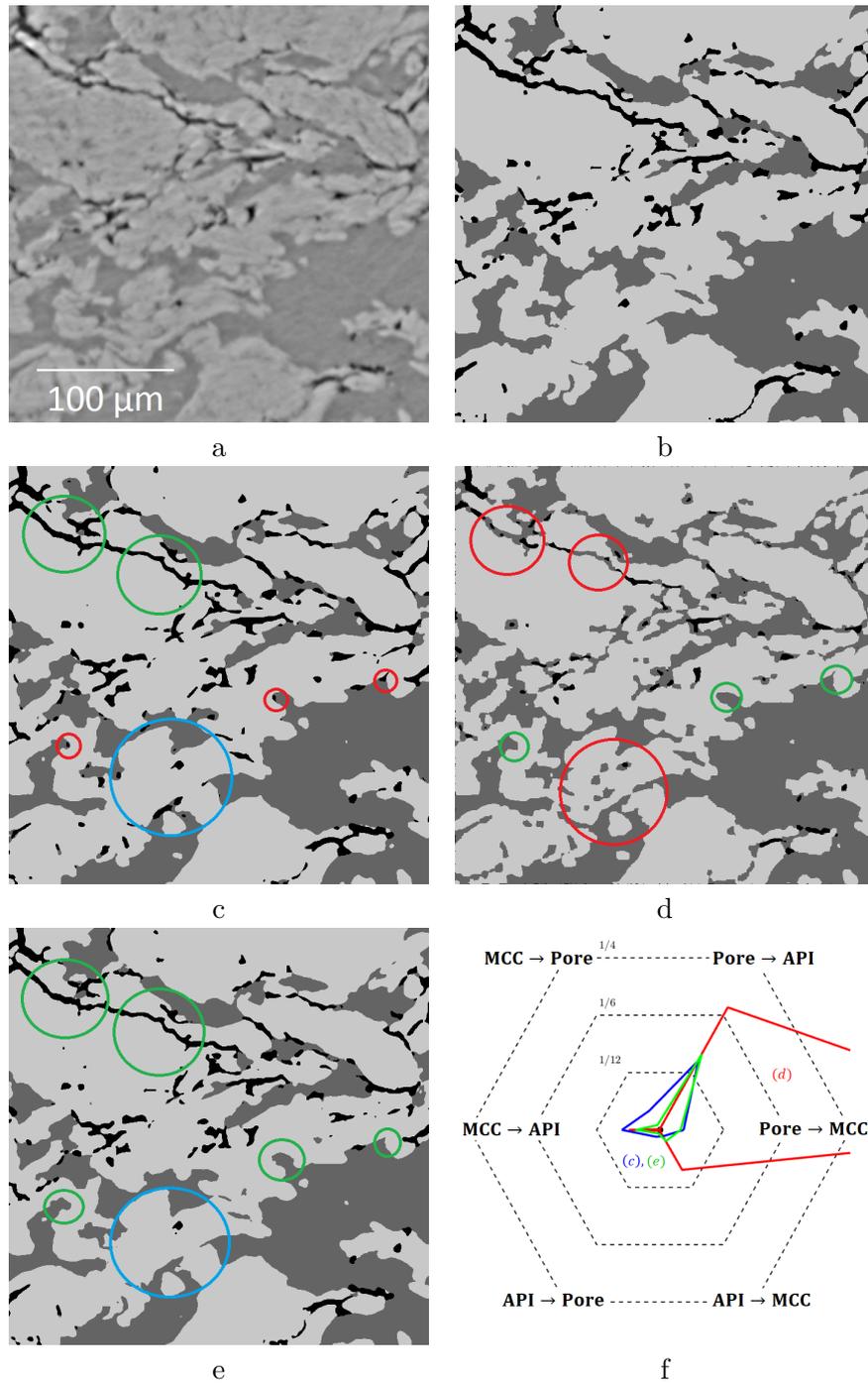


FIGURE 2. 2D slice of a cutout of greyscale image data in the xz -plane (a) and the corresponding manually trinarized slice (b) are shown. Different types of trinarization, i.e. the trinarization by the random forest algorithm (c), the watershed-based trinarization (d) and the hybrid approach (e) are represented. After trinarization, pores are visualized in black, API in dark grey and MCC in light grey. Examples for regions, where the corresponding trinarization algorithm works well are encircled in green and blue. Regions, where the algorithm leads to misclassifications, are in encircled in red. Numerical values of misclassifications are visualized by a cobweb graph (f), i.e. by hexagons, where each vertex represents a certain type of misclassification, e.g. pore voxels are classified as API (Pore \rightarrow API). In this case, the distance of a vertex to the origin is defined as the fraction of pore voxels classified as API. Note that using the watershed-based trinarization, the fraction of pore voxels classified as MCC is 0.55.

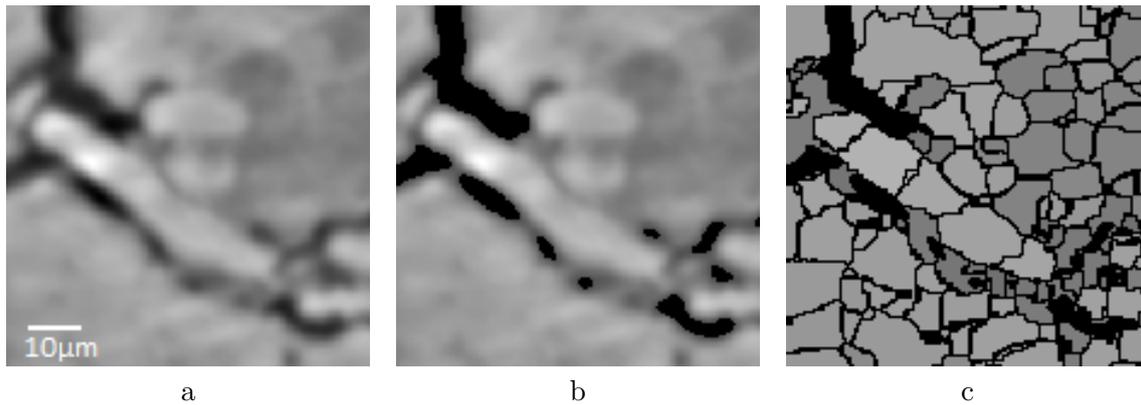


FIGURE 3. Visualization of the watershed segmentation at the example of a cutout of the same 2D slice, which is considered in Figure 2. At first the pores are determined (b, pores are represented in black) from the original greyscale image (a). Then, the image is partitioned in different watershed basins (c). The boundaries of watershed basins are represented in black and all voxels of each watershed basin are labeled with the average greyscale value of the basin.

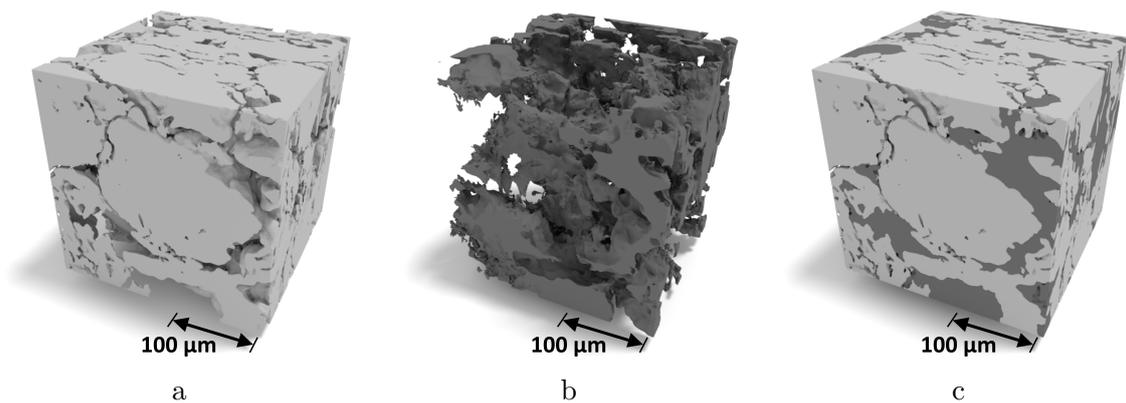


FIGURE 4. 3D visualization of MCC (a), API (b) and both components (c) determined by the hybrid approach combining a random forest algorithm with a watershed-based trinarization, where a cube of side length $306.6 \mu\text{m}$ is used for visualization.

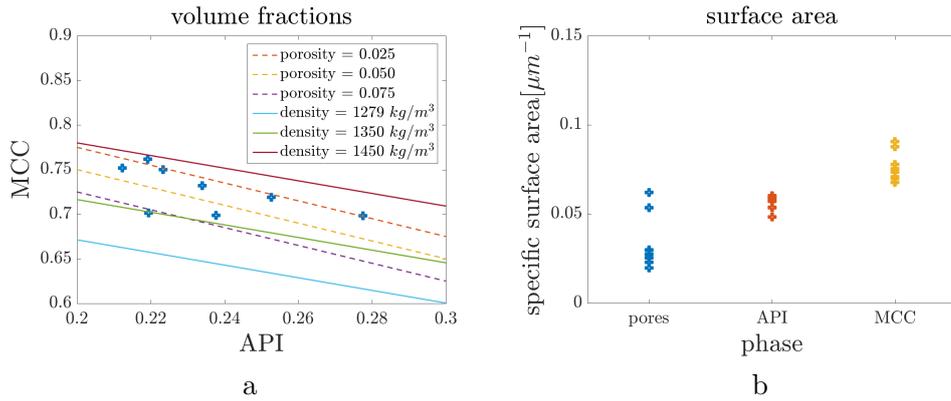


FIGURE 5. The volume fractions of API and MCC are represented by a blue dot for each of the eight cubic cutouts (a). Porosity as well as density of the cutouts are uniquely determined by volume fractions of API and MCC. The dashed/straight lines show volume fractions of MCC over the volume fractions of API for given values of porosity/density. Specific surface areas of the three phases (b).

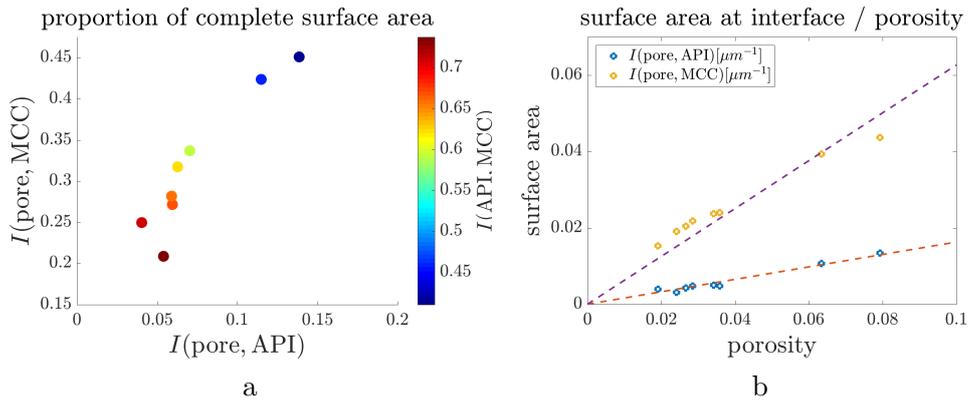


FIGURE 6. Surface areas of interfaces per unit cube are visualized: For the different interface areas per unit cube, their percentage of total surface area is visualized. The values of $I(\text{Pore, MCC})$ are plotted over $I(\text{Pore, API})$, while the value of $I(\text{API, MCC})$ is indicated by the color bar (a). The values of $I(\text{Pore, API})$ and $I(\text{Pore, MCC})$ correlate strongly with porosity (b).

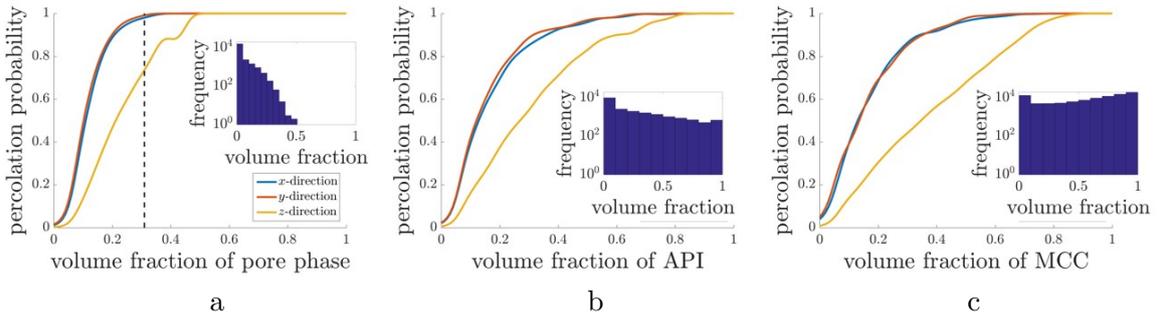


FIGURE 7. Local percolation probabilities estimated for pore phase (a), API (b) and MCC (c) based on sub-cubes with an edge length of 22 μm . Percolation probabilities are computed for x -, y -, and z -direction. The blue histograms show the frequency of volume fractions (of the corresponding phase) within the considered sub-cubes. For the estimation of local percolation probabilities of the pore space, we consider the values of $P(v)$ to be meaningful, which are located on the left side of the black dotted line.

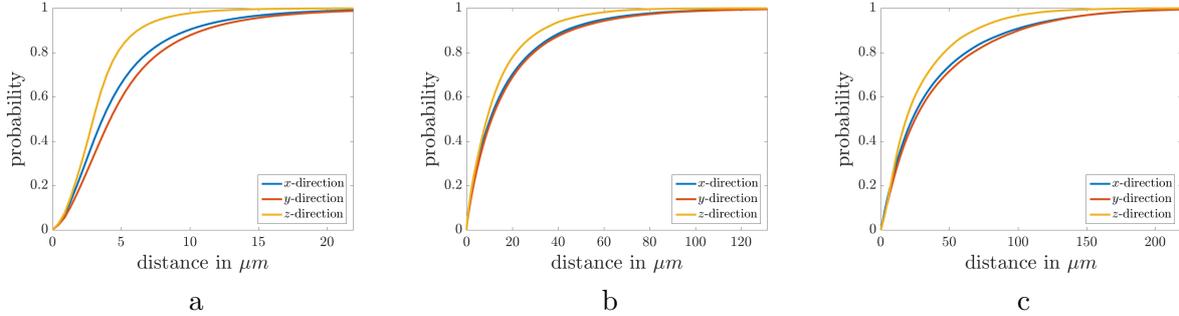


FIGURE 8. Mean chord length distribution functions estimated from image data. Chord lengths are computed for pores (a), API (b) and MCC (c).