

Quantifying the impact of 3D pore space morphology on soil gas diffusion in loam and sand

Benedikt Prifling^{1,*}, Matthias Weber^{1,*}, Nadja Ray², Alexander Prechtel³,
Maxime Phalempin⁴, Steffen Schlüter⁴, Doris Vetterlein⁴, Volker Schmidt¹

May 17, 2023

5 Effective diffusion is an important macroscopic property for assessing transport in porous
media. Numerical computations on segmented 3D CT images yield precise estimates for
diffusive properties. On the other hand, geometrical descriptors of pore space such as
porosity, specific surface area and further transport-related descriptors can be easily com-
puted from 3D CT images and are closely linked to diffusion processes. However, the
10 investigation of quantitative relationships between these descriptors and diffusive prop-
erties for a diverse range of porous structures is still ongoing. In the present paper, we
consider three different soil samples of each loam and sand for a total of six samples,
whose 3D microstructure is quantitatively investigated using univariate as well as bivari-
ate probability distributions of geometrical pore space descriptors. This information is
15 used for investigating microstructure-property relationships by means of empirically de-
rived regression formulas, where a particular focus is put on the differences between loam
and sand samples. Due to the analytical nature of these formulas, it is possible to obtain a
deeper understanding for the relationship between the 3D pore space morphology and the
resulting diffusive properties. In particular, it is shown that formulas existing so far in the
20 literature for predicting soil gas diffusion can be significantly improved by incorporating
further geometrical descriptors such as geodesic tortuosity, chord lengths, or constrictivity
of the pore space. The robustness of these formulas is investigated by fitting the regression
parameters on different data sets as well as by applying the empirically derived regression
formulas to data that is not used for model fitting. Among others, it turns out that a
25 formula based on porosity as well as mean and standard deviation of geodesic tortuosity
performs best with regard to the coefficient of determination and the mean absolute per-
centage error. Moreover, it is shown that regarding the prediction of diffusive properties
the concept of geodesic tortuosity is superior to geometric tortuosity, where the latter is
based on the creation of a skeleton of the pore space.

¹Institute of Stochastics, Ulm University, Ulm, Germany

²Mathematical Institute for Machine Learning and Data Science, Catholic University of Eichstätt-Ingolstadt, Ingolstadt, Germany

³Department of Mathematics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany

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

*Both authors contributed equally

30 1. Introduction

The three-dimensional morphology of the pore space has a strong impact on effective macroscopic properties of porous media [1, 2, 3]. Among others, diffusive properties are of great interest in a wide spectrum of applications ranging from lithium-ion batteries via solid-oxide fuel cells to civil engineering [2, 4, 5, 6, 3]. Likewise, the 3D microstructure of the pore space greatly affects water, gas and carbon
35 fluxes in soils [7]. A typical example of a soil structure mediated gas transport is the emission of greenhouse gases into the atmosphere [8]. Gas transport in soils is known to be primarily a diffusion dominated process. More precisely, diffusion of gas in soils is mainly governed by the so-called soil gas diffusivity, for which empirical relationships, *i.e.* regression formulas (also called analytical prediction formulas), have been developed using easily available soil bulk properties like the air filled porosity
40 [9, 10, 11] and then further improved by the integration of other soil properties such as soil bulk density [12, 13, 14] or specific surface area [15]. Moreover, it turned out that regression formulas of this type can be improved by incorporating further geometrical descriptors of the pore space [16]. This motivates a more detailed characterization of the pore space.

Due to recent efforts and technological advancements in the field of soil imaging, notably in X-ray
45 computed tomography (CT), it is possible to accurately visualize and characterize the soil pore space in three dimensions at high resolution. As a result, X-ray CT also offers the unprecedented opportunity to test the validity and relevance of incorporating physical descriptors of the pore space in empirical relationships aiming at predicting soil gas diffusivity. Previous studies relying on X-ray CT to derive gas transport parameters have shown promising results so far. In a study of no-tillage system in
50 a subtropical climate, significant correlations between the porosity derived with X-ray CT and the logarithm of relative gas diffusivity and air permeability were found [17]. In another study [18], the porosity obtained by means of X-ray CT was also shown to be well correlated with air permeability. Furthermore, in [19], it has been shown that there is a strong correlation of the lumped tortuosity-connectivity parameter obtained from standard gas measurements (as described in [20]) and the one
55 obtained from X-ray CT. Thus, based on the literature, it seems that X-ray CT can be considered as a powerful tool to derive gas transport parameters. It is however still unclear which combination of geometrical descriptors can be used to best predict effective macroscopic properties as effective diffusivity.

Likewise, the progress in computational power and sophisticated numerical methods enables the com-
60 putation of orientation-dependent diffusive properties from sound mathematical theory [21]. This is possible even on an extensive dataset of real and therefore quite complex geometries, which forms the basis for the quantitative investigation of microstructure-property relationships within this research. In particular, we consider the relationship between descriptors of 3D microstructure and diffusive properties.

In the present paper, we consider different kinds of regression formulas to predict diffusive properties
65 from geometric descriptors of 3D pore space. Applying this approach to data from two different soil textures, we quantitatively investigate the performance and robustness of the considered regression formulas. We thoroughly investigate geometrical descriptors of the 3D microstructure of the samples' pore space and analyze them in detail in the context of sand and loam. For this purpose, the loam and
70 sand samples under consideration as well as the corresponding tomographic imaging techniques are explained in Sections 2.1 and 2.2, respectively. Moreover, we provide details regarding the numerical computation of diffusive properties (Section 2.3) as well as various descriptors of the 3D pore space morphology (Section 2.4), which are used to characterize its 3D microstructure geometrically and with regard to its diffusive properties. Finally, in Section 2.5, our methods of model fitting and validation

75 are stated, which are used for assessing the predictive power of microstructure-property relationships. Then, in Section 3.1, a statistical analysis of 4608 subsamples is carried out to quantitatively investigate differences between loam and sand by means of univariate as well as bivariate probability distributions. Moreover, microstructure-property relationships for soil gas diffusion are established in Section 3.2, where the goal is to derive accurate (and possibly general) prediction formulas, *i.e.*, they do not
80 require individual sets of fitting parameters for the two soil textures. In contrast to other methods of data science as, for example, neural networks, the empirically derived prediction formulas establish an interpretable relationship between the 3D morphology of the pore space and its diffusive properties. In particular, in Section 4, we compare the predictive power of conventional models discussed in the literature with that of more sophisticated models which additionally include further geometrical
85 descriptors of the 3D pore space morphology. It turns out that the more accurate microstructure description achieved in this way allows us to surpass traditional prediction formulas with regard to the coefficient of determination as well as the mean absolute percentage error. Moreover, a quantitative comparison between two different notions of tortuosity, namely geodesic and geometric tortuosity, is carried out to investigate their suitability for establishing microstructure-property relationships. Section 5 concludes the paper and outlines possibilities for future research.
90

2. Materials and methods

2.1. Sample acquisition

The samples analyzed in this study were acquired within the framework of the plant growth experiments described in [22]. In particular, cylindrical columns of 23 cm in height and 7 cm in diameter
95 were packed with soil, where two different substrates (loam and sand) were used. The loam substrate was obtained from the upper 50 cm of a haplic Phaeozem soil profile, dried to 10 % gravimetric water content and then sieved to < 1 mm. The sand substrate constitutes a mix of 83.3 % quartz sand (WF 33, Quarzwerke Weferlingen, Germany) and 16.7 % of the sieved loam. Details on chemical and physical properties of these substrates are provided in [23]. The loam pots were packed homogeneously to
100 a bulk density of 1.26 g/cm³, whereas the sand pots were packed to a bulk density of 1.47 g/cm³. At the end of the growth period, a total of six samples (aluminum rings of 1.6 cm in height and diameter) were taken from 5, 10 and 15 cm depth (3 per substrate, one per depth) and were stored at 4 °C in sealed plastic bags prior to X-ray CT scanning.

2.2. X-ray computed tomography scanning and binarization

105 X-ray CT scanning was performed with an industrial μ CT scanner (X-TEK XTH 225, Nikon Metrology) including an Elmer-Perkin 1620 detector panel (1750 × 2000 px). The scanner was operated at 115 kV and a current of 85 μ A. A total of 2748 projections with an exposure time of 708 ms each were acquired during a full rotation of the samples. The obtained images were reconstructed into a three dimensional tomogram by means of the filtered back projection algorithm with the CT Pro 3D software (Nikon Metrology) [24]. During reconstruction, the grayscale range was normalized with a percentile
110 stretching method, which sets the darkest and brightest 0.2 % voxels to 0 and 255, respectively, and performs a linear stretching in between. The voxel size equals 10 μ m.

Denosing and contrast enhancement of the grayscale images were performed using a 2D non local

means filter ($\sigma = 15$) and an unsharp mask filter (radius=1 and weight=0.6) [25, 26], respectively. Subsequent binarization of the images was performed to distinguish between pore space and the soil matrix. The automatic threshold detection method presented in [27] was used for an adequate and objective choice of the global threshold.

For each texture and depth, the samples were truncated into stacks of 1536 slices with each 1024×1024 voxels. This coincides with cuboid of 15.36 mm height, 10.24 mm width and 10.24 mm length. For the purpose of the study, we further divided the cuboid into non-overlapping cubic cutouts with $128 \times 128 \times 128$ voxels (1.28 mm \times 1.28 mm \times 1.28 mm) each. Thus, we obtained 768 cubic cutouts for each depth and soil texture, which led to a total number of 4608 subsamples. Selected subsamples are shown in Figure 1.

Finally, isolated pores were removed, which led to only marginal changes with regard to the porosity, *i.e.*, the 3D binary images contained only few unconnected pores. The overall difference in porosity for loam was smaller than 2 % of the initial porosity, for sand even less than 0.5 %.

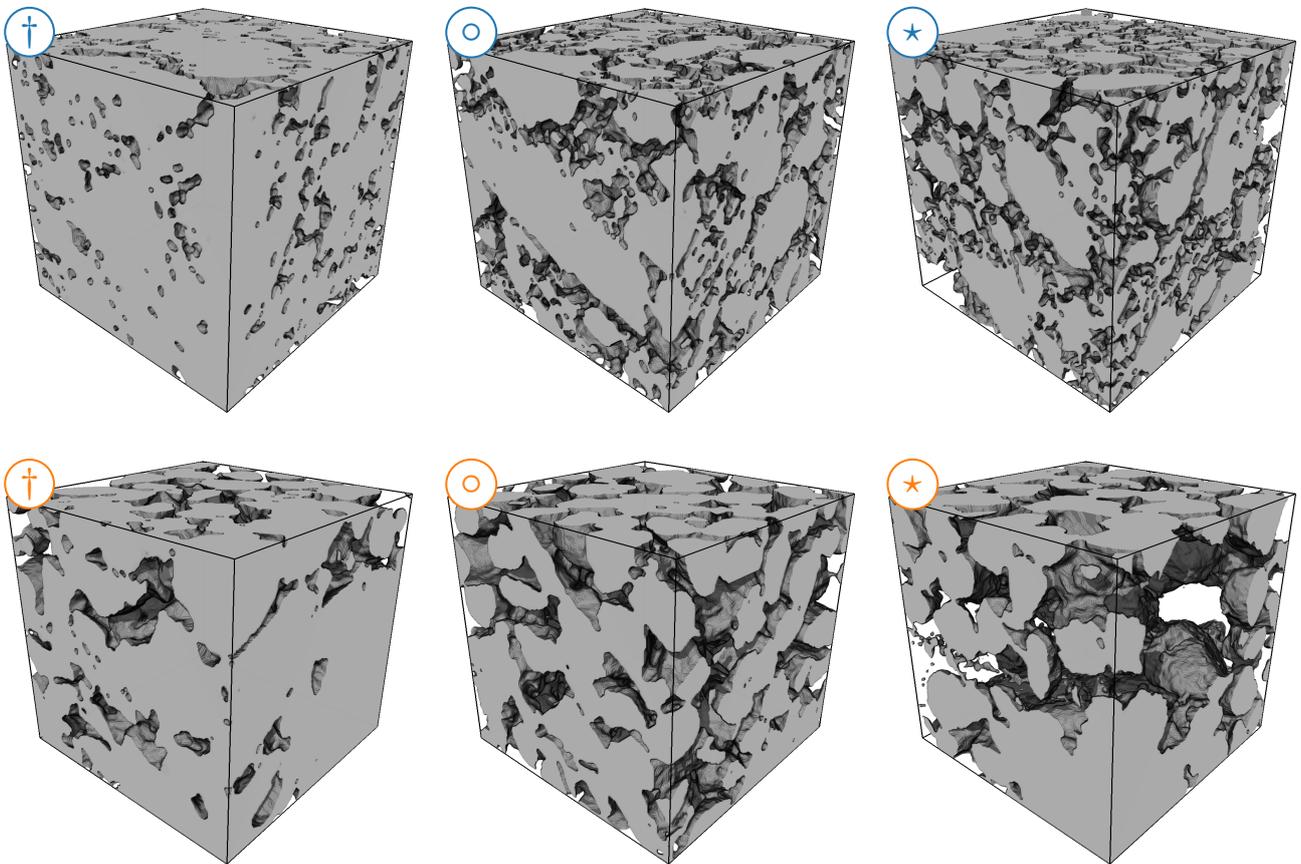


Figure 1: Three-dimensional renderings of selected loam (top row) and sand (bottom row) subsamples, where the solid phase is depicted in gray. The three columns correspond to minimum (left, †), mean (center, ◦) and maximum (right, ★) porosity.

2.3. Computation of diffusive properties

Diffusive transport of a chemical species with concentration field $c : \mathbb{R}^3 \rightarrow [0, \infty)$ through porous media is, in its simplest form, described by the Laplace equation

$$-\nabla \cdot (\mathbb{D} \nabla c) = 0, \quad (1)$$

130 where $\mathbb{D} \in \mathbb{R}^{3 \times 3}$ denotes the effective diffusion tensor, which is the essential input to the transport model. It contains all the information that is specific to the porous medium under consideration. However, it is quite difficult to characterize \mathbb{D} for (natural) porous media. If the underlying geometry is available, e.g. from binary CT data, one possibility to determine the tensor \mathbb{D} is to apply the results of homogenization theory [28, 21], which was originally derived for periodic settings.

135 In order to carry out the numerical simulations for the effective diffusivities by homogenization within a reasonable time frame, the subsamples have been coarsened by combining eight voxels into one (super-) voxel. If four of the original voxels were solid and four void, the resulting (super-) voxel was at random turned into a void or solid voxel, otherwise the majority defines the type of the resulting voxel. Hence, after performing this procedure, the dimensions of the subsamples are $64 \times 64 \times 64$ voxels (1.28 mm \times 1.28 mm \times 1.28 mm), where a voxel corresponds to $20^3 \mu\text{m}^3 = 8000 \mu\text{m}^3$.

For a connected, sufficiently smooth pore space $Y_p \subset Y = [0, 1.28 \text{ mm}]^3$ and complementing solid phase $Y_s = Y \setminus Y_p$, the entries of the diffusion tensor \mathbb{D} are given as

$$\mathbb{D}_{ij} = \frac{1}{V(Y)} \int_{Y_p} D_m (\partial_{y_i} \zeta_j(y) + \delta_{ij}) dy, \quad \text{for } i, j \in \{1, 2, 3\}, \quad (2)$$

with molecular diffusivity $D_m > 0$ and the following supplementary cell problems in ζ_j , $j \in \{1, 2, 3\}$:

$$\begin{cases} -\nabla_y \cdot (\nabla_y \zeta_j + e_j) = 0 & \text{in } Y_p, \\ (\nabla_y \zeta_j + e_j) \cdot \nu = 0 & \text{on } \partial Y_s, \\ \zeta_j \text{ is periodic in } y \text{ and has vanishing mean} \end{cases} \quad (3)$$

Here, $V(Y)$ denotes the volume of V , ν is the unit outer normal of the solid phase, e_j the unit vector in direction j , and δ_{ij} the Kronecker delta. Note that in the present setting, we assume zero diffusivity within the solid phase.

145 The partial differential equations given in (3) are solved within the connected pore space and discretized using the local discontinuous Galerkin (LDG) method [29, 30], which is implemented within the parallel finite element software, M++ [31]. The resulting linear systems are solved using a combination of the BiCGstab solver and the ILUT preconditioner [32] through Trilinos [33].

As a result of the computations described above, the 3×3 diffusion tensor \mathbb{D} has been determined for 150 all 4608 subsamples obtained from the X-ray CT scans of sand and loam, as described in Section 2.2. Further details regarding the analysis of these diffusion tensors are presented in [34]. In the following, the soil gas diffusion for the sand and loam subsamples will be described by the ratio of effective and intrinsic diffusivity which is sometimes called microstructure factor or M -factor [35]. In particular, we average over the three Cartesian axes directions such that the averaged M -factor is obtained from the diagonal entries of the diffusion tensor as $M = \frac{1}{3D_m} (\mathbb{D}_{11} + \mathbb{D}_{22} + \mathbb{D}_{33})$. This averaging is motivated by 155 the fact that neither the loam nor the sand samples show pronounced anisotropy effects [34].

2.4. Geometrical descriptors of pore space

In this section, various geometrical descriptors of the 3D pore space morphology and methods for their estimation from voxelized image data are presented. In Figure 2, two-dimensional versions of these descriptors are illustrated.

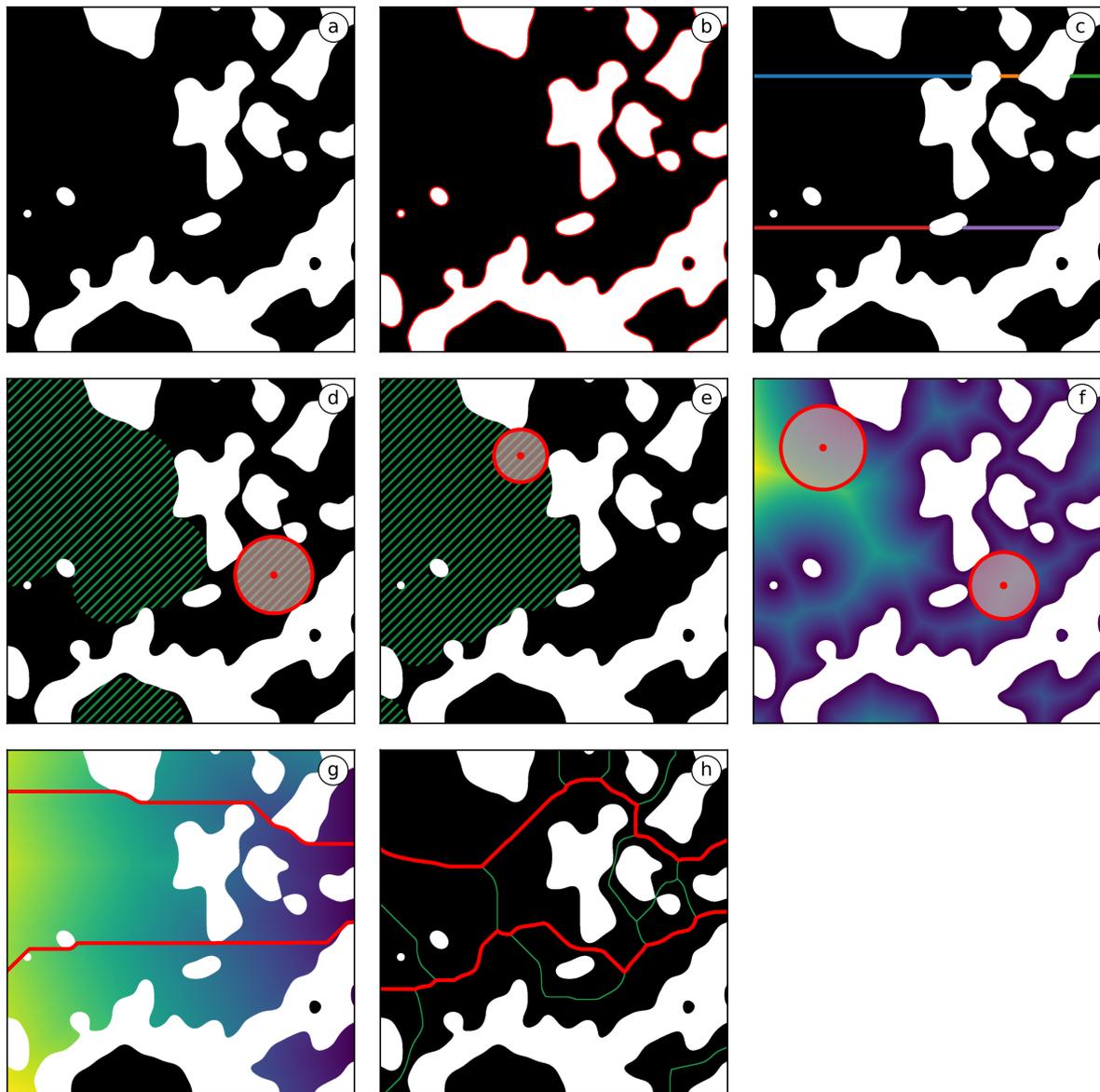


Figure 2: Schematic 2D illustration of microstructure descriptors: Exemplarily chosen 2D slice, where the pore space is depicted in black (a); surface between pore and solid phases, appearing as 1-dimensional object in the 2D slice (b); chords in horizontal direction for two exemplarily chosen vertical height levels (c); subsets of pore space used in the definitions of CPSD (d) and SMIP (e), respectively; shortest distances from pore space to soil (f); two exemplarily chosen shortest paths (from left to right) (g); skeleton of pore space and shortest paths along the skeleton (from left to right) depicted in green and red color, respectively (h).

Porosity ε One of the most fundamental descriptors of the 3D pore space morphology is the porosity $\varepsilon \in [0, 1]$. The estimation of this geometrical descriptor from binarized 3D image data is straightforward and carried out by computing the number of voxels assigned to the pore space (depicted in black in Figure 2a), which is divided by the total number of voxels of the entire sampling window.

165 **Specific surface area S** In addition to porosity, we consider the specific surface area of the pore space, denoted by S . It is defined as the surface area between the pore space and the solid phase divided by the volume of the sampling window. This quantity is then estimated from voxelized 3D image data using an approach presented in [36], which is based on a convolution of the image with a $2 \times 2 \times 2$ kernel avoiding the reconstruction of the actual surface. In Figure 2b, the interface between
170 pore and solid phases is depicted in red.

Mean chord length $\mu(C)$ A further descriptor of the 3D pore space morphology is its chord length distribution [37, 38], where a chord is a line segment that is completely contained in a predefined phase and can not be extended further without intersecting the complementary phase. Obviously, in general, the probability distribution of chord lengths depends on the orientation of the line segments.
175 We compute the chord length distribution of the pore space for the three Cartesian axes directions. In particular, for each of these three directions, we compute the mean value of the corresponding chord length distribution. In the following, the average of these three mean values, denoted by $\mu(C)$, is used for characterizing the 3D pore space morphology. In Figure 2c, five different chords are shown in horizontal direction and depicted in five different colors, for two exemplarily chosen vertical height
180 levels.

Constrictivity β In order to explain this descriptor, we first recall the concept of the continuous pore size distribution (CPSD) as well as of simulated mercury intrusion porosimetry (SMIP). The continuous pore size distribution is a function $\text{CPSD} : [0, \infty) \rightarrow [0, 1]$, where the value $\text{CPSD}(r)$ is given by the volume fraction of the pore space which can be covered by (possibly overlapping) spheres with radius
185 $r \geq 0$ (such that the spheres are completely contained in the pore space) [38, 39]. Furthermore, by r_{\max} the maximum radius $r > 0$ is denoted such that $\text{CPSD}(r) \geq \varepsilon/2$ where ε is the porosity. Note that the green shaded area in Figure 2d corresponds to that part of the pore space which can be covered by potentially overlapping spheres having the same radius as the red circle in Figure 2d.

The concept of SMIP is similar to that of CPSD, with the only difference that $\text{SMIP}(r)$ is now given by
190 the volume fraction of the pore space which can be covered by (potentially overlapping) spheres with radius r forming an intrusion from a predefined direction, see Figure 2e for an intrusion of spheres from left to right. Analogously to r_{\max} , by r_{\min} the maximum radius $r > 0$ is denoted such that $\text{SMIP}(r) \geq \varepsilon/2$. In general, r_{\min} depends on the direction of the intrusion. However, as already mentioned above, the lack of pronounced anisotropy effects in the loam and sand samples motivates
195 the following averaging. First, r_{\min} is computed for each of the three axes directions separately and, subsequently, the average of the three obtained values is used.

The constrictivity of the pore space is then defined as $\beta = \left(\frac{r_{\min}}{r_{\max}}\right)^2$. It is a measure for the strength of bottleneck effects and has been originally introduced in [40]. Since $r_{\min} \leq r_{\max}$ by definition, it holds that $\beta \in [0, 1]$, where $\beta \approx 1$ corresponds to a situation such that (almost) no constrictions within
200 the pore space exist. In [16, 41, 42], it has been shown that this quantity has a significant impact on effective macroscopic properties of porous media such as effective diffusivity or permeability.

Mean spherical contact distance $\mu(H)$ Here we consider the distribution of the shortest (contact) distances from randomly chosen points within the pore space to the solid phase. It is estimated from voxelized image data by an algorithm presented in [43], which is based on the computation of the Euclidean distance transform [39, 44]. The mean value of this distribution will be denoted by $\mu(H)$ in the following. In Figure 2f, the shortest distances from points in the pore space to the solid phase are visualized by (brighter and darker) colors, where the increase of brightness corresponds to an increase of contact distance. In particular, the radii of the red spheres in Figure 2f correspond to the shortest distances from their midpoints to the soil.

Mean value $\mu(\tau_g)$ and standard deviation $\sigma(\tau_g)$ of geodesic tortuosity A further transport-relevant quantity is the so-called geodesic tortuosity of the pore space. It describes the lengths and windedness of transport paths, which are completely contained in the pore space. From 3D image data, the distribution of (local) geodesic tortuosity is determined by computing the lengths of shortest paths from randomly selected pore voxels on a predefined starting plane to a parallel target plane, divided by the distance between those two planes, where shortest paths are computed using Dijkstra’s algorithm [45]. In Figure 2g, two exemplarily chosen shortest paths (from left to right) are shown in red, where the color of each pore space pixel corresponds to the extend of shortest distance to the target plane and the increase of brightness corresponds to an increase of path length. Usually, the starting and target planes are chosen orthogonal to the relevant transport direction. For the image data considered in the present paper, we compute the distribution of geodesic tortuosity with respect to each of the three Cartesian axes directions. The mean geodesic tortuosity $\mu(\tau_g)$ is then determined by averaging over all shortest path lengths divided by the distance between the starting and target planes. Analogously, the standard deviation $\mu(\sigma_g)$ is the empirical standard deviation of these normalized path lengths.

Note that in many applications, only $\mu(\tau_g)$ is considered in order to characterize the microstructure of porous media, despite the fact that $\sigma(\tau_g)$ contains further useful information about the 3D morphology of the pore space, see Section 3.2 for a detailed discussion. Furthermore, besides geodesic tortuosity, there exist several other concepts of tortuosity in the literature, see e.g. [46, 47] for a comprehensive overview. An example of such alternative tortuosity concepts is the so-called geometric tortuosity, which is described in the next paragraph.

Mean value $\mu(\tau_s)$ and standard deviation $\sigma(\tau_s)$ of geometric tortuosity In contrast to geodesic tortuosity described above, the concept of geometric tortuosity (sometimes called skeletonized tortuosity) is based on the computation of a skeleton of the pore space, see Figure 2h for an illustration where the skeleton is depicted in green color. In the present study, we compute the skeleton of the pore space by means of the Lee thinning [48, 49]. Similar to the concept of geodesic tortuosity, shortest paths along the transport direction are computed by Dijkstra’s algorithm [45], where in case of geometric tortuosity only voxels that belong to the skeleton of the pore space are allowed to be visited. Two such shortest paths (from left to right) along the skeleton of the pore space are shown in Figure 2h in red. Finally, we consider the distribution of the lengths of these shortest paths divided by the Euclidean distances between the starting and the target plane. Mean and standard deviation of this distribution are denoted by $\mu(\tau_s)$ and $\sigma(\tau_s)$, respectively. As in case of all direction-dependent microstructure characteristics considered in the present paper, these quantities are computed for each of the three Cartesian axis directions separately and then averaged.

Note that there exist various approaches for computing the skeleton of a 3D binary image, which in turn affects the computation of geometric tortuosity [39, 49].

245 **2.5. Model fitting and validation**

We now present statistical methods which will be used in Section 3 for a comprehensive structural analysis of the loam and sand samples as well as for establishing microstructure-property relationships, which quantitatively characterize the connection between geometrical descriptors of 3D pore space morphology and soil gas diffusion. In particular, in Section 3.2, regression formulas for the prediction of the M -factor are considered, which are based on different sets of descriptors of the 3D pore space morphology. In general, this approach leads to results that are easier to interpret compared to other methods from data science such as neural networks.

The regression formulas considered in Section 3.2 contain a small number of regression parameters, which are fitted by means of training data, *i.e.*, a fraction of the 4608 subsamples described in Section 2.2 is selected for determining “optimal” values for the regression parameters. In order to compute these parameters, the Levenberg-Marquardt algorithm [50] implemented in the python package `scipy` [51] is used. This algorithm solves the unconstrained nonlinear least-squares problem of finding an optimal set of parameter values for a given regression formula and training data.

To evaluate the performance of regression formulas, we select $k = 1024$ subsamples as test data, which have not been used for model fitting. Then, the goodness of fit is quantified by computing the coefficient of determination, denoted by R^2 , as well as the mean absolute percentage error, denoted by MAPE. These two quantities are defined as

$$R^2 = 1 - \frac{\sum_{j=1}^k (M_j - \widehat{M}_j)^2}{\sum_{j=1}^k (M_j - \bar{M})^2} \quad \text{and} \quad \text{MAPE} = \frac{100}{k} \sum_{j=1}^k \left| \frac{\widehat{M}_j - M_j}{M_j} \right|, \quad (4)$$

where M_1, \dots, M_k denote the M -factors computed by means of numerical simulations as described in Section 2.3 (ground truth), \bar{M} is the mean of M_1, \dots, M_k , and $\widehat{M}_1, \dots, \widehat{M}_k$ denote the M -factors obtained by an empirically derived regression formula.

For fitting the model parameters, we use all data of subsamples from 5 cm and 10 cm depth from both loam and sand (3072 subsamples). Model validation by means of R^2 and MAPE is performed on subsamples from 15 cm depth from both loam and sand (1536 subsamples). The parameters of all regression formulas considered in this study are fitted on the same (training) dataset, while validation of the fitted regression formulas is performed on subsamples from a depth which is not used for fitting. By performing model fitting and validation on different types of soil (loam and sand), we can show whether the best fitting regression formula is applicable for a general soil microstructure or only for structures very similar to those on which the model was fitted.

3. Results

275 **3.1. Statistical analysis of 3D pore space morphology**

The geometrical descriptors stated in Section 2.4 are used to quantitatively characterize the 3D microstructure of the soil subsamples considered in this paper. Figure 3 shows the histograms which

have been obtained for various geometrical descriptors of subsamples from loam and sand, respectively. It turned out that some descriptors, such as porosity ε and constrictivity β , can vary a lot between subsamples from the same soil texture at the chosen scale, whereas on average they differ only little between loam and sand, see Figures 3a and 3f. Thus, such descriptors might be less useful to distinguish between the textures of both soil textures, but rather useful for describing variations of individual samples from a given soil texture. On the other hand, descriptors like the specific surface area S and mean chord length $\mu(C)$ behave rather different for the two soil textures, see Figures 3b and 3c. Therefore, they seem to capture the morphological differences between the textures much better than ε and β .

A more detailed characterization of soil textures can be achieved by considering bivariate probability densities for pairs of descriptors, see Figure 4. While the univariate histograms of the mean chord length $\mu(C)$ obtained for subsamples of loam and sand show a certain overlapping, see Figure 3c, the joint (bivariate) probability densities of $\mu(C)$ and the mean spherical contact distance $\mu(H)$ allow to reliably distinguish between loam and sand, see Figure 4b.

When looking at Figures 3g and 3j, see also Figure 4 (e), we observe that the results obtained for the two notions of (geodesic and geometric) tortuosity are rather different. In particular, the mean geodesic tortuosity $\mu(\tau_g)$ is smaller for sand in comparison to the corresponding value obtained for loam (*i.e.* paths in the pore space of sand are shorter than those of loam), whereas the mean geometric tortuosity $\mu(\tau_s)$ is greater for sand. This is likely to be caused by the fact that sand comprises of larger pores than loam such that the paths along the coarse skeleton become comparatively long, whereas the finer pore structure observed in loam leads to a skeleton with a large number of nodes, which is closer to the actual axes of the pore space. The two notions of (geodesic and geometric) tortuosity will be further discussed in Section 4 with regard to their usefulness as geometrical descriptors for microstructure-property relationships.

A full analysis of the correlation coefficients between all pairs of geometrical descriptors considered in this paper is provided in the appendix, see Figure A.1. For selected pairs of geometrical descriptors, Figure 4 shows bivariate probability densities obtained by kernel density estimation as well as the corresponding Pearson correlation coefficients. With the exception of $(\varepsilon, \mu(\tau_g))$, the joint densities of all other pairs of geometrical descriptor shown in Figure 4 strongly depend on the considered soil texture, which once again highlights the structural differences between loam and sand. This suggests that any regression formula expressing the M -factor in terms of one of these pairs of geometrical descriptors would need to be adapted specifically to the distinct textures of sand and loam. Vice versa, regression formulas based on ε and $\mu(\tau_g)$ may be valid independent of the considered soil texture.

3.2. Regression formulas for the quantification of microstructure-property relationships

The scatter plots shown in Figure 5 illustrate the dependence of the M -factor on various pairs of geometrical descriptors. In this section, we consider several empirically derived regression formulas which will be used for the quantification of microstructure-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.

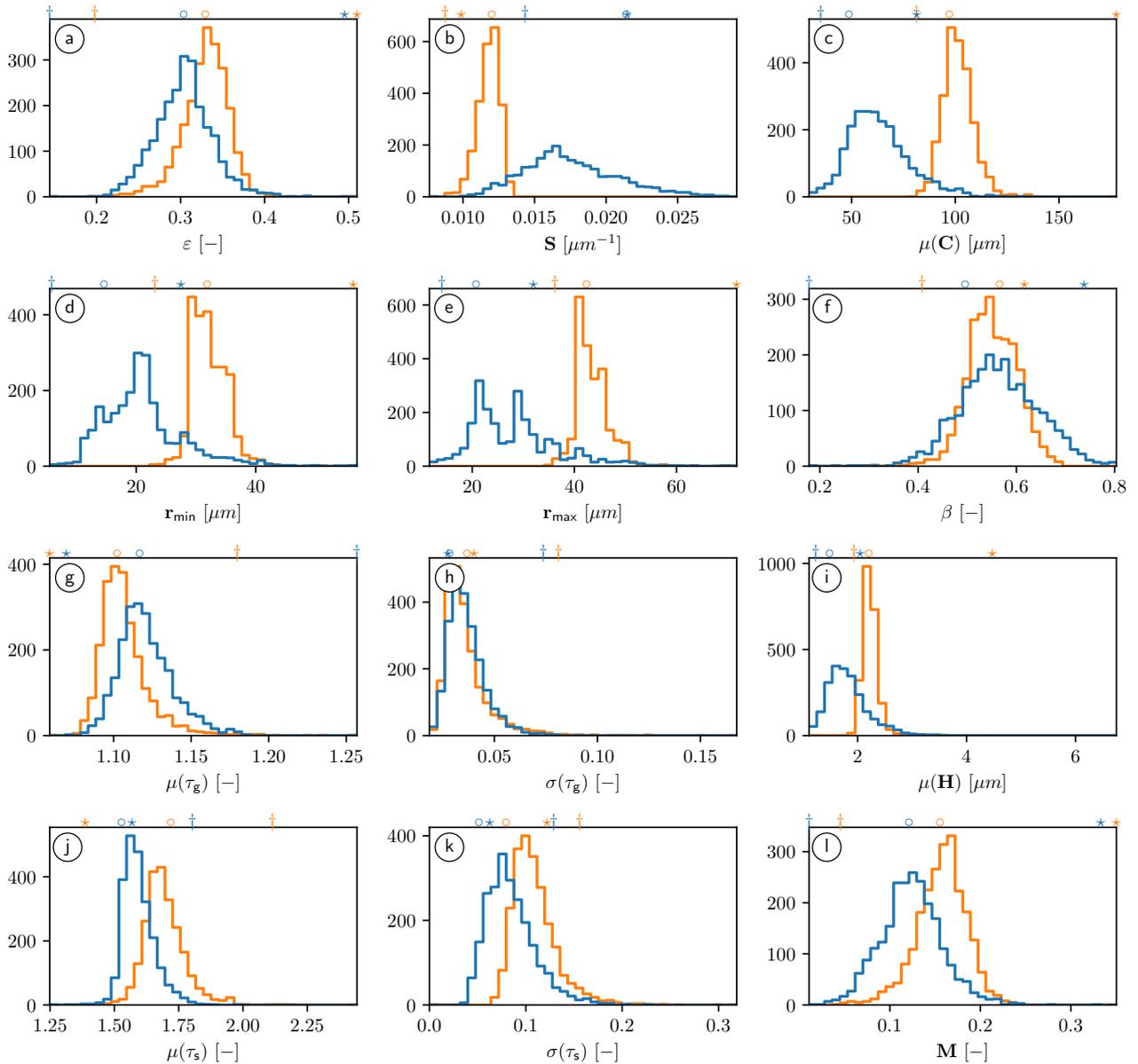


Figure 3: Histograms of geometrical descriptors – namely porosity (a), specific surface area (b), mean chord length (c), bottleneck radius r_{\min} (d), characteristic pore radius r_{\max} (e), constrictivity (f), mean and standard deviation of geodesic tortuosity (g,h), mean distance to solid phase (i), mean and standard deviation of geometric tortuosity (j) – and the M -factor (l) computed from tomographic image data for the $2 \cdot 3 \cdot 768 = 4608$ subsamples of loam (blue) and sand (orange), respectively. The symbols †, ○ and ★ indicate the values obtained for the subsamples shown in Figure 1.

A simple regression formula, which only involves the porosity ε , is given by

$$\widehat{M}_1 = \varepsilon^{c_1}, \quad \text{for some } c_1 \geq 1. \quad (5)$$

This power-law type of a regression formula is frequently used in the literature, where the most well-known examples are probably the Buckingham formula ($c_1 = 2$) [20], the Millington-Quirk formula ($c_1 = \frac{4}{3}$) [12], the formula derived by Marshall ($c_1 = \frac{3}{2}$) [11], and the work of Lai et al. ($c_1 = \frac{5}{3}$) [52]. The data considered in the present paper leads to $c_1 = 1.71$, which is right in between the values used

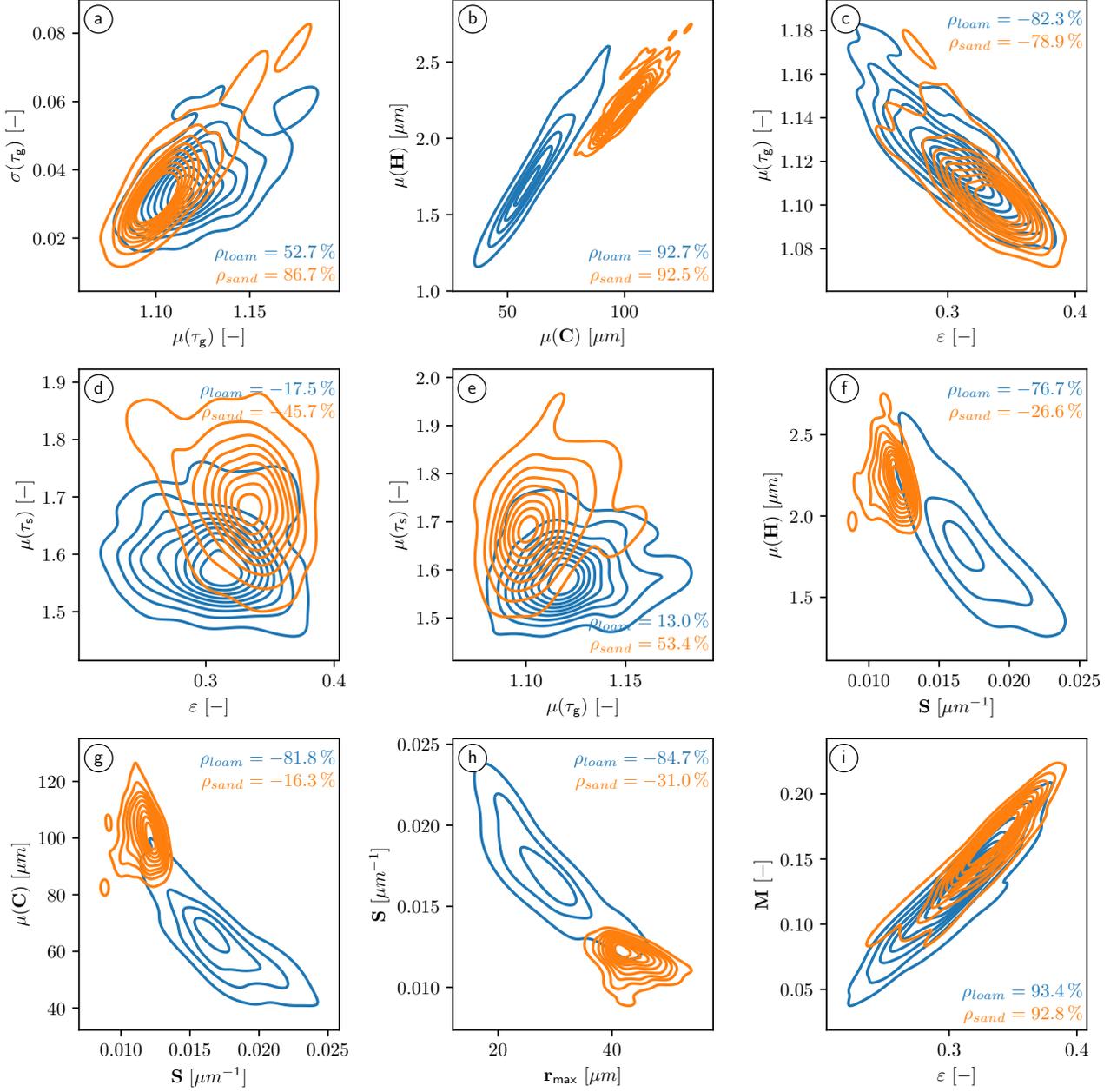


Figure 4: Bivariate probability densities for pairs of geometrical descriptors, visualized as contour plots based on kernel density estimation, which have been obtained for the $2 \cdot 3 \cdot 768 = 4608$ cubic subsamples of either loam (blue) or sand (orange). Furthermore, the values of Pearson's correlation coefficient ρ are given for each descriptor pair, separately computed for loam and sand.

in the Buckingham and Millington-Quirk formulas, but closest to that of Lai et al. [52].

Another simple regression formula is given by

$$\widehat{M}_2 = c_1 \varepsilon^{c_2}, \quad \text{for some } c_1, c_2 > 0, \quad (6)$$

325 which is used, for example, in [53] ($c_1 = 1.9, c_2 = 1.4$), [54] ($c_1 = 1.75, c_2 = 2.1$) and [55] ($c_1 = 5.25, c_2 = 3.36$). Moreover, the formula

$$\widehat{M}_3 = c_1 e^{c_2 \varepsilon}, \quad (7)$$

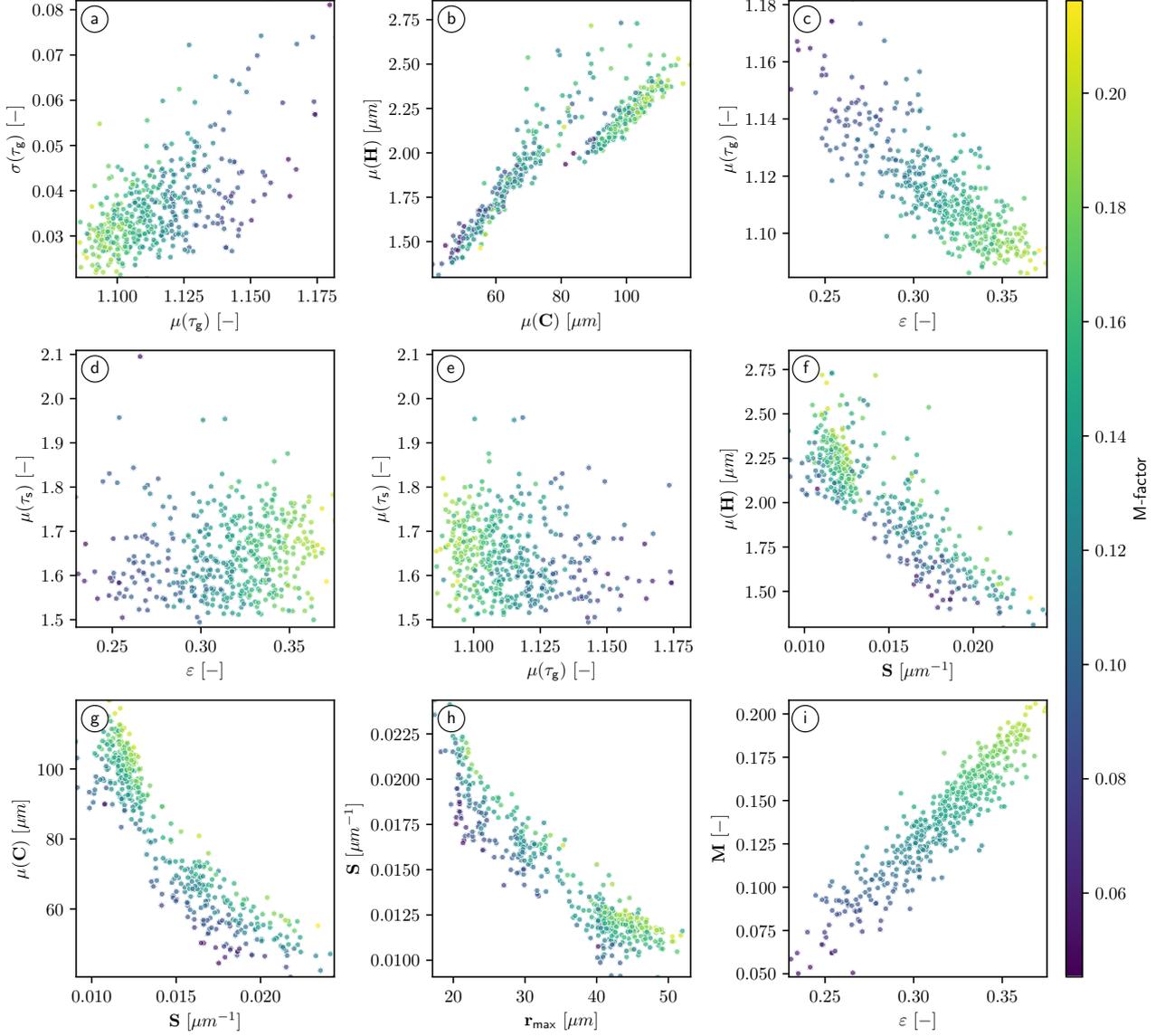


Figure 5: Scatter plots visualizing the dependence of the M -factor on various pairs of geometrical descriptors. The (brighter or darker) colors of the data points indicate whether the value of the corresponding M -factor is high or low. In this figure we do not distinguish between the two soil texture. For clarity of presentation, only randomly selected 10% of the 4608 cubic subsamples have been used to generate the 461 data points in each subfigure.

which has been introduced in [56] with $c_1 = 0.0085$ and $c_2 = 6.8$, is considered in the following. Besides, there are many other prediction formulas that only involve porosity, see e.g. [46, 57] for further details. However, a further consideration of this type of formulas is beyond the scope of the present paper, where a particular focus is now put on empirically derived regression formulas that – besides porosity – involve further (more sophisticated) descriptors of 3D pore space morphology such as the mean geodesic tortuosity $\mu(\tau_g)$ or the constrictivity β . In particular, we consider the formula

$$\widehat{M}_4 = \varepsilon^{c_1} \beta^{c_2} \mu(\tau_g)^{c_3}, \quad \text{for some } c_1 >, c_2 \geq 0, c_3 \leq 0, \quad (8)$$

which has been originally introduced in [58] (with $c_1 = 1.15, c_2 = 0.37, c_3 = -4.39$), and refitted to 63,000 virtually generated microstructures in [16] (with $c_1 = 1, c_2 = 0, c_3 = -8.45$).

335 Since Equation (8) does not hold in the dilute limit (*i.e.* for $\varepsilon \rightarrow 1$), the modified formula

$$\widehat{M}_5 = \varepsilon^{c_1 + c_2 \beta} \mu(\tau_{\mathbf{g}})^{c_3}, \quad \text{where } c_1 + c_2 \geq 0, c_3 \leq 0, \quad (9)$$

has been used in [42] (with $c_1 = 1.67, c_2 = -0.48, c_3 = -5.18$) and in [16] (with $c_1 = 1.25, c_2 = -1.25, c_3 = -7.82$), where the constrictivity β appears now in the exponent of the porosity ε .

Besides formulas based on porosity, constrictivity and mean geodesic tortuosity, further geometrical descriptors of 3D pore space morphology can be taken into account. For example, the formula

$$\widehat{M}_{6,j} = c_1 \mu(\tau_j)^{c_2} \sigma(\tau_j)^{c_3} \varepsilon^{c_4}, \quad \text{where } j \in \{\mathbf{g}, \mathbf{s}\}, \quad (10)$$

340 which has been introduced in [59] for $j = \mathbf{g}$ (with $c_1 = 0.06, c_2 = -2, c_3 = -0.6, c_4 = 1$) and also considered in [16] (with $c_1 = 1.18, c_2 = -9.17, c_3 = 0.03, c_4 = 1.02$), uses the standard deviation $\sigma(\tau_j)$ of (geodesic or geometric) tortuosity as an additional geometrical descriptor.

Moreover, we consider the formula

$$\widehat{M}_7 = \varepsilon^{c_1} S^{c_2} r_{\max}^{c_3}, \quad (11)$$

345 which has not yet been considered in the literature. In general, the specific surface area S has only rarely been used in the literature for the prediction of diffusive properties. However, a particular case of such a regression formula is presented in [15], which uses the formula $\widehat{M} = 1.1(\varepsilon - 0.039S^{0.52})$. However, this kind of regression formula can lead to negative values for the M -factor. In contrast, Equation (11) ensures that $\widehat{M}_7 \geq 0$.

Finally, we predict the M -factor by means of

$$\widehat{M}_8 = \varepsilon^{c_1} \left(\frac{\mu(C)}{\mu(H)} \right)^{c_2}, \quad (12)$$

350 which – besides porosity – contains the ratio of mean chord length $\mu(C)$ and the mean spherical contact distance $\mu(H)$ of a randomly chosen point within the pore space to the solid phase. The latter quantity can be regarded as some kind of shape information since this quotient in Equation (12) is larger for more elongated pores.

4. Discussion

355 First, we consider the case, where data gained for both loam as well as sand samples (taken in the depths of 5 and 10 cm) are used for training. Scatter plots visualizing the values obtained for the M -factor by numerical simulations versus those of the predicted M -factors $\widehat{M}_i, i = 1, \dots, 8$ obtained for the test data of either loam or sand is shown in Figure 6, where the parameters appearing in the prediction formulas given by Equations (5)–(12) have been fitted to the entire training data for loam and sand, see Table A.1. The resulting values obtained for the coefficient of determination (R^2) and the mean absolute percentage error (MAPE) can also be found in Figure 6. Furthermore, we compare the performance of the regression formulas given by Equations (5) – (12), where the model parameters are fitted to data considered in the present paper, with that of regression formulas which have been proposed previously in the literature, see Table A.2. Bear in mind that when using the latter formulas we do not refit the model parameters, but use the same parameter values as they have been derived
360 – based on different kinds of data and using different methodological approaches – in the original papers.

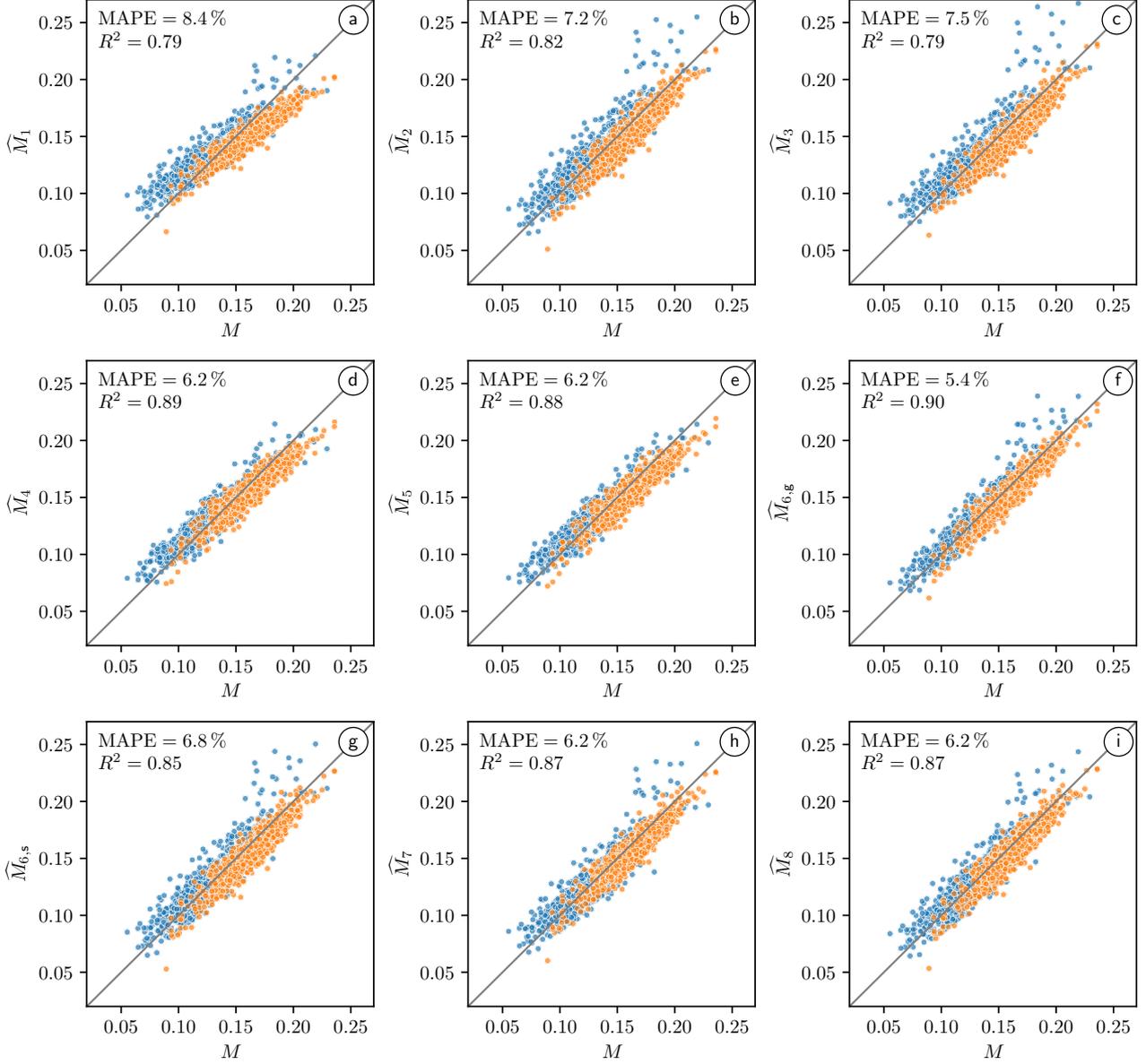


Figure 6: Scatter plots visualizing the values obtained for the M -factor by numerical simulations versus those of the predicted M -factors \widehat{M}_i , $i = 1, \dots, 8$ obtained for the test data of either loam (blue) or sand (orange). The parameters appearing in the prediction formulas given in Equations (5)–(12) have been fitted to the entire training data for loam and sand, see Table A.1. In particular, Equation (10), *i.e.* $\widehat{M}_{6,j}$, occurs twice since the concepts of both geodesic and geometric tortuosities are considered.

As expected, due to its simplicity, Equation (5) performs worst among all prediction formulas stated in Section 3.2. In this case, important structural details of the 3D pore space morphology are disregarded since only the porosity ε is used to characterize the morphology of the pore space. In case of Equation (5), one obtains $c_1 = 1.71$, when fitting it to the entire set of training data considered in the present paper. This value is roughly in the middle between 1.5 [11] and 2 [20], but quite close to the values of c_1 derived in [12] and [52]. When introducing an additional model parameter, as in Equations (6) and (7), the MAPE can be reduced to 7.2% and 7.5%, respectively. However, fitting Equation (6) to our soil data leads to quite different values for c_1 and c_2 compared to those obtained

in [55]. Due to the large difference in the values of the exponent c_2 of the porosity ε obtained in the present study and in [54, 55], respectively, it is not surprising that the predictive power of Equation (6) with $c_1 = 1.92$ and $c_2 = 2.29$, as derived from our soil data, is significantly better. On the other hand, note that the value which we obtained for c_1 in Equation (6) is similar to the one obtained in [53]. In case of results derived in [54], applying an already existing regression formula directly to the present soil data performs reasonably well (leading to a MAPE of 13.12%), even though refitting the regression parameters to the soil data considered in the present paper leads – as expected – to better results. With regard to Equation (7), it is interesting to point out that choosing $c_1 = 0.02$ and $c_2 = 6.88$ leads to a significantly better predictive power compared to that obtained in [56], even though their parameter values of $c_1 = 0.01$ and $c_2 = 6.8$ are quite close to the ones derived in the present paper.

A significant improvement of predictive power can be achieved by including further geometrical descriptors of the 3D pore space morphology into the regression formulas. However, fitting the model parameters in Equation (8) to our soil data leads to $c_2 = 0.04$, *i.e.*, the constrictivity β only plays a negligible role in the prediction of the M -factor. This effect has already been observed in [16], where 63,000 microstructures have been used for model fitting. Furthermore, directly applying the regression formulas derived in [16] and [58] to the present soil data leads to the best results among all (previously derived) formulas from the literature listed in Table A.2. In general, the validation scores R^2 and MAPE can be significantly improved by using Equation (8) compared to the prediction formulas given in Equations (5)–(7), which only use the porosity ε . This emphasizes the impact of the mean geodesic tortuosity $\mu(\tau_{\mathbf{g}})$ with regard to the prediction of diffusive mass transport in porous media. Interestingly, the value which we obtained in Equation (9) for the coefficient c_2 of constrictivity β is not equal to zero, see Table A.1. This matches with the results derived in [16] and [42]. Anyhow, it turned out that the predictive power of the regression formula given in Equation (9) is similar to that of the formula given in Equation (8), even though the value which we obtained for the exponent c_2 of β in Equation (8) is close to zero, *i.e.*, practically Equation (8) does not include the constrictivity β at all.

Note that the predictive power can be further increased when using Equation (10) for predicting the M -factor, where the regression formula given in Equation (10) is based on porosity, mean geodesic tortuosity as well as the standard deviation of geodesic tortuosity. In this case, for the validation scores R^2 and MAPE we obtained that $R^2 = 0.9$ and MAPE = 5.4%, which turns out to be the best result among all prediction formulas considered in the present paper. This indicates that not only the mean value $\mu(\tau_{\mathbf{g}})$ of shortest path lengths is of interest with regard to the prediction of diffusive mass transport, but also the corresponding standard deviation $\sigma(\tau_{\mathbf{g}})$. However, directly applying Equation (10) with the values of the regression parameters derived in [16] and [59] leads to a MAPE of 11.2% and 46.7%, respectively. This large difference can be explained by the corresponding values of the regression parameters, where in particular the values obtained for the exponents of $\mu(\tau_{\mathbf{g}})$ and $\sigma(\tau_{\mathbf{g}})$ differ significantly between the results of [16] ($c_2 = -9.17, c_3 = 0.03$), [59] ($c_2 = -2, c_3 = 0.6$) and our results ($c_2 = -6.07, c_3 = 0.07$) obtained in the present paper.

We also remark that Equation (11), which uses porosity, specific surface area as well as the characteristic pore radius r_{\max} , leads to similar results as Equations (8) and (9). Interestingly, using r_{\max} for predicting the M -factor turns out to be useful, even though Equation (8) fitted to our soil data does practically not make use of the constrictivity β , which is closely linked to r_{\max} . Finally, the M -factor can be predicted by means of Equation (12). This leads to $R^2 = 0.87$ and MAPE = 6.2%, which are similar values to those obtained for Equations (8), (9) and (11), see Table A.1. Thus, the dimensionless quantity in Equation (11) given by the ratio of the mean chord length $\mu(C)$ and the mean spherical contact distance $\mu(H)$, which characterizes the shape of pores, is – in combination

with porosity – well suited for predicting diffusive properties.

It is worth mentioning, that – in general – the question of the best geometrical descriptors with regard to the prediction of the M -factor is hard to answer, since certain geometrical descriptors that do not seem to be correlated with diffusive mass transport can be well suited for establishing microstructure-property relationships when they are combined with further microstructure characteristics. Moreover, for some features of 3D microstructures, several variants of geometrical descriptors are considered in the literature for one and the same structural feature. For example, as already mentioned in Section 2.4, there exist different notions of tortuosity, where it turned out that geodesic tortuosity is superior to geometric tortuosity with regard to the prediction of diffusive mass transport in soil, see the validation scores stated in Table A.1 for $\widehat{M}_{6,g}$ and $\widehat{M}_{6,s}$. This is likely to be caused by the fact that the skeleton of the pore space considered in the definition of $\widehat{M}_{6,s}$ takes the complex morphology of the pore space only partially into account. On the other hand, not surprisingly, all regression formulas for predicting the M -factor, whose parameters have been fitted to the soil data considered in the present paper, perform better with respect to both validation scores R^2 and MAPE than the corresponding regression formulas previously derived in the literature (and fitted to other types of image data), see Tables A.1 and A.2. This highlights the importance of using appropriately chosen data sets, which represent the nature of the 3D microstructure under consideration, for fitting the parameters of the regression formulas stated in Section 3.2.

Recall that so far we considered the case, where data gained from both loam as well as sand samples (taken in the depths of 5 and 10 cm) were used for training. Results regarding microstructure-property relationships, where only a subset of the training data (either for loam or sand) is used for fitting the regression parameters, can be found in the appendix, see Figures A.2, A.3 and A.4. As expected, the predictive power decreases in most cases when using only a subset of the training data for fitting compared to using the complete training data. However, using only data from sand for both fitting and validation often leads to even better results. This may be due to the fact that many descriptors as well as the M -factor show a smaller variation for sand than for loam. In addition, similar to the previously considered case, the concept of geodesic tortuosity leads to better results regarding the prediction of diffusive transport properties compared to geometric tortuosity. Moreover, it can be observed that Equation (10) seems to be quite robust in the sense that this regression formula leads to a large predictive power even if the corresponding regression parameters are fitted based on one soil texture and applied to the other soil texture. However, note that the regression formulas discussed in this section might only be valid to a limited extent if they are applied to other types of 3D microstructures, which are significantly different compared to the soil samples considered in the present paper.

In addition, the microstructure-property relationships in this study have been established on fairly small volumes ($1.28 \text{ mm} \times 1.28 \text{ mm} \times 1.28 \text{ mm}$). This had the favorable effect that the variability of the M -factor across all subsamples was quite large for packed soils and hence suitable for testing the accuracy of different models. However, not only diffusion but also some geometrical descriptors of 3D pore space morphology like tortuosity are scale-dependent [60]. Therefore, it will be tested in a forthcoming study to which extent these models hold for larger soil volumes.

5. Conclusion

In the present paper, we investigated the relationship between various geometrical descriptors of the 3D pore space morphology and diffusive transport properties for two soil different textures, namely loam and sand. For this purpose, sand and loam samples in 5, 10 and 15 cm depth have been imaged by means of X-ray computed tomography and segmented into two phases (pores and solid). The six binary 3D images obtained in this way have been partitioned into 768 non-overlapping subsamples each, which resulted in a total number of 4608 subsamples. For each of these subsamples, the 3D morphology of the pore space has then been characterized by means of various geometrical descriptors including, among others, the porosity as well as the specific surface area, mean geodesic tortuosity, constrictivity and mean spherical contact distance of the pore space. Besides this geometrical characterization of the pore space, diffusive transport properties have been computed by numerically solving the Laplace equation with inhomogeneous flux boundary conditions.

A comprehensive statistical analysis of the 3D pore space morphology has been carried out to quantify structural differences between sand and loam by determining univariate as well as bivariate probability distributions for individual geometrical descriptors and for pairs of them. Among others, it turned out that specific surface area and mean chord length of the pore space can be used to distinguish between loam and sand.

Moreover, microstructure-property relationships have been investigated for both types of soil, loam and sand, by means of parametric regression formulas. By using sophisticated geometrical descriptors of the 3D pore space morphology, which have not yet been considered in the context of soil gas diffusion, the predictive power of the regression formulas could be increased significantly. Among others, the considered descriptors of 3D microstructure include mean and standard deviation of geodesic tortuosity, constrictivity and mean chord length. In particular, it has been shown that with regard to the prediction of diffusive properties, the concept of geodesic tortuosity is superior compared to geometric tortuosity, which is based on skeletonization of the pore space. Furthermore, the robustness of the regression formulas has been investigated by fitting the regression parameters to one soil texture and applying the resulting regression formulas to the other soil texture. It turned out that the best performing regression formula is based on porosity as well as mean and standard deviation of geodesic tortuosity, leading to a mean absolute percentage error of about 5 %. While, in general, the predictive power of parametric regression formulas depends on the specific soil texture under consideration, the best regression formula performed well also in the case when fitted on one soil texture and applied to the other soil texture.

In summary, by our analysis we obtained an improved understanding of the relevance of certain structural features of 3D soil microstructures – which we measured by means of geometrical descriptors – for diffusive transport properties and we were able to predict them much more accurately compared to the precision of conventional regression formulas from the literature. A possible topic for future research is extending this analysis to structures at different length scales. Developing an appropriate multi-scale approach would allow for the prediction of diffusive properties for very large soil structures. Moreover, the present approach can be applied to more realistically developed soils, where pronounced anisotropy effects are to be expected.

6. Statements and Declarations

6.1. Acknowledgments

505 Nadja Ray and Alexander Prechtel thank Alice Lieu and Anne Hermann for supporting the data generation and handling of the effective diffusivities.

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 SCHM 997/33-1, 403640293, 403660839, 403801423).

510 6.3. Competing Interests

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

6.4. Author Contributions

515 Material preparation and data collection was performed by Maxime Phalempin and Steffen Schlüter. Nadja Ray and Alexander Prechtel elaborated the homogenized diffusion problems on the samples. Quantitative analysis and fitting of regression formulas were performed by Benedikt Prifling and Matthias Weber. The first draft of the manuscript was written by Benedikt Prifling and Matthias Weber. All authors contributed to writing and editing of the manuscript as well as the conception and design of the study.

6.5. Data Availability

520 The data sets generated during and/or analyzed during the present study are available in the Zenodo repository <https://doi.org/doi/10.5281/zenodo.7516228>.

References

- [1] M. Sahimi, *Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches*. Weinheim: Wiley-VCH, 2nd ed., 2011.
- 525 [2] J. Bear, *Modeling Phenomena of Flow and Transport in Porous Media*. Cham: Springer, 2018.

- [3] M. Blunt, *Multiphase Flow in Permeable Media: A Pore-Scale Perspective*. Cambridge: Cambridge University Press, 2017.
- [4] J. Newman and K. Thomas-Alyea, *Electrochemical Systems*. Hoboken: J. Wiley & Sons, 3rd ed., 2004.
- 530 [5] S. J. Cooper, M. Kishimoto, F. Tariq, R. S. Bradley, A. J. Marquis, N. P. Brandon, J. A. Kilner, and P. R. Shearing, “Microstructural analysis of an LSCF cathode using in situ tomography and simulation” *ECS Transactions*, vol. 57, 2671–2678, 2013.
- [6] M. K. Das, P. P. Mukherjee, and K. Muralidhar, *Modeling Transport Phenomena in Porous Media with Applications*. Cham: Springer, 2018.
- 535 [7] 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.
- [8] 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.
- 540 [9] R. Millington, “Gas diffusion in porous media” *Science*, vol. 130, no. 3367, 100–102, 1959.
- [10] H. L. Penman, “Gas and vapour movements in the soil: I. The diffusion of vapours through porous solids” *The Journal of Agricultural Science*, vol. 30, no. 3, 437–462, 1940.
- [11] T. Marshall, “The diffusion of gases through porous media” *Journal of Soil Science*, vol. 10, 79–82, 1959.
- 545 [12] R. J. Millington and J. P. Quirk, “Permeability of porous solids” *Transactions of the Faraday Society*, vol. 57, 1200–1207, 1961.
- [13] P. Moldrup, T. Olesen, J. Gamst, P. Schjønning, T. Yamaguchi, and D. Rolston, “Predicting the gas diffusion coefficient in repacked soil water-induced linear reduction model” *Soil Science Society of America Journal*, vol. 64, no. 5, 1588–1594, 2000.
- 550 [14] T. C. Deepagoda, P. Moldrup, P. Schjønning, L. W. de Jonge, K. Kawamoto, and T. Komatsu, “Density-corrected models for gas diffusivity and air permeability in unsaturated soil” *Vadose Zone Journal*, vol. 10, no. 1, 226–238, 2011.
- [15] P. Moldrup, T. Olesen, T. Komatsu, P. Schjønning, and D. Rolston, “Tortuosity, diffusivity, and permeability in the soil liquid and gaseous phases” *Soil Science Society of America Journal*,
555 vol. 65, no. 3, 613–623, 2001.
- [16] B. Prifling, M. Röding, P. Townsend, M. Neumann, and V. Schmidt, “Large-scale statistical learning for mass transport prediction in porous materials using 90,000 artificially generated microstructures” *Frontiers in Materials*, vol. 8, 786502, 2021.
- [17] T. S. da Silva, M. Pulido-Moncada, M. R. Schmidt, S. Katuwal, S. Schlüter, J. M. Köhne,

- 560 M. Mazurana, L. Juhl Munkholm, and R. Levien, “Soil pore characteristics and gas transport properties of a no-tillage system in a subtropical climate” *Geoderma*, vol. 401, 115222, 2021.
- [18] S. Katuwal, T. Norgaard, P. Moldrup, M. Lamandé, D. Wildenschild, and L. W. de Jonge, “Linking air and water transport in intact soils to macropore characteristics inferred from X-ray computed tomography” *Geoderma*, vol. 237, 9–20, 2015.
- 565 [19] S. Katuwal, E. Arthur, M. Tuller, P. Moldrup, and L. W. de Jonge, “Quantification of soil pore network complexity with X-ray computed tomography and gas transport measurements” *Soil Science Society of America Journal*, vol. 79, no. 6, 1577–1589, 2015.
- [20] E. Buckingham, *Contributions to our knowledge of the aeration of soils*. Washington: US Department of Agriculture, Bureau of Soils, 1904.
- 570 [21] 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.
- [22] E. Lippold, M. Phalempin, S. Schlüter, and D. Vetterlein, “Does the lack of root hairs alter root system architecture of zea mays?” *Plant Soil*, vol. 467, 267–286, 2021.
- [23] 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.
- 575 [24] J. Banhart, *Advanced Tomographic Methods in Materials Research and Engineering*. Oxford: Oxford University Press, 2008.
- 580 [25] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, (San Diego), 60–65, IEEE Computer Society, 2005.
- [26] W. Burger and M. Burge, *Digital Image Processing: An Algorithmic Introduction Using Java*. London: Springer, 2nd ed., 2016.
- 585 [27] N. Otsu, “A threshold selection method from gray-level histograms” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, 62–66, 1979.
- [28] U. Hornung, *Homogenization and Porous Media*. New York: Springer, 1997.
- [29] V. Aizinger, A. Rupp, J. Schütz, and P. Knabner, “Analysis of a mixed discontinuous Galerkin method for instationary Darcy flow” *Computational Geosciences*, vol. 22, 179–194, 2018.
- 590 [30] A. Rupp and P. Knabner, “Convergence order estimates of the local discontinuous Galerkin method for instationary Darcy flow” *Numerical Methods For Partial Differential Equations*, vol. 33, 1374–1394, 2017.
- [31] C. Wieners, “Distributed point objects. A new concept for parallel finite elements” in *Domain Decomposition Methods in Science and Engineering* (T. J. Barth, M. Griebel, D. E. Keyes, R. M.

- 595 Nieminen, D. Roose, T. Schlick, R. Kornhuber, R. Hoppe, J. Périaux, O. Pironneau, O. Widlund,
and J. Xu, eds.), (Berlin), 175–182, Springer, 2005.
- [32] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Philadelphia: Society for Industrial and Applied Mathematics, 2nd ed., 2003.
- [33] M. A. Heroux and J. M. Willenbring, “Trilinos Users Guide” Tech. Rep. SAND2003-2952, Sandia
600 National Laboratories, 2003.
- [34] A. Prechtel, A. Lieu, M. Phalempin, and R. Schulz, “Evaluating the effective diffusion in soil columns including biopores by homogenization” 2022. Preprint-Reihe Angewandte Mathematik 416, Friedrich-Alexander Universität Erlangen-Nürnberg, ISSN 2194-5127.
- [35] 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.
605
- [36] K. Schladitz, J. Ohser, and W. Nagel, “Measuring intrinsic volumes in digital 3D images” in *13th International Conference on Discrete Geometry for Computer Imagery* (A. Kuba, L. Nyúl, and K. Palágyi, eds.), 247–258, Springer, 2007.
- [37] G. Matheron, *Random Sets and Integral Geometry*. New York: J. Wiley & Sons, 1975.
610
- [38] J. Serra, *Image Analysis and Mathematical Morphology*. London: Academic Press, 1982.
- [39] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin: Springer, 2nd ed., 2013.
- [40] B. Münch and L. Holzer, “Contradicting geometrical concepts in pore size analysis attained with electron microscopy and mercury intrusion” *Journal of the American Ceramic Society*, vol. 91, no. 12, 4059–4067, 2008.
615
- [41] L. Holzer, D. Wiedenmann, B. Münch, L. Keller, M. Prestat, P. Gasser, I. Robertson, and B. Grobóty, “The influence of constrictivity on the effective transport properties of porous layers in electrolysis and fuel cells” *Journal of Materials Science*, vol. 48, 2934–2952, 2013.
- [42] M. Neumann., O. Stenzel, F. Willot, L. Holzer, and V. Schmidt, “Quantifying the influence of microstructure on effective conductivity and permeability: virtual materials testing” *International Journal of Solids and Structures*, vol. 184, 211–220, 2020.
620
- [43] J. Mayer, “A time-optimal algorithm for the estimation of contact distribution functions of random closed sets” *Image Analysis and Stereology*, vol. 23, 177–183, 2004.
- [44] C. R. Maurer, R. Qi, and V. Raghavan, “A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, 265–270, 2003.
625
- [45] D. Jungnickel, *Graphs, Networks and Algorithms*. Berlin: Springer, 3rd ed., 2008.

- [46] L. Holzer, P. Marmet, M. Fingerle, A. Wiegmann, M. Neumann, and V. Schmidt, *Tortuosity and Microstructure Effects in Porous Media: Classical Theories, Empirical Data and Modern Methods*. 2022. submitted.
- [47] B. Ghanbarian, A. G. Hunt, R. P. Ewing, and M. Sahimi, “Tortuosity in porous media: a critical review” *Soil science society of America journal*, vol. 77, no. 5, 1461–1477, 2013.
- [48] T. Lee, R. Kashyap, and C. Chu, “Building skeleton models via 3-D medial surface axis thinning algorithms” *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, 462–478, 1994.
- [49] P. K. Saha, G. Borgfors, and G. Sanniti di Baja, “A survey on skeletonization algorithms and their applications” *Pattern Recognition Letters*, vol. 76, 3–12, 2016.
- [50] J. J. Moré, “The Levenberg-Marquardt algorithm: Implementation and theory” in *Numerical Analysis* (G. A. Watson, ed.), vol. 630, 105–116, Heidelberg: Springer, 1978.
- [51] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental algorithms for scientific computing in python” *Nature Methods*, vol. 17, 261–272, 2020.
- [52] S.-H. Lai, J. M. Tiedje, and A. E. Erickson, “In situ measurement of gas diffusion coefficient in soils” *Soil Science Society of America Journal*, vol. 40, no. 1, 3–6, 1976.
- [53] J. A. Currie, “Gaseous diffusion in porous media. Part 2. – Dry granular materials” *British Journal of Applied Physics*, vol. 11, no. 8, 318–327, 1960.
- [54] J. A. Currie, “Gaseous diffusion in porous media. Part 3 – Wet granular materials” *British Journal of Applied Physics*, vol. 12, no. 6, 275–281, 1961.
- [55] A. R. Grable and E. G. Siemer, “Effects of bulk density, aggregate size, and soil water suction on oxygen diffusion, redox potentials, and elongation of corn roots” *Soil Science Society of America Journal*, vol. 32, no. 2, 180–186, 1968.
- [56] J. Richter and A. Großgebauer, “Untersuchungen zum Bodenlufthaushalt in einem Bodenbearbeitungsversuch. 2. Gasdiffusionskoeffizienten als Strukturmaße für Böden” *Zeitschrift für Pflanzenernährung und Bodenkunde*, vol. 141, no. 2, 181–202, 1978.
- [57] L. Shen and Z. Chen, “Critical review of the impact of tortuosity on diffusion” *Chemical Engineering Science*, vol. 62, no. 14, 3748–3755, 2007.
- [58] O. Stenzel, O. M. Pecho, L. Holzer, M. Neumann, and V. Schmidt, “Predicting effective conductivities based on geometric microstructure characteristics” *AIChE Journal*, vol. 62, 1834–1843, 2016.

- 665 [59] S. Barman, H. Rootzén, and D. Bolin, “Prediction of diffusive transport through polymer films from characteristics of the pore geometry” *AIChE Journal*, vol. 65, 446–457, 2019.
- [60] B. Ghanbarian, “Scale dependence of tortuosity and diffusion: Finite-size scaling analysis” *Journal of Contaminant Hydrology*, vol. 245, 103953, 2022.

A. Appendix

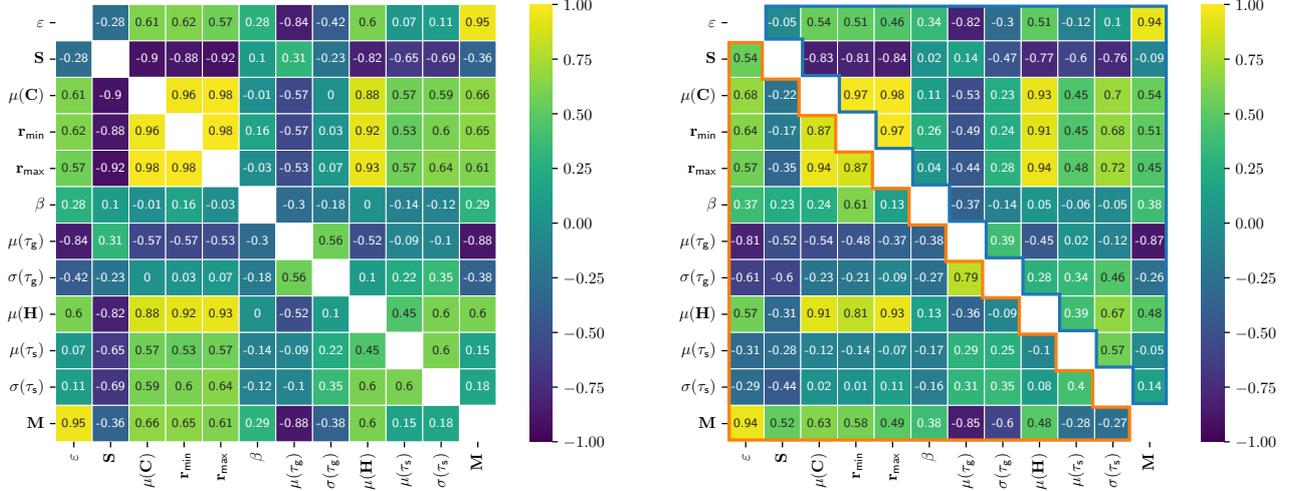


Figure A.1: Correlation coefficients between all pairs of geometrical descriptors (as well as the M -factor) considered in the present paper computed for the set of all 4608 subsamples (left) and for loam and sand individually (right), where in the latter case the upper right part of the matrix (blue frame) shows the values obtained for loam, the bottom left part (orange frame) those for sand.

	Regression formula	R^2	MAPE [%]
$\widehat{M}_1 =$	$\varepsilon^{1.71}$	0.79	8.40
$\widehat{M}_2 =$	$1.92\varepsilon^{2.29}$	0.82	7.23
$\widehat{M}_3 =$	$0.02e^{6.88\varepsilon}$	0.79	7.52
$\widehat{M}_4 =$	$\varepsilon^{1.06}\beta^{-0.04}\mu(\tau_g)^{-7.28}$	0.89	6.19
$\widehat{M}_5 =$	$\varepsilon^{1.17-0.18\beta}\mu(\tau_g)^{-7.00}$	0.88	6.23
$\widehat{M}_{6,g} =$	$2.19\mu(\tau_g)^{-6.07}\sigma(\tau_g)^{0.07}\varepsilon^{1.63}$	0.9	5.39
$\widehat{M}_{6,s} =$	$1.68\mu(\tau_s)^{0.23}\sigma(\tau_s)^{0.00}\varepsilon^{2.28}$	0.85	6.81
$\widehat{M}_7 =$	$\varepsilon^{2.60}S^{-0.61}r_{\max}^{-0.45}$	0.87	6.23
$\widehat{M}_8 =$	$\varepsilon^{2.23}(\mu(C)/\mu(H))^{0.16}$	0.87	6.22

Table A.1: 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.

	Regression formula	R^2	MAPE [%]	Reference
$\widehat{M}_1 =$	$\varepsilon^{2.00}$	-1.1	27.83	[20]
$\widehat{M}_1 =$	$\varepsilon^{1.50}$	-0.62	28.03	[11]
$\widehat{M}_1 =$	$\varepsilon^{1.33}$	-5.1	55.00	[12]
$\widehat{M}_2 =$	$1.90\varepsilon^{1.40}$	-62	172.81	[53]
$\widehat{M}_2 =$	$1.75\varepsilon^{2.10}$	0.53	13.12	[54]
$\widehat{M}_2 =$	$5.25\varepsilon^{3.36}$	-0.016	20.08	[55]
$\widehat{M}_3 =$	$0.01e^{6.80\varepsilon}$	-4.4	46.70	[56]
$\widehat{M}_4 =$	$\varepsilon^{1.15}\beta^{0.37}\mu(\tau_g)^{-4.39}$	0.76	8.56	[58]
$\widehat{M}_4 =$	$\varepsilon^{1.00}\beta^{0.00}\mu(\tau_g)^{-8.45}$	0.73	8.93	[16]
$\widehat{M}_5 =$	$\varepsilon^{1.67-0.48\beta}\mu(\tau_g)^{-5.18}$	0.084	17.79	[42]
$\widehat{M}_5 =$	$\varepsilon^{1.25-0.25\beta}\mu(\tau_g)^{-7.82}$	0.46	13.27	[16]
$\widehat{M}_{6,g} =$	$0.06\mu(\tau_g)^{2.00}\sigma(\tau_g)^{-0.60}\varepsilon^{1.00}$	-1.2	30.35	[59]
$\widehat{M}_{6,g} =$	$1.18\mu(\tau_g)^{-9.17}\sigma(\tau_g)^{0.03}\varepsilon^{1.02}$	0.6	11.22	[16]
$\widehat{M} =$	$1.1 (\varepsilon - 0.039S^{0.52})$	-43	146.67	[15]

Table A.2: Regression functions taken from the literature; validation scores R^2 and MAPE obtained for the validation data considered in the present paper.

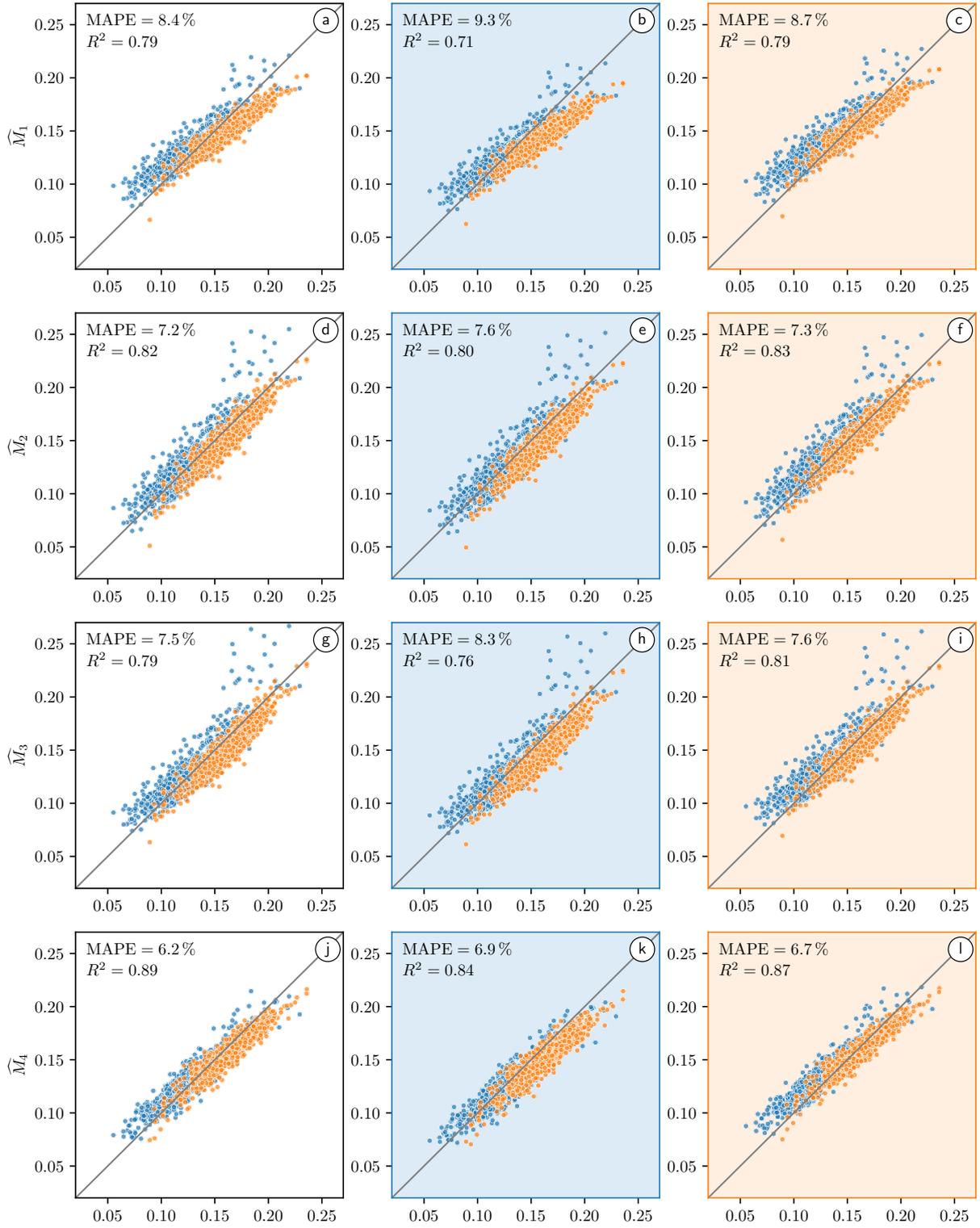


Figure A.2: M -factor obtained by numerical simulations versus the predicted M -factor \widehat{M}_i , $i = 1, \dots, 4$, obtained for the validation data, where different regression formulas (given by Equations (5), (6), (7) and (8); from top to bottom) have been used. The colors of data points represent 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.

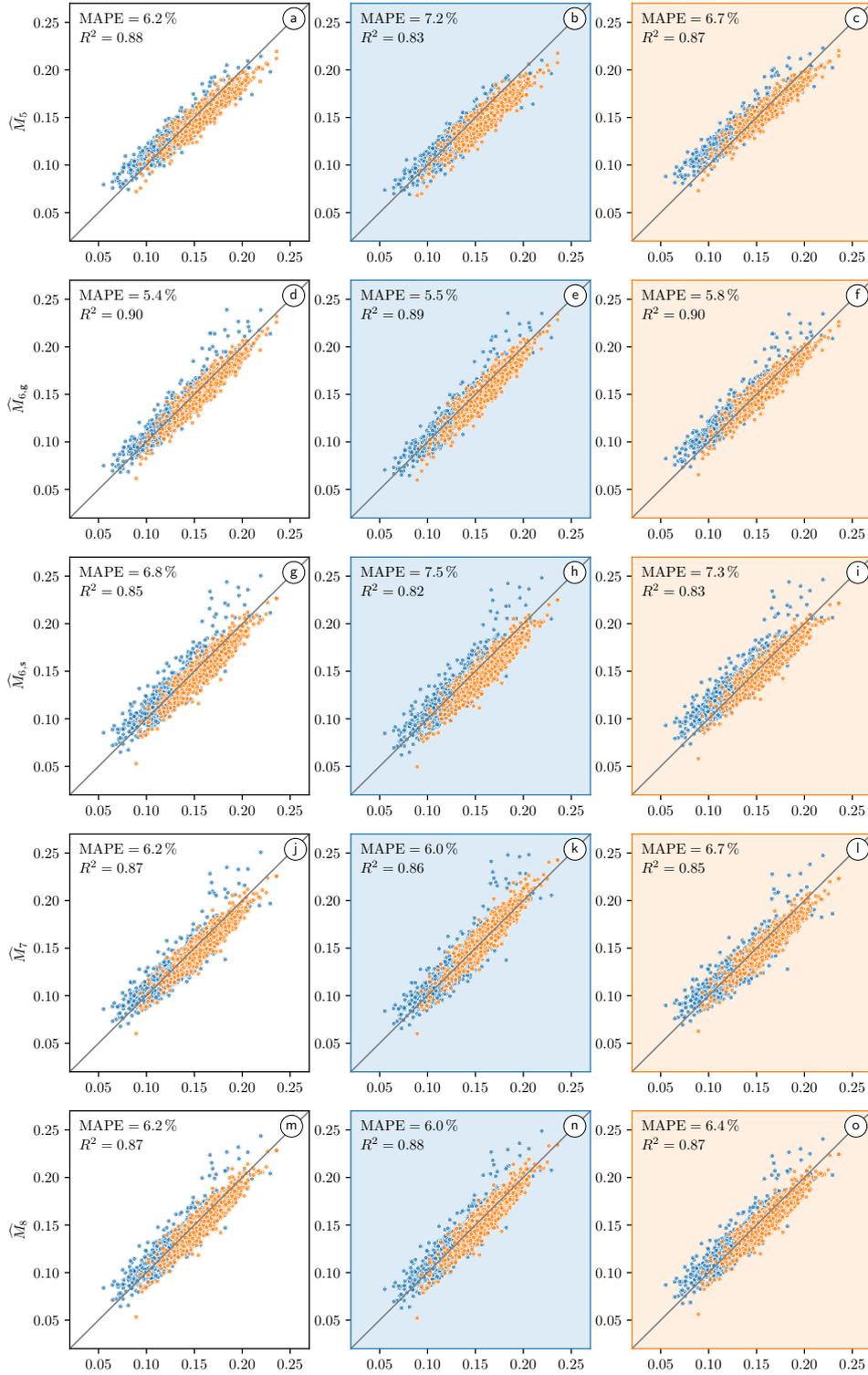


Figure A.3: M -factor obtained by numerical simulations versus the predicted M -factor \widehat{M}_i , $i = 5, \dots, 8$, obtained for the validation data, where different regression formulas (given by Equations (9), (10), (11), (12); from top to bottom) have been used. In particular, Equation (10) is 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.

$\widehat{M}_1 = \varepsilon^{c_1}$										$\widehat{M}_2 = c_1 \varepsilon^{c_2}$													
			R^2			MAPE [%]			c_1						R^2			MAPE [%]			(c_1, c_2)		
Fitting on	All	0.79	0.73	0.61	8.4	9.3	7.6	1.71	Fitting on	All	0.82	0.74	0.73	7.2	8.1	6.3	(1.92, 2.29)						
	Loam	0.71	0.80	0.28	9.3	7.7	10.8	1.75		Loam	0.80	0.77	0.62	7.6	7.6	7.7	(1.94, 2.32)						
	Sand	0.79	0.59	0.79	8.7	12.1	5.4	1.68		Sand	0.83	0.69	0.81	7.3	9.4	5.2	(1.60, 2.11)						
			Validation on									Validation on											
$\widehat{M}_3 = c_1 \varepsilon^{c_2 \varepsilon}$										$\widehat{M}_4 = \varepsilon^{c_1} \beta^{c_2} \mu(\tau_{\bar{g}})^{c_3}$													
			R^2			MAPE [%]			(c_1, c_2)						R^2			MAPE [%]			(c_1, c_2, c_3)		
Fitting on	All	0.79	0.72	0.63	7.5	8.2	6.9	(0.02, 6.88)	Fitting on	All	0.89	0.84	0.80	6.2	7.1	5.2	(1.06, -0.04, -7.28)						
	Loam	0.76	0.76	0.46	8.3	7.6	9.0	(0.01, 6.90)		Loam	0.84	0.86	0.64	6.9	6.4	7.5	(0.91, 0.08, -8.55)						
	Sand	0.81	0.66	0.76	7.6	9.8	5.4	(0.02, 6.36)		Sand	0.87	0.76	0.87	6.7	9.2	4.3	(1.21, 0.01, -5.13)						
			Validation on									Validation on											
$\widehat{M}_5 = \varepsilon^{c_1 + c_2 \beta} \mu(\tau_{\bar{g}})^{c_3}$										$\widehat{M}_{6,g} = c_1 \mu(\tau_{\bar{g}})^{c_2} \sigma(\tau_{\bar{g}})^{c_3} \varepsilon^{c_4}$													
			R^2			MAPE [%]			(c_1, c_2, c_3)						R^2			MAPE [%]			(c_1, c_2, c_3, c_4)		
Fitting on	All	0.88	0.85	0.79	6.2	6.9	5.5	(1.17, -0.18, -7.00)	Fitting on	All	0.90	0.87	0.82	5.4	5.7	5.0	(2.19, -6.07, 0.07, 1.63)						
	Loam	0.83	0.85	0.60	7.2	6.3	8.0	(1.21, -0.34, -7.86)		Loam	0.89	0.87	0.80	5.5	5.7	5.4	(1.94, -7.24, 0.03, 1.55)						
	Sand	0.87	0.76	0.87	6.7	9.2	4.2	(1.34, -0.18, -4.84)		Sand	0.90	0.82	0.87	5.8	7.2	4.4	(1.64, -4.70, 0.04, 1.57)						
			Validation on									Validation on											
$\widehat{M}_{6,s} = c_1 \mu(\tau_{\bar{s}})^{c_2} \sigma(\tau_{\bar{s}})^{c_3} \varepsilon^{c_4}$										$\widehat{M}_7 = \varepsilon^{c_1} S^{c_2} r_{\max}^{c_3}$													
			R^2			MAPE [%]			(c_1, c_2, c_3, c_4)						R^2			MAPE [%]			(c_1, c_2, c_3)		
Fitting on	All	0.85	0.77	0.76	6.8	7.6	6.0	(1.68, 0.24, 0.00, 2.29)	Fitting on	All	0.87	0.80	0.80	6.2	7.3	5.2	(2.60, -0.61, -0.45)						
	Loam	0.82	0.77	0.65	7.5	7.6	7.4	(1.63, 0.22, -0.03, 2.33)		Loam	0.86	0.76	0.81	6.0	7.4	4.5	(2.95, -0.86, -0.63)						
	Sand	0.82	0.67	0.81	7.5	9.8	5.2	(1.48, 0.03, -0.02, 2.11)		Sand	0.85	0.77	0.79	6.7	8.0	5.5	(2.79, -0.88, -0.72)						
			Validation on									Validation on											
$\widehat{M}_8 = \varepsilon^{c_1} (\mu(C)/\mu(H))^{c_2}$																							
			R^2			MAPE [%]			(c_1, c_2)														
Fitting on	All	0.87	0.81	0.80	6.2	7.1	5.3	(2.23, 0.16)															
	Loam	0.88	0.80	0.83	6.0	7.2	4.9	(2.31, 0.18)															
	Sand	0.87	0.80	0.81	6.4	7.5	5.2	(2.13, 0.13)															
			Validation on																				

Figure A.4: 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.