

# Correlating pore space morphology with soil gas diffusion for structured loam and sand, including stochastic 3D microstructure modeling

Benedikt Prifling<sup>1,\*</sup>, Matthias Weber<sup>1,\*</sup>, Maximilian Rötzer<sup>2</sup>, Nadja Ray<sup>3</sup>,  
, Alexander Prechtel<sup>2</sup>, Maxime Phalempin<sup>4</sup>,  
Steffen Schlüter<sup>4</sup>, Doris Vetterlein<sup>4</sup>, Volker Schmidt<sup>1</sup>

December 23, 2024

5

10

15

20

**Abstract.** Biogeochemical soil processes are closely linked to the structure of soil. In particular, nutrient transport depends on diffusivity and permeability within the soil's pore network. A deeper understanding of the relationship between microscopic soil structure and such effective macroscopic properties can be obtained by tomographic imaging combined with a quantitative analysis of soil morphology and numerical simulations of effective macroscopic properties. In a previous work it has been shown that different parametric regression formulas can be used to predict these relations for finely sieved soils of loam and sand. In the present paper, we validate these formulas and further extend their applicability to structured soils. In particular, 3D CT data of a total of six samples, consisting of three loam and three sand samples, are used as the basis for an extensive structural analysis. As expected, the performance of most regression formulas can be improved by specifically adjusting their parameters for the considered soil structures. However, it turns out that some regression formulas based on, e.g., tortuosity which were fitted for finely sieved soils still reliably predict diffusion for structured soils without adjusting their parameters. For additional validation and improvement of the considered regression formulas, artificially generated soil structures can be utilized. Therefore, a method for the generation of such structures via samples drawn from a parametric stochastic 3D microstructure model is outlined which may serve as a basis for further work.

**Keywords:** structure-property relationship, soil gas diffusion, 3D CT data, statistical image analysis, stochastic 3D modeling

25

---

<sup>1</sup>Institute of Stochastics, Ulm University, Ulm, Germany

<sup>2</sup>Department of Mathematics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany

<sup>3</sup>Mathematical Institute for Machine Learning and Data Science, Catholic University of Eichstätt-Ingolstadt, Ingolstadt, Germany

<sup>4</sup>Department of Soil System Science, Helmholtz-Centre for Environmental Research – UFZ, Halle, Germany

\*Both authors contributed equally

# 1 Introduction

The three-dimensional microstructure of many natural and technical materials plays a decisive role with regard to effective macroscopic properties [1, 2, 3]. One example of an effective macroscopic property is the effective diffusivity, which is of major importance for a wide spectrum of technical applications ranging from lithium-ion batteries to solid-oxide fuel cells [4, 5, 6, 7]. However, diffusion processes are not only of great importance in technical materials, but also in natural materials such as soils. In particular, it is well known that the 3D microstructure of the pore space has a profound impact on the transport of gas and water in soils [8, 9]. While many well-established relations formulate a dependence of the scalar diffusion coefficient only on the porous medium's porosity (see [10] for a review of different approaches), more precise structure-property for soil gas diffusion have been recently investigated based on 3D CT data of finely sieved soils [11]. More precisely, the parameters of analytical regression formulas, which link geometrical descriptors of 3D microstructure such as tortuosity or chord length distribution and diffusive properties computed via numerical simulations, have been calibrated in [11] to 3D image data of finely sieved loam and sand, as well as compared to structure-property relationships from the literature. The focus of the present paper is to use 3D image data of structured soils (again considering loam and sand as representatives of two different soil textures), where we first investigate to which extent the 3D morphology of structured soil differs from that of finely sieved soil. Moreover, the structure-property relationships derived in [11] are applied to and evaluated for structured soils. In doing so, we investigate to which extent these microstructure-property relationships hold for the present data of more realistic soil structures. This allows for a deeper, quantitative understanding of soil gas diffusion, which has a significant impact on nutrient supply and plant growth [8].

To develop even more general and robust quantitative structure-property relationships that generalize for various soil textures, stochastic 3D models can be used in order to create more comprehensive data sets for model calibration and validation than this is possible by tomographic image data. Thus, to illustrate this approach, the present paper also covers the development of a parametric stochastic 3D microstructure model, which is calibrated to 3D image data of structured sand. More precisely, excursion sets of random fields are used to model the 3D microstructure of the soil. Similar modeling ideas have been applied to generate digital twins of the 3D microstructure of lithium-ion batteries [12, 13], solid-oxide fuel cells [14, 15, 16], and gas-diffusion electrodes [17]. Model validation is carried out by comparing geometrical descriptors of simulated 3D structures and experimental image data, which have not been used for model calibration. This lays ground for generating virtual but realistic 3D soil structures just at the cost of computer simulations by systematic variation of the model parameters.

The outline of the present paper is as follows. First, the acquisition of six structured soil samples is described in Section 2.1, which are then imaged via 3D CT, see Section 2.2. Next, a brief description of the computation of diffusive properties (Section 2.3) and various geometrical descriptors (Section 2.4) is given. The results of a statistical analysis of the 3D pore space morphology of structured soils, in comparison to finely sieved soils, are provided in Section 3.1. Subsequently, quantitative structure-property relationships are investigated in Section 3.2, where once again a comparison with the results obtained in [11] is carried out. The calibration and validation of a parametric stochastic 3D microstructure model based on excursion sets of random fields is described in Section 4. A summary of the main results achieved in the present paper and an outlook to future research activities are given in Section 5.

## 2 Materials and methods

This section is focusing on the explanation of the acquisition of six structured soil samples (Section 2.1), their imaging via 3D CT (Section 2.2), the numerical calculation of diffusive properties of the underlying pore space (Section 2.3), and various geometrical descriptors of 3D microstructures (Section 2.4).

### 2.1 Sample acquisition

The six soil samples considered in the present paper are part of a previously published dataset on rhizosphere soil structure and its interplay with soil textures (i.e., loam versus sand) [18]. In short, the samples were taken in 2019 at the experimental station of the UFZ (Umweltforschungszentrum) in Bad Lauchstädt, Germany (51°22'0" N, 11°49'60" E). The structured soil has been sieved with a mesh size of 20 mm before filling into 75 cm deep excavation pits and subsequent compaction with a vibrating plate [19], followed by six months of natural soil settling. Aluminum rings of 5 cm in height and diameter were extracted at 6 and 16 cm depths with a set-up that allows to preserve the intact soil structure of the samples. Directly after sampling, the samples were brought to the laboratory and kept at 4° before X-ray CT scanning. Note that the finely sieved soil considered in [11] has been sieved with a mesh size of 1 mm. For more information on the experimental set-up and design, the reader is referred to [18, 19].

### 2.2 X-ray computed tomography scanning and binarization

X-ray CT scanning was performed with an industrial  $\mu$ -CT scanner (X-TEK XTH 225, Nikon Metrology) having an Elmer-Perkin 1620 detector panel (1750 × 2000 pixels). The obtained images were reconstructed into an 8-bit grayscale 3D tomogram having a voxel size of 25  $\mu$ m. The reconstruction was performed with a filtered back projection algorithm with the CT Pro 3D software (Nikon metrology). The conversion to 8-bit allowed saving considerable space without losing considerable information. During the 8-bit conversion, the grayscale range was normalized with a percentile stretching method. Further details regarding the number of projections, energy and current settings of the X-ray CT scanner are thoroughly given in [18].

The segmentation into pore space, where diffusion occurs, and the soil matrix, where no diffusion can take place, was performed by a global thresholding method, see Figure 1. For the choice of appropriate threshold values, the fuzzy c-means clustering method [20] implemented in Quantim v.4 was used. After the phase-based segmentation, the images were cropped in cuboids with a side length of 1000 voxels in order to exclude zones of the samples which could have potentially been disturbed during sampling. The binarized 3D image data of each of the six samples was then partitioned into 446 non-overlapping cutouts such that the pore space is completely connected in  $x$ -direction,  $y$ -direction and  $z$ -direction, respectively. Thus, a total of 2676 cutouts was considered. The latter condition was required to ensure the convergence of the numerical simulations conducted in Section 2.3.

Compared to the data investigated in [11], the present 3D CT images differ with regard to their voxel size. While this may impede the quantitative comparison between finely sieved soil considered in [11] and structured soil, both datasets form a suitable basis for the development and assessment of regression formulas for the prediction of diffusion from geometrical descriptors. In particular, we will see in Section 3 that despite pronounced structural differences, caused by the different kinds of soil and the different voxel

105 sizes, quite accurate structure-property relationships can be obtained by directly applying the regression formulas derived in [11] to the present data of structured soil.

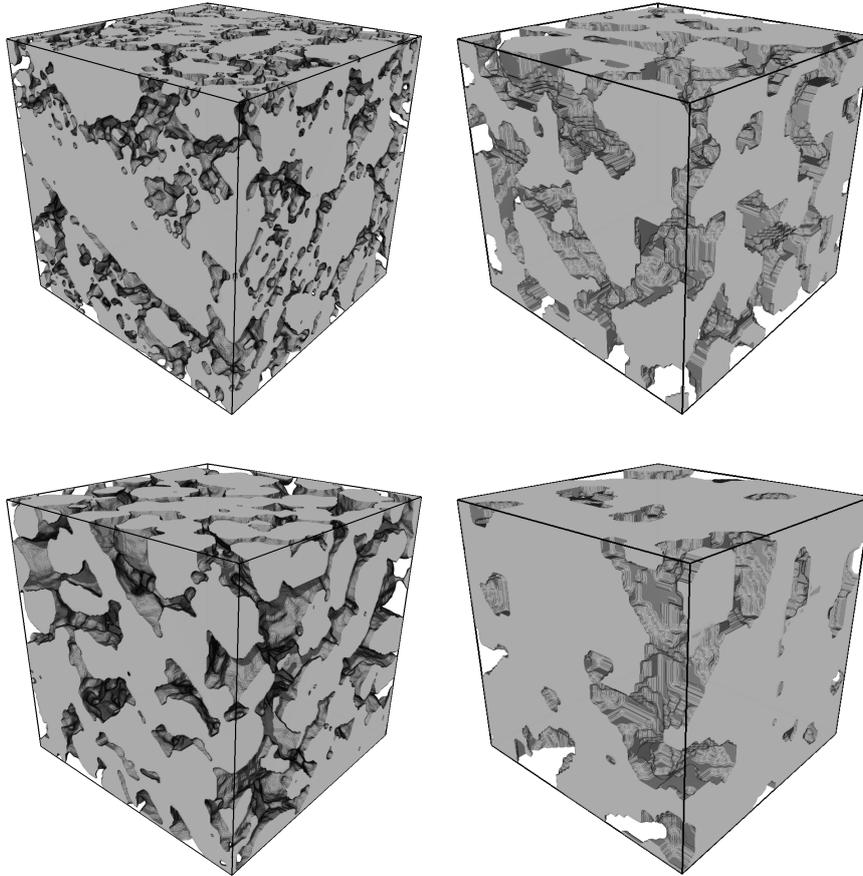


Figure 1: Three-dimensional renderings of selected loam (top row) and sand (bottom row) subsamples ( $1.28 \mu\text{m} \times 1.28 \mu\text{m} \times 1.28 \mu\text{m}$ ), where the solid phase is depicted in gray. Left: finely sieved soil as considered in [11] Right: structured soil considered in the present paper.

### 2.3 Computation of diffusive properties

For each of the cutouts described in Section 2.2, effective diffusive properties are computed by the method of periodic homogenization [21]. Further details regarding these computations are described in [11, 10]. As a result, we obtain a  $3 \times 3$  diffusion tensor  $\mathbb{D} = (\mathbb{D}_{i,j})_{i,j=1,2,3}$  for each cutout. A normalized characteristic that quantifies diffusive properties is the so-called M-factor, denoted by  $M$ , which is defined as the ratio of effective diffusivity and intrinsic diffusivity  $D_m$  [22]. Since all soils have been sieved prior to the field experiments [18, 19] there are no pronounced anisotropy effects in the present image data. The effective diffusivity is thus calculated by averaging the diagonal entries of  $\mathbb{D}$ , which leads to

115 
$$M = \frac{1}{3D_m} (\mathbb{D}_{11} + \mathbb{D}_{22} + \mathbb{D}_{33}).$$

### 2.4 Geometrical descriptors of pore space

The geometrical descriptors that are used in the present paper for quantifying the 3D pore space morphology and for establishing structure-property relationships are summarized in Table 1.

Geometrical descriptor	Symbol	Range	Unit
Volume fraction	$\varepsilon$	[0,1]	-
Specific surface area	$S$	[0, $\infty$ )	$\mu\text{m}^{-1}$
Mean chord length	$\mu(C)$	[0, $\infty$ )	$\mu\text{m}$
Characteristic bottleneck radius	$r_{\min}$	[0, $\infty$ )	$\mu\text{m}$
Characteristic pore size radius	$r_{\max}$	[0, $\infty$ )	$\mu\text{m}$
Constrictivity	$\beta$	[0,1]	-
Mean spherical contact distance	$\mu(H)$	[0, $\infty$ )	$\mu\text{m}$
Mean geodesic tortuosity	$\mu(\tau)$	[1, $\infty$ )	-
Standard deviation of geodesic tortuosity	$\sigma(\tau)$	[0, $\infty$ )	-

Table 1: Overview of geometrical descriptors of pore space.

Further details on geometrical descriptors of the pore space in soils, including additional references, as well as details regarding the computation of these descriptors from voxelized 3D image data can be found in [11]. Note that the mean and standard deviation of the geometric tortuosity, which is based on the skeleton of the pore space, has been considered in [11], but turned out to be not as useful as the concept of geodesic tortuosity with regard to the derivation of structure-property relationships. Thus, the geometric tortuosity of pore space is no longer considered in the present paper.

## 3 Correlating pore space morphology with soil gas diffusion

In this section, we present the results of a statistical analysis of the 3D pore space morphology of structured soils, comparing them to those obtained for finely sieved soils (Section 3.1), and we derive quantitative structure-property relationships for structured soils, where once again a comparison with corresponding relationships obtained in [11] is carried out (Section 3.2).

### 3.1 Statistical analysis of 3D pore space morphology

The 3D morphology of structured soil samples is investigated by means of the geometrical descriptors discussed in Section 2.4. The obtained results are compared to those achieved in [11] for finely sieved soils, see Figure 2. However, note that the voxel size of tomographic image data considered in [11] is equal to  $10\mu\text{m}$ , whereas the image data considered in the present paper have a coarser resolution with a voxel size of  $25\mu\text{m}$ , *i.e.*, the image data now examined for structured soils contains significantly fewer structural details than the higher-resolution image data for finely sieved soils considered in [11]. This fact has to be taken into account when comparing the results for the two sets of image data.

#### 3.1.1 Univariate distributions of single descriptors

It turns out that the local porosities, *i.e.*, the volume fractions  $\varepsilon$  of pore space computed from structured sand samples are significantly larger compared to the porosity values of structured loam, while this difference is much less pronounced for finely sieved soils, see Figure 2a. In case of the specific surface area

$S$  of the pore space, significant differences between loam and sand are apparent. The values obtained for structured sand samples are significantly larger compared to those for structured loam, where the opposite is true in case of finely sieved soil, see Figure 2b. The latter effect might be attributed to the different spatial resolutions of both types of data sets. The coarser resolution of image data considered in the present paper is also likely the reason for a significantly larger mean chord length  $\mu(C)$  of the pore space (for both structured loam and sand) compared to the values obtained in [11], since two or more chords within the pore space can now be detected as one single chord if the intermediate soil structure is too fine to be resolved with the coarser resolution. Interestingly, the pore space of the loam samples exhibits shorter mean chord lengths for finely sieved soil, whereas the opposite trend is observed with regard to structured soil, see Figure 2c.

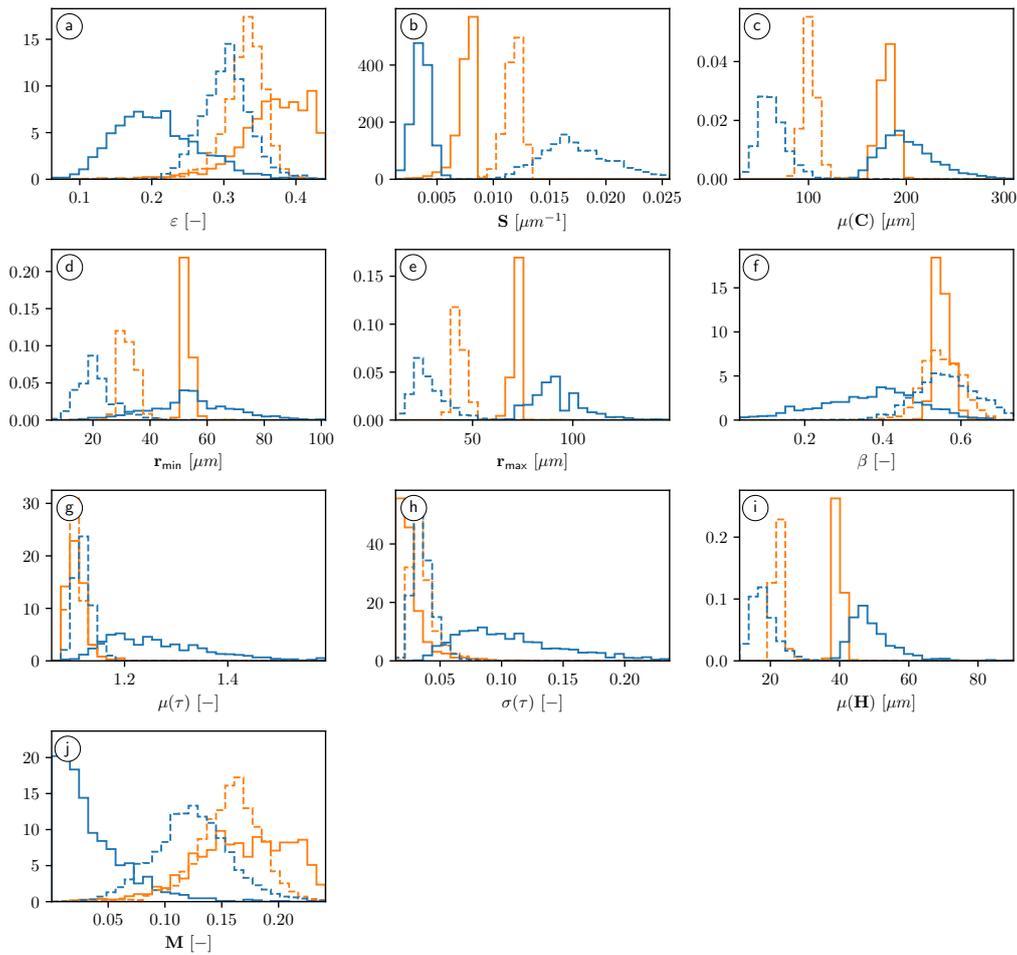


Figure 2: Histograms of geometrical and functional descriptors of pore space: Volume fraction (a), specific surface area (b), mean chord length (c), characteristic bottleneck radius (d), characteristic pore radius (e), constrictivity (f), mean and standard deviation of geodesic tortuosity (g,h), mean distance to solid phase (i), and the M-factor (j) computed from tomographic image data for the  $2 \cdot 3 \cdot 446 = 2676$  subsamples of loam (blue) and sand (orange), respectively. Solid lines represent structured soils while dashed lines correspond to finely sieved soils.

The coarsening effect due to a larger voxel size is also observed with regard to both characteristic radii  $r_{\min}$  and  $r_{\max}$ , see Figures 2d and 2e. With regard to constrictivity  $\beta$ , finely sieved soils showed rather similar distributions for both loam and sand, which is no longer true for structured soils. In particular, in the latter case there are pronounced differences regarding both the mean values and variances of

constrictivity within each sample, see Figure 2f. With regard to the mean geodesic tortuosity  $\mu(\tau)$ , both types of finely sieved soils exhibited very similar distributions with a mean around 1.1, see Figure 2g. This is still true for structured sand, but not for structured loam, which exhibits a broad distribution of mean geodesic tortuosities. The latter distribution might arise due to the coarser resolution that is not able to resolve narrow throats within the pore phase such that the shortest paths within the pore space now consist of significantly more detours. This would also explain the plots in Figure 2h showing the local distributions of  $\sigma(\tau)$ , where again the structured loam behaves differently compared to structured sand as well as both finely sieved soil types. In case of the mean distance  $\mu(H)$  to the solid phase, the behavior is similar to the mean chord length  $\mu(C)$ , see Figure 2i, which is intuitively clear since both geometrical descriptors quantify some kind of typical pore size.

The distributions of the local M-factor of structured soils show pronounced differences compared to those of finely sieved soils, see Figure 2j. More precisely, both types of structured soil show broader distributions of the local M-factor. Moreover, the distribution of the M-factor corresponding to loam is now skewed to the left with a majority of values below 0.05, indicating severe limitations for soil gas diffusion within the pore space. This is likely to be caused by the significant shift of the distribution of the (local) mean geodesic tortuosity  $\mu(\tau)$  to the right compared to finely sieved loam.

Overall, these investigations highlight the diversity in the descriptors depending on soil type and preparation. This observed diversity is beneficial for the development of generalized prediction formulas.

### 3.1.2 Bivariate distributions of descriptor pairs

Besides the univariate distributions discussed so far, we also consider bivariate distributions of selected pairs of geometrical and functional descriptors, see Figure 3.

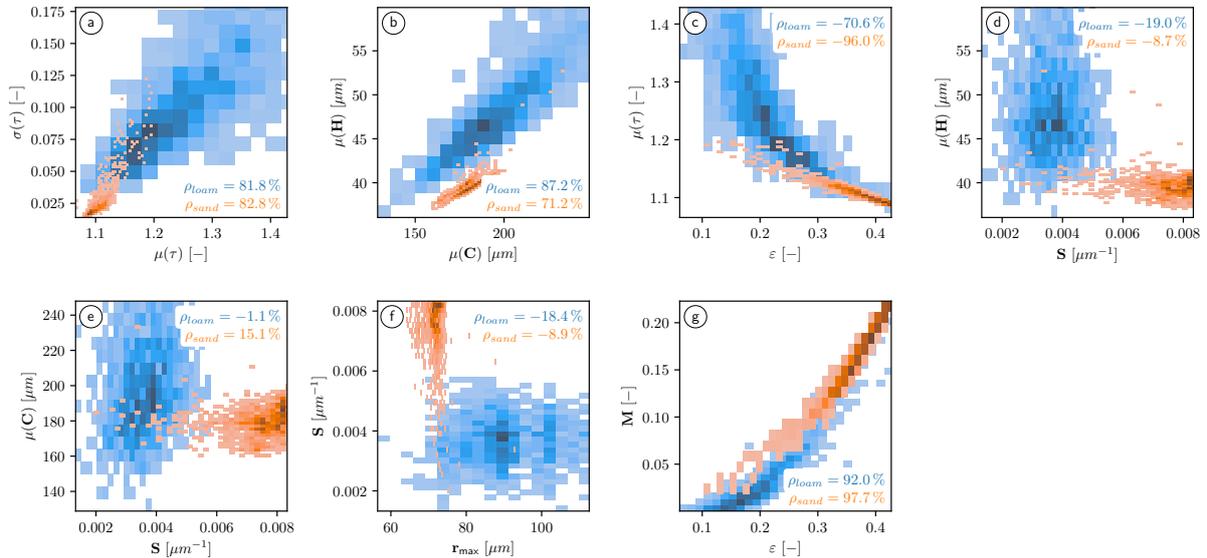


Figure 3: Bivariate probability densities for pairs of geometrical and functional descriptors, visualized as contour plots based on kernel density estimation, which have been obtained for the 2676 cubic subsamples of either structured loam (blue) or structured sand (orange). Furthermore, the values of Pearson's correlation coefficient  $\rho$  are given for each descriptor pair, separately computed for loam and sand.

It turns out that there is nearly no correlation between the specific surface area of the pore space  $S$  on the one hand, and  $\mu(H)$ ,  $\mu(C)$  and  $r_{\max}$  on the other hand, see Figures 3d - 3f. Note that in case of sand, this is similar to finely sieved soil, but the pronounced positive correlation between  $S$  and the other three geometrical descriptors mentioned above that has been observed in [11] for finely sieved loam is no longer observed for the present data. Moreover, there is a pronounced positive correlation between  $\mu(\tau)$  and  $\sigma(\tau)$ , which is similar for loam and sand, see Figure 3a. Note that there is also a pronounced positive correlation between  $\mu(C)$  and  $\mu(H)$ , where this correlation is larger for structured loam compared to structured sand, see Figure 3b. However, in case of finely sieved soils, this correlation has been even more pronounced for both loam and sand [11]. The most pronounced correlations can be observed between the porosity  $\varepsilon$  and the mean geodesic tortuosity  $\mu(\tau)$  as well as between  $\varepsilon$  and the M-factor  $M$ , see Figures 3c and 3g. In both cases, the strongest correlation is observed for structured sand with correlation coefficients of -0.96 and 0.977, respectively. In particular, the correlation coefficient between  $\varepsilon$  and  $M$  for structured sand, which is close to one, indicates that analytical formulas for the prediction of  $M$  from geometrical descriptors that only use the porosity  $\varepsilon$  as input are likely to lead to decent prediction accuracy. However, as will be shown in later in Section 3.2, combining the porosity with further more sophisticated geometrical descriptors allows for an even more accurate prediction of the M-factor.

### 3.2 Regression formulas for the quantification of structure-property relationships

We now investigate the question how well do the prediction formulas for the M-factor derived in [11] for finely sieved soil behave when they are directly applied to structured soil. Furthermore, we show how do the regression coefficients change compared to finely sieved soil when they are newly calibrated to the present tomographic image data measured for structured soil. An overview of the analytical prediction formulas for the M-factor considered in the following is given in Table 2.

Formula	Eq. number	Constraints	Values derived in [11]
$\widehat{M}_1 = \varepsilon^{c_1}$	(1)	$c_1 \geq 1$	1.71
$\widehat{M}_2 = c_1 \varepsilon^{c_2}$	(2)	$c_1, c_2 > 0$	(1.92, 2.29)
$\widehat{M}_3 = c_1 e^{c_2 \varepsilon}$	(3)	-	(0.02, 6.88)
$\widehat{M}_4 = \varepsilon^{c_1} \beta^{c_2} \mu(\tau)^{c_3}$	(4)	$c_1 >, c_2 \geq 0, c_3 \leq 0$	(1.06, -0.04, -7.28)
$\widehat{M}_5 = \varepsilon^{c_1 + c_2 \beta} \mu(\tau)^{c_3}$	(5)	$c_1 + c_2 \geq 0, c_3 \leq 0$	(1.17, -0.18, -7)
$\widehat{M}_6 = c_1 \mu(\tau)^{c_2} \sigma(\tau)^{c_3} \varepsilon^{c_4}$	(6)	-	(2.19, -6.07, 0.07, 1.63)
$\widehat{M}_7 = \varepsilon^{c_1} S^{c_2} r_{\max}^{c_3}$	(7)	-	(2.6, -0.61, -0.45)
$\widehat{M}_8 = \varepsilon^{c_1} \left( \frac{\mu(C)}{\mu(H)} \right)^{c_2}$	(8)	-	(2.23, 0.16)

Table 2: Overview of the considered analytical prediction formulas for the M-factor, including the constraints on the coefficients as well as the numerical values for the coefficients, which have been derived in [11] for finely sieved soil.

200 To obtain a first impression regarding the influence of geometrical descriptors of the pore space on the M-factor for structured soil, the scatter plots shown in Figure 4 illustrate the dependence of the M-factor on various pairs of geometrical descriptors.

Moreover, we consider several empirically derived regression formulas which will be used for the quantification of structure-property relationships and, in particular, for predicting the M-factor from the knowledge of an appropriately chosen vector of geometrical descriptors of the 3D pore space morphology. 205 A detailed evaluation of the performance of the regression formulas and their parameters when fitted to the data of structures soils can be found in the Appendix. For an overview, we refer to Figure 5. In summary, the formula given in Eq. (6), which was the best-performing regression formula in [11], as well as the formulas given in Eqs. (4) and (5) generalize (using the parameters originally derived in [11]) to tomographic image data of structured soils, see Figure 12 in the Appendix. The fit of these formulas 210 can be further increased by adjusting their parameters specifically for structured soils, see Figure 11 in the Appendix. In both cases, the formula given in Eq. (6) is still the best-performing or close to the best-performing formula. In particular, the prediction accuracy of rather simple prediction formulas, which only include the porosity as geometrical descriptor (i.e. the formulas in Eqs. (1), (2) and (3)) can be increased by incorporating the mean and standard deviation of geodesic tortuosity. Furthermore, the formulas that are solely based on porosity are less robust in a sense that using that different data sets (either finely sieved soil instead of structured soil or loam instead of sand) can not be described by those equations without re-calibrating the coefficients, see Figures 11 and 12. Interestingly, Eq. (7) is not the best-performing prediction formula despite a correlation coefficient of 0.94 between the specific surface area and the M-factor. Moreover, Eqs. (4) and (5), which both are based on the porosity, mean geodesic tortuosity and the constrictivity, are also performing well, even when using the coefficients obtained for 215 finely sieved soil and applying them to structured soil, see Figure 12.

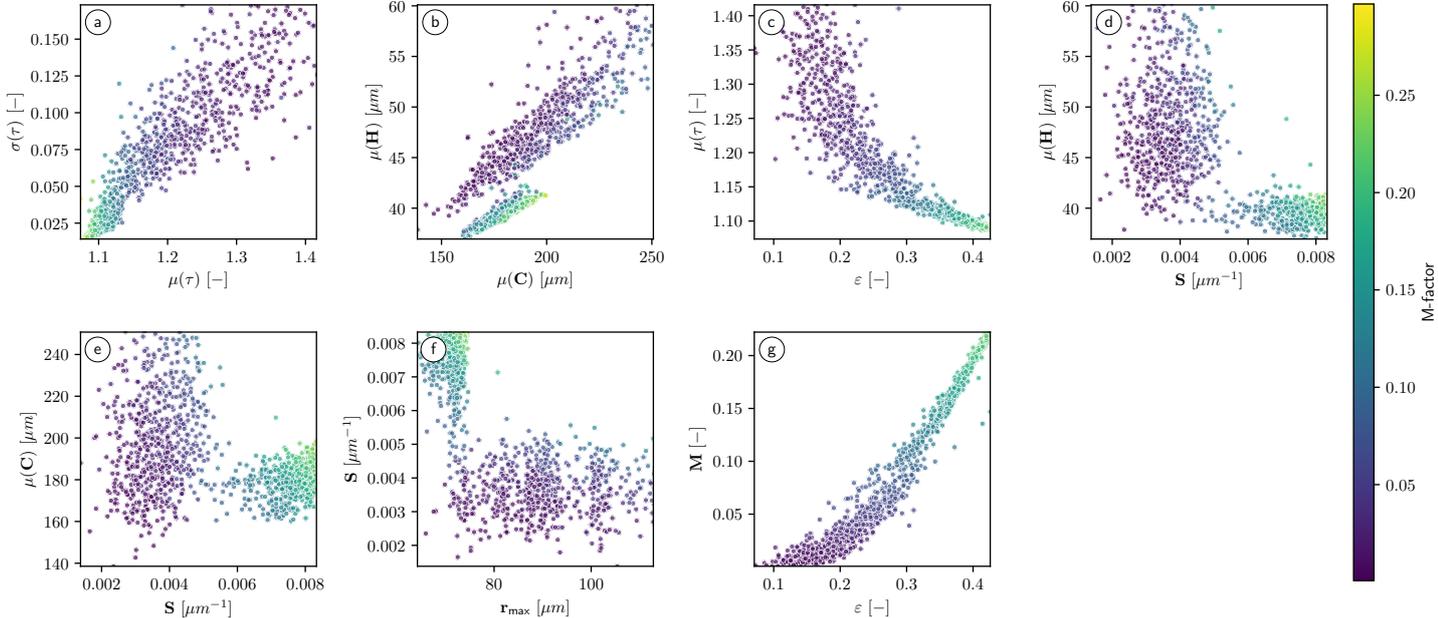


Figure 4: Scatter plots visualizing the dependence of the M-factor on various pairs of geometrical descriptors. The (brighter or darker) colors of the data points indicate whether the value of the corresponding M-factor is high or low. Note that in this figure we do not distinguish between the two soil textures which can be seen in Figure 3. For clarity of presentation, only randomly selected 10% of the 2676 cubic subsamples of structured loam and sand have been used to generate the 461 data points in each subfigure.

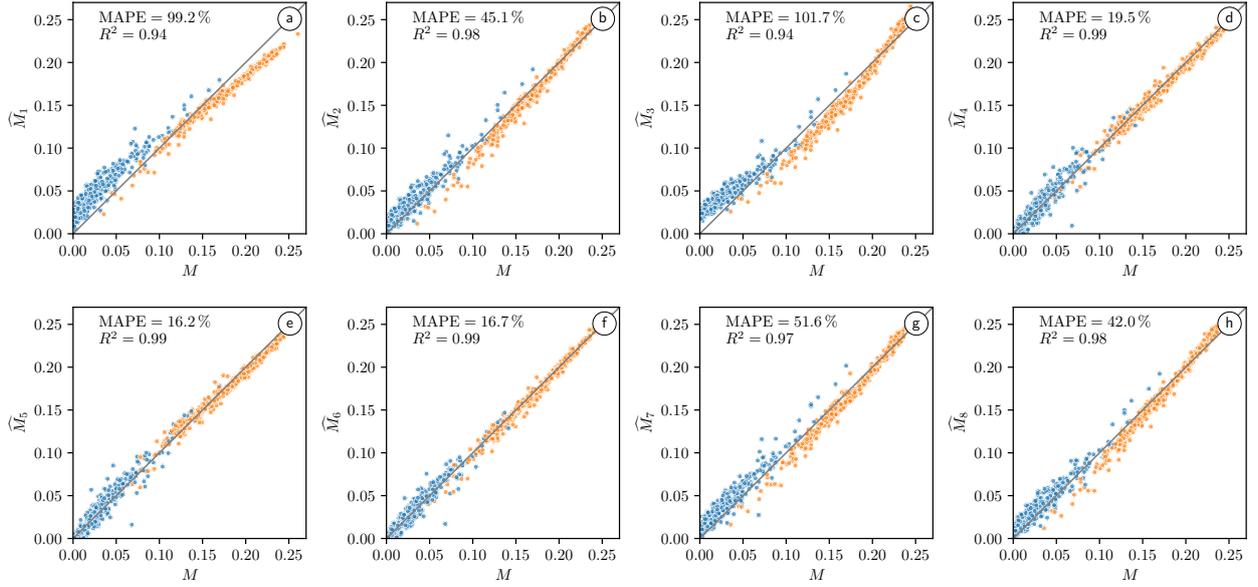


Figure 5: Scatter plots visualizing the values obtained for the M-factor  $M$  by numerical simulations versus those of the predicted M-factors  $\widehat{M}_i$ ,  $i = 1, \dots, 8$ , obtained for the test data of either structured loam (blue) or structured sand (orange). The parameters appearing in the prediction formulas given in Eqs. ((1))–((8)) have been fitted to the entire training data for loam and sand, see Table 4 in the Appendix.

## 4 Stochastic 3D microstructure modeling

As measured CT images are time and cost-intensive, they can only provide a limited amount of data – both in terms of a rather small number of measurements as well as a comparatively small structural variety. Thus, prediction formulas which are developed solely based on measured image data may have limited explanatory power for other soil textures. In the following, we present a stochastic 3D model for the microstructure of soils which may be used to generate a large variety of realistic soil structures to overcome the limitations of measured image data. More precisely, a parametric 3D model is calibrated to tomographic image data of structured sand. By a quantitative comparison of geometrical descriptors computed on measured as well as simulated soil structures, it is shown that the complex 3D morphology of structured sand can be characterized by a surprisingly small number of interpretable model parameters. Thus, the stochastic 3D model proposed in this section lays ground for future investigations, including virtual scenario analyzes by systematic variation of model parameters.

### 4.1 Model description

The modeling approach described in this section is motivated by the observation that the 3D image data of structured sand mainly shows two different classes of particles, namely a few large and many small ones. The stochastic microstructure model can be decomposed into three steps. First, we consider two independent Gaussian random fields  $X_1 = \{X_1(t), t \in \mathbb{R}^3\}$  and  $X_2 = \{X_2(t), t \in \mathbb{R}^3\}$ , where both random fields are assumed to be stationary and isotropic such that their distributions are uniquely characterized by two mean values  $\mu_1, \mu_2 \in \mathbb{R}$  and two covariance functions  $C_1, C_2 : [0, \infty) \rightarrow \mathbb{R}$ . Furthermore, we assume that the random fields  $X_1$  and  $X_2$  are normalized, *i.e.*, it holds that  $\mu_1 = \mu_2 = 0$  and  $C_1(0) = C_2(0) = 1$ ,

and that their covariance functions are given by

$$C_i(r) = \frac{2^{1-\nu_i}}{\Gamma(\nu_i)} \left( \sqrt{2\nu_i} \frac{r}{\rho_i} \right)^{\nu_i} K_{\nu_i} \left( \sqrt{2\nu_i} \frac{r}{\rho_i} \right),$$

for each  $r \geq 0$ , where  $\nu_i > 0$  and  $\rho_i > 0$  are some parameters for  $i \in \{1, 2\}$ . Here,  $\Gamma$  and  $K_{\nu_i}$  denote the Gamma function and the modified Bessel function of the second kind, respectively [23].

Next, for  $i \in \{1, 2\}$ , we consider the random sets  $\Xi_i = \{t \in \mathbb{R}^3 : X_i(t) \geq a_i\}$  as level sets (also called excursion sets) of the random fields  $X_1$  and  $X_2$ , where  $a_1, a_2 \in \mathbb{R}$ . Due to the Gaussian nature of the random fields  $X_1$  and  $X_2$ , the following analytical relationship between the volume fraction  $\varepsilon_i = \mathbb{P}(o \in \Xi_i)$  of  $\Xi_i$ , where  $o \in \mathbb{R}^3$  denotes the origin, and the threshold  $a_i$  is true for  $i \in \{1, 2\}$ . Namely, it holds that  $\varepsilon_i = 1 - \Phi(a_i)$  for  $i \in \{1, 2\}$ , where  $\Phi : \mathbb{R} \rightarrow [0, 1]$  denotes the cumulative distribution function of the univariate standard normal distribution [24].

Finally, we model the solid phase of structured soils as union  $\Xi = \Xi_1 \cup \Xi_2$  of the level sets  $\Xi_1$  and  $\Xi_2$ . Note that the independence of  $X_1$  and  $X_2$  implies that the random fields  $\Xi_1$  and  $\Xi_2$  are also independent. Thus, the volume fraction  $\varepsilon = \mathbb{P}(o \in \Xi)$  of  $\Xi$  is given by  $\varepsilon = \varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2$ .

## 4.2 Model calibration

To fit the model described in Section 4.1 to measured 3D image data, note that the volume fraction  $\varepsilon$  of the solid phase can be easily estimated by voxel counting [24]. Thus,  $\varepsilon$  is known and can be used to exploit the relationship  $\varepsilon_2 = \frac{\varepsilon - \varepsilon_1}{1 - \varepsilon_1}$ , which allows us to express  $\varepsilon_2$  in dependence of  $\varepsilon_1$ . Furthermore, this equation can be transformed to obtain the following relationship between the thresholds  $a_1$  and  $a_2$ :

$$a_2 = \Phi^{-1} \left( \frac{1 - \varepsilon}{\Phi(a_1)} \right). \quad (9)$$

This means that model calibration can be performed by means of the reduced parameter vector  $\theta = (a_1, \nu_1, \rho_1, \nu_2, \rho_2)$ , since  $a_2$  is obtained from  $a_1$  via Eq. (9), where the parameter vector  $\theta$  is chosen such that simulated 3D data drawn from the calibrated stochastic model are statistically similar to the observed sand structures. More precisely, the parameter vector  $\theta$  is fitted such that the chord length distributions within the solid phase as well as within the pore space match between measured and simulated data, as shown in Figure 6.

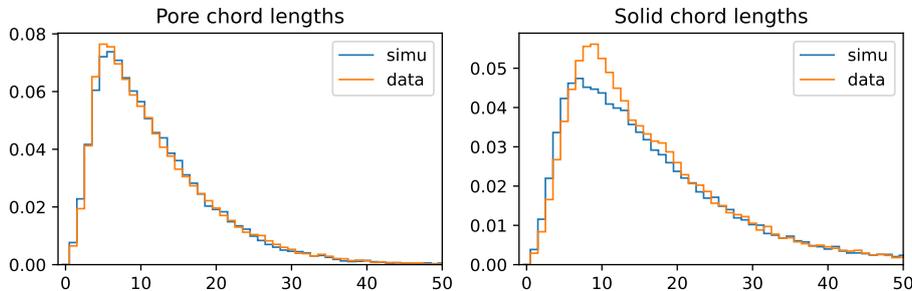


Figure 6: Left: Chord length distribution of pore space. Right: Chord length distribution of solid phase. In both graphs, image data of structured sand is considered, where orange corresponds to measured data and blue to a model realization.

Note that the underlying optimization procedure uses a cost function which is given by the sum of the quadratic distances between the probability densities of chord lengths of the solid phase and the pore space of measured and simulated structured sand, respectively. This cost function is chosen such that both pore and solid phase are taken into account equally. The cost function is numerically minimized via simulated annealing [25], leading to the model parameter  $\theta^* = (0.04, 44.6, 70.1, 82.1, 2.86)$ . For model fitting and the generation of artificial soil structures, the Gaussian random fields  $X_1$  and  $X_2$  introduced in Section 4.1 are simulated via the Fourier-based method described in [26].

### 4.3 Model validation

For a structural validation of the stochastic 3D model introduced in Section 4.1, we used the following approach. First, we have fitted the model in Section 4.2 to the complete tomographic image data (*i.e.*, not the small cutouts, which have been used for establishing structure-property relationships) of one of the two datasets of structured sand described in Section 2.1. Then, we simulated 3D structures of the same size using the fitted model described in Section 4.1. A visual comparison between tomographic image data and model realizations for structured sand is shown in Figure 7. Finally, we computed some of the geometric descriptors stated in Section 2.4, with the exception of mean chord lengths which were used for model calibration. As can be seen from Table 3, the simulated structures resemble the measured data quite well in terms of the considered geometric descriptors. This indicates the feasibility of further investigations of microstructure-property relationships by means of artificially generated soil structures.



Figure 7: 2D cross-sections of tomographic data of sand (left) and simulated data (right).

Descriptor	Measured	Simulated
Volume fraction $\varepsilon$	0.586	0.583
Specific surface area $S$	0.215	0.212
Mean geodesic tortuosity $\mu(\tau)$	1.090	1.098
Standard deviation of geodesic tortuosity $\sigma(\tau)$	0.0067	0.0096
Constrictivity $\beta$	0.514	0.452

Table 3: Overview of the geometrical descriptors computed for the pore space of measured and simulated data of structured sand.

## 5 Conclusion

285 In the present paper, 3D image data of structured soils, obtained by X-ray computed tomography, is used for establishing quantitative structure-property relationships as well as for calibrating a simple stochastic 3D model for structured sand, which is based on excursion sets of Gaussian random fields.

290 First, tomographic image data of the 3D morphology of structured soils, which consists of three loam and three sand samples, is investigated quantitatively by means of univariate and bivariate probability distributions of geometrical and functional descriptors. These distributions are compared to those of the 3D structure of finely sieved soils, which have been considered in [11]. Similar to the situation observed for finely sieved soils, significant differences between loam and sand are apparent within the structured soils. Moreover, differences between finely sieved and structured soils are found which can be partially attributed to a different resolution of the tomographic image data. Especially for the computed M-factors, 295 the difference between loam and sand is increased in structured soils compared to finely sieved soils.

In a second step, quantitative structure-property relationships are established via analytical regression formulas using a total of 2676 non-overlapping cutouts. For this purpose, the method of periodic homogenization is used to obtain diffusive transport properties. The regression formulas which were already established in [11] are used to predict the M-factor of pore space from geometric descriptors. It turns out 300 that the previously best-performing parametric regression formula outperforms other regression formulas also when fitted on the novel image data of structured soils. Although huge differences were observed between the values of geometrical descriptors and M-factors computed from image data of structured soils compared to those obtained for finely sieved soils, this formula reliably predicts the M-factor for data of structured soils, even when used with the parameters previously derived in [11] for finely sieved 305 soils. This suggests a universal applicability of this type of regression formula to various different types of structures without the need for adjustment of parameters.

To validate the applicability of the regression formulas considered in the present paper for further types of soils, stochastic 3D models may be used to generate digital twins of real soils, *i.e.*, artificial soils may be created in silico. In the present paper, a first step in this direction is taken by a stochastic 310 3D model which is calibrated to image data of structured sand, where the solid phase is modeled as the union of excursion sets of two independent Gaussian random fields. While being a relatively simple model, validation of the model with respect to various geometric descriptors of structured sand yields promising results. This encourages continued development of stochastic 3D models for more complex soil morphologies, like those of structured loam. In particular, structured loam does not exhibit two clear 315 length scales, which can be modeled via excursion sets of two independent Gaussian random fields.

## 6 Statements and Declarations

### 6.1 Acknowledgments

The authors acknowledge the computation of the geometrical descriptors for structured loam and sand samples by Eva Heincke.

## 320 6.2 Funding

This work was supported by the German Research Foundation (DFG) within the priority programme 2089 'Rhizosphere spatiotemporal organization - a key to rhizosphere functions' (grant numbers 426456278, 403640293, 403660839, 403801423).

## 6.3 Competing Interests

325 The authors have no relevant financial or non-financial interests to disclose.

## 6.4 Author Contributions

Material preparation and data collection was performed by Maxime Phalempin and Steffen Schlüter. Maximilian Rötzer, Nadja Ray and Alexander Prechtel elaborated the homogenized diffusion problems on the samples. Maximilian Rötzer conducted the necessary numerical simulations. Quantitative analysis and fitting of regression formulas as well as stochastic 3D microstructure modeling were performed by 330 Benedikt Prifling and Matthias Weber, who jointly prepared the first draft of the manuscript. All authors contributed to writing and editing of the manuscript as well as the conception and design of the study.

## 6.5 Data Availability

The data is available from the authors upon reasonable request.

## 335 References

- [1] S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. New-York: Springer, 2002.
- [2] F. Willot and S. Forest, eds., *Physics and Mechanics of Random Media: From Morphology to Material Properties*. Paris: Presses des MINES, 2018.
- 340 [3] M. Sahimi, *Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches*. Weinheim: Wiley-VCH, 2<sup>nd</sup> ed., 2011.
- [4] J. Bear, *Modeling Phenomena of Flow and Transport in Porous Media*. Cham: Springer, 2018.
- [5] J. Newman and K. Thomas-Alyea, *Electrochemical Systems*. Hoboken: J. Wiley & Sons, 3<sup>rd</sup> ed., 2004.
- 345 [6] M. K. Das, P. P. Mukherjee, and K. Muralidhar, *Modeling Transport Phenomena in Porous Media with Applications*. Cham: Springer, 2018.

- [7] M. Blunt, *Multiphase Flow in Permeable Media: A Pore-Scale Perspective*. Cambridge: Cambridge University Press, 2017.
- [8] E. Rabot, M. Wiesmeier, S. Schlüter, and H.-J. Vogel, “Soil structure as an indicator of soil functions: A review” *Geoderma*, vol. 314, 122–137, 2018.
- [9] K. A. Smith, T. Ball, F. Conen, K. E. Dobbie, J. Massheder, and A. Rey, “Exchange of greenhouse gases between soil and atmosphere: interactions of soil physical factors and biological processes” *European Journal of Soil Science*, vol. 54, no. 4, 779–791, 2003.
- [10] N. Ray, A. Rupp, R. Schulz, and P. Knabner, “Old and new approaches predicting the diffusion in porous media” *Transport in Porous Media*, vol. 124, 803–824, 2018.
- [11] B. Prifling, M. Weber, N. Ray, A. Prechtel, M. Phalempin, S. Schlüter, D. Vetterlein, and V. Schmidt, “Quantifying the impact of 3D pore space morphology on soil gas diffusion in loam and sand” *Transport in Porous Media*, vol. 149, 501–527, 2023.
- [12] B. Prifling, M. Ademmer, F. Single, O. Benevolenski, A. Hilger, M. Osenberg, I. Manke, and V. Schmidt, “Stochastic 3D microstructure modeling of anodes in lithium-ion batteries with a particular focus on local heterogeneity” *Computational Materials Science*, vol. 192, 110354, 2021.
- [13] L. S. Kremer, T. Danner, S. Hein, A. Hoffmann, B. Prifling, V. Schmidt, A. Latz, and M. Wohlfahrt-Mehrens, “Influence of the electrolyte salt concentration on the rate capability of ultra-thick NCM 622 electrodes” *Batteries & Supercaps*, vol. 3, 1172–1182, 2020.
- [14] B. Abdallah, F. Willot, and D. Jeulin, “Morphological modelling of three-phase microstructures of anode layers using SEM images” *Journal of Microscopy*, vol. 263, 51–63, 2016.
- [15] H. Moussaoui, J. Laurencin, Y. Gavet, G. Delette, M. Hubert, P. Cloetens, T. L. Bihan, and J. Debayle, “Stochastic geometrical modeling of solid oxide cells electrodes validated on 3D reconstructions” *Computational Materials Science*, vol. 143, 262–276, 2018.
- [16] M. Neumann, B. Abdallah, L. Holzer, F. Willot, and V. Schmidt, “Stochastic 3D modeling of three-phase microstructures for predicting transport properties: A case study” *Transport in Porous Media*, vol. 128, 179–200, 2019.
- [17] M. Neumann, M. Osenberg, A. Hilger, D. Franzen, T. Turek, I. Manke, and V. Schmidt, “On a pluri-Gaussian model for three-phase microstructures, with applications to 3D image data of gas-diffusion electrodes” *Computational Materials Science*, vol. 156, 325–331, 2019.
- [18] M. Phalempin, E. Lippold, D. Vetterlein, and S. Schlüter, “Soil texture and structure heterogeneity predominantly governs bulk density gradients around roots” *Vadose Zone Journal*, vol. 20, no. 5, e20147, 2021.
- [19] D. Vetterlein, E. Lippold, S. Schreiter, M. Phalempin, T. Fahrenkamp, F. Hochholdinger, C. Marcon, M. Tarkka, E. Oburger, M. Ahmed, M. Javaux, and S. Schlüter, “Experimental platforms for the investigation of spatiotemporal patterns in the rhizosphere — Laboratory and field scale” *Journal of Plant Nutrition and Soil Science*, vol. 184, no. 1, 35–50, 2021.

- [20] C. Jawahar, P. K. Biswas, and A. Ray, “Investigations on fuzzy thresholding based on fuzzy clustering” *Pattern Recognition*, vol. 30, no. 10, 1605–1613, 1997.
- 385 [21] U. Hornung, *Homogenization and Porous Media*. New York: Springer, 1997.
- [22] G. Gaiselmann, M. Neumann, V. Schmidt, O. Pecho, T. Hocker, and L. Holzer, “Quantitative relationships between microstructure and effective transport properties based on virtual materials testing” *AIChE Journal*, vol. 60, no. 6, 1983–1999, 2014.
- [23] C. Lantuéjoul, *Geostatistical Simulation: Models and Algorithms*. Berlin: Springer, 2002.
- 390 [24] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*. Chichester: J. Wiley & Sons, 3<sup>rd</sup> ed., 2013.
- [25] W. Press, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 2007.
- 395 [26] A. Lang and J. Potthoff, “Fast simulation of Gaussian random fields” *Monte Carlo Methods and Applications*, vol. 17, 195–214, 2011.

## 7 Appendix

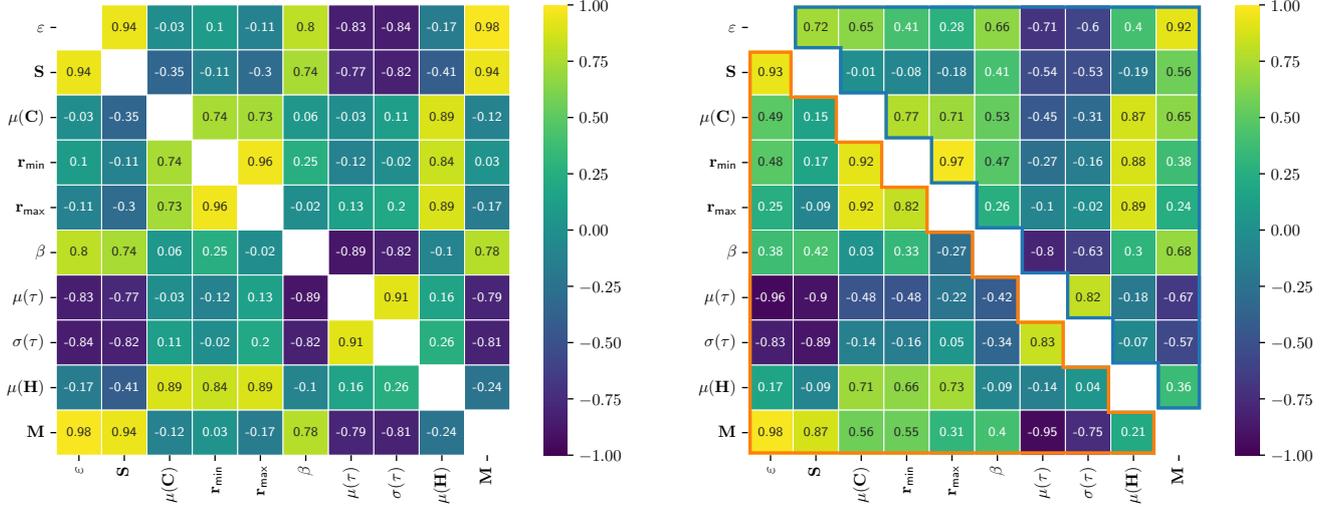


Figure 8: Correlation coefficients between all pairs of geometrical descriptors (as well as the M-factor) considered in the present paper, computed for the set of all  $2 \cdot 3 \cdot 446 = 2676$  subsamples (left) and for loam and sand individually (right). In the latter case, the upper right part of the matrix (blue frame) shows the values obtained for loam, the bottom left part (orange frame) those for sand.

	Formula	$R^2$	MAPE[%]
$\widehat{M}_1$	$\epsilon^{1.87}$	0.94	99.22
$\widehat{M}_2$	$1.95\epsilon^{2.53}$	0.98	45.14
$\widehat{M}_3$	$0.01e^{7.41\epsilon}$	0.94	101.71
$\widehat{M}_4$	$\epsilon^{1.26}\beta^{0.11}\mu(\tau_g)^{-8.27}$	0.99	15.41
$\widehat{M}_5$	$\epsilon^{1.22-0.02\beta}\mu(\tau_g)^{-9.23}$	0.99	16.21
$\widehat{M}_6$	$1.07\mu(\tau)^{-8.15}\sigma(\tau)^{-0.04}\epsilon^{1.22}$	0.99	16.72
$\widehat{M}_7$	$\epsilon^{3.32}S^{-1.06}r_{\max}^{-0.88}$	0.97	51.63
$\widehat{M}_8$	$\epsilon^{2.39}(\mu(C)/\mu(H))^{-0.31}$	0.97	57.52

Table 4: Parameter values of regression functions fitted on the entire training data; validation scores  $R^2$  and MAPE obtained for the validation data considered in the present paper.

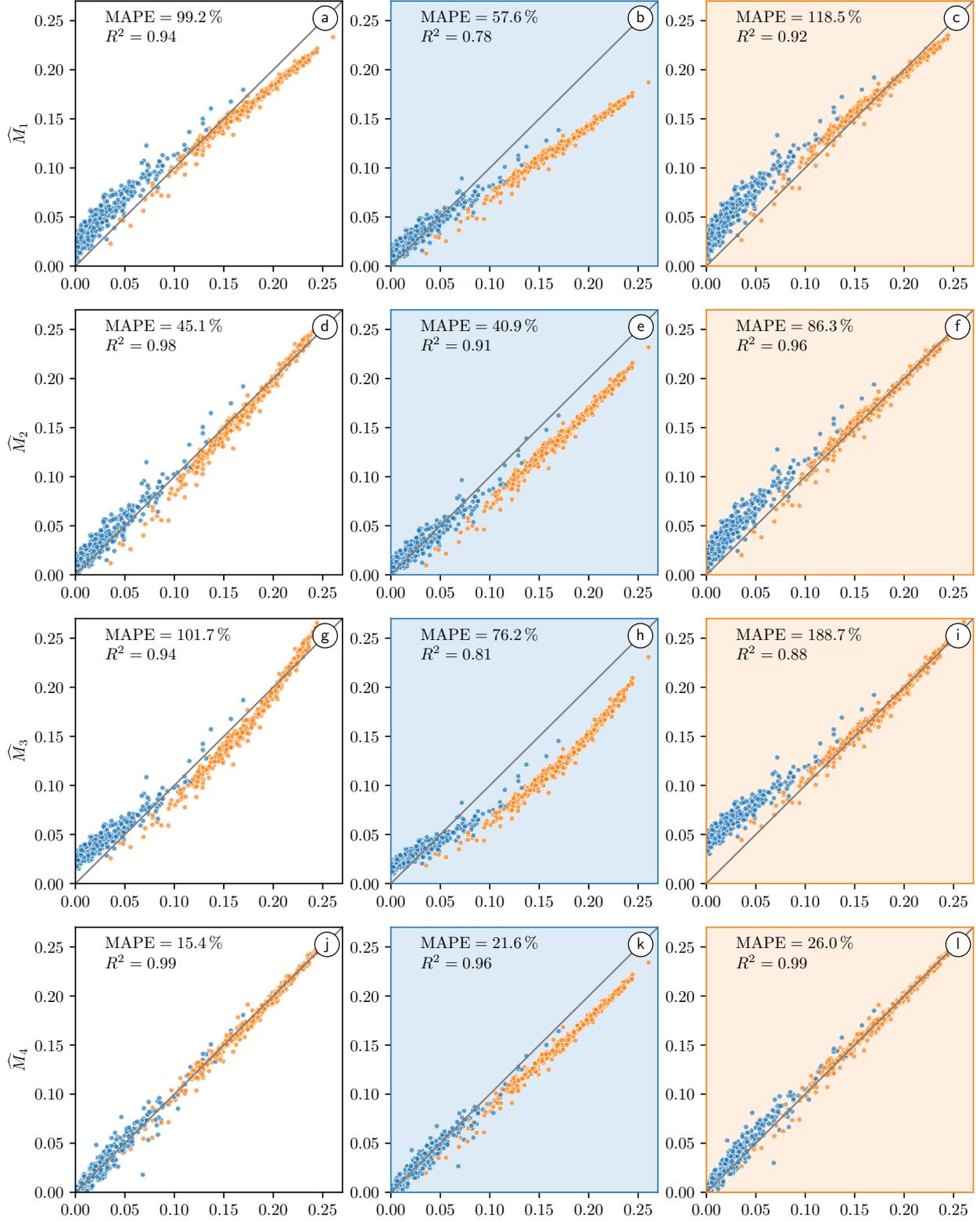


Figure 9: M-factor obtained by numerical simulations versus the predicted M-factor  $\widehat{M}_i, i = 1, \dots, 4$ , obtained for the validation data, where different regression formulas (given by Eqs. ((1)), ((2)), ((3)) and ((4)); from top to bottom) have been used. The colors of data points represent loam (blue) and sand (orange), respectively. The parameter values of regression formulas have been fitted on the entire training data (left), loam (center) and sand (right), respectively.

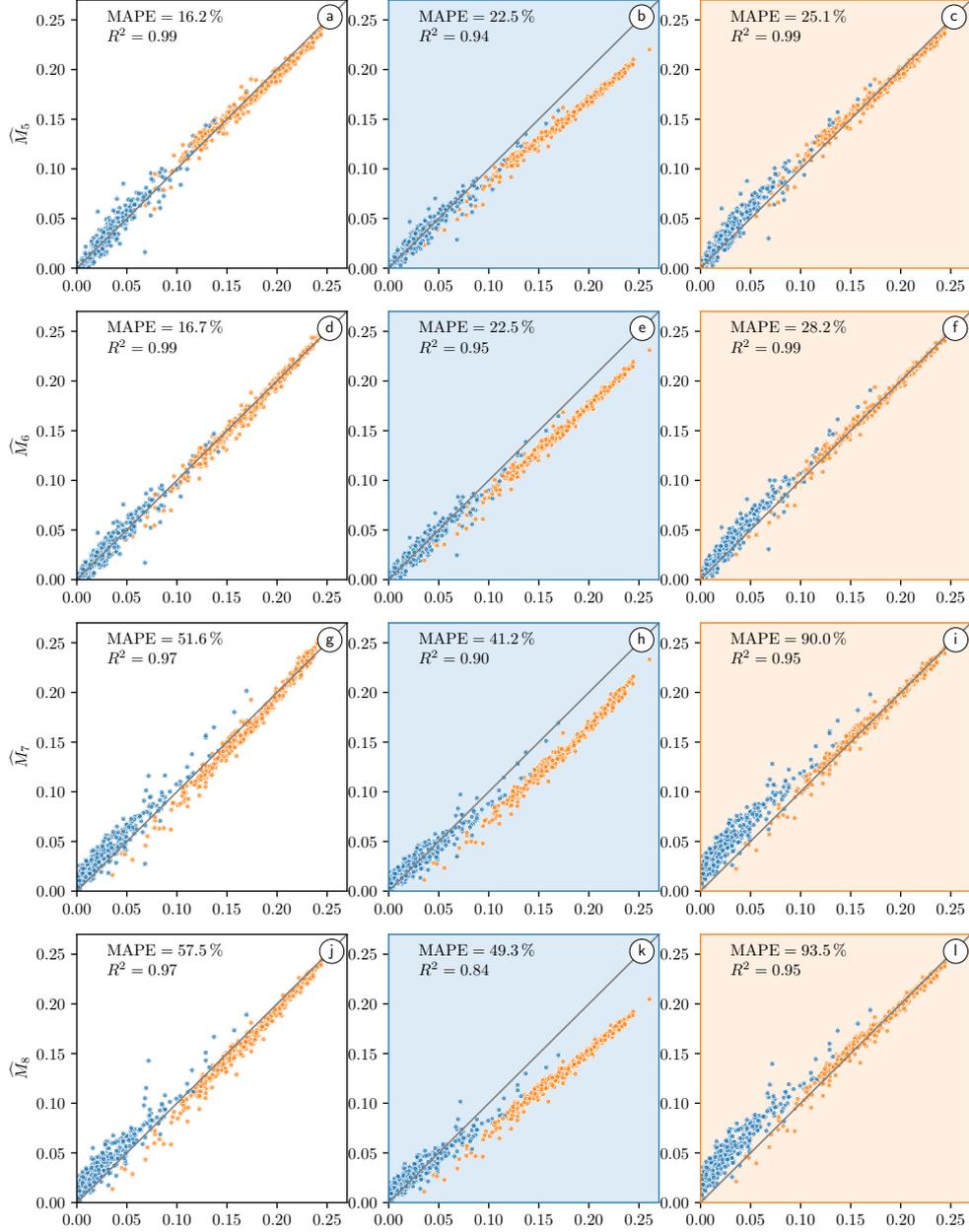


Figure 10: M-factor obtained by numerical simulations versus the predicted M-factor  $\widehat{M}_i, i = 5, \dots, 8$ , obtained for the validation data, where different regression formulas (given by Eqs. ((5)), ((6)), ((7)), ((8)); from top to bottom) have been used. In particular, Eq. ((6)) is used twice, once for geodesic tortuosity and once for geometric tortuosity. The colors of data points represent loam (blue) and sand (orange), respectively. The parameter values of regression formulas have been fitted on the entire training data (left), loam (center) and sand (right), respectively.



Figure 11: Validation scores  $R^2$  and MAPE obtained for the regression formulas considered in the present paper including the values for the corresponding regression parameters, where different data sets have been used for training/fitting and validation.

		$\widehat{M}_1 = \varepsilon^{c_1}$						$\widehat{M}_2 = c_1 \varepsilon^{c_2}$												
Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$c_1$	1.71	Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2)$	(1.92, 2.29)	
		0.86	-0.31	0.84	148.9	288.0	10.6					0.82	0.28	0.27	94.4	169.5	19.8			(1.92, 2.29)
		0.89	$-6 \cdot 10^{-2}$	0.93	134.3	262.5	6.8					0.85	0.36	0.39	88.7	160.0	17.9			(1.94, 2.32)
	Sand	0.82	-0.55	0.72	161.9	310.6	14.1	0.82	$7 \cdot 10^{-2}$		0.37	113.1	207.5	19.3	(1.60, 2.11)					
		All	Loam	Sand	All	Loam	Sand		All	Loam	Sand	All	Loam	Sand						
		Validation on structured soil						Validation on structured soil												
		$\widehat{M}_3 = c_1 e^{c_2 \varepsilon}$						$\widehat{M}_4 = \varepsilon^{c_1} \beta^{c_2} \mu(\tau_g)^{c_3}$												
Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2)$	(0.02, 6.88)	Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2, c_3)$	(1.06, -0.04, -7.28)	
		0.70	-0.51	$-7 \cdot 10^{-2}$	175.5	330.1	21.8					0.98	0.82	0.97	36.6	69.2	4.2			(1.06, -0.04, -7.28)
		0.75	-0.36	0.18	167.3	316.9	18.7					0.99	0.91	0.96	23.8	43.6	4.1			(0.91, 0.08, -8.55)
	Sand	0.69	-0.82	$5 \cdot 10^{-2}$	200.6	380.2	22.0	0.95	0.61		0.91	60.3	113.1	7.9	(1.21, 0.01, -5.13)					
		All	Loam	Sand	All	Loam	Sand		All	Loam	Sand	All	Loam	Sand						
		Validation on structured soil						Validation on structured soil												
		$\widehat{M}_5 = \varepsilon^{c_1+c_2} \beta \mu(\tau_g)^{c_3}$						$\widehat{M}_6 = c_1 \mu(\tau)^{c_2} \sigma(\tau)^{c_3} \varepsilon^{c_4}$												
Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2, c_3)$	(1.17, -0.18, -7.00)	Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2, c_3, c_4)$	(2.19, -6.07, 0.07, 1.63)	
		0.98	0.85	0.97	31.8	59.8	4.0					0.96	0.80	0.87	33.0	57.3	8.8			(2.19, -6.07, 0.07, 1.63)
		0.99	0.92	0.97	20.7	37.9	3.6					0.97	0.89	0.86	23.1	37.8	8.4			(1.94, -7.24, 0.03, 1.55)
	Sand	0.95	0.65	0.90	55.9	103.9	8.3	0.94	0.67		0.83	52.1	94.1	10.4	(1.64, -4.70, 0.04, 1.57)					
		All	Loam	Sand	All	Loam	Sand		All	Loam	Sand	All	Loam	Sand						
		Validation on structured soil						Validation on structured soil												
		$\widehat{M}_7 = \varepsilon^{c_1} S^2 r_{\max}^{c_3}$						$\widehat{M}_8 = \varepsilon^{c_1} (\mu(C)/\mu(H))^{c_2}$												
Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2, c_3)$	(2.60, -0.61, -0.45)	Fitting on finely sieved soil	All	$R^2$			MAPE [%]			$(c_1, c_2)$	(2.07, 0.29)	
		0.57	-0.73	-0.83	142.9	253.7	32.6					0.82	$9 \cdot 10^{-2}$	0.33	113.2	207.1	19.8			(2.07, 0.29)
		0.18	-1.39	-3.01	151.7	256.9	47.2					0.65	0.00	-0.74	109.8	189.6	30.5			(2.23, 0.45)
	Sand	0.56	-0.77	-0.84	148.6	264.9	33.0	0.83	0.20		0.33	104.8	190.8	19.3	(2.13, 0.32)					
		All	Loam	Sand	All	Loam	Sand		All	Loam	Sand	All	Loam	Sand						
		Validation on structured soil						Validation on structured soil												

Figure 12: Validation scores  $R^2$  and MAPE obtained for the regression formulas considered in the present paper including the values for the corresponding regression parameters. The data set of finely sieved soil described in [11] is used for training/fitting, whereas validation is carried out on image data for structured soil.