# Stochastic modeling and predictive simulations for the microstructure of organic semiconductor films processed with different spin coating velocities

**D Westhoff[1], J J van Franeker[2,3], T Brereton[1], D P Kroese[4], R A J Janssen[2] and V Schmidt[1]**

[1]Institute of Stochastics, Ulm University, Germany
[2]Molecular Materials and Nanosystems, Eindhoven University of Technology, The Netherlands
[3]Dutch Polymer Institute, Eindhoven, The Netherlands
[4]Department of Mathematics, The University of Queensland, Australia

E-mail: `daniel.westhoff@uni-ulm.de`, `J.J.v.Franeker@tue.nl`,
`timothy.brereton@uni-ulm.de`, `kroese@maths.uq.edu.au`,
`R.A.J.Janssen@tue.nl`, and `volker.schmidt@uni-ulm.de`

**Abstract.** A parametric stochastic model of the morphology of thin polymer:fullerene films is developed. This model uses a number of tools from stochastic geometry and spatial statistics. The fullerene-rich phase is represented by random closed sets and the polymer-rich phase is given by their complement. The model has three stages. First, a point pattern is used to model the locations of fullerene-rich domains. Second, domains are formed at these points. Third, the domains are rearranged to ensure a realistic configuration. The model is fitted to polymer:fullerene films produced using seven different spin coating velocities and validated using a variety of morphological characteristics. The model is then used to simulate morphologies corresponding to spin velocities for which no empirical data exists. The viability of this approach is demonstrated using cross-validation.

## 1. Introduction

Organic electronic devices have a number of attractive properties. They are cheap to produce and are also very flexible, which means they have a large number of potential applications; see, e.g., [1–4]. As a result, they have been the subject of considerable research. One of the barriers to developing a full understanding of organic electronic devices is that they can have highly irregular morphologies, which play a large role in determining their functional properties, such as charge generation and transport properties; see, e.g., [5] and [6]. Thus, in order to better understand and improve the performance of organic electronic devices, it is clearly desirable to carry out detailed empirical investigations of the relationship between processing parameters and the key morphological characteristics of the organic layer, which in turn relates to the device performance. It is also desirable to predict morphology and device performance from processing parameters directly. This is very difficult, however, as it is expensive and time consuming to obtain detailed data about material microstructures experimentally.

Stochastic modeling is a solution to the problem of generating sufficient data about material microstructures. It has been successfully used to investigate relationships between microstructure characteristics and functionality in a large number of settings; see, e.g., [7–9]. The idea is to develop a parametric stochastic model that describes key features of the microstructure of interest and that can be easily fitted to microstructure data obtained experimentally. The stochastic model can then be used to generate additional realizations of these microstructures. It is often possible to derive relatively simple empirical relationships between model parameters and processing parameters. Then, the parameters of the stochastic model can be systematically varied to produce realizations of material microstructures produced under different production settings. This allows microstructures to be generated even when the corresponding material has not (yet) been physically produced. It is then possible to search for optimal constellations of production parameters, a procedure called virtual materials design.

In this paper, we develop a stochastic model for the simulation of the morphologies of thin polymer:fullerene films, which are composed of fullerene-rich domains in a polymer-rich matrix. The films are produced using spin coating and the spin coating velocity has a significant influence on the microstructure. We fit the stochastic model to image data from real samples produced under seven different scenarios, corresponding to seven different choices of spin coating velocities. The stochastic model uses a point process to model the midpoints of fullerene-rich domains. The domains themselves are represented by unions of circles with random radii.

Having developed and verified the model, we then derive empirical relationships between model parameters and spin velocity, which is a key production parameter. Using these relationships, we demonstrate that our model can be used to create virtual thin-film morphologies. That is, morphologies corresponding to spin velocities for which we have no empirical data. We verify the effectiveness of this approach using cross-validation.

The paper is organized as follows. We describe the material and the empirical data sets in Section 2. We then develop the stochastic model in Section 3 and carry out validation of the model in Section 4. Finally we describe the process by which we generate virtual materials, and verify the output using cross-validation, in Section 5.

## 2. Description of material, imaging technique and binarization

The polymer:fullerene devices considered in this paper are composed of thin-films of a blend of two materials. The films are produced using spin coating. In this process, a polymer is mixed with a fullerene derivative in a common solvent, applied on a substrate, and spun at a certain rate during which the solvent evaporates. At a late stage of drying, the initial homogenous mix of solvent, polymer, and fullerene splits via spinodal demixing into two separate liquid phases. The first is a continuous, polymer-rich phase and the second phase an almost pure solution of the fullerene derivative. The two phases remain until the film has dried, resulting in droplet-shaped domains of the fullerene derivative in a polymer-rich continuous matrix. The size of these fullerene-rich domains is highly dependent on the spin coating velocity because the solvent evaporation rate changes with spin speed. The size of the domains formed during spin coating originates primarily from two effects. One is the characteristic length scale of spinodal demixing and the second is the subsequent growth of the domains by diffusion of fullerene until the film solidifies. For the material considered in this paper, cross-sectional transmission electron microscopy (TEM) images show that the droplet-like fullerene-rich domains extend through the whole film thickness, surrounded only by a thin polymer skin-layer, see [10]. Therefore, the most important morphological characteristics are contained in the 2-dimensional TEM images shown in Figure 1. This is why, in practice, each film can be considered as a single layer of fullerene-rich domains, which we call 'particles' in the following. For more information about polymer:fullerene blends, spin coating and phase separation; see, e.g., [10] and [11].

In this paper, the data sets we work with are images of thin films that consist of the polymer diketopyrrolopyrrole-quinquethiophene copolymer (PDPP5T) and the fullerene derivative [6,6]-phenyl-C71-butyric acid methyl ester ([70]PCBM) in a 1:2 weight ratio. Chloroform is used as a solvent. We consider thin films produced under seven different spin coating velocities, $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. In order to ensure that the data is representative, six different images were taken from different locations in one sample for each spin velocity. The resulting data sets are in the form of 2-dimensional gray scale images. The resolution of these images is $1024 \times 1024$ pixels. Each pixel corresponds to an area of approximately 11.75 nm$^2$. The imaging was done using low-magnification TEM on a Tecnai G2 Sphera TEM (FEI), which was operated at 200 kV. Figure 1 shows one image for each spin velocity.

In order to model the fullerene-rich domains, we first identify their positions and sizes. The image data is binarized using the Fiji software package [12]. A Gaussian
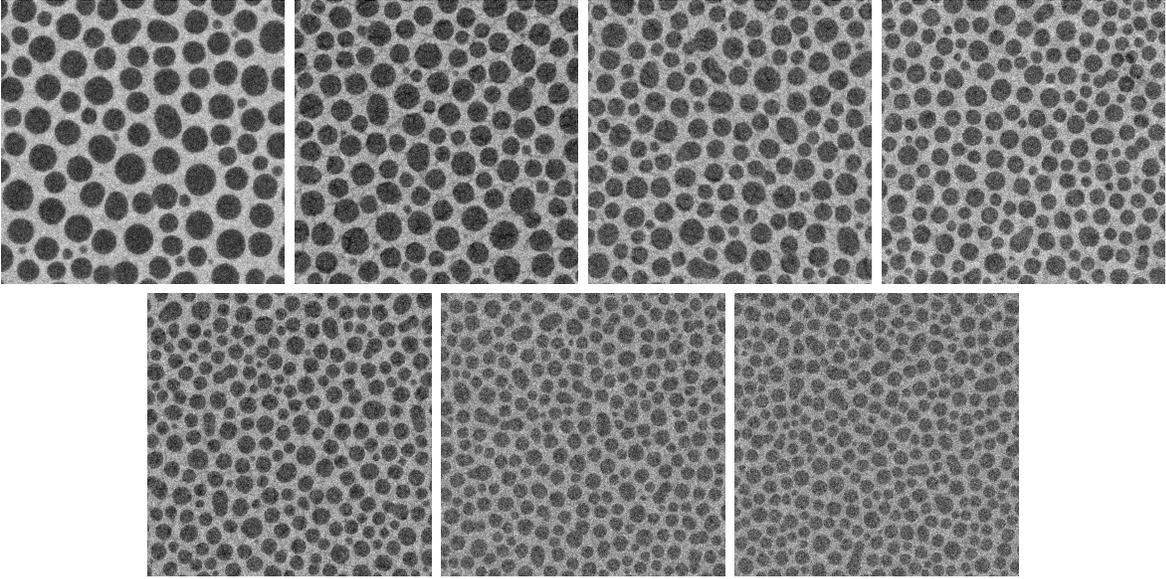
Figure 1: Experimental data sets. The darker areas represent the PCBM-rich domains and the brighter areas represent the polymer-rich matrix. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. The depicted area is $3.51 \times 3.51 \mu m$. The thin dark lines that can be discerned in some of the images and that seem to connect the PCBM clusters originate from height variations that occur at the bottom side of the film, possibly as a consequence of dewetting. The lines continue over the PCBM domains and are not related to the droplet formation as they can also be seen in films of pure polymer.

blur filter with standard deviation 3 is applied two times; see [13] for more information. Then, a global threshold for binarization is found using a modification of the isodata algorithm, where an iterative procedure is used to obtain the optimal threshold; see, e.g., [14]. The binarized data sets are displayed in Figure 2. The particles are then extracted via cluster detection using the Hoshen-Kopelman algorithm; see [15]. The location of each particle is given by its center of gravity and the size of each particle is measured by the number of pixels in the corresponding cluster. Figure 3 shows the centers of gravity of the particles shown in Figures 1 and 2.

## 3. Stochastic modeling approach

Our model of the polymer:fullerene films is two-dimensional, as we assume that the particles exist in a single layer. Because the model is only composed of two phases, it is sufficient to model one of the phases. We model the fullerene-rich domains in a bounded window, $W \subset \mathbb{R}^2$, with the complement representing the polymer-rich matrix. We represent the fullerene-rich domains as either single circles or unions of two equally sized circles. Each circle is described by a midpoint and a radius. Thus, the final output of our model is of the form $\{(S_i, R_i)\} \cup \{(S_j^{(1)}, R_j') \cup (S_j^{(2)}, R_j')\}$, where the $\{S_j\}$, $\{S_j^{(1)}\}$ and $\{S_j^{(2)}\}$ are the circle midpoints and the $\{R_i\}$ and $\{R_j'\}$ are the circle radii.

Our model has three stages. We first model the particle locations using a planar random point process; see Section 3.1. Having modeled the particle locations, we then assign a preliminary size to each particle. The particle sizes depend on the distances between neighboring particle midpoints. Based on the particle sizes and locations, we
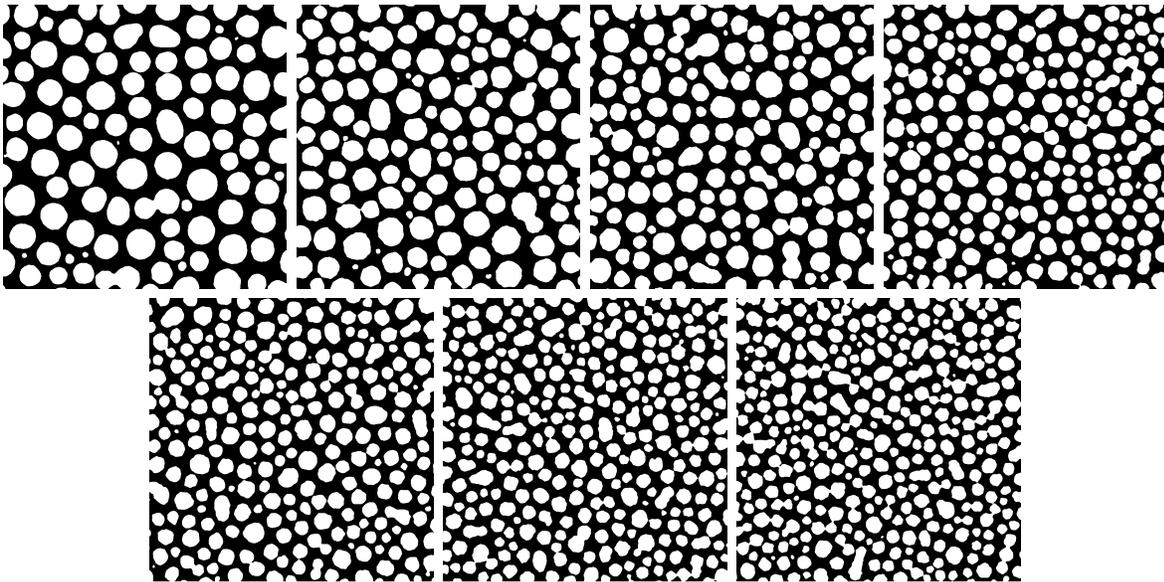
Figure 2: The binarized images corresponding to the images in Figure 1. White represents the fullerene-rich domains and black represents the polymer-rich matrix. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. The depicted area is $3.51 \times 3.51\mu m$.
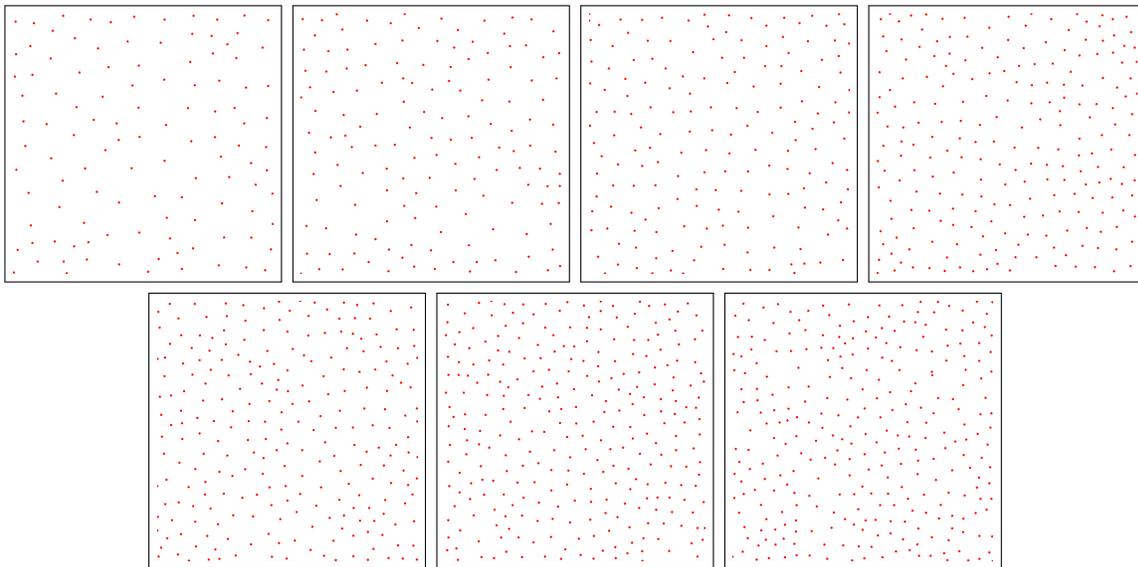


Figure 3: The centers of gravity of the particles shown in Figures 1 and 2. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. The depicted area is $3.51 \times 3.51\mu m$.

assign shapes to each particle. This may require changing the sizes of some particles. The procedure for generating the particle sizes and shapes is described in Section 3.2. Our procedure results in a configuration of particles that may overlap slightly. As we do not wish to allow overlapping particles, we then run an algorithm to iteratively change the configuration until no overlaps occur. This is described in Section 3.3.

### 3.1. Modeling the particle locations

Figure 3 shows the particle locations in seven of the experimental data sets, each corresponding to a different spin velocity. Note that there is a minimum distance between the points. This is because the particles cannot overlap. However, the distances between points vary considerably, as the particle sizes are not constant. The range of distances between particles also changes with the spin velocity. Note that the points are quite regular (that is, they are spread fairly evenly through space).

### 3.1.1. Hardcore point process with random radii

We use a hardcore point process with random radii to model the particle centers. This ensures that there is a minimal distance between neighboring points, but allows this distance to fluctuate. Our construction uses a generalization of a Matérn type II hardcore process in $\mathbb{R}^2$; see, e.g., [16]. We take a homogeneous Poisson process, $\{\widetilde{S}_j\}$, in $\mathbb{R}^2$ with intensity $\lambda > 0$. We then delete points as follows. Each point in $\{\widetilde{S}_j\}$ is marked with two independent and identically distributed (iid) marks: a random non-negative hardcore radius, $\widetilde{H}_j$, and an 'arrival time', $\widetilde{U}_j$, which is uniformly distributed on the interval $(0, 1)$. This results in a process $\widetilde{P} = \{(\widetilde{S}_j, \widetilde{H}_j, \widetilde{U}_j)\}$. A point, $\widetilde{S}_j$, is regarded as too close to another point and, thus, deleted if

$$\|\widetilde{S}_j - \widetilde{S}_k\| < \max\{\widetilde{H}_j, \widetilde{H}_k\} \text{ and } \widetilde{U}_j > \widetilde{U}_k \text{ for some } (\widetilde{S}_k, \widetilde{H}_k, \widetilde{U}_k) \in \widetilde{P},$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^2$. If we consider the points remaining in $\widetilde{P}$ after thinning, and record only their locations and radii, we have a new process, $P = \{(S_i, H_i)\}$. We can think of this, equivalently, as a random collection of circles in $\mathbb{R}^2$.

Because $P$ is made by thinning a Poisson process, it is not regular enough for our purposes. In particular, it is possible that large regions of space will not have any points in them. We correct for this by iteratively adding points to a replicate of $P$ as follows. We consider a sequence of independent replicates of $P$, $\{P^{(n)}\}_{n=1}^{\infty}$, restricted to a bounded simulation window $\widetilde{W} \subset \mathbb{R}^2$, that is considerably larger than $W$, the final sampling window. The radii, $\{H_j\}$, of the $\{P^{(n)}\}$ are normally distributed with mean $\mu_H$ and variance $\sigma_H^2$. Because the distances between particles are non-negative and because we want to avoid edge effects caused by very large particles outside $W$, the radii are truncated to the interval $(0, 2\mu_H)$. We begin by placing all the points from $P^{(1)}$ in the final point pattern. That is, we set $P^{\mathsf{fin}} = P^{(1)}$. We then add points from $P^{(2)}$ to $P^{\mathsf{fin}}$ in order to fill up the empty space. In order to preserve the hardcore properties of $P^{\mathsf{fin}}$, we only add a point, $(S_j, H_j) \in P^{(2)}$, to $P^{\mathsf{fin}}$ if

$$\|S_j - S_i\| > \max\{H_j, H_i\} \text{ for all } (S_i, H_i) \in P^{\mathsf{fin}}. \tag{1}$$

We proceed to add points from $P^{(3)}$ to $P^{\mathsf{fin}}$ so long as they satisfy condition (1). We continue in this manner, adding points from the $\{P^{(n)}\}$, until 99.9% of the simulation

window $\widetilde{W}$ is covered by the circles defined by the locations and radii of the points in $P^{\mathsf{fin}}$. Finally, we take the locations of the points in $P^{\mathsf{fin}}$, $\{S_i\}$, as the particle loctions. This process is illustrated in Figure 4.
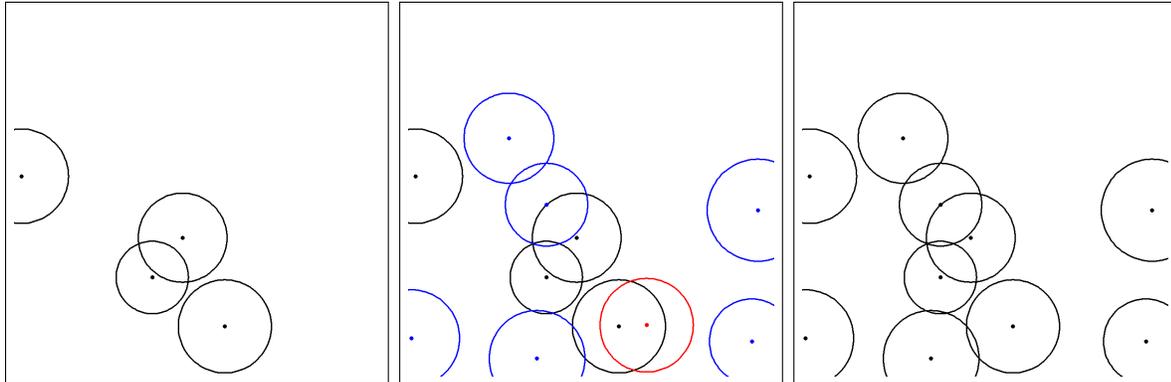


(a) Left: The first point pattern is generated. Center: The second point pattern is generated and each point is checked to make sure that it satisfies (1). The points to be added are shown in blue and those that will not are shown in red. Right: The point pattern after the points are added.
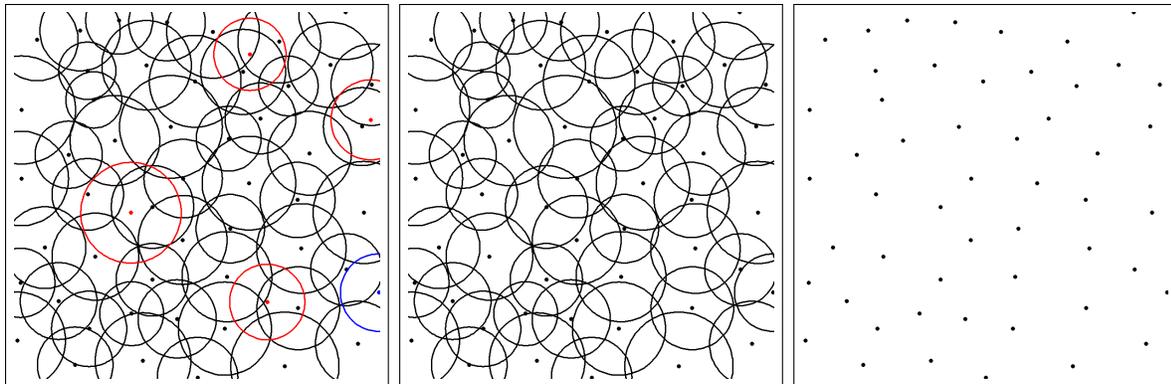


(b) Left: This procedure is continued many times. Center: We stop adding points when 99.9% of the simulation window is covered by circles. Right: The center points of the circles are kept as the particle locations.

Figure 4: An illustration of the iterative approach to generating the particle locations.

### 3.1.2. Fitting the particle location model

Our model for the particle locations has three parameters, $\mu_H, \sigma_H^2$ and $\lambda$. The intensity $\lambda$ of the Poisson process used to build the hardcore processes does not noticeably influence the final output of the model, so we set it rather arbitrarily to $\lambda = 5/\nu_2(W)$, where $\nu_2$ is the two-dimensional Lebesgue measure. In order to estimate $\mu_H$ and $\sigma_H^2$, which do play key roles, we consider the relationship between these parameters and the nearest neighbor distances within the resulting point pattern. More precisely, for each $S_i$ in $\{S_i\}$, we define the nearest neighbor distance

$$N_i = \min_{j \neq i} \|S_i - S_j\|.$$

There is clearly a close relationship between the nearest neighbor distances and the values of $\mu_H$ and $\sigma_H^2$ used to generate the points. In particular, a large $\mu_H$ should result

in a large expected value of $N_i$ and a large value of $\sigma_H^2$ should result in a large variance for $N_i$.

We observe nearest neighbor distances in the empirical data rather than $\mu_H$ and $\sigma_H^2$. More precisely, after appropriate edge correction, we observe $\{n_i\}$, a realization of $\{N_i\}$ in a subset of $W$. We wish to use information from $\{n_i\}$ to estimate $\mu_H$ and $\sigma_H^2$. It turns out that both $\mu_H$ and $\sigma_H^2$ are well estimated by affine combinations of $\bar{n}$, the empirical mean of the $\{n_i\}$ and $s_n^2$, the empirical variance of the $\{n_i\}$. That is,

$$(\mu_H, \sigma_H^2)^\intercal \approx A \cdot (\bar{n}, s_n^2)^\intercal + (b_1, b_2)^\intercal,$$

where $A \in \mathbb{R}^{2 \times 2}$. As we do not know the values of $\mu_H$ and $\sigma_H^2$ for the empirical data, we instead use synthetic data to estimate $A$. That is, we simulate 100 realizations of our model (and, thus, 100 realizations of $\{N_i\}$), for every combination of $\mu_H \in \{40, 50, \ldots, 100\}$ and $\sigma_H^2 \in \{60, 70, \ldots, 210\}$. For each of these realizations we calculate $\bar{n}$ and $s_n^2$. Figure 5, left, shows a linearly interpolated plot of the $\mu_H$ values corresponding to the observed values of $\bar{n}$ and $s_n^2$. Figure 5, right, shows a linearly interpolated plot of the $\sigma_H^2$ values corresponding to the observed values of $\bar{n}$ and $s_n^2$. Clearly, both $\mu_H$ and $\sigma_H^2$ are well represented by a linear model depending on $\bar{n}$ and $s_n^2$. We estimate $A$ and $(b_1, b_2)^\intercal$ using least squares, resulting in an equation of the form

$$\begin{pmatrix} \widehat{\mu}_H \\ \widehat{\sigma}_H^2 \end{pmatrix} = \begin{pmatrix} 0.8001 & 0.0525 \\ -0.6151 & 1.2531 \end{pmatrix} \begin{pmatrix} \bar{n} \\ s_n^2 \end{pmatrix} + \begin{pmatrix} 5.9554 \\ 27.9910 \end{pmatrix}.$$

The estimated equations give good fits to the data, with an adjusted $r^2$ of 0.9989 for the equation describing $\mu_H$ and an adjusted $r^2$ of 0.9209 for the equation describing $\sigma_H^2$, where $r^2$ is the coefficient of determination, with values close to 1 indicating a good fit of the regression model. Table 1 shows estimates of the parameters $\mu_H$ and $\sigma_H^2$ for the different spin velocities.
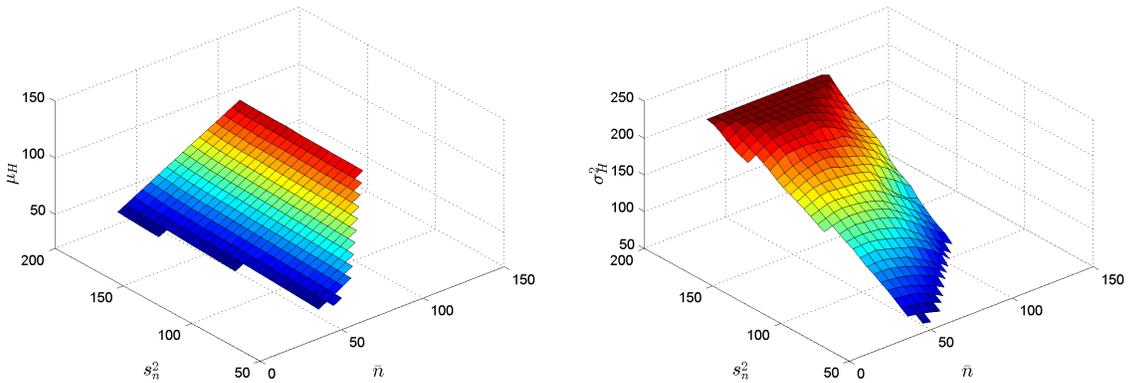


Figure 5: Left: A linearly interpolated plot of the $\mu_H$ values corresponding to the observed values of $\bar{n}$ and $s_n^2$. Right: A linearly interpolated plot of the $\sigma_H^2$ values corresponding to the observed values of $\bar{n}$ and $s_n^2$.

Table 1: Estimated hardcore radius parameters $\widehat{\mu}_H$ and $\widehat{\sigma}_H^2$ for the different spin velocities.

|  | 500rpm | 750rpm | 1000rpm | 1500rpm | 2000rpm | 3000rpm | 4000rpm |
|---|---|---|---|---|---|---|---|
| $\widehat{\mu}_H$ | 85.48 | 75.03 | 68.32 | 62.16 | 57.26 | 50.92 | 49.25 |
| $\widehat{\sigma}_H^2$ | 221.78 | 169.69 | 133.45 | 107.82 | 90.07 | 73.12 | 77.79 |

### 3.2. Modeling the size and shape of the particles

A distinctive feature of the data sets shown in Figures 1 and 2 is that all the particles seem to be well represented either by a single circle or by the union of two circles. This suggests that the particle shapes can be modeled by circles and unions of circles.

The locations of the particle centers also provide considerable information about the size and shape of the particles. In particular, if two particle centers are close together, then the sizes of the corresponding particles cannot be too large. In order to use this information, we consider the Voronoi tessellation induced by the particle centers. Given a point pattern $\{s_i\}$ in $\mathbb{R}^d$, a Voronoi tessellation divides the space into cells $\{C_i\}$, where the $i$th cell, $C_i$, corresponds to the $i$th point, $s_i$. The cells are defined such that $C_i$ contains all points that are closer to $s_i$ than to any other point of the point pattern. That is,

$$C_i = \{x \in \mathbb{R}^d, \|x - s_i\| \leq \|x - s_j\| \text{ for each } j \neq i\}.$$

For more information on Voronoi tessellations, see [17].
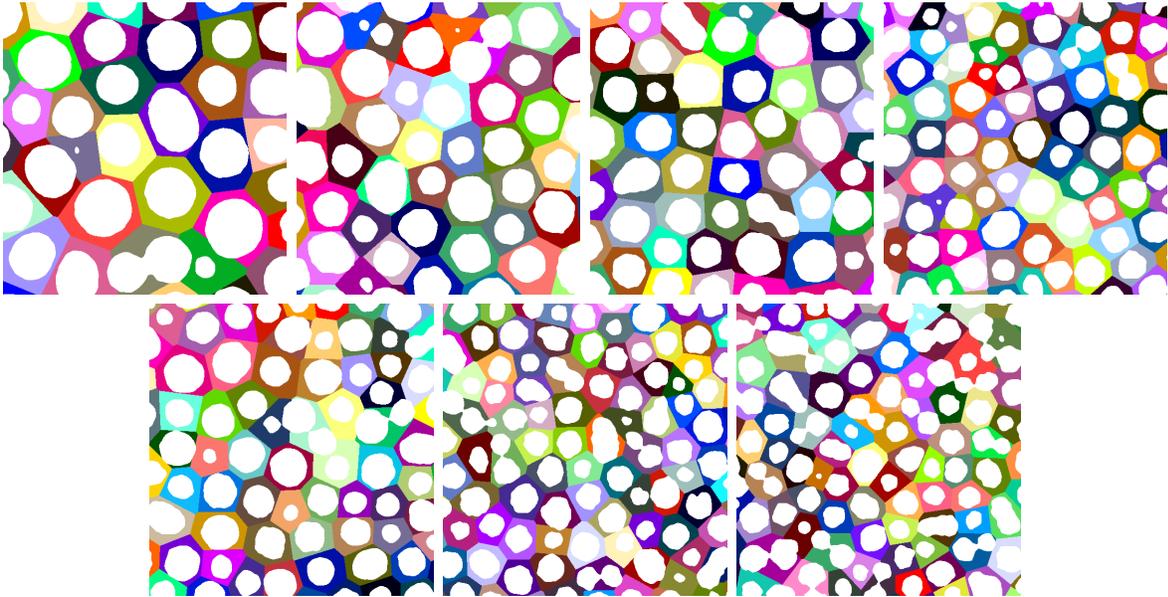


Figure 6: Voronoi tessellation on cutout of particle midpoints. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. The depicted area is about $1.9 \times 1.9 \mu m$.

Figure 6 depicts cutouts of the Voronoi tessellations corresponding to the experimental data shown in Figures 1, 2 and 3. Our approach to modeling particle size

and shape is based on three key observations. First, the particles are largely contained within their corresponding Voronoi cells. Second, those particles that extend beyond their Voronoi cells are best described using unions of circles, rather than single circles. Third, there is a strong relationship between the size of a Voronoi cell and the size of the corresponding particle; see Figure 7. These observations lead to a two step procedure for assigning sizes and shapes to the particles. In the first step, we treat all particles as circles and assign them random radii that depend on inter-particle distances. In the second step, we split some of the particles into unions of two equi-sized circles.

### 3.2.1. Assigning initial sizes to the particles

Based on the observations made above, we assign sizes to the particles that depend on the areas of their corresponding Voronoi cells. We measure the size of the particles in terms of their equivalent radii (where the equivalent radius of an object is the radius of a circle with the same area). Figure 7 shows a plot of the equivalent radii of the particles in the empirical data, $\{r_i\}$, against the logarithms of the areas of their corresponding Voronoi cells, $\{a_i\}$. These plots suggest that the radii can be described by linear regression models of the form

$$R_i = c_R + \beta_R \log a_i + \varepsilon_i, \tag{2}$$

where the $\{\varepsilon_i\}$ are assumed to be iid normal random variables with mean zero and variance $\sigma_R^2$. We estimate the values of $\beta_R$, $c_R$ and $\sigma_R^2$ for the various spin coating velocities using least-squares. The resulting estimates $\widehat{c}_R$, $\widehat{\beta}_R$ and $\widehat{\sigma}_R^2$ are shown in Table 2.
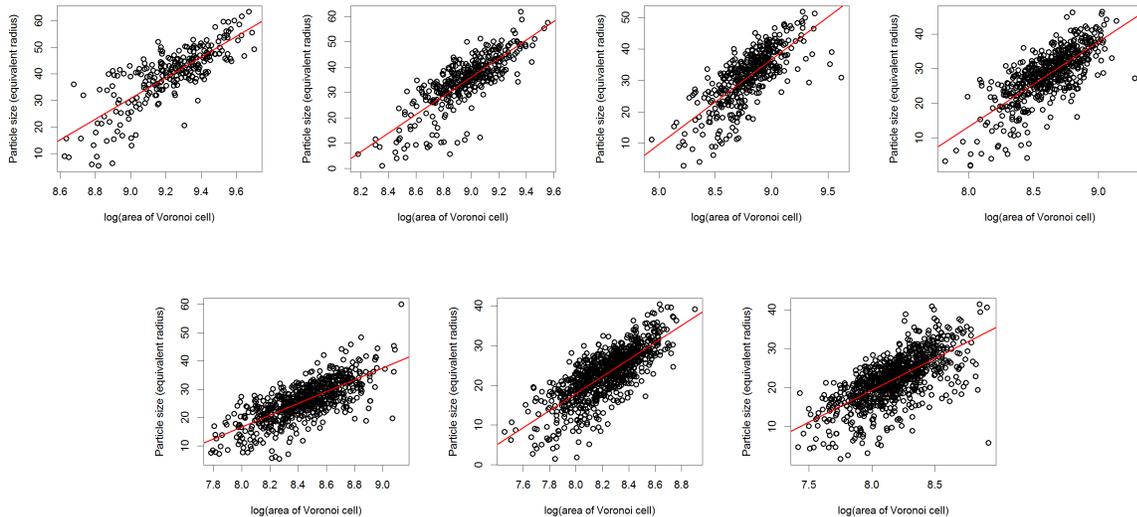


Figure 7: The equivalent radii of the particles plotted against the logarithms of the areas of the corresponding Voronoi cells, with lines of best fit shown. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

Table 2: Estimates of the intercept, $c_R$, and slope, $\beta_R$, together with the variance of the $\{\varepsilon_i\}$, $\sigma_R^2$, for a simple linear regression between logarithm of Voronoi cell size and particle size.

|  | 500rpm | 750rpm | 1000rpm | 1500rpm | 2000rpm | 3000rpm | 4000rpm |
|---|---|---|---|---|---|---|---|
| $\widehat{c}_R$ | -319.31 | -295.96 | -206.523 | -182.75 | -151.02 | -154.09 | -111.43 |
| $\widehat{\beta}_R$ | 38.89 | 36.88 | 27.02 | 24.50 | 20.95 | 21.50 | 16.34 |
| $\widehat{\sigma}_R^2$ | 41.76 | 37.74 | 32.29 | 24.81 | 22.35 | 16.06 | 20.66 |

We simulate the particle sizes using (2). More precisely, given the simulated particle locations, $\{s_i\}$, we construct the corresponding Voronoi tessellation, with cell areas given by $\{a_i\}$. We then simulate the radii of the particles as follows. We calculate the expected radius of the $i$th particle by $\mathbb{E}R_i = \widehat{c}_R + \widehat{\beta}_R \log a_i$. We draw the particle's radius, $R_i$, from a normal distribution with mean $\mathbb{E}R_i$ and variance $\sigma_R^2$. Because the particle radii must be non-negative and because we do not wish the area of a particle to be bigger than the area of its corresponding Voronoi cell, we truncate the radii to the interval $(0, \sqrt{a_i/\pi})$. At the end of this process, we have a collection of circles $\{(S_i, R_i)\}$.

### 3.2.2. Changing the particle shapes and modifying their sizes.

It is clear from Figures 1 and 2 that some particles are better represented by a union of two circles, rather than a single circle. Thus, we need to replace some of the circular particles we have generated with particles that consist of the union of two circles. We first identify the collection of particles, $\{(S_j, R_j)\} \subset \{(S_i, R_i)\}$, that we wish to change. We take the $\{(S_j, R_j)\}$ to be those circles which are not entirely contained within their corresponding cells, as this coincides with what is seen in the empirical data sets. Having identified the particles that we will replace, we proceed as follows. We replace each $(S_j, R_j)$ by a union of two circles $(S_j^{(1)}, R_j') \cup (S_j^{(2)}, R_j')$, where $S_j^{(1)}, S_j^{(2)}$ and $S_j$ are collinear and lie on a line parallel to the border of the Voronoi cell closest to $S_j$; see Figure 8, left. We determine $R_j'$ to be 0.95 multiplied by the distance from
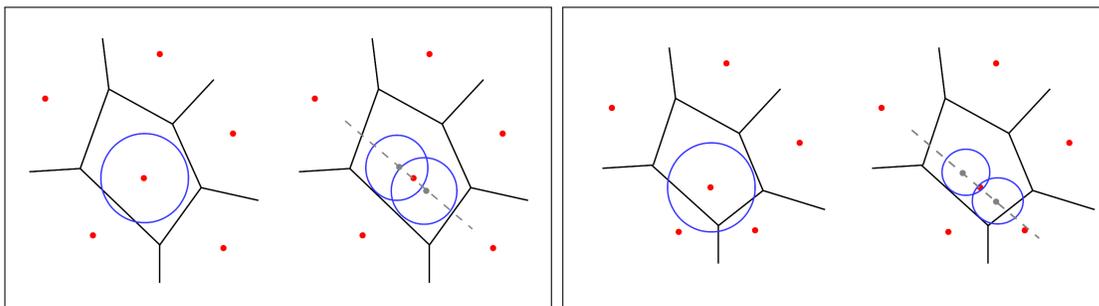


Figure 8: Schematic view of splitting algorithm. Left: Union of two circles is completely contained in the Voronoi cell. Right: Union of two circles still exceeds the Voronoi cell.

$S_j$ to its nearest edge. We place $S_j^{(1)}$ and $S_j^{(2)}$ equidistantly from $S_j$ such that the area of $(S_j^{(1)}, R_j') \cup (S_j^{(2)}, R_j')$ is the same as the area of $(S_j, R_j)$. Note that our approach

allows particles to extend beyond their Voronoi cells; see Figure 8, right. It may be the case that the area of $(S_j^{(1)}, R_j') \cup (S_j^{(2)}, R_j')$ is strictly less than the area of $(S_j, R_j)$ for all choices of $S_j^{(1)}$ and $S_j^{(2)}$. In this case, we draw $R_j$ again as in Section 3.2.1.

### 3.3. Rearranging the particles

We now have a system comprising of a set of single circles $\{(S_i, R_i)\}$ and a set of unions of circles $\{(S_j^{(1)}, R_j'), (S_j^{(2)}, R_j')\}$. However, it is possible that these objects may overlap because, when circles are split into two, the resulting object may exceed its Voronoi cell in the direction of splitting. In addition, particles can be either too close to each other or too far apart. For these reasons we use a force-biased algorithm, described in [18], to rearrange the particles until an acceptable configuration is found.

In order to quantify how close together and how far apart particles are allowed to be, we consider the nearest neighbor distances between particles. For the $i$th particle this quantity is given by $\widetilde{N}_i = \min_{j \neq i} d(P_i, P_j)$, where $d(\cdot, \cdot)$ is the minimal Euclidean distance between two particles. For each data set, we calculated $\widetilde{N}_{\min} = \min_i \widetilde{N}_i$, the minimum nearest neighbor distances between particles, and $\widetilde{N}_{\max} = \max_i \widetilde{N}_i$, the maximum nearest neighbor distances between particles. The minimum distance, $\widetilde{N}_{\min}$, is almost constant for all spin velocities, just varying between 2.0 and 3.0. Therefore, in the following, we always use the value $\widetilde{N}_{\min} = 2.5$. In contrast, the maximum distance, $\widetilde{N}_{\max}$, varies considerably. Table 3 shows the observed values of $\widetilde{N}_{\min}$ and $\widetilde{N}_{\max}$ for the different scenarios.

Table 3: Minimum and maximum distances between particles for each scenario.

|  | 500rpm | 750rpm | 1000rpm | 1500rpm | 2000rpm | 3000rpm | 4000rpm |
|---|---|---|---|---|---|---|---|
| $\widetilde{N}_{\min}$ | 3.00 | 2.83 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| $\widetilde{N}_{\max}$ | 30.00 | 24.02 | 21.95 | 22.13 | 21.02 | 18.79 | 18.00 |

The force-biased algorithm begins by placing a larger circle around each circle in the model. The radii of these outer circles are given by $R_i^{\mathsf{out}} = R_i + D_i$ for the $\{(S_i, R_i)\}$ and by $R_j^{\mathsf{out}'} = R_j' + D_j'$ for the $\{(S_j^{(1)}, R_j'), (S_j^{(2)}, R_j')\}$. To begin with, all the $\{D_i\}$ and $\{D_j'\}$ are set equal to $\widetilde{N}_{\max}$. This ensures that particles are not pushed further than $\widetilde{N}_{\max}$ away from one another. The end result of this is a system of outer circles $\{(S_i, R_i^{\mathsf{out}})\} \cup \{(S_j^{(1)}, R_j^{\mathsf{out}'}), (S_j^{(2)}, R_j^{\mathsf{out}'})\}$. At the same time, we assign a random shrinking factor, $C_i$ or $C_j'$, to each particle. These shrinking factors are drawn uniformly from $\{1/10, 2/10, \ldots, 9/10\}$. The particles are then collectively rearranged. This is done by first computing the 'forces' acting on each particle. These forces push the particle away from other particles whose inner circles overlap the outer circle(s) of the particle being considered. The forces are then added up to compute a net force acting on each particle. The vector valued net force acting on the $k$th particle in $\{(S_i, R_i)\}$, $F_k$, is given

by $F_k = F_k^{(1)} + F_k^{(2)}$, where

$$F_k^{(1)} = \sum_i \frac{R_k^{\text{out}} + R_i - \|S_k - S_i\|}{\|S_k - S_i\|}(S_k - S_i)\mathbb{1}\left(\|S_k - S_i\| < R_k^{\text{out}} + R_i\right)$$

and

$$F_k^{(2)} = \sum_j \sum_{n=1}^{2} \frac{R_k^{\text{out}} + R_j' - \|S_k - S_j^{(n)}\|}{\|S_k - S_j^{(n)}\|}(S_k - S_j^{(n)})\mathbb{1}\left(\|S_k - S_j^{(n)}\| < R_k^{\text{out}} + R_j'\right),$$

with $\mathbb{1}(\cdot)$ being the indicator function. The forces acting on the $\{(S_j^{(1)}, R_j')\}$ and $\{(S_j^{(2)}, R_j')\}$ are defined in the same manner. Then, for each $j$, the net forces acting on $\{(S_j^{(1)}, R_j')\}$ and $\{(S_j^{(2)}, R_j')\}$ are added together to give $F_j'$. As the modeling approach leads to a good starting configuration for the force-biased algorithm, we do not want to rearrange the particles too much. Thus, when a net force has a magnitude of greater than 10 pixels (about 34nm), it is scaled to a magnitude of 10 pixels. Having computed all the quantities, we then shift each particle by its corresponding net force. That is, for all $i$, we set $S_i = S_i + F_i$ and, for all $j$, we set $S_j^{(1)} = S_j^{(1)} + F_j'$ and $S_j^{(2)} = S_j^{(2)} + F_j'$. Once the particles have been rearranged, we check for overlaps of the outer circles with other circles. If there are still any overlaps, we set each $D_i = C_i D_i$ as long as it is bigger than $\widetilde{N}_{\text{min}}$ and each $D_j' = C_j' D_j'$ as long as it is bigger than $\widetilde{N}_{\text{min}}$. We then recompute the radii of the outer circles, recalculate the forces, and rearrange the particles again. We continue in this manner until no overlaps remain.

## 4. Model validation

We carry out validation of our stochastic model by comparing its output with the empirical data. We first compare the point patterns describing the particle centers in our model with those in the empirical data. We then compare key characteristics of the particles produced by our stochastic model with those observed in the empirical data.

### 4.1. Validation of point pattern model

Figure 9 shows the particle centers in seven realizations of our stochastic model, each corresponding to a different choice of spin velocity. The resulting point patterns appear to share many features with those observed in the empirical data (as shown in Figure 3). Likewise, the intensities of the points produced by the stochastic model agree closely with the intensities observed in the data sets, as can be seen in Table 4. In order to compare the point patterns more closely, we consider three second-order characteristics: the pair correlation function, the nearest neighbor distance distribution function, and the spherical contact distance distribution function. Because our model is stationary and isotropic (and the point patterns in the empirical data sets also appear to be) these characteristics are functions of one parameter. For a detailed discussion of second-order

characteristics of point processes see e.g. [16] and [19]. We estimate the characteristics for the stochastic model by taking pointwise averages based on 50 realizations of the model.
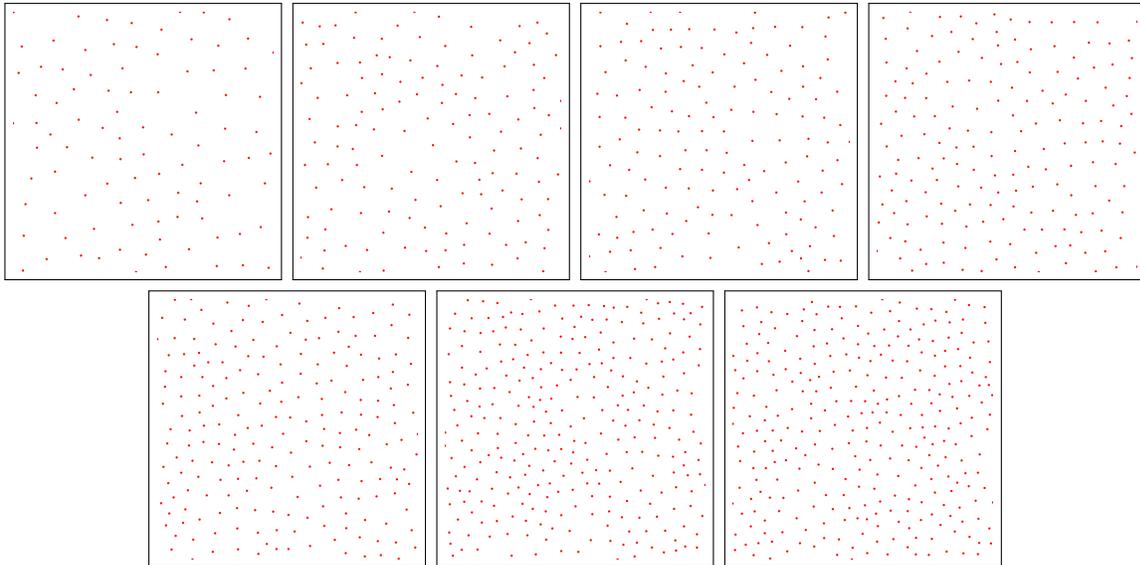


Figure 9: The centers of gravity of the particles generated by the stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. The depicted area is $3.51 \times 3.51 \mu m$.

Table 4: The estimated intensities of the particle centers for the experimental (top) and simulated (bottom) data.

|  | 500rpm | 750rpm | 1000rpm | 1500rpm | 2000rpm | 3000rpm | 4000rpm |
|---|---|---|---|---|---|---|---|
| $\lambda_E \cdot 10^4$ | 0.98 | 1.25 | 1.46 | 1.80 | 2.08 | 2.59 | 2.75 |
| $\lambda_S \cdot 10^4$ | 0.92 | 1.21 | 1.46 | 1.75 | 2.06 | 2.60 | 2.80 |

The pair correlation function, $g : [0, \infty) \to [0, \infty)$, is a measure of the distances between pairs of points. It is normalized so that a homogeneous Poisson process has $g(r) = 1$ for all $r \in [0, \infty)$. When $g(r_0) > 1$ for some $r_0 \geq 0$, it indicates that pairs of points with distance $r_0$ from one another occur more frequently than in the Poisson process and when $g(r_0) < 1$ for some $r_0 \geq 0$, it indicates that pairs of points with distance $r_0$ from one another occur less frequently than in the Poisson process. The estimated pair correlation functions for the stochastic model and empirical data sets are shown in Figure 10. The pair correlation functions are approximately zero for $r$ values in the interval $(0, 50)$, indicating that no points are closer than this distance to one another. This is a direct consequence of the non-overlapping particles. There is also a peak in each of the pair correlation functions, indicating that the spacing of points is more regular than in a Poisson process. The stochastic model clearly captures these features.

The nearest neighbor distance distribution function, $G : [0, \infty) \to [0, 1]$, is the cumulative distribution function of the distance from a point chosen at random in
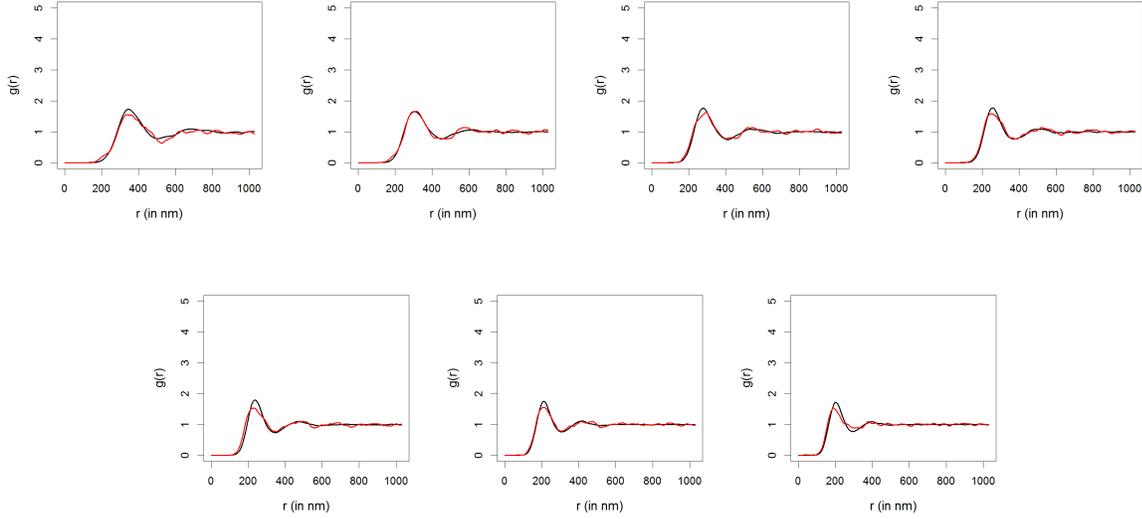
Figure 10: Pair correlation function, $g$. Red: Experimental data. Black: Stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

the point pattern to its nearest neighbor. The estimated nearest neighbor distance distribution functions are shown in Figure 11. These plots all show that the nearest neighbor of a point tends to lie at a distance of between 50 and 100. This reflects the hardcore nature of the particles. Again, the stochastic model describes this feature very well.
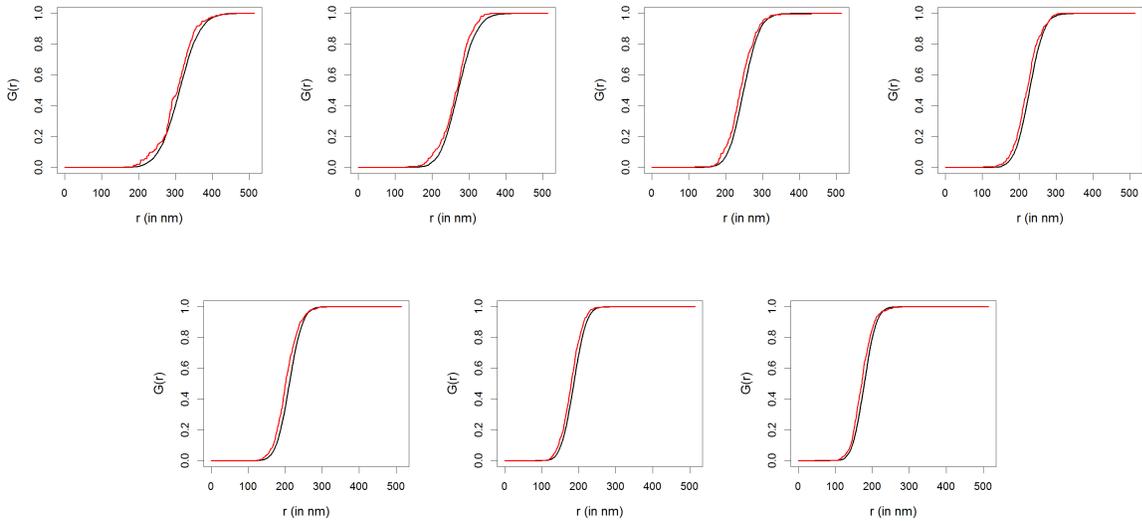


Figure 11: Nearest neighbor distance distribution function, $G$. Red: Experimental data. Black: Stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

The spherical contact distance distribution function $H : [0, \infty) \to [0, 1]$ is the cumulative distribution function of the distance from an arbitrary point in $\mathbb{R}^2$ to the

nearest point of the point pattern. The estimated spherical contact distance distribution functions are shown in Figure 12. Note that the spherical contact distance distribution functions corresponding to the stochastic model are almost identical to those of the empirical data sets.
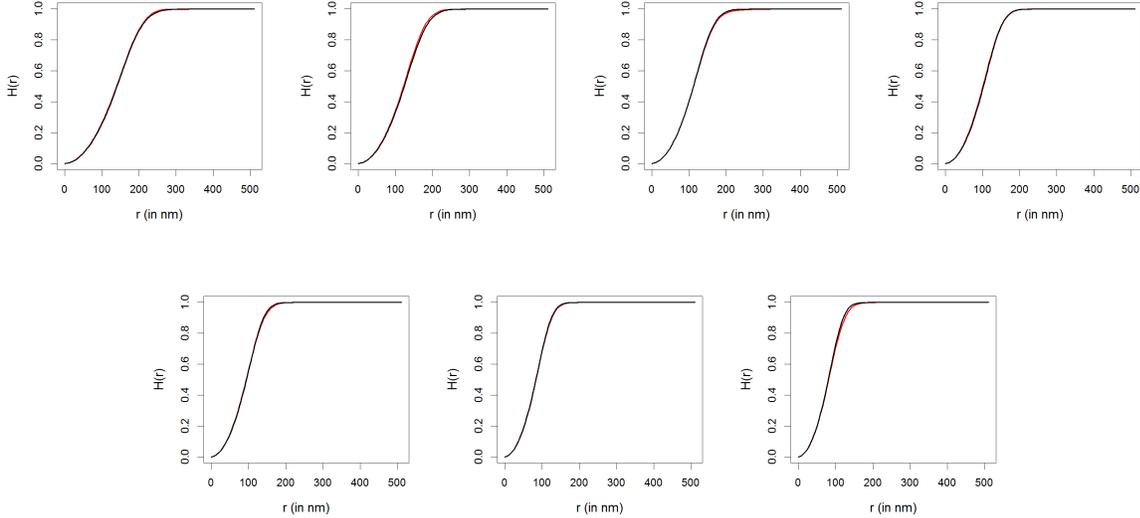


Figure 12: Spherical contact distance distribution function, $H$. Red: Experimental data. Black: Stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

## 4.2. Validation of final stochastic model

Figure 13 shows realizations of the stochastic model for each of the seven spin velocities. These are very similar to the images in Figure 2. We compare the output of the stochastic model with the empirical data by considering a number of important morphological characteristics. As above, we estimate the characteristics for the stochastic model by taking pointwise averages based on 50 realizations of the model.

We first consider the size distribution of the particles. Figure 14 shows the estimated particle size densities for the empirical data and stochastic models. Note that the stochastic model produces particles with similar sizes to those in the data sets. The stochastic model leads to unimodal particle size distributions, whereas for the experimental data we sometimes observe a further small peak or 'shoulder' for small radii. The origin and consequence of those is not yet investigated in detail, see, e.g., [11]. As we want to focus on the main structural properties of the considered thin-film morphologies, we neglect these small second peaks in our modeling approach. Analysis of further characteristics shows that this does not lead to different structural properties.

The $x$-direction chord length distribution function $C_x : [0, \infty) \to [0, 1]$ gives the distribution of the lengths of the intersections of an arbitrary horizontal line with the fullerene-rich domains. Because the stochastic model and data are both isotropic, it
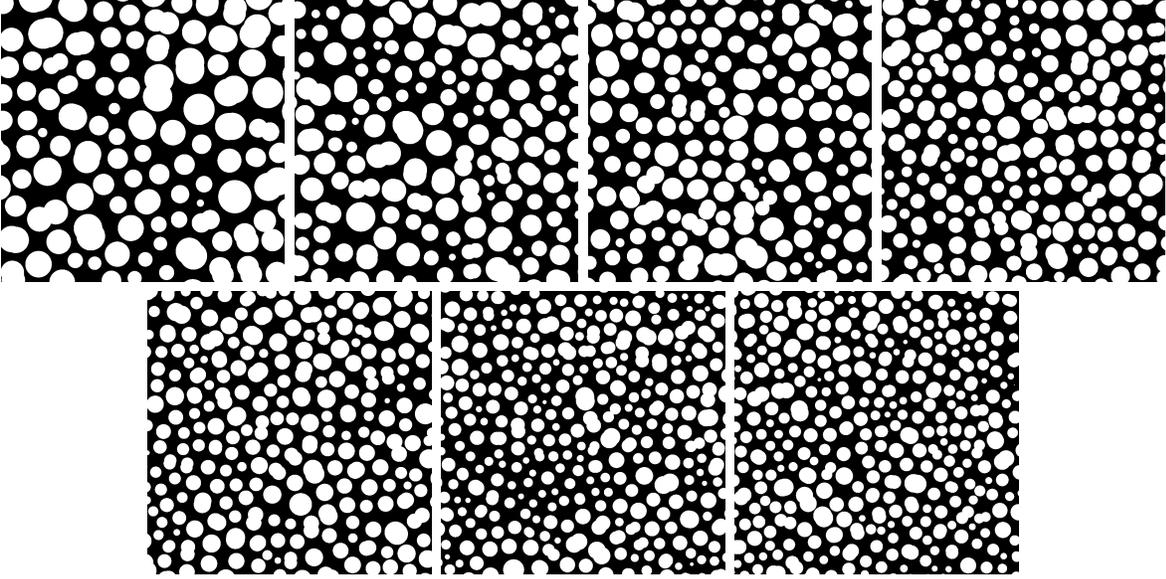
Figure 13: Realizations of the stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm. The depicted area is $3.51 \times 3.51\mu m$.
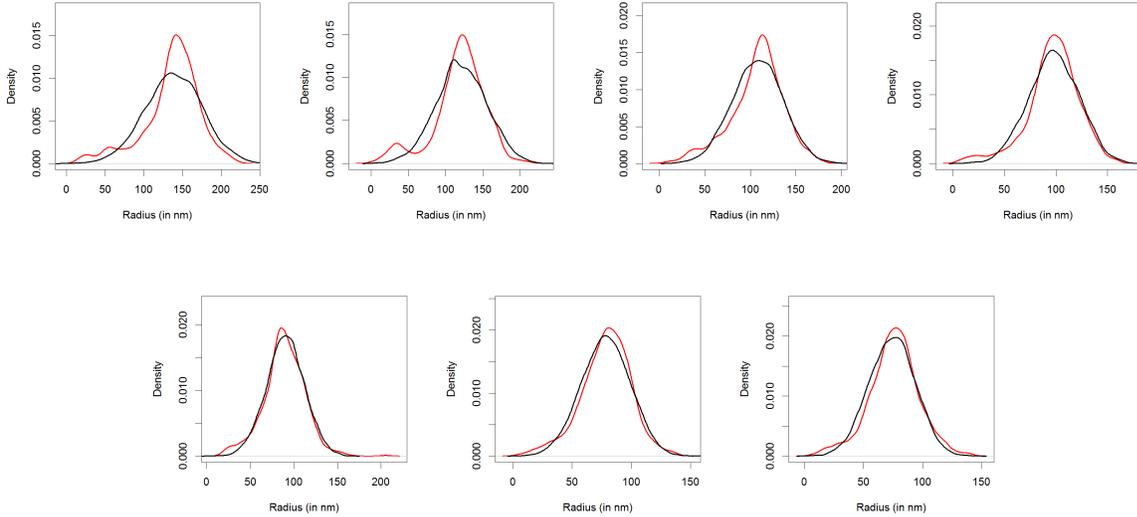


Figure 14: Particle size distribution. Red: Experimental data. Black: Stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

is sufficient to only consider horizontal lines. For more information on chord length distributions; see, e.g., [20]. The estimated chord length distribution functions are shown in Figure 15. There is almost no difference between the distribution functions estimated from the stochastic model and those estimated from the empirical data.

We also estimate the spherical contact distance distribution function, $G : [0, \infty) \to [0, 1]$, from the polymer-rich phase to the fullerene-rich phase. This gives the distribution of the distance from an arbitrary point in the polymer-rich phase to the closest point in
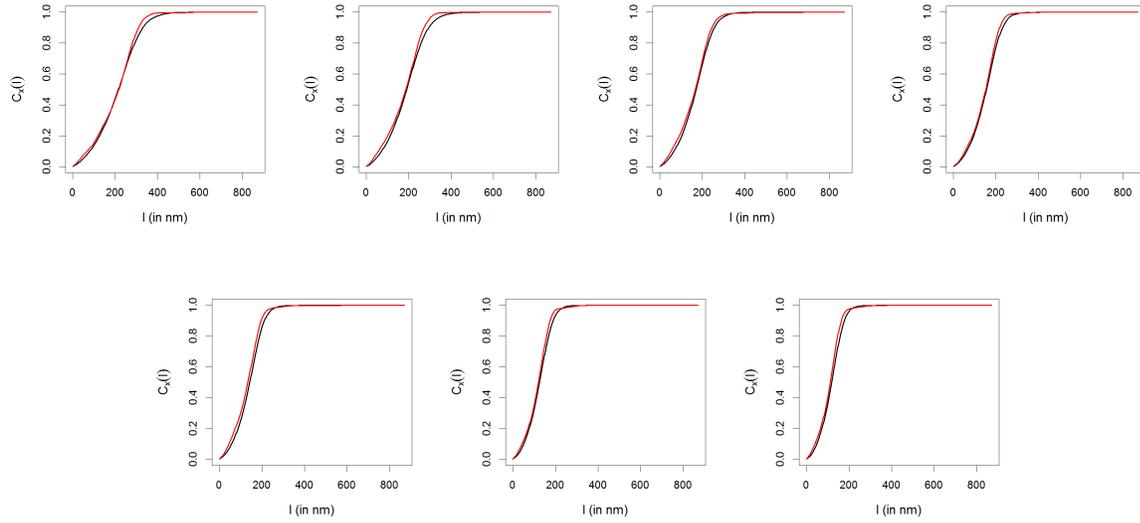
Figure 15: Chord length distribution, $C_x$. Red: Experimental data. Black: Stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

the fullerene-rich phase. For further information about this characteristic; see, e.g., [19]. The results can be found in Figure 16. A good coherence to the experimental data can be found.
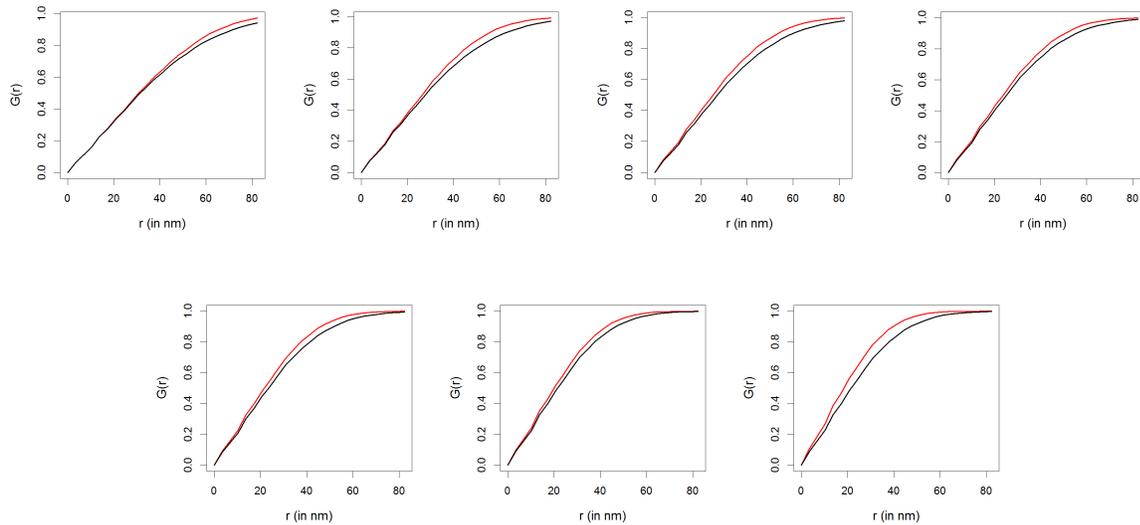


Figure 16: Spherical contact distance distributions, $G$, from the polymer-rich phase to the particle phase. Red: experimental data; black: stochastic model. Spin coating velocities from top left to bottom right: $v_1 = 500$rpm, $v_2 = 750$rpm, $v_3 = 1000$rpm, $v_4 = 1500$rpm, $v_5 = 2000$rpm, $v_6 = 3000$rpm and $v_7 = 4000$rpm.

## 5. Predictive simulations

Having developed a parametric stochastic model for polymer:fullerene films, we now wish to use this model to generate virtual microstructures of films for spin velocities that have not been physically considered (i.e., to carry out generation of 'virtual' thin-film morphologies). For this purpose, we fit equations that describe the parameters of our stochastic model in terms of spin velocity. Using these equations, we are able to choose parameter constellations that correspond to spin velocities that were not used to fit the model. This allows us to generate realizations of polymer:fullerene morphologies for which no experimental data are available. We carry out cross-validation to show that our approach generates highly realistic morphologies.

### 5.1. Fitting the predictive model

In order to be able to generate virtual morphologies, we need to express the parameters of our model as functions of spin velocity. We have estimates of each of the six parameters of our model — $\mu_H$, $\sigma_H^2$, $c_R$, $\beta_R$, $\sigma_R^2$ and $\widetilde{N}_{\mathsf{max}}$ — for seven different spin velocities. Figure 17 shows the values of the parameters as functions of the spin velocity values with curves of best fit. The curves are all of the form $c_0 \exp(\alpha_0 x) + c_1 \exp(\alpha_1 x)$ and the parameters $c_0, \alpha_0, c_1$ and $\alpha_1$ are fitted using non-linear least squares.
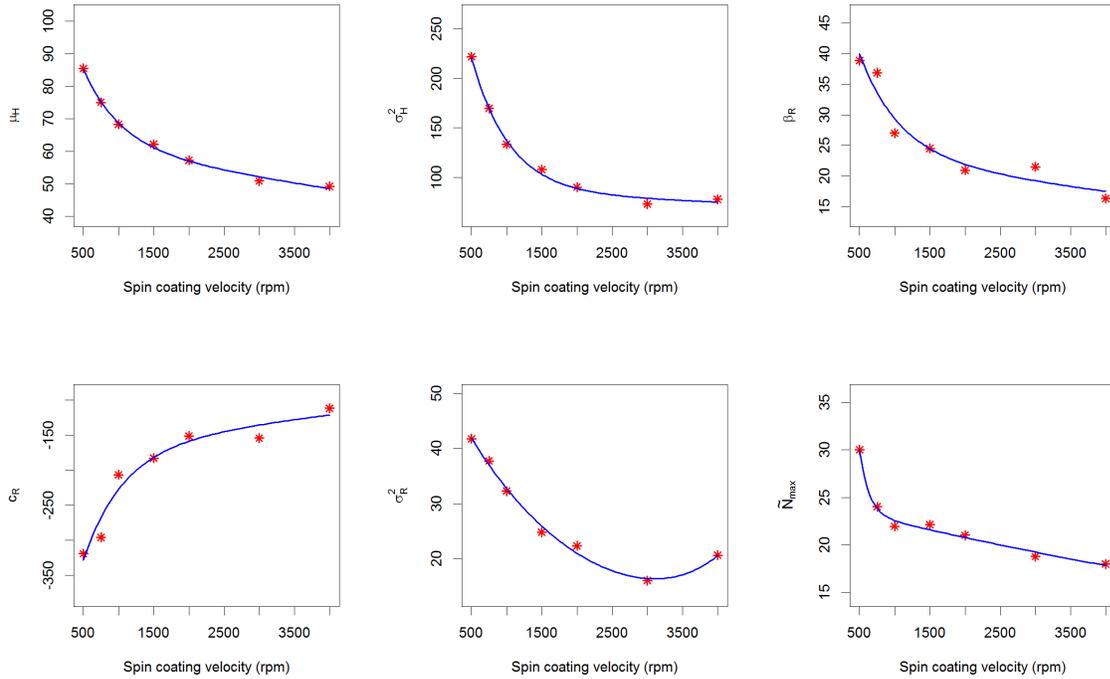


Figure 17: Plots of the parameters as functions of spin velocity with curves of best fit.

Note that the values of $\mu_H$, $\sigma_H^2$ and $\widetilde{N}_{\mathsf{max}}$ decrease with increasing spin velocities.

This is because higher spin velocities result in a larger number of smaller more tightly packed particles.
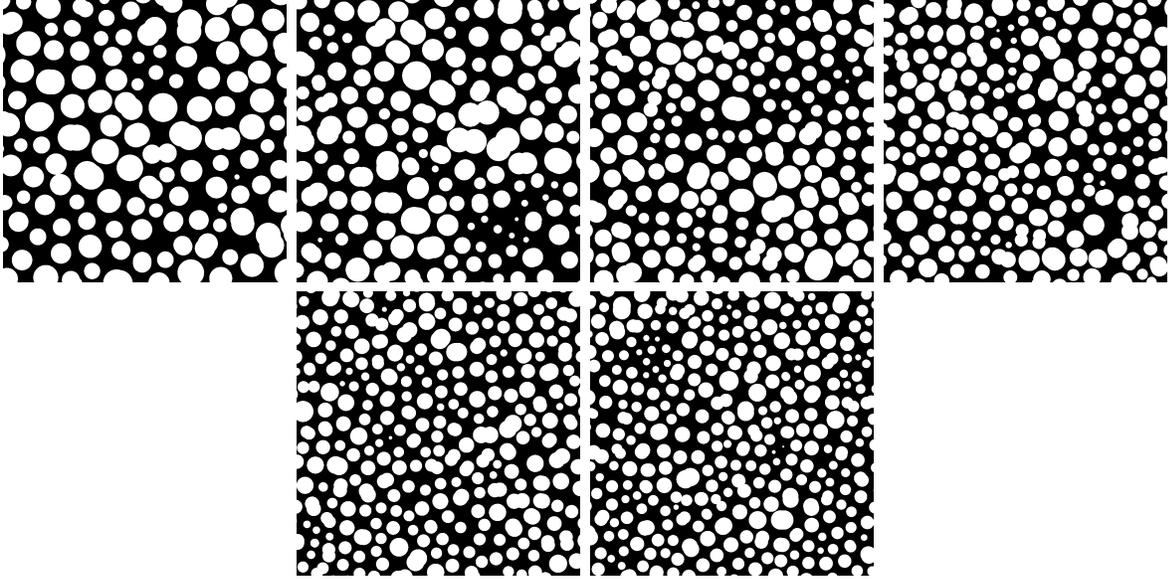


Figure 18: Predictive simulations for spin coating velocities $\tilde{v}_1 = 625$rpm, $\tilde{v}_2 = 875$rpm, $\tilde{v}_3 = 1250$rpm, $\tilde{v}_4 = 1750$rpm, $\tilde{v}_5 = 2500$rpm and $\tilde{v}_6 = 3500$rpm. The depicted area is $3.51 \times 3.51 \mu m$.

Using these fitted curves we are able to find appropriate values of the parameters for all spin velocities in the range from 500 to 4000 rpm. As an example, we use our model to generate morphologies corresponding to the spin coating velocities $\tilde{v}_1 = 625$rpm, $\tilde{v}_2 = 875$rpm, $\tilde{v}_3 = 1250$rpm, $\tilde{v}_4 = 1750$rpm, $\tilde{v}_5 = 2500$rpm and $\tilde{v}_6 = 3500$rpm. The output of the stochastic model corresponding to these spin velocities can be seen in Figure 18. Figure 19 shows the particle intensities and mean particle sizes for both the experimental data sets and the predictive simulations. Note that the values for the virtual microstructures seem consistent with the empirical values.

## 5.2. Validating the predictive model

We check the reliability of our predictive simulations using cross-validation. That is, we remove part of the empirical data, then use our model to predict this data. The results provide strong evidence that our method is able to reliably generate morphologies for a large range of spin velocities. More precisely, we remove the data points corresponding to the spin velocity $v_4 = 1500$rpm and estimate the curves describing the parameters as above. We then generate realizations of the stochastic model using the predicted parameters corresponding to a spin velocity of 1500rpm. We do the same for $v_6 = 3000$rpm. Figures 20 and 21 contain plots of a number of key morphological characteristics for the predictive simulations together with those for the empirical data and output from the stochastic model using the parameters fitted to all of the data sets. It is clear from these plots that the predictive simulations are able to accurately model key features of the corresponding morphologies.
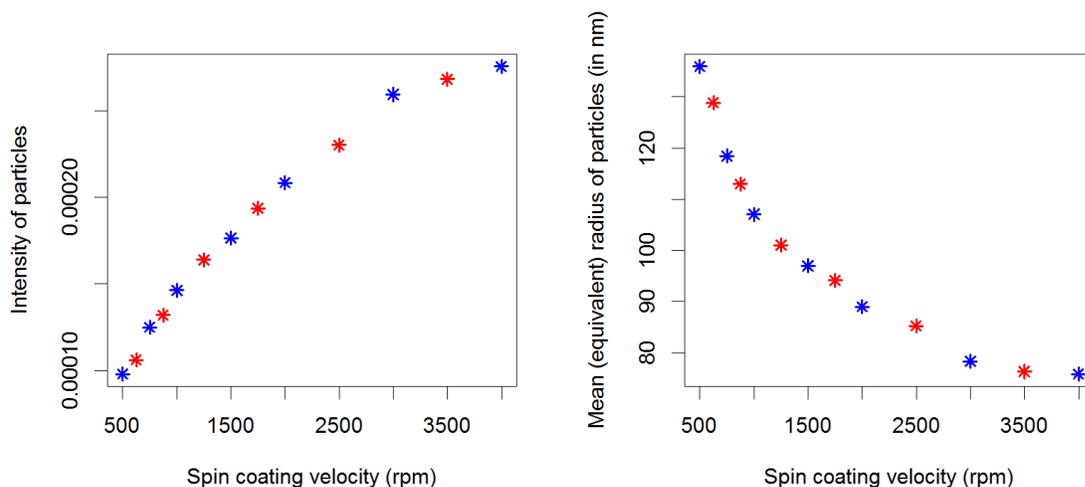
Figure 19: Intensity and mean size of particles in relation to spin coating velocities. Blue: Experimental data. Red: Predictive simulations.
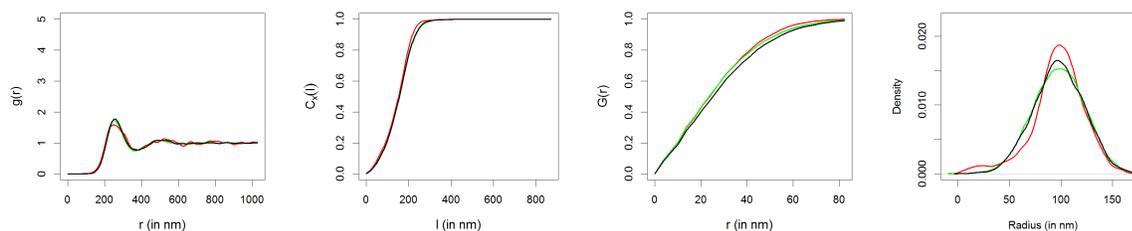


Figure 20: Characteristics of experimental and simulated morphologies corresponding to a spin velocity of 1500rpm. Red: Experimental data set. Black: Stochastic model with parameters fitted to the whole data set. Green: Stochastic model fitted without the data for $v_4 = 1500$rpm. From left to right: Pair correlation function, chord length distribution in $x$-direction, spherical contact distance distribution from polymer to particle and size distribution of particles.
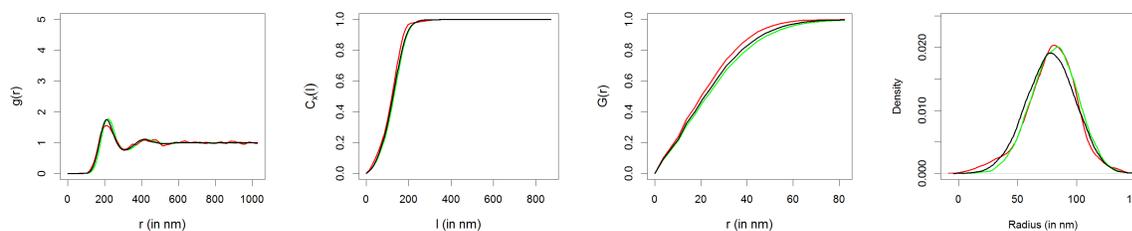


Figure 21: Characteristics of experimental and simulated morphologies corresponding to a spin velocity of 3000rpm. Red: Experimental data set. Black: Stochastic model with parameters fitted to the whole data set. Green: Stochastic model fitted without the data for $v_6 = 3000$rpm. From left to right: Pair correlation function, chord length distribution in $x$-direction, spherical contact distance distribution from polymer-rich phase to particles and size distribution of particles.

## 6. Conclusions

In this paper we have developed a parametric stochastic model for the morphology of thin polymer:fullerene films. We have fitted the parameters of this model to seven

different scenarios, corresponding to seven different spin coating velocities. We have considered a number of key morphological characteristics and shown that our model describes the data well. In addition, we have described an approach that allows us to find model parameters corresponding to arbitrary spin velocities and to use these parameters to generate virtual morphologies. We have demonstrated the viability of this approach using cross-validation.

The 2D-approach for modeling the microstructure of organic semiconductor films processed with different spin coating velocities presented in this contribution can already be used as input for realistic solar cell device modeling. The stochastic model developed in this contribution can be used to predict the radius of the PCBM-rich domains, which is important input for device models, see, e.g., [21]. In combination with a predictive model for the thickness of both the fullerene-rich domains and the polymer-rich matrix this could lead to predictive modeling tools for organic solar cells. This is why the stochastic 2D model developed in the present paper can be seen as a first step towards a corresponding microstructure model in three dimensions. The virtual (but realistic) microstructures generated by this 3D model can then be used to predict device performance, for example by numerical evaluation of quenching and charge transport properties of these virtual 3D microstructures, in dependence of the chosen spin velocity; see, e.g., [22]. This will be subject of our future research.

## Acknowledgments

## References

[1] B.C. Thompson and J.M.J. Fréchet. Polymer-fullerene composite solar cells. *Angewandte Chemie International Edition*, 47:58–77, 2008.

[2] N. Rehmann, D. Hertel, K. Meerholz, H. Becker, and S. Heun. Highly efficient solution-processed phosphorescent multilayer organic light-emitting diodes based on small-molecule hosts. *Applied Physics Letters*, 91:103507, 2007.

[3] L. Dou, J. You, Z. Hong, Z. Xu, G. Li, R. A. Street, and Y. Yang. 25th anniversary article: A decade of organic/polymeric photovoltaic research. *Advanced Materials*, 25:6642–6671, 2013.

[4] V. Khikhlovskyi, R. Wang, A. J. J. M. van Breemen, G. H. Gelinck, R. A. J. Janssen, and M. Kemerink. Nanoscale organic ferroelectric resistive switches. *The Journal of Physical Chemistry C*, 118:3305–3312, 2014.

[5] T. L. Benanti and D. Venkataraman. Organic solar cells: An overview focusing on active layer morphology. *Photosynthesis Research*, 87:73–81, 2006.

[6] P. K. Watkins, A. B. Walker, and G. L. B. Verschoor. Dynamical Monte Carlo modelling of organic solar cells: The dependence of internal quantum efficiency on morphology. *Nano Letters*, 9:1814–1818, 2005.

[7] O. Stenzel, C. Hirsch, V. Schmidt, T. Brereton, D.P. Kroese, B. Baumeier, and D. Andrienko. A general framework for consistent estimation of charge transport properties via random walks in random environments. *Multiscale Modeling and Simulation*, 12:1108–1134, 2014.

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

[9] M. Roland, A. Kruglova, G. Gaiselmann, T. Brereton, V. Schmidt, F. Muecklich, and S. Diebels. Numerical simulation and comparison of a real Al-Si alloy with virtually generated alloys. *Archive of Applied Mechanics*, Accepted, 2014.

[10] S. Kouijzer, J. J. Michels, M. van den Berg, V. S. Gevaerts, M. Turbiez, M. M. Wienk, and R. A. J. Janssen. Predicting morphologies of solution processed polymer:fullerene blends. *Journal of the American Chemical Society*, 135:12057–12067, 2013.

[11] J. J. van Franeker, D. Westhoff, M. Turbiez, M. M. Wienk, V. Schmidt, and R. A. J. Janssen. Controlling the dominant length scale of liquid-liquid phase separation in spin-coated organic semiconductor films. *Advanced Functional Materials*, in print, 2014. `doi:10.1002/adfm.201403392`.

[12] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, B. S. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9:676–682, 2012.

[13] W. Burger and M. J. Burger. *Digital Image Processing: An Algorithmic Introduction Using Java*. Springer, New York, 2010.

[14] T. W. Ridler and S. Calvard. Picture thresholding using an iterative selection method. *IEEE Transactions on Systems, Man and Cybernetics*, 8:630–632, 1978.

[15] J. Hoshen and R. Kopelman. Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. *Physical Review B*, 14:3438–3445, 1976.

[16] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modeling of Spatial Point Patterns*. John Wiley & Sons, Chichester, 2008.

[17] A. Okabe, B. Boots, K. Sugihara, and S.N. Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, Chichester, 2000.

[18] J. Mościński, M. Bargieł, Z. A. Rycerz, and P. W. M. Jacobs. The force-biased algorithm for the irregular close packing of equal hard spheres. *Molecular Simulation*, 3:201–212, 1989.

[19] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. John Wiley & Sons, Chichester, 2013.

[20] J. Ohser and F. Mücklich. *Statistical Analysis of Microstructures in Materials Science*. John Wiley & Sons, Chichester, 2000.

[21] D. Bartesaghi, M. Turbiez, and L. J. A. Koster. Charge transport and recombination in PDPP5T:[70]PCBM organic solar cells: The influence of morphology. *Organic Electronics*, 15:3191–3202, 2014.

[22] D. Bartesaghi and L. J. A. Koster. The effect of large compositional inhomogeneities on the performance of organic solar cells: A numerical study. *Advanced Functional Materials*, in print, 2014. `doi:10.1002/adfm.201402260`.