

Machine Learning for Revealing Spatial Dependence among Nanoparticles: Understanding Catalyst Film Dewetting via Gibbs Point Process Models

Ahmed Aziz Ezzat and Mostafa Bedewy*

Cite This: *J. Phys. Chem. C* 2020, 124, 27479–27494

Read Online

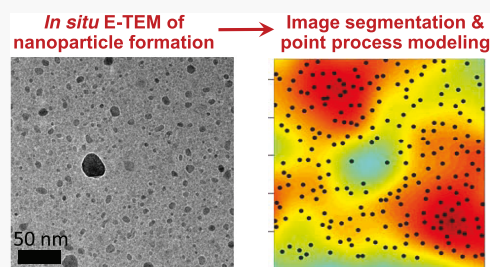
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: We combine *in situ* environmental transmission electron microscopy (E-TEM) with automated image processing and statistical machine learning to uniquely formulate interpretable mathematical models and accurate simulation tools for complex nanoscale phenomena involving coupled physical and chemical processes and interactions that are otherwise hard to model. In particular, there is a need for a better understanding, characterization, and prediction of the proximity effects among dense populations of metal nanocatalysts as they form and evolve over time. Here, we leverage point process theory, a branch of statistical machine learning, to “learn” the spatial dependencies among ensembles of adjacent alumina-supported iron nanoparticles from a time sequence of E-TEM images. We construct a set of point process models to make statistical inferences about the nature of spatial dependencies that govern the rapid formation, or “popping” of nanoparticles during thin film dewetting, concomitant with metal reduction in the presence of acetylene at 750 °C. We show that nanoparticles exhibit strong dispersion behavior, i.e., new nanoparticles pop in dispersed locations at a predictable distance from their existing territorial neighbors. We also show that Gibbs point processes adequately describe the pairwise interactions underlying such time-dependent spatial variations. Further, we build on our machine-learned models to develop a computational simulation tool capable of producing accurate spatiotemporal simulations of nanoparticle formation at finer time resolutions and larger spatial domains than those of experimental observations. This is a much needed capability to overcome current limitations in computational methods supporting the design, analysis, and control of the collective behavior of nanocatalyst populations.



1. INTRODUCTION

An increasing number of chemical conversion technologies depends on oxide-supported metal catalysts, such as iron (Fe), nickel (Ni), and copper (Cu).¹ In these heterogeneous catalyst systems, it has been abundantly shown that the support material, as well as the size, shape, composition, and chemical state of metal nanoparticles, determines the catalytic activity at different temperatures and chemical environments.^{2–4} Importantly, mounting experimental evidence indicates that interparticle distances also strongly influence catalyst stability and lifetime.^{2,5,6} Nevertheless, a comprehensive quantitative understanding of these collective effects is still largely missing.

For Al₂O₃-supported Fe catalyst, which is widely used for the synthesis of carbon nanotubes (CNTs) by chemical vapor deposition (CVD),⁷ Bedewy et al. have studied the population behavior of Fe nanoparticle catalytic activity by measuring the time evolution of CNT density during growth.^{8,9} In these densely packed conditions, where 10–100 s of billions of CNTs grow simultaneously per square centimeter, mechanical interactions among individual CNTs and their bundling behavior vary spatially.^{10–12} Moreover, the synergistic chemical coupling between neighboring areas of catalyst nanoparticles was shown to affect the areal density of catalytically active

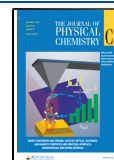
nanoparticles, as well as the collective growth rates of CNTs.¹³ Hence, revealing proximity effects extending from the nanoscale to the micron scale is crucial for controlling the morphology and properties of micropillars of vertically aligned CNTs that are needed for many applications including thermal interfaces¹⁴ and electrical interconnects.¹⁵

The most common approach for preparing substrate-bound catalyst nanoparticles for CNT forest growth is by solid-state dewetting of thin metal films,¹⁶ wherein atomic surface diffusion at high temperature drives film breakage and formation of nanoparticles.¹⁷ During this thermal treatment step, the chemical state of the catalyst typically also changes, which is revealed by employing various *in situ* surface characterization techniques including X-ray photoelectron spectroscopy (XPS),¹⁸ X-ray diffraction (XRD),¹⁹ electron diffraction (ED),²⁰ and electron energy loss spectroscopy

Received: August 25, 2020

Revised: November 16, 2020

Published: December 9, 2020



(EELS).²¹ In addition, real-time imaging using environmental transmission electron microscopy (E-TEM) enables characterizing the structural evolution of individual catalyst nanoparticles before and during catalytic activation.^{18,19,22–24} While these studies provide time-dependent information on catalyst evolution, they do not provide information on the spatial dependencies among ensembles of nanoparticles, either during their formation by dewetting or during the progressive activation of individual catalyst nanoparticles (i.e., during nucleation of individual CNTs from specific catalyst nanoparticles).

To study such collective behavior for Fe/Al₂O₃, we have previously used *in situ* E-TEM while adjusting the degree of magnification to enable simultaneously imaging hundreds of nanoparticles within the same field of view and combined these images with time-evolving ED and EELS measurements.^{11,20} ED results showed that catalyst reduction proceeds very slowly while heating in hydrogen at the low-pressure conditions of our experiments (40 mTorr). However, individual nanoparticles only started to appear abruptly, i.e., to “pop” in view, after introducing the acetylene gas, which accelerated the reduction of the catalyst. The sudden nature of nanoparticle formation by dewetting was also observed from *in situ* high-speed grazing-incidence small-angle X-ray scattering (GI-SAXS) experiments, in which the iron film was heated rapidly in ethylene.²⁵ These studies provided insights into the complex interplay between the chemical evolution (e.g., reduction and carbide formation) and the physical changes (e.g., nanoparticle formation and ripening) over time during the thermal treatment step that immediately precedes CNT nucleation and growth. Nevertheless, fully understanding spatial dependencies is largely limited by our ability to not only carry out these challenging *in situ*/operando experiments but also do comprehensive data analytics for quantifying and modeling proximity effects.

Starting from a time sequence of E-TEM images, this paper investigates the following set of scientific questions, which can only be addressed adequately by bringing machine learning to bear on, typically noisy, real-time electron microscopy measurements:

(Q1) Do nanoparticles pop uniformly over space? or do they in contrast exhibit spatial nonuniformity, wherein the formation of a nanoparticle is more probable in certain locations than others?

(Q2) Do nanoparticles exhibit spatial dependencies (proximity effects)? and if yes, what is the nature of these proximity effects (clustering versus dispersion behavior)?

(Q3) How to probabilistically model the spatial variations and dependencies so as to accurately describe, simulate, and predict the spatiotemporal evolution of nanoparticle formation during film dewetting?

To answer these questions, referred to hereinafter as Q1–Q3, this paper leverages point process theory—a branch of statistical machine learning—concerned with the probabilistic inference and prediction using *point pattern data*—a type of data represented as arrangements of points/events/objects observed in a spatial domain. The overarching aim of the point process theory is to make inference about the spatial variations and dependencies governing the formation/occurrence of such points in space by investigating systematic trends, clustering/dispersion patterns, spatial correlations, etc. Point pattern data is abundant in several real-world contexts. For instance, in climatology, modeling the locations of rain occurrence events

across vast geographical areas can help us better understand rainfall distribution across space and, hence, enable better atmospheric predictability.²⁶ In ecological sciences, spatiotemporal wildfire occurrences can be modeled and represented as point patterns to inform optimal firefighting resource deployment.²⁷

Other applications include but are not limited to forestry,²⁸ public health,^{29,30} and biological sciences.³¹ Baddeley et al. provide a comprehensive review of point process theory and its applications in science and engineering.³² In nanoscience, microscopy image data can be naturally regarded as point patterns (e.g., nanoparticles distributed over space). Nevertheless, the utility of point process theory in nanoscience is still highly underappreciated, notwithstanding recent efforts.^{33–38} This is likely due to the historical technical barriers associated with processing noisy electron microscopy images into a format that is accessible by point process models.³⁹ Here, we leverage point process theory as a methodical framework to formulate verifiable statistical tests of hypotheses, as well as physically interpretable statistical machine learning models, which enable us to reach statistically derived answers to the scientific questions Q1–Q3 listed above.

We start our analysis by leveraging a new image processing technique to extract location and size information of nanoparticles and then statistically demonstrate the spatial nonuniformity in nanoparticle popping during dewetting with a clear inverse-cubic relation between the areal density of nanoparticles and their sizes. Afterward, we statistically show that nanoparticles exhibit spatial dependencies among themselves, particularly a dispersion behavior among formation sites. In light of that, we show that Gibbs point processes, a class of point process models, adequately describe these proximity effects, which are shown to be time- and space-dependent. Finally, we leverage the probabilistic nature of these “learned” point process models to create spatiotemporal simulations that accurately capture and predict dewetting dynamics beyond the limited spatial and temporal ranges of experimental observations. This machine learning-based approach is powerful because such a complex process that involves coupling between chemical reduction and physical diffusion is otherwise exceedingly challenging to model and simulate.

2. METHODS

Our proposed statistical methods hinge on our ability to collect and transform the noisy raw E-TEM images into a format that is accessible by point process theory. In Section 2.1, we describe our approach to collect *in situ* E-TEM images for metal catalyst evolution during dewetting (Figure 1), followed by a unique image processing method to segment the images and extract important attributes such as nanoparticles’ locations and sizes over time (Section 2.2). We then conclude this section by listing the set of methods for modeling and simulation in Section 2.3, for which the details will be discussed in Section 3.

2.1. E-TEM Experiments. E-TEM images used here were experimentally collected using the FEI Titan 80-300 instrument at the Center for Functional Nanomaterials (Brookhaven National Laboratory). It is a field-emission TEM microscope that has aberration correction and is equipped with a heated sample holder and differential pumping apertures that enable *in situ* CVD studies in a microreactor. Samples were prepared by depositing the catalyst film (1 nm Fe on 10 nm Al₂O₃)

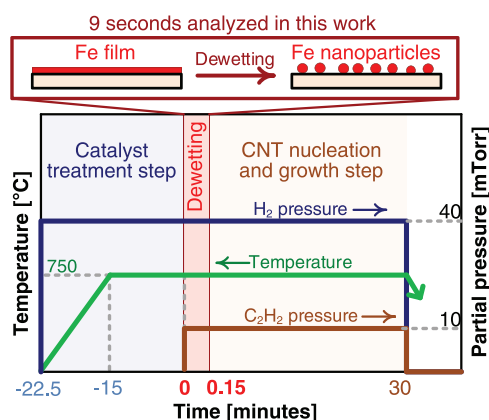


Figure 1. Schematic of the dewetting process studied here by E-TEM, along with the experimental recipe highlighting the 9 s of interest.

directly on commercially available TEM membranes (Norcada) with a 30 nm thickness.

After loading the samples inside the E-TEM, the system was pumped down to below 10^{-5} Torr (base pressure). According to the recipe shown in Figures 1 and S1, the first stage of the experiment was ramping-up temperature up to 750 °C in 40 mTorr of H_2 , followed by 15 min of dwell time at this temperature before introducing C_2H_2 , bringing the total pressure to 50 mTorr. Control of the partial pressures of both gases was maintained by using high precision leak valves. A microheater with a single-tilt sample holder was used to control temperature during the experiments. Thermal drift was negligible because we waited for 15 min after reaching the temperature of the experiment before collecting E-TEM images. Imaging parameters, such as focus and magnification, were adjusted before flowing the C_2H_2 gas into the microreactor.²⁰ A schematic of the CVD microreactor setup inside the E-TEM is shown in Figure S1.

The time sequence we use for all our analysis here starts at the point of introducing C_2H_2 . Hence, at time zero, only a few nanoparticles have already formed, as a result of the partial reduction and incomplete dewetting of the iron film after the

15 min of exposure to H_2 at 750 °C.²⁰ We collected 22 E-TEM images over the 9 s that started with C_2H_2 introduction, which are analyzed in this study. We exploited the presence of a large particle at the center of the image as a location reference to compensate for slight drift during imaging.

2.2. E-TEM Image Segmentation. Nowadays, E-TEM studies are capable of producing large amounts of image data that are not readily suitable for statistical analysis in their as-collected format. Hence, to realize the potential of point process theory, and of machine learning in general, a necessary step is to apply image segmentation to extract nanoparticles' outlines, centroids, and sizes from the E-TEM images. The output of the segmentation provides a time-evolving map of metal nanoparticle locations, areal densities, sizes, and statistical distributions thereof. Then, statistical point process theory is used to model and simulate the spatiotemporal evolution of nanoparticle formation to provide insights into the interfacial coupled phenomena underpinning the observed dynamics.

Existing image segmentation techniques include but are not limited to thresholding,⁴⁰ watershed transform,⁴¹ sliding band filter,⁴² active contour,⁴³ and iterative voting.⁴⁴ The high levels of background noise in E-TEM images make distinguishing nanoparticles from the noisy background an inherently challenging task, and it is infeasible to achieve using off-the-shelf software and classical segmentation approaches (see Figure S2). This calls for the use of advanced approaches that can borrow strength across multiple image segmentation techniques. In this paper, we adopt a modified version of the method proposed by Qian et al.,⁴⁵ which pools the results from two segmentation techniques based on intensity and gradient information, respectively. On the one hand, the intensity information is used to derive a set of segmentation results by applying a k-means clustering algorithm, which first separates the background from the foreground, followed by a watershed transform, which extracts the particle locations from the foreground. A second set of results, independent of the first one, is obtained via the gradient information by using an active contour procedure, which finds the foreground, followed by an iterative voting method, which determines the particle

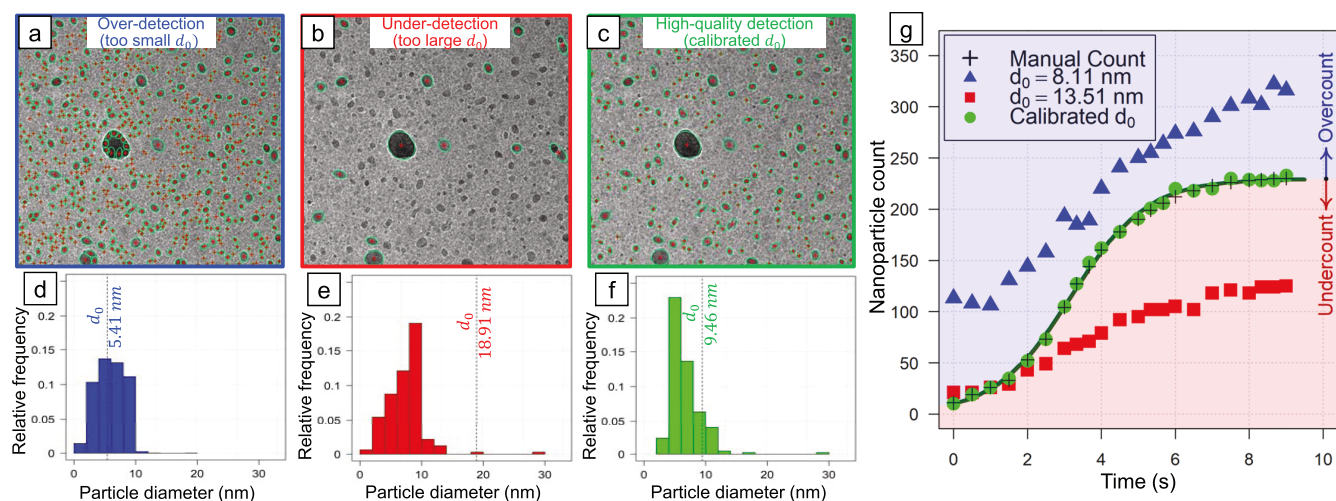


Figure 2. (a–c) Segmentation of E-TEM image #15 ($t = 6$ s) with values of $d_0 = 5.41$ nm (a), 18.91 nm (b), and 9.46 nm (c), showing the implications of a too small or a too large d_0 , leading to over- and underdetection, respectively (as compared to a calibrated image-specific $d_0 = 9.46$). (d–f) Histogram of nanoparticle size distribution for image segmentation shown in (a–c). (g) Validating the segmentation results for all images, as compared to a manual count from Bedewy et al.²⁰ The solid line is a sigmoidal model fit, as expressed in eq 1.

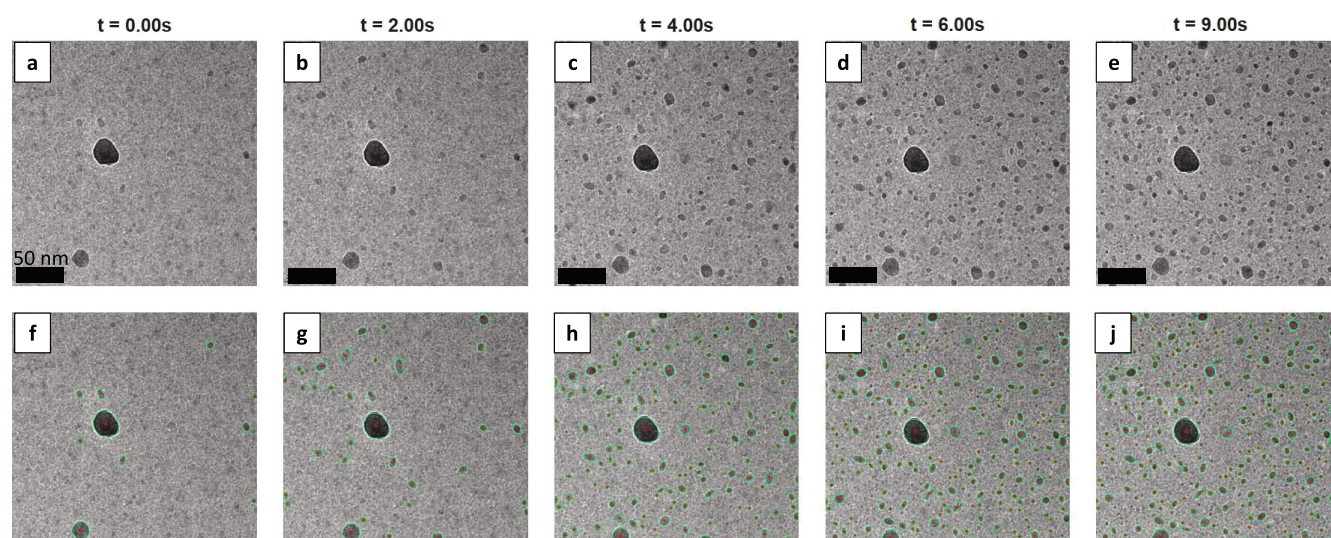


Figure 3. Top row (a–e): E-TEM images #1, 5, 10, 16, and 22, captured at $t = 0.00$, 2.00 , 4.00 , 6.50 , and 9.00 s, respectively. Bottom row (f–j): image segmentation results for the same images with an image-specific value of d_0 .

locations. A binary integer program is then formulated to resolve the conflict between the two sets and pool them into one combined set of segmentation results.

Hence, the image segmentation approach takes as input the raw E-TEM image, as well as one input parameter, d_0 , which is an initial estimate of the average particle diameter in the image. By implementing the method on our E-TEM images using various values of d_0 , we note that the quality of the segmentation results is highly sensitive to the selection of d_0 . Specifically, selecting a value for d_0 that is too large leads to a high number of false negatives (i.e., mistaking true particles for background noise). In contrast, a value for d_0 that is too small leads to a high number of false positives (i.e., over detection), as well as an increased likelihood of mistaking large-sized particles for an agglomerate of small overlapping particles (see Figure 2a–c). We postulate that this is an artifact of the temporal evolution of the nanoparticle size distribution (Figure 2d–f), which is continuously changing due to two factors: the formation of new, small-sized nanoparticles (driving down the true value of d_0) and the possible growth of existing nanoparticles (driving up the true value of d_0). In light of this temporally evolving nanoparticle size distribution, attempting to find a universal value for d_0 that works best for all E-TEM images is not a suitable approach. Instead, we seek an image-specific value for d_0 (hence, will be denoted as $d_0(j)$, $\forall j = 1, \dots, J$, where j is an image index, and J is the total number of images). We started by selecting a suitable value for $d_0(j = 1)$; then, for $j + 1$, $j = 1, \dots, 22$, we sampled a set of candidate values, which is compactly centered around $d_0(j)$, obtained the detection results for each candidate, and then selected the value that yields the best detection results, judged visually. As described below, the values of $d_0(j)$ selected for processing all images (Figure S3) were validated against a manual count (Figure 2g).

2.3. Data Analytics. The extracted information from image segmentation is used as direct input to the point process methods in this paper, which we categorize into methods for modeling and methods for simulation. The former entails training the point process models using the segmented data (as detailed in Sections 3.2–3.4), while in the latter, the trained models are used to probabilistically generate simulations at any

arbitrary time point or spatial window (as detailed in Section 3.5). For both sets of methods, we used several functionalities in the R package *spatstat*.⁴⁶ The simulations are generated using the Metropolis–Hastings algorithm, which is a Markov Chain Monte Carlo (MCMC) method, wherein the trained point process model is set as the Markov chain’s equilibrium distribution.⁴⁷

The interaction potentials obtained from the trained point process models are extrapolated to time steps where experimental data was not available. This is realized by training a spatiotemporal Gaussian process (GP) model, which is a nonparametric regression model that is widely used for modeling complex nonlinear response surfaces (interaction potentials in this paper).^{48,49} The underlying assumption in GPs is that the joint distribution of the model output (in our case, the interaction potential) at any finite set of inputs (e.g., interparticle distance and time) follows a multivariate normal distribution. This, along with their well-established predictive power, makes them mathematically tractable and suitable for many physics-driven applications.⁵⁰ In this paper, the GP model is fit by numerically maximizing the closed-form expression of the GP log-likelihood, using the command `nlm` in R. The point process model parameters are then estimated from the GP-learned interaction potentials using a nonlinear least-squares formulation, solved via the `NL2SOL` algorithm.⁵¹ Those learned parameters will be used as inputs to our simulation tool, as detailed in Section 3.5.

3. RESULTS AND DISCUSSION

We present the image segmentation results in Section 3.1 and then proceed with the statistical modeling for one E-TEM image (image #22 at 9 s) in Sections 3.2 and 3.3. To gain insights into process dynamics, we extend this analysis to all E-TEM images in Section 3.4. Finally, we build on those machine-learned models to create spatiotemporal simulations in Section 3.5, overcoming experimentally imposed limitations on spatial window and time resolution.

3.1. Time Evolution of Segmented Images. The top row of panels in Figure 3a–e shows the evolution of raw E-TEM images during the popping of nanoparticles by dewetting, while the bottom row (Figure 3f–j) shows the

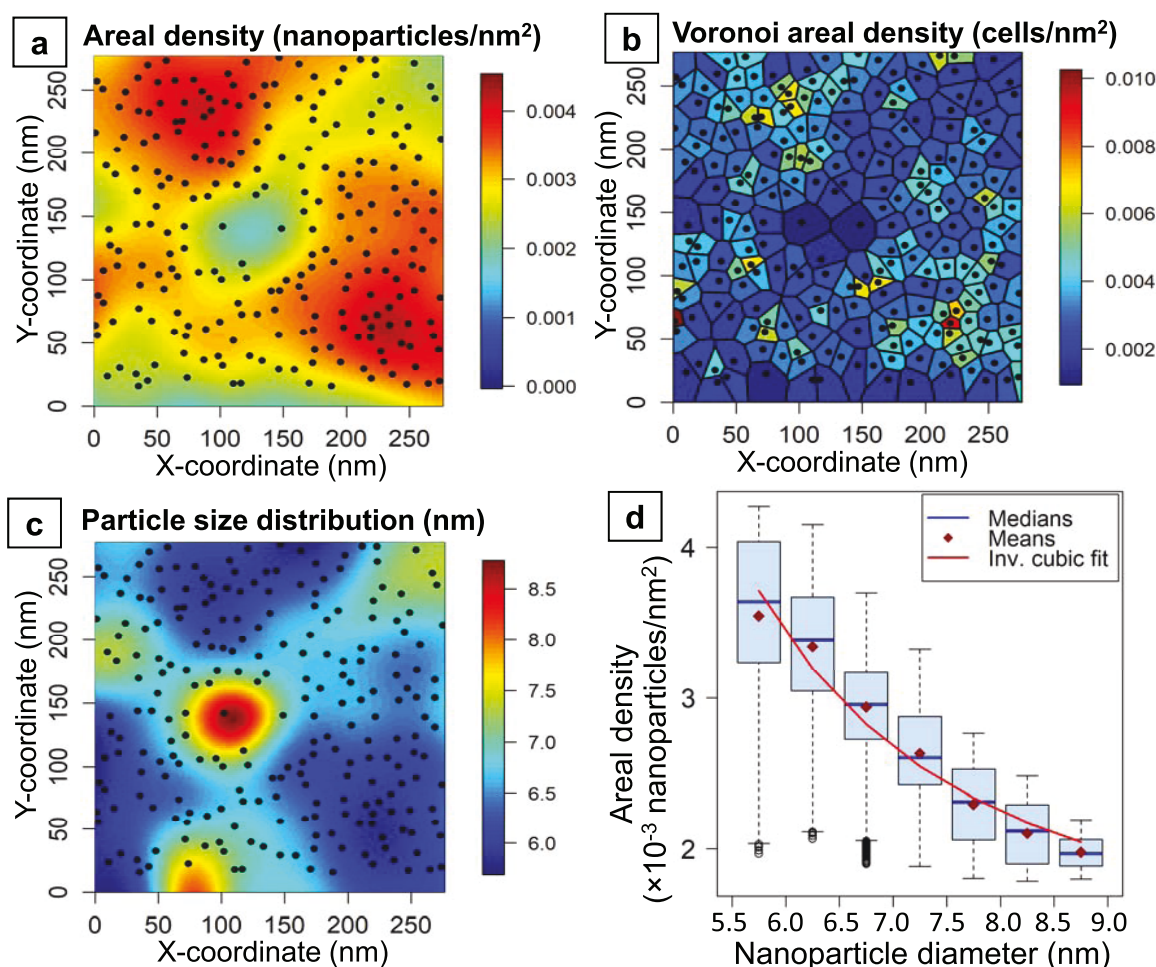


Figure 4. Nanoparticle areal density and sizes for E-TEM image #22 recorded at $t = 9.00$ s with 233 detected nanoparticles. (a) Heat map of Kernel smooth for first-order intensity (i.e., areal density of nanoparticles) overlaid on top of the observed point pattern. (b) Voronoi tessellation of nanoparticle locations with a heat map of areal density of cells (reciprocal of cell area). (c) Heat map of Kernel smooth for nanoparticle size distribution (based on the diameter of detected nanoparticles) overlaid on top of the observed point pattern $\mathbf{x}$. Note how nanoparticle density is minimal in the neighborhood of large-sized nanoparticles (e.g., the central part of the image), while density peaks in the vicinity of small-sized nanoparticles (e.g., the top left and bottom right regions of the image). (d) Box plot showing the inverse relation between the smoothed areal density of nanoparticles and their diameters, along with inverse-cubic fit of the form $0.001384 + 0.442524d^{-3}$.

segmentation results for the same images. The outcome of the segmentation procedure is a sequence of $n(j) \times 3$ data sets, where $j = 1, \dots, 22$ images. The first two columns of the j th data set correspond to the spatial locations of the nanoparticles (identified as horizontal and vertical axis coordinates), i.e., $\mathbf{X}(j) = \{\mathbf{x}_i(j) \in \mathbb{R}^2\}_{i=1}^{n(j)}$, where $\mathbb{R}^2$ is the two-dimensional real coordinate space. The third column represents the estimated area of each nanoparticle, i.e., $\mathbf{A}(j) = \{a_i(j) \in \mathbb{R}^+\}_{i=1}^{n(j)}$, where $\mathbb{R}^+$ is the set of positive real numbers. From these values of area, an estimated diameter for each nanoparticle is calculated, assuming circular geometry. This is denoted by $d_i(j) \in \mathbb{R}$. For all of the above, $j = 1, \dots, 22$ is the image index, and $i = 1, \dots, n(j)$ is the detected nanoparticle index within the j th image. Image segmentation results for all 22 images can be found in Figure S4.

Histograms of nanoparticle diameters for image #15 (at $t = 6.00$ s) are plotted in Figure 2d–f for the segmentation results shown in Figure 2a–c. It is noteworthy that the appropriate selection of d_0 is important not only for getting the correct count and locations of nanoparticles but also for extracting useful size information. Size distribution results for all images

are plotted in Figure S5. These results suggest that there are no significant changes to nanoparticle sizes after they abruptly appear in the field of view. Hence, coarsening of already-formed nanoparticles by Ostwald ripening^{52–54} does not dominate the size evolution over the 9 s studied here. Nevertheless, for longer time scales (over 30 min) at the high-temperature conditions of CNT growth by CVD, coarsening was observed.²⁰ Indeed, owing to the high temperatures of our experiments, the diffusion length of iron on the alumina surface is typically larger than the interparticle spacing.^{20,55,56} However, the rapid kinetics we observe here indicates that nanoparticle formation is dominated by fast diffusive transport and cluster coalescence.^{57,58}

To validate our segmentation results for all images (Figures 3 and S4), we compare the resulting detected nanoparticle count over time with that obtained from a manual count conducted in our earlier work on the same images (Figure 2g),²⁰ which showed a strong agreement. The nanoparticle popping dynamics is well described by a sigmoidal model (logistic function), as expressed in eq 1

$$n(t) = \frac{\alpha}{1 + \exp(-k(t - \tau))} \quad (1)$$

where $n(t)$ is the nanoparticle count at time t , α denotes the count at the end of the formation, τ is the time of the midpoint of the logistic curve, and k reflects the steepness of the curve. A nonlinear least-squares estimation yields the following values for the parameter estimates $\hat{\alpha} \approx 229$, $\hat{k} = .97$, and $\hat{\tau} = 3.18$. The constructed fit is shown as a solid line in Figure 2g, and it appears to fit the data well, with a normalized root-mean-squared deviation (NRMSD) of about 1.49%. Note that selecting a constant d_0 that has a small value (8.11 nm) not only resulted in a consistent and significant overcount for the number of nanoparticles but also distorted the S-shaped population behavior. On the other hand, selecting a constant d_0 that has a large value (13.51 nm) resulted in an undercount that became more significant with time.

We focus the rest of this paper on using the spatial locations of nanoparticle centroids to model spatial dependencies. While we use size data to derive useful insights to complement spatial point process models, nanoparticle size distributions are not explicitly included in the models.

3.2. Statistical Point Process Modeling and Analysis.

We first present our analysis based on one E-TEM image and then later extend the statistical methods to the remaining set of images. We choose to analyze the final image ($j = 22$, $t = 9.00$ s), shown in Figure 3e,3j, as it has the highest particle count, thus providing us with a rich data scenario. For now, we drop the image index j for notation simplicity.

We begin by formally defining the spatial point pattern resulting from the segmentation results of image #22. Specifically, let $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a spatial point pattern, where $\mathbf{x}_i \in \mathbb{R}^2$ is a two-dimensional point in space denoting the location of the i th detected particle in image #22. In point process theory, a spatial point pattern is generally defined as an unordered arrangement of points observed in a spatial window denoted by $\mathcal{W} \subset \mathbb{R}^2$, which is defined here by the boundaries of the E-TEM image ($\approx 277 \times 277$ nm²). Figure 4a shows the spatial point pattern $\mathbf{x}$ constructed using the outcome of segmenting image #22, where the black dots correspond to the detected nanoparticle centroids, while the background heat map denotes the areal density estimate of the nanoparticle count distribution. The total number of detected nanoparticles is $n = 233$ particles.

In point process modeling, both the number and the locations of the points are treated as random variables. Thus, we denote the underlying stochastic process governing the locations of the particles as $\{\mathcal{X}(\mathcal{B}): \mathcal{B} \subset \mathcal{W} \subset \mathbb{R}^2\}$, or for brevity $\mathcal{X}$, where $\mathcal{B}$ is a compact Lebesgue measurable subset of the spatial domain. We denote by $N(\mathcal{B})$ the random variable describing the particle count in a spatial region $\mathcal{B}$. As such, we can think of the observed point pattern $\mathbf{x}$ as a random realization of the underlying stochastic process $\mathcal{X}$.

Our goal is to make use of point process models to learn about the spatial variations and dependencies in $\mathcal{X}$ from its observed realization $\mathbf{x}$. Leveraging point process theory to answer the motivating scientific questions (Q1–Q3) entails the following:

(1) *Testing for spatial uniformity*: To formulate a statistical test for whether nanoparticles pop uniformly over space.

(2) *Testing for spatial dependence*: To formulate a statistical test for whether nanoparticles exhibit spatial dependencies (or proximity effects) among themselves.

(3) *Modeling and simulation*: To formulate a set of point process models that characterize spatial variations and dependencies (if any) to accurately describe, simulate, and predict nanoparticle formation during film dewetting.

3.2.1. Testing for Spatial Uniformity. A starting point to test the hypothesis of spatial uniformity is to investigate the so-called “first-order intensity function” denoted by $\lambda(\mathbf{s})$, where $\mathbf{s} \in \mathbb{R}^2$ refers to any arbitrary spatial location (whereas $\mathbf{x}_i$ refers to a location associated with an observed nanoparticle). The intensity function describes the areal density of nanoparticles and is used to estimate the expected particle count in a spatial subset of the domain, as expressed in eqs 2 and 3

$$\lambda(\mathbf{s}) = \lim_{|\mathcal{d}\mathbf{s}| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(\mathcal{d}\mathbf{s})]}{|\mathcal{d}\mathbf{s}|} \right\} \quad (2)$$

$$\mathbb{E}[N(\mathcal{B})] = \int_{\mathcal{B}} \lambda(\mathbf{s}) \, \mathrm{d}\mathbf{s}, \quad \mathcal{B} \subset \mathcal{W} \quad (3)$$

where $|\mathcal{d}\mathbf{s}|$ is the area of a small subset of $\mathcal{W}$. Figure 4a shows a kernel density smooth of $\lambda(\mathbf{s})$ using a Gaussian kernel with a 0.25 nanoparticles/nm² bandwidth, on top of its generating spatial point pattern. This plot illustrates that the areal density (or intensity) of nanoparticles exhibits considerable spatial variations, thus supporting the rejection of the uniform popping hypothesis in Q1. The same observation is noted in Figure 4b, which shows the Voronoi tessellation of the same point pattern, overlaid on the Voronoi-based intensity estimate (reciprocal of cell area). Still, a methodical statistical test is needed to assert that finding.

The essence of the statistical test for spatial uniformity is to investigate whether $\lambda(\mathbf{s})$ exhibits considerable variations over space, or in specific statistical terms, to test whether $\mathcal{X}$ has a “uniform” or “homogeneous” intensity. The formulation of this statistical test hinges on first proposing a statistical model for estimating $\lambda(\mathbf{s})$. Perhaps, the most basic statistical model for $\lambda(\mathbf{s})$ is the homogeneous Poisson point process (HPP) model. A HPP model implies the so-called complete spatial randomness (CSR) property, which enforces the following assumptions about the underlying stochastic point process. Assumption 1 (A1): $N(\mathcal{B}) \sim \text{Poisson}(\lambda|\mathcal{B}|)$, where λ is a constant intensity, and $|\mathcal{B}|$ is the size of the compact subset $\mathcal{B}$ of the spatial domain; Assumption 2 (A2): for any pair of disjoint compact spatial regions $\mathcal{B}_1$ and $\mathcal{B}_2$, $N(\mathcal{B}_1)$ and $N(\mathcal{B}_2)$ are independent random variables. HPP models are often used as benchmarks since, in reality, point pattern data rarely follow the CSR assumption, either due to a nonuniform intensity (violation of condition A1 above) or due to the dependencies between points (violation of condition A2 above), or both. If only condition A1 is violated, an alternative model to HPP is the inhomogeneous Poisson process (inh-PP), which is a generalization of the HPP model, where $\lambda(\mathbf{s})$ is assumed to change over space. As such, $N(\mathcal{B})$ would still follow a Poisson distribution but with a varying intensity, i.e., $N(\mathcal{B}) \sim \text{Poisson}\left(\int_{\mathcal{B}} \lambda(\mathbf{s}) \, \mathrm{d}\mathbf{s}\right)$, $\mathcal{B} \subset \mathcal{W}$.

We therefore propose a set of functional forms for $\lambda(\mathbf{s})$, yielding a set of inh-PP models (Table S1). A likelihood ratio test is then conducted⁵⁹ to test whether the gain in explanatory power introduced by capturing more complex spatial variations

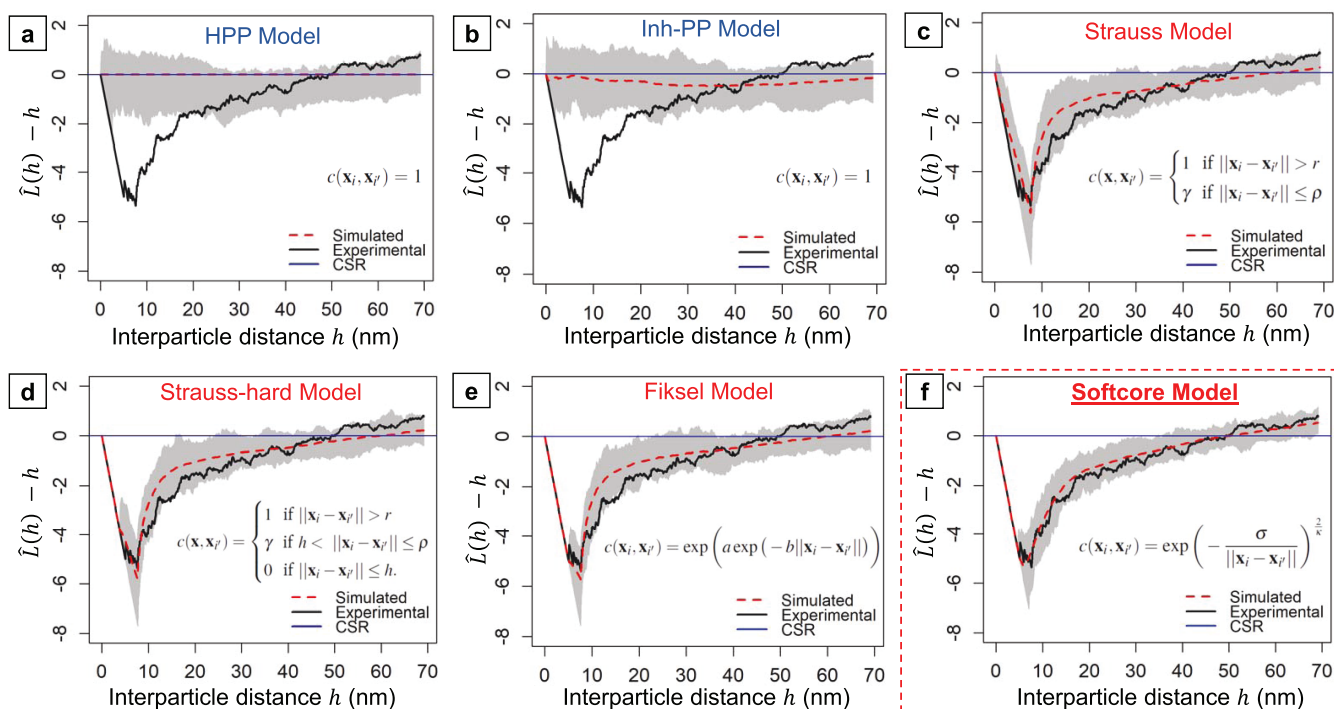


Figure 5. Point process modeling for spatial point pattern data extracted from E-TEM image #22 (at $t = 9.00$ s). (a, b) L -functions and simulation envelopes for Poisson models (HPP and inh-PP), indicating that intensity variations account for a small portion of CSR deviation—highlighting the importance of modeling spatial interactions. (c–f) L -functions and simulation envelopes for Gibbs models (Strauss, Strauss-hard, Fiksel, and Softcore), demonstrating that nanoparticles exhibit regularity (i.e., dispersion behavior), which gradually decays with interparticle distance. The Softcore model (f) appears to best capture this phenomenon. For all panels, number of simulations is set at 100.

via inh-PP, relative to the HPP model (the null hypothesis), is statistically significant. If it is, then the null hypothesis of CSR is rejected. Specifically, we fit 32 inh-PP models, each having a different parametric form that mainly depends on the spatial coordinates in $\mathbf{s} = (s_x, s_y)$, where s_x and s_y are the horizontal and vertical coordinates, respectively. The model selections are as follows: a log-linear trend in s_x , s_y , or both (3 models), a log-quadratic trend in s_x , s_y , or both (3 models), a log-cubic spline in s_x with the number of knots ranging from 3, ..., 15 (13 models), and a log-cubic spline in s_y with the number of knots ranging from 3, ..., 15 (13 models). This leads to a total number of 32 models. The specific functional forms for these models are expressed in Table S1. We then conduct a likelihood ratio test to evaluate the explanatory power of the fitted models relative to the HPP model. The p -values associated with each model are shown in Table S1. The smoothing spline model with dependence on s_y and 15 knots is shown to have the best explanatory power among all alternatives with a p -value of 0.0215, thus rejecting the null hypothesis of CSR and providing a *strong empirical evidence that nanoparticles do not form uniformly over space under our experimental conditions*.

3.2.2. Size Dependence of Areal Density. We are further interested in understanding the factors governing these spatial variations. Our conjecture is that such statistically significant variations are related to the nanoparticle size information. Specifically, Figure 4c shows a kernel smoother (Gaussian, bandwidth = 27 nm) to the associated nanoparticle size with each point, measured as the diameter (in nm) of the detected nanoparticles. By comparing Figure 4c to 4a, we easily observe that the heat map in Figure 4c resembles a reverse of that in Figure 4a. For example, the large central particle in our E-TEM

images (Figure 3) appears to be associated with a considerably lower density of nanoparticles within its immediate neighborhood. This highlights the inverse relationship between areal density (intensity) and nanoparticle size, suggesting that the likelihood of a nanoparticle forming in a particular location is dependent on the size of nanoparticles in its vicinity. This physical insight will be supported by our analysis of spatial dependence later.

We further quantify this inverse relation as shown in Figure 4d, which is a box plot for the kernel-smoothed areal density (shown in Figure 4a) vs the kernel-smoothed diameter (shown in Figure 4c) for all points in $\mathcal{W}$, i.e., for all points in the spatial window of the E-TEM image. For each bin of 0.5 nm width, we also plot the mean (in dark red) and the median (in dark blue) of the areal density. As seen in Figure 4d, an inverse-cubic function, specifically of the form $0.01384 + 442524d^{-3}$, represents a good fit for this data, with an NRMSE of 3.04%. This can be explained by the conservation of mass and volume of the metal comparing the flat nature of the initial film before dewetting with the presence of nanoparticles instead after dewetting. Accordingly, when a large nanoparticle forms, it consumes a larger volume of the metal film, which reduces the likelihood of formation of other nanoparticles nearby, i.e., the areal density is inversely proportional to nanoparticle volume (or inversely proportional to d^3). Moreover, the heteroscedasticity observed in Figure 4d highlights the dependence of variability in areal density on nanoparticle size, with smaller nanoparticles exhibiting a larger variability perhaps due to their increased dependence on randomness in surface texture of the as-deposited Fe film, as well as on defects in the support layer. These findings demonstrate the power of our statistical point process

modeling approach in “learning” physical/chemical phenomena without prior knowledge. Importantly, these findings are in agreement with the expected size distributions resulting from particle formation via the coalescence of highly mobile clusters in a low-density regime.^{60,61} Our results also suggest that there is planar strain, which was previously shown to widen the size distribution with an inverse power-law relation between the broadness of the size distribution and the areal number density during growth of interacting two-dimensional (2D) islands via diffusive mass transport.⁶²

3.2.3. Testing for Spatial Dependence. So far, our results revealed that nanoparticles do not pop uniformly over space. Now, we test whether these spatial variations are accompanied by spatial dependencies, i.e., whether the likelihood of a nanoparticle forming in a specific spatial location is dependent on the presence/locations of nanoparticles in its vicinity. Recall that deviations from the CSR scenario can be explained either by intensity variations (i.e., violation of A1), spatial dependencies (i.e., violation of A2), or both. While our test for spatial uniformity (above) shows that A1 is indeed violated, it does not give further information on the validity of A2.

Poisson models, be it HPP or inh-PP, only model first-order intensity variations but assume that particle counts in disjoint spatial regions are independent. Assessing spatial dependence requires going beyond first-order intensity modeling. A useful metric to use here is the so-called *L*-function, which is expressed by eq 4

$$L(h) = \sqrt{\frac{K(h)}{\pi}}, \quad h > 0 \quad (4)$$

where *h* is an interparticle separating distance, and *K*(*h*) is called the *K*-function and is defined by eq 5:

$$K(h) = \frac{\mathbb{E}[N_{\text{extra}}(h)]}{\lambda} \quad (5)$$

in which $N_{\text{extra}}(h)$ is the number of extra points within radius *h* of an arbitrary point. An empirical estimator of *L*(*h*), denoted by $\hat{L}(h)$, can be computed using the observed point pattern **x** as in eq 6

$$\hat{L}(h) = \sqrt{\frac{1}{\pi} \frac{1}{|\mathcal{W}|} \sum_i \sum_{i' \neq i} p(\mathbf{x}_i, \mathbf{x}_{i'}) \frac{\mathbb{I}[\|\mathbf{x}_i - \mathbf{x}_{i'}\| < h]}{\hat{\lambda}(\mathbf{x}_i) \hat{\lambda}(\mathbf{x}_{i'})}} \quad (6)$$

where $\hat{\lambda}(\cdot)$ is an estimate of $\lambda(\cdot)$ of eq 2, possibly using the parametric forms proposed earlier, $p(\mathbf{x}_i, \mathbf{x}_{i'})$ is an edge correction, $\|\cdot\|$ is the Euclidean norm, and $\mathbb{I}[\cdot]$ is the indicator function such that $\mathbb{I}[\|\mathbf{x}_i - \mathbf{x}_{i'}\| < h] = 1$ if $\|\mathbf{x}_i - \mathbf{x}_{i'}\| < h$, and 0 otherwise.

The expression in eq 6 forms the basis of a statistical test for spatial dependence. As shown in Figure 5, we compute an estimate of the *L*-function, denoted by $\hat{L}(h)$, using the observed point pattern data **x** (black solid line), and then compare it with an “envelope” of simulations randomly drawn from a Poisson model, be it HPP (Figure 5a) or inh-PP (Figure 5b) fitted to **x**. If $\hat{L}(h)$ significantly deviates from the envelope formed by the Poisson models, then there is a strong statistical evidence of spatial dependence.³¹ It is worth noting that, in theory, it can be shown that a HPP model satisfying the CSR assumption has $K_{\text{CSR}}(h) = \pi h^2$, and thus, substituting in eq 4, we have $L_{\text{CSR}}(h) = h$. A common transformation is to use $L(h) - h$ instead of *L*(*h*) as a test statistic for the spatial dependence test, as the CSR scenario (represented by the HPP

model) would be represented by a horizontal line at zero. This establishes an easy-to-visualize benchmark to assess the deviation of the observed point pattern **x** from the CSR scenario.

Figure 5a,b shows $\hat{L}(h) - h$ computed using the observed point pattern **x** (in black solid line) versus simulations from the HPP and inh-PP models, respectively, where the gray bands represent the envelope formed by the simulations randomly drawn from the fitted Poisson models. The blue horizontal line at zero is the theoretical $L_{\text{CSR}}(h) - h$ (where $L_{\text{CSR}}(h) = h$), while the red dashed line is the average of the simulations. The fundamental difference between Figure 5a,b is the use of the inh-PP, which captures the first-order intensity variations. In both, however, there is significant deviation between the empirical estimates and the simulation envelope, with little improvement offered by inh-PP over HPP. This suggests that the intensity variations account for a rather small portion of the deviation from the CSR scenario, and thus, a substantial portion of this deviation is yet to be explained by higher-order spatial dependence between nanoparticles. This motivates us to consider more advanced point process models, which, in contrast to HPP or inh-PP, are able to characterize spatial dependencies.

3.3. Beyond Poisson Models: Gibbs Point Processes.

The envelope tests conducted in Figure 5 convey more information than just asserting deviation from the CSR scenario. In particular, the values of $\hat{L}(h) - h$ tend to systematically exhibit negative values (i.e., $\hat{L}(h) - h < 0$) as seen in Figure 5, which means that the underlying point process is driven by a “dispersion” behavior since the expected count of nanoparticles located within a certain radius centered at an arbitrary nanoparticle is significantly smaller than that expected in a CSR setting (where independence between nanoparticles is assumed). This suggests that nanoparticles appear to exhibit a “territorial” behavior, preventing particles from forming in their immediate neighborhood, with the likelihood of a particle forming in close vicinity of other particles being smaller than farther away. This statistically learned conclusion asserts our finding in Figure 4d where nanoparticle formation/growth in a location influences the formation/growth likelihood of other nanoparticles in a spatial neighborhood. Interestingly, Figure 5 suggests that the strength of this dispersion interaction behavior gradually decays as the interparticle separating distance increases. Understandably, a Poisson model, be it HPP or inh-PP, cannot capture such behavior, due to its inherent spatial independence assumption (A2). Thus, we point our focus to another family of models called “Gibbs point process models”.

Unlike Poisson models, Gibbs point process models are specifically designed to capture pairwise spatial dependencies. In its most general form, the probability density of a Gibbs point process model can be written as in eq 7

$$f(\mathbf{x}) = \alpha \left[\prod_{i=1}^n \beta(\mathbf{x}_i) \right] \left[\prod_{i < i'} c(\mathbf{x}_i, \mathbf{x}_{i'}) \right] \quad (7)$$

where α is a normalizing constant, $\beta(\mathbf{s})$ is a first-order term to capture any spatial variations, and $c(\mathbf{x}_i, \mathbf{x}_{i'}) \in [0, 1]$ is a pairwise interaction term that models spatial dependencies. In fact, Gibbs models reduce to Poisson models if we set $c(\mathbf{x}_i, \mathbf{x}_{i'}) = 1$. As such, for $\beta(\cdot)$, we can use one of the parametric forms discussed previously for $\lambda(\mathbf{s})$ (Table S1) to capture the spatial variations. Different choices of $c(\cdot)$ can lead to well-defined

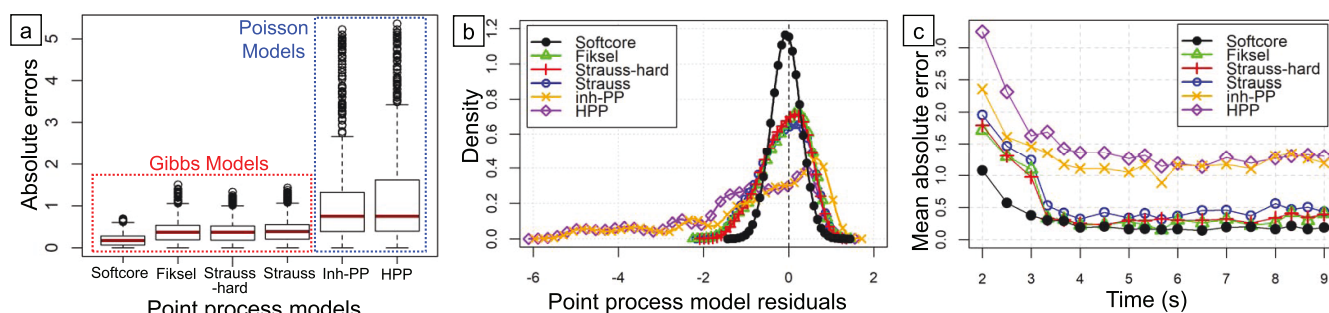


Figure 6. Quality of fit for all six point process models. (a) Boxplots of absolute errors for fitting data from E-TEM image # 22, $t = 9.00$ s, showing that Softcore is the best fit with the lowest absolute error. (b) Kernel density estimates for model residuals for image # 22, $t = 9.00$ s, showing a symmetric residual distribution for the Softcore model that is tightly centered around 0, further confirming the quality of this fit. (c) Mean absolute error (MAE) of all six point process models over time $t = 2.00, \dots, 9.00$ s, demonstrating that Softcore model consistently outperforms all other models.

statistical representations of Gibbs point process models. For instance, using a hard-thresholding procedure yields the Strauss model (Figure 5c)

$$c(\mathbf{x}_i, \mathbf{x}_{i'}) = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_{i'}\| > \rho \\ \gamma & \text{if } \|\mathbf{x}_i - \mathbf{x}_{i'}\| \leq \rho \end{cases} \quad (8)$$

where $\gamma \in (0, 1)$ is an interaction parameter indicating the strength of dependence such that $\gamma = 0$ corresponds to no interaction, and ρ is a scale parameter denoting the interparticle distance within which the spatial dependence is active.

A generalization of the Strauss model is the Strauss-hard model (Figure 5d), which applies a three-level thresholding mechanism instead of a bi-level threshold pattern. Within a radius ρ_1 , spatial dependence is maximal and nanoparticles are not allowed to form. In a distance range bounded between ρ_1 and ρ_2 , a spatial dependence with the strength specified by γ is enacted. At distances higher than ρ_2 , spatial interaction ceases to exist. This tri-level mechanism is expressed in eq 9

$$c(\mathbf{x}_i, \mathbf{x}_{i'}) = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_{i'}\| > \rho_2 \\ \gamma & \text{if } \rho_1 < \|\mathbf{x}_i - \mathbf{x}_{i'}\| \leq \rho_2 \\ 0 & \text{if } \|\mathbf{x}_i - \mathbf{x}_{i'}\| \leq \rho_1 \end{cases} \quad (9)$$

In contrast, Fiksel (Figure 5e) and Softcore (Figure 5f) models assume a soft-thresholding mechanism, wherein the spatial interaction gradually decays with the separating distance. The Fiksel model is expressed as in eq 10

$$c(\mathbf{x}_i, \mathbf{x}_{i'}) = \exp(a \exp(-b\|\mathbf{x}_i - \mathbf{x}_{i'}\|)) \quad (10)$$

where a and b are the range and length-scale parameters, respectively, a combination of which describes the strength and rate of decay of the pairwise interaction. The Softcore model, on the other hand, is expressed as in eq 11

$$c(\mathbf{x}_i, \mathbf{x}_{i'}) = \exp\left(-\frac{\sigma}{\|\mathbf{x}_i - \mathbf{x}_{i'}\|}\right)^2 / \kappa \quad (11)$$

where σ and κ are the length-scale and smoothness parameters denoting the range and shape of the interaction, respectively.

As shown in Figure 5, we find that all four Gibbs models (Strauss, Strauss-Hard, Fiksel, and Softcore) offer smaller deviations from the empirical L -function derived from the observed point pattern, relative to the two Poisson models

(HPP and inh-PP), highlighting the importance of spatial dependencies during nanoparticle formation. For the Strauss model, our estimates for interaction parameter γ and the scale parameter ρ were $\hat{\gamma} = .05$ and $\hat{\rho} = 8.00$ nm. The relatively small value of the scale parameter ρ , compared to the average diameter of nanoparticles (Figure 4c,d), indicates that a hard-thresholding mechanism in the Strauss model is not suitable to model the gradually decaying spatial dependence in nanoparticles. In fact, it appears that the model was only able to identify a neighborhood, which is almost equivalent to the boundaries of the particles as an area of strong spatial interaction and then cuts off to zero afterward, obviously overlooking an outer neighborhood in which interactions are still active. This observation is evident from Figure 5c, where the Strauss model, albeit offering substantial improvement over the Poisson models in Figure 5a,b, suffers from a relatively poor fit, especially at small separating distances. The parameter estimates for the Strauss-hard model were $\hat{\gamma} = .06$, $\hat{\rho}_1 = 7.59$ nm, and $\hat{\rho}_2 = 3.78$ nm, corresponding to the interaction strength, distance within which interaction is active, and distance within which the nanoparticle formation is prohibited, respectively. The parameter estimates suggest that the Strauss-hard model did not offer a substantial improvement over the Strauss model with a similarly too small interaction distance that is comparable to nanoparticle diameters.

On the other hand, Gibbs models that describe a gradual decay of interactions appear to provide significantly better fits to the data. Specifically, the Softcore model in Figure 5f outperforms all competing models, including the Fiksel model (Figure 5e). To further validate these conclusions, we quantitatively compare the models by computing the absolute deviation between the mean L -function computed by averaging the L -functions from 100 simulated point patterns generated from each model fit and that of the empirical L -function computed using the observed point pattern $\mathbf{x}$. Figure 6a shows the boxplots of the absolute errors for E-TEM image #22 ($t = 9$ s), which suggests that the Softcore model achieves the best overall performance compared to the other five models. Figure 6b confirms this finding by showing that the residuals from the Softcore model are both symmetric and tightly centered around 0, suggesting a satisfactory fit, especially when compared with other benchmarks.

In summary, we conclude that nanoparticles exhibit strong spatial dependencies among themselves, in particular, a dispersion behavior, under our experimental conditions, and that these dependencies gradually decay as a function of the interparticle

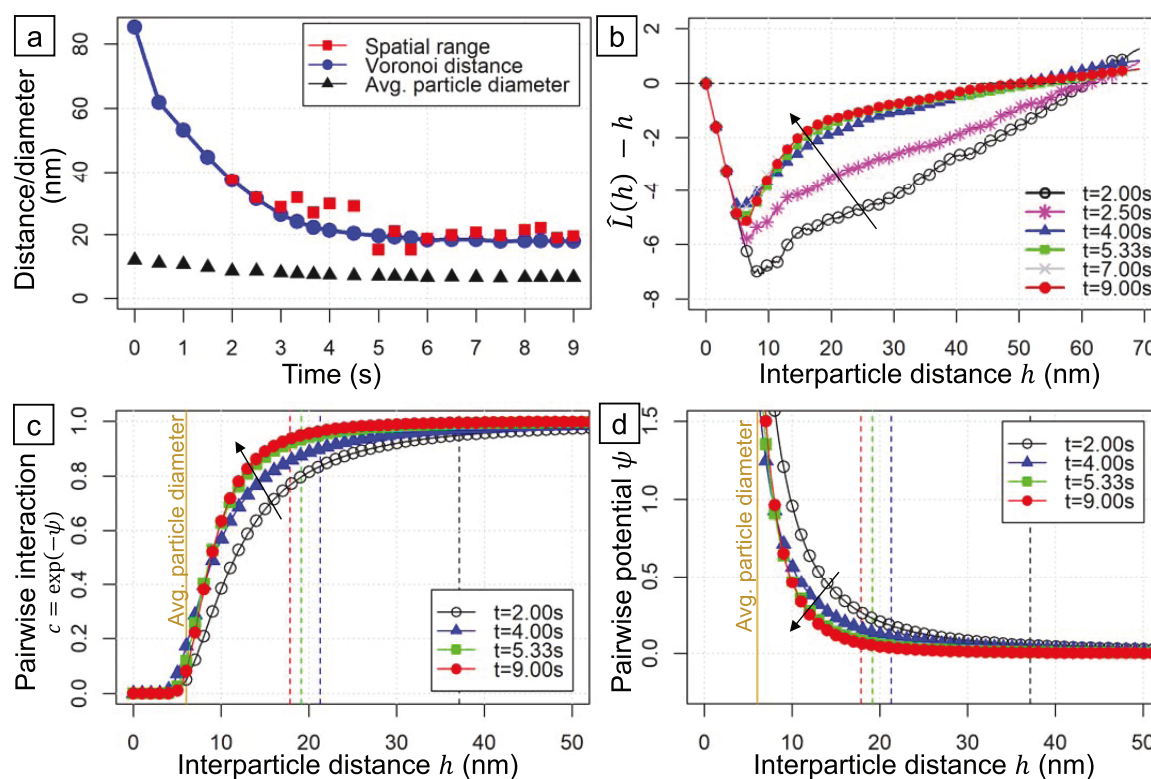


Figure 7. Dynamics of nanoparticle formation during film dewetting from Softcore modeling. (a) Time evolution of average nanoparticle diameter, average Voronoi distance, and the spatial range, defined as the interparticle distance of the 95th percentile of pairwise interaction c . (b) Time evolution of L -functions, estimated as the average of 100 simulations for modeling each E-TEM image ($t = 2$ – 9 s) using the Softcore model, as shown previously in Figure S6 for E-TEM image #22 (at $t = 9.00$ s). (c) Pairwise interaction as a function of interparticle distance at different times ($t = 2$ – 9 s). A value of 0 signifies maximal pairwise interaction, while a value of 1 denotes null interaction (i.e., independence). The vertical dashed lines denote Voronoi distances with each color corresponding to the respective time step. (d) Pairwise potential as a function of interparticle distance at different times ($t = 2$ – 9 s). A value higher than 0 signifies a dispersion behavior, with larger values indicating stronger dispersion.

distance. Hence, a Softcore model, a particular class of Gibbs processes, provides a powerful model to capture the dynamics of nanoparticle formation by dewetting.

3.4. Capturing Dewetting Dynamics via Gibbs Models. So far, all of the data analysis and statistical modeling were conducted on E-TEM image #22 ($t = 9$ s). We now extend our machine learning approach to the remaining sequence of E-TEM images. Hence, we restore the index $j = 1, \dots, 22$ for images taken at different points in time $t = 0, \dots, 9$ s. First, we show that the Softcore model is the most suitable Gibbs point process model for all E-TEM images over time, as illustrated in the consistently lower values of mean absolute error across the entire time span (Figure 6c), especially at early stages of the dewetting process, where the data deviates most from the CSR scenario (HPP model). This is important not only for providing insights into the time evolution of interactions among popping nanoparticles during dewetting but also for prescribing a methodical approach to simulate the spatiotemporal evolution of nanoparticle formation during dewetting using Gibbs point process models.

Analyzing the first four E-TEM images, i.e., images #1 ($t = 0.00$ s) through #4 ($t = 1.50$ s), shows that they had too few nanoparticles to obtain reliable parameter estimations for Gibbs model fits, as shown in Figure S6, where the large simulation envelope bands indicate a data size problem. Hence, for now, we primarily focus our point process model fitting on E-TEM images #5–22 ($2 \leq t \leq 9$). We will then later prescribe a method that leverages the model fits obtained over that time

span ($2 \leq t \leq 9$) to “learn” the parameters of the Softcore model at any arbitrary point in time either within the 2–9 s span or outside it. This will be explained in detail below and will form an indispensable component of our proposed computational simulation tool.

Moreover, we can gain valuable insights into the dynamics of nanoparticle formation during dewetting, including the early stage of the process ($0 \leq t < 2$ s), from the time evolution of the Voronoi tessellations of all images (Figures S7 and S8). Considering that the locations of nanoparticles are the sites for generating the Voronoi diagram, by definition all of the points enclosed within each Voronoi cell are closer to the generating sites (nanoparticle locations) than to other sites. Thus, the points on the sides of the convex polygon contour making up each cell represent equidistant loci between nearest neighbors, i.e., each vertex at which n number of cells meet is equidistant to the n number of nanoparticle sites generating those adjacent n cells. As such, we can use the distribution of Voronoi distances (Figure S9), calculated as the square root of Voronoi cell areas, as a suitable estimate of interparticle distances. In fact, Voronoi tessellations in general,⁶³ and Voronoi distances in particular,²⁵ have been previously used to analyze the spatial distribution of nanoparticles. As seen in Figure 7a, the mean Voronoi distance decays with time, owing to the progressive popping of new nanoparticles during dewetting until reaching an almost constant value slightly smaller than 20 nm toward the end when all of the nanoparticles form.

Starting from image #5 ($t = 2.00$ s), the exact same procedure previously described for using point process modeling to analyze E-TEM image #22 ($t = 9.00$ s) was implemented for the remaining 17 images. The only small difference being the selected trend for $\beta^t(\mathbf{x})$ in eq 7, which varied slightly from one E-TEM image to another. Superscript t , hereinafter, denotes a time-varying function or variable. Among the six fitted models, and across all images, the Softcore model achieves the best visual and quantitative performance, with an average percentage improvement in the mean absolute error of about 51% over the Strauss model and about 80% over the Inh-PP. This suggests that the Softcore model is a good modeling choice to explain the spatial distribution of catalyst nanoparticles, which exhibit strong dispersion behavior ($\hat{L}^t(h) - h < 0$) throughout the whole dewetting process, i.e., the existence of a nanoparticle at a certain location reduces the likelihood of other nanoparticles appearing in its vicinity. This can be explained by their constant competition on the limited resource, in this case the thin iron film, from which the instability-driven dewetting process draws the diffusing atoms making up each new nanoparticle.⁶⁴

We believe that the superiority of the Softcore model stems from its flexibility to model the gradual decay in interaction, which best describes the second-order variations in the underlying process. The time evolution of the L -function shown in Figure 7b indicates that the interaction both becomes weaker at later time steps and more local (i.e., the interaction distances become smaller). We best visualize this by plotting the pairwise interaction term $c^t(\cdot, \cdot)$ (Figure 7c), which can be used to describe the interaction strength or “potential” as a function of the interparticle distance. Specifically, we have $c^t(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\psi^t(\|\mathbf{x}_i - \mathbf{x}_j\|))$, such that $\psi^t(\cdot)$ is the potential function at time t , plotted in Figure 7d. These statistically derived mathematical formulations provide important physical insights into the complex dynamics of nanoparticle formation by thin film dewetting during chemical reduction and acetylene dissociation on the metal surface. Importantly, this experimentally learned model solves the long-standing issue of inability to use mean-field theory to derive a closed-form solution for the spatial interaction function between islands/nanoparticles in either 2D or three-dimensional (3D), although a simple exponential function adequately describes the case of one-dimensional (1D).^{65,66}

Figure 7d demonstrates that the collective behavior of nanoparticle popping exhibits a more abrupt decay of interaction potential ψ^t as the dewetting process progresses until it approaches zero in all cases at a different interparticle distance. A measure of such spatial range of interactions is calculated as the interparticle distance at which the pairwise interaction c^t reaches 0.95 (see Figure 7c). It is noteworthy that the values of this interparticle distance at the 95th percentile of pairwise interaction c^t are comparable to the average Voronoi distance and significantly larger than the average nanoparticle diameter, as shown in Figure 7a. This plot also illustrates that both the spatial range of pairwise interactions and the collective interparticle distances (estimated from Voronoi distance) decay during nanoparticle popping over time, while the average nanoparticle diameter only slightly changes. Importantly, these physically relevant quantities and mathematical formulas were learned through statistical modeling based on the experimental data used for training the models. Now that we have good model fits, we can

leverage the Softcore models for running predictive simulations with unprecedented capability of capturing the dynamics of chemically induced thin film dewetting, beyond the spatial and temporal limits imposed by E-TEM experiments.

3.5. Discrete Time Simulations via Gibbs Models. A major advantage of our machine learning approach based on point process models is that it enables running simulations with predictive power for either design or analysis purposes. Given the probabilistic nature of the Softcore model fits to the experimental data, one can simulate a set of realizations for point patterns that follow the same interaction functions as those in Figure 7c.

While useful, using only the model fits for the 18 E-TEM images analyzed earlier would restrict our predictive capability to the limited time points where experimental data are available, i.e., images #5–22 ($2 \leq t \leq 9$), because we do not have the interaction potentials or the fitted Softcore parameters at any arbitrary time. In practice, simulations become more impactful as a way to predict experimental outcomes, when the latter are not available, too costly, or infeasible to conduct. For instance, in our setting, we are exceedingly interested in simulating the earlier part of the dewetting process (0–2 s), or even beyond the 9 s end point, for which we do not have reliable model fits. Further, we are interested in creating discrete time simulations with finer sampling rates than the temporal resolution of imaging (e.g., fractions of a second). To build this computational simulation tool, which can predict the experimental outcome at any point in time, we therefore need a statistical “vehicle” that can predict the interaction potential at any arbitrary time step. We will denote this statistical model as the **Spatiotemporal Interaction Model or STIM** in short.

Spatiotemporal Interaction Model (STIM). STIM is a spatiotemporal statistical regression model, which models the pairwise interaction c^t , plotted in Figure 7c, as a function of two inputs: time t and interparticle distance h . Specifically, we use a spatiotemporal Gaussian process (GP) model⁴⁸ to construct this functional mapping for two reasons: (1) in machine learning, GPs are specifically tailored to model “correlations” in the input space,^{48,67,68} i.e., observations that are closer in the input space (time and interparticle distance) tend to have similar output values for the pairwise potential. This is particularly relevant for temporal processes like nanoparticle popping in thin film dewetting and (2) in the computer simulation literature, GPs are often used as efficient “surrogate models” for complex physics-driven phenomena due to their powerful predictive properties.^{50,69}

We trained STIM on a subset of the c^t functions obtained in 7 time steps, with 100 data points discretely sampled over the range of h (0–100 nm) at each t , thus yielding a total of 7 time steps $\times$ 100 data points = 700 training points. This training subset was sufficient to yield high predictive accuracy, as evidenced by a low value of NRMSE at 2.30%, calculated using the GP-based STIM predictions at the remaining time steps that were not used for model training. This high-quality predictive performance bolsters the accuracy of the simulations below.

Using STIM, we obtain predictions of the pairwise interaction functions at fine time increments. Figure 8a shows few predicted interaction functions outside the limited 2–9 s range, i.e., at $t = 0.5, 1, 1.5$, and 9.5 s. We further predict the pairwise interaction functions at 0.2 s time intervals, which we use to derive the spatial range of interactions, as shown in

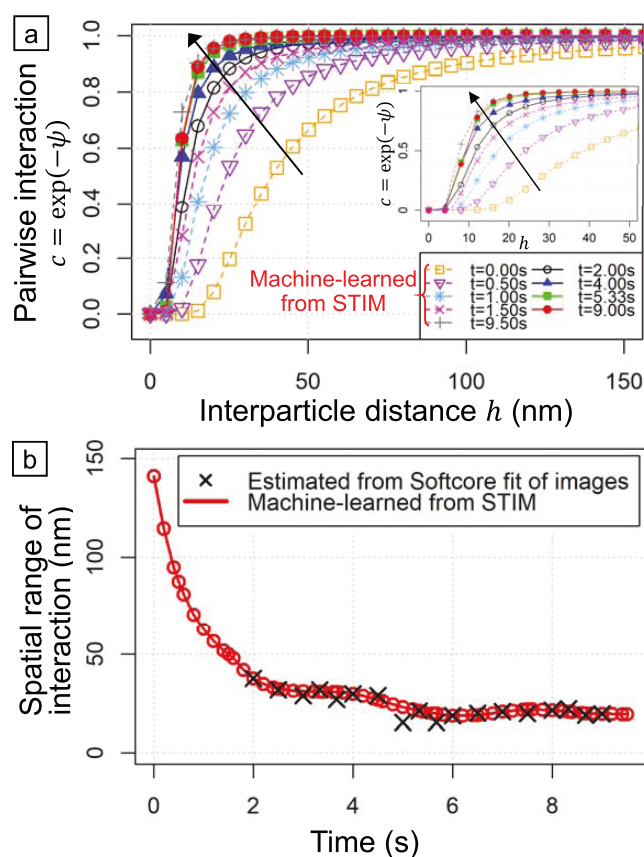


Figure 8. Spatiotemporal interaction modeling (STIM) results showing the comparison between machine-learned predictions from STIM and estimated values from Softcore fit of E-TEM images for (a) pairwise interactions and (b) spatial range. The inset in (a) is a close-up for a 0–50 nm range of h .

Figure 8b. Using these predicted interaction functions, we estimate the Softcore parameters (σ^t and κ^t) from eq 11 via nonlinear least squares, yielding estimated parameters for each time t . The time evolution of the predicted Softcore parameters $\hat{\sigma}^t$ and $\hat{\kappa}^t$ using STIM is plotted in Figure S10. Figure 8b illustrates the strong agreement between STIM predictions and those obtained from the Softcore fits of experimental E-TEM data by plotting the values for spatial range (interparticle distance at the 95th percentile of c^t) for each time step.

Simulation Results. Our computational simulation tool takes three inputs at any time point: (1) values for the Softcore parameters, predicted using STIM, (2) a nanoparticle count, predicted using our statistical particle count model (shown in Figure 2g), which models the S-shaped popping dynamics using a logistic function, and (3) a selection of the trend function for $\beta^t(\mathbf{x})$ of eq 7. The detailed steps of our simulation algorithm are presented in Table S2.

We present two sets of simulation results, corresponding to two case studies. In the first case study, we limit the spatial window to that imposed by the experiment ($\approx 277 \times 277 \text{ nm}^2$) but use STIM to extend the temporal resolution and range beyond those of the experiments. We call this case study **Temporally Extended Simulation**. In the second case study, referred to as **Spatially Extended Simulations**, we extend the spatial window, as well as temporal resolution and range.

Temporally Extended Simulations. Figure 9a–h and Video S1 show the time evolution of simulated nanoparticle locations and areal density heat maps from our computational simulation tool of nanoparticle formation by dewetting. More results for the Voronoi tessellations (with heat map of the reciprocal of cell area) are plotted in Figure S11. Figure 9i shows the evolution of the average L -functions of 25 simulations, which agree with those obtained from experimental data (Figure 7b). Importantly, our simulations are now not limited by the time resolution of our experimental E-TEM data, which was constrained to either 2 or 3 images per second. Instead, we ran our simulation with a smaller time step of 0.2 s. If desired, our simulation tool allows for even finer time resolution than 0.2 s. Moreover, using TIM, we are now able to predict the dynamics of the entire process even outside the time range for which we have experimental data.

From the 25 realizations, we extracted a statistic that we call the “strongest interaction”, which is calculated as the minimum $\hat{L}^t(h) - h$ value at each time point. Figure 9j shows the average value of this statistic over time (solid red line), while the lighter red lines denote the values of this statistic for the 25 realizations. We extracted the values of the same statistic from the experimental L -functions obtained from the observed point patterns at each time point of the 18 E-TEM images (at $t = 2$ –9 s) and plotted them on the same figure as black cross marks. The strong agreement between the simulated and experimental values validates the accuracy of our computational simulation tool and confirms its ability to capture the actual formation dynamics of nanoparticles during dewetting. Another statistic of interest is the average nearest-neighbor distance (NND) at time t , calculated as in eq 12

$$\text{NND}(t) = \frac{1}{n(t)} \sum_{i=1}^{n(t)} \min_{i' \neq i} \|\mathbf{x}_i - \mathbf{x}_{i'}\| \quad (12)$$

where $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ are two distinct nanoparticles. The time evolution of $\text{NND}(t)$ is shown in Figure 9k. This, along with the time evolution of Voronoi distance results shown in Figure S12, further illustrates our simulations’ ability to accurately represent the spatiotemporal distribution of nanoparticles observed experimentally from E-TEM.

It is worth noting here that for our simulation results in Figure 9, a single mathematical formulation for $\beta^t(\mathbf{x})$ in eq 7 was selected for simulating all time steps. Thus, the strong agreement between simulations and experiments highlights the robustness of our approach, i.e., simulations are not particularly dependent on the specific trend function as much as they are dependent on the Softcore parameters and nanoparticle count. This confirms our earlier finding (when discussing the statistical point process modeling) that the spatial dependencies are the major determinants of deviation from CSR assumptions.

Spatially Extended Simulations. We further conduct another set of simulations, where we use our computational simulation tool to mimic the behavior of nanoparticles for larger spatial areas. Since E-TEM images can be considered spatial samples or windows of a larger spatial area, an experimenter may be interested in predicting, or specifically extrapolating, the dynamics of nanoparticle formation in those unobserved experimental regions. Note here that the spatial window of $\approx 277 \times 277 \text{ nm}^2$ of our E-TEM images is a result of carefully adjusting the magnification during imaging. Considering that higher-magnification imaging conditions would

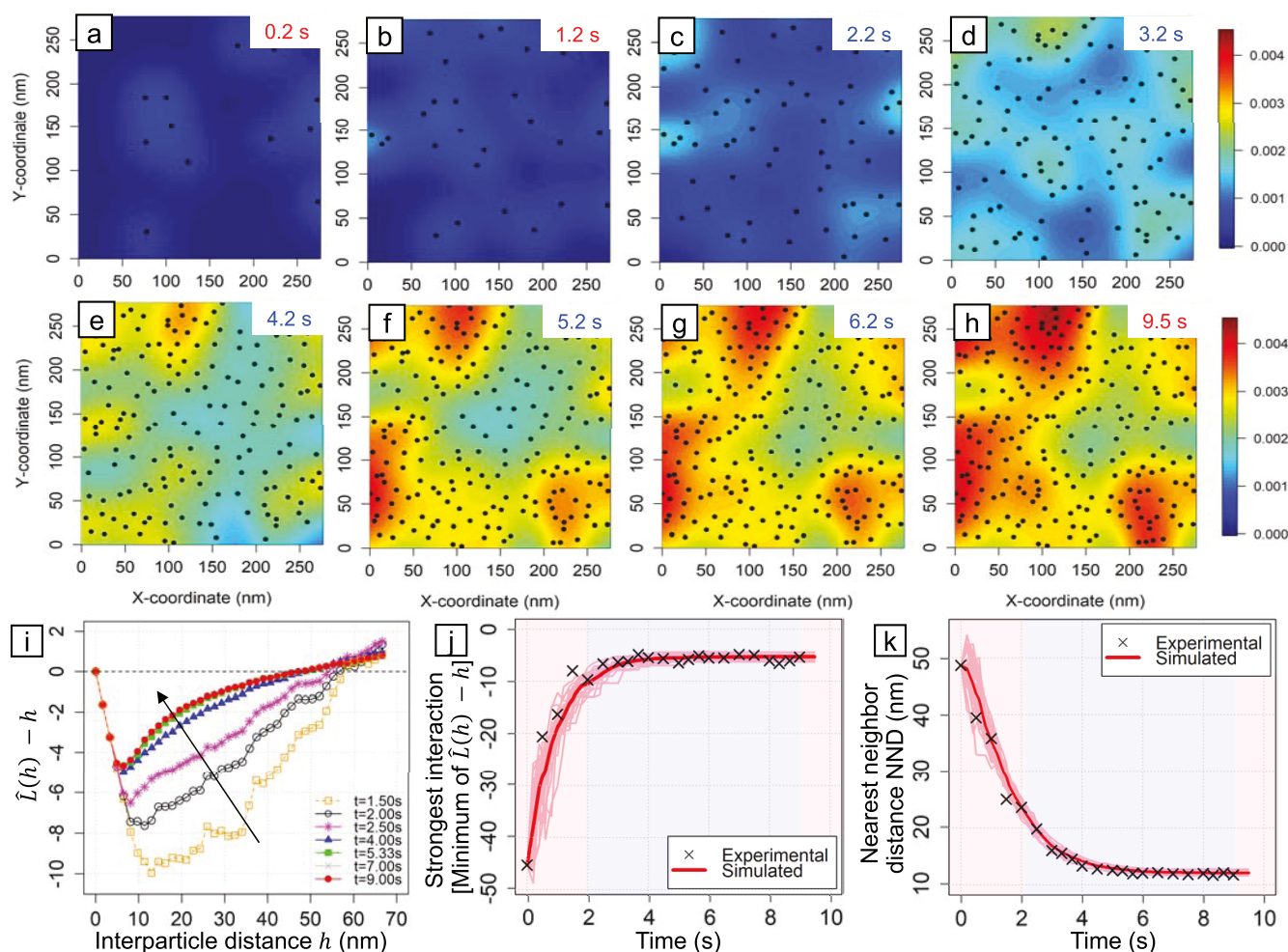


Figure 9. Simulation results for nanoparticle formation by dewetting: (a–f) spatial distribution of nanoparticles showing their locations and intensity (areal density) heat map for 0.2, 2.2, 3.2, 4.2, 5.2, and 9.5 s, respectively. (g) Time evolution of the value of the strongest interaction defined as the minimum value of $\hat{L}(h)$. (h) Time evolution of the estimated nearest-neighbor distance. Notice the strong agreement between the estimates in (g, h) from the Softcore model that are derived from experiments and those derived from simulations, for the time steps that were included in the 2–9 s range for which the Softcore model fits of experiments exist (shaded in blue), as well as for the time steps that are outside that time range (shaded in red).

provide a smaller number of point pattern data per image (i.e., unsuitable for obtained reliable fits for point process models), while lower-magnification conditions would make it more difficult to distinguish small nanoparticles from background noise (i.e., unsuitable for accurate detection of small nanoparticles), simulations that overcome this tradeoff limiting the possible spatial domain of experimental E-TEM data are desired.

To demonstrate this unique capability of spatially extending dewetting simulations, we conducted a case study wherein the spatial dimension of the window $\mathcal{W}$ is 4 times larger than that of the E-TEM image (i.e., the number of nanoparticles and window area are 16 times those observed in our E-TEM images), assuming the same Softcore model parameters and popping dynamics. Simulation results for the time evolution of simulated nanoparticle locations and areal density heat maps for one realization of this case study are shown in Figure S13 and Video S2. For comparison, we also ran another realization with the HPP model instead of Softcore, as shown in Figure S14. Importantly, the time evolution of the strongest interaction (i.e., the minimum value of $\hat{L}(h) - h$) and the average nearest-neighbor distance, shown in Figure S15,

demonstrate that the Softcore model simulations strongly agree with the experimental behavior as opposed to the HPP model, which significantly underestimates the strength of the dispersion interactions among nanoparticles during dewetting.

It is worth noting here that the extended simulations, for which the results are presented in Figures S13–S15, were carried out with a simulated initial state at $t = 0$ (by random sampling) and did not use the experimentally defined initial point pattern that was used for simulations in Figure 9. Moreover, a homogeneous mathematical formulation for $\beta^t(\mathbf{x})$ was adopted for all time steps. Despite these differences, the time evolution of the values of strongest interaction and nearest-neighbor distance for the Softcore simulations strongly matches experimental results, as shown in Figure S15. This finding further showcases the capabilities of our computational simulation tool. Hence, our machine-learned mathematical models not only mimic the actual experimental environment (and hence can be used as its surrogate) but further predict those dynamics at finer time scales and larger spatial areas, both of which may be experimentally infeasible or impractical to observe or capture.

We conclude this section by emphasizing that our computational simulation tool fills a significant gap in the literature, and we expect that it will prove to be of great utility, as there is currently no other accessible modeling approach that is capable of capturing both the coupled physical and chemical factors governing the dynamics of nanoparticle formation by thin film dewetting, wherein the combination of gas environment and high temperature results in metal oxide reduction and other surface chemical reactions (e.g., carbide formation). In particular, a simulation tool is needed to complement CNT nucleation and growth studies and modeling efforts aimed at capturing the mechanical interactions and bundling behavior among self-organizing CNTs,^{11,12} estimating forces ensuing during growth,^{10,70} and accurate modeling of the collective mechanical properties of CNT forest microstructures under compression.⁷¹ Moreover, the ability to understand and simulate the stochastic processes governing dewetting dynamics is crucial for catalyst design to achieve control on the size distribution, interparticle spacing, and areal density, which are sought after for high-performance electrical and thermal interfaces.^{15,72,73} In particular, recent research on spatial design of physical patterns for templating film dewetting, which is a capable approach for fabricating ordered catalyst nanoparticles,⁷⁴ will greatly benefit from our machine learning approach for simulating such spatially varying interactions during templated dewetting.

The machine learning-based methods presented in this paper represent a robust approach that can be further expanded to capture more physically/chemically relevant factors in its formulations. For example, running multiple E-TEM experiments with different film thicknesses, in different gas environments, or for different metal compositions can enable “learning” the effect of each of these physical/chemical factors explicitly in the model via physics-motivated intensity function parameterizations. Although one limitation of this approach is the inability to measure the time-evolving contact angles of nanoparticles from the bird-eye view E-TEM imaging described here, we can improve our modeling methodology by incorporating size evolution of the nanoparticles. Size information can be more rigorously taken into account through the so-called “marked point process models”, which have been recently used for diverse applications ranging from image processing⁷⁵ to climatology.⁷⁶ Constructing such models for nanomaterials science would open the door for an unprecedented data-driven ability to model and simulate complex interactions that are hard to capture by building model from first principles. This paradigm shift means that our ability to model a coupled nanoscale phenomenon is no longer limited by complexity of interactions, i.e., as long as we can obtain experimental data from E-TEM (or other *in situ* techniques), we can use statistical machine learning to model and simulate it in a physically interpretable manner.

4. CONCLUSIONS

Combining recent advances in *in situ* E-TEM with high-speed imaging enables generating high volumes of data over a broad spatial and temporal space. Hence, data science is set to play an instrumental role in creating a new comprehensive understanding of spatial and temporal evolution of nanoparticles with mathematical formulations that can capture hard to model coupled phenomena. In this study, we show that automated image processing of E-TEM images opens the door for extracting valuable point pattern data formats that provide

quantitative insights inaccessible by manual image analysis. We use point process theory, a branch of machine learning, to analyze the spatiotemporal behavior of nanoparticle formation by thin film dewetting. Our physically motivated statistical tests and models empirically demonstrate that nanoparticles do not form uniformly over space due to the inverse relation between areal density and nanoparticle size. More importantly, nanoparticles are shown to exhibit strong spatial dependence, particularly a dispersion behavior. Such dependencies appear to be distance- and time-dependent. The Softcore model, which is a Gibbs point process model, accurately describes this dispersion behavior, wherein the presence of a nanoparticle reduces the likelihood of others forming in its vicinity.

Combined with our proposed spatiotemporal interaction model (STIM), we have developed a computational simulation tool that can accurately mimic and further predict the spatiotemporal formation of nanoparticle ensembles at any arbitrary time point and any desired spatial domain, providing crucial insights to experimenters, whenever experiments are infeasible/impractical to conduct. Our work provides a robust foundation for future work on designing catalyst nanoparticle populations and leveraging their spatial interactions for controlling their collective behavior.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.0c07765>.

CVD microreactor for *in situ* study of catalyst nanoparticle formation by dewetting inside E-TEM; demonstration of inadequacy of off-the-shelf software packages for E-TEM image segmentation; time evolution of image-specific d_0 for all 22 E-TEM images; image segmentation results for all 22 E-TEM images; histograms of nanoparticle size distribution for all 22 segmented E-TEM images; formulations for $\lambda(s)$ with likelihood test ratio results for image #22; point process modeling results for the first four E-TEM images (i.e., at $t = 0.00$ – 2.00 s); Voronoi tessellations, density maps, and histograms of Voronoi distances for all 22 E-TEM images; algorithm for STIM; time evolution of the estimated values for the Softcore model parameters (both from fitting of experimental data and from STIM extrapolations); time evolution of Voronoi tessellation and Voronoi distances from simulation results; spatially and temporally extended simulation results using both softcore and HPP models with comparing the dynamics of strongest interaction and nearest-neighbor distance; videos of nanoparticle popping dynamics showing locations, areal density, and particle count as a function of time (PDF)

Time evolution of areal density heat maps and number of nanoparticles for our temporally extended simulations (first case study) (MP4)

Time evolution of areal density heat maps and number of nanoparticles for our spatially extended simulations (second case study) (MP4)

■ AUTHOR INFORMATION

Corresponding Author

Mostafa Bedewy – Department of Industrial Engineering,
Department of Chemical and Petroleum Engineering

Department of Mechanical Engineering and Materials
Science, University of Pittsburgh, Pittsburgh, Pennsylvania
15261, United States; orcid.org/0000-0003-4182-7533;
Email: mbedewy@pitt.edu

Author

Ahmed Aziz Ezzat – Department of Industrial & Systems
Engineering, Rutgers University, New Brunswick, New Jersey
08901-8554, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jpcc.0c07765>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank A. J. Hart (Massachusetts Institute of Technology), E. A. Stach (University of Pennsylvania), and D. N. Zakharov (Brookhaven National Laboratory). Financial support was provided by the National Science Foundation (NSF) under award number 1825772 (any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation). E-TEM work was carried out at the Center for Functional Nanomaterials, Brookhaven National Laboratory, which is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-SC0012704.

REFERENCES

- (1) Ro, I.; Resasco, J.; Christopher, P. Approaches for understanding and controlling interfacial effects in oxide-supported metal catalysts. *ACS Catal.* **2018**, *8*, 7368–7387.
- (2) Cuenya, B. R. Synthesis and catalytic properties of metal nanoparticles: Size, shape, support, composition, and oxidation state effects. *Thin Solid Films* **2010**, *518*, 3127–3150.
- (3) Cao, S.; Tao, F. F.; Tang, Y.; Li, Y.; Yu, J. Size- and shape-dependent catalytic performances of oxidation and reduction reactions on nanocatalysts. *Chem. Soc. Rev.* **2016**, *45*, 4747–4765.
- (4) van Deelen, T. W.; Mejía, C. H.; de Jong, K. P. Control of metal-support interactions in heterogeneous catalysts to enhance activity and selectivity. *Nat. Catal.* **2019**, *2*, 955–970.
- (5) Ono, L. K.; Roldan-Cuenya, B. Effect of interparticle interaction on the low temperature oxidation of CO over size-selected Au nanocatalysts supported on ultrathin TiC films. *Catal. Lett.* **2007**, *113*, 86–94.
- (6) Prieto, G.; Zečević, J.; Friedrich, H.; De Jong, K. P.; De Jongh, P. E. Towards stable catalysts by controlling collective properties of supported metal nanoparticles. *Nat. Mater.* **2013**, *12*, 34–39.
- (7) Yan, Y.; Miao, J.; Yang, Z.; Xiao, F.-X.; Yang, H. B.; Liu, B.; Yang, Y. Carbon nanotube catalysts: recent advances in synthesis, characterization and applications. *Chem. Soc. Rev.* **2015**, *44*, 3295–3346.
- (8) Bedewy, M.; Meshot, E. R.; Guo, H.; Verploegen, E. A.; Lu, W.; Hart, A. J. Collective mechanism for the evolution and self-termination of vertically aligned carbon nanotube growth. *J. Phys. Chem. C* **2009**, *113*, 20576–20582.
- (9) Bedewy, M.; Meshot, E. R.; Reinker, M. J.; Hart, A. J. Population growth dynamics of carbon nanotubes. *ACS Nano* **2011**, *5*, 8974–8989.
- (10) Bedewy, M.; Hart, A. J. Mechanical coupling limits the density and quality of self-organized carbon nanotube growth. *Nanoscale* **2013**, *5*, 2928–2937.
- (11) Balakrishnan, V.; Bedewy, M.; Meshot, E. R.; Pattinson, S. W.; Polsen, E. S.; Laye, F.; Zakharov, D. N.; Stach, E. A.; Hart, A. J. Real-Time Imaging of Self-Organization and Mechanical Competition in

Carbon Nanotube Forest Growth. *ACS Nano* **2016**, *10*, 11496–11504.

(12) Maschmann, M. R. Integrated simulation of active carbon nanotube forest growth and mechanical compression. *Carbon* **2015**, *86*, 26–37.

(13) Bedewy, M.; Farmer, B.; Hart, A. J. Synergetic chemical coupling controls the uniformity of carbon nanotube microstructure growth. *ACS Nano* **2014**, *8*, 5799–5812.

(14) Marconnet, A. M.; Panzer, M. A.; Goodson, K. E. Thermal conduction phenomena in carbon nanotubes and related nanostructured materials. *Rev. Mod. Phys.* **2013**, *85*, 1295.

(15) Esconjauregui, S.; Fouquet, M.; Bayer, B. C.; Ducati, C.; Smajda, R.; Hofmann, S.; Robertson, J. Growth of ultrahigh density vertically aligned carbon nanotube forests for interconnects. *ACS Nano* **2010**, *4*, 7431–7436.

(16) Bedewy, M. Data-driven understanding of collective carbon nanotube growth by in situ characterization and nanoscale metrology. *J. Mater. Res.* **2017**, *32*, 153–165.

(17) Thompson, C. V. Solid-state dewetting of thin films. *Annu. Rev. Mater. Res.* **2012**, *42*, 399–434.

(18) Hofmann, S.; et al. In situ observations of catalyst dynamics during surface-bound carbon nanotube nucleation. *Nano Lett.* **2007**, *7*, 602–608.

(19) Wirth, C. T.; Bayer, B. C.; Gamalski, A. D.; Esconjauregui, S.; Weatherup, R. S.; Ducati, C.; Baehz, C.; Robertson, J.; Hofmann, S. The phase of iron catalyst nanoparticles during carbon nanotube growth. *Chem. Mater.* **2012**, *24*, 4633–4640.

(20) Bedewy, M.; Viswanath, B.; Meshot, E. R.; Zakharov, D. N.; Stach, E. A.; Hart, A. J. Measurement of the dewetting, nucleation, and deactivation kinetics of carbon nanotube population growth by environmental transmission electron microscopy. *Chem. Mater.* **2016**, *28*, 3804–3813.

(21) He, M.; Liu, B.; Chernov, A. I.; Obratsova, E. D.; Kauppi, I.; Jiang, H.; Anoshkin, I.; Cavalca, F.; Hansen, T. W.; Wagner, J. B.; et al. Growth mechanism of single-walled carbon nanotubes on iron-copper catalyst and chirality studies by electron diffraction. *Chem. Mater.* **2012**, *24*, 1796–1801.

(22) Lin, M.; Ying Tan, J. P.; Boothroyd, C.; Loh, K. P.; Tok, E. S.; Foo, Y.-L. Direct observation of single-walled carbon nanotube growth at the atomistic scale. *Nano Lett.* **2006**, *6*, 449–452.

(23) Moseler, M.; Cervantes-Sodi, F.; Hofmann, S.; Csányi, G.; Ferrari, A. C. Dynamic catalyst restructuring during carbon nanotube growth. *ACS Nano* **2010**, *4*, 7587–7595.

(24) Yoshida, H.; Takeda, S.; Uchiyama, T.; Kohno, H.; Homma, Y. Atomic-scale in-situ observation of carbon nanotube growth from solid state iron carbide nanoparticles. *Nano Lett.* **2008**, *8*, 2082–2086.

(25) Meshot, E. R.; Verploegen, E.; Bedewy, M.; Tawfik, S.; Woll, A. R.; Green, K. S.; Hromalik, M.; Koerner, L. J.; Philipp, H. T.; Tate, M. W.; et al. High-speed in situ X-ray scattering of carbon nanotube film nucleation and self-organization. *ACS Nano* **2012**, *6*, 5091–5101.

(26) Jun, M.; Schumacher, C.; Saravanan, R. Global multivariate point pattern models for rain type occurrence. *Spatial Stat.* **2019**, *31*, No. 100355.

(27) Hering, A. S.; Bell, C. L.; Genton, M. G. Modeling spatio-temporal wildfire ignition point patterns. *Environ. Ecol. Stat.* **2009**, *16*, 225–250.

(28) Neeff, T.; Biging, G. S.; Dutra, L. V.; Freitas, C. C.; Dos Santos, J. R. Markov point processes for modeling of spatial forest patterns in Amazonia derived from interferometric height. *Remote Sens. Environ.* **2005**, *97*, 484–494.

(29) Martínez-Beneito, M. A.; Abellán, J. J.; López-Quílez, A.; Vanaclocha, H.; Zurriaga, O.; Jorques, G.; Fenollar, J. Source Detection in an Outbreak of Legionnaire's Disease. In *Case Studies In Spatial Point Process Modeling*; Springer: New York, 2006; Vol. 185, pp 169–182.

(30) Schoenberg, F. P.; Hoffmann, M.; Harrigan, R. J. A recursive point process model for infectious diseases. *Ann. Inst. Stat. Math.* **2019**, *71*, 1271–1287.

- (31) Diggle, P. J.; Eglén, S. J.; Troy, J. B. Modelling the Bivariate Spatial Distribution of Amacrine Cells. In *Case Studies in Spatial Point Process Modeling*; Springer: New York, 2006; pp 215–233.
- (32) Baddeley, A.; Turner, R. Modelling Spatial Point Patterns in R. In *Case Studies in Spatial Point Process Modeling*; Gregori, P.; Mateu, J.; Stoica, R.; Stoyan, D., Eds.; Springer: New York, 2006; Vol. 185, pp 23–74.
- (33) Zhou, Q.; Zeng, L.; Decicco, M.; Li, X.; Zhou, S. A comparative study on clustering indices for distribution uniformity of nanoparticles in metal matrix nanocomposites. *CIRP J. Manuf. Sci. Technol.* **2012**, *5*, 348–356.
- (34) Moreaud, M.; Jeulin, D.; Morard, V.; Revel, R. TEM image analysis and modelling: application to boehmite nanoparticles. *J. Microsc.* **2012**, *245*, 186–199.
- (35) Konomi, B. A.; Dhavala, S. S.; Huang, J. Z.; Kundu, S.; Huitink, D.; Liang, H.; Ding, Y.; Mallick, B. K.; et al. Bayesian object classification of gold nanoparticles. *Ann. Appl. Stat.* **2013**, *7*, 640–668.
- (36) Li, X.; Zhang, H.; Jin, J.; Huang, D.; Qi, X.; Zhang, Z.; Yu, D. Quantifying dispersion of nanoparticles in polymer nanocomposites through transmission electron microscopy micrographs. *J. Micro Nano-Manuf.* **2014**, *2*, No. 021008.
- (37) Dong, L.; Li, X.; Yu, D.; Zhang, H.; Zhang, Z.; Qian, Y.; Ding, Y. Quantifying nanoparticle mixing state to account for both location and size effects. *Technometrics* **2017**, *59*, 391–403.
- (38) Häbel, H.; Särkkä, A.; Rudemo, M.; Blomqvist, C. H.; Olsson, E.; Nordin, M. Colloidal particle aggregation in three dimensions. *J. Microsc.* **2019**, *275*, 149–158.
- (39) Park, C.; Ding, Y. Automating material image analysis for material discovery. *MRS Commun.* **2019**, *9*, 545–555.
- (40) Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66.
- (41) Lin, Y.-C.; Tsai, Y.-P.; Hung, Y.-P.; Shih, Z.-C. Comparison between immersion-based and toboggan-based watershed image segmentation. *IEEE Trans. Image Process.* **2006**, *15*, 632–640.
- (42) Quelhas, P.; Marcuzzo, M.; Mendonça, A. M.; Campilho, A. Cell nuclei and cytoplasm joint segmentation using the sliding band filter. *IEEE Trans. Med. Imaging* **2010**, *29*, 1463–1473.
- (43) Chan, T. F.; Vese, L. A. Active contours without edges. *IEEE Trans. Image Process.* **2001**, *10*, 266–277.
- (44) Parvin, B.; Yang, Q.; Han, J.; Chang, H.; Rydberg, B.; Barcellos-Hoff, M. H. Iterative voting for inference of structural saliency and characterization of subcellular events. *IEEE Trans. Image Process.* **2007**, *16*, 615–623.
- (45) Qian, Y.; Huang, J.; Ding, Y. Identifying multi-stage nanocrystal growth using insitu TEM video data. *IIEE Trans.* **2017**, *49*, 532–543.
- (46) Baddeley, A.; Rubak, E.; Turner, R. *Spatial point patterns: methodology and applications with R*; CRC Press, 2015.
- (47) Robert, C.; Casella, G. *Monte Carlo statistical methods*; Springer Science & Business Media, 2013.
- (48) Williams, C. K.; Rasmussen, C. E. *Gaussian processes for machine learning*; MIT Press, Cambridge, MA, 2006; Vol. 2.
- (49) Aziz Ezzat, A.; Tang, J.; Ding, Y. A model-based calibration approach for structural fault diagnosis using piezoelectric impedance measurements and a finite element model. *Struct. Health Monit.* **2020**, *19*, 1839–1855.
- (50) Santner, T. J.; Williams, B. J.; Notz, W. I.; Williams, B. J. *The design and analysis of computer experiments*; Springer, 2003; Vol. 1.
- (51) Dennis, J. E., Jr.; Gay, D. M.; Welsch, R. E. Algorithm 573: NL2SOL—an adaptive nonlinear least-squares algorithm. *ACM Trans. Math. Softw.* **1981**, *7*, 369–383.
- (52) Brune, H. Microscopic view of epitaxial metal growth: nucleation and aggregation. *Surf. Sci. Rep.* **1998**, *31*, 125–229.
- (53) Zinke-Allmang, M. Phase separation on solid surfaces: nucleation, coarsening and coalescence kinetics. *Thin Solid Films* **1999**, *346*, 1–68.
- (54) Voorhees, P. W. The theory of Ostwald ripening. *J. Stat. Phys.* **1985**, *38*, 231–252.
- (55) Amama, P. B.; Pint, C. L.; McJilton, L.; Kim, S. M.; Stach, E. A.; Murray, P. T.; Hauge, R. H.; Maruyama, B. Role of water in super growth of single-walled carbon nanotube carpets. *Nano Lett.* **2009**, *9*, 44–49.
- (56) Amama, P. B.; Pint, C. L.; Kim, S. M.; McJilton, L.; Eyink, K. G.; Stach, E. A.; Hauge, R. H.; Maruyama, B. Influence of alumina type on the evolution and activity of alumina-supported Fe catalysts in single-walled carbon nanotube carpet growth. *ACS Nano* **2010**, *4*, 895–904.
- (57) Viomy, J. L.; Beysens, D.; Knobler, C. M. Scaling description for the growth of condensation patterns on surfaces. *Phys. Rev. A* **1988**, *37*, 4965.
- (58) Rogers, T.; Elder, K.; Desai, R. C. Numerical study of the late stages of spinodal decomposition. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 9638.
- (59) Rao, C. R. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Cambridge Philos. Soc.* **1948**, *44*, 50–57.
- (60) Family, F.; Meakin, P. Scaling of the droplet-size distribution in vapor-deposited thin films. *Phys. Rev. Lett.* **1988**, *61*, 428.
- (61) Zinke-Allmang, M.; Feldman, L. C.; Van Saarloos, W. Experimental study of self-similarity in the coalescence growth regime. *Phys. Rev. Lett.* **1992**, *68*, 2358.
- (62) Liu, F.; Li, A. H.; Lagally, M. Self-assembly of two-dimensional islands via strain-mediated coarsening. *Phys. Rev. Lett.* **2001**, *87*, No. 126103.
- (63) Krejci, A. J.; Thomas, C. G.; Mandal, J.; Gonzalo-Juan, I.; He, W.; Stillwell, R. L.; Park, J.-H.; Prasai, D.; Volkov, V.; Bolotin, K. I.; et al. Using Voronoi tessellations to assess nanoparticle-nanoparticle interactions and ordering in monolayer films formed through electrophoretic deposition. *J. Phys. Chem. B* **2013**, *117*, 1664–1669.
- (64) Thiele, U. Open questions and promising new fields in dewetting. *Eur. Phys. J. E: Soft Matter* **2003**, *12*, 409–416.
- (65) Li, J.; Rojo, A.; Sander, L. M. Anomalous dimension and spatial correlations in a point-island model. *Phys. Rev. Lett.* **1997**, *78*, 1747.
- (66) Ammi, H. S.; Chame, A.; Touzani, M.; Benyoussef, A.; Pierre-Louis, O.; Misbah, C. Irreversible aggregation of interacting particles in one dimension. *Phys. Rev. E* **2005**, *71*, No. 041603.
- (67) Mahmoudi, M.; Aziz Ezzat, A.; Elwany, A. Layerwise anomaly detection in laser powder-bed fusion metal additive manufacturing. *J. Manuf. Sci. Eng.* **2019**, *141*, 031002.
- (68) Aziz Ezzat, A.; Jun, M.; Ding, Y. Spatio-temporal short-term wind forecast: A calibrated regime-switching method. *Ann. Appl. Stat.* **2019**, *13*, 1484–1510.
- (69) Aziz Ezzat, A.; Pourhabib, A.; Ding, Y. Sequential design for functional calibration of computer models. *Technometrics* **2018**, *60*, 286–296.
- (70) Hajilounzhad, T.; Ajiboye, D. M.; Maschmann, M. R. Evaluating the forces generated during carbon nanotube forest growth and self-assembly. *Materialia* **2019**, *7*, No. 100371.
- (71) Stein, I. Y.; Lewis, D. J.; Wardle, B. L. Aligned carbon nanotube array stiffness from stochastic three-dimensional morphology. *Nano-scale* **2015**, *7*, 19426–19431.
- (72) Lee, J.; Abdulhafez, M.; Bedewy, M. Decoupling Catalyst Dewetting, Gas Decomposition, and Surface Reactions in Carbon Nanotube Forest Growth Reveals Dependence of Density on Nucleation Temperature. *J. Phys. Chem. C* **2019**, *123*, 28726–28738.
- (73) Dee, N. T.; Li, J.; White, A. O.; Jacob, C.; Shi, W.; Kidambi, P. R.; Cui, K.; Zakharov, D. N.; Janković, N. Z.; Bedewy, M.; et al. Carbon-assisted catalyst pretreatment enables straightforward synthesis of high-density carbon nanotube forests. *Carbon* **2019**, *153*, 196–205.
- (74) Altomare, M.; Nguyen, N. T.; Schmuki, P. Templated dewetting: designing entirely self-organized platforms for photocatalysis. *Chem. Sci.* **2016**, *7*, 6865–6886.
- (75) Kim, D. W.; Aguilar, C.; Zhao, H.; Comer, M. L. Narrow Gap Detection in Microscope Images Using Marked Point Process Modeling. *IEEE Trans. Image Process.* **2019**, *28*, 5064–5076.
- (76) Xiao, S.; Kottas, A.; Sansó, B. Modeling for seasonal marked point processes: An analysis of evolving hurricane occurrences. *Ann. Appl. Stat.* **2015**, *9*, 353–382.