Change and Anomaly Detection with Sparse Representations

Giacomo Boracchi

Dipartimento di Elettronica, Informazione,

e Bioingegneria, Politecnico di Milano

giacomo.boracchi@polimi.it

Tampere University of Technology, Finland

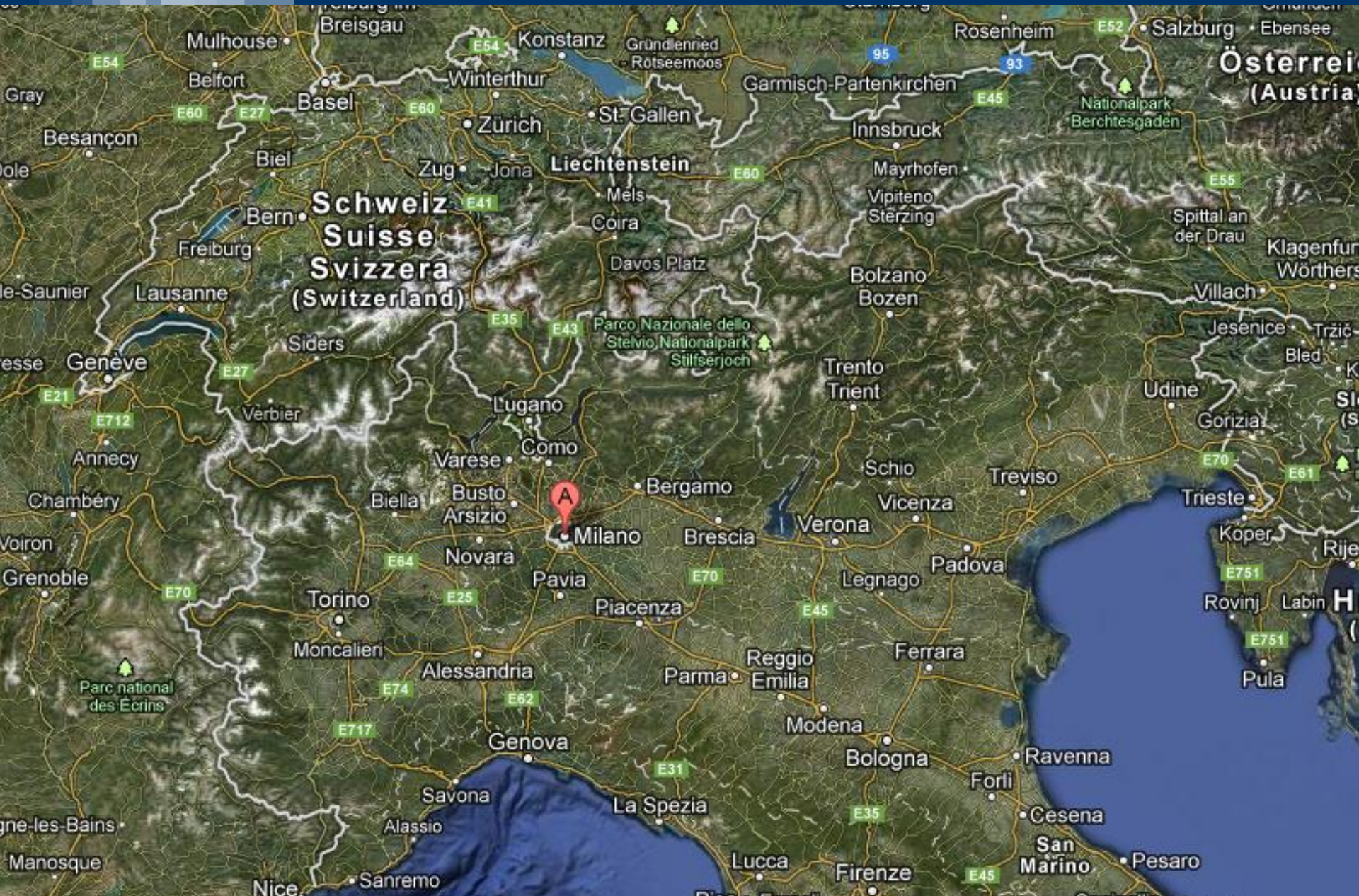
2° September 2014

AN ONGOING WORK WITH

Cesare Alippi, Diego Carrera (Polimi)

Brendt Wohlberg (Los Alamos National Laboratory)

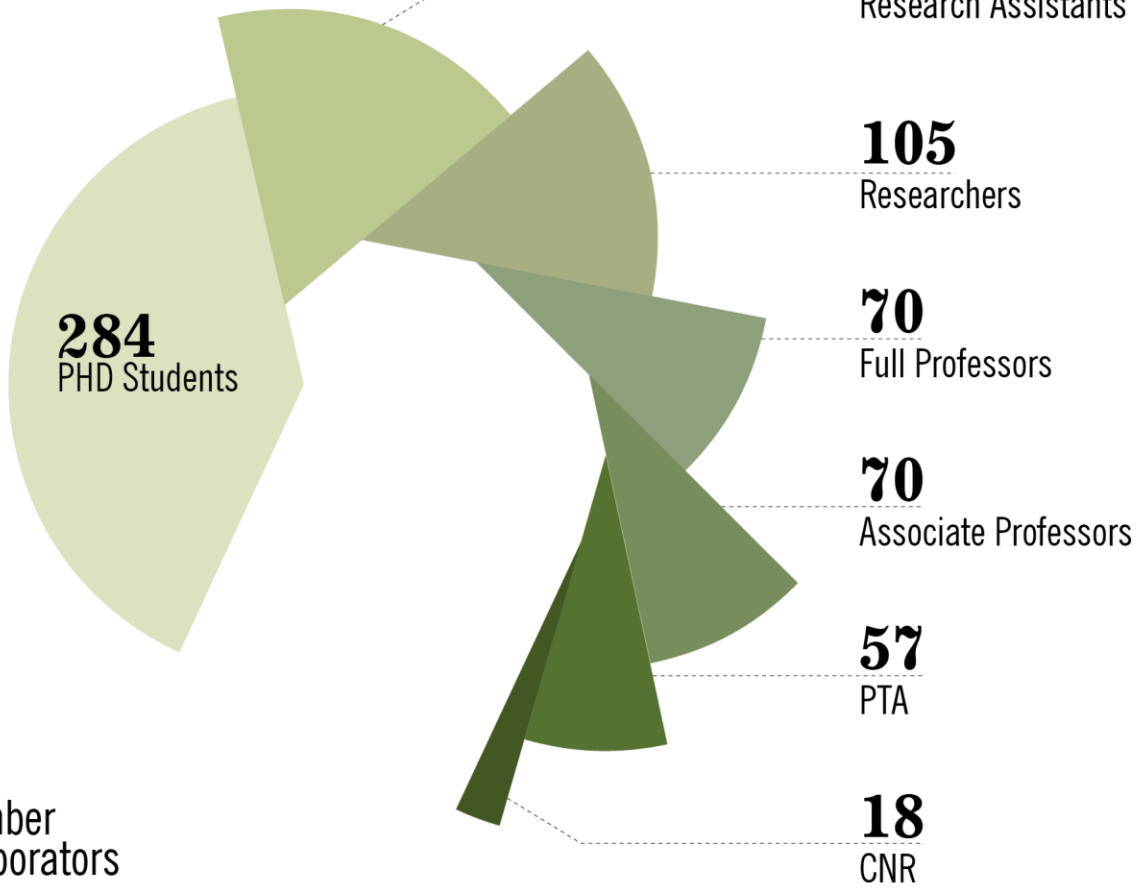
Politecnico di Milano



Dipartimento di Elettronica, Informazione e Bioingegneria

DEIB People

DEIB employs about 840 people organised as follows.



745
plus a large number of external collaborators

Outline

- The Problem Formulation
- Sparse Representations for Change/Anomaly Detection
  - Brief overview of Sparse Representations
  - Design of Change Indicators
- Experiments
  - Anomaly detection in SEM images
  - Change detection in streams of acoustic-emissinos
- Ongoing Works

PROBLEM FORMULATION

Data-Generating Process

- We assume that in **normal (stationary)** conditions, we observe data $\mathbf{s} \in \mathbb{R}^m$ drawn from a stochastic process $\mathcal{P}_N$

$$\mathbf{s} \sim \mathcal{P}_N$$

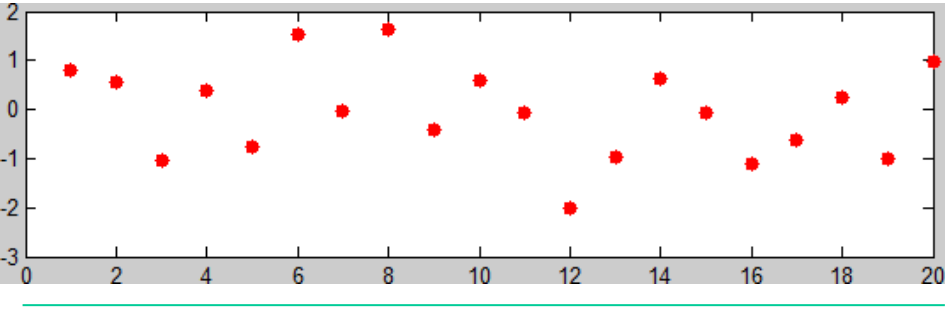
- We do not know the process, we only assume that data are i.i.d. realizations from $\mathcal{P}_N$.

Data-Generating Process

- We assume that in **normal (stationary)** conditions, we observe data $\mathbf{s} \in \mathbb{R}^m$ drawn from a stochastic process $\mathcal{P}_N$

$$\mathbf{s} \sim \mathcal{P}_N$$

- We do not know the process, we only assume that data are i.i.d. realizations from $\mathcal{P}_N$.

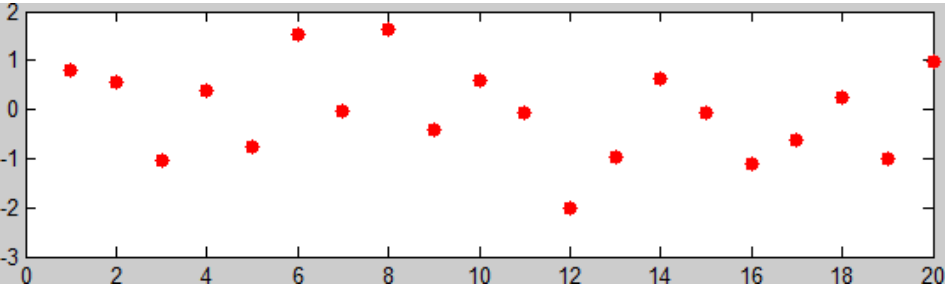


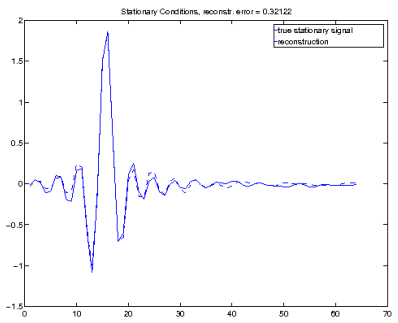
Data-Generating Process

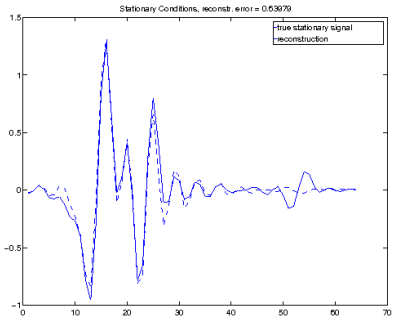
- We assume that in **normal (stationary)** conditions, we observe data $\mathbf{s} \in \mathbb{R}^m$ drawn from a stochastic process $\mathcal{P}_N$

$$\mathbf{s} \sim \mathcal{P}_N$$

- We do not know the process, we only assume that data are i.i.d. realizations from $\mathcal{P}_N$.





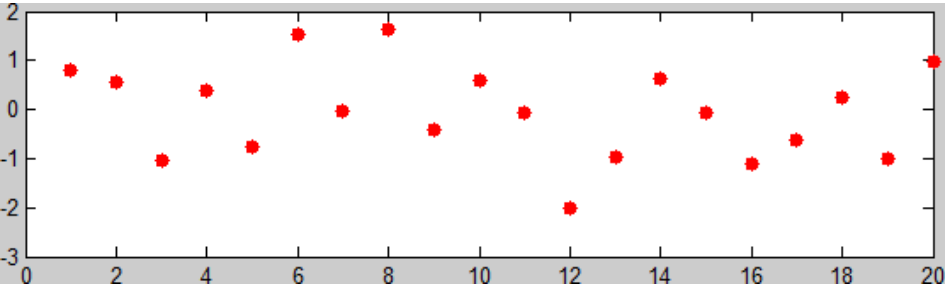


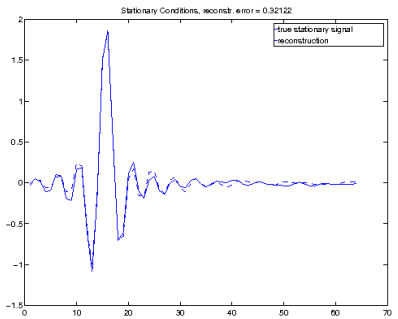
Data-Generating Process

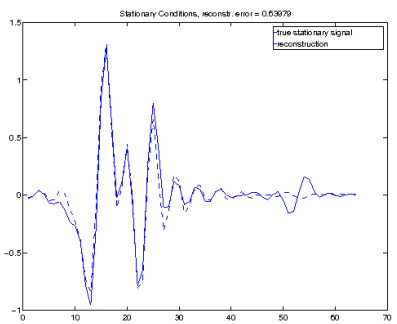
- We assume that in **normal (stationary)** conditions, we observe data $\mathbf{s} \in \mathbb{R}^m$ drawn from a stochastic process $\mathcal{P}_N$

$$\mathbf{s} \sim \mathcal{P}_N$$

- We do not know the process, we only assume that data are i.i.d. realizations from $\mathcal{P}_N$.







The change-detection problem

- The **change-detection problem** consists in monitoring a sequence of data (datastream), vectors of $\mathbb{R}^m$

$$\{\mathbf{s}_t\}_{t=1,\dots}$$

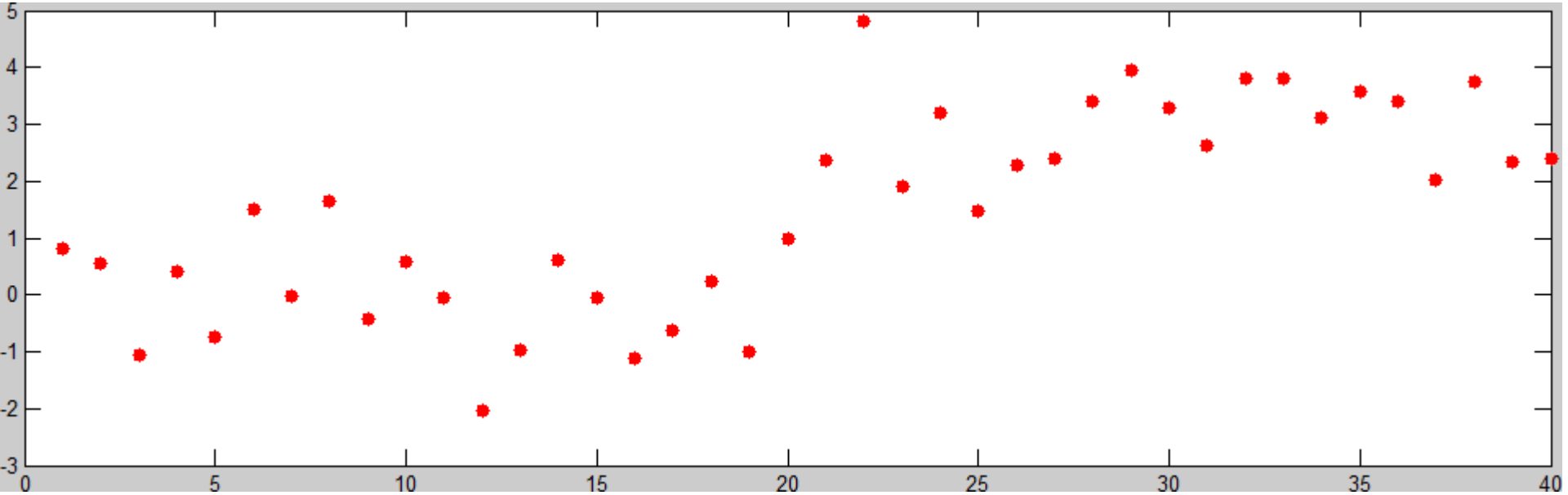
and determining when the data-generating process changes.

$$\mathbf{s}_t = \begin{cases} \mathbf{s}_t \sim \mathcal{P}_N & t < T^* \\ \mathbf{s}_t \sim \mathcal{P}_A & t \geq T^* \end{cases}$$

- **Unpredictability** of the change, $\mathcal{P}_A$ is unknown and sometimes also $\mathcal{P}_N$ is unknown.
- T^* is denoted the **change point**

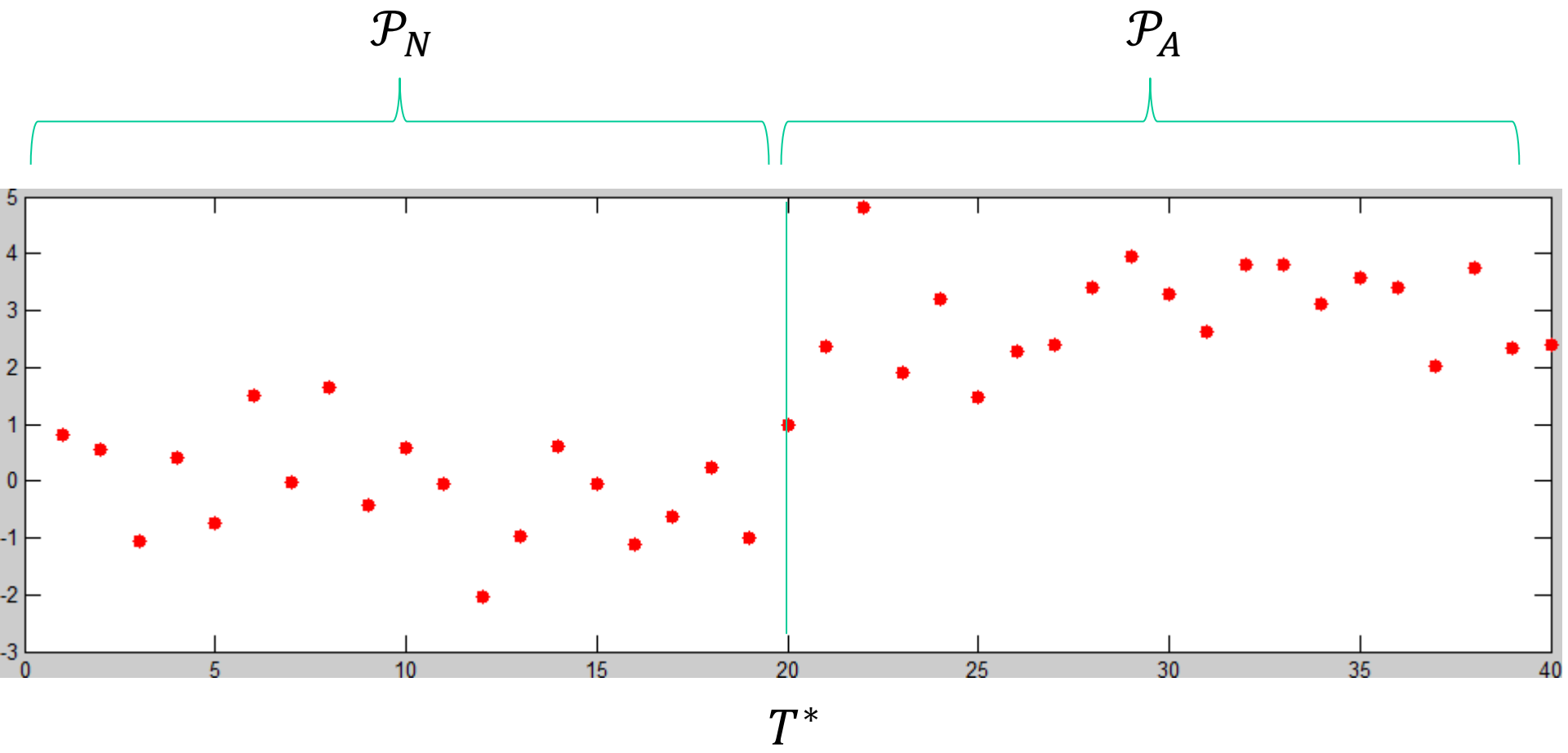
The change-detection problem

- There is a **temporal dimension** and we want do detect permanent shifts of the process



The change-detection problem

- There is a **temporal dimension** and we want do detect permanent shifts of the process



The Anomaly-Detection Problem

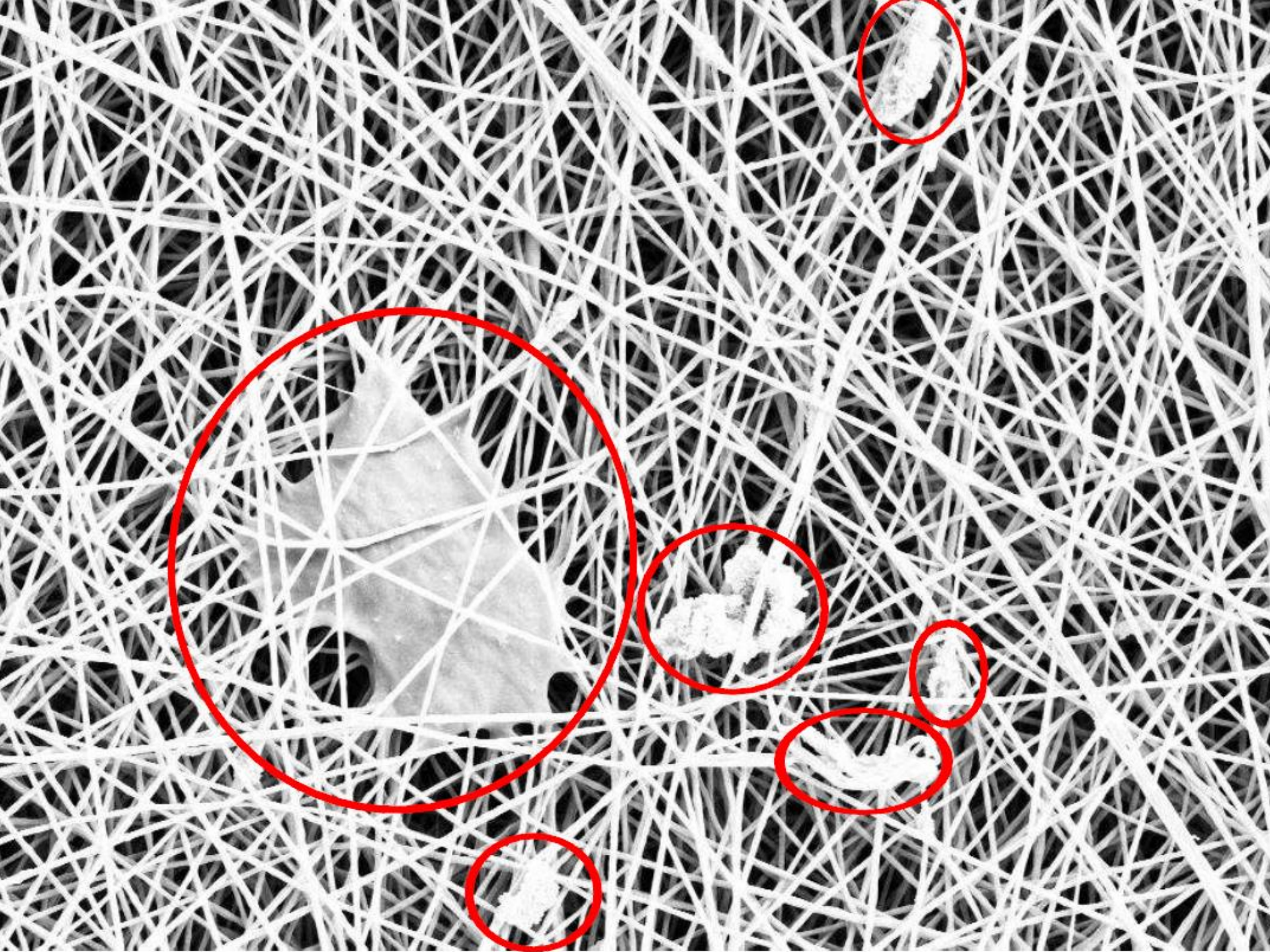
- In the anomaly-detection problem we assume a set of data (vectors of $\mathbb{R}^m$)

$$\{\mathbf{s}_i\}_{i=1,\ldots l}$$

and we want to detect **data that do not conform** to the **expected behavior** i.e., that are not likely to have been generated by $\mathcal{P}_N$.

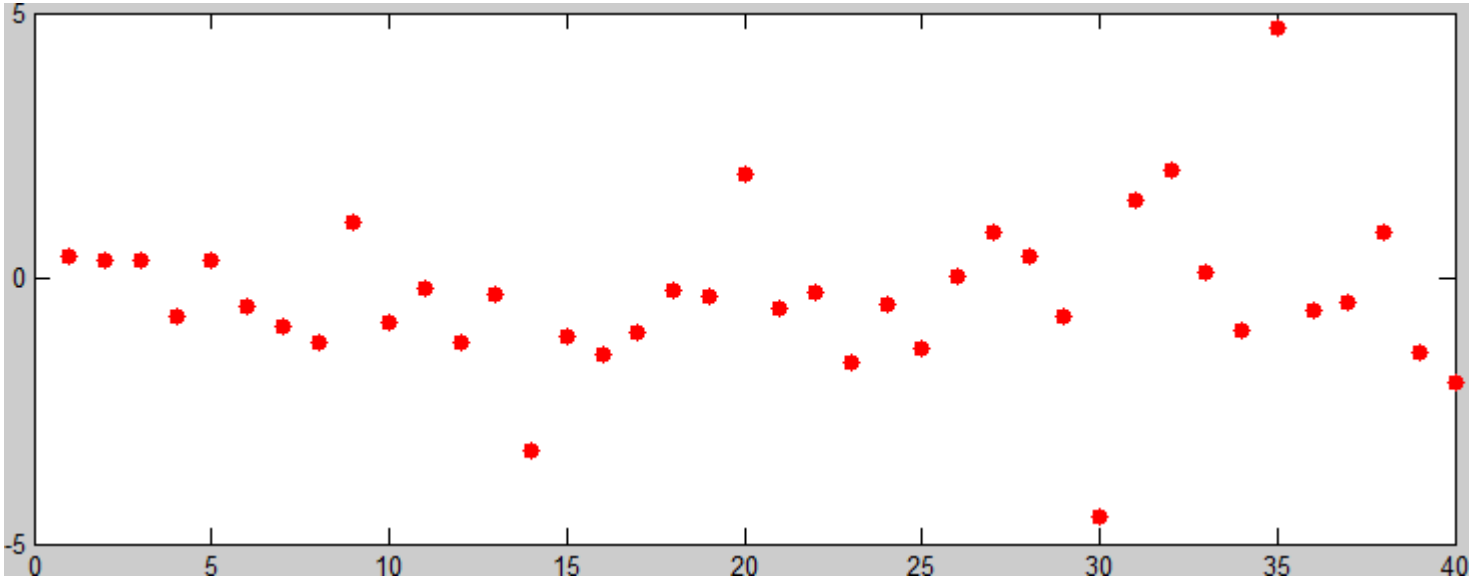
- Anomalies are also referred to **outliers**. "*An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism*" [Hawkins 1980]:

[Hawkins 1980] Hawkins, D. Identification of Outliers. Chapman and Hall, 1980.



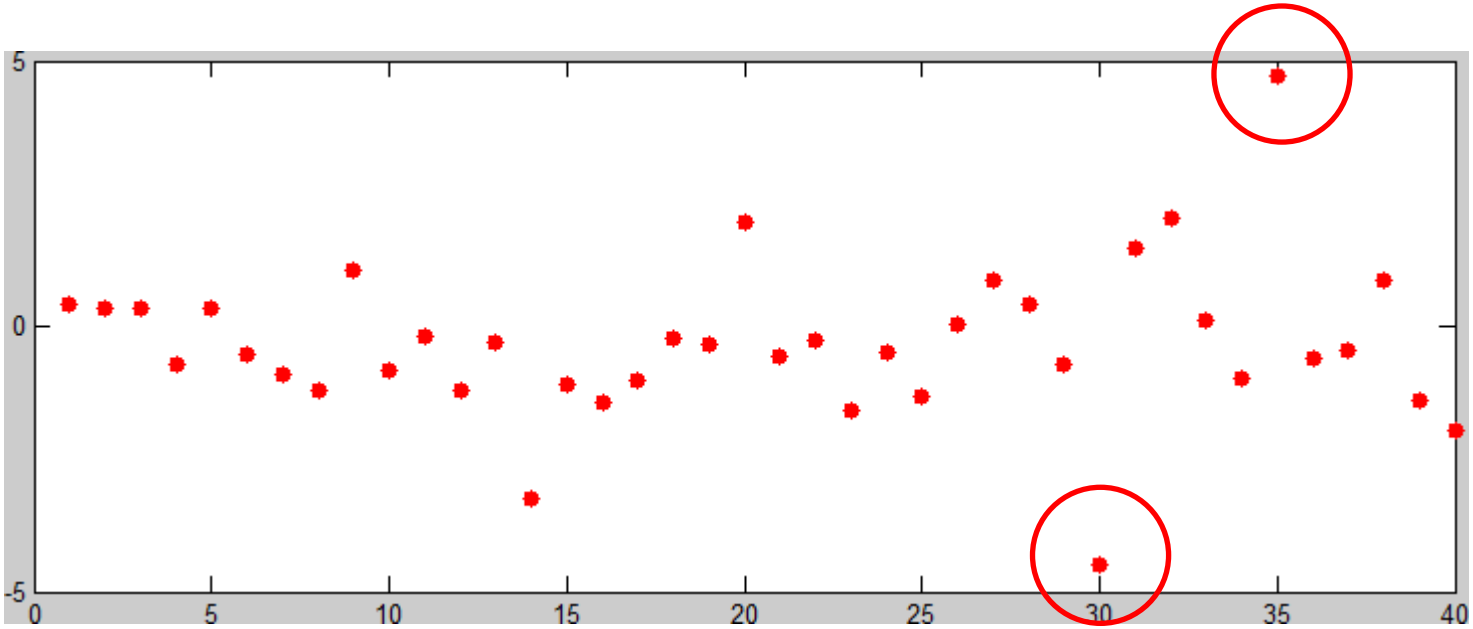
The Anomaly-Detection Problem

- In the anomaly-detection problem:
  - There is no an explicit temporal dimension.
  - Few anomaly might show up, the change cannot be considered as persisent.
  - Anomalies might comes from different processes.



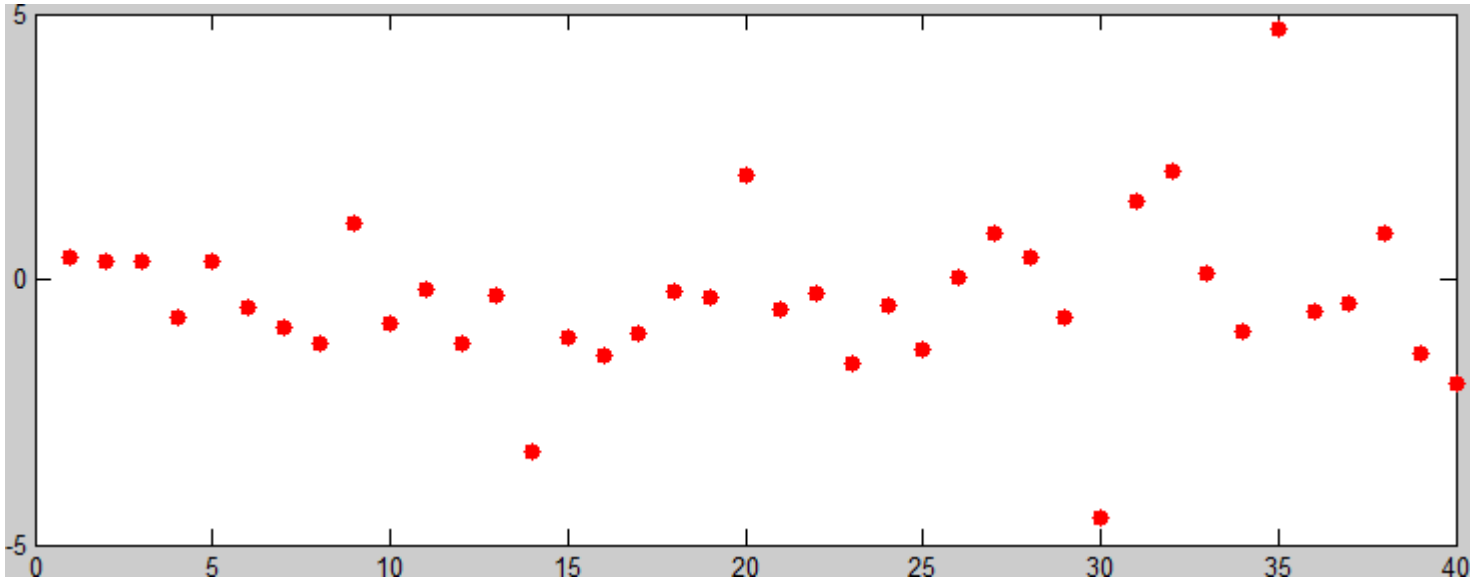
The Anomaly-Detection Problem

- In the anomaly-detection problem:
  - There is no an explicit temporal dimension.
  - Few anomaly might show up, the change cannot be considered as persisent.
  - Anomalies might comes from different processes.



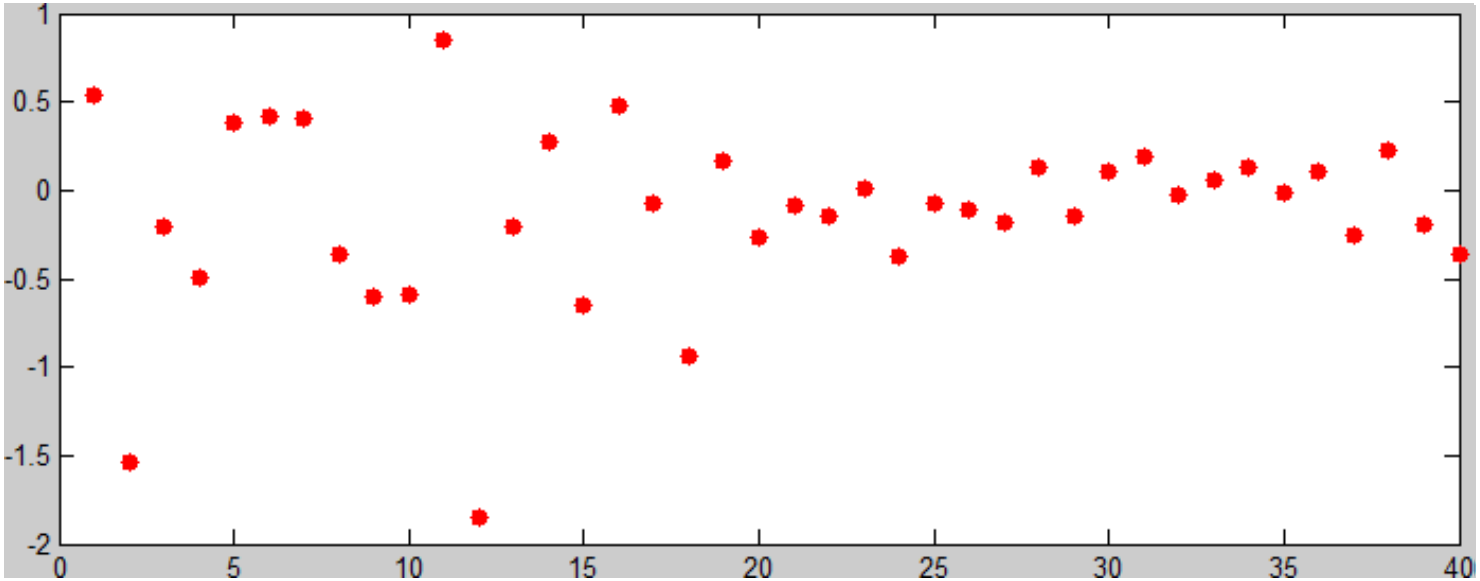
In the Random Variable World

- **Stationarity** means that a data are **i.i.d. realizations** of a random variable
- Not all outliers induce a process change



In the Random Variable World

- **Stationarity** means that a data are **i.i.d. realizations** of a random variable
- Not all process changes induce outliers



Detection Tools In the Random Variable World

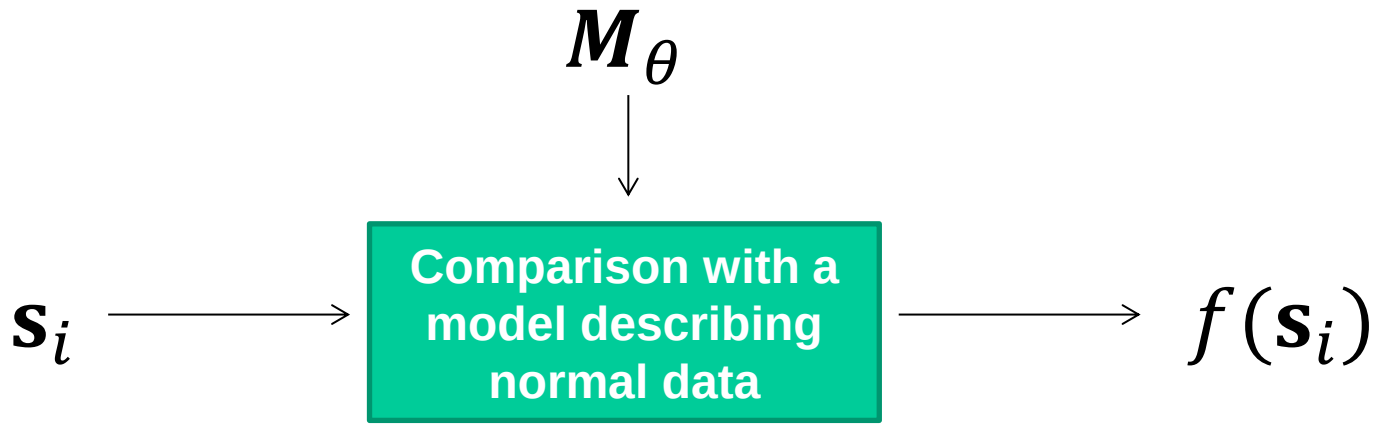
- Sequential Change-Detection Tests (CDTs) can be used for detecting changes in a datastream [Basseville 93]
  - Data are analyzed incrementally
  - Decisions are taken online considering in principle the whole past sequence

- Outlier detection methods:
  - Several statistical techniques have been developed ranging from graphical, confidence intervals-based, density-based and several others

[Basseville 93] M. Basseville and I. V. Nikiforov, Detection of abrupt changes: theory and application. Upper Saddle River, NJ, USA:Prentice-Hall, Inc., 1993.

And When Data Are Not Random Variables?

- Model-Based Approach
  - We learn a model to describe normal data (M_θ)
  - We measure the degree to which the model fits the data by means of change indicators $f(\mathbf{s}_i)$
  - We detect changes/anomalies by monitoring the change indicators as random variables



  - The most straightforward is $f(\mathbf{s}_i) = \|\mathbf{s}_i - R(\mathbf{s}_i, \boldsymbol{M}_\theta)\|_2$

A Semi-Supervised Learning Problem

- We assume that a **set of normal** data is provided for **training purposes:**
  - **Learning a suitable model** to compute the change indicators
  - **Learning the distribution** of **change indicators** in stationary conditions and run a change/anomaly detection algorithm on them
- **No examples of anomalies**/data generated after the **change** are instead **provided**
- In these settings, anomaly detection is also referred to as **novelty detetection** or **one class classification** [Pimentel 2014]

[Pimentel 2014] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," Signal Processing, vol. 99, pp. 215 – 249, June 2014.

SPARSE REPRESENTATIONS

for performing change/anomaly detection

Sparsity

- We say that a signal $\mathbf{s} \in \mathbb{R}^m$ is sparse w.r.t. to dictionary $D \in \mathbb{R}^{m \times n}$, i.e., a set of $\{\mathbf{d}_i\}_{i=1,..,n}$ vectors of $\mathbb{R}^m$

$$\exists\, \mathbf{x} \in \mathbb{R}^n \text{ s.t. } \mathbf{s} = \sum_{i=1}^{n} x_i \mathbf{d}_i$$

  and $\|\mathbf{x}\|_0 = L \ll n$.

  or equivalently, in matrix notation $\mathbf{s} = D\mathbf{x}$.

- Sparse signals in $\mathbb{R}^m$ live in a union of **low-dimensional subspaces** (each having dimension maximum L).
- When the (sorted) coefficients of $\mathbf{x}$ follow a power-law decay the signal is compressible, and can be accurately approximated by a sparse representation.

Sparse Representations

- **Sparse representations** has shown to be a very useful method for **constructing signal models**;
- The underlying assumption is that

$$\mathbf{s} = D\mathbf{x} + \boldsymbol{\nu}, \qquad \text{where } \boldsymbol{\nu} \text{ is a noise term}$$

- Where $D \in \mathbb{R}^{n \times m}$ is the **dictionary**, whose columns are called **atoms**, $\mathbf{x}$ are the coefficients which are assumed to be sparse, i.e., $\|\mathbf{x}\|_0 \ll n$
- There are efficient tools for computing $\hat{\mathbf{x}}$, the sparse approximation of a signal $\mathbf{s}$ w.r.t. a given dictionary D

$$\hat{\mathbf{s}} = D\hat{\mathbf{x}} \text{ and } \hat{\mathbf{s}} \approx D\mathbf{x}$$

in a sense that $\|D\mathbf{x} - \hat{\mathbf{s}}\|_2$ is small

- This operation is referred to as the **sparse coding**

Sparse Coding

- Sparse coding solving the following constrained problem

$$\text{P0:} \quad \hat{\mathbf{x}}_{\mathbf{0}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \ \|D\mathbf{x} - \mathbf{s}\|_{\mathbf{2}} \ \text{s.t.} \ \|\mathbf{x}\|_0 \leq L$$

- Exact solutions are computationally intractable.
- Typically solved by means of Greedy Algoritms, such as the Orthogonal Matching Pursuit (OMP).
- Solving this problem actually corresponds to projecting the observed data into the union of subspaces (determined by at most L atoms).

Sparse Coding (cnt)

- Sparse coding solving the following unconstrained problem

$$\text{P1:} \quad \hat{\mathbf{x}}_{\mathbf{1}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \; J_\lambda(\mathbf{x}, D, \mathbf{s})$$

where the functional is

$$J_\lambda(\mathbf{x}, D, \mathbf{s}) = \|D\mathbf{x} - \mathbf{s}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

- The sparsity requirement is relaxed by a penalization term on the ℓ_1- norm of the coefficients
- Under some conditions the solution of $\mathrm{P0}$ and $\mathrm{P1}$ do coincide
- This is a Basis Pursuit Denoising (BPDN) problem: there are several optimization methods in the literature.
- We adopt Alternating Direction Method of Multipliers (ADMM)

Dictionary Learning

- It is possible to learn a dictionary $\widehat{D}$ that provides sparse approximation for a set of training data $T \in \mathbb{R}^{m,l}$.
- Solution is a joint optimization over the dictionary and coefficients of a sparse repr. of the training matrix T

$$\widehat{D} = \underset{D \in \mathbb{R}^{m \times n}, X \in \mathbb{R}^{n \times l}}{\operatorname{argmin}} \|DX - T\|_F$$

such that $\|\mathbf{x}_k\|_0 \leq L, \forall k$

  - Greedy solutions are obtained by alternating optimization over the dictionary atoms and the sparse representations of the training set
  - We consider here the KSVD [Aharon 06] that uses OMP as sparse coding stage

[Aharon 06] M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," Transactions on Signal Processing vol. 54, no. 11, November 2006, pp. 4311–4322.

Learning a Dictionary for Modeling Stationarity

- Learning $\widehat{D}$ corresponds to learning the union of subpaces where signals in T - generated in stationarity - live.
- In order to measure the extent to which a given signal $\mathbf{s}$ is **consistent with the stationary conditions** we compute the **sparse coding** of $\mathbf{s}$ w.r.t. $\widehat{\boldsymbol{D}}$

$$\mathbf{s} \rightarrow \hat{\mathbf{s}}, \quad \text{where } \hat{\mathbf{s}} = D\hat{\mathbf{x}} \text{ and } \hat{\mathbf{s}} \approx \mathbf{s}$$

- We need suitable **change-indicators** that **quantitatively assess,** in the sparse domain, how close $\mathbf{s}$ is to stationary signals.
  - In the specific case of sparse representations, the **change indicators** have to take into account both **accuracy** and **sparsity** of the representation

Change Indicators

- The **following change indicators** have been considered:
  - When solving $\mathrm{P0}$ the reconstruction error

    $e(\mathbf{s}) = \left\|\mathbf{s} - \widehat{D}\hat{\mathbf{x}}_{\mathbf{0}}\right\|_{\mathbf{2}}$, being $\hat{\mathbf{x}}_{\mathbf{0}}$ the solution of $\mathrm{P0}$
  - When solving $\mathrm{P1}$, the value of the functional

    $f(\mathbf{s}) = \left\|\mathbf{s} - \widehat{D}\hat{\mathbf{x}}_{\mathbf{1}}\right\|_{\mathbf{2}} + \lambda\|\hat{\mathbf{x}}_{\mathbf{1}}\|_{\mathbf{1}}$, being $\hat{\mathbf{x}}_{\mathbf{1}}$ the solution of $\mathrm{P1}$
  - When solving $\mathrm{P1}$, jointly the sparsity and the error

    $g(\mathbf{s}) = [\left\|\mathbf{s} - \widehat{D}\hat{\mathbf{x}}_{\mathbf{1}}\right\|_{\mathbf{2}};\ \lambda\|\hat{\mathbf{x}}_{\mathbf{1}}\|_{\mathbf{1}}]$, being $\hat{\mathbf{x}}_{\mathbf{1}}$ the solution of $\mathrm{P1}$

Anomaly Detection from Change Indicators

- We treat **change indicators** computed from i.i.d. stationary data as **random variables**.
- We define **high-density regions** for the empirical distribution of change indicators from T
- In case of 1D-change indicators, an high-density region is

$$\mathcal{I}_{\alpha}^{e} = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$$

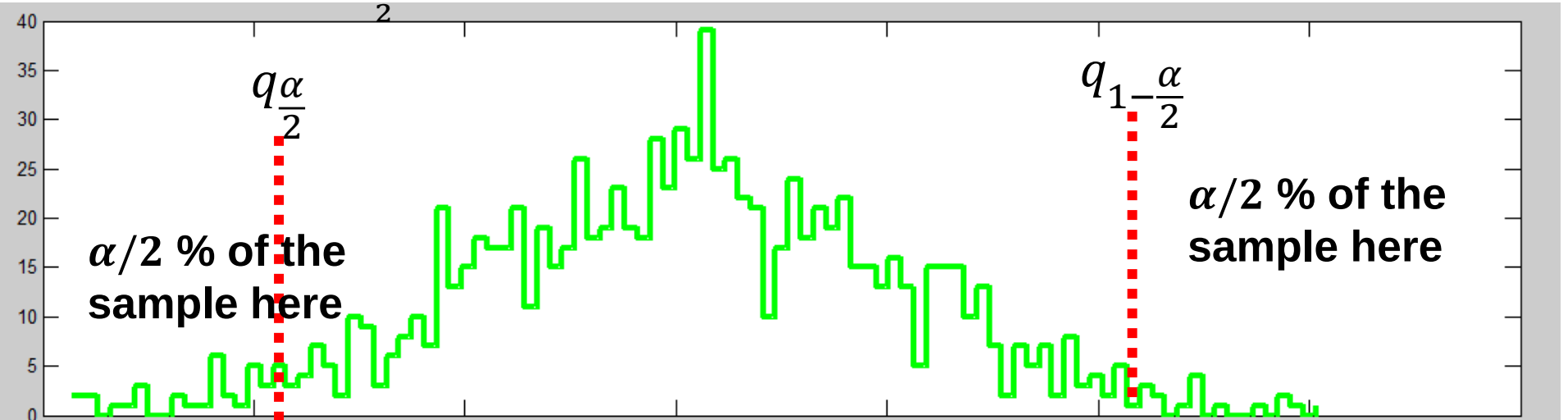
where $q_{\frac{\alpha}{2}}$ is the $\alpha/2$ quantile of the empirical distribution

Anomaly Detection from Change Indicators

- We treat **change indicators** computed from i.i.d. stationary data as **random variables**.
- We define **high-density regions** for the empirical distribution of change indicators from T
- In case of 1D-change indicators, an high-density region is

$$\mathcal{I}_{\alpha}^{e} = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$$

where $q_{\frac{\alpha}{2}}$ is the $\alpha/2$ quantile of the empirical distribution



Anomaly Detection from Change Indicators

- We treat **change indicators** computed from i.i.d. stationary data as **random variables**.
- We define **high-density regions** for the empirical distribution of change indicators from T
- In case of 1D-change indicators, an high-density region is

$$\mathcal{I}_{\alpha}^{e} = [q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}}]$$

  where $q_{\frac{\alpha}{2}}$ is the $\alpha/2$ quantile of the empirical distribution

- We detect anomalies as data yilding change indicators, out of high-density regions (outliers)

$$e(\mathbf{s}) \notin \mathcal{I}_{\alpha}^{e}$$

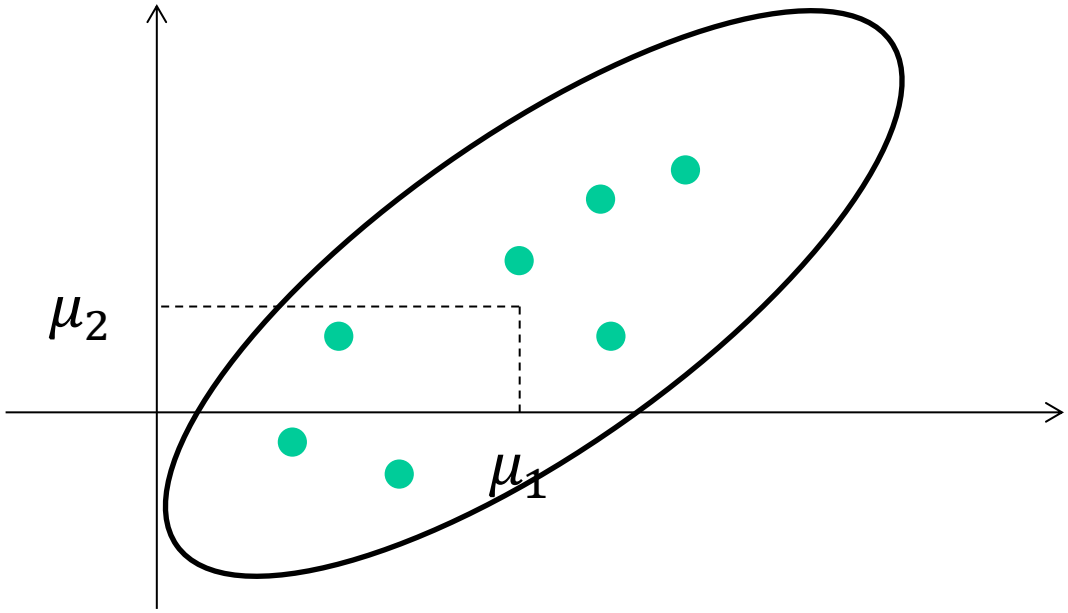
- The same for change indicators $f(\cdot)$

Anomaly Detection from 2D Change Indicators

- For the bivariate indicator $g(\cdot)$ we build a confidence region

$$R_\gamma = \left\{ \xi \in \mathbb{R}^2, \text{s.t.} \sqrt{(\xi - \mu)' \Sigma^{-1} (\xi - \mu)} \leq \gamma \right\}$$

where μ and Σ are the sample mean and sample covariance of the change indicators from T.



Anomaly Detection from 2D Change Indicators

- For the bivariate indicator $g(\cdot)$ we build a confidence region

$$R_\gamma = \left\{\xi \in \mathbb{R}^2, \text{s.t.} \sqrt{(\xi-\mu)'\Sigma^{-1}(\xi-\mu)} \leq \gamma\right\}$$

  where μ and Σ are the sample mean and sample covariance of the change indicators from T.

- The Chebyshev's inequality says that a normal patch falls outside R_γ with probability $\leq 2/\gamma^2$

- Anomalies are detected as

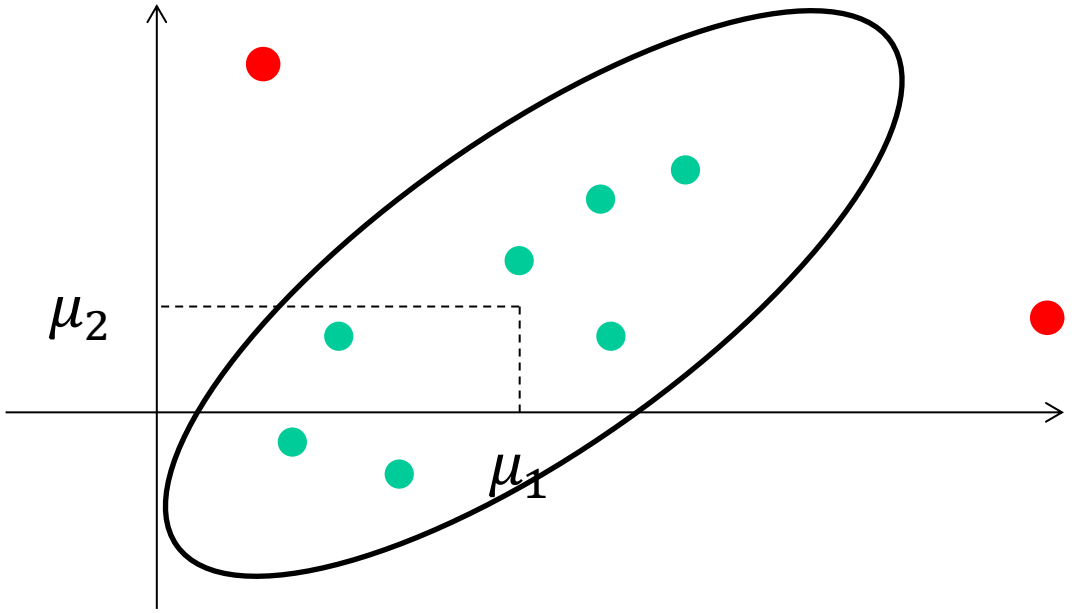
$$\mathbf{s} \text{ s.t. } \sqrt{(\boldsymbol{g}(\mathbf{s})-\mu)'\Sigma^{-1}(\boldsymbol{g}(\mathbf{s})-\mu)} > \gamma$$

Anomaly Detection from 2D Change Indicators

- For the bivariate indicator $g(\cdot)$ we build a confidence region

$$R_\gamma = \left\{ \xi \in \mathbb{R}^2, \text{s.t.} \sqrt{(\xi - \mu)' \Sigma^{-1} (\xi - \mu)} \leq \gamma \right\}$$

where μ and Σ are the sample mean and sample covariance of the change indicators from T.



Change Detection from Change Indicators

- We adopt Change-Point Methods (CPMs) [Hawkins 2003]
- CPMs are hypothesis test to assess weather a finite sequence $\{x_t\}_{t=1,\dots,N}$ contains a change point, i.e.,

$$\begin{cases} H_0\text{: "all data in the sequence are i.i.d."} \\ H_1\text{: "the sequence contains a change point."} \end{cases}$$

- A change-point is a point such that

$$x_t = \begin{cases} x_t \sim \phi_o & t < T^* \\ x_t \sim \phi_1 & t \geq T^* \end{cases}$$

[Hawkins 2003] D. M. Hawkins, P. Qiu, and C. W. Kang, "The changepoint model for statistical process control," Journal of Quality Technology, vol. 35, No. 4, pp. 355–366, 2003.

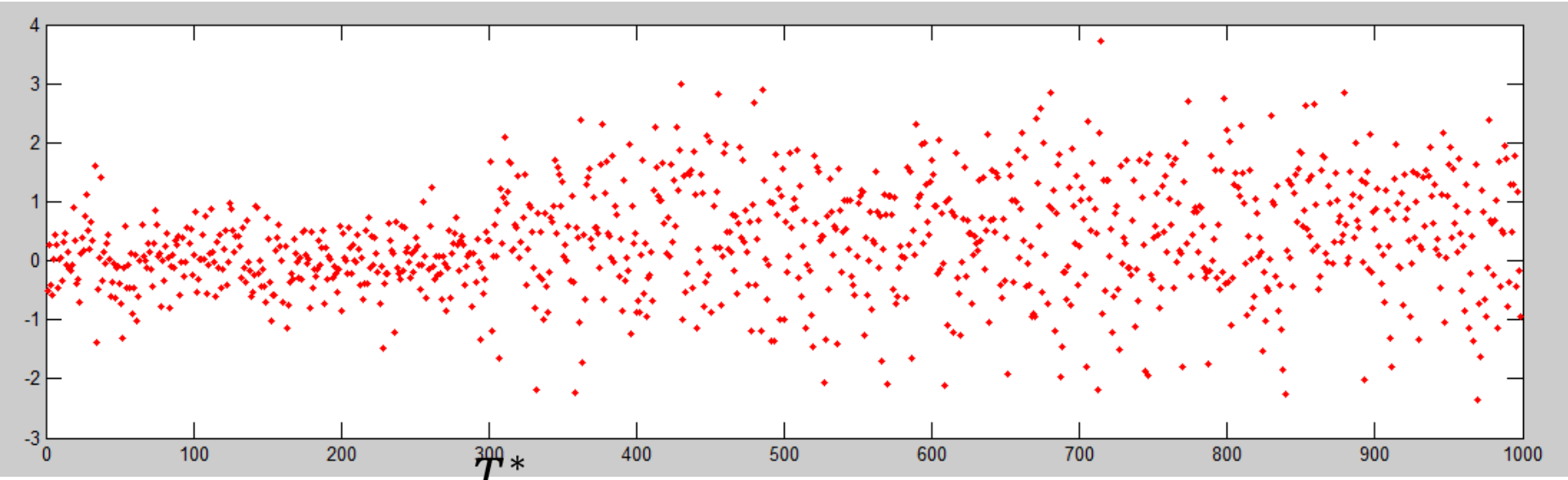
Change-Point Methods

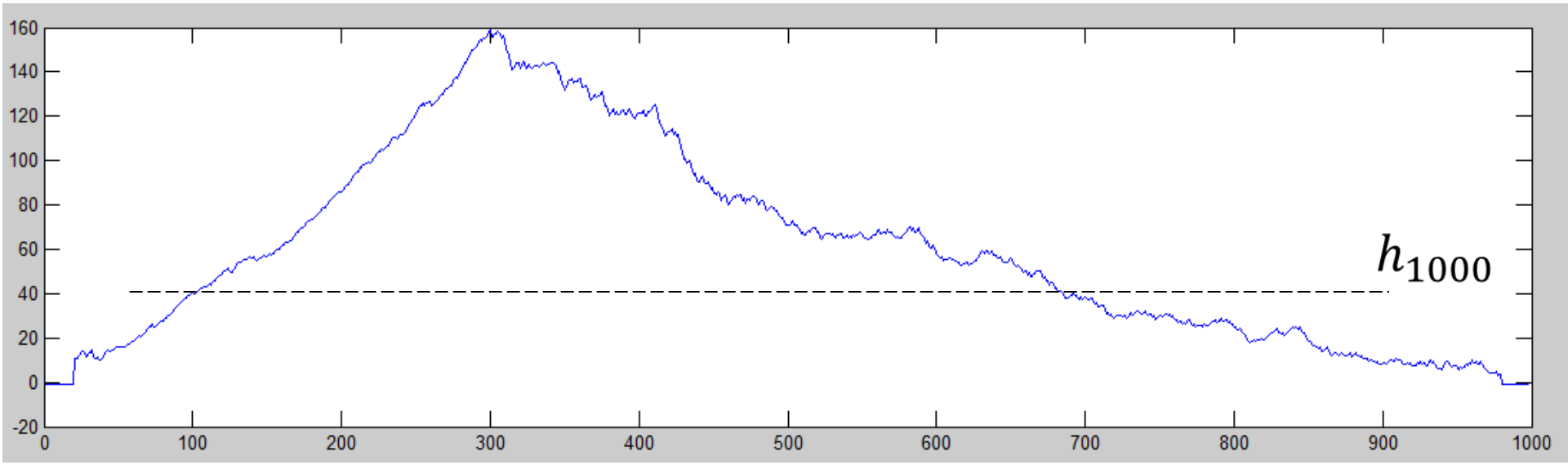
- Each point S in the sequence $\{x_t\}_t = 1, \ldots$ is considereded as a perspective change point
- The two sets

$$A_S = \{x_t, t < S\}$$
$$B_S = \{x_t, S \leq t \leq N\}$$

are compared by means of a **suitable statistic** $\mathcal{T}$

- When the **partitioning** corresponding the **largest value** of the **statistics** yields **enough statistical evidence for claiming the change**, the sequence contains a change point


4
3
2
1
0
-1
-2
-3
0
100
200
300
400
500
600
700
800
900
1000
$T*$


160
140
120
100
80
60
40
20
0
-20
0
100
200
300
400
500
600
700
800
900
1000
h_{1000}
τ

Change Detection from Change Indicators

- Optimized implementations of CPMs have been recently presented to operate **online in a nonparametric** manner
- We adopt the CPM in [Ross 2011] based on the Lepage Test Statistic
- The **Lepage** test Statistic detects **changes in the scale and location** of an unknown random variable
- To monitor a of data $\{\mathbf{x}_i\}_{i=1,\dots}$ we compute, at each new arrival the change indicator $e(\mathbf{x}_i)$ and use CPM of Lepage test statistic to detect changes in the location and scale of the change indicators

[Ross 2011] G. J. Ross, D. K. Tasoulis, and N. M. Adams, "Nonparametric monitoring of data streams for changes in location and scale," Technometrics, vol. 53, no. 4, pp. 379–389, 2011.

Alternative Approaches

- There are not so many solution for change/anomaly detection using sparse representation
- In the CS scenario there are quite a few works concerning
  - the signal detection problem, see references in [Alippi 2013]
  - and other methods assuming known changes
- Sparse representations have been though used for discriminative tasks such as classifications

[Alippi 2014] C. Alippi, G. Boracchi, and B. Wohlberg, "Change detection in streams of signals with sparse representations," ICASSP 2014 , pp. 5252 – 5256.

Other Solutions

- In [Adler 2013] the anomaly detection is performed during the sparse coding. The following model is consider

$$\mathbf{s} = D\mathbf{x} + \mathbf{a} + \boldsymbol{\nu} \quad \text{where } \boldsymbol{\nu} \text{ is a noise term}$$

  and $\mathbf{a}$ collects all the components of $\mathbf{s}$ that cannot be sparsely approximated.

- Sparse coding is performed solving the following problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{s} - D\boldsymbol{x} - \boldsymbol{a}\|_2 + \|\boldsymbol{x}\|_1 + \|\boldsymbol{a}\|_2$$

- Normal patches: $\|\boldsymbol{a}\|_2$ is negligible, anomalous patches: $\|\boldsymbol{a}\|_2$ is large.

- Anomalies detected comparing $\|\boldsymbol{a}\|_2$ against a threshold

[Adler 2013] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," in Proc. of IEEE MLSP, September 2013,

EXPERIMENTS

Performing change/anomaly detection using sparse representations

Anomaly detection in images

- Data are 8×8 patches extracted from textured images characterized by a specific structure

Test on Synthetic Images

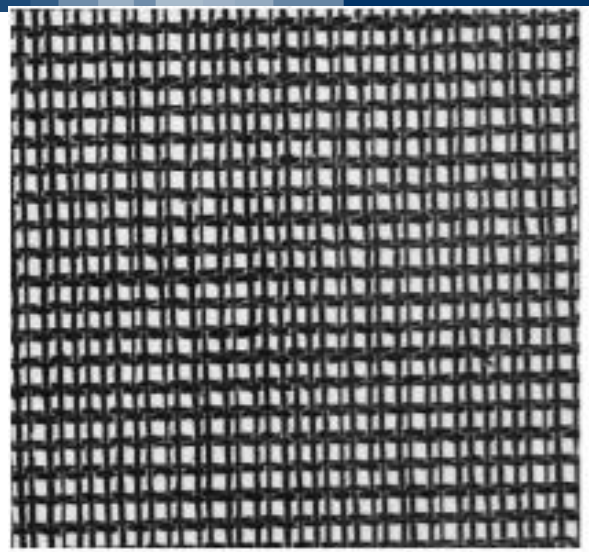


Image 1

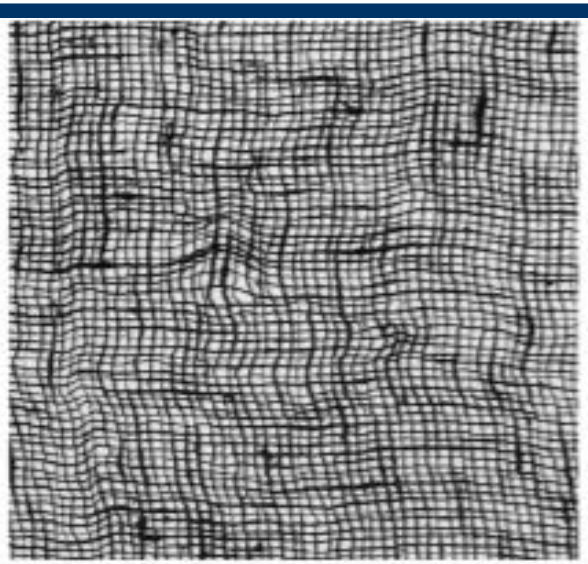


Image 2

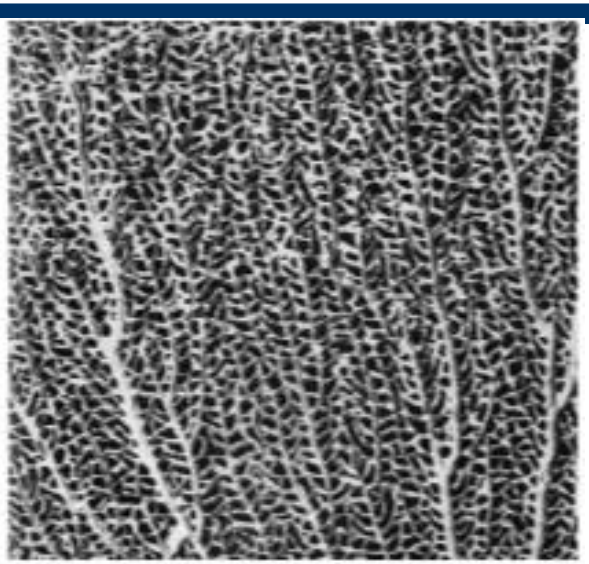


Image 3

Image 4

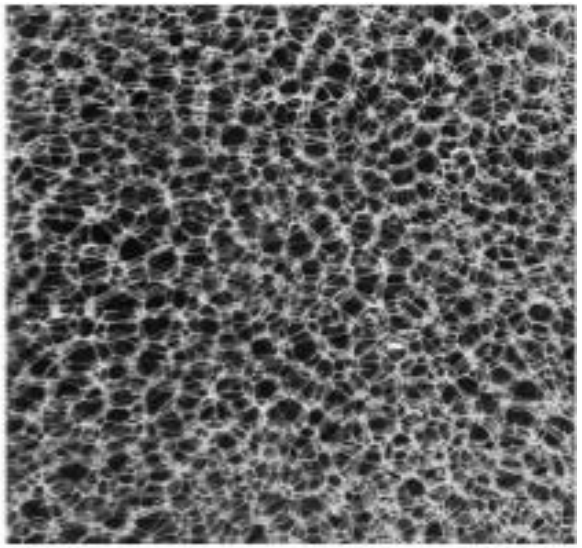


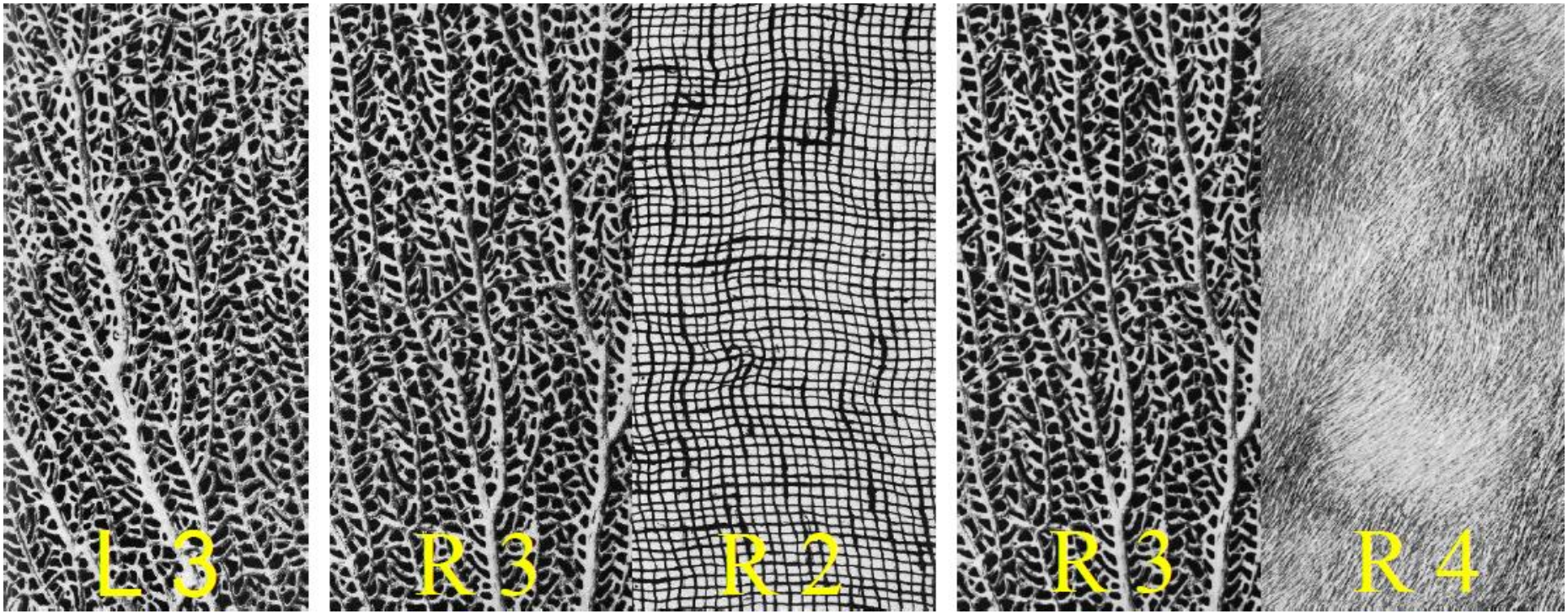
Image 5

Anomaly detection in images

- Data are 8×8 patches extracted from textured images characterized by a specific structure
- Anomaly detection problems are simulated by assembling test images that contains patches from different texture
  - The left half of each image is used to learn $\widehat{D}$
  - The right half is used for testing and juxtaposed with other half images

Test Images

Test images



We learn a dictionary from here

Anomaly detection in images

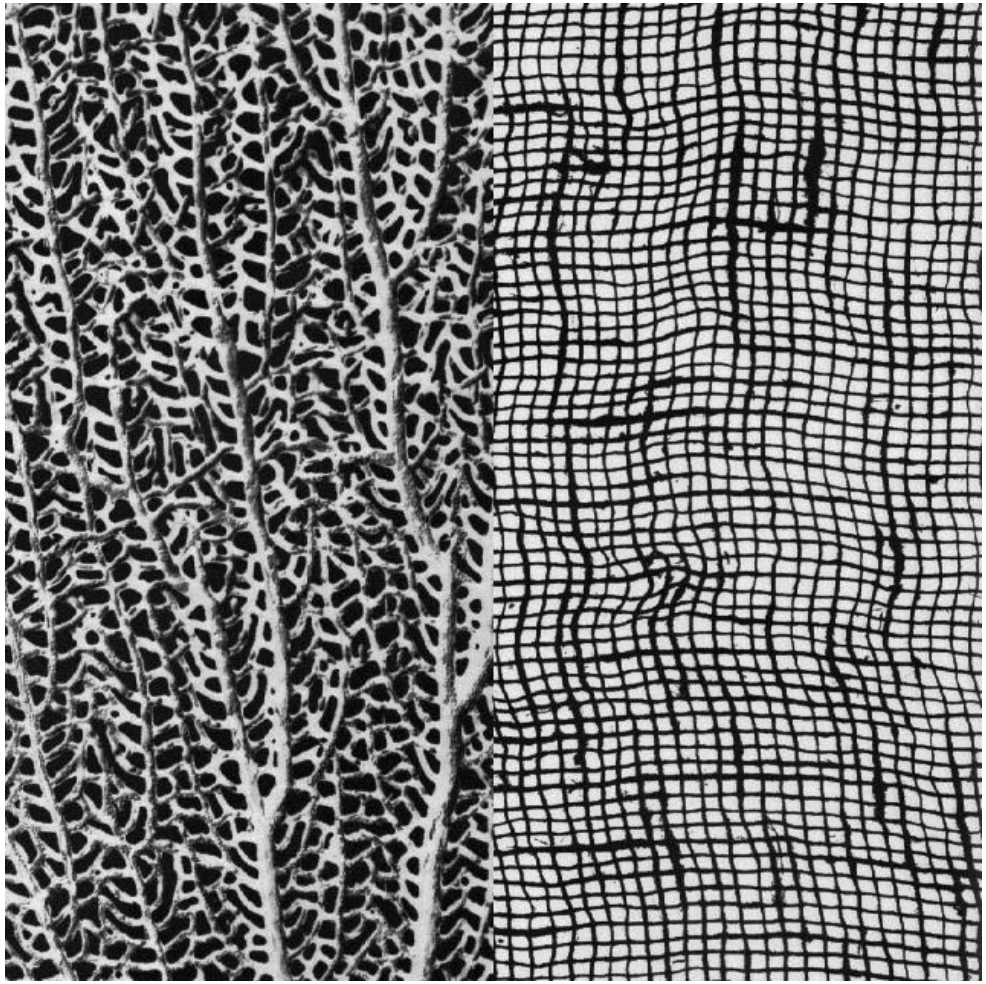
- Data are 8×8 patches extracted from textured images characterized by a specific structure
- Anomaly detection problems are simulated by syntetically creating test images gathering patches from different texture
- Each patch is **pre-processed** by subtracting its mean
- **No post-processing** to aggregate decision spatially is performed
- For further details, please refer to [Boracchi 2014]

[Boracchi 2014] Giacomo Boracchi, Diego Carrera, Brendt Wohlberg «Anomaly Detection in Images By Sparse Representations» SSCI 2014

Figures of Merit

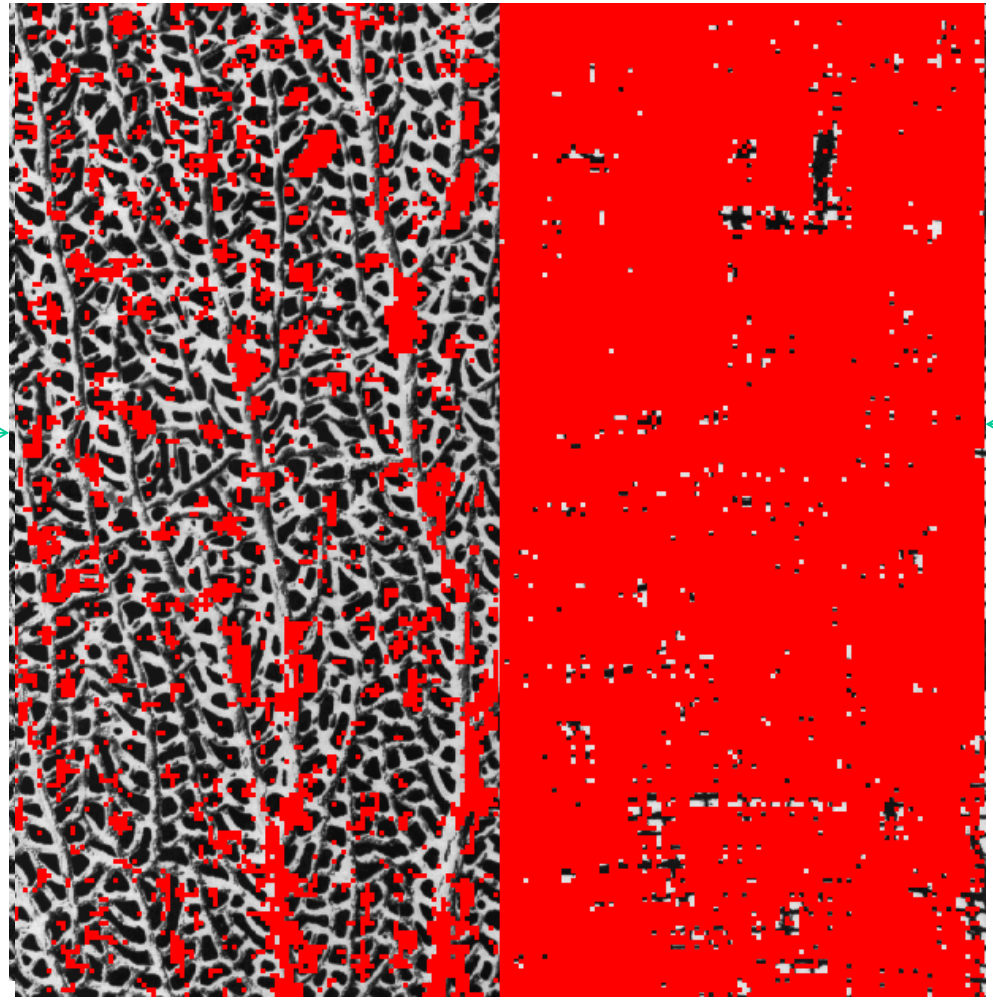
- FPR: the false positive rate, i.e. the percentage of normal patches labelled as abnormal
- TPR: the false negative rate, i.e., the percentage of anomalies correctly detected

Figures of Merit



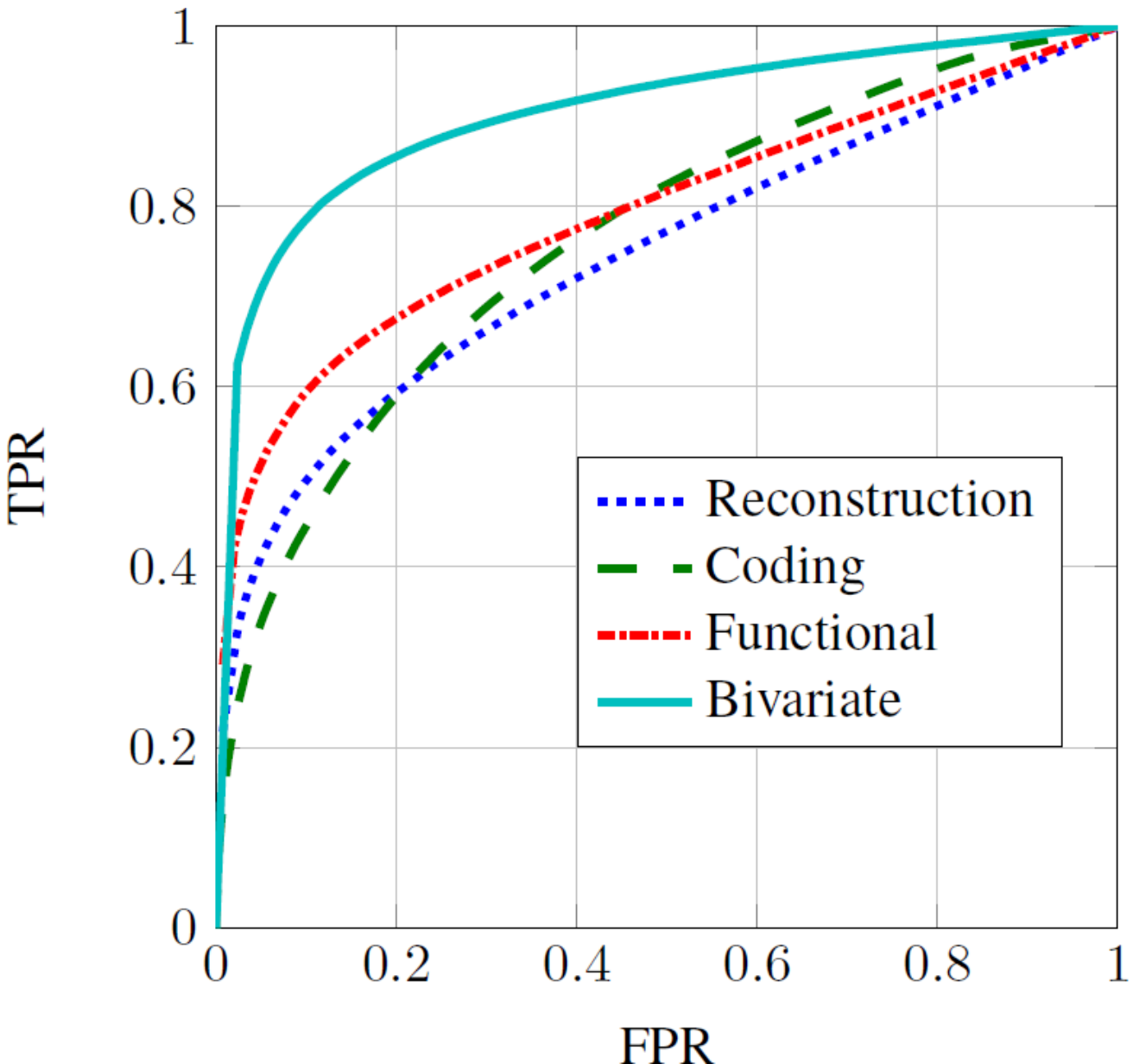
Figures of Merit

False Positives



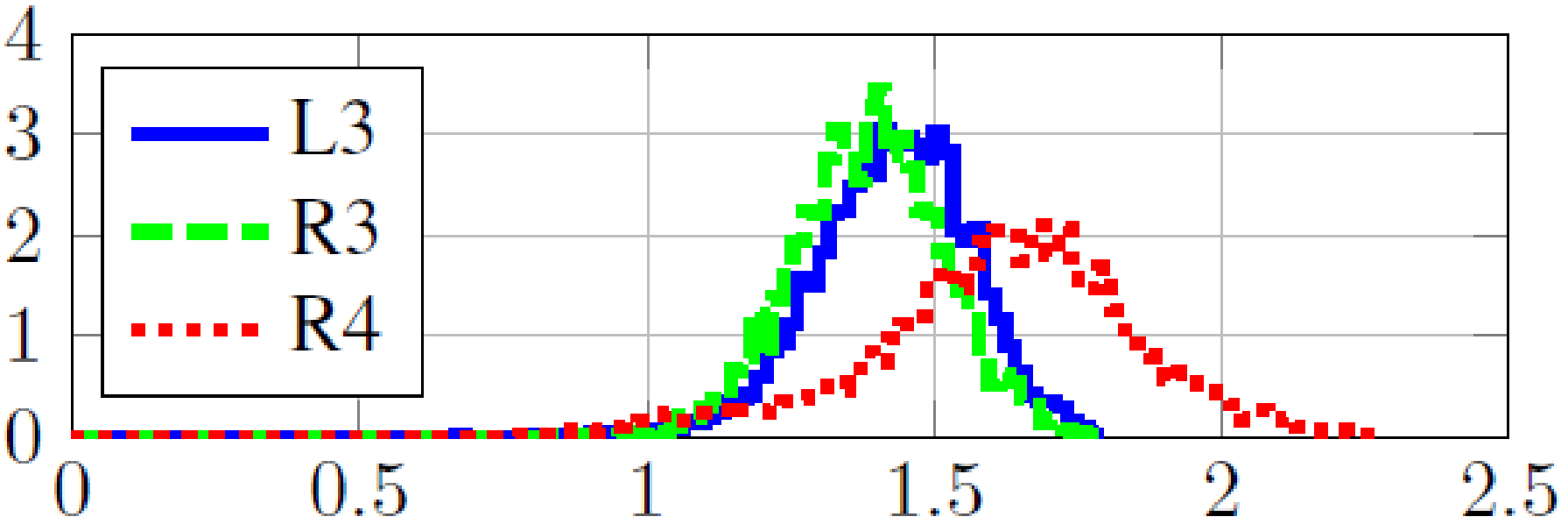
True Positives

Performance evaluation of the considered indicators

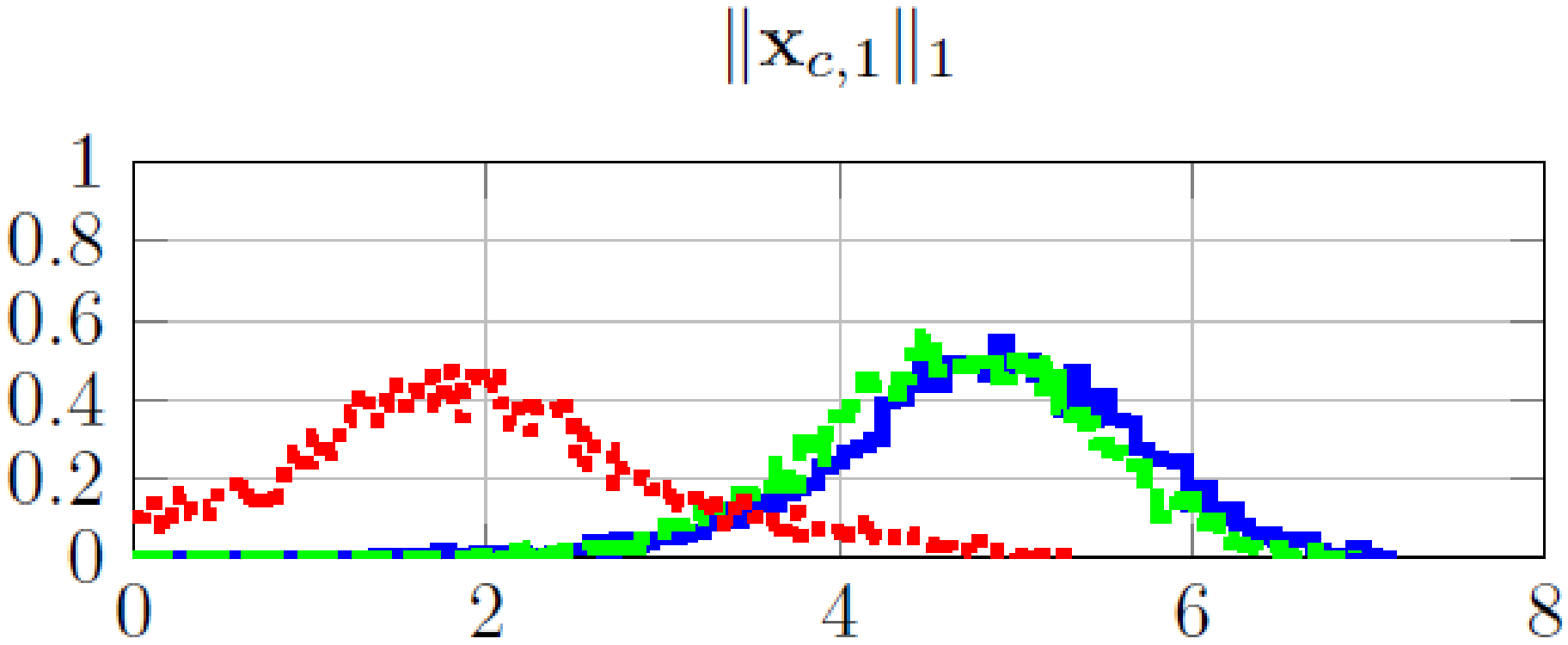


Why such a performance gap?

$$\|\widehat{\mathbf{D}}\mathbf{x}_{c,1} - \mathbf{s}_c\|_2$$

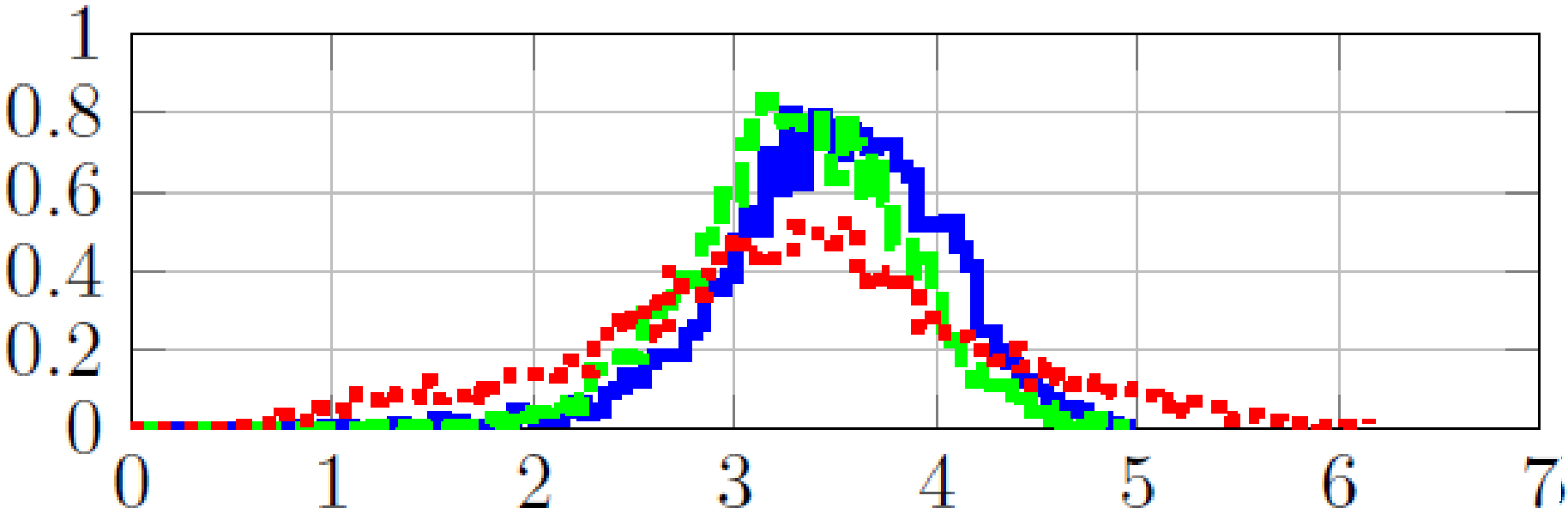


Why such a performance gap?



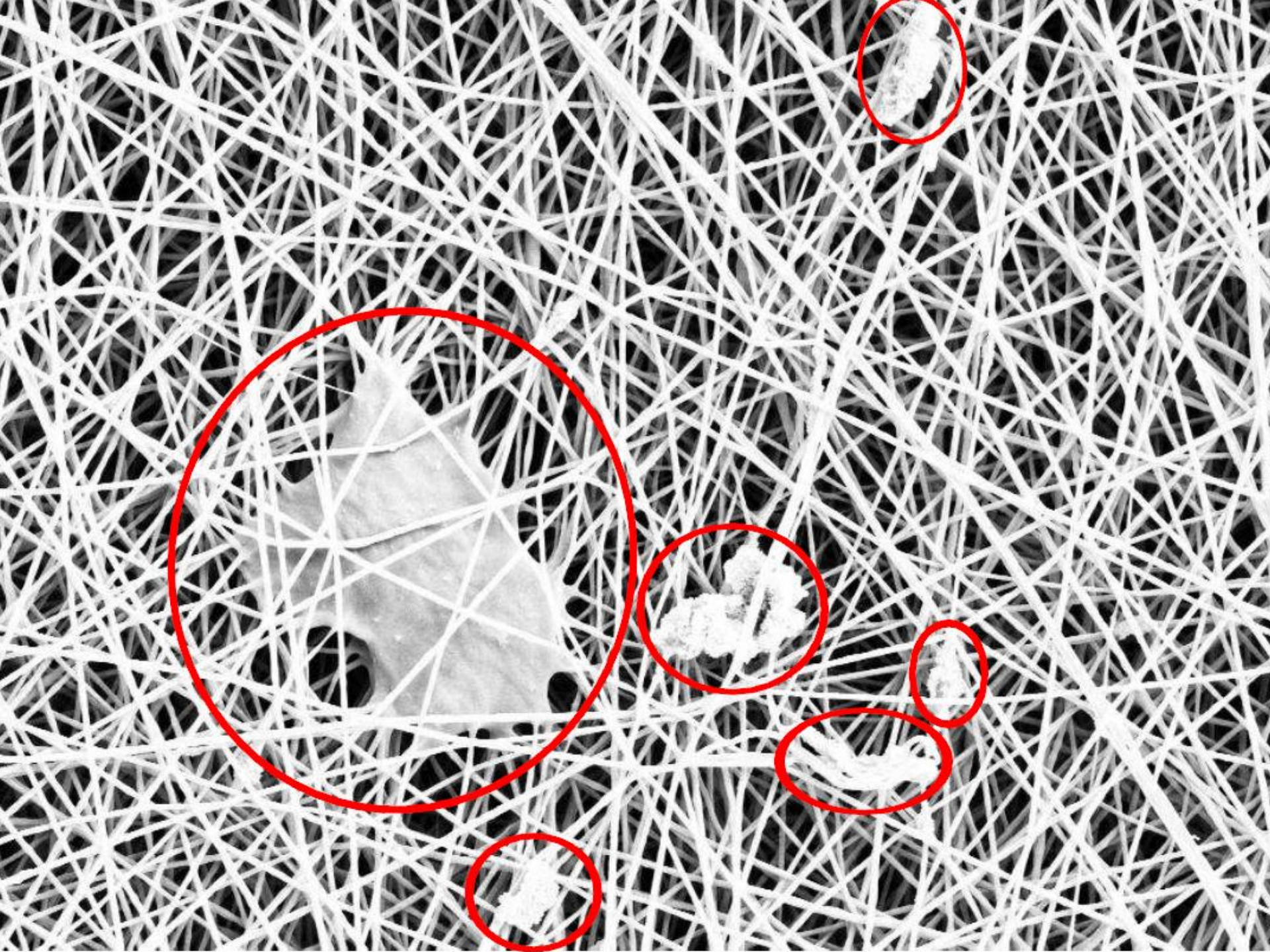
Why such a performance gap?

$$\|\widehat{\mathbf{D}}\mathbf{x}_{c,1} - \mathbf{s}_c\|_2 + \lambda\|\mathbf{x}_{c,1}\|_1$$



Anomaly detection in SEM images

- Problem Description: we consider the production of nanofibrous materials by an electrospinning process
- An scanning electron microscope (SEM) is used to monitor the production process and detect the presence of
  - Beads
  - Films
- Detecting anomalies and assessing how large they are is very important for supervising the monitoring process

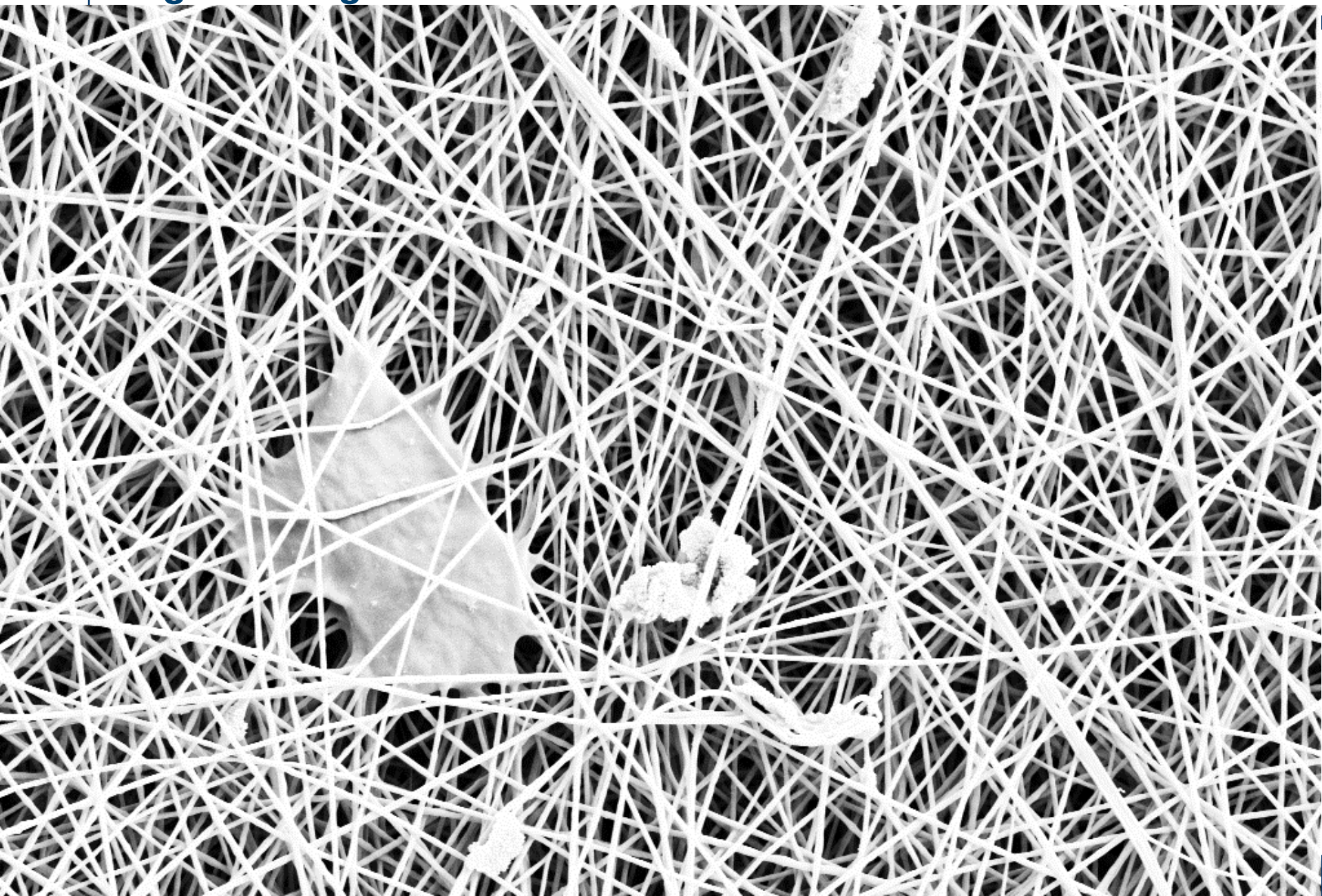


Anomaly detection in SEM images

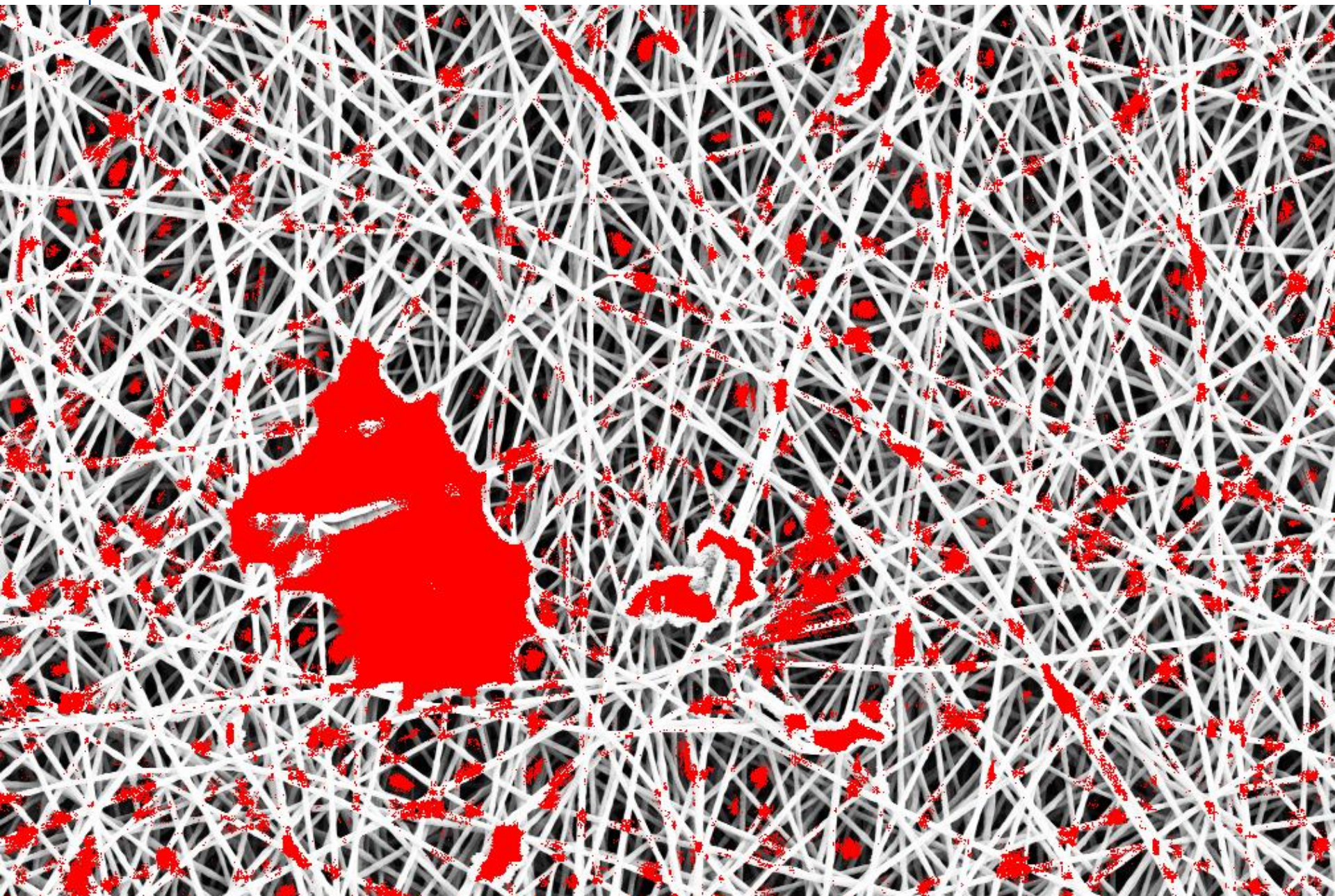
- Problem Description: we consider the production of nanofibrous materials by an electrospinning process
- An scanning electron microscope (SEM) is used to monitor the production process and detect the presence of
  - Beads
  - Films
- Detecting anomalies and assessing how large they are is very important for supervising the monitoring process
- All the anomaly detection methods have been manually tuned to operate at its best performance
- Further details can be found in [Boracchi 2014]

[Boracchi 2014] Giacomo Boracchi, Diego Carrera, Brendt Wohlberg «Anomaly Detection in Images By Sparse Representations» SSCI 2014

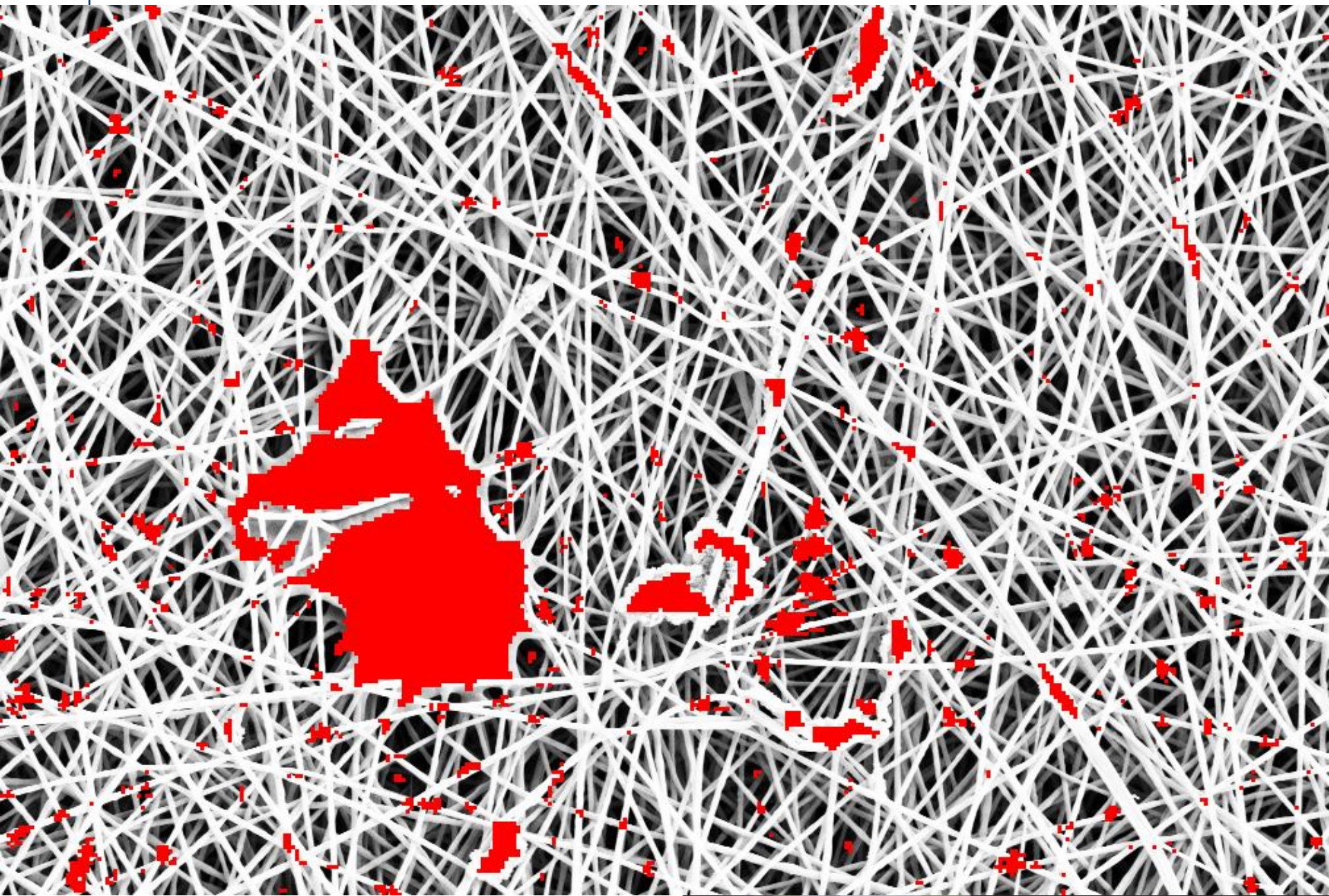
Original Image



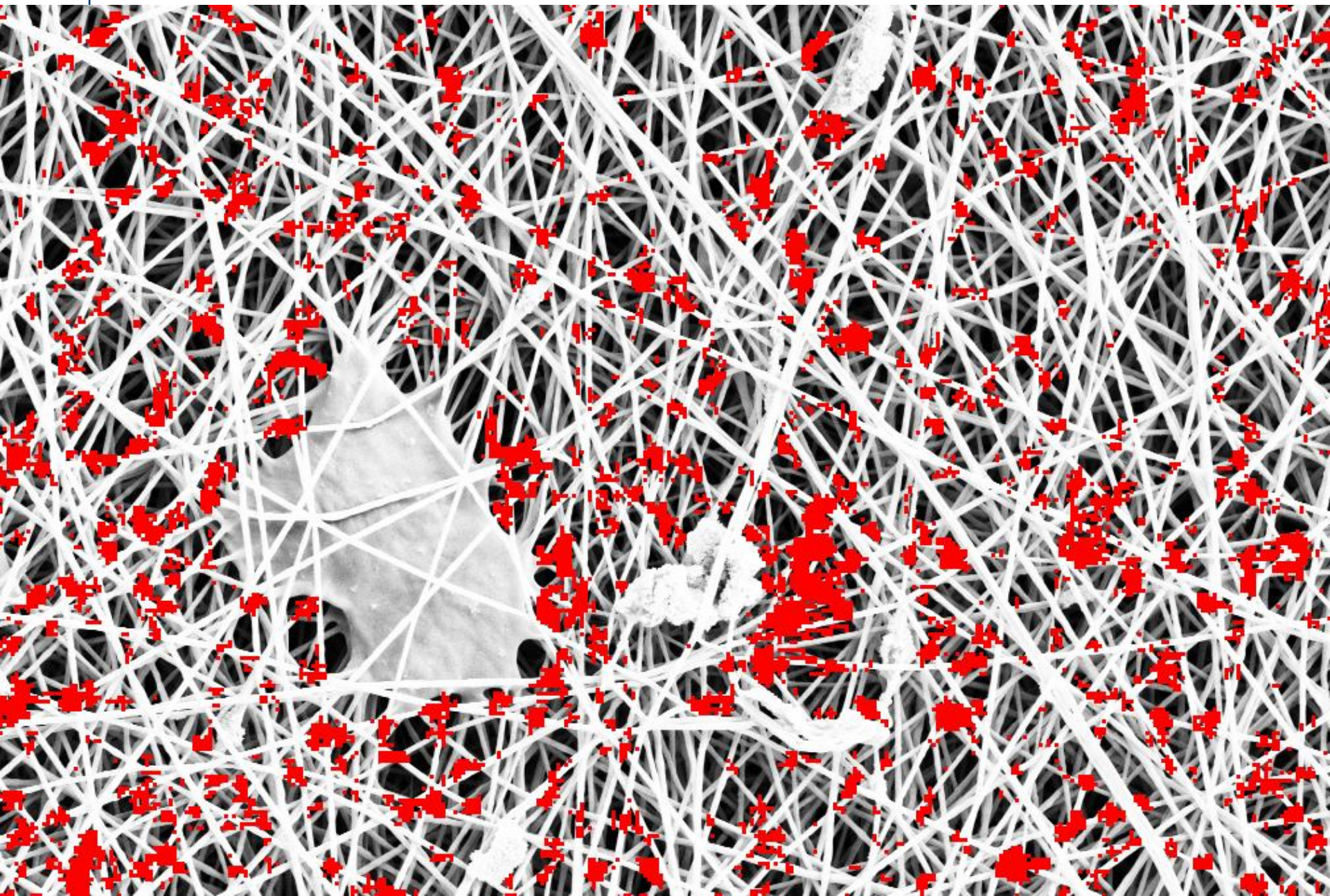
Anomaly detection by means of $e(\cdot)$



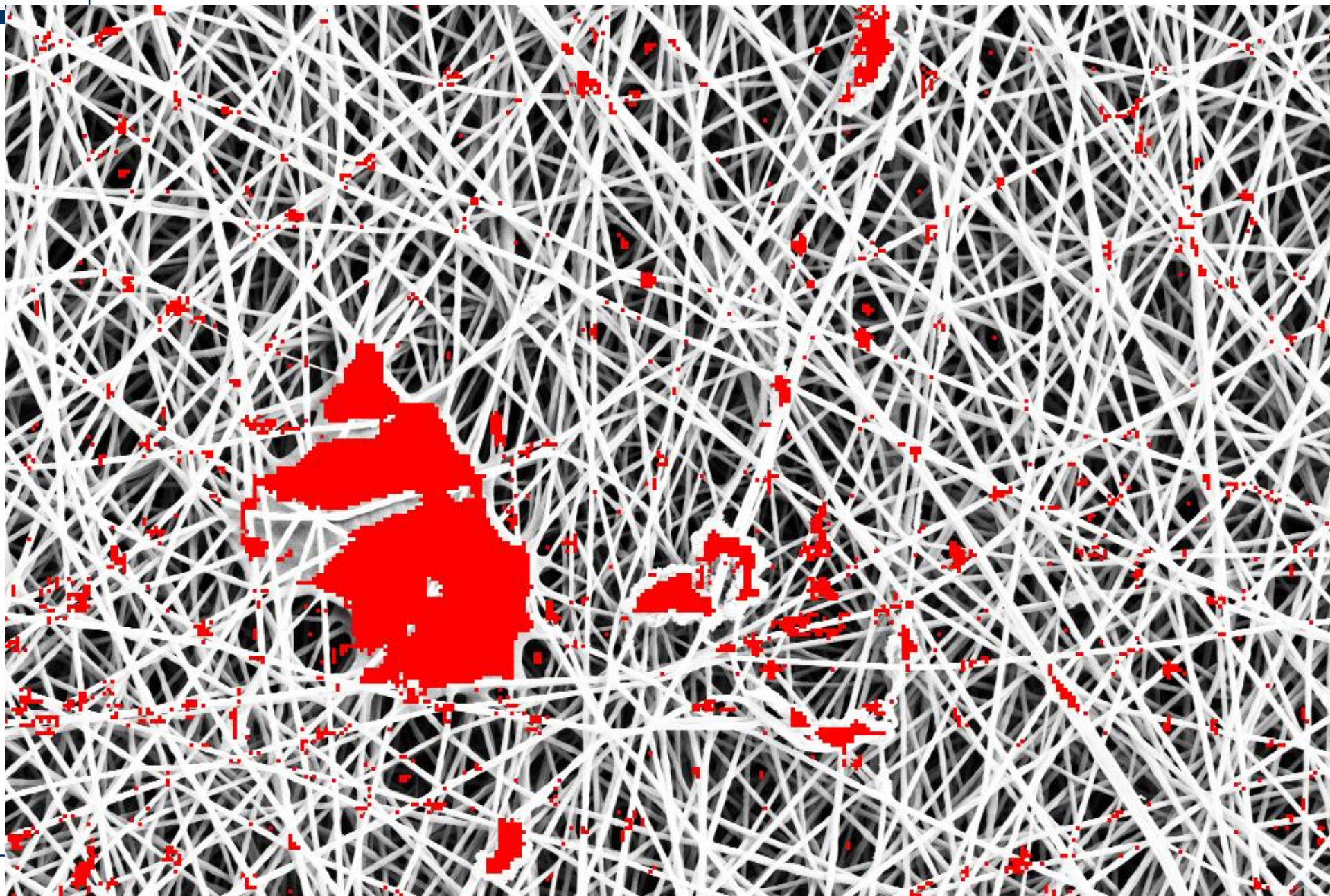
Anomaly detection by means of $f(\cdot)$



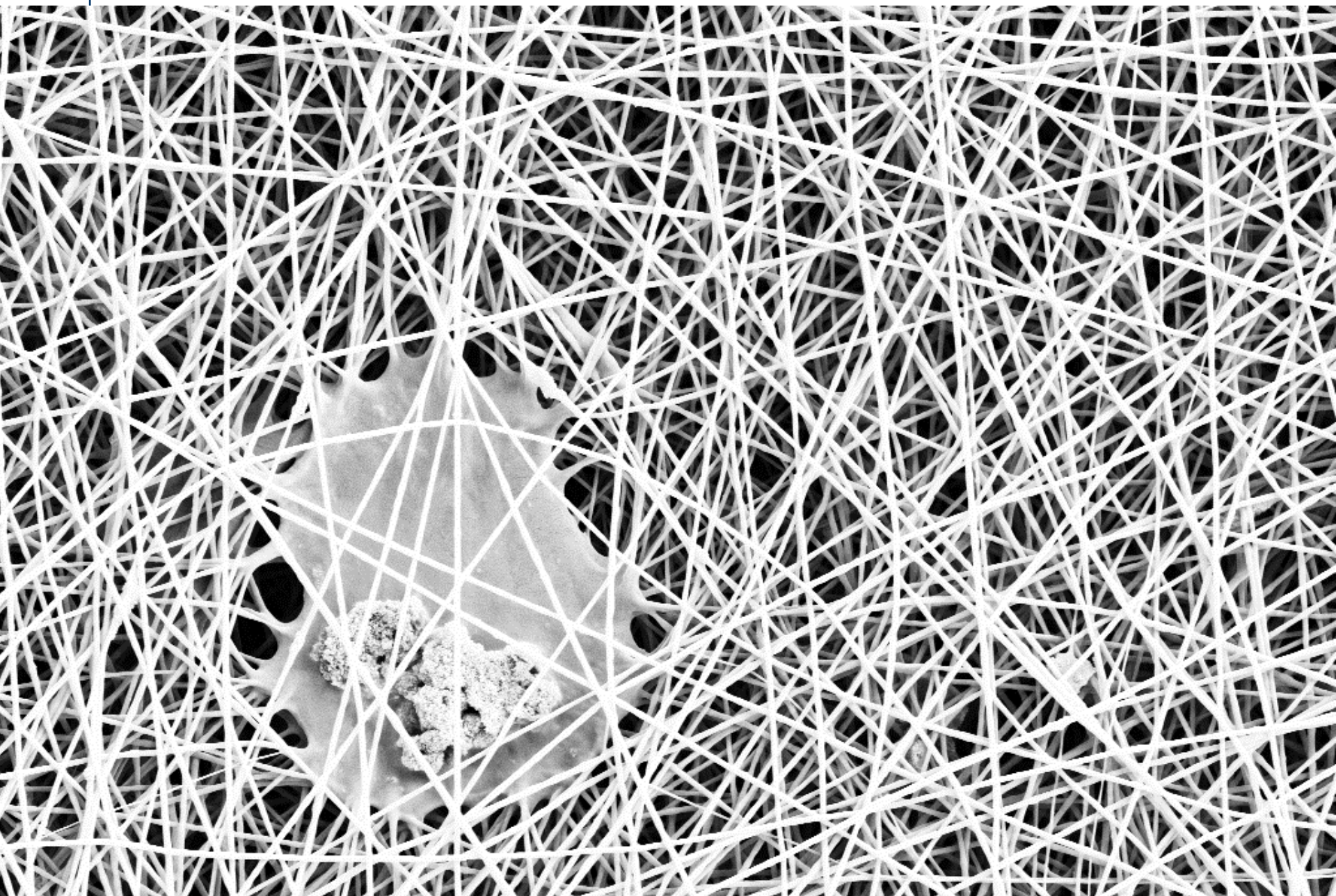
Anomaly detection by means of [Adler 2013]



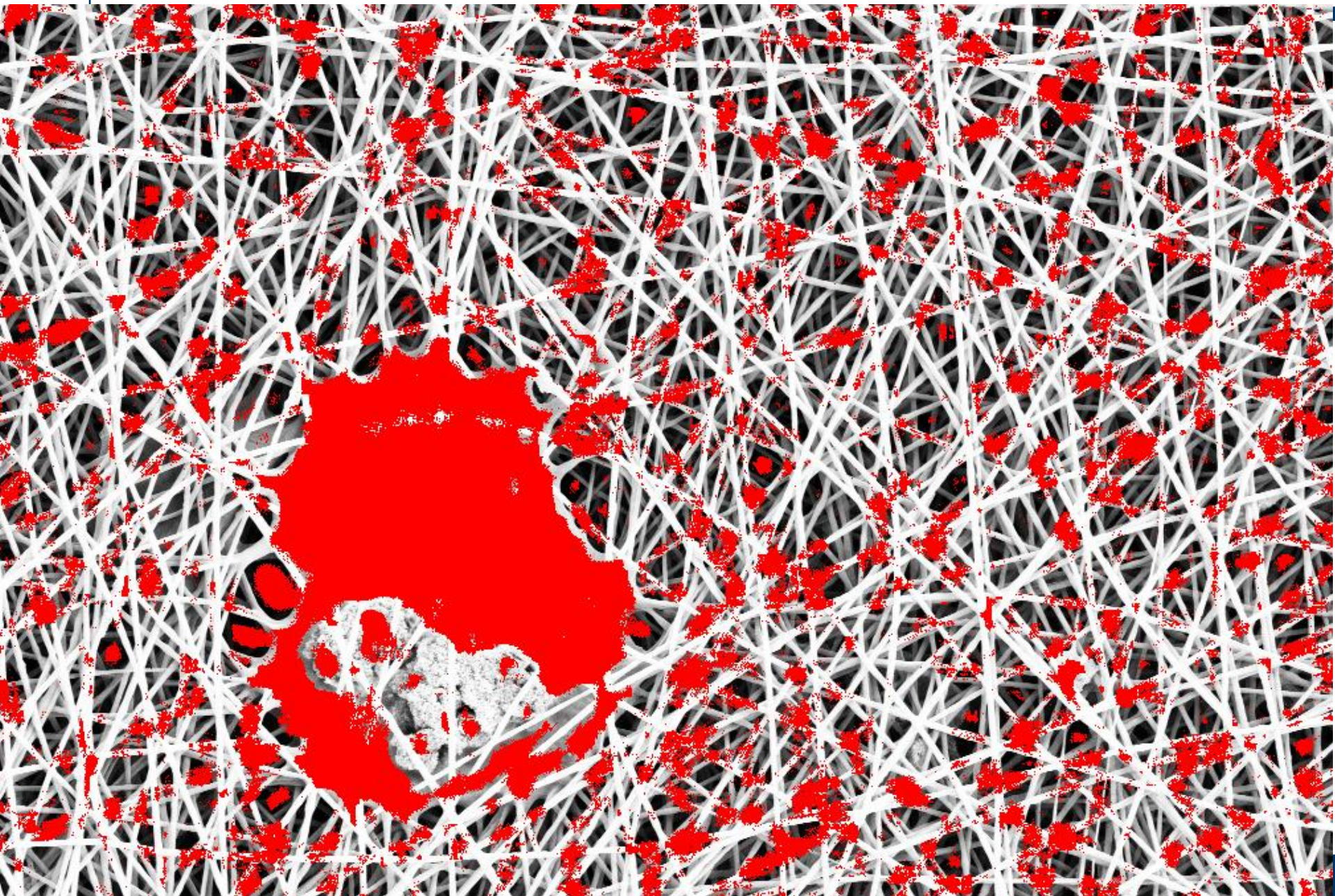
Anomaly detection by means of $g(\cdot)$



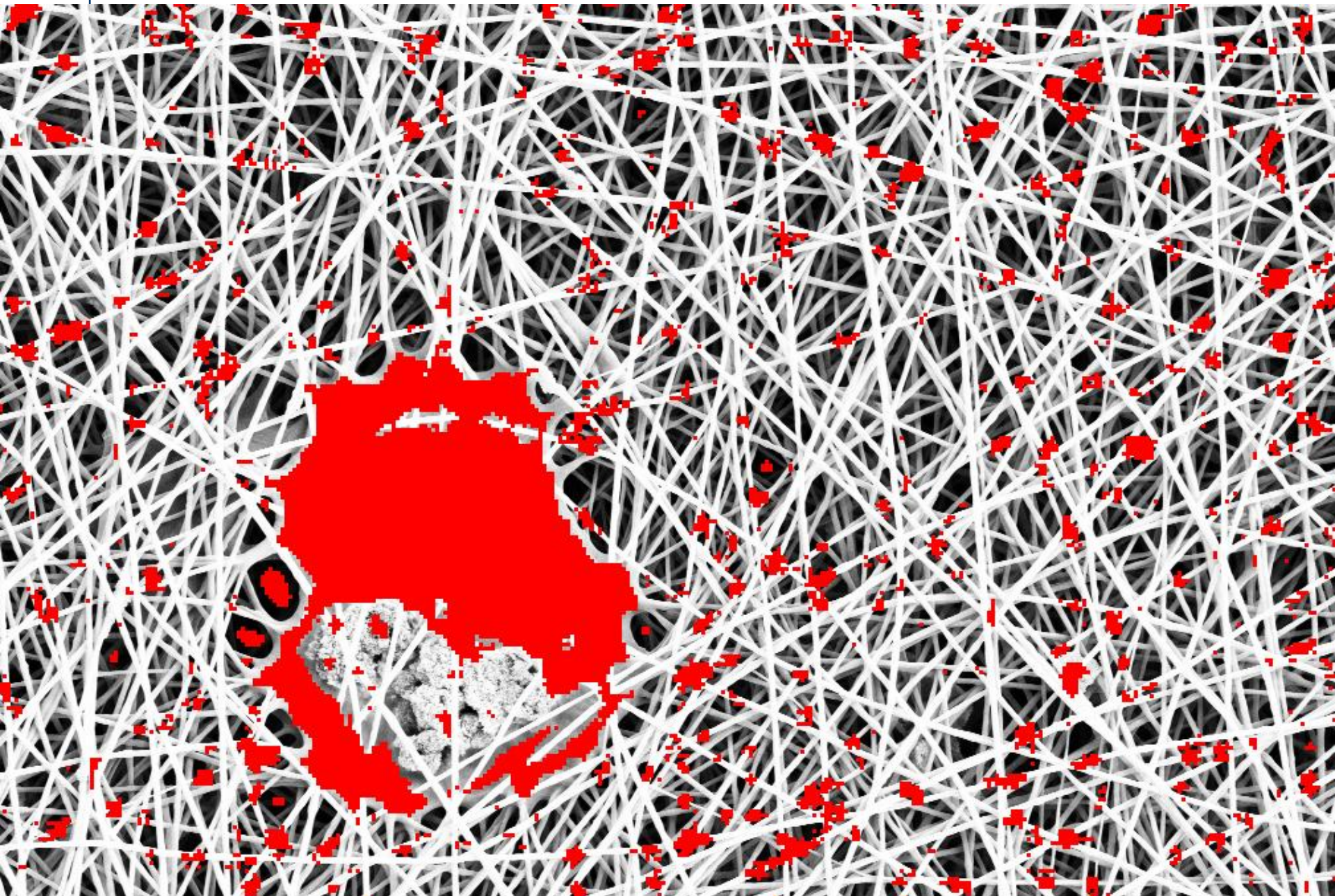
Original Image



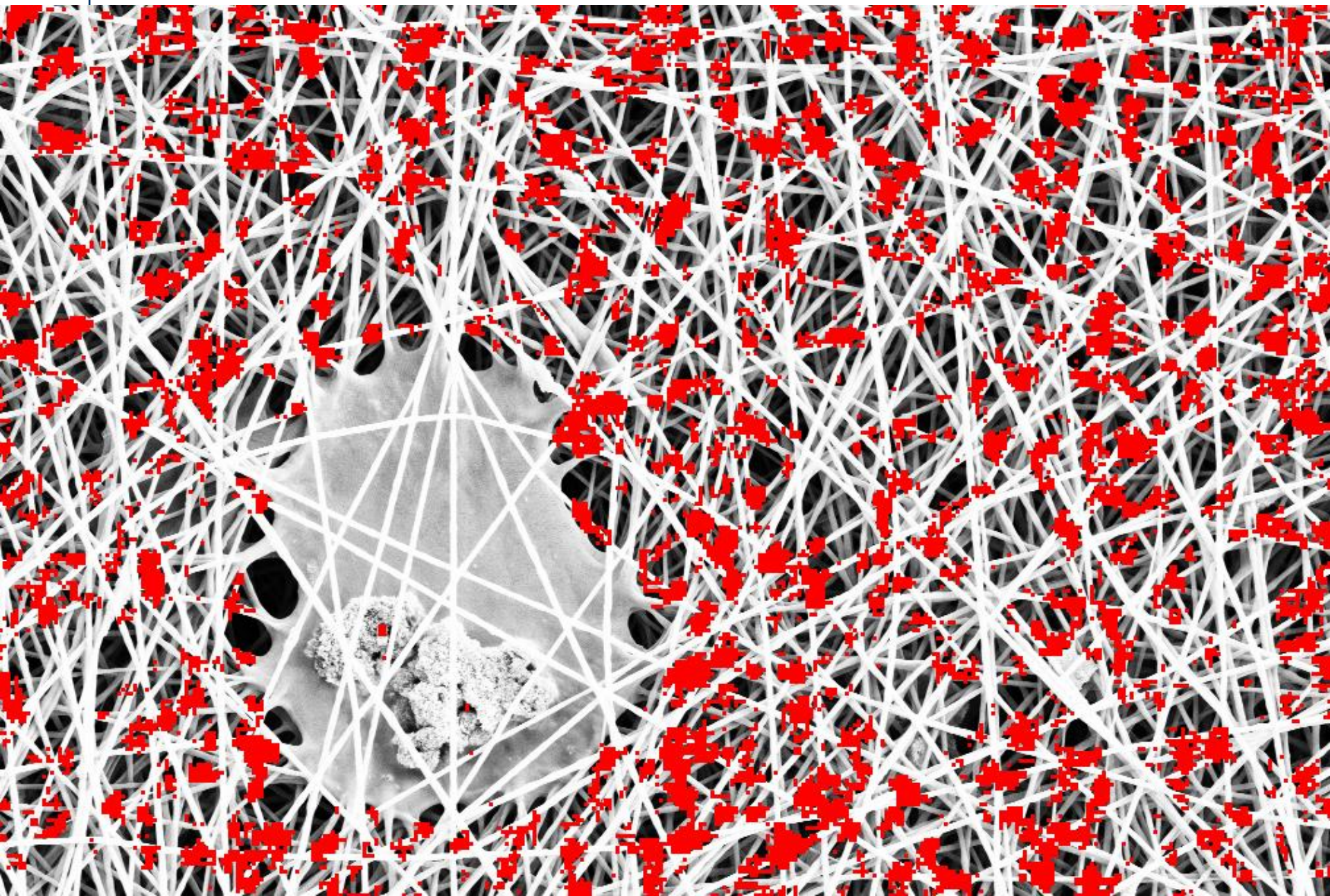
Anomaly detection by means of $e(\cdot)$



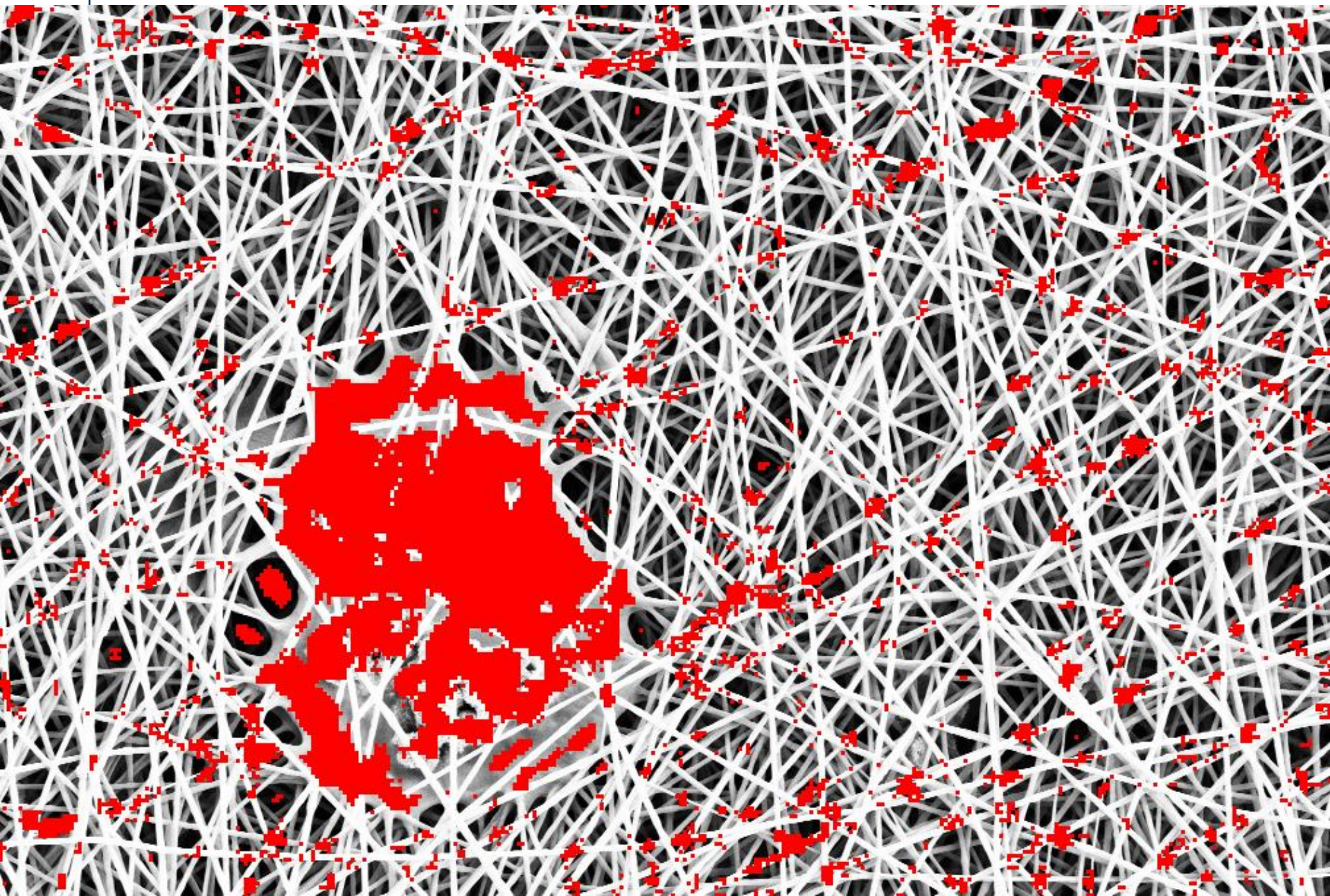
Anomaly detection by means of $f(\cdot)$



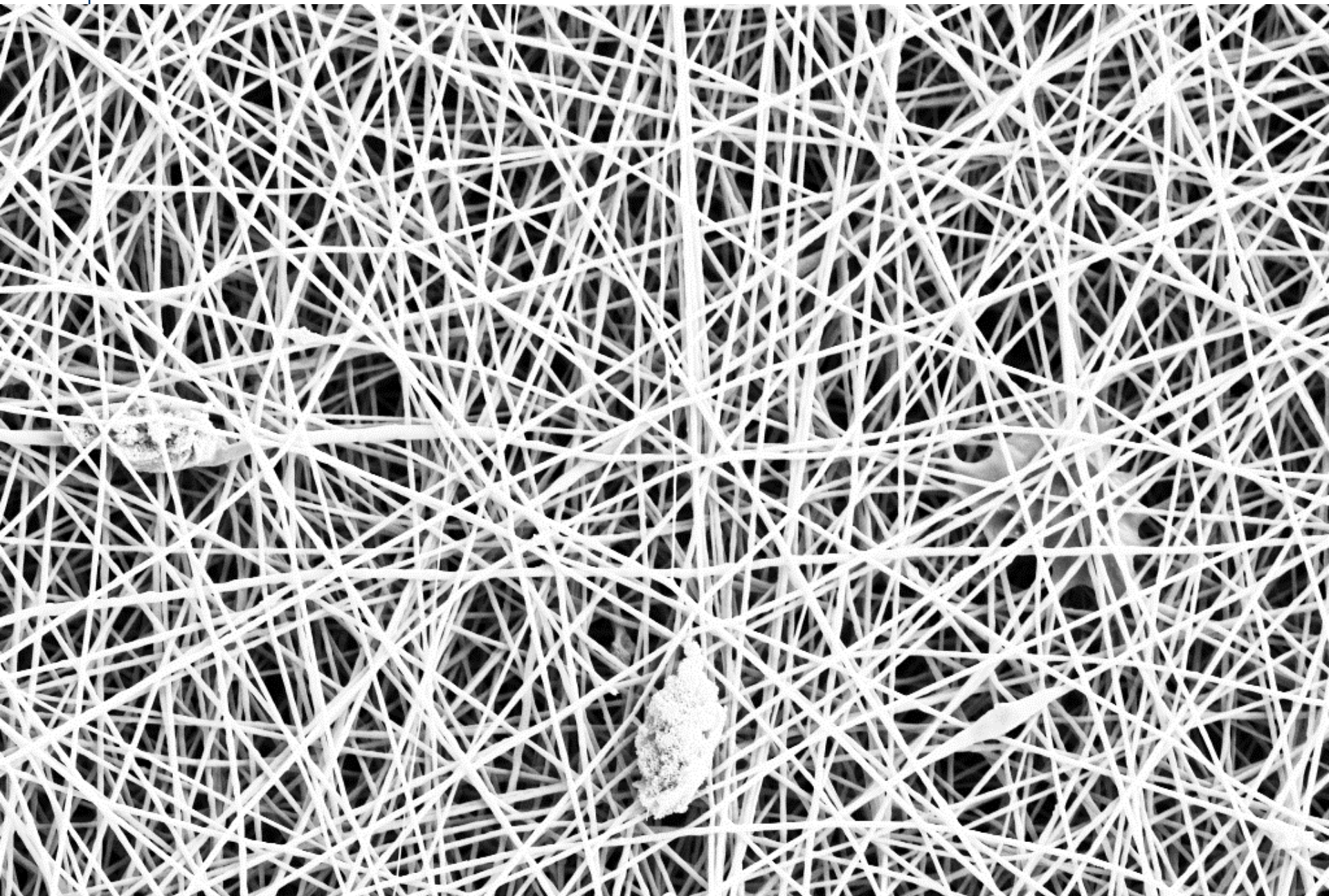
Anomaly detection by means of [Adler 2013]



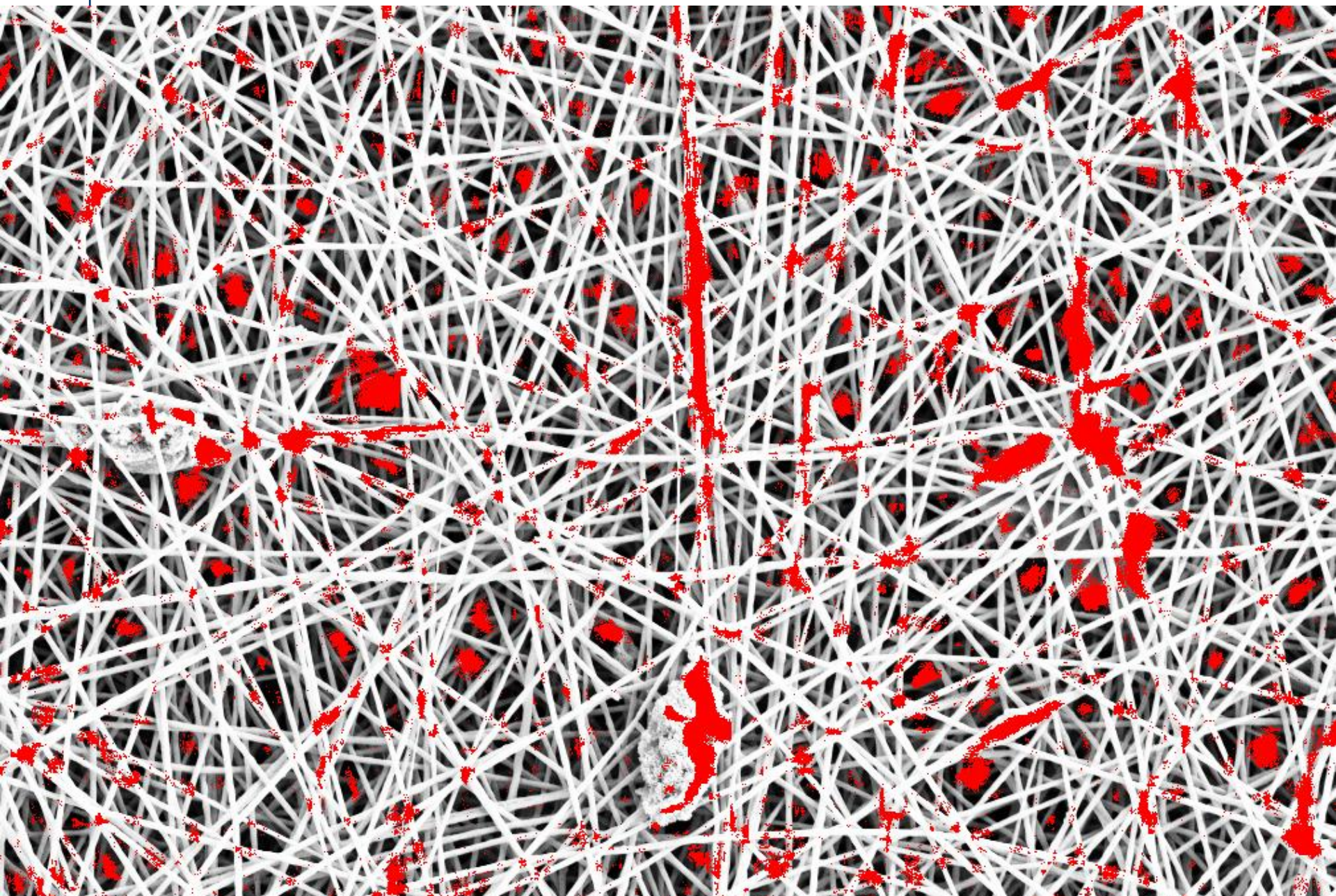
Anomaly detection by means of $g(\cdot)$



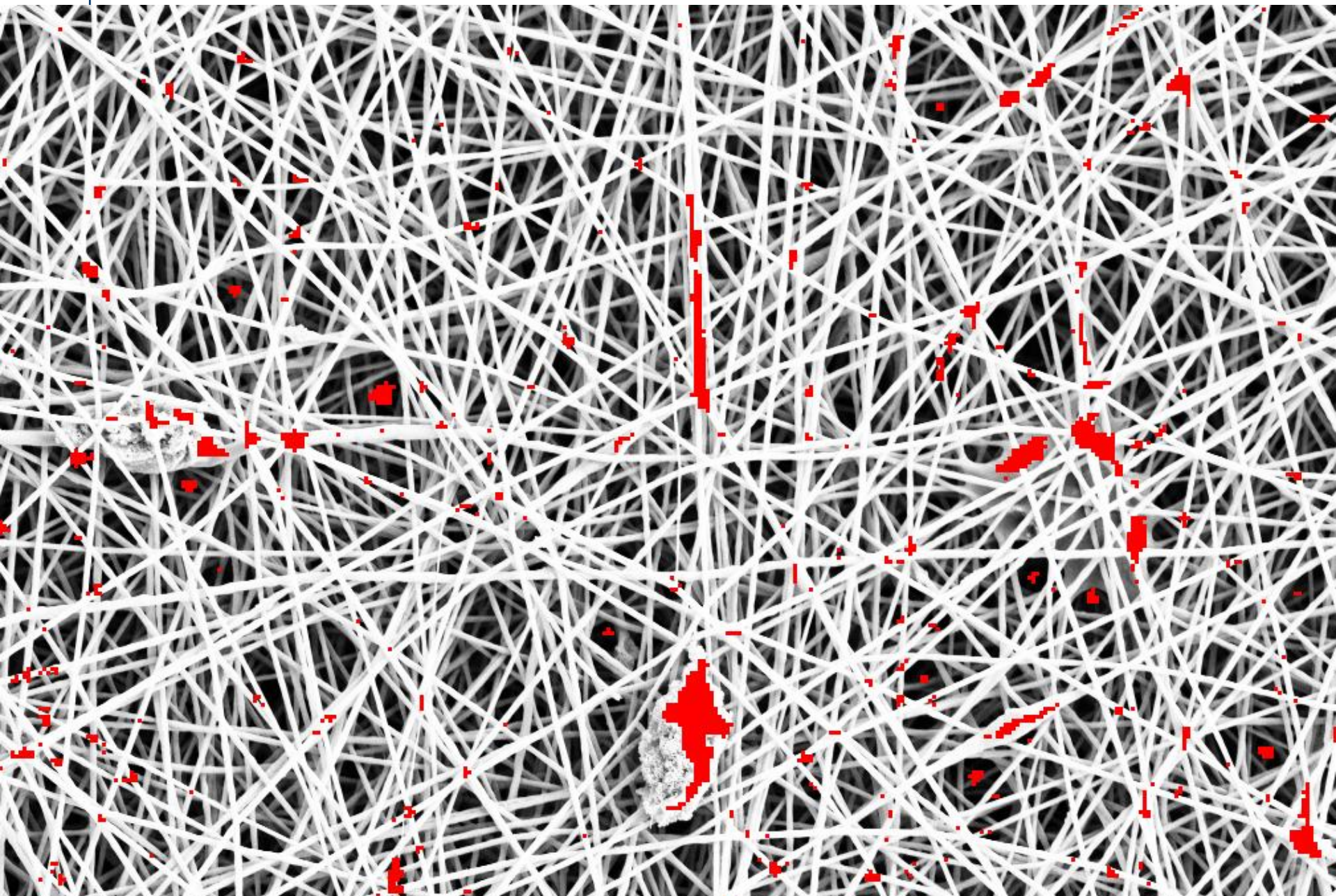
Original Image



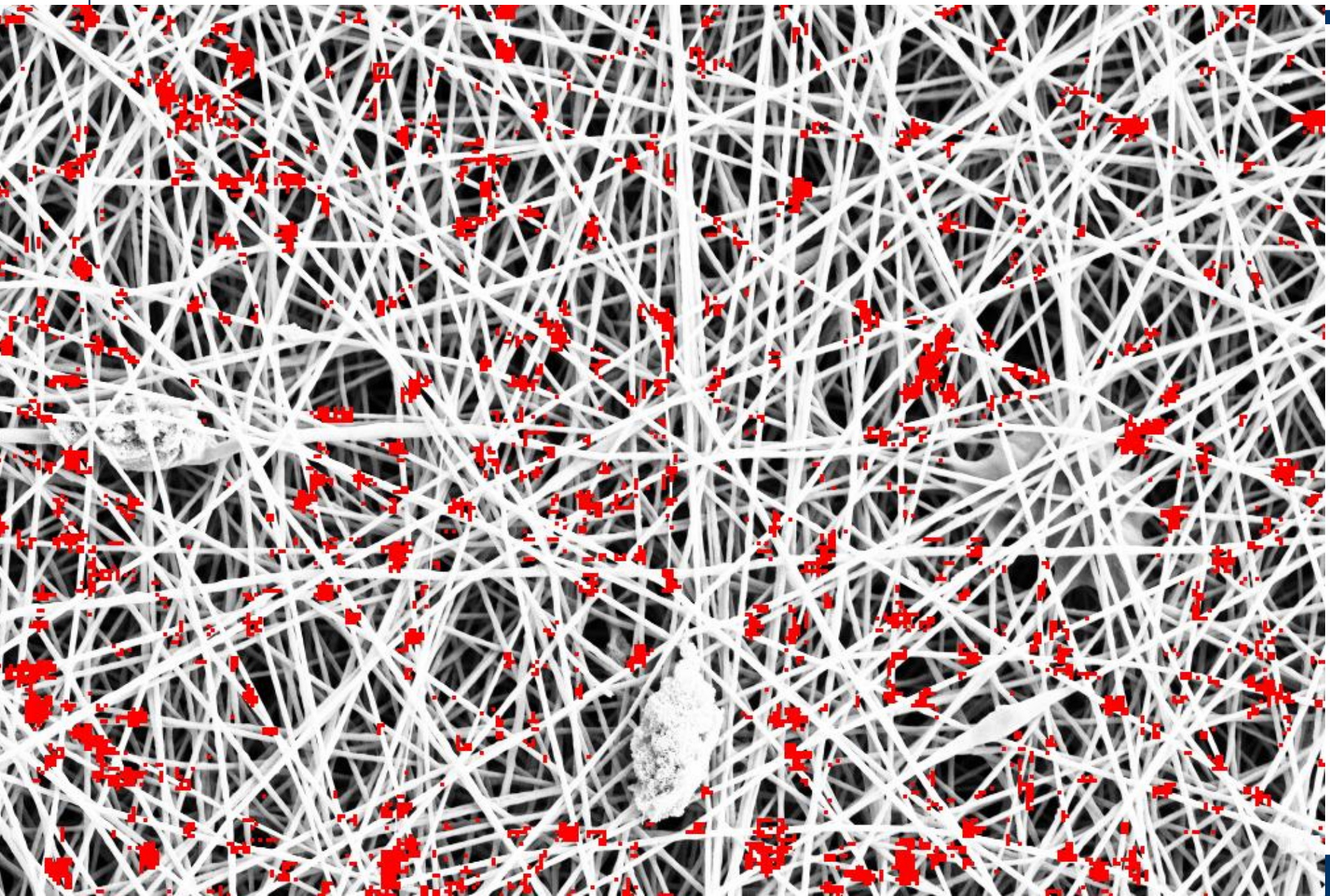
Anomaly detection by means of $e(\cdot)$



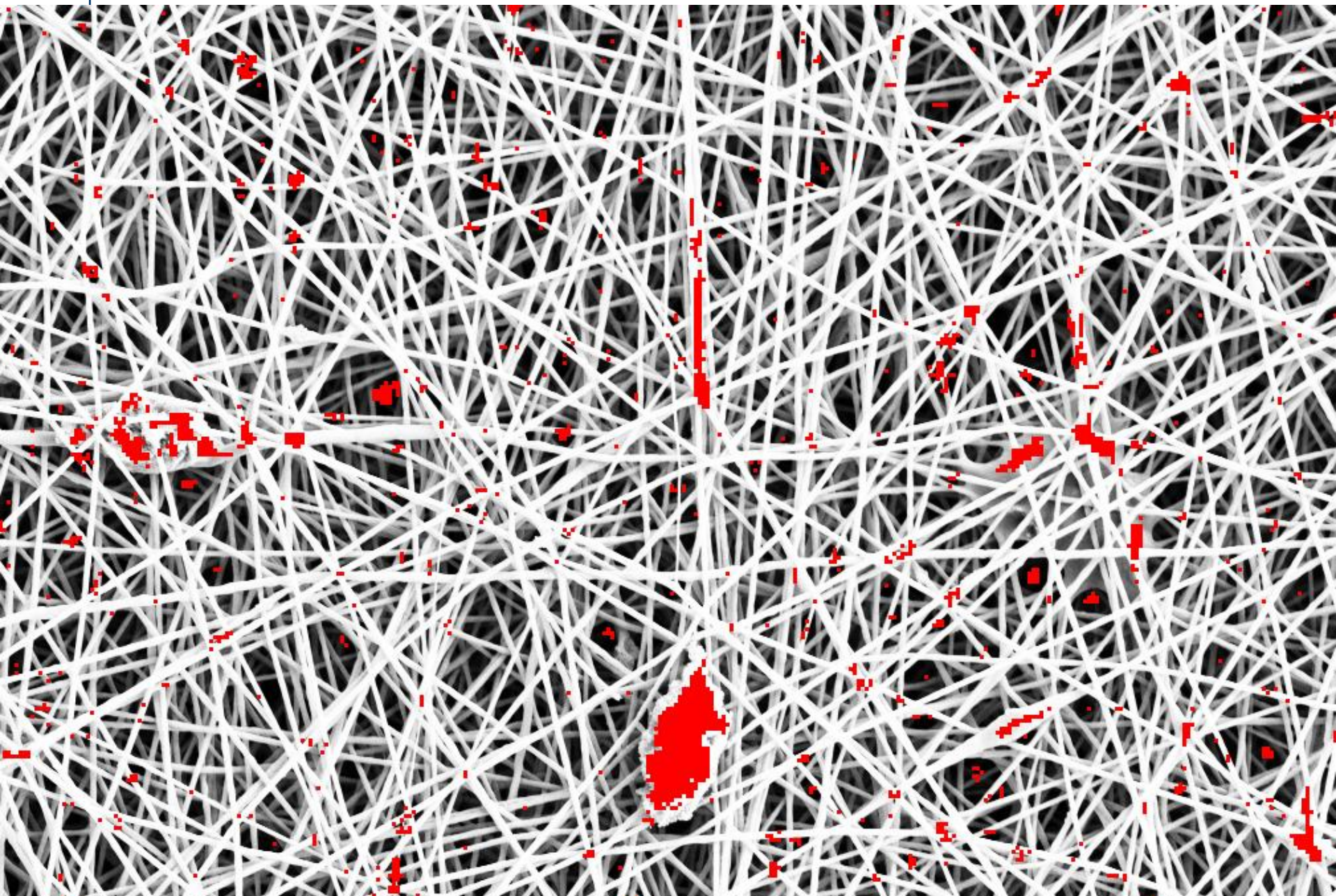
Anomaly detection by means of $f(\cdot)$



Anomaly detection by means of [Adler 2013]



Anomaly detection by means of $g(\cdot)$



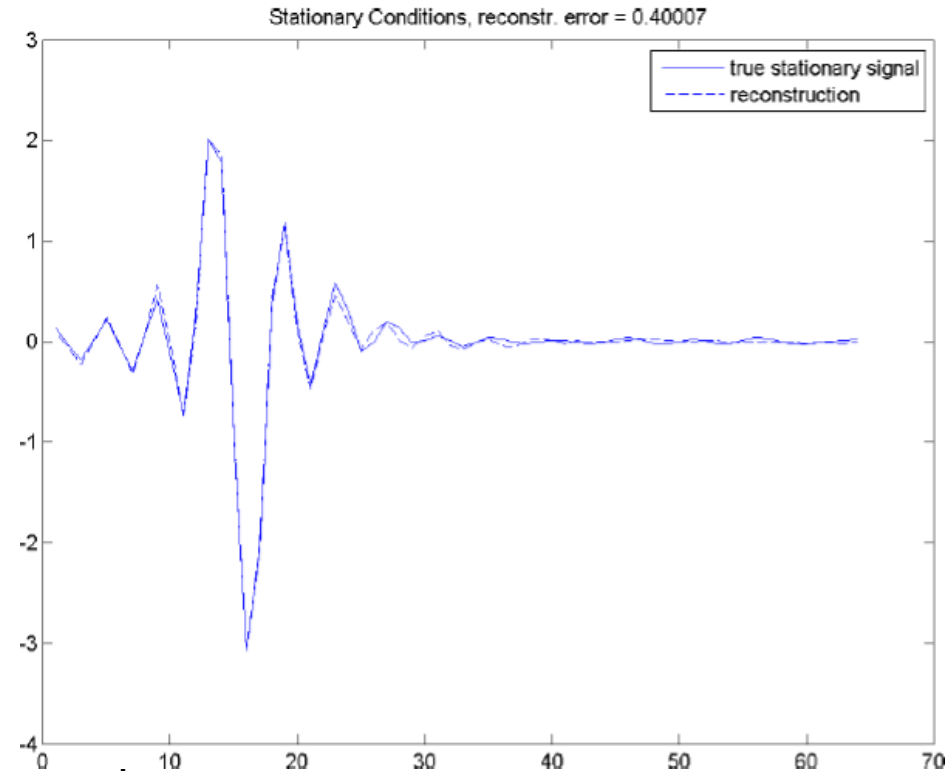
Environmental Monitoring

- We consideres acoustic emissions acquired by an wired/wireless sensor networks meant to monitor a rock faces
- 64 samples signals acquired at 2 KHz by a MEMS.
- Anomalies have been synthetically modified by randomly adding a DB4 wavelet basis atom

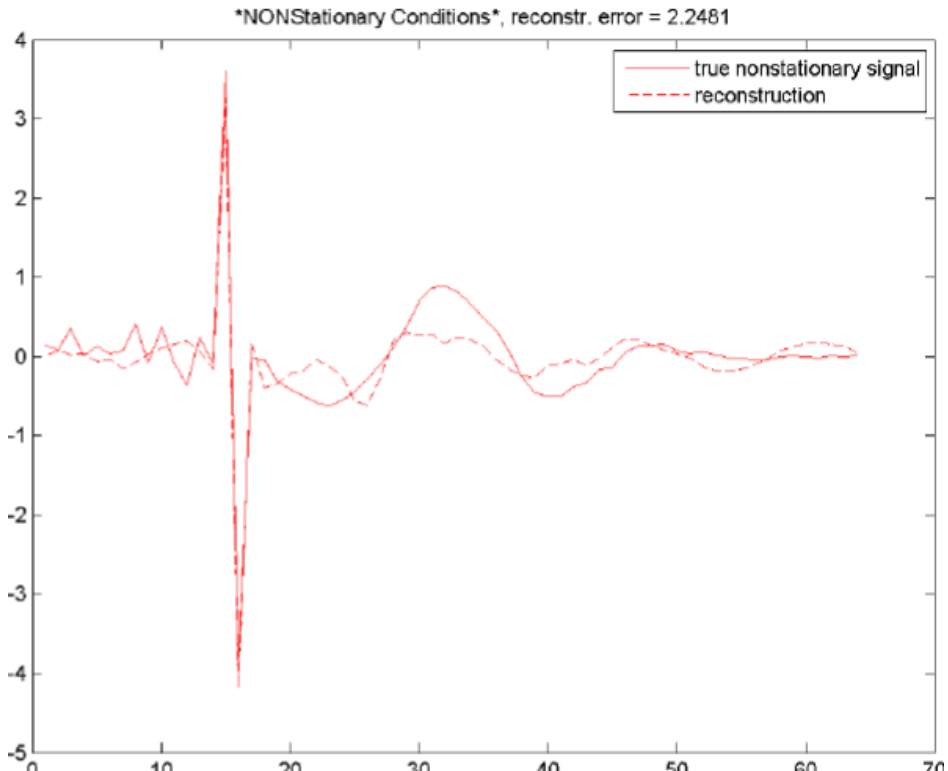
Environmental Monitoring

- We consideres acoustic emissions acquired by an wired/wireless sensor networks meant to monitor a rock faces
- 64 samples signals acquired at 2 KHz by a MEMS.

Example of Original Bursts



Example of Bursts Modified adding atoms from D_1



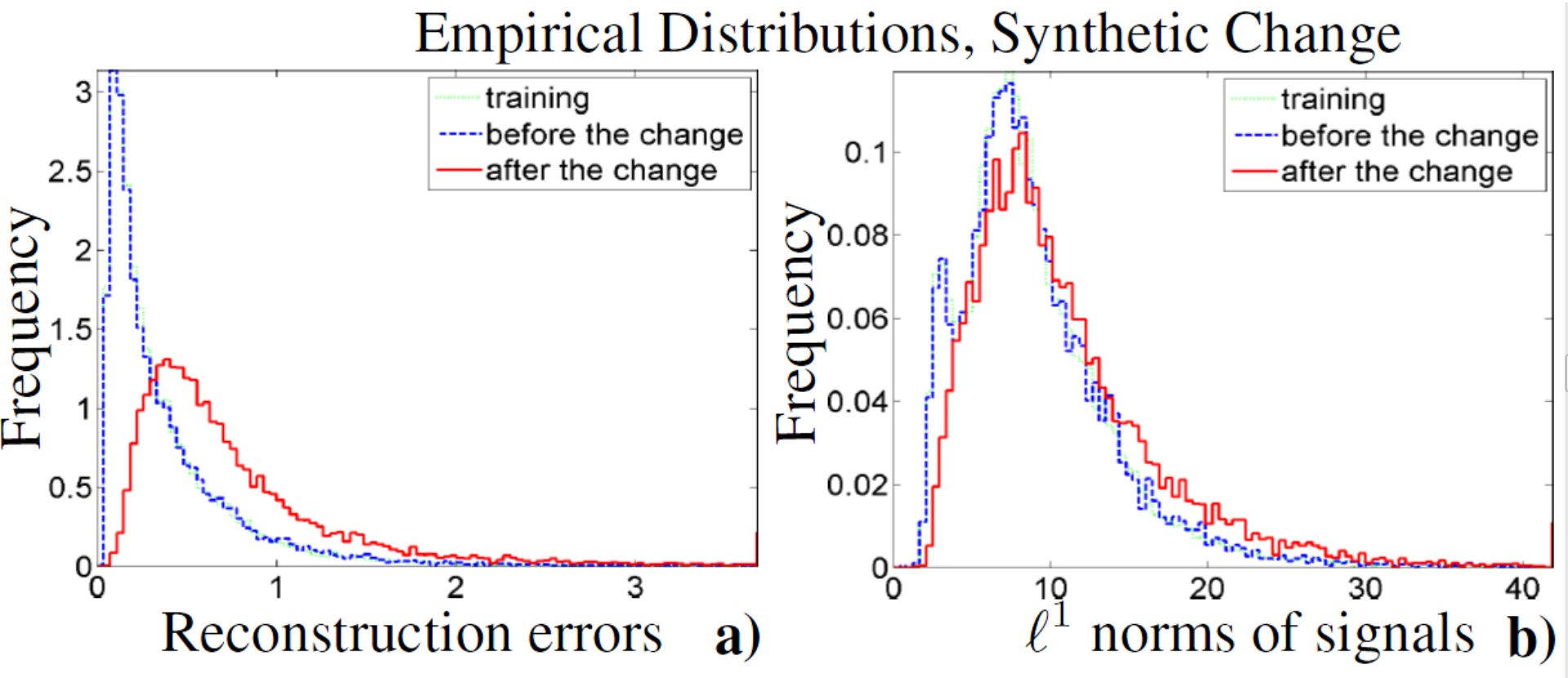
Environmental Monitoring

- We consideres acoustic emissions acquired by an wired/wireless sensor networks meant to monitor a rock faces
- 64 samples signals acquired at 2 KHz by a MEMS.
- Anomalies have been synthetically modified by randomly adding a DB4 wavelet basis atom
- We perform change detection by means of the Lepage CPM using the $e(\cdot)$ change indicator.
- We synthetically generate sequences containing 500 signals before and after the change
- Further details on available in [Alippi 2014]

[Alippi 2014] C. Alippi, G. Boracchi, and B. Wohlberg, "Change detection in streams of signals with sparse representations," ICASSP 2014 , pp. 5252 – 5256.

Distribution of the change indicators

- To show the detectablity of the change we plot the empirical distribution of change indicator before and after the change.
- And compare it with the distirbution of $\|\mathbf{x}_i\|_2$ and $\|\mathbf{x}_i\|_1$



Distribution of the change indicators

- To show the detectablity of the change we plot the empirical distribution of change indicator before and after the change.
- And compare it with the distirbution of $\|\mathbf{x}_i\|_2$ and $\|\mathbf{x}_i\|_1$
- Change-detection performance using CPM are in line with the detectability of the change
  - Using $e(\cdot)$ all the changes are detected with no false positive with an average detection delay of 25 samples
  - Using $\|\mathbf{x_i}\|_1$ delay increased at 124, with 33% of FN
  - Using $\|\mathbf{x_i}\|_2$ no detections

CONCLUDING REMARKS

Ongoing Works

- Our preliminary investigation shows that **sparse representation** allows to build effective models for performing change/anomaly detection
- Sparse representations provide models able to fit data generating processes that in stationary conditions yield heterogenous signals (e.g. belonging to different classes): Atoms of $\widehat{D}$ might be from different *classes.*
- Ongoing works include:
  - the study of customized dictionary learning metods for performing change/anomaly detection
  - the application of the proposed system to other application domains such as EGC analysis to detect arrhythmia.