Detecting anomalies in fibre systems using 3-dimensional image data

Denis Dresvyanskiy¹ · Tatiana Karaseva¹ · Vitalii Makogin² · Sergei Mitrofanov¹ · Claudia Redenbach³ · Evgeny Spodarev²

Received: 15 July 2019 / Accepted: 4 January 2020 / Published online: 28 January 2020
© The Author(s) 2020

Abstract

We consider the problem of detecting anomalies in the directional distribution of fibre materials observed in 3D images. We divide the image into a set of scanning windows and classify them into two clusters: homogeneous material and anomaly. Based on a sample of estimated local fibre directions, for each scanning window we compute several classification attributes, namely the coordinate wise means of local fibre directions, the entropy of the directional distribution, and a combination of them. We also propose a new spatial modification of the Stochastic Approximation Expectation-Maximization (SAEM) algorithm. Besides the clustering we also consider testing the significance of anomalies. To this end, we apply a change point technique for random fields and derive the exact inequalities for tail probabilities of a test statistic. The proposed methodology is first validated on simulated images. Finally, it is applied to a 3D image of a fibre reinforced polymer.

Keywords Anomaly detection · Classification · Fibre composite · Directional distribution · Change point problem · Entropy · SAEM algorithm

This research was supported by the German Ministry of Education and Research (BMBF) through the project “AniS”, the DFG Research Training Group GRK 1932, DFG Grant 390879134, as well as the DAAD scientific exchange program “Strategic Partnerships”.

✉ Vitalii Makogin
vitalii.makogin@uni-ulm.de

Denis Dresvyanskiy
DresvyanskiyDenis@yandex.ru

Tatiana Karaseva
tatyankaraseva@yandex.ru

Sergei Mitrofanov
sergeimitrofanov95@gmail.com

Claudia Redenbach
redenbach@mathematik.uni-kl.de

Evgeny Spodarev
evgeny.spodarev@uni-ulm.de

¹ Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, Russian Federation 660037

² Institut für Stochastik, Universität Ulm, 89069 Ulm, Germany

³ Fachbereich Mathematik, Technische Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany

1 Introduction

Fibre composites, e.g., fibre reinforced polymers or high performance concrete, are an important class of functional materials. Physical properties of a fibre composite such as elasticity or crack propagation are influenced by its microstructure characteristics including the fibre volume fraction, the size or the direction distribution of the fibres. Therefore, an understanding of the relations between the fibre geometry and macroscopic properties is crucial for the optimisation of materials for certain applications. During the last years, micro computed tomography (μ CT) has proven to be a powerful tool for the analysis of the three-dimensional microstructure of materials.

In the compression moulding process of glass fibre reinforced polymers, the fibres order themselves inside the raw material as a result of mechanical pressure. During this process, deviations from the requested direction may occur, creating undesirable fibre clusters and/or deformations. These inhomogeneities are characterized by abrupt changes in the direction of the fibres, and their detection is studied in this paper.

The problem of detecting change points in random sequences, (multivariate) time series, panel and regression data has a long history, see the books Basseville and Niki-

forov (1993); Brodsky and Darkhovsky (2000); Carlstein et al. (1994); Chen and Gupta (2012); Csörgö and Horváth (1997); Wu (2005). Changes to be detected may concern the mean, variance, correlation, spectral density, etc. of the (stationary) sequence $\{X_k, k \geq 0\}$. This kind of change detection has been considered by various authors starting with Page (1954). Sen and Srivastava (1975) considered tests for a change in the mean of a Gaussian model. An overview can also be found in Brodsky and Darkhovsky (1993). The CUSUM procedure, Bayesian approaches as well as maximum likelihood estimation are often used. Scan statistics come also into play naturally, see e.g. Brodsky and Darkhovsky (1999, 2000).

First approaches to change point analysis for random fields (or measures) have been developed in the papers Bucchia (2014), Bucchia and Heuser (2015), Cao and Worsley (1999), Chambaz (2002), Hahubia and Mnatsakanov (1996), Jarušková and Piterbarg (2011), Kaplan (1990, 1992), Lai (2008), Müller and Song (1994), Ninomiya (2004), Sharipov et al. (2016), Siegmund and Yakir (2008), Siegmund and Worsley (1995), see also the review in Brodsky and Darkhovsky (1999, Section 2, D) and Brodsky and Darkhovsky (2000, Chapter 6). The involved methods include M-estimation, minimax methods for risks, the geometric tube method, some nonparametric and Bayesian techniques. However, much is still to be done in this relatively new area of research.

In this paper, we develop a change point test for m -dependent random fields. In the spirit of the book Brodsky and Darkhovsky (2000), it uses inequalities for tail probabilities of suitable test statistics. It is applied to the mean and the entropy of the local directional distribution of fibres observed in a 3D image of a fibre composite obtained by micro computed tomography. Characteristics are estimated in a moving scanning window that runs over the observed material sample, cf. Alonso-Ruiz and Spodarev (2017, 2018). Our main task is to detect areas with anomalous spreading of the fibres. Even though we focus on anomalies in fibres' directions, our method will work with any local characteristic of fibres with values in a (compact) Riemannian manifold such as fibre length or mean curvature.

If an anomaly is present, its location is detected using a new spatial modification of the Stochastic Approximation Expectation Maximization (SAEM) algorithm (see Korolev (2007) for a review of Expectation Maximization (EM) algorithms for the separation of components in a mixture of Gaussian distributions, and Laurent et al. (2018) for a recent paper). It allows for spatial clustering of the whole fibre material into a “normal” and an “anomaly” zone.

The paper is organized as follows. In Sect. 2, we introduce the stochastic model of a fibre process. In Sect. 3, we describe the procedure of generating the sample data, introduce the mean of local directions as well as their entropy.

There, we compare two methods for entropy estimation: plug-in and nearest neighbor statistics. In Sect. 4, we consider the detection of anomalies as a change point problem for the corresponding m -dependent random fields. In Sect. 5, we localize the anomalous region of fibres solving a clustering problem for multivariate random fields. For this purpose, we propose a new spatial modification of SAEM algorithm, which decreases the diffuseness of clusters. In Sect. 6, we apply our methods to 3D images of simulated (Sect. 6.1) and real (Sect. 6.2) fibre materials and compare their performance. For the sake of legibility, some of the proofs and tables have been left out of the main part of this paper. They can be found in the Appendix.

2 Problem setting

In this section, we give some basic definitions and results for fibre processes. For more details, see, for example, the book Chiu et al. (2013). In 3-dimensional Euclidean space, a fibre γ is a simple curve $\{\gamma(t) = (\gamma_1(t), \gamma_2(t), \gamma_3(t)), t \in [0, 1]\}$ of finite length satisfying the following assumptions:

- $\{\gamma(t), t \in [0, 1]\}$ is a C^1 -smooth function.
- $\|\gamma'(t)\|_3^2 > 0$ for all $t \in [0, 1]$, where $\|\gamma'(t)\|_3^2 = |\gamma_1'(t)|^2 + |\gamma_2'(t)|^2 + |\gamma_3'(t)|^2$.
- A fibre does not intersect itself.

The collection of fibres forms a *fibre system* ϕ if it is a union of at most countably many fibres $\gamma^{(i)}$, such that any compact set is intersected by only a finite number of fibres, and $\gamma^{(i)}((0, 1)) \cap \gamma^{(j)}((0, 1)) = \emptyset$, if $i \neq j$, i.e., the distinct fibres may have only end-points in common. The length measure corresponding to the fibre system ϕ (and denoted by the same symbol) is defined by

$$\phi(B) = \sum_{\gamma^{(i)} \in \phi} h(\gamma^{(i)} \cap B)$$

for bounded Borel sets $B \in \mathcal{B}(\mathbb{R}^3)$, where $h(\gamma \cap B) = \int_0^1 \mathbb{I}\{\gamma(t) \in B\} \sqrt{\|\gamma'(t)\|_3^2} dt$ is the length of fibre γ in window B . Then $\phi(B)$ is the total length of fibre pieces in the window B .

Definition 1 A *fibre process* Φ is a random element with values in the set $\mathbb{D}$ of all fibre systems ϕ with σ -algebra $\mathcal{D}$ generated by sets of the form $\{\phi \in \mathbb{D} : \phi(B) < x\}$ for all bounded Borel sets B and real numbers x . The distribution P of a fibre process is a probability measure on $[\mathbb{D}, \mathcal{D}]$. The fibre process Φ is said to be *stationary* if it has the same distribution as the translated fibre process $\Phi_x = \Phi + x$ for all $x \in \mathbb{R}^3$.

For classification needs we consider an abstract fibre characteristic w . Let $(E, \mathcal{E}, \sigma)$ be a measurable space where E is a (compact) Riemannian manifold equipped with a metric ρ . Let $w(x) \in E$ be some characteristic of a fibre at point $x \in \mathbb{R}^3$, assuming that exactly one fibre of Φ passes through x . Then a weighted random measure Ψ can be defined by

$$\Psi(B \times L) = \int_B \mathbb{I}\{w(x) \in L\} \Phi(dx)$$

for bounded $B \in \mathcal{B}(\mathbb{R}^3)$ and $L \in \mathcal{E}$. Thus, $\Psi(B \times L)$ is the total length of all fibre pieces of Φ in B such that their characteristic w lies in range L .

As classifying characteristics w we can, for instance, choose the fibres' local direction (with E being the sphere $\mathbb{S}^2$), their length or curvature (both with $E = \mathbb{R}_+$). In this article we focus on local directions of fibres, but the results can easily be applied to other choices of w .

If the fibre process Φ is stationary then the *intensity measure* of Ψ can be written as $\mathbb{E}\Psi(B \times L) = \lambda|B|f(L)$, where $\lambda > 0$ is called the *intensity* of Ψ , $|\cdot|$ is the Lebesgue measure in $\mathbb{R}^3$ and f is a probability measure on $\mathbb{S}^2$ which is called the *directional distribution* of fibres. The distribution f is the fibre direction distribution in the typical fibre point, hence length-weighted. In what follows, $|A|$ is either the cardinality of a finite set A or the Lebesgue measure of A , if A is uncountable and measurable.

Let $\oplus$ and $\ominus$ be the dilation (erosion, resp.) operation on images as introduced e.g. in Chiu et al. (2013). Assume that we observe a dilated version $\mathcal{E} = \Phi \oplus B_r$ of Φ within a window $W = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, $a_i < b_i, i = 1, 2, 3$, where B_r is the ball of radius $r > 0$ centered at the origin. In our setting, we assume that the fibres' length is significantly larger than their diameter $2r$. Moreover, we assume that there is $\varepsilon > 0$ such that $\mathcal{E}$ is morphologically closed w.r.t. B_ε , i.e., $(\mathcal{E} \oplus B_\varepsilon) \ominus B_\varepsilon = \mathcal{E}$. This condition ensures that the local fibre direction is uniquely defined in each point within $\mathcal{E}$.

We would like to test the hypothesis

H_0 : Φ is stationary with intensity λ and directional distribution f vs.

H_1 : There exists a compact set $A \subset W$ with $|A| > 0$ and $|W \setminus A| > 0$ such that

$$\begin{aligned} & \frac{1}{\lambda|A|} \mathbb{E} \int_A \mathbb{I}\{w(x) \in \cdot\} \Phi(dx) \\ & \neq \frac{1}{\lambda|W \setminus A|} \mathbb{E} \int_{W \setminus A} \mathbb{I}\{w(x) \in \cdot\} \Phi(dx). \end{aligned}$$

If H_1 holds true, the region A is called an *anomaly region*. In the following, we discuss how to test the hypothesis H_0 and how to detect the anomaly region A .

Fig. 1 Construction of scanning windows W_l

3 Data and clustering criteria

We assume that the dilated fibre system $\mathcal{E} \cap W$ is observed as a 3D greyscale image. Several methods for estimating the local fibre direction $w(x)$ in each fibre pixel $x \in \mathcal{E}$ are discussed in Wirjadi et al. (2016). We use the approach based on the Hessian matrix that is implemented in the 3D image analysis software tool MAVI Fraunhofer (2005). The smoothing parameter σ required by the method is chosen as $\sigma = \hat{r}$, where $\hat{r}$ is an estimate of the (constant) thickness of the typical fibre. In the simulated samples, it is known. For the real data, it is obtained manually from the images.

We divide the observation window W into small cubes $\tilde{W}_i$ (see Fig. 1) of the same size, whose edge length Δ equals three times the fibre diameter. The principal axis $\hat{w}_i$ of local directions (e.g. Wirjadi et al. 2016) in each $\tilde{W}_i$, here referred to as “average local direction”, is then computed using the function *SubfieldFibreDirections* in MAVI.

Let $J_T = \{\mathbf{i} = (i_1, i_2, i_3), i_1 = \overline{1, n_1}, i_2 = \overline{1, n_2}, i_3 = \overline{1, n_3}\}$ be the regular grid of cubes $\tilde{W}_i$. Some of the cubes $\tilde{W}_i$ may not contain enough fibre voxels to obtain a reliable estimate of the local fibre direction $\hat{w}_i$. Let $J \subset J_T$ be the subset of indices of cubes which allow for such an estimation. For each point $\mathbf{i} \in J$ denote by $X_i = (x_i, y_i, z_i)^T$ the average local direction estimated from $\tilde{W}_i$. We assume that our fibres are non-oriented and can be then transformed such that $X_i \in \mathbb{S}_+^2$, where $\mathbb{S}_+^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, z \in$

Fig. 2 Visualization of simulated material with anomaly zone

$[0, 1]$ is a hemisphere. The size of this sample is $N = |J| \leq n_1 n_2 n_3$.

Our main task is to determine the anomaly regions or, in other words, to classify the set of points J into two clusters corresponding to the “homogeneous material” and the “anomaly” (one of these clusters can be empty) (Fig. 2). To do so, we combine M^3 of the small cubes $\tilde{W}_i$ (having edge length Δ) to a larger cube W_1 , such that the 3D image W is divided into larger non-empty cubic observation windows (see Fig. 1). The larger cubes have side length $M\Delta$ and the corresponding grid of the larger cubes is denoted by $J_W = \{\mathbf{l} = (l_1, l_2, l_3) : l_1 = \overline{1, m_1}, l_2 = \overline{1, m_2}, l_3 = \overline{1, m_3}\}$. The set of indexes of non-empty cubes $\tilde{W}_i$ within W_1 is denoted by $S_1 = \{(i_1, i_2, i_3) \in J, \tilde{W}_{(i_1, i_2, i_3)} \subset W_1\}$. For each window $W_1, \mathbf{l} \in J_W$, we estimate the entropy and the mean of the local directions, based on the estimates $\{\hat{w}_i, \mathbf{i} \in S_1\}$ as described below. The choice of the size of W_1 is dictated by statistical estimation procedures described in Sects. 3.3 and 6.

3.1 Mean of local fibres direction

The vector $M_1 = (M_{x,1}, M_{y,1}, M_{z,1})^T$ is calculated for each window $W_1, \mathbf{l} \in J_W$ as the coordinate-wise sample mean of local directions (MLD):

$$M_{x,1} = \frac{1}{|S_1|} \sum_{\mathbf{i} \in S_1} \hat{w}_{i_1}, \quad M_{y,1} = \frac{1}{|S_1|} \sum_{\mathbf{i} \in S_1} \hat{w}_{i_2},$$

$$M_{z,1} = \frac{1}{|S_1|} \sum_{\mathbf{i} \in S_1} \hat{w}_{i_3}.$$

Note that $|S_1| \leq M^3$ and normally $|S_1| \approx M^3$.

3.2 The entropy of the directional distribution

The entropy of a random variable is a certain measure of the diversity/concentration of its range. Let $\mathbf{P}$ be a probability distribution of a random element X on an abstract measurable phase-space $(\mathbb{X}, \sigma)$. The value

$$E_X = - \int_{\Omega} \ln(\mathbf{P}(X(\omega))) \mathbf{P}(d\omega) \quad (1)$$

is called the *Shannon (differential) entropy* of X . If $\mathbf{P}$ is absolutely continuous with respect to some measure σ then there exists the Radon–Nikodym derivative (or density) $f = \frac{d\mathbf{P}}{d\sigma}$, and the entropy of X has the following form

$$E_f := E_X = - \int_{\mathbb{X}} \ln(f(x)) f(x) \sigma(dx). \quad (2)$$

In what follows, we assume that the random variable X is absolutely continuous with density $f : \mathbb{X} \rightarrow \mathbb{R}_+$. In our problem setting, X corresponds to the local fibre direction, $\mathbb{X}$ is the sphere $\mathbb{S}^2$ and σ is the spherical surface area measure on $\mathbb{S}^2$. Since our fibres are non-oriented ($X \in \mathbb{S}_+^2$), we may consider even local direction densities f on the whole sphere $\mathbb{S}^2$ where appropriate, instead of a density f defined on $\mathbb{S}_+^2$. However, choosing another classifying characteristic w will lead to a different measurable space $(\mathbb{X}, \sigma)$.

3.3 Entropy estimation

In the literature, there is a large number of papers devoted to non-parametric entropy estimation for i.i.d. random vectors in $\mathbb{R}^D$, see e.g. the review in Beirlant et al. (1997) and references in Bulinski and Dimitrov (2019). We will dwell upon two important estimates: the plug-in and the nearest neighbor ones.

3.3.1 Plug-in estimator of entropy

For simplicity, define the plug-in estimator for directional distributions on the sphere $\mathbb{S}^2$ with even densities $f : f(x) = f(-x), x \in \mathbb{S}^2$. Its general definition on compact manifolds $\mathbb{X}$ can be found e.g. in Alonso-Ruiz and Spodarev (2017).

For a directional distribution density f , take the kernel estimator $\hat{f}_B(\cdot)$ on a window $B \subseteq J_T$ of the form

$$\hat{f}_B(y) = \frac{1}{|B|} \sum_{\mathbf{i} \in B \cap J} \frac{|\sin \rho(y, X_{\mathbf{i}})|}{h^2 \rho(y, X_{\mathbf{i}})} K\left(\frac{\rho(y, X_{\mathbf{i}})}{h}\right), \quad (3)$$

where $h > 0$ denotes the bandwidth, $K : \mathbb{R}_+ \rightarrow \mathbb{R}$ is a kernel function and $\rho : \mathbb{S}^2 \times \mathbb{S}^2 \rightarrow \mathbb{R}_+$ is a geodesic metric

Table 1 The plug-in entropy estimator for the uniform directional distribution on $\mathbb{S}^2$

Sample size	Mean	Variance	MSE
500,000	2.476	2.848×10^{-6}	0.003
375,000	2.456	5.034×10^{-6}	0.006
250,000	2.418	6.293×10^{-6}	0.013
125,000	2.309	7.245×10^{-6}	0.049
62,500	2.099	2.739×10^{-5}	0.187
12,500	0.981	8.917×10^{-5}	2.403
6250	0.359	1.251×10^{-4}	4.718
1250	0.008	4.695×10^{-4}	6.36

given by $\rho(\mathbf{x}, \mathbf{y}) = \arccos\langle \mathbf{x}, \mathbf{y} \rangle$, $\mathbf{x}, \mathbf{y} \in \mathbb{S}^2$, where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product in $\mathbb{R}^3$.

Then the *plug-in* estimator of E_f in the window W_1 is given by

$$\widehat{E}_{f,1} = -\frac{1}{|S_1|} \sum_{\mathbf{i} \in S_1} \ln \widehat{f}_{B+\mathbf{i}}(X_{\mathbf{i}}), \quad (4)$$

where $B \subset J_T$ is the sub-window and $B + \mathbf{i} := \{\mathbf{j} + \mathbf{i}, \mathbf{j} \in B\}$ denotes the translation of B .

For homogeneous marked Poisson point processes, the plug-in estimator $\widehat{E}_f$ as above is considered in Alonso-Ruiz and Spodarev (2017). See also Alonso-Ruiz and Spodarev (2018) for the context of Boolean models of line segments. We also made an attempt to apply this method to our 3D image data. But we met difficulties which basically come from the relatively small amount of data available. Namely, $\widehat{E}_f$ needs a large number of points in sub-windows B during the estimation of f together with a large number of such sub-windows. Let us illustrate these difficulties on a simple example.

Example 1 Consider the uniform distribution on the sphere $\mathbb{S}^2$, i.e., the density is $f(x) = \frac{1}{4\pi}$, $x \in \mathbb{S}^2$. We generate a sample from this distribution and estimate its entropy E_f using the plug-in estimator (4) with $|B| = |\mathbb{S}^2|^{1/9}$ (as in Alonso-Ruiz and Spodarev 2018). The results are presented in Table 1. Moreover, we run the procedure 100 times and compare the obtained values with the exact value of the entropy

$$E_f = - \int_{\mathbb{S}^2} \frac{1}{4\pi} \ln \frac{1}{4\pi} \sigma(dx) = \ln 4\pi \approx 2.53.$$

Obviously, the bias of $\widehat{E}_f$ is too large with less than 62,000 entries, which is in accordance with Miller (1993) stating the impracticability of plug-in entropy estimates for samples in higher dimensions. There are 430,741 entries in J_T for the

real data (Fig. 13) and 463,537 entries in RSA fibre data (Fig. 5), that allows us to subdivide the images only into 4 non-intersecting regions with more than 100,000 cubes W_1 . In other words, in order to test the hypotheses H_0 vs. H_1 with test statistics based on estimated entropy (4) we have a sample of size 4, which is too small, compare Sect. 4, inequality (18). There, for $m = 1$ the minimal sample size $|W|$ must be 1000.

3.3.2 Nearest neighbor estimator of entropy

In order to overcome the above difficulties we apply another estimator of E_f introduced in the paper by Kozachenko and Leonenko (1987). We call this estimator “Dobrushin estimator” because its main idea is due to Dobrushin (1958). The estimator from Kozachenko and Leonenko (1987) cannot be applied directly, because it is designed for random vectors in a d -dimensional Euclidean space which is flat. In our setting, the phase space $(\mathbb{S}^2, \sigma)$ is a manifold of positive constant curvature with geodesic metric ρ . Therefore, we take a version of Dobrushin estimator for the case of an d -dimensional compact Riemannian manifold $\mathbb{X}$ with a geodesic metric ρ and Hausdorff measure σ .

For defining this estimator, the following results will be useful. Denote by $B_\delta(x)$ the ball in $(\mathbb{X}, \rho)$ with radius $\delta > 0$ and center x , i.e., $B_\delta(x) = \{y \in \mathbb{X} : \rho(x, y) \leq \delta\}$. Since a ρ -ball and a Euclidean d -dimensional ball are bi-Lipschitz equivalent, d coincides with the Hausdorff dimension of the manifold $\mathbb{X}$, see Falconer (2003, Corollary 2.4). Furthermore, for σ -almost all points $x \in \mathbb{X}$ the Lebesgue density theorem is true, i.e.,

$$\sigma(B_\delta(x)) \sim c\delta^d, \text{ as } \delta \rightarrow 0+, \quad (5)$$

where $d = \dim_H X$ is the Hausdorff dimension of $\mathbb{X}$ and $c = 2^d D_X > 0$, where D_X is the Hausdorff density of $\mathbb{X}$, see Falconer (2003, Proposition 4.1.5.1) and Rogers (1998, Theorem 30).

Definition 2 Let $(\xi_1, \dots, \xi_N)$ be a sample of i.i.d. $\mathbb{X}$ -valued random elements with continuous density function $f : \mathbb{X} \rightarrow \mathbb{R}_+$. Denote by ρ_i the distance to the nearest neighbor of ξ_i , $i = \overline{1, N}$, i.e.,

$$\rho_i = \min_{j=\overline{1, N}, j \neq i} \{\rho(\xi_i, \xi_j)\}.$$

Define the statistic

$$\widehat{E} = \frac{d}{N} \sum_{i=1}^N \ln \rho_i + \ln(c(N-1)) + \gamma, \quad (6)$$

where c and d are defined by (5) and

$$\gamma = - \int_0^\infty (\ln y) e^{-y} dy \approx 0.5772$$

Table 2 The Dobrushin estimator (6) for the uniform directional distribution on the sphere

Sample size	Mean	Variance	MSE
125	2.51	0.02	0.02
64	2.50	0.03	0.02

is Euler's constant. The statistic $\hat{E}$ is called *nearest neighbor (Dobrushin) estimator* of the entropy.

It coincides with the nearest-neighbour entropy estimate given in Penrose and Yukich (2013, p. 2169) with the only difference that in Penrose and Yukich (2013) Euclidean distances between ξ_i are used instead of geodesic distances ρ_i . The L_2 -consistency of $\hat{E}$ is proven in Penrose and Yukich (2013, Theorem 2.4) for i.i.d. samples $(\xi_1, \dots, \xi_N)$ as above with bounded density f of compact support.

In fact, a large class of parametric distributions on a sphere, including the Fisher, the Watson or the Angular Central Gaussian distribution, has bounded densities with compact support.

Remark 1 In many problems of probability theory, limit theorems for independent observations remain true for weakly dependent data. Since the fibre materials are weakly dependent (the fibers have a finite length), we can assume that the entropy and mean local directions are weakly dependent as well. The proof of consistency of (6) for weakly dependent ξ_i is non-trivial and goes beyond the scope of this paper.

Remark 2 Our data sets consist of straight fibres which are longer than the edge length Δ of small observation windows $\tilde{W}$. Such fibres yield several almost equal values of average local directions X_i . This leads to very small values of a distance to the nearest neighbor ρ_i and, consequently, to the large negative bias of $\hat{E}$ which is computed using $\log \rho_i$. Trying to eliminate this bias, we propose to use the following version of (6)

$$\hat{E} = \frac{d}{\sum_{i=1}^N \mathbb{I}\{\rho_i > \rho_0\}} \sum_{i=1}^N \mathbb{I}\{\rho_i > \rho_0\} \ln \rho_i + \ln c + \ln \left(\sum_{i=1}^N \mathbb{I}\{\rho_i > \rho_0\} - 1 \right) + \gamma \quad (7)$$

with penalty value $\rho_0 = 0.01$ found by computational tuning.

In order to test the accuracy of the Dobrushin estimator, we have generated 100 samples from the uniform directional distribution on $\mathbb{S}^2$. We have computed the Dobrushin statistic and compared it with the exact entropy value $\ln(4\pi) \approx 2.53$. The results are presented in Table 2.

Based on these results, we conclude that the Dobrushin estimator (6) is quite accurate for small sample sizes. Even

for a sample with 64 entries the entropy is estimated much better than by the plug-in method.

4 Change point detection in random fields

To test the hypothesis H_0 against H_1 , we check the existence of anomaly regions in a realization of an m -dependent geometric random field $\{s_k, k \in W\}$. Here we follow the ideas from Brodsky and Darkhovsky (1993), where change point problems for mixing random fields on general parametric (disorder) regions were considered. The field $\{s_k, k \in W\}$ will be chosen in a way such that the hypothesis H_0 implies that it is stationary, whereas H_1 means the presence of a region $I_\theta \subset W$ with different mean value of s_k . Later in our application to fibre materials in Sect. 4.3, we assume the anomaly region to be a box Bucchia and Wendler (2017).

4.1 Random fields with inhomogeneities in mean

Let $\{\tilde{\xi}_k, k \in \mathbb{Z}^3\}$ be an integrable stationary real-valued random field with $\mu = \mathbb{E}\tilde{\xi}_k$. Denote by ξ_k the centered field $\xi_k = \tilde{\xi}_k - \mu, k = (k_1, k_2, k_3) \in \mathbb{Z}^3$. Moreover, we assume that $\{\xi_k, k \in \mathbb{Z}^3\}$ is m -dependent, i.e., ξ_k and ξ_l are independent if $\max_{i=1,2,3} |k_i - l_i| > m, k = (k_1, k_2, k_3) \in \mathbb{Z}^3, l = (l_1, l_2, l_3) \in \mathbb{Z}^3$. Let Θ be a finite parameter space. For every $\theta \in \Theta$ we define subsets of anomalies $I_\theta \subset \mathbb{Z}^3$ completely determined by a parameter θ . Then for some $\theta_0 \in \Theta$ we observe

$$s_k = \tilde{\xi}_k + h \mathbb{I}\{k \in I_{\theta_0}\}, k \in W, \quad (8)$$

where $W = [1, M_1] \times [1, M_2] \times [1, M_3] \cap \mathbb{Z}^3$, and $h \in \mathbb{R}$ is fixed and unknown. Assume that $I_\theta \subset W$ for every $\theta \in \Theta$. Denote $I_\theta^c = W \setminus I_\theta$.

Let Θ_0 correspond to the values of θ for anomalies which we consider as significant, i.e., they are neither extremely small nor represent the majority of the data. Formally, for $\gamma_0, \gamma_1 \in (0, 1), \gamma_0 < \gamma_1$, we let

$$\Theta_0 = \{\theta \in \Theta : \gamma_0 |W| \leq |I_\theta| \leq \gamma_1 |W|\}.$$

Then $\Theta_1 = \Theta \setminus \Theta_0$ corresponds to extremely small or large anomalies, i.e.,

$$\Theta_1 = \{\theta \in \Theta : |I_\theta| < \gamma_0 |W|, \text{ or } |I_\theta^c| < (1 - \gamma_1) |W|\}.$$

4.2 Testing the change of expectation

Now we can formulate the change point hypotheses for the random field $\{s_k, k \in W\}$ with respect to its expectation as follows.

$H'_0 : \mathbb{E}s_k = \mu$ for every $k \in W$ (i.e. $h = 0$) vs.
 $H'_1 : \text{There exists } \theta_0 \in \Theta_0 \text{ such that } \mathbb{E}s_k = \mu + h, k \in I_{\theta_0}, h \neq 0, \text{ and } \mathbb{E}s_k = \mu, k \in I_{\theta_0}^c.$

Consider the following change-in-mean statistic for the sample $S = \{s_k, k \in W\}$:

$$\begin{aligned} Z(\theta) &= \frac{1}{|I_\theta|} \sum_{k \in I_\theta} s_k - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} s_k \\ &= \frac{1}{|I_\theta|} \sum_{k \in I_\theta} (\xi_k + \mu + h \mathbb{I}\{k \in I_{\theta_0}\}) \\ &\quad - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} (\xi_k + \mu + h \mathbb{I}\{k \in I_{\theta_0}\}) \\ &= \frac{1}{|I_\theta|} \sum_{k \in I_\theta} \xi_k - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} \xi_k + h \left(\frac{|I_\theta \cap I_{\theta_0}|}{|I_\theta|} - \frac{|I_\theta^c \cap I_{\theta_0}|}{|I_\theta^c|} \right). \end{aligned} \quad (9)$$

$$(10)$$

In order to test H'_0 vs. H'_1 we use the test statistic

$$\begin{aligned} T_W(S) &= \max_{\theta \in \Theta_0} |Z(\theta)| \\ &= \max_{\theta \in \Theta_0} \left| \frac{1}{|I_\theta|} \sum_{k \in I_\theta} s_k - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} s_k \right|. \end{aligned} \quad (11)$$

We reject H'_0 if $T_W(S)$ exceeds the critical value y_α . Let us find such $y_\alpha > 0$ via the probability of the 1st-type error $\mathbf{P}_{H'_0}(\max_{\theta \in \Theta_0} |Z(\theta)| \geq y_\alpha) \leq \alpha$. It holds

$$\begin{aligned} \mathbf{P}_{H'_0} \left(\max_{\theta \in \Theta_0} |Z(\theta)| \geq y_\alpha \right) &= \mathbf{P}_{H'_0} \left(\max_{\theta \in \Theta_0} \left| \eta(\theta) + h \left(\frac{|I_\theta \cap I_{\theta_0}|}{|I_\theta|} - \frac{|I_\theta^c \cap I_{\theta_0}|}{|I_\theta^c|} \right) \right| \geq y_\alpha \right) \\ &= \mathbf{P}(|\max_{\theta \in \Theta_0} \eta(\theta)| \geq y_\alpha), \end{aligned}$$

where

$$\begin{aligned} \eta(\theta) &:= Z(\theta) - \mathbb{E}Z(\theta) \\ &= \frac{1}{|I_\theta|} \sum_{k \in I_\theta} \xi_k - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} \xi_k, \theta \in \Theta \end{aligned}$$

is the centered field $Z(\theta)$.

Thus, we find the bounds for tail probabilities of the random variable $\max_{\theta \in \Theta_0} |\eta(\theta)|$. To do so, we use the ideas from Heinrich (1990) to get the following bounds for m -dependent random fields. For the sake of generality, our results are formulated in $\mathbb{Z}^D$, $D \in \mathbb{N}$.

Theorem 1 Let $\{\xi_k, k \in \mathbb{Z}^D\}$ be a stationary real-valued m -dependent centered random field and $\{b_k, k \in \mathbb{Z}^D\}$ be real

numbers. Assume that there exist $H, \sigma > 0$ such that

$$\mathbb{E}|\xi_k^p| \leq \frac{p!}{2} H^{p-2} \sigma^2, \quad p = 2, 3, \dots \quad (12)$$

Then for any $W \subset \mathbb{Z}^D$, $|W| < \infty$ we have

$$\begin{aligned} \mathbf{P} \left(\left| \sum_{k \in W} \xi_k b_k \right| \geq y \right) &\leq 2 \exp \left(- \frac{y^2}{4m^D \sigma^2 \|b\|_2^2} \right) \\ &\quad \times \mathbb{I} \left\{ 0 < y \leq \frac{\sigma^2 \|b\|_2^2}{H \|b\|_\infty} \right\} \\ &\quad + 2 \exp \left(- \frac{y}{2Hm^D \|b\|_\infty} + \frac{\sigma^2 \|b\|_2^2}{4H^2 m^D \|b\|_\infty^2} \right) \\ &\quad \times \mathbb{I} \left\{ y > \frac{\sigma^2 \|b\|_2^2}{H \|b\|_\infty} \right\}, \end{aligned} \quad (13)$$

where $\|b\|_\infty = \max_{k \in W} |b_k|$ and $\|b\|_2^2 = \sum_{k \in W} b_k^2$.

Theorem 2 Let $\{\xi_k, k \in \mathbb{Z}^3\}$ be a stationary real-valued m -dependent centered random field and there exist $H, \sigma > 0$ such that inequality (12) holds. If $0 < \gamma_0 < \gamma_1 < 1/2$, then

$$\begin{aligned} \mathbf{P} \left(\max_{\theta \in \Theta_0} |\eta(\theta)| \geq y \right) &\leq \sum_{\theta \in \Theta_0: |I_\theta^c| \leq \frac{\sigma^2 |W|}{yH}} 2e^{-\frac{y^2}{4m^3 \sigma^2} \frac{|I_\theta^c| |I_\theta|}{|W|}} \\ &\quad + \sum_{\theta \in \Theta_0: |I_\theta^c| > \frac{\sigma^2 |W|}{yH}} 2e^{-\frac{y}{2Hm^3} |I_\theta| + \frac{\sigma^2 |W|}{4H^2 m^3 |I_\theta^c|} |I_\theta|}. \end{aligned} \quad (14)$$

Hence, we reject H'_0 if the test statistic $T_W(S) \geq y_\alpha$, where critical value y_α is the minimum positive number such that

$$\begin{aligned} \sum_{\theta \in \Theta_0: |I_\theta^c| \leq \frac{\sigma^2 |W|}{y_\alpha H}} 2e^{-\frac{y_\alpha^2}{4m^3 \sigma^2} \frac{|I_\theta^c| |I_\theta|}{|W|}} &+ \sum_{\theta \in \Theta_0: |I_\theta^c| > \frac{\sigma^2 |W|}{y_\alpha H}} 2e^{-\frac{y_\alpha}{2Hm^3} |I_\theta| + \frac{\sigma^2 |W|}{4H^2 m^3 |I_\theta^c|} |I_\theta|} \leq \alpha. \end{aligned} \quad (15)$$

Remark 3 In the case of a Gaussian random field $\{\xi_k, k \in \mathbb{Z}^3\}$, we can put $H = \sigma$ in (12) and (15), where $\mathbb{E}\xi_0^2 \leq \sigma^2$. Indeed, $\mathbb{E}|\xi_0|^p \leq \sigma^{p-2} \sigma^2 \mathbb{E}|\xi_0|^p$, $p \geq 2$, where $\xi \sim N(0, 1)$.

If the field $\{\xi_k, k \in \mathbb{Z}^3\}$ is bounded, i.e., $|\xi_k| \leq M_0$ a.s., then $\mathbb{E}|\xi_k|^p \leq \mathbb{E}(\xi_k^2 |\xi_k|^{p-2}) \leq M_0^{p-2} \mathbb{E}\xi_k^2 \leq M_0^{p-2} \sigma^2$. Therefore, we put $H = M_0$ in (12) and (15).

Remark 4 Our hypotheses H'_0 and H'_1 are formulated to detect the change of the mean in random field. Under additional assumptions, that the field is centered, a similar approach can be used to detect the changes in the variance of the field if one considers the mean square deviation from the mean of the field. To detect the changes in the spectral density of the field, an alternative approach has to be designed.

4.3 Comparison with other methods

The bounds for tail probabilities of $T_W(S)$ derived in Brodsky and Darkhovsky (1993) under some mixing assumptions contain the mixing coefficients. Unfortunately, it is not realistic to estimate these coefficients or to give natural bounds for them.

The tail probabilities of the maximum of a stationary Gaussian field are considered in the paper Siegmund and Worsley (1995). Although our statistic $\{Z(\theta), \theta \in \Theta_0\}$ forms a Gaussian random field, it is not stationary.

The approach described in Bucchia (2014), Bucchia and Wendler (2017) is based on the limiting behaviour of T_W as the size of the observation window W grows to infinity. We are limited in the use of this approach because of the insufficient size of W and high computational costs. Moreover, the bounds for tail probabilities are not precise in this case.

The asymptotic behaviour of the CUSUM statistics on multivariate time series is also widely studied. In our setting, one can construct the corresponding time series by averaging suitable characteristics (entropy and local direction) over slices orthogonal to the coordinate axes and apply recent methods of multivariate change point analysis Schwaar (2017). However, the m -dependence and the insufficient length of the constructed series (approx. 10–20 time lags) falsify the first kind error.

All in all, the merit of our approach based on m -dependent random fields (in comparison with other mentioned methods) is that it allows for (somewhat conservative) non-asymptotic statistical tests of the “change point in the mean”-hypothesis for moderate size image data.

4.4 Change point detection in simulated random fields

In this section, we study the behaviour of the test statistic $T_W(S)$ given in (11) and probabilities of 1st-type error $\mathbf{P}_{H'_0}(T_W(S) \geq y_\alpha)$ with respect to different values of σ^2 and m .

The form of T_W allows to test the existence of the anomaly regions of arbitrary form and arbitrary number of connected components. On the other hand, we need to decrease the value $|\Theta|$ up to a feasible quantity for computational reasons [see bounds (34) and (35)]. Let $W = [1, M_1] \times [1, M_2] \times [1, M_3] \cap \mathbb{N}^3$. We fix $\gamma_0 = 0.05$ and $\gamma_1 = 0.5$, as the anomaly

should not cover the majority of the window. In this paper we restrict I_θ to be a single rectangular parallelepiped of the form $[1 + \Delta_0 i_1, 1 + \Delta_0 i_1 + \Delta_1 l_1] \times [1 + \Delta_0 i_2, 1 + \Delta_0 i_2 + \Delta_1 l_2] \times [1 + \Delta_0 i_3, 1 + \Delta_0 i_3 + \Delta_1 l_3]$. Then the parametric set of significant anomaly regions is given by

$$\begin{aligned} \Theta_0 = \{ & (1 + \Delta_0 i_1, 1 + \Delta_0 i_2, 1 + \Delta_0 i_3, \Delta_1 l_1, \Delta_1 l_2, \Delta_1 l_3), \\ & \text{where } (i_1, i_2, i_3, l_1, l_2, l_3) \in \mathbb{N}_0^6, \\ & 1 + \Delta_0 i_j + \Delta_1 l_j \leq M_j, l_j \geq L_M, 1 \leq j \leq 3, \\ & \gamma_0 \leq \frac{|I_\theta \cap J|}{M_1 M_2 M_3} \leq \gamma_1 \} \end{aligned} \quad (16)$$

The offset parameters Δ_0 and Δ_1 as well as the parameter L_M controlling the minimal edge length of the cuboids have to be chosen by the user.

Assuming the m -dependence for our observations, we do not know the exact value of m . Hence, we need to impose some restrictions on the field ξ . First, if we know a-priori the maximum length of a typical fiber we can immediately obtain the bound for m . Second, we can estimate the covariance function of the random field $\{s_k, k \in W\}$ and assess m as the range when this empirical covariance is sufficiently close to zero. From relation (35) with $\frac{\sigma^2}{H(1-\gamma_1)} < y < H$ we obtain the following approximate bound for an admissible m :

$$\begin{aligned} 2|\Theta_0| \exp\left(-\frac{y}{4Hm^3} \gamma_0 |W|\right) & \leq \alpha \\ \Rightarrow m^3 & \leq \frac{\gamma_0}{4 \log(2|\Theta_0|/\alpha)} |W|. \end{aligned} \quad (17)$$

For example, for $|\Theta_0| = 10^4$, $\alpha = 0.05$, $\gamma_0 = 0.05$, one gets

$$m \leq \frac{1}{10} |W|^{1/3} \text{ or } |W| \geq 10^3 m^3. \quad (18)$$

Let us now compare the empirical probability of the error of the 1-st type with the bounds (15) for $\mathbf{P}_{H'_0}(T_W(\cdot) \geq y_\alpha)$. We generate 3000 realizations of a Gaussian centered m -dependent ($m = 10$) random field $\{Y_k, k \in W\}$ with $W = [1, 80]^3 \cap \mathbb{N}^3$ (which is matched to the considered data sets) and $Y_k \sim N(0, 1)$. The dependence is modelled as follows: random variables Y_{l+mk} , $k \geq 0$ are independent, and $Y_{l+mk} = Y_{l+mk}$, $k \in \mathbb{N}^3$ for all $l \in \{1, \dots, m\}^3$. We take $\Delta_0 = \Delta_1 = 8$, $\gamma_0 = 0.05$, and $\gamma_1 = 0.5$. In this case, $|\Theta_0| = 11,954$. Based on the simulated sample of values of the test statistic $T_W(Y)$ we compute the empirical critical value $\hat{y}_\alpha = 0.6413$ for $\alpha = 0.05$. From comparison of $\hat{y}_\alpha$ with critical values y_α based on inequality (15) with $H = \sigma$ (presented in Table 3), we see that even under the exact value of $\sigma^2 = 1$, critical values y_α are quite conservative. For example, $y_\alpha = 0.7197$ for $m = 7$ is still greater than $\hat{y}_\alpha = 0.6413$ generated for $m = 10$. Therefore, we can

Table 3 Critical values $y_{0.05}$ based on inequality (15) for different values of m and σ

σ^2	$m = 10$	$m = 9$	$m = 8$	$m = 7$	$m = 6$
1	1.0757	1.0493	0.8793	0.7197	0.5711
4	2.1513	2.0985	1.7587	1.4395	1.1423
8	3.0424	2.9677	2.4871	2.0357	1.6154
σ^2	$m = 5$	$m = 4$	$m = 3$	$m = 2$	
1	0.4345	0.3109	0.2019	0.1099	
4	0.8689	0.6218	0.4039	0.2198	
8	1.2289	0.8793	0.5711	0.3109

Fig. 3 Tail probabilities $T_W(S)$ (black) and the values of bound (14) for several values of m

use critical values from inequality (15) with m smaller than its real value.

Analyzing the values in Table 3 (cf. also Table 7 in the Appendix for its extended version) we see that the critical values are more sensitive with respect to misspecification of m rather than σ . This fact can be analytically shown with the help of formulas (34) and (35), that is, $y_\alpha \propto \sigma m^{3/2}$ or $y_\alpha \propto \sigma m^3$ for $H = \sigma$, when other parameters are fixed.

To illustrate how tight bound (14) is, we compute the tail probabilities $\mathbf{P}(T_W(S) \geq y)$, $y \geq 0$ based on our simulated samples. The graph of $\mathbf{P}(T_W(S) \geq y)$ together with the values of bound (14) (with $\sigma = 1$ and $m = 10, 9, 8, 7, 6$) are presented in Fig. 3. One can see that estimate (14) with the correct value of $m = 10$ is conservative and for the value $m = 7$ it is quite tight. The lower values of $m \leq 6$ lead to the wrong critical values, which also holds for Tables 3 and 7.

Now we simulate the m -dependent ($m = 10$) Gaussian random field Y_k , $k \in W$ with anomaly zones and investigate the power of our change point test. The simulation procedure is described above. We consider three anomaly zones:

$I_{\theta_1} = [1, 30] \times [1, 30] \times [1, 30] \cap \mathbb{N}^3$ ($|I_{\theta_1}|/|W| \approx 0.05$),
 $I_{\theta_2} = [1, 40] \times [1, 40] \times [1, 80] \cap \mathbb{N}^3$ ($|I_{\theta_2}|/|W| = 0.25$),
 $I_{\theta_3} = [1, 40] \times [1, 70] \times [1, 80] \cap \mathbb{N}^3$, ($|I_{\theta_3}|/|W| \approx 0.44$).
Then we build the observable samples $S_k^i(h) = Y_k + h\mathbb{I}\{k \in I_{\theta_i}\}$, $k \in W$, $i = 1, 2, 3$ with $h \in [-1, 1]$. Since the law of $\{Y_k, k \in \mathbb{Z}^3\}$ is invariant with respect to translations, the power of the test depends in fact on the size of I_{θ_i} only.

Let $\alpha = 0.05$. Then the empirical critical value is $\hat{y}_\alpha = 0.6413$. The critical values y_α obtained from (15) with $\sigma = 1$, $m = 10$, and $m = 7$ are 1.0757 and 0.7197, respectively. We simulate $S_k^i(h)$ 3000 times in order to compute the power $\mathbf{P}(T_W(S^i)(h) \geq y_\alpha)$, $h \in [-1, 1]$, $i = 1, 2, 3$. We present results on Fig. 4 for anomaly regions I_{θ_1} , I_{θ_2} , and I_{θ_3} separately. One can mention that the power of our test increases with the volume of the anomaly zone.

5 Cluster based anomaly detection

For processes $\mathcal{E}$ of thick fibres introduced in Sect. 2, the evidence of an anomaly is tested by applying the test of Sect. 4 to the random field $\{s_k, k \in W\}$ of estimated local mean or entropy of the chosen fibre characteristic w . In this paper, $w(x)$ is the average direction vector of the fibres of $\mathcal{E}$ at $x \in W$ or one of its coordinates (introduced in Sect. 3 as $\hat{w}_i$).

Assume that the anomaly test presented in Sect. 4 rejected the hypothesis H'_0 (and hence H_0), i.e., we have an evidence of an anomalous fibre behaviour in the rectangular subregion I_{θ_0} of our image data. Now we are interested in a more accurate estimate of the geometry of this anomaly. The search for an anomaly region in a 3D image can be interpreted as a problem of splitting the volume of the image into two disjoint clusters: *homogeneous material* and *anomaly*.

In our problem setting, the volume under investigation, $\bigcup_{i \in J_W} W_i$, is a union of scanning windows W_i with meaningful local direction information. Each of them yields the clustering attributes mean of local directions (MLD) and entropy. We need to classify all the windows W_i as either belonging to the homogeneous material or the anomaly. For this purpose, a spatial version of the Stochastic Expectation Maximization algorithm is used.

5.1 Spatial modification of a Stochastic Approximation Expectation Maximization (SAEM) algorithm

We assume that under the alternative H_1 (see page 2), fibres in the material may have two different distributions f_0 and f_1 of local directions. Therefore, the distribution of clustering attributes is a mixture of the distributions f_0 and f_1 .

The Expectation-Maximization algorithm (EM) is commonly used to separate modes in a finite mixture of distributions, cf. Korolev (2007) for a review. It is an iterative

Fig. 4 The power of the change point test for anomaly regions I_{θ_1} (left), I_{θ_2} (center), and I_{θ_3} (right). Black color corresponds to values of $\mathbf{P}(T_W(S^i)(h) \geq 1.0757)$ and blue color corresponds to values of

$\mathbf{P}(T_W(S^i)(h) \geq 0.7197)$, i.e., for y_α from (15) with $m = 10$ and $m = 7$ respectively. Values of $\mathbf{P}(T_W(S^i)(h) \geq \hat{y}_\alpha)$, the highest possible in our test procedure, are indicated by red color

procedure consisting of two steps: Expectation (Estimation) and Maximization. In general, one assumes that the probability law under study is a mixture of k distributions from the same parametric family. In the first step, the hidden parameters of the sample distribution, i.e., the weights of the mixture components, are estimated, while in the second step the resulting parameters are updated by maximizing the likelihood function.

Since the EM algorithm belongs to the so-called “greedy” algorithms, that is, it converges to the first local optimum that has been found, a modification that compensates this deficiency should be used. One way out is a random “shaking” of observations in each iteration. This method is the basis of the *Stochastic EM (SEM) algorithm* (cf. Gorshenin et al. 2017; Korolev 2007).

The SEM algorithm works relatively fast in comparison with other methods, and its results are non-sensitive to an initial approximation. Random perturbations on the parameter space in the S-step guarantee the convergence to the global maximum of the likelihood function and help to avoid unstable local maxima. On the other hand, the outputs of the SEM algorithm form a Markov chain and the final solution is its stationary distribution. To avoid this additional problem we use a modification called SAEM (Stochastic Approximation of EM) algorithm which brings together advantages of both EM and SEM approaches, e.g. Celeux and Diebolt (1992).

Assume that the observable distribution has a density of the form

$$\varphi_\delta(x) = \beta\varphi(x, \delta_1) + (1 - \beta)\varphi(x, \delta_2), \quad x \in \mathbb{R}^d,$$

where $\varphi(x, \delta_i)$ is a multivariate Gaussian density with unknown parameter $\delta_i = (\mu_i, \Sigma_i)$, $i = 1, 2$ and $\beta \in [0, 1]$. Here μ_i is the mean and Σ_i is the covariance matrix of Gaussian component $i = 1, 2$. The combined unknown parameter is $\delta = (\beta, \delta_1, \delta_2)$. We call $\varphi(\cdot, \delta_1)$ and $\varphi(\cdot, \delta_2)$ the *first* and the *second component of the mixture*, respectively.

For each observation x_1 , $\mathbf{l} \in J_W$, we define a new variable $y_1 = \mathbb{I}\{x_1 \text{ belongs to the first component}\}$. Therefore, we have two samples: observable $\mathbf{x} = \{x_1, \mathbf{l} \in J_W\}$ and unobservable $\mathbf{y} = \{y_1, \mathbf{l} \in J_W\}$. Then the log-likelihood function equals

$$\begin{aligned} \ln L(\delta, \mathbf{x}, \mathbf{y}) &= \sum_{\mathbf{l} \in J_W} [y_1 \ln(\beta\varphi(x_1, \delta_1)) + (1 - y_1) \ln((1 - \beta)\varphi(x_1, \delta_2))] \\ &= v_1 \ln \beta + v_2 \ln(1 - \beta) + \sum_{\mathbf{l} \in J_W: y_1=1} \ln \varphi(x_j, \delta_1) \\ &\quad + \sum_{\mathbf{l} \in J_W: y_1=0} \ln \varphi(x_j, \delta_2), \end{aligned}$$

where $v_1 = \sum_{\mathbf{l} \in J_W} y_1$ denotes the number of observations belonging to the first mixture component and $v_2 = m_1 m_2 m_3 - v_1$ observations belong to the second one.

Assume that we know the a posteriori probability $q_1^{(k-1)}$, $\mathbf{l} \in J_W$, that x_1 belongs to the first component $\varphi(\cdot, \delta_1)$, where $k - 1$ is the iteration number. Let us describe the EM part. During the M-Step we obtain new estimates of the parameters $\hat{\delta}_1^{(k)} = (\mu_1^{(k)}, \Sigma_1^{(k)})$, $\hat{\delta}_2^{(k)} = (\mu_2^{(k)}, \Sigma_2^{(k)})$, $\hat{\beta}^{(k)}$ by

$$\begin{aligned} \mu_1^{(k)} &= \frac{\sum_{\mathbf{l} \in J_W} q_1^{(k-1)} x_1}{\sum_{\mathbf{l} \in J_W} q_1^{(k-1)}}, \\ \Sigma_1^{(k)} &= \frac{\sum_{\mathbf{l} \in J_W} q_1^{(k-1)} (x_1 - \mu_1^{(k)})(x_1 - \mu_1^{(k)})^T}{\sum_{\mathbf{l} \in J_W} q_1^{(k-1)}}, \end{aligned} \quad (19)$$

$$\mu_2^{(k)} = \frac{\sum_{\mathbf{l} \in J_W} (1 - q_1^{(k-1)}) x_1}{\sum_{\mathbf{l} \in J_W} (1 - q_1^{(k-1)})}, \quad (20)$$

$$\Sigma_2^{(k)} = \frac{\sum_{\mathbf{l} \in J_W} (1 - q_1^{(k-1)}) (x_1 - \mu_2^{(k)})(x_1 - \mu_2^{(k)})^T}{\sum_{\mathbf{l} \in J_W} (1 - q_1^{(k-1)})}, \quad (21)$$

$$\hat{\beta}^{(k)} = \frac{1}{m_1 m_2 m_3} \sum_{\mathbf{l} \in J_W} q_{\mathbf{l}}^{(k-1)}. \quad (22)$$

In the E-step we compute the new probabilities based on (19)–(22) as

$$q_{\mathbf{l}}^{k,EM} = \frac{\hat{\beta}^{(k)} \varphi(x_{\mathbf{l}}, \hat{\delta}_1^{(k)})}{\hat{\beta}^{(k)} \varphi(x_{\mathbf{l}}, \hat{\delta}_1^{(k)}) + (1 - \hat{\beta}^{(k)}) \varphi(x_{\mathbf{l}}, \hat{\delta}_2^{(k)})}, \quad \mathbf{l} \in J_W. \quad (23)$$

In the SEM-part we act in a different way. In the S-step we generate independent Bernoulli-distributed random variables $y_{\mathbf{l}}^{(k)} \in \{0, 1\}$ with probabilities $\mathbf{P}(y_{\mathbf{l}}^{(k)} = 1) = q_{\mathbf{l}}^{(k-1)}, \mathbf{l} \in J_W$.

During the M-Step we get $v_1^{(k)} = \sum_{\mathbf{l} \in J_W} y_{\mathbf{l}}^{(k)}$ and the estimates $\hat{\delta}_1^{(k)} = (\mu_1^{(k)}, \Sigma_1^{(k)})$, $\hat{\delta}_2^{(k)} = (\mu_2^{(k)}, \Sigma_2^{(k)})$, $\hat{\beta}^{(k)}$ by

$$\begin{aligned} \mu_1^{(k)} &= \frac{\sum_{\mathbf{l} \in J_W: y_{\mathbf{l}}^{(k)}=1} x_{\mathbf{l}}}{v_1^{(k)}}, \\ \Sigma_1^{(k)} &= \frac{\sum_{\mathbf{l} \in J_W: y_{\mathbf{l}}^{(k)}=1} (x_{\mathbf{l}} - \mu_1^{(k)})(x_{\mathbf{l}} - \mu_1^{(k)})^T}{v_1^{(k)}}, \end{aligned} \quad (24)$$

$$\begin{aligned} \mu_2^{(k)} &= \frac{\sum_{\mathbf{l} \in J_W: y_{\mathbf{l}}^{(k)}=0} x_{\mathbf{l}}}{v_2^{(k)}}, \\ \Sigma_2^{(k)} &= \frac{\sum_{\mathbf{l} \in J_W: y_{\mathbf{l}}^{(k)}=0} (x_{\mathbf{l}} - \mu_2^{(k)})(x_{\mathbf{l}} - \mu_2^{(k)})^T}{v_2^{(k)}}, \end{aligned} \quad (25)$$

$$\hat{\beta}^{(k)} = \frac{v_1^{(k)}}{m_1 m_2 m_3} = \frac{1}{m_1 m_2 m_3} \sum_{\mathbf{l} \in J_W} y_{\mathbf{l}}^{(k)}. \quad (26)$$

In the E-Step, we compute the updated probabilities $q_{\mathbf{l}}^{(k,SEM)}$ based on (24)–(26) by relation (23).

The essential idea of the SAEM algorithm is to mix $q_{\mathbf{l}}^{(k,EM)}$ and $q_{\mathbf{l}}^{(k,SEM)}$ in iteration step k as

$$q_{\mathbf{l}}^{(k)} = \lambda_k q_{\mathbf{l}}^{(k,SEM)} + (1 - \lambda_k) q_{\mathbf{l}}^{(k,EM)}, \quad \mathbf{l} \in J_W, \quad (27)$$

which gives the a posteriori probabilities for the next $(k+1)$ th iteration. Here in (27), $\{\lambda_k, k \geq 1\}$ is a sequence of positive real numbers $\lambda_k \in (0, 1)$ decreasing to zero. We stop the SAEM algorithm after the k -th iteration if $\sum_{\mathbf{l} \in J_W} |q_{\mathbf{l}}^{(k-1)} - q_{\mathbf{l}}^{(k)}| \leq \varepsilon$, where ε is some threshold. In the following, the choice of the parameters is a result of experimental tuning to our image data yielding good practical results. In particular, we use $\lambda_k = \frac{50}{50+k^2}, k \geq 1$ and $\varepsilon = 0.0001$ in our computations.

When the SAEM algorithm stops in the k_0 -th iteration, we obtain the values $\{q_{\mathbf{l}}^{(k_0)}, \mathbf{l} \in J_W\}$ which indicate that $x_{\mathbf{l}}$ belongs to the first component if $q_{\mathbf{l}}^{(k_0)} > 1/2$ and to the second one, otherwise.

Applying the above SAEM algorithm to our image data yields diffuseness in the resulting clusters (see Fig. 9). To avoid this, we propose a smoothing modification (*Spatial SAEM*), which takes the spatial location of the sample data into account. Let us describe the new *Spatial step*.

Let SAEM stop after k_0 iterations. For each sample entry $x_{\mathbf{l}}, \mathbf{l} = (l_1, l_2, l_3) \in J_W$ we define the coordinate $v_{\mathbf{l}} = ((l_1 - 1)M\Delta, (l_2 - 1)M\Delta, (l_3 - 1)M\Delta)$, that is a vertex of the cube $W_{\mathbf{l}}$. In each further iteration, i.e. for $k > k_0$, Bernoulli random variables $y_{\mathbf{l}}^{(k)} \in \{0, 1\}, \mathbf{l} \in J_W$ with success probability $q_{\mathbf{l}}^{(k_0)}, \mathbf{l} \in J_W$ are simulated. Now $y_{\mathbf{l}}^{(k)}$ classifies $x_{\mathbf{l}}$ in such a way that $y_{\mathbf{l}}^{(k)} = 0$ indicates that $x_{\mathbf{l}}$ belongs to the first component and the second one, otherwise. Then, we compute the number $a_{\mathbf{l}}^{(k)}$ of neighbors of $x_{\mathbf{l}}$ belonging to the same cluster as $x_{\mathbf{l}}$. Neighborhood is defined in terms of the r -neighborhood of $v_{\mathbf{l}}$ such that

$$a_{\mathbf{l}}^{(k)} = \sum_{\mathbf{i} \in J_W, \mathbf{i} \neq \mathbf{l}} \mathbb{I}(y_{\mathbf{i}}^{(k)} = y_{\mathbf{l}}^{(k)}, \|v_{\mathbf{l}} - v_{\mathbf{i}}\|_{\infty} \leq r), \quad r > 0. \quad (28)$$

If all $a_{\mathbf{l}}^{(k)}$ are greater than or equal to a certain threshold a (for our image data, $a = 3$ is used) we call the classification $y_{\mathbf{l}}^{(k)}, \mathbf{l} \in J_W$, *admissible* and move to the next iteration. Otherwise, for sample entries $x_{\mathbf{l}}$ with $a_{\mathbf{l}}^{(k)}$ being less than a , we change $y_{\mathbf{l}}^{(k)}$, hence, the class of $x_{\mathbf{l}}$. If the new set $y_{\mathbf{l}}^{(k)}, \mathbf{l} \in J_W$, is admissible, then we pass to iteration $k + 1$. If it is not, we resimulate $y_{\mathbf{l}}^{(k)}, \mathbf{l} \in J_W$, until $a_{\mathbf{l}}^{(k)} \geq a$ for all $\mathbf{l} \in J_W$.

The smoothing procedure stops when K admissible classifications have been generated ($K = 1000$ is used). Since the classifications $y^{(k)}, k \geq 1$ form a Markov chain tending to a stationary distribution $q = \{q_{\mathbf{l}}, \mathbf{l} \in J_W\}$, we take the last K realizations of $y^{(k)}$ for the estimation of q . The final a posteriori probabilities $q_{\mathbf{l}}$ in the space of all admissible classifications are computed over the sample of $\{y_{\mathbf{l}}^{(k)}, \mathbf{l} \in J_W\}, k_0 \leq k \leq k_0 + K$ by

$$q_{\mathbf{l}} = \frac{1}{K} \sum_{k=k_0}^{k_0+K} y_{\mathbf{l}}^{(k)}, \quad \mathbf{l} \in J_W. \quad (29)$$

We also get estimates of the components' weights $(\hat{\beta}, 1 - \hat{\beta})$ by (22). If $\hat{\beta} \geq 0.5$ we say that the second component corresponds to the “anomaly”. The observation window $W_{\mathbf{l}}$ thus belongs to the zone of homogeneous material if $q_{\mathbf{l}} \geq 0.5$ and to the anomaly zone if $q_{\mathbf{l}} < 0.5$. If $\hat{\beta} < 0.5$ the roles of the components are swapped.

Fig. 5 Visualization of simulated layered RSA fibre data, $2000 \times 2000 \times 2100$ voxels

Fig. 6 Visualization of simulated homogeneous RSA fibre data, $2000 \times 2000 \times 2100$ voxels

6 Application to 3D image data of fibre materials

6.1 Simulated data

First, we illustrate the use of the methods from Sects. 4 and 5 on simulated 3D fibre images. We choose a *random sequential absorption* (RSA) model that randomly adds fibres to the existing material, such that they do not intersect each other, cf. Andrä et al. (2014); Redenbach and Vecchio (2011). Figure 5 shows simulated RSA fibre data in an image of $2000 \times 2000 \times 2100$ voxels. The sample exhibits three layers, where fibres differ in their local directional distribution. Each layer has a thickness of 700 voxels and contains 82,474 fibres with a constant radius of 4 voxels and length of 100 voxels. Fibre directions are distributed according to a special case of the Angular Central Gaussian distribution described in Franke et al. (2016). In the two outer layers, the preferred direction is the x -direction and the concentration parameter is $\beta = 0.1$ resulting in a high concentration of the fibres along the main direction. In the middle layer, which is considered the anomaly region, the preferred direction is the y -direction and the fibres are less concentrated ($\beta = 0.5$).

Additionally, we investigate a homogeneous RSA data set where no anomalies should be detected. The data set consists of an image of $2000 \times 2000 \times 2100$ voxels. Here, the concentration parameter of the fibre direction distribution is $\beta = 0.1$ in the whole sample. The preferred direction is the x -direction. The fibre radius is 4 voxels, the fibre length is

100 voxels. A visualisation of a realisation of this model is shown in Fig. 6.

Now we apply the change point analysis of Sect. 4 to the random fields of mean local directions and entropy estimates for the homogeneous and layered RSA data.

To do so, we transform the data of average local directions $X_{\mathbf{k}}, \mathbf{k} \in J$. In order to avoid cancelling due to averaging, we build for each coordinate x, y, z , the samples $\tilde{x}_{\mathbf{k}}, \tilde{y}_{\mathbf{k}}, \tilde{z}_{\mathbf{k}}$, such that their entries lie in the hemispheres $\mathbb{S}_x^2 = \{(x, y, z) \in \mathbb{S}, x \geq 0\}$, $\mathbb{S}_y^2 = \{(x, y, z) \in \mathbb{S}, y \geq 0\}$, $\mathbb{S}_z^2 = \{(x, y, z) \in \mathbb{S}, z \geq 0\}$, respectively, i.e., $\tilde{x}_{\mathbf{k}} = |x|_{\mathbf{k}}, \tilde{y}_{\mathbf{k}} = |y|_{\mathbf{k}}, \tilde{z}_{\mathbf{k}} = |z|_{\mathbf{k}}$.

The sample of the estimated entropy values $\hat{E}_{\mathbf{k}}$ of the directional distribution of fibres $X_{\mathbf{i}}, \mathbf{i} \in J$ in the windows $W_{\mathbf{k}}, \mathbf{k} \in J_W$ is built by estimator (7) over the transformed directions $\tilde{X}_{\mathbf{i}} \in \mathbb{S}_+^2, \mathbf{i} \in J$. (Fig. 7)

We apply the results of Sect. 4 to the random fields $s_{\mathbf{k}} = \tilde{x}_{\mathbf{k}}, \tilde{y}_{\mathbf{k}}$, or $\tilde{z}_{\mathbf{k}}$, and $\hat{E}_{\mathbf{k}}$, which yields the following four pairs of hypotheses of (H'_0, H'_1) -type.

- $H_0^x : \mathbb{E}\tilde{x}_{\mathbf{k}} = \mu_x$ for every $\mathbf{k} \in W$ vs.
- $H_1^x : \exists \theta_0 \in \Theta_0$ such that $\mathbb{E}\tilde{x}_{\mathbf{k}} = \mu_x + h_x, \mathbf{k} \in I_{\theta_0}$ and $\mathbb{E}\tilde{x}_{\mathbf{k}} = \mu_x, \mathbf{k} \in I_{\theta_0^c}, h_x \neq 0$;
- $H_0^y : \mathbb{E}\tilde{y}_{\mathbf{k}} = \mu_y$ for every $\mathbf{k} \in W$ vs.
- $H_1^y : \exists \theta_0 \in \Theta_0$ such that $\mathbb{E}\tilde{y}_{\mathbf{k}} = \mu_y + h_y, \mathbf{k} \in I_{\theta_0}$ and $\mathbb{E}\tilde{y}_{\mathbf{k}} = \mu_y, \mathbf{k} \in I_{\theta_0^c}, h_y \neq 0$;
- $H_0^z : \mathbb{E}\tilde{z}_{\mathbf{k}} = \mu_z$ for every $\mathbf{k} \in W$ vs.
- $H_1^z : \exists \theta_0 \in \Theta_0$ such that $\mathbb{E}\tilde{z}_{\mathbf{k}} = \mu_z + h_z, \mathbf{k} \in I_{\theta_0}$ and $\mathbb{E}\tilde{z}_{\mathbf{k}} = \mu_z, \mathbf{k} \in I_{\theta_0^c}, h_z \neq 0$;
- $H_0^E : \mathbb{E}\hat{E}_{\mathbf{k}} = \mu_E$ for every $\mathbf{k} \in J_W$ vs.

Table 4 Change point test for mean local directions of homogeneous RSA data

Attribute	Sample var.	Test statistics	p -value
$\tilde{x}$	0.04360	0.0344	1.00
$\tilde{y}$	0.03743	0.0130	1.00
$\tilde{z}$	0.03749	0.0146	1.00
$\tilde{E}$	0.08984	0.0942	1.00

$$- H_1^E : \exists \theta_0 \in \Theta_0 \text{ such that } \mathbb{E} \hat{E}_{\mathbf{k}} = \mu_E + h_E, \mathbf{k} \in I_{\theta_0} \text{ and } \mathbb{E} \hat{E}_{\mathbf{k}} = \mu_E, \mathbf{k} \in I_{\theta_0^c}, h_E \neq 0.$$

Since we test only 4 hypotheses simultaneously, we stick to the classical Bonferroni method, e.g. we test each direction and entropy separately with significance level $\frac{1}{4}\alpha$.

Before running the algorithms we need to choose the right size of scanning windows. From the initial layered and homogeneous RSA images with $2000 \times 2000 \times 2100$ voxels we obtain $83 \times 83 \times 87$ small windows $\tilde{W}_i$ with $24 \times 24 \times 24$ voxels each, and 463,537 and 460,559 nonempty entries, respectively.

Due to the model parameters (fibre length of 100 voxels corresponds to 5 points in W), we can assume the random field $\tilde{X}_{\mathbf{k}}$ to be m -dependent with $m = 5$ and $\sigma^2 = 0.2$, $M_0 = \frac{1}{2}$. For the mean local directions $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$, the parametric set Θ_0 is constructed with $\Delta_0 = \Delta_1 = 8$, $\gamma_0 = 0.05$, $\gamma_1 = 0.5$, and $L_M = 22$ in (16), $|\Theta_0| = 39,395$.

We point out that the samples $(\tilde{x}, \tilde{y}, \tilde{z})$ and $\hat{E}$ have different sizes due to the construction described in Sect. 3. Therefore, the parameters m , σ^2 and the parameter set Θ_0 in (16) for $\hat{E}$ differ from the ones for $\tilde{x}, \tilde{y}, \tilde{z}$.

For the sample of estimated entropy values $\hat{E}_{\mathbf{k}}, \mathbf{k} \in J_W$, we have $m = 1$, $\sigma^2 = 0.5$ and the parametric set Θ_0 is constructed with $\Delta_0 = \Delta_1 = 2$, $\gamma_0 = 0.05$, $\gamma_1 = 0.5$, and $L_M = 4$ in (16), $|\Theta_0| = 16,536$. The Entropy values approximately have a normal distribution, so we put $H = \sigma$ in (15).

The computed statistics (given in (11)) $T_W(\tilde{x}), T_W(\tilde{y}), T_W(\tilde{z}), T_W(\hat{E})$ and the corresponding p -values from relation (15) are presented in Table 4 for the homogeneous and in Table 5 for the layered RSA data. Thus, there is no evidence to reject $H_0^x, H_0^y, H_0^z, H_0^E$ in the homogeneous case. The described test allows to claim that there is an anomaly region in the layered RSA image data, because we reject H_0^x, H_0^y , and H_0^E , but have no evidence to reject H_0^z .

Therefore, the choice of the mean of local directions attribute for the change point analysis in our problem with layered fibre image data is reasonable. Moreover, depending on the data (e.g. containing whirlpools of fibres) it may be better to choose entropy or other attributes to test for other types of anomalies.

Table 5 Change point test for mean local directions of layered RSA data

Attribute	Sample var.	Test statistics	p -value
$\tilde{x}$	0.10592	0.44036	4.6×10^{-30}
$\tilde{y}$	0.10948	0.43163	2.8×10^{-23}
$\tilde{z}$	0.06151	0.18764	0.301
$\tilde{E}$	0.3583	1.07030	0.00

Fig. 7 Histogram of frequencies of the local entropy of fibre directions and two separated Gaussian probability density functions found by the spatial SAEM algorithm. Layered RSA fibre data, $\sigma = 0.5986$

It follows from Penrose and Yukich (2013, Theorem 2.4.) that the Dobrushin estimator of the entropy of i.i.d. vectors on a C^1 -smooth manifold is asymptotically Gaussian. Although the RSA fibre data do not satisfy the i.i.d. assumption of mutual independence of fibre locations and directions, the estimated local directional entropy $\hat{E}$ for the homogeneous data seems to have a unimodal distribution, see Fig. 8. Assuming that the Gaussian distribution provides a reasonable approximation also in this case, we apply the 3σ -rule with σ^2 being the sample variance of $\hat{E}$, compare Alonso-Ruiz and Spodarev (2018), to find anomaly regions in both Figs. 5 and 6.

One can see that all centered entropy values lie in the interval $[-3\sigma; 3\sigma]$, so the 3σ -method does not distinguish between homogeneous (Fig. 6) and inhomogeneous (Fig. 5) images. We conclude that the 3σ -rule for anomaly detection does not work well if the anomaly regions are large enough to produce histograms of the clustering attribute with many modes.

The fact that the distribution of $\hat{E}$ seems to have two modes might indicate that it is a mixture of two Gaussian distributions. So we apply the Spatial SAEM algorithm from Sect. 5.1 to separate these modes. By an empirical study, a scanning window W_1 consisting of $5 \times 5 \times 5$ small windows $\tilde{W}_{\mathbf{k}}$ was selected, i.e. $M = 5$. The choice of W_1 is justified by the minimal sample size needed to the entropy estimation. Additionally, we put $r = \Delta M$ in (28) by default.

Fig. 8 Histogram of frequencies of the local entropy of fibre directions fitted Gaussian probability density. Homogeneous RSA fibre data, $\sigma = 0.2997$

Fig. 9 Anomaly detection in layered RSA fibre data using the local entropy: SAEM algorithm

First, let us consider clustering based on the attribute entropy. In the layered data, the Spatial SAEM algorithm finds two clusters and determines the distributional parameters for them. The clustering results are presented in Figs. 9 and 10. Green labels denote the centers of scanning windows corresponding to the homogeneous fibre material and blue labels mark objects belonging to the anomalous region. As expected, the Spatial SAEM algorithm found no anomaly for the homogeneous RSA fibre data.

We also ran the Spatial SAEM algorithm with the attributes mean of local direction (MLD) and a vector combining entropy and MLD. The results for the layered data are presented in Figs. 11 and 12. One can see that combination of both attributes gives a more reliable result.

Remark 5 The problem of clustering a fibre material into homogeneity and anomaly zones using vector-valued cluster attributes can be solved by a variety of other clustering methods, see the books Everitt et al. (2011); Hennig et al. (2016); Wierchoń and Kłopotek (2018) for an overview. In

Fig. 10 Anomaly detection in layered RSA fibre data using the local entropy: spatial SAEM algorithm

Fig. 11 Anomaly detection with spatial SAEM algorithm in the layered RSA fibre data using mean of local directions

Fig. 12 Anomaly detection with spatial SAEM algorithm in the layered RSA fibre data using local entropy and mean of local directions attributes

Fig. 13 3D image of a glass fibre reinforced composite material. $970 \times 1469 \times 1217$ voxels, voxel spacing: $4 \mu\text{m}$

Fig. 14 3D image of a glass fibre reinforced composite material. The part of the data containing an anomaly ($970 \times 700 \times 660$ voxels)

addition to the results reported here, we tried the recent AWC algorithm Efimov et al. (2019). However, the spatial SAEM approach yields better results, cf. Dresvyanskiy et al. (2019). Moreover, it does not require a complex parameter tuning and operates fast.

We also tried to use a principal axis of fibre directions as a classification attribute in the described SAEM algorithm. But the results are worse than the ones for the MLD attribute. This effect can be explained by the fact that the distribution of principal axes has its support on a unit sphere and thus cannot be a mixture of Gaussian distributions in $\mathbb{R}^3$. Therefore, the estimates in the M-step (19)–(21), (24), (25) have to be modified, cf. Franke et al. (2016). This task goes, however, beyond the scope of the present article.

6.2 Real glass fibre reinforced polymer

Now we apply our anomaly detection approach to a 3D-image of a glass fibre reinforced polymer. The images are provided by the Institute for Composite Materials (IVW) in Kaiserslautern, see Fig. 13. For a detailed description of the material we refer to Wirjadi et al. (2014).

We apply the change point analysis from Sect. 4 to real data with $970 \times 1469 \times 1217$ voxels and the estimated radius of 3 voxels. We obtain $64 \times 97 \times 80$ small windows $\tilde{W}_i$ with $15 \times 15 \times 15$ voxels (Figs. 13 and 14).

For mean local directions, the parametric set Θ_0 is constructed with $\Delta_0 = \Delta_1 = 8$, $\gamma_0 = 0.05$, $\gamma_1 = 0.5$ and $L_M = 22$ in (16), which gives $|\Theta_0| = 33,004$. To choose the suitable value of m for m -dependence we need additional investigation. Under the hypotheses H_0^x , H_0^y , H_0^z the random fields $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ are assumed to be stationary, so that we can estimate their covariance functions. We use the standard approach and estimate e.g. $\rho_x(\mathbf{h}) = \text{Cov}(\tilde{x}_1, \tilde{x}_{1+\mathbf{h}})$ as

Table 6 Change point test for mean local directions of real data

Attribute	Sample var.	Test statistics	p -value
$\tilde{x}$	0.04589	0.15995	1.00
$\tilde{y}$	0.06795	0.44733	2.1×10^{-10}
$\tilde{z}$	0.07982	0.43383	1.3×10^{-6}
E	0.30126	0.46811	3.96×10^{-8}

$$\hat{\rho}(\mathbf{h}) = \frac{1}{|K| - 1} \sum_{\mathbf{k} \in K} (\tilde{x}_{\mathbf{k}} - \bar{x}_0) (\tilde{x}_{\mathbf{k}+\mathbf{h}} - \bar{x}_{\mathbf{h}}),$$

where $K = \{(k_1, k_2, k_3) \in \mathbb{N}^3, 1 \leq k_1 \leq M_1 - h_1, 1 \leq k_2 \leq M_2 - h_2, 1 \leq k_3 \leq M_3 - h_3\}$, and $\bar{x}_0$ and $\bar{x}_{\mathbf{h}}$ are the sample means of $\tilde{x}$ over the index ranges K and $K + \mathbf{h}$ respectively. To visualize $\hat{\rho}_x$ we compute its maximum values in the following way:

$$\hat{\rho}_{x,\max}(i) = \max_{1 \leq k_1, k_2, k_3 \leq i} (\hat{\rho}_x(i, k_2, k_3), \hat{\rho}_x(k_1, i, k_3), \hat{\rho}_x(k_1, k_2, i)), i \geq 1.$$

The values of $\hat{\rho}_{x,\max}$, $\hat{\rho}_{y,\max}$, $\hat{\rho}_{z,\max}$ are given in Fig. 15. We choose the value of m in such a way that $\hat{\rho}_{x,\max}(i) \leq \varepsilon_0$, $\hat{\rho}_{y,\max}(i) \leq \varepsilon_0$, $\hat{\rho}_{z,\max}(i) \leq \varepsilon_0$, for all $i \geq m$, where ε_0 is a threshold. Here we use $\varepsilon_0 = 0.04$ and obtain $m = 9$. We have $H = M_0 = 0.5$ and assume that $\sigma^2 = 0.2$.

Moreover, due to simulation experiments in Sect. 4, we compute critical values for the change-point statistic $T_W(\cdot)$ and p -values from inequality (15) with $m = 7$.

For the random field of estimated local entropies we have $m = 1$ and the parametric set Θ_0 is constructed with $\Delta_0 = \Delta_1 = 2$, $\gamma_0 = 0.05$, $\gamma_1 = 0.5$ and $L_M = 4$ in (16), which gives $|\Theta_0| = 12,366$. We assume that $\sigma^2 = 0.5$ and $H = \sigma$. The result of our change point analysis is presented in Table 6.

Fig. 15 Values of empirical covariance $\hat{\rho}_{x,max}$ (black color), $\hat{\rho}_{y,max}$ (red color), $\hat{\rho}_{z,max}$ (blue color)

Fig. 16 Anomaly detection in the fibre image (Fig. 13) using local entropy

Our change point test detects the evidence of anomaly regions in real fibre data at significance level $\alpha = 8.4 \times 10^{-10}$.

Similarly to the case of RSA data, the detection of anomalies by the 3σ -rule gives meaningless results. The Spatial SAEM algorithm works much better. Its results are presented in Figs. 16, 17, and 18, where the color labelling of points is the same as for the simulated data.

We conclude that the Spatial SAEM algorithm with attributes entropy, MLD, and a combination of both produces adequate results.

Fig. 17 Anomaly detection in the fibre image (Fig. 13) using mean of local direction

Fig. 18 Anomaly detection in the fibre image (Fig. 13) using local entropy and mean of local direction

Since the images in Figs. 17 and 18 look very similar at first glance, we investigate the results of the Spatial SAEM anomaly detection in more detail. We separate a part of the 3D image (Fig. 14) into 9 layers and present the result of clustering for the 1st and 5th layer for the entropy in Fig. 19, for MLD in Fig. 20, and for the combination of both in Fig. 21.

One can observe that the Spatial SAEM algorithm with local entropy attribute detects vortices of fibers in the material as anomaly regions. This is natural since a vortex exhibits a large diversity of fibre directions, and the entropy is a measure of such diversity. The Spatial SAEM anomaly detection using the mean of local fibre directions identifies the central

Fig. 19 Spatial SAEM clustering according to entropy

Fig. 20 Spatial SAEM clustering according to mean of local directions

Fig. 21 Spatial SAEM clustering according to a combination of entropy and MLD

part of the image (cf. Fig. 20) as an anomaly region, where the directions of fibres differ from the average throughout the image. Finally, the Spatial SAEM approach using both clustering attributes identifies both vortices of fibres and layers of fibres with principally different main direction, cf. Fig. 21.

Acknowledgements Open Access funding provided by Projekt DEAL. We are grateful to Dr. Stefanie Schwaar from Fraunhofer ITWM for valuable discussions and to Jan Niedermeyer for simulating the RSA data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Here we provide the proof to Theorems 1 and 2.

Proof of Theorem 1 Using the Markov inequality we have for any $u > 0$ that

$$\begin{aligned} \mathbf{P}\left(\left|\sum_{k \in W} \xi_k b_k\right| \geq y\right) &= \mathbf{P}\left(\sum_{k \in W} \xi_k b_k \geq y\right) \\ &+ \mathbf{P}\left(\sum_{k \in W} \xi_k (-b_k) \geq y\right) \leq \frac{\mathbb{E} \exp(u \sum_{k \in W} \xi_k b_k)}{\exp(uy)} \\ &+ \frac{\mathbb{E} \exp(u \sum_{k \in W} \xi_k (-b_k))}{\exp(uy)}. \end{aligned} \quad (30)$$

Denote by $W(l) = \{l + mi \in W | i \in \mathbb{Z}^D\}$ for $l \in \{1, \dots, m\}^D$. It follows from Hölder's inequality and m -dependence that

$$\begin{aligned} \mathbb{E} \exp\left(u \sum_{k \in W} \xi_k b_k\right) &= \mathbb{E} \left(\prod_{l \in \{1, \dots, m\}^D} \exp\left(u \sum_{k \in W(l)} \xi_k b_k\right) \right) \\ &\leq \prod_{l \in \{1, \dots, m\}^D} \left(\mathbb{E} \exp\left(m^D u \sum_{k \in W(l)} \xi_k b_k\right) \right)^{1/m^D} \end{aligned}$$

$$\begin{aligned} &= \prod_{l \in \{1, \dots, m\}^D} \left(\prod_{k \in W(l)} \mathbb{E} e^{m^D u \xi_k b_k} \right)^{1/m^D} \\ &= \prod_{k \in W} \left(\mathbb{E} e^{m^D u b_k \xi_k} \right)^{1/m^D}. \end{aligned} \quad (31)$$

From Taylor's expansion we have for $u \in \left[0, \frac{1}{2Hm^D |b_k|^{-1}}\right]$

$$\begin{aligned} \mathbb{E} e^{m^D u b_k \xi_k} &= 1 + m^D u b_k \mathbb{E} \xi_k + \frac{1}{2} m^{2D} u^2 b_k^2 \mathbb{E} \xi_k^2 \\ &+ m^{2D} u^2 b_k^2 \sum_{p \geq 3} \frac{1}{p!} m^{D(p-2)} u^{p-2} b_k^{p-2} \mathbb{E} \xi_k^p \\ &\leq 1 + \frac{1}{2} m^{2D} u^2 b_k^2 \sigma^2 + \frac{1}{2} m^{2D} u^2 b_k^2 \sigma^2 \sum_{p \geq 3} (Hm^D u |b_k|)^{p-2} \\ &= 1 + \frac{1}{2} \frac{m^{2D} u^2 b_k^2 \sigma^2}{1 - Hm^D u |b_k|} \\ &\leq 1 + m^{2D} u^2 b_k^2 \sigma^2 \leq \exp(m^{2D} u^2 b_k^2 \sigma^2). \end{aligned} \quad (32)$$

Combining (31) and (32) we continue for the first term in (30) with the following bound for $0 \leq u \leq (2Hm^D \max_{k \in W} |b_k|)^{-1}$:

$$\begin{aligned} e^{-uy} \prod_{k \in W} \left(\mathbb{E} e^{m^D u b_k \xi_k} \right)^{1/m^D} &\leq \exp\left(-uy + m^D u^2 \sigma^2 \sum_{k \in W} b_k^2\right). \end{aligned} \quad (33)$$

The minimum of (33) is achieved for $u = y/(2m^D \sigma^2 \|b\|_2^2)$. Moreover, bound (33) is valid for the second term in (30), too. Therefore, for $0 \leq y \leq \frac{\sigma^2 \|b\|_2^2}{H \|b\|_\infty}$ we have

$$\mathbf{P}\left(\left|\sum_{k \in W} \xi_k b_k\right| \geq y\right) \leq 2 \exp\left(-\frac{y^2}{4m^D \sigma^2 \|b\|_2^2}\right).$$

For $y > \frac{\sigma^2 \|b\|_2^2}{H \|b\|_\infty}$ we put $u = (2Hm^D \|b\|_\infty)^{-1}$ in (33) and obtain

$$\begin{aligned} \mathbf{P}\left(\left|\sum_{k \in W} \xi_k b_k\right| \geq y\right) &\leq 2 \exp\left(-\frac{y}{2Hm^D \|b\|_\infty} + \frac{1}{4H^2 m^D \|b\|_\infty^2} \sigma^2 \|b\|_2^2\right). \end{aligned}$$

This completes the proof. $\square$

We apply Theorem 1 to $\eta(\theta)$, $\theta \in \Theta_0$, $D = 3$. First, we rewrite $\eta(\theta)$ as

$$\eta(\theta) = \frac{1}{|I_\theta|} \sum_{k \in W} \xi_k \mathbb{I}\{k \in I_\theta\} - \frac{1}{|I_\theta^c|} \sum_{k \in W} \xi_k \mathbb{I}\{k \in I_\theta^c\}$$

$$= \sum_{k \in W} \xi_k b_k(\theta), \theta \in \Theta_0,$$

where

$$b_k(\theta) = \frac{\mathbb{I}\{k \in I_\theta\}}{|I_\theta|} - \frac{\mathbb{I}\{k \in I_\theta^c\}}{|I_\theta^c|}.$$

Corollary 1 Let $|I_\theta| \leq |I_\theta^c|$ for $\theta \in \Theta_0$, then under the conditions of Theorem 1 we have that

$$\begin{aligned} \mathbf{P}(|\eta(\theta)| \geq y) &\leq 2 \exp\left(-\frac{y^2}{4m^3\sigma^2} \frac{|I_\theta^c||I_\theta|}{|W|}\right) \mathbb{I}\left\{0 < y \leq \frac{\sigma^2|W|}{H|I_\theta^c|}\right\} \\ &\quad + 2 \exp\left(-\frac{y}{2Hm^3}|I_\theta| + \frac{\sigma^2|W|}{4H^2m^3|I_\theta^c|}|I_\theta|\right) \\ &\quad \times \mathbb{I}\left\{y > \frac{\sigma^2|W|}{H|I_\theta^c|}\right\}. \end{aligned}$$

Proof From the definition of $b_k(\theta)$ we have

$$\begin{aligned} \|b(\theta)\|_2^2 &= \sum_{k \in W} \left(\frac{\mathbb{I}\{k \in I_\theta\}}{|I_\theta|} - \frac{\mathbb{I}\{k \in I_\theta^c\}}{|I_\theta^c|} \right)^2 \\ &= \frac{1}{|I_\theta|} + \frac{1}{|I_\theta^c|} = \frac{|W|}{|I_\theta^c||I_\theta|}, \end{aligned}$$

and

$$\|b(\theta)\|_\infty = \max_{k \in W} |b_k(\theta)| = \max\left(\frac{1}{|I_\theta|}, \frac{1}{|I_\theta^c|}\right) = \frac{1}{|I_\theta|}.$$

Then the statement of the corollary follows directly from (13). $\square$

Proof of Theorem 2 The statement of the theorem follows immediately from Corollary 1, since

$$\mathbf{P}\left(\max_{\theta \in \Theta_0} |\eta(\theta)| \geq y\right) \leq \sum_{\theta \in \Theta_0} \mathbf{P}(|\eta(\theta)| \geq y). \quad \square$$

Simplifying the result of Theorem 2, we get that (14) is bounded by

$$\begin{aligned} 2 \left| \left\{ \theta \in \Theta_0 : |I_\theta^c| \leq \frac{\sigma^2|W|}{yH} \right\} \right| e^{-\frac{y^2}{4m^3\sigma^2}|W|\gamma_0(1-\gamma_0)} \\ + 2 \left| \left\{ \theta \in \Theta_0 : |I_\theta^c| > \frac{\sigma^2|W|}{yH} \right\} \right| e^{-\frac{y}{4Hm^3}|W|\gamma_0}. \end{aligned}$$

Particularly, if $y < \frac{\sigma^2}{H(1-\gamma_0)}$ then

$$\mathbf{P}\left(\max_{\theta \in \Theta_0} |\eta(\theta)| \geq y\right) \leq 2|\Theta_0| e^{-\frac{y^2}{4m^3\sigma^2}|W|\gamma_0(1-\gamma_0)}, \quad (34)$$

Table 7 Critical values $y_{0.05}$ computed from (15)

$m \downarrow, \sigma^2 \rightarrow$	0.4	0.6	0.8	1	1.2	1.4	1.6	2	3	4	5
1	0.0246	0.0301	0.0347	0.0389	0.0426	0.0460	0.0492	0.0550	0.0673	0.0778	0.0869
2	0.0695	0.0852	0.0983	0.1099	0.1204	0.1301	0.1390	0.1554	0.1904	0.2198	0.2458
3	0.1277	0.1564	0.1806	0.2019	0.2212	0.2389	0.2554	0.2856	0.3498	0.4039	0.4515
4	0.1967	0.2408	0.2781	0.3109	0.3406	0.3679	0.3933	0.4397	0.5385	0.6218	0.6952
5	0.2748	0.3365	0.3886	0.4345	0.4759	0.5141	0.5496	0.6144	0.7526	0.8689	0.9715
6	0.3612	0.4424	0.5108	0.5711	0.6256	0.6758	0.7225	0.8077	0.9893	1.1423	1.2771
7	0.4552	0.5575	0.6437	0.7197	0.7884	0.8516	0.9104	1.0178	1.2466	1.4395	1.6094
8	0.5561	0.6811	0.7865	0.8793	0.9633	1.0404	1.1123	1.2435	1.5231	1.7587	1.9663
9	0.6636	0.8128	0.9385	1.0493	1.1494	1.2415	1.3272	1.4839	1.8174	2.0985	2.3462
10	0.6803	0.8332	0.9621	1.0757	1.1784	1.2727	1.3606	1.5212	1.8631	2.1513	2.4052
11	0.6979	0.8547	0.9870	1.1035	1.2088	1.3056	1.3958	1.5606	1.9113	2.2069	2.4674
12	0.7796	0.9548	1.1025	1.2326	1.3503	1.4584	1.5591	1.7432	2.1350	2.4652	2.7562
13	0.9911	1.2138	1.4016	1.5671	1.7166	1.8542	1.9822	2.2162	2.7142	3.1341	3.5041
14	1.2379	1.5161	1.7506	1.9572	2.1440	2.3158	2.4757	2.7679	3.3900	3.9145	4.3765
15	1.5225	1.8647	2.1531	2.4073	2.6370	2.8484	3.0450	3.4044	4.1696	4.8146	5.3829
16	1.8478	2.2630	2.6131	2.9216	3.2004	3.4569	3.6955	4.1317	5.0603	5.8431	6.5328
17	2.2163	2.7144	3.1343	3.5043	3.8388	4.1463	4.4326	4.9558	6.0696	7.0086	7.8358
18	2.6309	3.2222	3.7206	4.1598	4.5568	4.9219	5.2618	5.8829	7.2050	8.3196	9.3016

Extended version of Table 3

and if $y > \frac{\sigma^2}{H(1-\gamma_1)}$ then

$$\mathbf{P} \left(\max_{\theta \in \Theta_0} |\eta(\theta)| \geq y \right) \leq 2 |\Theta_0| e^{-\frac{y}{4Hm^3} \gamma_0 |W|}. \quad (35)$$

References

- Alonso-Ruiz, P., Spodarev, E.: Estimation of entropy for Poisson marked point processes. *Adv. Appl. Probab.* **49**(1), 258–278 (2017)
- Alonso-Ruiz, P., Spodarev, E.: Entropy-based inhomogeneity detection in fiber materials. *Methodol. Comput. Appl. Probab.* **20**(4), 1223–1239 (2018)
- Andrä, H., Gurka, M., Kabel, M., Nissle, S., Redenbach, C., Schladitz, K., Wirjadi, O.: Geometric and mechanical modeling of fiber-reinforced composites. In: *Proceedings of the 2nd International Congress on 3D Materials Science*, pp. 35–40. Springer (2014)
- Basseville, M., Nikiforov, I.: *Detection of Abrupt Changes: Theory and Application*. Prentice Hall Information and System Sciences Series. Prentice Hall Inc, Englewood Cliffs, NJ (1993)
- Beirlant, J., Dudewicz, E.J., Györfi, L., van der Meulen, E.C.: Non-parametric entropy estimation: an overview. *Int. J. Math. Stat. Sci.* **6**(1), 17–39 (1997)
- Brodsky, B.E., Darkhovsky, B.S.: *Nonparametric Methods in Change-point Problems*. Mathematics and its Applications, vol. 243. Kluwer Academic Publishers Group, Dordrecht (1993)
- Brodsky, B.E., Darkhovsky, B.S.: Problems and methods of probabilistic diagnostics. *Avtomat. i Telemekh.* **60**(8), 3–50 (1999)
- Brodsky, B.E., Darkhovsky, B.S.: *Non-parametric Statistical Diagnosis. Mathematics and Its Applications*, vol. 509. Kluwer Academic Publishers, Dordrecht (2000). Problems and methods
- Bucchia, B.: Testing for epidemic changes in the mean of a multiparameter stochastic process. *J. Stat. Plan. Inference* **150**, 124–141 (2014)
- Bucchia, B., Heuser, C.: Long-run variance estimation for spatial data under change-point alternatives. *J. Stat. Plan. Inference* **165**, 104–126 (2015)
- Bucchia, B., Wendler, M.: Change-point detection and bootstrap for Hilbert space valued random fields. *J. Multivar. Anal.* **155**, 344–368 (2017)
- Bulinski, A., Dimitrov, D.: Statistical estimation of the Shannon entropy. *Acta. Math. Sin. Engl. Ser.* **35**(1), 17–46 (2019)
- Cao, J., Worsley, K.J.: The detection of local shape changes via the geometry of Hotelling's T^2 fields. *Ann. Stat.* **27**(3), 925–942 (1999)
- Carlstein, E., Müller, H.G., Siegmund, D. (eds.): *Change-Point Problems*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 23. Institute of Mathematical Statistics, Hayward, CA (1994). Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College, South Hadley, MA, July 11–16 (1992)
- Celeux, G., Diebolt, J.: A stochastic approximation type em algorithm for the mixture problem. *Stoch. Stoch. Rep.* **41**(1–2), 119–134 (1992)
- Chambaz, A.: Detecting abrupt changes in random fields. *ESAIM Probab. Stat.* **6**, 189–209 (2002). New directions in time series analysis (Luminy, 2001)
- Chen, J., Gupta, A.K.: *Parametric Statistical Change Point Analysis*, 2nd edn. Birkhäuser/Springer, New York (2012). With applications to genetics, medicine, and finance
- Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J.: *Stochastic Geometry and Its Applications*. Wiley Series in Probability and Statistics, 3rd edn. Wiley, Chichester (2013)
- Csörgö, M., Horváth, L.: *Limit Theorems in Change-Point Analysis*. Wiley Series in Probability and Statistics. Wiley, Chichester (1997)
- Dobrushin, R.L.: A simplified method of experimentally evaluating the entropy of a stationary sequence. *Theory Probab. Its Appl.* **3**(4), 428–430 (1958)
- Dresvyanskiy, D., Karaseva, T., Mitrofanov, S., Redenbach, C., Schwaar, S., Makogin, V., Spodarev, E.: Application of clustering methods to anomaly detection in fibrous media. *IOP Conference Series: Materials Science and Engineering* **537**(2), 022001 (2019)
- Efimov, K., Adamyan, L., Spokoyny, V.: Adaptive nonparametric clustering. *IEEE Trans. Inf. Theory* **65**(8), 4875–4892 (2019)
- Everitt, B.S., Landau, S., Leese, M., Stahl, D.: *Cluster Analysis*. Wiley Series in Probability and Statistics, 5th edn. Wiley, Chichester (2011)
- Falconer, K.: *Fractal Geometry. Mathematical Foundations and Applications*, 2nd edn. Wiley, Hoboken, NJ (2003)
- Franke, J., Redenbach, C., Zhang, N.: On a mixture model for directional data on the sphere. *Scand. J. Stat.* **43**(1), 139–155 (2016)
- Fraunhofer, I.T.W.M.: Department of Image Processing: MAVI—modular algorithms for volume images. <http://www.mavi-3d.de> (2005)
- Gorshenin, A.K., Korolev, V.Y., Tursunbaev, A.M.: Median modifications of the EM-algorithm for separation of mixtures of probability distributions and their applications to the decomposition of volatility of financial indexes. *J. Math. Sci. (N.Y.)* **227**(2), 176–195 (2017)
- Hahubia, T., Mnatsakanov, R.: On the mode-change problem for random measures. *Georgian Math. J.* **3**(4), 343–362 (1996)
- Heinrich, L.: Some bounds of cumulants of m -dependent random fields. *Math. Nachr.* **149**, 303–317 (1990)
- Hennig, C., Meila, M., Murtagh, F., Rocci, R. (eds.): *Handbook of Cluster Analysis. Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL (2016)
- Jarušková, D., Piterbarg, V.I.: Log-likelihood ratio test for detecting transient change. *Stat. Probab. Lett.* **81**(5), 552–559 (2011)
- Kaplan, E.I.: On the change-point problem for random fields. *Teor. Veroyatnost. i Primenen.* **35**(2), 353–358 (1990)
- Kaplan, E.I.: Convergence of estimates for partitions in the change point problem for random fields. *Teor. Imovir. ta Mat. Statist.* **47**, 34–39 (1992)
- Korolev, V.Y.: *EM-Algorithm, Its Modifications and Their Use in the Problem of Decomposing the Mixtures of Probability Distributions*. IPIRAS, Moscow (2007)
- Kozachenko, L.F., Leonenko, N.N.: A statistical estimate for the entropy of a random vector. *Problemy Peredachi Informatsii* **23**(2), 9–16 (1987)
- Lai, T.L.: Saddlepoint approximations and boundary crossing probabilities for random fields and their applications. In: *Third International Congress of Chinese Mathematicians. Part 1, 2*, AMS/IP Studies in Advanced Mathematics, 42, pt. 1, vol. 2, pp. 29–39. American Mathematical Society, Providence, RI (2008)
- Laurent, B., Marteau, C., Maugis-Rabusseau, C.: Multidimensional two-component Gaussian mixtures detection. *Ann. Inst. Henri Poincaré Probab. Stat.* **54**(2), 842–865 (2018)
- Miller, W.D.: Quasi-Heyting algebras: a new class of lattices, and a foundation for nonclassical model theory with possible computational applications. ProQuest LLC, Ann Arbor, MI (1993). Thesis (Ph.D.), Kansas State University
- Müller, H.G., Song, K.S.: Cube splitting in multidimensional edge estimation. In: *Change-Point Problems* (South Hadley, MA, 1992), IMS Lecture Notes Monograph Series, vol. 23, pp. 210–223. Institute of Mathematical Statistics, Hayward, CA (1994)
- Ninomiya, Y.: Construction of conservative test for change-point problem in two-dimensional random fields. *J. Multivar. Anal.* **89**(2), 219–242 (2004)

- Page, E.S.: Continuous inspection schemes. *Biometrika* **41**, 100–115 (1954)
- Penrose, M.D., Yukich, J.E.: Limit theory for point processes in manifolds. *Ann. Appl. Probab.* **23**(6), 2161–2211 (2013)
- Redenbach, C., Vecchio, I.: Statistical analysis and stochastic modelling of fibre composites. *Compos. Sci. Technol.* **71**, 107–112 (2011)
- Rogers, C.A.: *Hausdorff Measures*. Cambridge University Press, Cambridge (1998)
- Schwaar, S.: Asymptotics for change-point tests and change-point estimators. Doctoral thesis, Technische Universität Kaiserslautern (2017). <http://nbn-resolving.de/urn:nbn:de:hbz:386-kluedo-45991>
- Sen, A., Srivastava, M.S.: On tests for detecting change in mean. *Ann. Stat.* **3**, 98–108 (1975)
- Sharipov, O., Tewes, J., Wendler, M.: Sequential block bootstrap in a Hilbert space with application to change point analysis. *Can. J. Stat.* **44**(3), 300–322 (2016)
- Siegmund, D., Yakir, B.: Detecting the emergence of a signal in a noisy image. *Stat. Interface* **1**(1), 3–12 (2008)
- Siegmund, D.O., Worsley, K.J.: Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Ann. Stat.* **23**(2), 608–639 (1995)
- Wierzchoń, S.T., Kłopotek, M.A.: *Modern Algorithms of Cluster Analysis*. Studies in Big Data, vol. 34. Springer, Cham (2018)
- Wirjadi, O., Godehardt, M., Schladitz, K., Wagner, B., Rack, A., Gurka, M., Nissle, S., Noll, A.: Characterization of multilayer structures of fiber reinforced polymer employing synchrotron and laboratory X-ray CT. *Int. J. Mater. Res.* **105**(7), 645–654 (2014)
- Wirjadi, O., Schladitz, K., Easwaran, P., Ohser, J.: Estimating fibre direction distributions of reinforced composites from tomographic images. *Image Anal. Stereol.* **35**(3), 167–179 (2016)
- Wu, Y.: Inference for Change-Point and Post-Change Means After a CUSUM test. *Lecture Notes in Statistics*, vol. 180. Springer, New York (2005)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.