Big data for microstructure-property relationships: a case study of predicting effective conductivities

Ole Stenzel[1,*], Matthias Neumann[2,*,**], Omar Pecho[1], Lorenz Holzer[1] and Volker Schmidt[2]

[1] Institute of Computational Physics, ZHAW, CH-8400 Winterthur, Switzerland

[2] Institute of Stochastics, Ulm University, D-89069 Ulm, Germany

[*] Both authors have equally contributed to the present paper

[**] Corresponding author. Phone: +49 731 50 23617. Fax: +49 731 50 23649. Email: matthias.neumann@uni-ulm.de

Abstract

The analysis of big data is changing industries, businesses and research since large amounts of data are available nowadays. In the area of microstructures, acquisition of (3D tomographic image) data is difficult and time-consuming. It is shown that large amounts of data representing the geometry of virtual, but realistic 3D microstructures can be generated using stochastic microstructure modeling. Combining the model output with physical simulations and data mining techniques, microstructure-property relationships can be quantitatively characterized. Exemplarily, we aim to predict effective conductivities given the

microstructure characteristics volume fraction, mean geodesic tortuosity and constrictivity. Therefore, we analyze 8119 microstructures generated by two different stochastic 3D microstructure models. This is - to the best of our knowledge - by far the largest set of microstructures that has ever been analyzed. Fitting artificial neural networks, random forests and classical equations, the prediction of effective conductivities based on geometric microstructure characteristics is possible.

Keywords: *Big data, effective conductivity, geodesic tortuosity, microstructure characteristics, predictive simulation, stochastic microstructure modeling.*

1 Introduction

Data is the new oil. The analysis of big data is changing industries, businesses and research. Big data is also used in order to advance materials research [1] aiming an accelerated systematic design of functional materials [2] like (organic) solar cells, fuel cells and batteries. This includes the identification of new molecules with desired properties as well as the optimization of micro- or nanostructures, i.e. the spatial arrangement of materials components, which have a large influence on the functional properties of these materials, see [3] and the references therein.

In order to optimize microstructures in functional materials, the relation-

ship between structural characteristics and functional properties has to be understood quantitatively, which is often not the case or just for some special types of simple structures [3]. The progress of 3D imaging during the last decades enables the computation of well-defined microstructure characteristics from real data, which can be compared to effective properties that are either measured experimentally or simulated with numerical models [4, 5, 6, 7]. Although this approach allows a direct investigation of the relationship between microstructure and effective properties, it is limited due to the high costs of 3D imaging.

Thus, virtual materials testing (VMT), i.e. the combination of stochastic microstructure models (SMM) with numerical simulations of physical processes, was used in [8] and [9] to investigate the quantitative relationship between microstructure characteristics and effective conductivity in porous materials. The use of SMM allows us to generate virtual microstructures in short time, where certain microstructure characteristics can be varied systematically. The virtual microstructures are used as an input for finite element modeling (FEM) where the corresponding effective conductivities are simulated. The generation of virtual microstructures leads to big data and thus, the microstructure-property relationships can be considered as a statistical learning problem.

In [9] it was shown that effective conductivity σ_{eff} of porous microstructures can be approximately predicted by three microstructure characteristics, which are volume fraction ε of the solid phase, its mean geodesic tortuosity

τ_{geod} and a certain constriction factor β, using the equation

$$\sigma_{\text{eff}} = \sigma_0 \frac{\varepsilon^{1.15}\,\beta^{0.37}}{\tau_{\text{geod}}^{4.39}}, \tag{1}$$

where σ_0 denotes the intrinsic conductivity of the bulk material without microstructure limitation. This empirical relationship has been established on the basis of 43 virtual microstructures, where the corresponding effective conductivities have been computed with the software GeoDict [10]. In the present paper, we consider 8119 virtual microstructures and use neural networks and random forests to predict the effective conductivity given the structural properties. Although it is more difficult to interpret these new prediction formulas in comparison with Equation (1), they increase accuracy of the prediction. The effective conductivity can now be predicted with a prediction error of less than 9% instead of 13.6%, which is the prediction error when applying Equation (1) to the 8119 virtual microstructures. This shows that combining stochastic microstructure modeling, physical computations and data mining is a powerful and helpful approach to establish quantitative microstructure-property relationships. This concept, outlined in Figure 1, is not restricted to conductive transport processes, but can - in principle - be applied to establish all kinds of microstructure-property relationships.

The concept, however, has the disadvantage that the effective conductivities of the virtual microstructures can not be compared with experimental measurements. For a proper validation, one needs to prepare real samples,

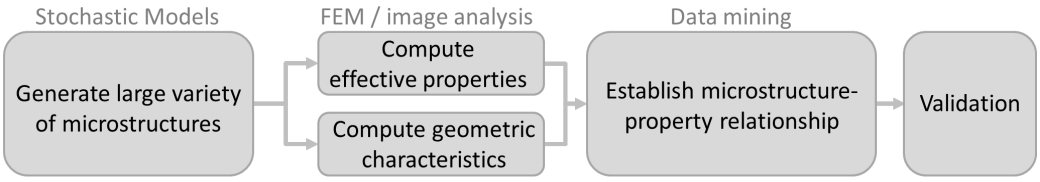

Figure 1: Combination of VMT with statistical learning to analyze microstructure-property relationships with big data.

measure the effective conductivities experimentally, do 3D imaging of the samples and then compare experimentally measured with predicted conductivities. Such a comparison is expensive in costs and time and can only be performed for a small number of samples. This was done in [8] to validate the VMT approach, where a reasonably good agreement between predicted and measured conductivities was found. Validation of simulating effective conductivity by GeoDict can be found in [11].

Moreover, it has to be emphasized that we do not want to replace experimental 3D imaging by the in-silico approach of VMT. It can be understood as an additional tool which makes 3D imaging more powerful in tailoring new microstructures with specific properties. In order to create virtual, but realistic microstructures, a certain SMM is fitted to experimental microstructures such that the model creates statistically equivalent microstructures. Then, model parameters of the SMM can be correlated to production parameters of the microstructures, in order to suggest production parameters that lead to a certain type of microstructure, see e.g. [12]. In many materials, e.g. in fuel cells, various different transport processes take place simultaneously which

makes microstructure optimization difficult: each type of transport process may prefer a different microstructure. Thus, for a successful microstructure optimization, quantitative microstructure-property relationships must be established not only for conductive transport but also for other kinds of transport processes, e.g. for effective permeability or mechanical stress-strain curves.

This paper is organized as follows. In Section 2, we present the SMM (Section 2.1) used for the generation of virtual microstructures, their geometric characteristics (volume fraction, constrictivity, mean geodesic tortuosity, Section 2.2), the considered transport processes (Section 2.3) as well as the predictive models from statistical learning (Section 2.4). The results are presented and discussed in Section 3, where the proposed microstructure-property relationships are validated by experimental data (Section 3.3), too. Conclusions are presented in Section 4.

2 Data & Methods

2.1 Stochastic microstructure modeling

With increasing availability of highly-resolved image data stochastic microstructure modeling becomes a frequently used tool in materials science [3]. During the last years a number of stochastic micro- and nanostructure models has been created for specific types of microstructures in organic solar cells, Li-ion batteries and fuel cells, see e.g. [13] and the references therein.

In general, an SMM uses tools from stochastic geometry [14] to generate virtual, random microstructures whose properties can be adjusted by the model parameters. To develop an SMM, a purposive combination of random variables is used to model spatial data, like point configurations, spatial networks or random sets. The generation of a virtual microstructure typically requires little computational effort and therefore many different microstructures can be simulated in short time.

A simple example for an SMM is the Boolean model with spherical grains, see e.g. [15], where possibly overlapping spheres are distributed completely at random in space (2D or 3D) with a predefined distribution of radii. The influence of model parameters on transport properties has been recently investigated in [16] for Boolean models with more general grains.

For our case study, we consider two SMM that generate different types of microstructures: the stochastic spatial graph model (SSGM) introduced in [8] and a simplified version of the multiscale sphere model (MSM) introduced in [17]. By means of the SSGM microstructures within a wide range of different values for volume fraction, mean geodesic tortuosity and constrictivity can be generated. We additionally incorporate the MSM into our investigation since it was fitted to image data of real microstructures. Moreover, considering two models instead of one reduces the errors introduced by the model type.

2.1.1 Stochastic spatial graph model

The stochastic spatial graph model (SSGM) is based on a random spatial graph that is randomly dilated [8], see Figure 2. The model has a large flexibility to generate microstructures with different volume fractions, mean geodesic tortuosities and constrictivities. All microstructures realized by the SSGM are completely connected by definition. Using the SSGM, 3900 microstructures with different structural characteristics have been generated for the present study.

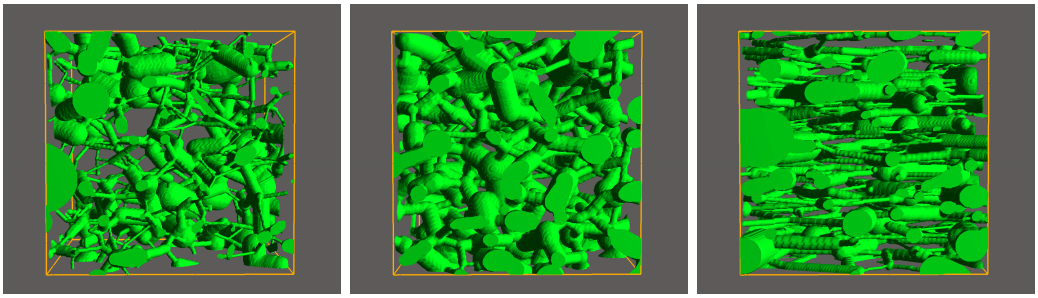

Figure 2: Virtual microstructures generated by the SSGM.

2.1.2 Multi-scale sphere model

The second SMM is the multi-scale sphere model (MSM) from [12, 17]. It follows a completely different approach in comparison to the SSGM described in Section 2.1.1. It is based on a random, anisotropic arrangement of spheres. The midpoints of the spheres follow a Markov-chain of 2D point processes. The model has two components: a macro-scale component and a micro-scale component that adds structural complexity. In the present paper, we only use

the macro-scale model of the MSM. Examples of realizations are displayed in Figure 3. In total, we consider 2131 microstructures where the sphere system is the transport phase and 2088 microstructures where transport takes place in the complement of the sphere system. Since only the connected (non-isolated) part of the considered material phase contributes to transport, a post-processing is applied where all material is removed that is not connected with both, in-let and out-let plane.

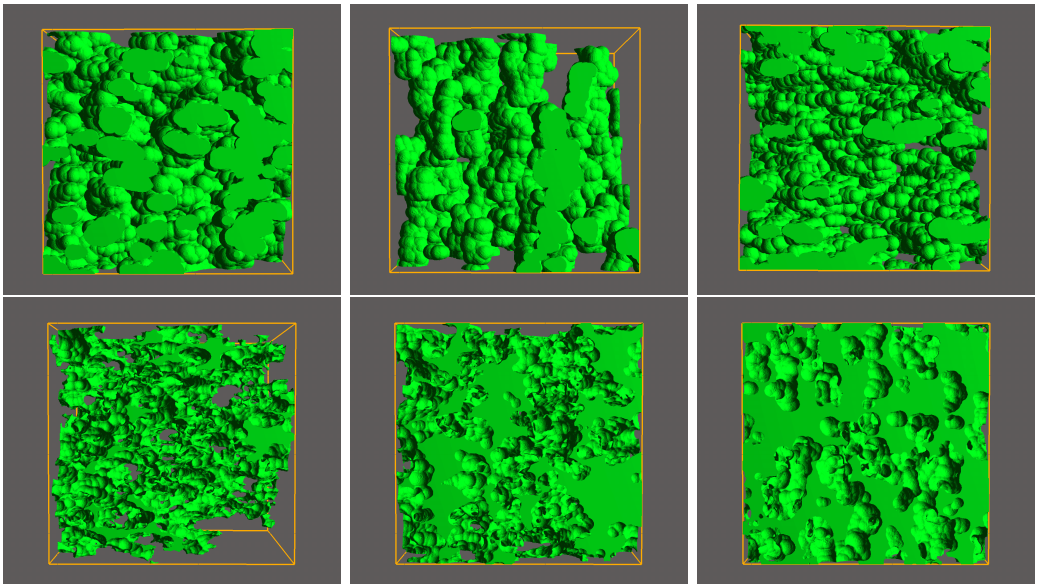

Figure 3: Virtual microstructures generated by a simplified version of the MSM. Top row: microstructures generated by a sphere system. Bottom row: microstructures generated as complementary phase of a sphere system.

2.2 Geometric characteristics

In [8] and [9] VMT has shown that three microstructure characteristics of the conducting phase carry significant information with respect to σ_{eff}. These microstructure characteristics are volume fraction ε, mean geodesic tortuosity τ_{geod} and constrictivity β. Note that these micostructure characteristics can be defined by means of expectations with respect to the underlying stochastic model, see e.g. [14], or they can be estimated from a given microstructure. In the present paper we use the latter way, because given a certain microstructure we are interested in the influence of $\varepsilon, \tau_{\text{geod}}$ and β on σ_{eff}.

The volume fraction ε is estimated by the ratio of the volume of the transporting phase divided by the total volume of the 3D image. The influence of winded transport paths of the conducting phase is described by the mean geodesic tortuosity τ_{geod}, which is defined as the ratio of the expected shortest path lengths from inlet- to outlet-plane over the material thickness. Thereby, the shortest path lengths (in terms of geodesic distance [18]) in transport direction from inlet- to outlet-planes are computed within the voxel space that represents the transporting phase (see the left-hand side of Figure 4). To determine τ_{geod} we consider an average of geodesic tortuosities computed for all voxels of the transporting phase in the inlet-plane. Obviously, it holds that $\tau_{\text{geod}} \geq 1$ and higher values of τ_{geod} indicate more winded pathways.

Besides the windedness of transport paths through the material, narrow constrictions of the conducting phase, quantified by the so-called constrictiv-

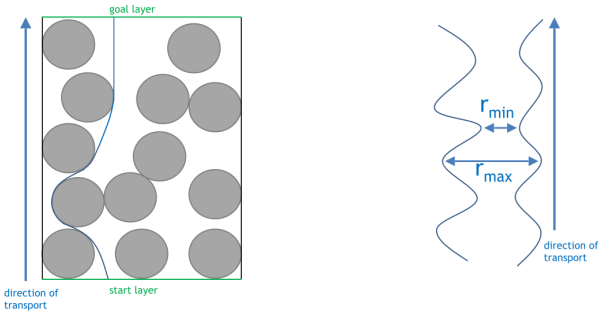

Figure 4: Concept of geodesic tortuosity τ_{geod} (left) and concept of constrictivity $\beta = (r_{\min}/r_{\max})^2$ (right).

ity β, have a strong influence on σ_{eff}. Constrictivity is defined as

$$\beta = \left(\frac{r_{\min}}{r_{\max}} \right)^2, \tag{2}$$

where, heuristically speaking, $r_{\min}$ indicates the radius of the characteristic bottleneck and $r_{\max}$ indicates the radius of the characteristic bulge, see the right-hand side of Figure 4. More precisely, $r_{\max}$ is the 50% quantile of the continuous pore size distribution (c-PSD) and $r_{\min}$ is the 50% quantile of the MIP pore size distribution, which is based on a geometrical simulation of mercury intrusion porosimetry (MIP), introduced in [19]. Constrictivity takes values between 0 and 1, where values close to 0 indicate strong bottleneck effects while values close to 1 indicate that there are no bottlenecks at all.

For details regarding these structural characteristics and their estimation from 3D image data, the reader is referred to [9]. The 8119 microstructures generated by the aid of SSGM and MSM cover a wide range of values for the

characteristics $\varepsilon, \tau_{\text{geod}}$ and β, see Section 3.

2.3 Conductive transport

As in [8] and [9], we consider conductive transport processes within composite materials, where only one phase is conducting. The electric charge transport is described by Ohm's law

$$J = -\sigma \frac{\mathrm{d}U}{\mathrm{d}x} \tag{3}$$

and

$$\frac{\mathrm{d}U}{\mathrm{d}t} = \sigma \frac{\mathrm{d}^2U}{\mathrm{d}x^2}, \tag{4}$$

where J is the current density, σ is the conductivity, U is the electric potential, and t is time. Assuming constant boundary conditions, such systems converge to an equilibrium which is described by the Laplace equation

$$\frac{\mathrm{d}^2U}{\mathrm{d}x^2} + \frac{\mathrm{d}^2U}{\mathrm{d}y^2} + \frac{\mathrm{d}^2U}{\mathrm{d}z^2} = 0, \tag{5}$$

where x, y and z denote the coordinates in the 3D Euclidean space.

Since transport only takes place in one phase, the geometry of the microstructure reduces the intrinsic conductivity σ_0 of the material to the effective conductivity σ_{eff}, i.e.

$$\sigma_{\text{eff}} = \sigma_0 M \tag{6}$$

for some $0 \leq M \leq 1$. The influence of the microstructure on the effective conductivity is described by the factor M. Our goal is to validate the prediction of the M-factor based on the geometric characteristics $\varepsilon, \tau_{\text{geod}}$ and β, which has been derived in [9]. Moreover, we improve the prediction formula using methods from statistical learning, i.e. by neural networks and random forests. For each of the 8119 synthetic microstructures, the effective conductivity and the associated M-factor are determined by numerical simulation using the software GeoDict [10].

2.4 Statistical learning

Neural networks and random forests are two methods from statistical learning that can be used for non-linear regression [20]. Both methods are used to predict the M-factor of virtual 3D microstructures by the corresponding values of $\varepsilon, \tau_{\text{geod}}$ and β. Basically leaned on [20], we give a short description of neural networks and random forests. In both cases, an output variable $Y \in \mathbb{R}$ is predicted by an input vector $X \in \mathbb{R}^p$ consisting of p features, where $p \in \mathbb{N}$. In our case we have $X = (\varepsilon, \tau_{\text{geod}}, \beta)$ and $Y = \log_2(M)$. Since the computed M-factors vary over several orders of magnitude, a better fit is obtained by putting the M-factors on a $\log_2$-scale.

Neural networks are two-stage regression models. Here we use a single hidden layer network. For prediction of Y, the vector X is mapped to the hidden layer, which is a vector $Z \in \mathbb{R}^L, L \in \mathbb{N}$, where for each $l \in \{1, \ldots, L\}$ we have $Z_l = \sigma\left(\alpha_{0,l} + \sum_{i=1}^{p} \alpha_{i,l} X_i\right)$ for a parameter matrix $\alpha = (\alpha_{i,j}) \in$

$\mathbb{R}^{(p+1)\times L}$ and some function $\sigma : \mathbb{R} \longrightarrow \mathbb{R}$. Here we choose σ as the sigmoid function, i.e. $\sigma(t) = (1 + e^{-t})^{-1}$ for each $t \in \mathbb{R}$. The predictor $\widehat{Y}$ of Y is finally constructed by a linear combination of the entries of Z, to be more precise $\widehat{Y} = \min\{\max\{\widehat{Y}^*, 0\}, 1\}$ with

$$\widehat{Y}^* = \theta_0 + \sum_{i=1}^{L} \theta_i Z_i \tag{7}$$

for some parameter vector $\theta \in \mathbb{R}^{L+1}$. In order to fit the parameters α and θ, we minimize the mean squared error (MSE) between $\widehat{Y}^*$ and Y by the Matlab implementation [21] of the Levenberg-Marquardt backpropagation algorithm [22], where the initial values are determined by the Nguyen-Widrow algorithm [23]. During the fitting procedure, data is divided completely at random into training data (70%), validation data (15%) and test data (15 %). Training data is directly used to fit α and θ, whereas validation data is used to define a stopping criterium for the Levenberg-Marquardt algorithm [20]. The dimension L of the hidden layer is chosen such that the mean absolute percentage error (MAPE) of test data is minimized. For this purpose, we average over 200 random subdivisions, where data is divided into training data, validation data and test data. Altogether, the described division of data avoids overfitting by the neural network.

Random forests [24] are further regression models from statistical learning, which are based on so-called regression trees. The predictor $\widehat{Y}$ of Y obtained by a single regression tree is a linear combination of indicators, i.e.

$$\widehat{Y} = \sum_{m=1}^{M} c_m \mathbb{1}_{x \in R_m}, \tag{8}$$

for an appropriate partition $\mathcal{R} = \{R_1, \ldots, R_M\}$ of $\mathbb{R}^p$, where $\mathbb{1}_{x \in R} = 1$ if $x \in R$ and $\mathbb{1}_{x \in R} = 0$ otherwise for each $R \subset \mathbb{R}^p$. Beginning with $\mathcal{R} = \{\mathbb{R}^p\}$ the partition $\mathcal{R}$ is refined iteratively. In each iteration, all regions are split into two half-spaces such that by an optimal choice of coefficients c_m the MSE can be minimized. The refinement is stopped when each region contains a predefined minimum number of observations of X. For our purpose this minimum number is set to 5 as recommended in [20]. In random forests averaging over randomized regression trees improves the prediction. Randomization takes place in two different ways. To fit the individual regression trees, different random subsets of the input vector are chosen. Moreover, $k < p$ features of X, denoted by $i_1, \ldots, i_k$, are chosen at random for each splitting of a region. Then, splitting is only possible along one of the axes $i_1, \ldots, i_k$. Usually $k = \sqrt{p}$ is used. This procedure allows a variance reduction of the predictor $\widehat{Y}$, caused by averaging of single regression trees as well as by the described randomization. For prediction of the M-factor, we choose $k = 2$. Similar to neural networks we divide data into training data (70%) and test data (30 %) completely at random. The number of trees used for averaging is chosen such that the MSE of test data does not decrease significantly for a larger number of trees. As in the case of neural networks, we consider 200 random subdivisions to determine the number of trees. In

order to fit and simulate random forests, we use the randomForest-package [25] of the statistical software R [26].

3 Results & Discussion

Simulations, which are based on the stochastic models presented in Section 2.1, provide 8119 virtual 3D microstructures. For each of these virtual 3D microstructures, we compute the geometric microstructure characteristics $\varepsilon, \tau_{\text{geod}}$ and β, described in Section 2.2, as well as the corresponding M-factor, see Section 2.3.

3.1 Characteristics of simulated virtual 3D microstructures

Figure 5 shows that the generated virtual 3D microstructures cover a wide range of constellations for $\varepsilon, \tau_{\text{geod}}$ and β. For small values of ε, many microstructures are generated, the transport paths of which are more than 1.5 longer than the materials thickness, that is $\tau_{\text{geod}} \geq 1.5$. The unflexibility of the models regarding τ_{geod} for large values of ε is not surprising, since - excluding pathological counterexamples - the mean length of transport paths through the material decreases strongly with increasing volume fractions.

Most values of β are in the interval $[0, 0.8]$ and for virtual microstructures with $\varepsilon \in [0.4, 0.7]$ the corresponding constrictivities take nearly all values between 0.05 and 0.7. While higher values of constrictivity are observed in

virtual microstructures generated by the MSM (blue and red dots in Figure 5), the correlation between ε and β is less strong in the SSGM (black dots in Figure 5). The SSGM was especially developed for varying the considered microstructure characteristics as independently as possible [8].

The right-hand side of Figure 5 shows that the simulated M-factors of the virtual 3D microstructures cover the whole range between 0 and 1. All M-factors are below the upper bound $M \leq \varepsilon^{1.15}$ (green dashed line) resulting from the empirically derived prediction formula, see Equation (1). Note that a rigorous upper bound for M is given by $M \leq \varepsilon$ [3].

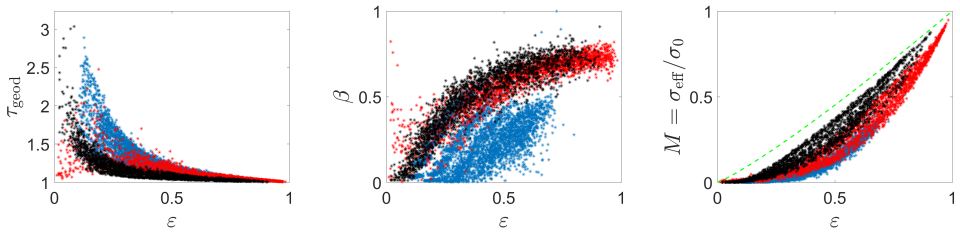

Figure 5: Characteristics of the 8119 virtual 3D microstructures generated by the SSGM (blue), the MSM (red) and the complement of the MSM (black). The plots show mean geodesic tortuosity τ_{geod} (left), constrictivity β (center) and M-factor vs. volume fraction ε.

3.2 Prediction of M-factor by geometric microstructure characteristics

On the basis of the simulated microstructures we validate the prediction formula derived in [9]. Furthermore, we present the predictions obtained by neural networks and random forests, which are fitted to simulated data as it

is described in Section 2.4. Figure 6 shows scatter plots of computed and predicted M-factors, while the MAPE as well as the coefficient of determination R^2 are listed in Table 1.

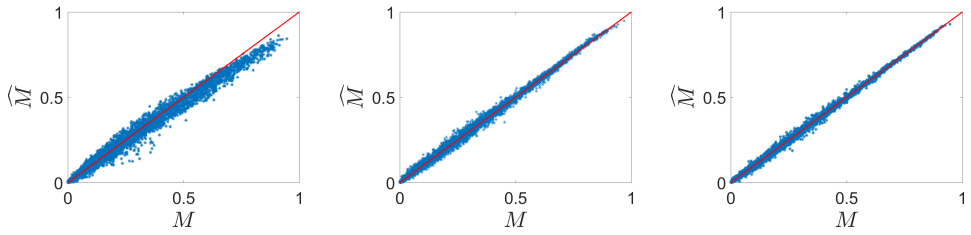

Figure 6: Scatter plots of computed and predicted M-factors using the prediction formula from Equation (1) (left), neural networks (center) and random forests (right), where the identity function is added in each plot (red lines).

For the prediction formula given in Equation (1), the MAPE corresponding to the 8119 virtual microstructures is 13.6%, while the MAPE was 19.6% for the virtual microstructures considered in [9]. The reason for this smaller value of MAPE is that the microstructures analyzed in the present paper are less extreme, i.e. they have a larger average M-factor than those in [9]. Altogether, the formula given in Equation (1) offers a good prediction of the M-factor, see Figure 6 (left), which is also indicated by a high coefficient of determination R^2. However, the formula seems to systematically underestimate the M-factor for values above 0.7, i.e. for materials with a high volume fraction.

Using neural networks and random forests the prediction of the M-factor can be improved. Fitting a single hidden layer neural network leads to a hidden layer of size $L = 20$. The MAPE for the test data is 8.94%, while

Table 1: MAPE and coefficient of determination R^2 of the different prediction models. Note that the prediction formula from Equation (1) was not fitted to the data simulated in the present study. Thus, the complete data can be considered as test data in this case.

Model	MAPE (training data)	MAPE (test data)	R^2
Prediction formula	-	13.6%	0.984
Neural network	8.20%	8.94%	0.997
Random forest	3.99%	8.47%	0.999

$R^2 = 0.997$. For prediction by a random forest we average over 500 trees and obtain a MAPE of 8.47%, which is slightly better than prediction by neural networks. Also $R^2 = 0.999$ shows a better prediction by random forests. Note that random forests, in contrast to neural networks, have a much smaller MAPE for training data than test data, see Table 1.

Random forests and neural networks offer a much lower prediction error than the formula given by Equation (1), see Table 1. Thus, for prediction purposes, random forests or neural networks should be used from our point of view. Both methods are equivalent in terms of their prediction accuracy. However, random forests and neural networks are extremely difficult if not impossible to interpret. Thus, it is difficult to explain why a microstructure has a certain M-factor. The big advantage of the prediction formula from Equation (1) is that it allows us to explain how $\varepsilon, \tau_{\text{geod}}$ and β influence the M-factor. In short, we propose to use neural networks and random forests for prediction and Equation (1) for explanation.

Using neural networks or random forests, the MAPE of test data is smaller

than 9%. This means that the considered volume-averaged characteristics $\varepsilon, \tau_{\text{geod}}$ and β carry significant information about effective conductivity, but certainly not all information. One possibility to further improve the prediction accuracy would be to consider the active volume fraction instead of the connected volume fraction. Imagine a microstructure that is completely connected yet has many dead-ends which are not used for transport. Then considering active volume (connected volume minus 'dead-end'-volume) instead of connected volume should further increase prediction accuracy. The precise mathematical definition and computation of active volume, however, is challenging and subject of current research.

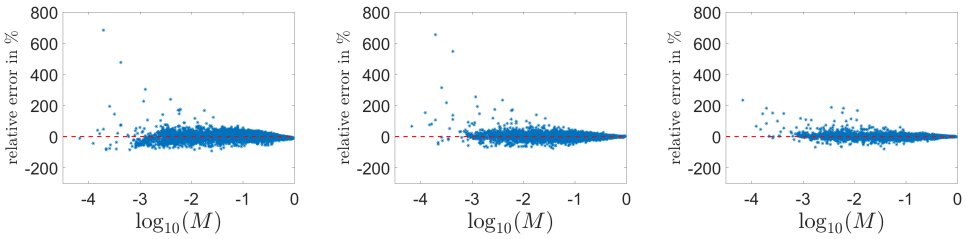

Figure 7: Computed M-factors on a $\log_{10}$-scale vs. relative prediction errors. Predictions are obtained by the prediction formula from Equation (1) (left), neural networks (center) and random forests (right).

Considering Figure 6, it seems that all three prediction models work well for all microstructures without any exceptions. However, for all three methods, the prediction error increases for decreasing M-factors and extreme errors occur for very small M-factors (below 10^{-2}), see Figure 7. Note that the errors are less extreme when random forests are used for prediction of the M-factor.

Interestingly, all extreme errors overestimate the M-factor, i.e. the corresponding microstructures have a smaller M-factor than predicted. These extreme deviations are caused by microstructures, which are close to their percolation threshold, i.e. eroding the microstructure a little bit would eliminate connectivity. The microstructures have a low connectivity and much of the volume is not used for transport ('dead-end' volume). Measuring active volume instead of connected volume could lead to a better prediction of the M-factor.

3.3 Validation with experimental microstructures

In order to validate our method, we compare M-factors predicted by the three different methods with computed M-factors (using GeoDict) for different 3D image data obtained by FIB-SEM tomography. For this purpose, the same data sets are considered, which have also been used in [9] for validation of Equation (1). In total we have 10 images, where six of them representing anodes in solid oxide fuel cells (SOFC) consisting of pores, nickel (Ni) and yttrium-stabilized zirconia (YSZ) [27] and four of them represent porous membranes used as liquid junctions in pH-Sensors [28]. In the SOFC anodes electric conduction takes place in the Ni phase and ionic conduction in the YSZ phase, while liquid electrolyte diffusion occurs in the pores of the membranes of pH-Sensors. Note that mathematically the concept of effective diffusivity is the same as the concept of effective conductivity and the M-factor can be analogously defined for diffusion processes. For more

information about the experimental data, see [9] and the references therein.

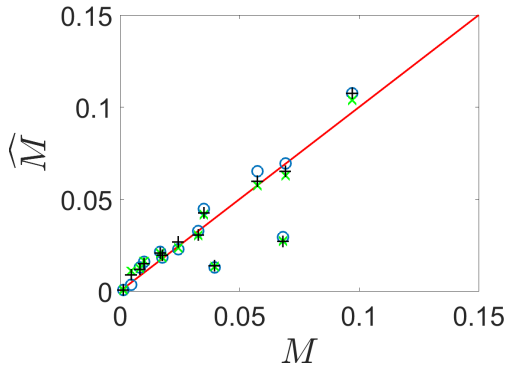

Figure 8: Computed M-factors and the corresponding predictions $\widehat{M}$ for experimental image data. Predictions have been performed by Equation (1) (blue circles), by neural networks (green crosses) and random forests (black plus signs).

In Figure 8 the M-factors computed by numerical simulation on the image data sets are compared to the predictions by Equation (1), neural networks and random forests. In general, the prediction fits the simulated M-factors nicely, where the results obtained from statistical learning are slightly worse than those obtained by the prediction formula. The MAPE is 28.0% for the prediction formula, 33.8% for the neural network and 30.3% for the random forest. However, note that only 16 values of the M-factors are considered. Thus, there is no need to withdraw the conclusion from Section 3.2 based on more than 8000 virtual microstructures, which is that methods from statistical learning improve the prediction of the M-factor by ε, τ_{geod} and β.

Figure 8 shows two outliers which can be explained as follows: The two data points represent electric conductivity in the Ni phase of SOFC anodes

that were exposed to harsh conditions, which led to strong microstructure alteration (i.e. Ni-agglomeration). It was shown in [9] and [27] that due to the strong alteration, the representative volume is much larger than the observation window that can be obtained by FIB-tomography. Therefore the analyses based on these two 3D-data sets suffer from a high uncertainty. For all other data points the predictions are reasonably well. From the validation with experimental microstructures we can conclude that the stochastic models are realistic enough to use them in order to derive predictors for effective conductivity.

4 Conclusion

In the present paper, we investigate microstructure-property relationships for conductive transport processes using 8119 virtual microstructures generated by SMM. Effective conductivity is predicted by the three microstructure characteristics volume fraction, mean geodesic tortuosity and constrictivity. The interpretable prediction formula proposed in [9] yields a prediction error of 13.6%, which can be considered as a further validation of this prediction formula since only 43 virtual microstructures have been used to derive it. Random forests and neural networks which are difficult to interpret yield smaller prediction errors of less than 9%, where in all cases the prediction becomes unstable for microstructures at their percolation threshold.

Validation with experimental microstructures shows that the generated

virtual microstructures are sufficiently realistic to derive prediction models for effective conductivity. Overall, the present paper points out that the combination of stochastic microstructure modeling with physical computations and data mining techniques is a powerful tool to establish quantitative microstructure-property relationships. These relationships enable the identification of improved microstructures with respect to effective conductivity. The method itself is not restricted to conduction processes and can also be used to investigate relationships between microstructure characteristics and other functional properties, like e.g. effective permeability or mechanical stress-strain curves.

Supplementary material

The fitted neural network as well as the fitted random forest are provided as supplementary material. The code can be used to predict the M-factor for given volume fraction, mean geodesic tortuosity and constrictivity.

References

[1] White A. Big data are shaping the future of materials science. MRS Bulletin. 2013;38:594–595.

[2] Kalidindi SR, De Graef M. Materials data science: current status and future outlook. Annual Review of Materials Research. 2015;45:171–193.

[3] Torquato S. Random Heterogeneous Materials: Microstructure and Macroscopic Properties. New York: Springer. 2013.

[4] Doyle M, Fuller TF, Newman J. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. Journal of the Electrochemical Society. 1993;140(6):1526–1533.

[5] Holzer L, Wiedenmann D, Münch B, Keller L, Prestat M, Gasser P, Robertson I, Grobéty B. The influence of constrictivity on the effective transport properties of porous layers in electrolysis and fuel cells. Journal of Materials Science. 2013;48:2934–2952.

[6] Shikazono N, Kanno D, Matsuzaki K, Teshima H, Sumino S, Kasagi N. Numerical assessment of SOFC anode polarization based on three-dimensional model microstructure reconstructed from FIB-SEM images. Journal of the Electrochemical Society. 2010;157(5):B665–B672.

[7] Tippmann S, Walper D, Balboa L, Spier B, Bessler WG. Low-temperature charging of lithium-ion cells part I: Electrochemical modeling and experimental investigation of degradation behavior. Journal of Power Sources. 2014;252:305–316.

[8] Gaiselmann G, Neumann M, Pecho OM, Hocker T, Schmidt V, Holzer L. Quantitative relationships between microstructure and effective transport properties based on virtual materials testing. AIChE Journal. 2014; 60(6):1983–1999.

[9] Stenzel O, Pecho OM, Neumann M, Schmidt V, Holzer L. Predicting effective conductivities based on geometric microstructure characteristics. AIChE Journal. 2016;62:1834–1843.

[10] GeoDict. www.geodict.com. 2014.

[11] Becker J, Flückiger R, Reum M, Büchi FN, Marone F, Stampanoni M. Determination of material properties of gas diffusion layers: experiments and simulations using phase contrast tomographic microscopy. Journal of The Electrochemical Society. 2009;156(10):B1175–B1181.

[12] Stenzel O, Koster L, Thiedmann R, Oosterhout SD, Janssen RAJ, Schmidt V. A New Approach to Model-Based Simulation of Disordered Polymer Blend Solar Cells. Advanced Functional Materials. 2012; 22(6):1236–1244.

[13] Neumann M, Schmidt V. Stochastic 3D modeling of amorphous microstructures - a powerful tool for virtual materials testing. Proceedings of the VII European Congress on Computational Methods in Applied Sciences and Engineering. 2016;Paper-ID 8172.

[14] Chiu SN, Stoyan D, Kendall WS, Mecke J. Stochastic Geometry and its Applications. Chichester: J. Wiley & Sons, 3rd ed. 2013.

[15] Molchanov I. Statistics of the Boolean Model for Practitioners and Mathematicians. Chichester: J. Wiley & Sons. 1997.

[16] Scholz C, Wirner F, Klatt MA, Hirneise D, Schröder-Turk GE, Mecke K, Bechinger C. Direct relations between morphology and transport in Boolean models. Physical Review E. 2015;92(4):043023.

[17] Stenzel O, Hassfeld H, Thiedmann R, Koster LJA, Oosterhout SD, van Bavel SS, Wienk MM, Loos J, Janssen RAJ, Schmidt V. Spatial modeling of the 3D morphology of hybrid polymer-ZnO solar cells, based on electron tomography data. The Annals of Applied Statistics. 2011; 5:1920–1947.

[18] Soille P. Morphological Image Analysis: Principles and Applications. New York: Springer. 2003.

[19] Münch B, Holzer L. Contradicting geometrical concepts in pore size analysis attained with electron microscopy and mercury intrusion. Journal of the American Ceramic Society. 2008;91(12):4059–4067.

[20] Friedman J, Hastie T, Tibshirani R. The Elements of Statistical Learning. New York: Springer, 2nd ed. 2008.

[21] MATLAB 2015b. The MathWorks. wwwmatlabcom. 2015;.

[22] Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks. 1994; 5(6):989–993.

[23] Nguyen D, Widrow B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights.

In: IJCNN International Joint Conference on Neural Networks. IEEE. 1990; pp. 21–26.

[24] Breiman L. Random forests. Machine learning. 2001;45(1):5–32.

[25] Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2(3):18–22.

[26] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2015.

[27] Pecho OM, Stenzel O, Gasser P, Neumann M, Schmidt V, Hocker T, Flatt RJ, Holzer L. 3D microstructure effects in Ni-YSZ anodes: Prediction of effective transport properties and optimization of redox-stability. Materials. 2015;8(9):5554–5585.

[28] Holzer L, Stenzel O, Pecho OM, Ott T, Boiger G, Gorbar M, De Hazan Y, Penner D, Schneider I, Cervera R, Gasser P. Fundamental relationships between 3D pore topology, electrolyte conduction and flow properties: Towards knowledge-based design of ceramic diaphragms for sensor applications. Materials & Design. 2016;99:314–327.