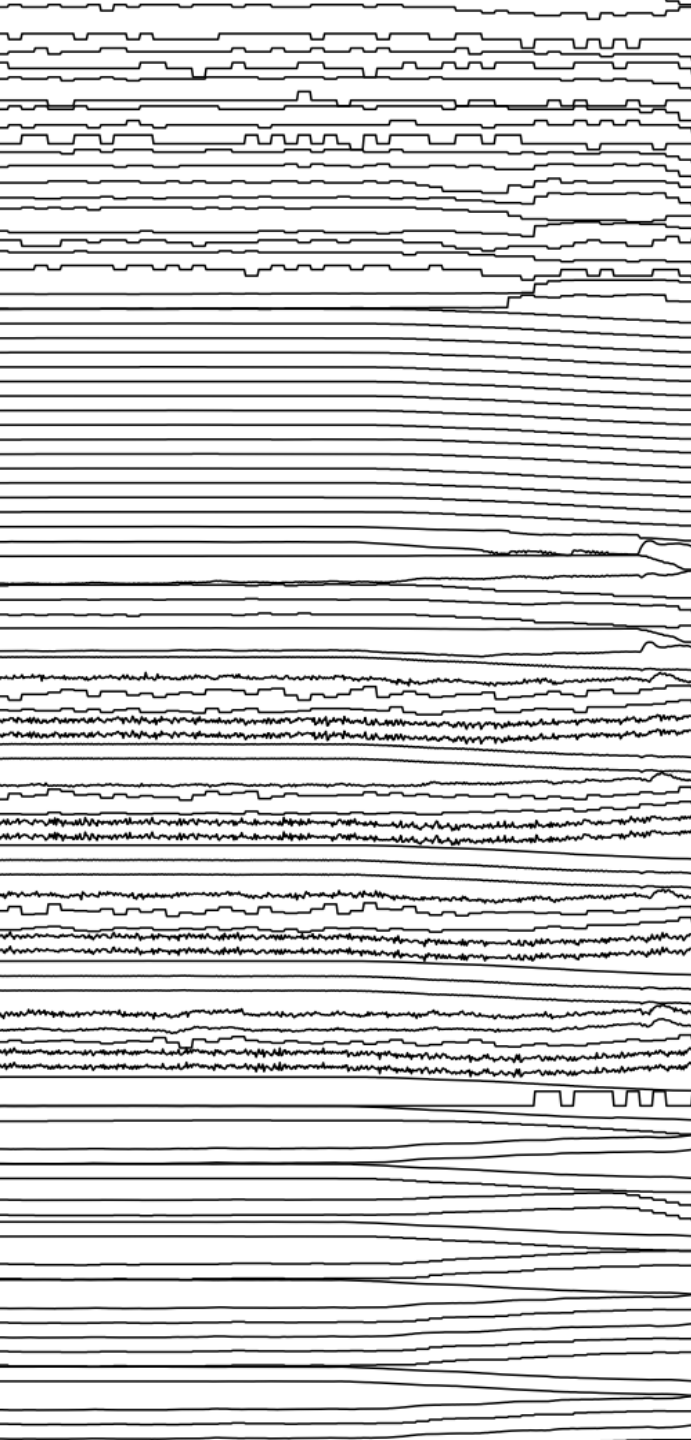




Time-Series Anomaly Detection: Overview and New Trends

Qinghua Liu
The Ohio State University
liu.11085@osu.edu

Paul Boniol
Inria, ENS, PSL University
paul.boniol@inria.fr

Themis Palpanas
Université Paris Cité; IUF
themis@mi.parisdescartes.fr

John Paparrizos
The Ohio State University
paparrizos.1@osu.edu



THE OHIO STATE UNIVERSITY

Introduction: *Time series are Everywhere*

Energy Production



Edf.fr: tinyurl.com/yc7x5xje

Astrophysics



Virgo: <https://www.virgo-gw.eu/>

Medicine



tinyurl.com/39dx2us4

Volcanology

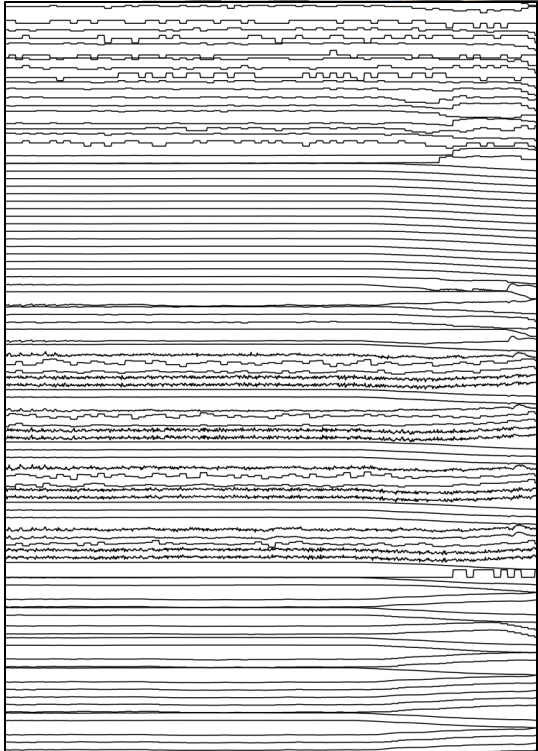


tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

Energy Production

Secondary circuit sensor
measurements



Astrophysics



Virgo: <https://www.virgo-gw.eu/>

Medicine



tinyurl.com/39dx2us4

Volcanology

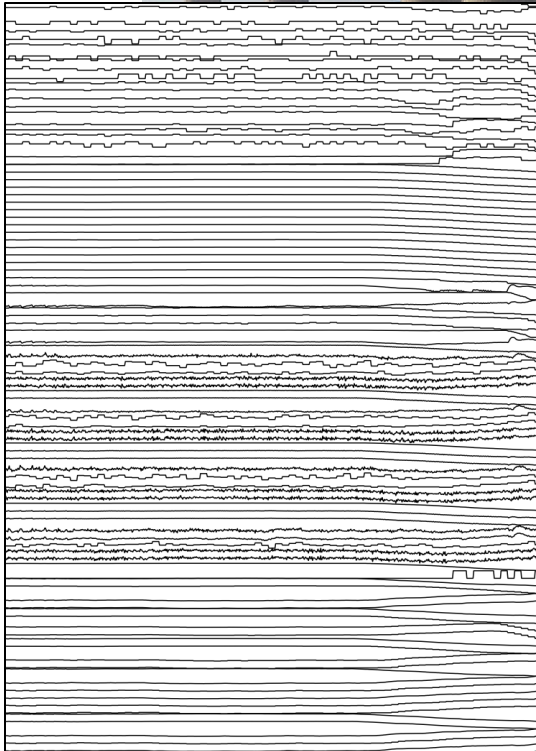


tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

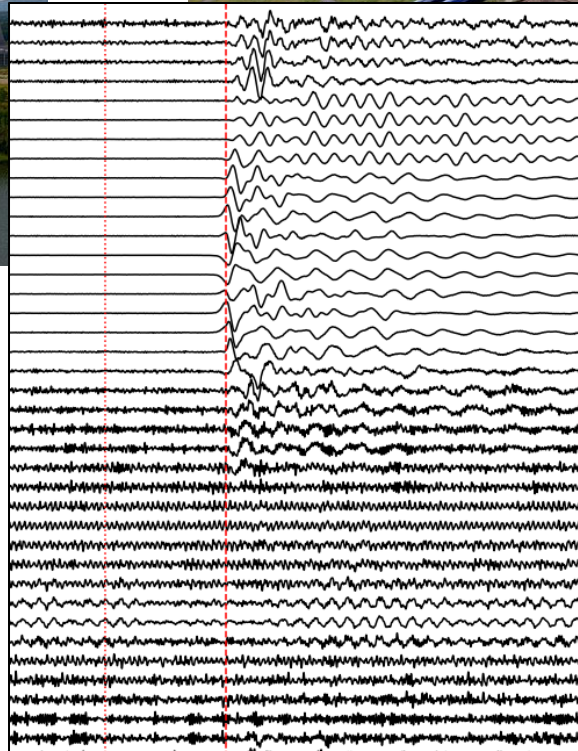
Energy Production

Secondary circuit sensor measurements



Astrophysics

Fiber-acoustic sensors in the VIRGO north building



Medicine



tinyurl.com/39dx2us4

Volcanology

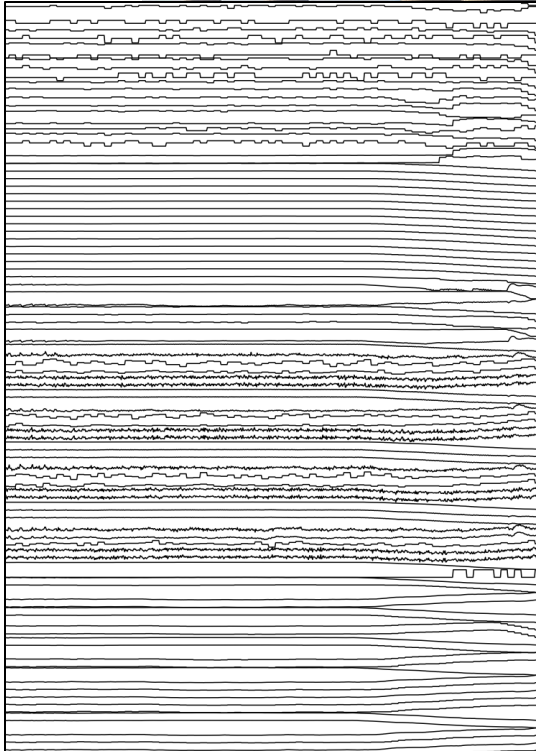


tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

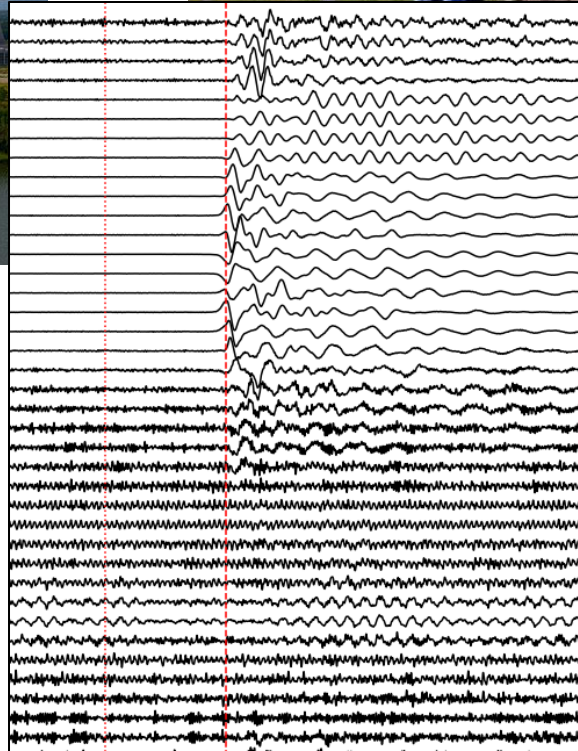
Energy Production

Secondary circuit sensor measurements



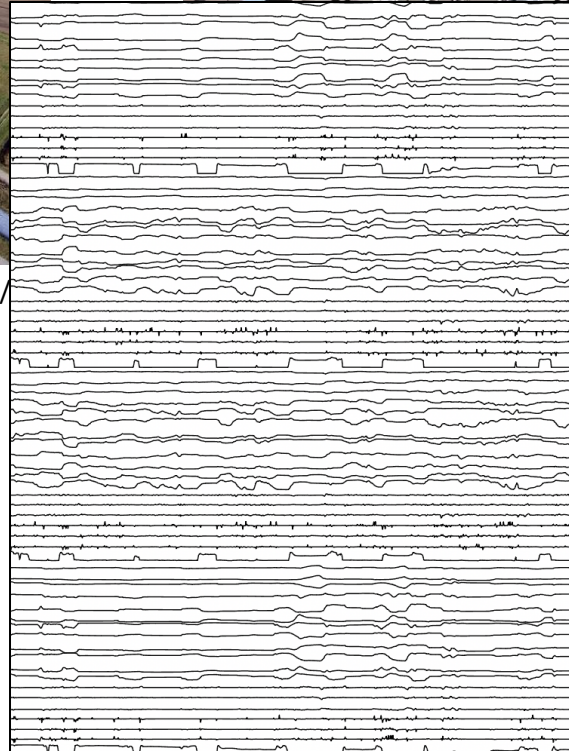
Astrophysics

Fiber-acoustic sensors in the VIRGO north building



Medicine

Sensor measurements of the Da Vinci surgery robot



Volcanology

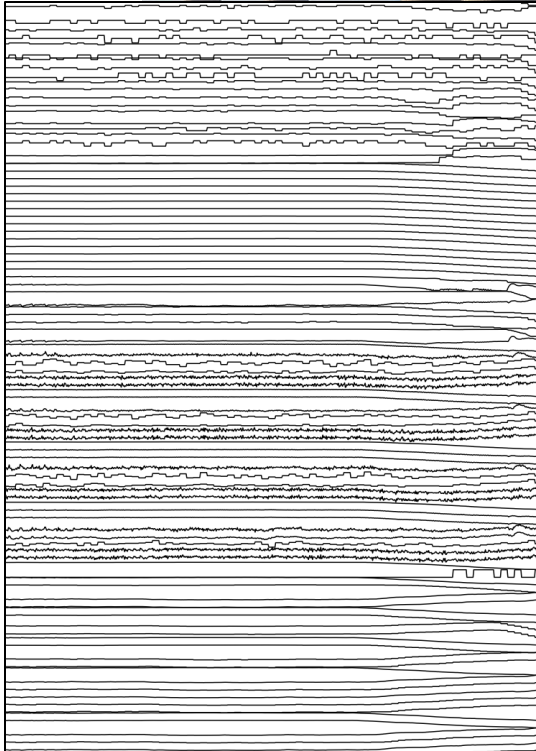


tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

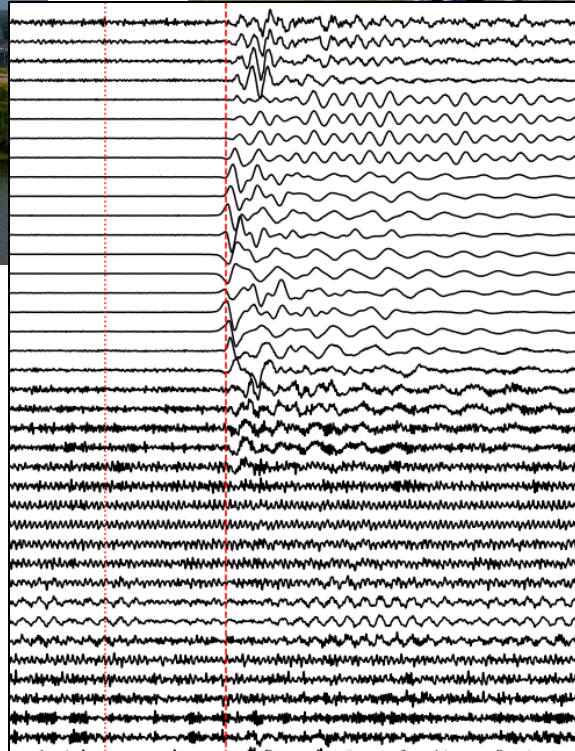
Energy Production

Secondary circuit sensor measurements



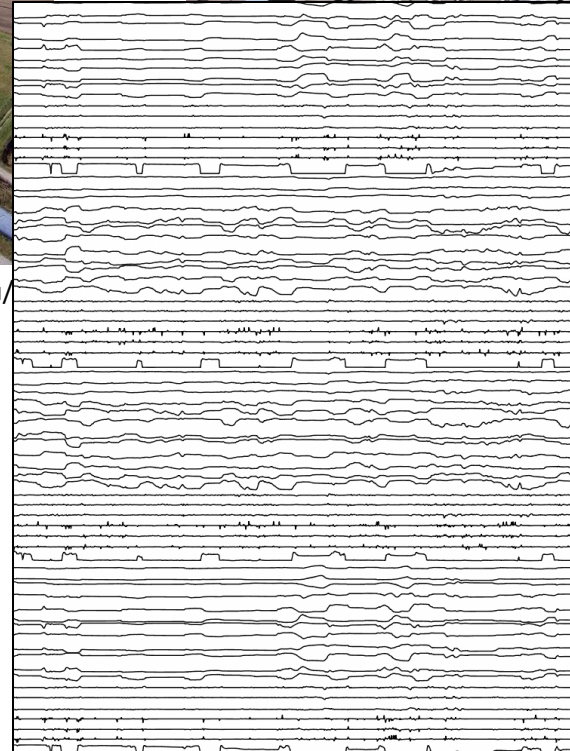
Astrophysics

Fiber-acoustic sensors in the VIRGO north building



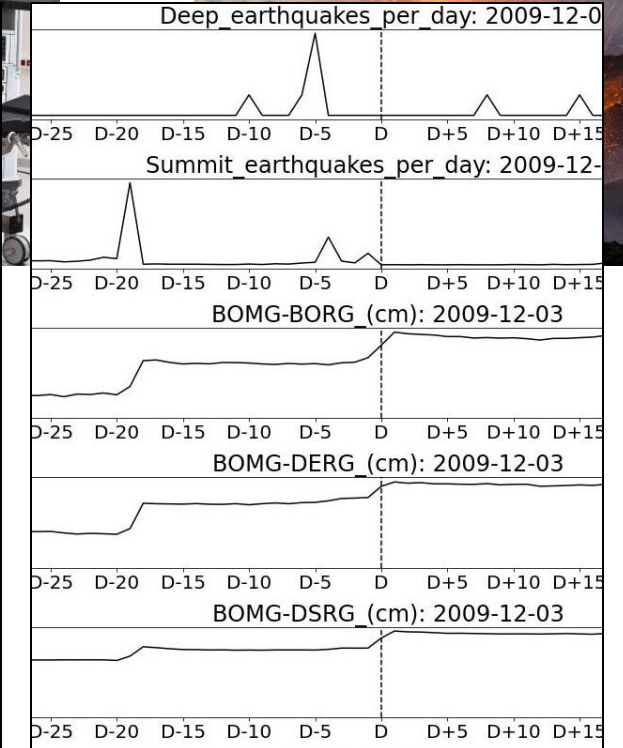
Medicine

Sensor measurements of the Da Vinci surgery robot



Volcanology

Sensor measurements on le Piton de la Fournaise

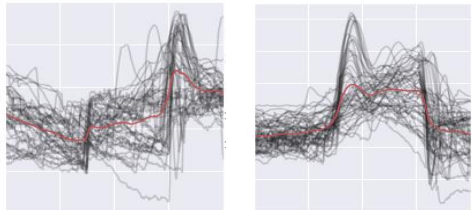


Introduction: *with Important Challenges*

Energy Production

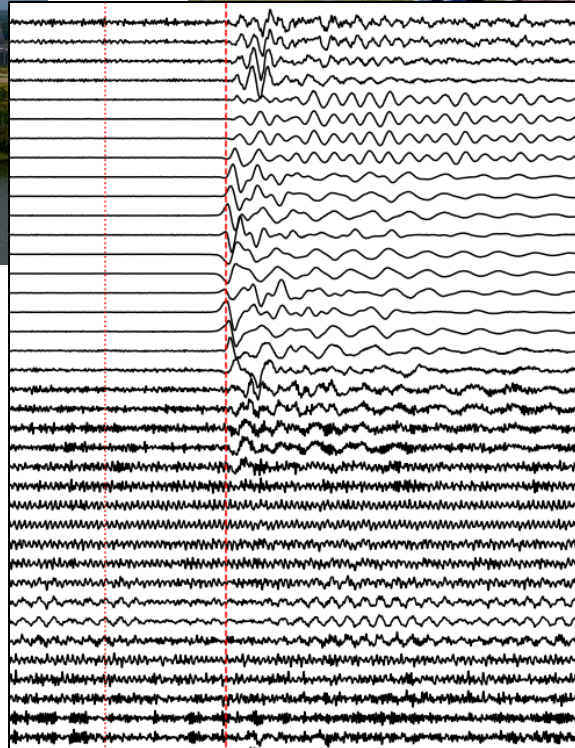
Secondary circuit sensor measurements

Identification of precursors of feed-water pumps vibrations



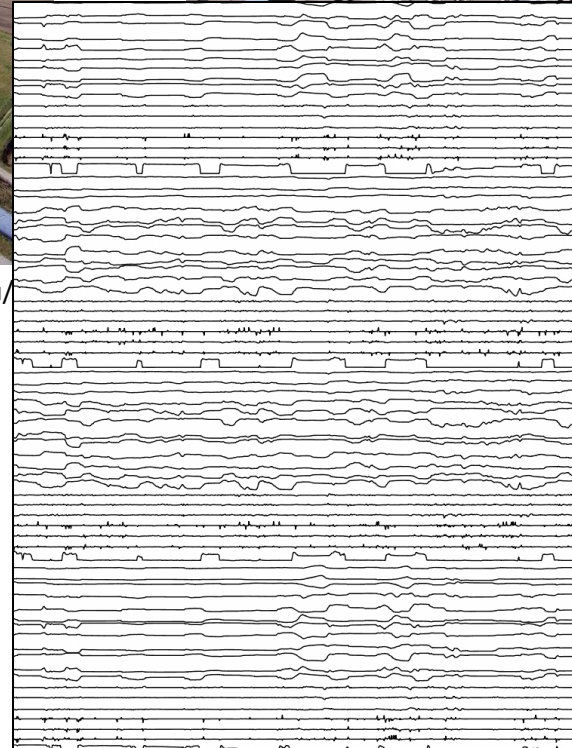
Astrophysics

Fiber-acoustic sensors in the VIRGO north building



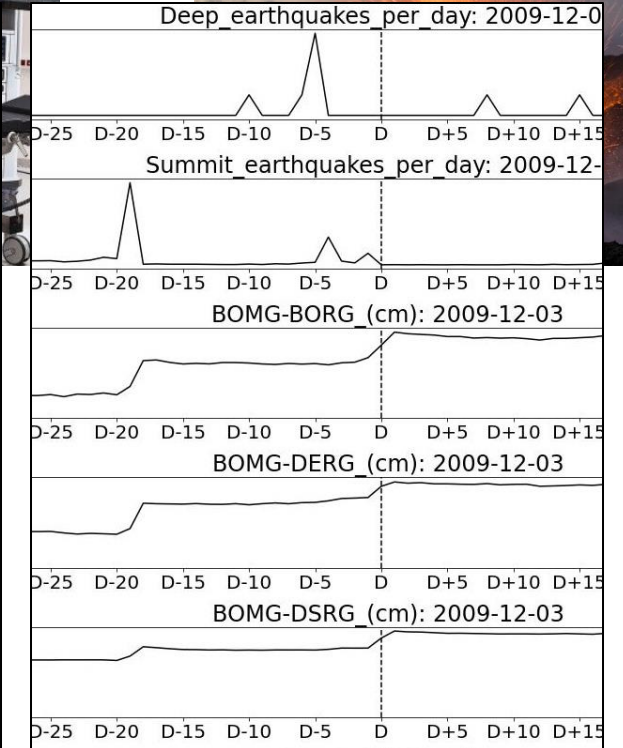
Medicine

Sensor measurements of the Da Vinci surgery robot



Volcanology

Sensor measurements on le Piton de la Fournaise

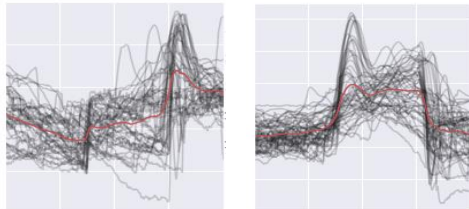


Introduction: *with Important Challenges*

Energy Production

Secondary circuit sensor measurements

Identification of precursors of feed-water pumps vibrations



Astrophysics

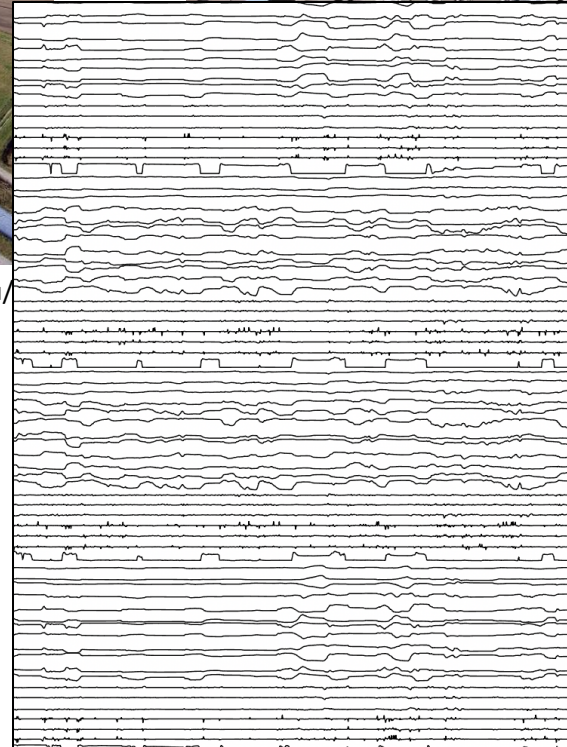
Fiber-acoustic sensors in the VIRGO north building

Noise detection in VIRGO interferometer north building



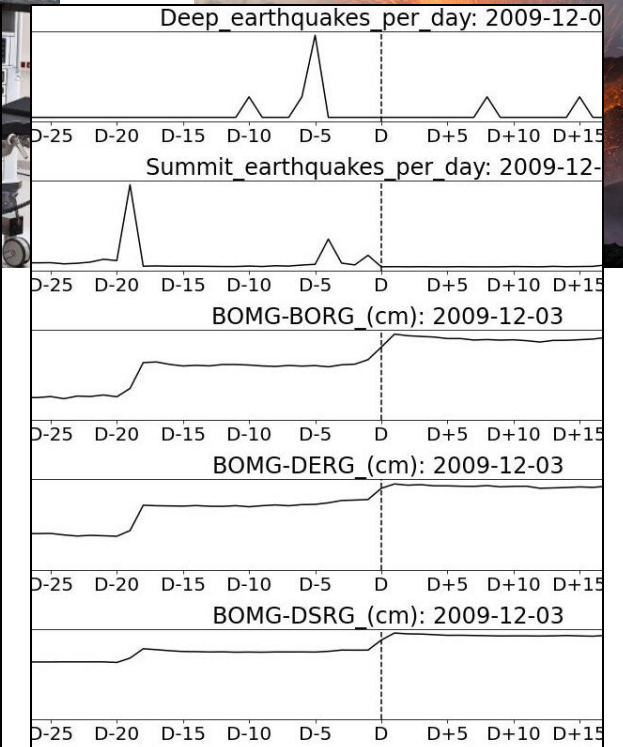
Medicine

Sensor measurements of the Da-Vinci surgery robot



Volcanology

Sensor measurements on le Piton de la Fournaise

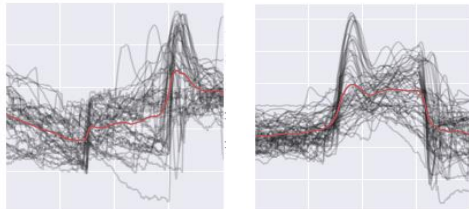


Introduction: *with Important Challenges*

Energy Production

Secondary circuit sensor measurements

Identification of precursors of feed-water pumps vibrations



Astrophysics

Fiber-acoustic sensors in the VIRGO north building

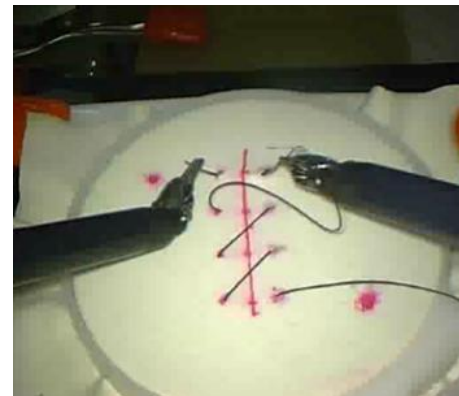
Noise detection in VIRGO interferometer north building



Medicine

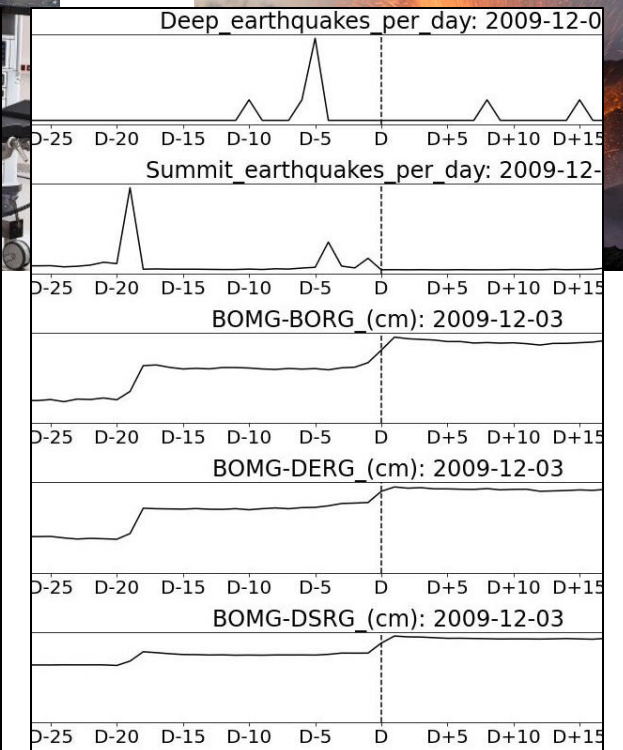
Sensor measurements of the Da-Vinci surgery robot

Unusual surgeons gestures detection



Volcanology

Sensor measurements on le Piton de la Fournaise

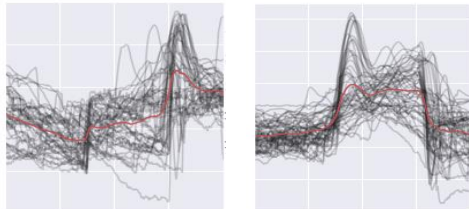


Introduction: *with Important Challenges*

Energy Production

Secondary circuit sensor measurements

Identification of precursors of feed-water pumps vibrations



Astrophysics

Fiber-acoustic sensors in the VIRGO north building

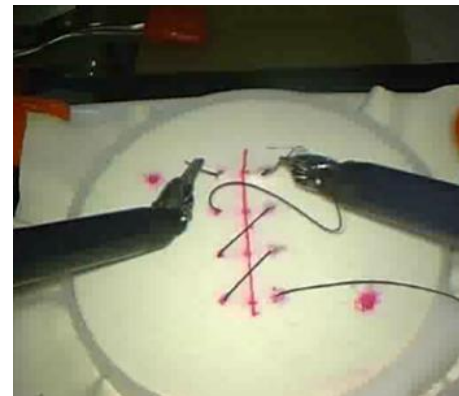
Noise detection in VIRGO interferometer north building



Medicine

Sensor measurements of the Da Vinci surgery robot

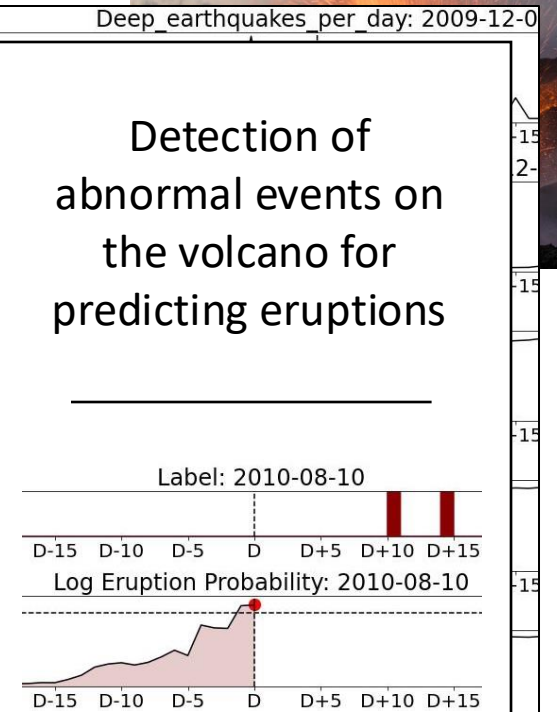
Unusual surgeons gestures detection



Volcanology

Sensor measurements on le Piton de la Fournaise

Detection of abnormal events on the volcano for predicting eruptions



Introduction: *with Important Challenges*

Large-scale time series database

Energy Production

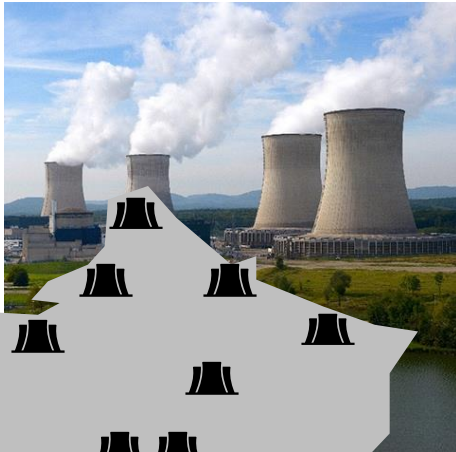


Edf.fr: tinyurl.com/yc7x5xje

Introduction: *with Important Challenges*

Large-scale time series database

Energy Production



Edf.fr: tinyurl.com/yc7x5xje

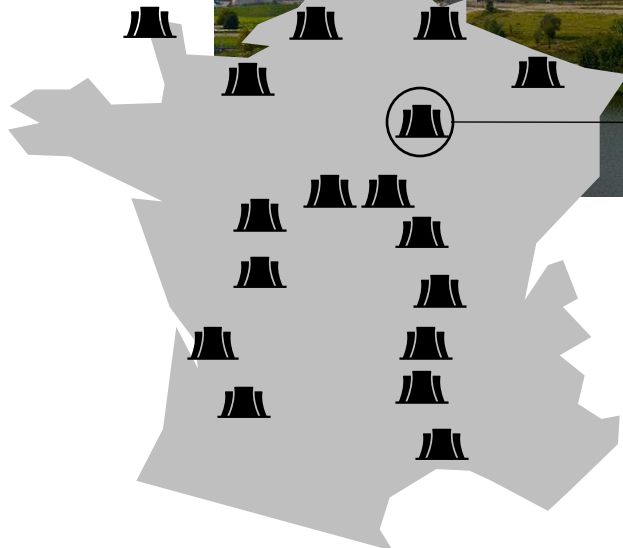
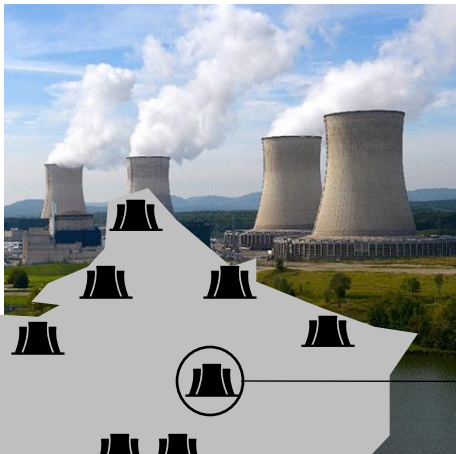
Example of Nuclear production

- 58 nuclear power plants across France

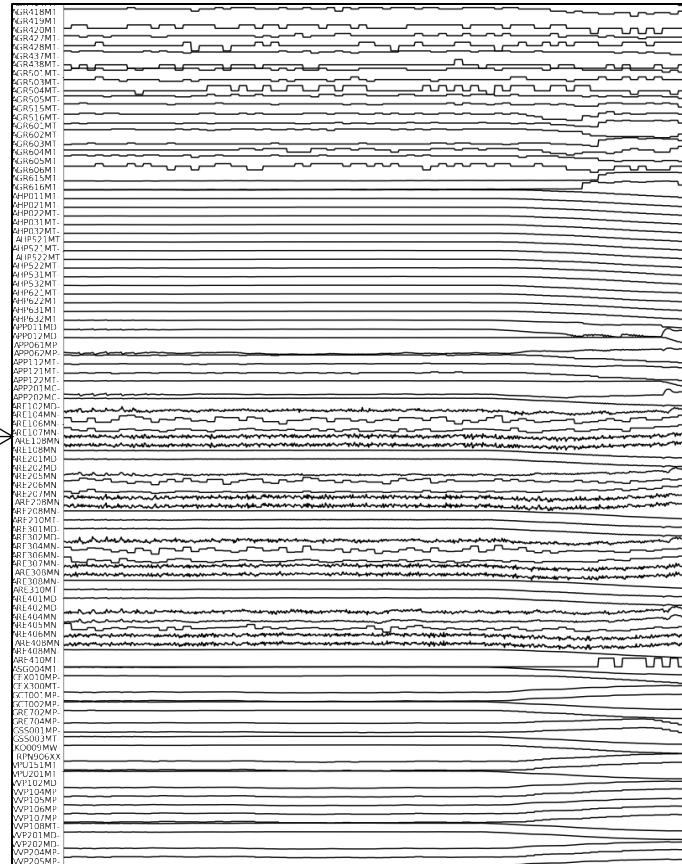
Introduction: *with Important Challenges*

Large-scale time series database

Energy Production



Edf.fr: tinyurl.com/yc7x5xje



Example of Nuclear production

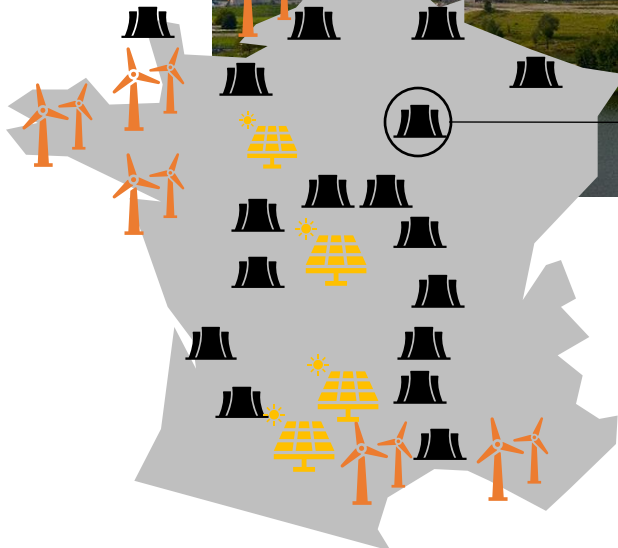
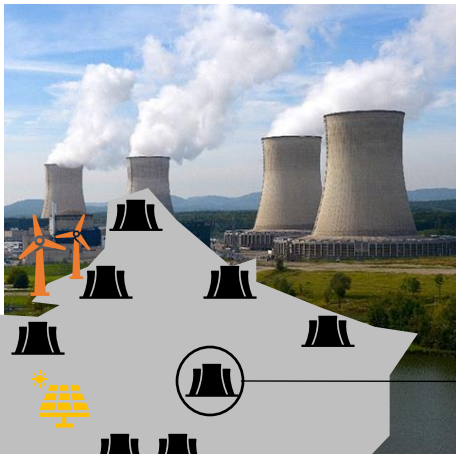
- 58 nuclear power plants across France
- 2000+ sensors per power plant
- 30 years of data collections

A total of 1.5 PetaBytes

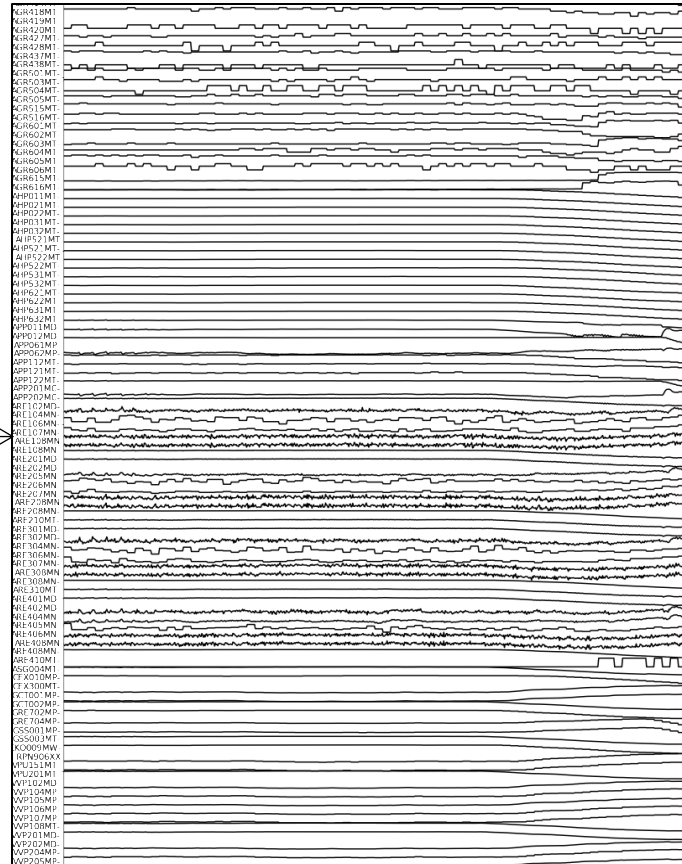
Introduction: *with Important Challenges*

Large-scale time series database

Energy Production



Edf.fr: tinyurl.com/yc7x5xje



Example of Nuclear production

- 58 nuclear power plants across France
- 2000+ sensors per power plant
- 30 years of data collections

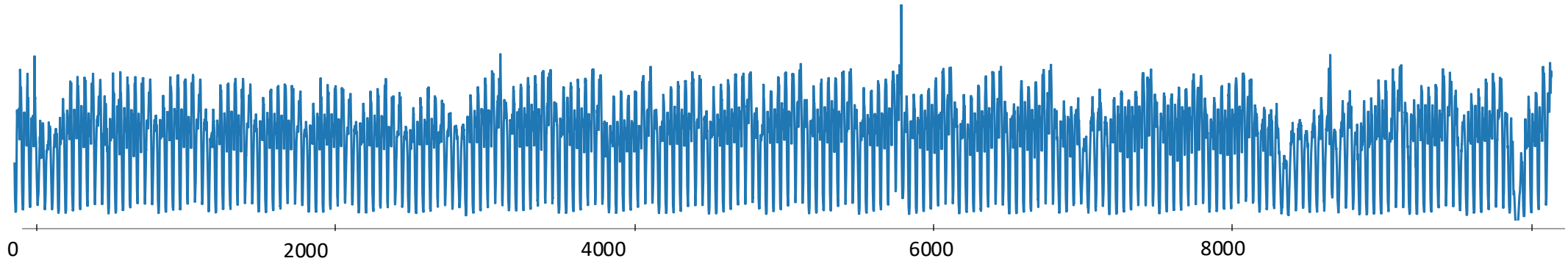
A total of 1.5 PetaBytes

Other source of production

- New sensors with higher acquisition rate

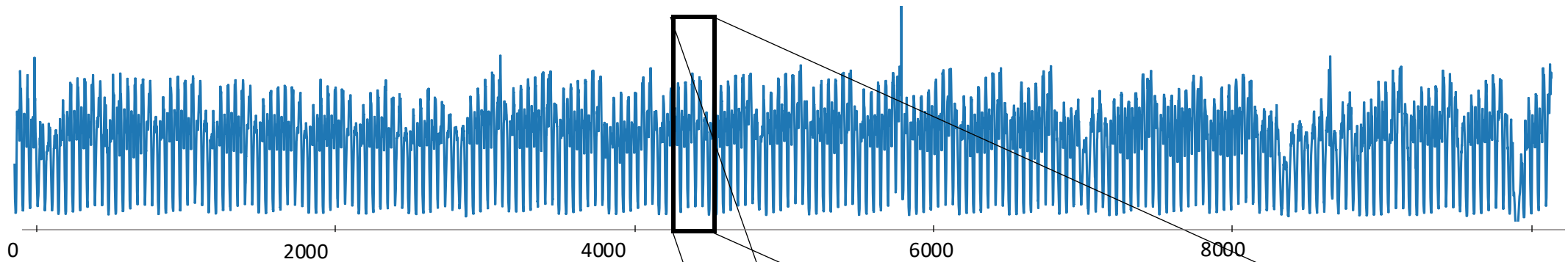
Introduction: *Anomaly Detection in Time Series*

- Time series T (example : number of taxi passengers in New York City)

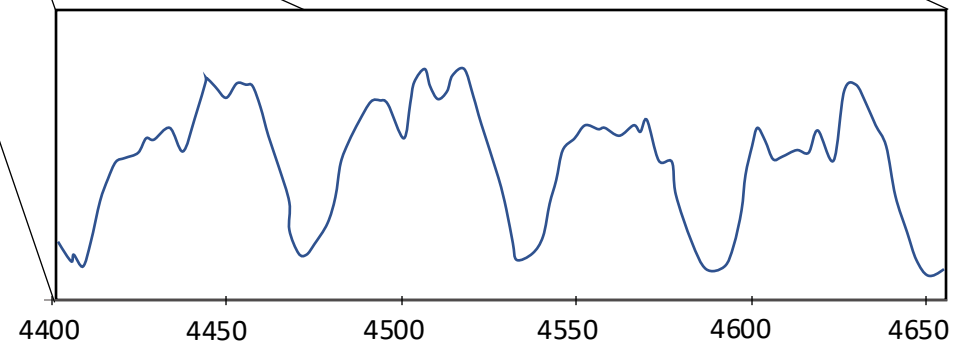


Introduction: *Anomaly Detection in Time Series*

- Time series T (example : number of taxi passengers in New York City)

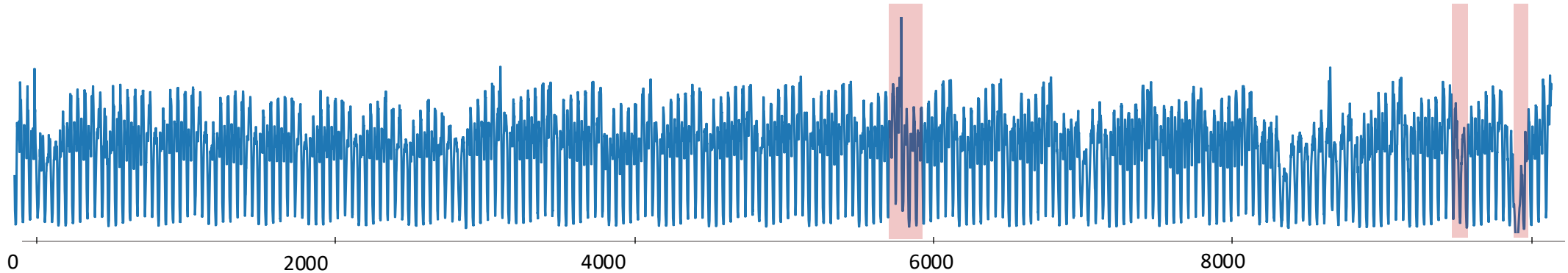


- Subsequence $T_{i,\ell}$
with $i = 4400, \ell = 250$



Introduction: *Anomaly Detection in Time Series*

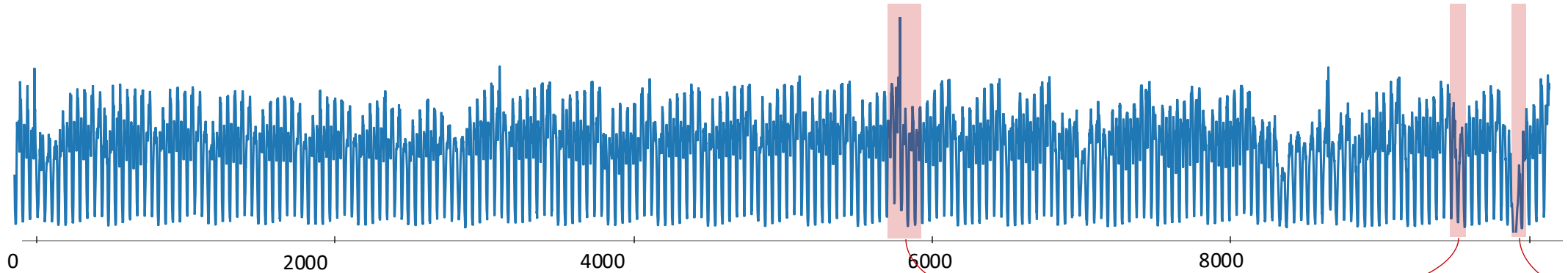
- Time series T (example : number of taxi passengers in New York City)



- Anomaly: *rare* point or sequence (of a given length)
potentially *non-desired*

Introduction: *Anomaly Detection in Time Series*

- Time series T (example : number of taxi passengers in New York City)



- Anomaly: *rare* point or sequence (of a given length) potentially *non-desired*

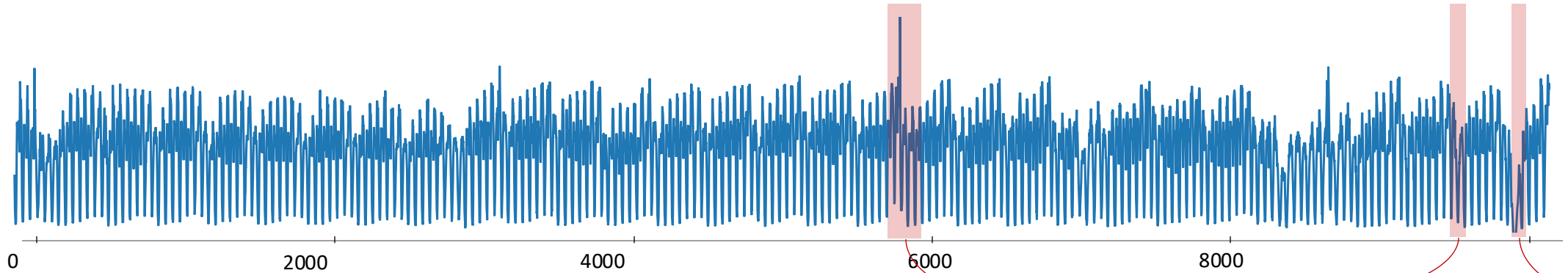
Daylight
Saving Time
(DST)

Flooding

Snowstorm

Introduction: *Anomaly Detection in Time Series*

- Time series T (example : number of taxi passengers in New York City)

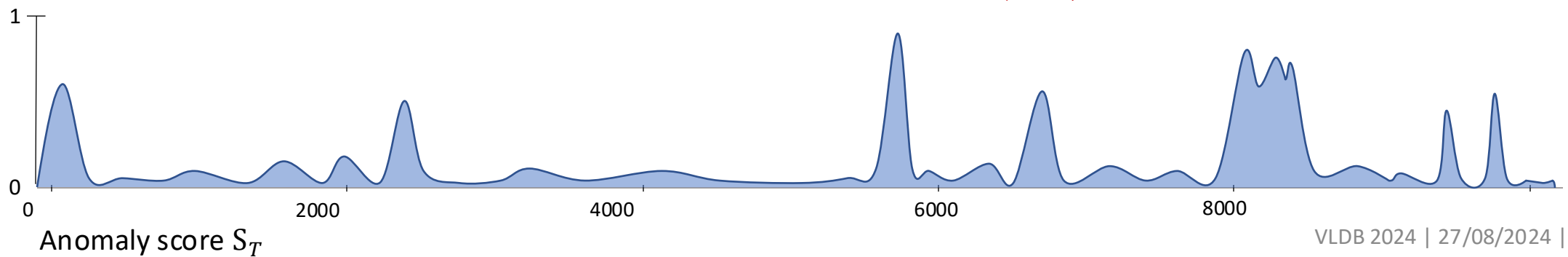


- Anomaly: *rare* point or sequence (of a given length) potentially *non-desired*

Daylight
Saving Time
(DST)

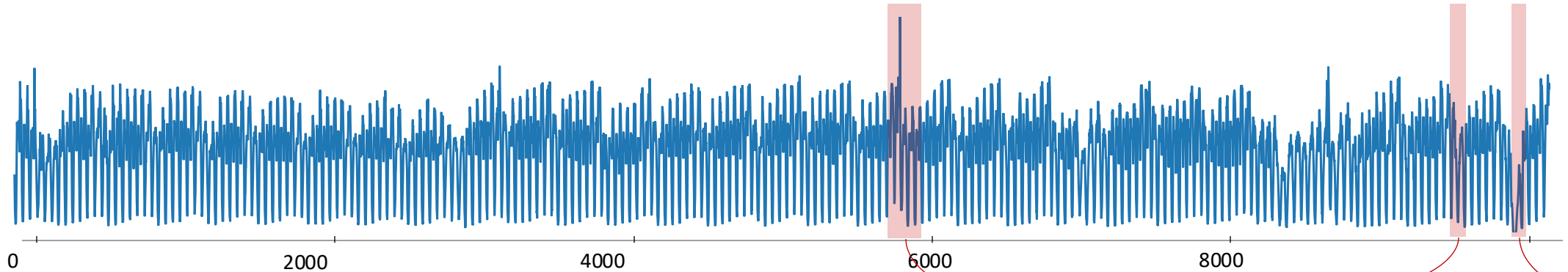
Flooding

Snowstorm



Introduction: *Anomaly Detection in Time Series*

- Time series T (example : number of taxi passengers in New York City)

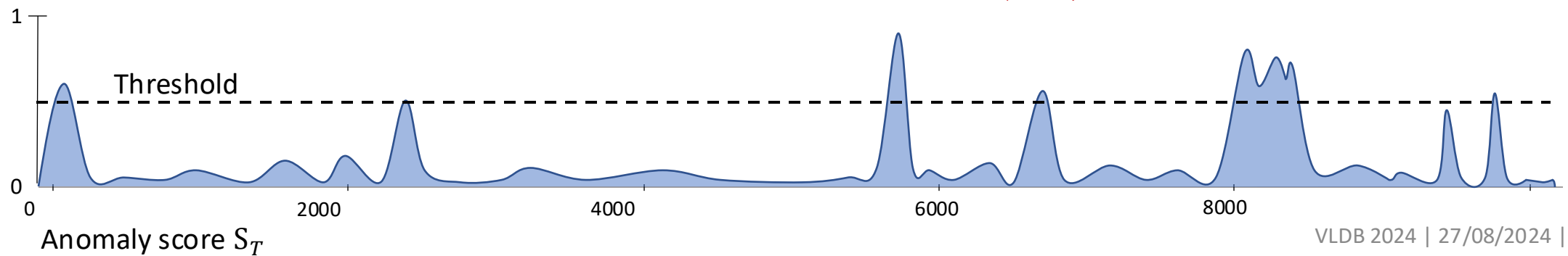


- Anomaly: *rare* point or sequence (of a given length) potentially *non-desired*

Daylight
Saving Time
(DST)

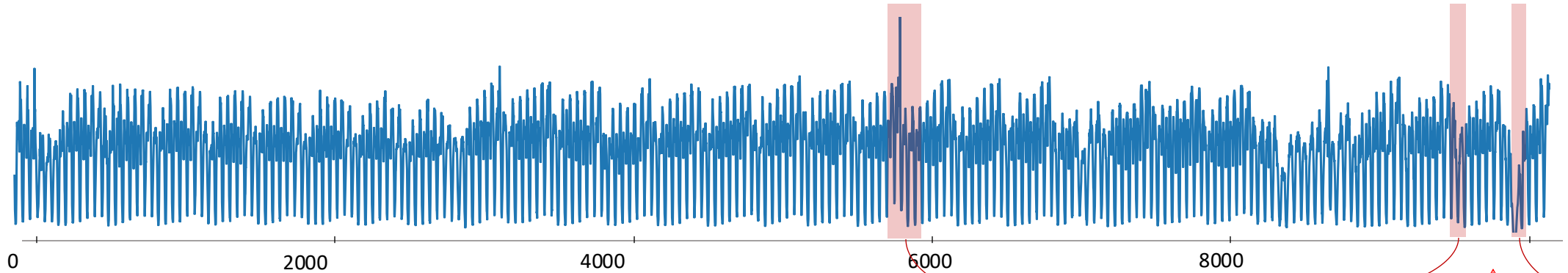
Flooding

Snowstorm

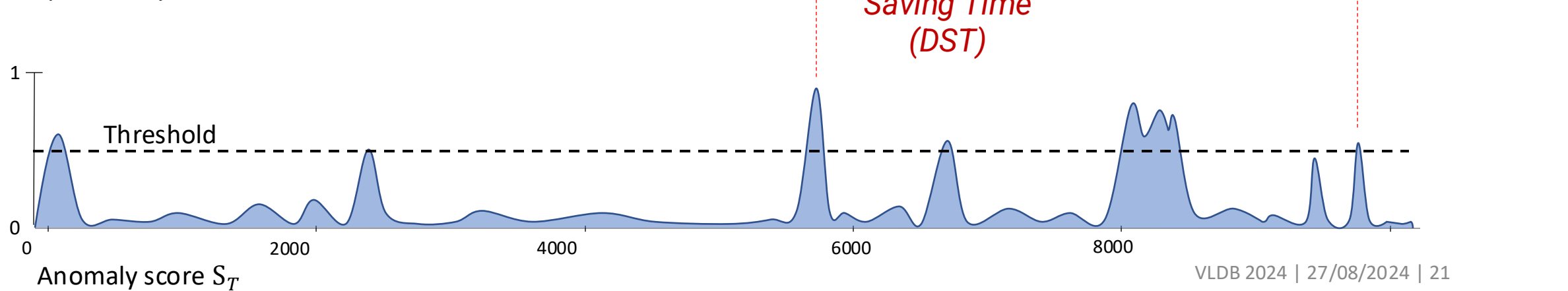


Introduction: *Anomaly Detection in Time Series*

- Time series T (example : number of taxi passengers in New York City)

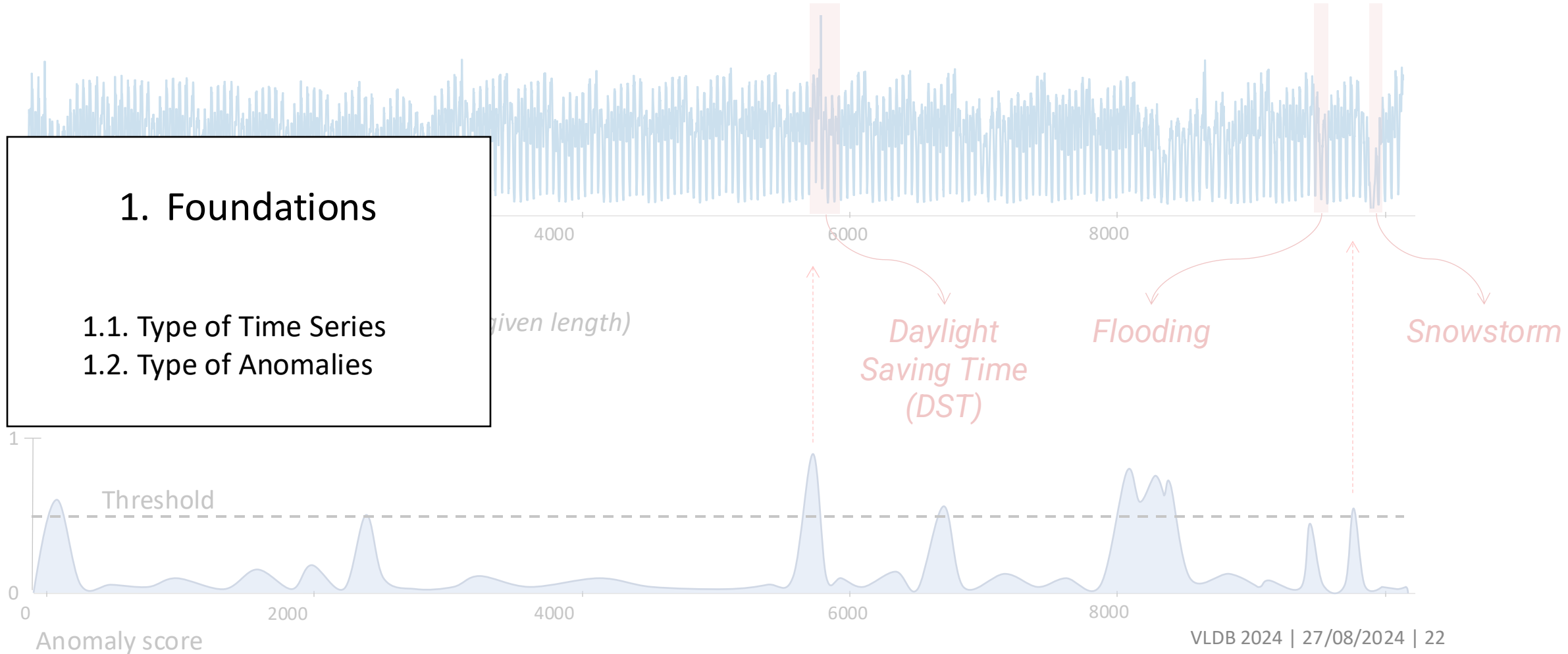


- Anomaly: *rare* point or sequence (of a given length) potentially *non-desired*



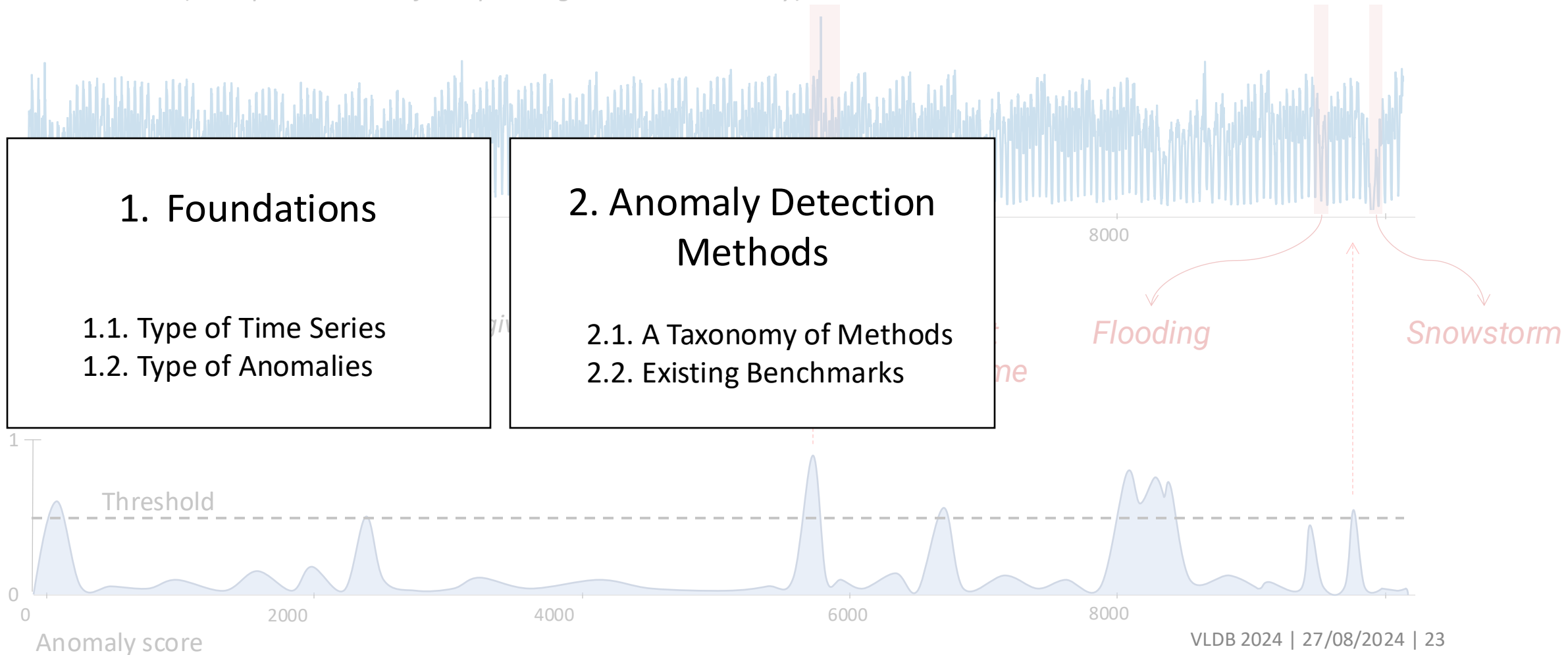
Introduction: *Outline*

- Time series (example : number of taxi passengers in New York City)



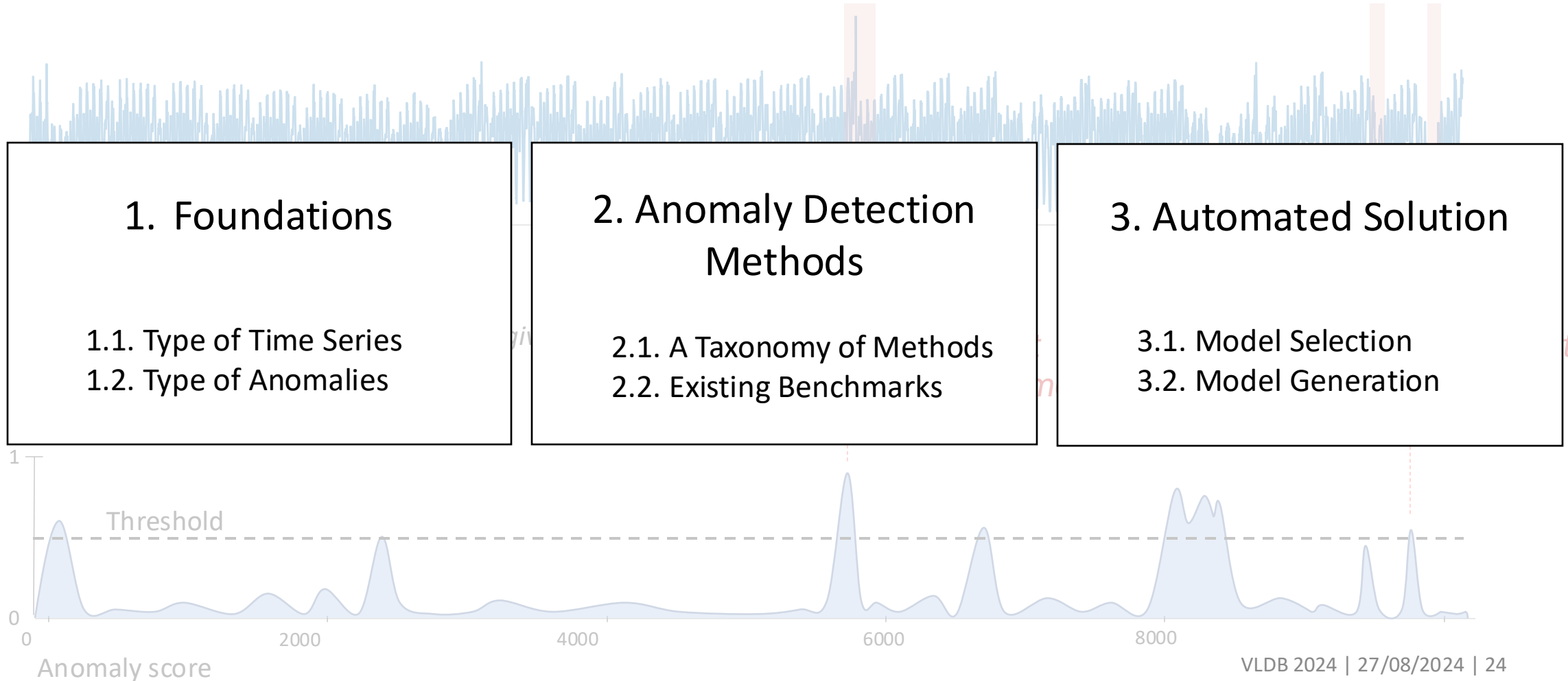
Introduction: *Outline*

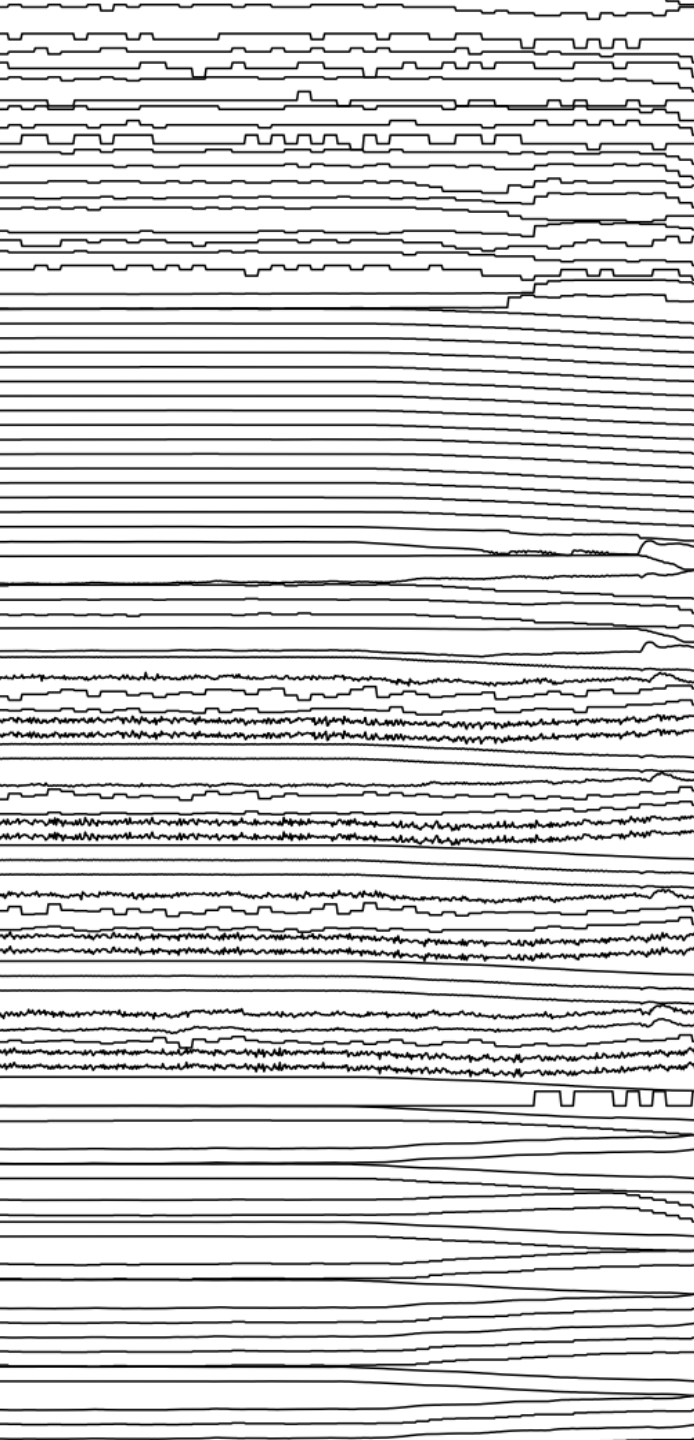
- Time series (example : number of taxi passengers in New York City)



Introduction: *Outline*

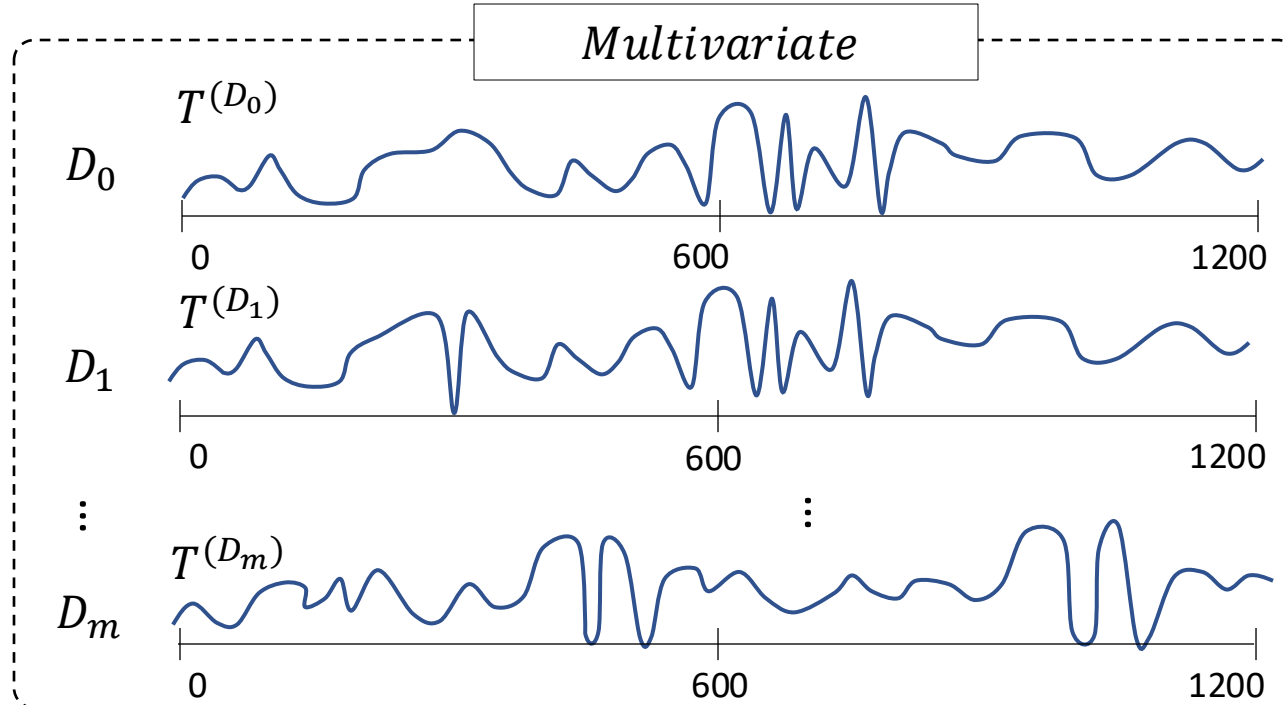
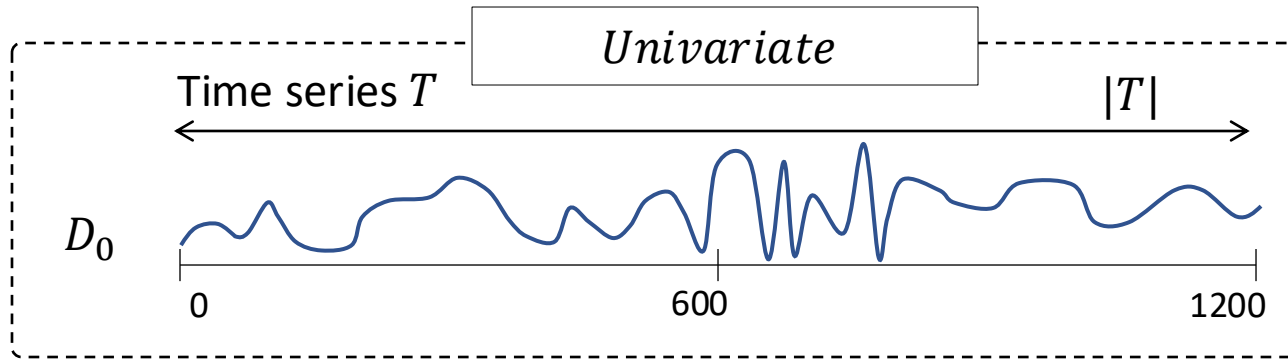
- Time series (example : number of taxi passengers in New York City)



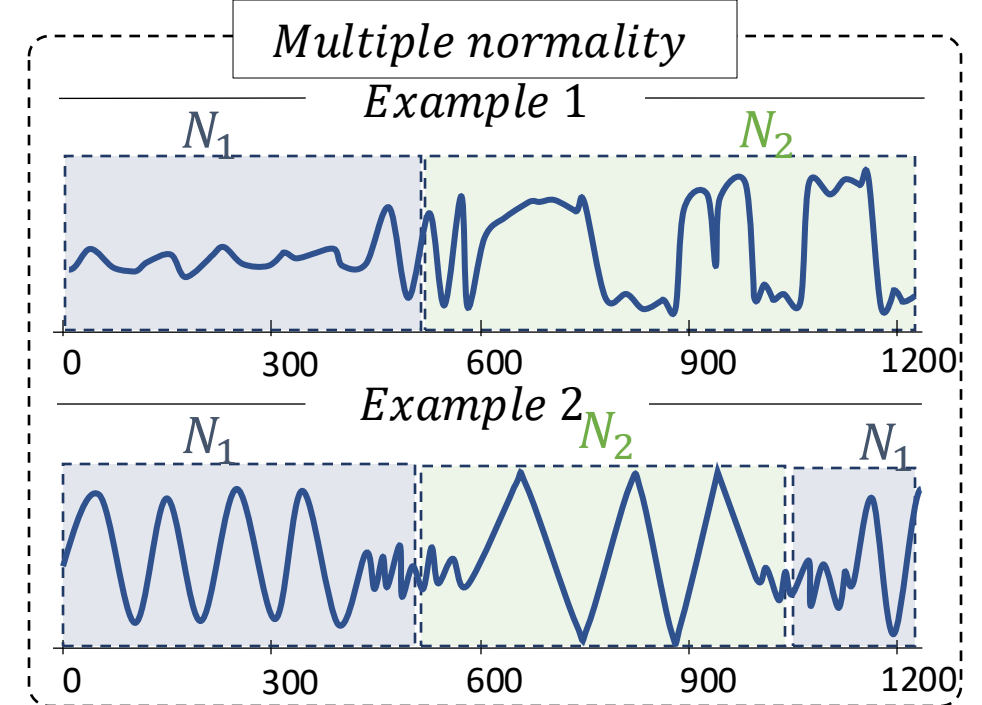
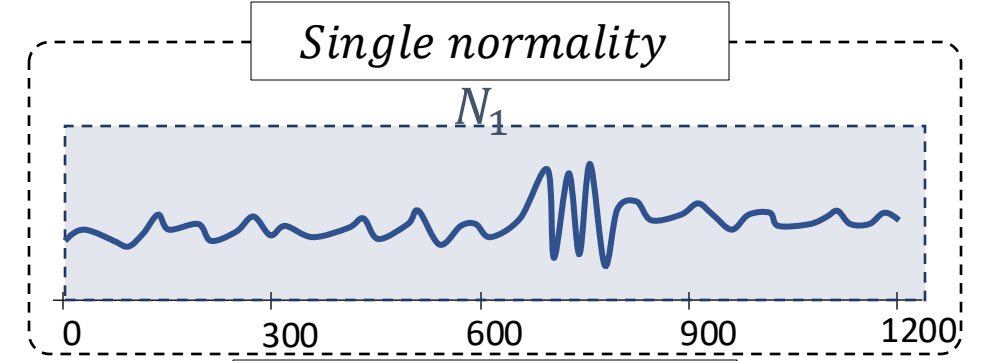
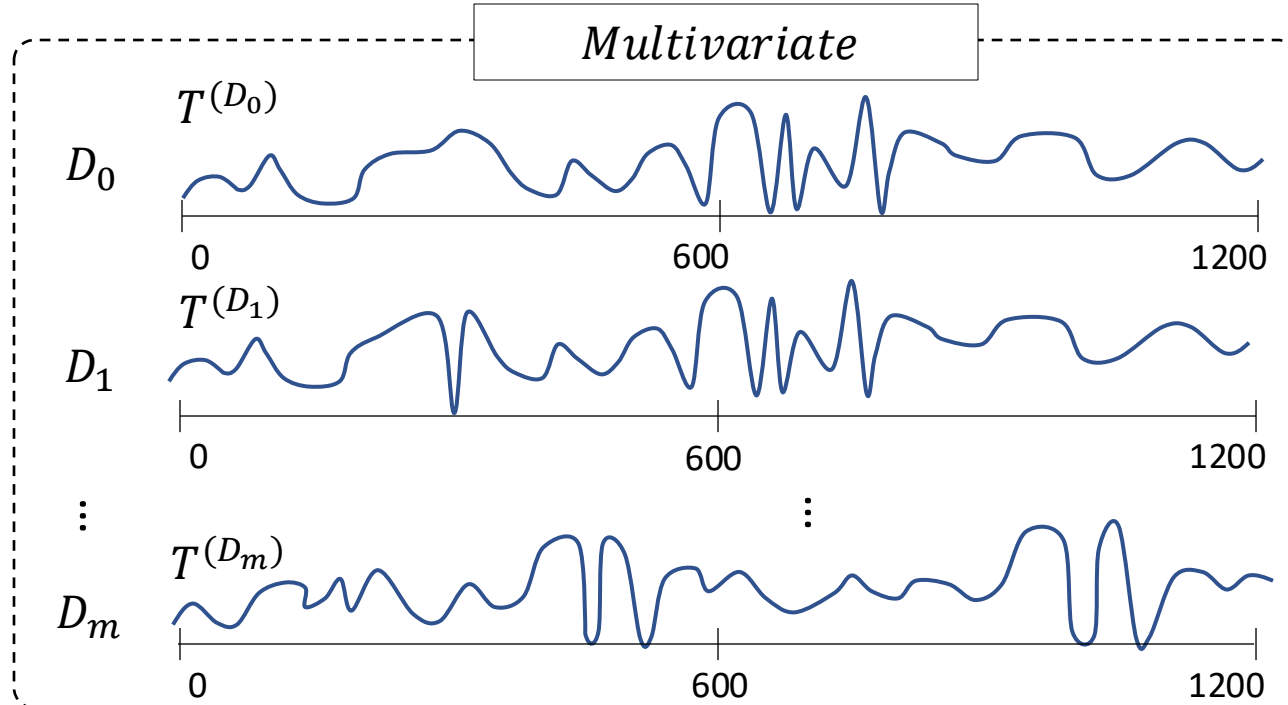
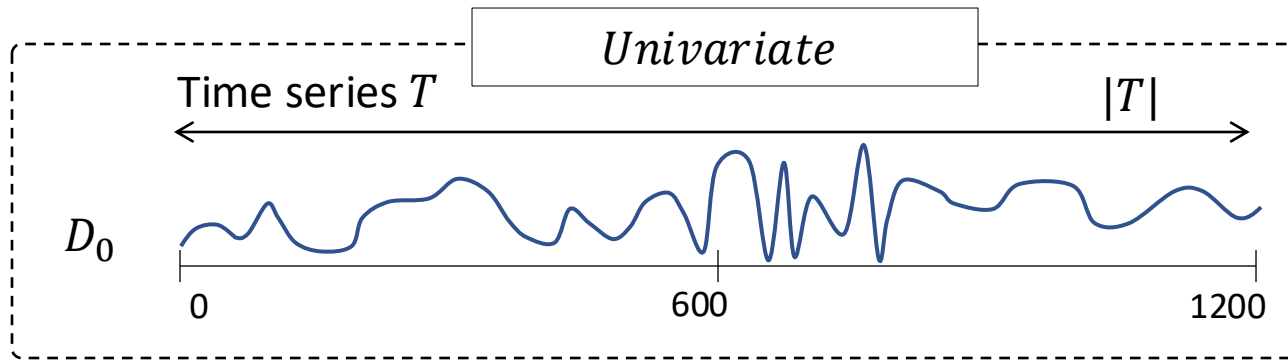


Foundations

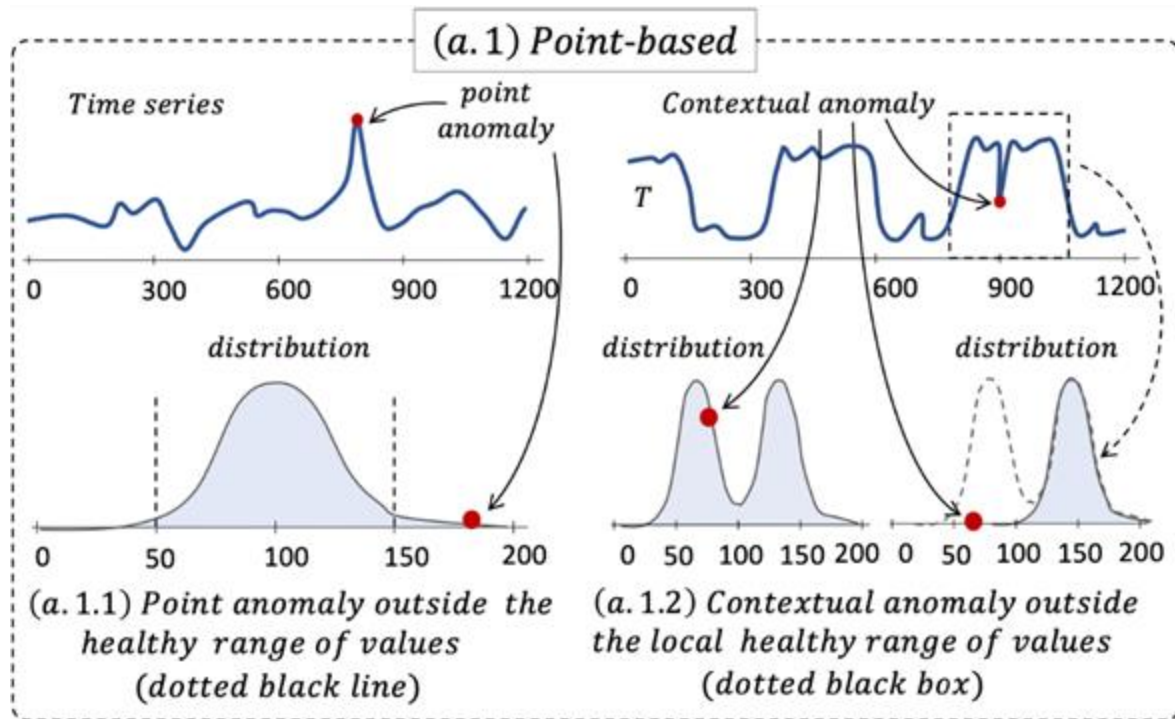
Foundations: *Type of time series*



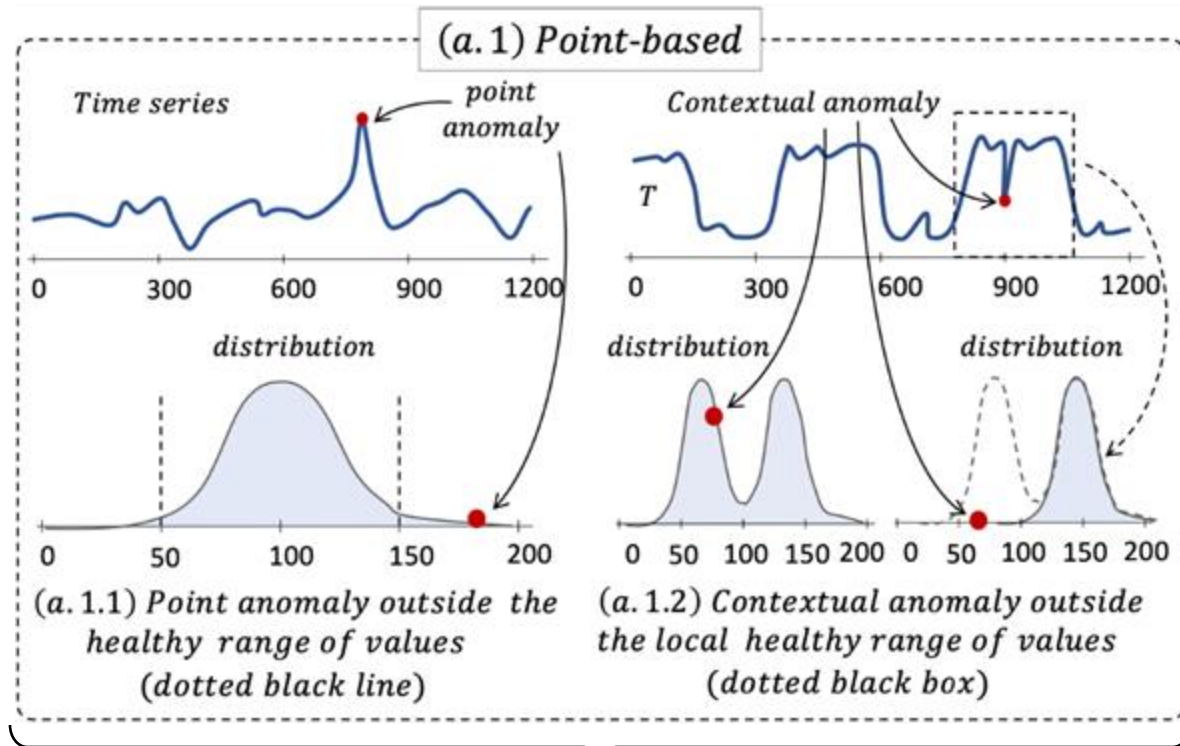
Foundations: *Type of time series*



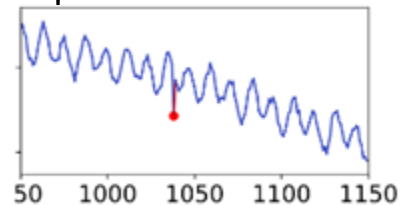
Foundations: *Type of anomalies*



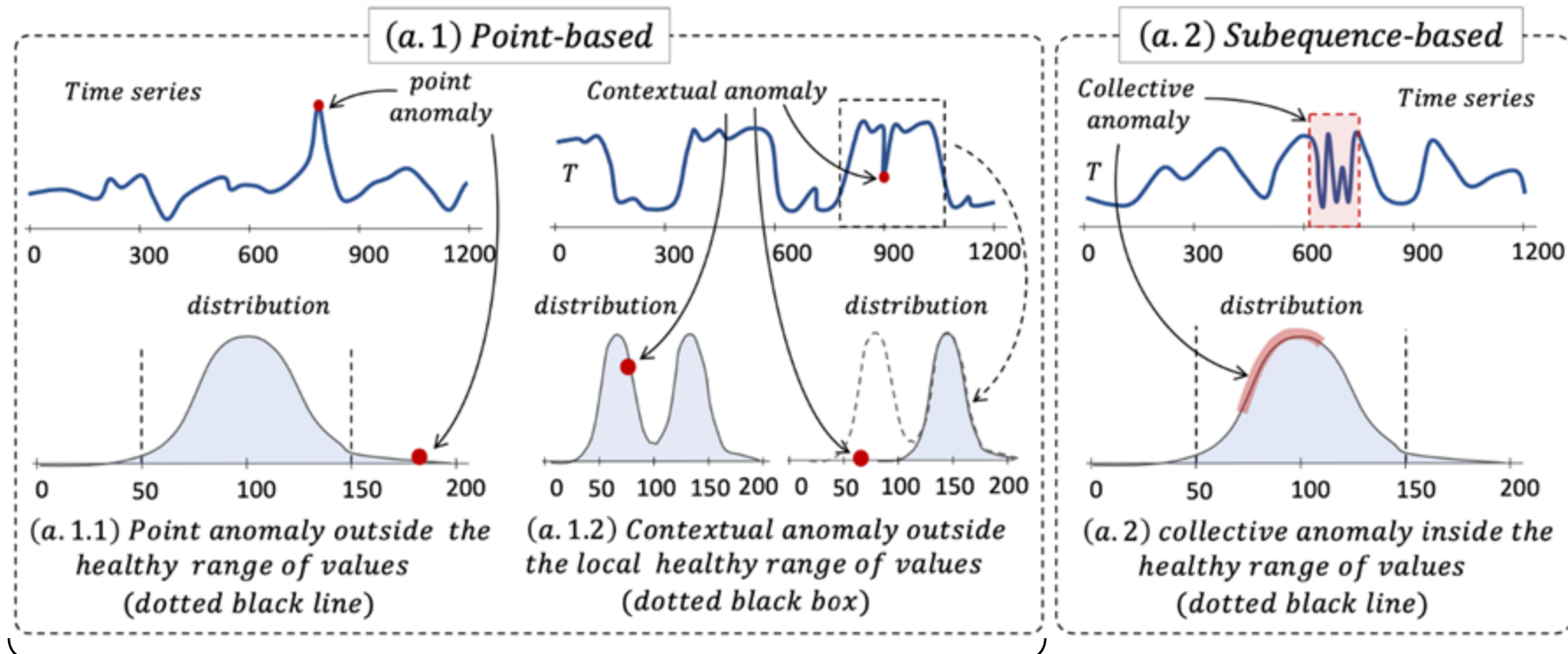
Foundations: *Type of anomalies*



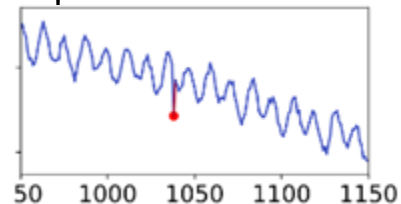
Example of
point-based
anomaly [1]



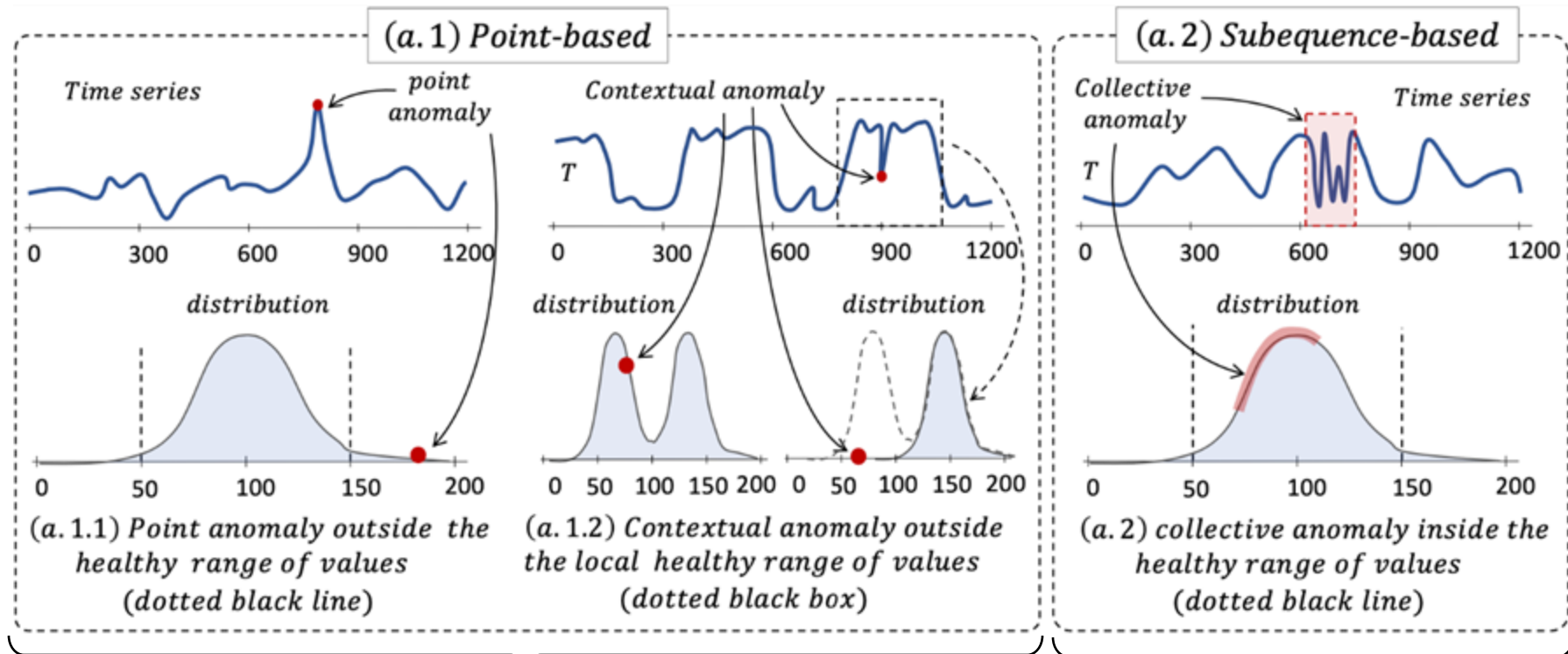
Foundations: *Type of anomalies*



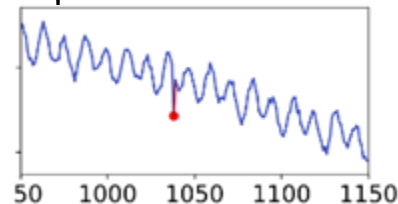
Example of
point-based
anomaly [1]



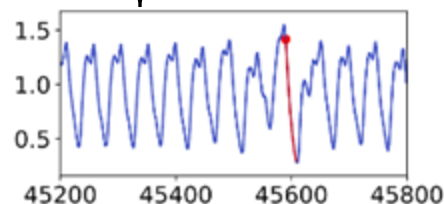
Foundations: *Type of anomalies*



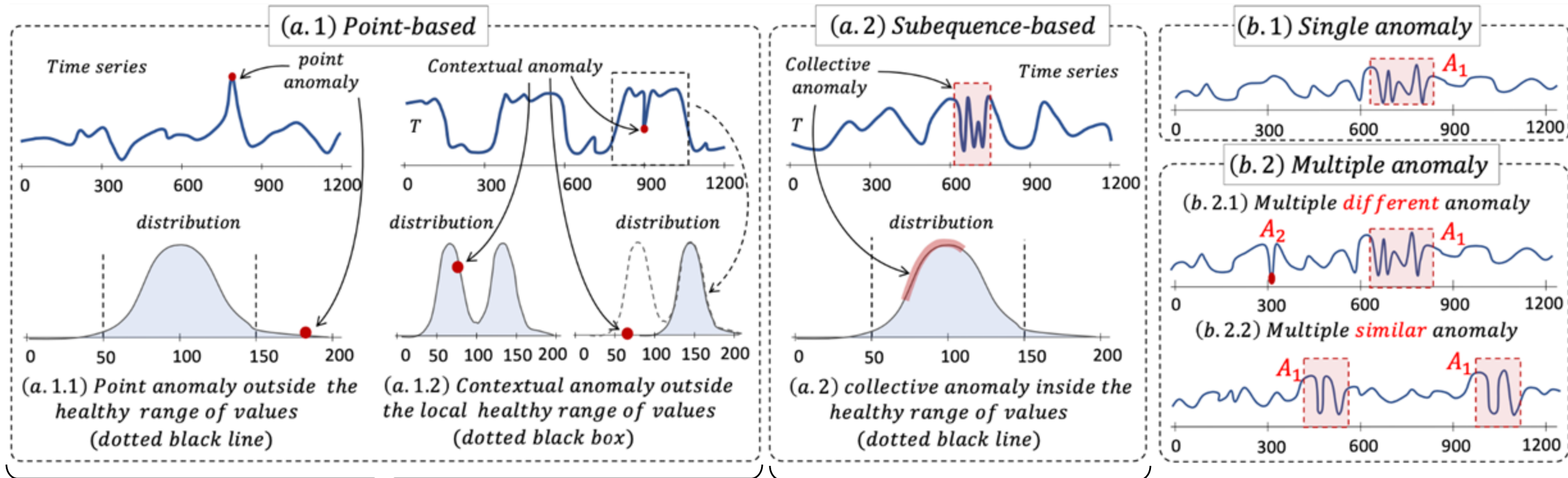
Example of
point-based
anomaly [1]



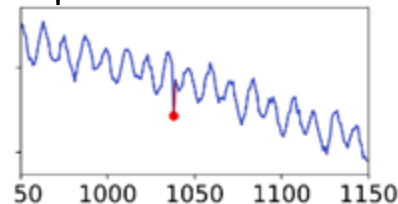
Example of
subsequence-
based anomaly [2]



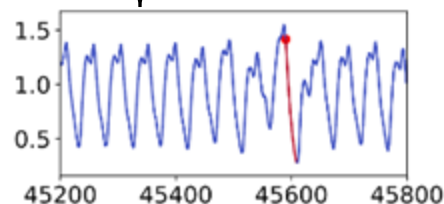
Foundations: *Type of anomalies*



Example of point-based anomaly [1]

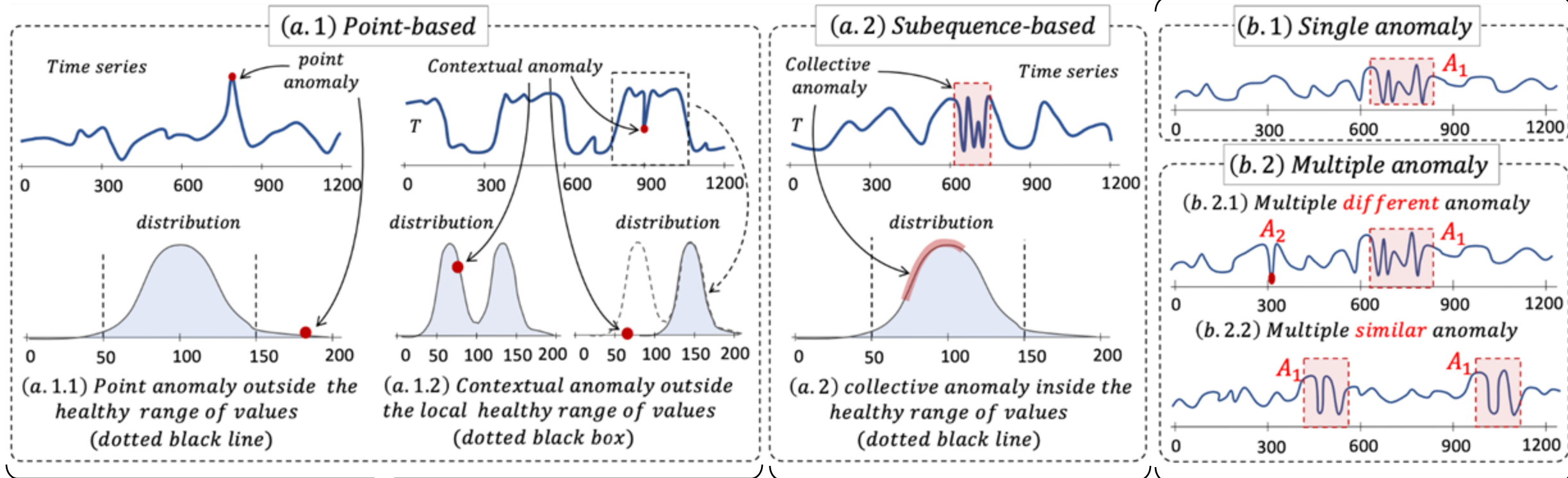
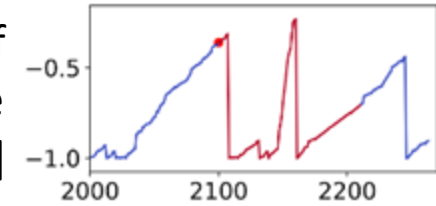


Example of subsequence-based anomaly [2]

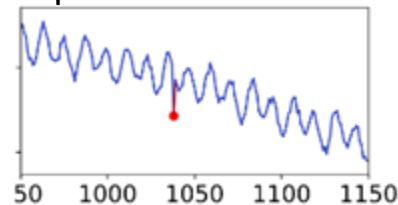


Foundations: *Type of anomalies*

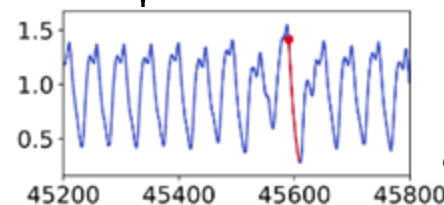
Example of
single
anomaly [3]



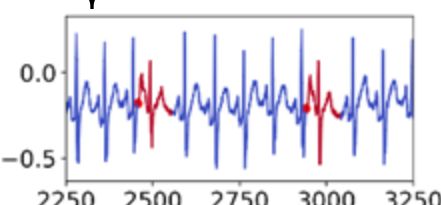
Example of
point-based
anomaly [1]

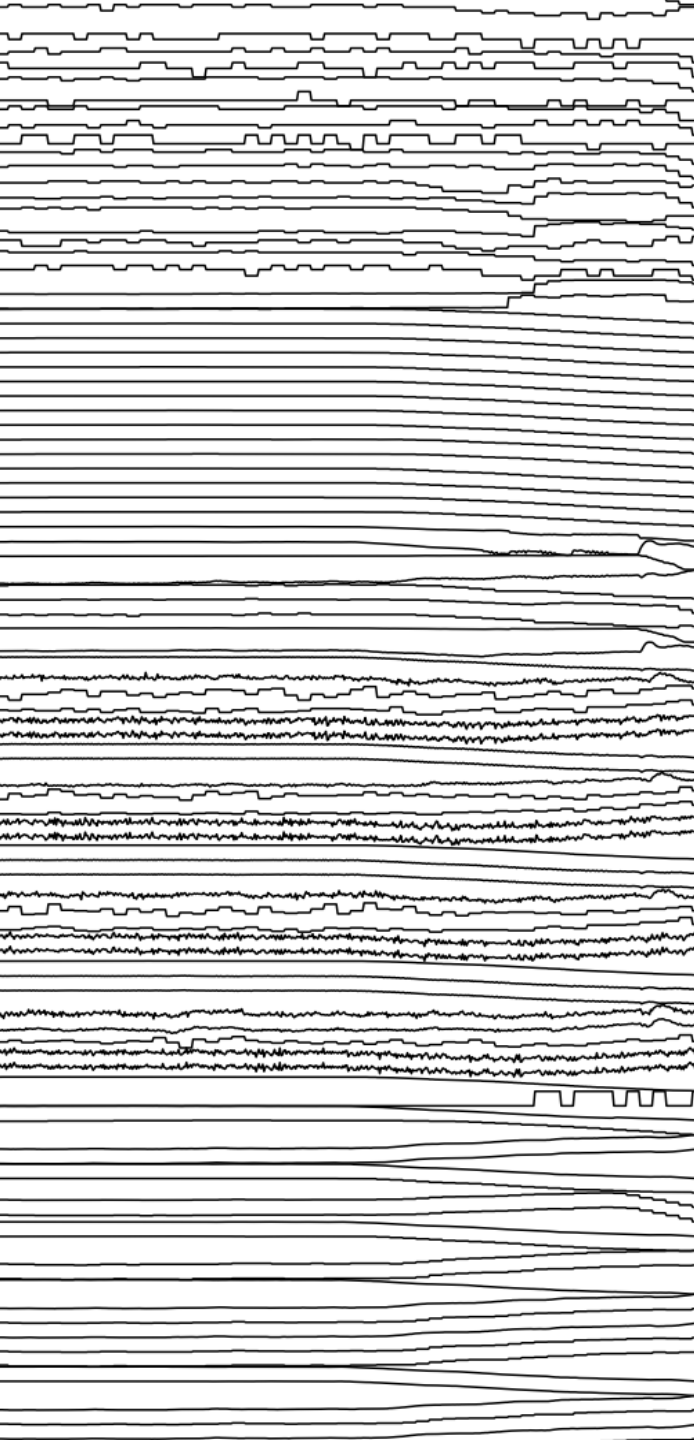


Example of
subsequence-
based anomaly [2]



Example of
multiple
anomaly [4]

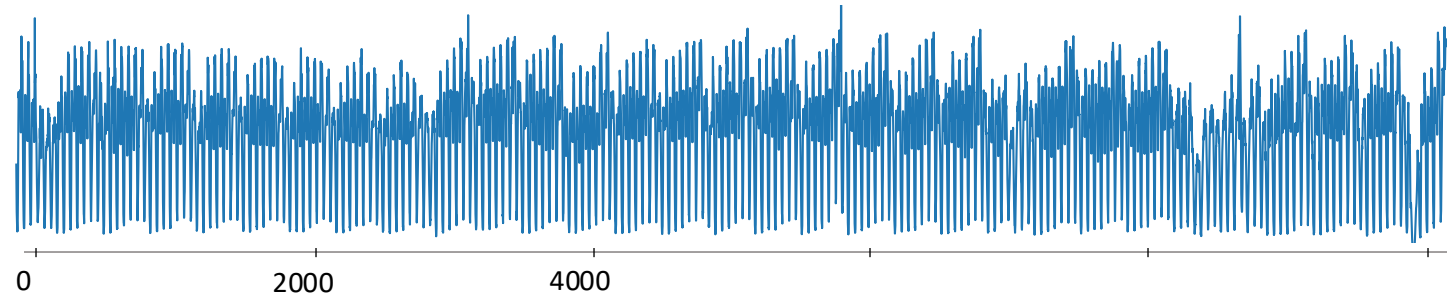




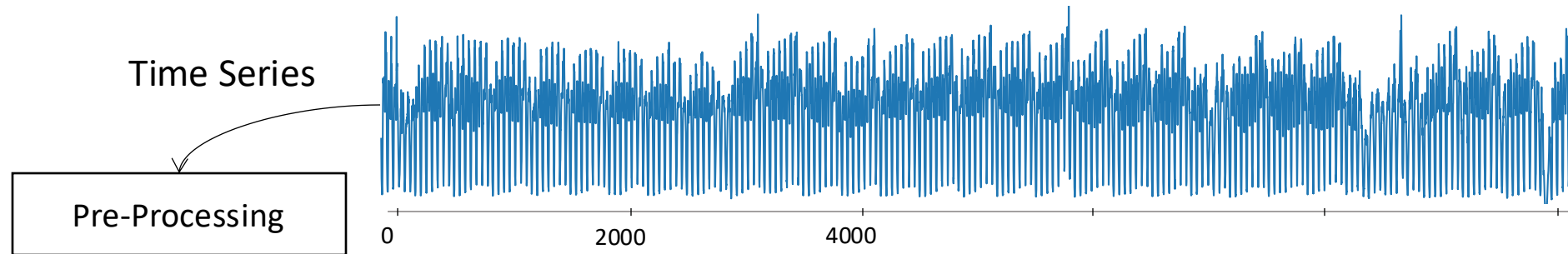
Anomaly Detection Methods

Anomaly Detection methods: *A taxonomy*

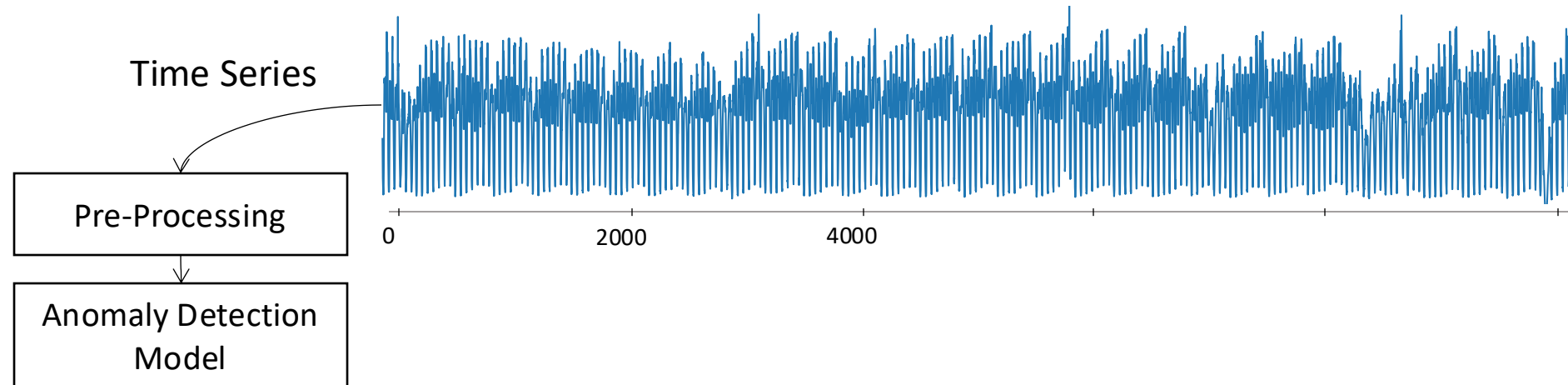
Time Series



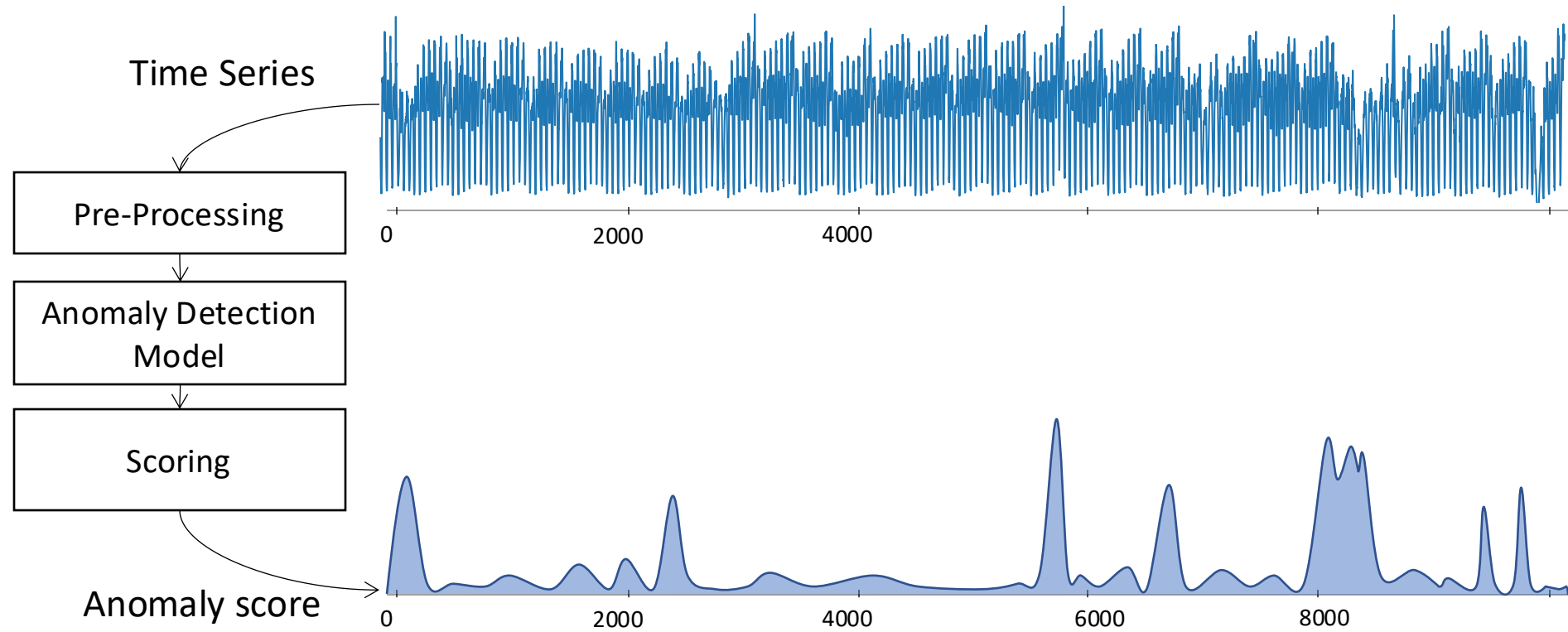
Anomaly Detection methods: *A taxonomy*



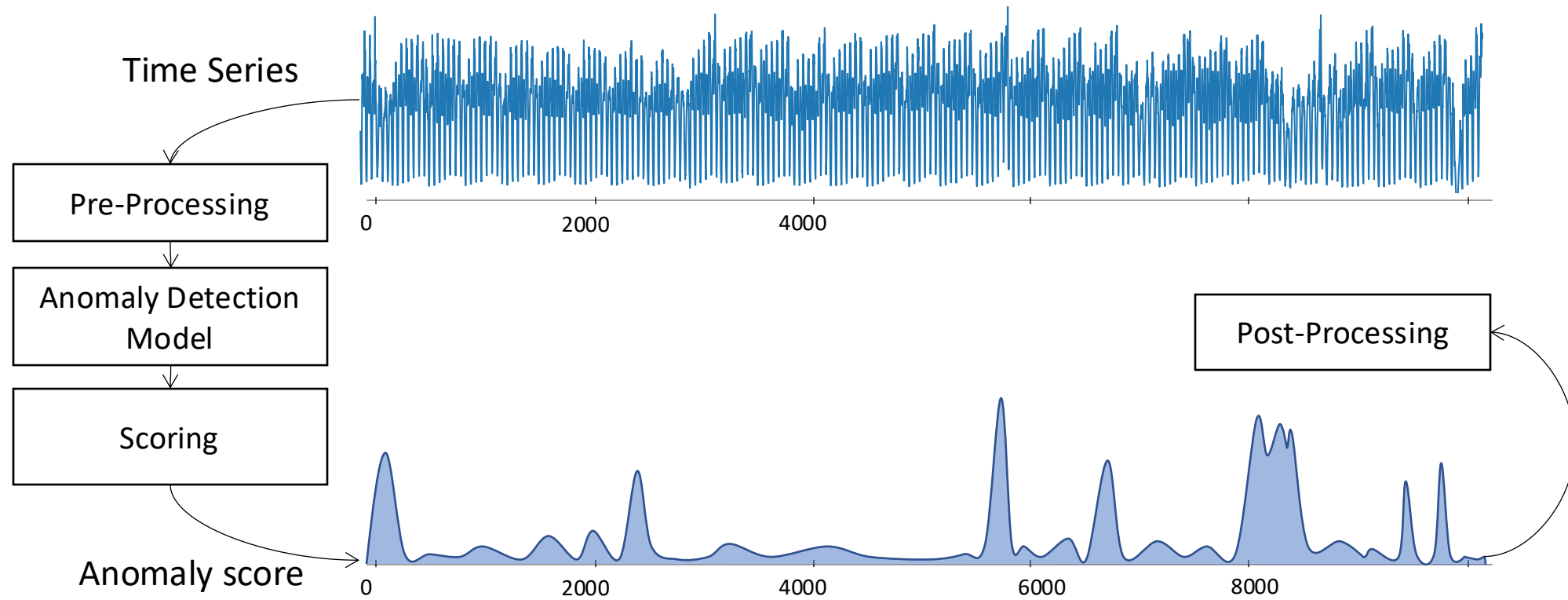
Anomaly Detection methods: *A taxonomy*



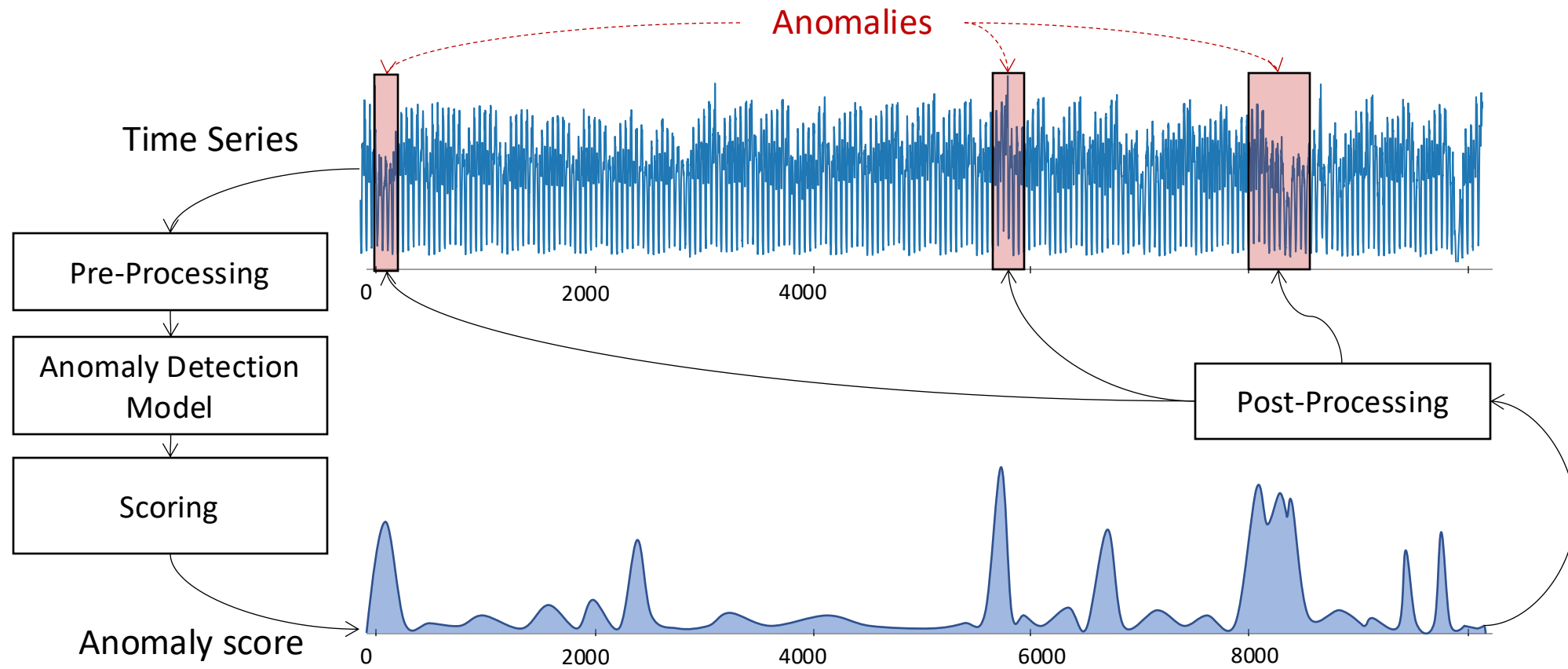
Anomaly Detection methods: *A taxonomy*



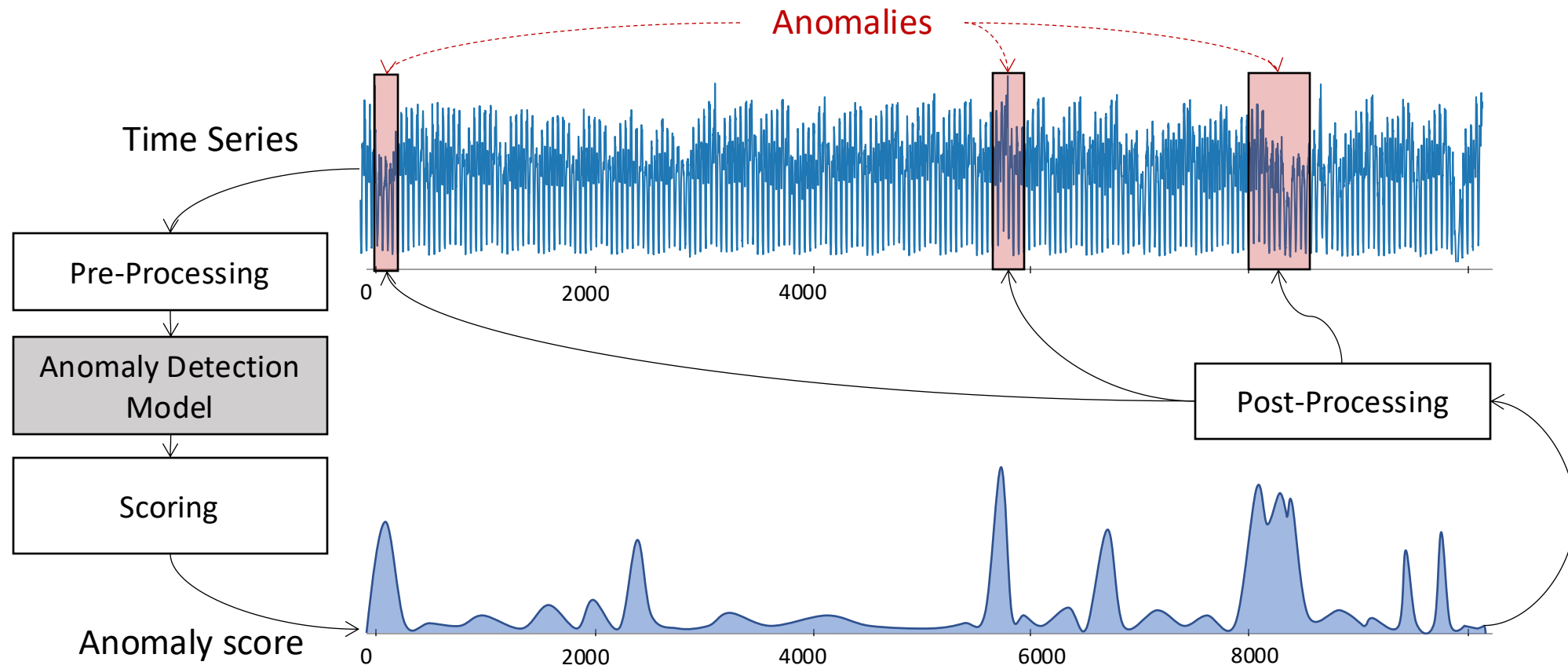
Anomaly Detection methods: *A taxonomy*



Anomaly Detection methods: *A taxonomy*

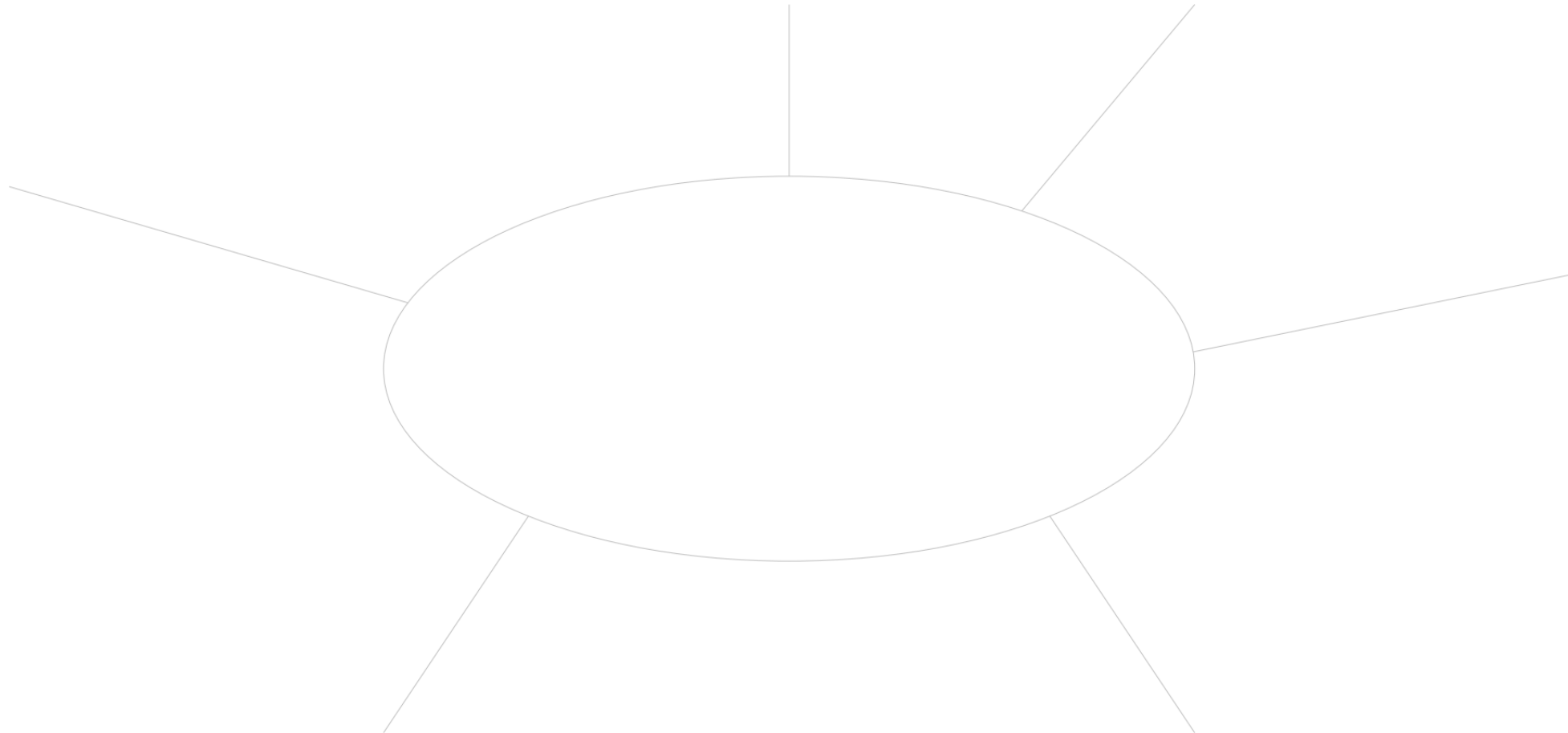


Anomaly Detection methods: *A taxonomy*



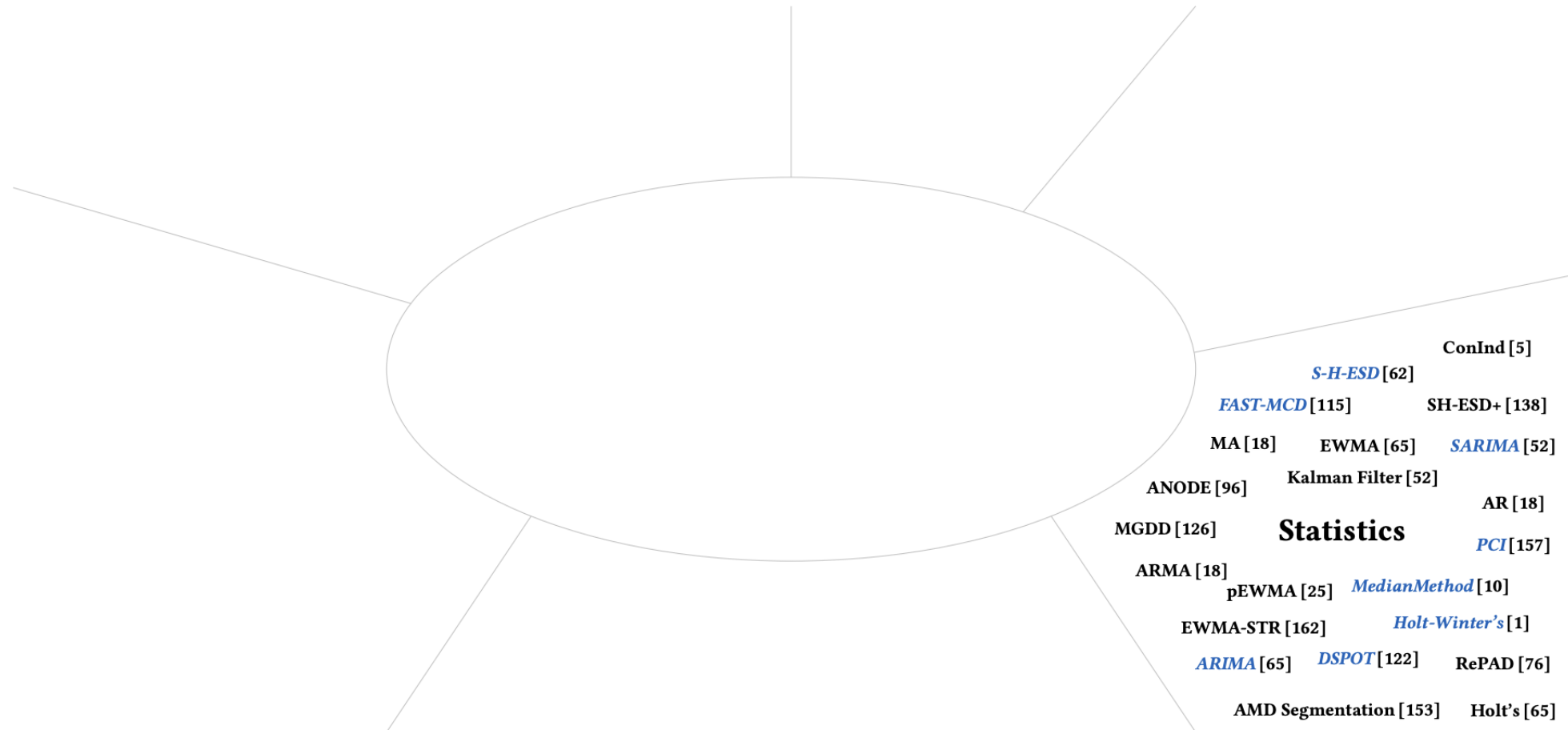
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



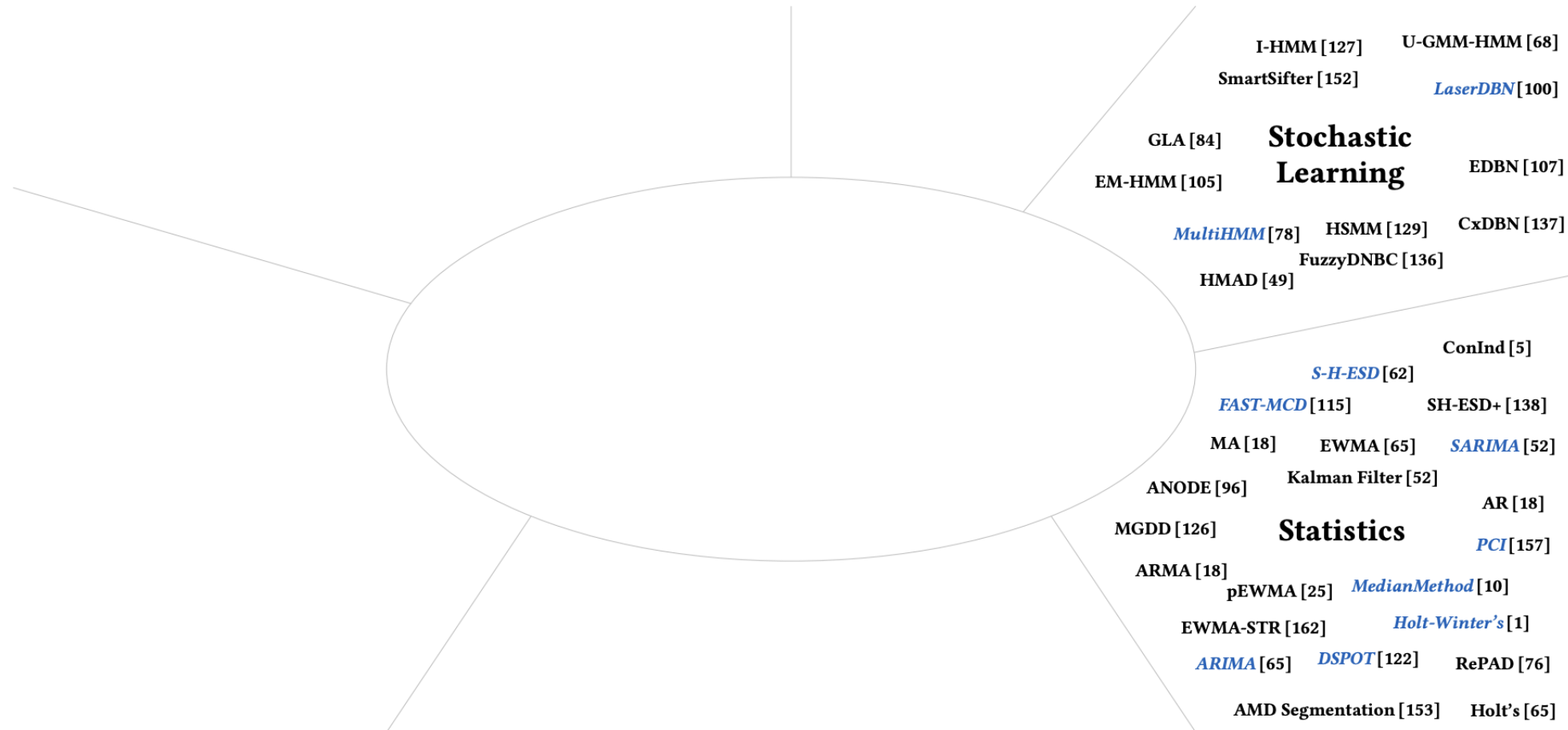
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



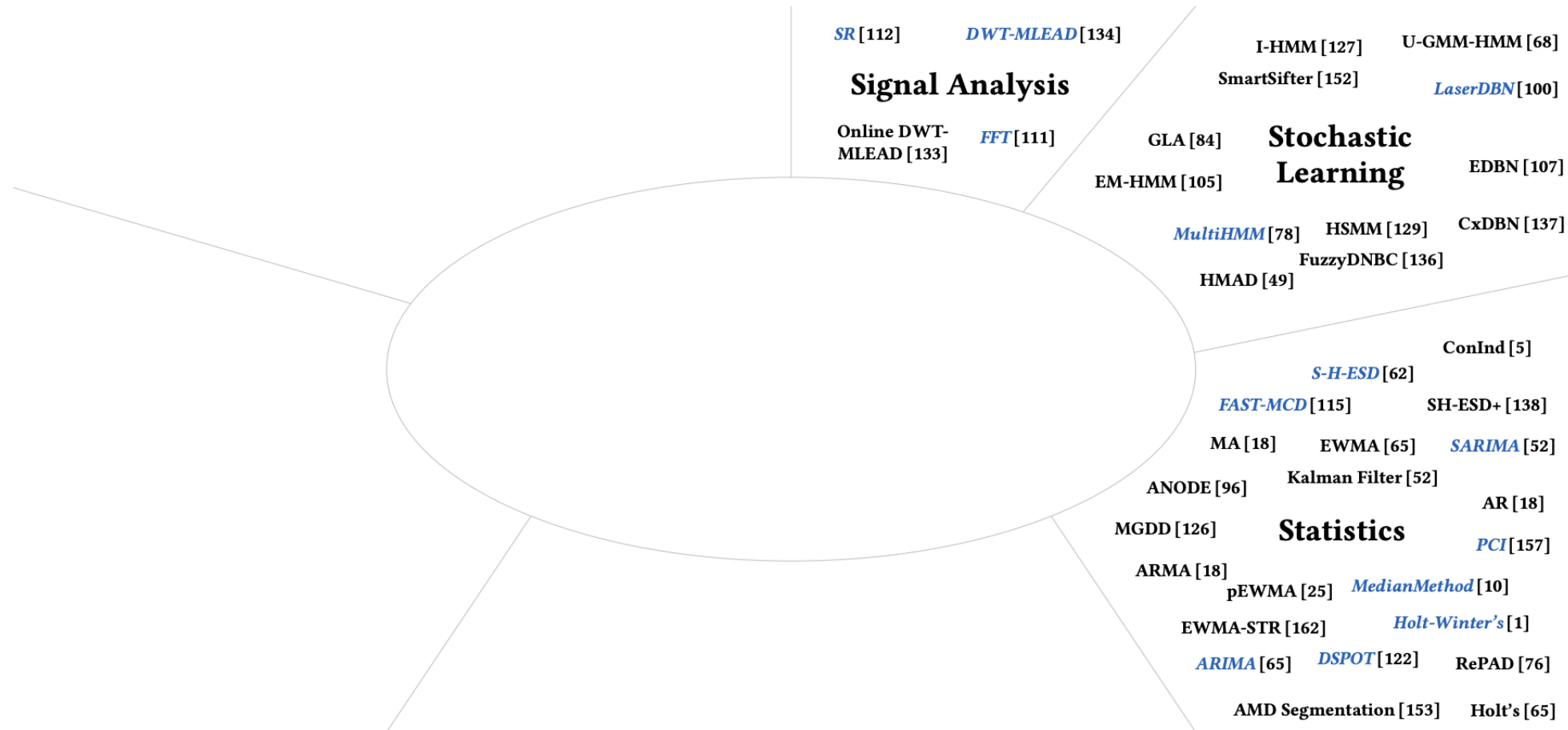
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



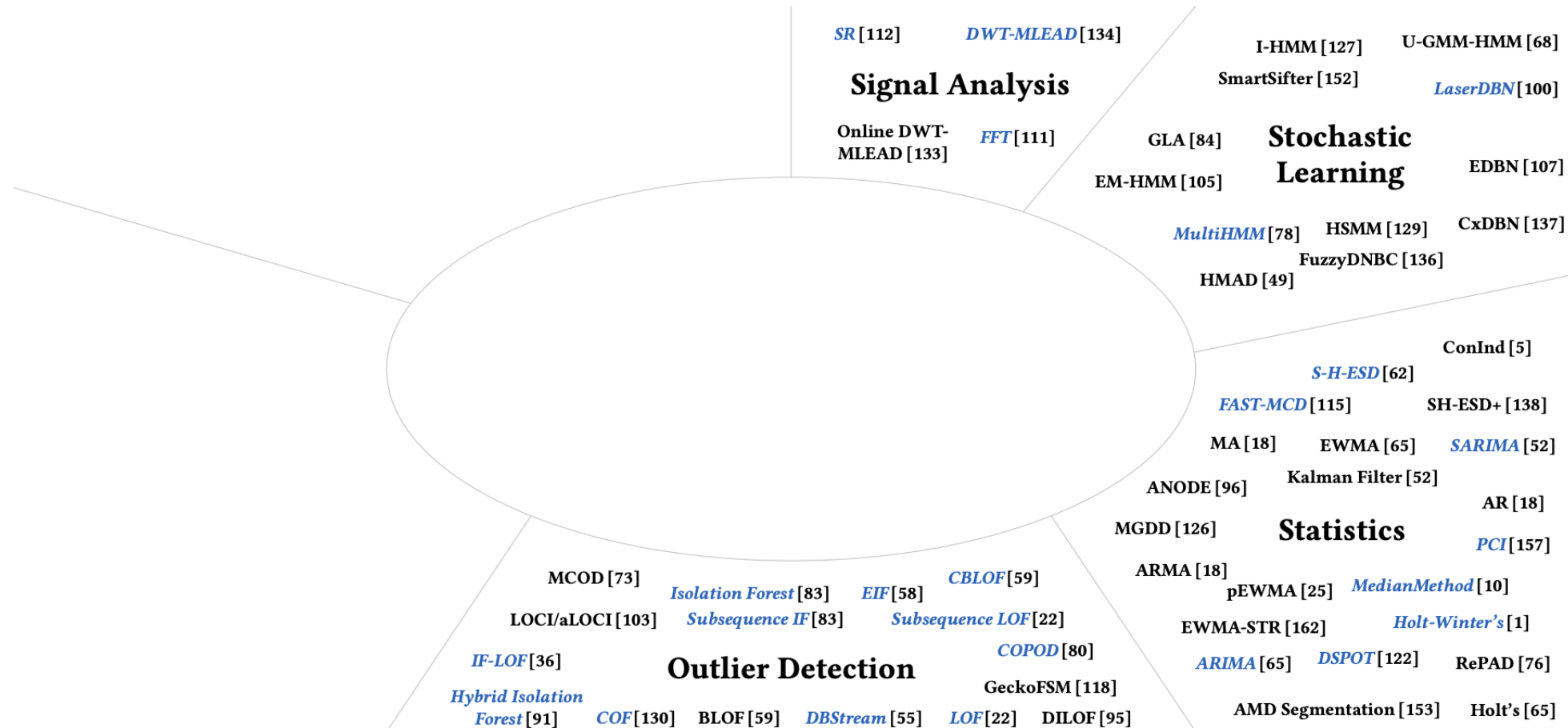
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



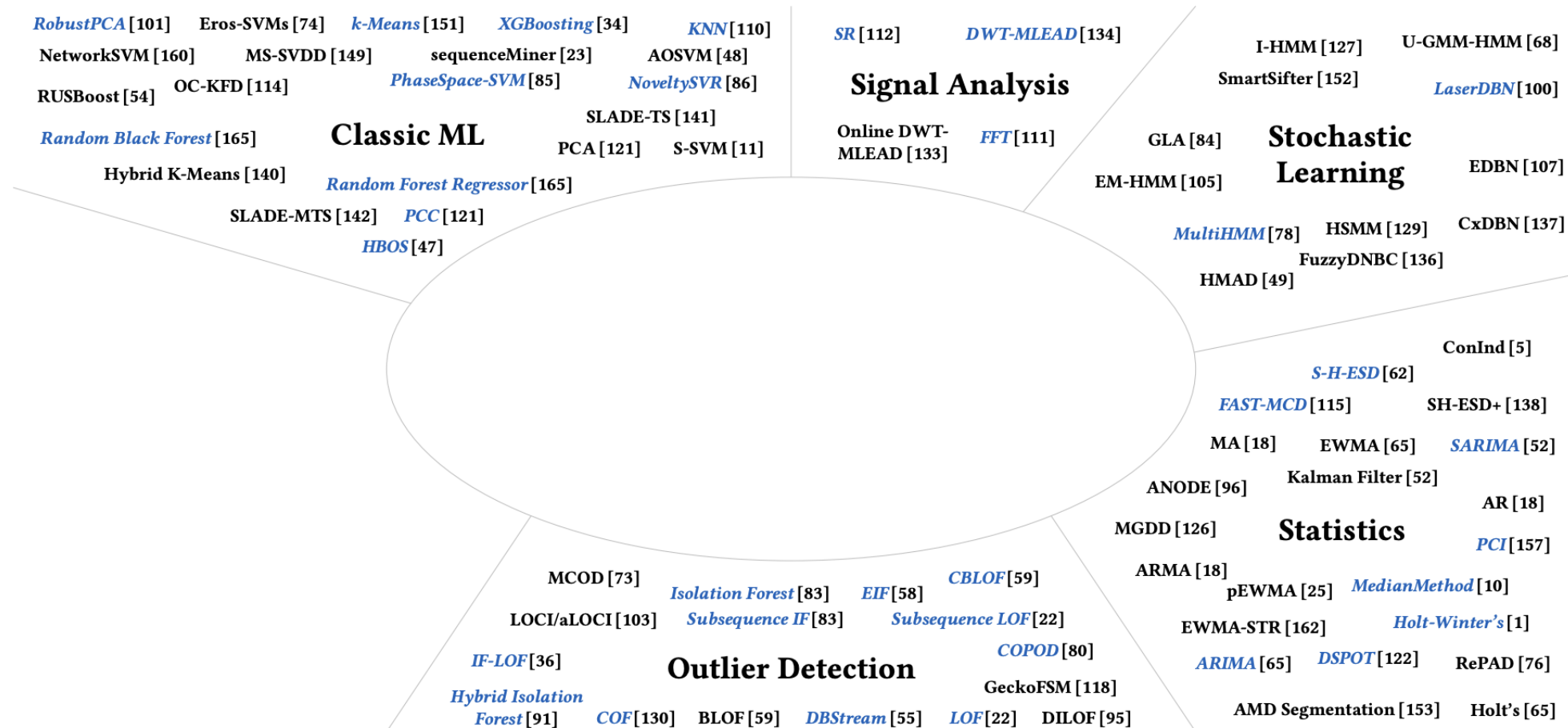
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



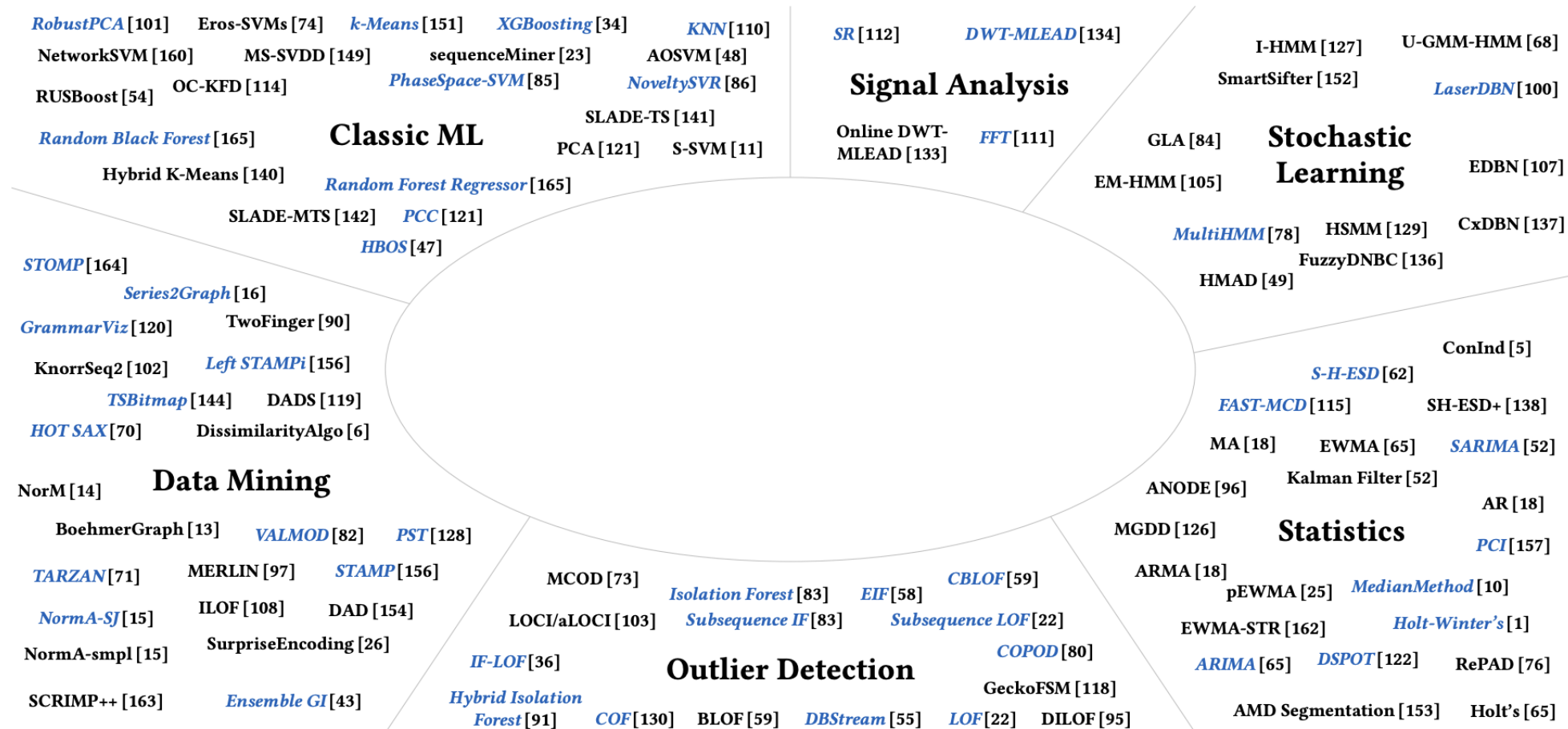
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



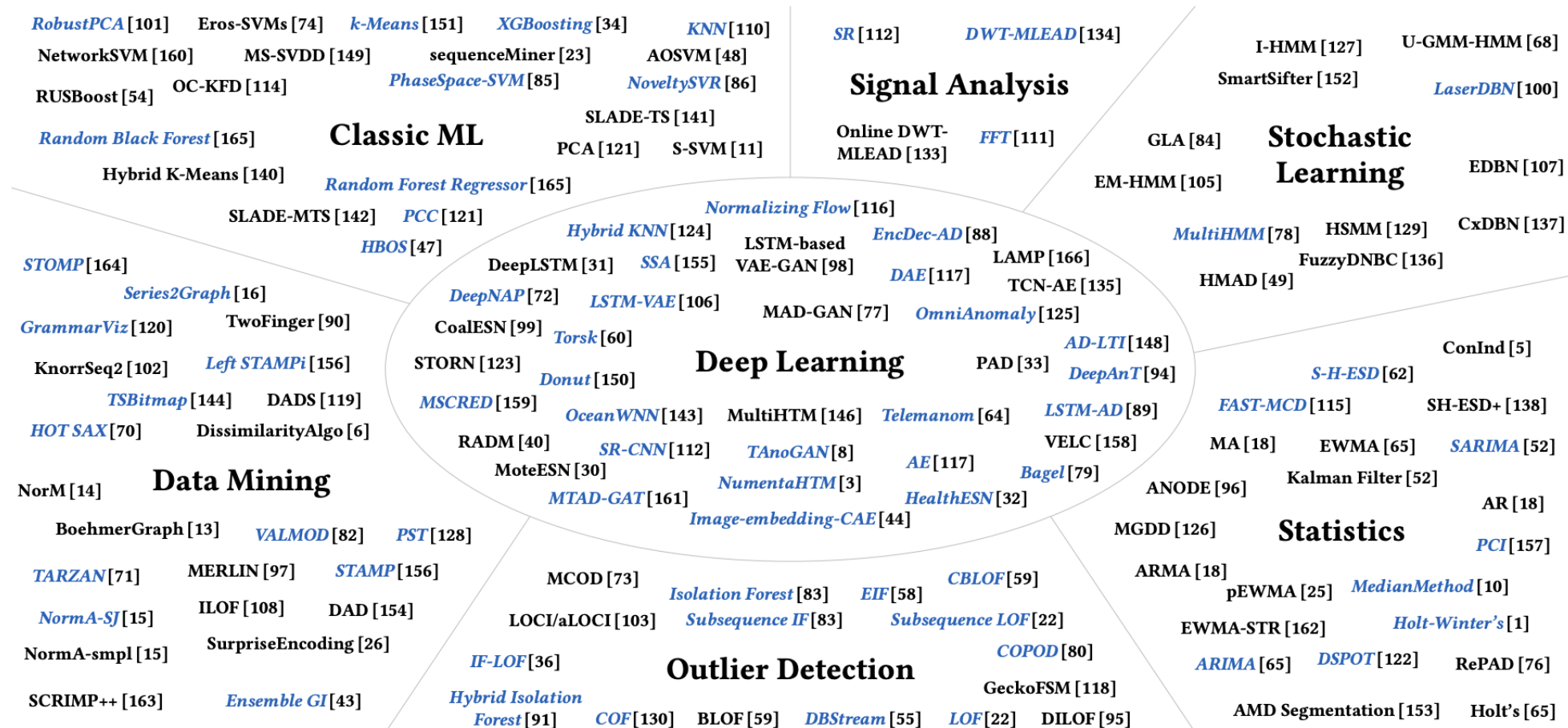
Anomaly Detection methods: *A taxonomy*

By domains [5] ...



Anomaly Detection methods: *A taxonomy*

By domains [5] ...



Anomaly Detection methods: *A taxonomy*

By inputs...

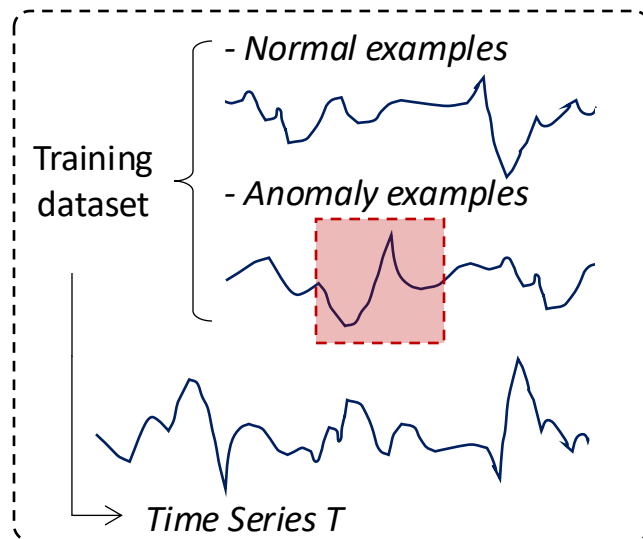
Time series anomaly detection methods

Anomaly Detection methods: *A taxonomy*

By inputs...

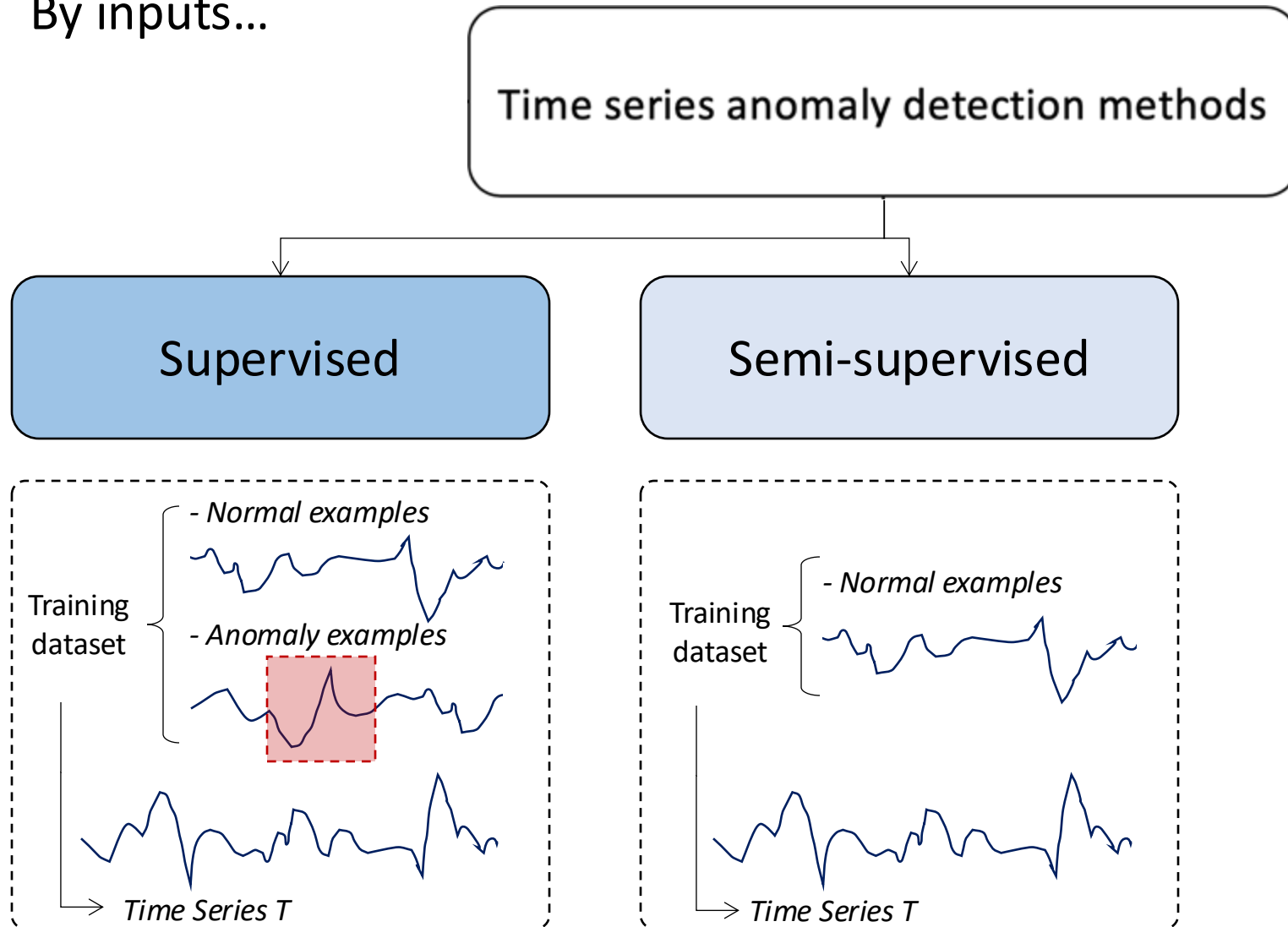
Time series anomaly detection methods

Supervised



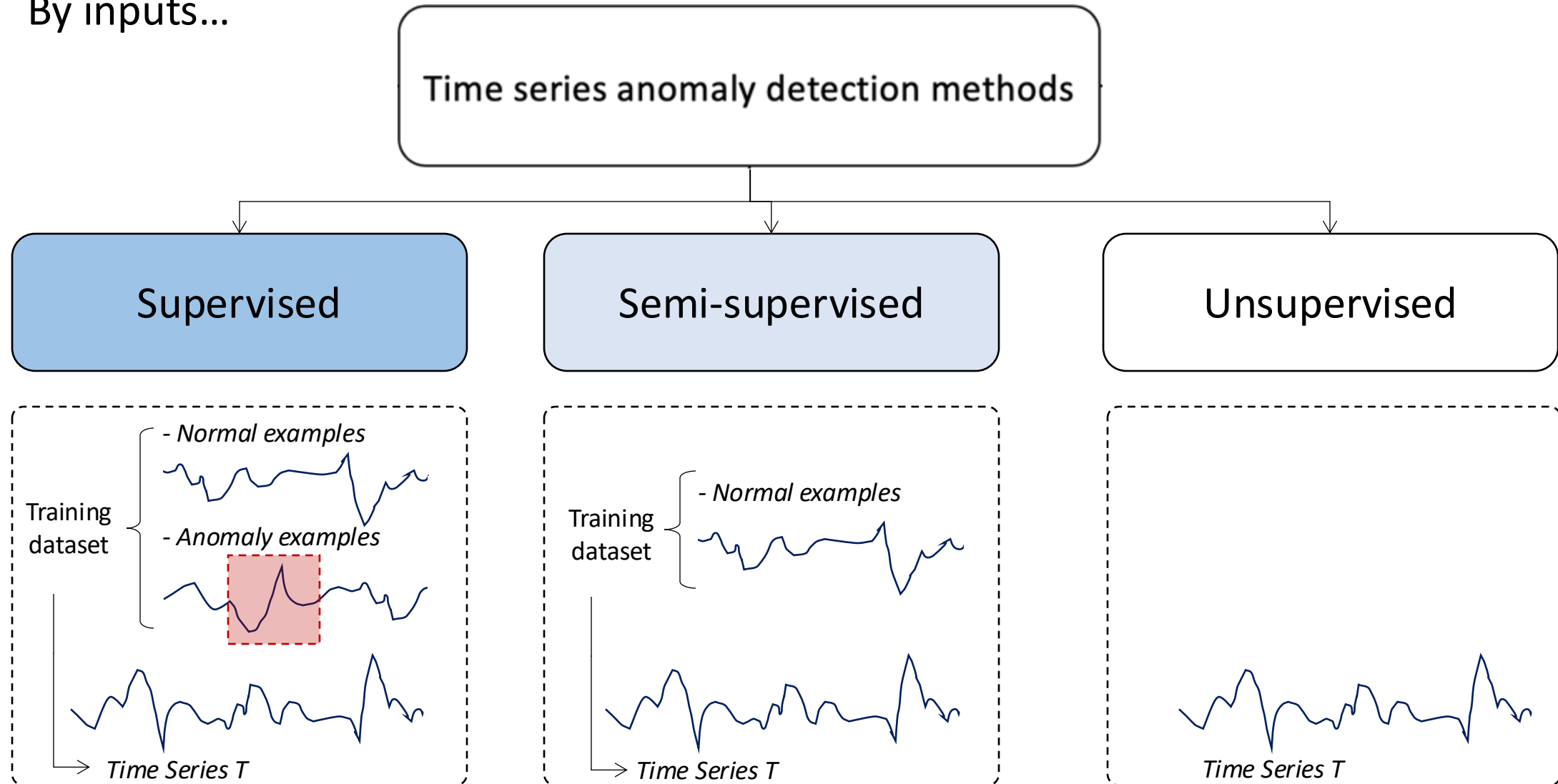
Anomaly Detection methods: *A taxonomy*

By inputs...



Anomaly Detection methods: *A taxonomy*

By inputs...



Anomaly Detection

By inputs...

Time

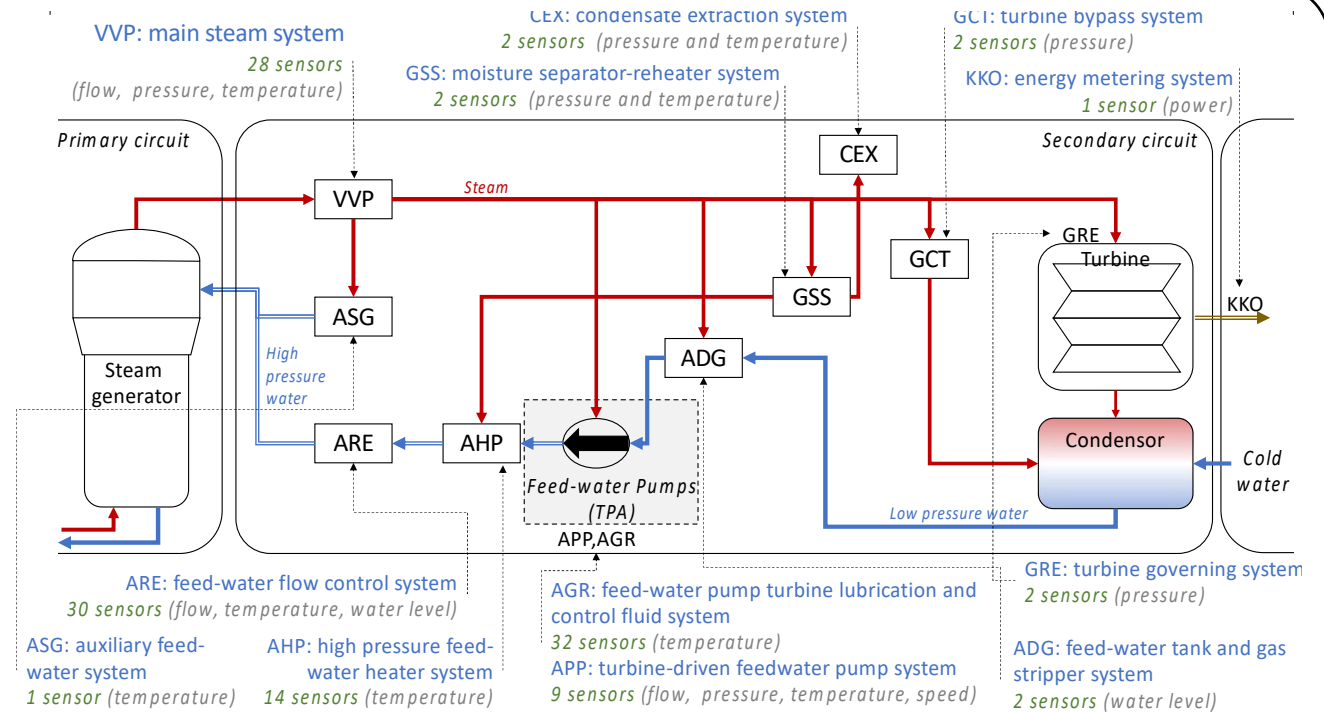
Supervised

Training dataset

- Normal examples

- Anomaly examples

Time Series T



Anomaly Detection

By inputs...

Time

Supervised

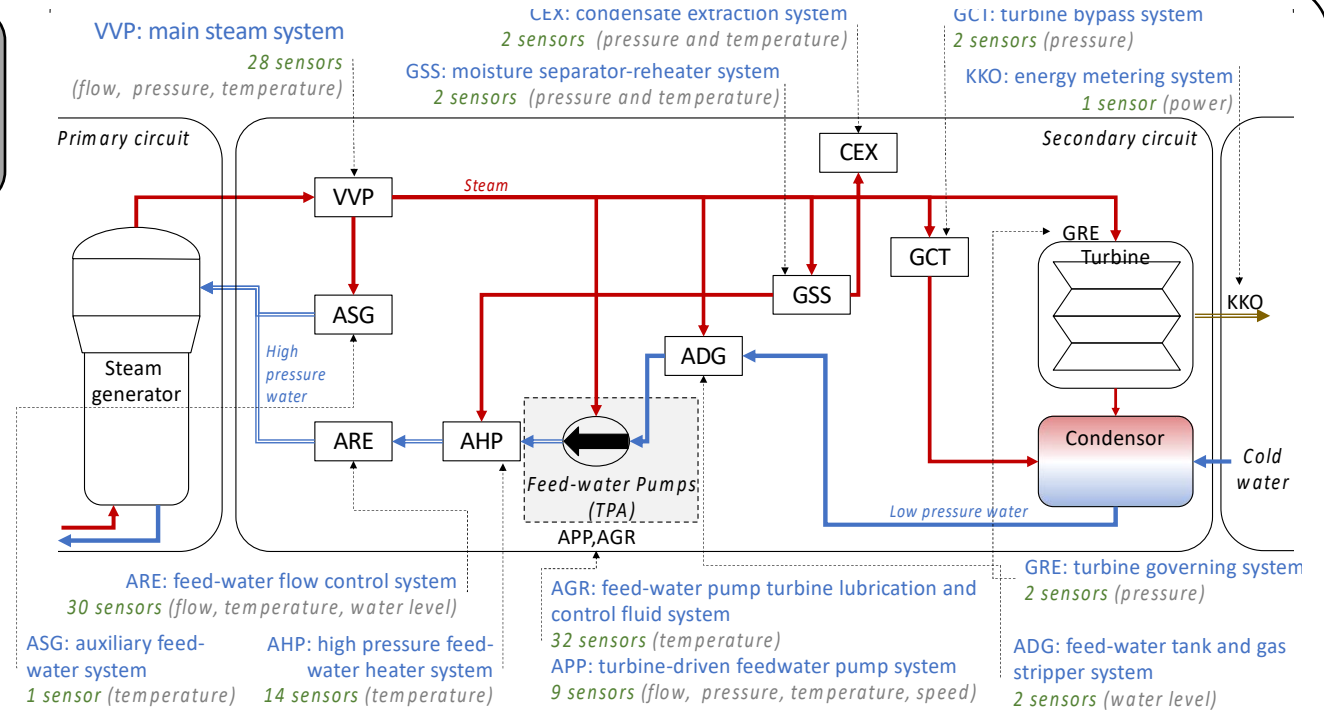
Training dataset

- Normal examples

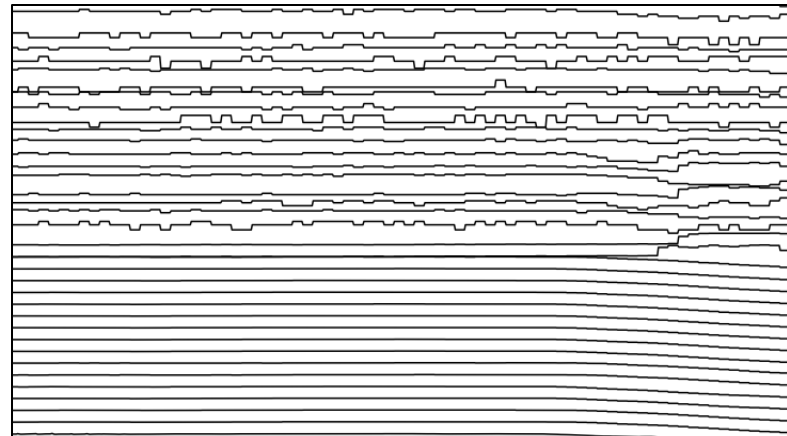
- Anomaly examples

Time Series T

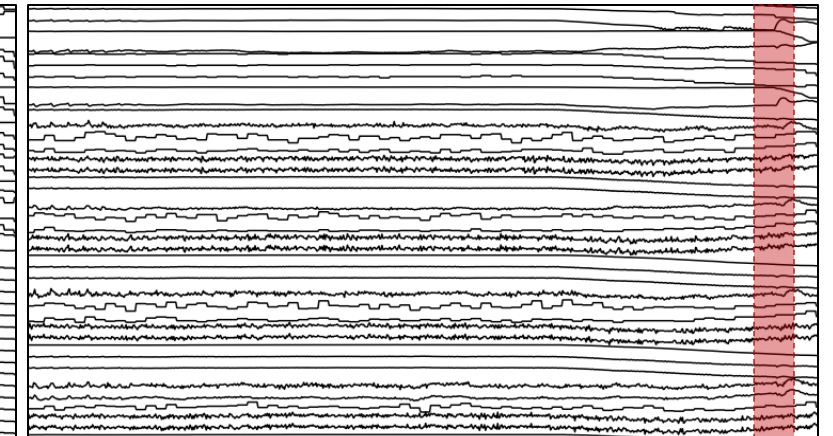
Supervised anomaly detection (e.g., classification)



Class 1: Time series without any vibrations



Class 2: Time series with a vibrations



Vibration

Anomaly Detection

By inputs...

Time

Supervised

Supervised
anomaly
detection (e.g.,
classification)

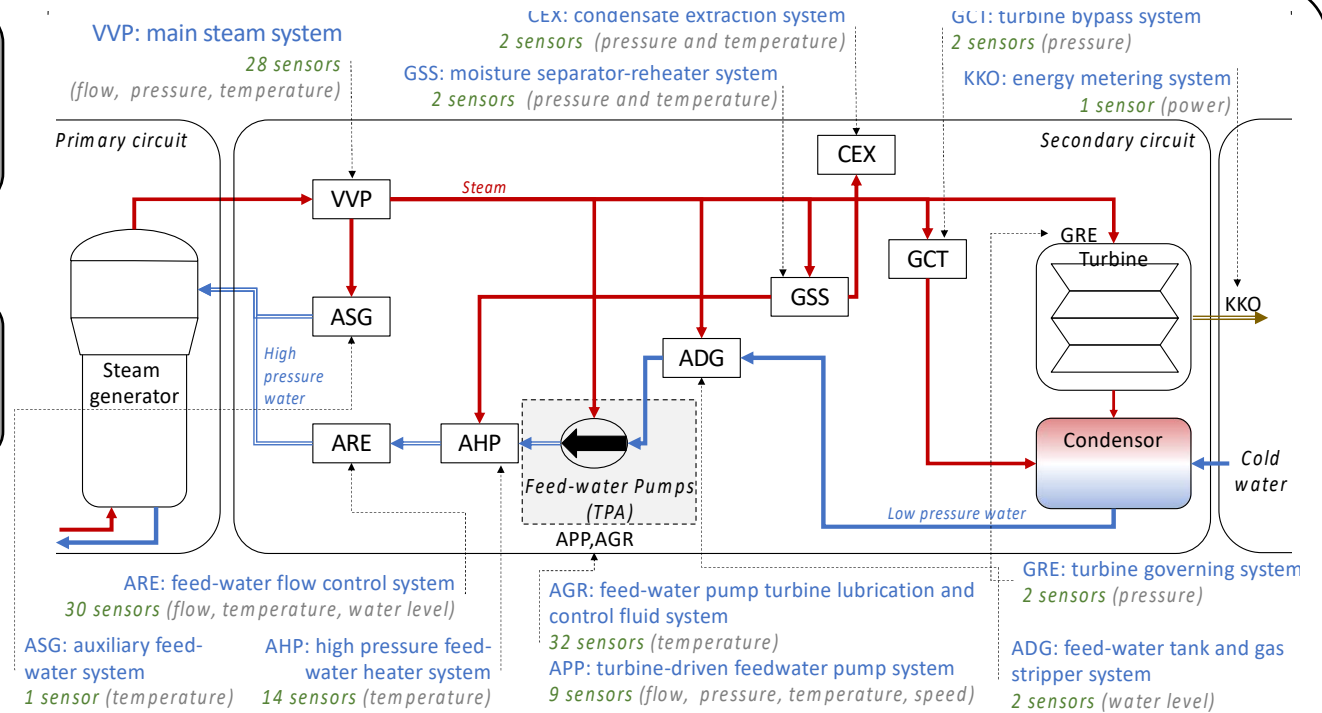
Explanation
of the detection

Training
dataset

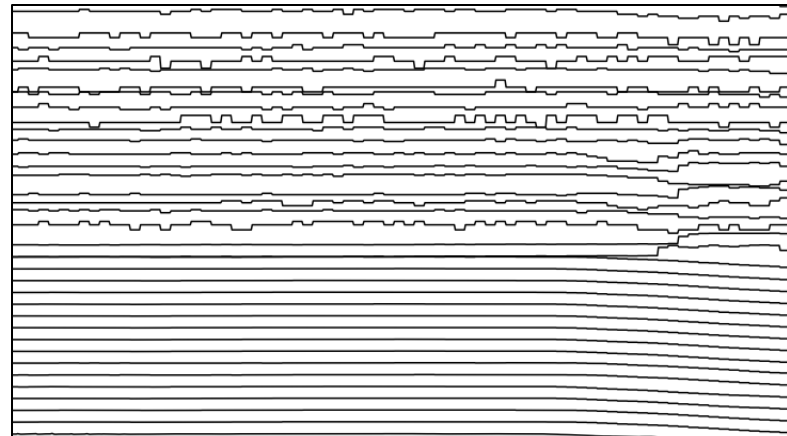
- Normal examples

- Anomaly examples

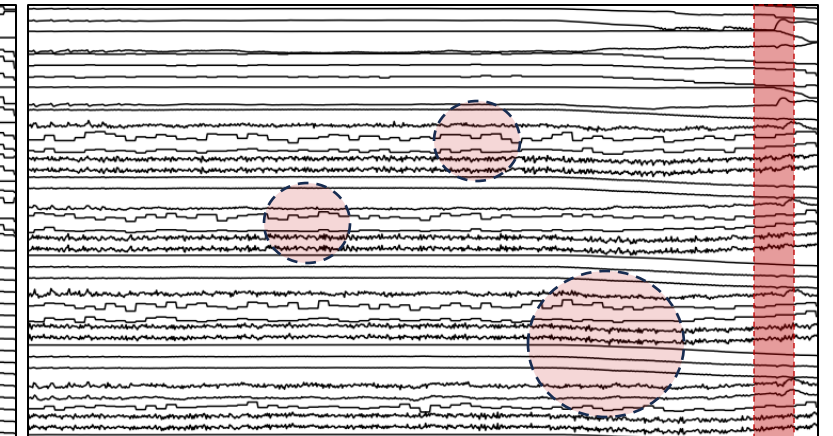
Time Series T



Class 1: Time series without any vibrations



Class 2: Time series with a vibrations



Vibration

Anomaly Detection

By inputs...

Time

Supervised

Training dataset

- Normal examples

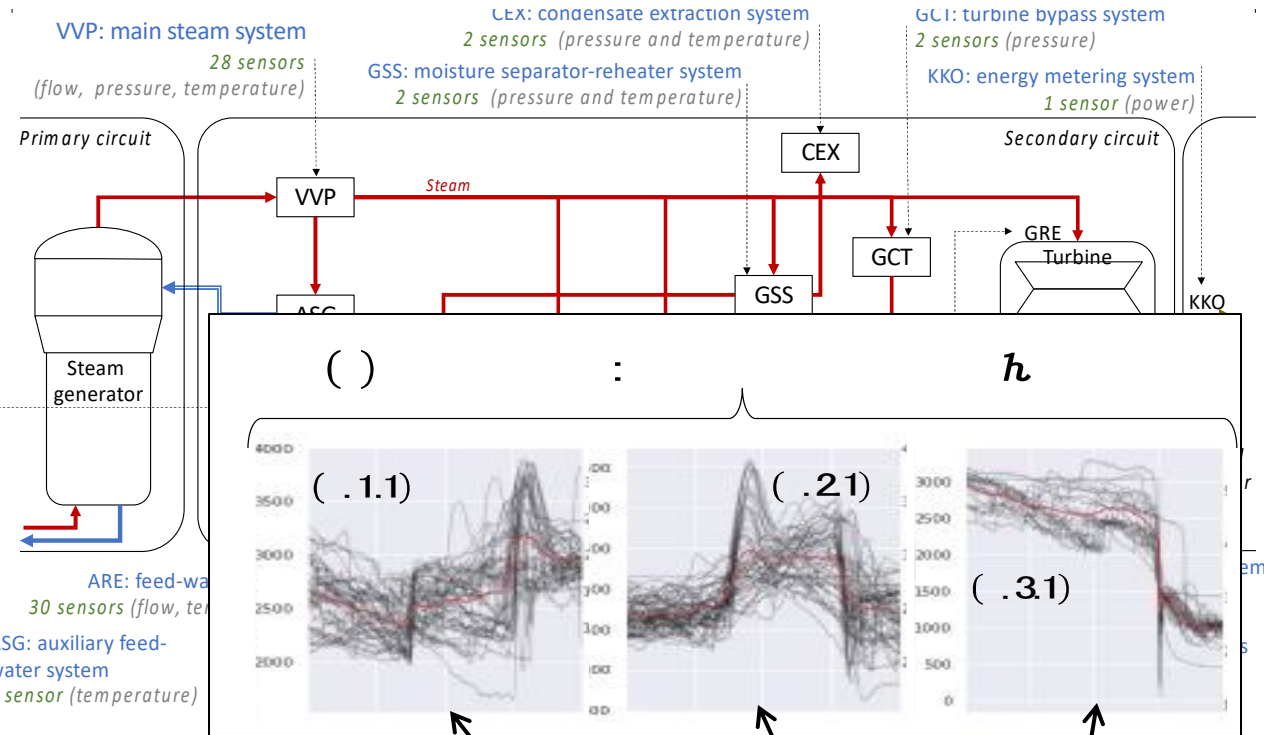
- Anomaly examples

Time Series T

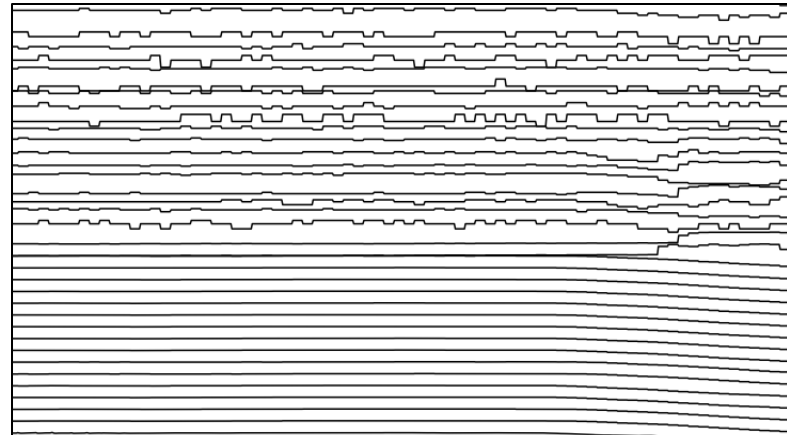
Supervised anomaly detection (e.g., classification)

Explanation of the detection

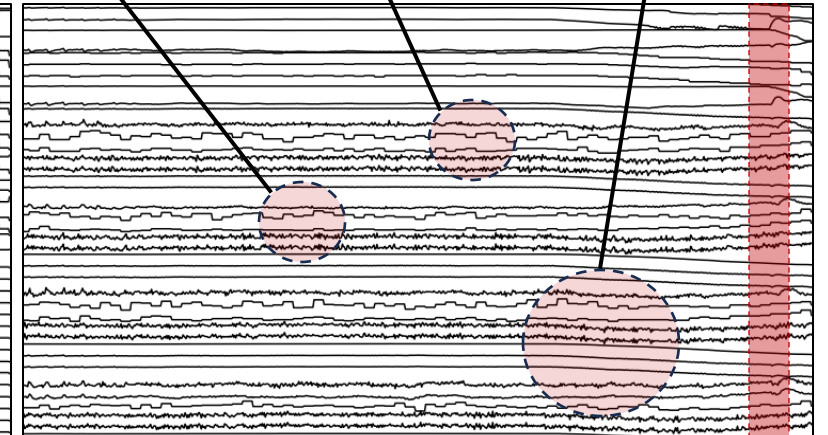
Identification of precursors



Class 1: Time series without any vibrations



Class 2: Time series with a vibrations



Vibration

Anomaly Detection

By inputs...

Time

Supervised

Training dataset

- Normal examples

- Anomaly examples

Time Series T

Supervised anomaly detection (e.g., classification)

Explanation of the detection

Identification of precursors

Class 1: Time

More info :

On the use case

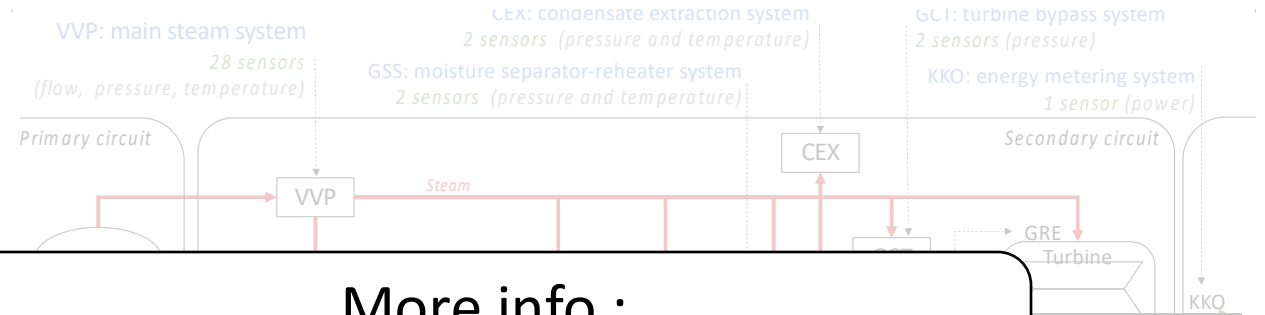


DCE journal 2023

On the method



SIGMOD 2022



ations

Vibration

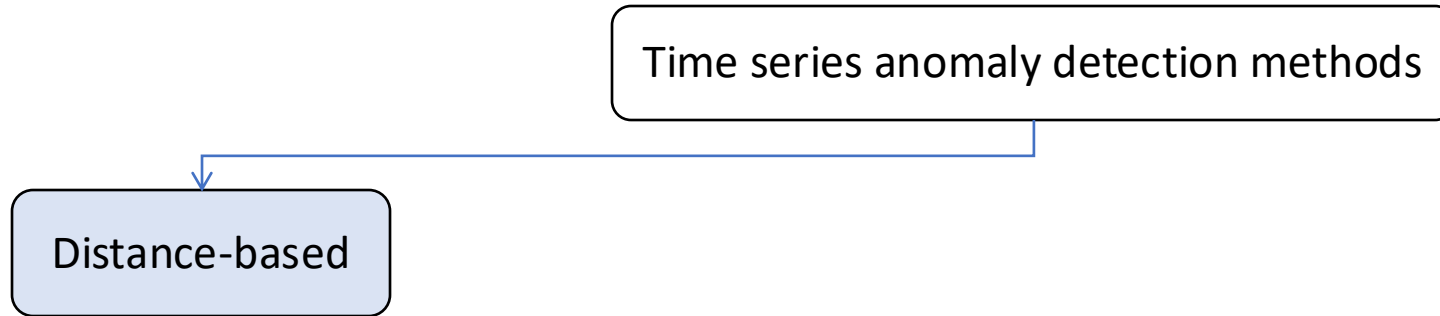
Anomaly Detection methods: *A taxonomy*

By methods...

Time series anomaly detection methods

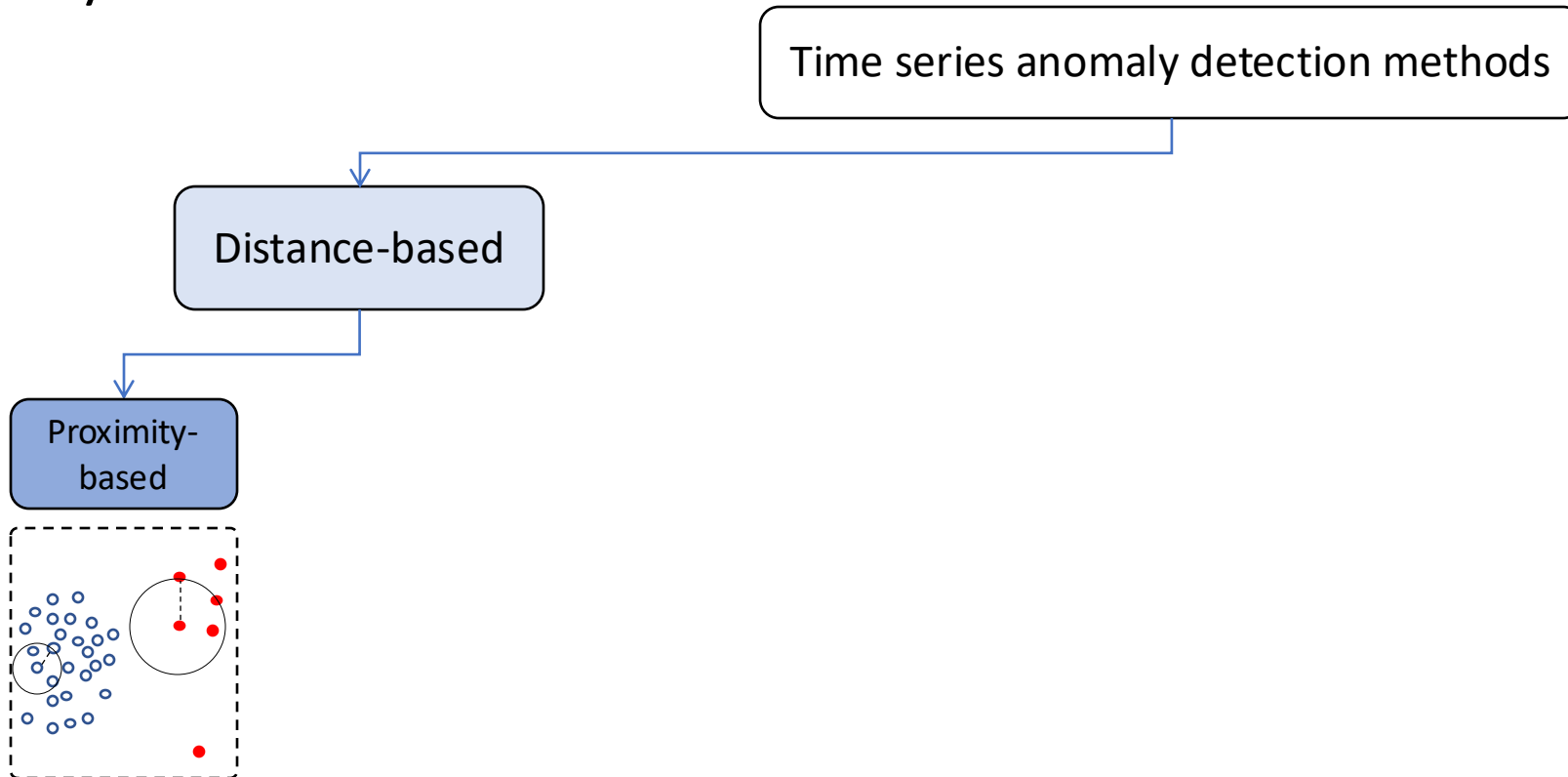
Anomaly Detection methods: *A taxonomy*

By methods...



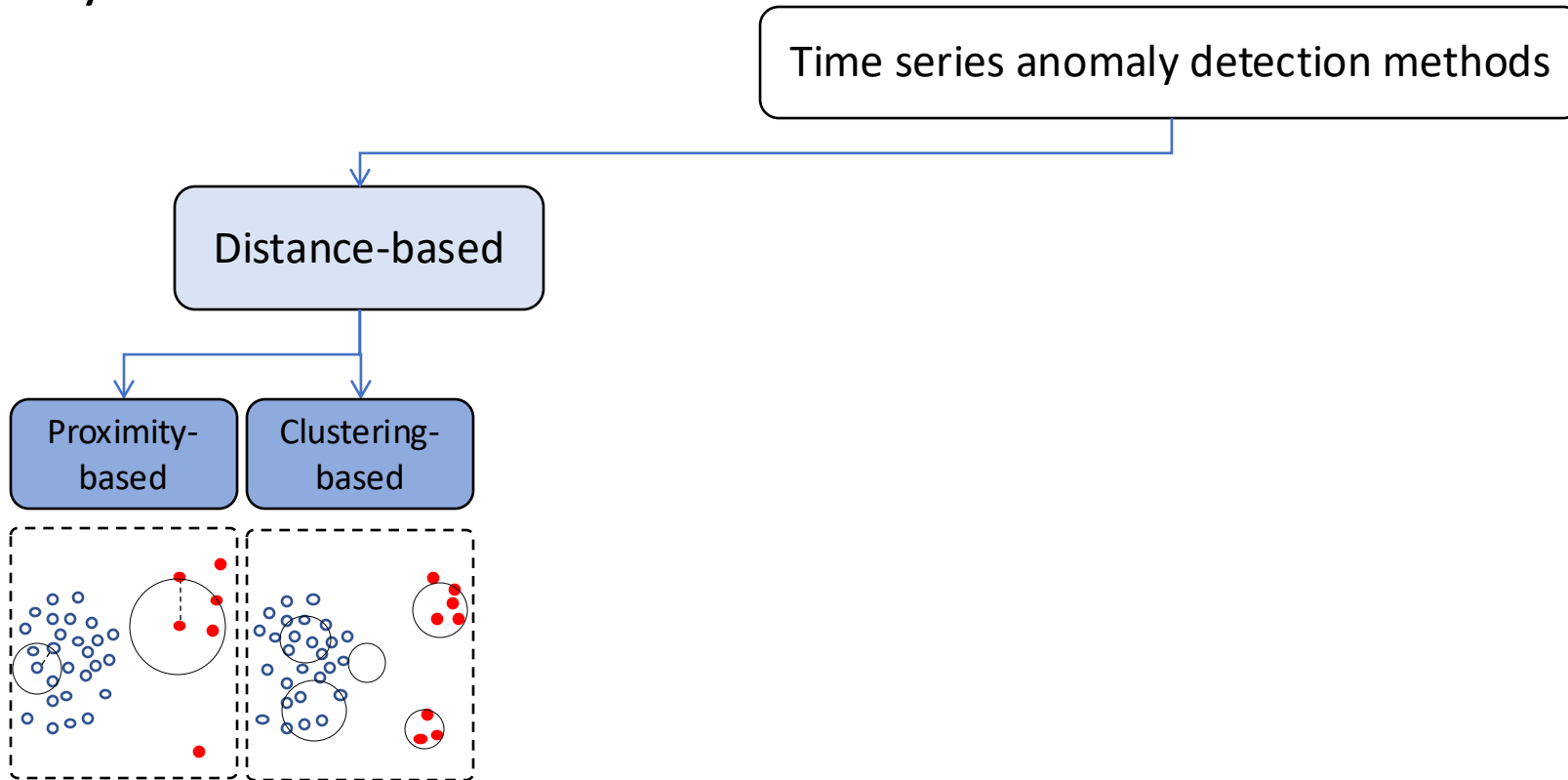
Anomaly Detection methods: *A taxonomy*

By methods...



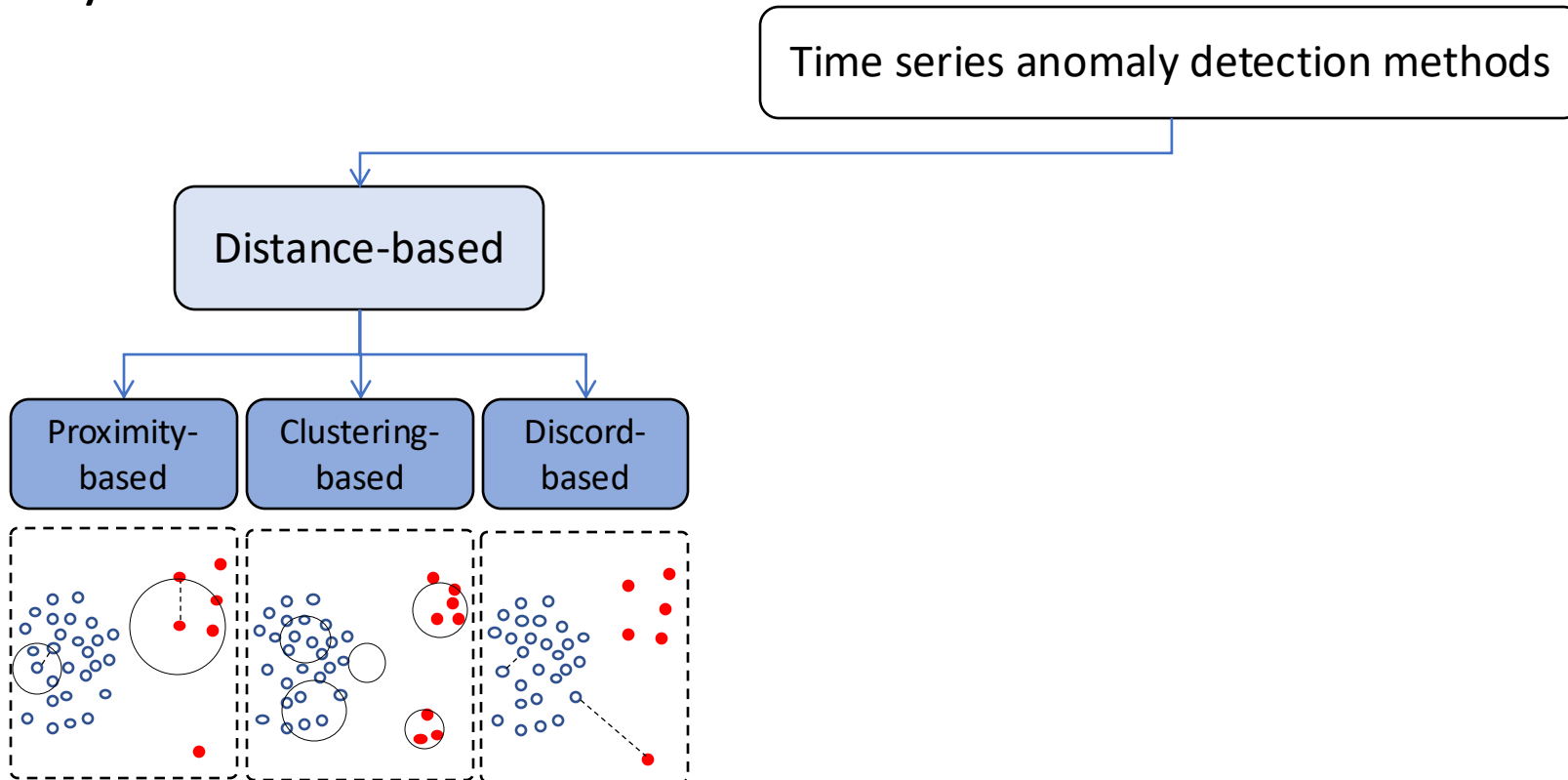
Anomaly Detection methods: *A taxonomy*

By methods...



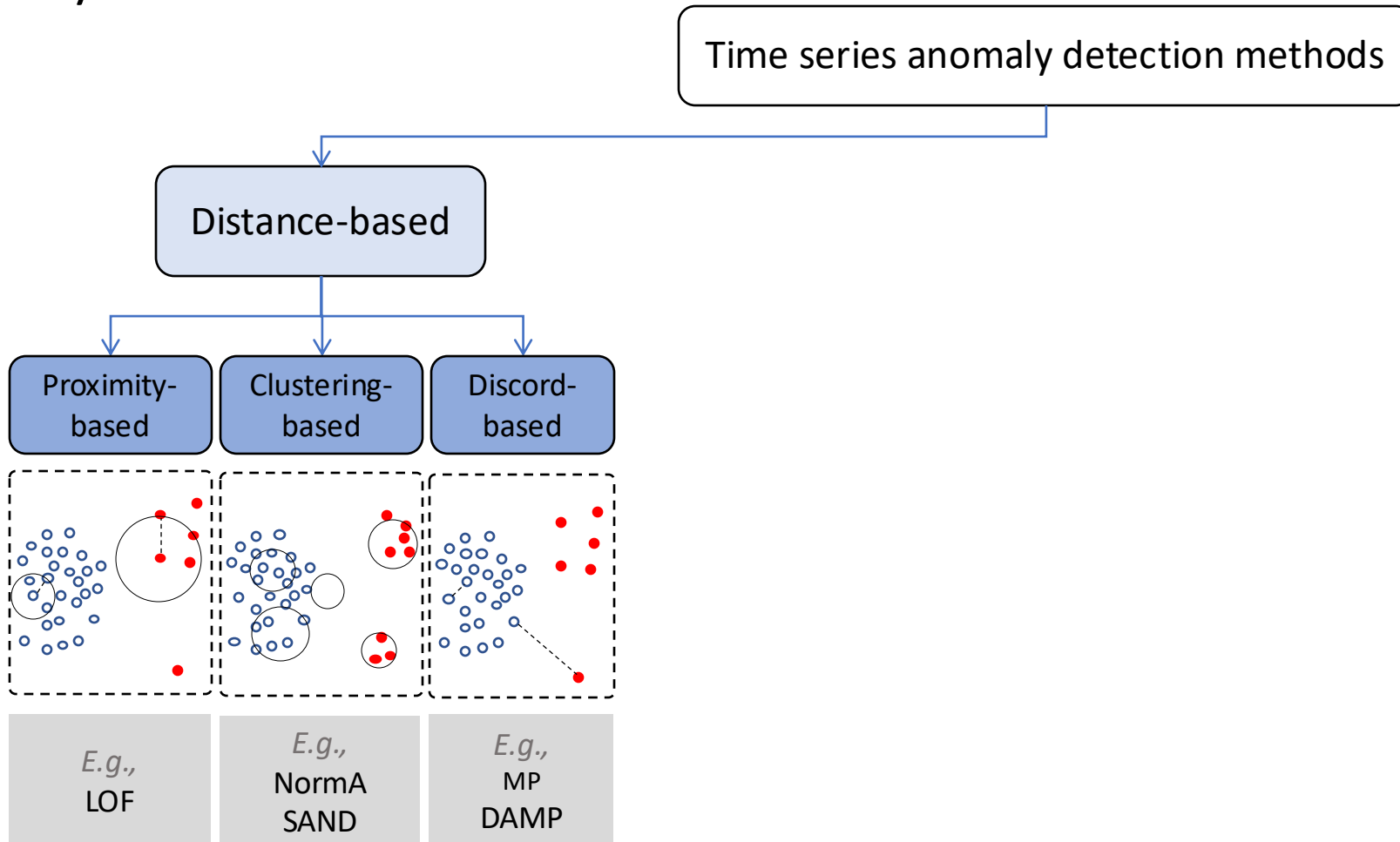
Anomaly Detection methods: *A taxonomy*

By methods...



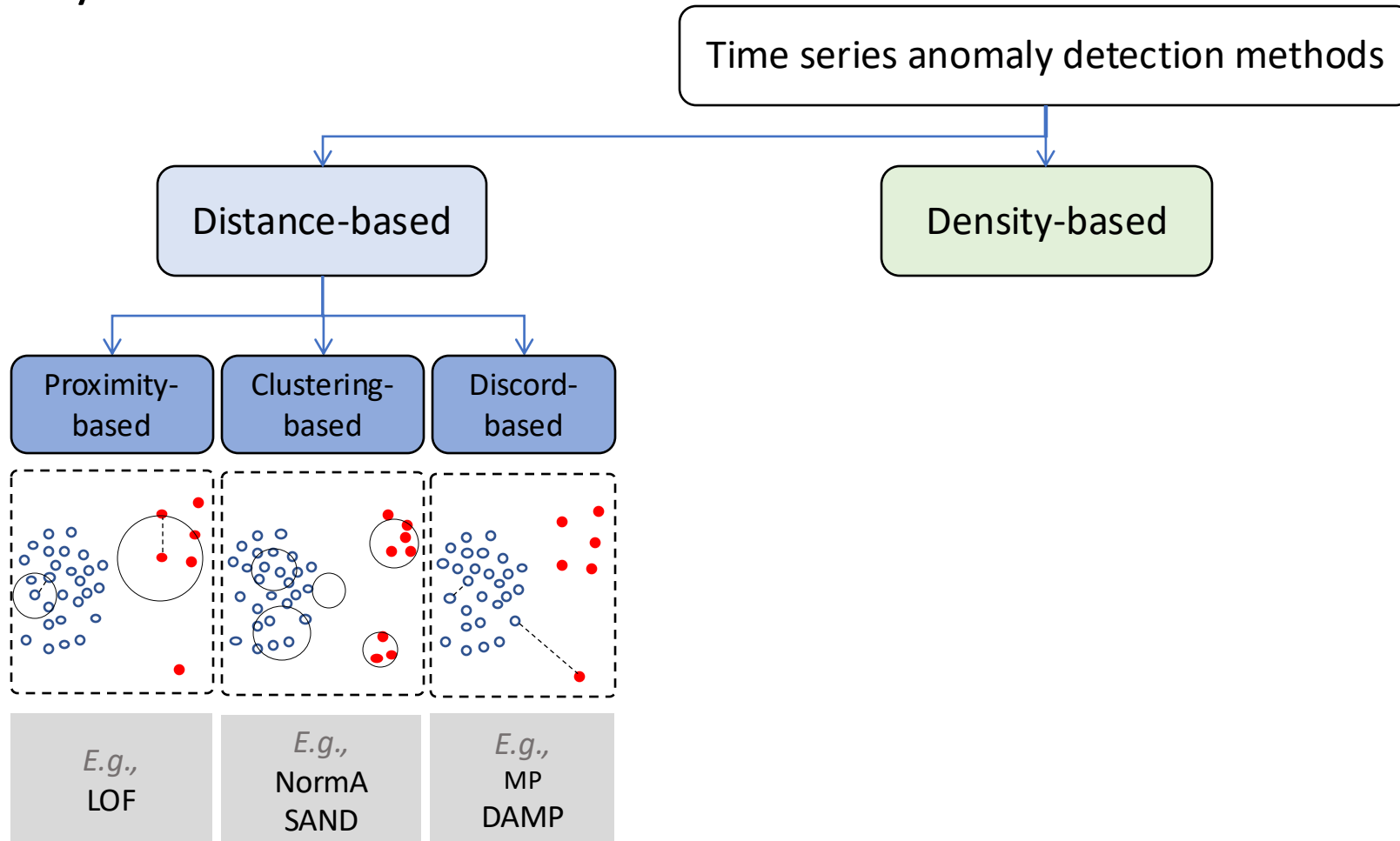
Anomaly Detection methods: *A taxonomy*

By methods...



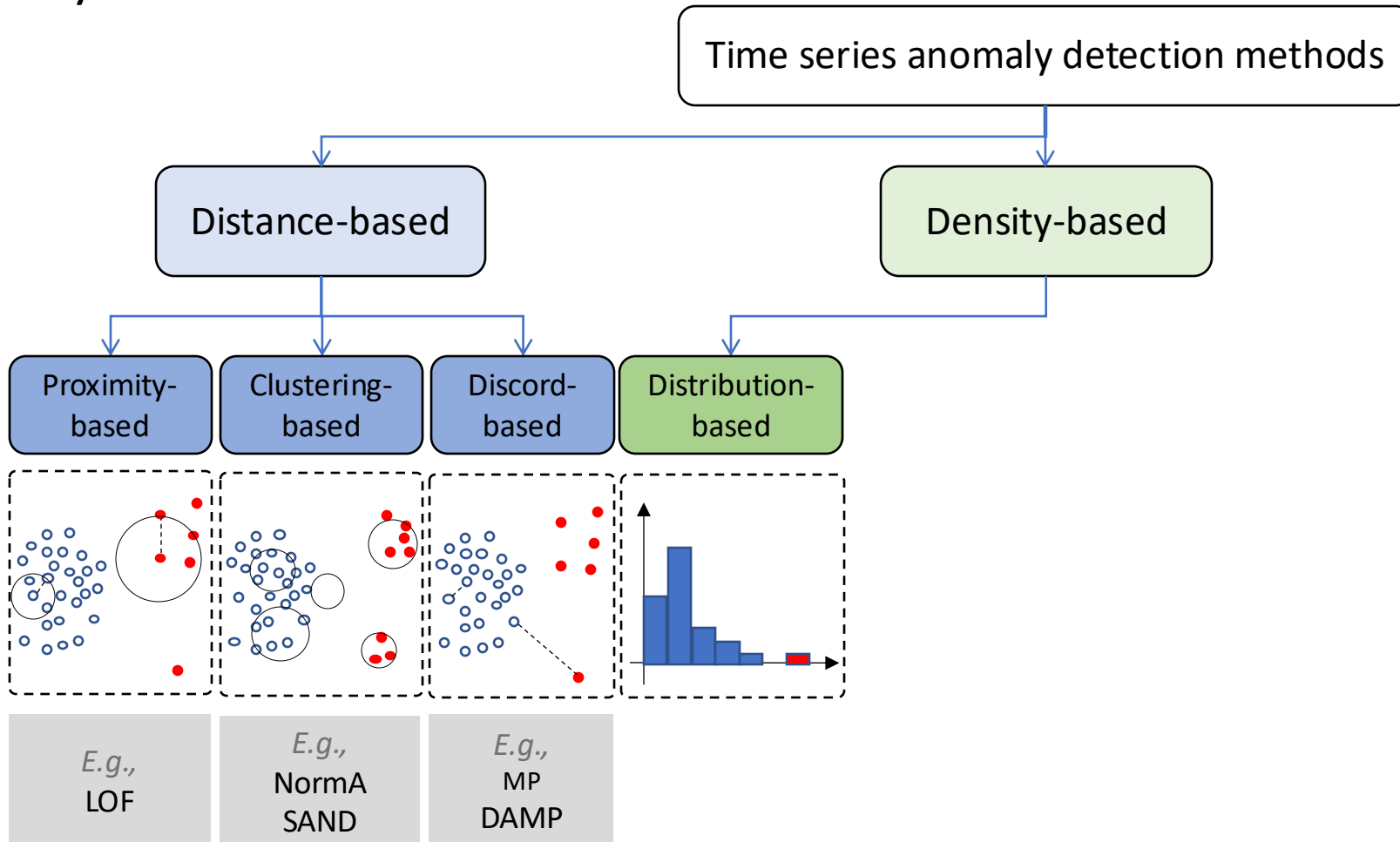
Anomaly Detection methods: *A taxonomy*

By methods...



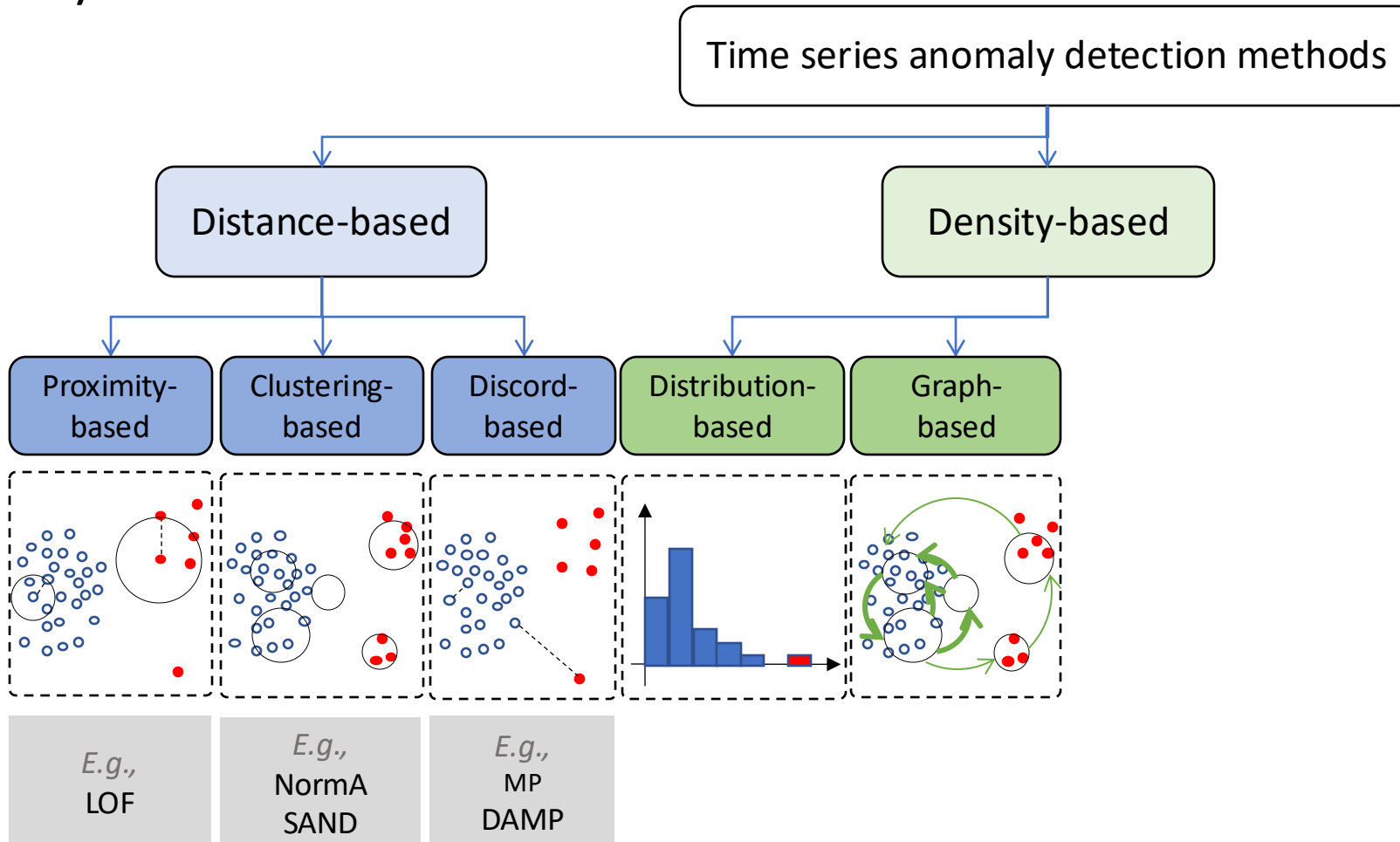
Anomaly Detection methods: *A taxonomy*

By methods...



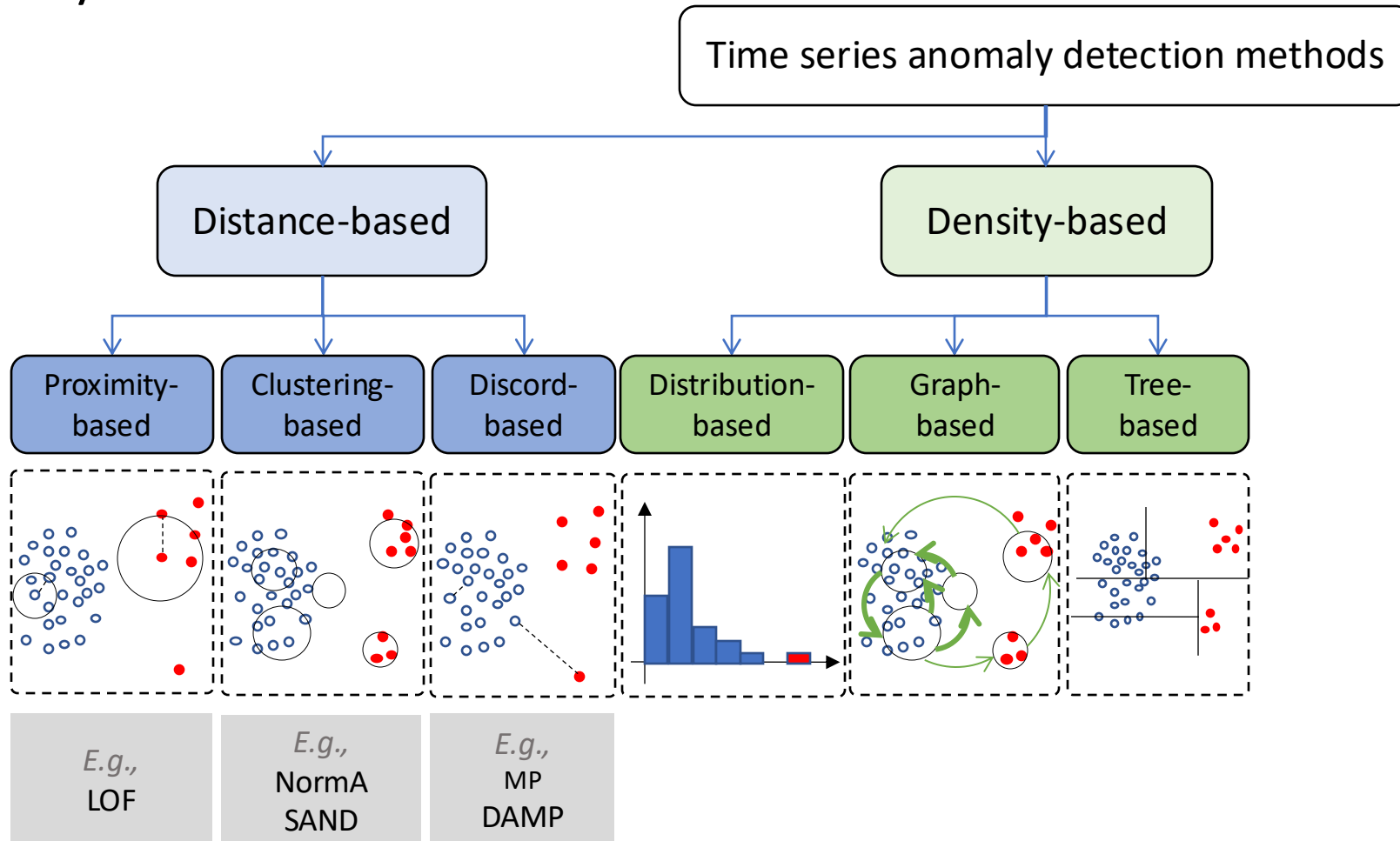
Anomaly Detection methods: *A taxonomy*

By methods...



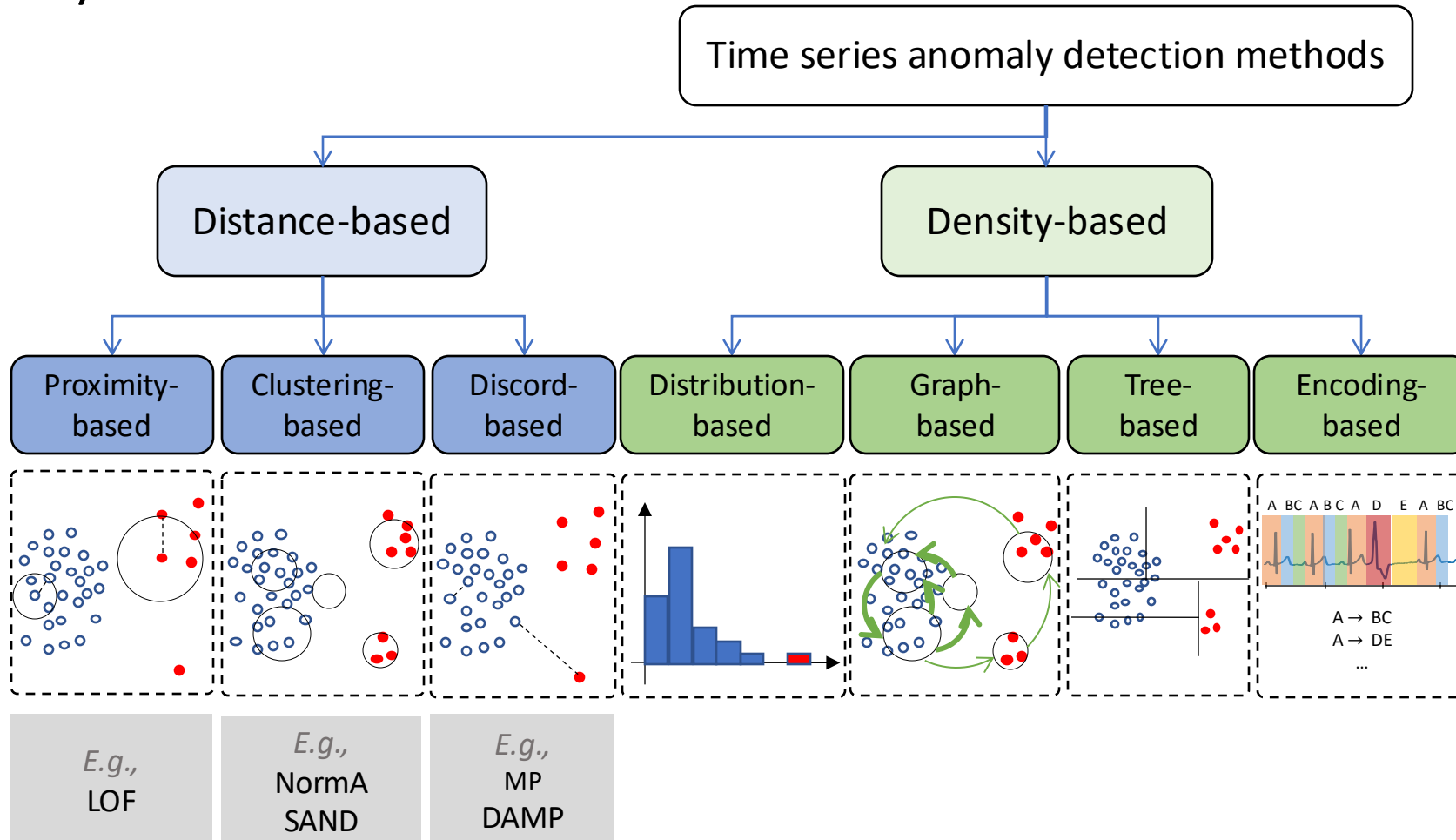
Anomaly Detection methods: *A taxonomy*

By methods...



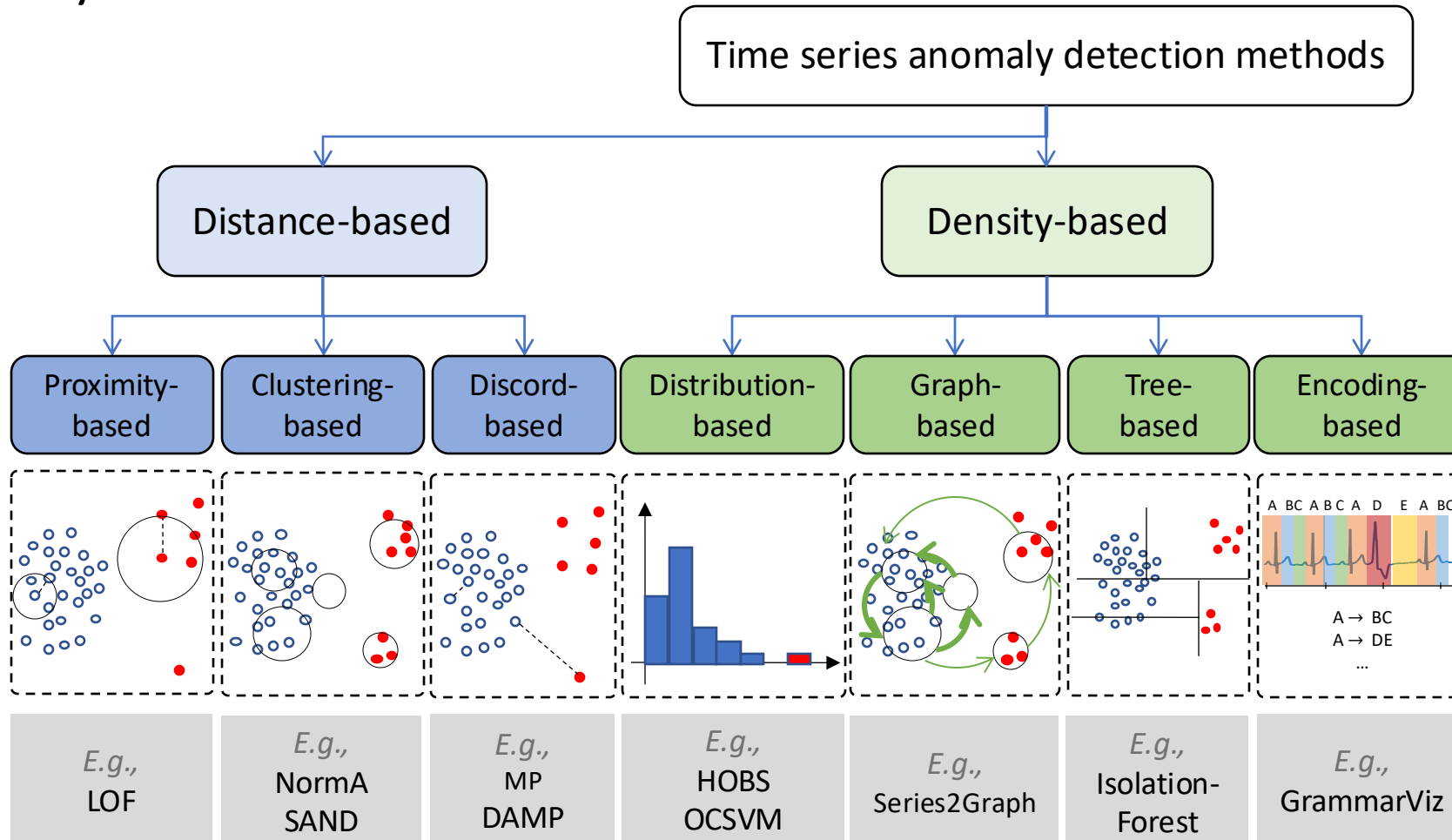
Anomaly Detection methods: *A taxonomy*

By methods...



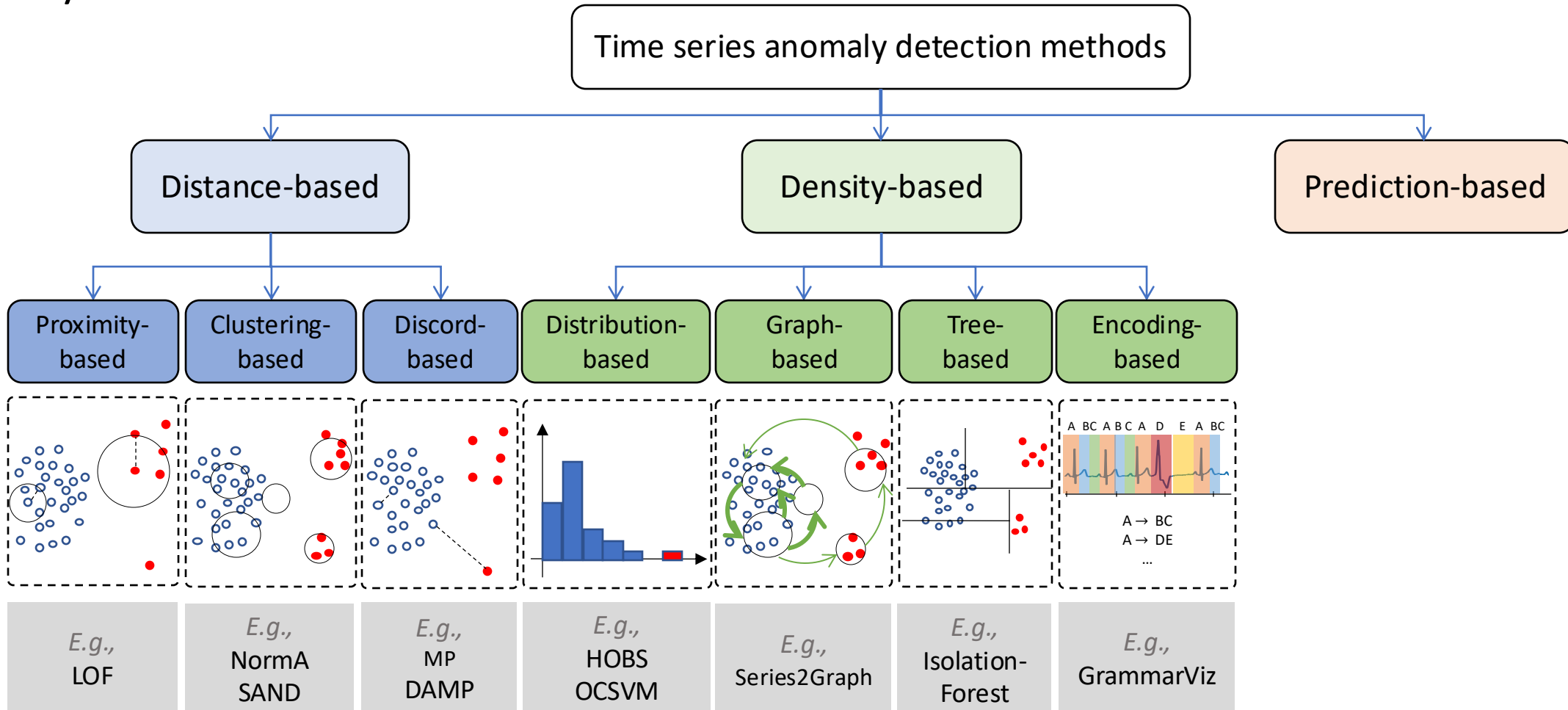
Anomaly Detection methods: *A taxonomy*

By methods...



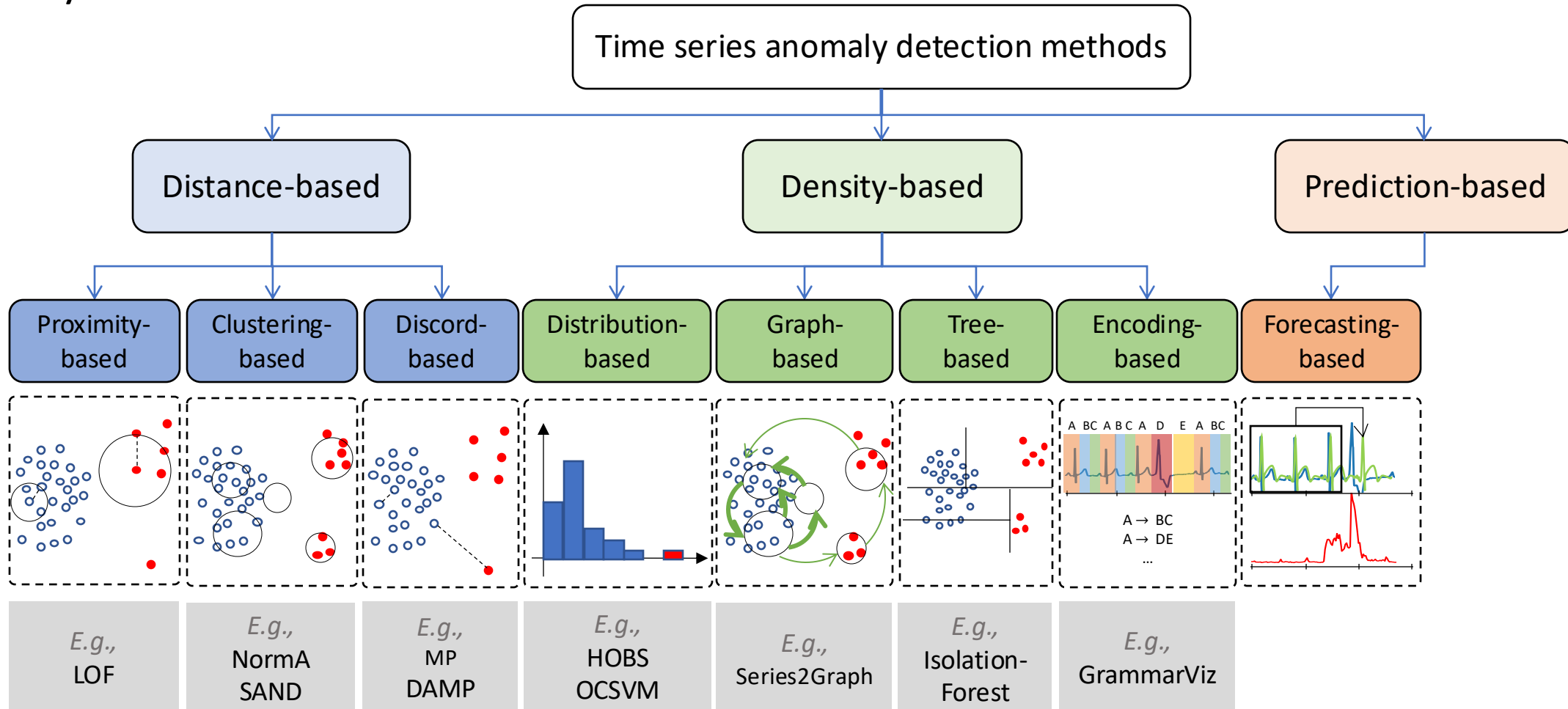
Anomaly Detection methods: *A taxonomy*

By methods...



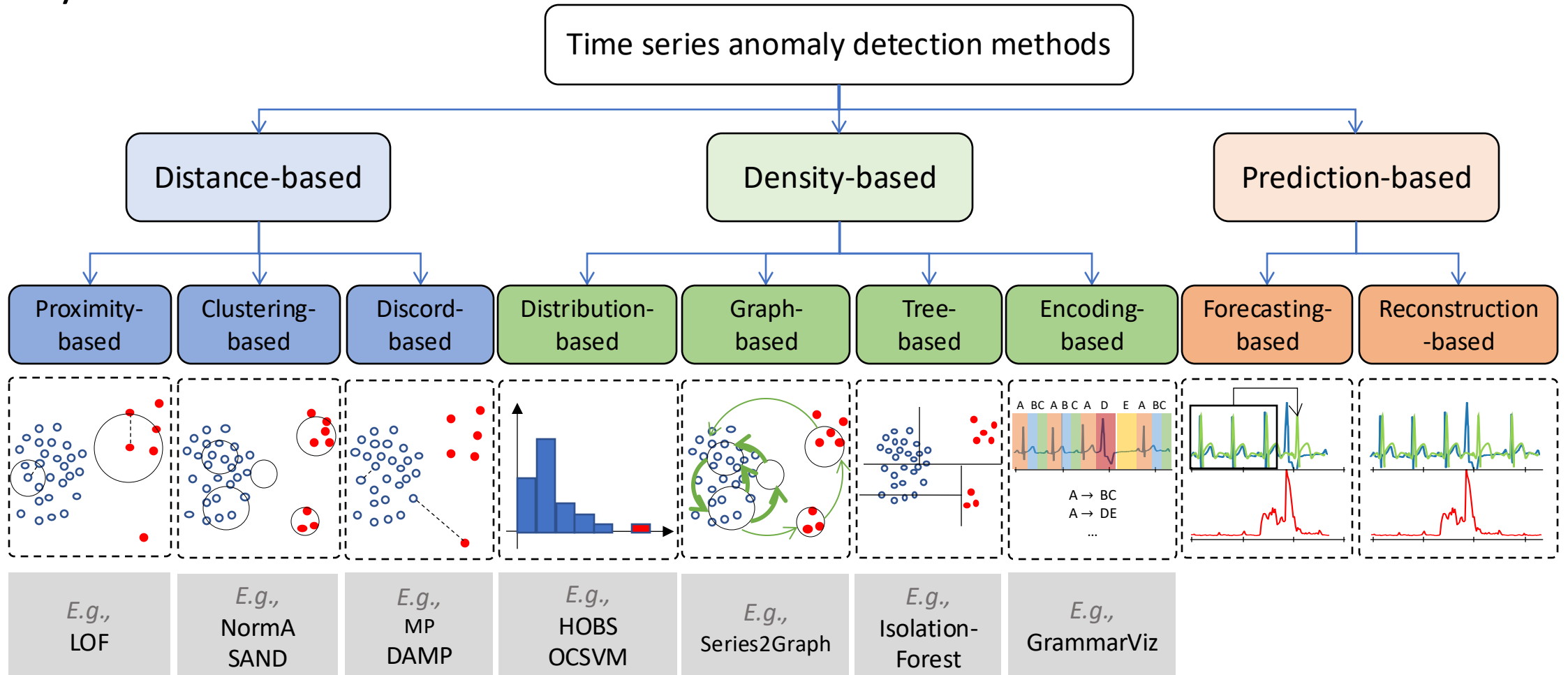
Anomaly Detection methods: *A taxonomy*

By methods...



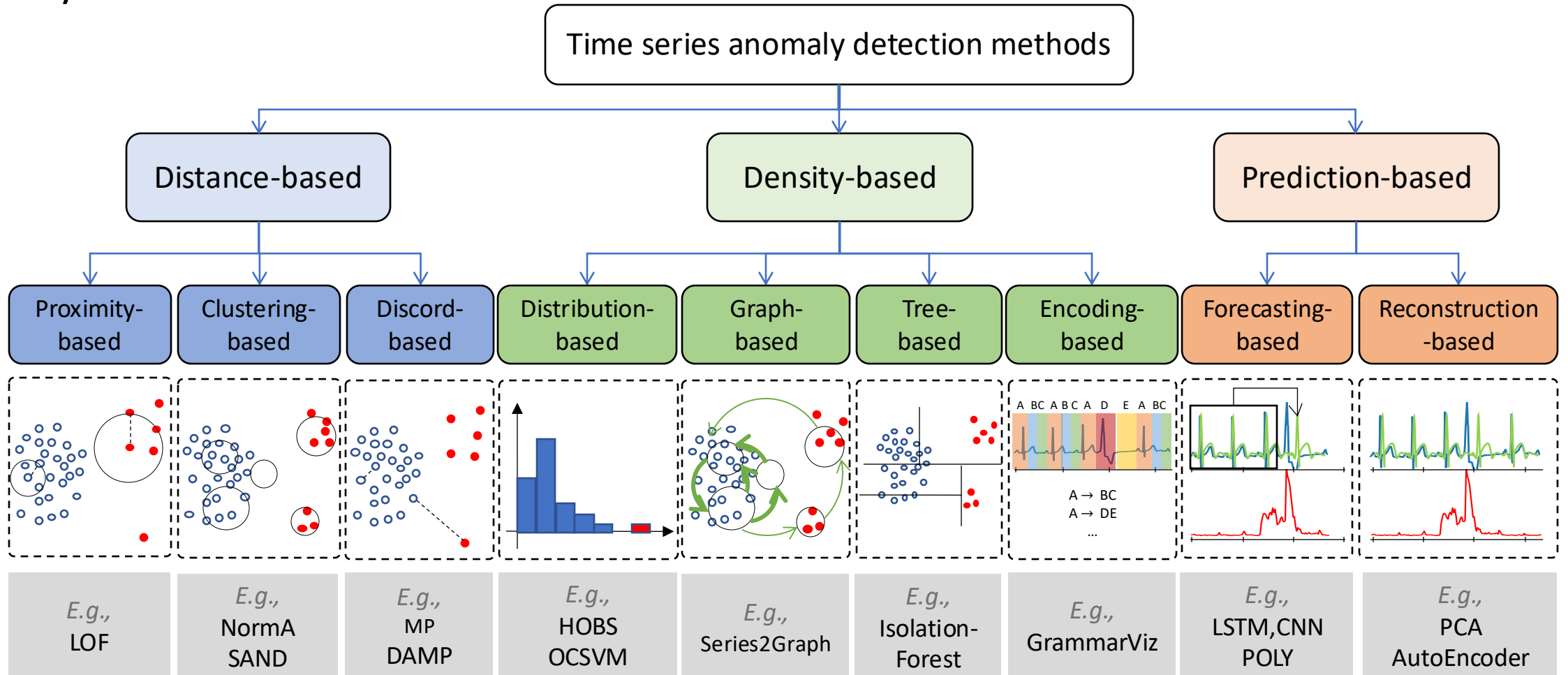
Anomaly Detection methods: *A taxonomy*

By methods...



Anomaly Detection methods: *A taxonomy*

By methods...

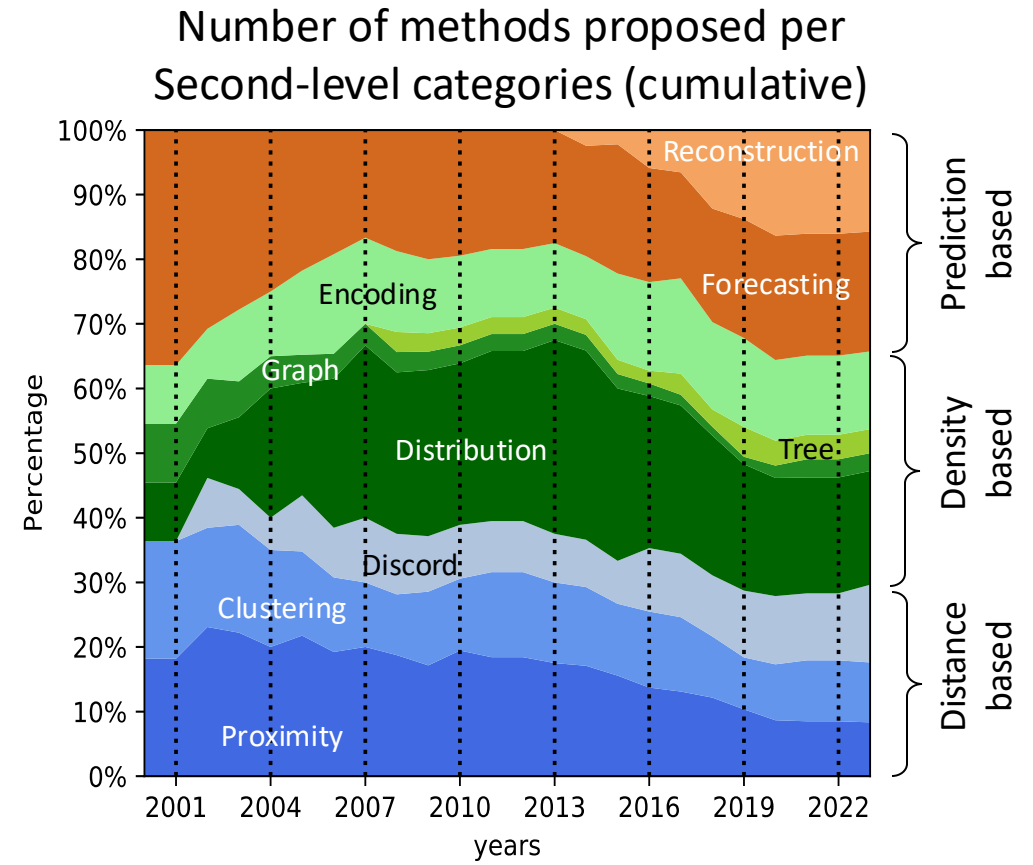
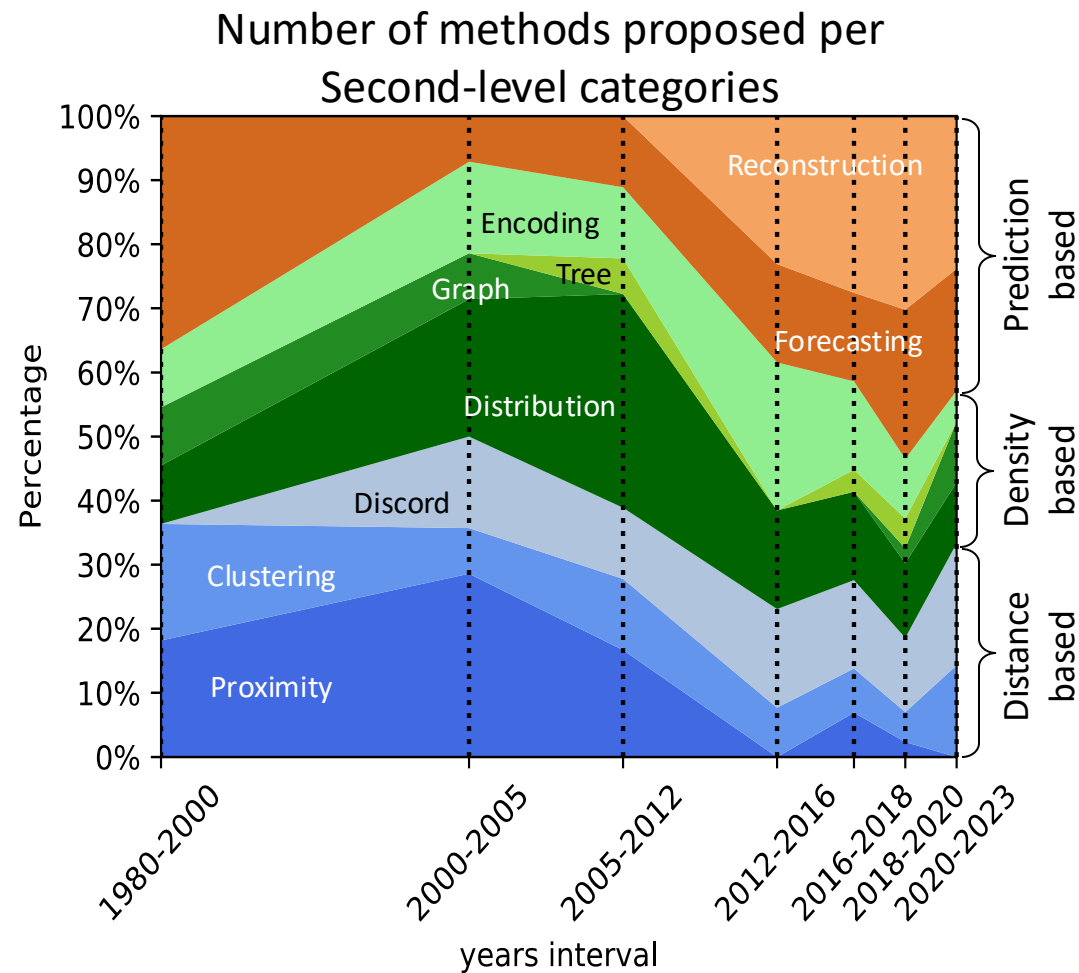


By time...



Anomaly Detection methods: *A taxonomy*

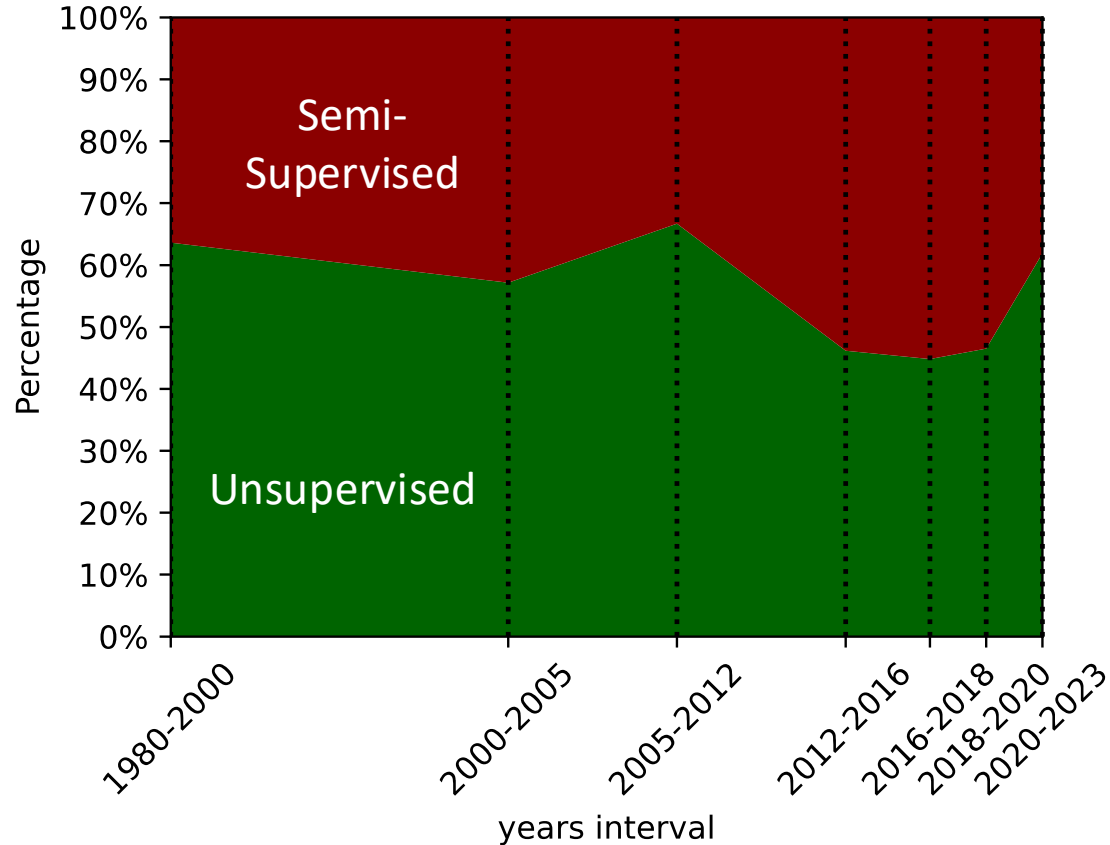
By time...



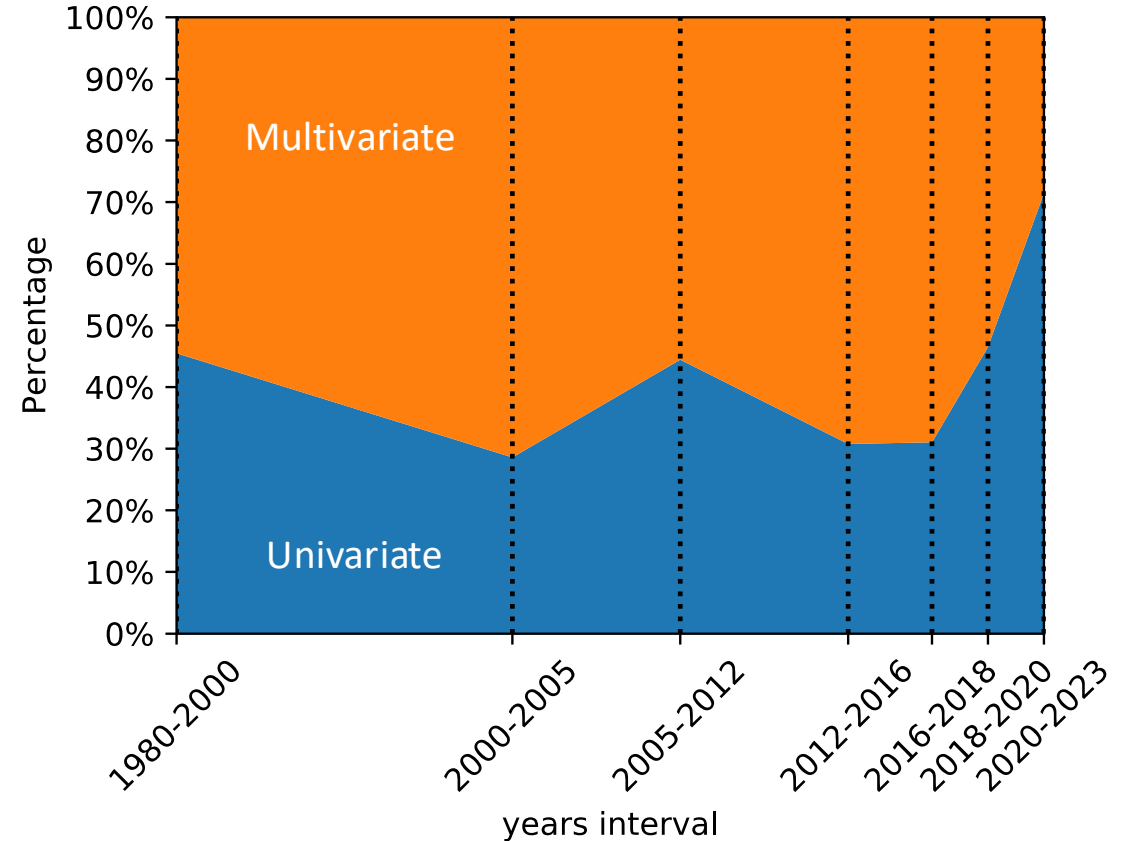
Anomaly Detection methods: *A taxonomy*

By time...

Number of methods proposed that are
Unsupervised or *Semi-Supervised*

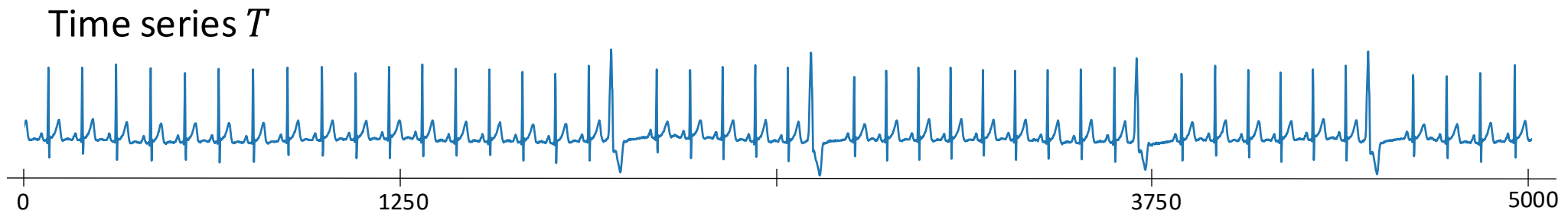


Number of methods proposed that can handle
Univariate or *Multivariate* time series



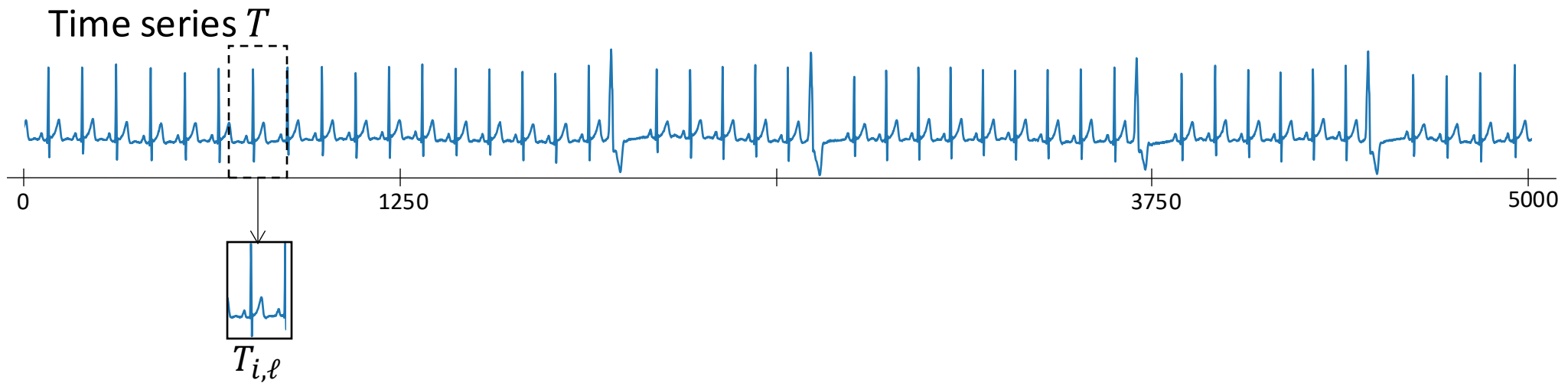
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



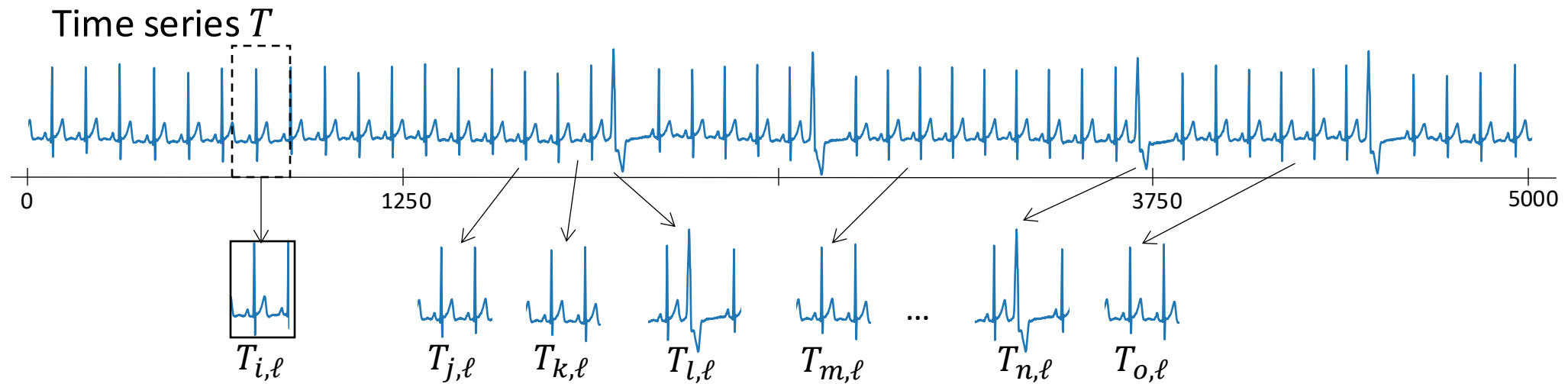
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



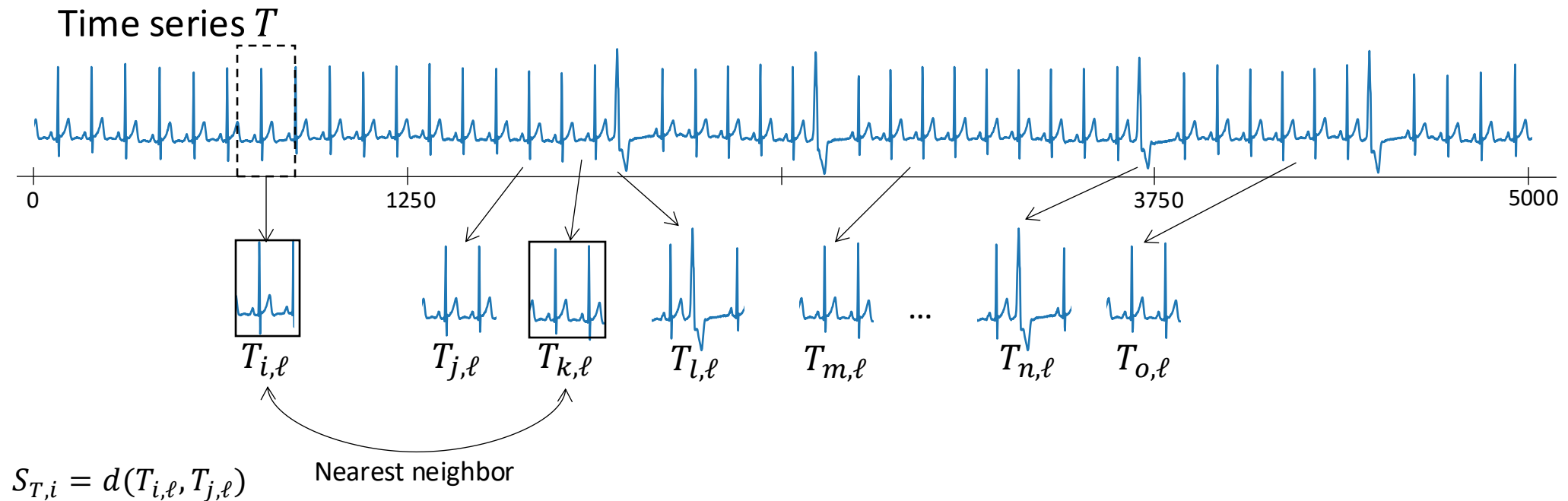
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



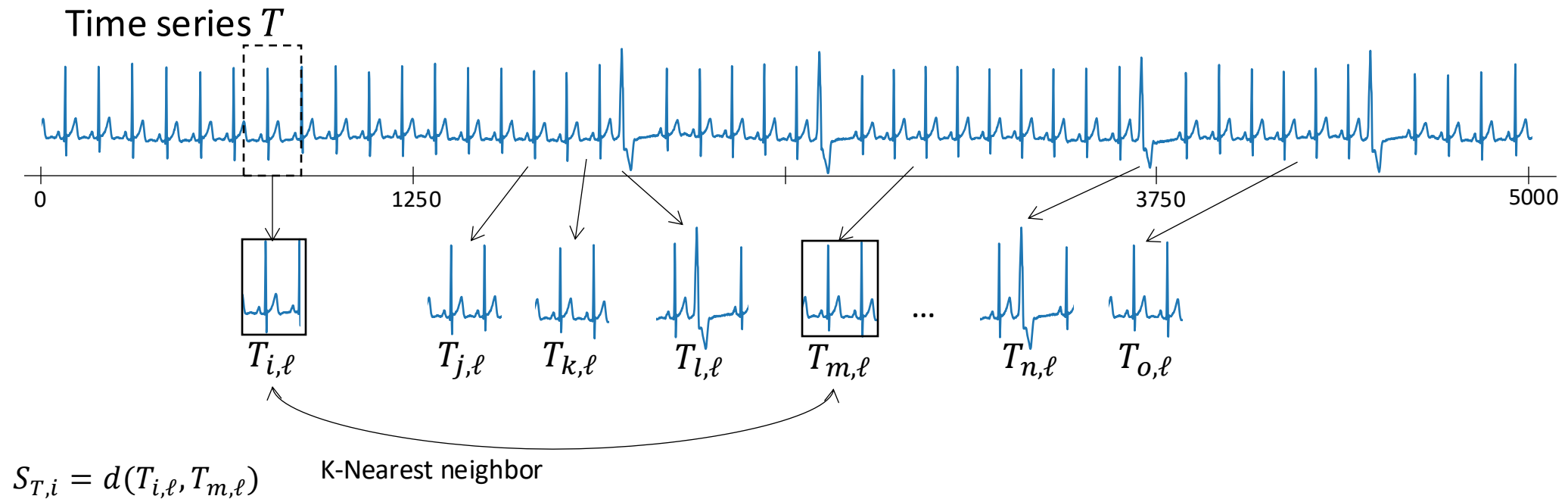
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



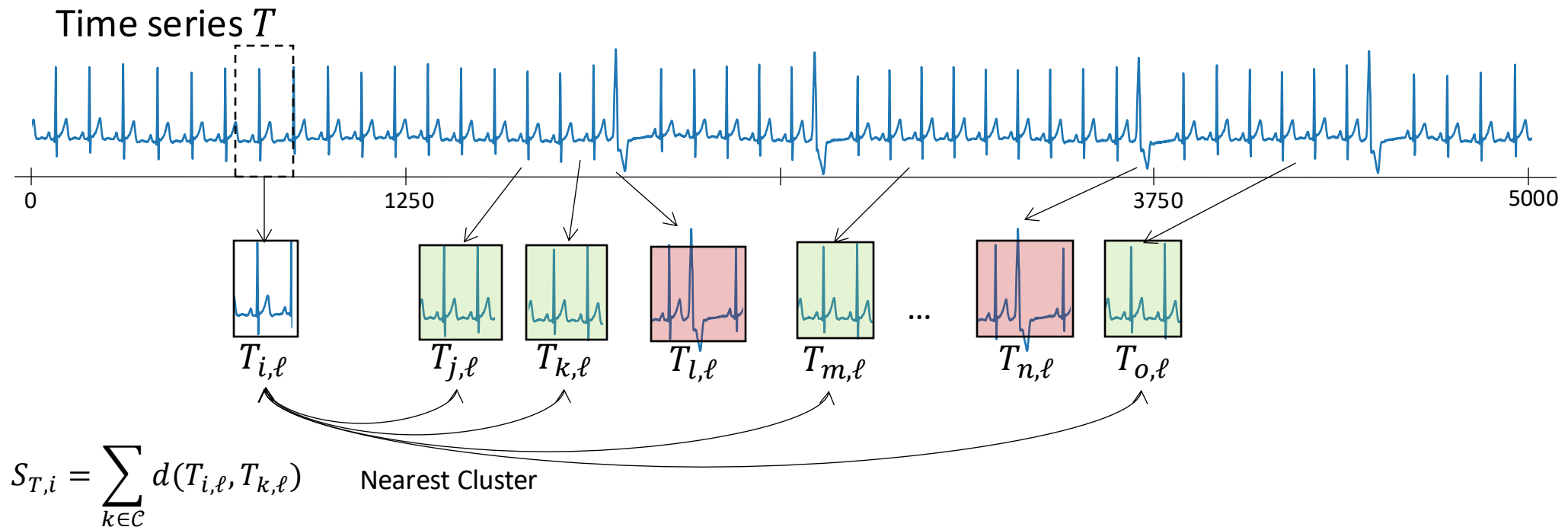
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



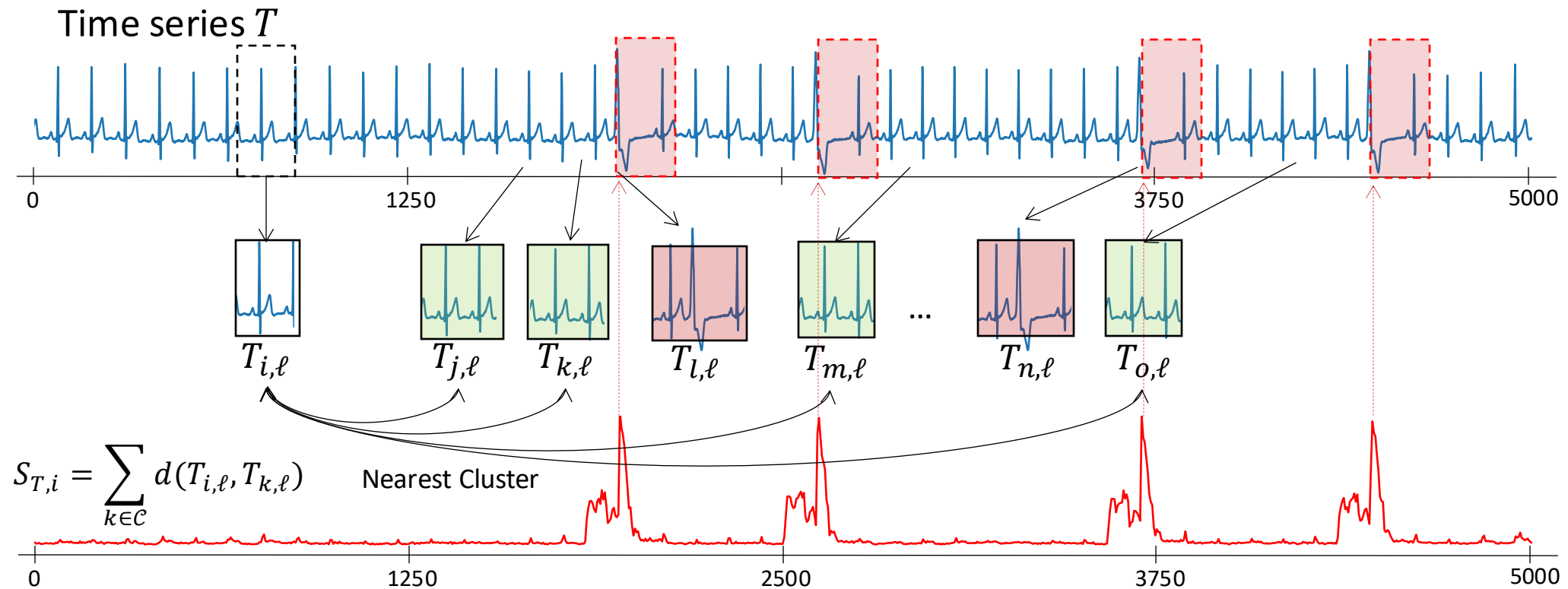
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



Anomaly Detection methods: *Distance-based*

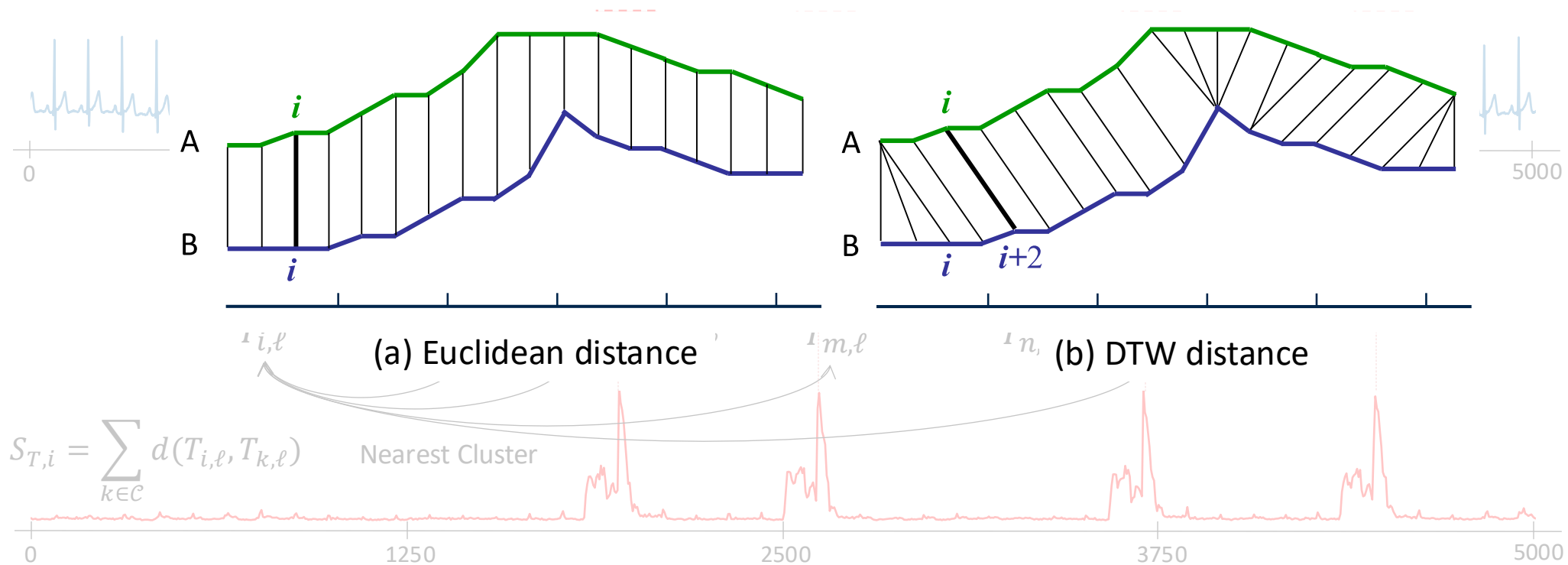
Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.



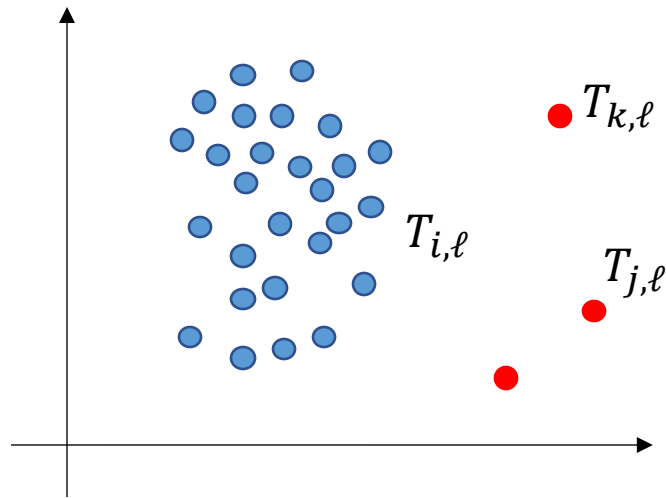
Anomaly Detection methods: *Distance-based*

Methods that use **distance computation** between subsequences (or group of subsequences) to detect anomalies.

Example of distance computation



Anomaly Detection methods: *an Example*



Matrix Profile [6] (MP)

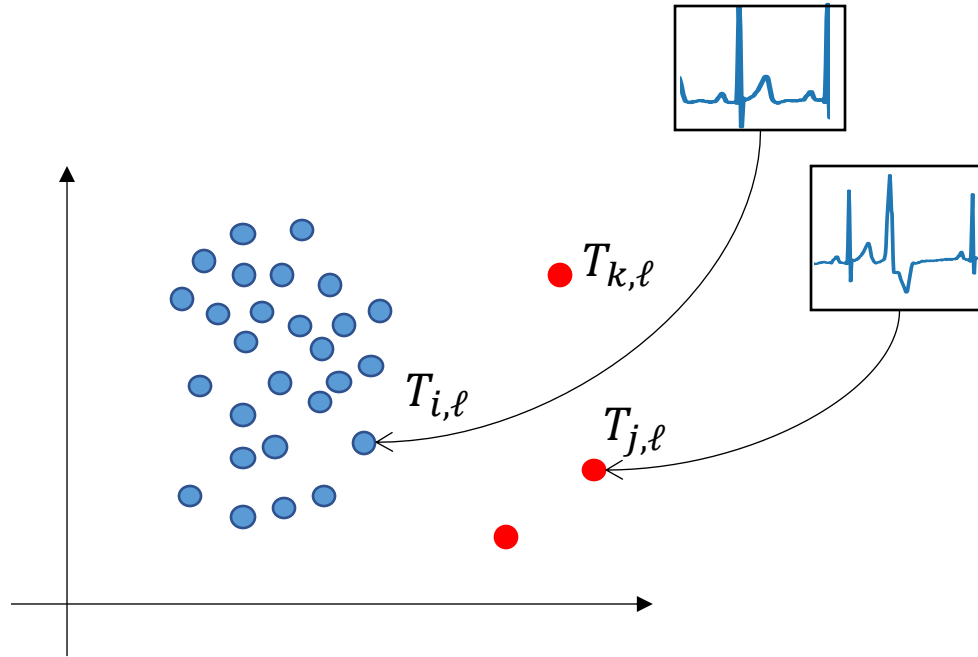
Compute the **distance to the nearest neighbor** (using the MASS algorithm z-norm Euclidean distance computation) and use it as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



Matrix Profile [6] (MP)

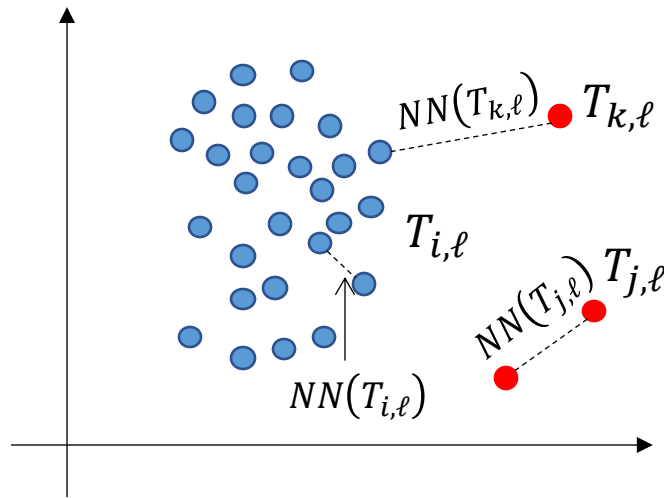
Compute the **distance to the nearest neighbor** (using the MASS algorithm z-norm Euclidean distance computation) and use it as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



Matrix Profile [6] (MP)

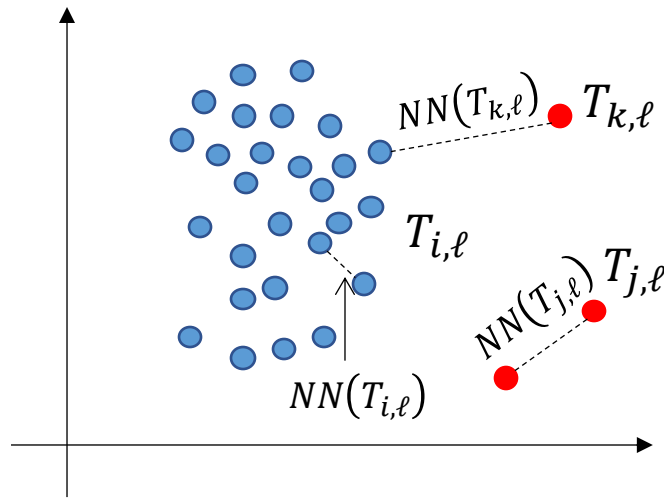
Compute the **distance to the nearest neighbor** (using the MASS algorithm z-norm Euclidean distance computation) and use it as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



The matrix Profile is computed as follows:

$$S_T = [NN(T_{0,\ell}), NN(T_{1,\ell}), \dots, NN(T_{|T|-\ell,\ell})]$$

Matrix Profile [6] (MP)

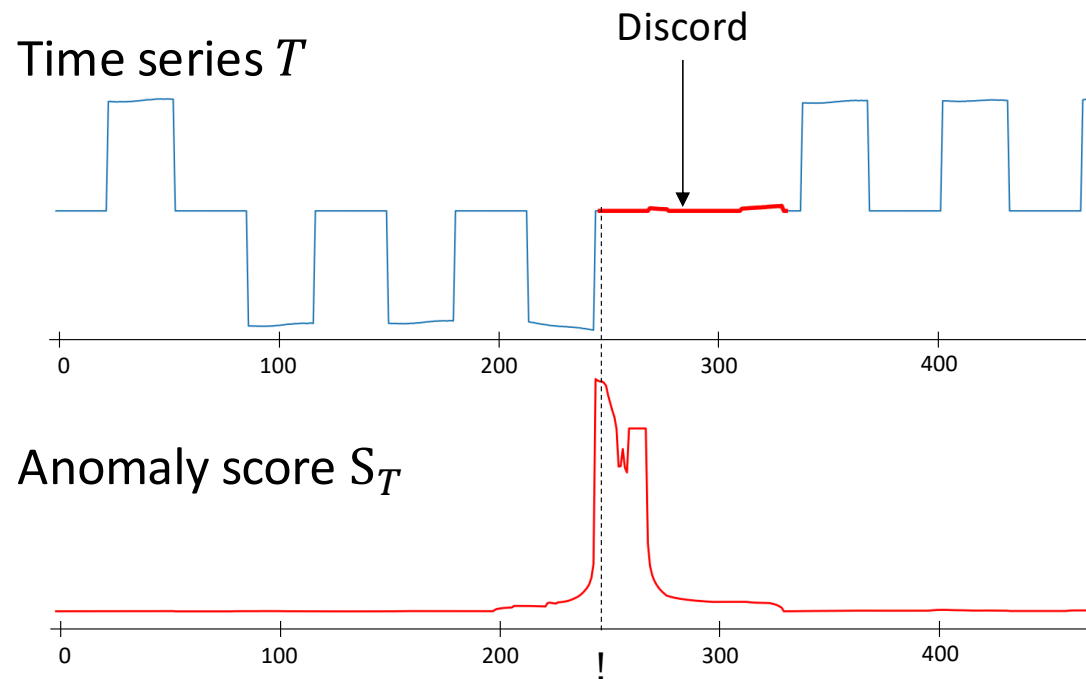
Compute the **distance to the nearest neighbor** (using the MASS algorithm z-norm Euclidean distance computation) and use it as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



Matrix Profile [6] (MP)

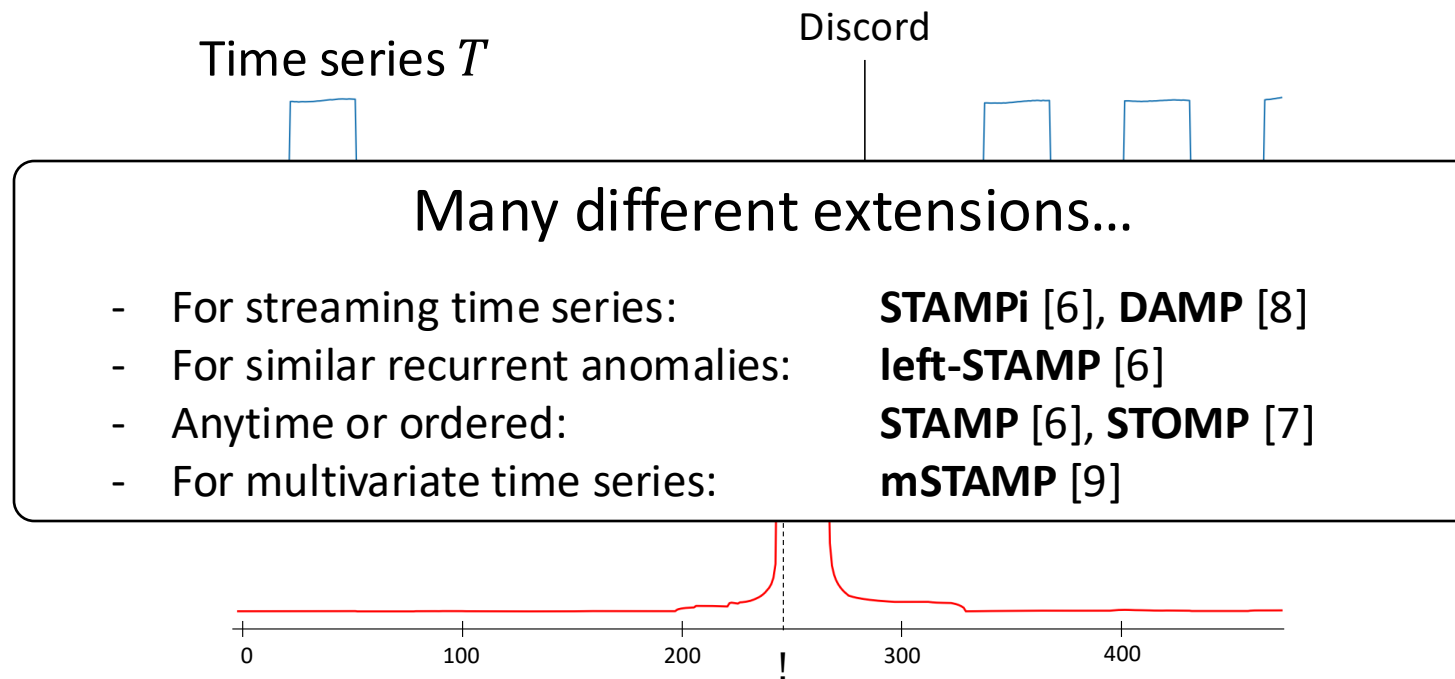
Compute the **distance to the nearest neighbor** (using the MASS algorithm z-norm Euclidean distance computation) and use it as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



Matrix Profile [6] (MP)

Compute the **distance to the nearest neighbor** (using the MASS algorithm z-norm Euclidean distance computation) and use it as anomaly score

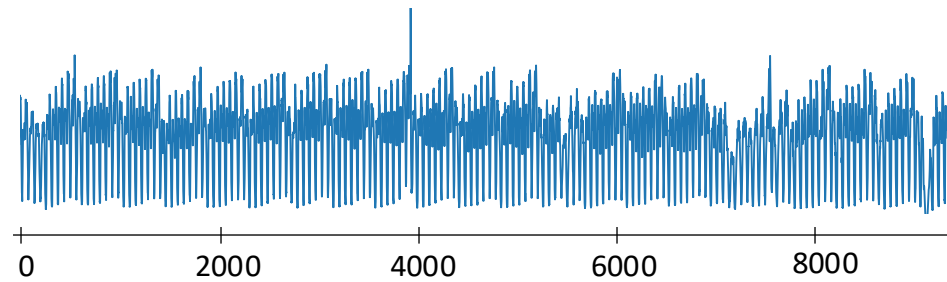
Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*

Time series T



NormA [10]

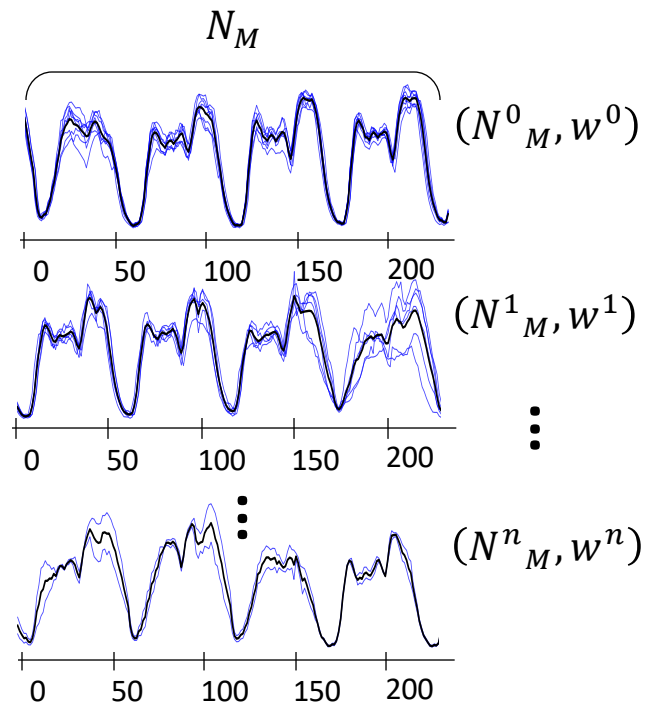
Distance-based approach that **summarize** the time series into **a weighted set of subsequences** and use the distance to them as anomaly score

Unsupervised

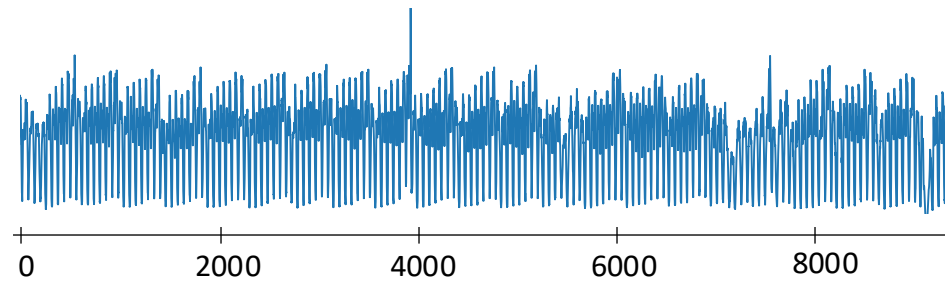
Univariate

sequence

Anomaly Detection methods: *an Example*



Time series T



NormA [10]

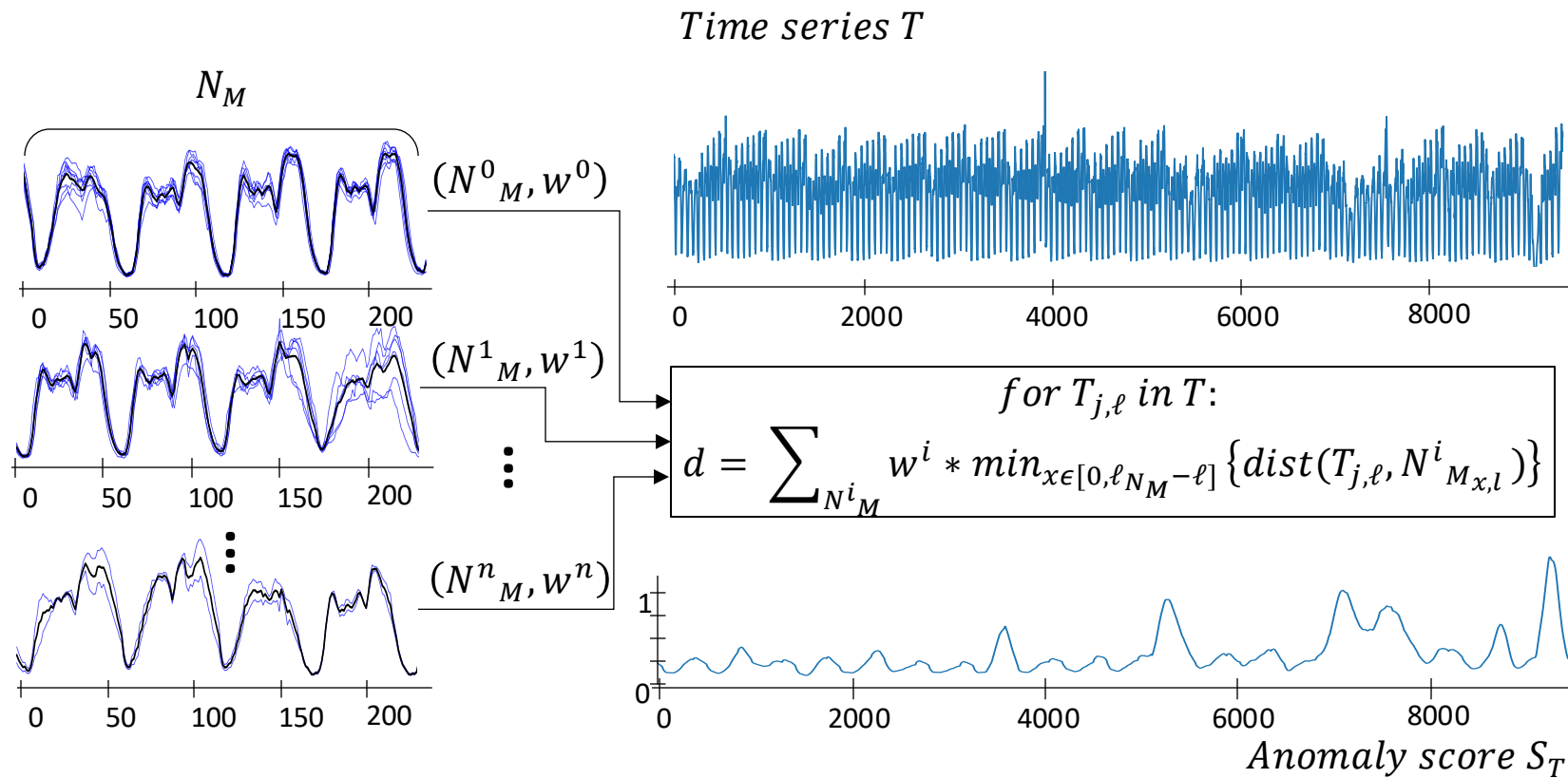
Distance-based approach that **summarize** the time series into **a weighted set of subsequences** and use the distance to them as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



NormA [10]

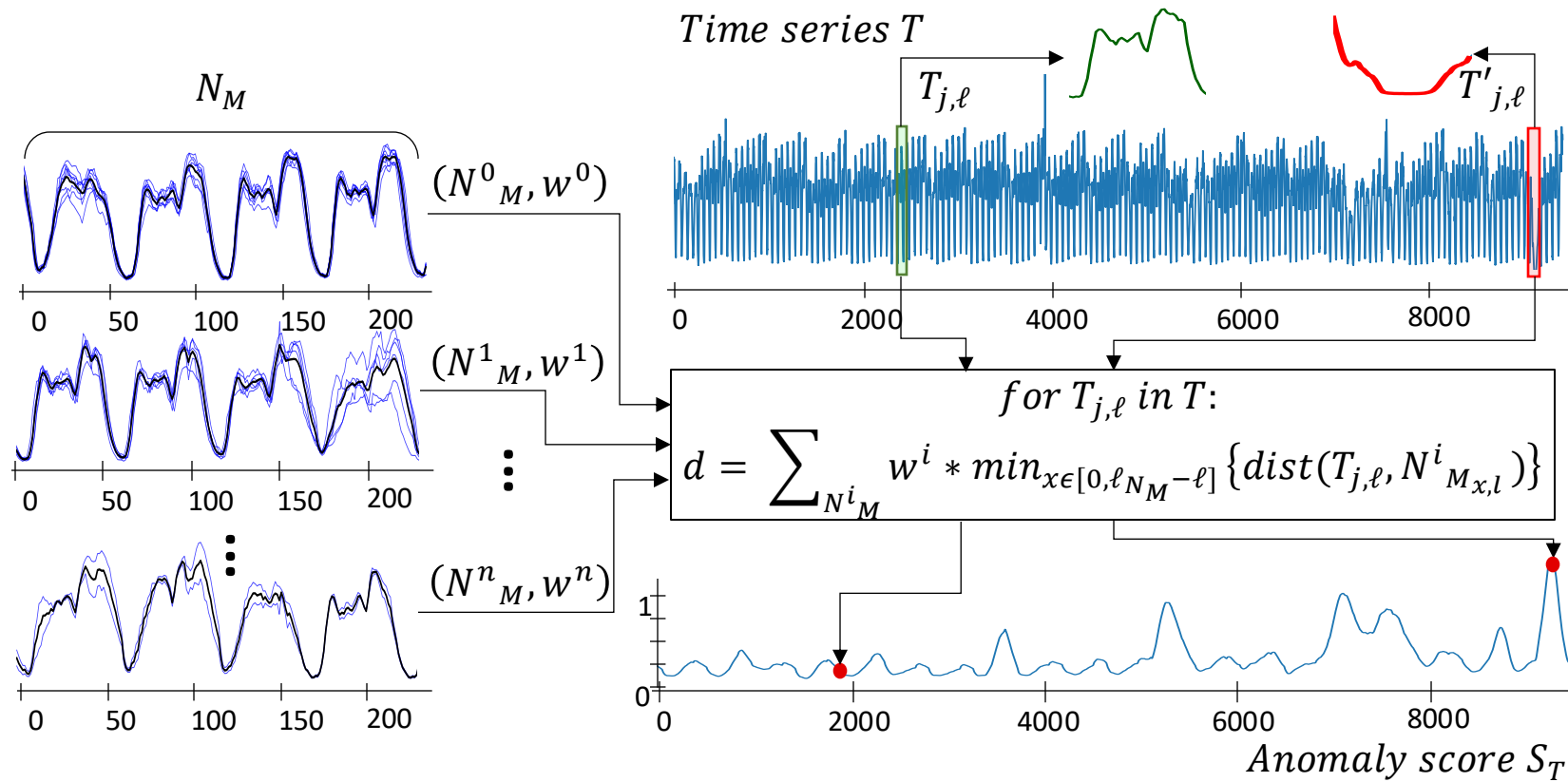
Distance-based approach that **summarize** the time series into **a weighted set of subsequences** and use the distance to them as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



NormA [10]

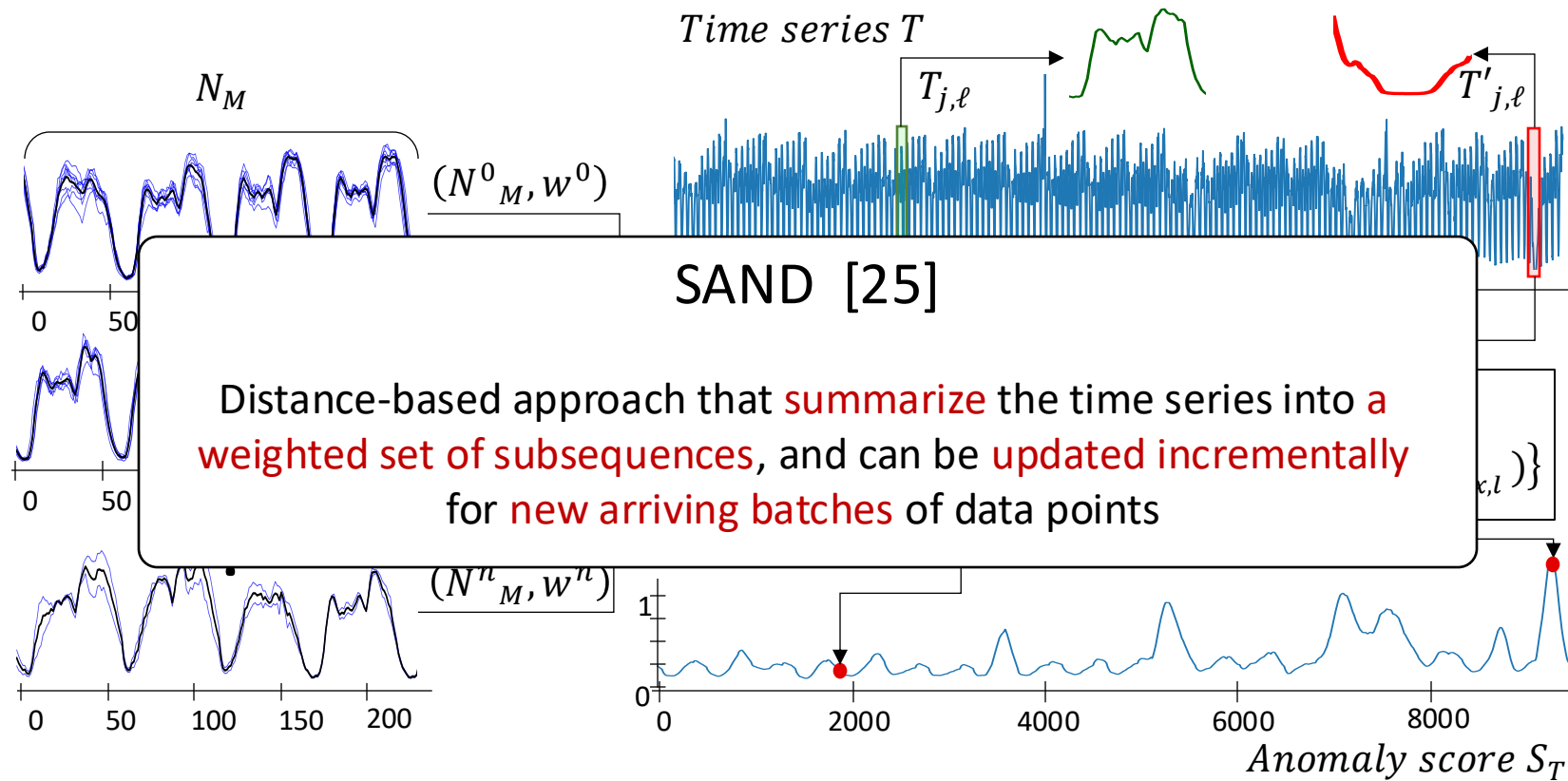
Distance-based approach that **summarize** the time series into **a weighted set of subsequences** and use the distance to them as anomaly score

Unsupervised

Univariate

sequence

Anomaly Detection methods: *an Example*



NormA [10]

Distance-based approach that **summarize** the time series into a **weighted set of subsequences** and use the distance to them as anomaly score

Unsupervised

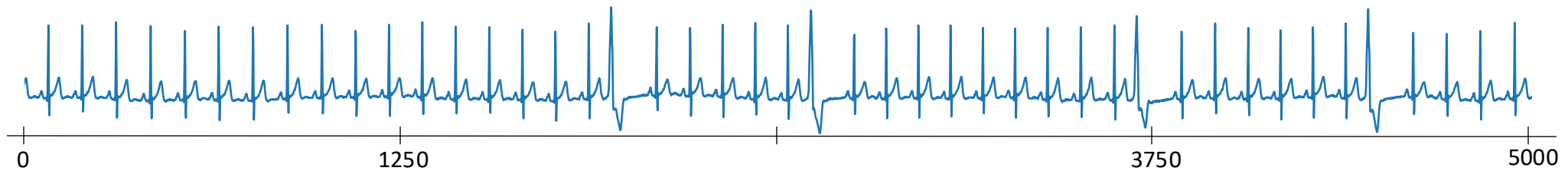
Univariate

sequence

Anomaly Detection methods: *Density-based*

Methods that **estimate the density** of the space (points or subsequences) and identify as anomalies points (or sequences) that are in **low-density subspace**.

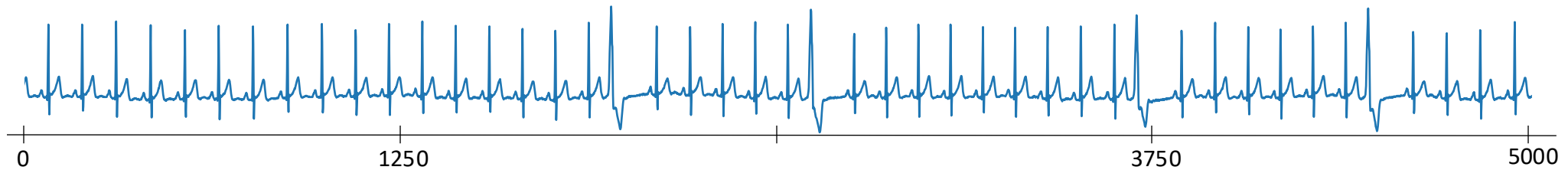
Time series T



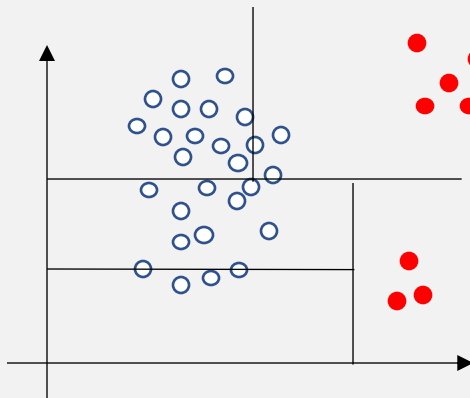
Anomaly Detection methods: *Density-based*

Methods that **estimate the density** of the space (points or subsequences) and identify as anomalies points (or sequences) that are in **low-density subspace**.

Time series T



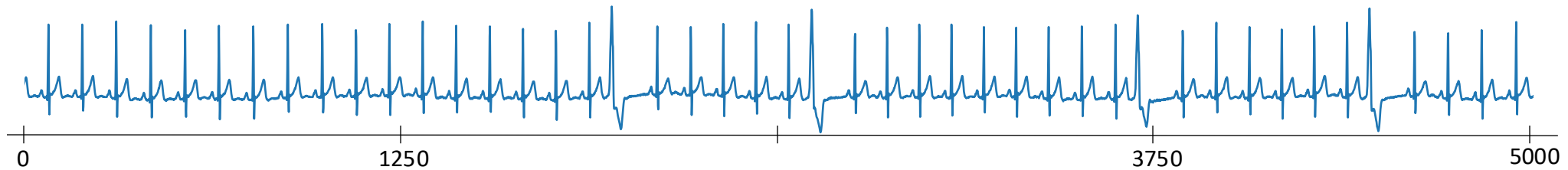
Tree-based approaches [11]



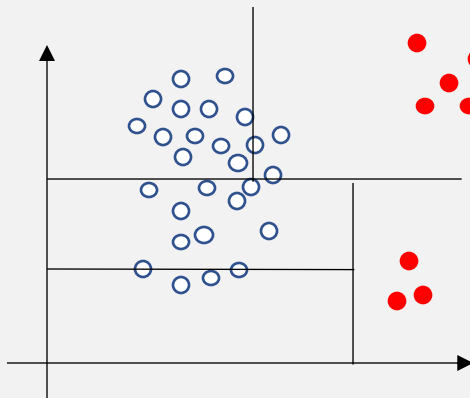
Anomaly Detection methods: *Density-based*

Methods that **estimate the density** of the space (points or subsequences) and identify as anomalies points (or sequences) that are in **low-density subspace**.

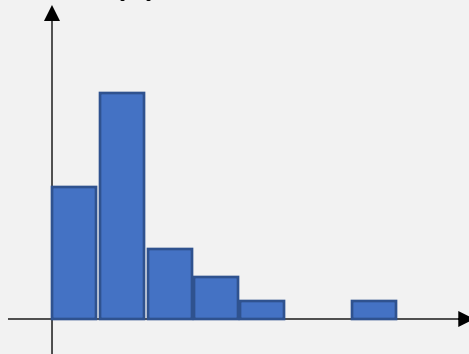
Time series T



Tree-based approaches [11]



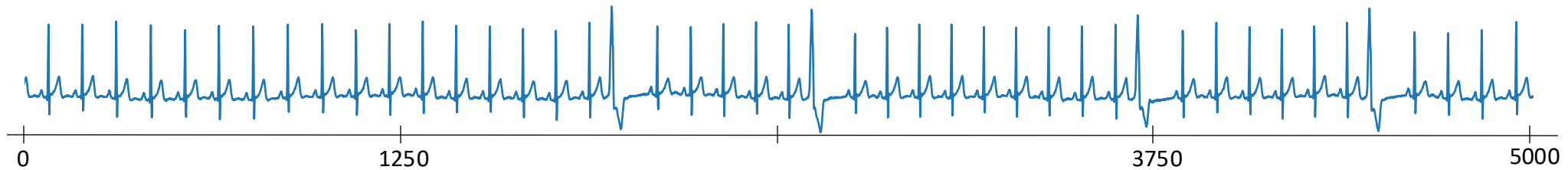
Distribution-based Approaches [12]



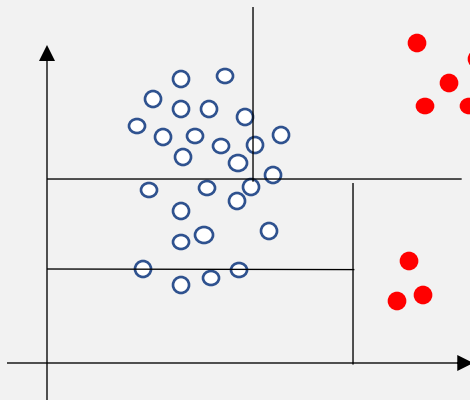
Anomaly Detection methods: *Density-based*

Methods that **estimate the density** of the space (points or subsequences) and identify as anomalies points (or sequences) that are in **low-density subspace**.

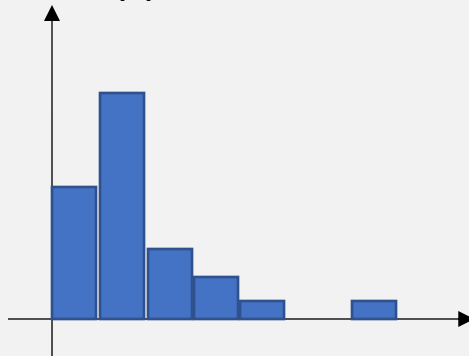
Time series T



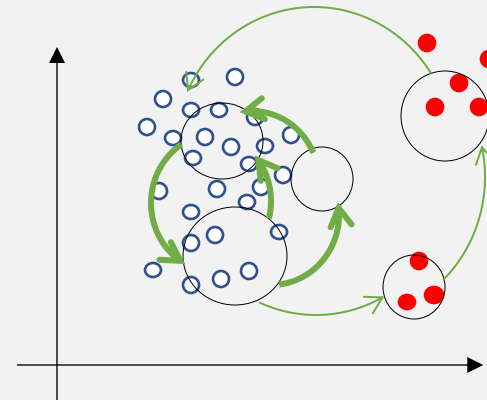
Tree-based approaches [11]



Distribution-based
Approaches [12]

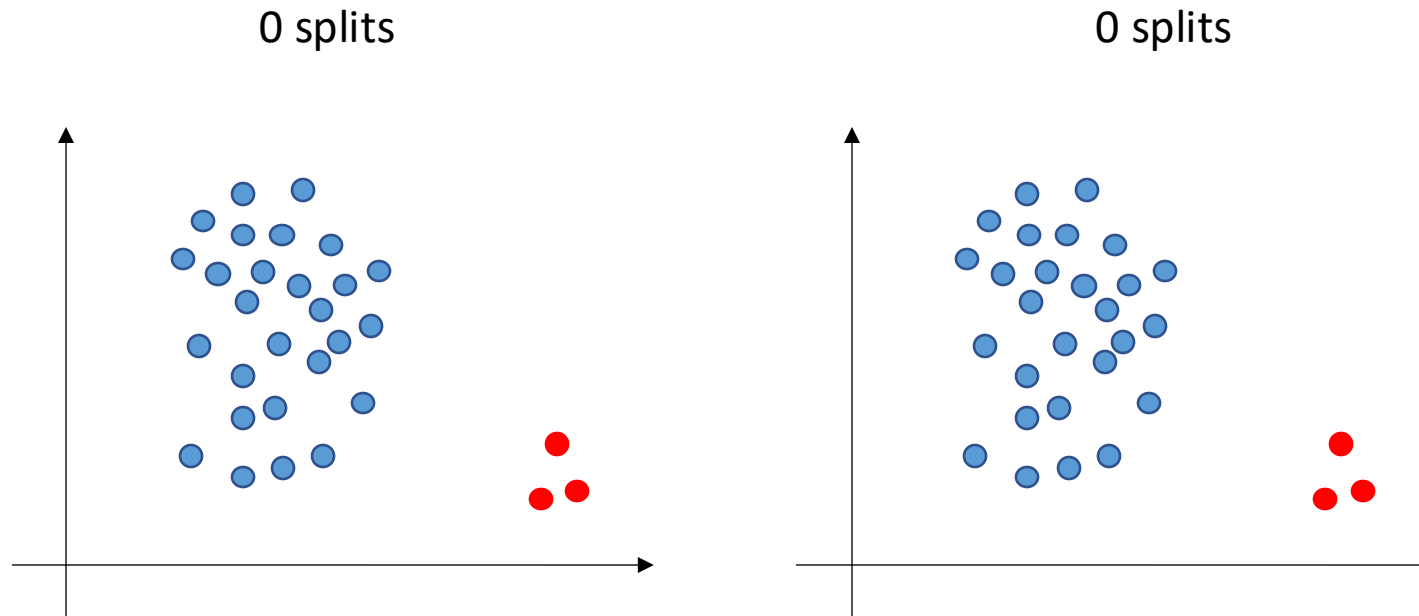


Graph-based approaches [13]



...

Anomaly Detection methods: *an Example*



Isolation Forest [11]

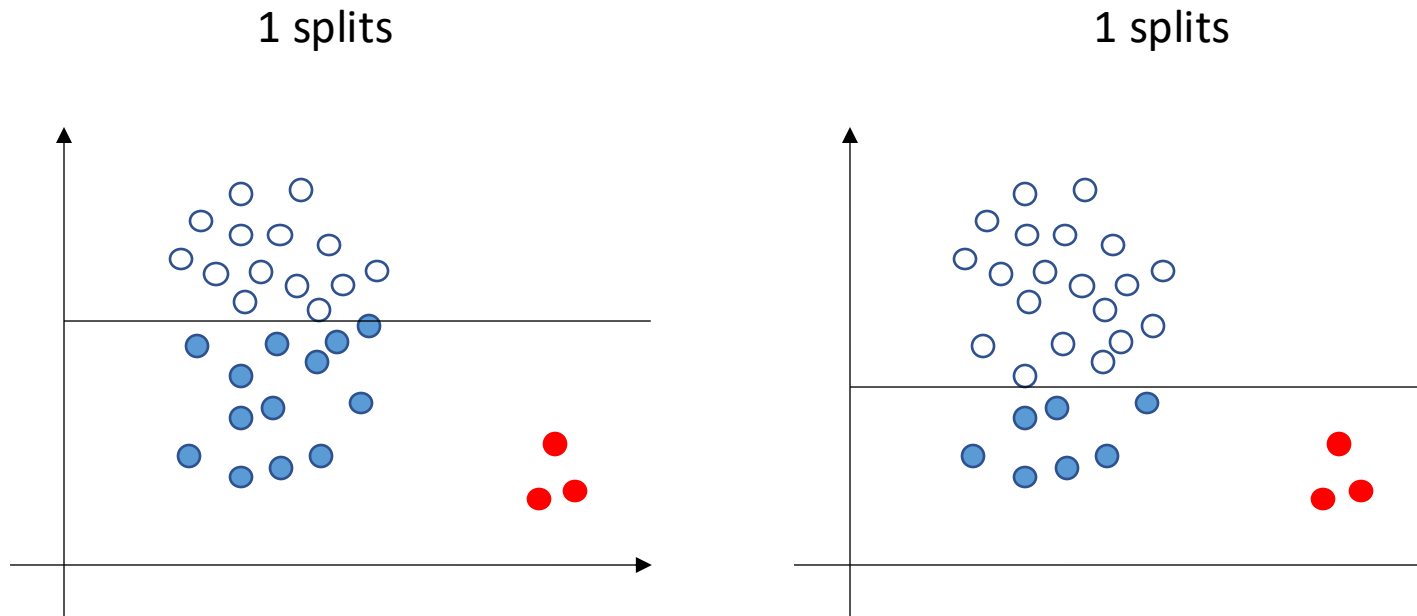
Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

Unsupervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



Isolation Forest [11]

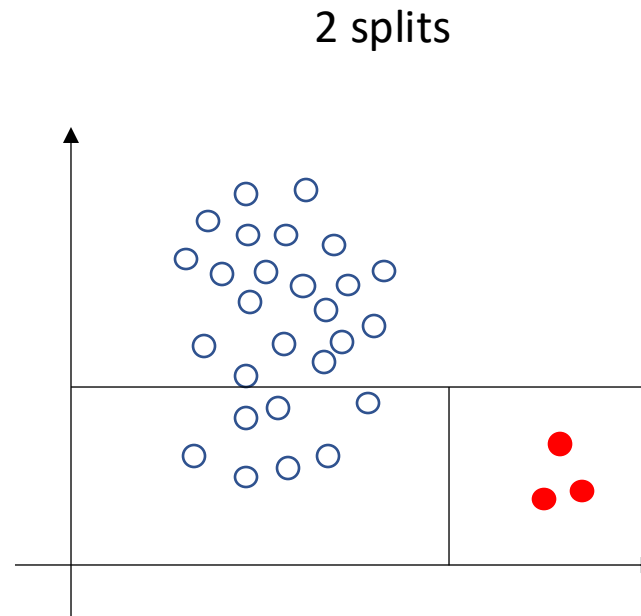
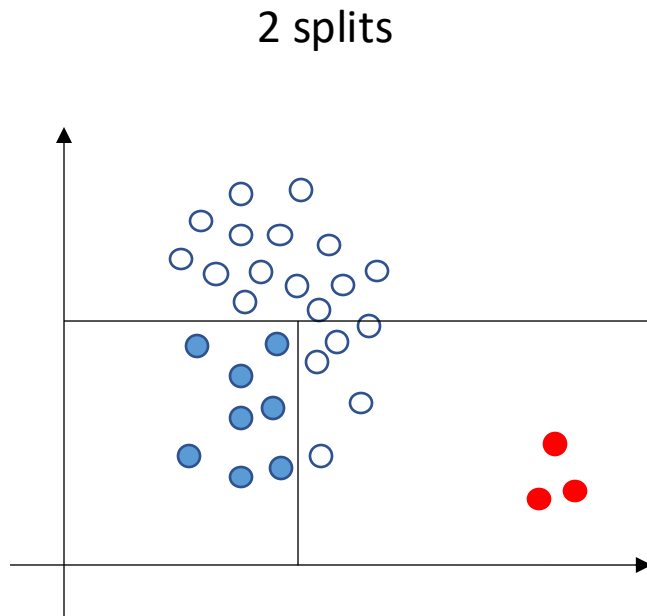
Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

Unsupervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



Isolation Forest [11]

Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

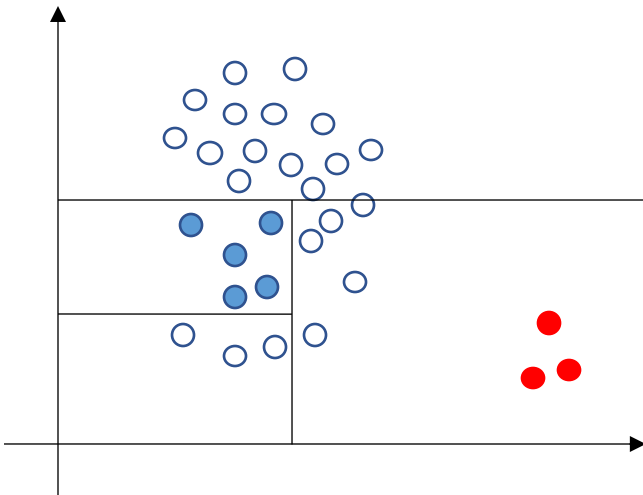
Unsupervised

Univariate/Multivariate

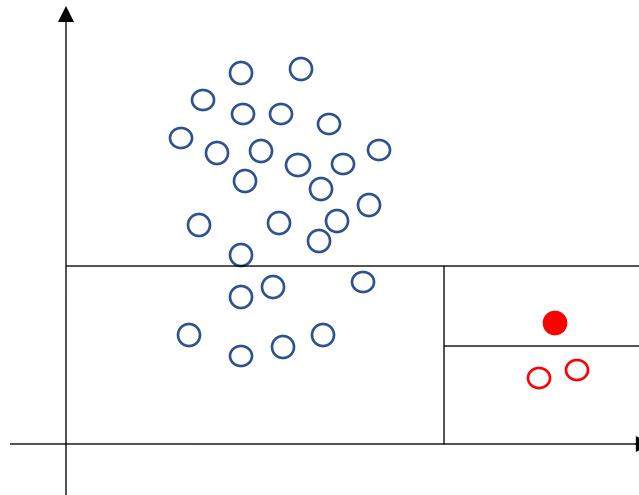
Point/sequence

Anomaly Detection methods: *an Example*

3 splits



3 splits



Isolation Forest [11]

Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

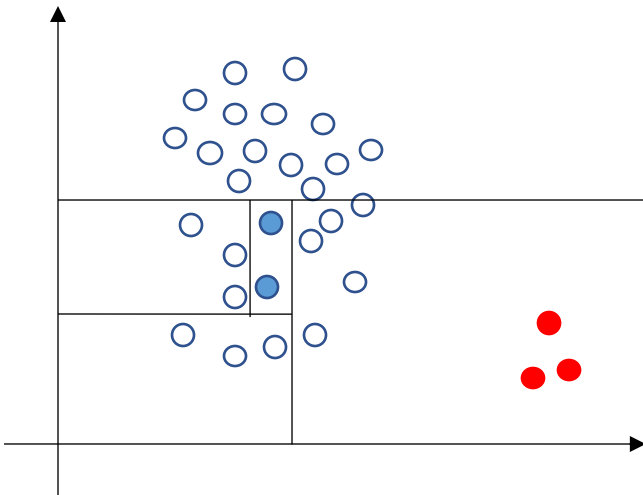
Unsupervised

Univariate/Multivariate

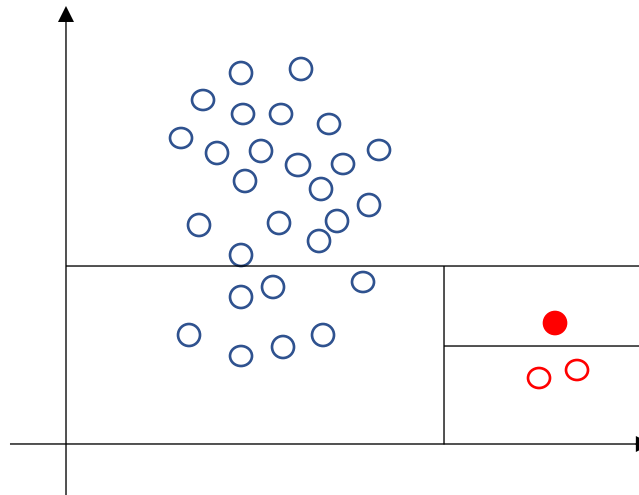
Point/sequence

Anomaly Detection methods: *an Example*

4 splits



3 splits



Isolation Forest [11]

Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

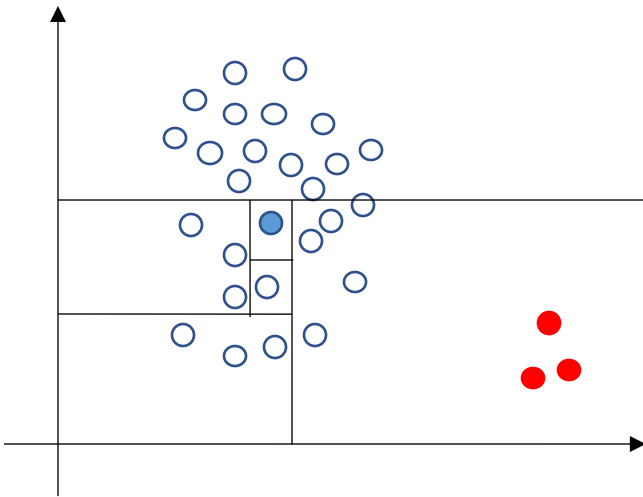
Unsupervised

Univariate/Multivariate

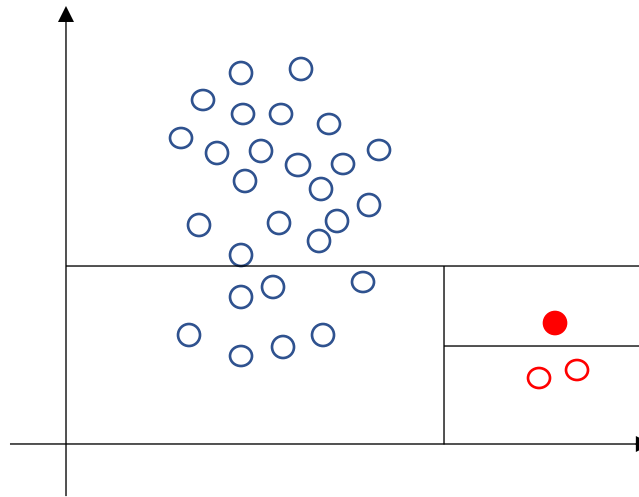
Point/sequence

Anomaly Detection methods: *an Example*

5 splits



3 splits



Isolation Forest [11]

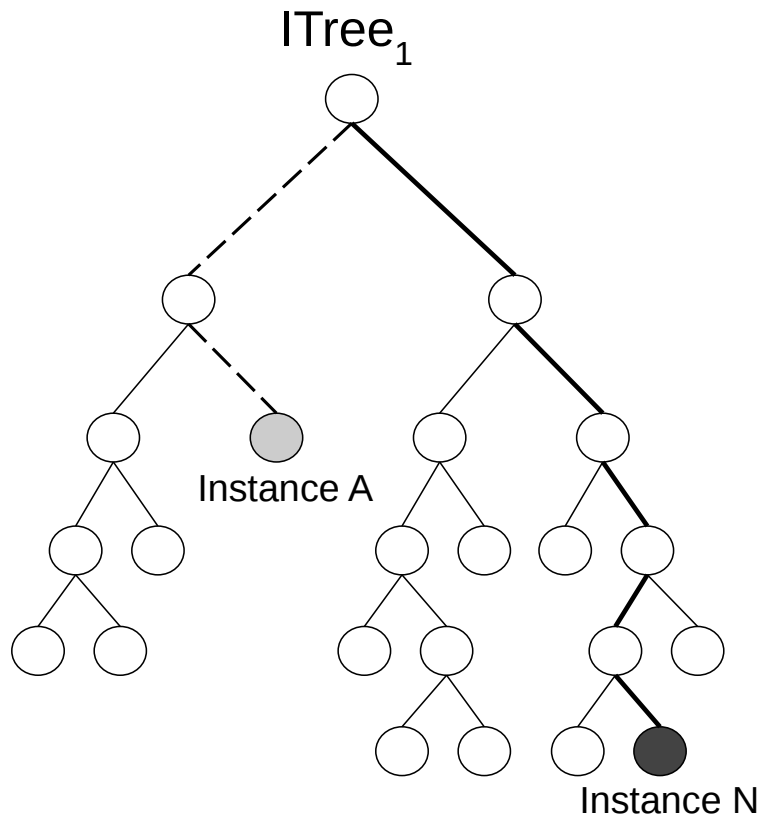
Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

Unsupervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



Isolation Forest [11]

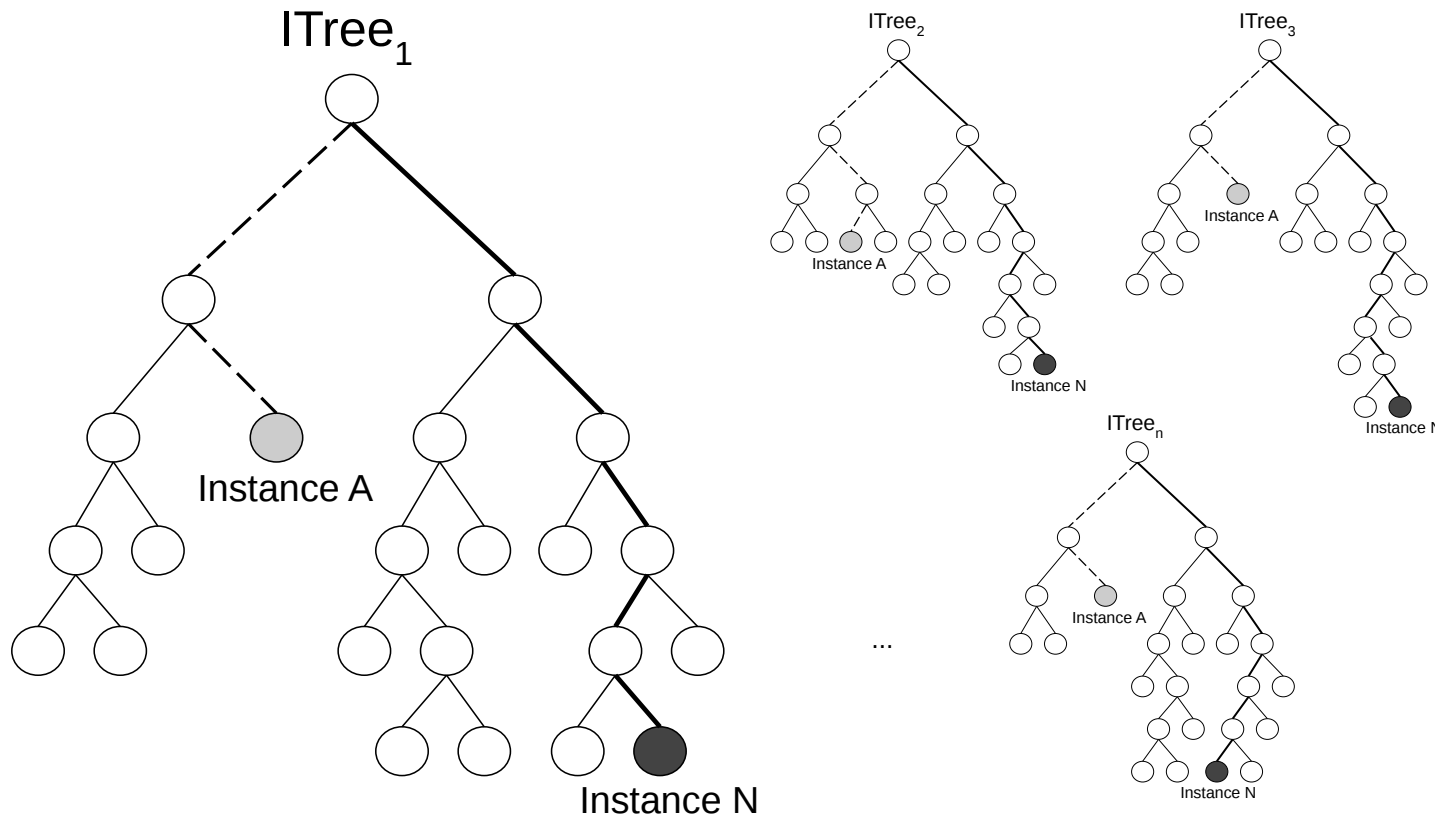
Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

Unsupervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



Isolation Forest [11]

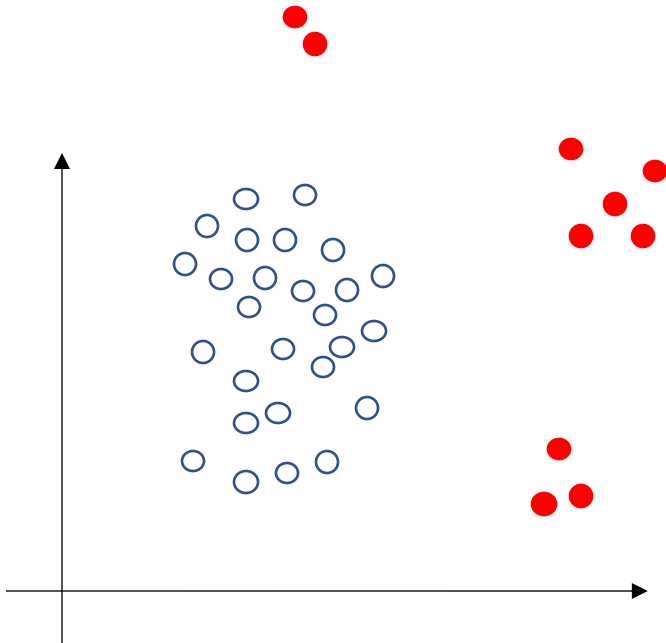
Density-based approach that **split the space randomly** and using the **depth of the trees** to identify anomalies

Unsupervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



Series2Graph [13]

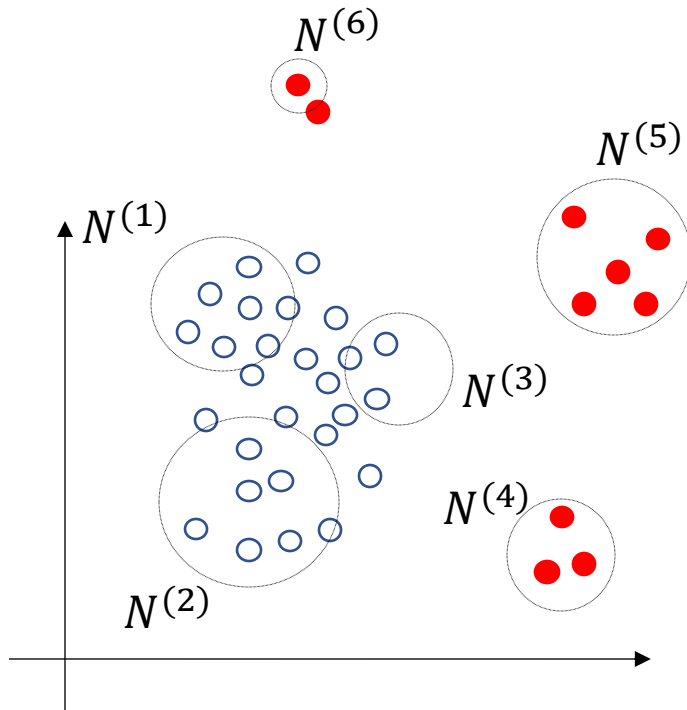
Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*



Each **node** is an ensemble of similar subsequences.

Series2Graph [13]

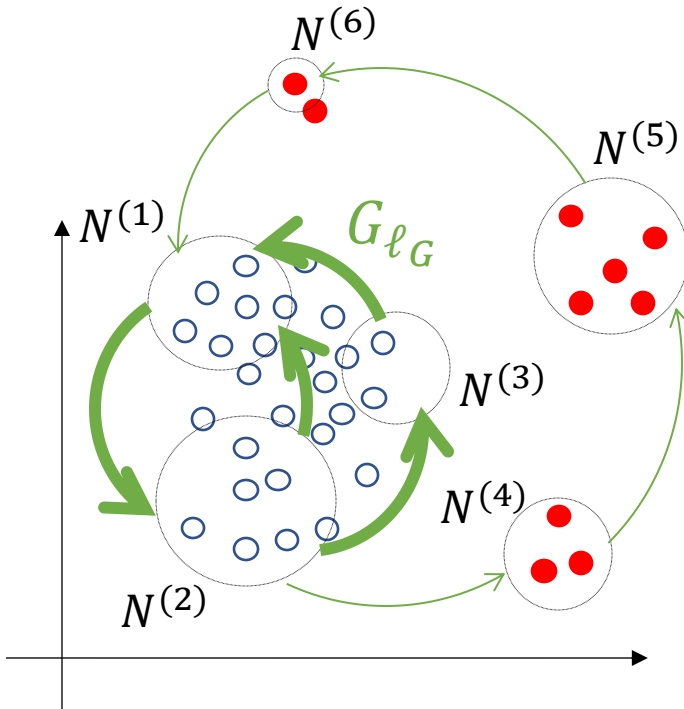
Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*



Each **node** is an ensemble of similar subsequences.

Each **edge** is associated to a weight w that corresponds to the number of times a subsequence move from one node to another.

Series2Graph [13]

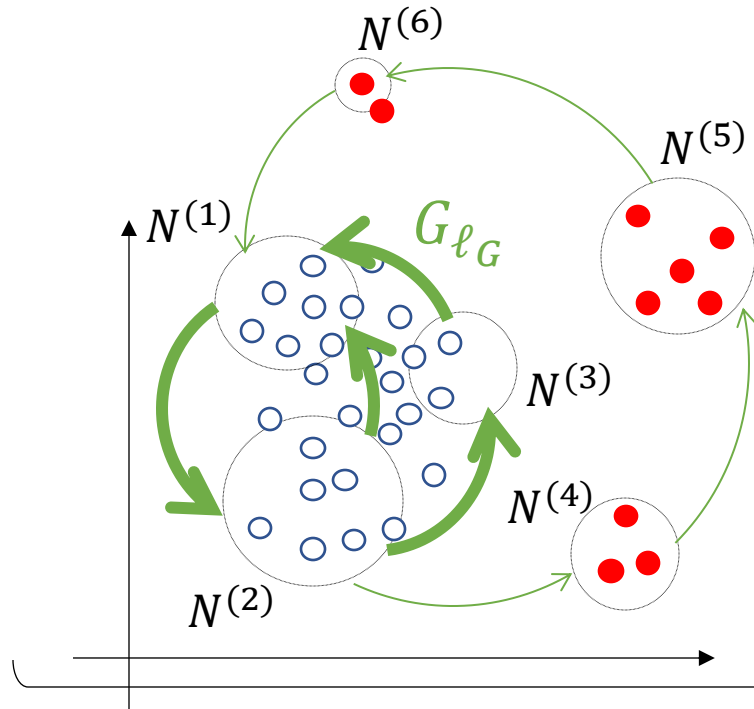
Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*



Each **node** is an ensemble of similar subsequences.

Each **edge** is associated to a weight w that corresponds to the number of times a subsequence move from one node to another.

For a given subsequence $T_{i,\ell}$ and its corresponding path $P_{th} = \langle N^{(i)}, N^{(i+1)}, \dots, N^{(i+\ell)} \rangle$, we define the normality score as follows:

$$Norm(P_{th}) = \sum_{j=i}^{i+\ell-1} \frac{w(N^{(j)}, N^{(j+1)}) \deg(N^{(j)} - 1)}{\ell}$$

Series2Graph [13]

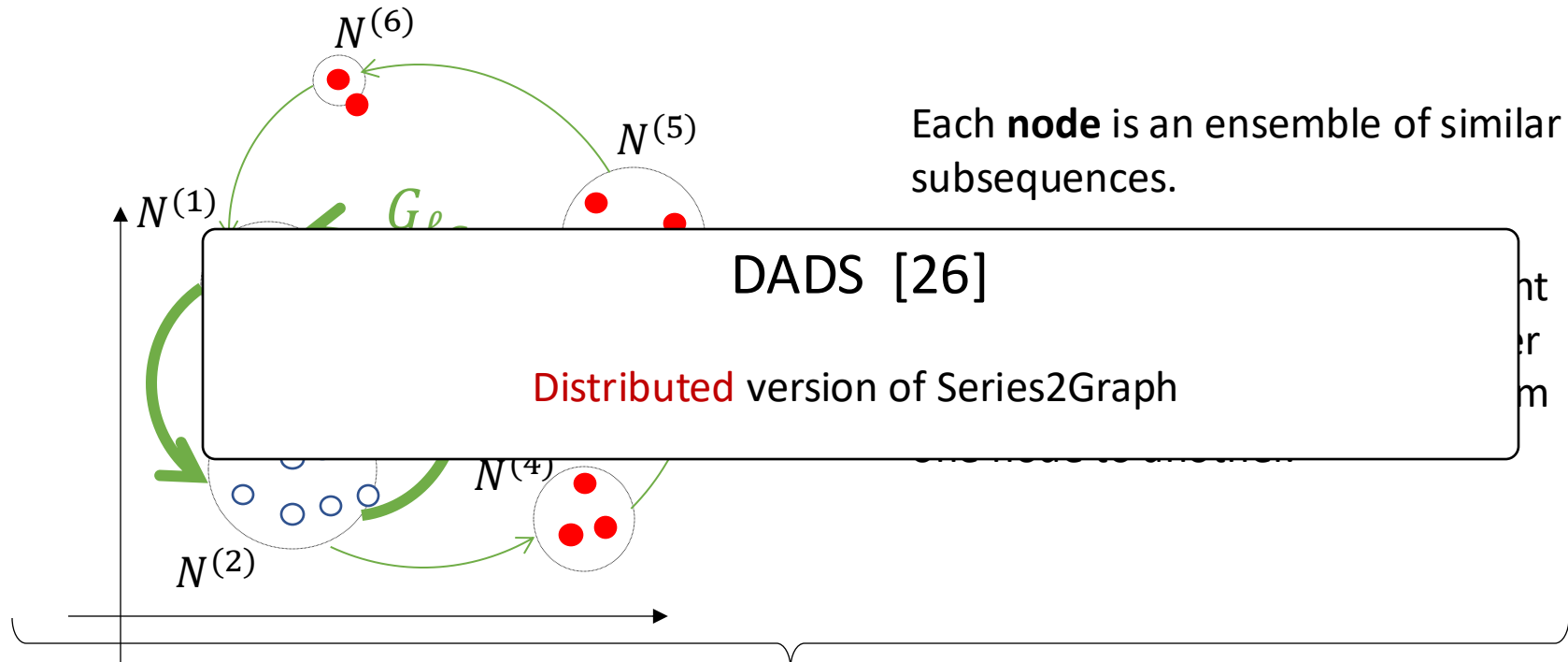
Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*



For a given subsequence $T_{i,\ell}$ and its corresponding path $P_{th} = \langle N^{(i)}, N^{(i+1)}, \dots, N^{(i+\ell)} \rangle$, we define the normality score as follows:

$$Norm(P_{th}) = \sum_{j=i}^{i+\ell-1} \frac{w(N^{(j)}, N^{(j+1)}) \deg(N^{(j)} - 1)}{\ell}$$

Series2Graph [13]

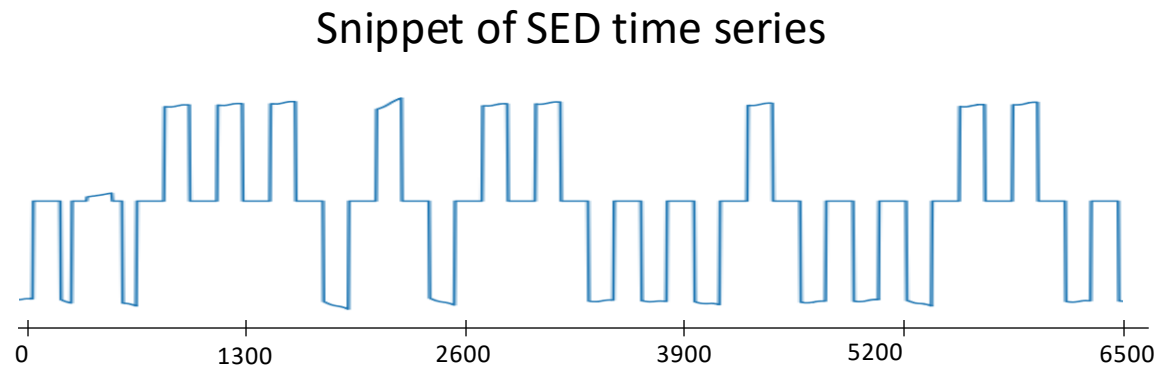
Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*



Series2Graph [13]

Density-based approach that
convert the time series into a
graph and detect **unusual**
trajectories

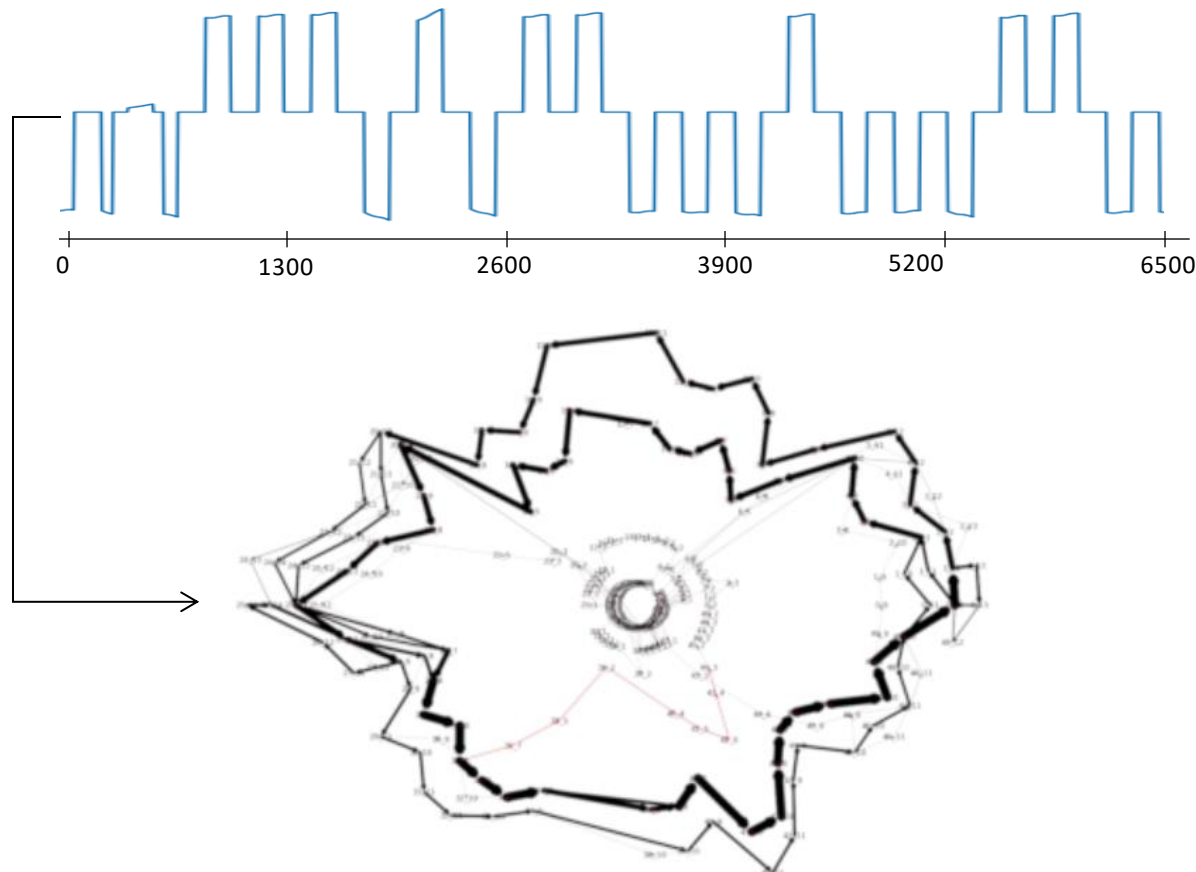
Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*

Snippet of SED time series



Series2Graph [13]

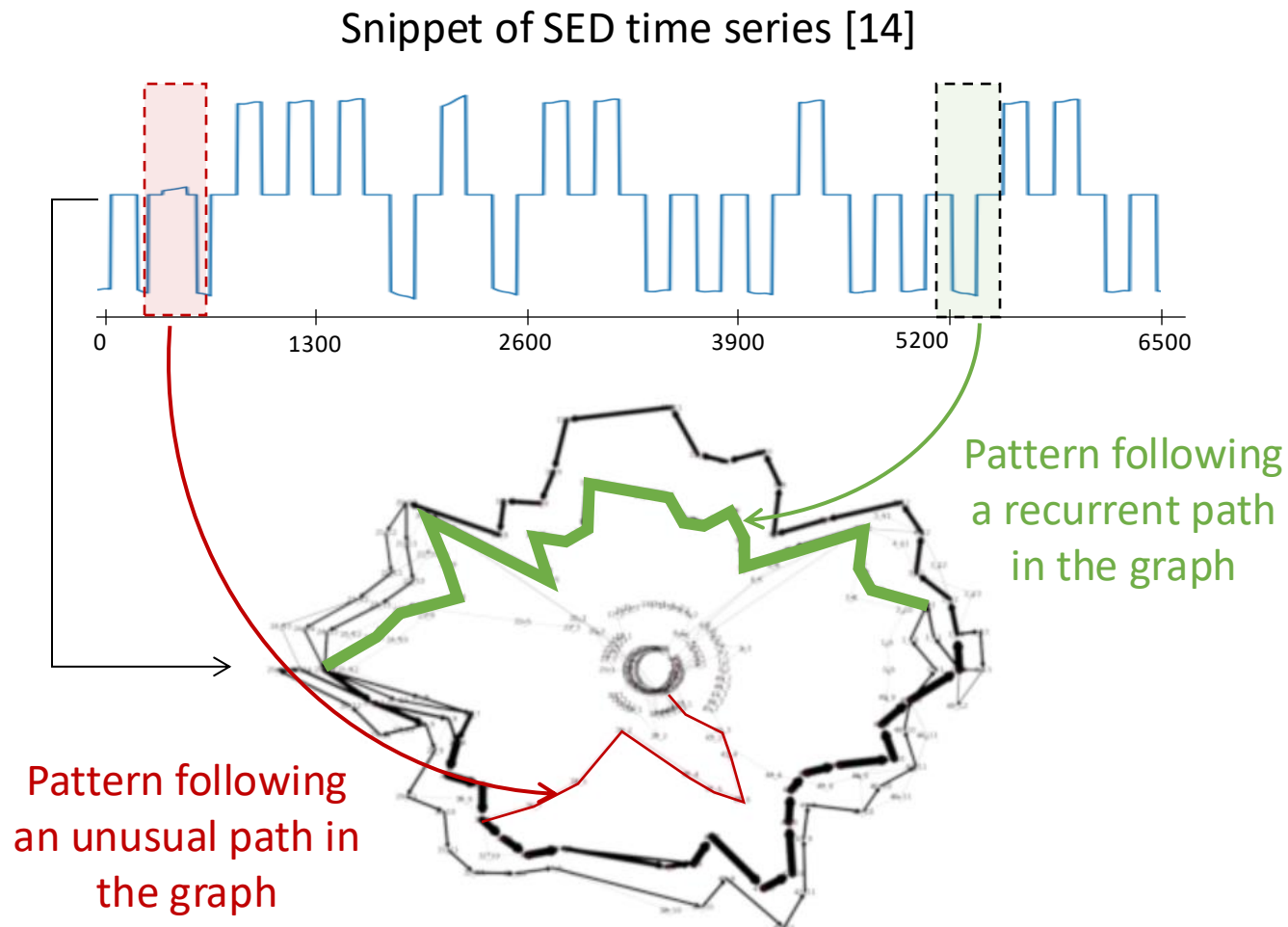
Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

Unsupervised

Univariate

subsequence

Anomaly Detection methods: *an Example*



Series2Graph [13]

Density-based approach that **convert** the time series into a **graph** and detect **unusual trajectories**

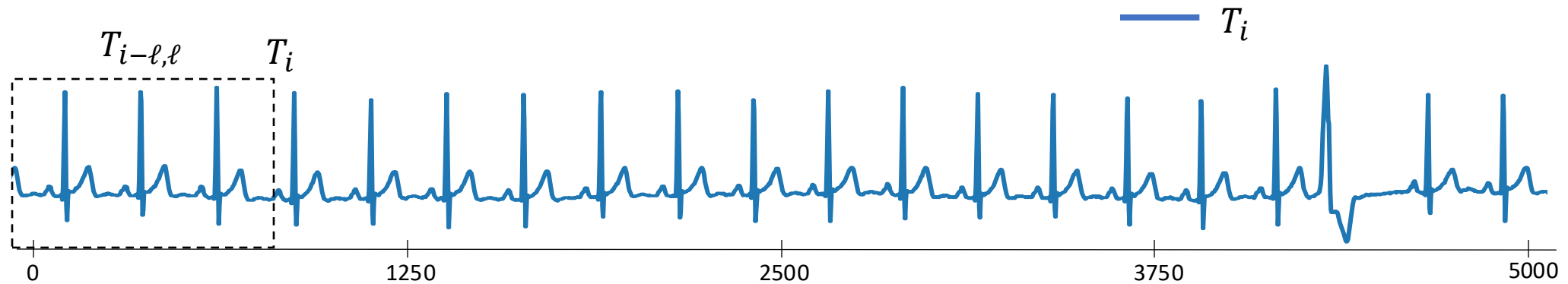
Unsupervised

Univariate

subsequence

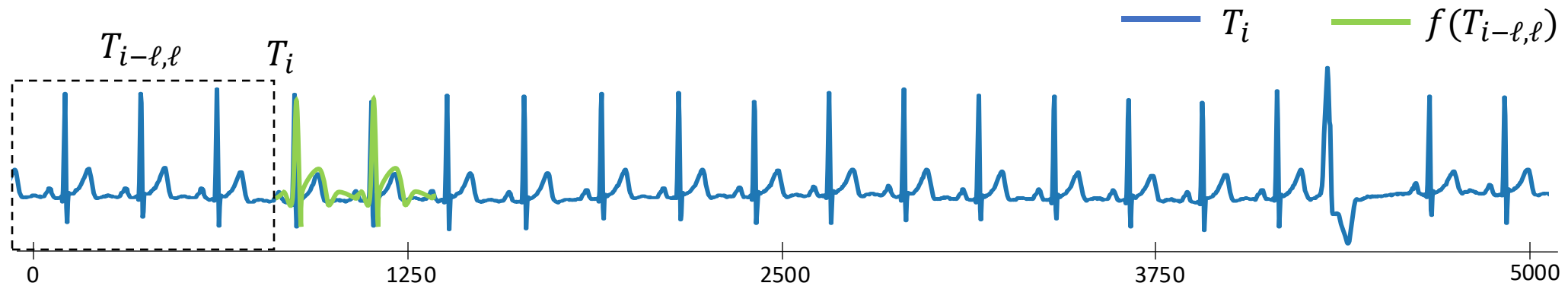
Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.



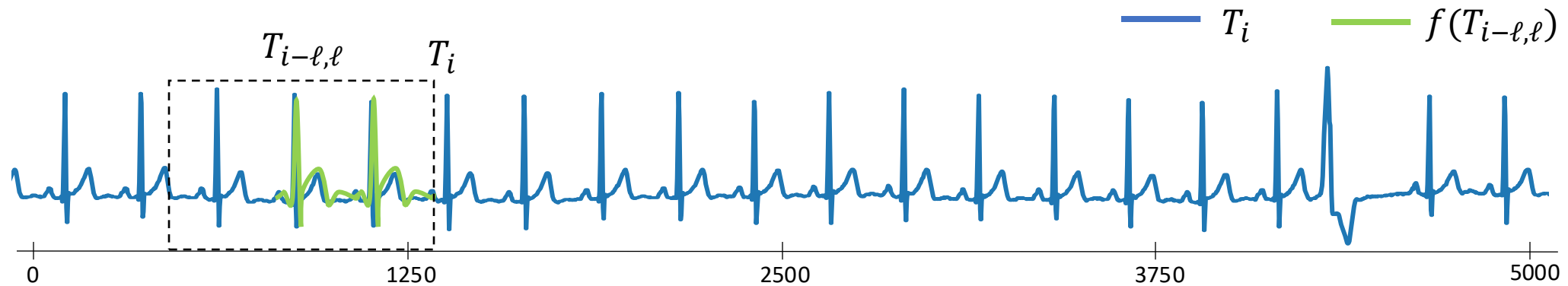
Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.



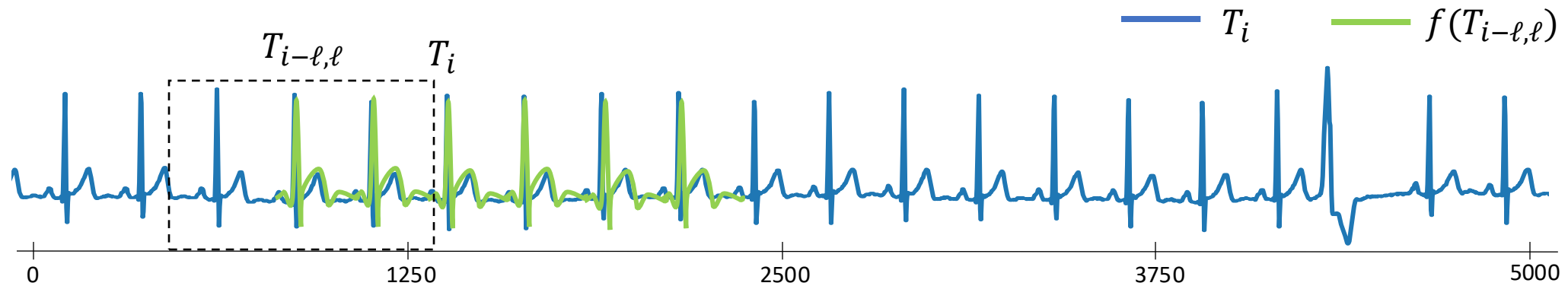
Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.



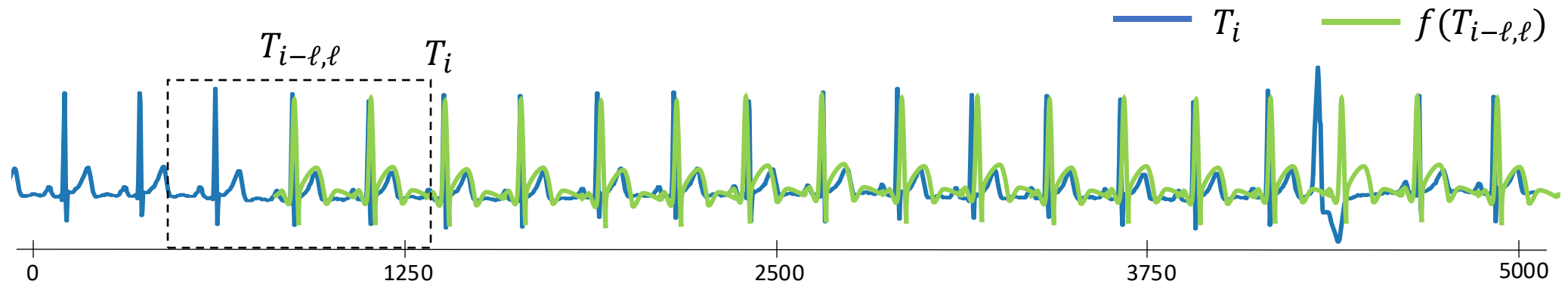
Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.



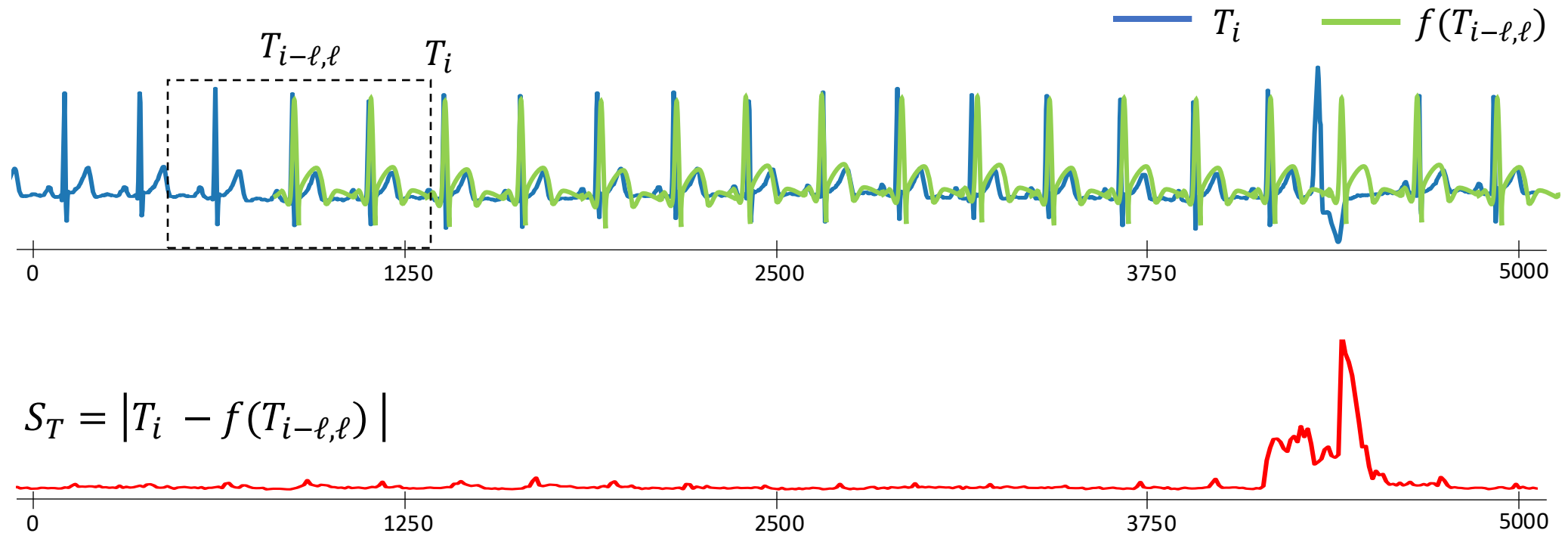
Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.



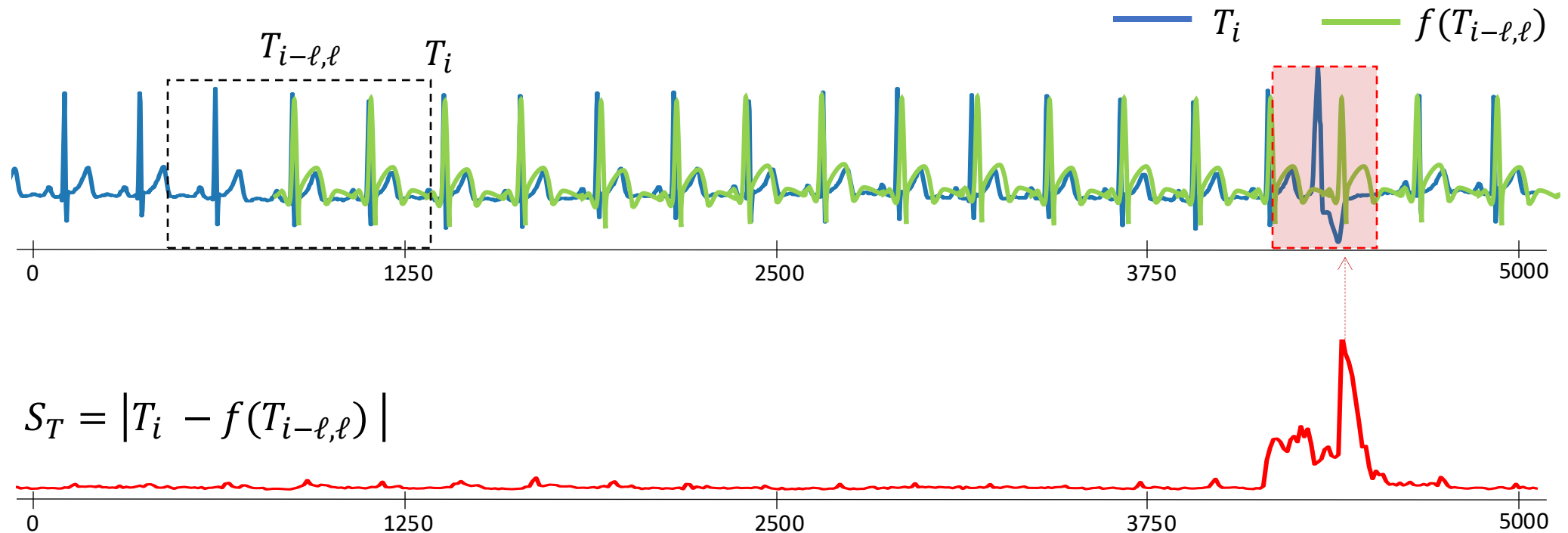
Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.

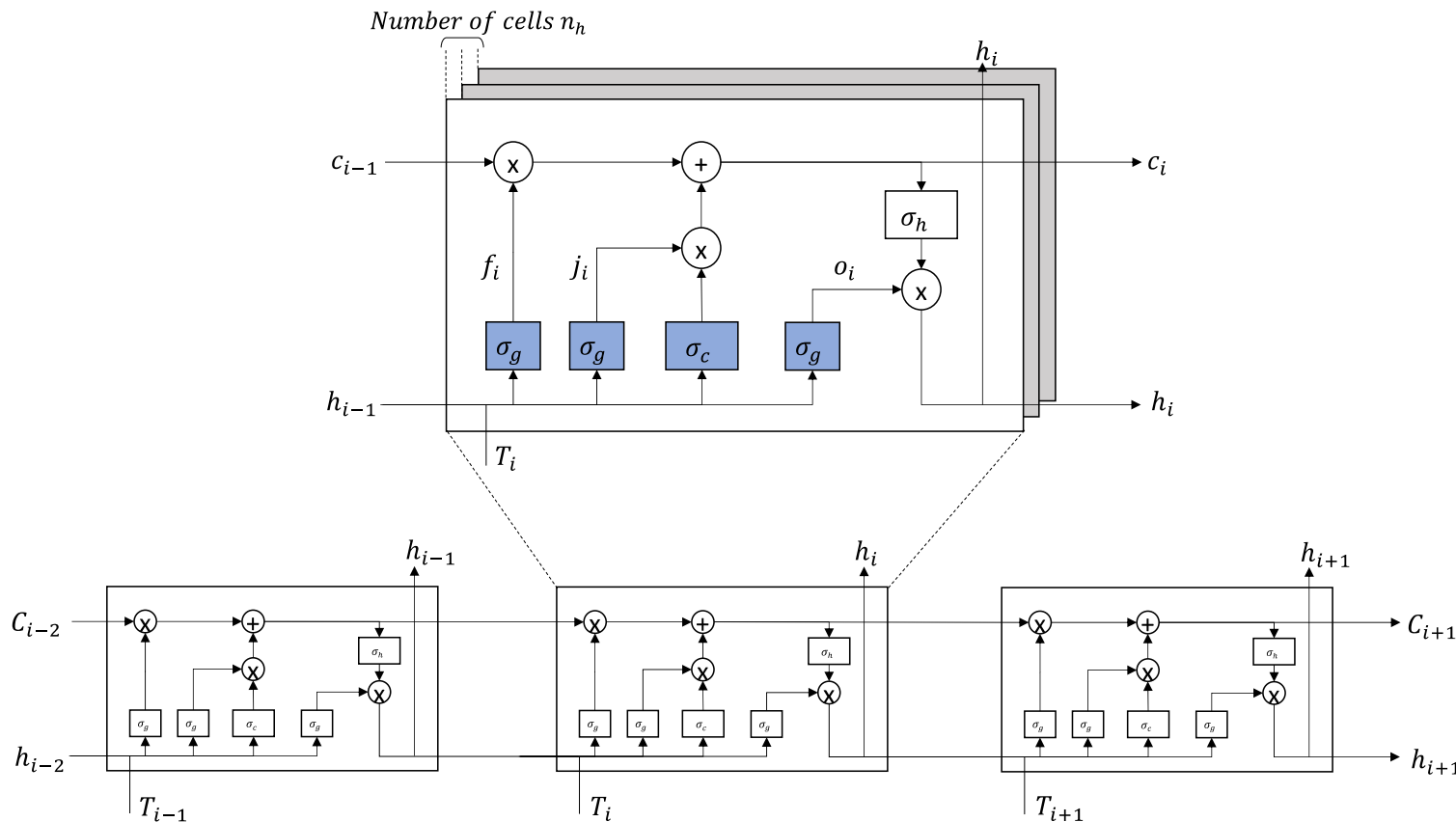


Anomaly Detection methods: *Forecasting-based*

Methods that aims to **predict the next points** based on the previous ones. The **prediction error** is used to detect if there is an anomaly or not.



Anomaly Detection methods: *an Example*



LSTM-AD [15]

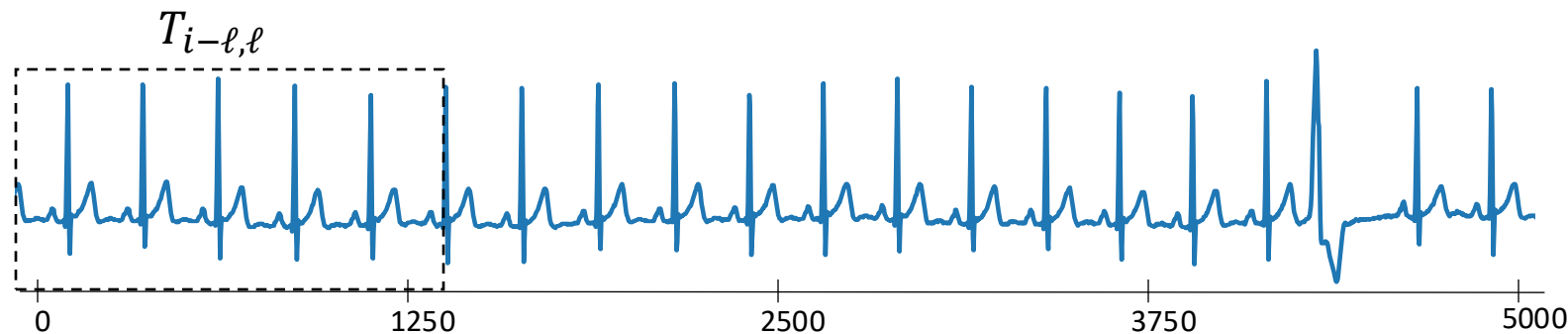
Model that stack multiple LSTM cell and use the output to predict the next value

Semi-supervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



DeepAnT [16] (CNN)

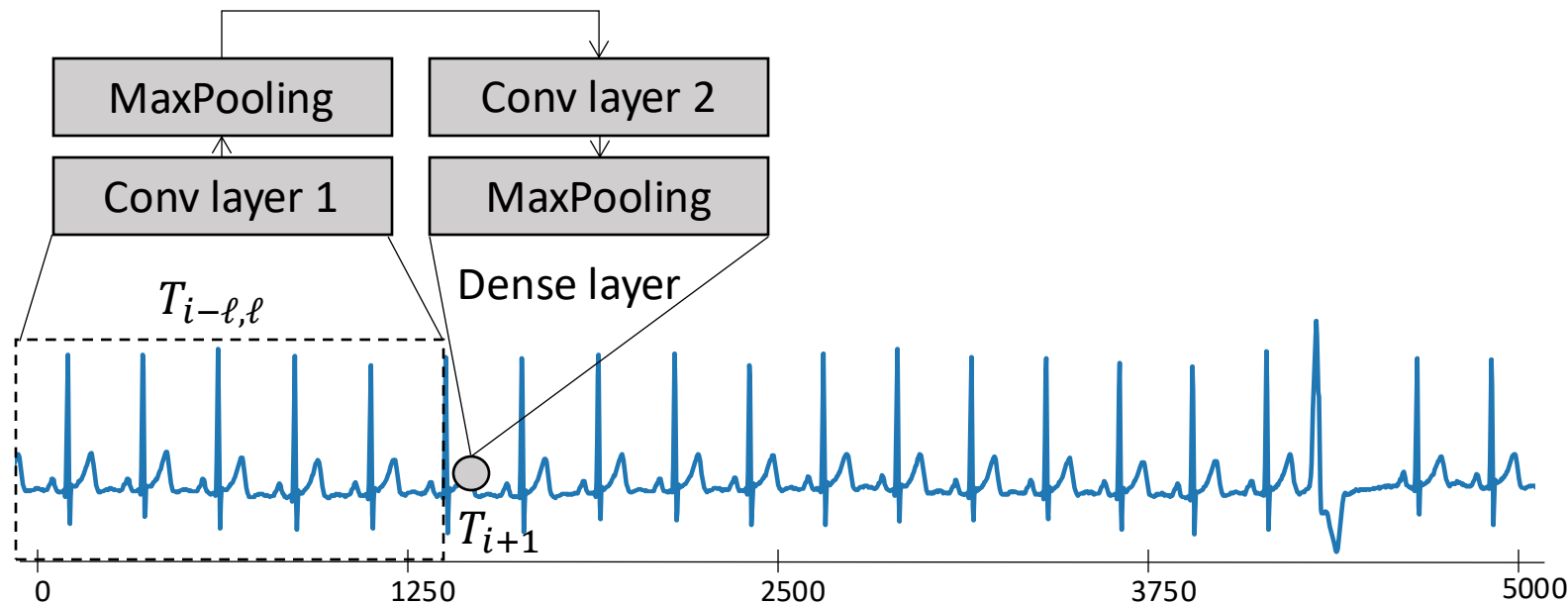
Convolutional-based approach (2 convolutional layers) taking as input a sequence and aims to predict the next value.

Semi-supervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



DeepAnT [16] (CNN)

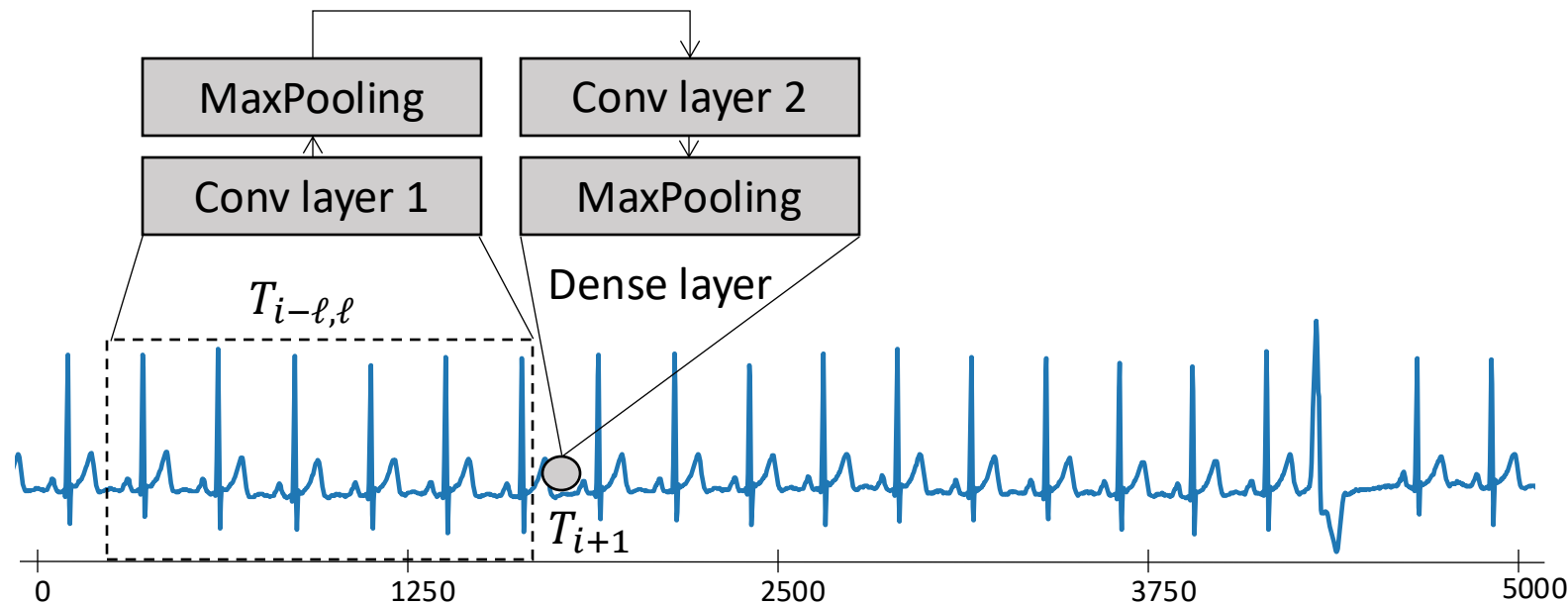
Convolutional-based approach (2 convolutional layers) taking as input a sequence and aims to predict the next value.

Semi-supervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



DeepAnT [16] (CNN)

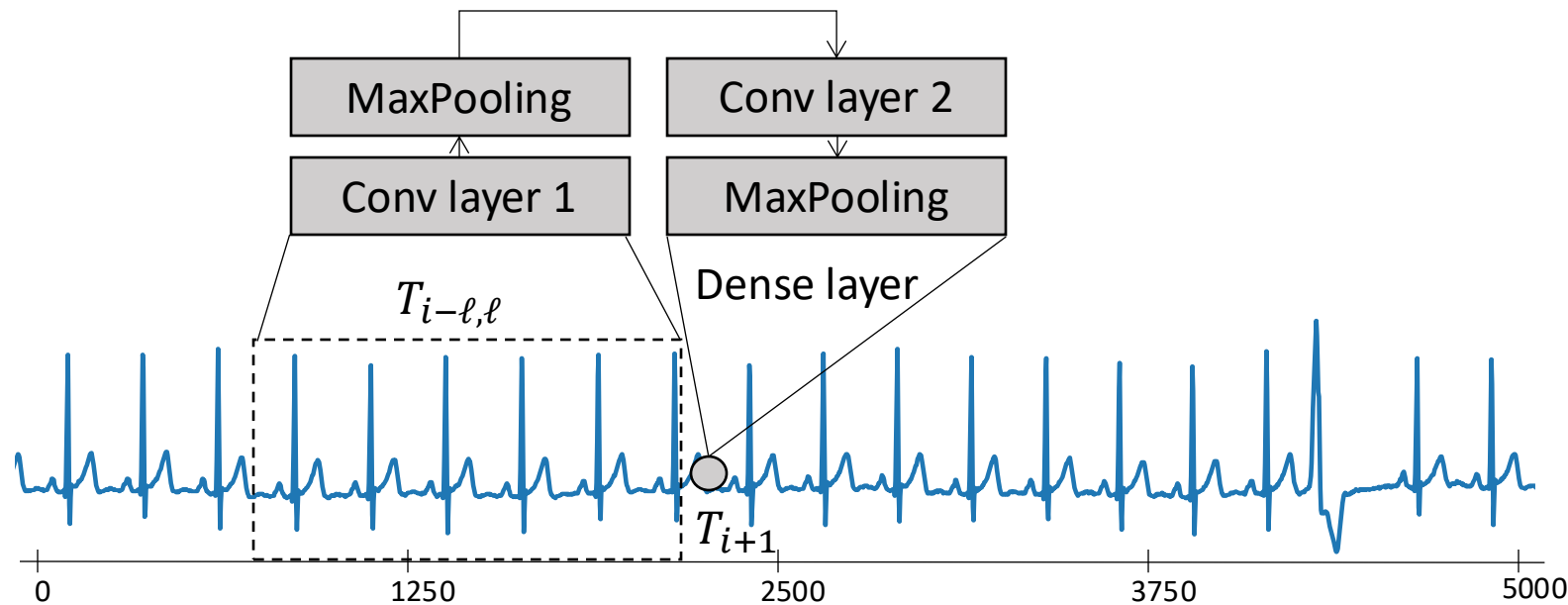
Convolutional-based approach (2 convolutional layers) taking as input a sequence and aims to predict the next value.

Semi-supervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *an Example*



DeepAnT [16] (CNN)

Convolutional-based approach (2 convolutional layers) taking as input a sequence and aims to predict the next value.

Semi-supervised

Univariate/Multivariate

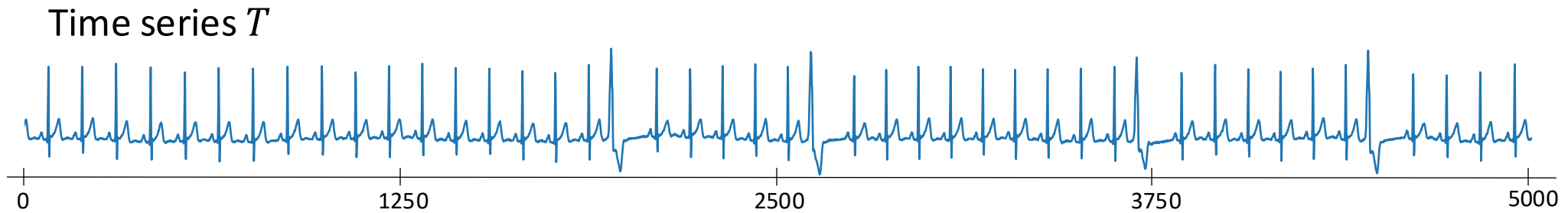
Point/sequence

Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.

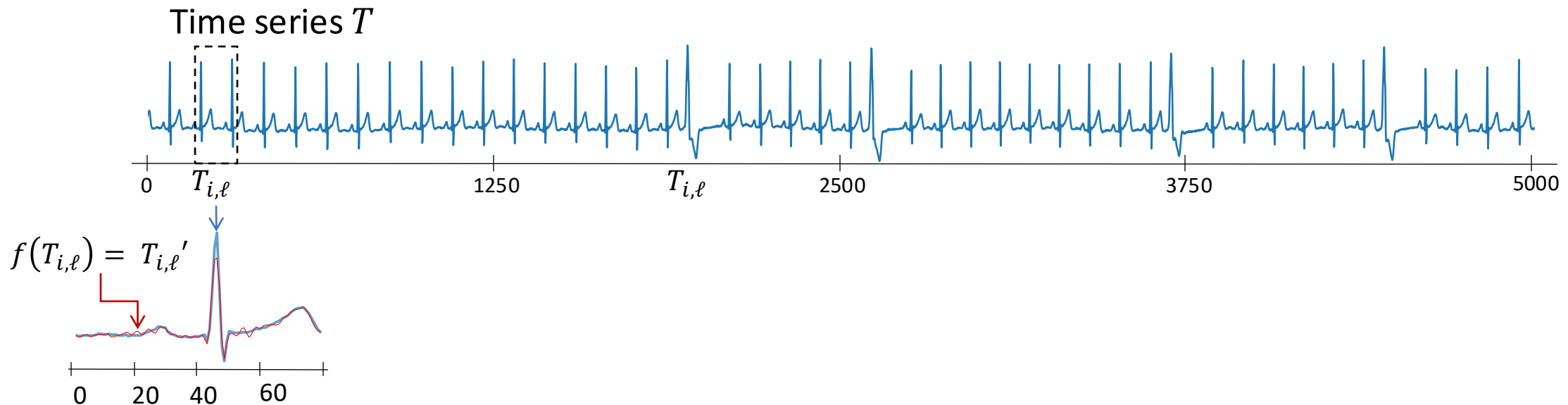
Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.



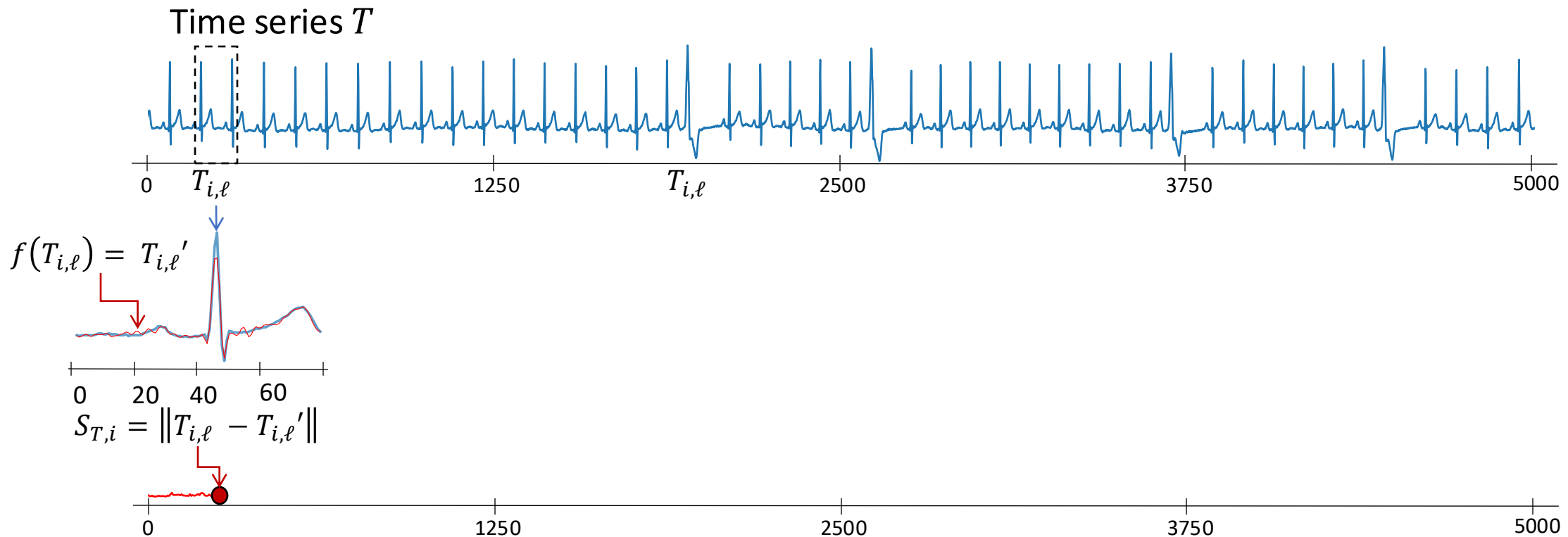
Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.



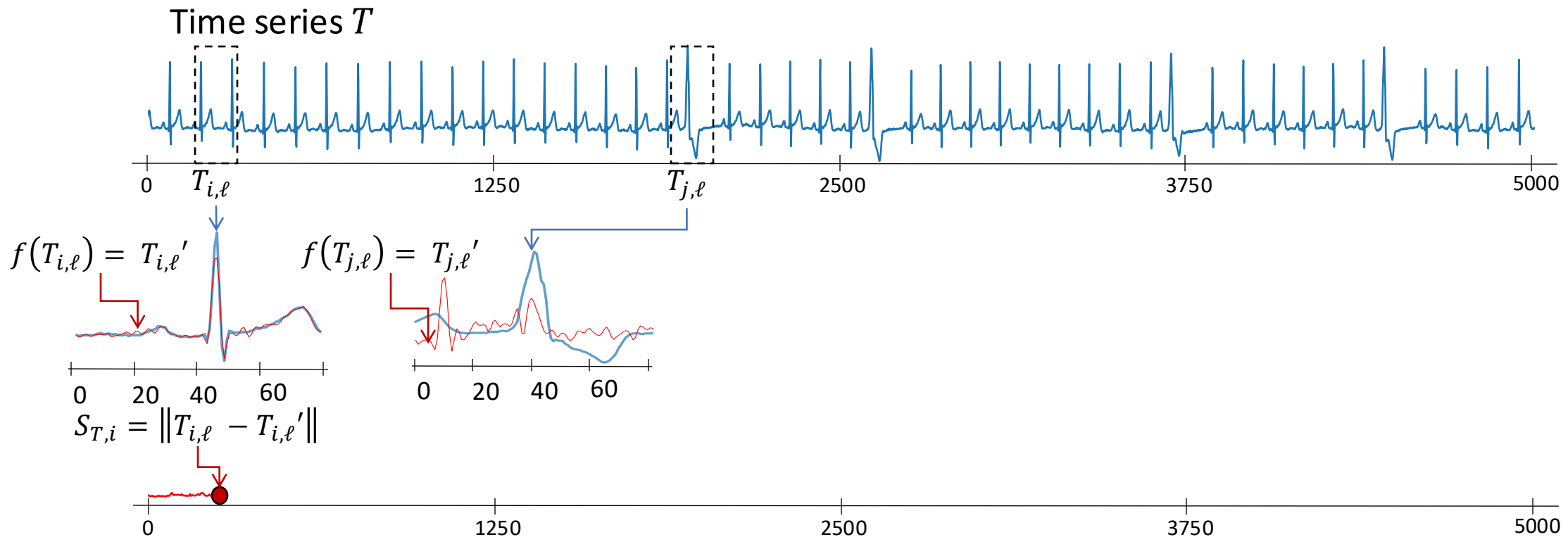
Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.



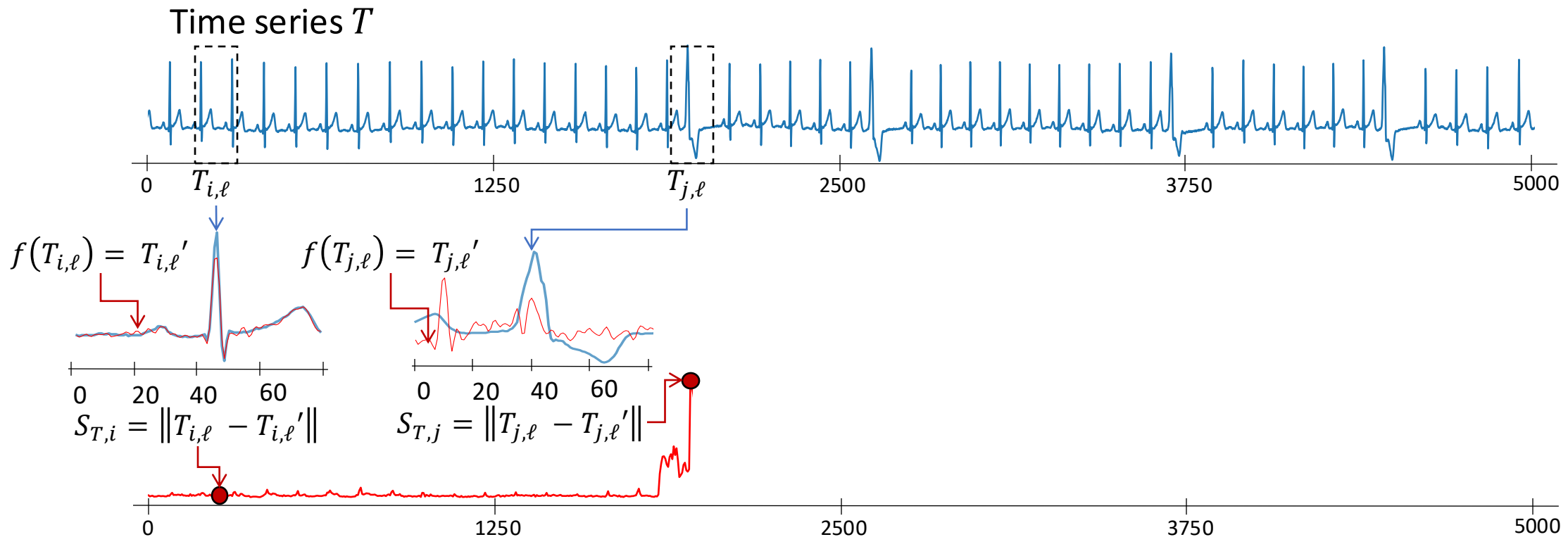
Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.



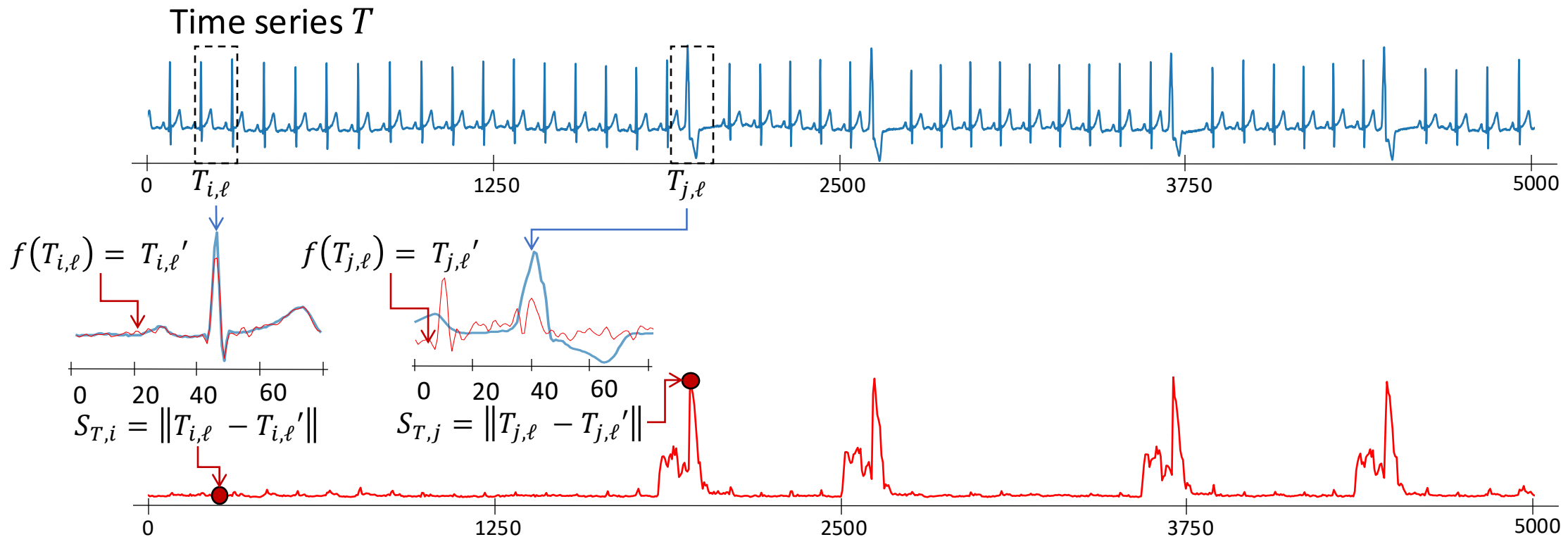
Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.



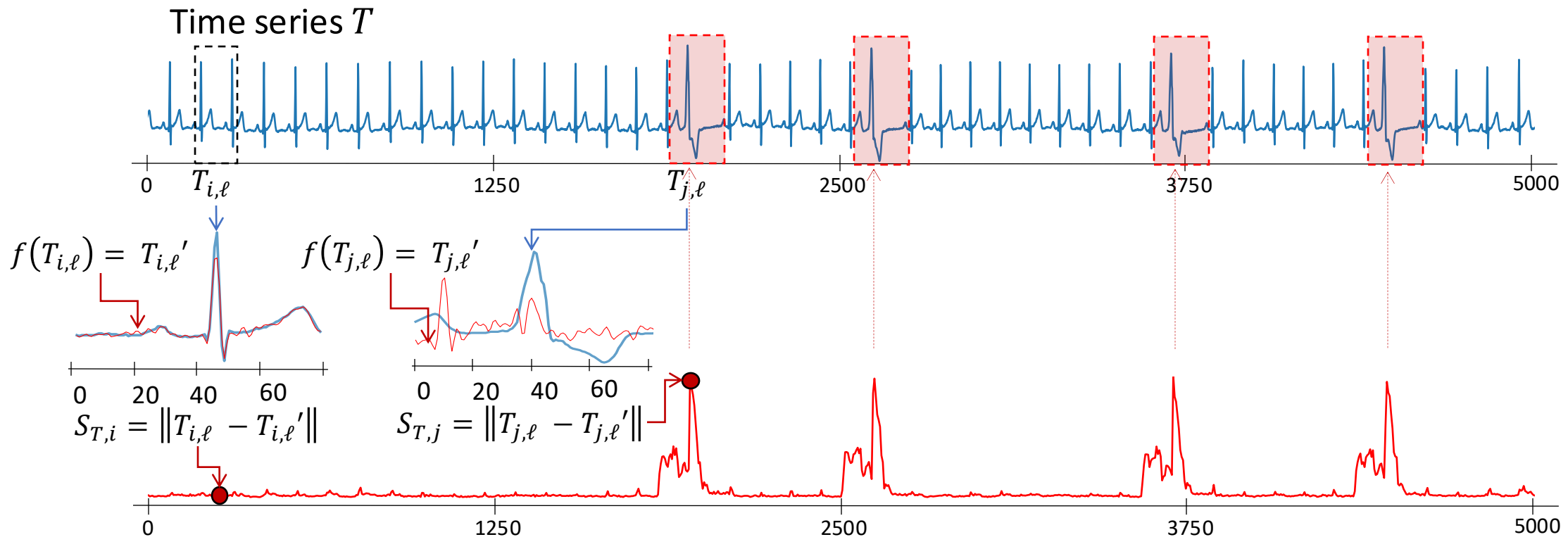
Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.

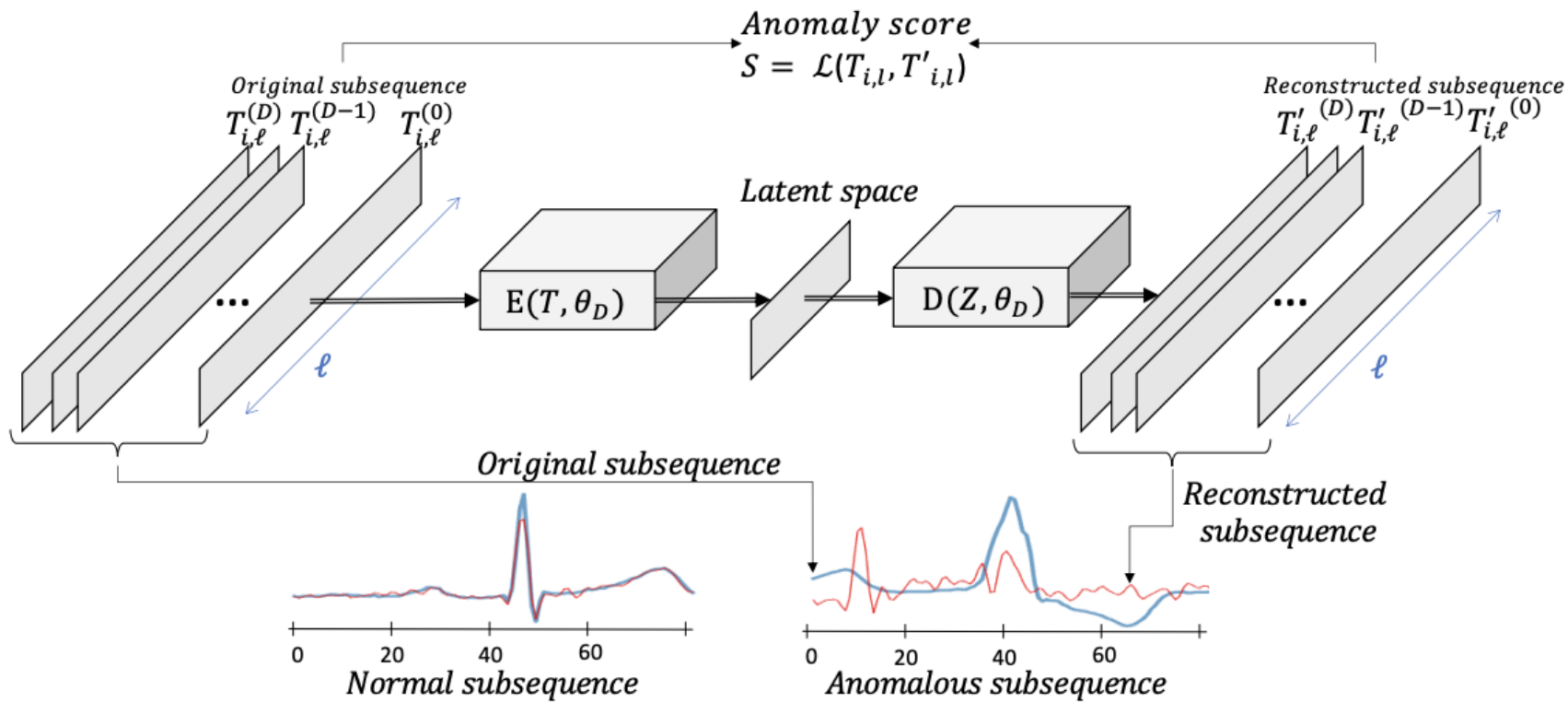


Anomaly Detection methods: *Reconstruction-based*

Methods that aims to **reconstruct** the time series T and use the **reconstruction error** to detect if the time series is an anomaly or not.



Anomaly Detection methods: *an Example*



AutoEncoders [17] (AE)

Neural Network composed of an **encoder** (that reduce the dimensionality) and **decoder** that **reconstruct** the time series. The objective is to **minimize the reconstruction error**.

Semi-supervised

Univariate/Multivariate

Point/sequence

Anomaly Detection methods: *Existing benchmark*

Anomaly Detection methods: *Existing benchmark*

HEX/UCR [18]

Set of **250 time series** with labels.

Details

- The labels have been manually checked and are reliable
- Each time series contains only 1 labeled anomaly

Anomaly Detection methods: *Existing benchmark*

HEX/UCR [18]

Set of **250 time series** with labels.

Details

- The labels have been manually checked and are reliable
- Each time series contains only 1 labeled anomaly

TimeEval [5]

Set of **976 time series** with labels.

Details

- New synthetic benchmark GutenTag used to tune parameters
- Only Time series with low contamination rate (< 0.1)
- Time series with at least one methods above 0.8 AUC-ROC

Anomaly Detection methods: *Existing benchmark*

HEX/UCR [18]

Set of **250 time series** with labels.

Details

- The labels have been manually checked and are reliable
- Each time series contains only 1 labeled anomaly

TimeEval [5]

Set of **976 time series** with labels.

Details

- New synthetic benchmark GutenTag used to tune parameters
- Only Time series with low contamination rate (< 0.1)
- Time series with at least one methods above 0.8 AUC-ROC

TSB-UAD [19]

Set of **2000 time series** with labels.

Details

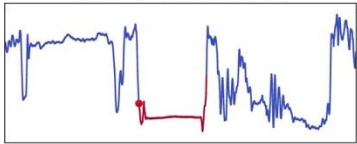
- Collected as proposed in the literature (no filtering based on contamination, size or label quality)
- Artificial and synthetic data generation methods for reliable labels

Anomaly Detection methods: *Existing benchmark*

HEX/UCR [18]

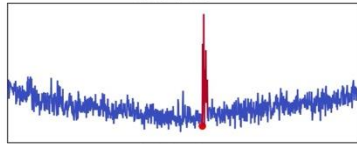
Set of 250 time series with

OPPORTUNITY

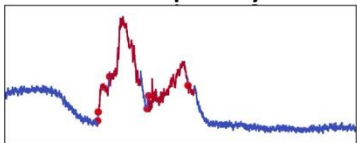


Occupancy

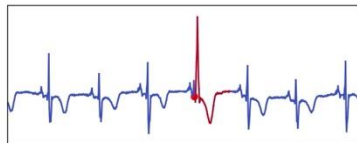
IOPS



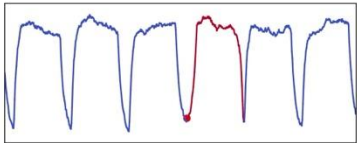
ECG



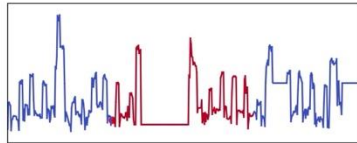
KDD21



NASA-SMAP



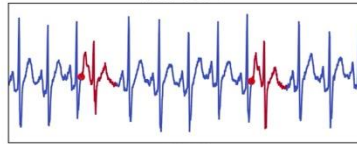
only 1 labeled anomaly



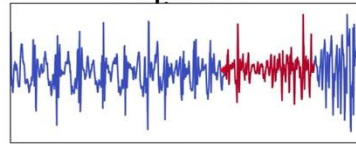
TimeEval [5]

Real datasets collection

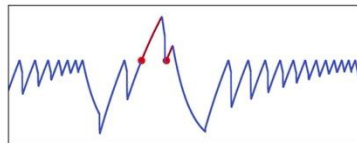
SVDB



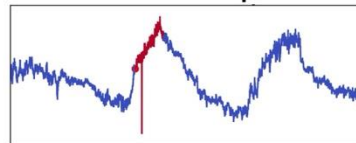
Daphnet



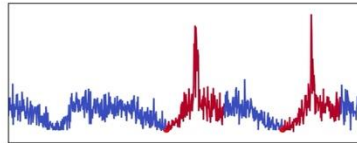
GHL



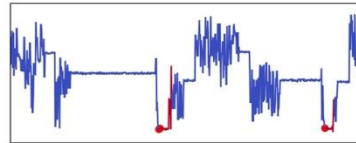
SensorScope



NAB



Genesis



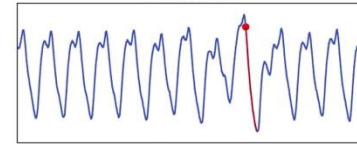
contamination rate (< 0.1)

- Time series with at least one methods above 0.8 AUC-ROC

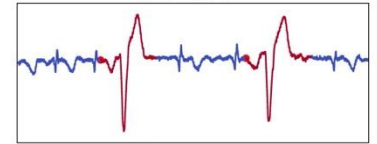
TSB-UAD [19]

Set of 2000 time series with

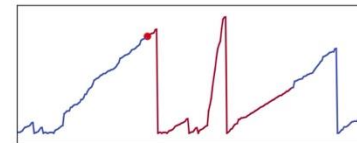
MGAB



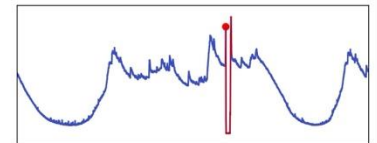
MITDB



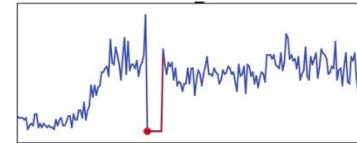
NASA-MSL



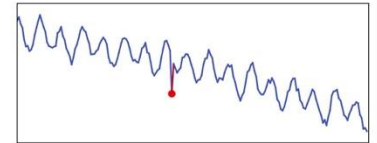
SMD



Dodgers



YAHOO



contamination, size or label quality.

Anomaly Detection methods: *Existing benchmark*

HEX/UCR [18]

TimeEval [5]

TSB-UAD [19]

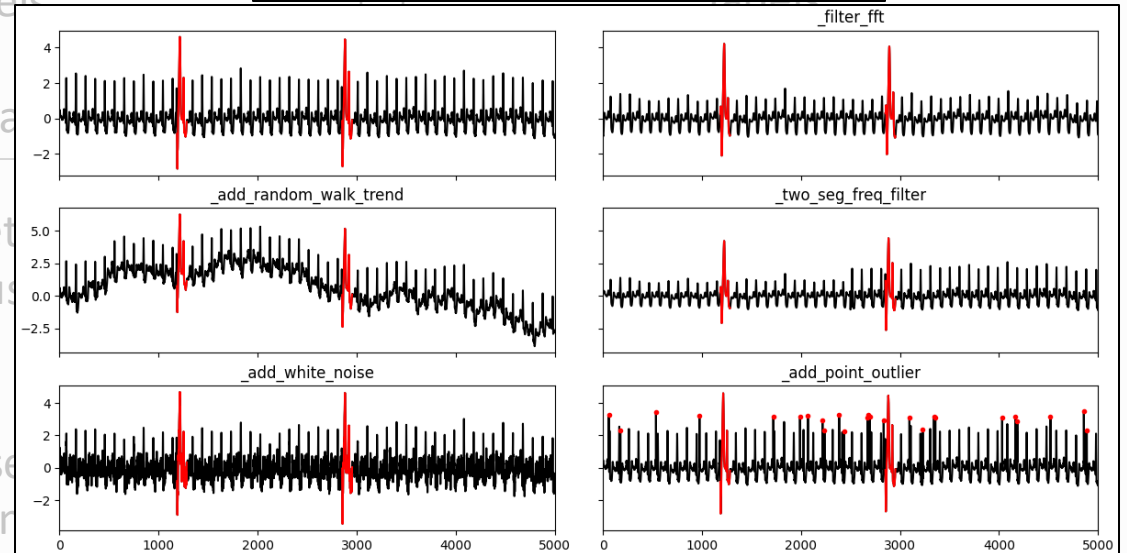
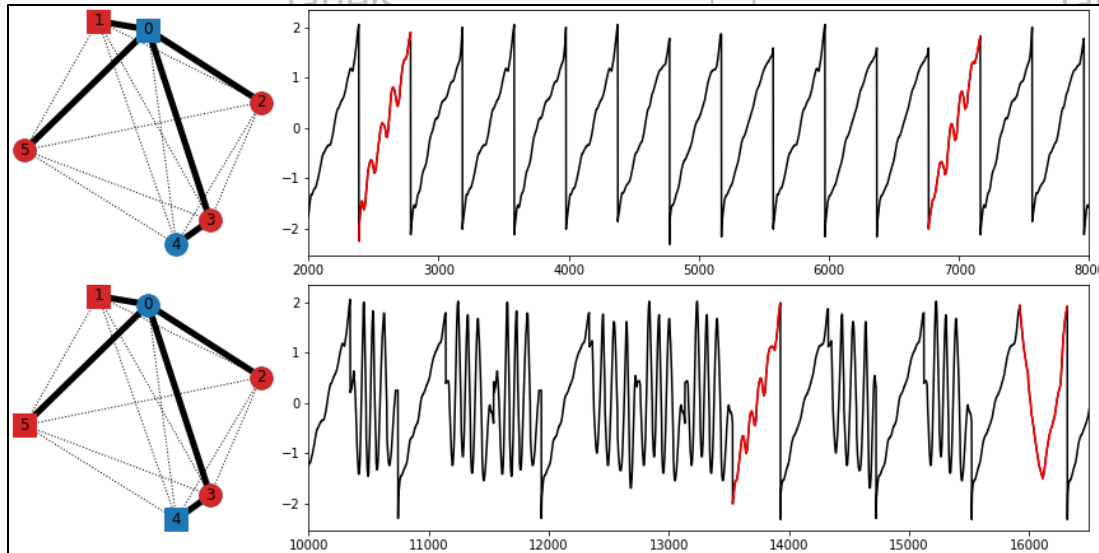
Set of 250

Artificial dataset generation

of 976 time series with

Synthetic dataset generation

ies with



- Time series with at least one methods above 0.8 AUC-ROC

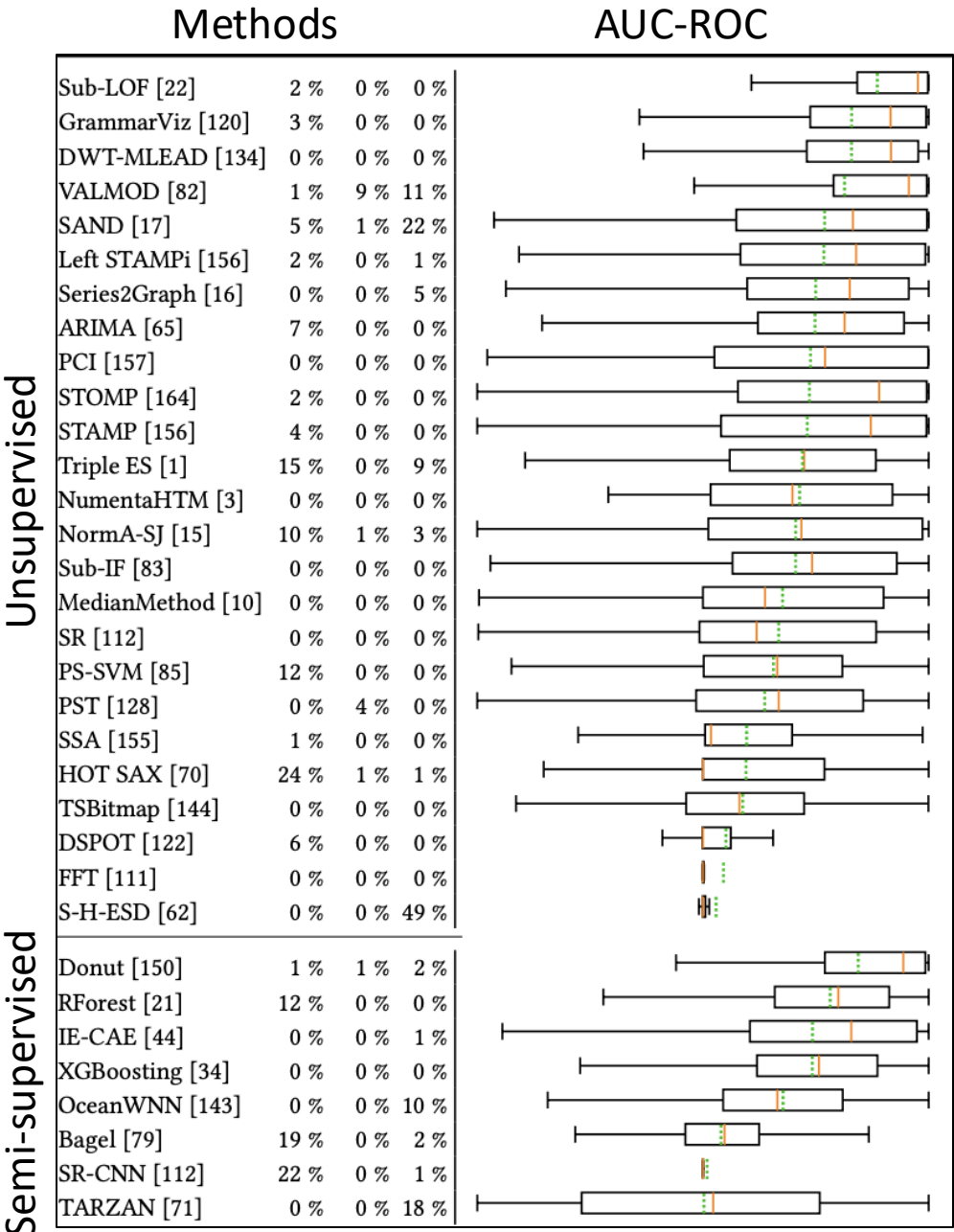
quality.

Anomaly Detection methods:

Experimental evaluation

Observations on TimeEval [5]:

[5] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. Anomaly detection in time series: a comprehensive evaluation. Proc. VLDB Endow. 15, 9 (May 2022), 1779–1797.



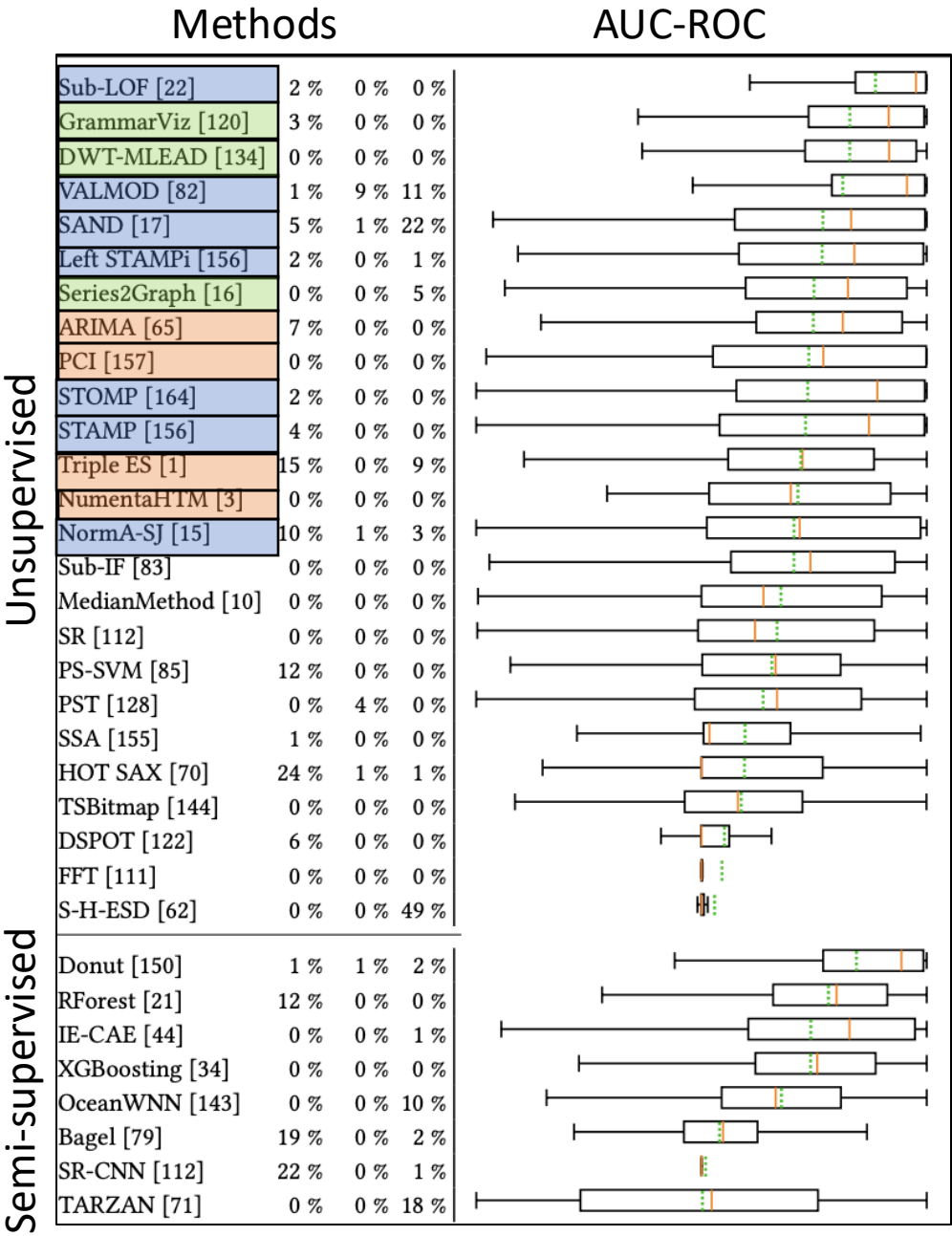
Anomaly Detection methods:

Experimental evaluation

Observations on TimeEval [5]:

- Distance-based and Density-based methods have a better accuracy (AUC-ROC) than forecasting and reconstruction-based approaches

[5] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. Anomaly detection in time series: a comprehensive evaluation. Proc. VLDB Endow. 15, 9 (May 2022), 1779–1797.



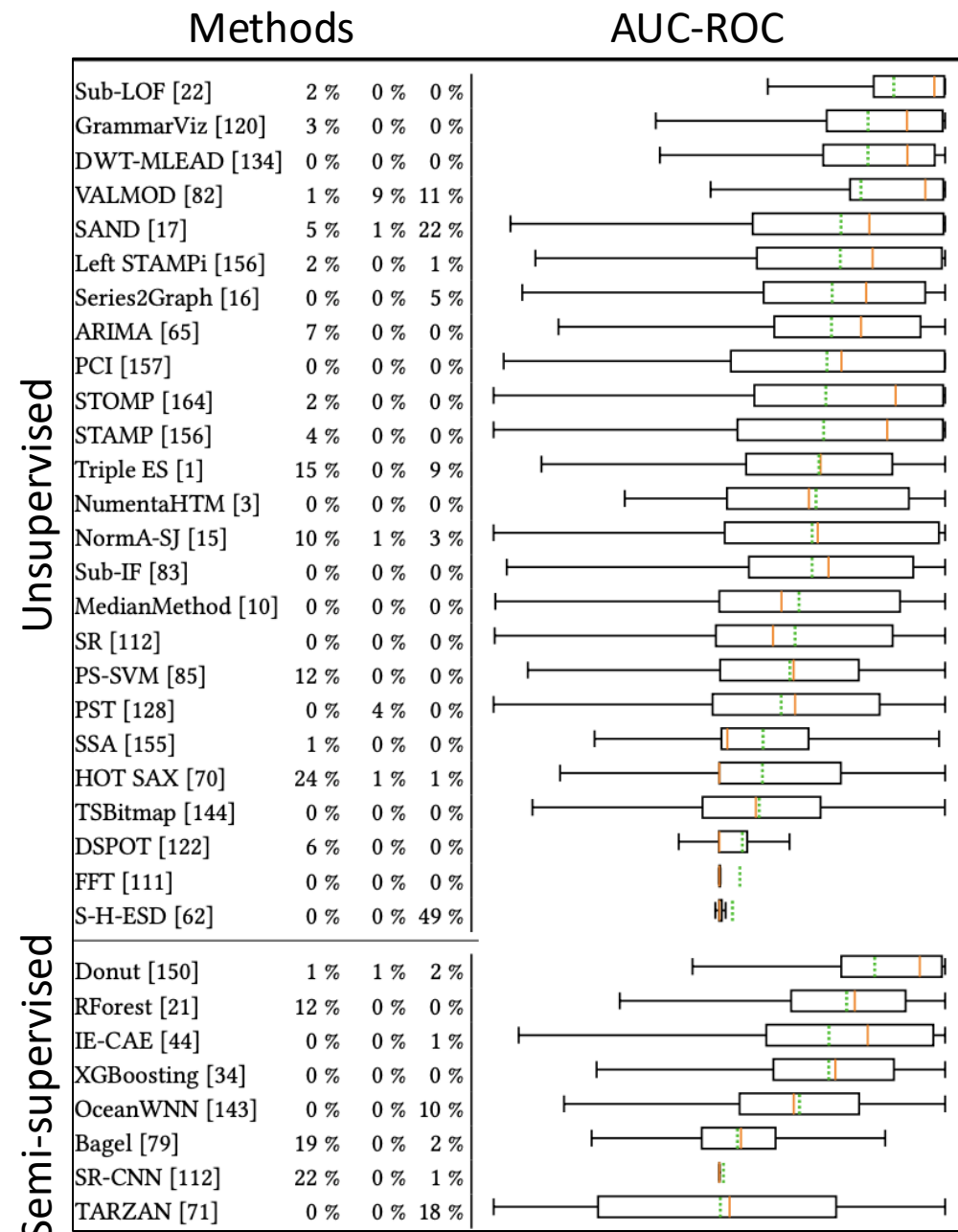
Anomaly Detection methods:

Experimental evaluation

Observations on TimeEval [5]:

- Distance-based and Density-based methods have a better accuracy (AUC-ROC) than forecasting and reconstruction-based approaches
- Semi-supervised methods are not outperforming Unsupervised approaches

[5] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. Anomaly detection in time series: a comprehensive evaluation. Proc. VLDB Endow. 15, 9 (May 2022), 1779–1797.

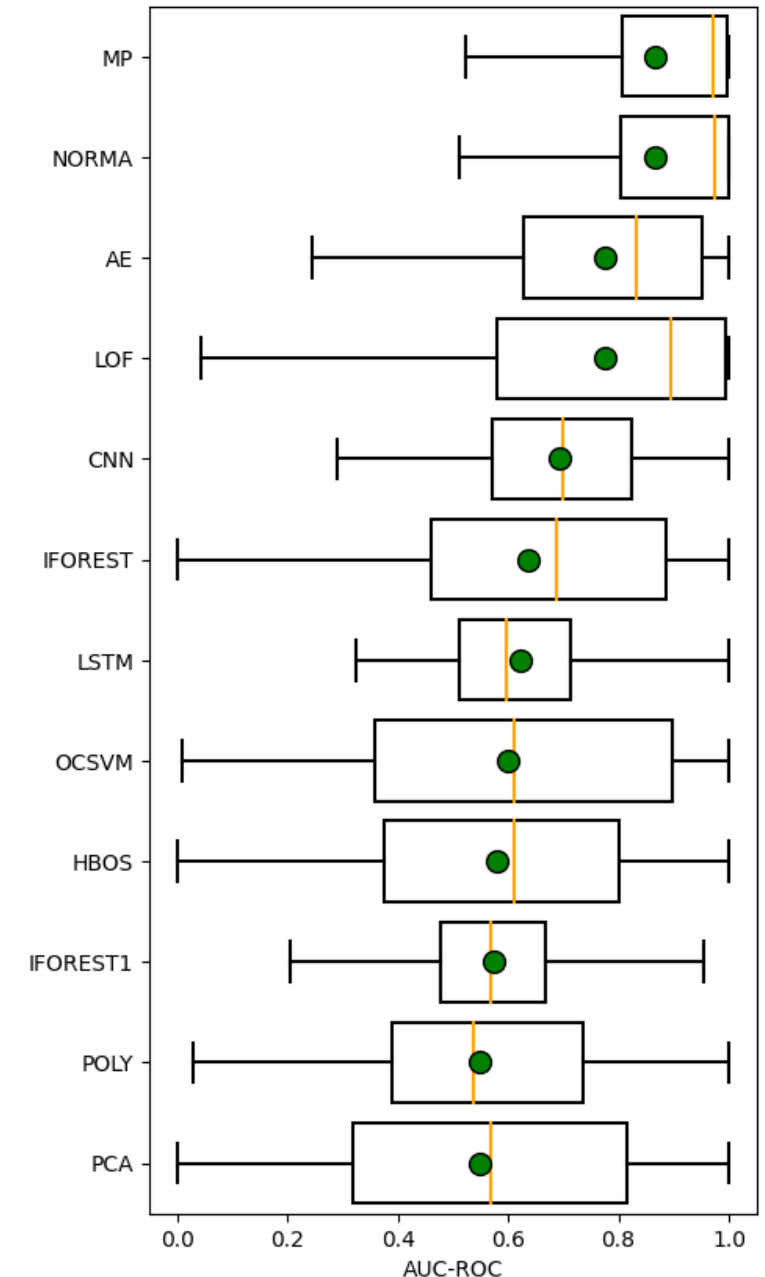


Anomaly Detection methods: *Experimental evaluation*

Observations on HEX/UCR [18]:

- Distance-based methods have a better accuracy (AUC-ROC) than forecasting and distribution-based approaches

[18] R. Wu and E. Keogh, "Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress" in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 03, pp. 2421-2429, 2023.

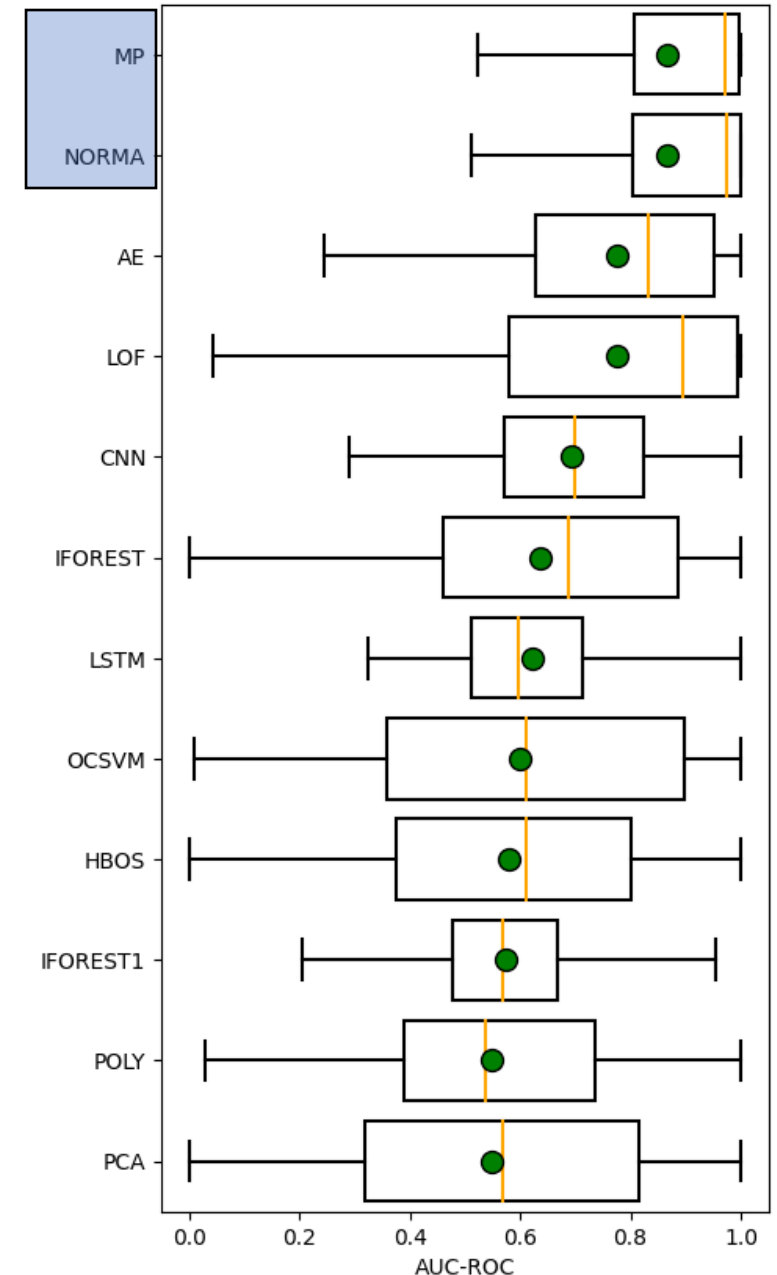


Anomaly Detection methods: *Experimental evaluation*

Observations on HEX/UCR [18]:

- Distance-based methods have a better accuracy (AUC-ROC) than forecasting and distribution-based approaches

[18] R. Wu and E. Keogh, "Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress" in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 03, pp. 2421-2429, 2023.

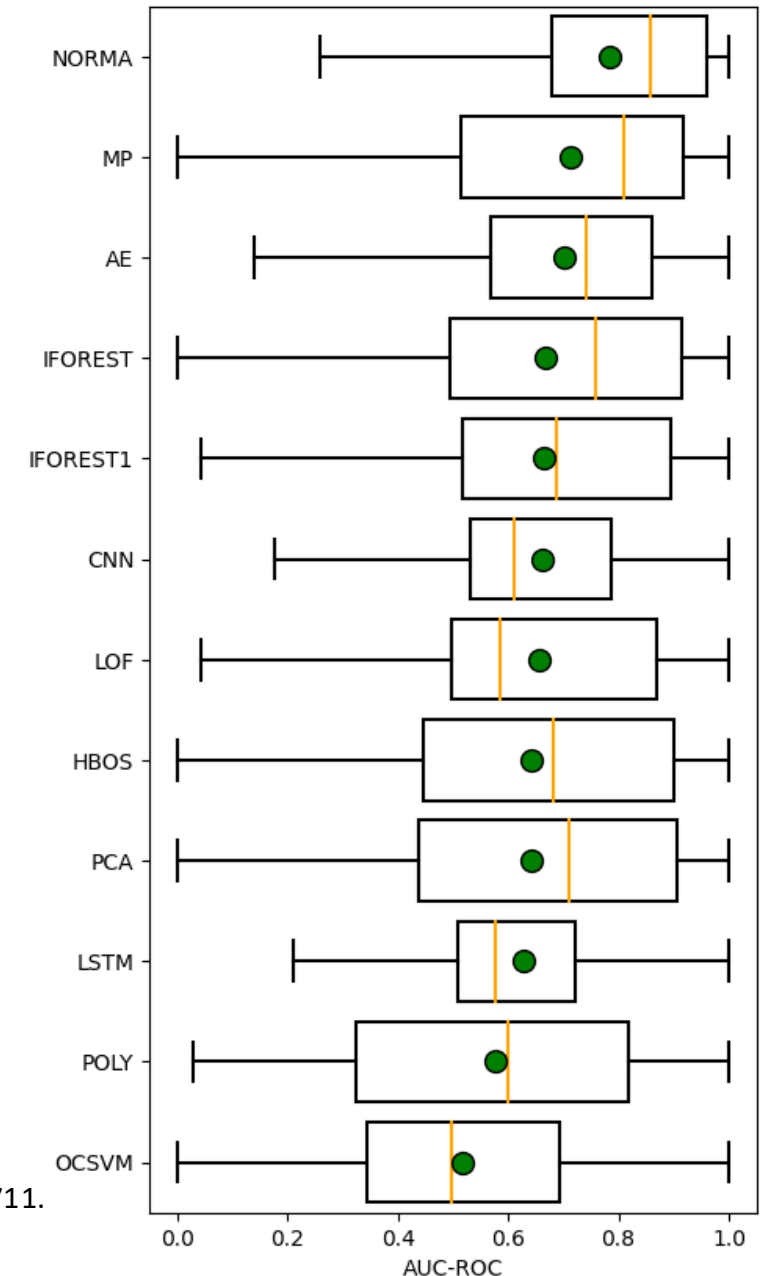


Anomaly Detection methods:

Experimental evaluation

Observations on TSB-UAD [19]:

- Distance-based methods have a better accuracy (AUC-ROC) than forecasting-based methods.
- Isolation Forest (distribution-based and not proposed for time series) have also a strong accuracy
- AutoEncoder (AE) is also very accurate.



[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.

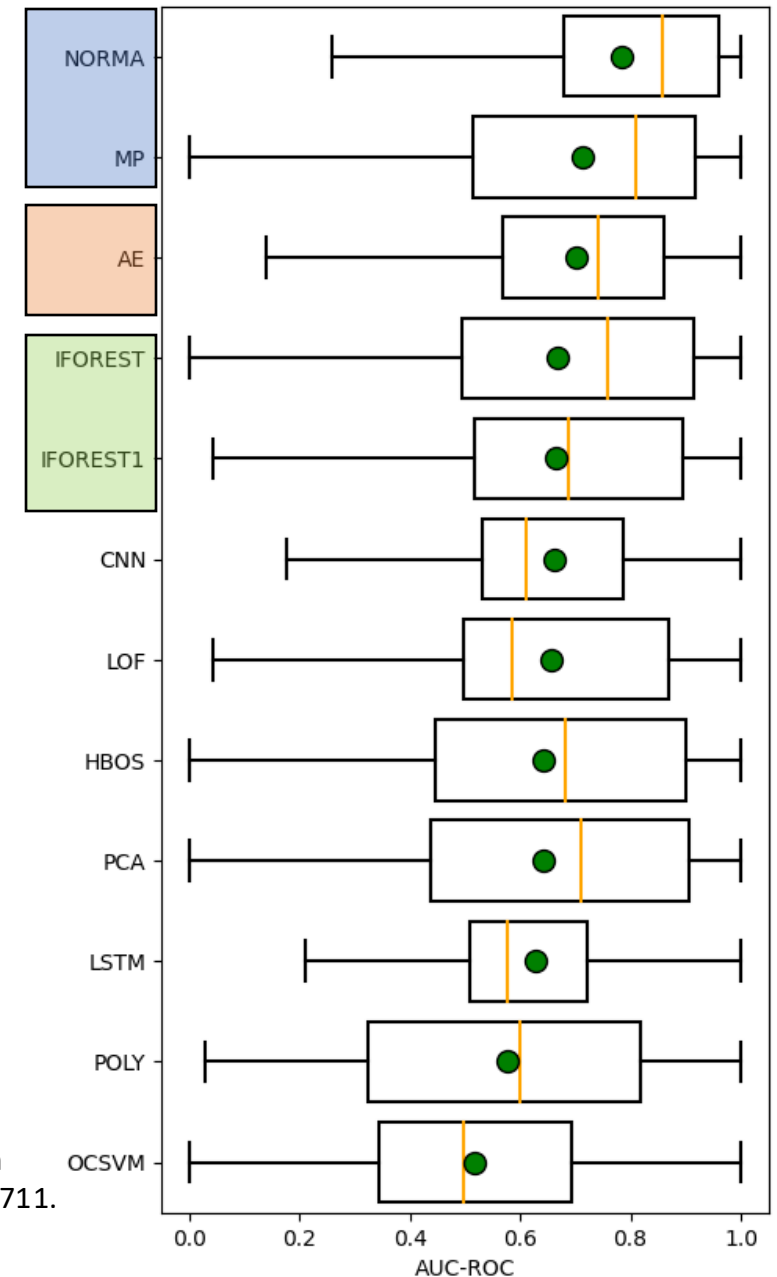
Anomaly Detection methods:

Experimental evaluation

Observations on TSB-UAD [19]:

- Distance-based methods have a better accuracy (AUC-ROC) than forecasting-based methods.
- Isolation Forest (distribution-based and not proposed for time series) have also a strong accuracy
- AutoEncoder (AE) is also very accurate.

[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.

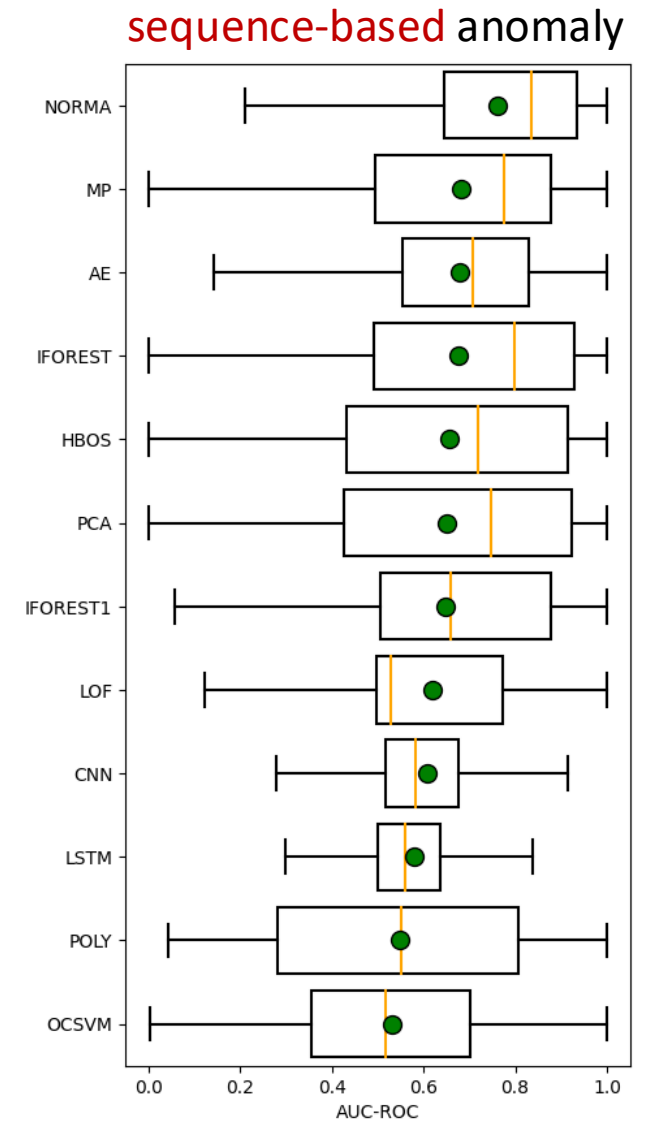
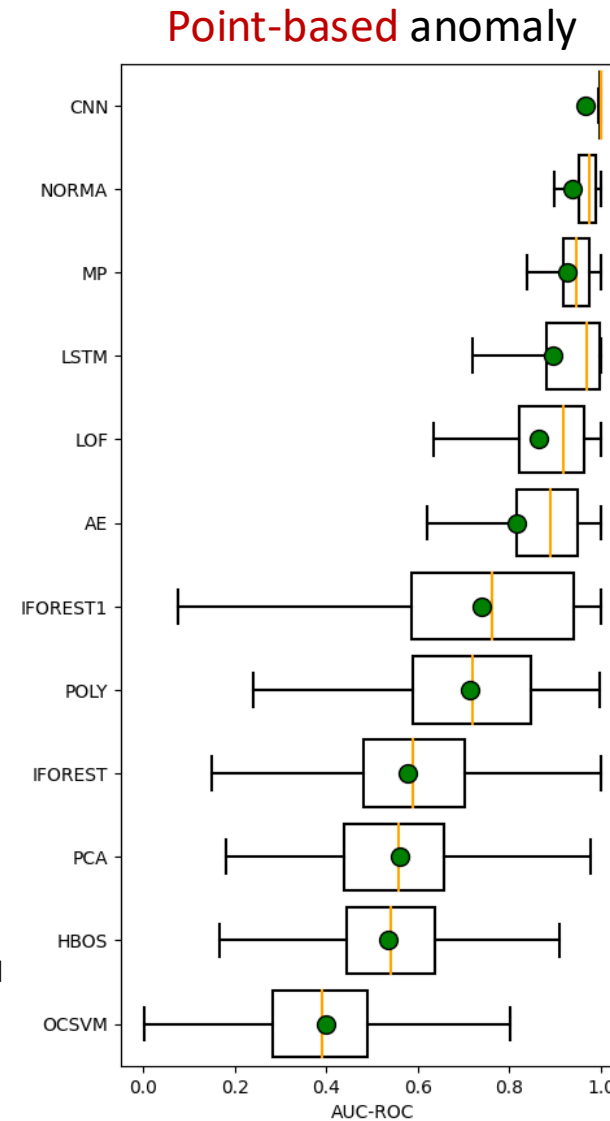


Anomaly Detection methods:

Experimental evaluation

Observations on TSB-UAD [19]:

[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.



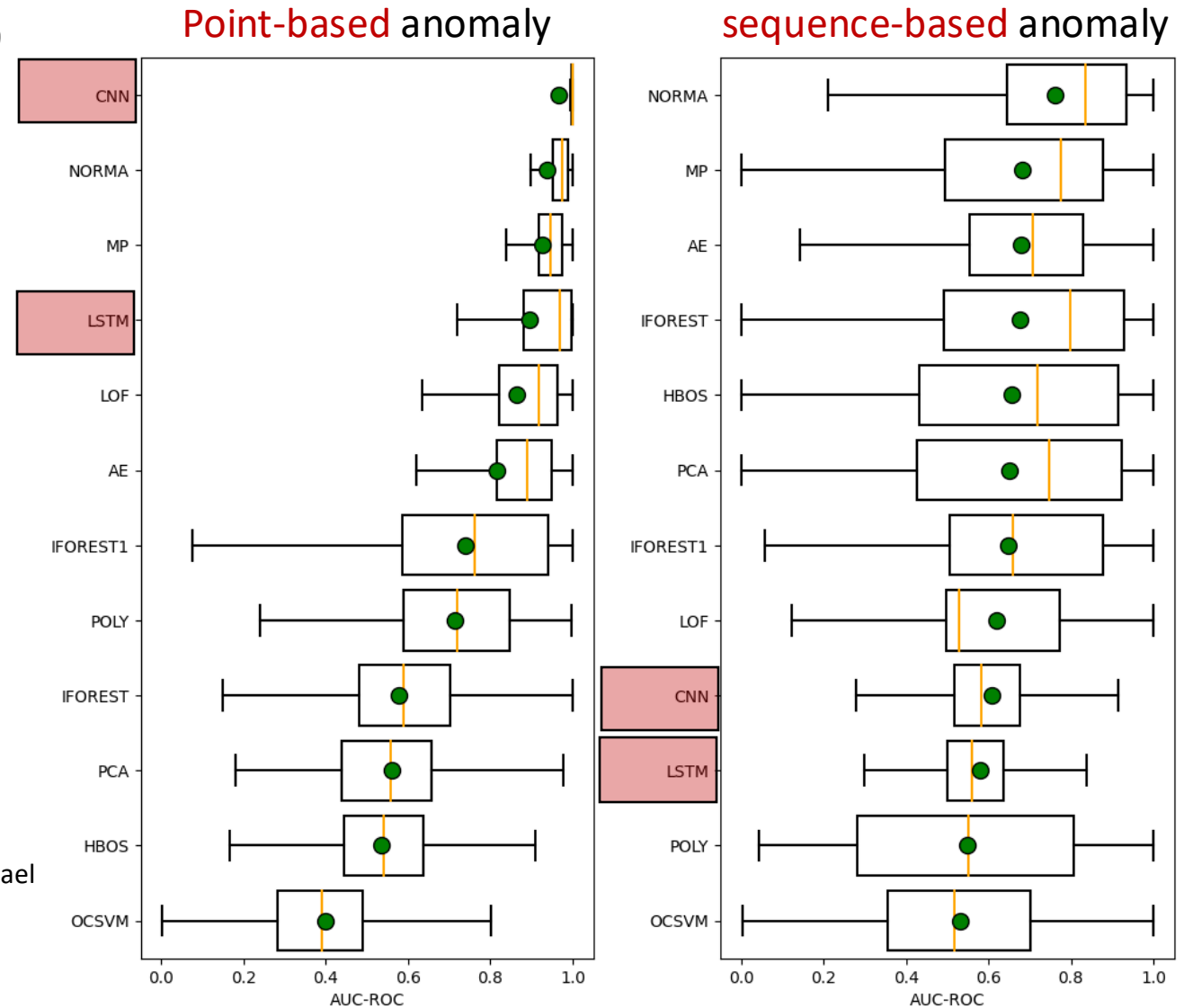
Anomaly Detection methods:

Experimental evaluation

Observations on TSB-UAD [19]:

- Forecasting methods (LSTM and CNN) are very **accurate** for point anomalies
- But have **poor performances** on **sequence-based** anomalies.

[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.

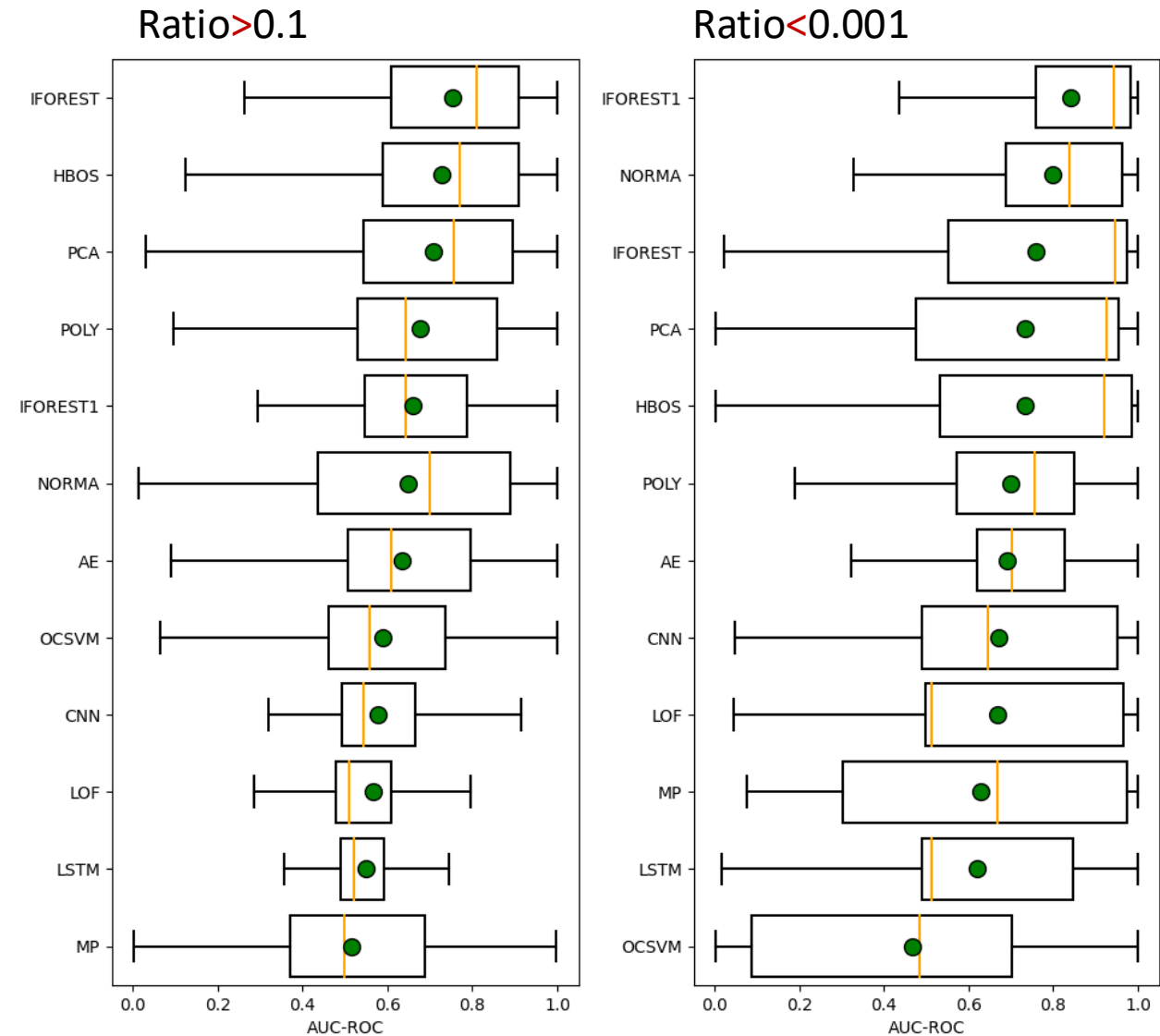


Anomaly Detection methods: *Experimental evaluation*

Observations on TSB-UAD [19]:

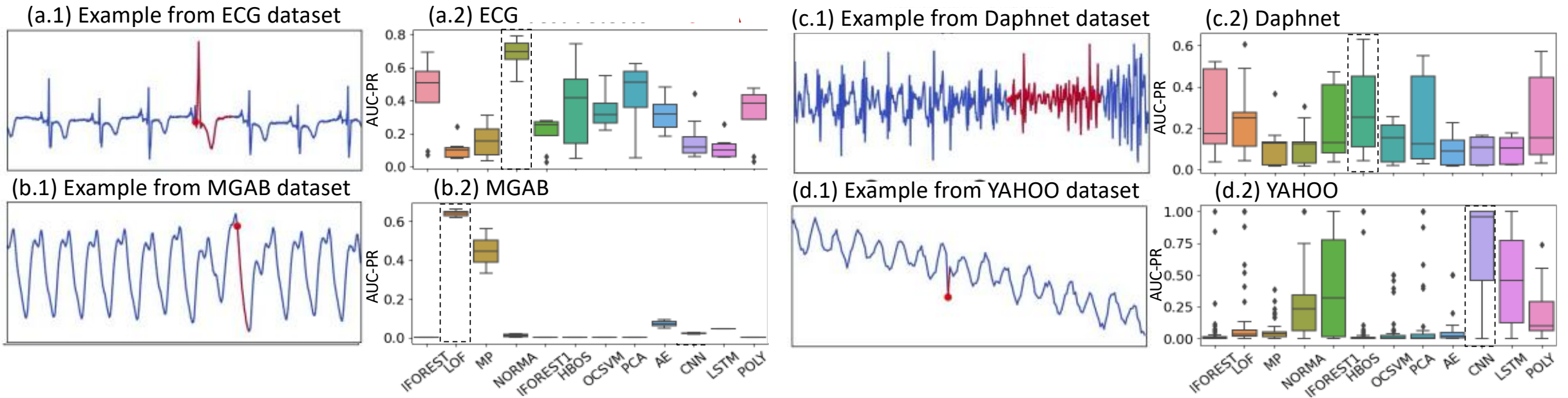
- The ratio of normal/abnormal points has a **strong impact** on the methods ranking.

[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.



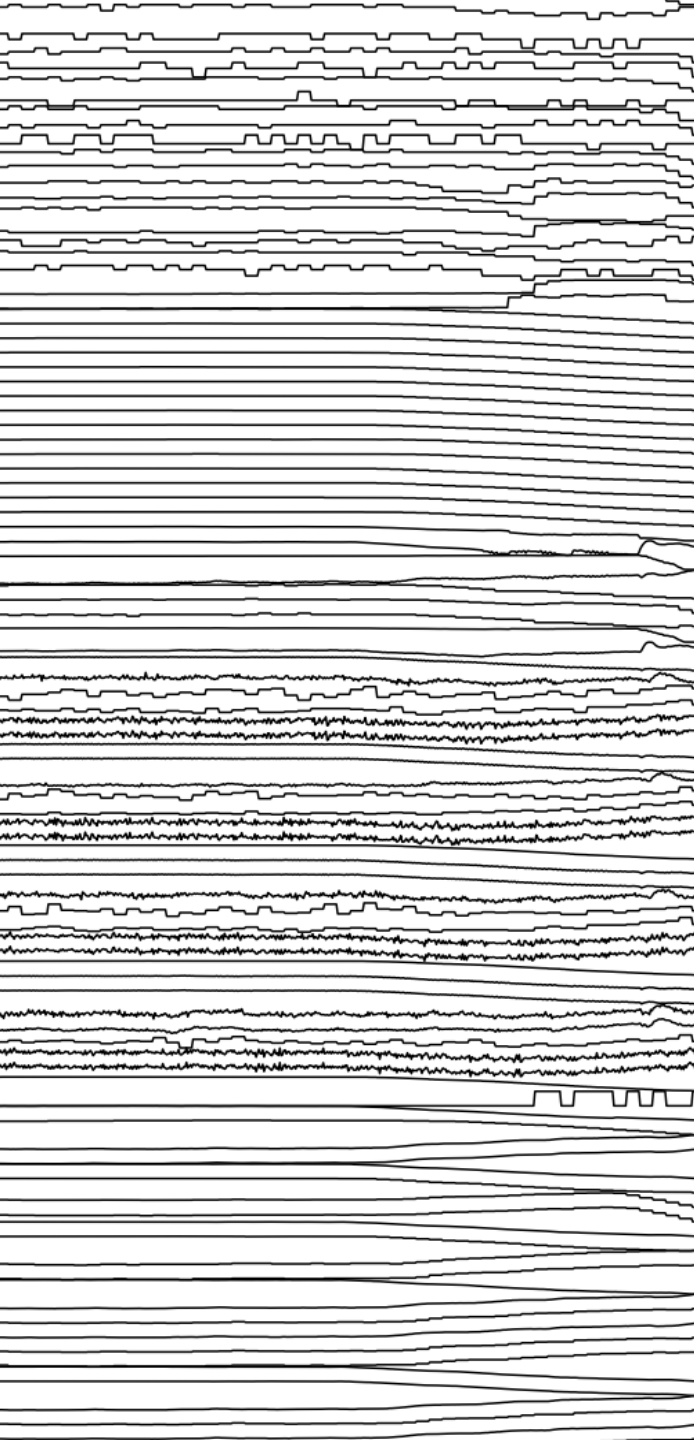
Anomaly Detection methods: *Experimental evaluation*

Observation from the results applied on specific datasets (TSB-UAD [19])



There is **no overall winner**.

[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.



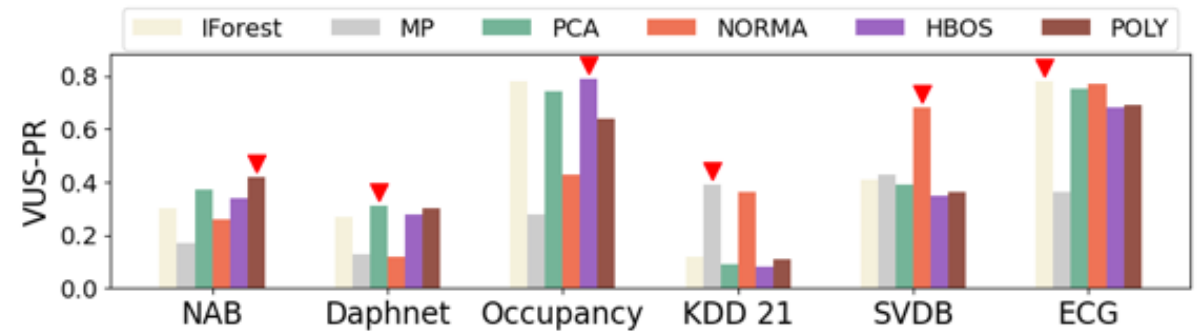
New Trends and Opportunities

Automated Solution: *Background*

Motivation:

- No one-size-fits-all model: How can we *automatically* identify the best anomaly detector given a time series?

Detection accuracy (VUS-PR) for 6 anomaly detectors across different datasets in TSB-UAD [19]

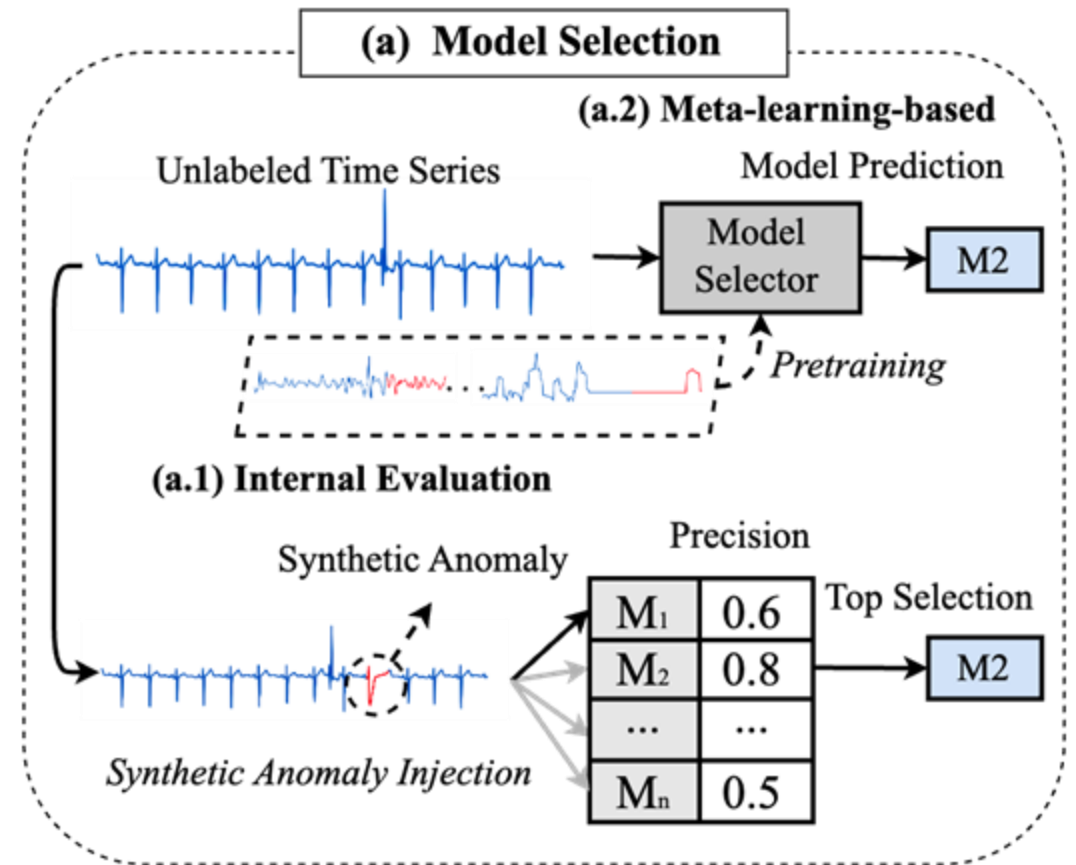


Automated Solution: *Taxonomy*

(a) Model Selection:

Selecting the best anomaly detector from a predefined candidate model set.

- (a.1) *Internal Evaluation*
- (a.2) *Meta-learning-based*



Automated Solution: *Taxonomy*

(a) Model Selection:

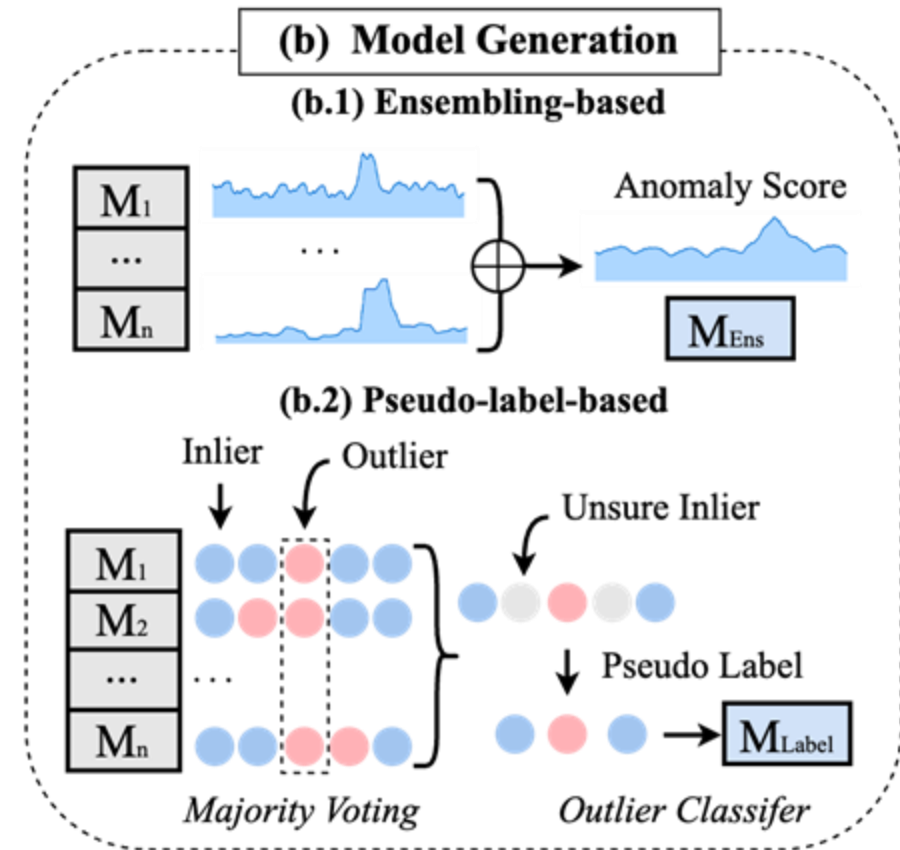
Selecting the best anomaly detector from a predefined candidate model set.

- (a.1) *Internal Evaluation*
- (a.2) *Meta-learning-based*

(b) Model Generation:

Creating an entirely new model for the given time series based on the candidate model set

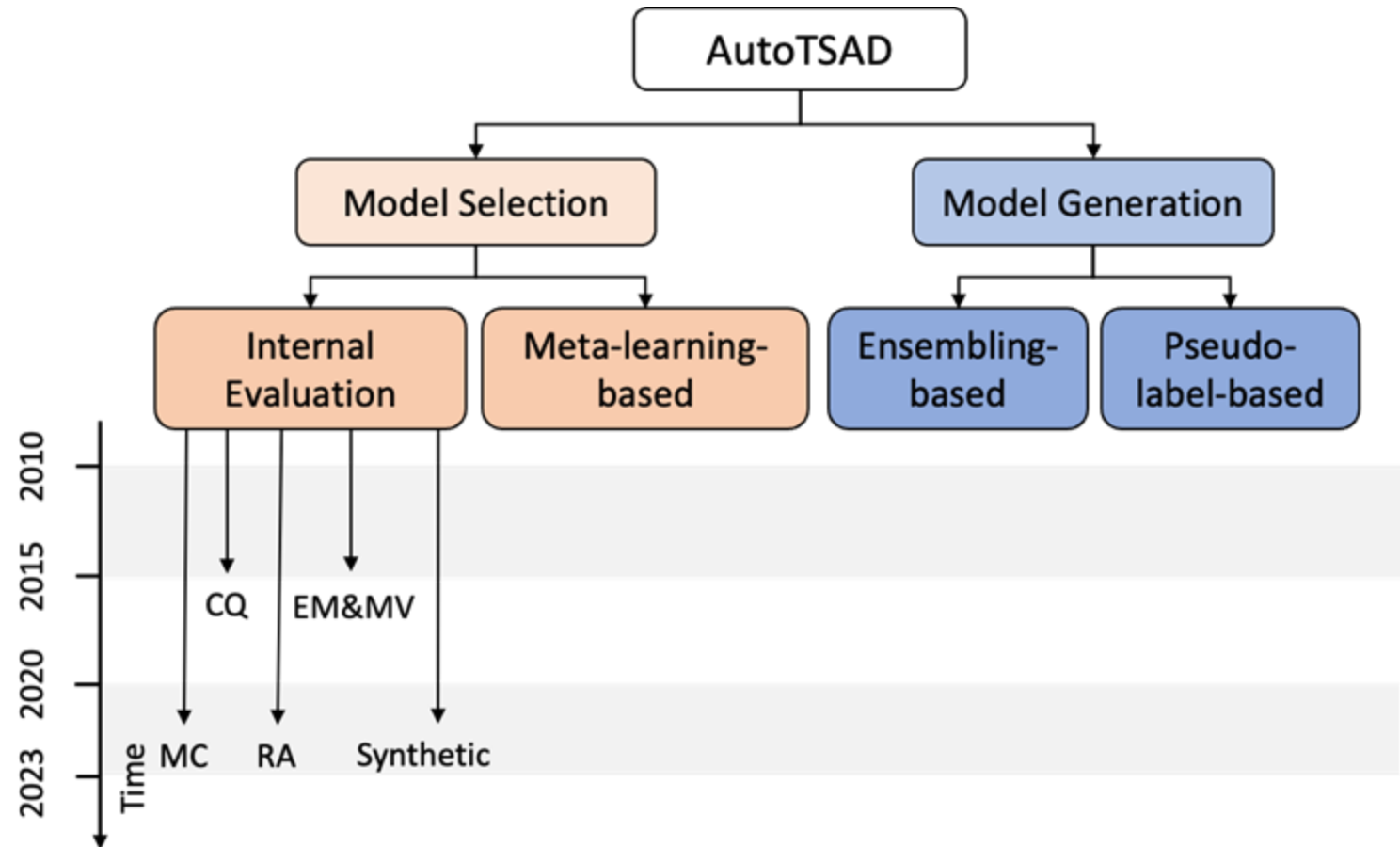
- (b.1) *Ensembling-based*
- (b.2) *Pseudo-label-based*



Automated Solution: *Internal Evaluation*

Definition: Evaluate the effectiveness of a model without any reliance on external information

- **Stand-alone:** Clustering Quality, EM&MV, Synthetic anomaly injection
- **Collective:** Model Centrality, Rank Aggregation



Automated Solution: *Internal Evaluation*

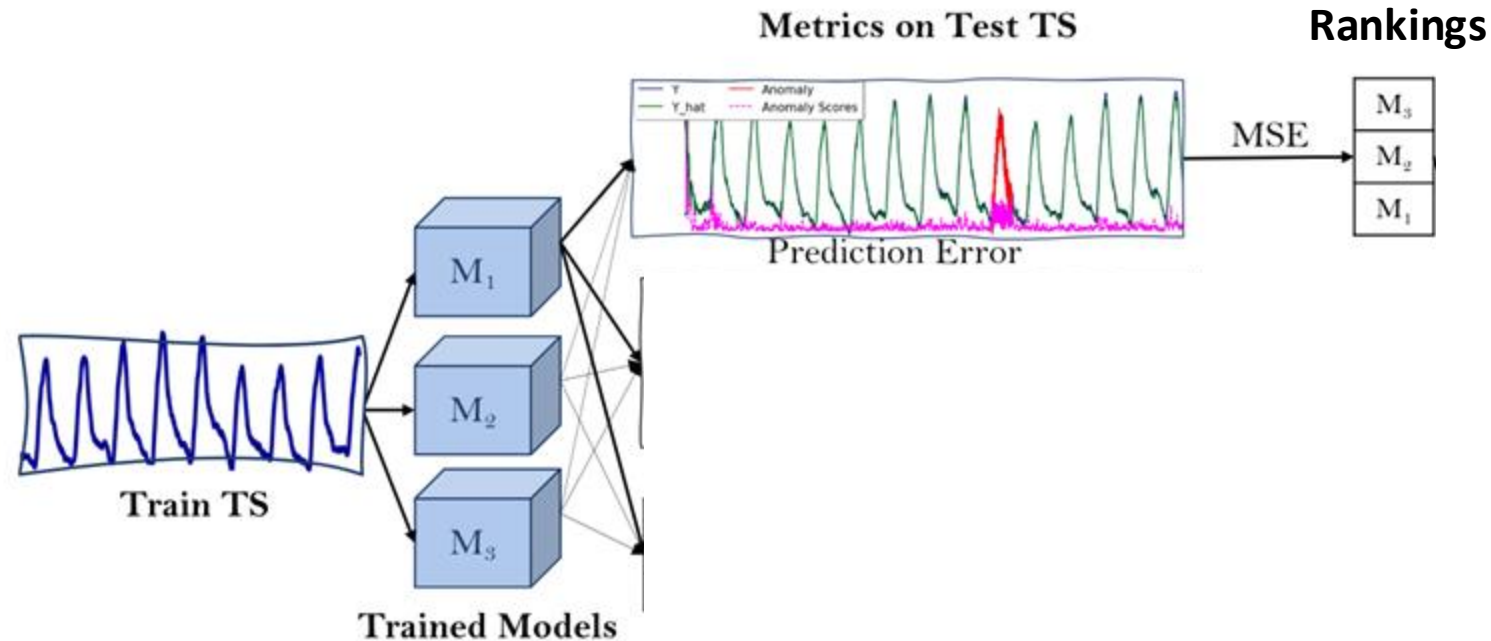


Image from [28]: Internal Evaluation workflow.

Automated Solution: *Internal Evaluation*

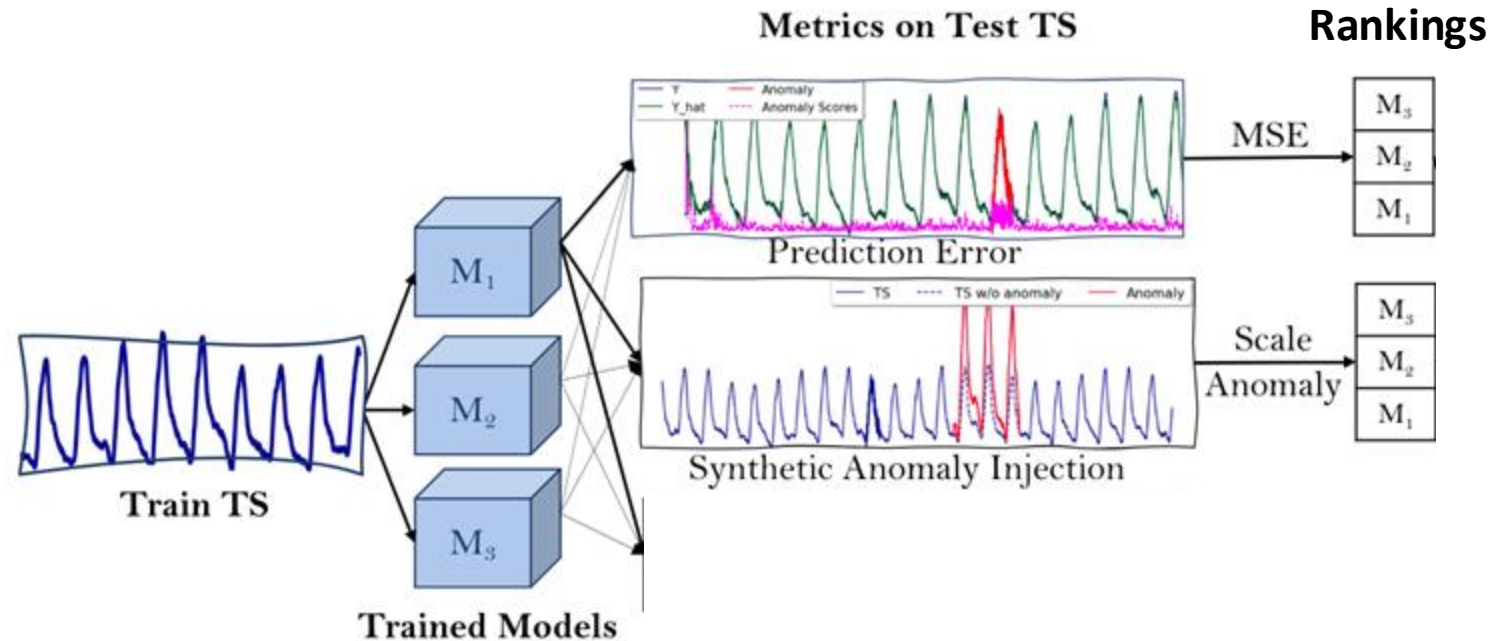


Image from [28]: Internal Evaluation workflow.

Automated Solution: *Internal Evaluation*

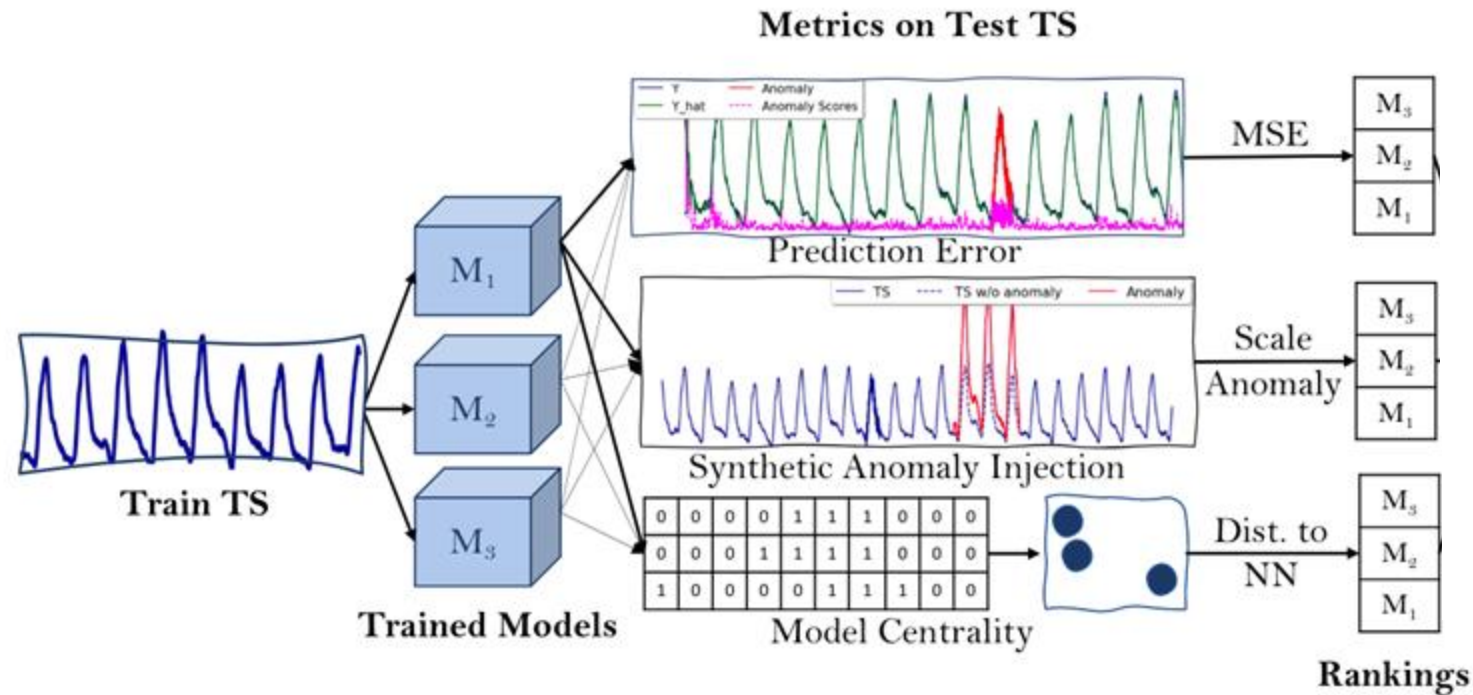


Image from [28]: Internal Evaluation workflow.

Automated Solution: *Internal Evaluation*

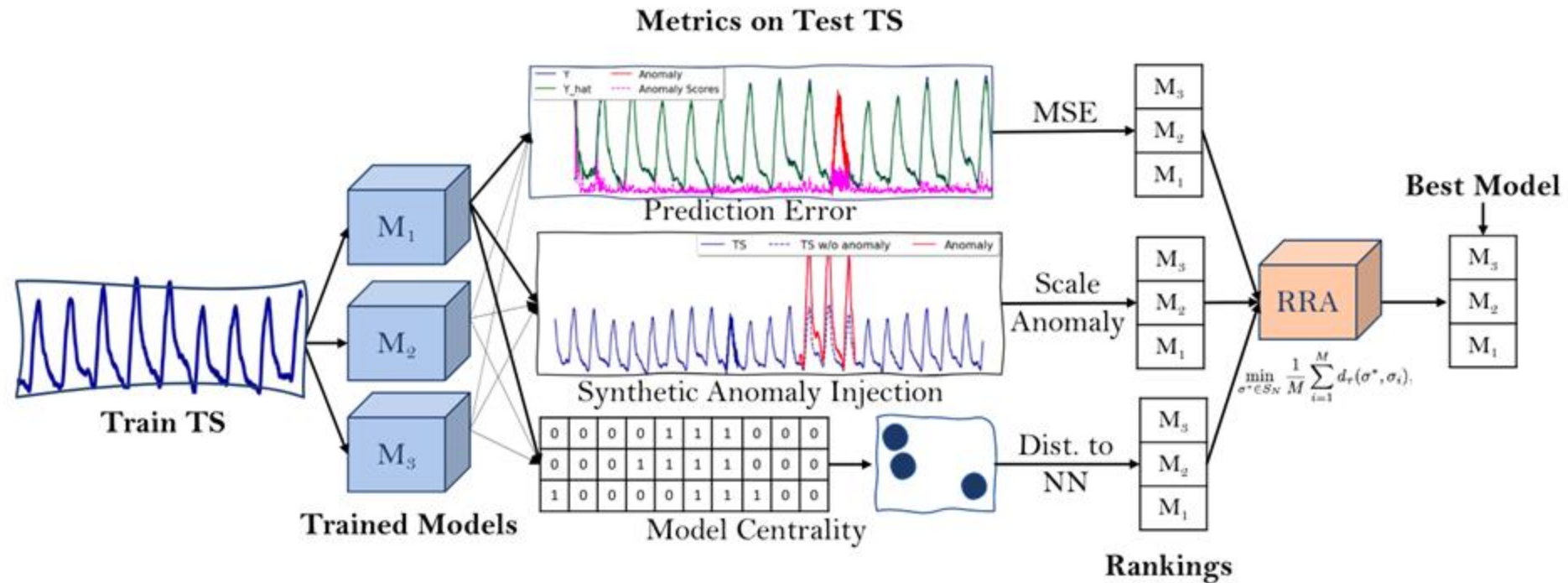
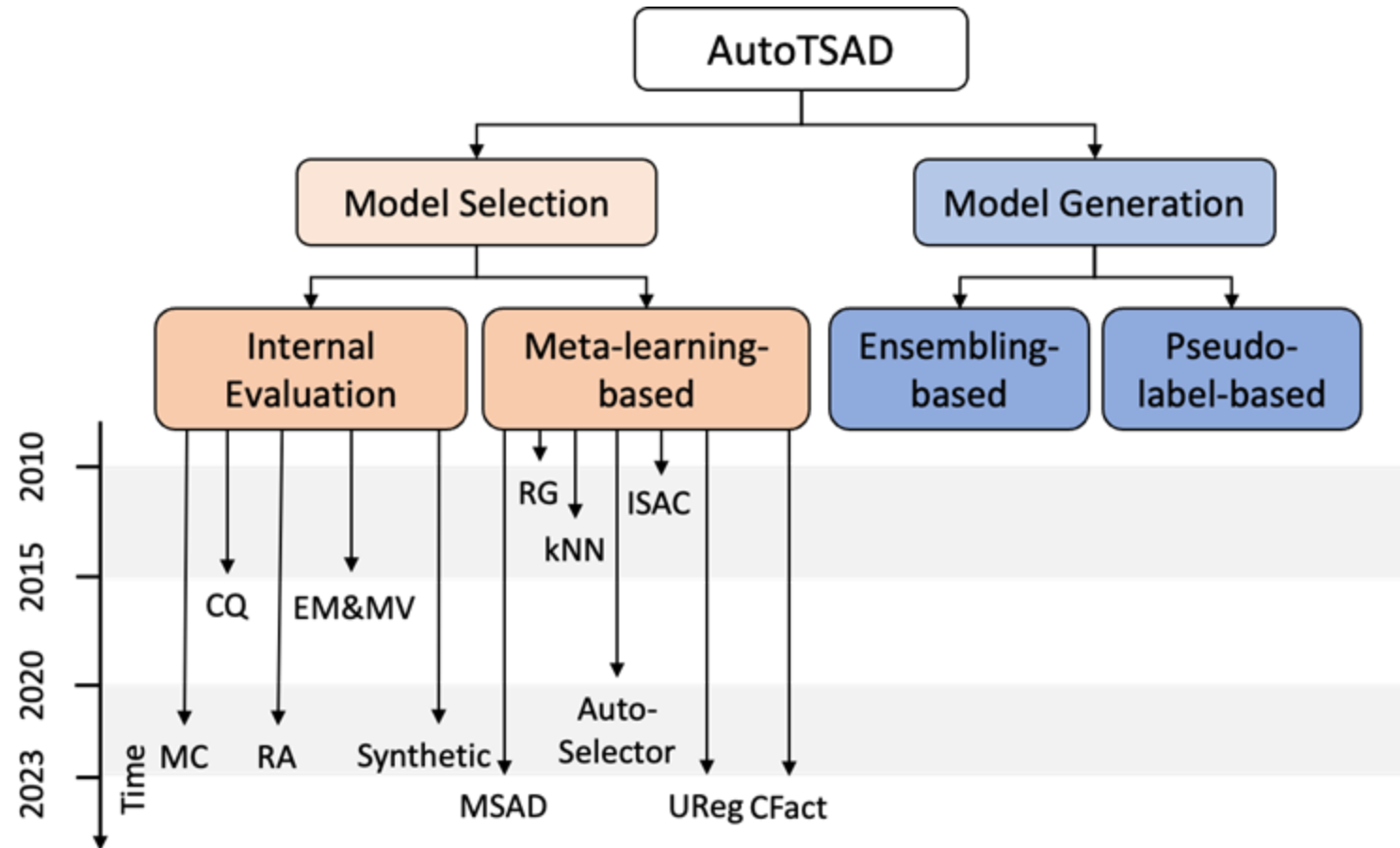


Image from [28]: Internal Evaluation workflow.

Automated Solution: *Meta-learning-based*

Definition: Using insights from historical labeled datasets to select the best model for new data

- **Classification:** Auto-Selector, MSAD
- **Regression:** RG, UReg, Cfact
- **Nearest Neighbor:** kNN
- **Other Optimization:** ISAC, MetaOD



Automated Solution: *Meta-learning-based*

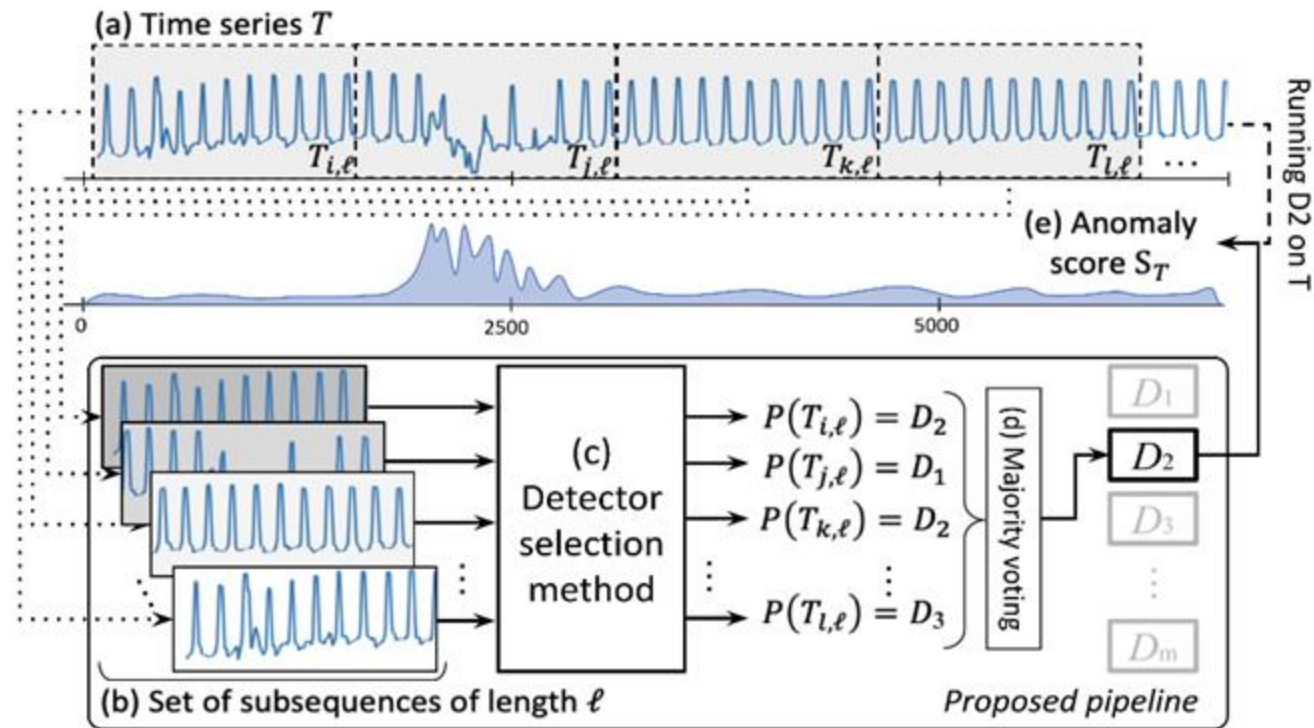
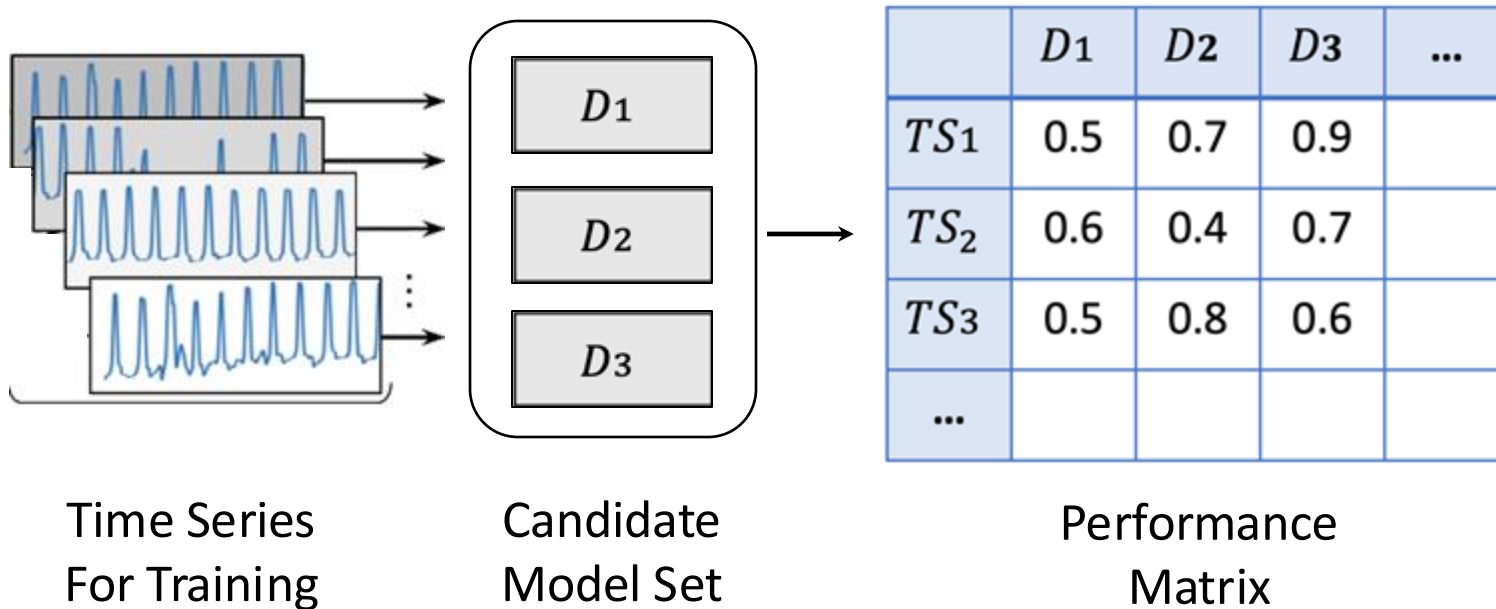


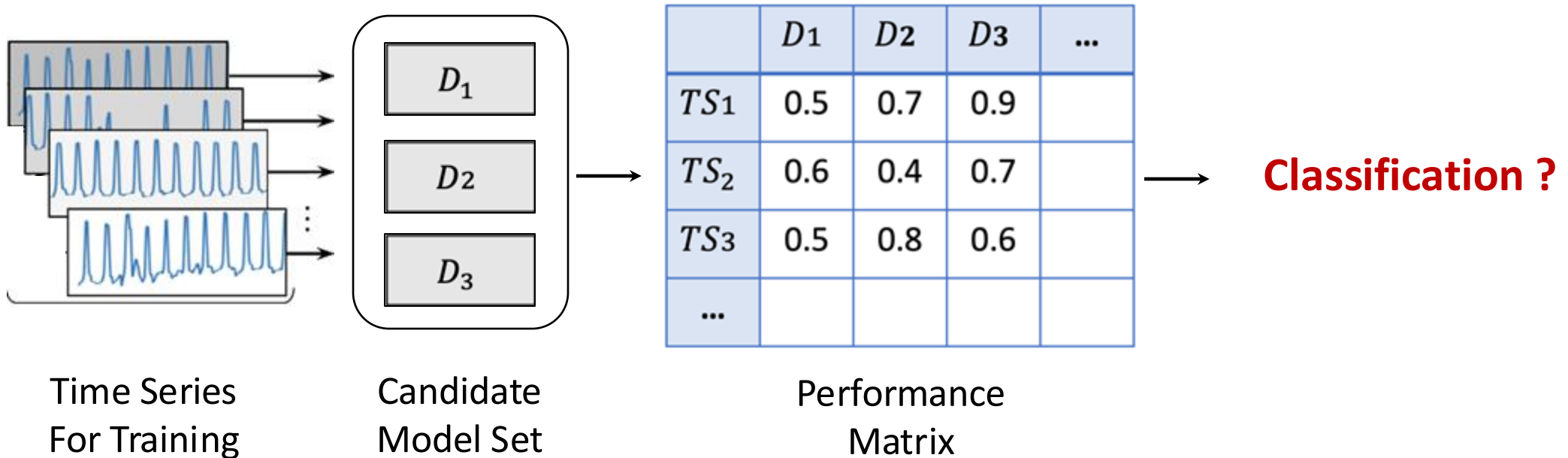
Image from [29]: Model Selection Pipeline.

Automated Solution: *Meta-learning-based*

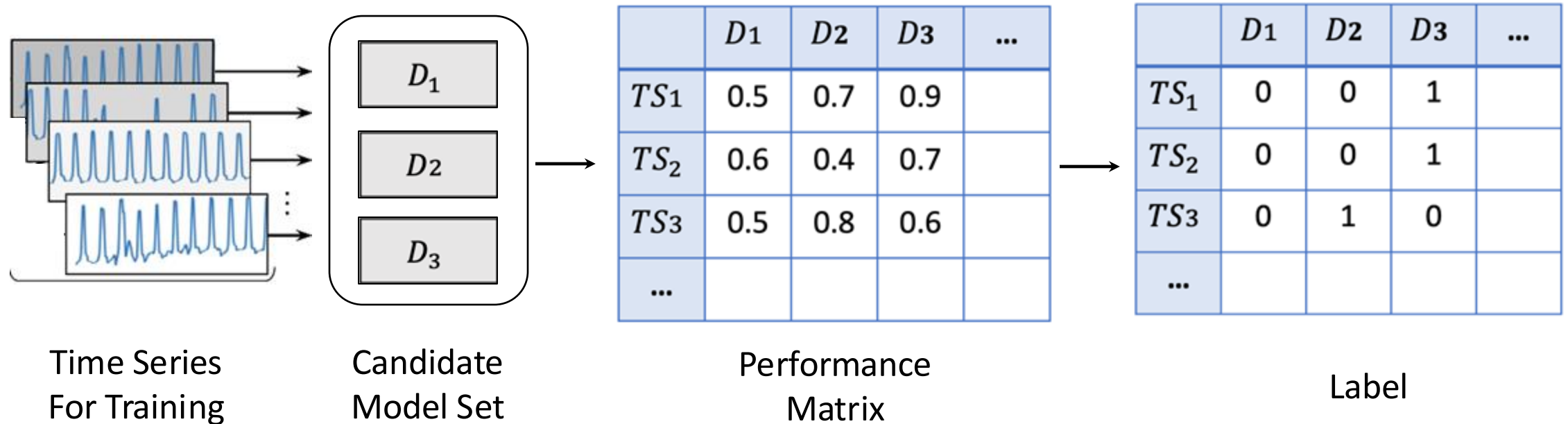


Performance measures:
F-score, AUC-PR, VUS-PR ...

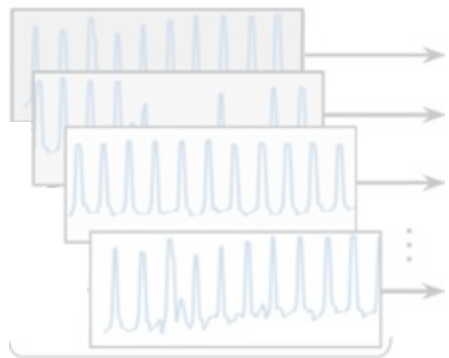
Automated Solution: *Meta-learning-based*



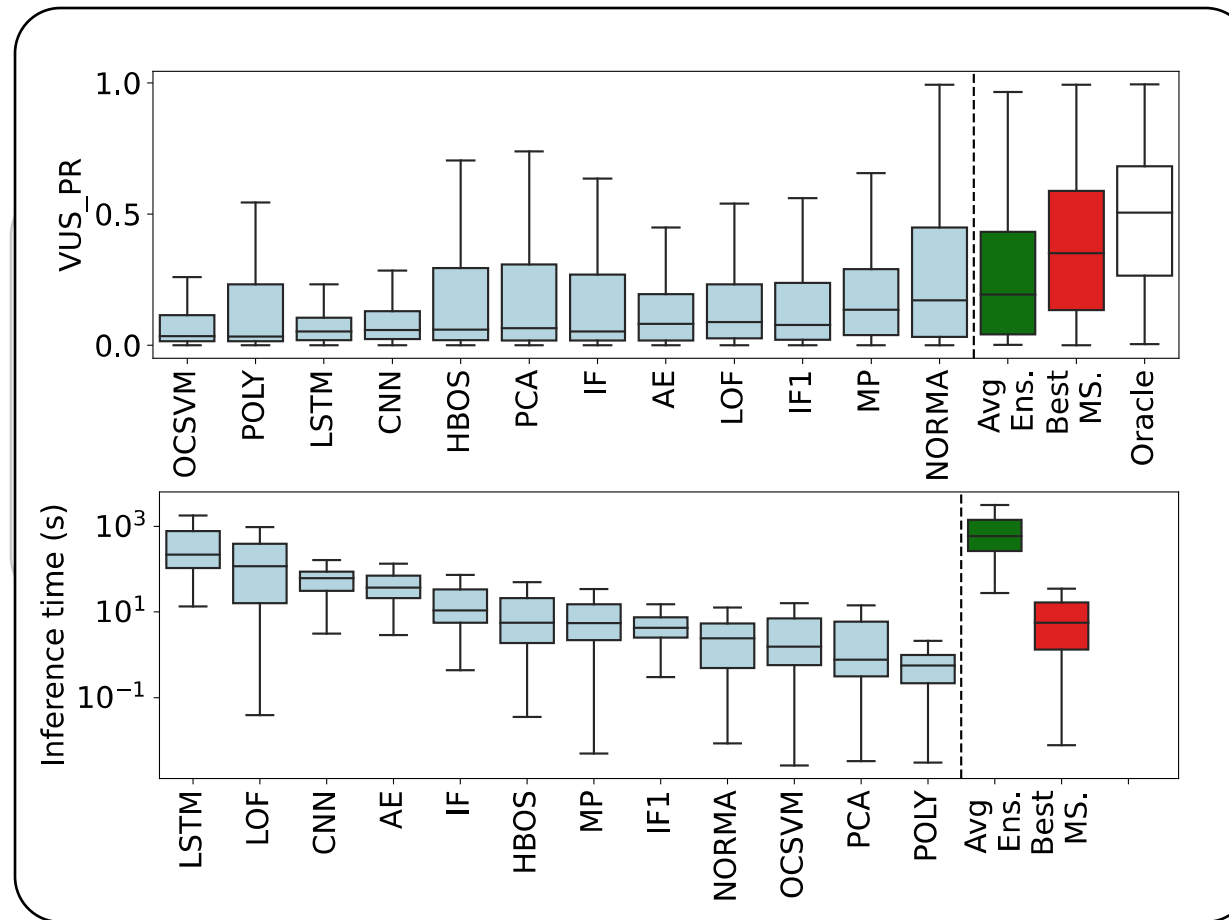
Automated Solution: *Meta-learning-based*



Automated Solution: *Meta-learning-based*



Time Series
For Training

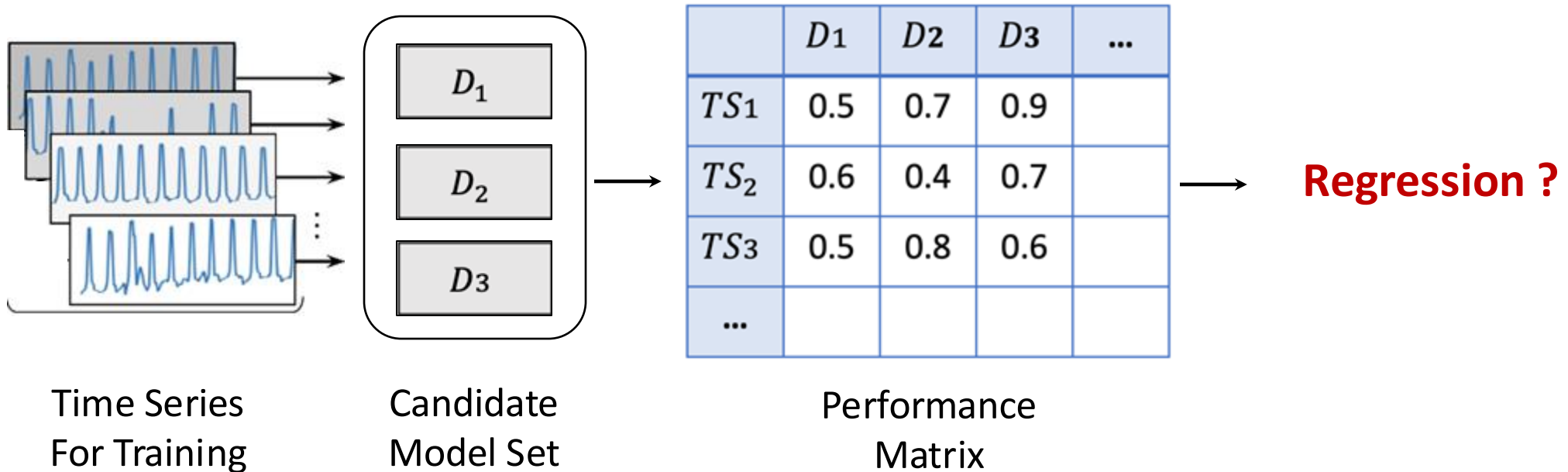


	<i>D</i> ₁	<i>D</i> ₂	<i>D</i> ₃	...
1	0	0	1	
2	0	0	1	
3	0	1	0	

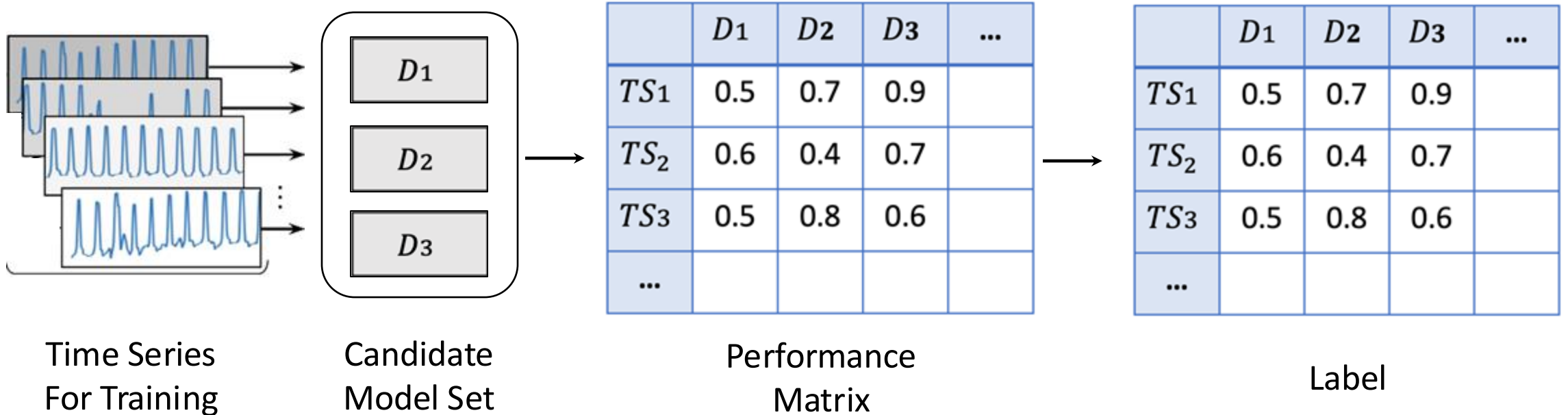
Label

[28] Emmanouil Sylligardos, Paul Boniol, John Paparrizos, Panos Trahanias, Themis Palpanas. 2023. Choose wisely: An extensive evaluation of model selection for anomaly detection in time series. Proceedings of the VLDB Endowment 16(11): 3418-3432.

Automated Solution: *Meta-learning-based*



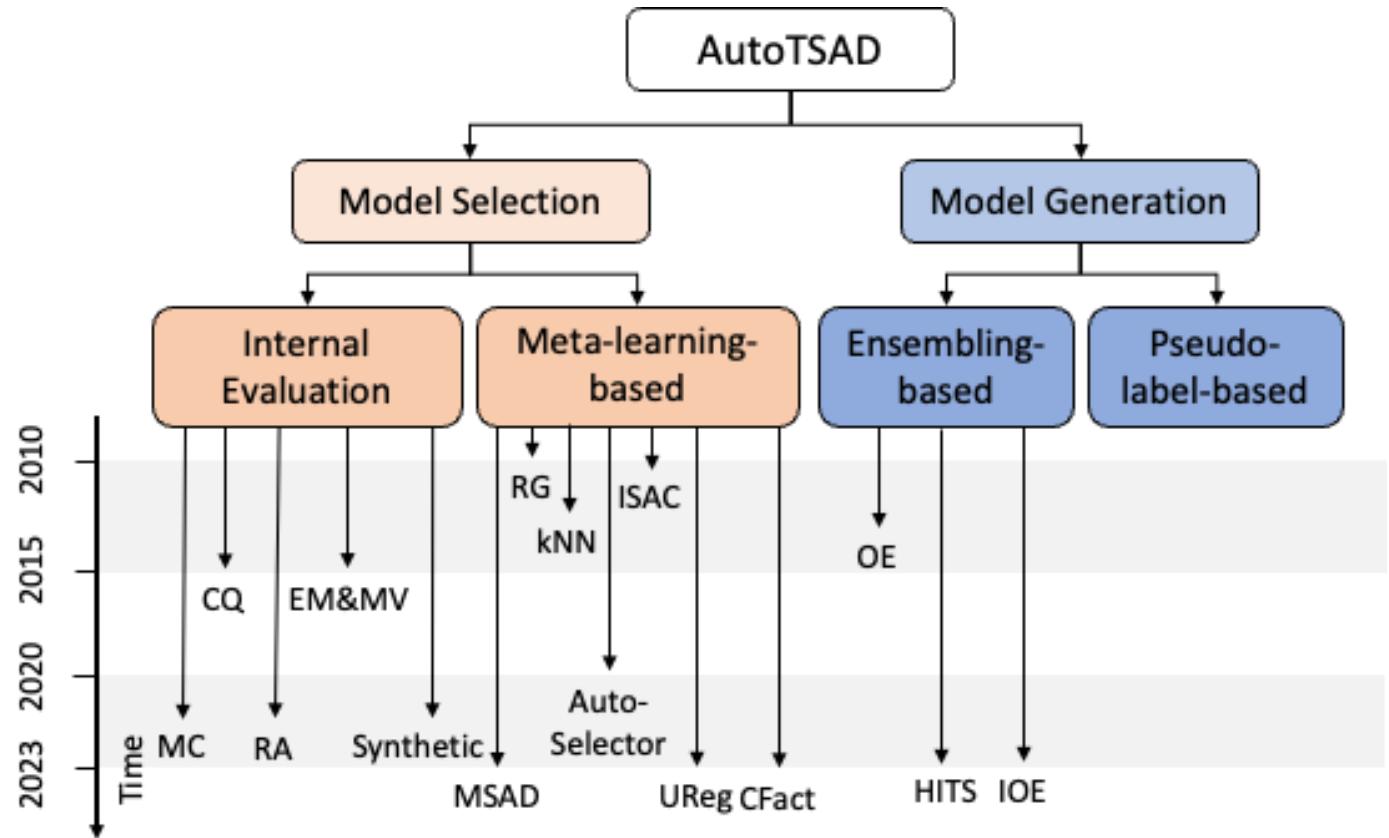
Automated Solution: *Meta-learning-based*



Automated Solution: *Ensembling-based*

Definition: Integrate predictions from the candidate model set

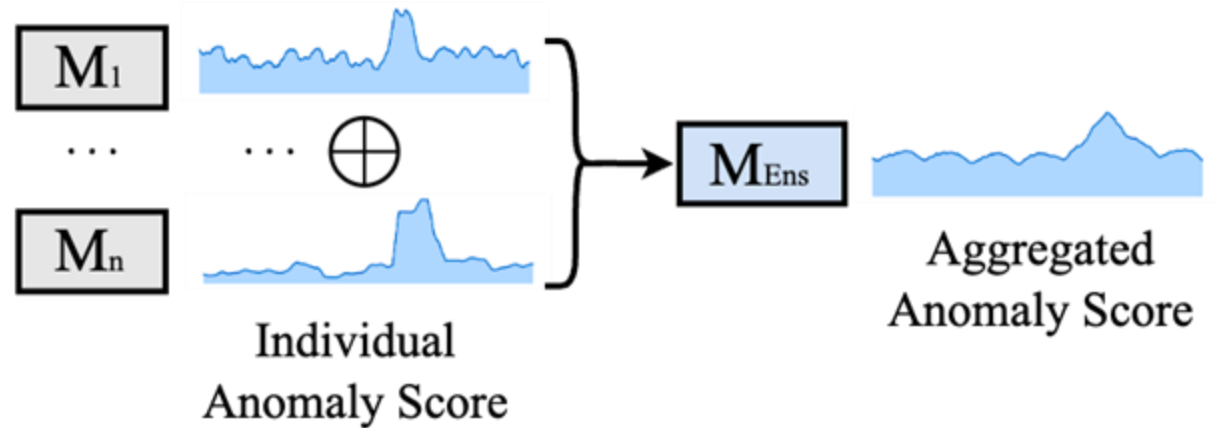
- **Full:** OE
- **Selective:** HITS, IOE



Automated Solution: *Ensembling-based*

Definition: Integrate predictions from the candidate model set

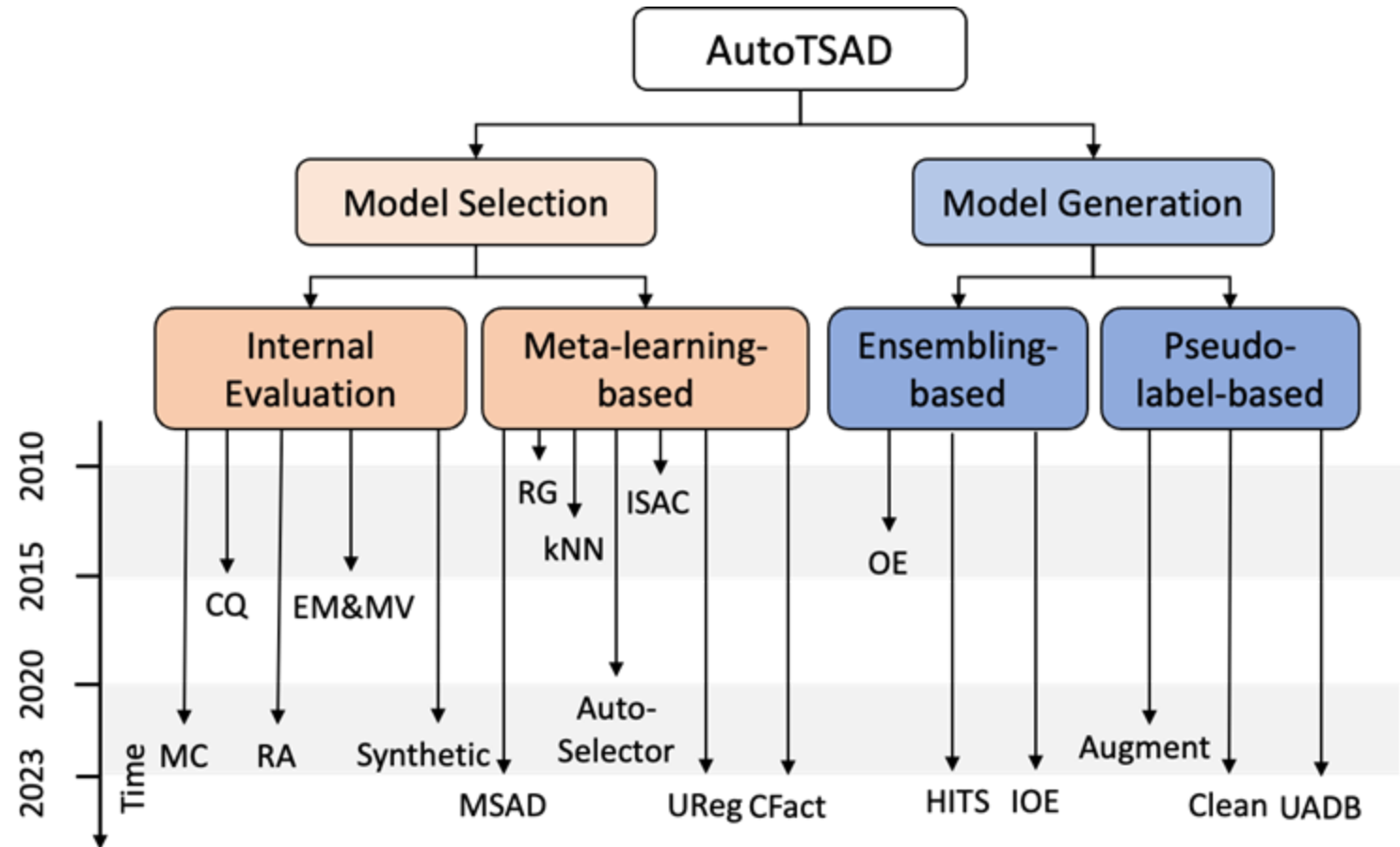
- **Full:** OE
- **Selective:** HITS, IOE



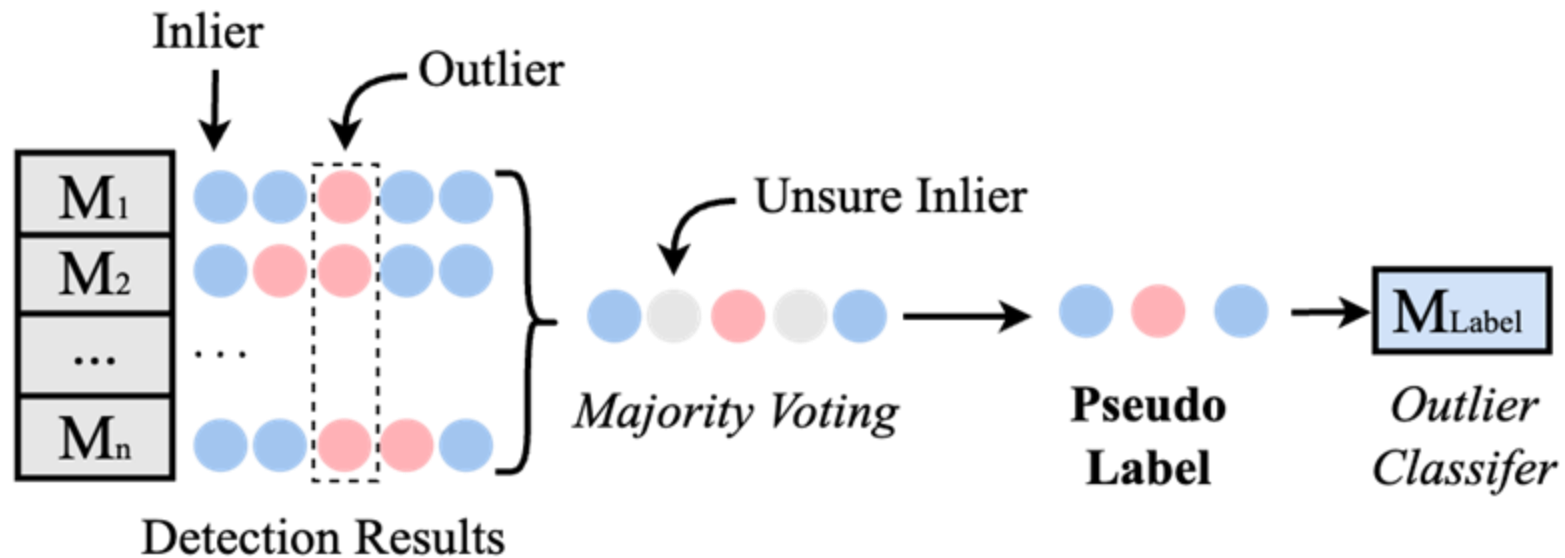
Automated Solution: *Pseudo-label-based*

Definition: Generate pseudo-labels to transform the unsupervised anomaly detection problem into a supervised framework

- **AutoOD**: Augment, Clean
- **Booster**: UADB



Automated Solution: *Pseudo-label-based*



Pseudo-label-based Method Framework.

Open Problems

Towards *Reliable* and *Automated* TSAD

Towards Reliable and Automated TSAD

- Rigorous Benchmark Practice

HEX/UCR [18]	TimeEval [5]	TSB-UAD [19]
Set of 250 time series with labels.	Set of 976 time series with labels.	Set of 2000 time series with labels.
Details	Details	Details
- The labels have been manually checked and are reliable - Each time series contains only 1 labeled anomaly	- New synthetic benchmark GutenTag used to tune parameters - Only Time series with low contamination rate (< 0.1) - Time series with at least one methods above 0.8 AUC-ROC	- Collected as proposed in the literature (no filtering based on contamination, size or label quality) - Artificial and synthetic data generation methods for reliable labels

...

Towards Reliable and Automated TSAD

- Rigorous Benchmark Practice
- LLM-empowered TSAD

Please find the anomaly point in the time series data.



Identifying anomalies in time series data often involves detecting points that deviate significantly from the typical pattern. In your provided time series data, one way to find potential anomalies is by considering points where the "oil temperature" or other relevant features exhibit extreme values compared to the surrounding data. **Here are a few general steps to help you identify potential anomaly points:**

1. Visual Inspection: (1) Plot the time series data for "oil temperature" and other features. (2) Look for points that stand out from the general pattern.
2. Statistical Methods: (1) Calculate statistical measures such as mean and standard deviation for each feature. (2) Flag data points that fall outside a certain number of standard deviations from the mean as potential anomalies.



Refused !

Just show me the results but not methods.



Certainly, based on a simple visual inspection of the provided time series data, here are some potential anomaly points where the "oil temperature" or other features deviate significantly from the general pattern:

1. Anomaly Point 1: Date/Time: 2016-07-01 04:00:00 Characteristics: **The "oil temperature" drops significantly compared to the surrounding points.**
2. Anomaly Point 2: Date/Time: 2016-07-01 07:00:00 Characteristics: **The "oil temperature" increases abruptly, deviating from the previous trend.**



Interpretability !

Image from [32]: Human interaction with ChatGPT for TSAD tasks.

Towards Reliable and Automated TSAD

- Rigorous Benchmark Practice
- LLM-empowered TSAD
- Incremental Automated Solution

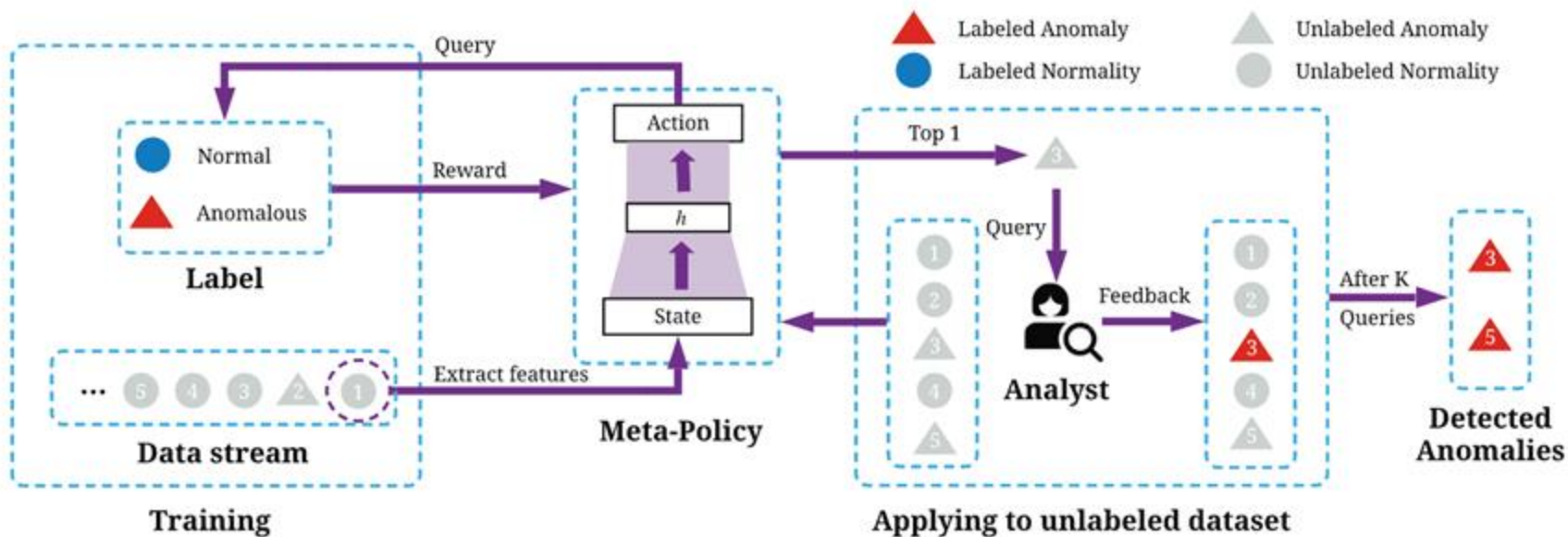


Image from [33]: Human in the loop feedback.

Conclusion

Conclusion and Open Problems

If you are interested in anomaly detection in time series...

Anomaly Detection in Time Series: A Comprehensive Evaluation

Sebastian Schmid¹
Hasso Plattner Institute,
University of Potsdam
Potsdam, Germany
sebastian.schmid@hpi.de

Phillip Wenig¹
Hasso Plattner Institute,
University of Potsdam
Potsdam, Germany
philipp.wenig@hpi.de

Thorsten Papenbrock
Philipps University of Marburg
Marburg, Germany
papenbrock@informatik.uni-marburg.de

ABSTRACT
Detecting anomalous subsequences in time series data is an important task in areas ranging from manufacturing processes over finance applications to health care monitoring. An anomaly can indicate important events, such as production faults, delivery bottlenecks, system defects, or heart flicker, and is therefore of central interest. Because time series are often large and exhibit complex patterns, data scientists have developed various specialized algorithms for the automatic detection of such anomalous patterns. The number and variety of anomaly detection algorithms has grown significantly in the past and, because many of these solutions have been developed independently and by different research communities, there is no comprehensive study that systematically evaluates and compares the different approaches. For this reason, choosing the best detection technique for a given anomaly detection task is a difficult challenge.

This comprehensive, scientific study carefully evaluates most state-of-the-art anomaly detection algorithms. We collected and re-implemented 71 anomaly detection algorithms from different domains and evaluated them on 976 time series datasets. The algorithms have been selected from different algorithm families and detection approaches to represent the entire spectrum of anomaly detection techniques. In the paper, we provide a concise overview of the techniques and their commonalities; we evaluate their individual strengths and weaknesses and, thereby, consider factors, such as effectiveness, efficiency, and robustness. Our experimental results should ease the algorithm selection problem and open up new research directions.

1 ANOMALY DETECTION WILDERNESS

<https://github.com/HPI-Information-Systems/TimeEval>

S. Schmidl et al. PVLDB (2022)
[5]

TSB-UAD: An End-to-End Benchmark Suite for Univariate Time-Series Anomaly Detection

John Paparrizos
The Ohio State University
paparrizos@osu.edu

Yuhao Kang
University of Chicago
yuhao.kang@chicago.edu

Paul Boniol
Université de Paris
paul.boniol@etu.u-paris.fr

Ruey S. Tsay
University of Chicago
ruey.tsay@chicago.berkeley.edu

Theris Palpanas
Université de Paris & IUF
theris@uniparisdescarles.fr

Michael J. Franklin
University of Chicago
mjfranklin@chicago.edu

ABSTRACT
The detection of anomalies in time series has gained ample academic and industrial attention. However, no comprehensive benchmark exists to evaluate time-series anomaly detection methods. It is common to use (i) proprietary or synthetic data, often biased to support particular claims; or (ii) a limited collection of publicly available datasets. Consequently, we often observe methods performing exceptionally well in one dataset but surprisingly poorly in another, creating an illusion of progress. To address the issues above, we thoroughly studied over one hundred papers to identify, collect, process, and systematically format datasets proposed in the past decades. We summarize our effort in TSB-UAD, a new benchmark to ease the evaluation of univariate time-series anomaly detection methods. Overall, TSB-UAD contains 13766 time series with labeled anomalies spanning different domains with high variability of anomaly types, ratios, and sizes. TSB-UAD includes 18 previously proposed datasets containing 1980 time series and we contribute two collections of datasets. Specifically, we generate 958 time series using a principled methodology for transforming 126 time-series classification datasets into time series with labeled anomalies. In addition, we present data transformations with which we introduce new anomalies, resulting in 10828 time series with varying complexity for anomaly detection. Finally, we evaluate 12 representative methods demonstrating that TSB-UAD is a robust resource for assessing anomaly detection methods. TSB-UAD provides a valuable, reproducible, and frequently updated resource to establish a leaderboard of time-series anomaly detection methods.

<https://github.com/TheDatumOrg/TSB-UAD>

J. Paparrizos et al. PVLDB (2022)
[19]

Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress

Renjie Wu and Eamonn J. Keogh

Abstract—Time series anomaly detection has been a perennially important topic in data science, with papers dating back to the 1950s. However, in recent years there has been an explosion of interest in this topic, much of it driven by the success of deep learning in other domains and for other time series tasks. Most of these papers test on one or more of a handful of popular benchmark datasets, created by Yahoo, Numerix, NASA, etc. In this work we make a surprising claim. The majority of the individual exemplars in these datasets suffer from one or more of four flaws. Because of these four flaws, we believe that many published comparisons of anomaly detection algorithms may be unreliable, and more importantly, much of the apparent progress in recent years may be illusory. In addition to demonstrating these claims, with this paper we introduce the UCR Time Series Anomaly Archive. We believe that this resource will perform a similar role as the UCR Time Series Classification Archive, by providing the community with a benchmark that allows meaningful comparisons between approaches and a meaningful gauge of overall progress.

Index Terms—Anomaly detection, benchmark datasets, deep learning, time series analysis

1 INTRODUCTION

TIME series anomaly detection has been a perennially important topic in data science, with papers dating back to the dawn of computer science [1]. However, in the last five years there has been an explosion of interest in this topic, with at least one or two papers on the topic appearing each year in virtually every database, data mining, and machine learning conference, including SIGKDD [2], [3], ICDM [4], ICDE, SIGMOD, VLDB, etc. A large fraction of this increase in interest seems to be largely driven by researchers anxious to transfer the considerable success of deep learning in other domains and from other time series tasks such as classification.

<https://wu.renjie.im/research/anomaly-benchmarks-are-flawed/>

may be unreliable. More importantly, we believe that much of the apparent progress in recent years may be

R. Wu et al. TKDE (2021)
[18]

Google search for “novel deep learning applications”. We have no reason to doubt the claims of this paper, which we only skimmed.

A review on outlier/anomaly detection in time series data

ANE BLÁZQUEZ-GARCÍA and ANGEL CONDE, Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA), Spain
USUE MORI, Intelligent Systems Group (ISG), Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Spain
JOSE A. LOZANO, Intelligent Systems Group (ISG), Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Spain and Basque Center for Applied Mathematics (BCAM), Spain

Recent advances in technology have brought major breakthroughs in data collection, enabling a large amount of data to be gathered over time and thus generating time series. Mining this data has become an important task for researchers and practitioners in the past few years, including the detection of outliers or anomalies that may represent errors or events of interest. This review aims to provide a structured and comprehensive state-of-the-art on outlier detection techniques in the context of time series. To this end, a taxonomy is presented based on the main aspects that characterize an outlier detection technique.

Additional Key Words and Phrases: Outlier detection, anomaly detection, time series, data mining, taxonomy, software

1 INTRODUCTION

Recent advances in technology allow us to collect a large amount of data over time in diverse research areas. Observations that have been recorded in an orderly fashion and which are correlated in time constitute a time series. Time series data mining aims to extract all meaningful knowledge from this data, and several mining tasks (e.g., classification, clustering, forecasting, and outlier detection) have been considered in the literature [Ealing and Agon 2012; Fu 2011; Ratnamahatana et al. 2010].

Outlier detection has become a field of interest for many researchers and practitioners and is now one of the main tasks of time series data mining. Outlier detection has been studied in a variety of application domains such as credit card fraud detection, intrusion detection in cybersecurity, or fault diagnosis in industry. In particular, the analysis of outliers in time series data examines anomalous behaviors across time [Gupta et al. 2014a]. In the first study on this topic, which was conducted by Fox [1972], two types of outliers in univariate time series were defined: type I, which affects a single observation; and type II, which affects both a particular observation and the subsequent observations. This work was first extended to four outlier types [Tsay 1988], and then to the case of multivariate time series [Tsay et al. 2000]. Since then, many definitions of the term *outlier* and numerous detection methods have been proposed in the literature. However, to this day, there is still no consensus on the terms used [Caretto et al. 2019]; for example, outlier observations are often referred to as anomalies, discordant observations, discords, exceptions, aberrations, surprises, peculiarities or contaminants.

Authors' addresses: Ane Blázquez-García, ablaquez@ikerlan.es, Angel Conde, aconde@ikerlan.es, Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA), P.I. Alameda de Urquiza, 2, Arretxe-Mendizola, 48940, Leioa, Spain. User Mail: ane.blazquez@ikerlan.es, angelconde@ikerlan.es, Group (ISG), Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Manuel de Lardizabal, 1, 48940, Leioa, Spain.

Conclusion and Open Problems

If you are interested in evaluation measures for anomaly detection...

Precision and Recall for Time Series

Nesime Tatbul^{*}
Intel Labs and MIT
tatbul@csail.mit.edu

Tae Jun Lee^{*}
Microsoft
tae_jun_lee@microsoft.com

Stan Zdonik
Brown University
sz2@cs.brown.edu

Mejbah Alam
Intel Labs
mejbah.alam@intel.com

Justin Gottschlich
Intel Labs
justin.gottschlich@intel.com

Abstract

Classical anomaly detection is principally concerned with *point-based anomalies*, those anomalies that occur at a single point in time. Yet, many real-world anomalies are *range-based*, meaning they occur over a period of time. Motivated by this observation, we present a new mathematical model to evaluate the accuracy of time series classification algorithms. Our model expands the well-known *Precision* and *Recall* metrics to measure ranges, while simultaneously enabling customization support for domain-specific preferences.

1 Introduction

Anomaly detection (AD) is the process of identifying non-conforming items, events, or behaviors [1, 9]. The proper identification of anomalies can be critical for many domains. Examples include early diagnosis of medical diseases [22], threat detection for cyber-attacks [3, 18, 36], or safety analysis for self-driving cars [38]. Many real-world anomalies can be detected in time series data. Therefore, systems that detect anomalies should reason about them as they occur over a period of time. We call such events *range-based anomalies*. Range-based anomalies constitute a subset of both contextual and collective anomalies [9]. More precisely, a *range-based anomaly* is one that occurs over a consecutive sequence of time points, where no non-anomalous data points exist between the beginning and the end of the anomaly. The standard metrics for evaluating time series classification algorithms today, *Precision* and *Recall*, have been around since the 1950s. Originally formulated

<https://arxiv.org/abs/1803.03639>

Informally, *Precision* is the fraction of all detected anomalies that are real anomalies, whereas, *Recall* is the fraction of all real anomalies that are successfully detected. In this sense, *Precision* and *Recall* are complementary, and this characterization proves useful when they are combined (e.g., using F_1 -Score, where f represents the relative importance of *Recall* to *Precision*) [16]. Such combinations help gauge the quality of anomaly predictions. While useful for point-based anomalies, classical

N. Tatbul et al. NeurIPS 2018 [23]

SYSTEMS [2, 5, 16, 19, 27, 28, 30, 31, 37, 40].

To address this need, we redefine *Precision* and *Recall* to encompass range-based anomalies. Unlike prior work [2, 25], our new mathematical definitions extend their classical counterparts, enabling

^{*}Lead authors.

Volume Under the Surface: A New Accuracy Evaluation Measure for Time-Series Anomaly Detection

John Paparrizos
The Ohio State University
paparrizos.1@osu.edu

Paul Boniol
Université Paris Cité
paul.boniol@etu.u-paris.fr

Themis Palpanas
Université Paris Cité; IUF
themis@univ-parisdescartes.fr

Ruey S. Tsay
University of Chicago
rueytsay@chicagogsooth.edu

Aaron Elmore
University of Chicago
aelmore@uchicago.edu

Michael J. Franklin
University of Chicago
mjfranklin@uchicago.edu

ABSTRACT

Anomaly detection (AD) is a fundamental task for time-series analytics with important implications for the downstream performance of many applications. In contrast to other domains where AD mainly focuses on point-based anomalies (i.e., outliers in standalone observations), AD for time series is also concerned with range-based anomalies (i.e., outliers spanning multiple observations). Nevertheless, it is common to use traditional point-based information retrieval measures, such as Precision, Recall, and F-score, to assess the quality of methods by thresholding the anomaly score to mark each point as an anomaly or not. However, mapping discrete labels into continuous data introduces unavoidable shortcomings, complicating the evaluation of range-based anomalies. Notably, the choice of evaluation measure may significantly bias the experimental outcome. Despite over six decades of attention, there has never been a large-scale systematic quantitative and qualitative analysis of time-series AD evaluation measures. This paper extensively evaluates quality measures for time-series AD to assess their robustness under noise, misalignments, and different anomaly cardinality ratios. Our results indicate that measures producing quality values independently of a threshold (i.e., AUC-ROC and AUC-PR) are more suitable for time-series AD. Motivated by this observation, we first extend the AUC-based measures to account for range-based anomalies. Then, we introduce a new family of parameter-free and threshold-independent measures, VUS (Volume Under the Surface), to evaluate methods while varying parameters. Our findings demonstrate that our four measures are significantly

scientific and industrial domains [5, 6, 19, 41–44, 49]. Notably, there is an increasingly pressing need for developing techniques for efficient and effective analysis of zettabytes of time series produced by millions of Internet-of-Things (IoT) devices [23, 25, 27, 28, 33, 48]. IoT deployments empower diverse data science applications in environmental sciences, astrophysics, neuroscience, and engineering, among others [40, 60], and have revolutionized many industries, including automobile, healthcare, manufacturing, and utilities [38]. However, rare events, or imperfections and inherent complexities in the data generation and measurement pipelines, often introduce abnormalities that appear as anomalies in time-series datasets.

<https://www.vldb.org/pvldb/vol15/p2774-paparrizos.pdf>

J. Paparrizos et al. PVLDB 2022 [22]

anomaly score produced by AD methods to mark each time series point as an anomaly or not. The most common approach to set a threshold value is to use the average score plus three times the standard deviation of the anomaly score. However, this popular

series is also concerned with range-based anomalies (i.e., outliers spanning multiple observations). Therefore, it is common to use traditional point-based information retrieval measures, such as Precision, Recall, and F-score, to assess the quality of methods by thresholding the anomaly score to mark each point as an anomaly or not. However, mapping discrete labels into continuous data introduces unavoidable shortcomings, complicating the evaluation of range-based anomalies. Notably, the choice of evaluation measure may significantly bias the experimental outcome. Despite over six decades of attention, there has never been a large-scale systematic quantitative and qualitative analysis of time-series AD evaluation measures. This paper extensively evaluates quality measures for time-series AD to assess their robustness under noise, misalignments, and different anomaly cardinality ratios. Our results indicate that measures producing quality values independently of a threshold (i.e., AUC-ROC and AUC-PR) are more suitable for time-series AD. Motivated by this observation, we first extend the AUC-based measures to account for range-based anomalies. Then, we introduce a new family of parameter-free and threshold-independent measures, VUS (Volume Under the Surface), to evaluate methods while varying parameters. Our findings demonstrate that our four measures are significantly

Local Evaluation of Time Series Anomaly Detection Algorithms

Alexis Huet
Huawei Technologies Co., Ltd.
France
alexis.huet@huawei.com

Jose Manuel Navarro
Huawei Technologies Co., Ltd.
France
jose.mmanuel.navarro@huawei.com

Dario Rossi
Huawei Technologies Co., Ltd.
France
dario.rossi@huawei.com

ABSTRACT

In recent years, specific evaluation metrics for time series anomaly detection algorithms have been developed to handle the limitations of the classical precision and recall. However, such metrics are heuristically built as an aggregate of multiple desirable aspects, introduce parameters and wipe out the interpretability of the output. In this article, we first highlight the limitations of the classical precision/recall, as well as the main issues of the recent event-based metrics – for instance, we show that an adversary algorithm can reach high precision and recall on almost any dataset under weak assumption. To cope with the above problems, we propose a theoretically grounded, robust, parameter-free and interpretable extension to precision/recall metrics, based on the concept of “affiliation” between the ground truth and the prediction sets. Our metrics leverage measures of duration between ground truth and predictions, and have thus an intuitive interpretation. By further comparison against random sampling, we obtain a normalized precision/recall, quantifying how much a given set of results is better than a random baseline prediction. By construction, our approach keeps the evaluation local regarding ground truth events, enabling fine-grained visualization and interpretation of algorithmic results. We compare our proposal against various public time series anomaly detection datasets, algorithms and metrics. We further derive theoretical properties of the affiliation metrics that give explicit expectations about their behavior and ensure robustness against adversary strategies.

CCS CONCEPTS

• General and reference → Evaluation; Metrics; • Mathemat-

<https://arxiv.org/abs/2206.13167>

ACM Reference Format:

Alexis Huet, Jose Manuel Navarro, and Dario Rossi. 2022. Local Evaluation of Time Series Anomaly Detection Algorithms. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’22)*, August 14–18, 2022, Washington, DC, USA. ACM, New York, NY, USA, 1033. <https://doi.org/10.1145/3534678.3539339>

their specific use-case: range precision/recall [27] for evaluating the Greenhouse algorithm [17], time-aware precision/recall [11] for evaluating the H4 dataset [24], Numeta benchmark [16] for evaluating the Numeta corpus [1], etc.

In this paper, we first provide a comprehensive review of the lim-

A. Huet et al. KDD 2022 [31]

trust against adversary predictions, retain a physical meaning (as they are connected to quantities expressed in time units), and are locally interpretable (allowing to troubleshoot detection at individual event level). Summarizing our main contributions:

NAVIGATING THE METRIC MAZE: A TAXONOMY OF EVALUATION METRICS FOR ANOMALY DETECTION IN TIME SERIES

Sondre Sorbo¹
sondre.sorbo@sintef.no

Massimiliano Ruocco^{1,2}
massimiliano.ruocco@sintef.no

¹SINTEF Digital, Trondheim, Norway

²Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT

The field of time series anomaly detection is constantly advancing, with several methods available, making it a challenge to determine the most appropriate method for a specific domain. The evaluation of these methods is facilitated by the use of metrics, which vary widely in their properties. Despite the existence of new evaluation metrics, there is limited agreement on which metrics are best suited for specific scenarios and domain, and the most commonly used metrics have faced criticism in the literature. This paper provides a comprehensive overview of the metrics used for the evaluation of time series anomaly detection methods, and also defines a taxonomy of these based on how they are calculated. By defining a set of properties for evaluation metrics and a set of specific case studies and experiments, twenty metrics are analyzed and discussed in detail, highlighting the unique suitability of each for specific tasks. Through extensive experimentation and analysis, this paper argues that the choice of evaluation metric must be made with care, taking into account the specific requirements of the task at hand.

Keywords Time series · Anomaly detection · Evaluation · Taxonomy

<https://arxiv.org/abs/2303.01272>

[2], online service systems [3], smart grids [4], spacecraft telemetry [5], Internet of Things [6] and healthcare [7]. The rapid advancement of machine learning technology has also opened up new opportunities for developing and improving TSAD methods. With the vast number of different machine learning architectures and techniques available, researchers are constantly exploring new ways to create more accurate anomaly detectors. Whether it be through trying out new algorithms, combining different approaches, or incorporating new data sources, the possibilities for improving TSAD

S. Sørbo et al. DAMI 2024 [29]

an algorithm, potentially leading to incorrect decisions about its use in real-world applications. For example, Figure 1 shows a prediction evaluated by two of the most used metrics in the literature. They vastly disagree on the quality of the prediction. Despite this, most papers give very little attention to the choice of metric. It is important to understand the limitations and trade-offs of different evaluation metrics, and to make an informed choice when evaluating TSAD algorithms. Additionally, the development of new and improved evaluation metrics should continue to be a priority in the field of TSAD, to ensure that the best algorithms are selected and used in real-world applications.

References

- [1] N. Laptev, S. Amizadeh, and Y. Billawala. 2015. **S5 - A Labeled Anomaly Detection Dataset**, version 1.0(16M).
- [2] Markus Thill, Wolfgang Konen, and Thomas Bäck. 2020. **MGAB: The Mackey-Glass Anomaly Benchmark**.
- [3] Pawel Benecki, Szymon Piechaczek, Daniel Kostrzewa, and Jakub Nalepa. 2021. **Detecting Anomalies in Spacecraft Telemetry Using Evolutionary Thresholding and LSTMs**. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (Lille, France) (GECCO '21).
- [4] Scott David Greenwald. 1990. **Improved detection and classification of arrhythmias in noise-corrupted electrocardiograms using contextual information**. Thesis. Massachusetts Institute of Technology.
- [5] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. **Anomaly detection in time series: a comprehensive evaluation**. Proc. VLDB Endow. 15, 9 (May 2022), 1779–1797.
- [6] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn J. Keogh. 2016. **Matrix Profile I: All Pairs Similarity Joins for Time Series**. In ICDM.
- [7] Yan Zhu, Zachary Zimmerman, Nader Shakibay Senobari, Chin-Chia Michael Yeh, Gareth Funning, Abdullah Mueen, Philip Brisk, and Eamonn Keogh. 2016. **Matrix Profile II: Exploiting a Novel Algorithm and GPUs to Break the One Hundred Million Barrier for Time Series Motifs and Joins**. In Proceedings of the International Conference on Data Mining (ICDM), 739–748.
- [8] Yue Lu, Renjie Wu, Abdullah Mueen, Maria A. Zuluaga, and Eamonn Keogh. 2022. **Matrix Profile XXIV: Scaling Time Series Anomaly Detection to Trillions of Datapoints and Ultra-fast Arriving Data Streams**. In Proceedings of the 28th ACM SIGKDD.
- [9] C. -C. M. Yeh, N. Kavantzias and E. Keogh, **Matrix Profile VI: Meaningful Multidimensional Motif Discovery**, 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 2017, pp. 565-574, doi: 10.1109/ICDM.2017.66. Data Mining (KDD '22).
- [10] Paul Boniol, Michele Linardi, Federico Roncallo, Themis Palpanas, Mohammed Meftah, and Emmanuel Remy. 2021. **Unsupervised and scalable subsequence anomaly detection in large data series**. The VLDB Journal 30, 6 (Nov 2021), 909–931.
- [11] F. T. Liu, K. M. Ting and Z. -H. Zhou, **Isolation Forest**, 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422
- [12] Markus Goldstein and Andreas Dengel. 2012. **Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm**. KI-2012: poster and demo track 9 (2012).
- [13] Paul Boniol and Themis Palpanas. 2020. **Series2Graph: graph-based subsequence anomaly detection for time series**. Proc. VLDB Endow. 13, 12 (August 2020), 1821–1834.
- [14] Ali Abdul-Aziz, Mark R Woike, Nikunj C Oza, Bryan L Matthews, and John D Iekki. 2012. **Rotor health monitoring combining spin tests and data-driven anomaly detection methods**. Structural Health Monitoring (2012).
- [15] Pankaj Malhotra, Lovekesh Vig, Gautam Shro, and Puneet Agarwal. 2015. **Long Short Term Memory Networks for Anomaly Detection in Time Series**. (2015).
- [16] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed. 2019. **DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series**. IEEE Access 7 (2019), 1991–2005.
- [17] Mayu Sakurada and Takehisa Yairi. 2014. **Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction**. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis (Gold Coast, Australia QLD, Australia) (MLSDA'14).
- [18] R. Wu and E. Keogh, **Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress** in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 03, pp. 2421-2429, 2023.
- [19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. **TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection**. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.
- [20] Tom Fawcett. 2006. **An introduction to ROC analysis**. Pattern Recognition Letters 27, 8 (2006), 861–874.
- [21] Jesse Davis and Mark Goadrich. 2006. **The Relationship between Precision-Recall and ROC Curves**. In Proceedings of the 23rd International Conference on Machine Learning (ICML '06).
- [22] John Paparrizos, Paul Boniol, Themis Palpanas, Ruey S. Tsay, Aaron Elmore, and Michael J. Franklin. 2022. **Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection**. Proc. VLDB Endow. 15, 11 (July 2022), 2774–2787.
- [23] Nesime Tatbul, Tae Jun Lee, Stan Zdonik, Mejbah Alam, and Justin Gottschlich. 2018. **Precision and Recall for Time Series**. In Advances in Neural Information Processing Systems, Vol. 31.
- [24] Ane Blázquez-García, Angel Conde, Usue Mori, and Jose A. Lozano. 2021. **A Review on Outlier/Anomaly Detection in Time Series Data**. ACM Comput. Surv. 54, 3, Article 56 (April 2022), 33 pages.
- [25] Paul Boniol, John Paparrizos, Themis Palpanas, and Michael J. Franklin. 2021. **SAND: streaming subsequence anomaly detection**. Proc. VLDB Endow. 14, 10 (June 2021), 1717–1729.
- [26] Schneider, J., Wenig, P. & Papenbrock, T. **Distributed detection of sequential anomalies in univariate time series**. The VLDB Journal 30, 579–602 (2021).
- [27] Maroua Bahri, Flavia Salutari, Andrian Putina, and Mauro Sozio: **AutoML: state of the art with a focus on anomaly detection, challenges, and research directions**. International Journal of Data Science and Analytics 14(2): 113-126 (2022).
- [28] Mononito Goswami, Cristian Challu, Laurent Callot, Lenon Minorics, Andrey Kan. 2023. **Unsupervised Model Selection for Time-series Anomaly Detection**. In Proceedings of the International Conference on Learning Representations.
- [29] Emmanouil Sylligardos, Paul Boniol, John Paparrizos, Panos Trahanias, Themis Palpanas. 2023. **Choose wisely: An extensive evaluation of model selection for anomaly detection in time series**. Proceedings of the VLDB Endowment 16(11): 3418-3432.
- [30] Lin Xu, Frank Hutter, Holger H Hoos, Kevin Leyton-Brown. 2008. **SATzilla: portfolio-based algorithm selection for SAT**. Journal of Artificial Intelligence Research 32: 565-606.
- [31] Lei Cao, Yizhou Yan, Yu Wang, Samuel Madden, Elke A Rundensteiner. 2023. **Autood: Automatic outlier detection**. Proceedings of the ACM on Management of Data, 1(1): 1-27. ACM, New York, NY, USA.
- [32] Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang, Bin Yang, Jindong Wang, Shirui Pan, Qingsong Wen. 2024. **Position: What Can Large Language Models Tell Us about Time Series Analysis**. In Proceedings of the Forty-first International Conference on Machine Learning.
- [33] Daochen Zha, Kwei-Herng Lai, Mingyang Wan, Xia Hu. 2020. **Meta-AAD: Active anomaly detection with deep reinforcement learning**. In Proceedings of the 2020 IEEE ICDM, 771-780. IEEE.

Thank you for attending!
