ulm university universität ulm

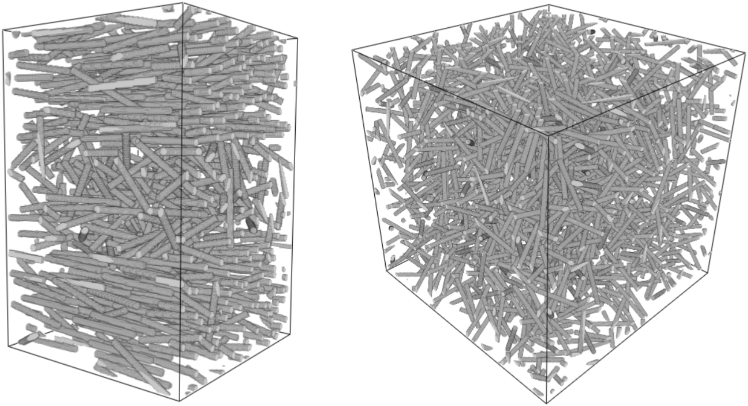
Change-point methods for anomaly detection in fibrous media

joint work with E. Spodarev, C. Redenbach, D. Dresvyanskiy, T. Karaseva and S. Mitrofanov

Vitalii Makogin | 10.10.2019 | Institute of Stochastics, Ulm University

Fibre materials

Random sequential adsorption (RSA)



Visualization of layered RSA image with 200 × 200 × 300 voxels (left) and homogeneous RSA image with 500 × 500 × 500 voxels (right), radius 4

Problem setting

- A fibre γ is a simple curve $\{\gamma(t), t \in [0,1]\}$ in $\mathbb{R}^3$ of finite length.
- The collection of fibres forms a fibre system ϕ.
- The length measure $\phi(B) = \sum_{\gamma \in \phi} h(\gamma \cap B)$, where h is the length of fibre in window $B \subset \mathbb{R}^3$.
- A fibre process Φ is a random element with values in the set $\mathbb{D}$ of all fibre systems ϕ with σ-algebra D generated by sets of the form $\{\phi \in \mathbb{D} : \phi(B) < x\}$.

Classification

- Let $w(x)$ be some characteristic of a fibre at point x : fibre local direction, curvature, etc.
- A weighted random measure $\Psi(B \times L) = \int_B \mathbb{1}\{w(x) \in L\}\Phi(dx)$.
- If the fibre process Φ is stationary, then $\mathbf{E}\Psi(B \times L) = \lambda |B| f(L)$, where λ is called the intensity of Ψ,
- A probability measure f on $\mathbb{S}^2$ is called the directional distribution of fibres.

Testing

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

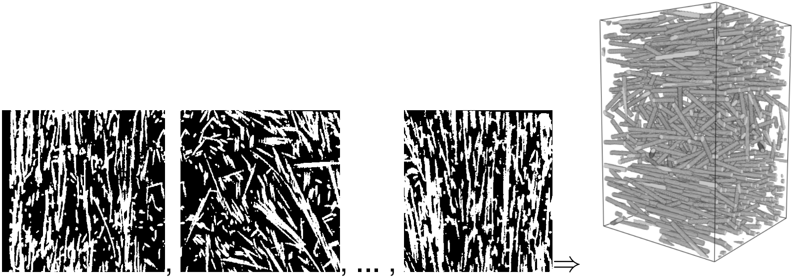
H_1 : There exists a compact set $A \subset W$ with $|A| > 0$ and $|W \setminus A| > 0$ such that

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

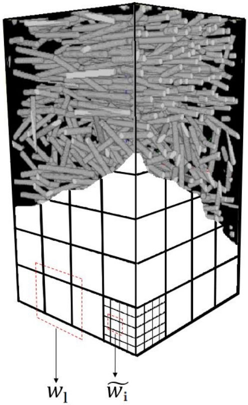
If H_1 holds true, the region A is called an anomaly region.

Data

- The dilated fibre system $\Phi \oplus B_r \cap W$ in window $W \subset \mathbb{R}^3$ is observed as a 3D greyscale image.
- Reconstruction of μCT image by MAVI: Modular Algorithms for volume Images, provided by Fraunhofer ITWM.



Local directions and clustering criteria



- A separation of fibres (and estimation of their directions) requires large computational resources for 3D images.
- An estimation of local direction is much faster but produces dependent sample.
- In each $\widetilde{W}_l$, the “average local direction” is computed using SubfieldFibreDirection in MAVI.
- We group $\widetilde{W}_l$ in classification windows W_l.
- For each W_l we assign a classification attribute: entropy, average direction.

Entropy estimation

- Entropy of an absolutely continuous $\mathbb{S}^2-$valued random variable with density f is $E_X = -\int_{\mathbb{S}^2} \log(f(x))f(x)\sigma(dx)$, where σ is the spherical surface measure on $\mathbb{S}^2$.
- Plug-in estimators required large samples. In simulations for uniform distribution on a sphere $N > 50^3$.
- Nearest neighbour estimator (KL-estimator, 1987)

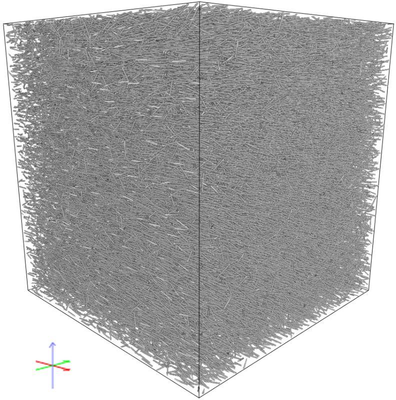
$$\hat{E} = \frac{d}{N}\sum_{i=1}^{N} \log \rho_i + \log(c(N-1)) + \gamma.$$

- We propose modification of the nearest neighbour estimator:

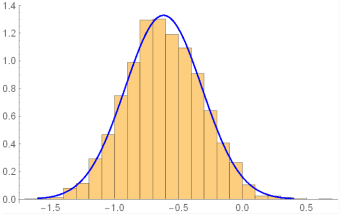
$$\hat{E}_M = \frac{d}{N_M}\sum_{i=1}^{N} \mathbb{1}\{\rho_i > \rho_0\} \log \rho_i + \log(c(N_M-1)) + \gamma,$$

where $N_M = \sum_{i=1}^{N} \mathbb{1}\{\rho_i > \rho_0\}$, ρ_0 is a penalty value.

Entropy estimation:
homogeneous RSA (random sequential adsorption) image

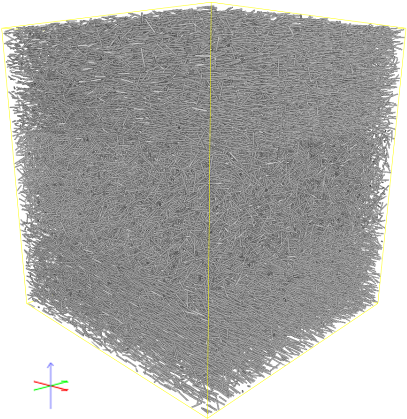


$2000 \times 2000 \times 2100$ voxels

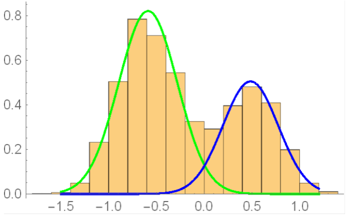


Histogram of frequencies of the local entropy

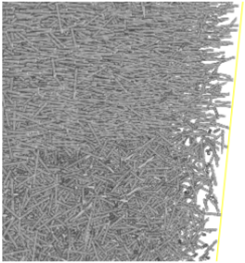
Entropy estimation: layered RSA image



$2000 \times 2000 \times 2100$ voxels



Histogram of frequencies of the local entropy



Random fields with inhomogeneities in mean

- Let be $\{\xi_k, k \in \mathbb{Z}^3\}$ an integrable, centered, stationary, real-valued random field.
- $\{\xi_k, k \in \mathbb{Z}^3\}$ is $m-$dependent, and there exist $H, \sigma > 0$ such that $\mathbf{E}|\xi_k|^p \leq \frac{p!}{2} H^{p-2}\sigma^2, p = 2, 3, \ldots$
- Let Θ be a finite parametric space. For every $\theta \in \Theta$ we define subspace of anomalies $I_\theta \subset \mathbb{Z}^3$.
- For some $\theta_0 \in \Theta$ we observe
$s_k = \xi_k + \mu + h\mathbb{1}\{k \in I_{\theta_0}\}, k \in W$.
- Let Θ_0 correspond to the significant anomalies, i.e, for $\gamma_0, \gamma_1 \in (0, 1)$ 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 the extremely small or large anomalies, i.e,
$\Theta_1 = \{\theta \in \Theta : |I_\theta| < \gamma_0|W|, \text{ or } |I_\theta| > \gamma_1|W|\}$.

Testing the change of expectation

- The change-point hypotheses for the random field $\{s_k, k \in W\}$ with respect to its expectation

H_0 : $\mathbf{E}s_k = \mu$ for every $k \in W$ (i.e. $h = 0$) vs.

H_1 : There exists $\theta_0 \in \Theta_0$ such that
$\mathbf{E}s_k = \mu + h, k \in I_{\theta_0},\ h \neq 0$, and $\mathbf{E}s_k = \mu, k \in I_{\theta_0^c}$.

- Analogue of CUSUM statistics

$$Z(\theta) = \frac{1}{|I_\theta|} \sum_{k \in I_\theta} s_k - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} s_k$$

$$= \frac{1}{|I_\theta|} \sum_{k \in I_\theta} \xi_k - \frac{1}{|I_\theta^c|} \sum_{k \in I_\theta^c} \xi_k + h \left(\frac{|I_\theta \cap I_{\theta_0}|}{|I_\theta|} - \frac{|I_\theta^c \cap I_{\theta_0}|}{|I_\theta^c|} \right).$$

- Test statistics: $T_W = \max_{\theta \in \Theta_0} |Z(\theta)|$.

Test statistics

- Critical values y_α via the probability of the 1st-type error:

$$\mathbf{P}_{H_0}(T_W \geq y_\alpha) = \mathbf{P}\left(\max_{\theta\in\Theta_0}\left|\frac{1}{|I_\theta|}\sum_{k\in I_\theta}\xi_k - \frac{1}{|I_\theta^c|}\sum_{k\in I_\theta^c}\xi_k\right| \geq y_\alpha\right) \leq \alpha.$$

- Tail probabilities

$$\mathbf{P}_{H_0}(T_W \geq y) \leq \sum_{\theta\in\Theta_0:|I_\theta^c|\leq\frac{\sigma^2|W|}{yH}} 2\exp\left(-\frac{y^2}{4m^3\sigma^2}\frac{|I_\theta^c||I_\theta|}{|W|}\right)$$

$$+ \sum_{\theta\in\Theta_0:|I_\theta^c|>\frac{\sigma^2|W|}{yH}} 2\exp\left(-\frac{y}{2Hm^3}|I_\theta| + \frac{\sigma^2|W|}{4H^2m^3|I_\theta^c|}|I_\theta|\right).$$

Tail probabilities

- If ξ_k's are Gaussian, then $H = \sigma$. If $|\xi_k| \leq M$ then $H = M$.
- Particularly, if $y < \frac{\sigma^2}{H(1-\gamma_0)}$ then

$$\mathbf{P}_{H_0}(T_W \geq y) \leq 2\,|\Theta_0|\exp\left(-\frac{y^2}{4m^3\sigma^2}|W|\gamma_0(1-\gamma_0)\right),$$

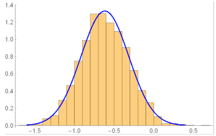
and if $y > \frac{\sigma^2}{H(1-\gamma_1)}$ then

$$\mathbf{P}_{H_0}(T_W \geq y) \leq 2\,|\Theta_0|\exp\left(-\frac{y}{4Hm^3}\gamma_0|W|\right).$$

Simulated data

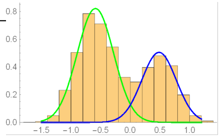
Homogeneous RSA data:

Attr.	$\lvert\Theta_0\rvert$	Var.	Test stat.	$p-$value
$\tilde{x}$	39395	0.04360	0.0344	1.00
$\tilde{y}$	39395	0.03743	0.0130	1.00
$\tilde{z}$	39395	0.03749	0.0146	1.00
$\widetilde{E}$	16536	0.08984	0.0942	1.00

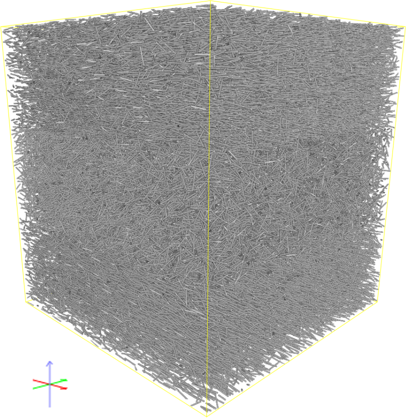


Layered RSA data:

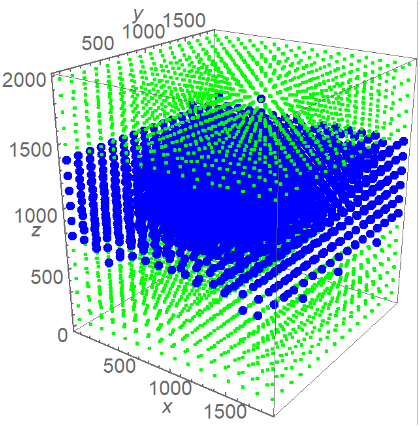
Attr.	$\lvert\Theta_0\rvert$	Var.	Test stat.	$p-$value
$\tilde{x}$	39395	0.10592	0.44036	4.6×10^{-30}
$\tilde{y}$	39395	0.10948	0.43163	2.8×10^{-23}
$\tilde{z}$	39395	0.06151	0.18764	0.301
$\widetilde{E}$	16536	0.3583	1.07030	0.00



Classification by spatial SAEM algorithm: layered RSA image.



2000 × 2000 × 2100 voxels



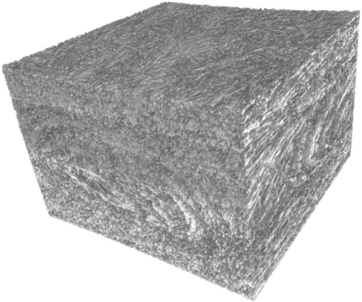
Spatial SAEM classification by entropy and mean local directions

Real glass fibre reinforced polymer

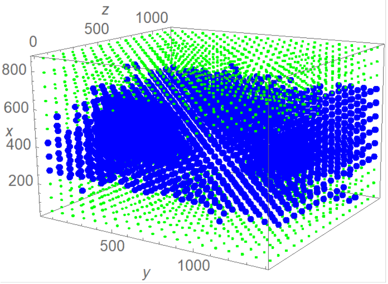
- The images are provided by the Institute for Composite Materials (IVW) in Kaiserslautern: $970 \times 1469 \times 1217$ voxels, the estimated radius of 3 voxels. We obtain $64 \times 97 \times 80$ small windows $\widetilde{W}_i$ with $15 \times 15 \times 15$ voxels.
- Change point testing:

Attr.	H	σ^2	m	$\lvert\Theta_0\rvert$	Var.	Test stat.	$p-$value
$\tilde{x}$	0.5	0.2	7	33004	0.04589	0.15995	1.00
$\tilde{y}$	0.5	0.2	7	33004	0.06795	0.44733	2.1×10^{-10}
$\tilde{z}$	0.5	0.2	7	33004	0.07982	0.43383	1.3×10^{-6}
E	0.7071	0.5	1	12366	0.30126	0.46811	3.96×10^{-8}

Classification by spatial SAEM algorithm: real data.



970 × 1469 × 1217 voxels



Spatial SAEM classification by entropy and mean local directions

Preprint: *Detecting anomalies in fibre systems using 3-dimensional image data. arXiv:1907.06988.*

Support:

- German Ministry of Education and Research (BMBF): the project "AniS"
- DFG Research Training Group GRK 1932;
- DFG grant SP 971/10-1;
- DAAD scientific exchange program "Strategic Partnerships".

Thank you for your attention!