RESEARCH ARTICLE

Parameter Optimization on a Tessellation Model for Crack Pattern Simulation

ROBERTO LEÓN[1], ELIZABETH MONTERO[2], (Member, IEEE), AND WERNER NAGEL[3]

[1]Department of Computer Science, Universidad Técnica Federico Santa María, Santiago 8940897, Chile
[2]Department of Computer Science, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile
[3]Institut für Stochastik, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany

Corresponding author: Elizabeth Montero (elizabeth.montero@usm.cl)

ABSTRACT The capabilities of a parametric model for crack patterns simulation are presented. Planar tessellations are partitions of the plane into convex polygons (called cells) without overlapping. The Voronoi tessellations and Poisson line tessellations are the most prominent models; however, to model crack patterns, it is more appropriate to deal with tessellations that are generated by a cell division process. We describe the STIT tessellation as a reference model for crack patterns and introduce several modifications. Having described a variety of 40 parametric models and appropriate simulation algorithms, we delineate and specify tuning methods to optimize the adaption of the model to real crack pattern data. An example of a metalized polydimethylsiloxane demonstrates the capability of our approach. The results indicate that this approach yields a considerable improvement in modeling compared to previous studies.

INDEX TERMS Crack pattern, random tessellation, STIT tessellation, spatial statistics, metaheuristic tuning methods.

I. INTRODUCTION

Crack patterns appear and are studied in nanotechnology, materials science, soft matter, and geology. Their length scales can vary widely, but they share some essential features. This motivates the development of models for such structures. In several research areas, mathematical models are developed for such patterns. Often, the features of these models can be studied – up to now – by simulation only. Crack pattern simulation is an approach designed to emulate the generation and evolution of structures where different kinds of failures, fissures, or cracks arise. Recently, such models are also of growing interest in the context of machine learning, see [1] and the references therein.

In the present paper, we start with the STIT tessellation as a reference model and consider modifications to it. This was already commenced in [2]. Now we add some further model refinements, which we call ASA ('avoid small angles') and RD ('roundness optimization').

The aim is to provide parametric models such that the parameters have a geometric meaning and that efficient

optimization methods can adapt a model to an actual structure as well as possible.

First, we describe the planar STIT tessellation and then explain its Monte Carlo simulation.

In the second part of the paper, we introduce the mentioned modifications of the STIT model, which are more flexible and thus potentially allow for a better adaption to actual crack patterns. Furthermore, we define the statistics for random tessellations used to evaluate a model's adequacy.

For this variety of parametric models, a tuning process is described, which allows for the best adaption of a model with respect to given optimization criteria.

In the last part of the paper, we consider data of a real crack pattern in a thin chromium/gold film on an elastomer (polydimethylsiloxane, PDMS). A tensile stress in the layered material leads to film cracking. We check how well certain statistical features can be fitted by the introduced models and the optimization of parameter values.

II. MATERIAL AND METHODS

The material we refer to in the present study is a real crack pattern in a thin chromium/gold film on an elastomer (polydimethylsiloxane, PDMS), see Figure 13. And we

The associate editor coordinating the review of this manuscript and approving it for publication was Mingbo Zhao.

assume that the methods presented here can be applied to a wide variety of crack patterns.

Various mathematical models have already been developed for the simulation of crack patterns. The literature contains numerous examples of cracking and fracturing, ranging from geology and materials science to soft matter and nanotechnology. Often, these papers focus on the genesis and physics of an individual crack, e.g., [3], and [4]. However, there are also approaches to whole crack patterns, e.g., [5], [6], [7], [8], [9], [10], and [11]. A promising approach is based on stochastic geometry, particularly on models for random tessellations. A tessellation (mosaic) is a division of the plane into polygons or of the three-dimensional space into polyhedra [12]. A motivation for modeling crack structures is to study the relationship between the geometric structure on one side and the physical properties on the other. Thus, models including quantitative parameters which have a geometric or physical meaning are of particular interest. This can also potentially support the design of new materials.

A. THE STIT MODEL

Denoting by $\mathbb{R}^2$ the Euclidean plane and by $\mathcal{P}$ the set of all two-dimensional convex polygons, a subset $T \subset \mathcal{P}$ is called a **tessellation** if:
  (i) the polygons fill the plane, that means $\bigcup_{z \in T} z = \mathbb{R}^2$,
  (ii) the polygons do not overlap; more precisely, for all $z, z' \in T$, if $z \neq z'$, then $\operatorname{int} z \cap \operatorname{int} z' = \emptyset$, where $\operatorname{int} z$ denotes the topological interior of z,
  (iii) T is locally finite, that means the set $\{z \in T : z \cap C \neq \emptyset\}$ is finite for all compact sets $C \subset \mathbb{R}^2$.
A random tessellation is a random variable with values in the set of all tessellations, see also [12].

The polygons forming a tessellation are called **cells**. The Voronoi tessellations and Poisson line tessellations are the most prominent models. To model crack patterns, it is more appropriate to deal with tessellations that are generated by a consecutive division of cells – a cell division process. Inspired by [13], we focus on the class of processes that can be characterized by
  L The rule for the random **lifetime** of a cell that is the time between the birth of a cell by division of a mother cell and the division of the cell.
  D The rule for the random **division** of a cell at the end of its lifetime.

For such models, we will assume that an extant cell's lifetime and division depend only on the cell itself, neither the adjacent cells nor the history of the division process.

The STIT model in a Euclidean space of arbitrary dimension – a random tessellation whose distribution is **ST**able under the operation **IT**eration of tessellations – was first introduced in [14]. This stochastic stability is an essential property that allows for many theoretical results. Here we consider the two-dimensional case only.

The main ingredient of a STIT tessellation is a directional distribution, that is, a probability distribution φ on the interval $[0, \pi)$. Up to a scaling factor, the choice of φ determines a STIT tessellation in the following way:

Let $\mathcal{H}$ denote the set of all lines in the plane. A line $H = H(\alpha, r) \in \mathcal{H}$ is parameterized by its normal direction $\alpha \in [0, \pi)$ and its signed distance r from the origin, where the distance has a positive sign if the intersection of the line with its orthogonal subspace is in the upper half-plane. For a line, H that does not contain the origin, denote by H^+ and H^- the closed half-planes generated by H, where H^+ contains the origin. In our context, the random lines contain the origin with probability zero. For a set $B \subset \mathbb{R}^2$ denote by $[B] := \{H \in \mathcal{H} : H \cap B \neq \emptyset\}$ the set of all lines intersecting B.

Based on φ, a translation invariant measure Θ on $\mathcal{H}$ is determined by

$$\int_{\mathcal{H}} f(H)\Theta(dH) = \int_{[0,\pi)} \int_{\mathbb{R}} f(H(\alpha, r))\, dr\, \varphi(d\alpha) \quad (1)$$

for any non-negative measurable function $f : \mathcal{H} \to [0, \infty)$, see Section 4.4 of [15]. An interpretation of this formula and its application in the Monte Carlo simulation is given below. Throughout the paper, it is assumed that φ is not concentrated on a single value. This guarantees that the constructed object is a random tessellation; see [15].

For a convex polygon $z \in \mathcal{P}$ we define a probability distribution $\Theta_{[z]}$ on the set $[z]$ of all lines which intersect z by

$$\Theta_{[z]}(\cdot) := \Theta(\cdot \cap [z])/\Theta([z]).$$

An informal description of the STIT tessellation process is given by the following specifications of the lifetime distribution L and of the division rule D:
  • **L-STIT:** A cell z has a random lifetime that is exponentially distributed with parameter $\Theta([z])$.
  • **D-STIT:** At the end of its lifetime, z is divided by a random line H with law $\Theta_{[z]}$, independent of the lifetime, and conditionally independent, given z, of all the dividing lines used before.
More precisely, we describe the planar STIT tessellations in a convex polygon $W \subset \mathbb{R}^2$ called a window.

Let $\mathbb{N} := \{1, 2, \ldots\}$ denote a set of positive integers. The birth time of a cell z is denoted by $\beta(z)$.

Definition 1: Let $\underline{\tau} = (\tau_n : n \in \mathbb{N})$ be a sequence of independent and identically distributed (i.i.d.) random variables, exponentially distributed with parameter 1. The STIT tessellation process $(Y_{t,W} : t \geq 0)$ in W, driven by the measure Θ is defined by
  (a) *Initial setting.*
      $Y_{t,W} = \{W\}$ *for* $0 \leq t < \tau_1/\Theta([W])$, $\beta(W) = 0$ *and* $z_1 = W$.
  (b) *Recursion.*
      For $t > 0$, let be $Y_{t,W} = \{z_{i_1}, \ldots, z_{i_n}\}$, that is $\beta(z_{i_k}) < t$ and $\beta(z_{i_k}) + \tau_{i_k}/\Theta([z_{i_k}]) \geq t$ for $k = 1, \ldots, n$.
      Define $i^* \in \{i_1, \ldots, i_n\}$ as the index of the cell which is the next to be divided by

$$t_{i^*} = \beta(z_{i^*}) + \tau_{i^*}/\Theta([z_{i^*}])$$

$$= \min \left\{ \beta(z_{i_k}) + \tau_{i_k} / \Theta([z_{i_k}]) : k = 1, \ldots, n \right\}.$$

Then $Y_{t,W}$ remains constant until the jump at time t_{i*}, when the process jumps, by division of the cell z_{i*} into the state

$$Y_{t_{i*},W} = \left(\{z_{i_1}, \ldots, z_{i_n}\} \setminus \{z_{i*}\} \right) \cup \{z_{2n}, z_{2n+1}\}$$

with $z_{2n} = z_{i*} \cap H_{i*}^+$, $z_{2n+1} = z_{i*} \cap H_{i*}^-$, where H_{i*} is a random line with law $\Theta_{[z_{i*}]}$ and independent of $\underline{\tau}$ and conditionally independent, given z_{i*}, of the $n-1$ dividing lines used before.

The birth time of the dividing line and the new cells is $\beta(H_{i*}) = \beta(z_{2n}) = \beta(z_{2n+1}) := t_{i*}$.

This definition looks rather involved, but its advantage is that it provides an algorithm for the STIT construction.

Note that $(Y_{t,W} : t \geq 0)$ is a random *process*, and for all $t > 0$, the value $Y_{t,W}$ is the cutout for the window W, of a STIT tessellation of the whole plane.

To see how the law $\Theta([z])$ for a cell division must be implemented correctly, note the following. For a convex polygon z and a fixed direction $0 \leq \alpha < \pi$, let be $h_0(z, \alpha) < h_1(z, \alpha)$ the two values of the signed distances of the tangential lines to z with normal direction α, which are values of the support function of z. Formally, $h_0(z, \alpha) := \min\{x \cos \alpha + y \sin \alpha : (x, y) \in z\}$ and $h_1(z, \alpha) := \max\{x \cos \alpha + y \sin \alpha : (x, y) \in z\}$. Hence, a line $H(\alpha, r)$ divides z, that is $H(\alpha, r) \in [z]$, if and only if $h_0(z, \alpha) < r < h_1(z, \alpha)$. The width (or breadth) of z in direction α is $b(z, \alpha) := h_1(z, \alpha) - h_0(z, \alpha)$ which is the distance of the two supporting (tangential) lines to z with the normal direction α. The maximum width of z is denoted by $b_{\max}(z) := \max_{0 \leq \alpha < \pi} b(z, \alpha)$, and the minimum width of z is denoted by $b_{\min}(z) := \min_{0 \leq \alpha < \pi} b(z, \alpha)$. By $1\{\cdot\}$, we denote the indicator function, which is 1, if the condition in $\{\cdot\}$ is satisfied and 0 otherwise. For $0 < \alpha_0 \leq \pi$ and a convex polygon z, equation (1) yields

$$\Theta \left(\{H(\alpha, r) \in [z] : 0 \leq \alpha < \alpha_0\} \right)$$
$$= \int_{[0, \pi)} 1\{0 \leq \alpha < \alpha_0\} \, b(z, \alpha) \, \varphi(\mathrm{d}\alpha) \quad (2)$$

and in particular for $\alpha_0 = \pi$,

$$\Theta([z]) = \int_{[0, \pi)} b(z, \alpha) \, \varphi(\mathrm{d}\alpha). \quad (3)$$

Hence, $\Theta([z])$ can be understood as the φ-weighted mean width of z. If φ is the uniform distribution on $[0, \pi)$ then $\Theta([z]) = P(z)/\pi$, where P denotes the perimeter.

Remark 1: By (3), the directional distribution of a line that intersects a polygon z is not φ itself, but it is φ endowed with the density $b(z, \alpha)/\Theta([z])$. This must be taken into account when a dividing line is generated in the Monte Carlo simulation of a STIT tessellation. Given a direction α, the position (shift or translation) of the dividing line is uniformly distributed in the respective interval $(h_0(z, \alpha), h_1(z, \alpha))$

When a cell z_i is divided, it is replaced by its two daughter cells $z_i \cap H_i^+$ and $z_i \cap H_i^-$ with birth time $\beta(z_i \cap H_i^+) = \beta(z_i \cap H_i^-) = \beta(z_i) + \tau_i / \Theta([z_i])$.

A crucial feature of the construction is that, by (b) of Definition 1, the lifetime of a cell z_i is $\tau_i / \Theta([z_i])$, which means that it is exponentially distributed with parameter $\Theta([z_i])$. Thus by (3), the distribution of the lifetime of a cell depends on its size – smaller cells have a longer expected lifetime than larger ones. In the particular case when the measure Θ is isotropic, that is, φ is the uniform distribution on $[0, \pi)$, the parameter of the exponential lifetime distribution of a cell z is $\Theta([z]) = P(z)/\pi$, the perimeter of z divided by π.

Remark 2: Even if this construction is performed in a fixed and bounded window W, it provides a spatially consistent distribution in the following sense. Let $W' \subset W$ be a convex polygon, $(Y_{t,W'} : t > 0)$ and $(Y_{t,W} : t > 0)$ the STIT tessellation processes generated in W' and W, respectively. The symbol $\overset{D}{=}$ stands for the identity of distributions of random variables. Then for all $t > 0$ we have that

$$Y_{t,W'} \overset{D}{=} \{z \cap W' : z \in Y_{t,W}, W' \cap \text{int } z \neq \emptyset\}, \quad (4)$$

which means that the restriction of $Y_{t,W}$ to W' has the same distribution as $Y_{t,W'}$, see [14]. This consistency property yields that, for any $t > 0$, there exists a spatially stationary (or homogeneous, which means the invariance of the distribution under translations of the Euclidean plane) random tessellation Y_t of $\mathbb{R}^2$ such that the restriction of Y_t to W has the same distribution as $Y_{t,W}$ for all polygons $W \in \mathcal{P}$. Note that this spatial consistency is lost when the STIT model is modified, and hence the distribution of the tessellation generated in a window depends on the choice of this window, see [16].

The realizations of planar STIT tessellations suggest that they can be potential models for crack or fracture patterns. However, already tentative studies in [17] or [18] indicated that the STIT tessellations are not appropriate for some real patterns. This is not surprising because the STIT model emerged from purely mathematical ideas. Thus an adaption of the cell division model to such crack patterns is necessary. We have chosen a "phenomenological approach", aiming to model the geometric appearance, not focusing on the physics of crack formation.

B. MODIFICATIONS OF THE STIT MODEL

Some modifications of the STIT model were introduced in [2].

In all the models, at first, a **directional distribution** φ on $[0, \pi)$ has to be fixed. The following directional distributions are used in the simulation study:

- **ISO:** The *isotropic distribution*, which means that φ is the uniform distribution on the interval $[0, \pi)$.
- **DISCR:** A *discrete distribution* with finitely many directions, that is $\varphi = \sum_{i=1}^{k} p_i \delta_{\alpha_i}$, $2 \leq k$, with $0 \leq \alpha_i < \pi$ and probabilities $0 < p_i < 1$, $\sum_{i=1}^{k} p_i = 1$. By δ_α we denote the Dirac measure concentrated on a single value $0 \leq \alpha < \pi$, that is $\delta_\alpha(B) = 1$ if $\alpha \in B$ and 0 otherwise, for a subset $B \subseteq [0, \pi)$. In the generated

tessellation, segments with normal directions $\alpha_1, \ldots, \alpha_k$ appear. Once the number k of directions is chosen, for example, $2 \leq k \leq 32$, one has to indicate the values α_i as the radian divided by π; for example, the input 0.5 means the angle $0.5\,\pi = 90°$. And for each α_i, its probability p_i has to be chosen as well.

- **DDISCR:** A *disturbed discrete distribution*, that is $\varphi = \left(\sum_{i=1}^{k} p_i \delta_{\alpha_i} \right) * \psi$, where ψ is the elliptic distribution with parameter b_{ellip}, see ELLIP below. The $*$ denotes the convolution of measures. This means that a random 'disturbance' with an elliptic distribution is added to a direction chosen from a discrete distribution DISCR.
- **RECT:** The discrete distribution with horizontal and vertical directions only, and both with the same probability, which means that $\varphi = 0.5\,\delta_0 + 0.5\,\delta_{\pi/2}$, which is a particular case of DISCR. The tessellation consists of random rectangles.
- **DRECT:** The discrete distribution RECT which is disturbed by the elliptic distribution ELLIP with parameter b_{ellip}. This is a particular case of DDISCR.
- **ELLIP:** An *elliptic distribution*. Consider an ellipse with a horizontal half-axis of length 1 and a vertical half-axis of length $0 < b_{\text{ellip}} < 1$. Then the cumulative distribution function of φ is defined as

$$F_\varphi(\alpha) = \frac{\text{area of ellipse sector } [0, \alpha]}{\text{half area of the ellipse}}, \quad 0 \leq \alpha < \pi.$$

Notice that we obtain the isotropic distribution ISO for the value $b_{\text{ellip}} = 1$.

Then, based on the directional distribution φ, a **division rule** is chosen.

- **D-STIT:** For a cell z, the direction α of the dividing line is generated according to the distribution $b(z, \alpha)\varphi / \Theta([z])$ which is the distribution φ, endowed with the density $b(z, \alpha)/\Theta([z])$, see Remark 1. Then, the signed distance from the origin of the dividing line is uniformly distributed in the interval $(h_0(z, \alpha), h_1(z, \alpha))$.
- **D-GAUSS:** For a cell z, the direction α of the dividing line is generated similarly for D-STIT. Then, the signed distance r from the origin is distributed in the interval $(h_0(z, \alpha), h_1(z, \alpha))$ according to a truncated (to this interval) Gaussian distribution with a standard deviation $\sigma\,b(z, \alpha) = \sigma\,(h_1(z, \alpha) - h_0(z, \alpha))$ and mean $0.5\,(h_0(z, \alpha) + h_1(z, \alpha))$. The parameter σ is in the range $0 < \sigma < 1$.

In contrast to D-STIT, the division rule D-GAUSS takes into account the observation that the cracks in real structures tend to be 'central' when they divide a cell. In D-GAUSS, the parameter $\sigma > 0$ has to be chosen, and the smaller its value, the more concentrated to the center of the interval $[h_0(z, \alpha), h_1(z, \alpha)]$ is the random position of the dividing line.

An alternative rule to L-STIT was introduced for the **lifetime distribution**:

- **L-STIT:** A cell z has a random lifetime that is exponentially distributed with parameter $\Theta([z])$.

- **L-AREA:** The lifetime of a cell z is exponentially distributed with parameter $A(z)$, where A denotes the area.

The rule L-AREA is motivated by observation in [17], which indicates that the coefficient of variance of the area, CV(A), see III-A, of STIT tessellations, is much larger than that one of data from crack patterns. Therefore, it seems reasonable to base the lifetime rule on the area rather than on the width or perimeter of the cells. Simulation results of [2] clearly shows that CV(A) is indeed significantly reduced when L-AREA is applied.

Whereas there are numerous theoretical results for STIT tessellations, the modified models can – up to now – be investigated by simulation studies only.

By Remark 1, the simulation of the direction α of the line dividing a cell z requires special attention because it is not correct to generate the direction of the dividing line directly from φ. In our simulations, we apply a rejection method. For a given cell z, generate a direction α according to φ and a random number d which is uniformly distributed in the interval $[h_0(z, \alpha), h_0(z, \alpha) + b_{\max}(z)]$. The simulated line $H(\alpha, d)$ is *accepted* as a dividing line if $d < h_1(z, \alpha)$, otherwise, it is *rejected*. Recall that $h_0(z, \alpha) < h_1(z, \alpha)$ are the two signed distances of the tangential lines to z with normal direction α.

To simulate the lifetime of an extant cell z, the integral in (3) is discretized and replaced by a sum using 128 equidistant angles in $[0, \pi)$.

In our simulation study, the window W is a square of side length $a \in \mathbb{N}$, and the process of consecutive cell division runs to a fixed time $t_{\text{STOP}} \in \mathbb{N}$, at which the construction stops. To obtain a reasonable number of cells in a simulation, the value of t_{STOP} should be chosen in an appropriate relation to a. Furthermore, because the models with D-GAUSS are not spatially consistent, the dependence on the choice of a is essential. Therefore, we decided first to launch a cell division process with D-STIT and L-STIT until time t_{STOP} and then to switch to D-GAUSS and the chosen L-rule for a time interval of length t_{MOD}, see (ii) at the end of Subsection II-C.

Some examples of simulations for different directional distributions and lifetime rules with D-STIT and D-GAUSS are shown in Figure 1 and 2 respectively.

C. NEW MODIFICATIONS OF THE STIT MODEL

Observations of real crack structures appearing in material, see Figure 3 motivate further modifications of the L- and D-rules.

In the present paper, we introduce alternatives to D-STIT and D-GAUSS. The aim is to generate cells with larger values of the mean isoperimetric quotient (or roundness) RD and the mean aspect ratio AR, see Subsection III-A.

- **D-RDMIN:** For a cell z and a direction α, the signed parameter r of the dividing line is chosen in the interval $(h_0(z, \alpha), h_1(z, \alpha))$ such that for the daughter cells z_1 and z_2, that are generated by the division of z, the value of $\min\{RD(z_1), RD(z_2)\}$ is maximized.
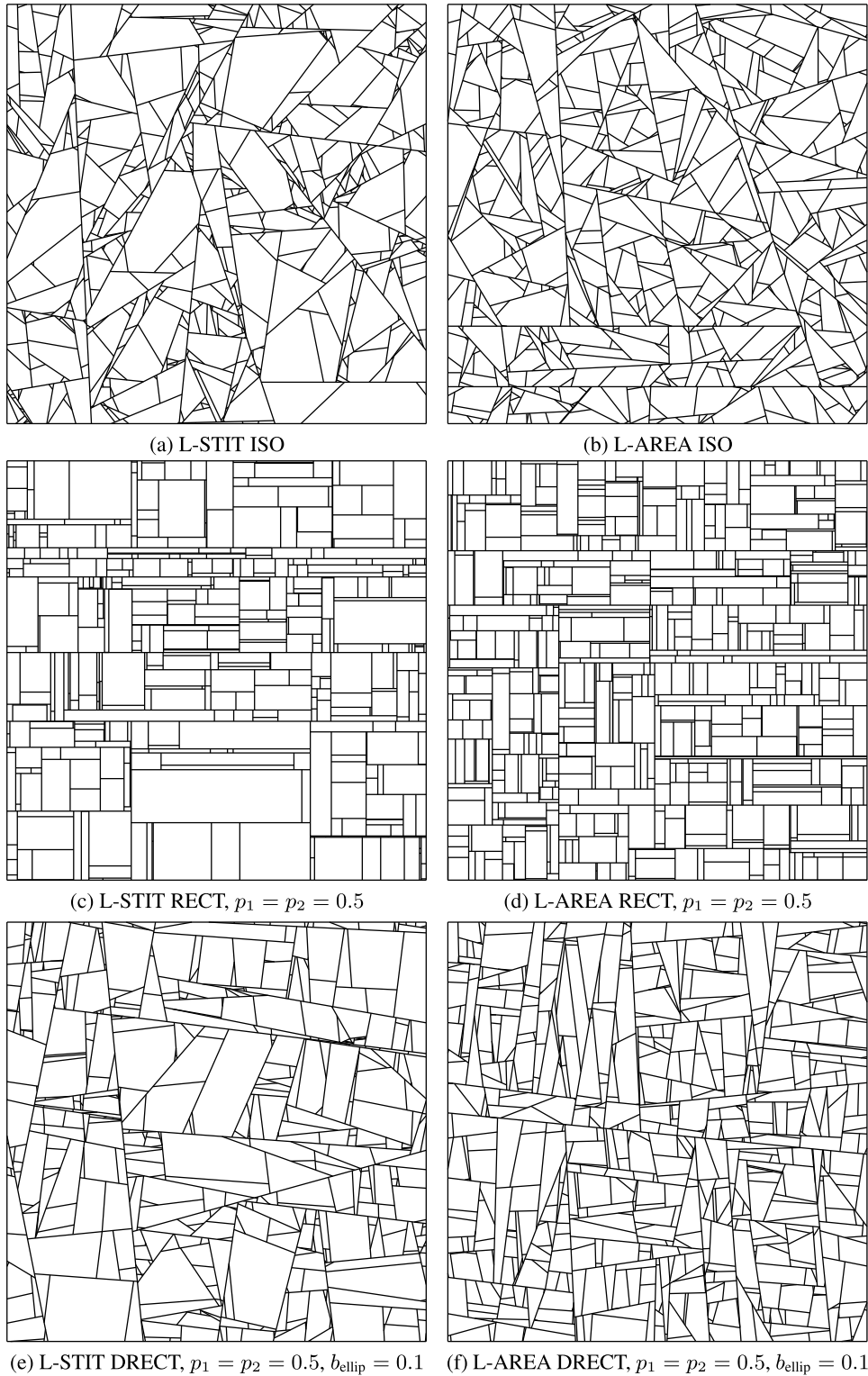
(a) L-STIT ISO

(b) L-AREA ISO

(c) L-STIT RECT, $p_1 = p_2 = 0.5$

(d) L-AREA RECT, $p_1 = p_2 = 0.5$

(e) L-STIT DRECT, $p_1 = p_2 = 0.5$, $b_{\text{ellip}} = 0.1$

(f) L-AREA DRECT, $p_1 = p_2 = 0.5$, $b_{\text{ellip}} = 0.1$

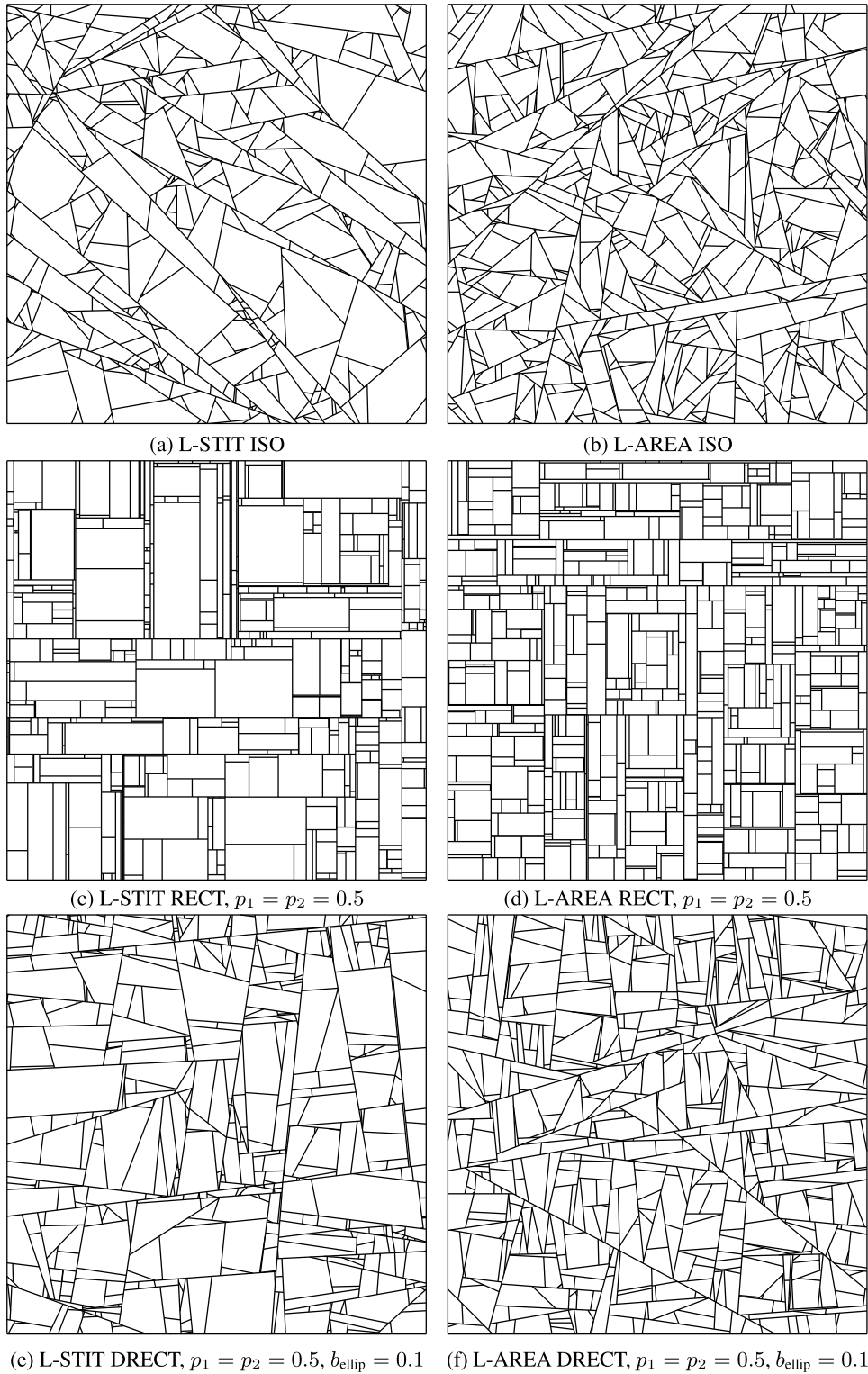**FIGURE 1.** Simulations of tessellations with division rule D-STIT and two different lifetime rules: L-STIT (a, c, and e) with $t_{\text{STOP}} = 40$, and L-AREA (b, d, and f) with $t_{\text{STOP}} = 650$.

- **D-RDSSQ:** Analogous to D-RDMIN, but here the value of the sum of squares $RD(z_1)^2 + RD(z_2)^2$ is maximized.

For a given direction α, consider a family of M parallel lines h_j for $j = 1, \ldots, M$ that divide a cell z. Let z_1^j

and z_2^j be the cells generated by the line h_j and $RD(z_1^j)$ and $RD(z_2^j)$ the roundness of each cell, respectively. Then, in the case of D-RDMIN the line h_j is chosen, where $\min\{RD(z_1^j), RD(z_2^j)\}$ is maximal. In the case of D-RDSSQ,

(a) L-STIT ISO

(b) L-AREA ISO

(c) L-STIT RECT, $p_1 = p_2 = 0.5$

(d) L-AREA RECT, $p_1 = p_2 = 0.5$

(e) L-STIT DRECT, $p_1 = p_2 = 0.5$, $b_{\text{ellip}} = 0.1$

(f) L-AREA DRECT, $p_1 = p_2 = 0.5$, $b_{\text{ellip}} = 0.1$

FIGURE 2. Simulations of tessellations with division rule D-GAUSS with $\sigma = 0.1$, and two different lifetime rules: L-STIT (a, c, and e) with $t_{\text{STOP}} = 40$, and L-AREA (b, d, and f) with $t_{\text{STOP}} = 650$. In each simulation $t_{\text{GAUSS}} = 40$.

the line h_j is chosen, where $RD(z_1^j)^2 + RD(z_2^j)^2$ is maximal. The method is illustrated in Figure 4. Panel (a) shows the division by a particular line. Such a division is considered

for all the lines shown in (b). The blue line observed in (c) is the optimal dividing line such that $\min\{RD(z_1^j), RD(z_2^j)\}$ is maximized.

(a) Ceramic surface, [17].

(b) Ceramic plate, [19].
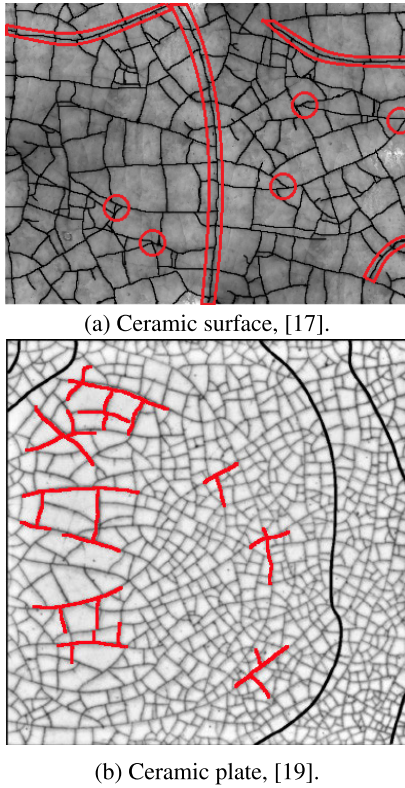
FIGURE 3. Real crack patterns.
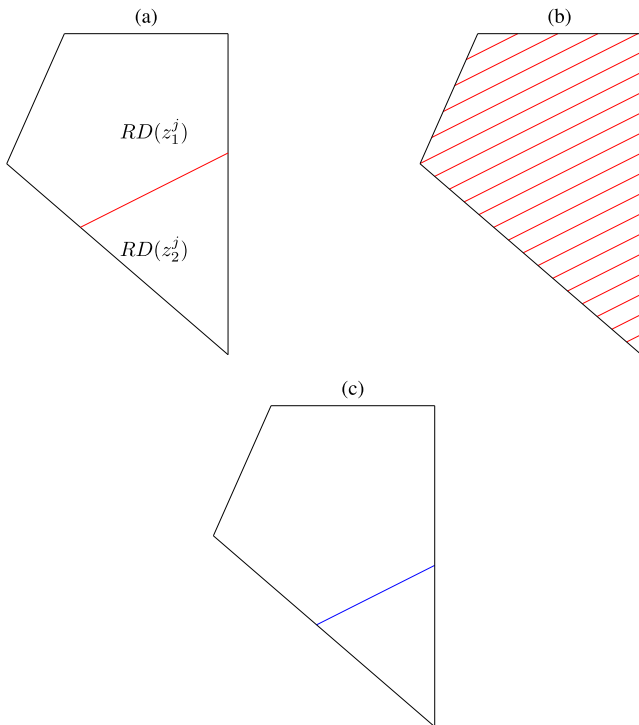
FIGURE 4. Roundness optimization. (a) Computing roundness $RD(z_1^j)$ and $RD(z_2^j)$ for each cell. (b) A family of different dividing lines. (c) The dividing line with the best roundness considering D-RDMIN.

Furthermore, all these division rules can be augmented by a rejection method which we call ''avoid small angles'' (ASA). This modification is introduced to approximate the
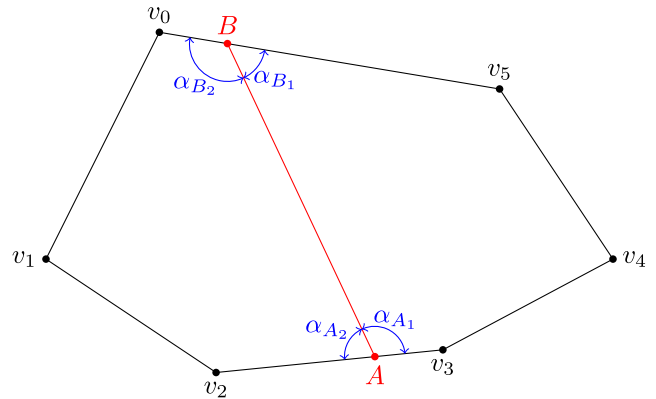
FIGURE 5. A potential dividing line cuts a cell. Four angles are analyzed in the rejection method ASA. The potential dividing line is accepted if the four angles are greater or equal than the parameter ω.

phenomenon that a new crack tends to meet an already existing crack almost orthogonally; see Figure 3.

- **ASA:** A parameter $\omega \in (0, \pi/2)$ is chosen. A potential dividing line generated by some D-rule is accepted if and only if all the four angles $\alpha_{A_1}, \alpha_{A_2}, \alpha_{B_1}, \alpha_{B_2}$, indicated in Figure 5 are greater or equal ω. If a potential dividing line is rejected, then a new potential line has to be generated according to the respective D-rule.

Thus, for example, besides D-STIT or D-GAUSS, we also consider D-STIT-ASA or D-GAUSS-ASA, respectively. The cracks in real structures are often curved, which is not incorporated in our models – the ASA method is a compromise. Note that the rejection method in ASA can cause a change in the directional distribution φ, and the resulting directional distribution of dividing segments can be quite different from φ. Some simulations with D-STIT-ASA and D-GAUSS-ASA are shown in Figures 6 and 7 respectively, and they indicate the effect of increasing the lower bound ω.

If we consider the two L-rules and the two D-rules D-STIT and D-GAUSS with and without ASA, and the D-rules D-RDMIN-ASA and D-RDSSQ-ASA, then there are 12 models. And for each model, one has the choice of a directional distribution φ, see (1). A survey is given in Figure 10. Our preliminary studies have shown that it is not worthwhile to consider D-RDMIN and D-RDSSQ without ASA.

The modifications of STIT lose the spatial consistency property (4), see [16]. Therefore, in an arbitrarily given window W, one cannot start the cell division process appropriately, such that (4) is satisfied. Being aware of this problem, one can choose one of the following options:

(i) Nonetheless launching the division procedure in W with the chosen L- and D-rules. The division process stops at time $t_{\text{MOD}} \in \mathbb{N}$.

(ii) Launching the division process with L-STIT and D-STIT until a time $t_{\text{STOP}} \in \mathbb{N}$ and then, putting the clocks back to 0, the simulation of the modified model is launched until time $t_{\text{MOD}} \in \mathbb{N}$, but only in those cells of the generated STIT tessellation which do not intersect the boundary of the window W. The

(a) $\omega = 0$

(b) $\omega = 0.1\,\pi$

(c) $\omega = 0.2\,\pi$

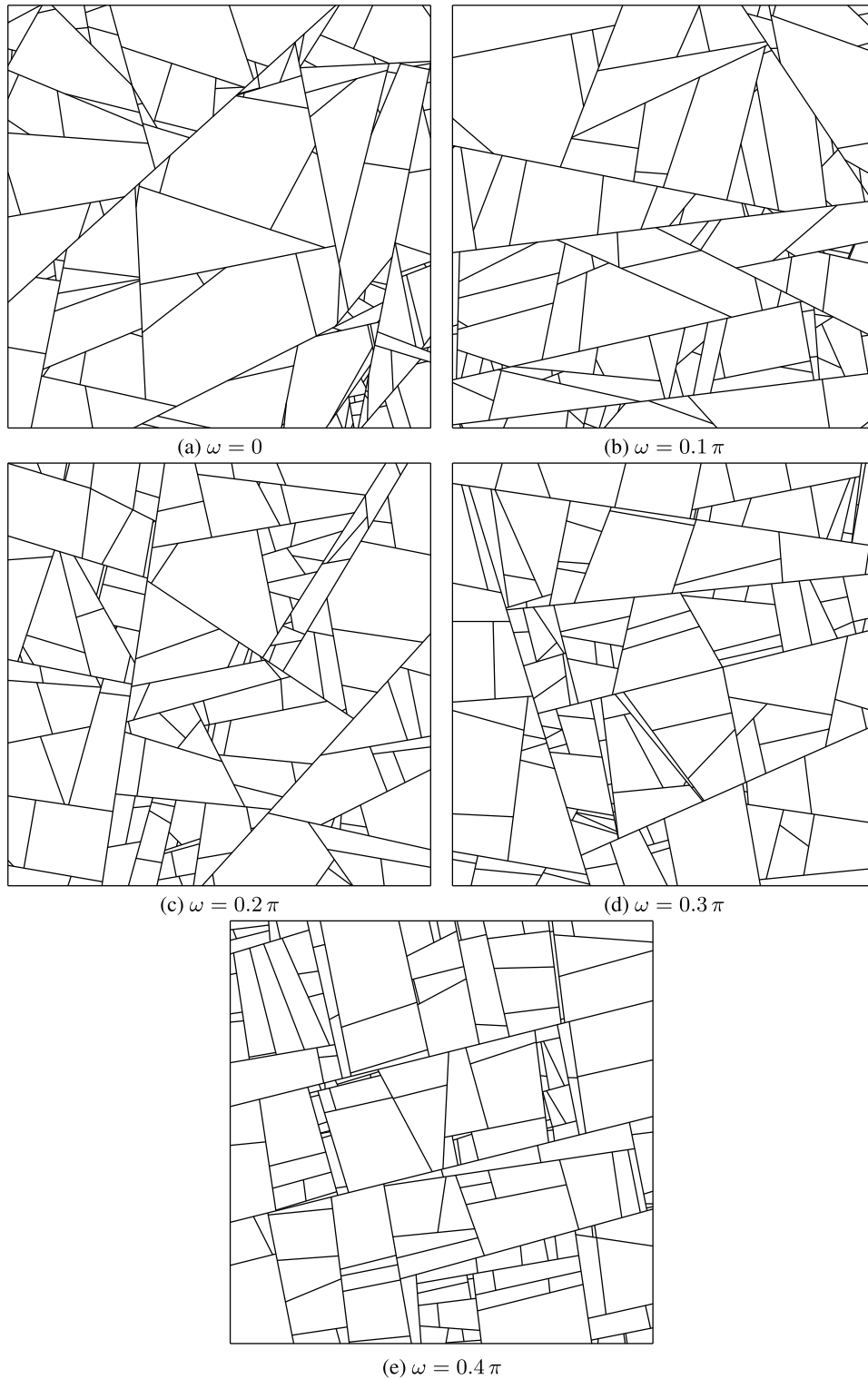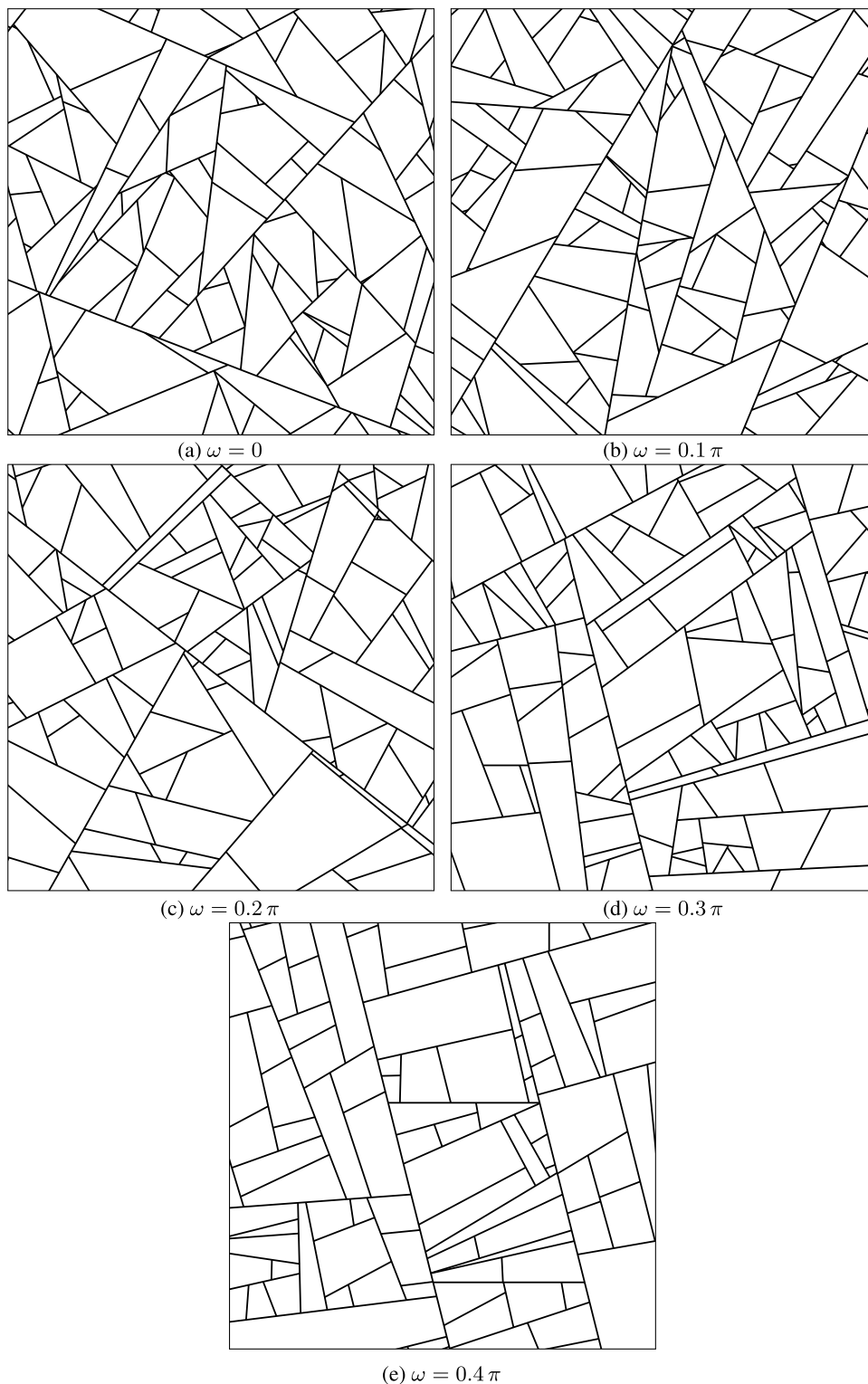(d) $\omega = 0.3\,\pi$

(e) $\omega = 0.4\,\pi$

FIGURE 6. Simulations of tessellations with L-rule LSTIT, division rule D-STIT-ASA, directional distribution DRECT, $t_{\text{STOP}} = 20$, $b_{\text{ellip}} = 0.2$ and different lower bounds ω.

cells that intersect the boundary of W have not been taken into account further. Note that until time t_{STOP}, the simulation yields an initial tessellation that is not intended. In order to see the effect of the modification, the value of t_{MOD} should be large compared to t_{STOP}.

(iii) Launching the division process in a much larger window $W'' \supset W$ with the chosen L- and D-rules. This can attenuate the systematic error appearing inside W.

In the present study, we apply (i), with the exception of D-GAUSS, where we use (ii) and thus t_{STOP} and t_{GAUSS} :=

FIGURE 7. Simulations of tessellations with L-rule LAREA, division rule D-GAUSS-ASA, directional distribution DRECT, $t_{STOP} = 60$, $t_{GAUSS} = 180$, $\sigma = 0.2$, $b_{ellip} = 0.1$ and different lower bounds ω.

t_{MOD} as additional parameters. The values of t_{STOP} and t_{MOD}, which we have chosen in our simulations, are given in Tables 1, 2, 3 and 4.

Some simulations with D-RDMIN-ASA and D-RDSSQ-ASA are shown in Figure 8 and 9.

III. THEORY

A. STATISTICS OF THE CELLS WHICH ARE USED FOR COMPARISON

In order to adapt and evaluate a model that is suggested with respect to some real data, the choice of criteria for the

goodness-of-fit is crucial. This is not obvious in complex geometric structures such as tessellations and can be rather involved. In our study, we have focused on a few instructive and feasible features of the cells z of a tessellation, namely the *area $A(z)$*, the *isoperimetric quotient or roundness or sphericity*

$$RD(z) := 4\pi A(z)/P^2(z), \quad \text{where } P \text{ is the perimeter,}$$

and *the aspect ratio*

$$AR(z) = b_{\min}(z)/b_{\max}(z).$$

The isoperimetric quotient RD and the aspect ratio AR are invariant with respect to scaling, and they provide information about the shapes of the cells.

The entities used for assessing the adequacy of a model should be invariant with respect to scaling in space. Therefore, the mean values are estimated for RD and AR. For the area, we estimate the coefficient of variation CV, which is scale-invariant and expresses the variability of cell areas in a tessellation, which means that a small value of CV indicates that the structure is more 'homogeneous'; that is, the differences between the individual cell areas are relatively small.

The estimation of a cell statistic based on a realization of a tessellation in a bounded window is biased. Edge effects cause this. As the modifications of STIT are not spatially consistent, the edge effects cannot be treated exactly – even for cells completely contained inside the window. Therefore, we cannot remove the bias, only reduce it. To do this, we apply an *edge correction*, inspired by the Miles-Lantuejoul method [12], [20]. Any cell that does not intersect the window's boundary is given a weight, which is proportional to the reciprocal probability that this cell is completely contained in the window. This compensates for the different chances of the cells to appear completely inside the window. In a square window of side length a, the size of the window, and parallel to the horizontal and vertical axes, the weight of a cell $z \subset W$ is

$$w(z) = \frac{1}{(a - b(z, 0))(a - b(z, \pi/2))}.$$

The formula to estimate the mean area of the cells in a given simulation is

$$\text{MEAN}(A) = \frac{1}{\sum_{z \subset W} w(z)} \sum_{z \subset W} w(z) A(z),$$

the estimated mean squared error is

$$\text{MSE}(A) = \frac{1}{\sum_{z \subset W} w(z)} \sum_{z \subset W} w(z) (A(z) - \text{MEAN}(A))^2,$$

the estimated standard deviation is

$$\text{SD}(A) = \sqrt{\text{MSE}(A)},$$

and the estimated coefficient of variation is

$$\text{CV}(A) = \text{SD}(A)/\text{MEAN}(A).$$

The corresponding formulae are obtained for the other entities by simply replacing the symbol A by P, RD, $b_{\max}$, $b_{\min}$, and AR, respectively.

Summarizing our approach for model adaption by a tuning process is based on the following three statistical entities for tessellations.

- **CV(A):** The estimated coefficient of variation of the cell area, see (III-A).
- **MEAN(RD):** The estimated mean isoperimetric quotient.
- **MEAN(AR):** The estimated mean aspect ratio.

B. PARAMETRIC MODEL TUNING PROCESS

We consider the space of parametric models shown in Figure 10. Our aim is to find models and the respective parameters which yield the best approximation to the image in Figure 13. The quality of the approximation is measured by

$$f_1 = \left| \text{Target}_{\text{CV}(A)} - \text{CV}(A) \right| \tag{5}$$

$$f_2 = \left| \text{Target}_{\text{MEAN}(RD)} - \text{MEAN}(RD) \right| \tag{6}$$

$$f_3 = \left| \text{Target}_{\text{MEAN}(AR)} - \text{MEAN}(AR) \right| \tag{7}$$

The Target values are listed in Table 6. For a given stochastic model and fixed parameters and a fixed window size a, the values of CV(A), MEAN(RD), and MEAN(AR) are constants. The purpose is either to minimize one of the quality functions, which is referred to as the mono-objective approach, or to find models with parameters that for all three quality functions are Pareto solutions or minimax solutions – the multi-objective approach. This describes the tuning problems we deal with, and the methods ParamILS and MO-ParamILS described below are used to perform this tuning. If all the models and parameters were considered, we would have theoretical results to determine CV(A), MEAN(RD), and MEAN(AR) exactly; we would have these pure tuning problems. The exact values of CV(A), MEAN(RD) and MEAN(AR) are not known for the models considered; with the only exceptions of CV(A)=1.9836 for isotropic STIT tessellations; CV(A)=3 for the STIT tessellation with RECT; CV(A)=1 for L-AREA, D-STIT with RECT, see [21]. Therefore, the unknown values are estimated based on simulations of the models. We emphasize that these simulations are an auxiliary procedure (in the 'background'), and they are *not* a part of the tuning, and we do *not* address a tuning of these simulation algorithms.

The use of a parameter tuning method to understand and analyze the capabilities of some tessellation models described in Section II is now presented.

A parameter tuning process can be defined as determining the best parameter values of a given tessellation model according to some given quality measures. Figure 11 shows the main components of a tuning process.

A parameter tuning process requires a tuning method and a model to be tuned. In our case, the models being tuned are the tessellation models described in the previous sections. They communicate each time the tuning method requires measuring the quality of a set of parameter values c,
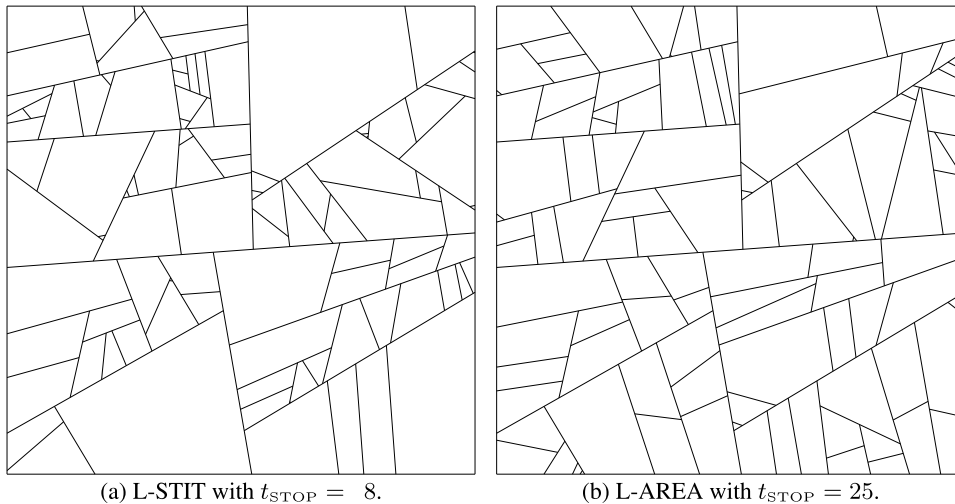
(a) L-STIT with $t_{\text{STOP}} = 8$.

(b) L-AREA with $t_{\text{STOP}} = 25$.

FIGURE 8. Simulations of tessellations with division rule D-RDMIN-ASA and the two different lifetime rules: L-STIT (a) and L-AREA (b). Both simulations with window size $a = 2$, DRECT with $p_1 = p_2 = 0.5$, $\omega = 0.25\,\pi$ and $b_{\text{ellip}} = 0.2$.
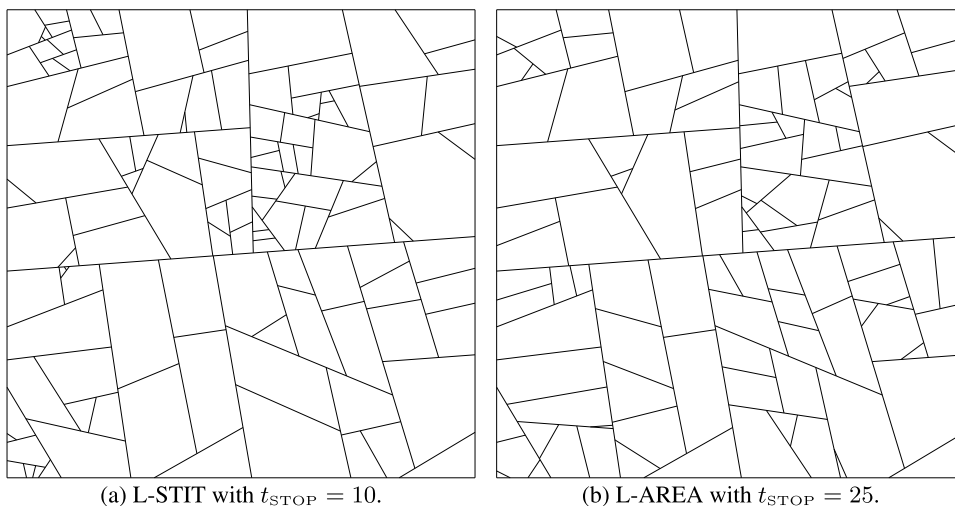

(a) L-STIT with $t_{\text{STOP}} = 10$.

(b) L-AREA with $t_{\text{STOP}} = 25$.

FIGURE 9. Simulations of tessellations with division rule D-RDSSQ-ASA and the two different lifetime rules: L-STIT (a) and L-AREA (b). Both simulations with window size $a = 2$, DRECT with $p_1 = p_2 = 0.5$, $\omega = 0.25\,\pi$ and $b_{\text{ellip}} = 0.2$.

usually referred to as parameter configuration. The measured *quality*(*es*) is then returned to the tuning method. With this, the tuning method computes the corresponding gain measure. There are several types of tuning approaches in the literature [22]. In [23], an updated review and classification of tuning methods can be found.

The **input** of a parameter tuning process is a tuning scenario composed by:

- A set of parameters to be tuned. A tuning method requires the names and the specific sets of possible values each parameter can take. In our experiments, we tuned the parameters listed in Tables 1 to 5. The window size is set to $a = 1$ except in the models D-RDMIN-ASA and D-RDSSQ-ASA, where the size is $a = 2$.

For example, Table 1 summarizes the four-dimensional parameter space considered when the D-STIT model

is tuned. In square brackets, the initial value for each parameter is shown.

The parameters *lifetime* and *direction* are categorical parameters. Categorical parameters have finite domains with no distance metric or ordering between values. The parameter *lifetime* refers to the L-rule applied in the simulation (L-STIT or L-AREA). The parameter *direction* refers to the directional distributions applied to the simulation; see Subsection II-B. Furthermore, *tstop* and *bellip* are numerical parameters; that is, their domains are subsets of $\mathbb{N}$ or of $\mathbb{R}$. For these, selecting a finite set of relevant values is necessary. In Table 1, eleven values are considered for *tstop*, and four values are considered for *bellip*. In Table 1, the parameter *bellip* is a so-called *conditional parameter* because it is relevant only when the parameter *direction* is set to DRECT or ELLIP. In Table 2, which refers to
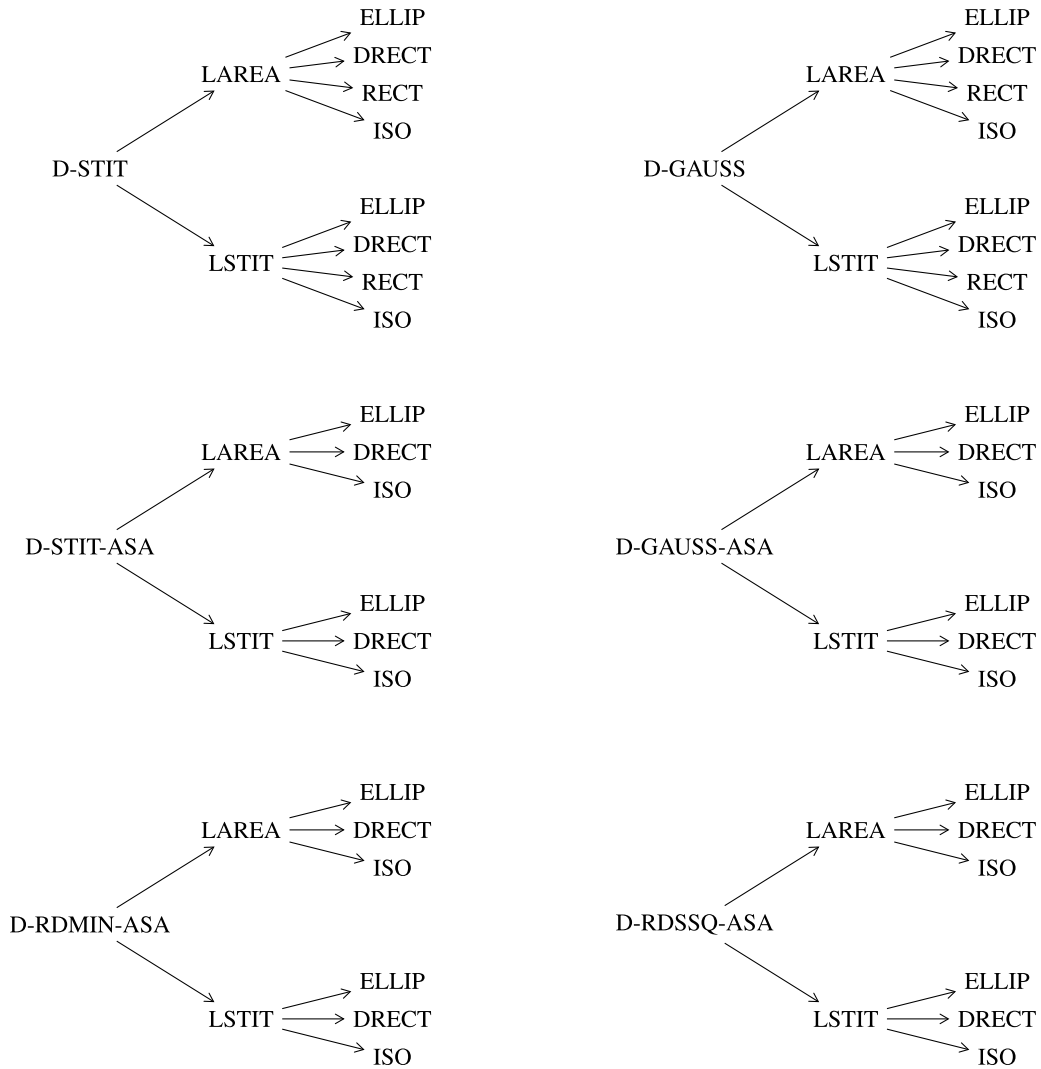
FIGURE 10. Combining the division and lifetime rules with the directional distributions. In total, we consider 40 models.
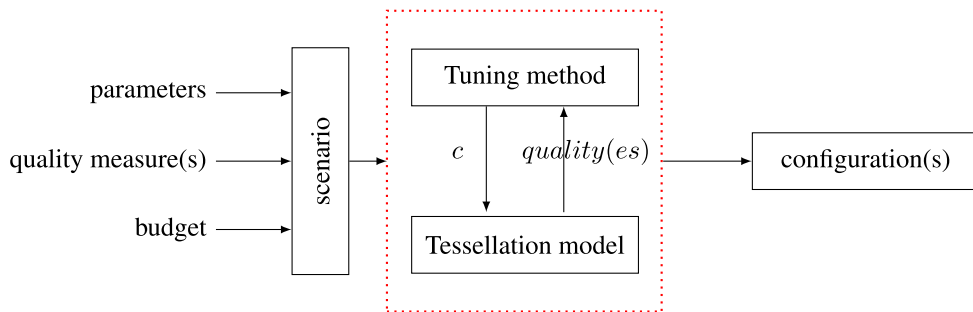

FIGURE 11. Tuning process components of a tessellation model.

D-GAUSS, the additional parameters *tmod* and *sigma* are considered. The first one refers to the construction time for the division process with D-STIT and the second one is the standard deviation σ of the Gaussian distribution. The parameters in Table 3 are analogous to Table 1 but also considering *omega* as the parameter which refers to the angle in the rejection method of

ASA, see Subsection II-C. The parameters in Table 4 are analogous to Table 2 but also consider *omega*. The parameters in Table 5 are analogous to Table 1, but the directional distribution RECT is not considered because ASA is dispensable if, a priori, only right angles appear. There are two construction times *tmod*; one is considered for lifetime rule L-STIT (*tmod1*), and the other one is
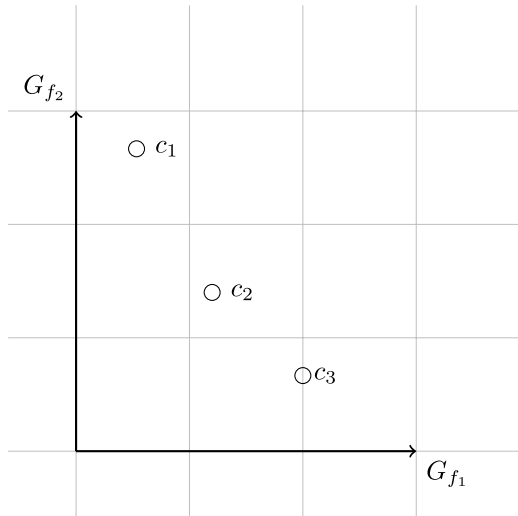
FIGURE 12. Multi-objective parameter tuning approach.

TABLE 1. Parameter search space definitions: D-STIT.

$lifetime \in \{LSTIT, LAREA\}[LSTIT]$
$direction \in \{ISO, RECT, DRECT, ELLIP\}[RECT]$
$tstop \in \{128, 192, 256, 384, 512, 768, 1024, 1536, 2048, 3072, 5096\}[128]$
$bellip \in \{0.05, 0.1, 0.15, 0.2\}[0.05]$

$Conditionals :$
$bellip \mid direction \in \{DRECT, ELLIP\}$

TABLE 2. Parameter search space definitions: D-GAUSS.

$lifetime \in \{LSTIT, LAREA\}[LSTIT]$
$direction \in \{ISO, RECT, DRECT, ELLIP\}[RECT]$
$tstop \in \{128, 192, 256, 384, 512, 768, 1024, 1536, 2048, 3072, 5096\}[128]$
$bellip \in \{0.05, 0.1, 0.15, 0.2\}[0.05]$
$sigma \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2,$
        $1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}[1.0]$
$tmod \in \{128, 192, 256, 384, 512, 768, 1024, 1536, 2048, 3072, 5096\}[128]$

$Conditionals :$
$bellip \mid direction \in \{DRECT, ELLIP\}$

TABLE 3. Parameter search space definitions: D-STIT-ASA.

$lifetime \in \{LSTIT, LAREA\}[LSTIT]$
$direction \in \{ISO, DRECT, ELLIP\}[DRECT]$
$tstop \in \{128, 192, 256, 384, 512, 768, 1024, 1536, 2048, 3072, 5096\}[128]$
$bellip \in \{0.05, 0.1, 0.15, 0.2\}[0.1]$
$omega\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}[0.25]$
$Conditionals :$
$bellip \mid direction\{DRECT, ELLIP\}$

TABLE 4. Parameter search space definitions: D-GAUSS-ASA.

$lifetime \in \{LSTIT, LAREA\}[LSTIT]$
$direction \in \{ISO, DRECT, ELLIP\}[DRECT]$
$tstop \in \{128, 192, 256, 384, 512, 768, 1024, 1536, 2048, 3072, 5096\}[128]$
$bellip \in \{0.05, 0.1, 0.15, 0.2\}[0.1]$
$sigma \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2,$
        $1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}[1.0]$
$tmod \in \{128, 192, 256, 384, 512, 768, 1024, 1536, 2048, 3072, 5096\}[128]$
$omega \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45\}[0.25]$
$Conditionals :$
$bellip \mid direction \in \{DRECT, ELLIP\}$

TABLE 5. Parameter search space definitions: D-RDMIN-ASA and D-RDSSQ-ASA.

$lifetime \in \{LSTIT, LAREA\}[LSTIT]$
$direction \in \{ISO, DRECT, ELLIP\}[DRECT]$
$tmod1 \in \{26, 28, 30, 32, 34, 36, 38, 40\}[32]$
$tmod2 \in \{300, 320, 340, 360, 380, 400, 420, 440, 460, 480, 500\}[400]$
$bellip \in \{0.05, 0.1, 0.15, 0.2\}[0.05]$
$omega \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}[0.25]$
$Conditionals :$
$bellip \mid direction \in \{DRECT, ELLIP\}$
$tmod1 \mid lifetime \in \{LSTIT\}$
$tmod2 \mid lifetime \in \{LAREA\}$

TABLE 6. Statistical data for Fig. 13.

	MEAN	SD	CV
Area	$403.515 \, \mu m^2$	$319.725 \, \mu m^2$	0.794
Perimeter	$75.075 \, \mu$ m	$31.185 \, \mu$ m	0.415
Isoper. Quotient	0.787	0.109	0.138
Min Width	$18.375 \, \mu$ m	$8.295 \, \mu$ m	0.451
Max Width	$28.56 \, \mu$ m	$11.445 \, \mu$ m	0.401
Aspect Ratio	0.644	0.127	0.198

considered for lifetime rule L-AREA (*tmod2*). These parameters are chosen differently in order to generate approximately the same number of cells (approximately 2000 cells).

The values for *sigma* are based on the preceding analysis for CV(*A*) presented in [2]. The values for *tstop* and *tmod* are based on the values of t_{STOP}, and t_{GAUSS} used in [2], respectively. The discretization of the values is selected by simple inspection according to different simulations considering the number of cells. In the case of *bellip*, the variation of CV(*A*) is not significant

for $b_{\text{ellip}} > 0.2$, see [2], therefore, we consider only $b_{\text{ellip}} \leq 0.2$. The parameter *omega* is bounded by $0 < \omega < \pi/2$. For the D-rules D-RDMIN and D-RDSSQ, we consider *tmod1* and *tmod2* values with L-rules L-STIT and L-AREA, respectively.

- Quality measures. The tuning algorithms use quality measures to evaluate the quality of a specific parameter configuration. In our case, three quality measures, given in (5), (6) and (7), are considered, which are based on target values for CV(*A*), MEAN(*RD*), and MEAN(*AR*). Given the stochastic nature of the tessellation models, tuning methods estimate the "real" performance of parameter configuration using gain measures $G_q(c)$, where *q* indicates the associated quality measures. The gain measure summarizes the performance of several simulations for a parameter configuration *c*. The most common gain measures use the average (which is the arithmetic mean) and the median. For example, the gain measure associated with parameter configuration *c* executed *n* times considering the average coefficient of variation of the area as quality measure computed by $G_q(c) = av(q, n)$.

- A budget. The budget of a tuning process limits the computational effort invested. It could be measured as the maximum execution time and the maximum number of runs. In our study, we have chosen a maximum budget of 10,000 simulations of tessellation models.
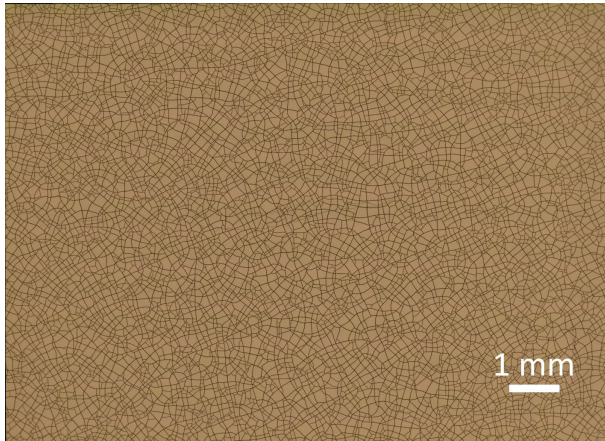
FIGURE 13. Optical microscopy image of cracking of metalized (Cr/Au – 20/100 nm) PDMS, reproduced from [2].

TABLE 7. Mono-objective results. 500 simulations are executed with the respective parameters for each division rule and target. The average of the respective target value is shown in the last three columns.

Division Rule	Lifetime	Dir. Dist.	Parameters					CV(A)	MEAN(RD)	MEAN(AR)
			t_{STOP}	t_{MOD}	σ	b_{ellip}	ω	0.794	0.787	0.644
D-STIT	L-AREA	RECT	192					**0.973**	0.530	0.336
	L-AREA	RECT	192					0.973	**0.530**	0.336
	L-AREA	ISO	768					1.190	0.520	**0.428**
D-GAUSS	L-AREA	RECT	192	5096	0.2			**0.793**	0.610	0.402
	L-STIT	RECT	128	3072	0.1			1.283	**0.655**	0.444
	L-AREA	ISO	128	5096	0.1			0.766	0.576	**0.479**
D-STIT-ASA	L-AREA	ISO	768				0.35	**0.905**	0.593	0.450
	L-STIT	ISO	1024				0.25	0.952	**0.600**	0.509
	L-STIT	ISO	128				0.25	1.575	0.601	**0.507**
D-GAUSS-ASA	L-AREA	RECT	384	5096	0.2		0.3	**0.796**	0.606	0.399
	L-STIT	ISO	128	1536	0.1		0.35	1.130	**0.657**	0.522
	L-STIT	ISO	128	1536	0.1		0.25	1.168	0.638	**0.546**
D-RDMIN-ASA	L-AREA	DRECT	480			0.1	0.15	**0.790**	0.653	0.506
	L-AREA	DRECT	340			0.05	0.15	0.703	**0.665**	0.493
	L-STIT	ISO	28				0.25	1.455	0.640	**0.556**
D-RDSSQ-ASA	L-AREA	DRECT	480			0.1	0.15	**0.819**	0.672	0.533
	L-STIT	DRECT	28			0.05	0.15	1.139	**0.689**	0.526
	L-STIT	ISO	36				0.2	1.602	0.647	**0.563**

The **output** of the parameter tuning process is a set of one or more parameter configurations that optimize the gain. In our analysis, we use two tuning methods:

- The first tuning method used is ParamILS.
  ParamILS, see [24], is an iterated local algorithm searching for better quality parameter configurations in the neighborhood of the current one. The FocusedILS version of ParamILS increases the number of simulations of tessellation models when more accurate comparisons are required. The output of ParamILS is the best parameter configuration found under the restriction of the given budget.
  In our experiments, we used ParamILS to tune the tessellation models in {D-STIT, D-GAUSS, D-STIT-ASA, D-GAUSS-ASA, RD-MIN-ASA, RD-SSQ-ASA} considering three quality measures based on a target value for CV(A), MEAN(RD), MEAN(AR)}. In this case, we aim to find an appropriate tessellation model for a real crack structure as it is given, for example, in Figure 13. The quality measure for the coefficient of variation of the area, mean of roundness, and mean aspect ratio are shown in Equations (5), (6) and (7) respectively. Gains for these quality measures are computed by the averages $G_{f_1}(c) = av(f_1, n)$, $G_{f_2}(c) = av(f_2, n)$ and $G_{f_3}(c) = av(f_3, n)$.
  For each tuning procedure, the best parameter configuration found was stored. A total of $6 \times 3 = 18$ ParamILS runs were executed. The same budget of 10,000 tessellation model simulations was considered in all these tuning procedures.
- The second tuning method is MO-ParamILS.
  The MO-ParamILS method, see [25], is a multi-objective parameter tuning method based on the ParamILS approach by [26]. A multi-objective tuning method uses a set of quality measures – and their associated gains – concurrently to evaluate parameter configurations. The intended output of a multi-objective

tuning method is a set of high-gain and well-distributed non-dominated parameter configurations.

The dominance relationship can be defined considering a multi-objective tuning problem that minimizes all of the m gain measures. In this case a parameter configuration c_1 dominates a parameter configuration c_2 iff $G_i(c_1) \leq G_i(c_2), \forall i \in \{1, \ldots, m\}$, and there is at least a gain measure i where $G_i(c_1) < G_i(c_2)$. All these parameter configurations that are not dominated by any other configuration are known as *non-dominated* parameter configurations. Furthermore, the *Pareto optimal set* can be defined as the set of non-dominated configurations, and the *Pareto front* is the set of all values of the gain measures on the Pareto optimal set.

MO-ParamILS executes an iterated local search process that works with an archive of non-dominated parameter configurations. At each iteration, a single parameter configuration from the current archive is selected and submitted to a sequence of s 1-exchange that changes the value of *one parameter* each time; see [25]. The resulting configuration is stored as a new archive from which the next local search phase starts.

In this paper, we use MO-ParamILS to obtain a Pareto front regarding the gains associated with quality measures in equations (5), (6) and (7).

Figure 12 shows an example of a set of three non-dominated parameter configurations. For simplicity, the example shows only two of the three gain measures. In this case, Configuration 1 allows the corresponding tessellation model to obtain the best average of f_1 but the worst average of f_2. Configuration 3 is the best in the average of f_2 but the worst in the average of f_1. Configuration 2 shows a balance concerning these two gain indicators. We can observe that improving a measure can generate a loss in another one. In our experiments, we expect to observe this behavior between different parameter sets in the tessellation models. Furthermore, we can determine each measure's minimum and maximum values that
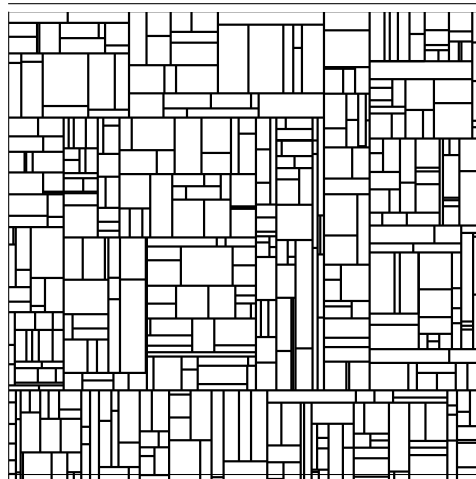
(a) D-GAUSS CV(A) = 0.796, MEAN(RD) = 0.612, MEAN(AR) = 0.403
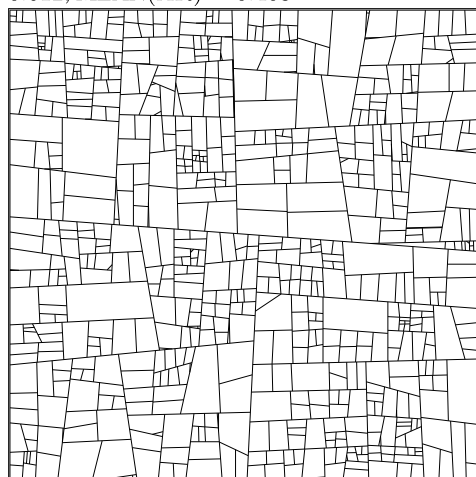
(b) D-RDSSQ-ASA CV(A) = 1.161, MEAN(RD) = 0.693, MEAN(AR) = 0.530
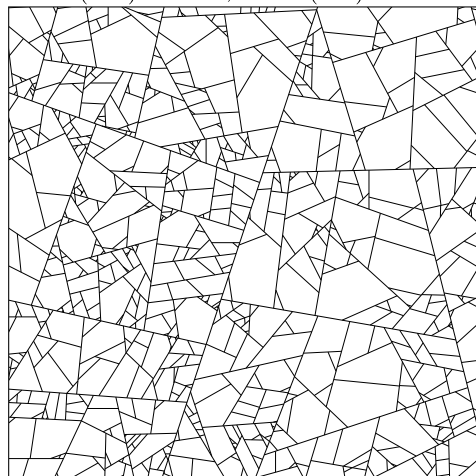
(c) D-RDSSQ-ASA CV(A) = 1.515, MEAN(RD) = 0.645, MEAN(AR) = 0.563

FIGURE 14. Simulations for the best parameters using mono-objective tuning.

TABLE 8. Multi-objective results: 500 simulations are executed with the respective parameters for each division rule and target. In the columns CV(A), MEAN(RD), and MEAN(AR) are shown the average of the respective target value.

Division Rule	Lifetime	Dir. Dist.	Parameters					CV(A) 0.794	MEAN(RD) 0.787	MEAN(AR) 0.644	M	N
			t_{STOP}	t_{MOD}	σ	b_{ellip}	ω					
D-STIT	L-AREA	RECT	192					**0.982**	0.528	0.334	1	16
	L-AREA	RECT	128					0.960	**0.530**	0.336	3	
	L-AREA	ISO	2048					1.200	0.520	**0.428**	16	
	L-AREA	DIRECT	128			0.1		1.106	0.509	0.382	8	
D-GAUSS	L-AREA	RECT	192	5096	0.2			**0.793**	0.610	0.402	1	54
	L-AREA	RECT	128	5096	0.1			0.698	**0.650**	0.440	50	
	L-AREA	ISO	192	5096	0.1			0.770	0.575	**0.478**	16	
	L-AREA	DIRECT	128	5096	0.1	0.2		0.765	0.582	0.475	20	
D-STIT-ASA	L-AREA	ISO	768				0.35	**0.906**	0.593	0.449	1	25
	L-STIT	ISO	128				0.3	1.484	**0.602**	0.481	22	
	L-STIT	ISO	256				0.25	1.575	0.601	**0.507**	25	
	L-AREA	ISO	128				0.25	0.958	0.598	0.506	14	
D-GAUSS-ASA	L-AREA	DIRECT	384	5096	0.2	0.2	0.35	**0.797**	0.626	0.478	1	39
	L-STIT	DIRECT	128	2048	0.1	0.2	0.35	1.081	**0.657**	0.514	33	
	L-AREA	ISO	256	5096	0.1		0.25	0.749	0.635	**0.543**	20	
	L-AREA	ISO	128	5096	0.1		0.2	0.755	0.618	0.527	25	
D-RDMIN-ASA	L-AREA	DIRECT	500			0.1	0.15	**0.790**	0.653	0.506	1	33
	L-AREA	DIRECT	300			0.05	0.15	0.705	**0.666**	0.494	14	
	L-STIT	ISO	36				0.2	1.524	0.626	**0.542**	33	
	L-AREA	DIRECT	320			0.1	0.2	0.772	0.658	0.510	10	
D-RDSSQ-ASA	L-AREA	DIRECT	480			0.1	0.15	**0.819**	0.672	0.533	1	18
	L-STIT	DIRECT	28			0.05	0.15	1.136	**0.688**	0.525	13	
	L-STIT	ISO	28				0.2	1.587	0.647	**0.563**	18	
	L-AREA	DIRECT	320			0.1	0.2	0.776*	0.688*	0.545*	6	

In our experiments, a multi-objective tuning process was executed for each model of the set {D-STIT, D-GAUSS, D-STIT-ASA, D-GAUSS-ASA, RD-MIN-ASA, RD-SSQ-ASA} considering the three gain measures associated to equations (5), (6) and (7) all together. The tessellation models implementation is available in [27].

IV. CALCULATION

The crack structure shown in Figure 13 was obtained by metalizing polydimethylsiloxane (PDMS), which was already presented and studied in [2]. The image analysis result is shown in Table 6. The statistical data of the image is reported in [2]. The area $A(z)$ of the cells is estimated by pixel counting in z, i.e., $\hat{A}(z) = nc^2$, where n is the pixel number, and c is the side length of a pixel. Even if the cells are not all convex, the accuracy of $\hat{A}(z)$ can be approximated as follows

$$A(z) - \frac{c}{12}P(z) \lesssim \hat{A}(z) \lesssim A(z) + \frac{c}{12}P(z) + \frac{\pi^2}{144} \quad (8)$$

where $P(z)$ is the perimeter. The perimeter $P(z)$ estimation is usually based on a discrete version of one of Crofton's intersection formulae [28], where the discretization is induced by sampling the continuous set z on a square point lattice of spacing c. For more details, see [2]. The minimal and maximal widths are estimated using the convex hull C of the set of pixels belonging to z. The convex hull is effectively determined using Graham's scan algorithm [29].

We are interested in the CV of Area, mean of Isoperimetric Quotient (roundness), and mean of Aspect Ratio. Thus, we consider the following target values for the tuning process:

$$\text{Target}_{\text{CV}(A)} = 0.794$$

$$\text{Target}_{\text{MEAN}(RD)} = 0.787$$

$$\text{Target}_{\text{MEAN}(AR)} = 0.644$$

V. RESULTS

A. MONO-OBJECTIVE APPROACH

Table 7 shows results for the mono-objective optimization. We fixed a division rule, indicated in the left column. For this,
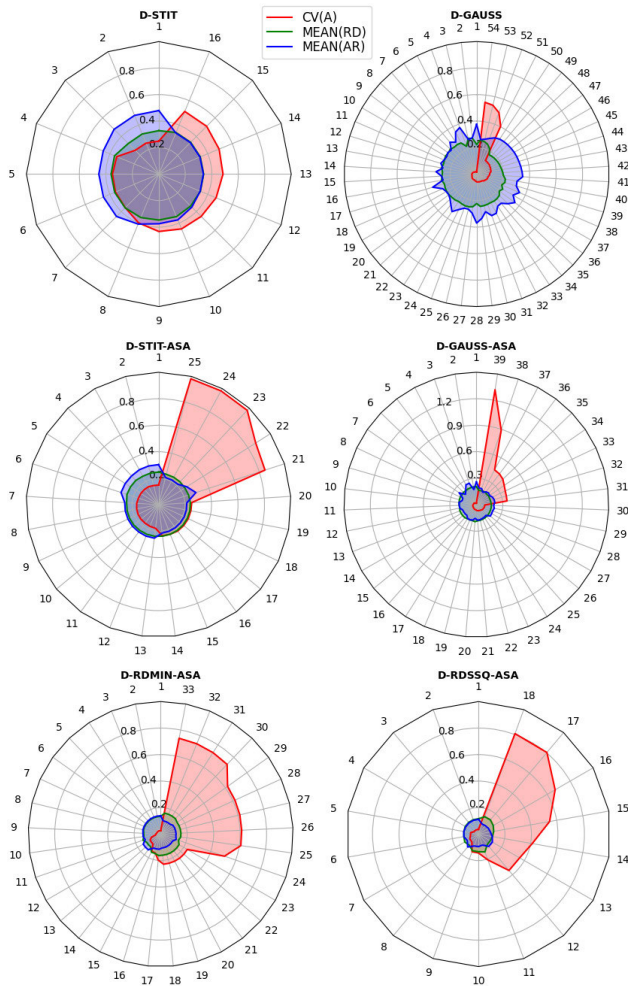
each tessellation model can obtain. Or we can choose a minimax solution, see Subsection V-B.

FIGURE 15. Multi-objective radial chart.

we optimized the following parameters: lifetime distribution, directional distribution (with their respective quantitative parameters), and the times t_{STOP} and t_{MOD}, respectively. The values emphasized by boldface indicate which was the target parameter of optimization, and they show the best value found. Thus, for example, in the second line of the table, the value 0.531 is, with restriction to D-STIT, the best approximation for the target MEAN(RD) = 0.787, and L-AREA, directional distribution RECT reach it with t_{STOP} = 192.

Fig. 14 shows simulations for the best parameters concerning CV(A), MEAN(RD), and MEAN(AR), respectively.

B. MULTI-OBJECTIVE APPROACH

Table 8 shows results for the multi-objective optimization. The three objectives are CV(A), MEAN(RD), and MEAN(AR). We fixed a division rule, indicated in the left column. For this, we tuned the following parameters: lifetime distribution, directional distribution (with their respective quantitative parameters), and the times t_{STOP} and t_{MOD}, respectively. All the lines show values of triples that belong to the Pareto front. The values emphasized by boldface

indicate that in this line, a triple of the Pareto front was chosen, providing the best approximation of the target value. Thus, for example, in the second line of the table, the triple (0.960, 0.530, 0.336) is an element of the Pareto front (under the restriction to D-STIT). Among all elements of this Pareto front, this triple yields the best approximation for the target MEAN(RD) = 0.787, and L-AREA reaches it, with directional distribution RECT and t_{STOP} = 128.

The column N of Table 8 indicates the number of parameter configurations – in the respective discretized parameter spaces described in Tables 1 to 5 – which belong to the Pareto front. For each division rule, these N elements of the Pareto front were sorted by increasing order considering the absolute distance to the CV(A) target value. The column M indicates the numbers of the configurations with the best approximation to one of the three target values. For example, in the division rule D-GAUSS, there are N = 54 Pareto front configurations, and the best approximations for the targets CV(A), MEAN(RD), and MEAN(AR) correspond to the configurations 1, 50, and 16, respectively. Fig. 16 (a), (b), (c) show simulations for the best parameters concerning CV(A), MEAN(RD), and MEAN(AR), respectively.

The radial charts in Figure 15 illustrate the complete records of the N Pareto front values for the six division rules and the respective discretized parameter spaces. Each radial line is labeled with the solution number. The red, green, and blue thick lines indicate the absolute distance to the CV(A), MEAN(RD), and MEAN(AR) target values, respectively. See formulae (5), (6), (7).

For example, the radial chart for D-STIT-ASA shows the N = 25 Pareto front values of the triple (CV(A), MEAN(RD), MEAN(AR)). As given in Table 8, the solutions with numbers 1, 22, and 25 are those for which one of the objectives reaches the best approximation among all Pareto front values. Solutions 22 and 25 are extremely bad regarding CV(A). This example also shows that the choice of a 'good-compromise'-solution within the Pareto front should not be restricted to those particular solutions emphasized in Table 8.

The minimax principle is a standard criterion for a choice within the Pareto front. In our context, a minimax solution is a parameter configuration for which $\max\{f_1, f_2, f_3\}$ is minimized. For each considered division rule, the minimax solution is shown in the respective fourth row in Table 8. For instance, the configuration M = 20 is the minimax solution for the D-GAUSS division rule. Comparing the minimax solutions for the different division rules, we find the *overall minimax solution* by D-RDSSQ-ASA, L-AREA, DRECT, t_{STOP} = 320, b_{ellip} = 0.1, and ω = 0.2. This yields CV(A) = 0.776, MEAN(RD) = 0.688 and MEAN(AR) = 0.545, see the last line of Table 8. Fig. 16 (d) shows a simulation for this overall minimax solution.

In the previous study, [2], a modeling of the same crack data was performed, and it was based on simulations and an intuitive search in a smaller class of models (in particular without RDSSQ and without ASA) without a
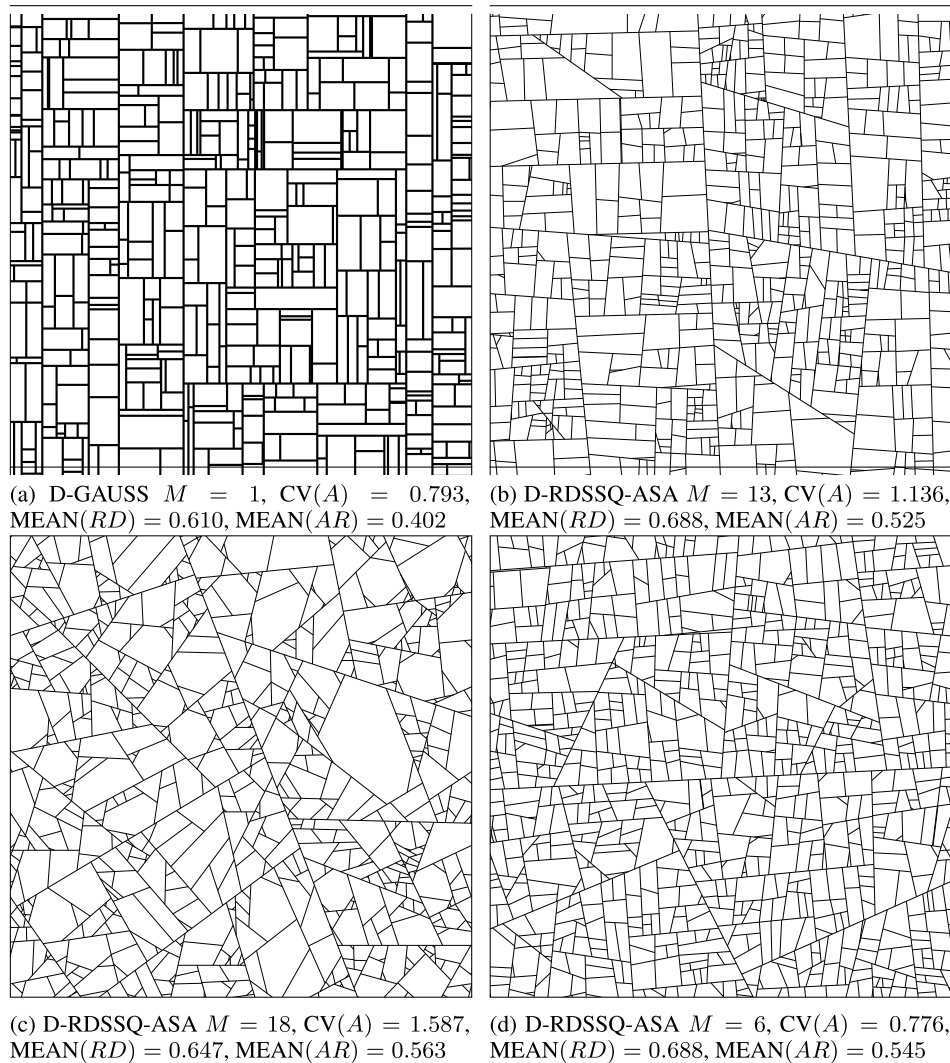
(a) D-GAUSS $M = 1$, $CV(A) = 0.793$, $MEAN(RD) = 0.610$, $MEAN(AR) = 0.402$

(b) D-RDSSQ-ASA $M = 13$, $CV(A) = 1.136$, $MEAN(RD) = 0.688$, $MEAN(AR) = 0.525$

(c) D-RDSSQ-ASA $M = 18$, $CV(A) = 1.587$, $MEAN(RD) = 0.647$, $MEAN(AR) = 0.563$

(d) D-RDSSQ-ASA $M = 6$, $CV(A) = 0.776$, $MEAN(RD) = 0.688$, $MEAN(AR) = 0.545$

FIGURE 16. Simulations for the best parameters using multi-objective tuning. For the indicated D-rules, *M* refers to the last but one column of Table 8.

formal optimization. There, it was suggested to choose D-GAUSS, L-AREA, DRECT, $b_{ellip} = 0.1$, $\sigma = 0.1$, $t_{STOP} = 500$, and $t_{MOD} = 2\,000$ with the result $CV(A) = 0.789$, $MEAN(RD) = 0.576$ and $MEAN(AR) = 0.439$.

VI. CONCLUSION

In the present paper, the variety of models introduced in [2] is extended, and a systematic search for a best-adapted model to real crack pattern data by a tuning process is performed.

The results, in particular the overall minimax solution, see the last line of Table 8, indicate that the extension of the considered models by the division rules ASA and RDSSQ, as well as the application of a tuning method, yield an improved model adaption to real crack data. This concerns mainly a better approximation of the shape of the cells, quantitatively expressed by MEAN(*RD*) and MEAN(*AR*). And also, intuitively, panel (d) of Figure 16 seems to be a reasonable approximation to the structure shown in Figure 13. In [2] the parameters to fit the model are chosen by simple inspection. In our current work, the parameters are found by

well-known tuning procedures, which means an advantage if we want to change the real data we are using.

However, our study is restricted to the analysis of *single* cells, not regarding, for example, relations between adjacent cells or, more generally, the arrangement of cells in the plane. A standard approach to this topic is based on the pair correlation function; see [30], [31], and [32].

Furthermore, boundaries emerge in the division of cells *curved*. To our knowledge, there is not yet a feasible parametric model.

We assume that the methods presented here can be applied to a wide variety of crack patterns and a wide range of micro to macro structures.

REFERENCES

[1] E. O'Reilly and N. M. Tran, "Stochastic geometry to generalize the Mondrian process," *SIAM J. Math. Data Sci.*, vol. 4, no. 2, pp. 531–552, Jun. 2022.

[2] R. León, W. Nagel, J. Ohser, and S. Arscott, "Modeling crack patterns by modified STIT tessellations," *Image Anal. Stereol.*, vol. 39, no. 1, pp. 33–46, Apr. 2020.

[3] A. G. Evans and J. W. Hutchinson, "Mixed mode cracking in layered materials," *Adv. Appl. Mech.*, vol. 29, pp. 63–191, Jan. 1991.

[4] M. D. Thouless, A. G. Evans, M. F. Ashby, and J. W. Hutchinson, "The edge cracking and spalling of brittle plates," *Acta Metallurgica*, vol. 35, no. 6, pp. 1333–1341, Jun. 1987.

[5] Z. C. Xia and J. W. Hutchinson, "Crack patterns in thin films," *J. Mech. Phys. Solids*, vol. 48, nos. 6–7, pp. 1107–1131, Jun. 2000.

[6] F. Boulogne, F. Giorgiutti-Dauphiné, and L. Pauchard, "Surface patterns in drying films of silica colloidal dispersions," *Soft Matter*, vol. 11, no. 1, pp. 102–108, 2015.

[7] P. Nandakishore and L. Goehring, "Crack patterns over uneven substrates," *Soft Matter*, vol. 12, no. 8, pp. 2233–2492, 2016.

[8] A. Hafver, E. Jettestuen, M. Kobchenko, D. K. Dysthe, P. Meakin, and A. Malthe-Sørenssen, "Classification of fracture patterns by heterogeneity and topology," *Europhys. Lett.)*, vol. 105, no. 5, p. 56004, Mar. 2014.

[9] A. Kumar, R. Pujar, N. Gupta, S. Tarafdar, and G. U. Kulkarni, "Stress modulation in desiccating crack networks for producing effective templates for patterning metal network based transparent conductors," *Appl. Phys. Lett.*, vol. 111, no. 1, Jul. 2017, Art. no. 013502.

[10] R. Seghir and S. Arscott, "Controlled mud-crack patterning and self-organized cracking of polydimethylsiloxane elastomer surfaces," *Sci. Rep.*, vol. 5, no. 1, p. 14787, Oct. 2015.

[11] G. Hernandez, R. Leon, L. Salinas, and E. Dimnet, "A fragmentation model with neighborhood interaction," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 4, 2012, Art. no. 046214.

[12] S. Chiu, D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications* (Wiley Series in Probability and Statistics), 3rd ed. Chichester, U.K.: Wiley, 2013.

[13] R. Cowan, "New classes of random tessellations arising from iterative division of cells," *Adv. Appl. Probab.*, vol. 42, no. 1, pp. 26–47, Mar. 2010.

[14] W. Nagel and V. Weiss, "Crack tessellations: characterization of stationary random tessellations stable with respect to iteration," *Adv. Appl. Probab.*, vol. 27, no. 4, pp. 73–78, 2005.

[15] R. Schneider and W. Weil, *Stochastic and Integral Geometry*. Cham, Switzerland: Springer, 2008.

[16] W. Nagel and E. Biehler, "Consistency of constructions for cell division processes," *Adv. Appl. Probab.*, vol. 47, no. 3, pp. 640–651, Sep. 2015.

[17] W. Nagel, J. Mecke, J. Ohser, and V. Weiss, "A tessellation model for crack patterns on surfaces," *Image Anal. Stereol.*, vol. 27, no. 2, pp. 73–78, 2008.

[18] L. Mosser and S. Matthai, "Tessellations stable under iteration: Evaluation of application as an improved stochastic discrete fracture modeling algorithm," in *Proc. Int. Discrete Fract. Netw. Eng. Conf.*, 2014, p. 163.

[19] S. Bohn, L. Pauchard, and Y. Couder, "Hierarchical crack pattern as formed by successive domain divisions," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 71, no. 4, 2005, Art. no. 046214.

[20] J. Serra, *Image Analysis and Mathematical Morphology*. London, U.K.: Academic, 1982.

[21] S. Martínez and W. Nagel, "Tessellation-valued processes that are generated by cell division," *J. Appl. Probab.*, vol. 2023, pp. 1–19, Nov. 2023.

[22] E. Montero, M.-C. Riff, and B. Neveu, "A beginner's guide to tuning methods," *Appl. Soft Comput.*, vol. 17, pp. 39–51, Apr. 2014.

[23] A. E. Eiben and S. K. Smit, "Parameter tuning for configuring and analyzing evolutionary algorithms," *Swarm Evol. Comput.*, vol. 1, no. 1, pp. 19–31, Mar. 2011.

[24] F. Hutter, H. H. Hoos, and T. Stützle, "Automatic algorithm configuration based on local search," in *Proc. 22nd Conf. Artif. Intell.*, 2007, pp. 1152–1157.

[25] A. Blot, H. H. Hoos, L. Jourdan, M.-É. Kessaci-Marmion, and H. Trautmann, "MO-ParamILS: A multi-objective automatic algorithm configuration framework," in *Learning and Intelligent Optimization*, P. Festa, M. Sellmann, and J. Vanschoren, Eds. Cham, Switzerland: Springer, 2015, pp. 32–47.

[26] F. Hutter, H. H. Hoos, K. Leyton-Brown, and T. Stuetzle, "ParamILS: An automatic algorithm configuration framework," *J. Artif. Intell. Res.*, vol. 36, pp. 267–306, Oct. 2009.

[27] R. León. (2023). *Crack Pattern*. Accessed: Mar. 24, 2023. [Online]. Available: https://github.com/rleonphd/crackPattern.git

[28] J. Ohser, B. Steinbach, and C. Lang, "Efficient texture analysis of binary images," *J. Microsc.*, vol. 192, no. 1, pp. 20–28, Oct. 1998.

[29] R. L. Graham, "An efficient algorithm for determining the convex hull of a finite planar set," *Inf. Process. Lett.*, vol. 1, no. 4, pp. 132–133, 1972.

[30] V. Weiss, J. Ohser, and W. Nagel, "Second moment measure and K-function for planar STIT tessellations," *Image Anal. Stereol.*, vol. 29, no. 2, pp. 121–131, 2010.

[31] C. Redenbach and C. Thäle, "On the arrangement of cells in planar STIT and Poisson line tessellations," *Methodol. Comput. Appl. Probab.*, vol. 15, no. 3, pp. 643–654, Sep. 2013.

[32] C. Redenbach and C. Thäle, "Second-order comparison of three fundamental tessellation models," *Statistics*, vol. 47, no. 2, pp. 237–257, Apr. 2013.

ROBERTO LEÓN received the Ph.D. degree in computer engineering from Universidad Técnica Federico Santa María, Chile. He is currently an Academician in computing science with Universidad Técnica Federico Santa María. His research interests include scientific computing, numerical simulation, and computational geometry.

ELIZABETH MONTERO (Member, IEEE) received the Ph.D. degree from the University of Nice Sophia-Antipolis, France. She is currently an Academician in computing science with Universidad Técnica Federico Santa María, Chile. Her research interests include the foundations and application of heuristic search methods, parameter setting problems, automatic algorithm generation, and their applications to combinatorial optimization problems. She has published over 50 technical papers in high-level heuristic search conferences, such as GECCO, PPSN, and CEC, and over 20 Q1 and Q2 WoS papers.

WERNER NAGEL is currently a Mathematician with the University of Jena, Germany. He also works in stochastic geometry, in particular on random tessellations and their application to crack structures and to machine learning. He has published more than 50 papers, mainly in mathematical journals.

● ● ●