

Copula-based modeling and simulation of 3D systems of curved fibers by isolating intrinsic fiber properties and external effects

Matthias Weber,^{1,*} Andreas Grießer,² Dennis Mosbach,² Erik Glatt,²
Andreas Wiegmann,² and Volker Schmidt¹

October 5, 2023

¹ Institute of Stochastics, Ulm University, Helmholtzstraße 18, 89069 Ulm, Germany

² Math2Market GmbH, Richard-Wagner-Straße 1, D-67655, Kaiserslautern, Germany

* corresponding author, matthias.weber@uni-ulm.de

Abstract

In this paper we lay the foundation for data-driven 3D analysis of virtual fiber systems with respect to their microstructure and functionality. In particular, we develop a stochastic 3D model for systems of curved fibers similar to nonwovens, which is fitted to tomographic image data. By systematic variations of model parameters, efficient computer-based scenario analyses can be performed to get a deeper insight how effective properties of this type of functional materials depend on their 3D microstructure. In a first step, we consider single fibers as polygonal tracks which can be modeled by a third-order Markov chain. For constructing the transition function of the Markov chain, we formalize the intuitive notions of intrinsic fiber properties and external effects and build a copula-based transition function such that both aspects can be varied independently. Using this single-fiber model, in a second step we derive a model for the entire fiber system observed in a bounded sampling window and fit it to two different 3D datasets of nonwovens measured by CT imaging. Considering various geometric descriptors of the 3D microstructure related to effective properties of the pore space, we evaluate the goodness of model fit by comparing geometric descriptors of the 3D morphology of model realizations with those of tomographic image data.

Keywords: Stochastic modeling, curved fiber, polygonal track, Markov chain, transition function, copula, nonwoven material, 3D microstructure, tomographic image data, segmentation, convolutional neural network

1 Introduction

A wide range of applications from fuel cell technology [28], hygiene products [19] and filtration [10] heavily depend on the usage of fiber-based materials like nonwovens. Here, various properties of nonwovens like diffusivity or wettability must be tailored to meet specific needs. Improving on these properties by means of traditional manufacturing and testing is expensive and time-consuming. However, when the microstructure of a nonwoven is known, numerical methods can be used for in-silico investigation of effective properties. Combined with a stochastic microstructure model, this can be used to perform virtual materials testing by creating and investigating realistic structures which yet have not been considered before. By this approach, insight into the relationship between microstructure geometry and effective properties can be obtained at a low cost which can then be used to streamline the procedure for developing materials with enhanced properties [15, 27, 30].

Various stochastic microstructure models for nonwovens have been proposed in the literature which are used for different purposes like filter media [1, 2, 29, 37], gas-diffusion layers in fuel cells [9, 28, 38] and general applications [5, 18, 21, 22, 36], while other models are concerned with specific fiber properties like orientation and distance [4, 14]. In [34], we recently developed a stochastic model for creating single fibers similar to those seen in various nonwoven materials. This model was based on representing fibers using the so-called Frenet-Serret formulas, but did not include any information about the global appearance of the fiber system. In the present paper, we adapt this model to account for what one might intuitively describe as intrinsic properties of fibers as well as external effects acting upon fibers. Roughly speaking, we formalize these notions by considering all those properties as intrinsic which do not affect the global appearance of fibers, where we consider as external effects only information about the global orientation of parts of fibers based on the distance of the fiber segments to the upper and lower boundaries of the material. Combining both intrinsic properties and external effects allows for the application of the developed model to describe entire fiber systems.

We assume that each fiber can be considered as a third-order Markov chain [26] and model the transition function as a product of two (conditional) univariate probability densities, one of which represents the intrinsic properties of fibers and the other one represents the external effects. The underlying (unconditional) bi- and trivariate probability densities are in turn modeled by means of copulas [7, 17], which provide a powerful tool for stochastic modeling of non-Gaussian correlated vector data. Together with a simple model for the generation of initial points, this Markov chain constitutes the basis of a model for entire fiber systems which we fit to two different 3D datasets of nonwovens measured by CT imaging, see Figure 1, and processed using a convolutional neural network [11, 12] to extract the center-lines of fibers. Considering various geometric descriptors of the 3D microstructure related to effective properties of the pore space, we evaluate the goodness of model fit by comparing geometric descriptors of the

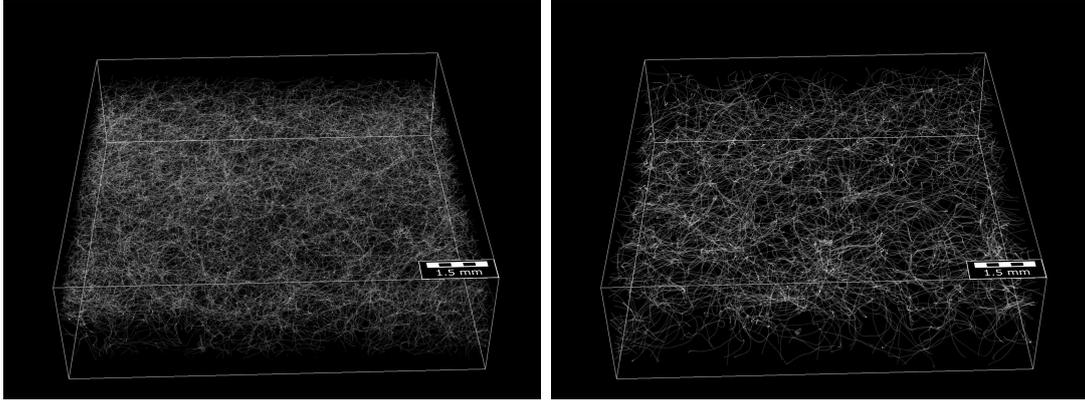


Figure 1: 3D view of the gray value image of Sample 1 (left) and Sample 2 (right).

3D morphology of model realizations with those of tomographic image data. For the two different material samples considered in this paper, Figure 2 shows both segmented 3D images of measured fiber systems along with realizations of the respective fitted models. The overall visual impression of Figure 2 is that measured and simulated fiber systems are quite similar for each of the two samples, but rather different between the two samples. This qualitative assessment will be specified later in the paper.

The rest of the paper is organized as follows. We first describe the methods used for constructing and validating the model as well as the underlying measured data. In particular, we describe the stochastic 3D model for single curved fibers and entire fiber systems, which is fitted to tomographic image data. Furthermore, we evaluate the goodness of model fit by comparing geometric descriptors of the 3D morphology of model realizations with those of CT data. Finally, we discuss the obtained results and outline possible further research.

2 Materials and methods

2.1 Description of nonwoven material

We investigated two samples of nonwoven materials in this paper. The size of Sample 1 is $11.2\text{mm} \times 11.2\text{mm} \times 1.9\text{mm}$ and that of Sample 2 is $10.8\text{mm} \times 10.8\text{mm} \times 2.1\text{mm}$. Sample 1 consists of a single type of circular Polyethylen-Polyethylenterephthalat (PE-PET) fibers. The material has a solid volume fraction of 1.4% and an average fiber diameter of 18.12 micron. Sample 2 consists of a single type of circular Polyethylenterephthalat (PET) fibers. Its solid volume fraction of is 0.71% and fibers have an average diameter of 22.65 micron. Both samples consist of staple fibers with lengths of a few millimeters

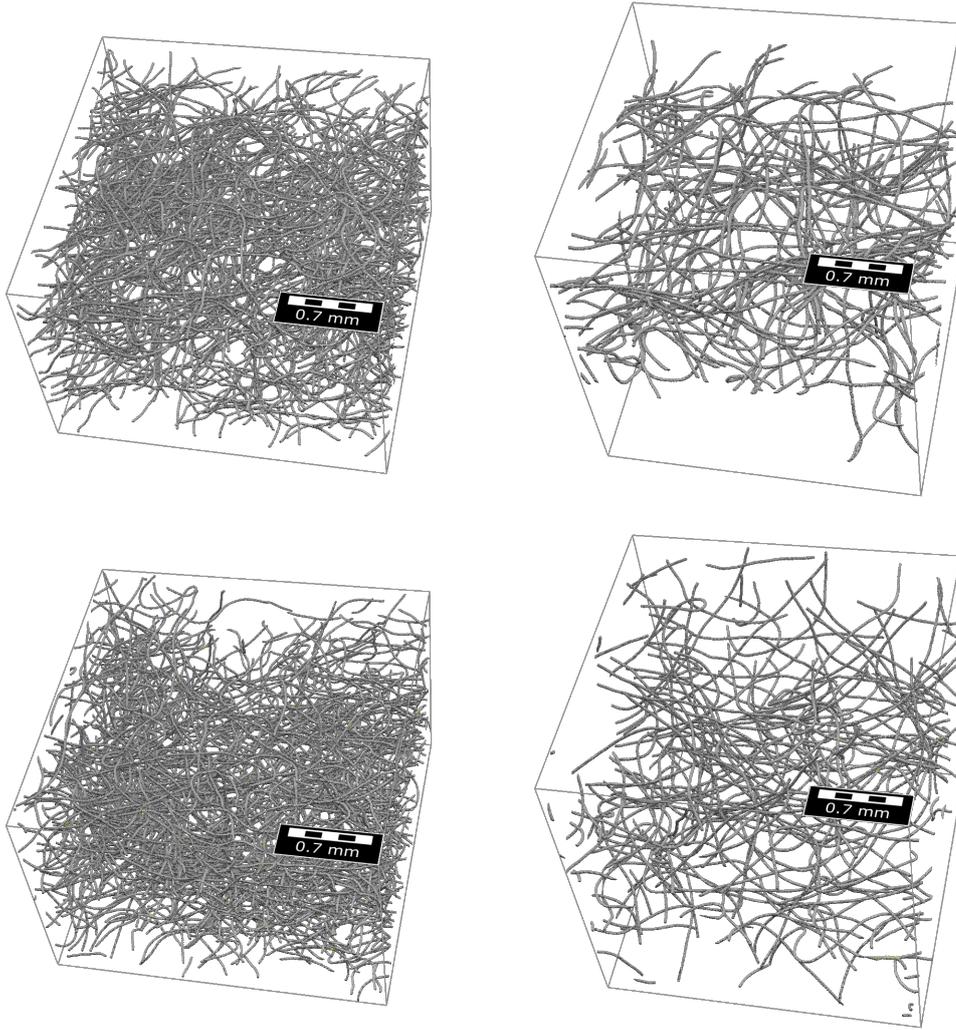


Figure 2: Cutouts of measured (top) and simulated (bottom) fiber systems for Sample 1 (left) and Sample 2 (right). See also Figure 1.

and monodisperse diameters.

These nonwovens were specifically created to investigate the curvature of fiber systems. The usual air bonding step was omitted to simplify the identification of individual fibers. Similar materials are commonly used in hygiene products, making the results of this research relevant for such applications. By employing these unique nonwovens, we can explore the inherent curvature properties of fiber systems and gain valuable insights on how to model these types of fiber systems.

2.2 Data acquisition

The two nonwoven materials were scanned in a Scanco MicroCT50 micro-CT scanner at a resolution of 2 micron and an energy of 45KeV. Overall 3600 projections were taken using Off-Axis scanning. Each projection was integrated for 4.5 seconds. For reconstruction the standard Feldkamp–Davis–Kress (FDK) [8] algorithm was used. A visualization of the gray value images is shown in Figure 1. We use GeoDict [3] to perform the necessary image processing steps and segmenting the scans into fibers and background.

To analyze the fiber system within the nonwoven, the individual fibers are identified and their geometry is analyzed. This is done with the FiberFind module of GeoDict. The algorithm used in that module is described in [12]. As discussed there, this method provides good results for nonwoven structures, especially those that are used in hygiene products. It uses a neural network to identify the center line of each fiber in the segmented micro-CT scan. This neural network was pre-trained on artificial models of nonwoven structures to avoid manually labeling of training data and is included in GeoDict. As a neural network architecture, a 3D U-Net [6] is used. The center lines of the fibers labeled by the neural network are utilized to construct a graph representation of the complete fiber network. In the graph, endpoints are investigated to check if they are true fiber endpoints. If fiber fragments are found that are well aligned and the endpoints are close to each other the separation is considered an error and the fragments are reconnected. After these corrections, each connected component of the graph represents exactly one fiber. The connected components are then used to create analytic representations of the fibers, where each fiber is represented by a list of piece-wise linear segments and respective diameters. The diameters are determined by averaging the center-line values of an euclidean distance transform inside the fibers. Using these analytic representation, various properties of the fibers, including orientation, curvature, and length, can easily be computed.

2.3 Model for nonwoven materials

In the following, we develop a stochastic microstructure model for curved fiber systems. This model is based on representing fibers as polygonal tracks which are then modeled by a third-order Markov chain. We choose a specific approach to model the transition function of this Markov chain and select and fit the underlying parametric distribution families in a subsequent step. Finally, we validate the model by simulating artificial fiber systems and statistically comparing them to data of measured fiber systems.

2.3.1 Third-order Markov chain for modeling the trajectories of single fibers

To describe our model for entire fiber systems, we start by modeling single curved fibers. For this purpose, we adapt the stochastic 3D model previously proposed in [34]. We assume that fibers exhibit a circular cross-section with a constant radius $\delta > 0$ and assume that the center-lines of fibers can be modeled as finite cutouts of an infinite random polygonal track, i.e., as parts of a sequence of random vectors $\dots, P_{-2}, P_{-1}, P_0, P_1, P_2, \dots : \Omega \rightarrow \mathbb{R}^3$ such that $|P_{n+1} - P_n| = c$ for each integer $n \in \mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ for some constant segment length $c \in \mathbb{R}$, where $|\cdot|$ denotes the Euclidean norm. Choosing a constant segment length hugely simplifies the definition of our model and by appropriately choosing c , all fibers observed in measured data can be represented with reasonable accuracy. Moreover, we assume that the random sequence $\{P_n, n \in \mathbb{Z}\}$ satisfies the following three conditions.

- (i) $\{P_n, n \in \mathbb{Z}\}$ forms a third-order Markov chain [26], i.e., for each $n \in \mathbb{Z}$ and any finite subset $I \subset \mathbb{Z}$ with $\max\{i : i \in I\} < n - 3$, the random vector P_n is conditionally independent of the random vectors $\{P_i, i \in I\}$, under the condition that the values of $P_{n-3}, P_{n-2}, P_{n-1}$ are given. Note that this assumption is justified by the statistical analysis of measured and simulated image data performed later on in this paper.
- (ii) We furthermore assume that the Markov chain $\{P_n, n \in \mathbb{Z}\}$ is stationary, i.e., for any finite subset $I \subset \mathbb{Z}$ and for each $k \geq 1$, the distributions of the sequences $\{P_i, i \in I\}$ and $\{P_{i+k}, i \in I\}$ coincide. The latter assumption is motivated by the fact that from a statistical point of view the morphological properties of fibers do not change along their trajectories.
- (iii) Finally, we assume that the stationary Markov chain $\{P_n, n \in \mathbb{Z}\}$ is reversible, i.e., for each $k \geq 1$, the distributions of (P_0, P_1, \dots, P_k) and $(P_0, P_{-1}, \dots, P_{-k})$ coincide. This assumption is motivated by the circumstance that the statistical behavior of fiber trajectories is the same regardless of whether we traverse them forwards or backwards.

The conditions (i) - (iii) stated above imply that the distribution of the random sequence $\{P_n, n \in \mathbb{Z}\}$ is fully characterized by the joint distribution of the four random vectors P_1, P_2, P_3, P_4 . In the following we assume that this distribution possesses a probability density, which will be denoted by $f_{P_1, P_2, P_3, P_4} : \mathbb{R}^{4 \cdot 3} \rightarrow [0, \infty)$. Thus, modeling the random sequence $\{P_n, n \in \mathbb{Z}\}$ boils down to modeling the probability density f_{P_1, P_2, P_3, P_4} of (P_1, P_2, P_3, P_4) .

2.3.2 Isolating intrinsic fiber properties and external effects

Intuitively, one may be inclined to distinguish between intrinsic properties of the fibers and external effects, which may influence the morphological properties of fibers. More formally, we will model the probability density f_{P_1, P_2, P_3, P_4} by a product of three probability densities of independent random vectors which represent the position and orientation of fibers, intrinsic properties of fibers and external effects, respectively. For this purpose, we are going to denote the x -, y - and z -coordinates of P_n by X_n , Y_n and Z_n for each $n \in \mathbb{Z}$ and use a similar notation for a realization p_n of the random vector $P_n = (X_n, Y_n, Z_n)$, i.e., $p_n = (x_n, y_n, z_n) \in \mathbb{R}^3$.

As mentioned above, the model for f_{P_1, P_2, P_3, P_4} consists of three components. First we consider the location of P_1 and the orientation of the vector (P_1, P_2) , which can be expressed in terms the angle between $P_2 - P_1$ and the unit vector $(1, 0)$ (pointing along the x -axis), where, for now, we set aside the z -coordinates of P_1 and P_2 . This is, we model the joint density $f_{X_1, Y_1, A_0} : \mathbb{R}^2 \times [-\pi, \pi] \rightarrow [0, \infty)$ of X_1, Y_1 and $A_0 = \text{ang}((X_2, Y_2) - (X_1, Y_1))$, where $\text{ang}(p)$ denotes the angle between $p = (p_x, p_y) \in \mathbb{R}^2$ and the unit vector $(1, 0)$. We define $\text{ang}(p) < 0$ if p is below the x -axis, and $\text{ang}(p) \geq 0$ otherwise. Note that in computer science, the mapping $p \mapsto \text{ang}(p)$ is often referred to as the four-quadrant inverse tangent (denoted by atan2), i.e., $\text{ang}(p) = \text{atan2}(p_y, p_x)$. Note that, while the distribution of A_0 observed in measured image data is not exactly uniform, for simplicity of the model we assume A_0 to be uniformly distributed on the interval $(-\pi, \pi)$.

Then, we model the behavior of the fibers in z -direction, which captures the most important external effects for the data considered in the present paper. As the fiber systems are bounded in z -direction by some (lower and upper) parallel planes, fibers should not pierce through these planes, but, instead, be redirected towards the interior between the planes. We will model these external effects by appropriately modeling the z -coordinates Z_1, Z_2, Z_3, Z_4 of P_1, P_2, P_3, P_4 . In particular, we assume that the random sequence $\{Z_n, n \in \mathbb{Z}\}$ forms a stationary second-order Markov chain. This implies that modeling the joint density $f_{Z_1, Z_2, Z_3, Z_4} : \mathbb{R}^4 \rightarrow [0, \infty)$ of Z_1, Z_2, Z_3, Z_4 only requires modeling the joint density $f_{Z_1, Z_2, Z_3} : \mathbb{R}^3 \rightarrow [0, \infty)$ of (Z_1, Z_2, Z_3) , because it holds that

$$f_{Z_1, Z_2, Z_3, Z_4}(z_1, z_2, z_3, z_4) = f_{Z_1, Z_2}(z_1, z_2) f_{Z_3|Z_1=z_1, Z_2=z_2}(z_3) f_{Z_4|Z_1=z_2, Z_2=z_3}(z_4) \quad (1)$$

for any $z_1, z_2, z_3, z_4 \in \mathbb{R}$, where $f_{Z_3|Z_1=z_i, Z_2=z_{i+1}} : \mathbb{R} \rightarrow [0, \infty)$ is the conditional density of Z_3 under the condition that $Z_1 = z_i$ and $Z_2 = z_{i+1}$ for $i = 2, 3$ which is given by

$$f_{Z_3|Z_1=z_i, Z_2=z_{i+1}}(z_j) = \frac{f_{Z_1, Z_2, Z_3}(z_i, z_{i+1}, z_j)}{\int_{-\infty}^{\infty} f_{Z_1, Z_2, Z_3}(z_i, z_{i+1}, z) dz} \quad \text{for any } z_i, z_{i+1}, z_j \in \mathbb{R}.$$

It is clear that, instead of modeling f_{Z_1, Z_2, Z_3} , we may equivalently choose to model the joint density of $Z_1, Z_2 - Z_1$ and $Z_3 - Z_2$ which we will use in the following.

Finally, to capture intrinsic fiber properties, we consider the curvature of the fibers which is directly related to the angles between neighboring fiber segments $P_n - P_{n-1}$ and $P_{n+1} - P_n$. To account for features like persistence of curvature, we do not model these angles independently from each other, but by means of another stationary (first-order) Markov chain. More precisely, as we need to keep these angles independent of the z -coordinates Z_1, Z_2, Z_3, Z_4 of P_1, P_2, P_3, P_4 , we only model the angles between the projections of consecutive fiber segments onto the x - y -plane. Thus, we consider the angles $A = \mathcal{A}((X_3, Y_3) - (X_2, Y_2), (X_2, Y_2) - (X_1, Y_1))$ and $B = \mathcal{A}((X_4, Y_4) - (X_3, Y_3), (X_3, Y_3) - (X_2, Y_2))$. Here, $\mathcal{A}: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow [-\pi, \pi]$ denotes the (signed) angle between two vectors in the x - y -plane, i.e., $\mathcal{A}(s_1, s_2) = \text{ang}(s_2) - \text{ang}(s_1)$ for any $s_1, s_2 \in \mathbb{R}^2$, where $\text{ang}(s_i)$ is defined as described above. The joint density $f_{A,B}: [-\pi, \pi]^2 \rightarrow [0, \infty)$ of A and B will be the third building stone in our model for f_{P_1, P_2, P_3, P_4} .

In summary, we model the probability density f_{P_1, P_2, P_3, P_4} of P_1, \dots, P_4 as product of the probability densities f_{X_1, Y_1, A_0} , f_{Z_1, Z_2, Z_3, Z_4} and $f_{A, B}$, i.e., for any $p_1, p_2, p_3, p_4 \in \mathbb{R}^3$ with $|p_2 - p_1| = |p_3 - p_2| = |p_4 - p_3| = 1$ we put

$$\begin{aligned} f_{P_1, P_2, P_3, P_4}(p_1, p_2, p_3, p_4) &= f_{(X_1, Y_1), A_0}((x_1, y_1), \text{ang}((x_2, y_2) - (x_1, y_1))) \\ &\quad \times f_{Z_1, Z_2, Z_3, Z_4}(z_1, z_2, z_3, z_4) f_{A, B}(\mathcal{A}(s_2^{(x, y)}, s_1^{(x, y)}), \mathcal{A}(s_3^{(x, y)}, s_2^{(x, y)})), \end{aligned} \quad (2)$$

where $s_i^{(x, y)} = (x_{i+1}, y_{i+1}) - (x_i, y_i)$ for $i = 1, 2, 3$. Note that using Eq. (1), the second factor on the right-hand side of Eq. (2) can be written as product of uni- and bivariate probability densities. Having this in mind, the transition function $f_{P_4|P_1=p_1, P_2=p_2, P_3=p_3}: \mathbb{R}^3 \rightarrow [0, \infty)$ of the Markov chain $\{P_n, n \in \mathbb{Z}\}$ is given by

$$\begin{aligned} f_{P_4|P_1=p_1, \dots, P_3=p_3}(p_4) &= f_{Z_3 - Z_2|Z_1=z_2, Z_2 - Z_1=z_3 - z_2}(z_4 - z_3) \\ &\quad \cdot f_{B|A=\mathcal{A}(s_2^{(x, y)}, s_1^{(x, y)})}(\mathcal{A}(s_3^{(x, y)}, s_2^{(x, y)})) \end{aligned} \quad (3)$$

for any $p_1, p_2, p_3, p_4 \in \mathbb{R}^3$ with $|p_2 - p_1| = |p_3 - p_2| = |p_4 - p_3| = c$, where c is the constant segment length, $f_{B|A=\mathcal{A}(s_2^{(x, y)}, s_1^{(x, y)})}: [-\pi, \pi] \rightarrow [0, \infty)$ is the conditional density of B given that $A = \mathcal{A}(s_2^{(x, y)}, s_1^{(x, y)})$, and $f_{Z_3 - Z_2|Z_1=z_2, Z_2 - Z_1=z_3 - z_2}: [-1, 1] \rightarrow [0, \infty)$ is the conditional density of $Z_3 - Z_2$ given that $Z_1 = z_2$ and $Z_2 - Z_1 = z_3 - z_2$.

2.3.3 Selecting parametric families of probability distributions

For utilizing the approach to modeling the transition function $f_{P_4|P_1=p_1, P_2=p_2, P_3=p_3}$ given in Eq. (3), we have to decide on how to model the tri- and bivariate joint densities $f_{Z_1, Z_2 - Z_1, Z_3 - Z_2}$ and $f_{A, B}$ of $(Z_1, Z_2 - Z_1, Z_3 - Z_2)$ and (A, B) , respectively. For this, we separately incorporate knowledge about the measured image data for Sample 1 and Sample 2, respectively, into our model.

Let $N_F \geq 1$ denote the number of fibers $\{F_i, 1 \leq i \leq N_F\}$ with lengths $L_i = |F_i|$, which have been extracted from the corresponding tomographic image, each of which consists of $L_i + 1$ vertices $\{p_{i,k}, 1 \leq k \leq L_i + 1\}$ for $i \in \{1, \dots, N_F\}$. Then, we consider the set of all quadruples $Q_i = \{(p_{i,k}, p_{i,k+1}, p_{i,k+2}, p_{i,k+3}), 1 \leq k \leq L_i - 2\}$ of consecutive vertices along each fiber F_i . Furthermore, we take the union set $Q = \cup_{1 \leq i \leq N_F} Q_i$ of all quadruples and, for simplicity, rewrite their indices as $Q = \{(p_{1,j}, p_{2,j}, p_{3,j}, p_{4,j}), 1 \leq j \leq N\}$ with $N = \#(Q)$, where $\#(Q)$ denotes the cardinality of Q . Finally, we compute the corresponding angles $\{(a_j, b_j), 1 \leq j \leq N\}$ and z -coordinates $\{(z_{1,j}, z_{2,j}, z_{3,j}), 1 \leq j \leq N\}$. Based on this data, we can choose appropriate types of distributions for $f_{Z_1, Z_2 - Z_1, Z_3 - Z_2}$ and $f_{A, B}$. By examining the tomographic image data considered in this paper, it becomes obvious that in both cases, i.e. for $(Z_1, Z_2 - Z_1, Z_3 - Z_2)$ and (A, B) , the components of these random vectors can neither be considered as normally distributed, see Figure 3, nor independent, see Figure 4. Thus, we make use of copulas [17, 24] for modeling these distributions, which provide a flexible approach to parametric modeling of multivariate non-Gaussian distributions. In particular for modeling the joint distribution of $(Z_1, Z_2 - Z_1, Z_3 - Z_2)$, we employ an R-vine copula [7, 17].

Recall that due to the stationarity of the Markov chain $\{P_n, n \in \mathbb{Z}\}$, only three different marginal distributions need to be modeled. This is, we need to model the marginal distributions of A, Z_1 and $Z_2 - Z_1$. Based on two different samples of measured image data, we model all three univariate distributions (of A, Z_1 and $Z_2 - Z_1$) by generalized normal distributions [23]. Further details on this parametric family of distributions can be found in the appendix. As described there, we chose this family to get a common best fit for both samples. This means that for choosing, e.g., a family of distributions for A , we fit a wide variety of parametric families of univariate distributions to the values of a obtained for both samples, where we used tools provided by the python package SciPy [32] and, for each parametric family, compute the likelihood function as a measure for the goodness-of-fit. Then, we choose the distribution family for which the maximum of the likelihood functions among both datasets is largest, see Figure 3.

The dependence structure of the random vector (A, B) will be modelled by a bivariate copula. Similar to the procedure described above for choosing univariate marginal distributions, we fit a wide range of parametric copula families to the respective data from both measured samples using the python package pyvinecopulib [31]. Then, we choose the copula which maximizes the goodness-of-fit among both samples with respect to the likelihood function. Thereby, we use the Student's t-copula to model the dependence structure of (A, B) , see Figure 4.

For modeling the dependence structure of $(Z_1, Z_2 - Z_1, Z_3 - Z_2)$, we employ an R-vine copula as described in the appendix. For this, we need to model three bivariate copulas, namely for the distributions of $(Z_3 - Z_2, Z_1)$ and $(Z_2 - Z_1, Z_1)$ as well as for the conditional distribution of $(Z_3 - Z_2, Z_2 - Z_1)$ conditioning on Z_1 . Note that, as described in the appendix, this choice of an R-vine copula allows for the simulation of $Z_3 - Z_2$, under the condition that the values of Z_1 and $Z_2 - Z_1$ are given. Using the

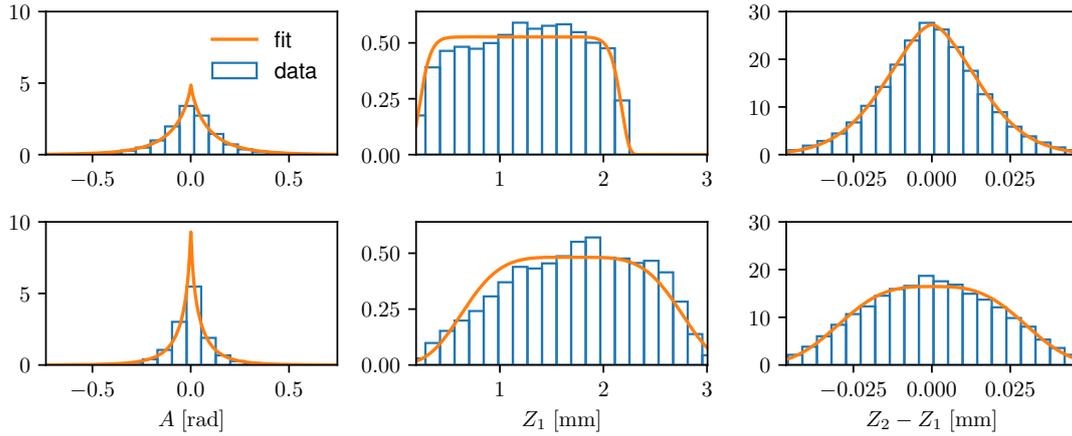


Figure 3: Empirical and fitted distributions of A , Z_1 and $Z_2 - Z_1$ for Sample 1 (top) and Sample 2 (bottom). Note that B adheres to the same distribution as A , and $Z_3 - Z_2$ has the same distribution as $Z_2 - Z_1$.

same approach as before, we choose to model the dependence structures of $(Z_3 - Z_2, Z_1)$ and $(Z_2 - Z_1, Z_1)$ by Clayton copulas and that of $(Z_3 - Z_2, Z_2 - Z_1)$, conditional on Z_1 , by a Student's t-copula.

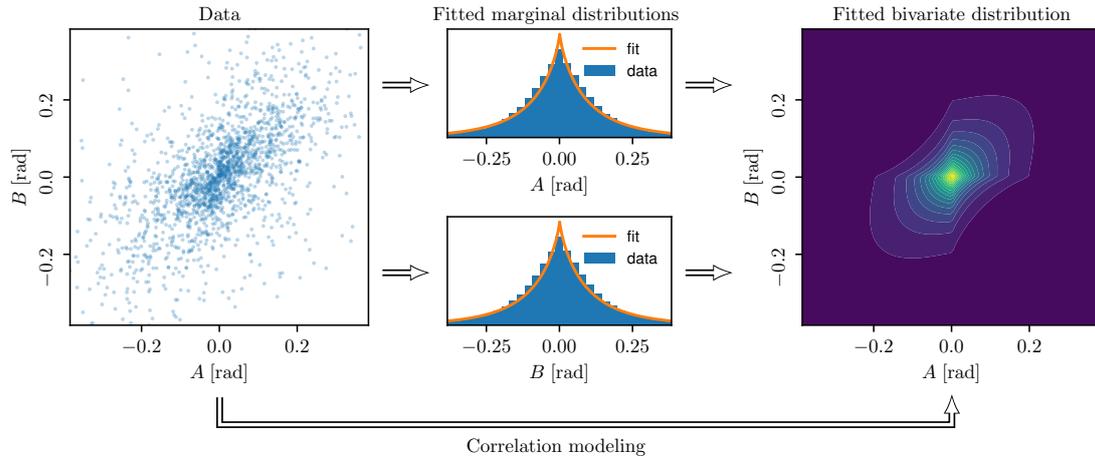


Figure 4: Left: Empirical distribution of (A, B) for Sample 1, shown as a scatter plot of measured pairs of (A, B) . Center: Fitting generalized normal distributions to the marginal distributions of A and B , respectively. Right: Modeling the correlation structure of A and B using a Student's t copula.

2.3.4 Fitting model parameters

After having selected a parametric family of univariate probability distributions for the (real-valued) random variables A , Z_1 , $Z_2 - Z_1$ and parametric copula families for the two-dimensional random vectors (A, B) , $(Z_3 - Z_2, Z_1)$ and $(Z_2 - Z_1, Z_1)$ as well as for the conditional distribution of $(Z_3 - Z_2, Z_2 - Z_1)$ conditioning on Z_1 , we fit the parameters of these univariate distributions and copulas, again using the data $\{(a_j, b_j), 1 \leq j \leq N\}$ and $\{(z_{1,j}, z_{2,j}, z_{3,j}), 1 \leq j \leq N\}$ described above.

Note that fitting the parameters of the univariate distributions is performed separately for Sample 1 and Sample 2, respectively, by maximum-likelihood estimation provided by SciPy. Figure 3 shows the fitted distributions and the underlying measured data. The parameters of the fitted univariate distributions are given in Table 1.

	Distribution	Sample	Parameters	Location	Scale	Mean	Var
A [rad]	gen. normal	1	0.81	0	0.09	0	0.04
		2	0.68	0	0.04	0	0.02
Z_1 [mm]	gen. normal	1	14.71	1.20	0.98	1.20	0.31
		2	4.38	1.69	1.14	1.69	0.43
$Z_2 - Z_1$ [mm]	gen. normal	1	1.69	0	0.02	0	0
		2	3.12	0	0.03	0	0

Table 1: Parameters of the fitted univariate distributions. Note that due to symmetry, we would expect a mean (i.e., location) of 0 for the distributions of A and $Z_2 - Z_1$, which is in fact (almost) the case for the fitted values. The distribution of Z_1 essentially dictates the extent of the structures in z -direction, where the mean values corresponds to roughly half the thickness of the samples. Note that B adheres to the same distribution as A , and $Z_3 - Z_2$ has the same distribution as $Z_2 - Z_1$. The units given in the left columns apply also to the location and scale parameters.

Analogously, using pyvinecopulib, the copula parameters are fitted separately for Sample 1 and Sample 2, respectively, based on maximum-likelihood estimation, see Table 2.

2.3.5 Simulating the entire fiber system

For simulating the entire fiber system, we fix a bounding cuboid $W = [0, x_{\max}] \times [0, y_{\max}] \times [0, z_{\max}]$ for some $x_{\max}, y_{\max}, z_{\max} > 0$. While we may choose x_{\max} and y_{\max} arbitrarily large, z_{\max} directly relates to the distribution of Z_1 which should be theoretically limited to a bounded interval. In practice, however, the best fitting distribution is unbounded and thus, we may choose, e.g., the 99.9%-quantile of this distribution as

	Copula	Sample	Parameters	Rotation
(A, B)	Student's t	1	(0.66, 3.15)	0
		2	(0.82, 3.11)	0
$(Z_3 - Z_2, Z_1)$	Clayton	1	0.05	90
		2	0.06	90
$(Z_2 - Z_1, Z_1)$	Clayton	1	0.02	90
		2	0.03	90
$(Z_3 - Z_2, Z_2 - Z_1) Z_1$	Student's t	1	(0.93, 2.05)	0
		2	(0.99, 5.00)	0

Table 2: Parameters of the fitted copulas. The results stated in the last three rows are used to build a trivariate R-vine copula for $(Z_1, Z_2 - Z_1, Z_3 - Z_2)$. Parameters for Student's t-copulas are given as (ρ, ν) .

z_{\max} . As a further parameter, we choose the total length $L > 0$ of the fiber system within W .

Note that we do not explicitly model the lengths of single fibers and assume that all fibers span the whole sampling window W which will be guaranteed by the following simulation procedure. This assumption seems largely justified for the considered nonwovens.

To avoid edge-effects, for some $r > 0$ we repeatedly simulate fibers in a larger window $\widetilde{W} = [0 - r, x_{\max} + r] \times [0 - r, y_{\max} + r] \times [0, z_{\max}]$. This means, we start at some initial vertices p_1, p_2, p_3 and simulate $\{p_i, i = 1, 2, \dots\}$ according to the Markov chain approach described above until the fiber leaves \widetilde{W} . Then, we consider those parts of the simulated fiber which lie within the original bounding window W , i.e., $\{p_{k_j-1}, \dots, p_{l_j+1}\}, j \geq 0$, where $1 < k_j < l_j$, such that $0 \leq x_i \leq x_{\max}$ and $0 \leq y_i \leq y_{\max}$ for all $i \in \{k_j, \dots, l_j\}$ and $x_i \notin [0, x_{\max}]$ or $y_i \notin [0, y_{\max}]$ for both $i = k_j - 1$ and $i = l_j + 1$. These parts then only start and end at the boundary of W . We stop simulation when the total length of these parts reaches the value of L . Note that we did not extend W in z -direction as the Markov chain $\{Z_n, n \in \mathbb{Z}\}$, which controls the z -coordinates, ensures that z_i belongs to the interval $[0, z_{\max}]$ for all $i = 1, 2, \dots$

We choose the initial vertices p_1, p_2, p_3 by simulating the z -coordinates z_1, z_2, z_3 of p_1, p_2, p_3 using the (known) joint distribution of $Z_1, Z_2 - Z_1, Z_3 - Z_2$. Furthermore, for the first point $p_1 = (x_1, y_1, z_1)$, we may simply choose (x_1, y_1) uniformly distributed in $[0, x_{\max}] \times [0, y_{\max}]$. Then, to generate (x_2, y_2) , we simulate the angle between $(x_2, y_2) - (x_1, y_1)$ and the x -axis by drawing a sample from the uniform distribution on the interval $[-\pi, \pi]$. Finally, to generate (x_3, y_3) , we simulate the angle between $(x_3, y_3) - (x_2, y_2)$ and $(x_2, y_2) - (x_1, y_1)$ which is drawn from the known distribution of A . Together with the already fixed values of z_1, z_2, z_3 , we can then easily compute the actual values of $p_2 = (x_2, y_2, z_2)$ and $p_3 = (x_3, y_3, z_3)$.

For drawing a point p_{n+1} conditional on p_{n-2}, p_{n-1}, p_n , we employ the transition function given in Eq. (3). For this, we independently simulate (x_{n+1}, y_{n+1}) and z_{n+1} using the fitted distributions described above. Calculating the angle a between $(x_n, y_n) - (x_{n-1}, y_{n-1})$ and $(x_{n-1}, y_{n-1}) - (x_{n-2}, y_{n-2})$ and then drawing from B conditional on $A = a$ using the probability density $f_{B|A=a}$ results in the angle between $(x_{n+1}, y_{n+1}) - (x_n, y_n)$ and $(x_n, y_n) - (x_{n-1}, y_{n-1})$. Furthermore, we draw the value of $z_{n+1} - z_n$ from the probability density $f_{Z_3-Z_2|Z_1=z_{n-1}, Z_2-Z_1=z_n-z_{n-1}}$ of $Z_3 - Z_2$ conditional on $Z_1 = z_{n-1}$ and $Z_2 - Z_1 = z_n - z_{n-1}$, and then set z_{n+1} accordingly. If z_{n+1} would be outside of the interval $[0, z_{max}]$, we reject it and draw a new value from the conditional distribution of $Z_3 - Z_2$.

Fitting and simulation of the copula-based model described above were implemented in python using, among others, the packages SciPy [32] and a slightly adapted version of pyvinecopulib [31].

2.4 Validation measures

Various geometric descriptors of the 3D microstructure of porous media play an important role for transport within the pore space. These descriptors are thus especially well suited to compare the goodness-of-fit between measured image data and model realizations. Assume that we observe a porous material (in our case, a fiber system) in 3D within a bounding cuboid $W = [0, x_{max}] \times [0, y_{max}] \times [0, z_{max}]$ and with some transport direction. In the data at hand, we take the z -axis as the transport direction, i.e., we assume that transport through the nonwoven (as opposed to within) is of interest. For reference, we define a “starting plane”, e.g., at $z = 0$ where transport starts and an opposing “end plane”, e.g., at $z = z_{max}$ where transport ends. In the following, we give a short overview over the geometric descriptors considered in the present paper. For more details and further applications, see e.g. [25].

The geodesic tortuosity is a measure for windedness of transport pathways (within the pore space). For some point $p = (x, y, z) \in W$ with $z = 0$ on the starting plane which is contained in the pore space (i.e., not on a fiber), we compute the distance l of the shortest path from p to the end plane through the pore space using Dijkstra’s algorithm on the voxel grid. By definition, it holds that $l \geq d$, where d denotes the distance between starting and end planes, and we call the fraction l/d the tortuosity of the shortest path starting at p . The distribution of these values for randomly selected points p is considered as the tortuosity distribution.

The so-called constrictivity is a measure for the strength of bottleneck effects along transport pathways. It is given as a fraction $\frac{r_{min}}{r_{max}} \in [0, 1]$. Here, r_{max} is the largest radius such that at least 50% of pore space can be covered by (potentially overlapping) spheres of radius r_{max} . Furthermore, r_{min} is the largest radius such that at least 50% of

pore space can be covered by spheres of radius r_{min} intruding the pore space from the starting plane. This is, a point $p \in W$ inside the pore space is covered by such spheres if there exists a point p_c and a path from the starting plane to p_c such that $|p_c - p| \leq r_{min}$ and the distance from any point on said path to the solid phase (i.e., the fibers) is larger than r_{min} .

The spherical contact distance for a given point $p \in W$ within the pore space is the shortest distance from p to any point within the solid phase (i.e., the fibers). We consider the distribution of these shortest distances for randomly selected points p .

3 Results

Recall that we fitted the copula-based model, which we consider in this paper, to two different measured datasets of nonwoven structures, and we simulated artificial fiber systems using the fitted models. For comparability, we chose the simulation window, the total length of fibers and fiber diameter to match the values of the measured structures given in the Materials section. Figure 2 illustrates the measured fiber systems along with the simulated structures. Visual comparison of these fiber systems shows a good agreement of the overall morphological properties of measured and simulated fiber systems. In addition, we quantitatively evaluate the agreement between measured and simulated image data, where, among others, we use geometric descriptors of the pore space morphology as explained in the Methods section. In the following, the term "measured" refers to the experimentally measured nonwovens while "simulated" describes the structures obtained from the proposed model. All properties presented in the following are calculated from both types of data in the same way.

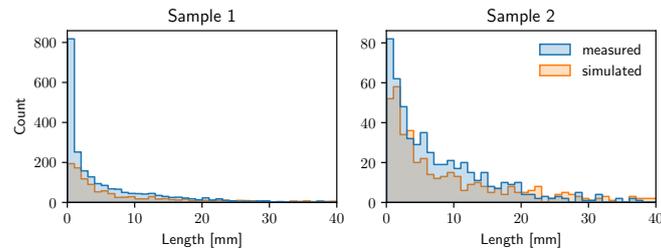


Figure 5: Distribution of fiber lengths, computed from measured image data (blue) and model-based simulations (orange) of Sample 1 (left) and Sample 2 (right).

Considering the bounding window of each of these four datasets, we can compute the specific fiber length, i.e. the total fiber length per unit volume and the porosity of the structure, as shown in Table 3. While these descriptors are closely linked with each other, porosity may not necessarily decrease with increasing number of fibers when fibers are

overlapping. Note that, while overlapping fibers are not present in the measured data, we model fibers independently of each other. Thus, we cannot guarantee non-overlapping fibers in the simulations. If non-overlapping fibers are necessary, a suitable displacement algorithm may be applied to the simulated structures. Comparing measured and simulated systems with respect to these geometric descriptors shows a good agreement. Note that the specific fiber lengths of measured and simulated systems are almost identical, which would be expected by definition of the model. The number of fibers is closely linked to the fiber length distribution as shown in Figure 5. While both the number of fibers and the fiber length distribution match closely between measured and simulated data for Sample 2, fibers tend to be significantly longer in the simulated data when considering Sample 1 which is also reflected in the lower number of simulated fibers, see Table 3. This is probably due to the fact that the measured data of Sample 1 has almost four times the specific fiber length compared to Sample 2 which makes image segmentation challenging and may lead to oversegmentation, i.e., splitting fibers into shorter fragments. Recall that we did not explicitly model the lengths of single fibers and assumed that each fiber spans the whole sampling window.

Sample	Data	Specific fiber length [cm ⁻²]	Number of fibers	Porosity
1	measured	4165	2189	0.987
1	simulated	4165	1308	0.987
2	measured	1081	530	0.997
2	simulated	1084	481	0.997

Table 3: Specific fiber length, number of fibers, and porosity, computed from measured image data and model-based simulations of Samples 1 and 2.

Moreover, we computed the solid volume fraction of these structures for slices parallel to the x - y -plane at different heights $z \in [0, z_{\max}]$. Figure 6 shows the results which we obtained for measured and simulated image data. Especially for Sample 2, they seem to be in good agreement with each other.

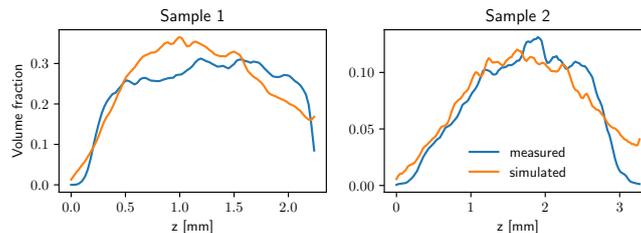


Figure 6: Solid volume fraction for slices at given z -position, computed from measured image data (blue) and model-based simulations (orange) of Sample 1 (left) and Sample 2 (right).

For further validation, we estimated the distribution of the spherical contact distance d to the fiber system from a randomly selected location within the pore space, see the first column of Figure 7. For both Sample 1 and Sample 2, the obtained results show a good agreement between simulated and measured structures. Furthermore, we computed the distribution of geodesic tortuosity τ of pore space which is a descriptor of the windedness of pathways, see the second column of Figure 7. Finally, for assessing local heterogeneity, we divided each structure into $20 \times 20 \times 15$ cutouts and computed the mean geodesic tortuosity $\mu(\tau)$ as well as the so-called constrictivity β of the pore space on each cutout separately. The third and fourth columns of Figure 7 show histograms of the values which we obtained for these pore space descriptors. It turned out that the overall agreement between all these distributions is quite good.

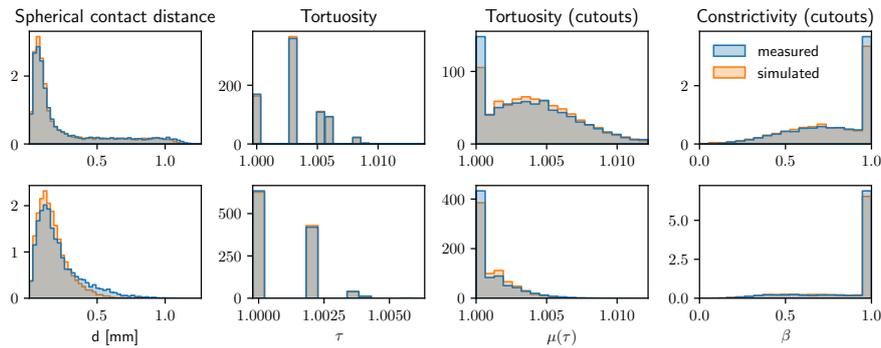


Figure 7: Distribution of geometric descriptors of the pore space morphology, computed from measured image data (blue) and model-based simulations (orange) of Sample 1 (top) and Sample 2 (bottom). Note that as tortuosity is measured on a voxel grid, only finitely many different values of tortuosity can exist which causes the void spaces between the bins in the distributions of tortuosity.

4 Discussion

In this paper we laid the foundation for data-driven 3D analysis of virtual fiber systems with respect to their microstructure and functionality. In particular, we developed a stochastic 3D model for systems of curved fibers similar to nonwovens, which is fitted to tomographic image data. By systematic variations of model parameters, efficient computer-based scenario analyses can be performed to get a deeper insight how effective properties of this type of functional materials depend on their 3D microstructure.

The modeling approach was based on representing fibers as polygonal tracks which were then modeled by means of third-order Markov chains. By choosing the transition function of these Markov chains to be the product of two (conditional) probability densities,

we managed to isolate the influence of intrinsic fiber properties from that of external effects. Modeling these two influencing factors separately yielded comprehensible (conditional) probability densities which can be directly related to different aspects of the geometry of single fibers as well as fiber systems. For a precise description of the underlying multivariate distributions, we chose copulas which minimized the required number of parameters. By construction, the model could be easily adapted to various types of structures which we exemplified by fitting the model to two different measured datasets of nonwovens. Artificial structures simulated by these fitted models resembled the measured datasets with respect to various geometric descriptors of the pore space morphology which are related to macroscopic physical properties of the underlying material and were not used for model fitting. Moreover, the presented model is robust with respect to oversegmentation of the data used for fitting.

As fitting the model and drawing realizations can be performed efficiently with relatively low computational costs, the model is well suited for the development of a framework for virtual materials testing and further investigation of the relationship between microstructure geometry and effective macroscopic properties. This will be investigated in a forthcoming paper [35].

5 Utilized numerical tools

We used the python programming language to implement the model, simulate the structures, compute geometric descriptors and prepare most of the figures for this manuscript. Notably, the following libraries were used: NumPy [13], SciPy [32], a slightly modified version of pyvinecopulib [31], Numba [20] for accelerated execution and Matplotlib [16] and seaborn [33] for creating most of the the figures.

GeoDict [3] was used to numerically compute the permeability and for 3D rendering of selected structures.

6 Data availability

The data generated and analyzed during this study, i.e., the center-lines used to fit the model and the center-lines of the simulated structures are available from the corresponding author upon reasonable request.

7 Code availability

All formulations and algorithms necessary to reproduce the results of this study are described in the “Results” and “Methods” sections.

8 Acknowledgements

We thank Adrien Grenier, Dirk Saevecke, Tom Dufresne, and The Procter & Gamble Company for assisting with this research.

9 Author contributions

Extraction of fibers from 3D CT images was performed by A.G. and D.M. The model was implemented and fitted by M.W. The initial draft of the paper was written by M.W. All authors discussed the results and contributed to writing the manuscript. A.W., E.G. and V.S. designed and supervised the research.

10 Competing interests

The authors declare no competing interests.

References

- [1] S. Abishek, A. King, R. Mead-Hunter, V. Golkarfard, W. Heikamp, and B. Mullins. Generation and validation of virtual nonwoven, foam and knitted filter (separator/coalescer) geometries for CFD simulations. *Separation and Purification Technology*, 188:493–507, 2017.
- [2] M. Azimian, C. Kühnle, and A. Wiegmann. Design and optimization of fibrous filter media using lifetime multipass simulations. *Chemical Engineering & Technology*, 41(5):928–935, 2018.

- [3] J. Becker, F. Biebl, M. Boettcher, L. Cheng, F. Frank, E. Glatt, A. Grießer, S. Linden, D. Mosbach, A. Neundorf, C. Wagner, A. Weber, R. Westerteiger, and A. Wiegmann. GeoDict Software. <https://doi.org/10.30423/release.geodict2023>, 2023.
- [4] J. P. Chiverton, O. Ige, S. J. Barnett, and T. Parry. Multiscale shannon’s entropy modeling of orientation and distance in steel fiber micro-tomography data. *IEEE Transactions on Image Processing*, 26(11):5284–5297, 2017.
- [5] J. P. Chiverton, A. Kao, M. Roldo, and G. Tozzi. Volumetric simulation of nanofibres and 2D SEM and 3D XCT imaging processes. In *Medical Image Understanding and Analysis: 24th Annual Conference, MIUA 2020, Oxford, UK, July 15-17, 2020, Proceedings 24*, pages 436–445. Springer, 2020.
- [6] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, Cham, 2016. Springer International Publishing.
- [7] C. Czado. *Analyzing Dependent Data with Vine Copulas*. Springer, 2019.
- [8] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of The Optical Society of America A-optics Image Science and Vision*, 1:612–619, 1984.
- [9] G. Gaiselmann, D. Froning, C. Tötze, C. Quick, I. Manke, W. Lehnert, and V. Schmidt. Stochastic 3D modeling of non-woven materials with wet-proofing agent. *International Journal of Hydrogen Energy*, 38(20):8448–8460, 2013.
- [10] C. Geerling, M. Azimian, A. Wiegmann, H. Briesen, and M. Kuhn. Designing optimally-graded depth filter media using a novel multiscale method. *AIChE Journal*, 66(2):e16808, 2020.
- [11] A. Grießer, R. Westerteiger, C. Wagner, and A. Wiegmann. FiberFind: Machine learning-based segmentation and identification of individual fibers in μ CT images of fibrous media. In *International Conference on Tomography of Materials & Structures 2019, Cairns, Australia*, 2019.
- [12] A. Grießer, R. Westerteiger, E. Glatt, H. Hagen, and A. Wiegmann. Identification and analysis of fibers in ultra-large X-ray scans of nonwoven textiles using deep learning. *Journal of the Textile Institute*, published online (<https://doi.org/10.1080/00405000.2022.2145429>), 2022.
- [13] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cour-

- napeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020.
- [14] H. Herrmann, E. Pastorelli, A. Kallonen, and J.-P. Suuronen. Methods for fibre orientation analysis of X-ray tomography images of steel fibre reinforced concrete (SFRC). *Journal of Materials Science*, 51:3772–3783, 2016.
- [15] X. Huang, Q. Zhou, J. Liu, Y. Zhao, W. Zhou, and D. Deng. 3D stochastic modeling, simulation and analysis of effective thermal conductivity in fibrous media. *Powder Technology*, 320:397–404, 2017.
- [16] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [17] H. Joe. *Dependence Modeling with Copulas*. Chapman and Hall/CRC, 2014.
- [18] H. Kallel and K. Joulain. Design and thermal conductivity of 3D artificial cross-linked random fiber networks. *Materials & Design*, 220:110800, 2022.
- [19] J. Kroutilova, M. Maas, Z. Mecl, T. Wagner, F. Klaska, and P. Kasarkova. Bulky nonwoven fabric with enhanced compressibility and recovery, 2020. Patent WO2020/103964.
- [20] S. K. Lam, A. Pitrou, and S. Seibert. Numba: A LLVM-based python JIT compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM '15*, New York, NY, USA, 2015. Association for Computing Machinery.
- [21] N. Mao, S. J. Russell, and B. Pourdeyhimi. Characterisation, testing and modelling of nonwoven fabrics. In S. J. Russell, editor, *Handbook of Nonwovens*, pages 401–514. Elsevier, 2007.
- [22] A. Moghadam, S. H. Yousefi, H. V. Tafreshi, and B. Pourdeyhimi. Characterizing nonwoven materials via realistic microstructural modeling. *Separation and Purification Technology*, 211:602–609, 2019.
- [23] M. Nardon and P. Pianca. Simulation techniques for generalized Gaussian densities. *Journal of Statistical Computation and Simulation*, 79(11):1317–1329, 2009.
- [24] R. B. Nelsen. *An Introduction to Copulas*. Springer, 2006.
- [25] B. Prifling, M. Weber, N. Ray, A. Prechtel, M. Phalempin, S. Schlüter, D. Vet-

- terlein, and V. Schmidt. Quantifying the impact of 3D pore space morphology on soil gas diffusion in loam and sand. *Transport in Porous Media*, published online: <https://doi.org/10.1007/s11242-023-01971-z>, 2023.
- [26] A. E. Raftery. A model for high-order Markov chains. *Journal of the Royal Statistical Society: Series B (Methodological)*, 47(3):528–539, 1985.
- [27] M. Schneider. The sequential addition and migration method to generate representative volume elements for the homogenization of short fiber reinforced plastics. *Computational Mechanics*, 59(2):247–263, 2017.
- [28] V. P. Schulz, J. Becker, A. Wiegmann, P. P. Mukherjee, and C.-Y. Wang. Modeling of two-phase behavior in the gas diffusion medium of PEFCs via full morphology approach. *Journal of the Electrochemical Society*, 154(4):B419, 2007.
- [29] P. Soltani, M. Zarrebini, R. Laghaei, and A. Hassanpour. Prediction of permeability of realistic and virtual layered nonwovens using combined application of X-ray μ CT and computer simulation. *Chemical Engineering Research and Design*, 124:299–312, 2017.
- [30] D. Venkateshan, M. Tahir, H. Vahedi Tafreshi, and B. Pourdeyhimi. Modeling effects of fiber rigidity on thickness and porosity of virtual electrospun mats. *Materials & Design*, 96:27–35, 2016.
- [31] Vinecopulib. Vinecopulib/pyvinecopulib: A python library for vine copula models. <https://github.com/vinecopulib/pyvinecopulib>, 2023. Accessed: 2023-04-20.
- [32] 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*, 17:261–272, 2020.
- [33] M. L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.
- [34] M. Weber, A. Griebner, E. Glatt, A. Wiegmann, and V. Schmidt. Modeling curved fibers by fitting R-vine copulas to their Frenet representations. *Microscopy and Microanalysis*, 29(1):155–165, 2023.
- [35] M. Weber, B. Prifling, A. Griebner, E. Glatt, A. Wiegmann, and V. Schmidt. Inves-

tigating microstructure-property relationships of nonwovens by model-based virtual materials testing. 2023. Working paper (under preparation).

- [36] A. Wiegmann. Effective properties of nonwoven textiles from microstructure simulations. In L. L. Bonilla, M. Moscoso, G. Platero, and J. M. Vega, editors, *Progress in Industrial Mathematics at ECMI 2006*, pages 708–712. Springer, 2008.
- [37] A. Wiegmann, S. Rief, and A. Latz. Computer models of nonwoven geometry and filtration simulation. *Filtration News*, 26(4):65–69, 2007.
- [38] N. Zamel, X. Li, J. Shen, J. Becker, and A. Wiegmann. Estimating effective thermal conductivity in carbon paper diffusion media. *Chemical Engineering Science*, 65(13):3994–4006, 2010.

Supplementary Information

Copula-based modeling and simulation of 3D systems of curved fibers by isolating intrinsic fiber properties and external effects

Matthias Weber,^{1,*} Andreas Grießer,² Dennis Mosbach,² Erik Glatt,²
Andreas Wiegmann,² and Volker Schmidt¹

¹ Institute of Stochastics, Ulm University, Helmholtzstraße 18, 89069 Ulm, Germany

² Math2Market GmbH, Richard-Wagner-Straße 1, D-67655, Kaiserslautern, Germany

* corresponding author, matthias.weber@uni-ulm.de

A. Appendix

A.1. Python package for fitting univariate probability distributions

In the context of the present paper, we need to model various univariate data by means of parametric families of probability distributions. The python package SciPy [12] provides tools for using—and especially for fitting—over 100 different distribution families. We can directly exploit this package to fit a specific parametric probability distribution to a given set of observations. Moreover, to choose the best suited distribution for the dataset, we fit all distribution families available in SciPy to the data, where we use the likelihood function as a measure for the goodness-of-fit for each fitted distribution. Based on these values, we choose the best fitting distribution for the given dataset, i.e., we choose the distribution family for which the maximum of the likelihood functions is largest. By this approach, we decide which parametric families we use to model the marginal distributions of the random vectors comprised in the copula-based model considered in the present paper.

However, note that we base this decision on two datasets. While we may not assume that a property follows the same type of distribution for both datasets, we would like our model to be able to compare the different morphologies of both datasets with each other. Thus, we would like to find a distribution family which (using different parameters) captures the empirical distribution of the considered property for both datasets.

It turned out that for all univariate data considered in this paper, the generalized normal distribution [7] provides the best fit. This is a parametric family of univariate distributions with parameters $\mu \in \mathbb{R}, \alpha, \beta > 0$, whose density $f_{\mu, \alpha, \beta} : \mathbb{R} \rightarrow [0, \infty)$ is given by

$$f_{\mu, \alpha, \beta}(x) = \frac{\alpha, \beta}{2\alpha\Gamma(1/\beta)} \exp\left(-(|x - \mu|/\alpha)^\beta\right), \quad \text{for all } x \in \mathbb{R},$$

where μ , α and β are known as location, scale and shape parameters, respectively, and $\Gamma : [0, \infty) \rightarrow [0, \infty)$ denote the gamma function. For our purposes, we need to determine the distribution function and quantile function of the generalized normal distribution which are computed numerically using the python package SciPy [12].

A.2. Sklar's representation formula for multivariate distributions

For modeling multivariate probability distributions, different approaches exist. While some univariate distribution families like normal, Student's t [4], or von Mises [8] distributions have multivariate counterparts with parametric probability densities, parametrically describing the multivariate distribution of arbitrary random vectors (X_1, \dots, X_d) for any given integer $d \geq 2$ requires more advanced theory. This is indeed the case for the multivariate distributions considered in the present paper. For modeling such multivariate probability distributions, so-called copulas can be exploited [6, 9]. Using copulas, the univariate marginal distributions of a random vector can be chosen separately and fitted directly to the data. Then, in a second step, the dependence structure of the data can be modeled by means of a copula, which turns the marginal distributions into the desired multivariate distribution.

A d -variate copula is a d -variate (cumulative) distribution function $C : [0, 1]^d \rightarrow [0, 1]$ with uniform marginal distributions on the unit interval $[0, 1]$. According to Sklar's representation formula [6, 9], for any d -variate distribution function $F : \mathbb{R}^d \rightarrow [0, 1]$ with marginal distribution functions $F_1, \dots, F_d : \mathbb{R} \rightarrow [0, 1]$ a d -variate copula $C : [0, 1]^d \rightarrow [0, 1]$ exists such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad \text{for all } x_1, \dots, x_d \in \mathbb{R}. \quad (1)$$

Note that in the case when the joint distribution function $F : [0, 1]^d \rightarrow [0, 1]$ of a certain random vector (X_1, \dots, X_d) has a probability density $f : \mathbb{R}^d \rightarrow [0, \infty)$, this leads to an equivalent version of Eq. (1) in terms of the probability density $c : [0, 1]^d \rightarrow [0, \infty)$

corresponding to C . Namely, then it holds that

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) f_1(x_1) \dots f_d(x_d) \quad \text{for all } x_1, \dots, x_d \in \mathbb{R}, \quad (2)$$

where $f_i : \mathbb{R} \rightarrow [0, \infty)$ is the density corresponding to F_i for $i \in \{1, \dots, d\}$.

Furthermore, the transformed components $F_1(X_1), \dots, F_d(X_d)$ of the random vector (X_1, \dots, X_d) are then uniformly distributed on $[0, 1]$. Thus, after fitting the marginal density $f_i : \mathbb{R} \rightarrow [0, \infty)$ to a dataset $(x_i^{(1)}, \dots, x_i^{(N)}) \in \mathbb{R}^N$ for each $i = 1, \dots, d$ as described above, and after computing the cumulative distribution functions F_1, \dots, F_d , which is typically done by numerical integration of f_1, \dots, f_d , respectively, the transformed data $(F_1(x_1^{(j)}), \dots, F_d(x_d^{(j)})), 1 \leq j \leq N$, can be used to fit a parametric probability density $c : [0, 1]^d \rightarrow [0, \infty)$.

Various parametric models for copulas are considered in the literature [6, 9], many of which are so-called Archimedean copulas $C : [0, 1]^d \rightarrow [0, 1]$ with

$$C(u_1, \dots, u_d) = \psi^{-1}(\psi(u_1) + \dots + \psi(u_d)) \quad \text{for all } u_1, \dots, u_d \in [0, 1],$$

where the so-called generator function $\psi : [0, 1] \rightarrow [0, \infty)$ is a continuous, strictly decreasing convex function such that $\psi(1) = 0$ and ψ^{-1} is the pseudo-inverse of ψ . Examples of Archimedean copulas which we use in the present paper include the Clayton copula with generator function $\psi_\theta : [0, 1] \rightarrow (0, \infty)$ given by

$$\psi_\theta(u) = \frac{u^{-\theta} - 1}{\theta} \quad \text{for all } u \in [0, 1]$$

and some parameter $\theta \in [-1, \infty) \setminus \{0\}$.

Furthermore, some copulas correspond to known multivariate probability distributions. This is due to the fact that for any d -variate cumulative distribution function $F : \mathbb{R}^d \rightarrow [0, 1]$ as considered above, it can be easily shown that by the function $C : [0, 1]^d \rightarrow [0, 1]$ with

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad \text{for all } u_1, \dots, u_d \in [0, 1]$$

a d -variate copula is given which can be used to model F . Commonly used copulas derived in this way include the Student's t-copula which is based on the multivariate Student's t-distribution. In the 2-dimensional case, this distribution is defined by its probability density $f_{\nu, \mu, \Sigma} : \mathbb{R}^2 \rightarrow [0, \infty)$ function with

$$f_{\nu, \mu, \Sigma}(x) = \frac{\Gamma[(\nu + 2)/2]}{\Gamma(\nu/2)\nu\pi |\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(x - \mu)^\top \Sigma^{-1}(x - \mu) \right]^{-(\nu+2)/2} \quad \text{for all } x \in \mathbb{R}^2,$$

for some degree of freedom $\nu > 0$, location vector $\mu \in \mathbb{R}^2$ and positive-definite scale matrix $\Sigma \in \mathbb{R}^{2 \times 2}$. Note that for the definition of the copula, μ and the diagonal elements of Σ are irrelevant and can be set to 0 and 1, respectively. This leaves the bivariate Student's t-copula with two parameters: $\nu > 0$ for the degree of freedom, and $\rho > 0$ for the non-diagonal entries of Σ of the underlying bivariate t-distribution.

A.3. R-vine copulas

While especially Archimedean copulas can easily be used for any dimension $d \geq 2$, many multivariate distributions in higher dimensions can not be expressed by means of Archimedean copulas. Then, so-called R-vine copulas provide a flexible approach which can be used for modeling arbitrary multivariate distributions. A general theory and some applications of R-vine copulas can be found, e.g., in [1, 2, 3, 6]. In the present paper, we use an approach similar to R-vine copulas to model the trivariate probability densities which have been used in the model.

Let $X = (X_1, X_2, X_3)$ be a three-dimensional random vector with joint probability density $f: \mathbb{R}^3 \rightarrow [0, \infty)$, marginal densities $f_1, f_2, f_3: \mathbb{R} \rightarrow [0, \infty)$ and corresponding cumulative distribution functions $F_1, F_2, F_3: \mathbb{R} \rightarrow [0, 1]$. Then, using Eq. (2) and Bayes' theorem of probability calculus, the trivariate density f can be written as

$$\begin{aligned} f(x_1, x_2, x_3) &= c_{2,3|X_1=x_1}(F_{2|X_1=x_1}(x_2), F_{3|X_1=x_1}(x_3)) \\ &\quad \times c_{2,1}(F_2(x_2), F_1(x_1)) c_{3,1}(F_3(x_3), F_1(x_1)) \\ &\quad \times f_1(x_1) f_2(x_2) f_3(x_3), \end{aligned} \quad (3)$$

for all $(x_1, x_2, x_3) \in \mathbb{R}^3$ such that $f(x_1, x_2, x_3) > 0$. Here, $c_{i,1}$ is the copula density corresponding to the distribution of (X_i, X_1) for $i = 2, 3$, and $c_{2,3|X_1=x_1}$ is that corresponding to the distribution of (X_2, X_3) conditional on $X_1 = x_1$. Furthermore, $F_{i|X_1=x_1}$ denotes the distribution function of X_i conditional on $X_1 = x_1$ for $i = 2, 3$.

For computational feasibility, one generally assumes that the bivariate copula density $c_{2,3|X_1=x_1}$ does not depend on the specific value x_1 of X_1 , denoting it by $c_{2,3|1}$, see [5]. We also remark that the order of the entries X_1, X_2, X_3 of the random vector (X_1, X_2, X_3) can be changed, i.e., on the right-hand side of Eq. (3) one can condition on $X_2 = x_2$ or $X_3 = x_3$, instead of conditioning on $X_1 = x_1$. In this way, the right-hand side of Eq. (3) can be adjusted to suit the needs for further use of this copula representation formula.

A.4. Fitting of copula densities

For fitting a multivariate probability density $f: \mathbb{R}^d \rightarrow [0, \infty)$ to data, we need to fit the copula density $c: [0, 1]^d \rightarrow [0, 1]$ appearing on the right-hand side of Eq. (2), where we exploit the C-library `vinecopulib` and its python bindings (`pyvinecopulib`) [11] for choosing an appropriate copula model and calibrating it to given data. For a large variety of parametric copula families, this package provides tools for data fitting, simulation and further purposes.

For $d = 2$, similar to choosing marginal distributions as described above, we choose bivariate copula families which perform equally well for two different datasets. More

precisely, for fitting the copula density $c : [0, 1]^2 \rightarrow [0, 1]$, after choosing appropriate marginal distribution functions F_1 and F_2 for a dataset $(x_1^{(1)}, x_2^{(1)}), \dots, (x_1^{(N)}, x_2^{(N)})$ and transforming the data into the dataset $(u_1^{(1)}, u_2^{(1)}), \dots, (u_1^{(N)}, u_2^{(N)})$, where $u_j^{(i)} = F_j(x_j^{(i)})$ for $i = 1, \dots, N$ and $j = 1, 2$, we use the tools provided by vinecopulib to fit various copula models to $(u_1^{(1)}, u_2^{(1)}), \dots, (u_1^{(N)}, u_2^{(N)})$. In particular, for each dataset extracted from Sample 1 and Sample 2, respectively, and for each bivariate copula model provided by vinecopulib, we compute the log-likelihood function and finally choose the copula such that the minimum of the maximum log-likelihood among both datasets is maximized. Similar to the method explained above for choosing the marginal distributions, this ensures that the behavior of both datasets can be captured equally well using the same copula, albeit with different parameters.

Note that also in the three-dimensional case ($d = 3$), for fitting a trivariate probability density $f : \mathbb{R}^3 \rightarrow [0, \infty)$ to data, we only need to fit bivariate copulas, using the representation formula given in Eq. (3). For example, for fitting the (unconditional) copula density $c_{2,1}$ appearing on the right-hand side of Eq. (3), after choosing appropriate marginal distribution functions F_2 and F_1 for a dataset $(x_2^{(1)}, x_1^{(1)}), \dots, (x_2^{(N)}, x_1^{(N)})$ and transforming the data into the dataset $(u_2^{(1)}, u_1^{(1)}), \dots, (u_2^{(N)}, u_1^{(N)})$, where $u_j^{(i)} = F_j(x_j^{(i)})$ for $i = 1, \dots, N$ and $j = 1, 2$, where we again use the tools provided by vinecopulib to fit various copulas, but now to the dataset $(u_2^{(1)}, u_1^{(1)}), \dots, (u_2^{(N)}, u_1^{(N)})$. Moreover, as in the two-dimensional case described above, for each dataset extracted from Sample 1 and Sample 2, respectively, and for each bivariate copula model provided by vinecopulib, we compute the log-likelihood function and choose the copula such that the minimum of the maximum log-likelihood among both datasets is maximized. The other (unconditional) copula density $c_{3,1}$ on the right-hand side of Eq. (3) is fitted to data in the same way. For fitting the (conditional) copula density $c_{2,3|1}$ on the right-hand side of Eq. (3), we consider the dataset $(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}), \dots, (x_1^{(N)}, x_2^{(N)}, x_3^{(N)})$ and transform it into the dataset $(u_1^{(1)}, u_2^{(1)}, u_3^{(1)}), \dots, (u_1^{(N)}, u_2^{(N)}, u_3^{(N)})$, as described above for the two-dimensional case. Then, using so-called h -functions, we further transform the latter dataset into the dataset $(\tilde{u}_2^{(1)}, \tilde{u}_3^{(1)}), \dots, (\tilde{u}_2^{(N)}, \tilde{u}_3^{(N)})$, where $\tilde{u}_j^{(i)} = h_{j,1}(u_j^{(i)}, u_1^{(i)})$ for $i = 1, \dots, N$ and $j = 2, 3$ and $h_{j,1}(u_j, u_1) = \mathbb{P}(U_j \leq u_j \mid U_1 = u_1)$ for $j = 2, 3$. Here, (U_j, U_1) denotes a two-dimensional random vector with (joint) probability density $c_{j,1}$ for $j = 2, 3$. Finally, the copula density $c_{2,3|1}$ is fitted to the dataset $(\tilde{u}_2^{(1)}, \tilde{u}_3^{(1)}), \dots, (\tilde{u}_2^{(N)}, \tilde{u}_3^{(N)})$, as described above.

A.5. Drawing samples from conditional probability densities

For simulation purposes, we need methods to draw samples from (conditional) distributions modeled by a copula approach as given in Eqs. (1) and (3), where we use the tools provided by vinecopulib [11].

For conditional sampling, we first consider the two-dimensional case where, according to Eq. (1), the joint distribution function of (X_1, X_2) is given by the marginal distribution functions $F_1: \mathbb{R} \rightarrow [0, 1]$ and $F_2: \mathbb{R} \rightarrow [0, 1]$ of X_1 and X_2 , respectively, and the copula $C: [0, 1]^2 \rightarrow [0, 1]$ describing the joint distribution function of $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$. Drawing a value x_2 from X_2 conditional on $X_1 = x_1$ can be performed by drawing a value u_2 from U_2 conditional on $U_1 = F_1(x_1)$ and setting $x_2 = F_2^{-1}(u_2)$. For this, we employ the inverse of the so-called first h-function $h: [0, 1]^2 \rightarrow [0, 1]$ corresponding to the copula C which is defined as

$$h(u_1, u_2) = \mathbb{P}(U_2 \leq u_2 \mid U_1 = u_1) \quad \text{for any } u_1, u_2 \in [0, 1], \quad (4)$$

where the inverse $h^{-1}: [0, 1]^2 \rightarrow [0, 1]$ of this function is computed with respect to the second argument, i.e., $h^{-1}(u_1, v)$ is the v -quantile of the distribution of U_2 conditional on $U_1 = u_1$. For some copulas, analytical methods to evaluate the h -functions and their inverse exist while for others, numerical methods can be used. For the copulas used in our model, they are implemented in `vinecopulib` [11]. Now, we can apply the inversion method for simulating from the distribution of U_2 conditional on $U_1 = u_1$ [10]. This is, drawing a sample $x_2 \in \mathbb{R}$ from X_2 conditional on $X_1 = x_1$ amounts to drawing a sample $v \in [0, 1]$ from the (standard) uniform distribution on the unit interval $[0, 1]$ and computing

$$x_2 = F_2^{-1}(h^{-1}(F_1(x_1), v)).$$

We use this approach to sample the angle B conditional on the value of A .

In the three-dimensional case, we need to sample a random variable X_3 conditional on two random variables X_1 and X_2 , where the joint density of (X_1, X_2, X_3) is given by the copula representation as stated in Eq. (3).

For doing this, we use the so-called Rosenblatt transform [6] which, in the three-dimensional case, maps a sample (x_1, x_2, x_3) of (X_1, X_2, X_3) onto a sample (r_1, r_2, r_3) of (R_1, R_2, R_3) , where R_1, R_2, R_3 are independent random variables which are uniformly distributed on $[0, 1]$ such that

$$\begin{aligned} r_1 &= F_1(x_1), \\ r_2 &= F_{2|X_1=x_1}(x_2), \\ r_3 &= F_{3|X_1=x_1, X_2=x_2}(x_3), \end{aligned}$$

where $F_1: \mathbb{R} \rightarrow [0, 1]$ denotes the cumulative distribution function of X_1 and $F_{2|X_1=x_1}$ and $F_{3|X_1=x_1, X_2=x_2}$ denote the conditional distribution functions of X_2 conditional on X_1 , and X_3 conditional on X_1 and X_2 , respectively. When the corresponding densities are non-zero, the inverse Rosenblatt transform can be obtained by

$$\begin{aligned} x_1 &= F_1^{-1}(r_1), \\ x_2 &= F_{2|X_1=x_1}^{-1}(r_2), \\ x_3 &= F_{3|X_1=x_1, X_2=x_2}^{-1}(r_3). \end{aligned} \quad (5)$$

As R_1, R_2, R_3 are independent, drawing from X_3 conditional on $X_1 = x_1$ and $X_2 = x_2$ can be performed by drawing r_3 from the uniform distribution on $[0, 1]$ and then computing x_3 using the inverse Rosenblatt transform, see Eq. (5). As the joint density of (X_1, X_2, X_3) can be expressed by the pair-copula representation as stated in Eq. (3), the expression for x_3 given above can be written as

$$x_3 = F_3^{-1}(h_{1,3}^{-1}(r_1, h_{2,3|1}^{-1}(r_2, r_3))), \quad (6)$$

where $r_1 = F_1(x_1)$, $r_2 = h_{1,2}(r_1, F_2(x_2))$ and $h_{1,3}, h_{1,2}, h_{2,3|1}: [0, 1]^2 \rightarrow [0, 1]$ are h-functions similar to Eq. (4) given by

$$\begin{aligned} h_{1,2}(u_1, u_2) &= \mathbb{P}(U_2 \leq u_2 \mid U_1 = u_1), \\ h_{1,3}(u_1, u_3) &= \mathbb{P}(V_3 \leq u_3 \mid V_1 = u_1), \\ h_{2,3|1}(u_2, u_3) &= \mathbb{P}(W_3 \leq u_3 \mid W_2 = u_2) \end{aligned}$$

for $u_1, u_2, u_3 \in [0, 1]$. Here, the random vectors (U_2, U_1) , (V_3, V_1) and (W_2, W_3) are distributed according to the copula densities $c_{2,1}$, $c_{3,1}$, $c_{2,3|1}$ appearing in Eq. (3), respectively. As in the two-dimensional case, the inverse h-functions in Eq. (6) are computed with respect to the second argument and are implemented in `pyvinecopulib` [11]. We use this approach for simulating $Z_3 - Z_2$ conditional on the values of $Z_2 - Z_1$ and Z_1 .

References

- [1] K.-M. Aigner, P. Schaumann, F. von Loeper, A. Martin, V. Schmidt, and F. Liers. Robust DC optimal power flow with modeling of solar power supply uncertainty via R-vine copulas. *Optimization and Engineering*, published online: <https://doi.org/10.1007/s11081-022-09761-0>, 2022.
- [2] C. Czado. *Analyzing Dependent Data with Vine Copulas*. Springer, 2019.
- [3] O. Furat, T. Kirstein, T. Leißner, K. Bachmann, J. Gutzmer, U. A. Peucker, and V. Schmidt. Multidimensional characterization of particle morphology and mineralogical composition using CT data and R-vine copulas. *Minerals Engineering*, submitted, <https://doi.org/10.48550/arXiv.2301.07587>, 2023.
- [4] A. Genz and F. Bretz. *Computation of Multivariate Normal and t Probabilities*, volume 195. Springer Science & Business Media, 2009.
- [5] I. H. Haff, K. Aas, and A. Frigessi. On the simplified pair-copula construction—simply useful or too simplistic. *Journal of Multivariate Analysis*, 101(5):1296–1310, 2010.

- [6] H. Joe. *Dependence Modeling with Copulas*. Chapman and Hall/CRC, 2014.
- [7] M. Nardon and P. Pianca. Simulation techniques for generalized Gaussian densities. *Journal of Statistical Computation and Simulation*, 79(11):1317–1329, 2009.
- [8] A. Navarro, J. Frellsen, and R. Turner. The multivariate generalised von Mises distribution: Inference and applications. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 2394–2400, 2017.
- [9] R. B. Nelsen. *An Introduction to Copulas*. Springer, 2006.
- [10] C. P. Robert, G. Casella, and G. Casella. *Monte Carlo Statistical Methods*, volume 2. Springer, 1999.
- [11] Vinecopulib. Vinecopulib/pyvinecopulib: A python library for vine copula models. <https://github.com/vinecopulib/pyvinecopulib>, 2023. Accessed: 2023-04-20.
- [12] 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*, 17:261–272, 2020.