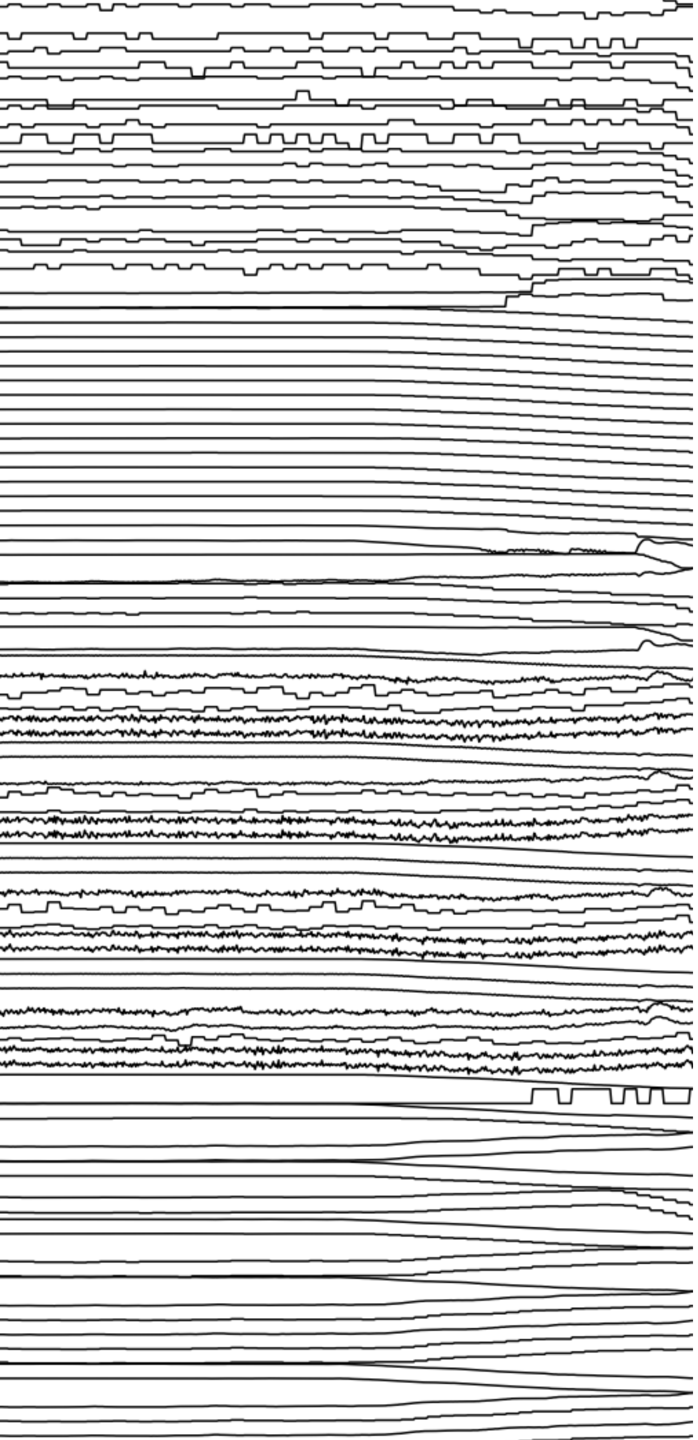


Advances in Time-Series Anomaly Detection:

Algorithms, Benchmarks, and Evaluation Measures

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

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

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

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

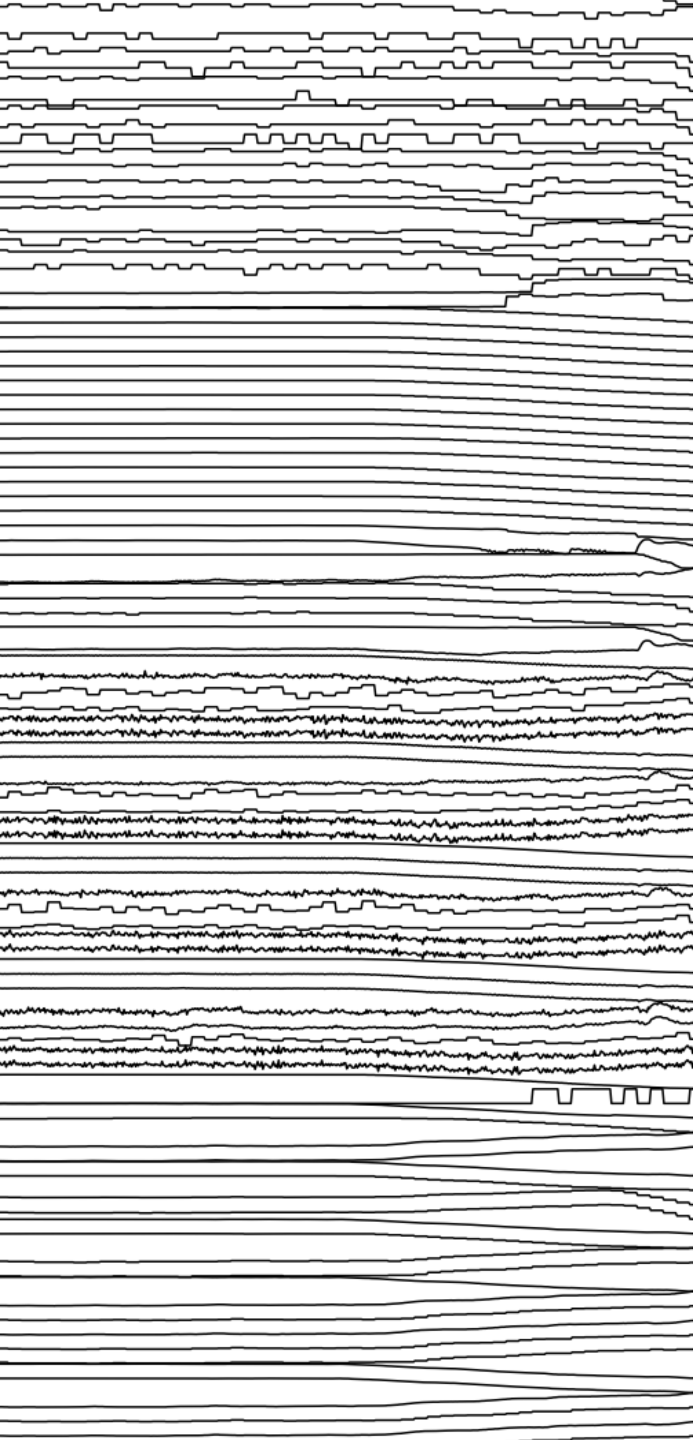
THE OHIO STATE UNIVERSITY

Inria

ENS
ÉCOLE NORMALE SUPÉRIEURE

Université Paris Cité

💡 Table of Contents



Part 1: Introduction, Motivation and Foundations

Introduction: *Time series are Everywhere*

Energy Production

Edf.fr: tinyurl.com/yc7x5xje

Astrophysics

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

Medicine

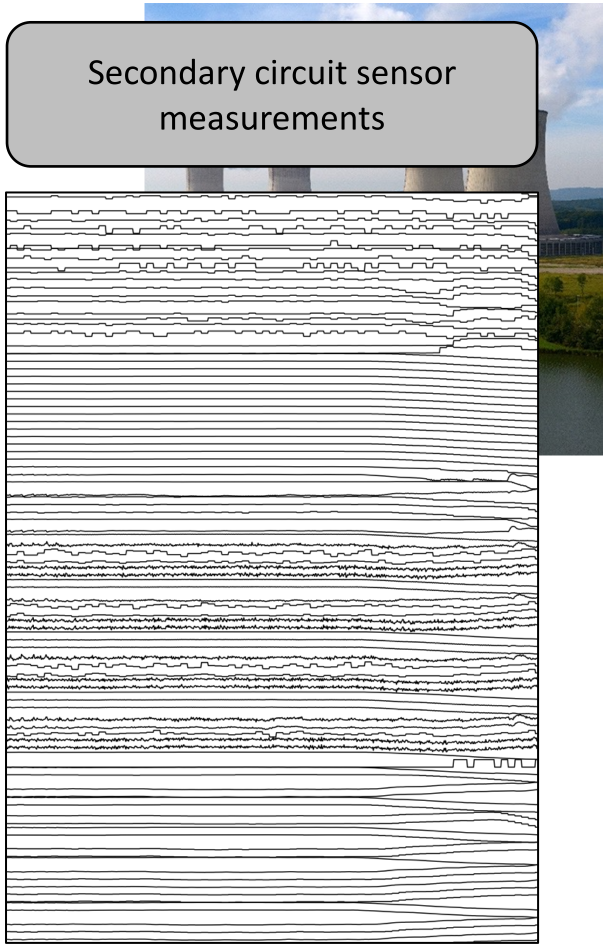
tinyurl.com/39dx2us4

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

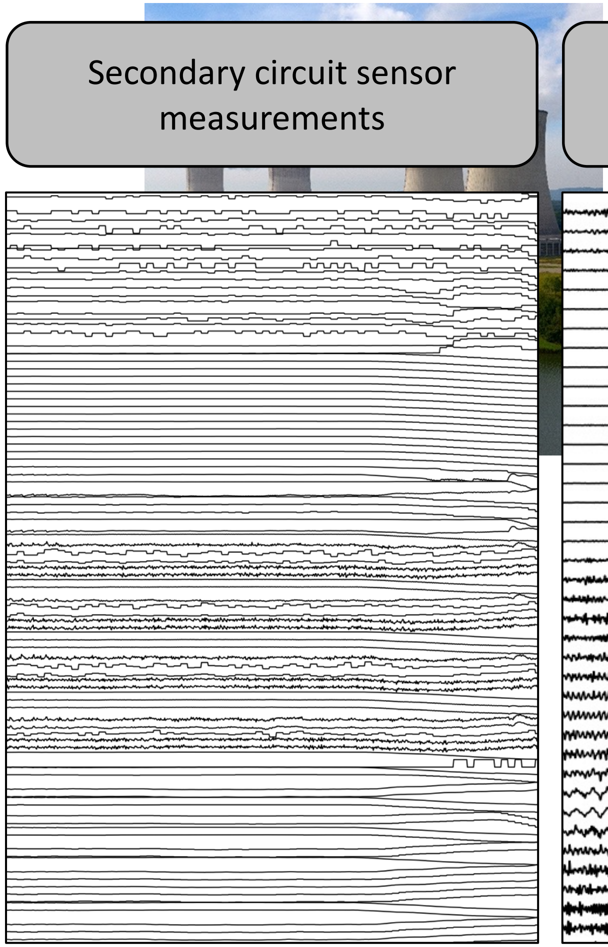
Volcanology

tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

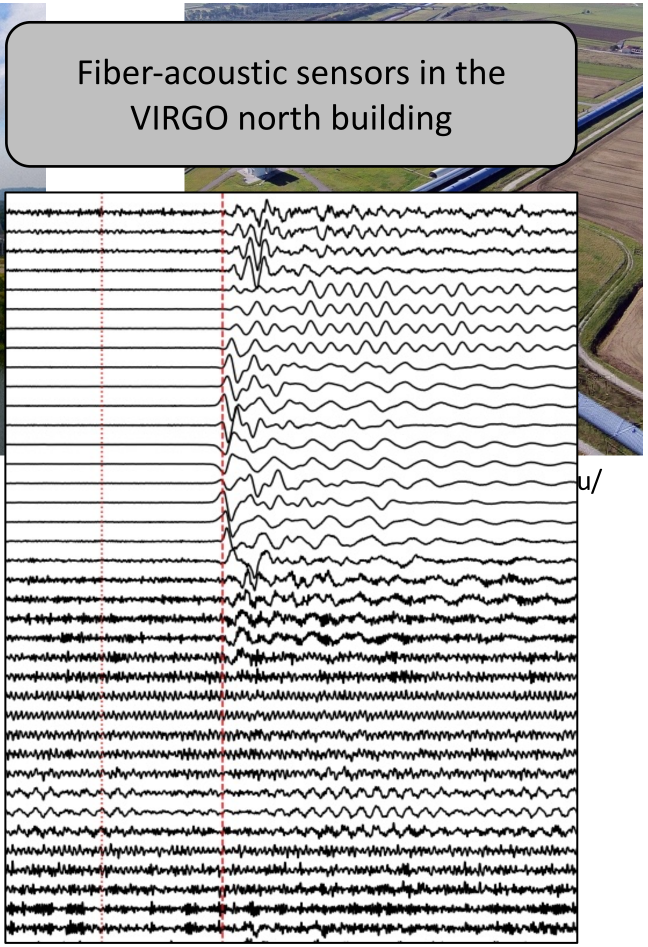
Energy Production

Secondary circuit sensor measurements



Astrophysics

Fiber-acoustic sensors in the VIRGO north building



u/

Medicine

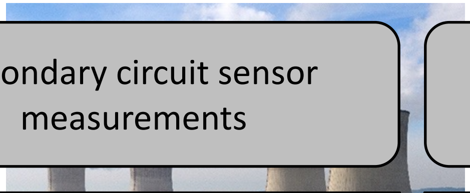
tinyurl.com/39dx2us4

Volcanology

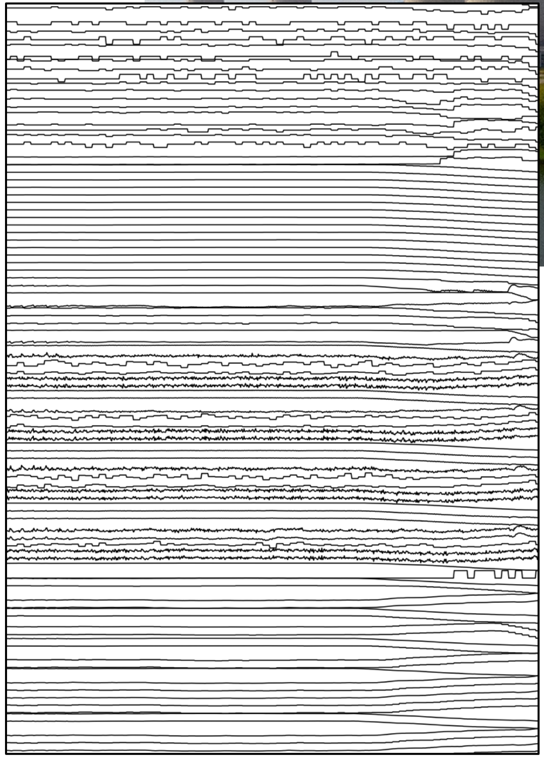
tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

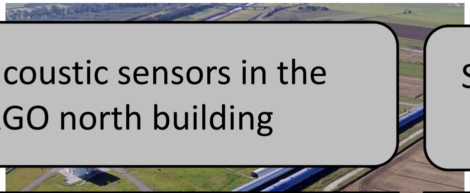
Energy Production



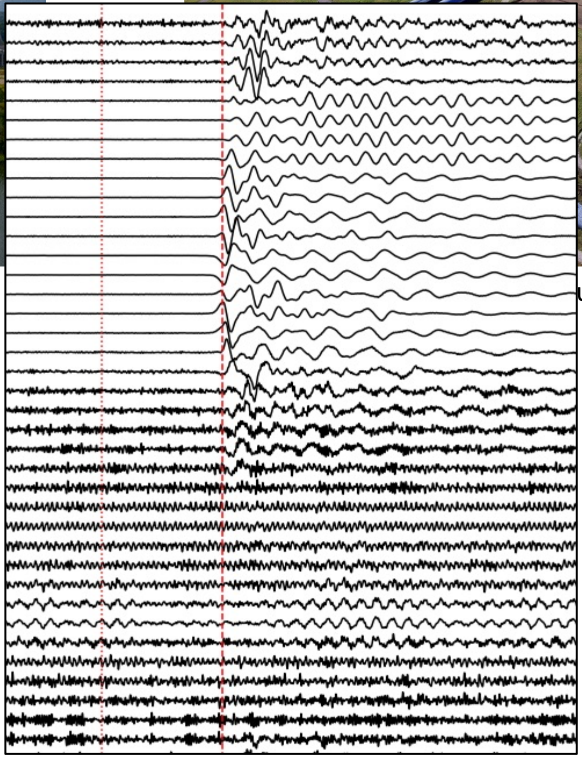
Secondary circuit sensor measurements



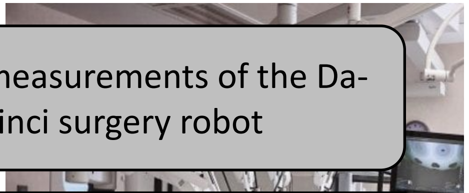
Astrophysics



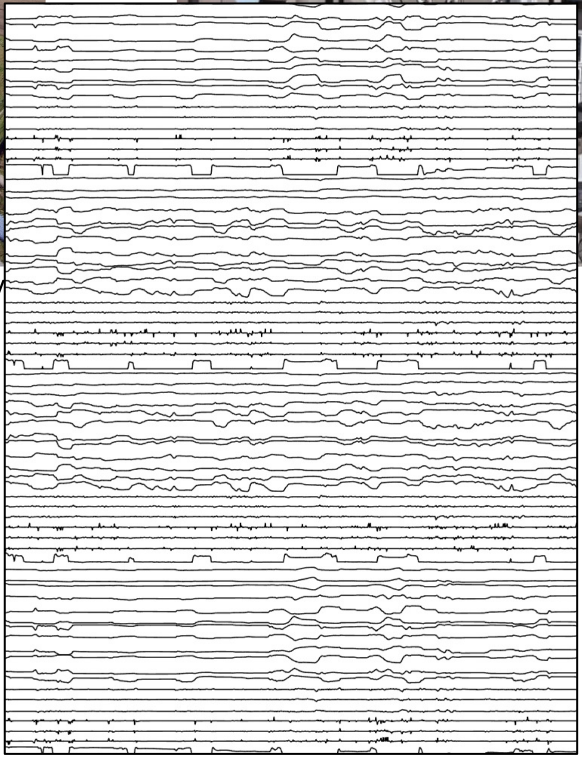
Fiber-acoustic sensors in the VIRGO north building



Medicine



Sensor measurements of the Da-Vinci surgery robot



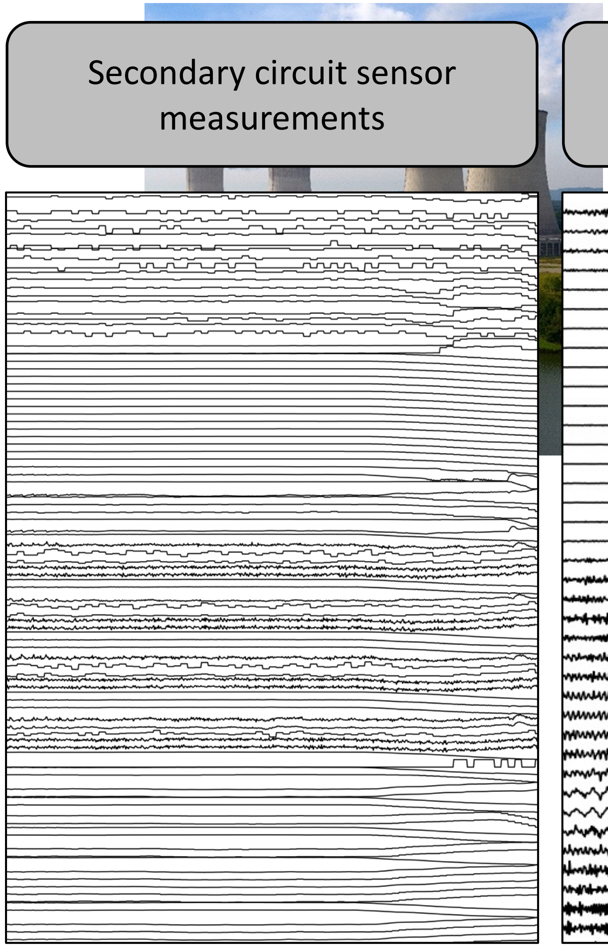
Volcanology

tinyurl.com/ybcttmfz

Introduction: *Time series are Everywhere*

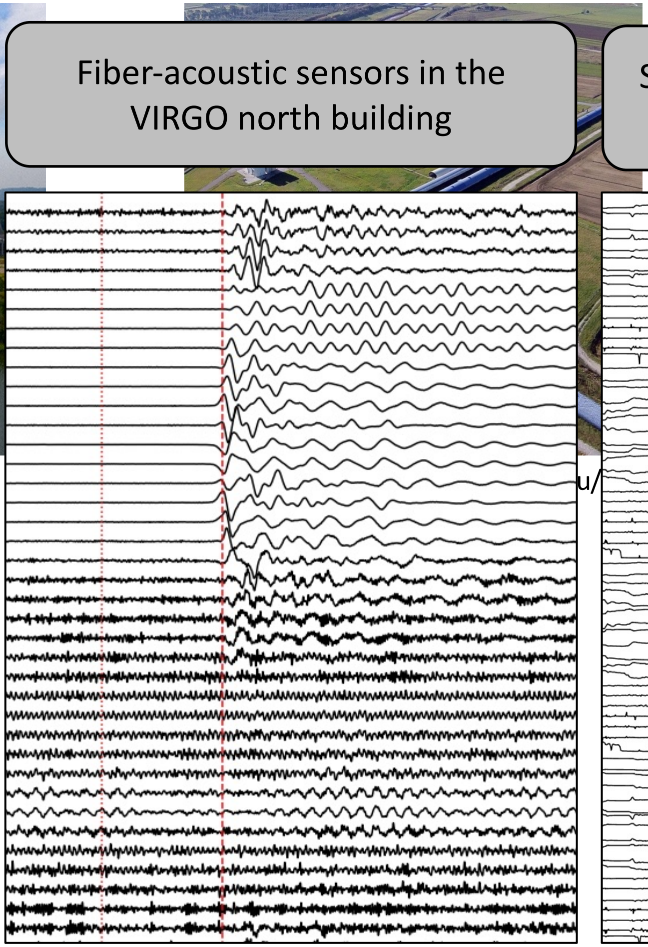
Energy Production

Secondary circuit sensor measurements



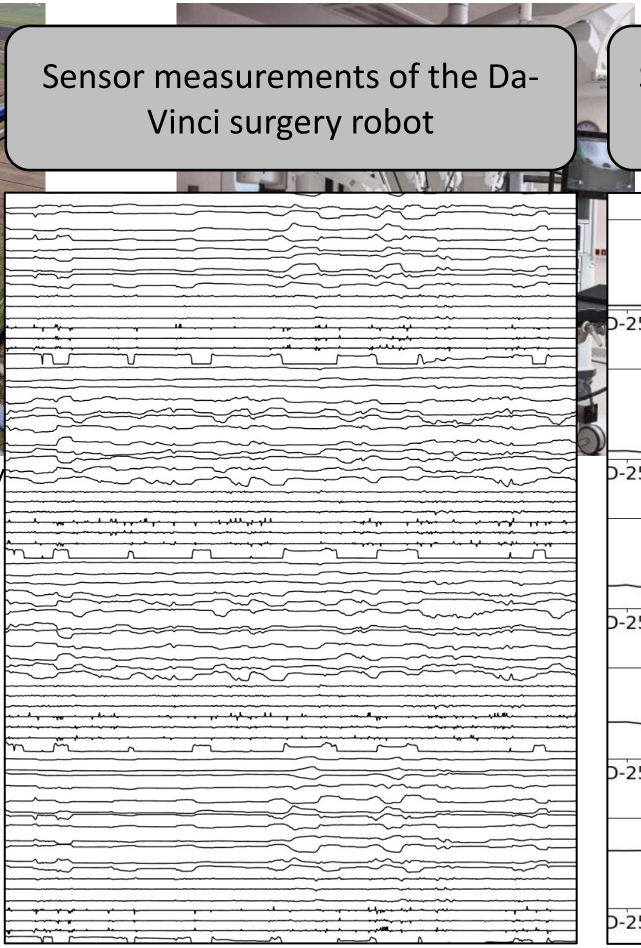
Astrophysics

Fiber-acoustic sensors in the VIRGO north building



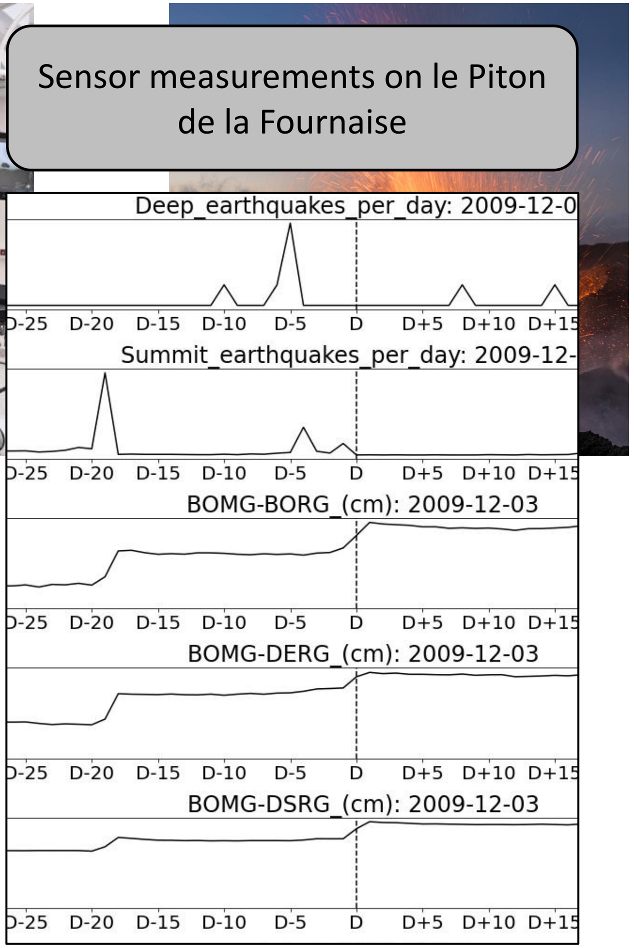
Medicine

Sensor measurements of the Da-Vinci surgery robot



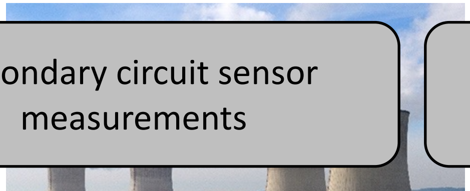
Volcanology

Sensor measurements on le Piton de la Fournaise

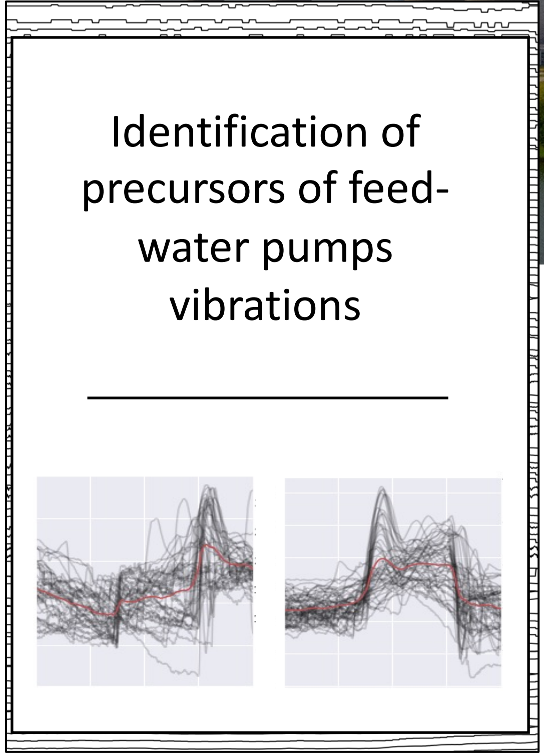


Introduction: *with Important Challenges*

Energy Production

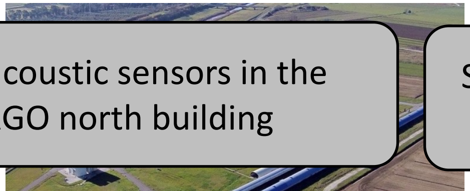


Secondary circuit sensor measurements

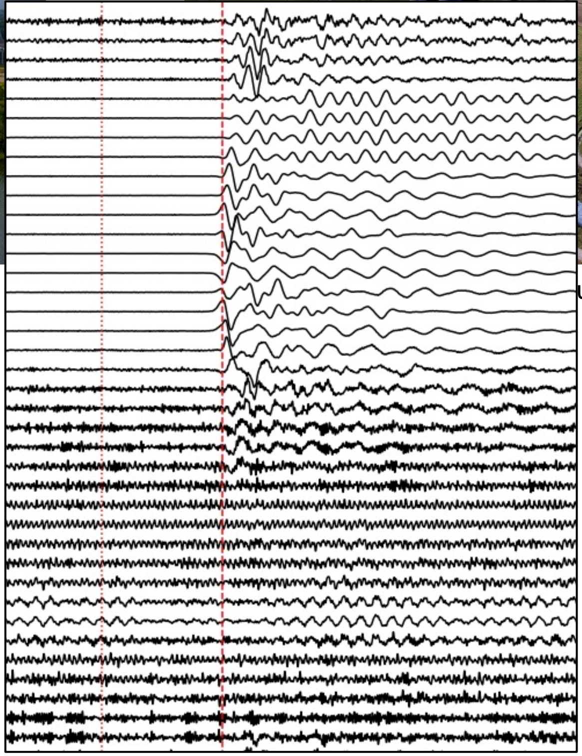


Identification of precursors of feed-water pumps vibrations

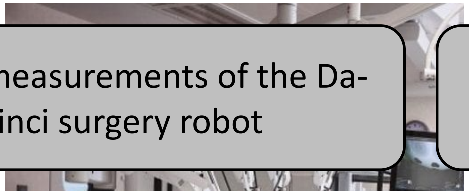
Astrophysics



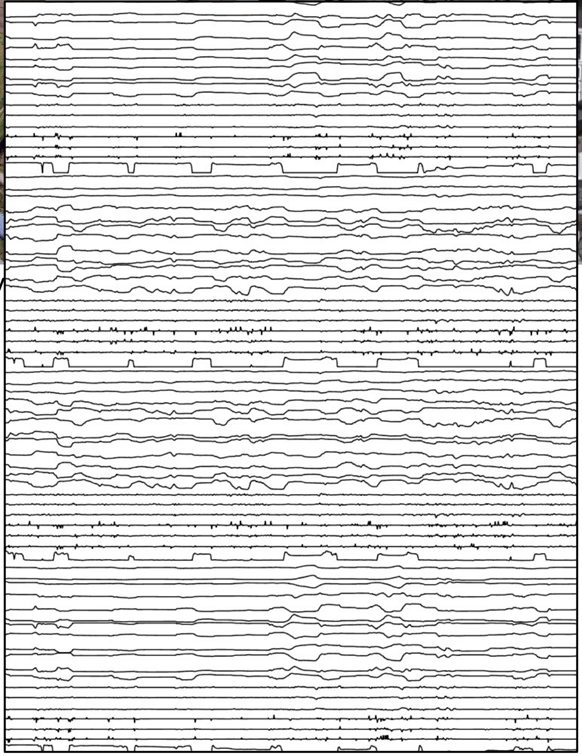
Fiber-acoustic sensors in the VIRGO north building



Medicine

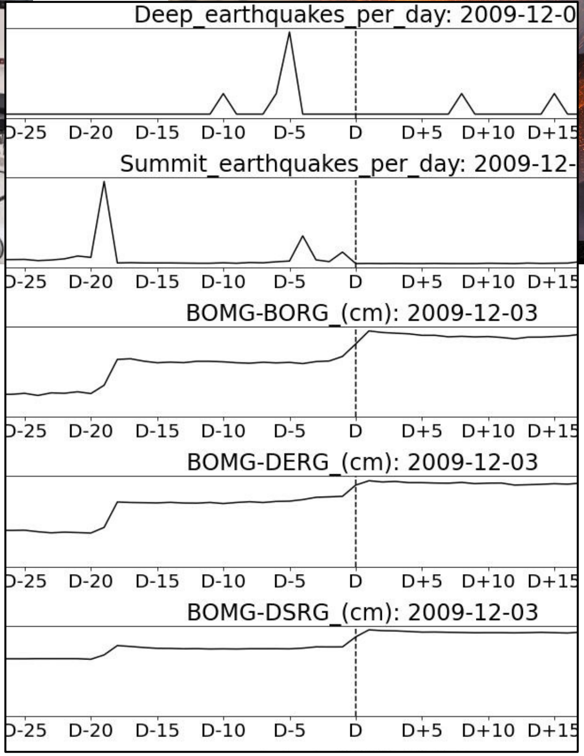


Sensor measurements of the Da-Vinci surgery robot



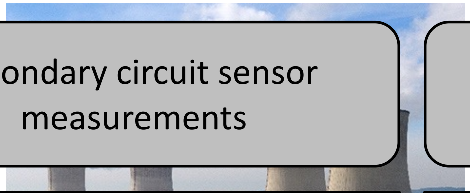
Volcanology

Sensor measurements on le Piton de la Fournaise

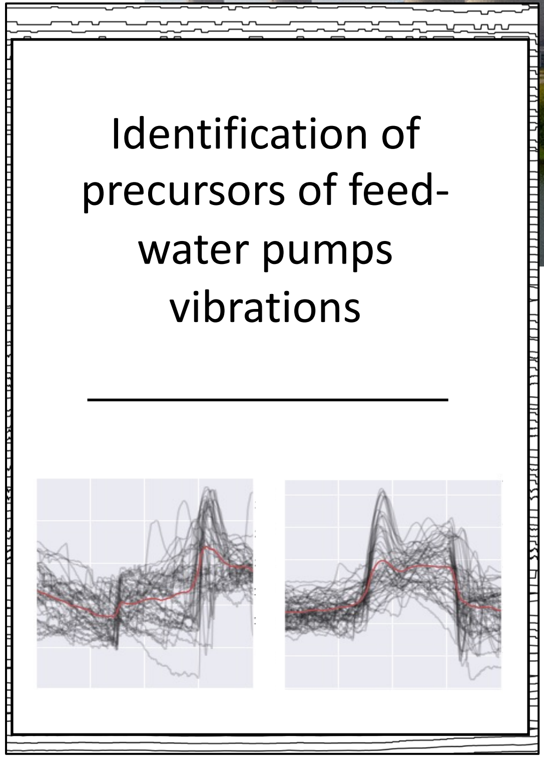


Introduction: *with Important Challenges*

Energy Production

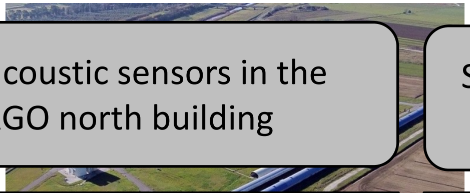


Secondary circuit sensor measurements

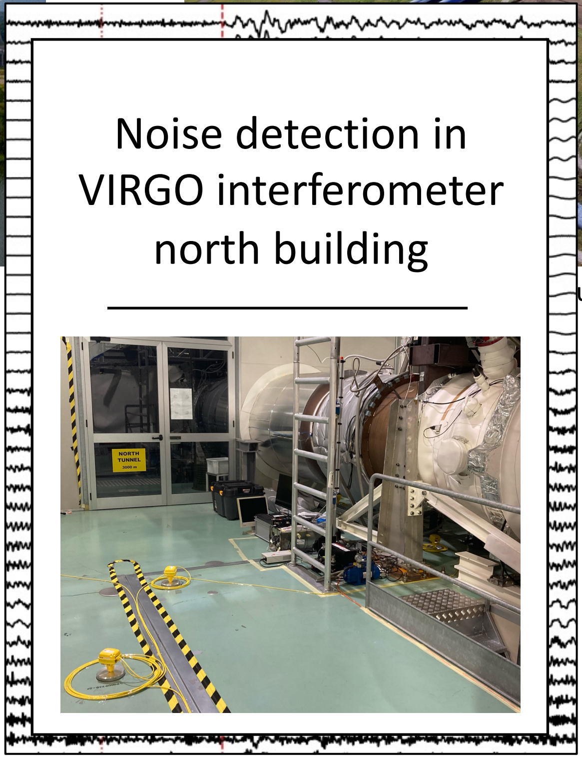


Identification of precursors of feed-water pumps vibrations

Astrophysics

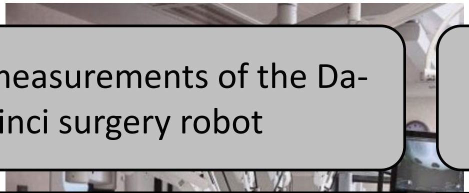


Fiber-acoustic sensors in the VIRGO north building

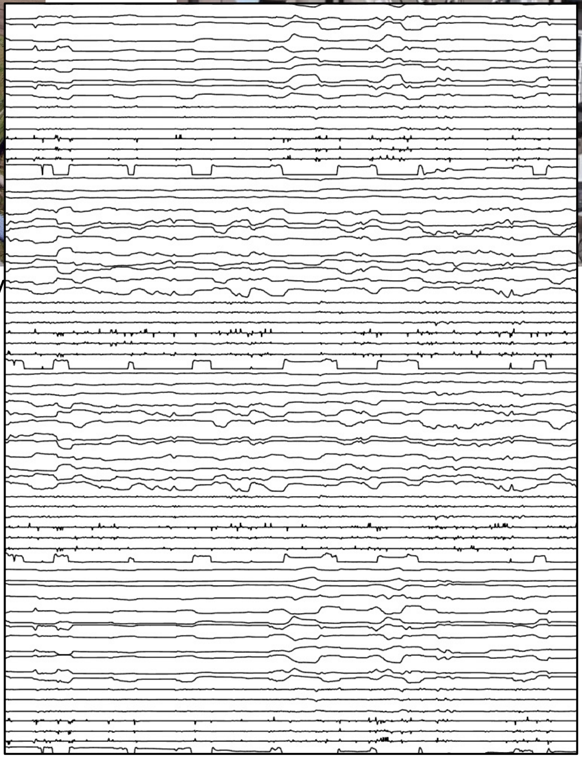


Noise detection in VIRGO interferometer north building

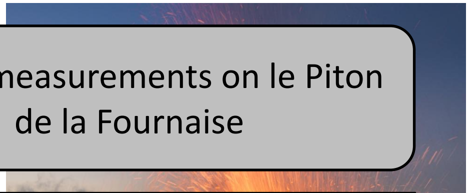
Medicine



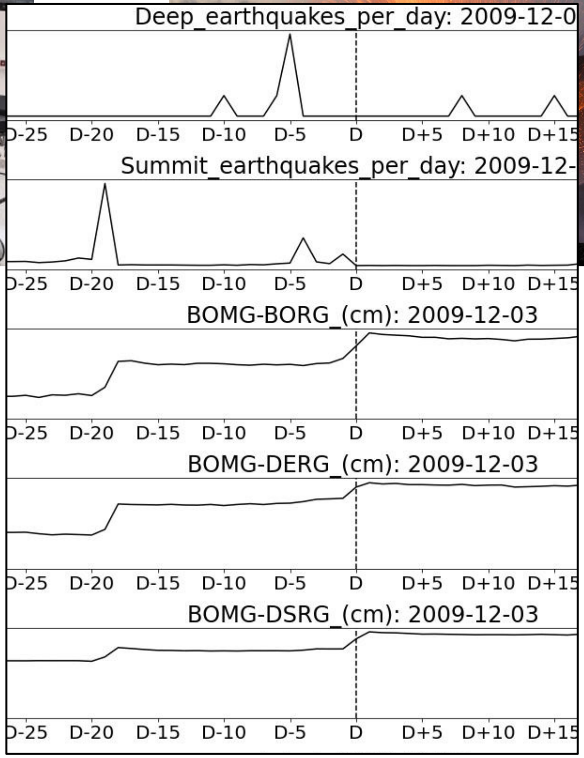
Sensor measurements of the Da-Vinci surgery robot



Volcanology



Sensor measurements on le Piton de la Fournaise

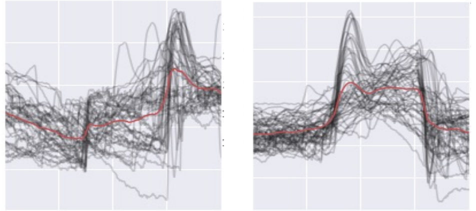


Introduction: *with Important Challenges*

Energy Production

Secondary circuit sensor measurements

Identification of precursors of feed-water pumps vibrations



Astrophysics

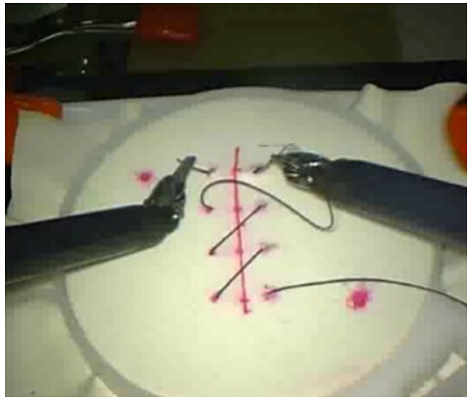
Fiber-acoustic sensors in the VIRGO north building

Noise detection in VIRGO interferometer north building

Medicine

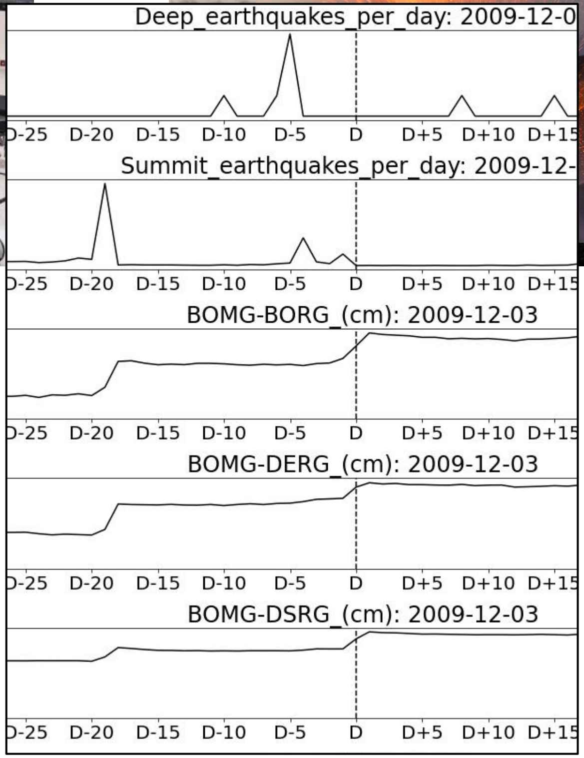
Sensor measurements of the Da-Vinci surgery robot

Unusual surgeons gestures detection



Volcanology

Sensor measurements on le Piton de la Fournaise

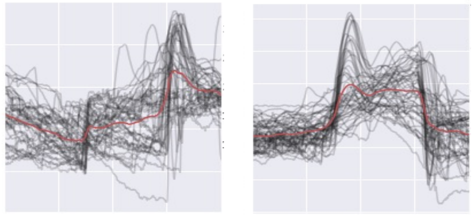


Introduction: *with Important Challenges*

Energy Production

Secondary circuit sensor measurements

Identification of precursors of feed-water pumps vibrations



Astrophysics

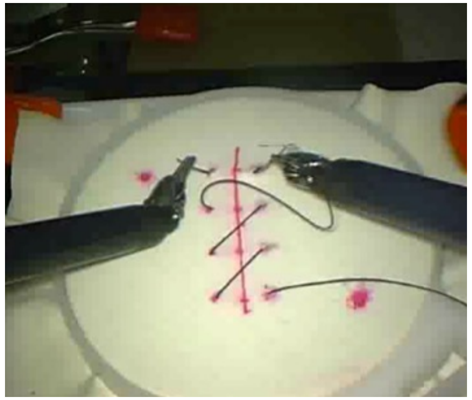
Fiber-acoustic sensors in the VIRGO north building

Noise detection in VIRGO interferometer north building

Medicine

Sensor measurements of the Da-Vinci surgery robot

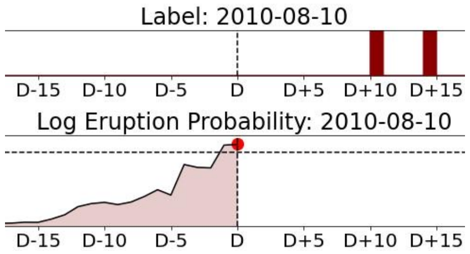
Unusual surgeons gestures detection



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

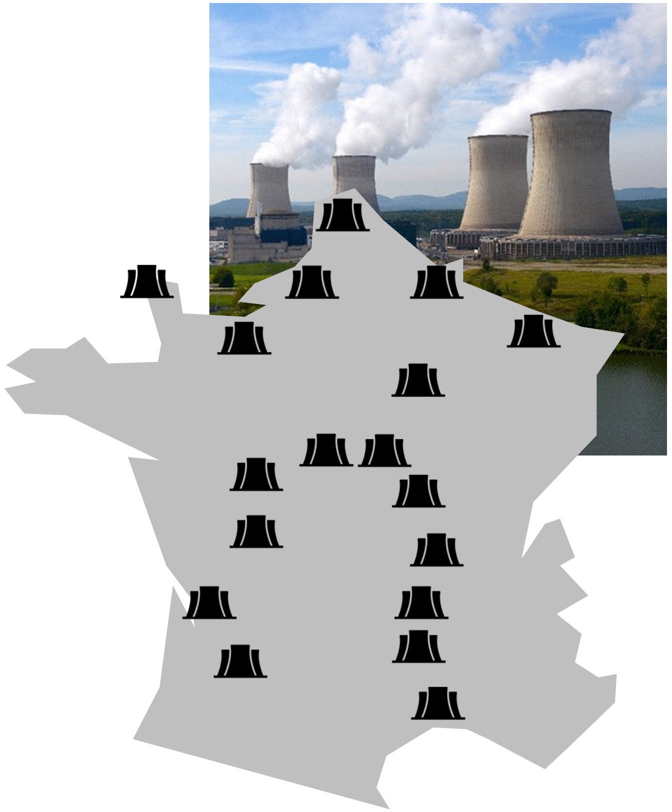
Energy Production

Edf.fr: tinyurl.com/yc7x5xje

Introduction: *with Important Challenges*

Large-scale time series database

Energy Production



Edf.fr: tinyurl.com/yc7x5xje

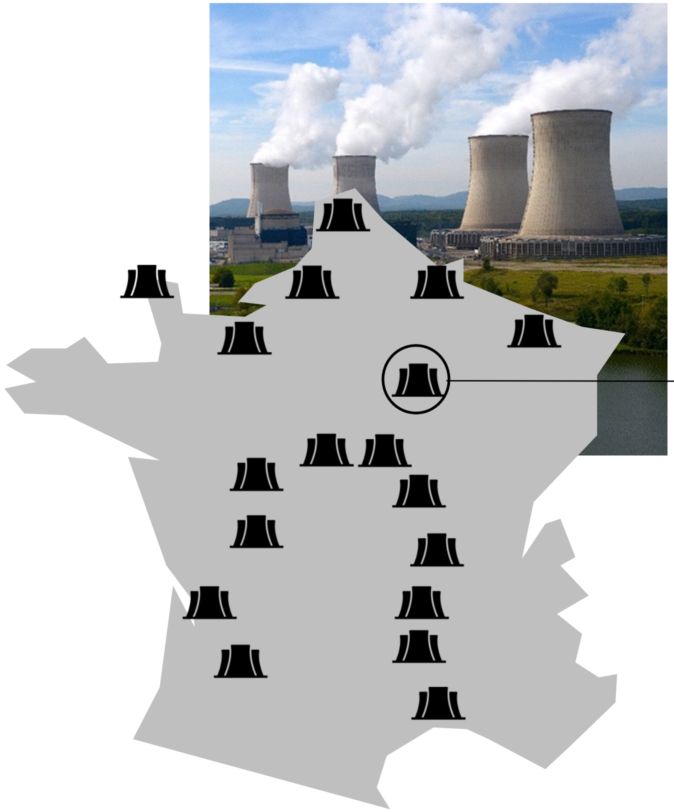
Example of Nuclear production

- 58 nuclear power plants across France

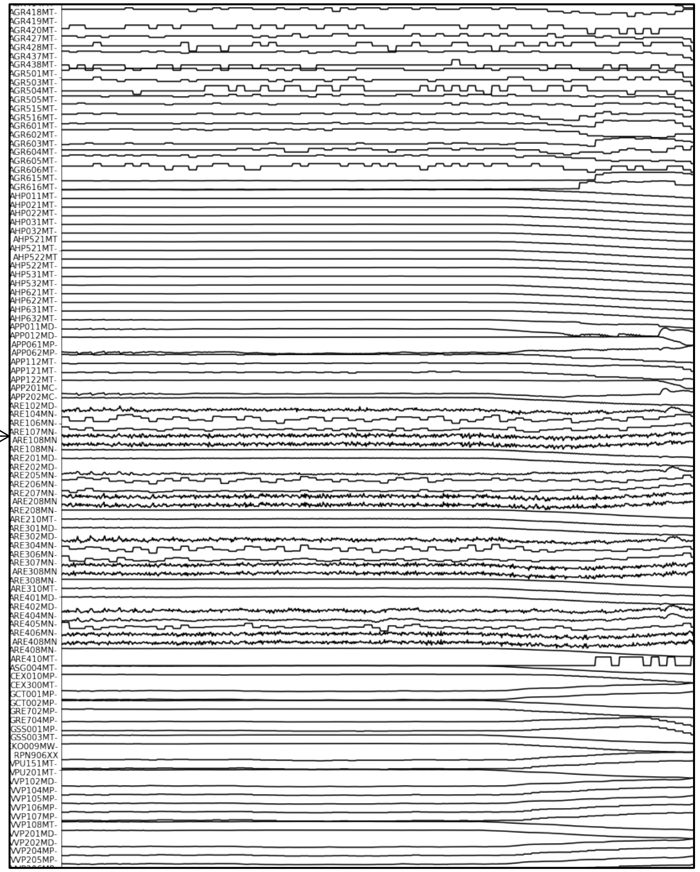
Introduction: *with Important Challenges*

Large-scale time series database

Energy Production



Edf.fr: tinyurl.com/yc7x5xje



Example of Nuclear production

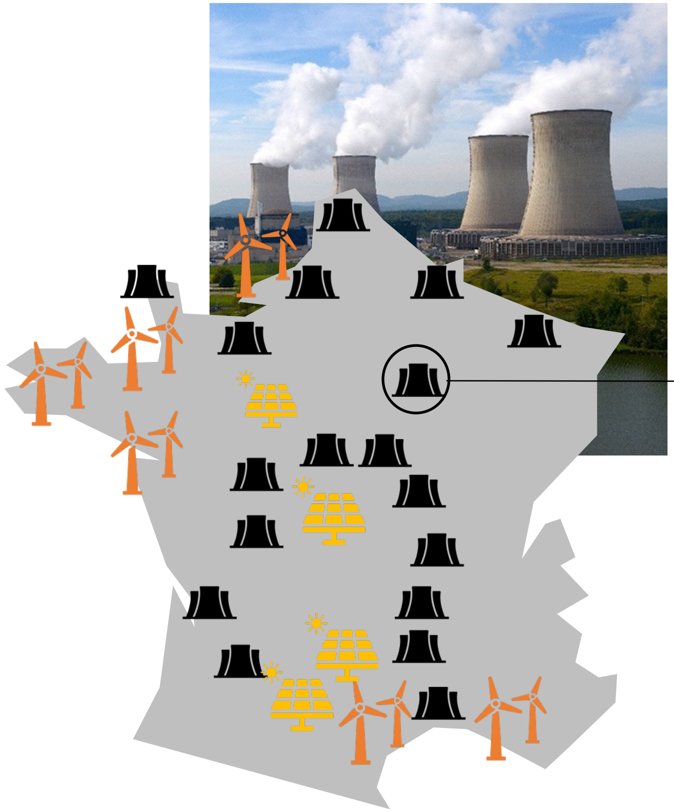
- 58 nuclear power plants across France
- 2000+ sensors per power plant
- 30 years of data collections

A total of 500 TeraBytes

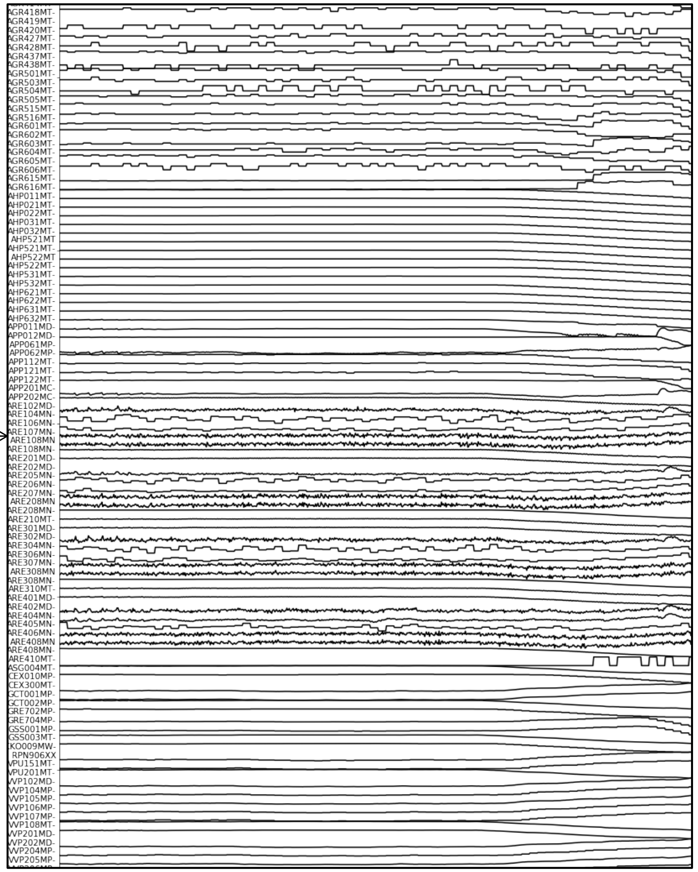
Introduction: *with Important Challenges*

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

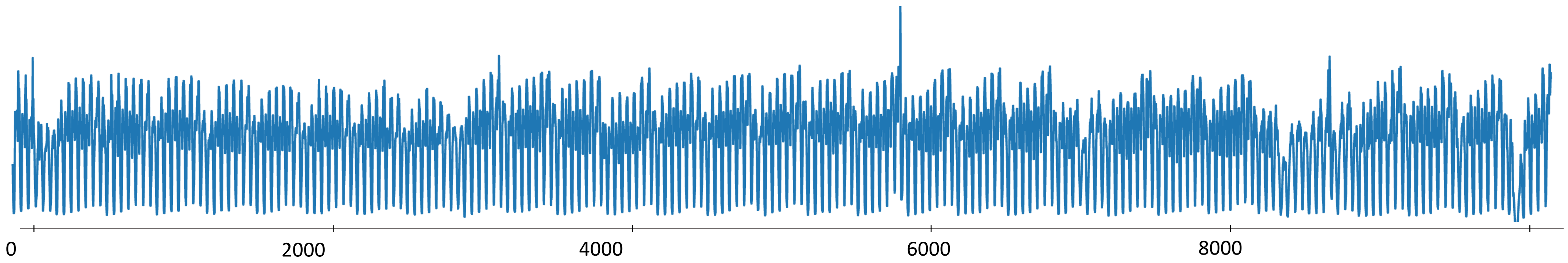
A total of 500 TeraBytes

Other source of production

- New sensors with higher acquisition rate

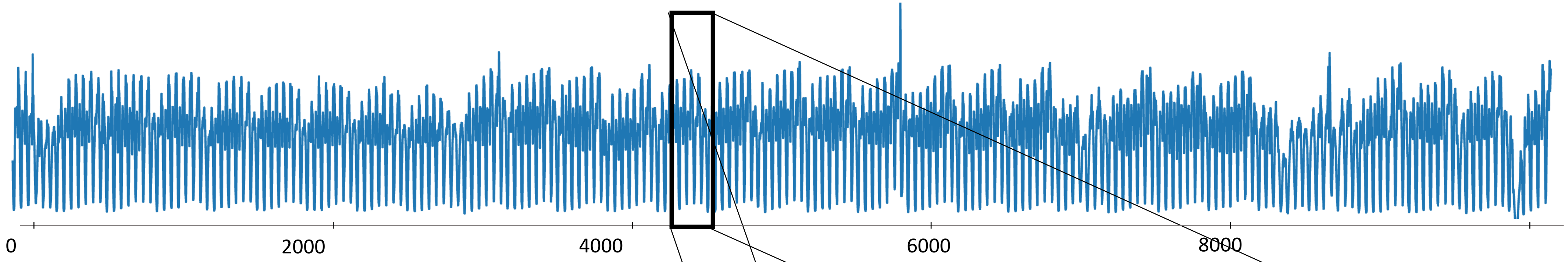
Introduction: *Anomaly Detection in Time Series*

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

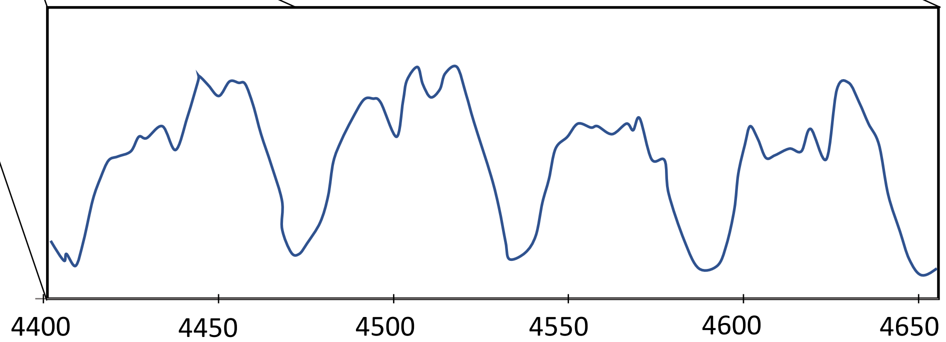


Introduction: *Anomaly Detection in Time Series*

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

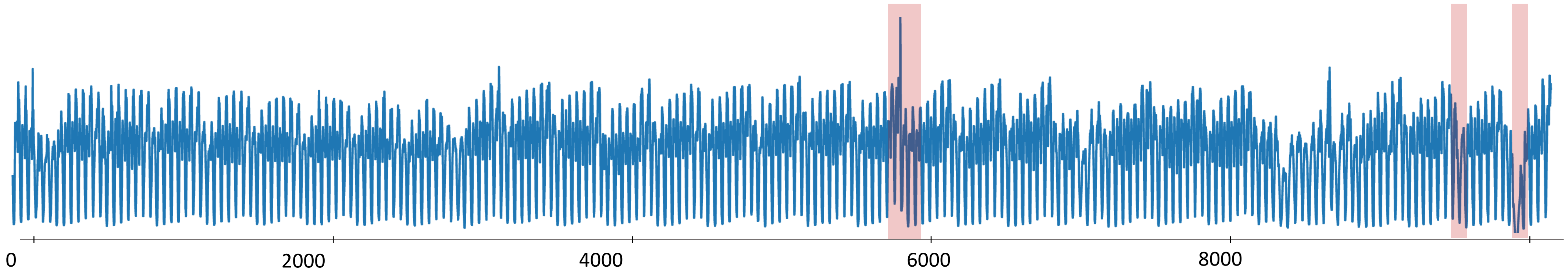


- Subsequence $T_{i,\ell}$
  with $i = 4400, \ell = 250$



Introduction: *Anomaly Detection in Time Series*

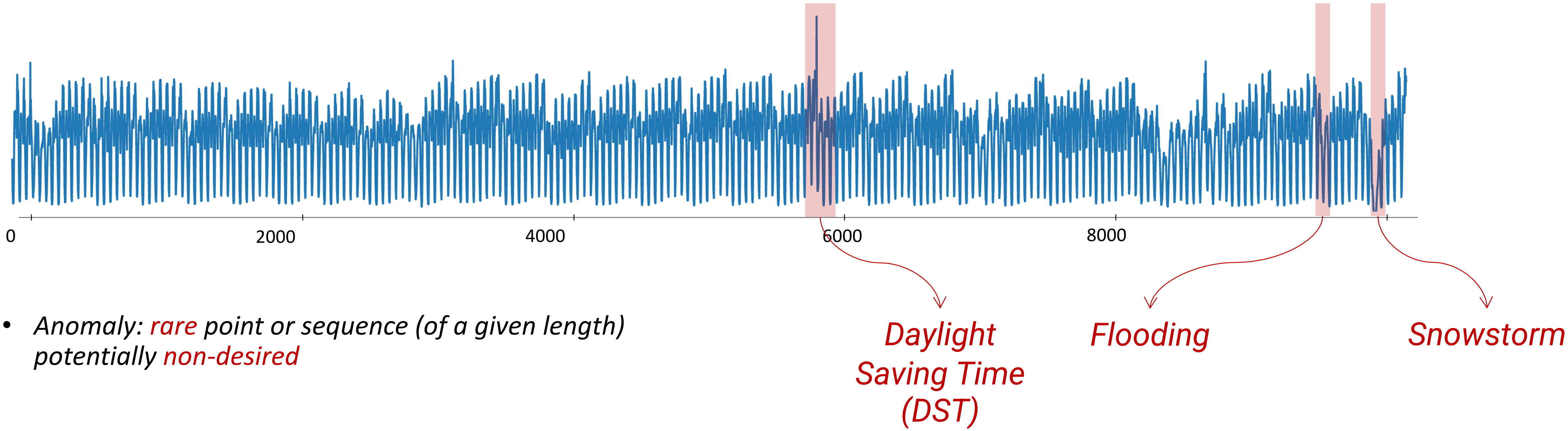
- Time series T *(example : number of taxi passengers in New York City)*



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

Introduction: *Anomaly Detection in Time Series*

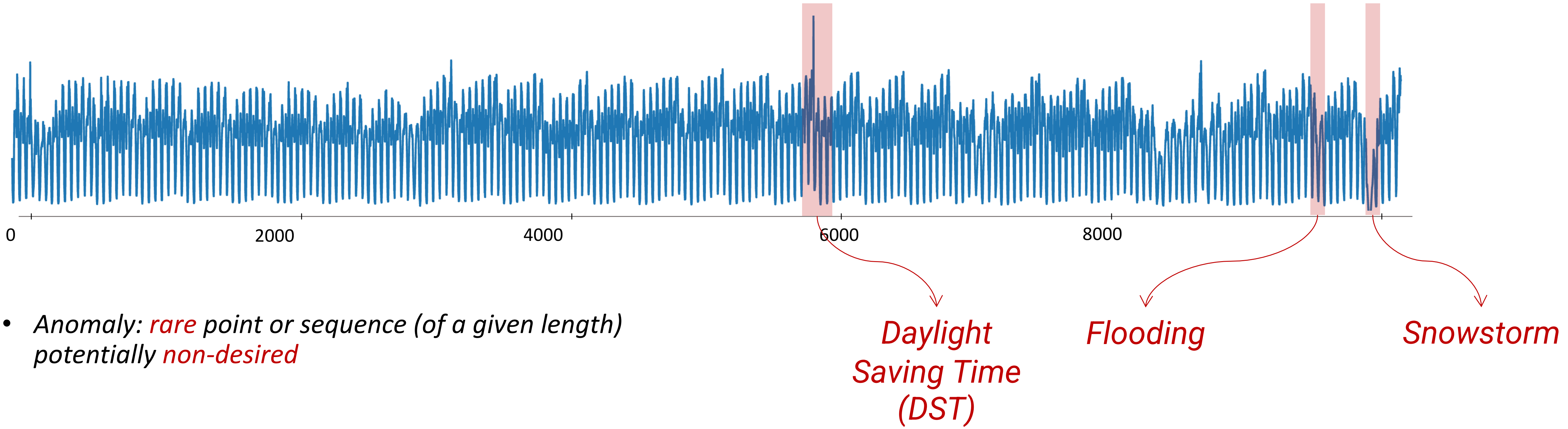
- Time series T *(example : number of taxi passengers in New York City)*



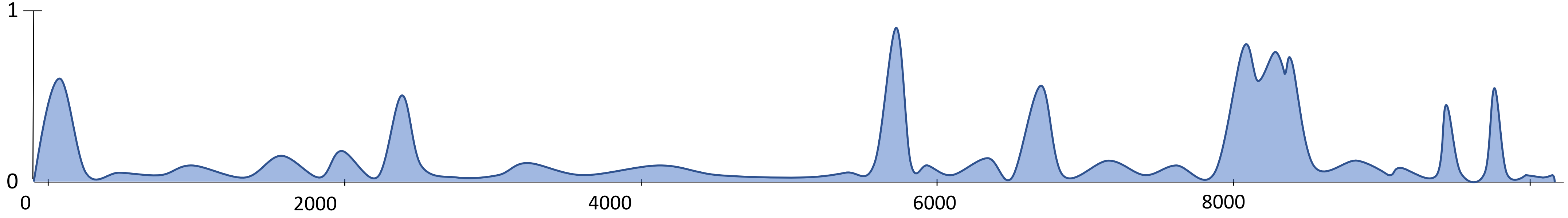
- *Anomaly: rare point or sequence (of a given length) potentially non-desired*

Introduction: *Anomaly Detection in Time Series*

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



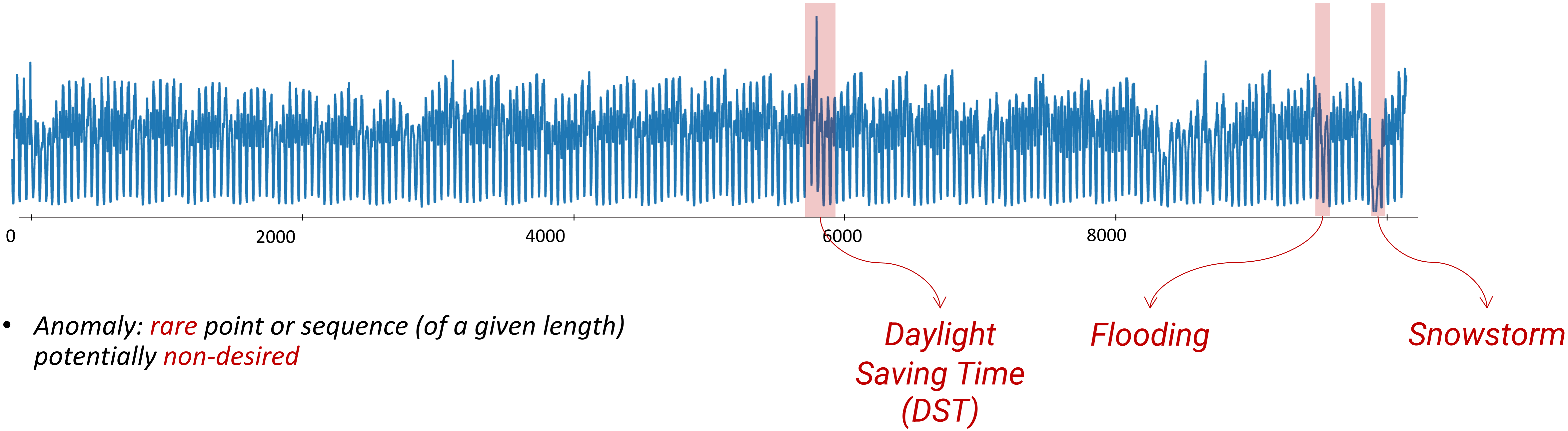
- *Anomaly: rare point or sequence (of a given length) potentially non-desired*



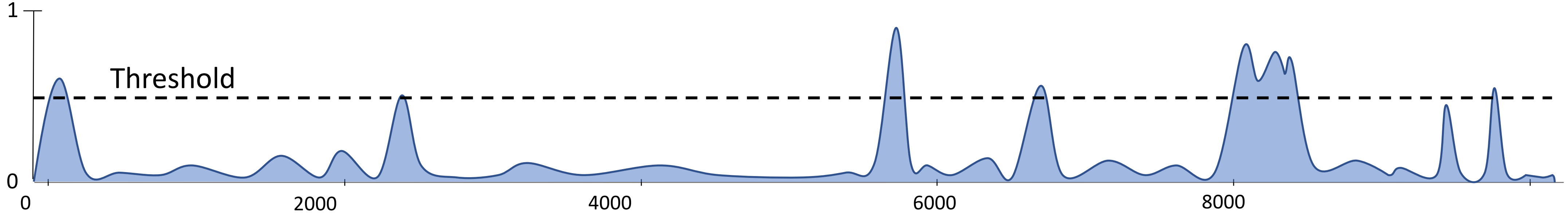
Anomaly score S_T

Introduction: *Anomaly Detection in Time Series*

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



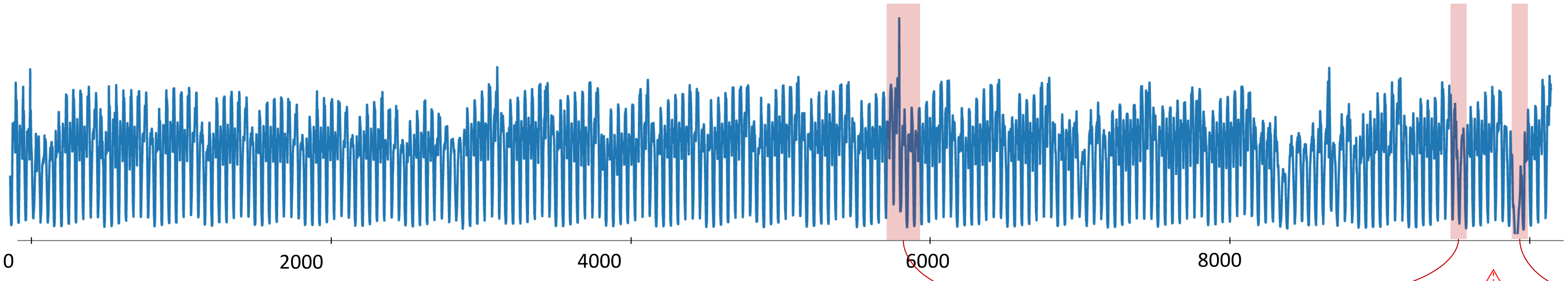
- *Anomaly: rare point or sequence (of a given length) potentially non-desired*



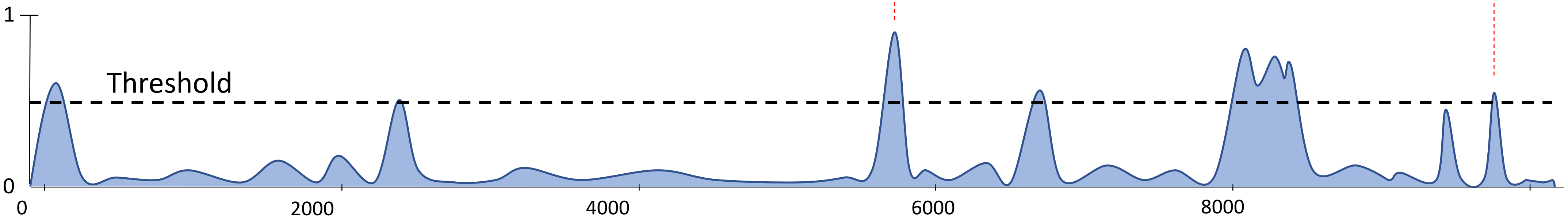
Anomaly score S_T

Introduction: *Anomaly Detection in Time Series*

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



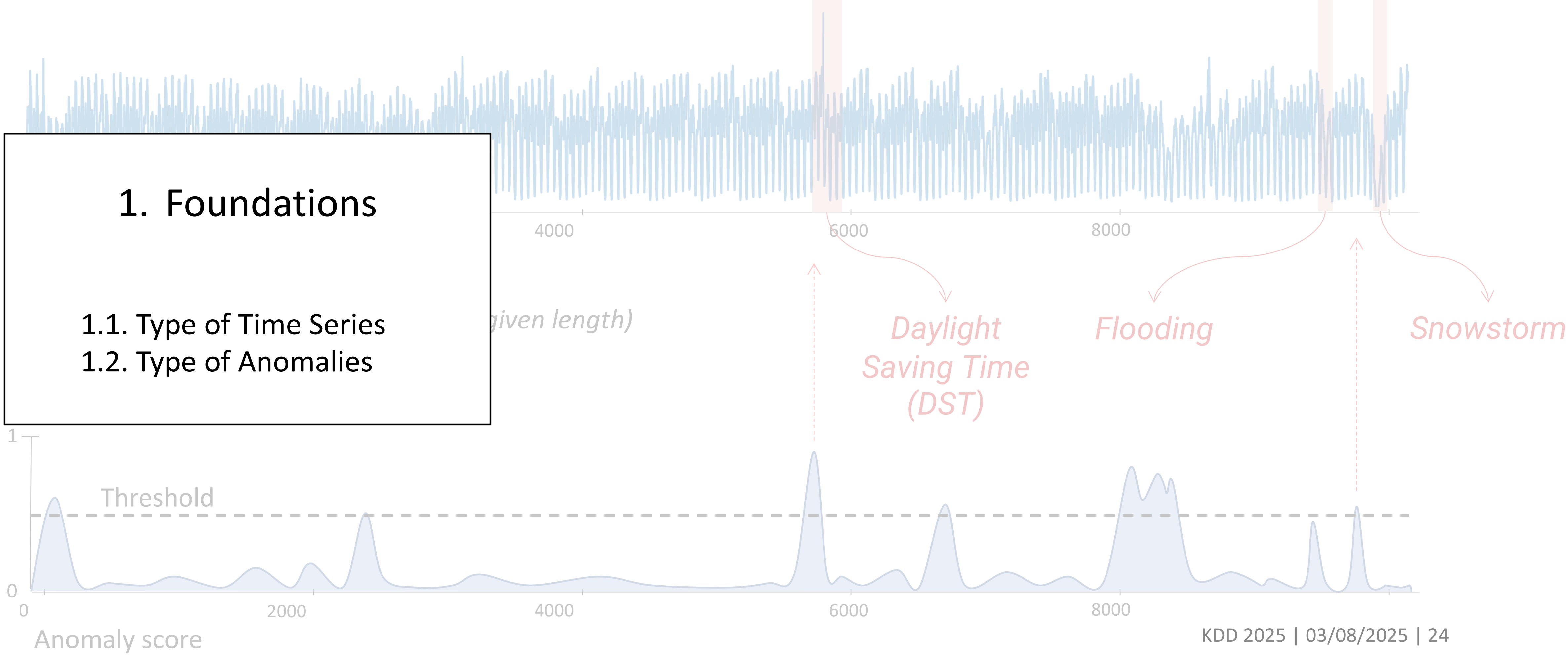
- *Anomaly: rare point or sequence (of a given length) potentially non-desired*



Anomaly score S_T

Introduction: *Outline*

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

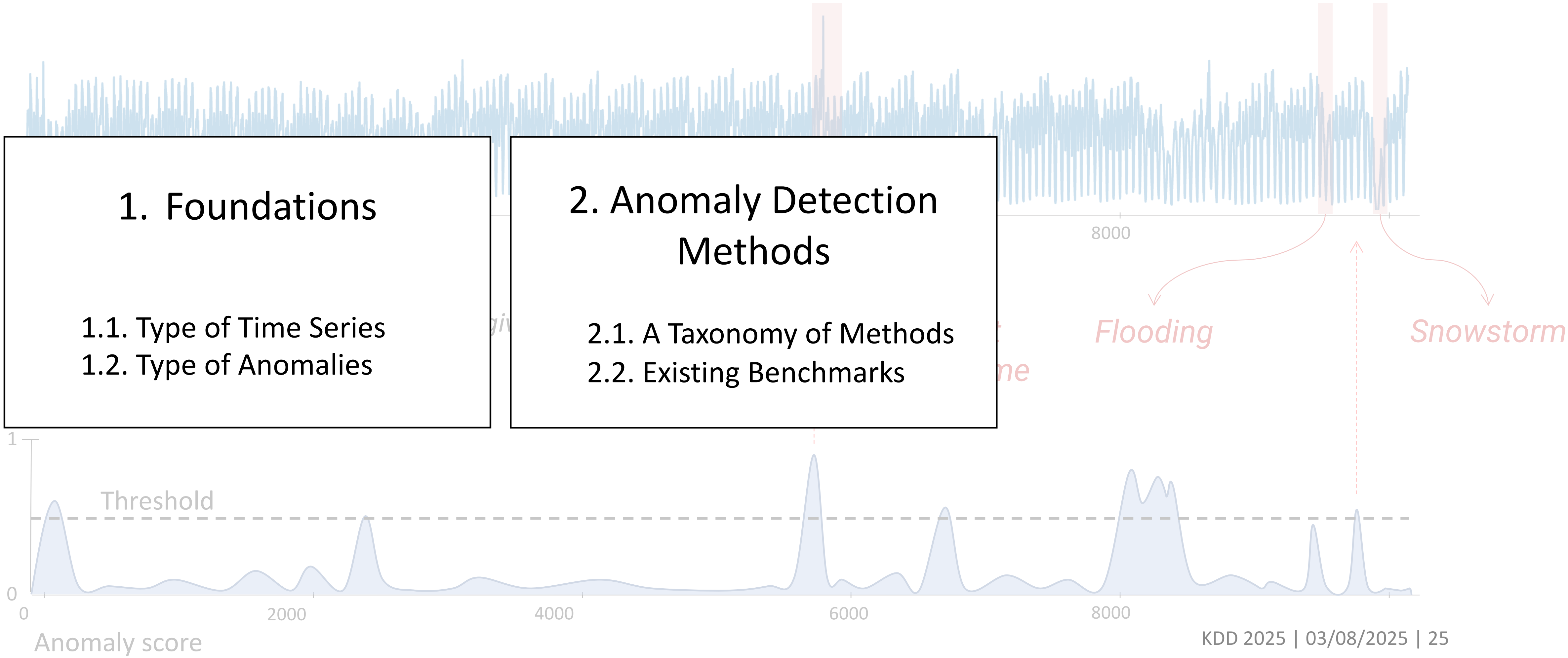


1. Foundations

1.1. Type of Time Series
1.2. Type of Anomalies

Introduction: *Outline*

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



1. Foundations

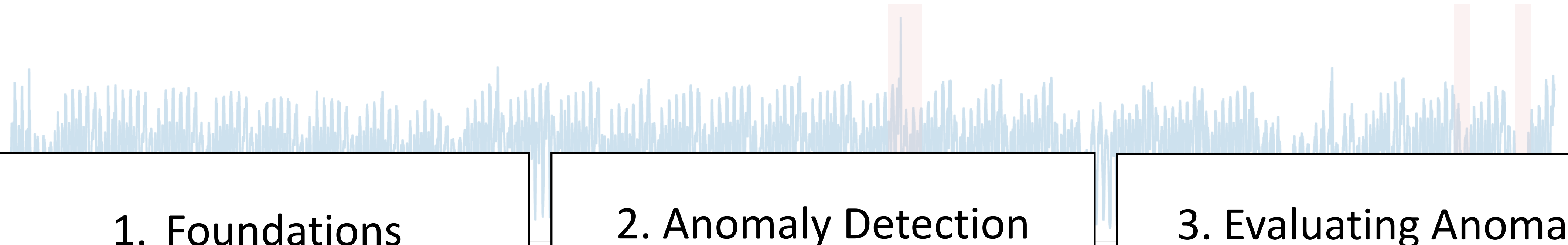
1.1. Type of Time Series
1.2. Type of Anomalies

2. Anomaly Detection Methods

2.1. A Taxonomy of Methods
2.2. Existing Benchmarks

Introduction: *Outline*

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



1. Foundations

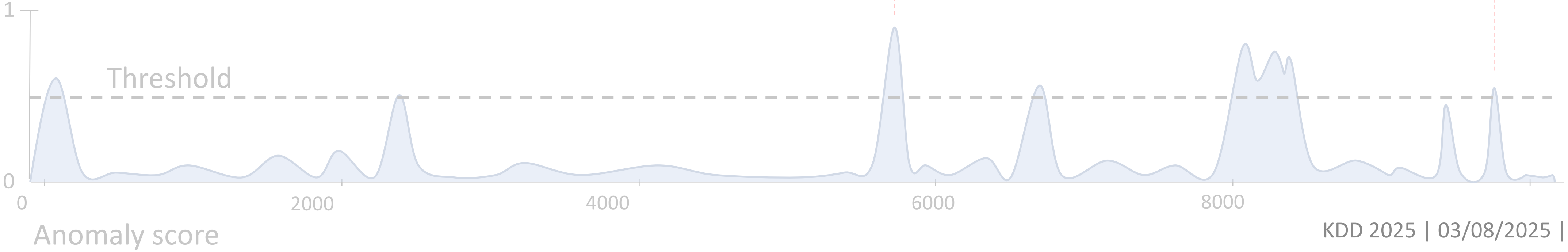
1.1. Type of Time Series
1.2. Type of Anomalies

2. Anomaly Detection Methods

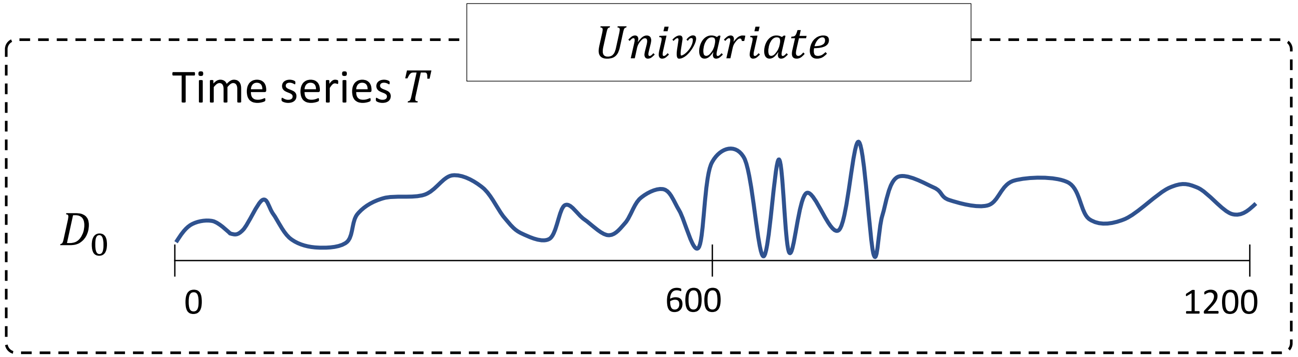
2.1. A Taxonomy of Methods
2.2. Existing Benchmarks

3. Evaluating Anomaly Detection

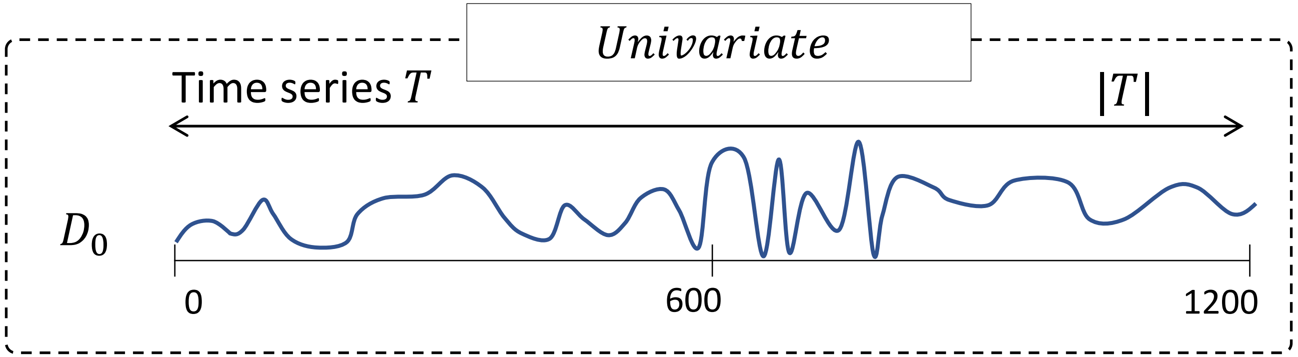
3.1. Threshold-based
3.2. Time series labeling issues



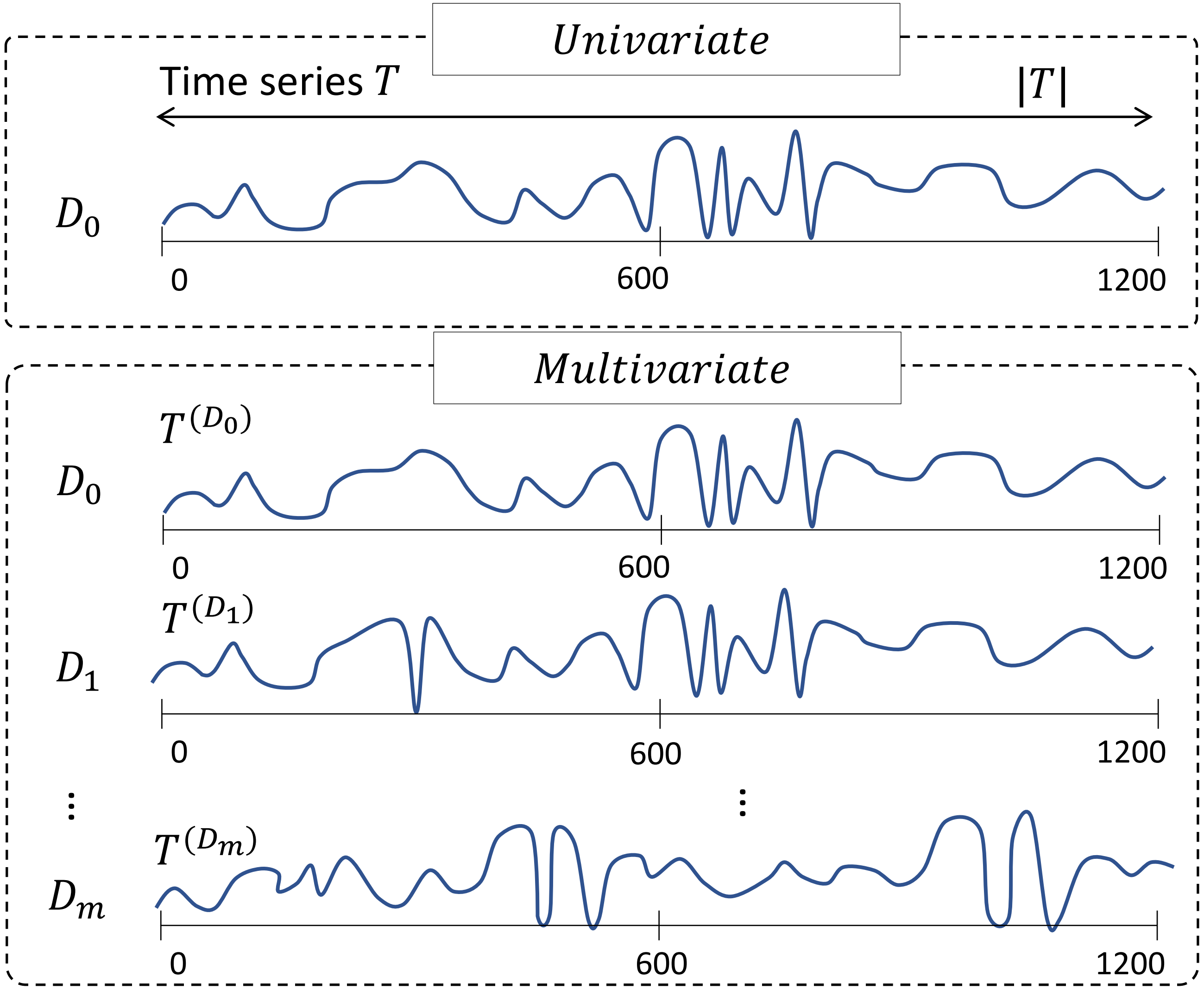
Foundations: *Type of time series*



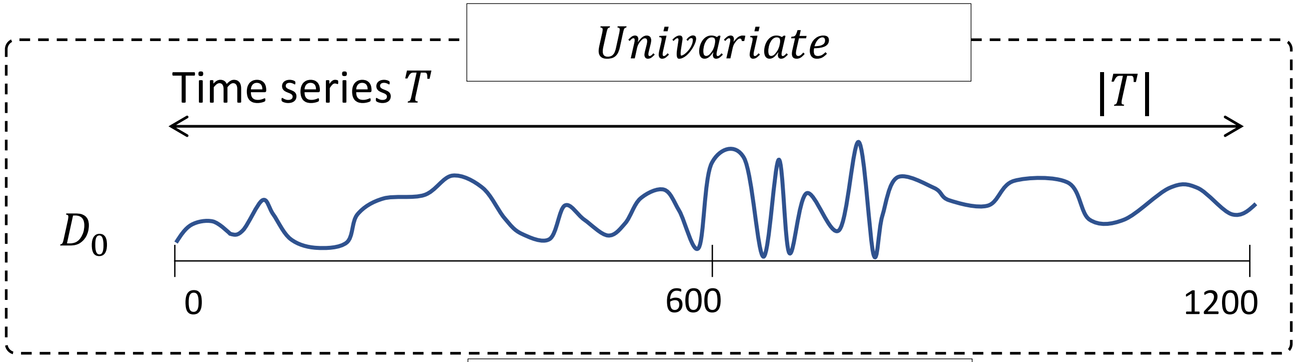
Foundations: *Type of time series*

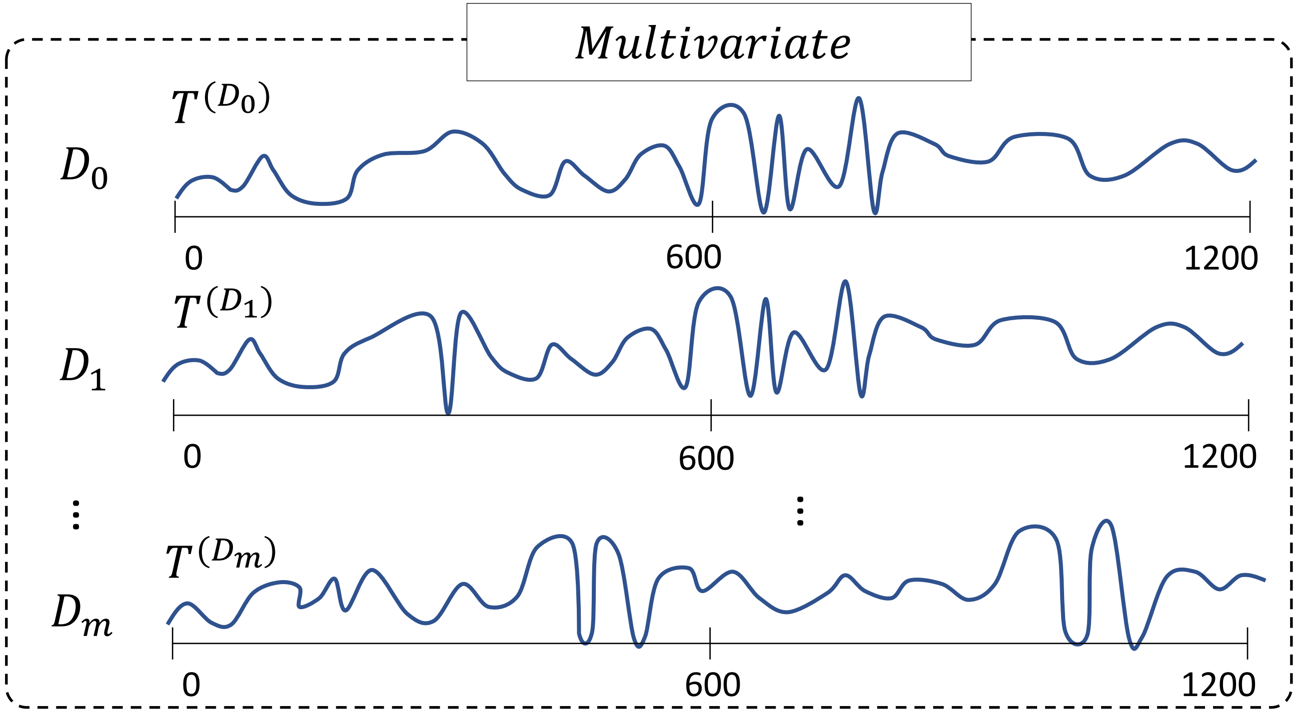


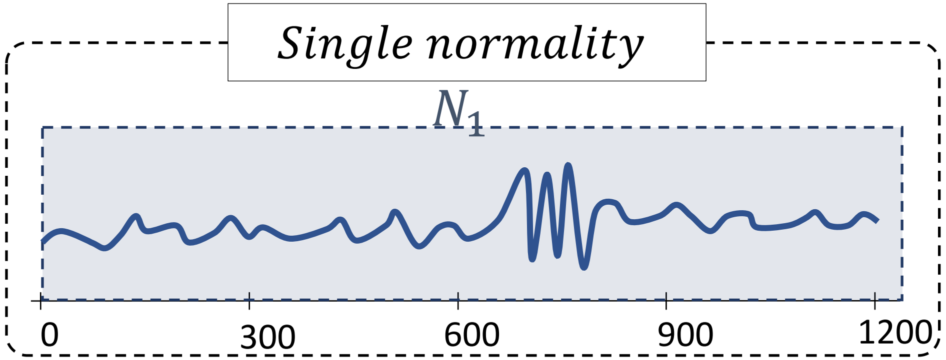
Foundations: *Type of time series*



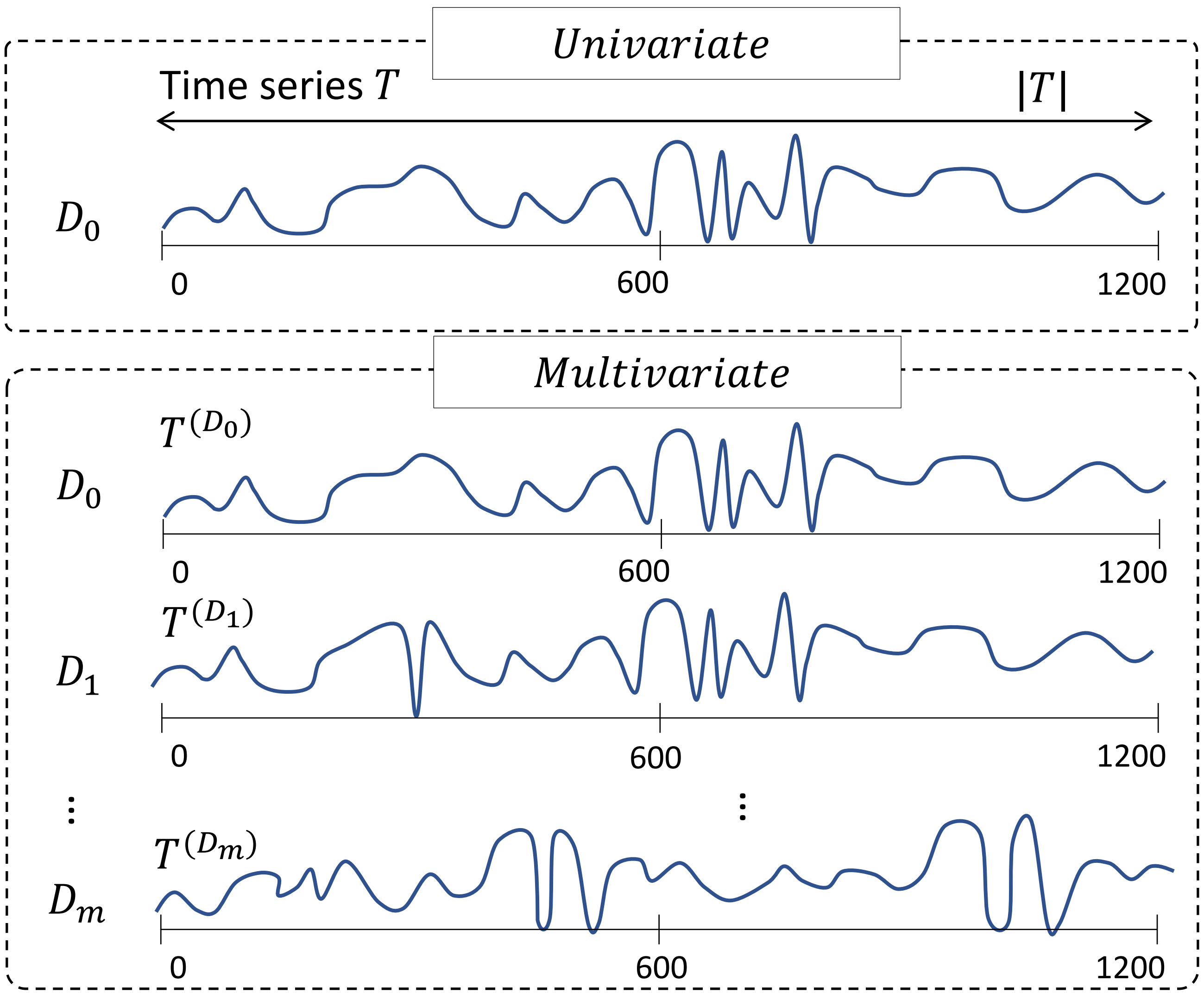
Foundations: *Type of time series*

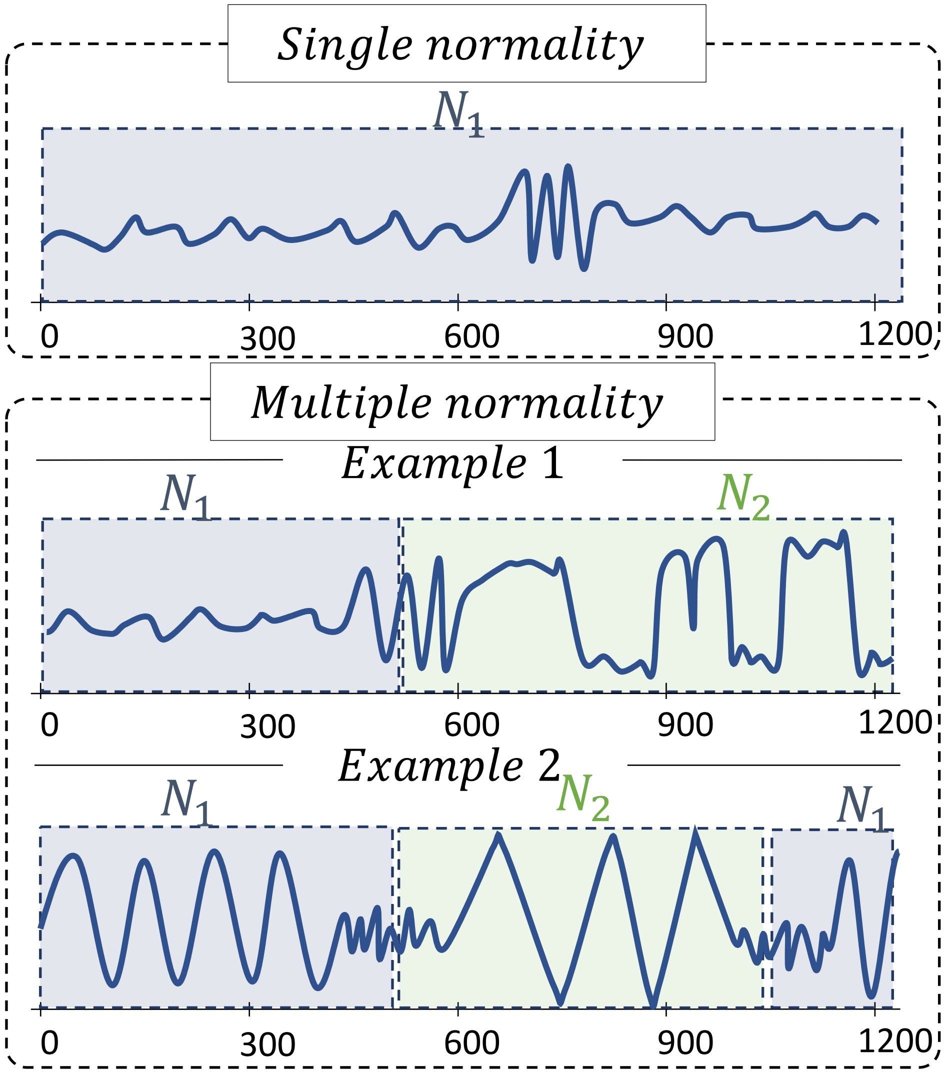




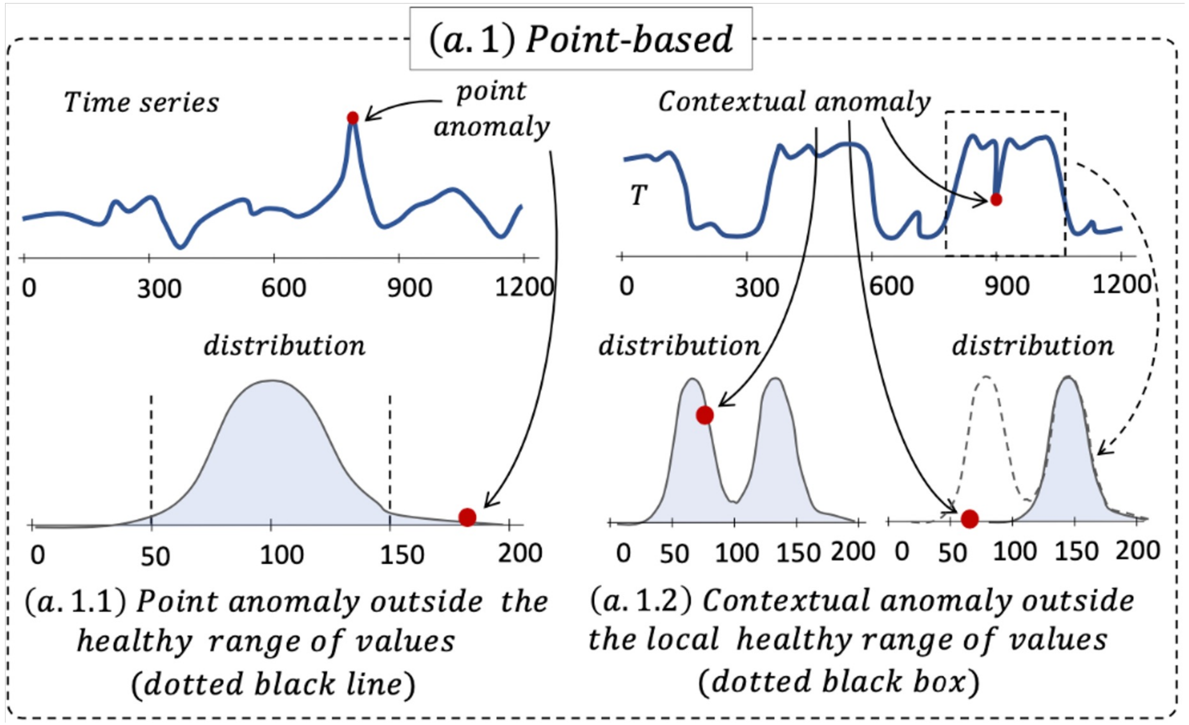


Foundations: *Type of time series*

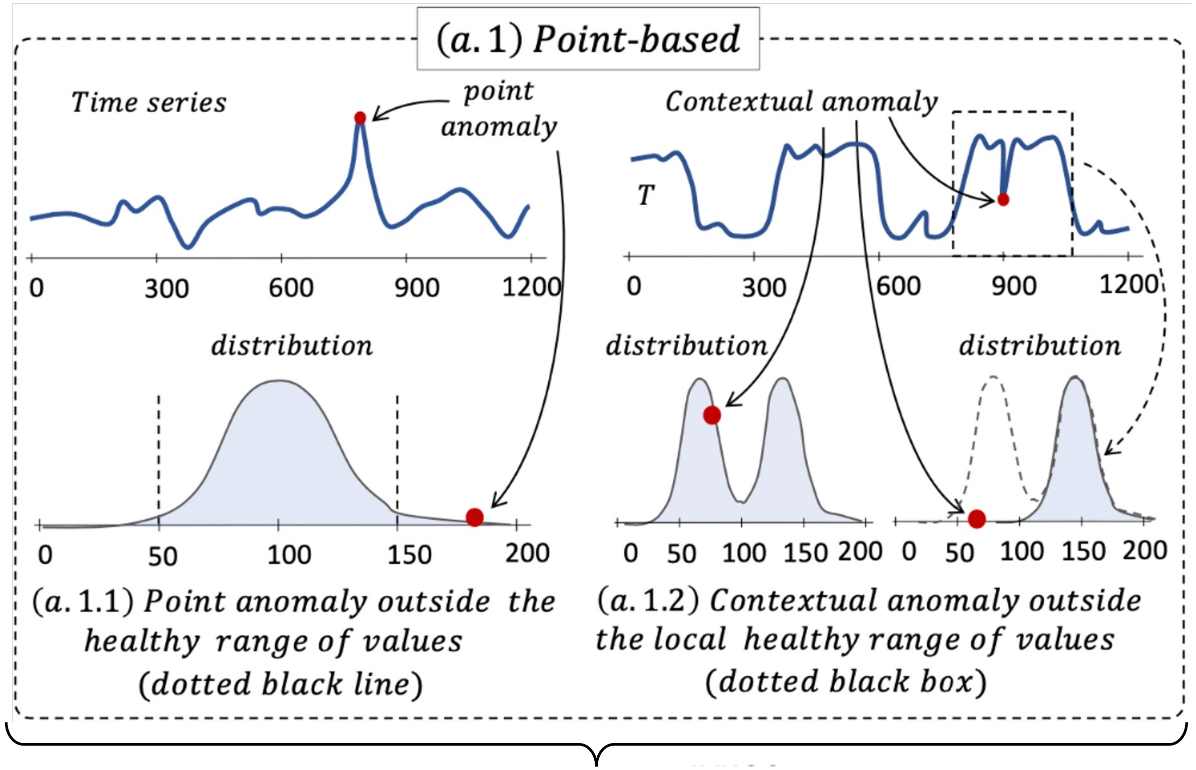




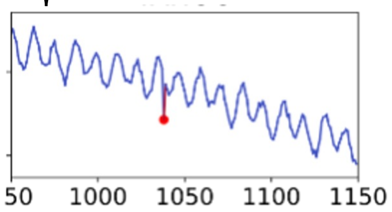
Foundations: *Type of anomalies*



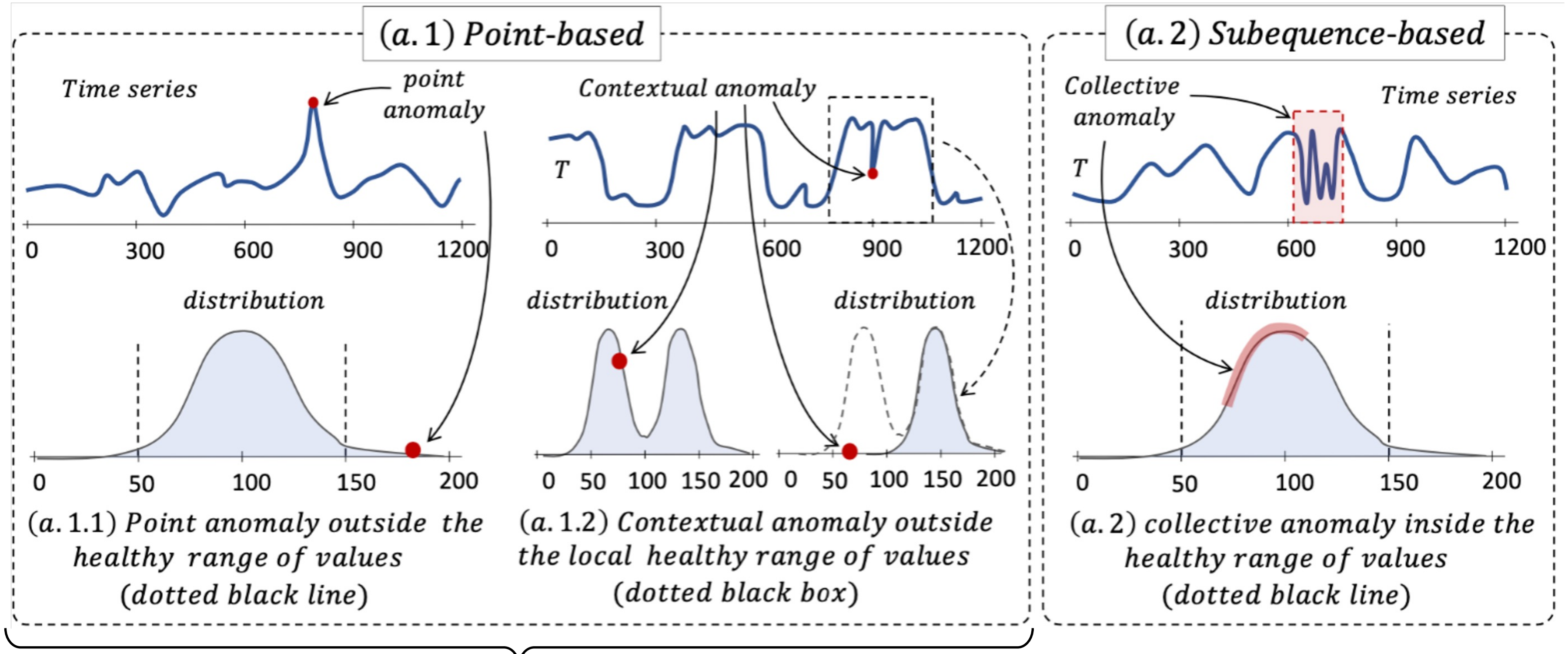
Foundations: *Type of anomalies*



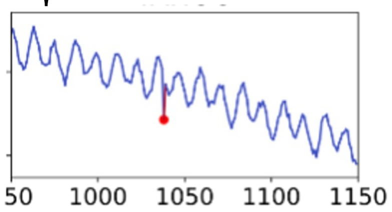
Example of point-based anomaly [1]



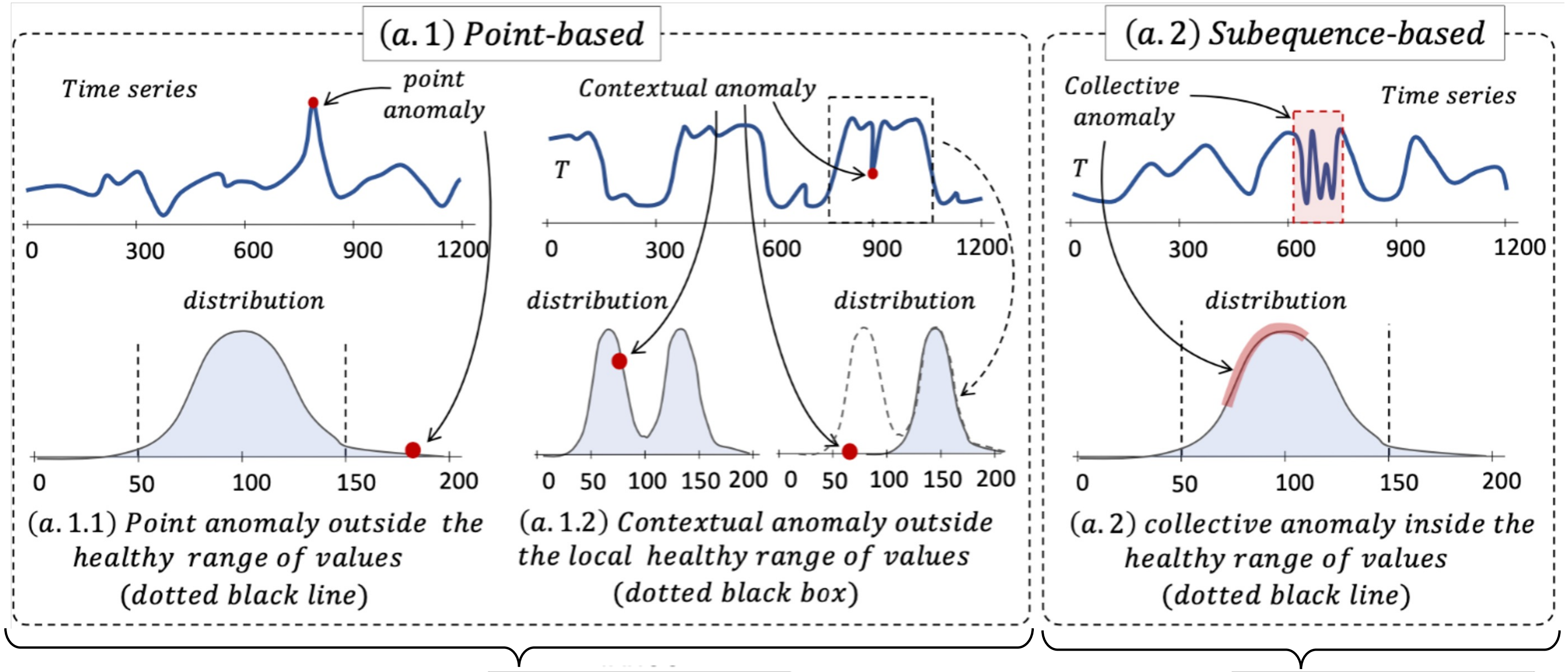
Foundations: *Type of anomalies*



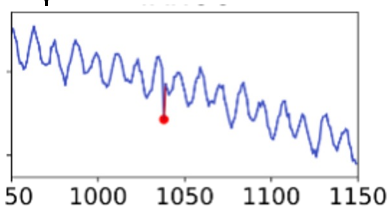
Example of point-based anomaly [1]



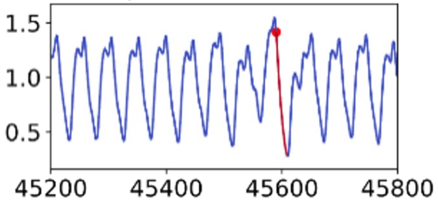
Foundations: *Type of anomalies*



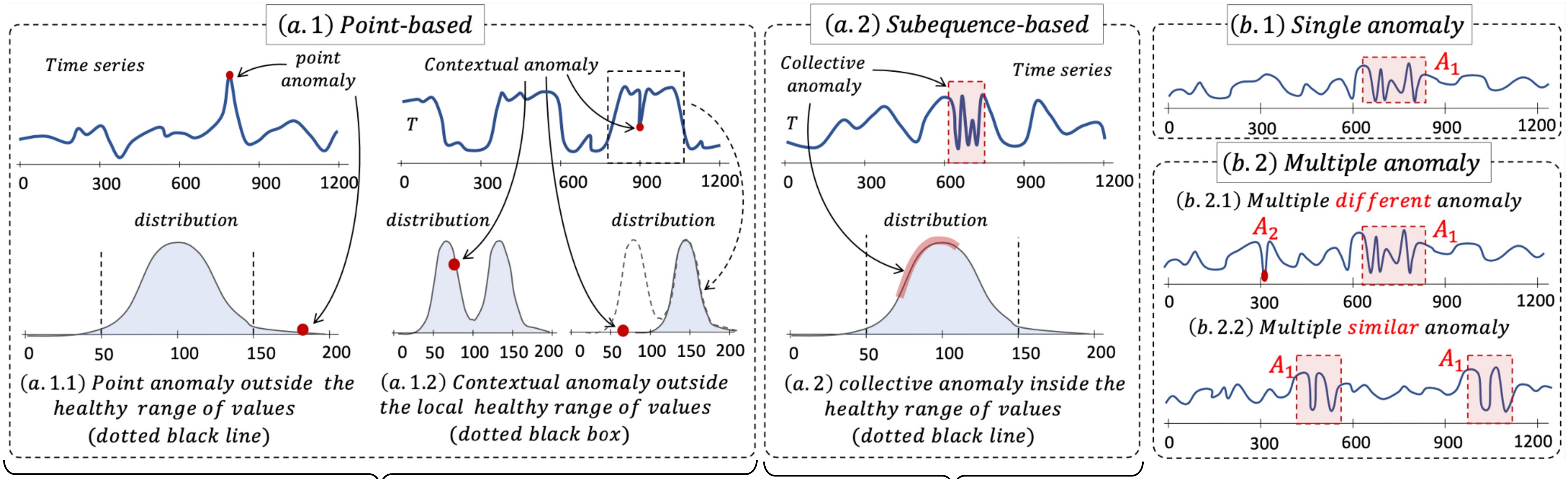
Example of point-based anomaly [1]



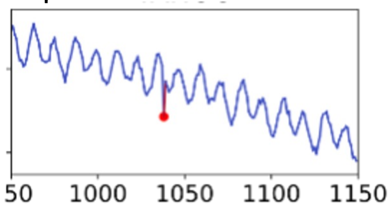
Example of subsequence-based anomaly [2]



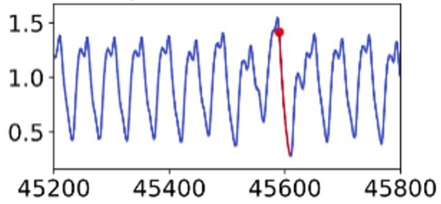
Foundations: *Type of anomalies*



Example of point-based anomaly [1]

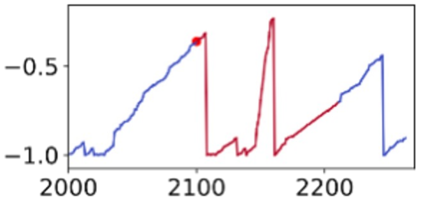


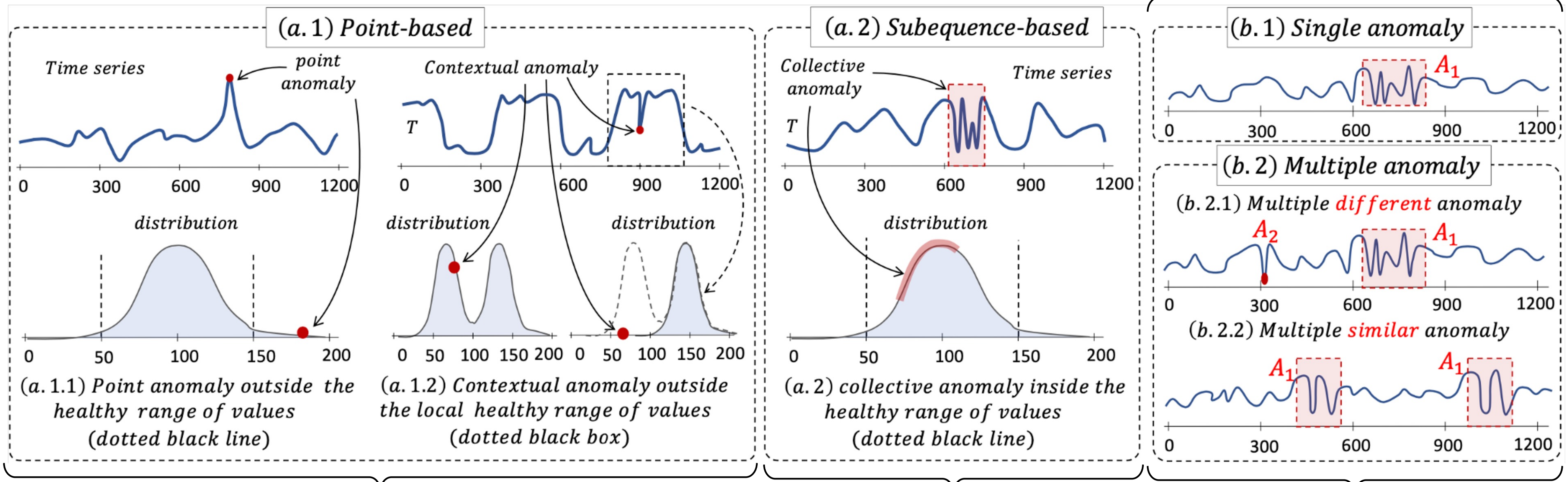
Example of subsequence-based anomaly [2]



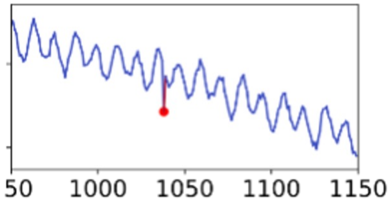
Foundations: *Type of anomalies*

Example of single anomaly [3]

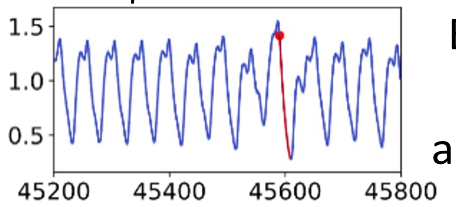




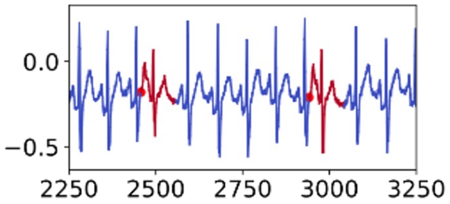
Example of point-based anomaly [1]



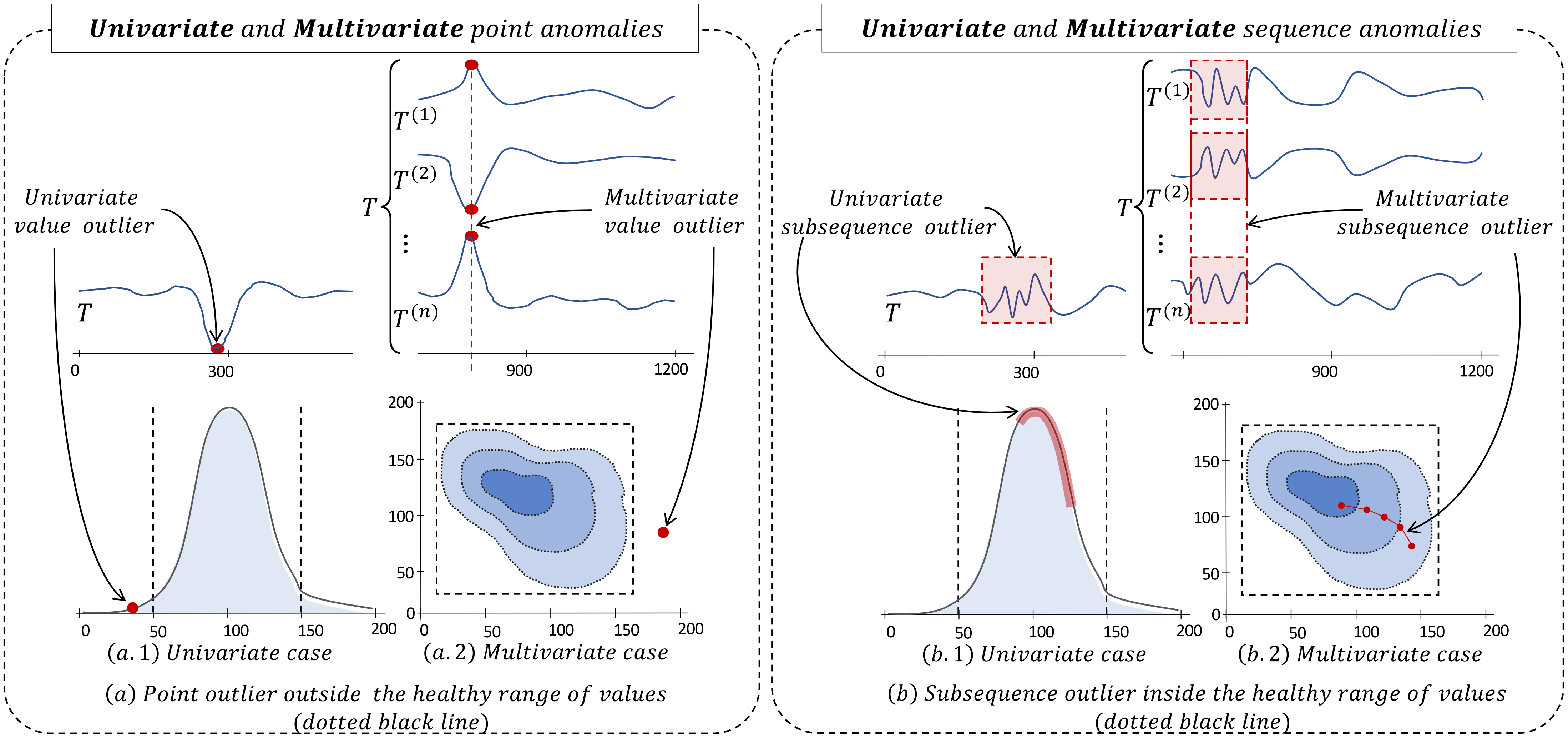
Example of subsequence-based anomaly [2]

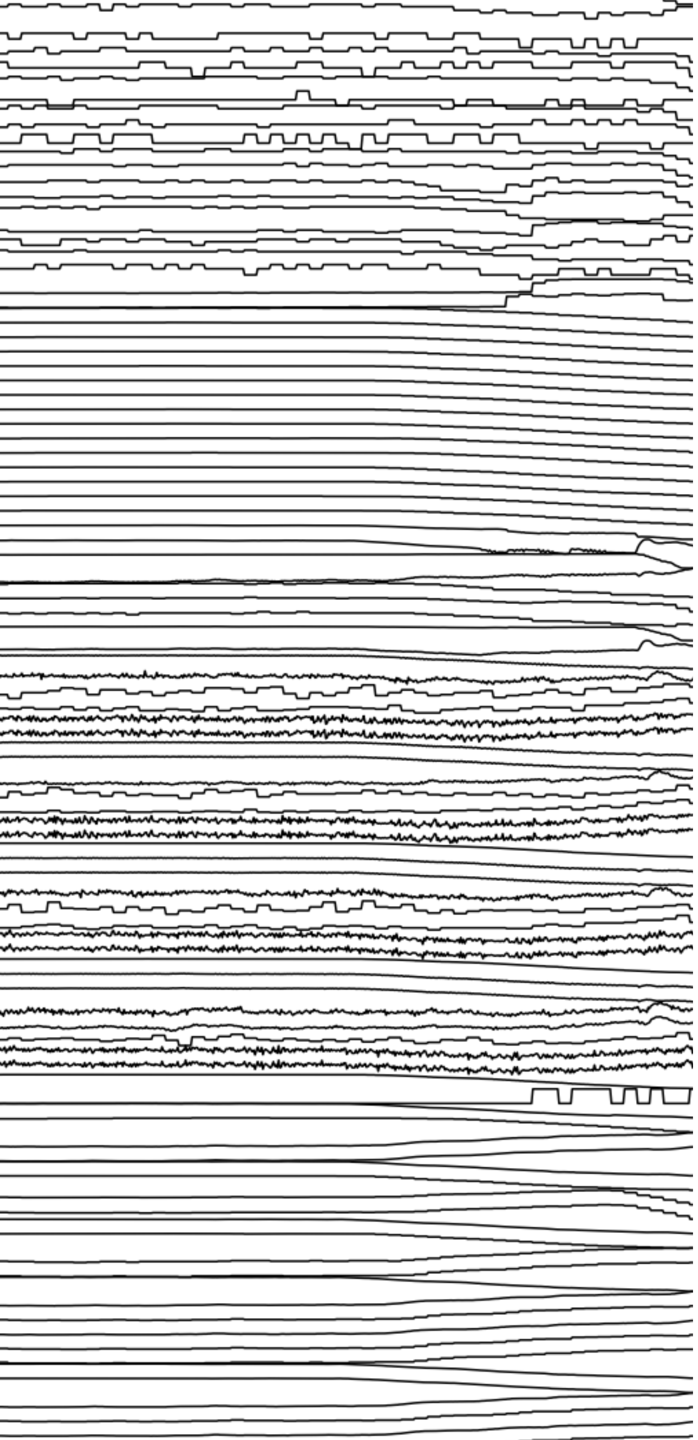


Example of multiple anomaly [4]



Foundations: *Type of anomalies*

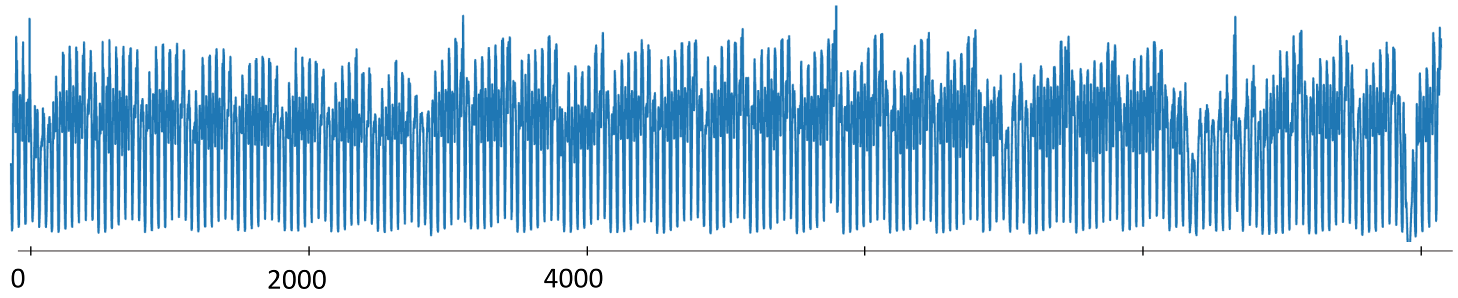




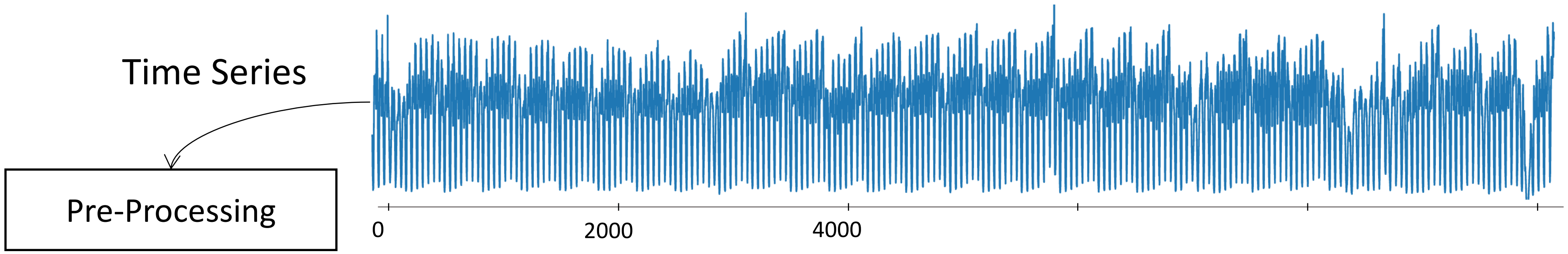
Part 2: Taxonomy of Anomaly Detection Methods

Anomaly Detection methods: *A taxonomy*

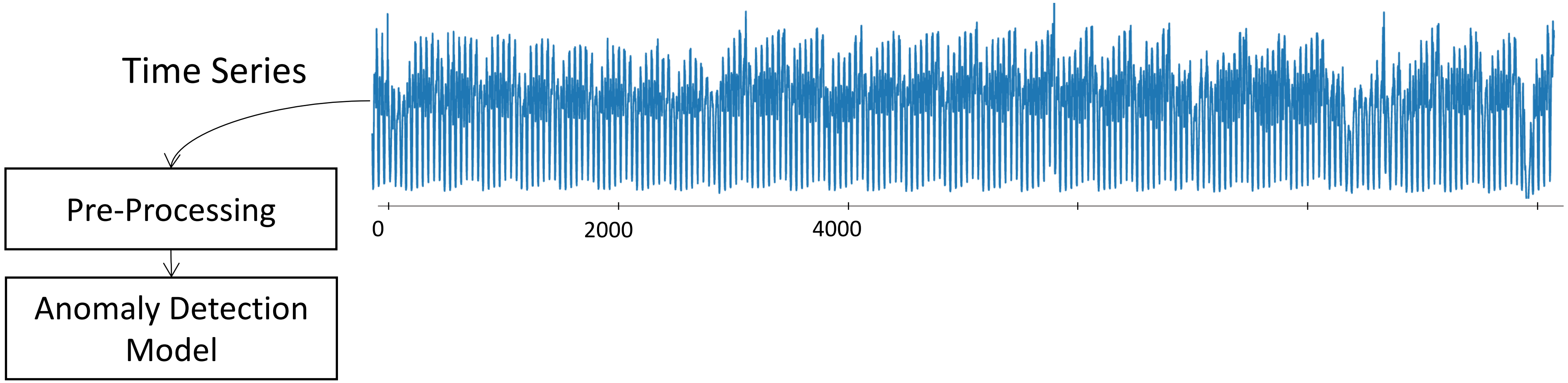
Time Series



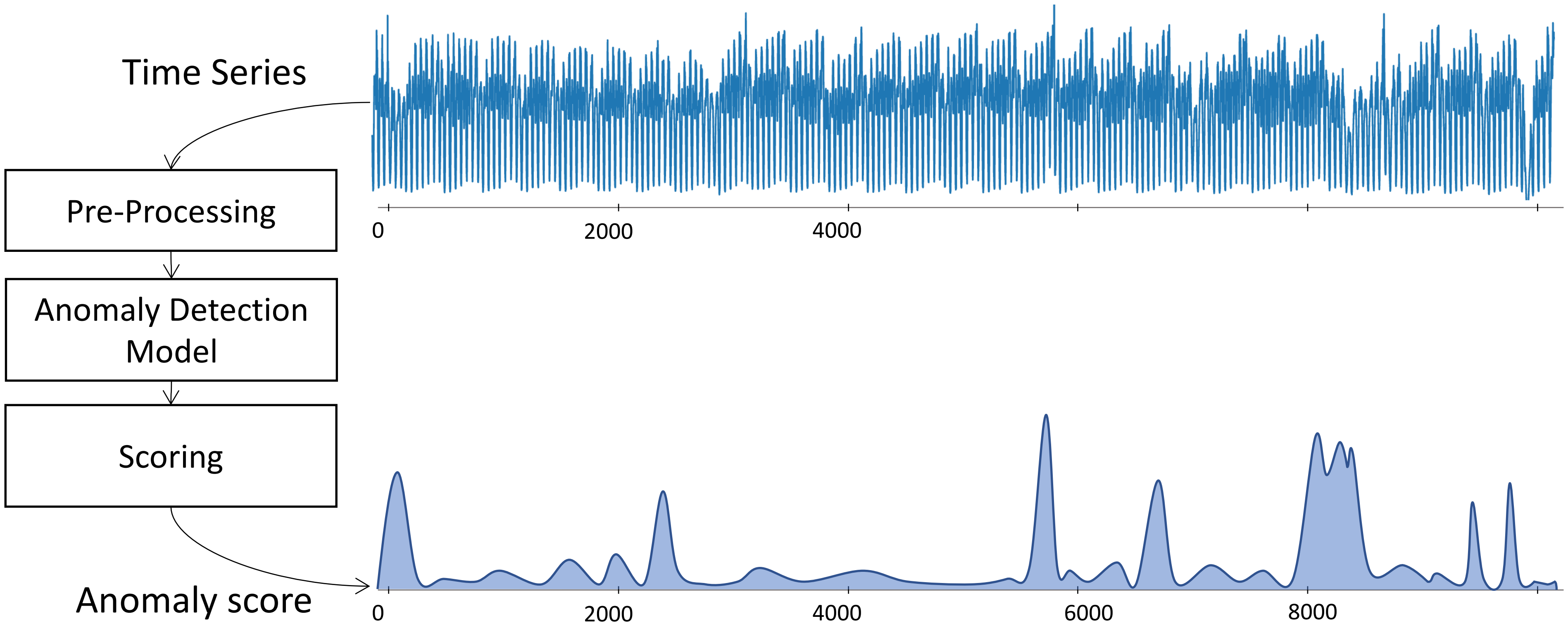
Anomaly Detection methods: *A taxonomy*



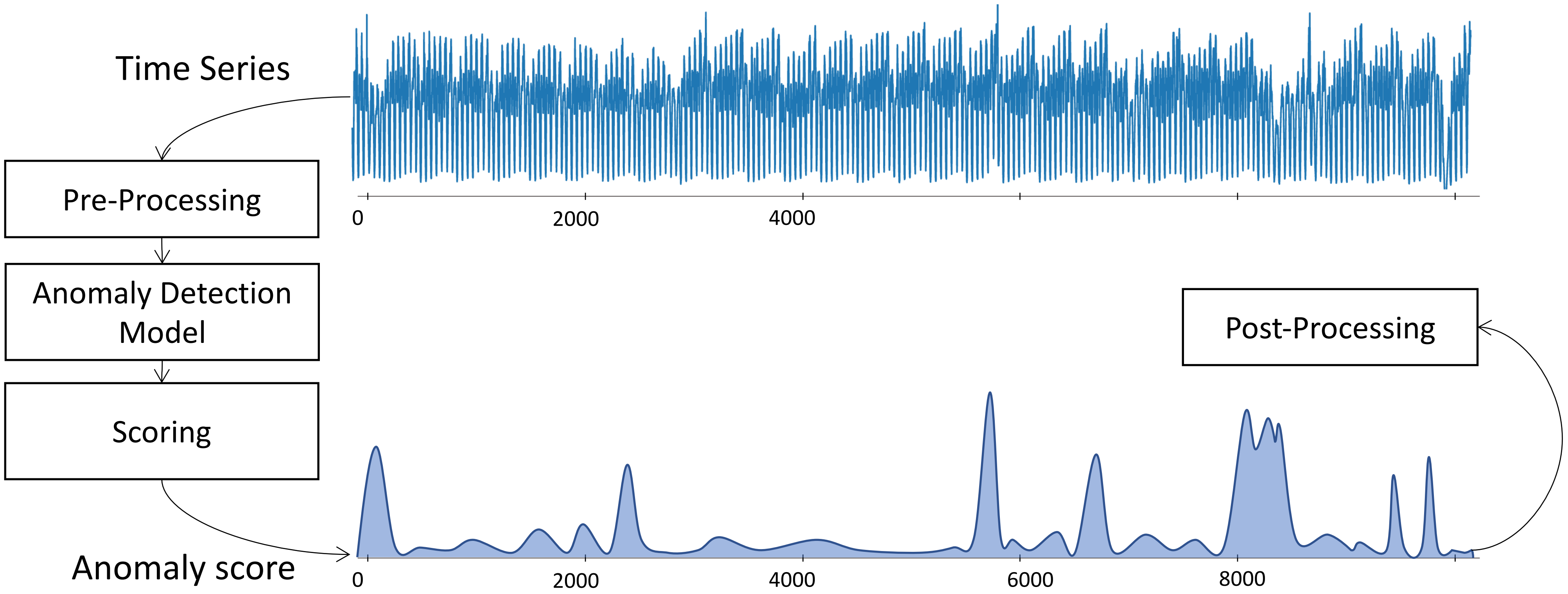
Anomaly Detection methods: *A taxonomy*



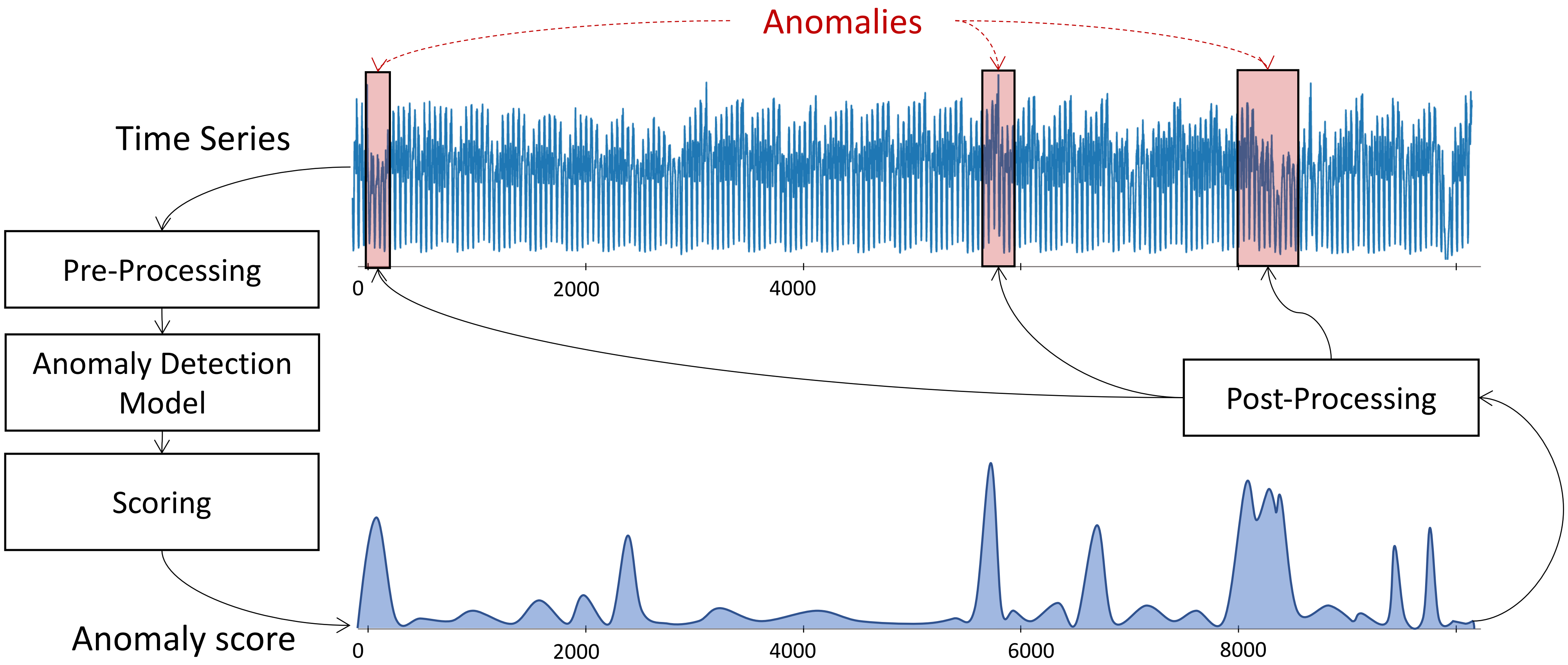
Anomaly Detection methods: *A taxonomy*



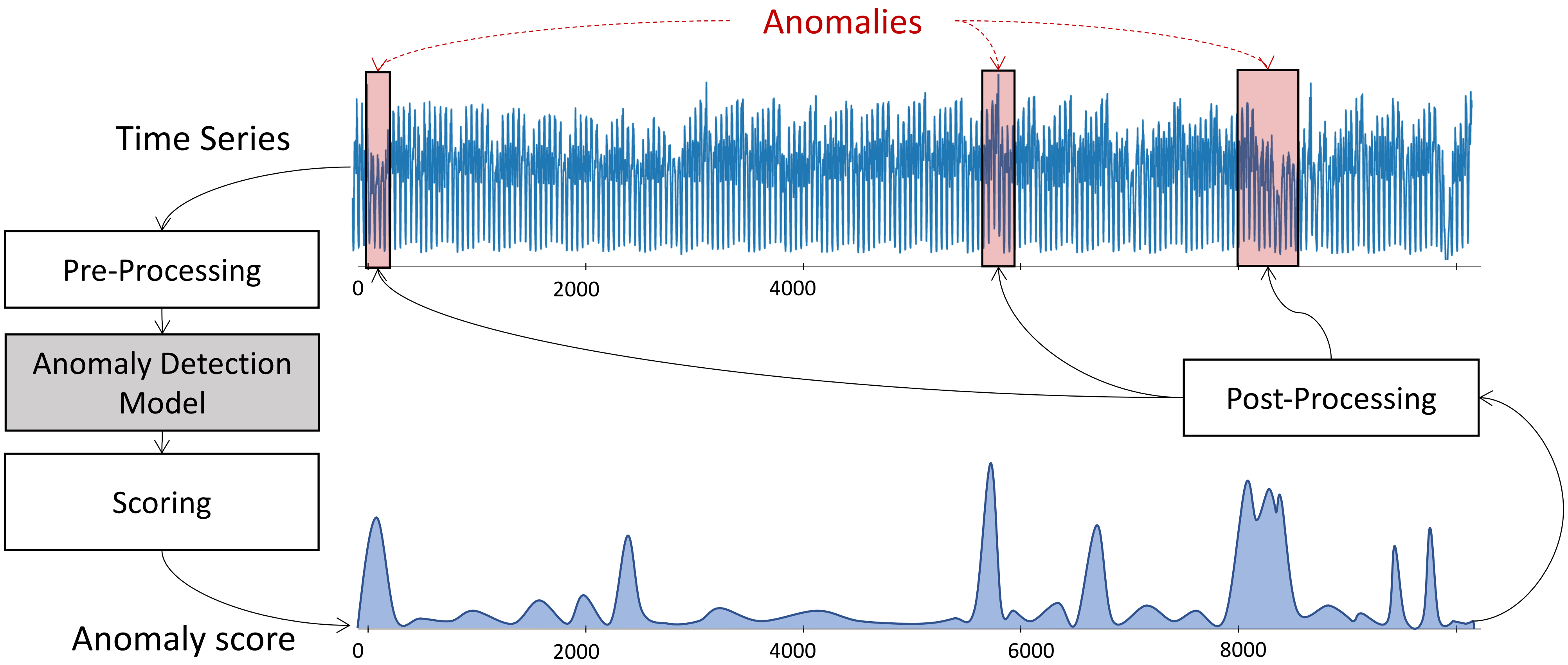
Anomaly Detection methods: *A taxonomy*



Anomaly Detection methods: *A taxonomy*

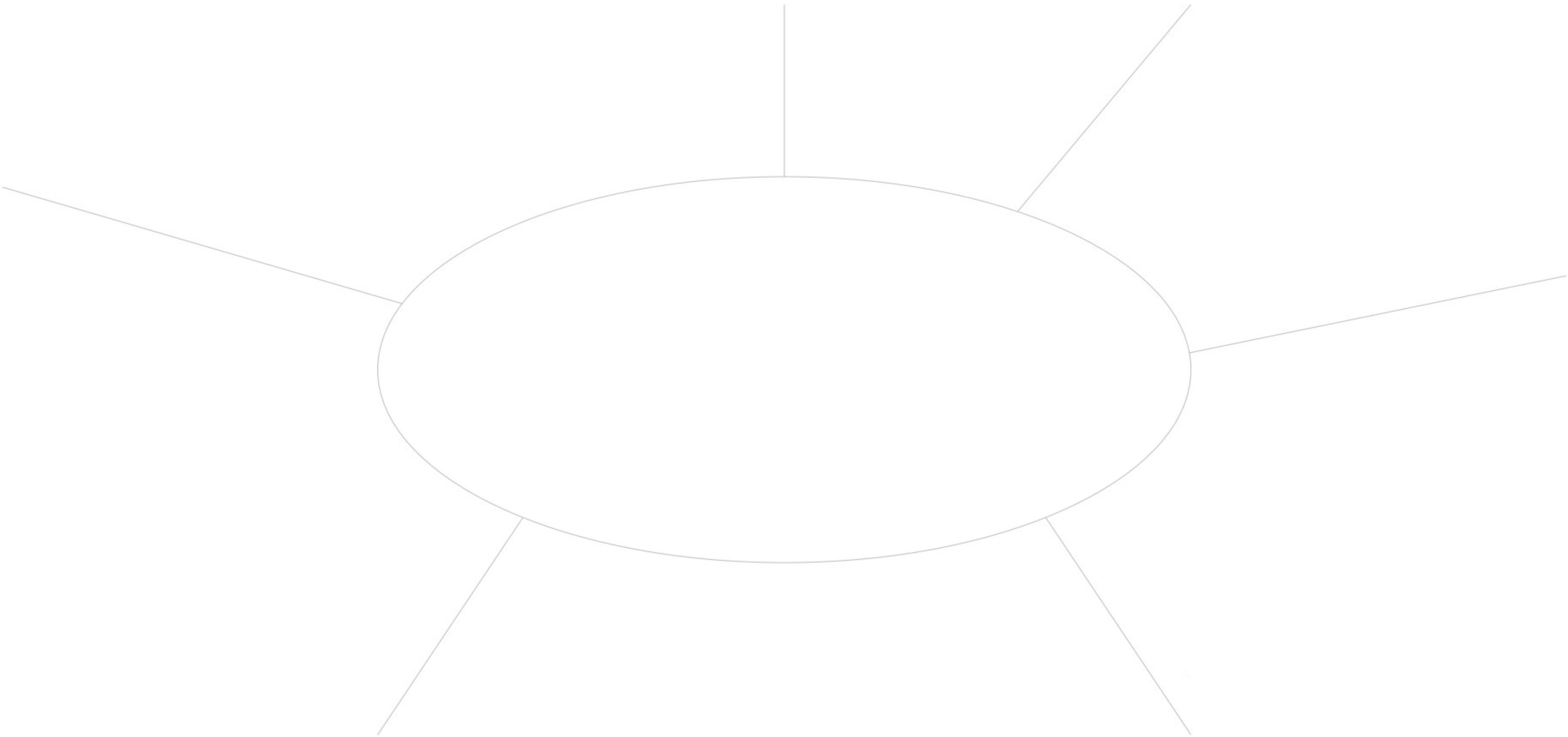


Anomaly Detection methods: *A taxonomy*



Anomaly Detection methods: *A taxonomy*

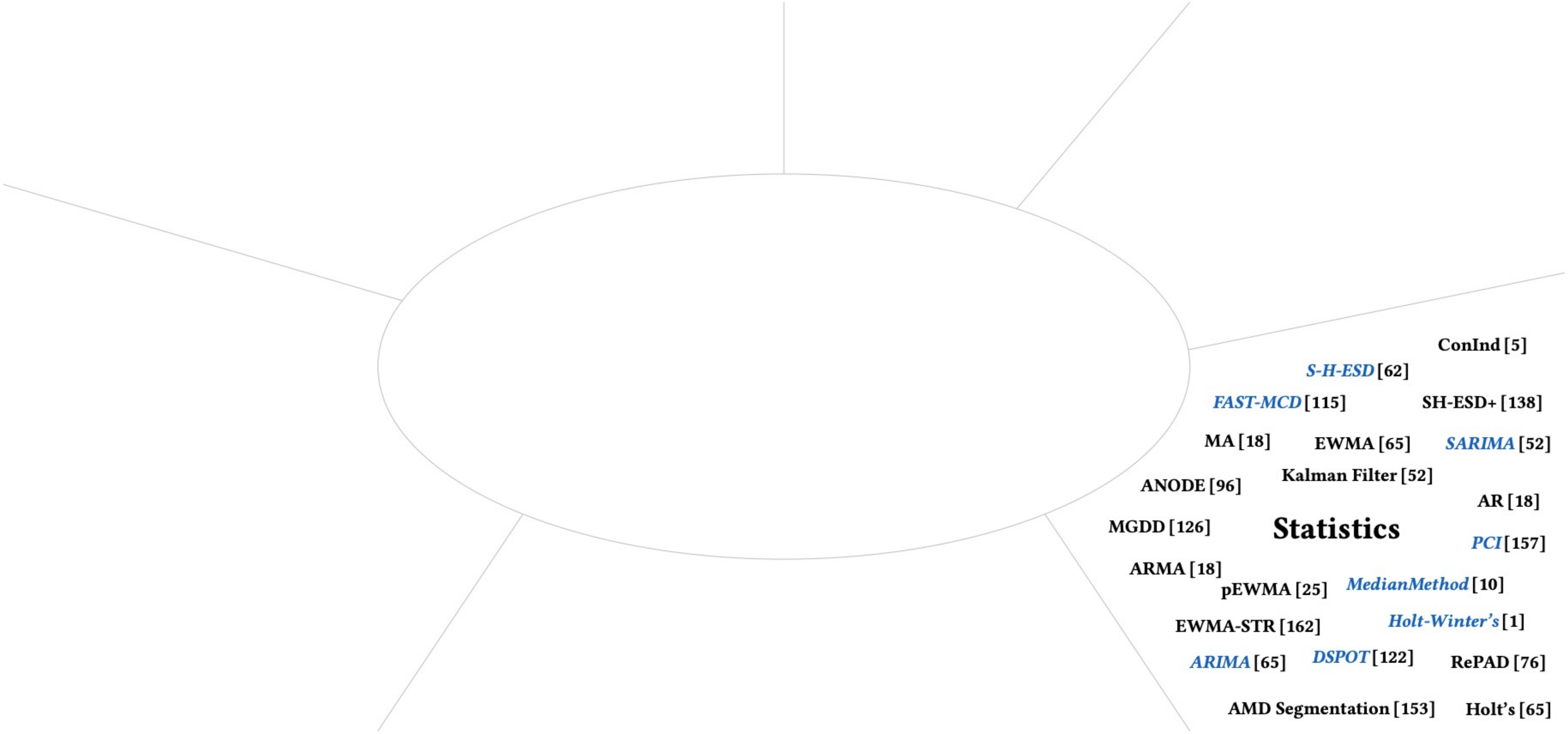
By domains [5] …



[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: *A taxonomy*

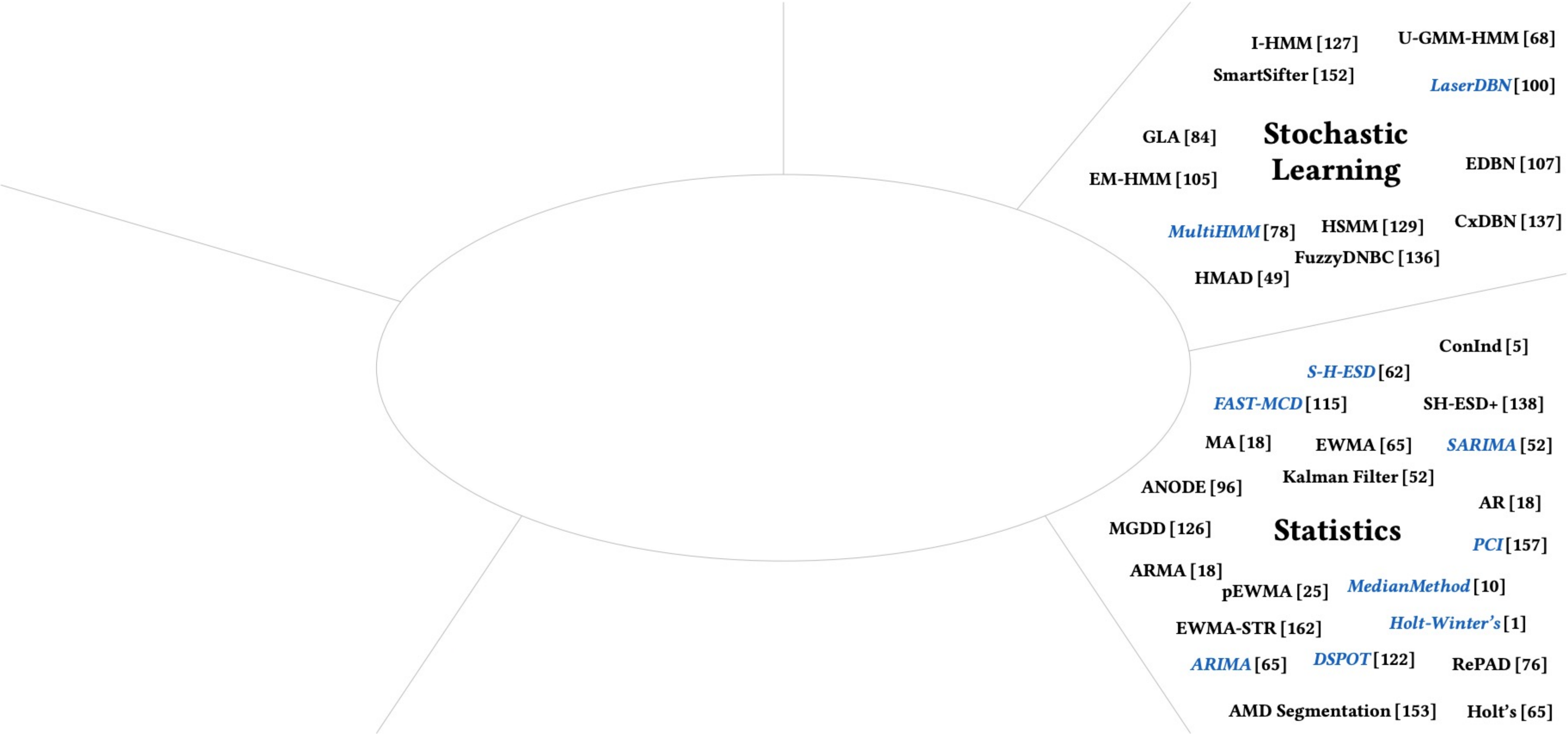
By domains [5] …



[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: *A taxonomy*

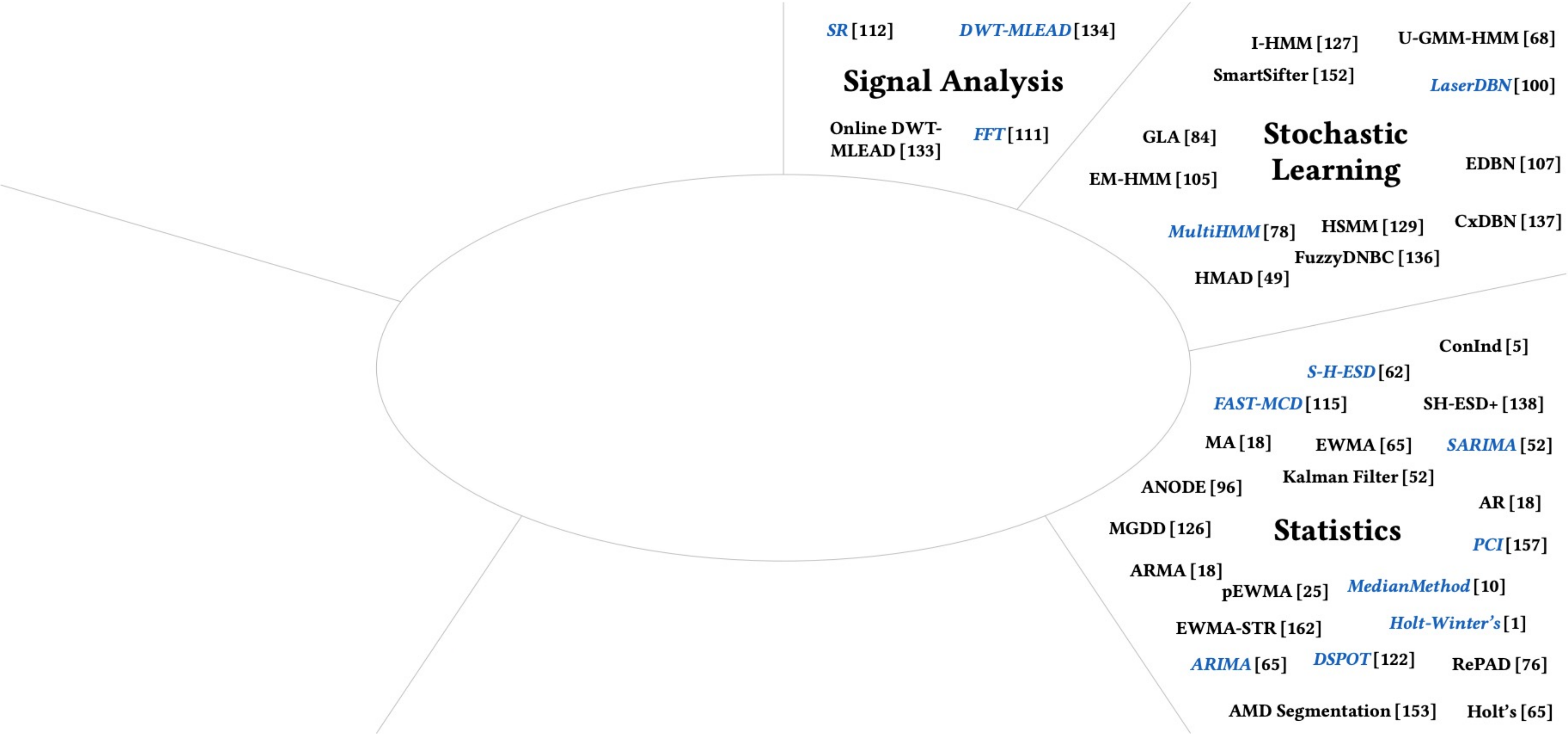
By domains [5] …



[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: *A taxonomy*

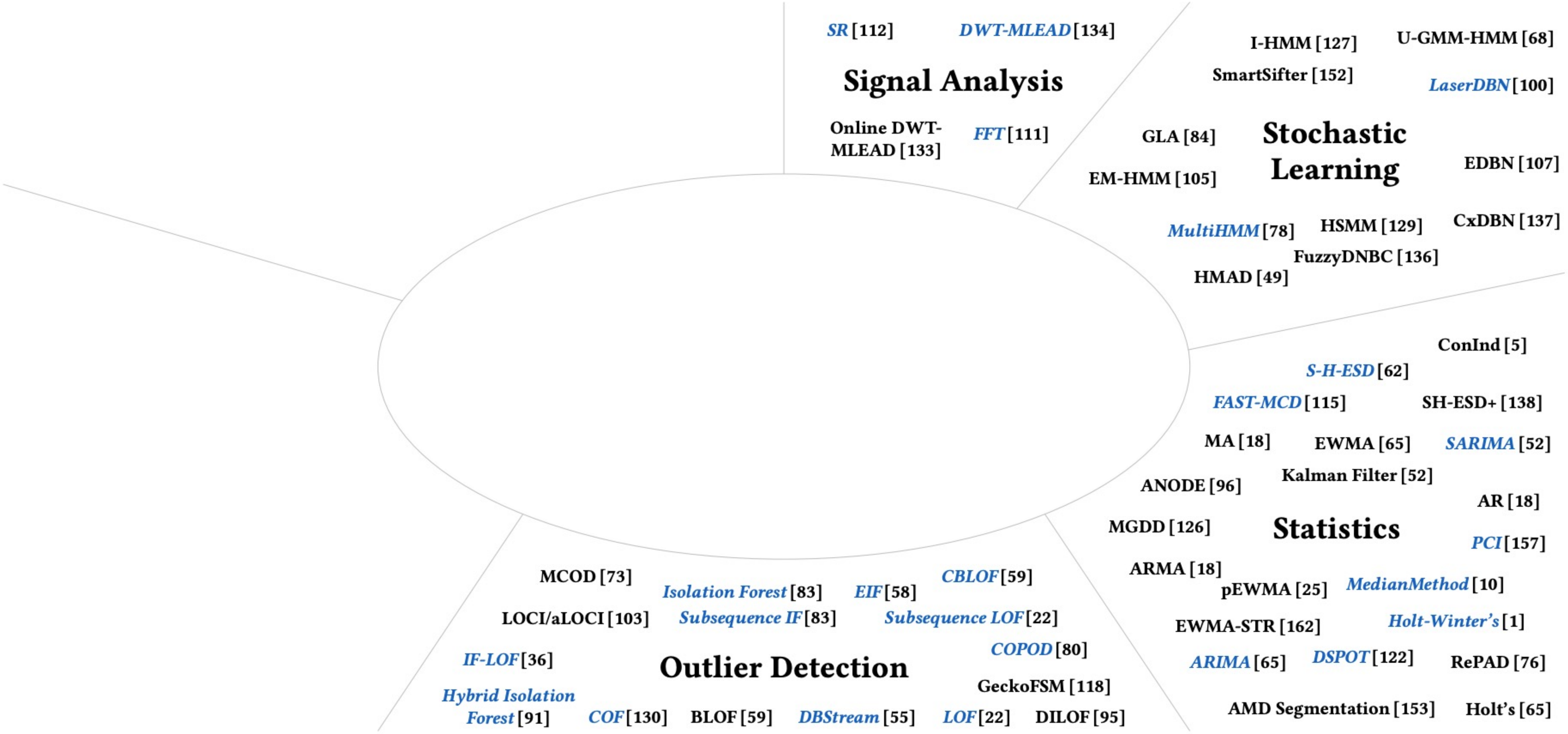
By domains [5] …



[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: *A taxonomy*

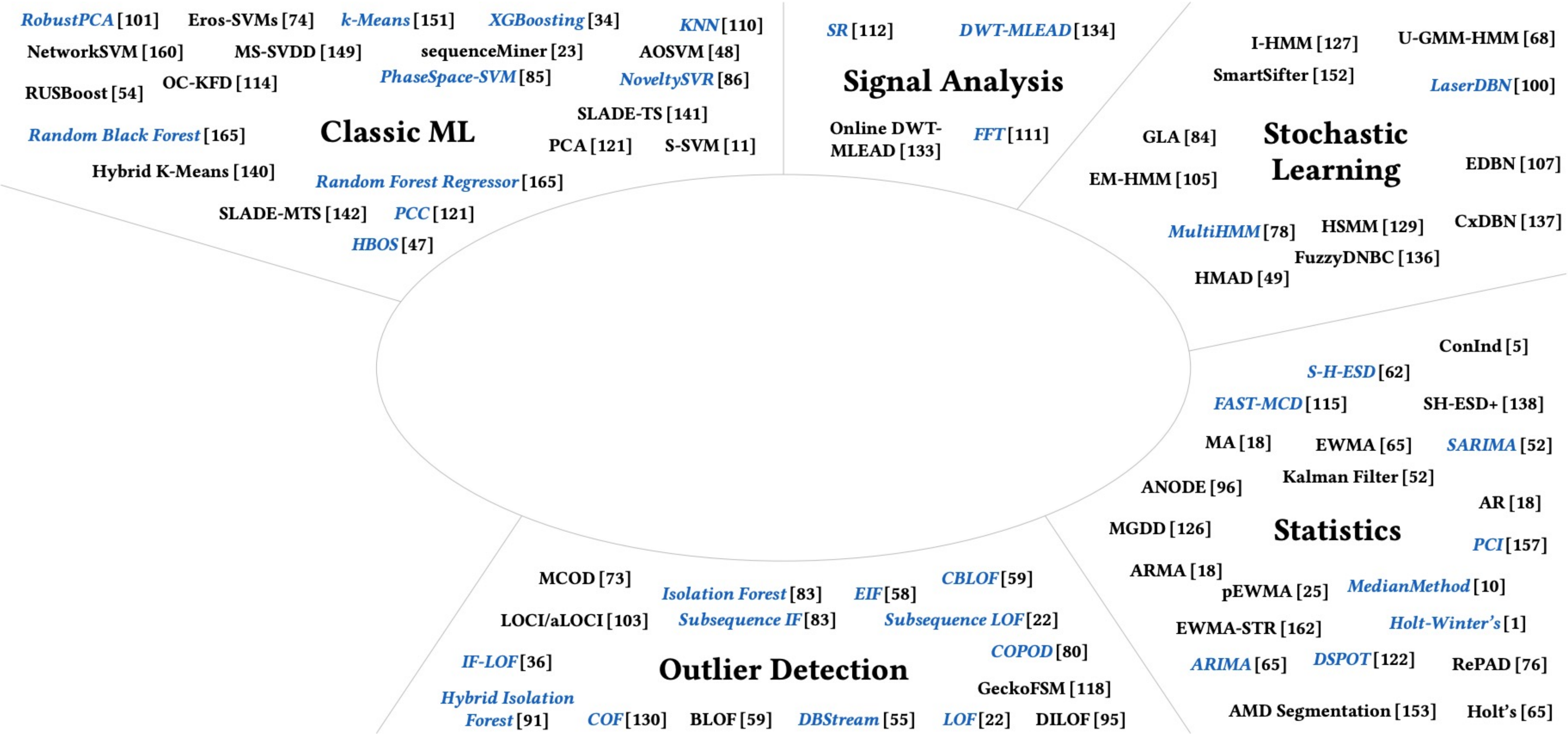
By domains [5] …



[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: *A taxonomy*

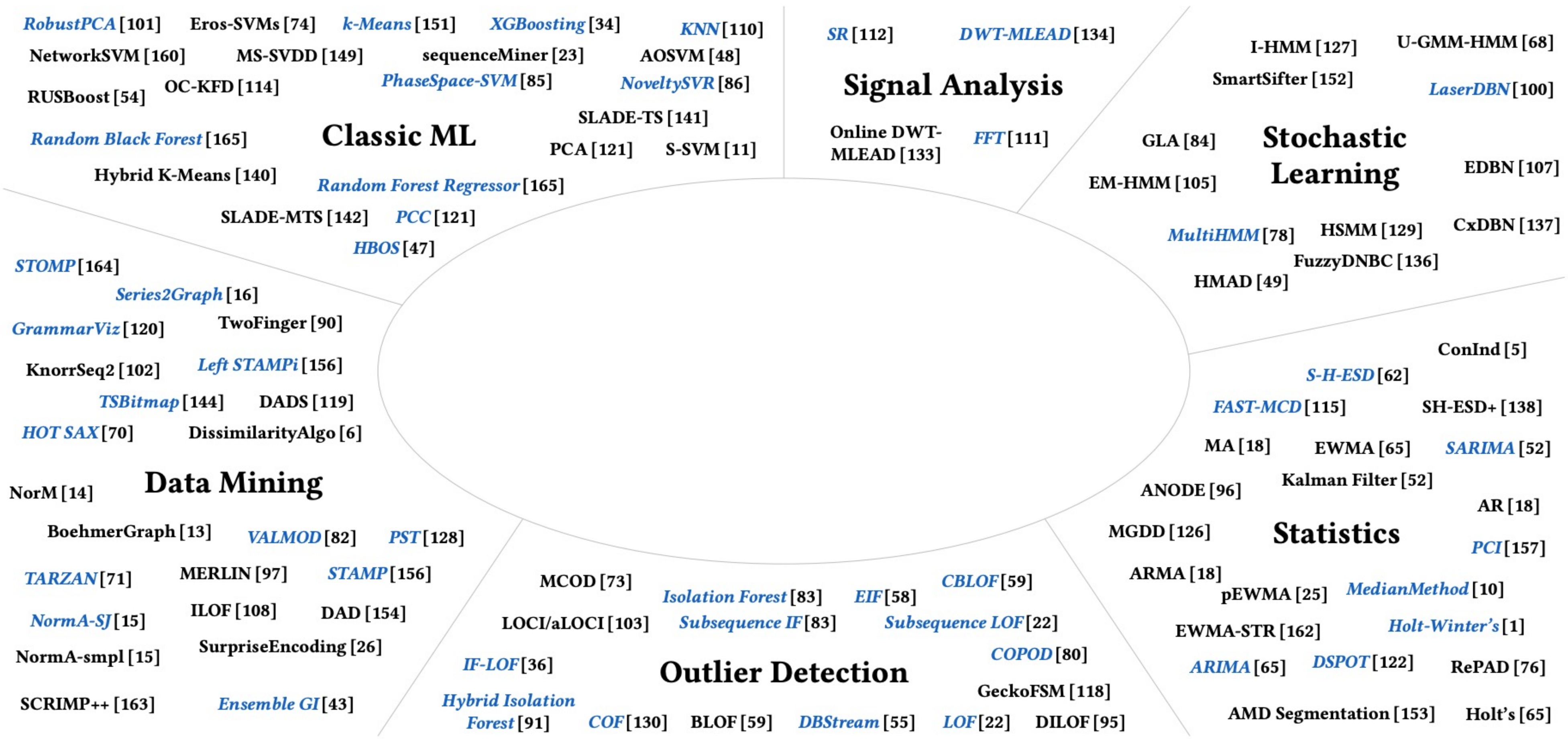
By domains [5] …



[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: *A taxonomy*

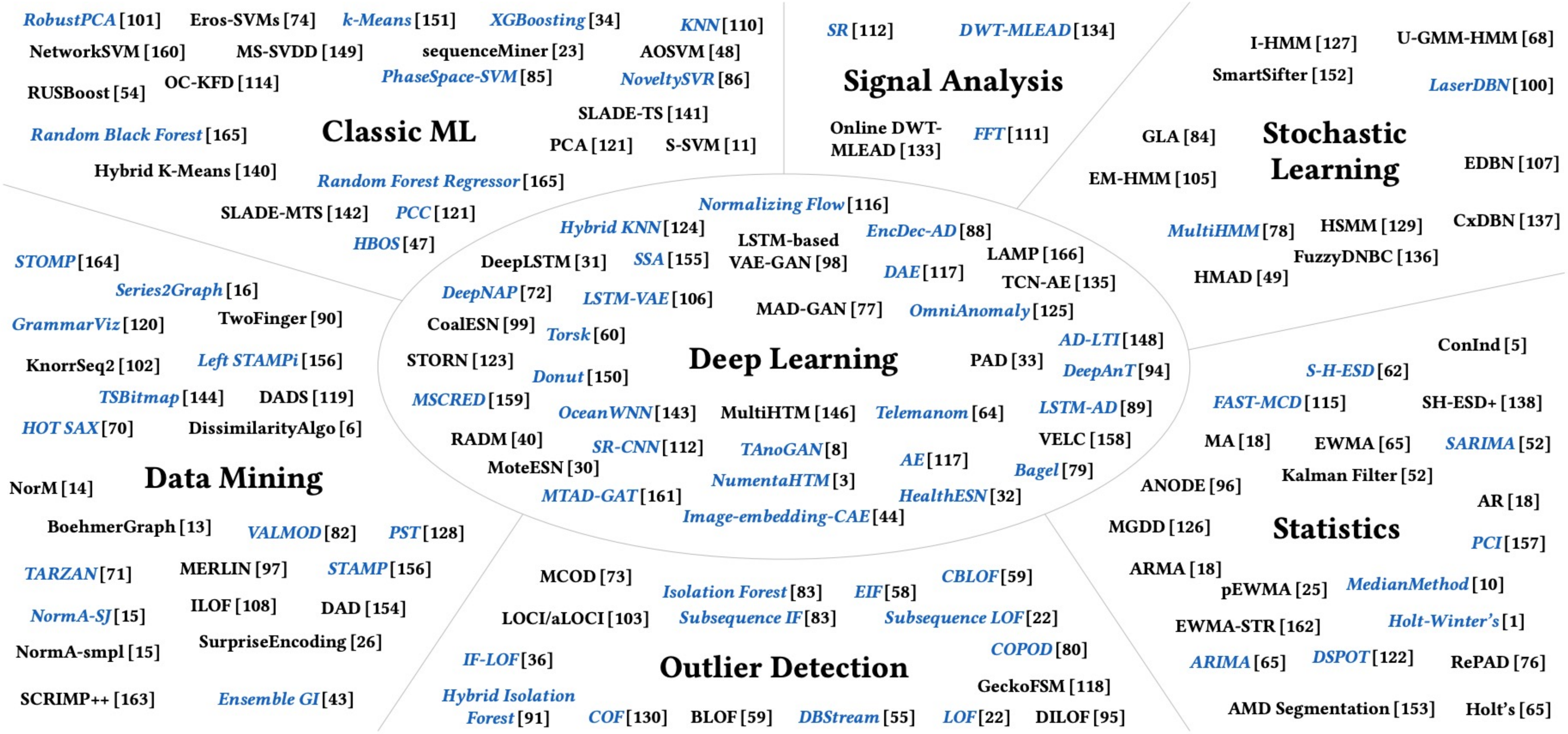
By domains [5] …



[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: *A taxonomy*

By domains [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: *A taxonomy*

By inputs…

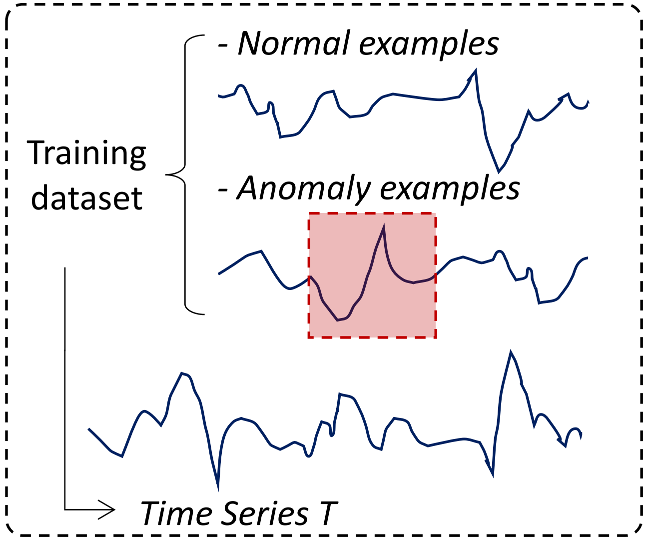
TSAD Methods

Anomaly Detection methods: *A taxonomy*

By inputs…

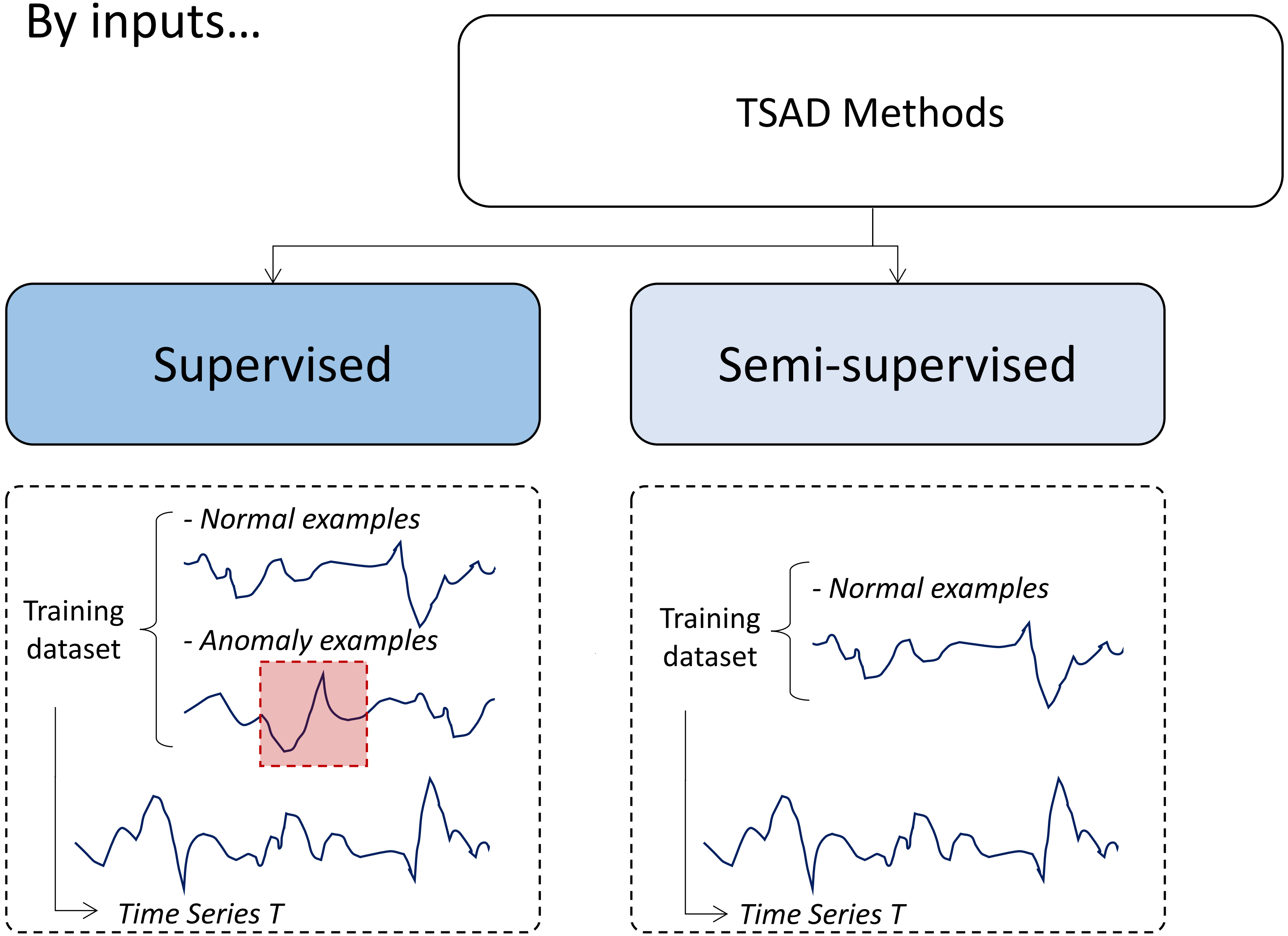
TSAD Methods

Supervised



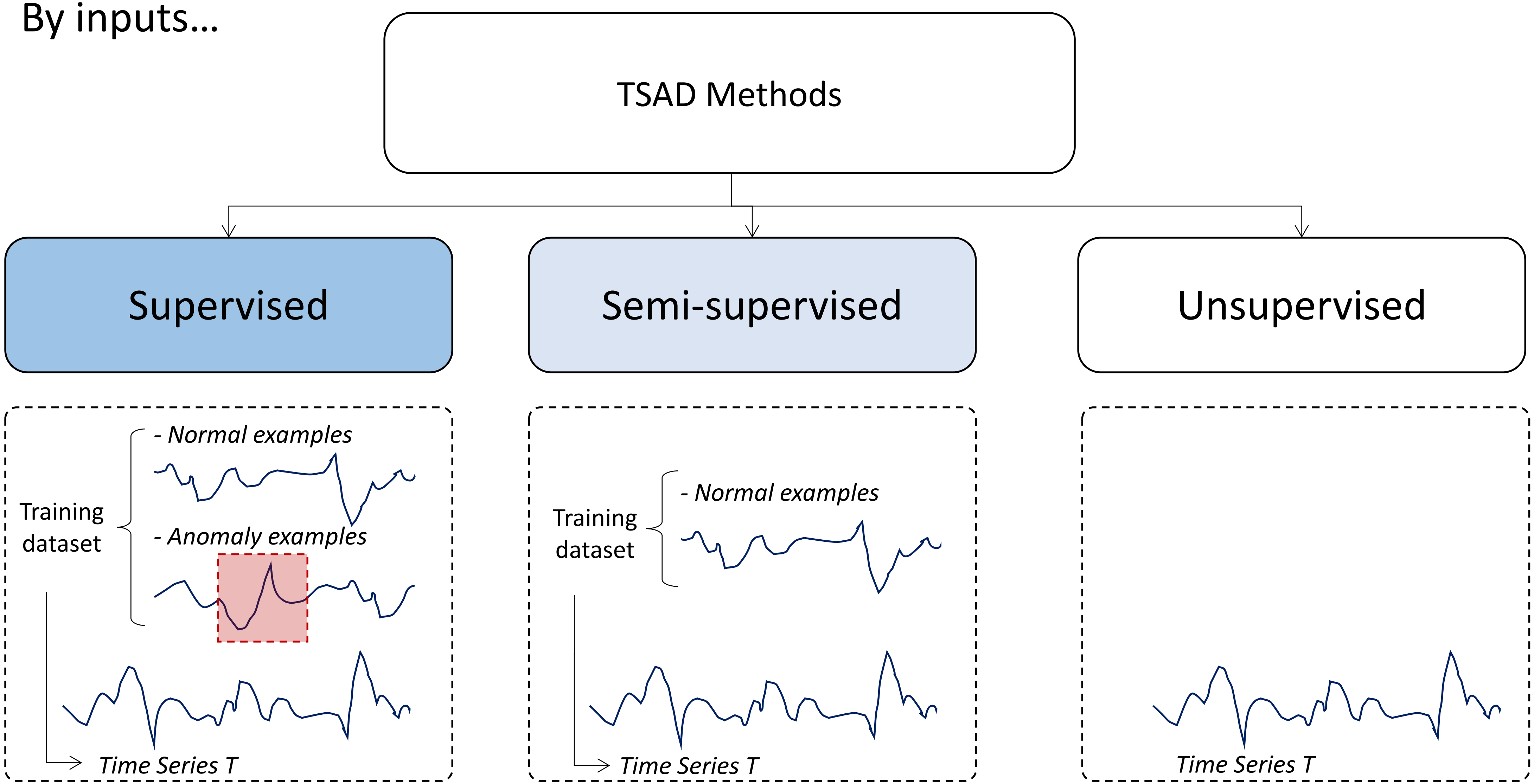
Anomaly Detection methods: *A taxonomy*

By inputs…



Anomaly Detection methods: *A taxonomy*

By inputs…



Anomaly Detection methods: *A taxonomy*

By methods…

TSAD Methods

Anomaly Detection methods: *A taxonomy*

By methods…

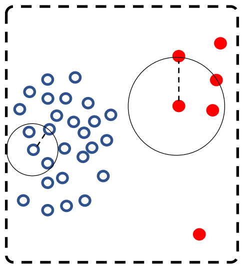
Anomaly Detection methods: *A taxonomy*

By methods…

TSAD Methods

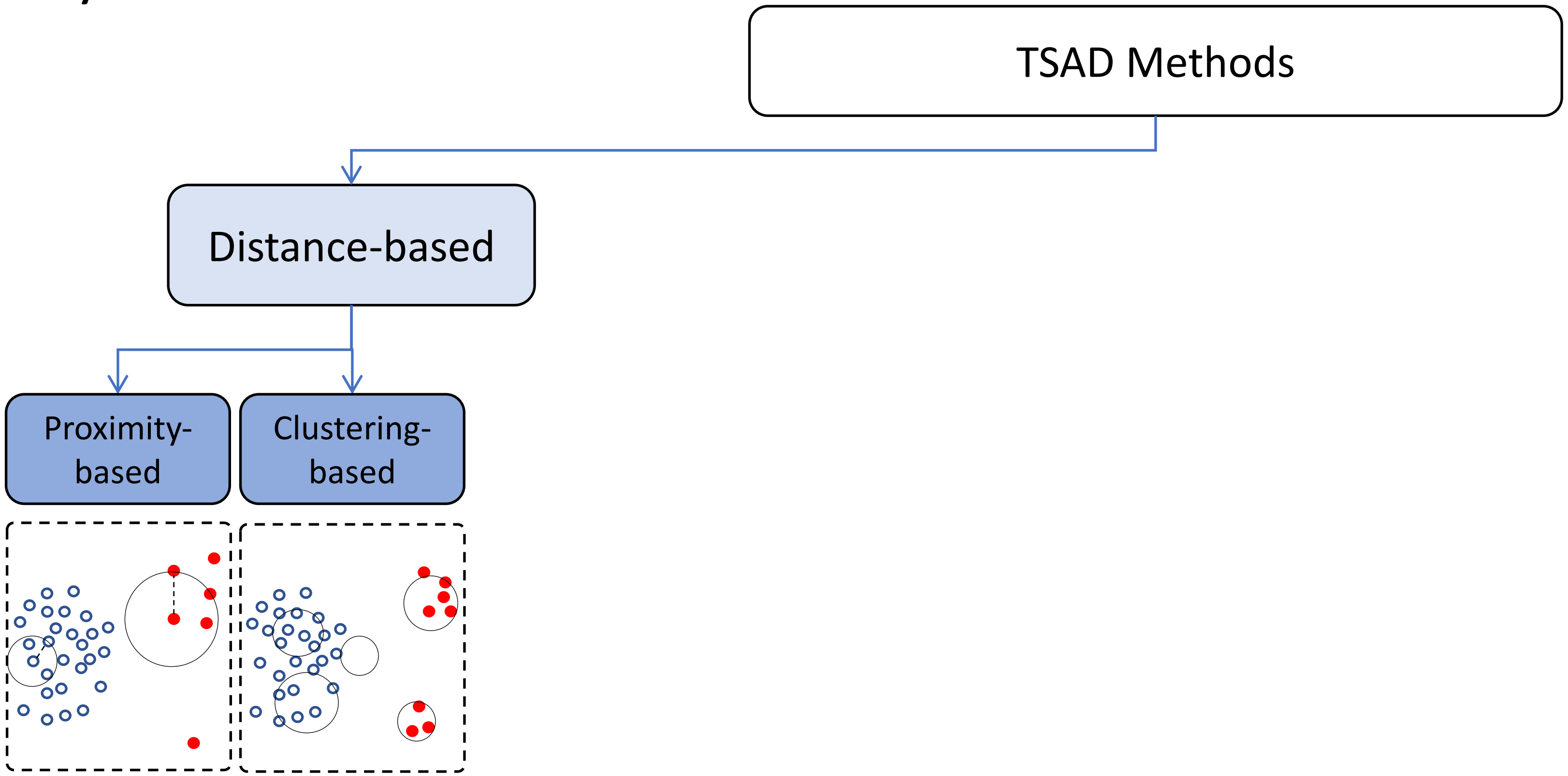
Distance-based

Proximity-based



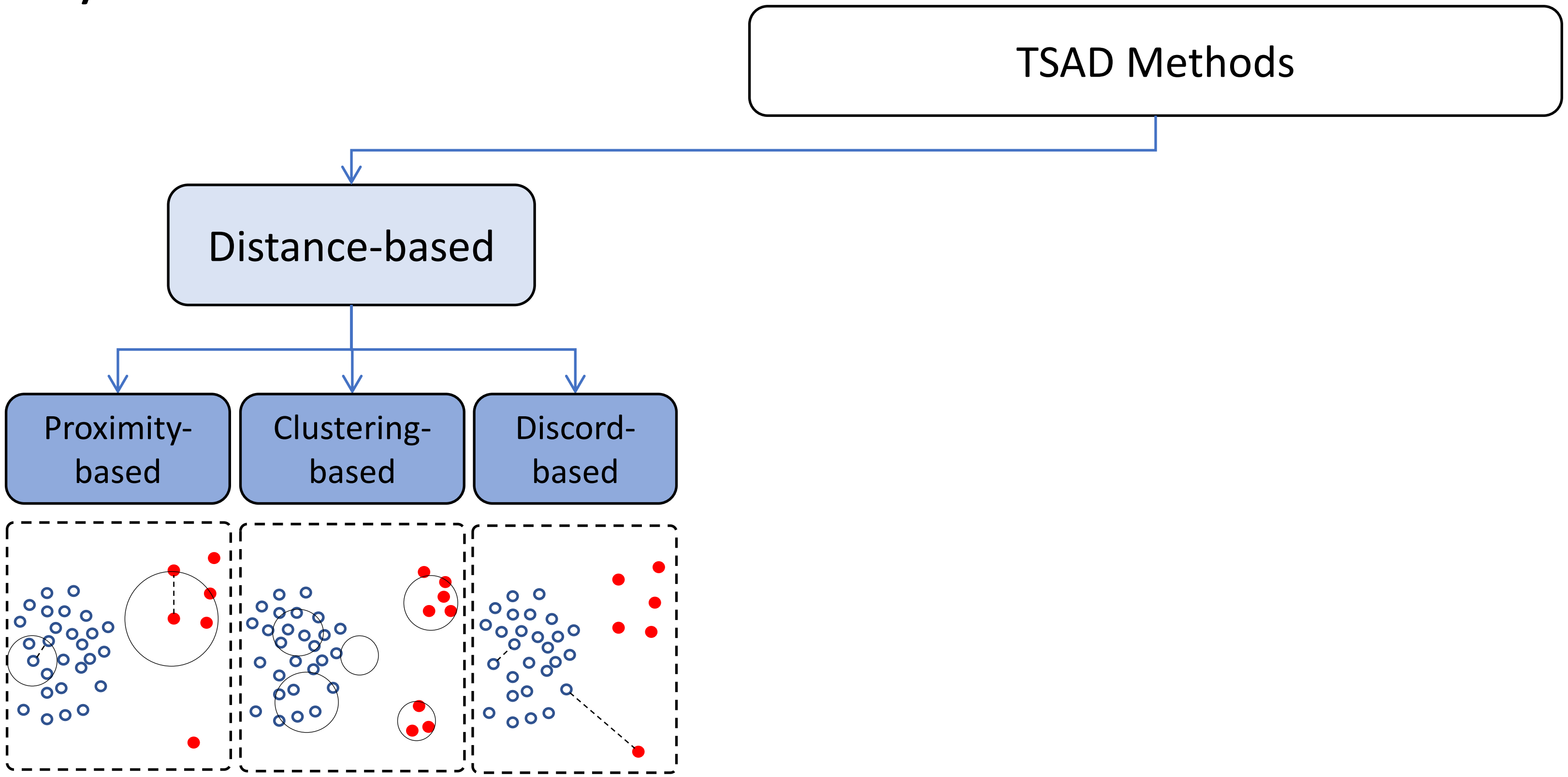
Anomaly Detection methods: *A taxonomy*

By methods…



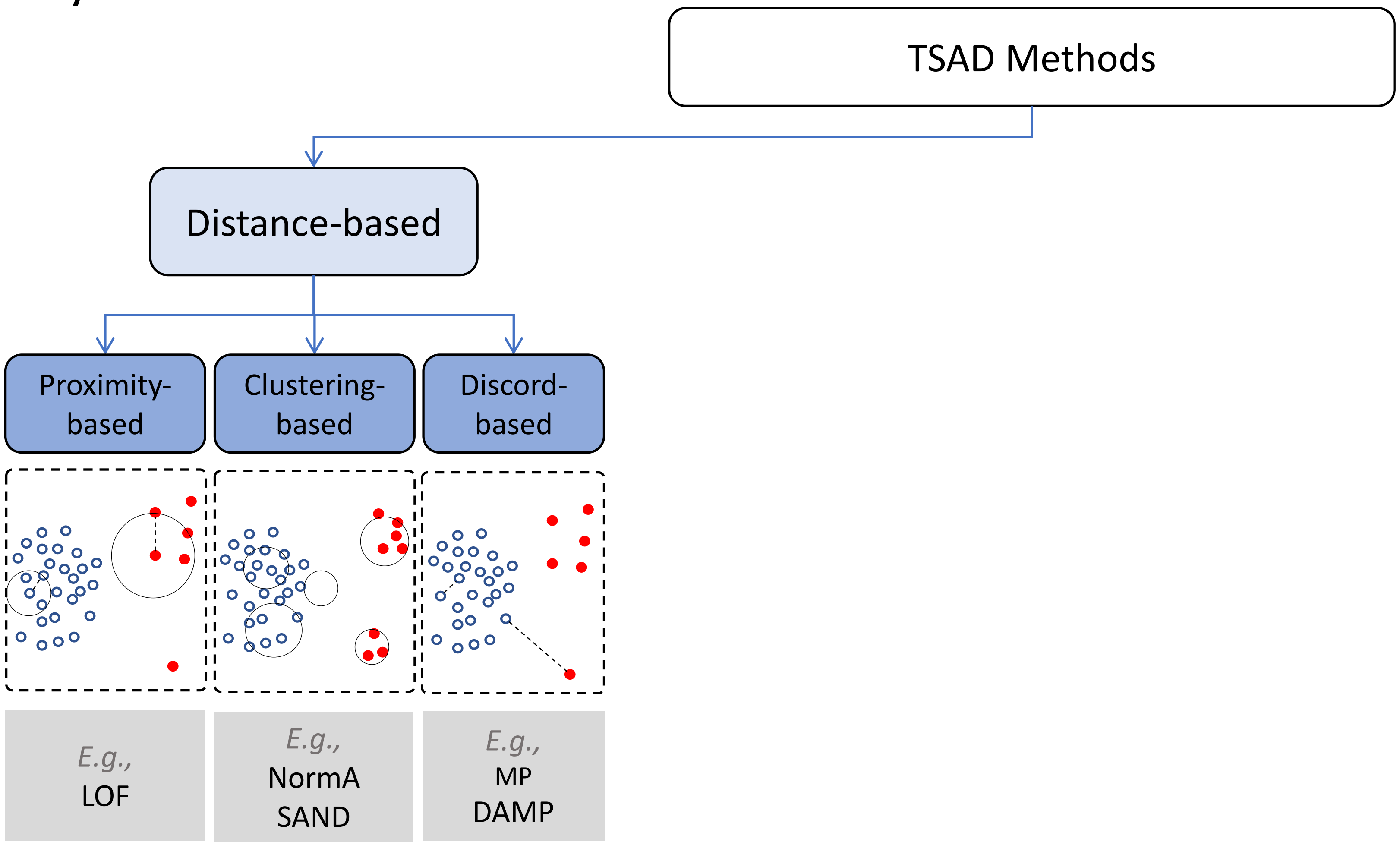
Anomaly Detection methods: *A taxonomy*

By methods…



Anomaly Detection methods: *A taxonomy*

By methods…

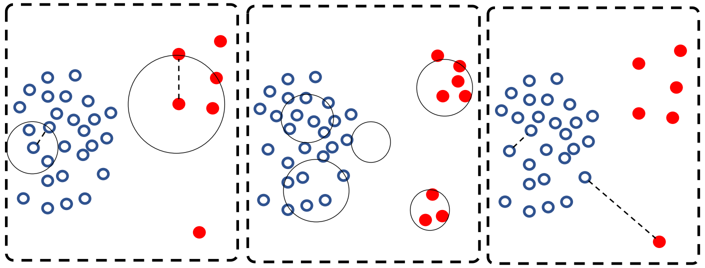


Anomaly Detection methods: *A taxonomy*

By methods…

TSAD Methods

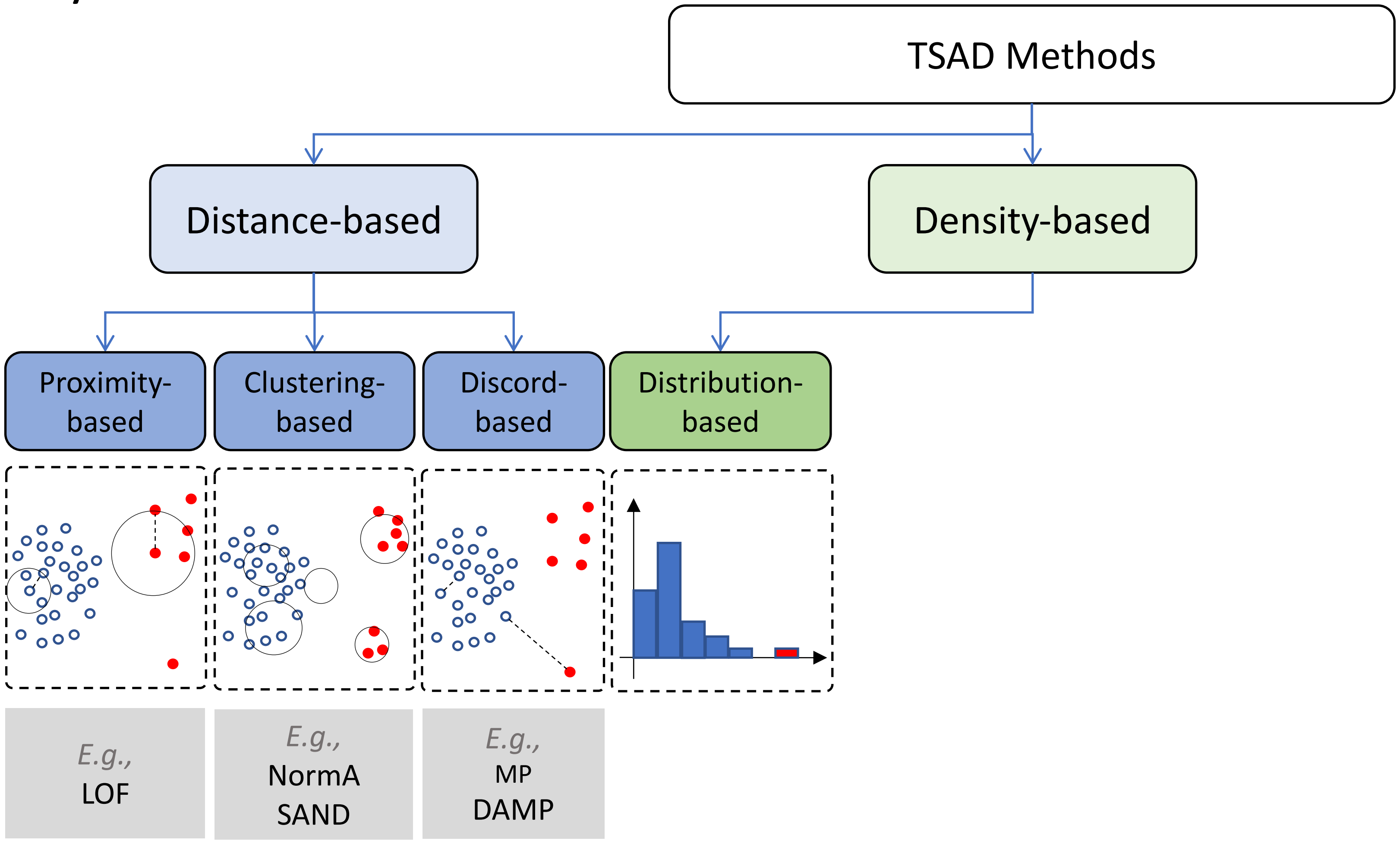
- Distance-based
  - Proximity-based
  - Clustering-based
  - Discord-based
- Density-based



Proximity-based	Clustering-based	Discord-based
E.g., LOF	*E.g.,* NormA SAND	*E.g.,* MP DAMP

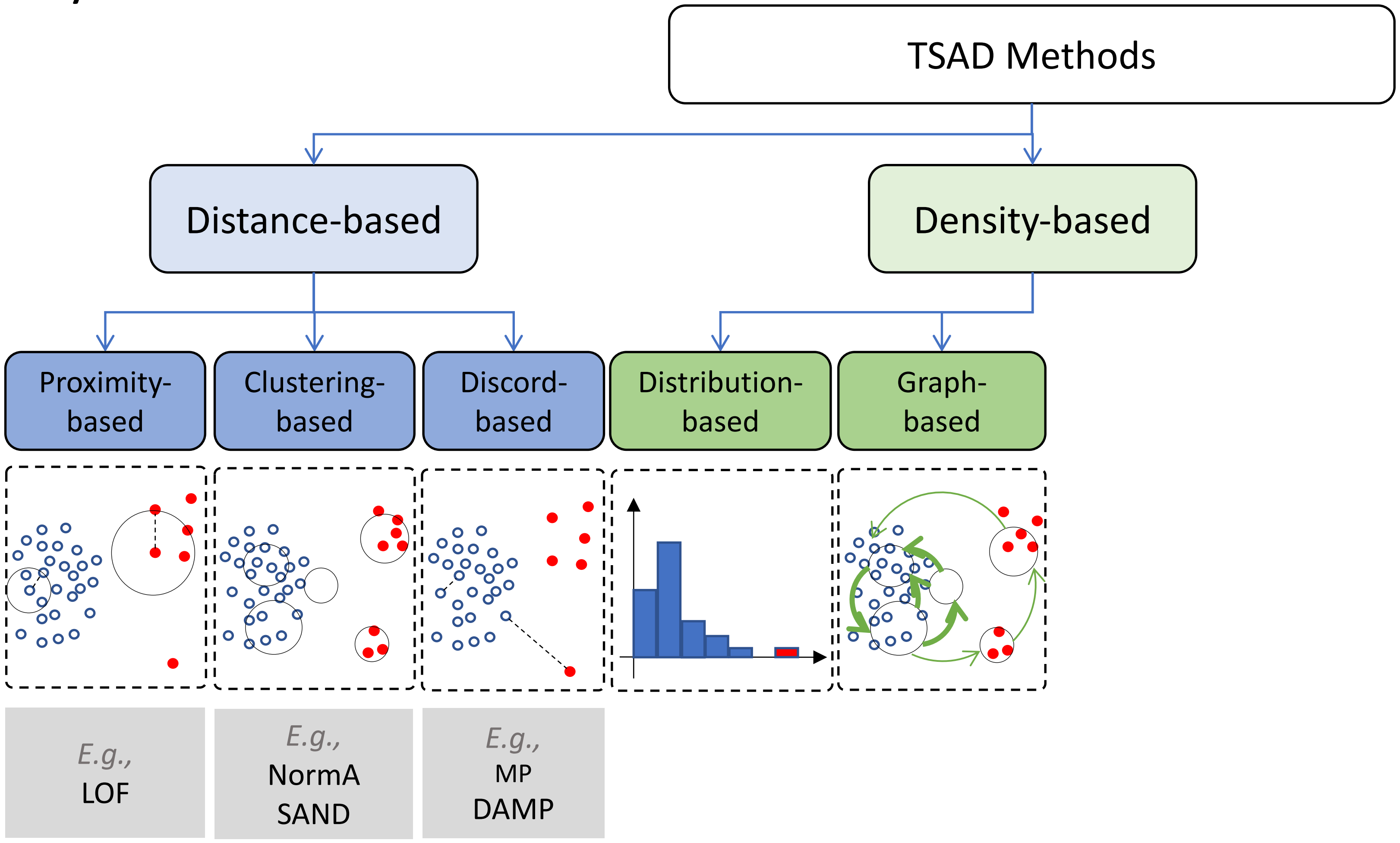
Anomaly Detection methods: *A taxonomy*

By methods…



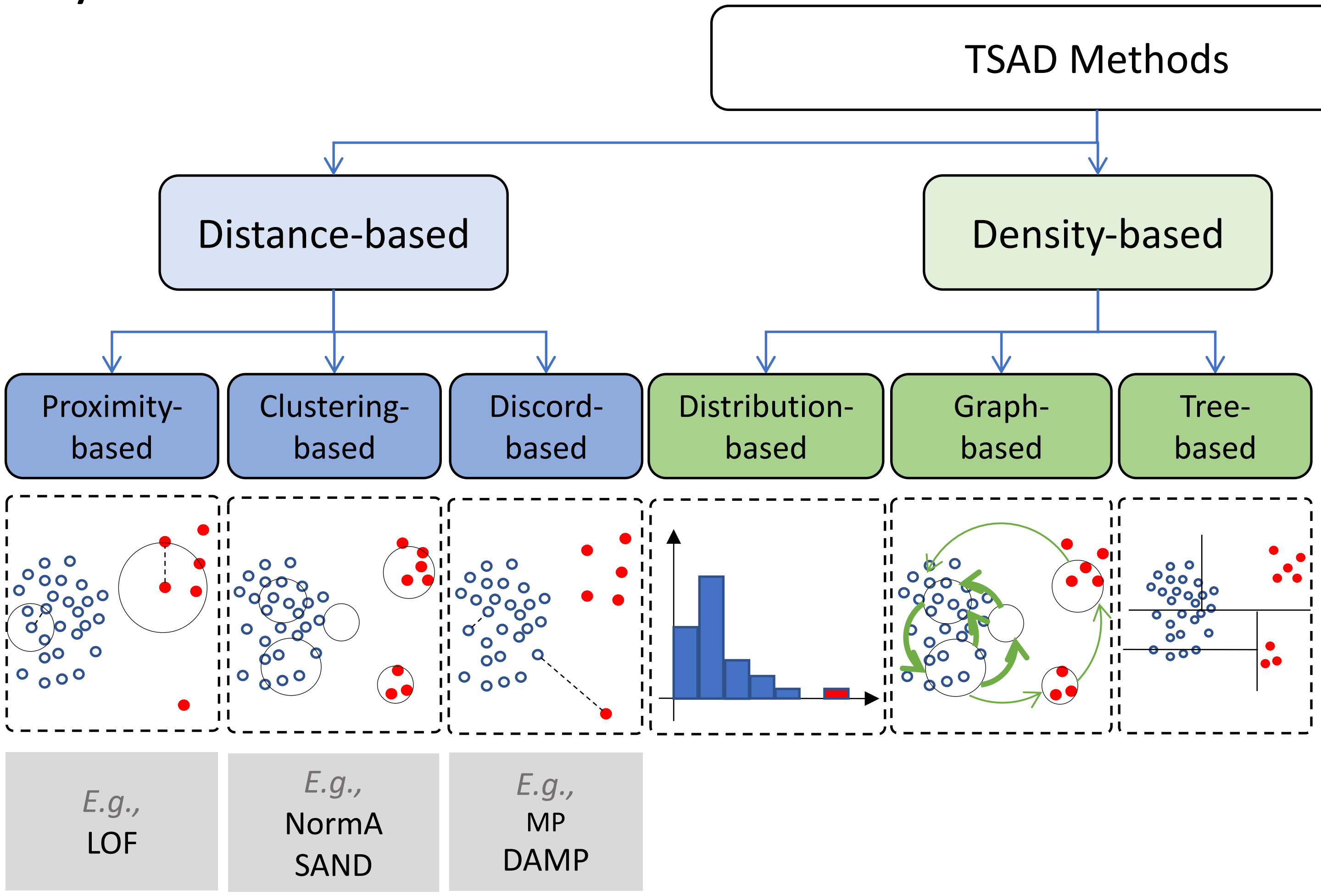
Anomaly Detection methods: *A taxonomy*

By methods…



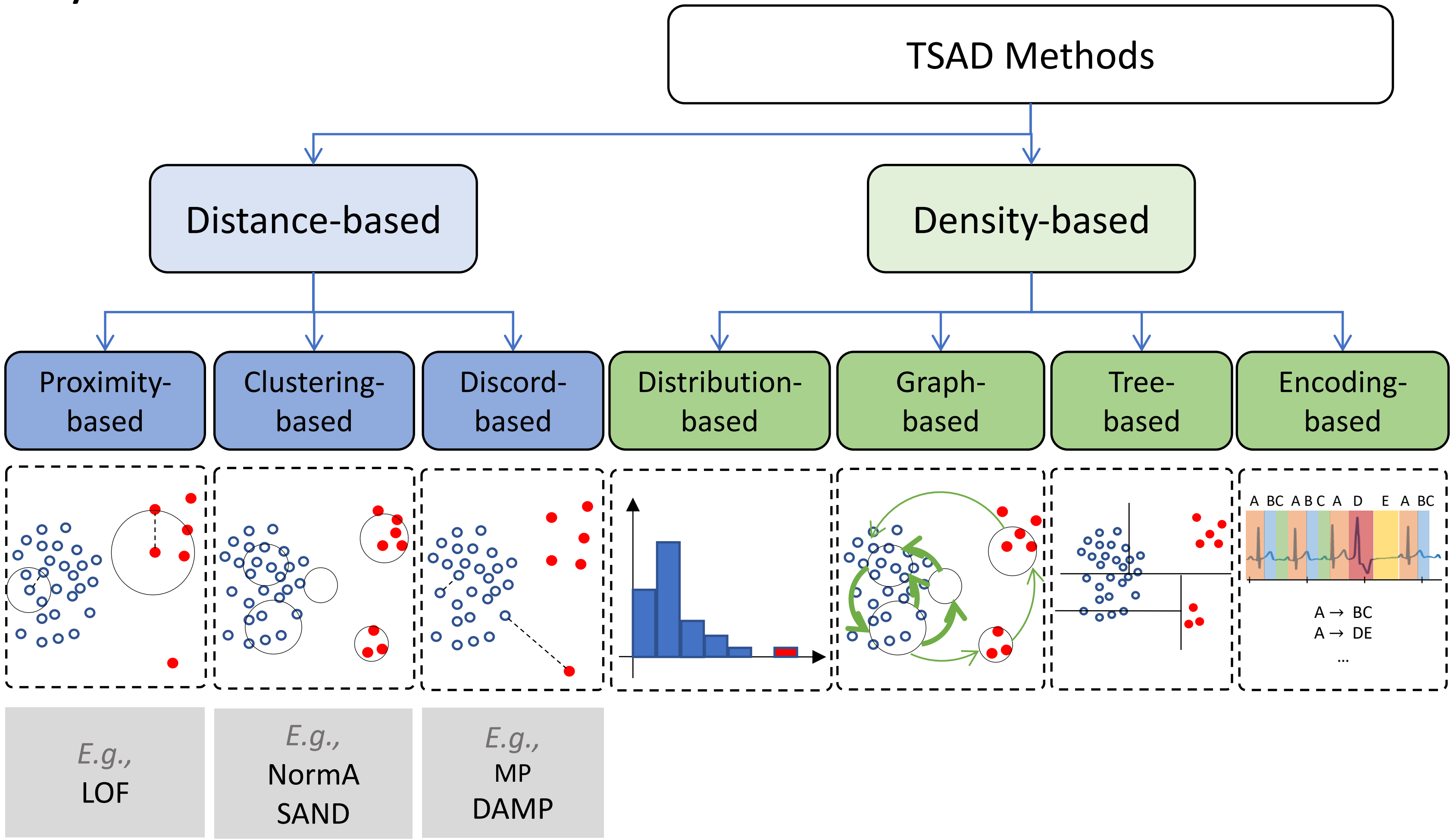
Anomaly Detection methods: *A taxonomy*

By methods…



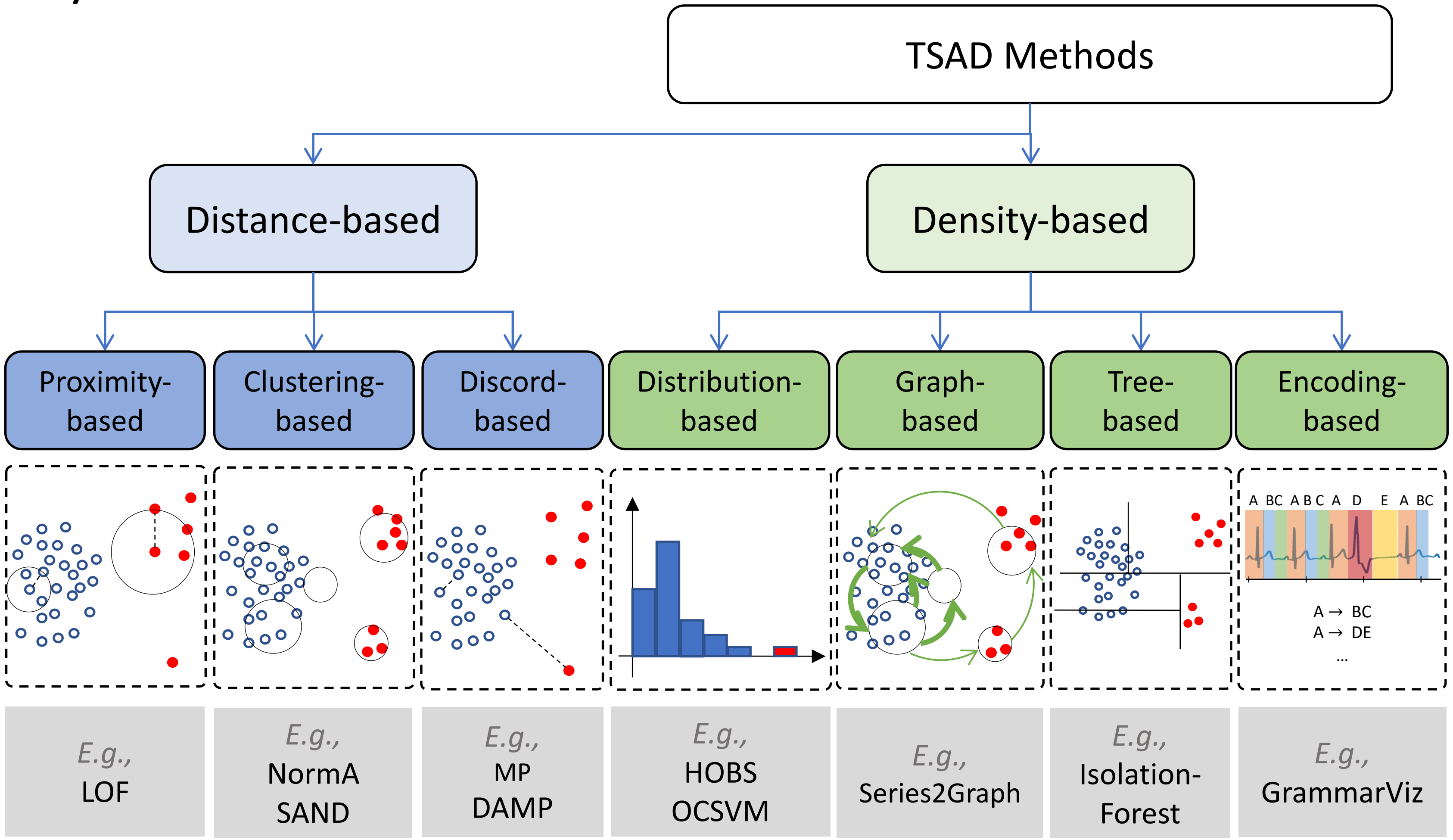
Anomaly Detection methods: *A taxonomy*

By methods…



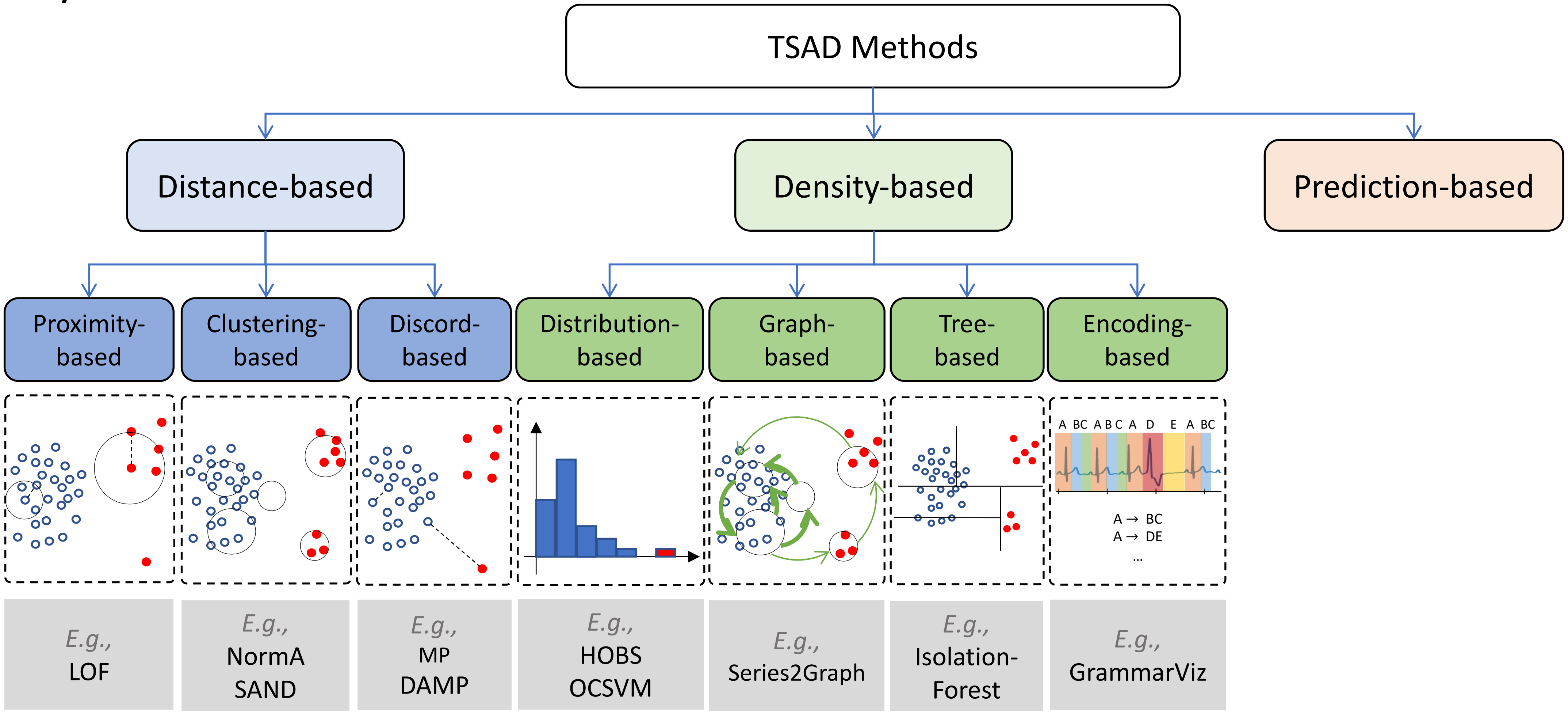
Anomaly Detection methods: *A taxonomy*

By methods…



Anomaly Detection methods: *A taxonomy*

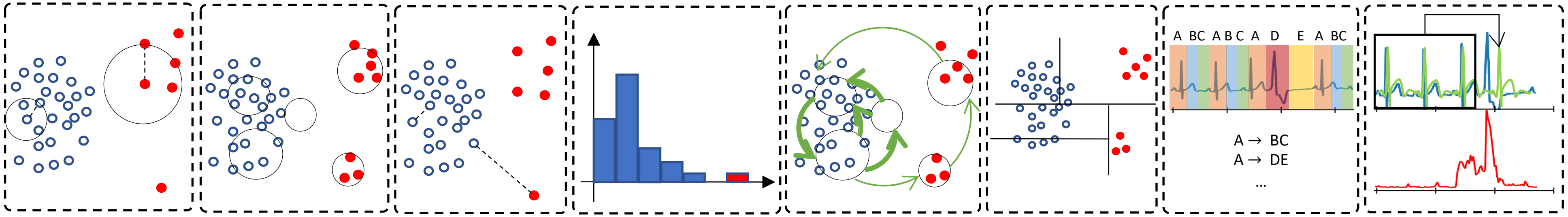
By methods…



Anomaly Detection methods: *A taxonomy*

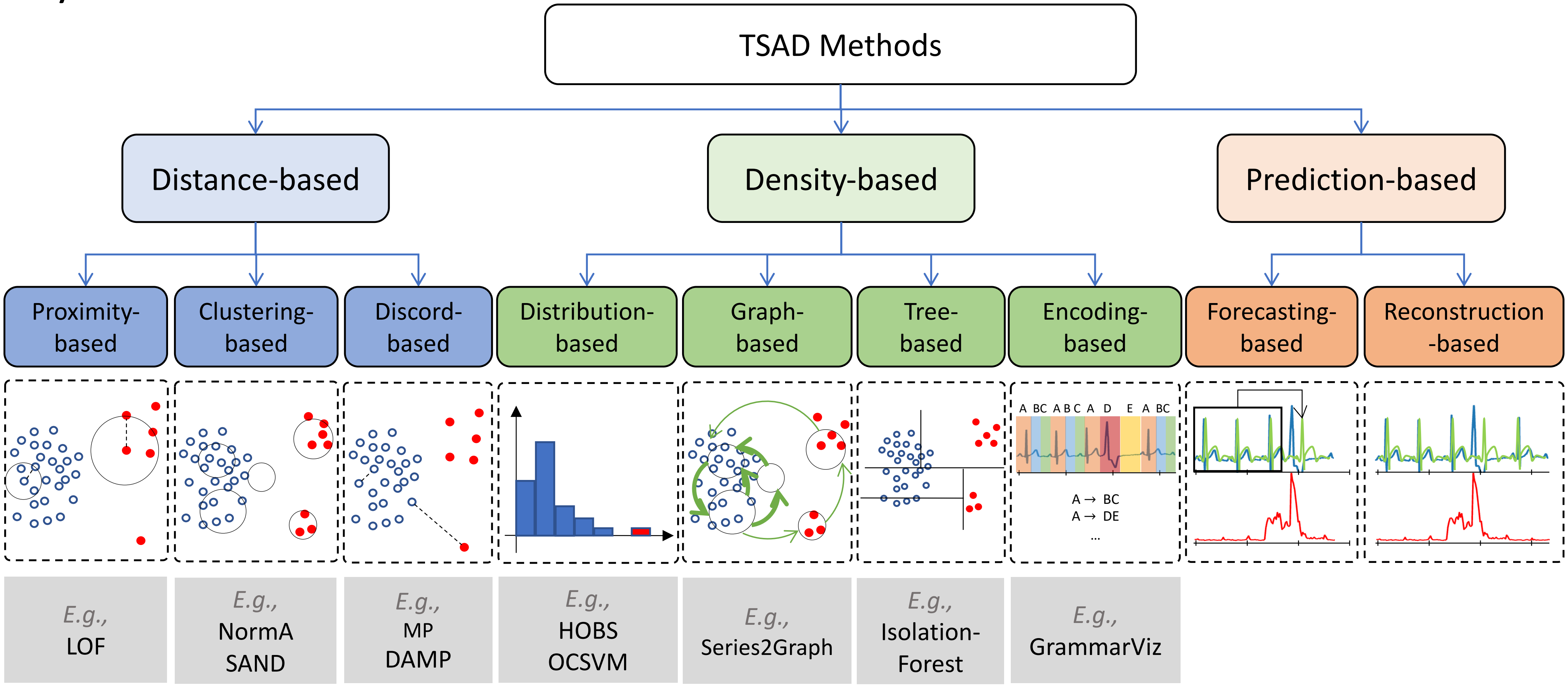
By methods…

- TSAD Methods
  - Distance-based
    - Proximity-based — *E.g.,* LOF
    - Clustering-based — *E.g.,* NormA, SAND
    - Discord-based — *E.g.,* MP, DAMP
  - Density-based
    - Distribution-based — *E.g.,* HOBS, OCSVM
    - Graph-based — *E.g.,* Series2Graph
    - Tree-based — *E.g.,* Isolation-Forest
    - Encoding-based — *E.g.,* GrammarViz
  - Prediction-based
    - Forecasting-based



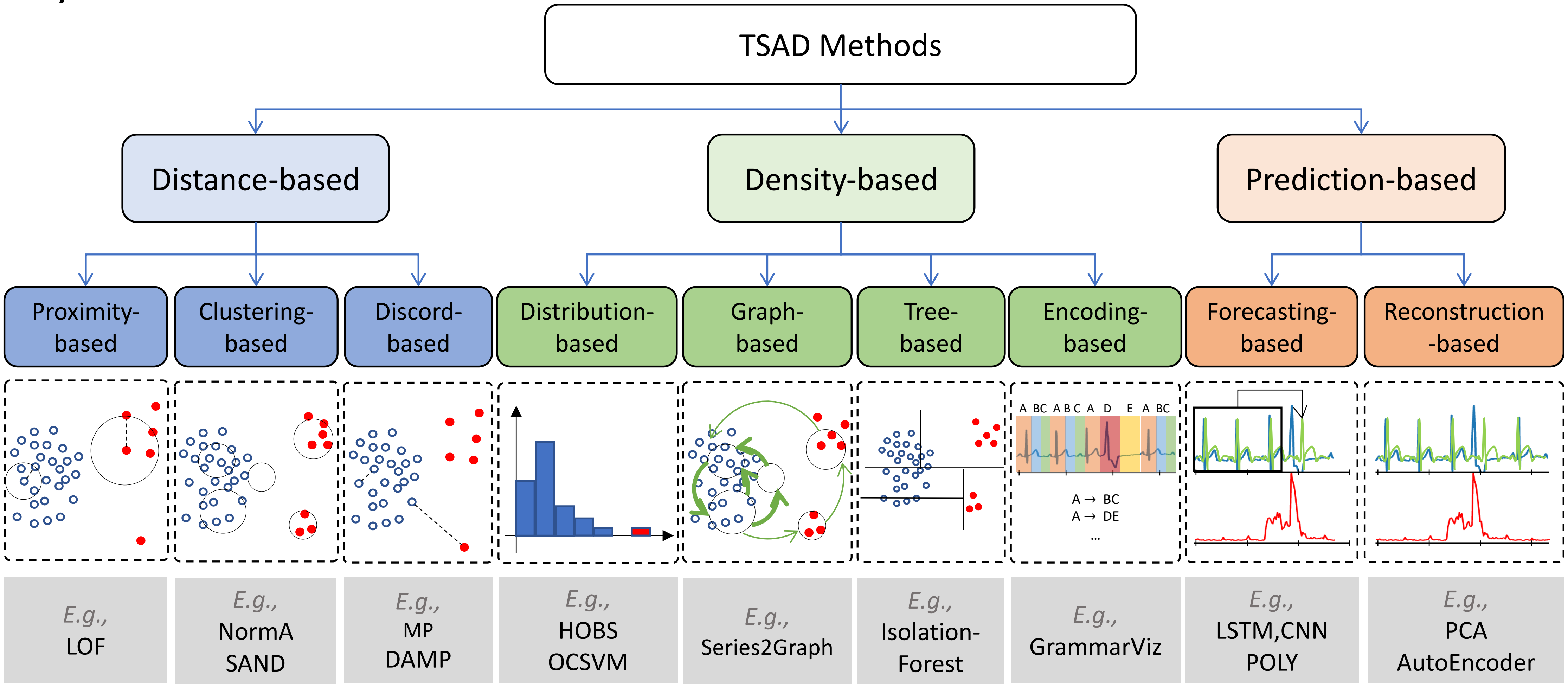
Anomaly Detection methods: *A taxonomy*

By methods…



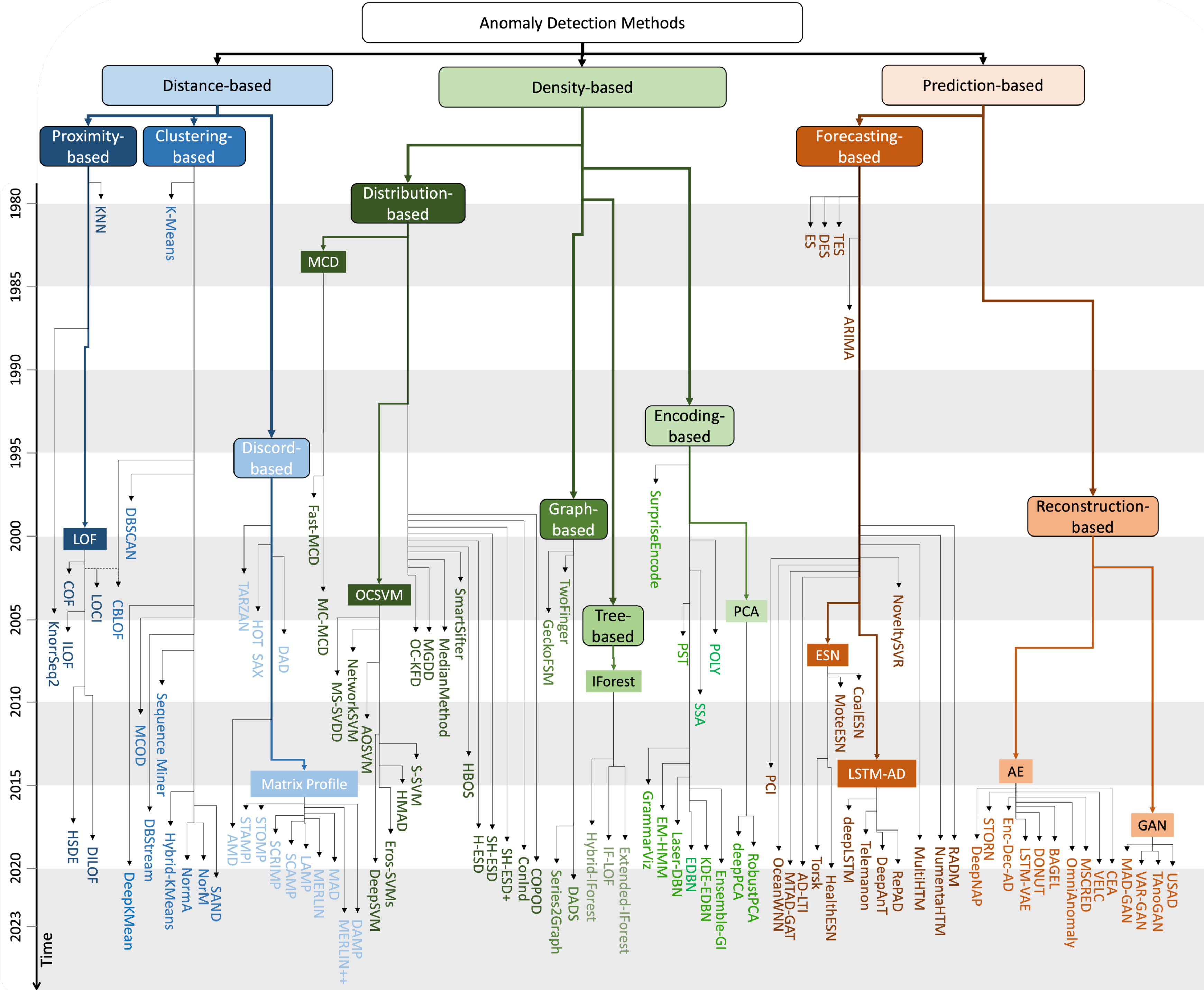
Anomaly Detection methods: *A taxonomy*

By methods…



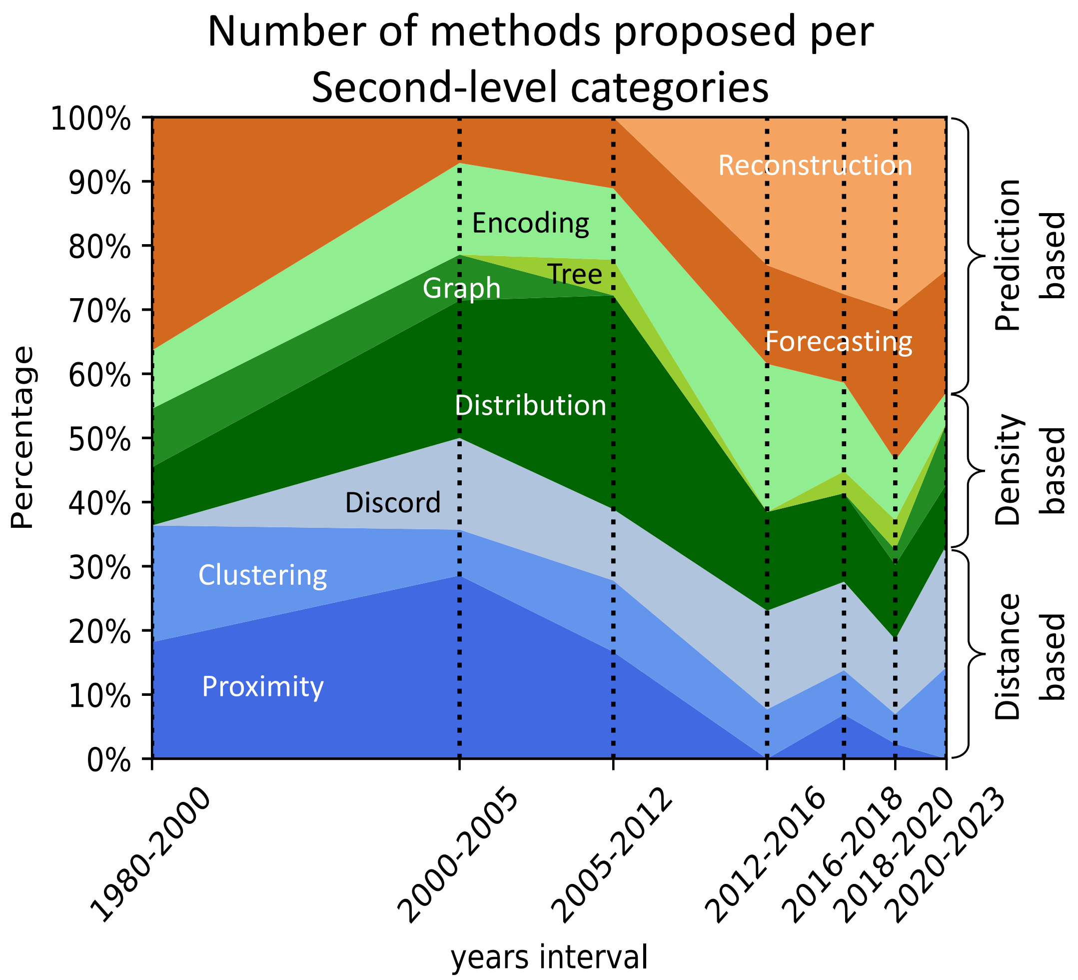
Anomaly Detection methods: *A taxonomy*

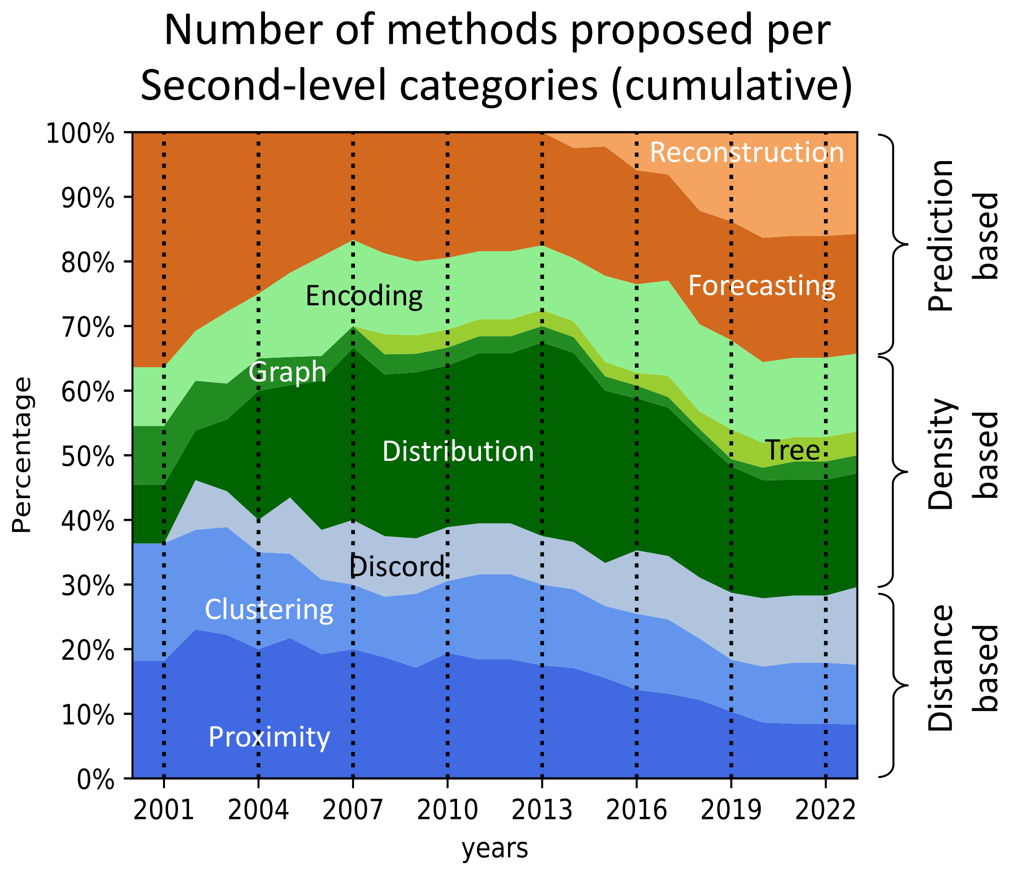
By time…



Anomaly Detection methods: *A taxonomy*

By time…

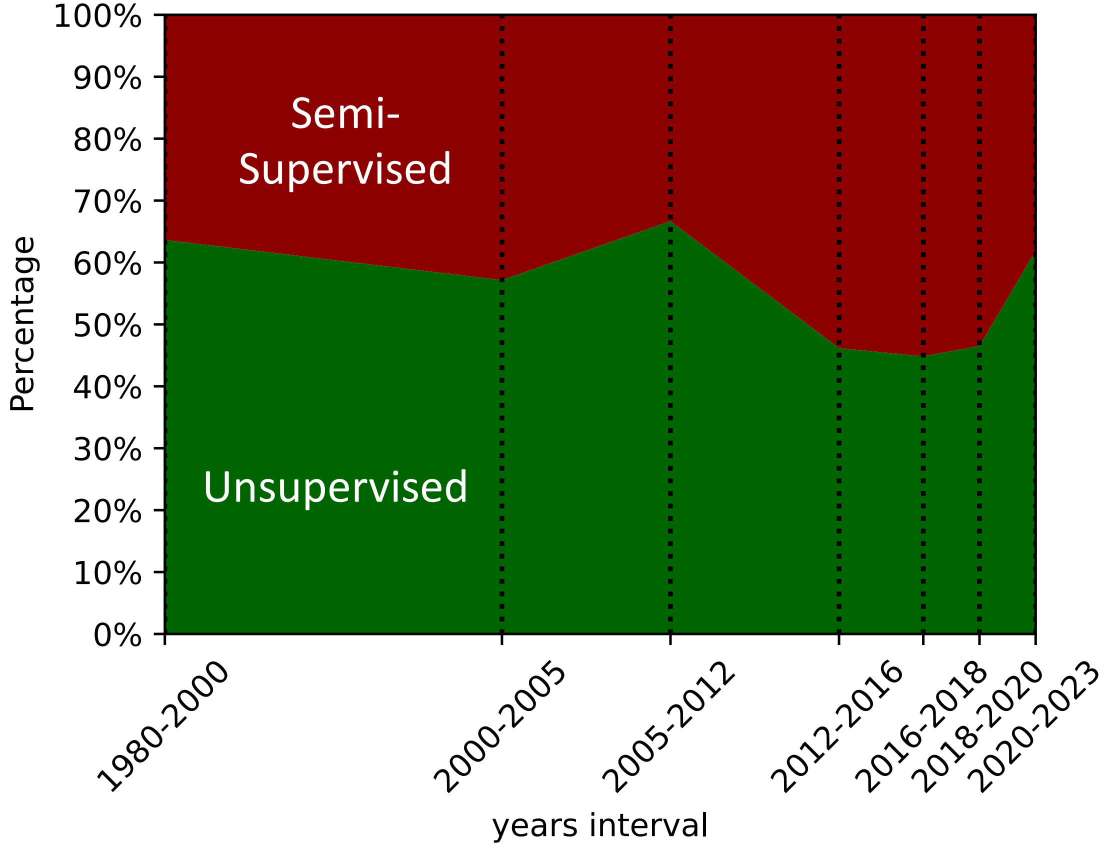




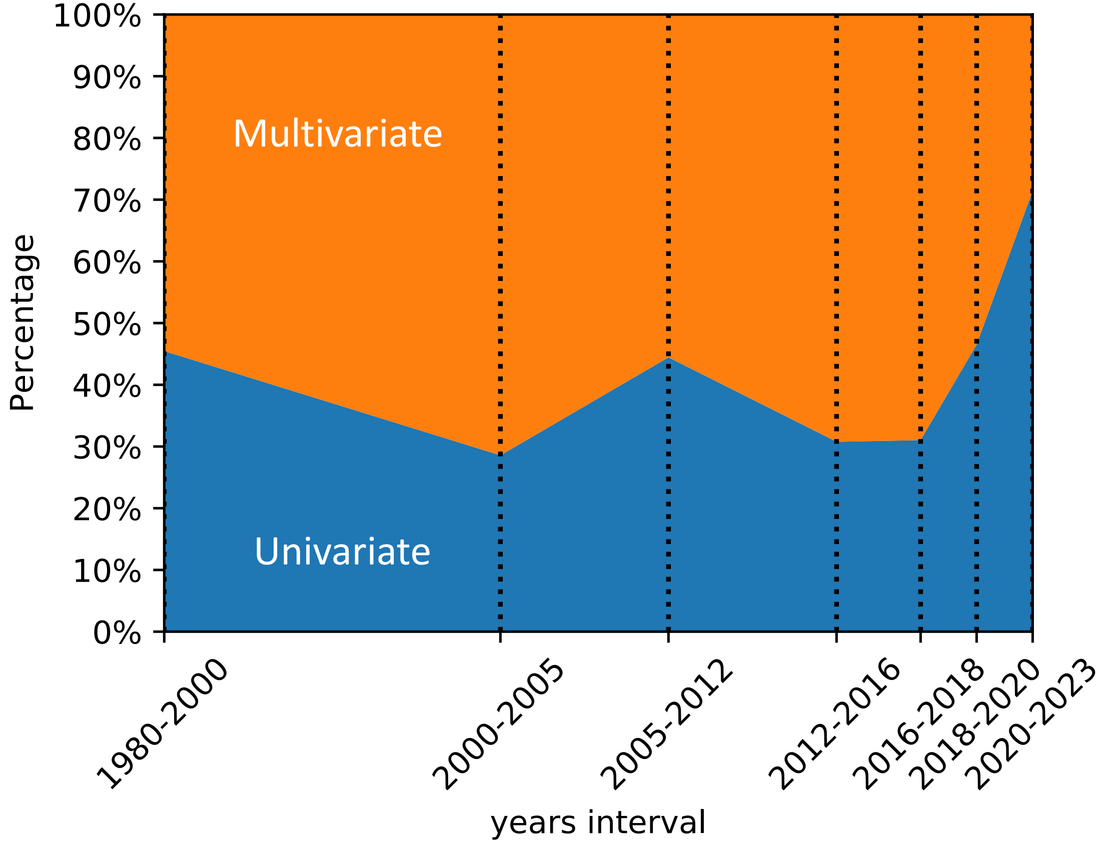
Anomaly Detection methods: *A taxonomy*

By time…

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

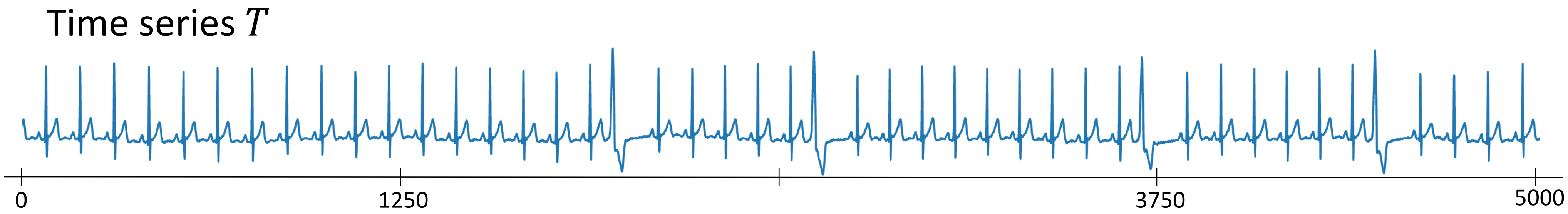


Number of methods proposed that can handle *Univariate* or *Multivariate* time series



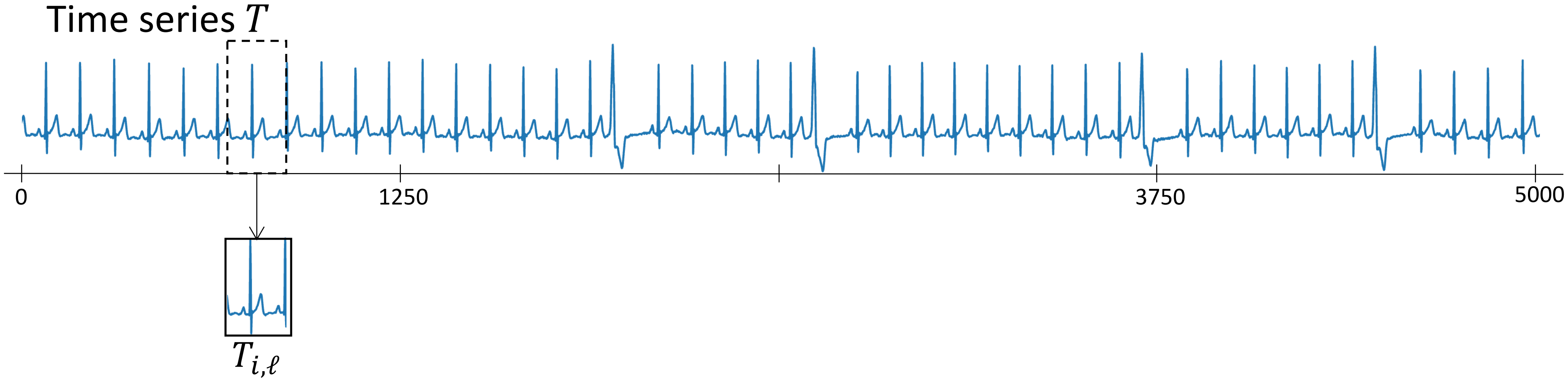
Anomaly Detection methods: *Distance-based*

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



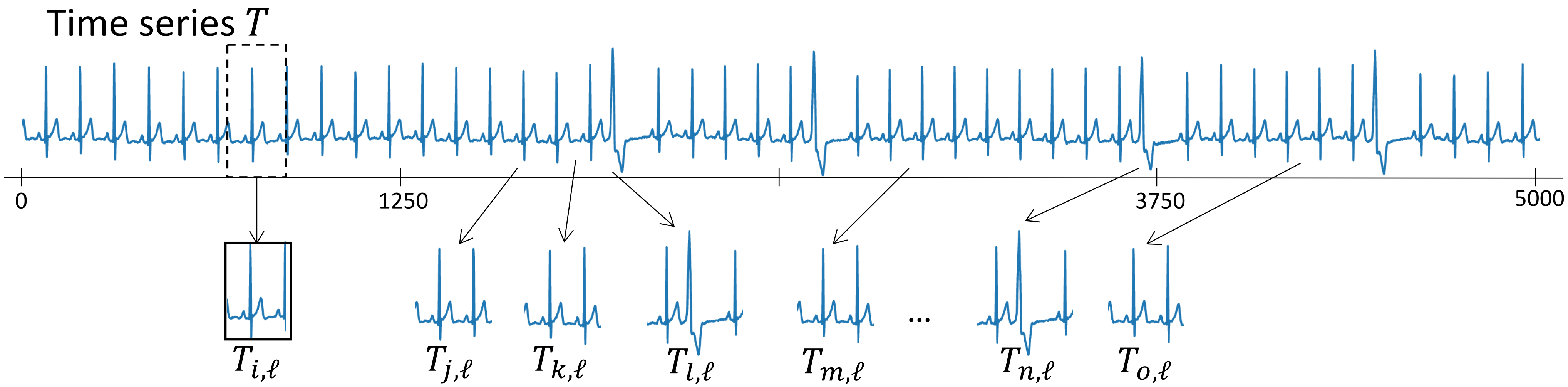
Anomaly Detection methods: *Distance-based*

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



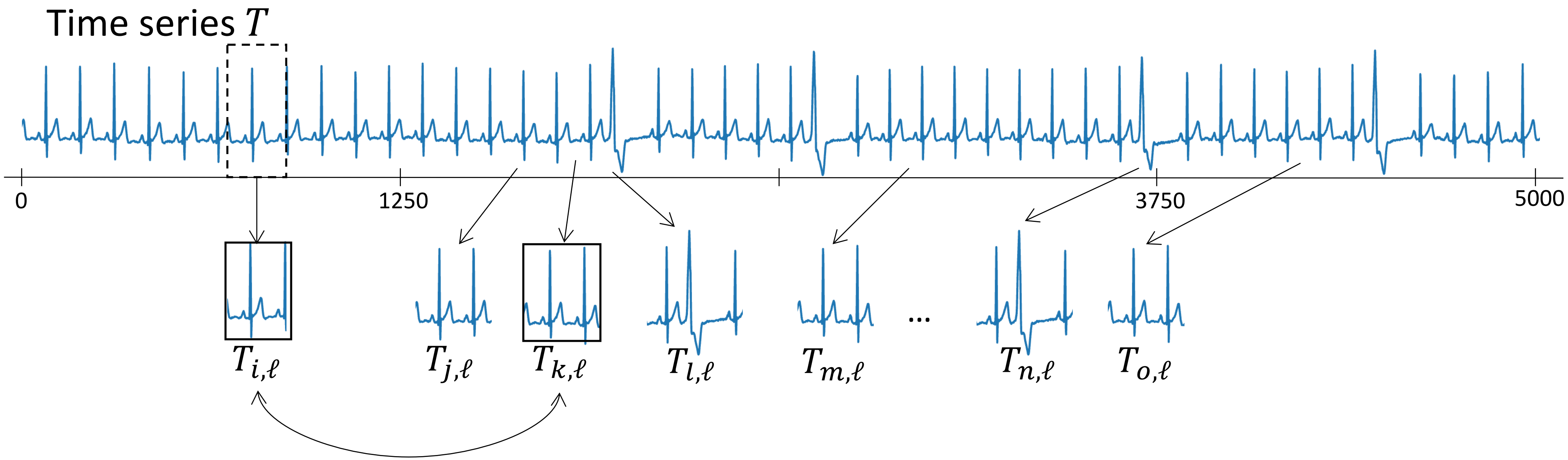
Anomaly Detection methods: *Distance-based*

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



Anomaly Detection methods: *Distance-based*

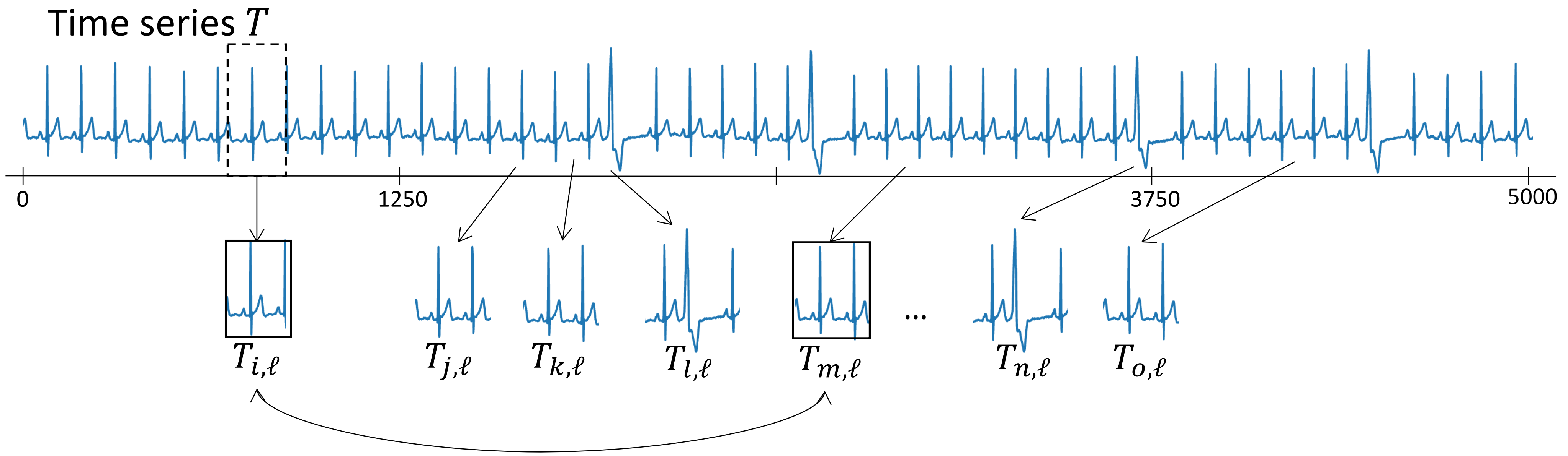
Methods that use distance computation between subsequences (or group of subsequences) to detect anomalies.



$S_{T,i} = d(T_{i,\ell}, T_{j,\ell})$

Anomaly Detection methods: *Distance-based*

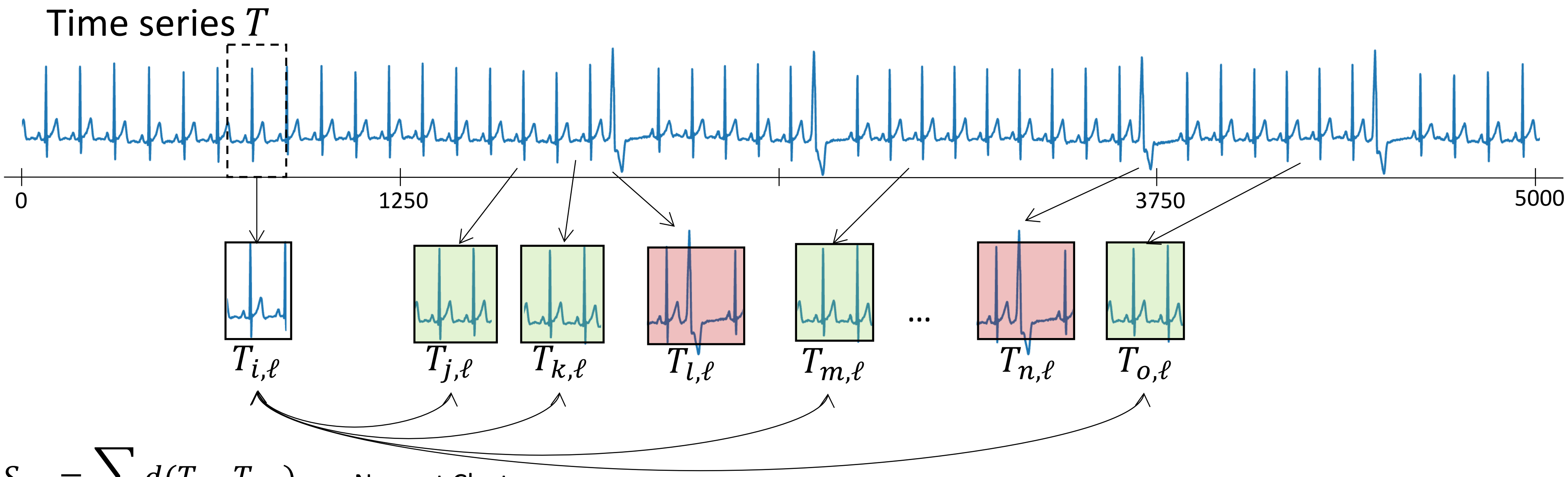
Methods that use distance computation between subsequences (or group of subsequences) to detect anomalies.



$S_{T,i} = d(T_{i,\ell}, T_{m,\ell})$

Anomaly Detection methods: *Distance-based*

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

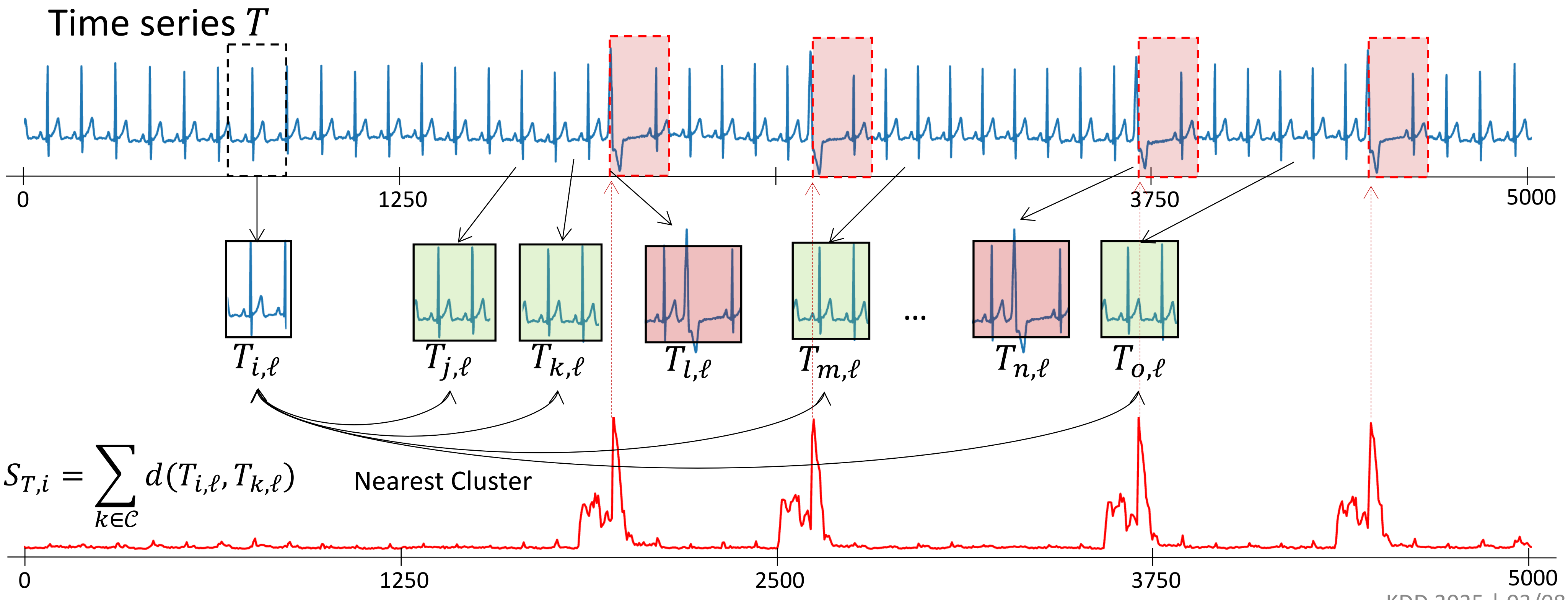


$$S_{T,i} = \sum_{k \in \mathcal{C}} d(T_{i,\ell}, T_{k,\ell})$$

Nearest Cluster

Anomaly Detection methods: *Distance-based*

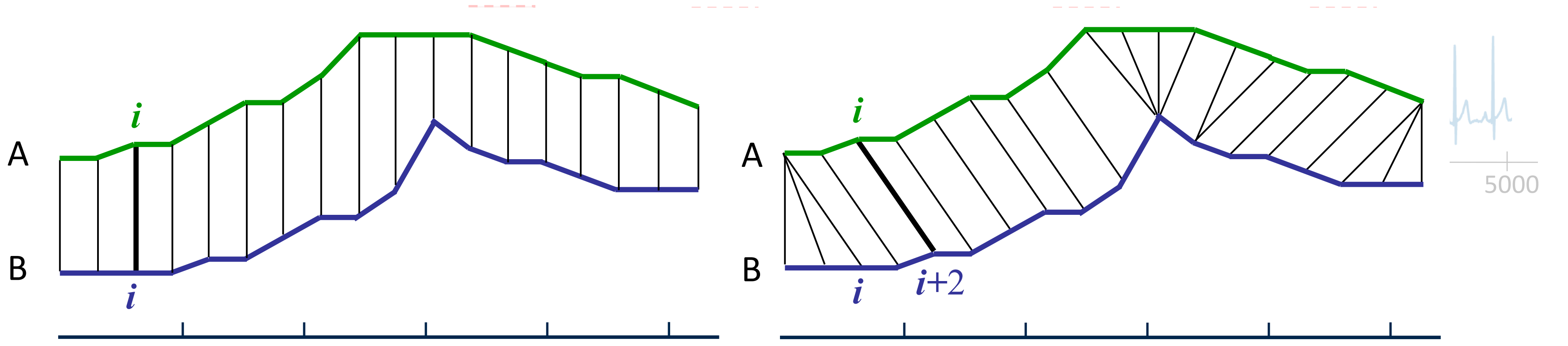
Methods that use distance computation between subsequences (or group of subsequences) to detect anomalies.



Anomaly Detection methods: *Distance-based*

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

Example of distance computation

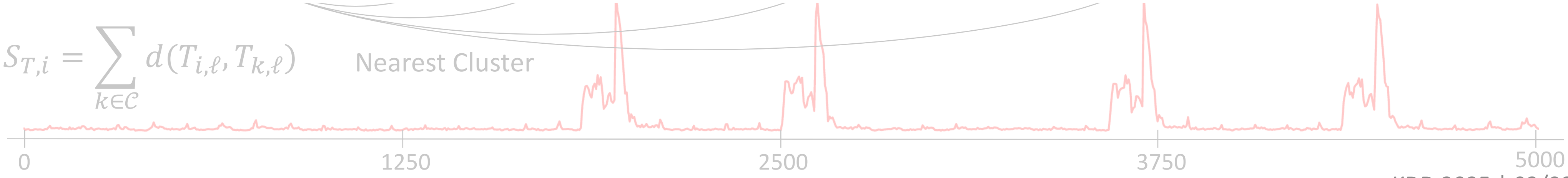


(a) Euclidian Distance

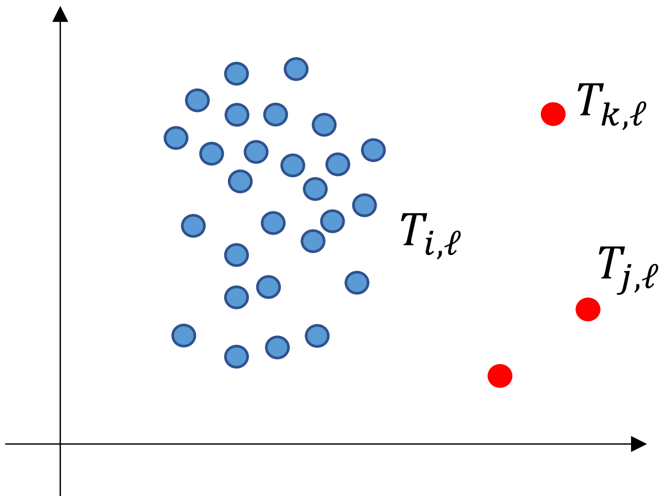
(b) DTW distance

$$S_{T,i} = \sum_{k \in \mathcal{C}} d(T_{i,\ell}, T_{k,\ell})$$

Nearest Cluster



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

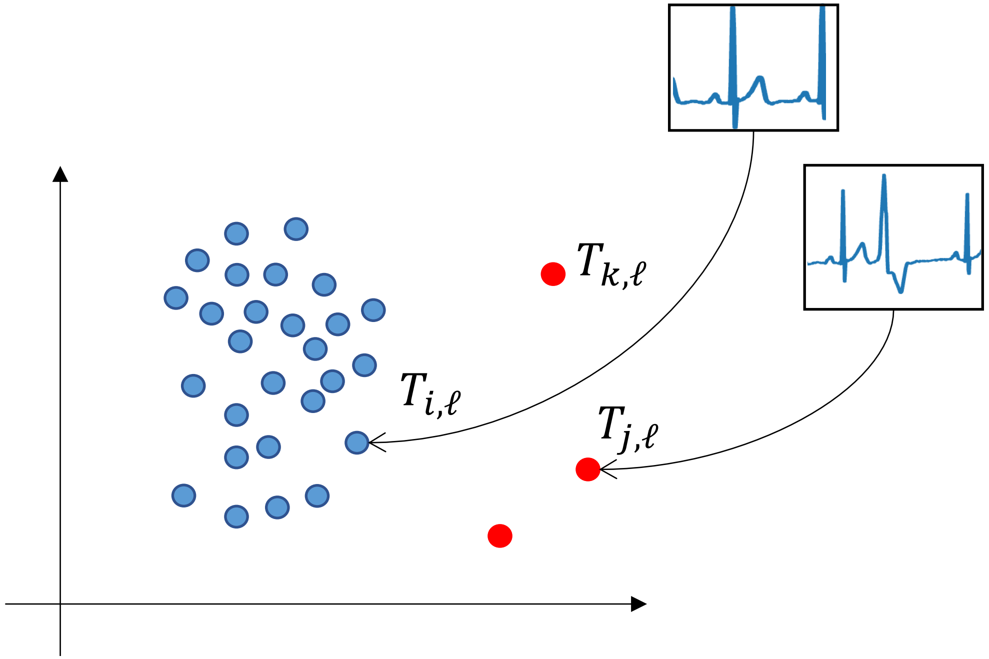
Unsupervised

Univariate

sequence

[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 Prole I: All Pairs Similarity Joins for Time Series. In ICDM.

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

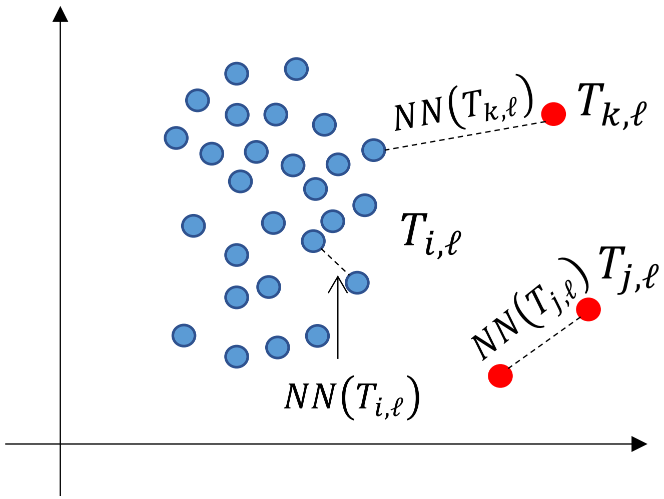
Unsupervised

Univariate

sequence

[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 Prole I: All Pairs Similarity Joins for Time Series. In ICDM.

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

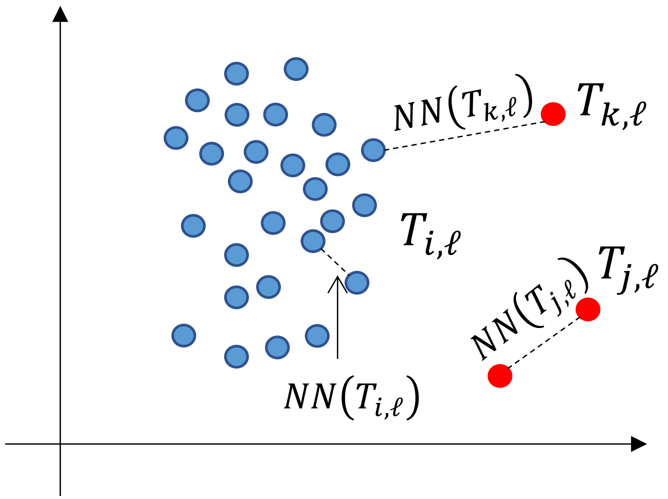
Unsupervised

Univariate

sequence

[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 Prole I: All Pairs Similarity Joins for Time Series. In ICDM.

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

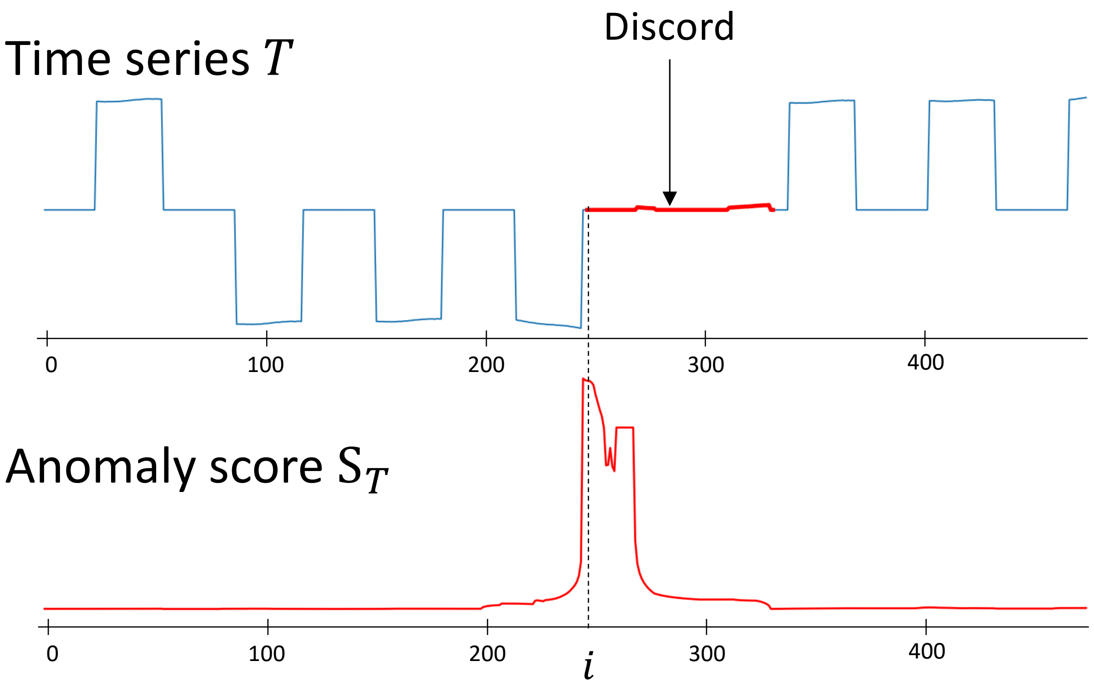
Unsupervised

Univariate

sequence

[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 Prole I: All Pairs Similarity Joins for Time Series. In ICDM.

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

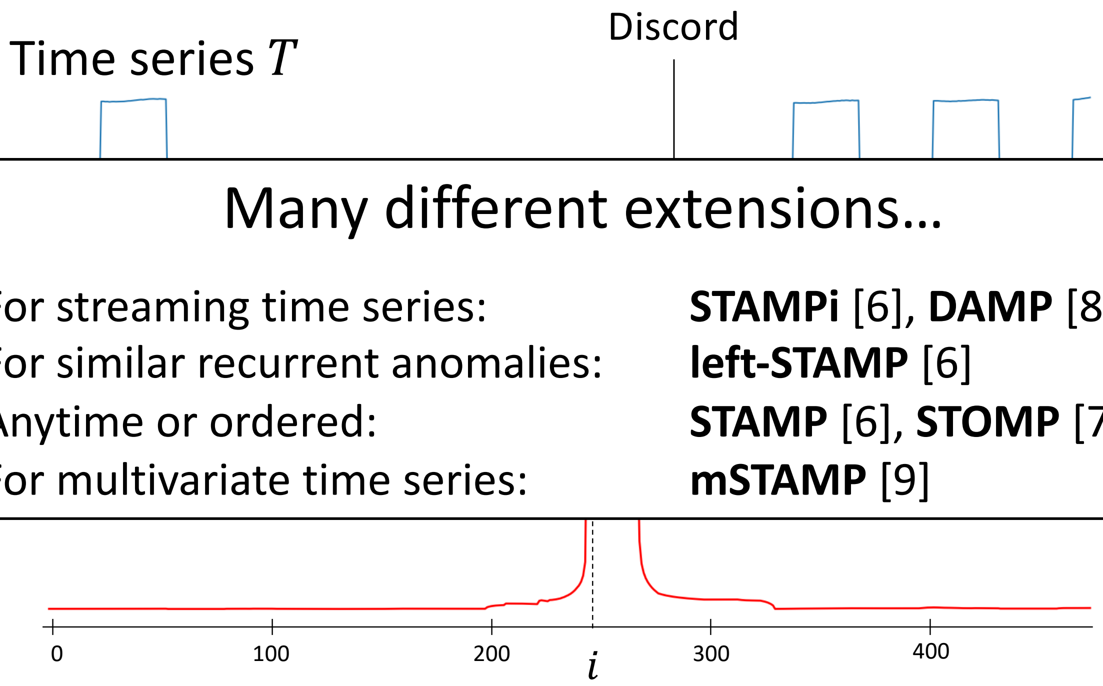
Unsupervised

Univariate

sequence

[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 Prole I: All Pairs Similarity Joins for Time Series. In ICDM.

Anomaly Detection methods: *an Example*



Many different extensions…

- For streaming time series:	**STAMPi** [6], **DAMP** [8]
- For similar recurrent anomalies:	**left-STAMP** [6]
- Anytime or ordered:	**STAMP** [6], **STOMP** [7]
- For multivariate time series:	**mSTAMP** [9]

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

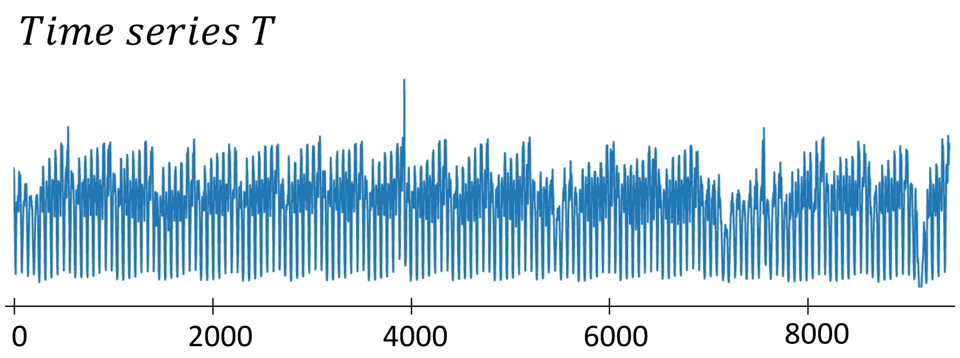
Unsupervised

Univariate

sequence

[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 Prole I: All Pairs Similarity Joins for Time Series. In ICDM.

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

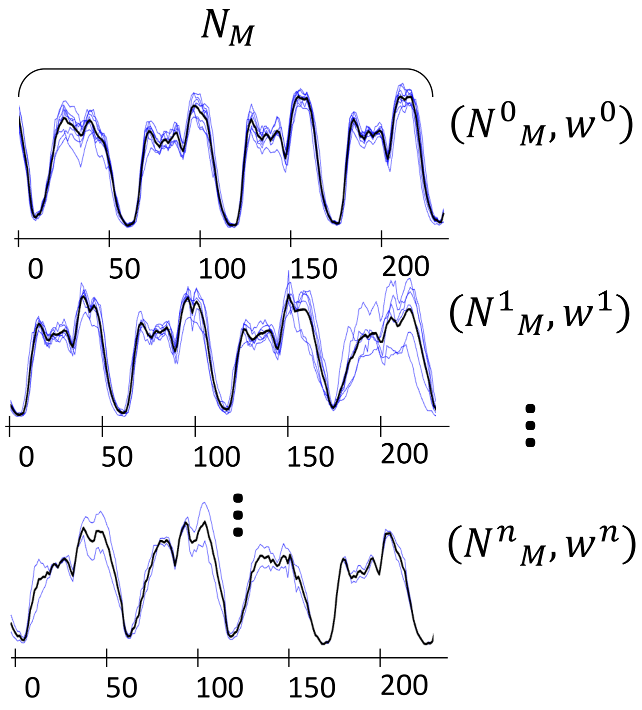
Unsupervised

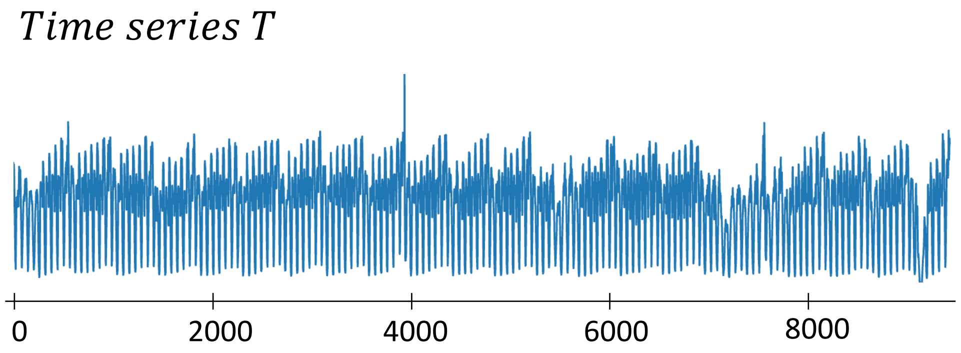
Univariate

sequence

[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.

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

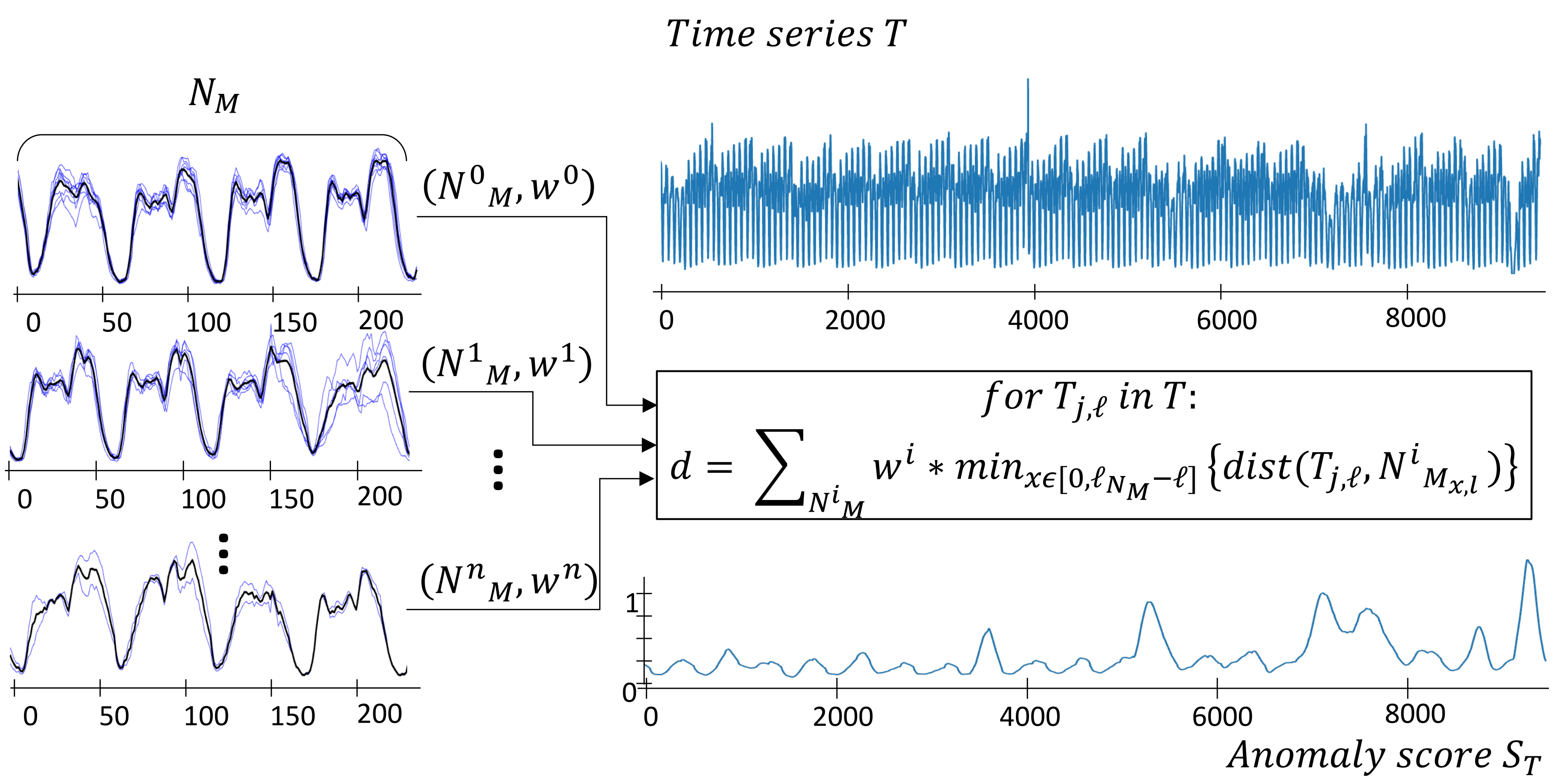
Unsupervised

Univariate

sequence

[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.

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

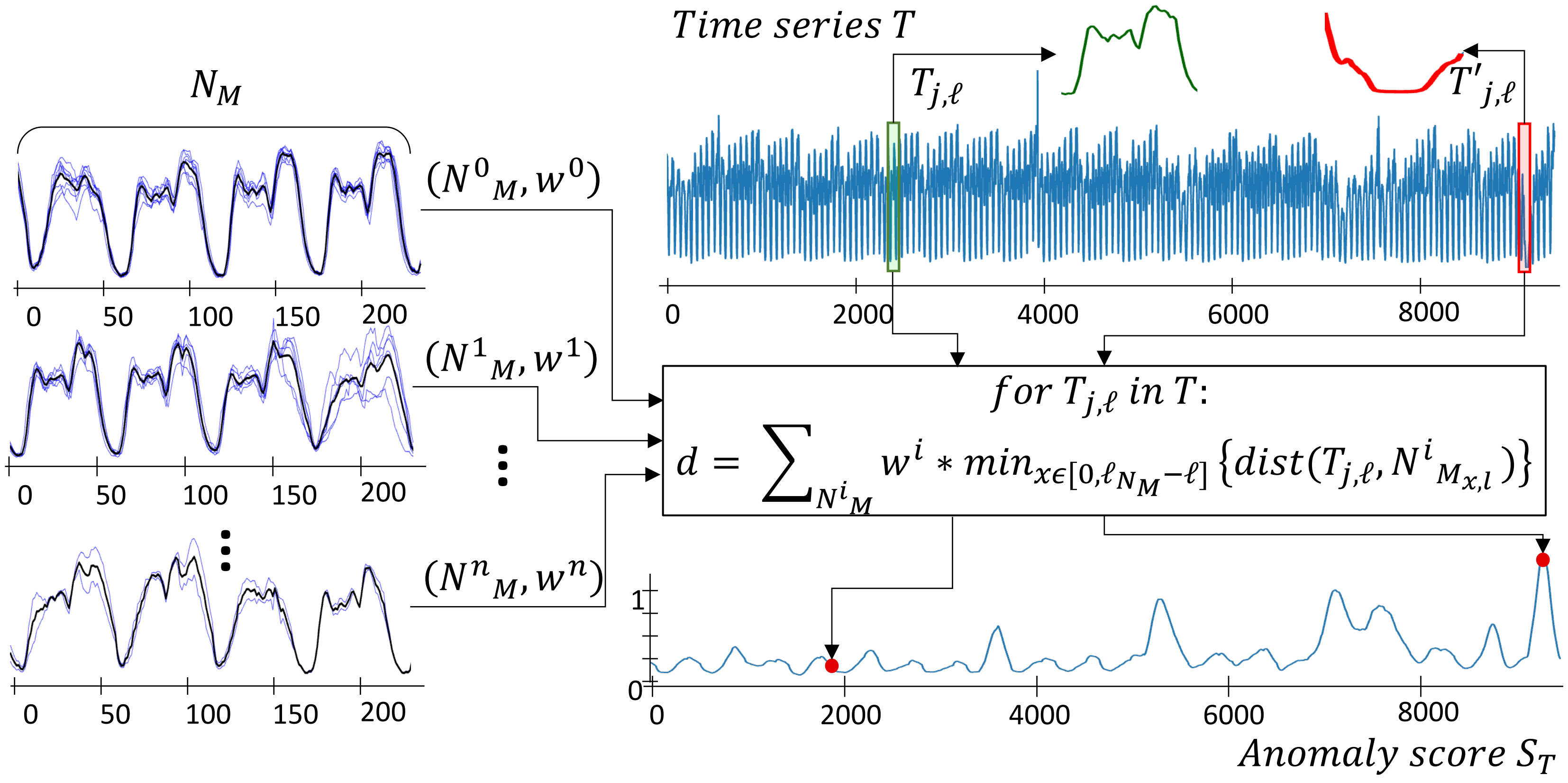
Unsupervised

Univariate

sequence

[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.

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

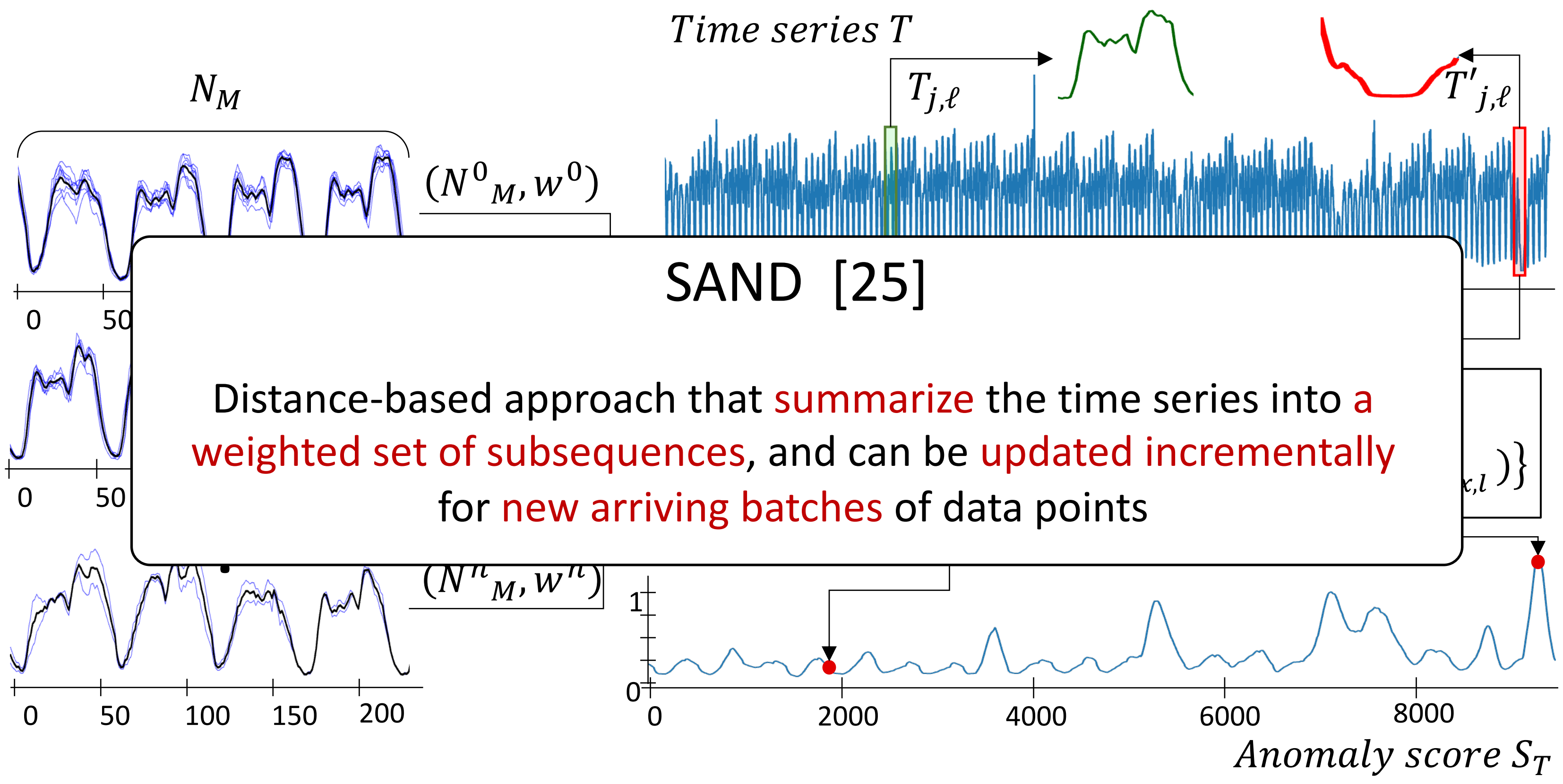
Unsupervised

Univariate

sequence

[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.

Anomaly Detection methods: *an Example*



SAND [25]

Distance-based approach that summarize the time series into a weighted set of subsequences, and can be updated incrementally for new arriving batches of data points

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

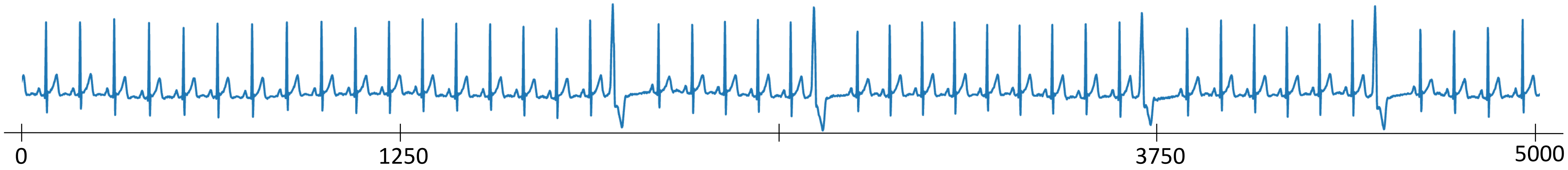
sequence

[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.

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.

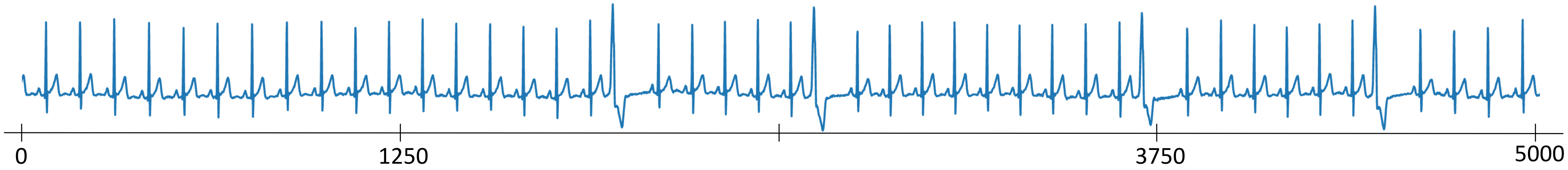
Time series T



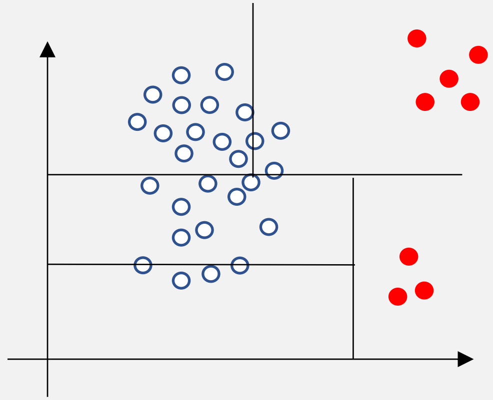
Anomaly Detection methods: *Density-based*

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



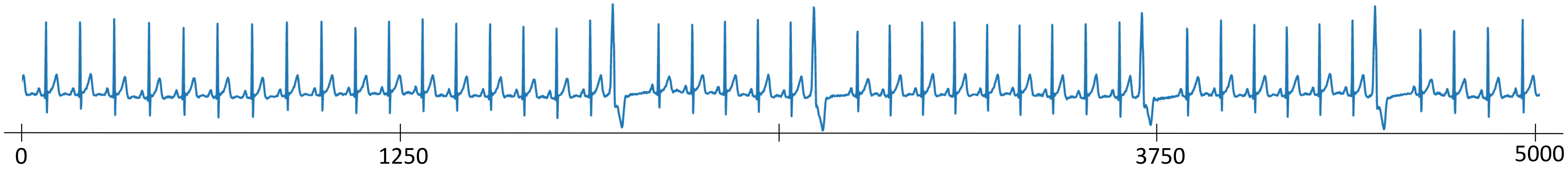
Tree-based approaches [11]



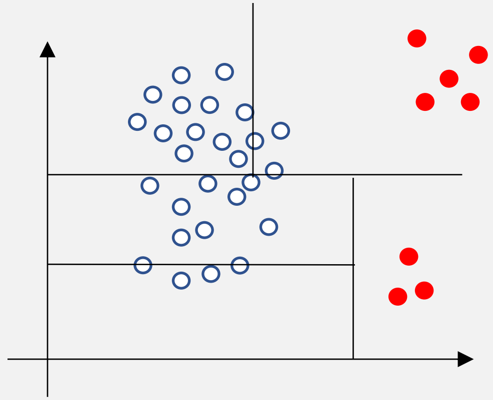
Anomaly Detection methods: *Density-based*

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

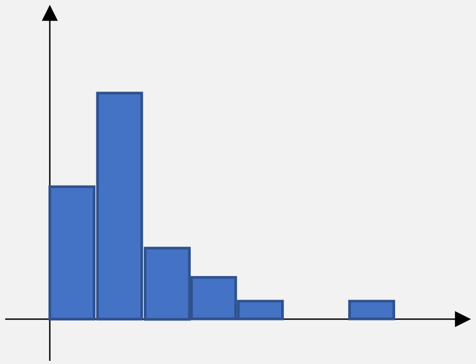
Time series T



Tree-based approaches [11]



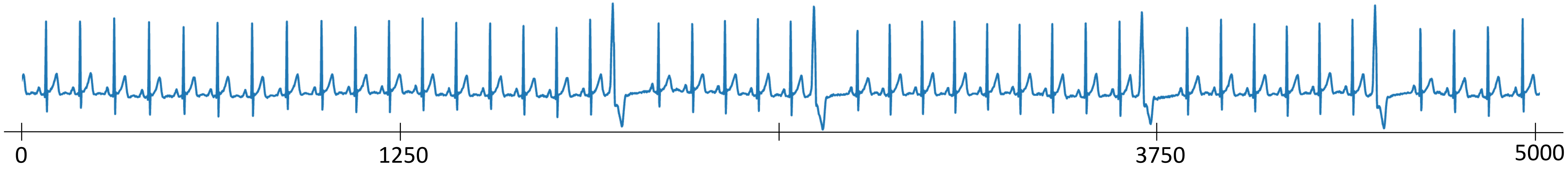
Distribution-based Approaches [12]



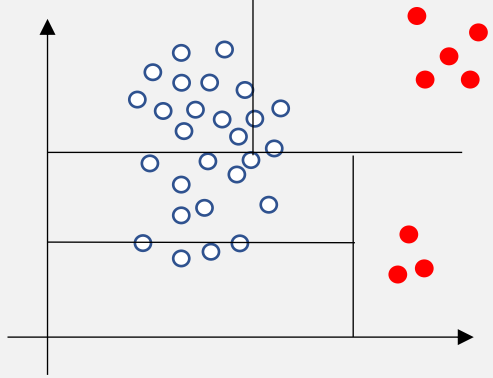
Anomaly Detection methods: *Density-based*

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

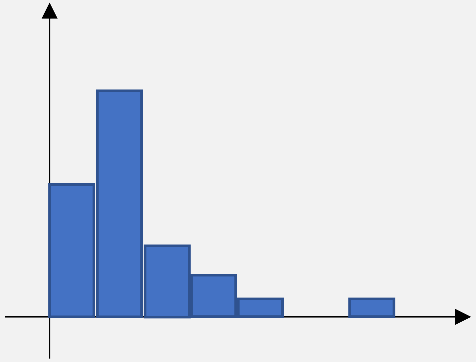
Time series T



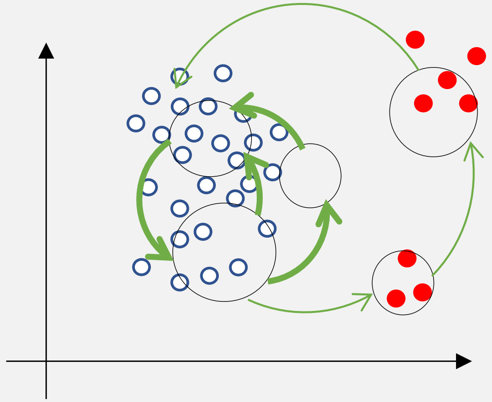
Tree-based approaches [11]



Distribution-based Approaches [12]



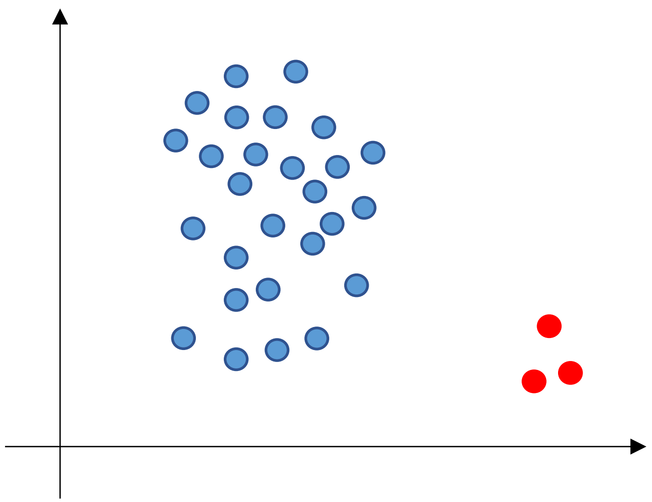
Graph-based approaches [13]



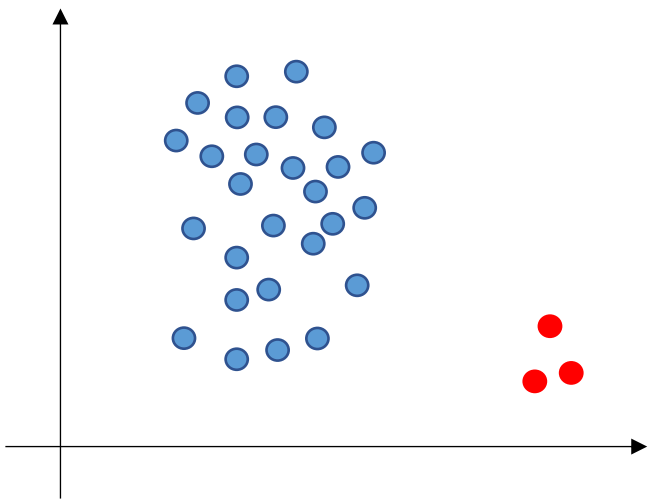
...

Anomaly Detection methods: *an Example*

0 splits



0 splits



Isolation Forest [11]

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

Unsupervised

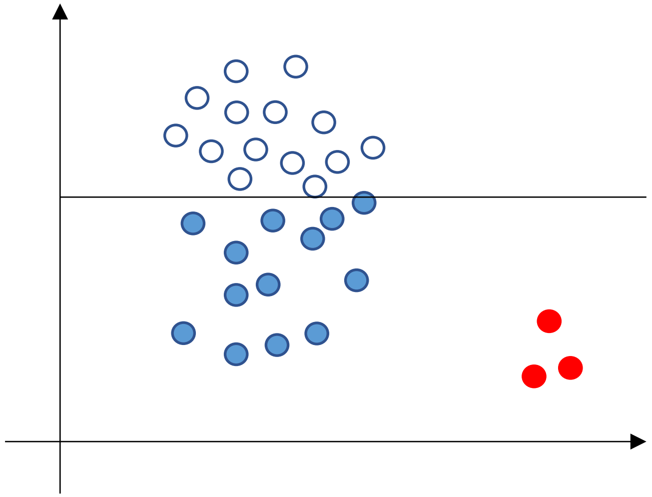
Univariate/Multivariate

Point/sequence

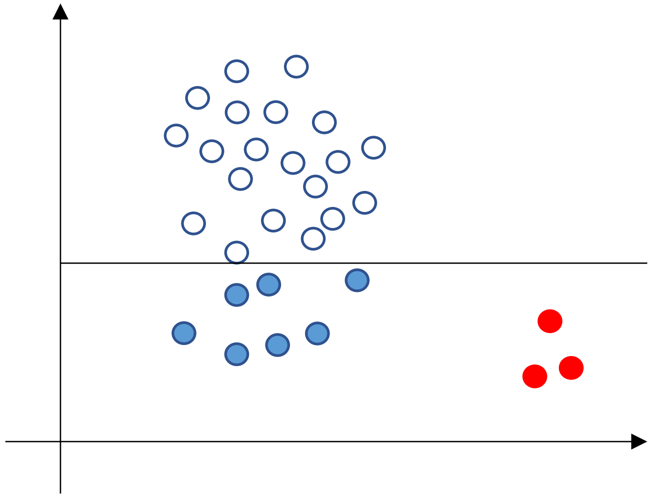
[11] F. T. Liu, K. M. Ting and Z. -H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422

Anomaly Detection methods: *an Example*

1 splits



1 splits



Isolation Forest [11]

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

Unsupervised

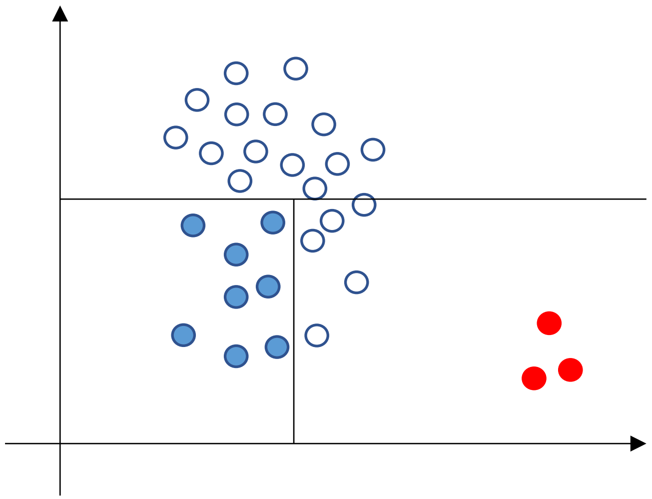
Univariate/Multivariate

Point/sequence

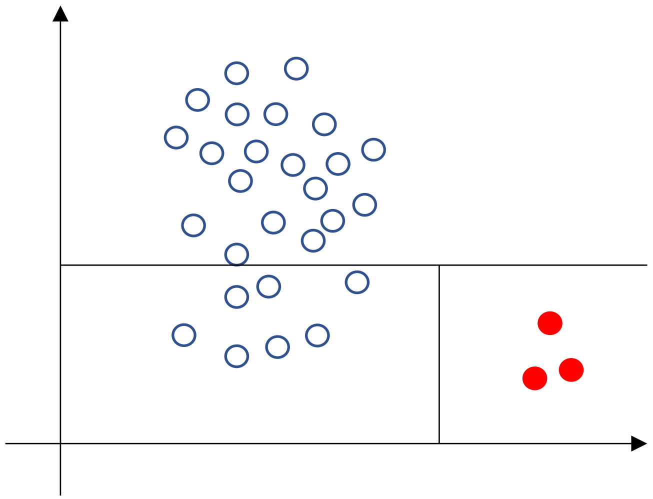
[11] F. T. Liu, K. M. Ting and Z. -H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422

Anomaly Detection methods: *an Example*

2 splits



2 splits



Isolation Forest [11]

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

Unsupervised

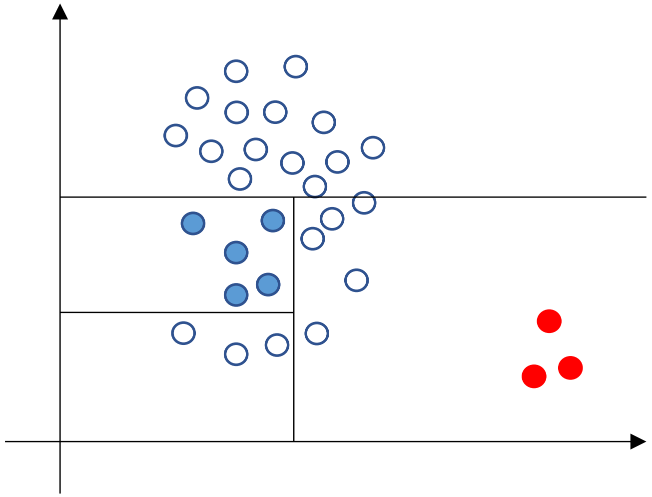
Univariate/Multivariate

Point/sequence

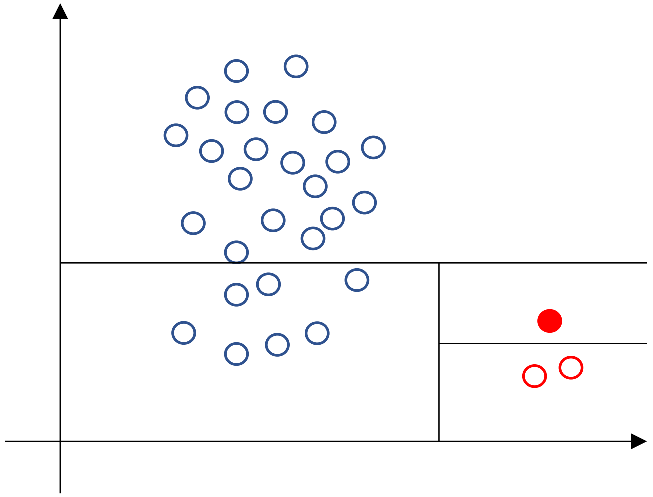
[11] F. T. Liu, K. M. Ting and Z. -H. Zhou, "Isolation Forest," 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422

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

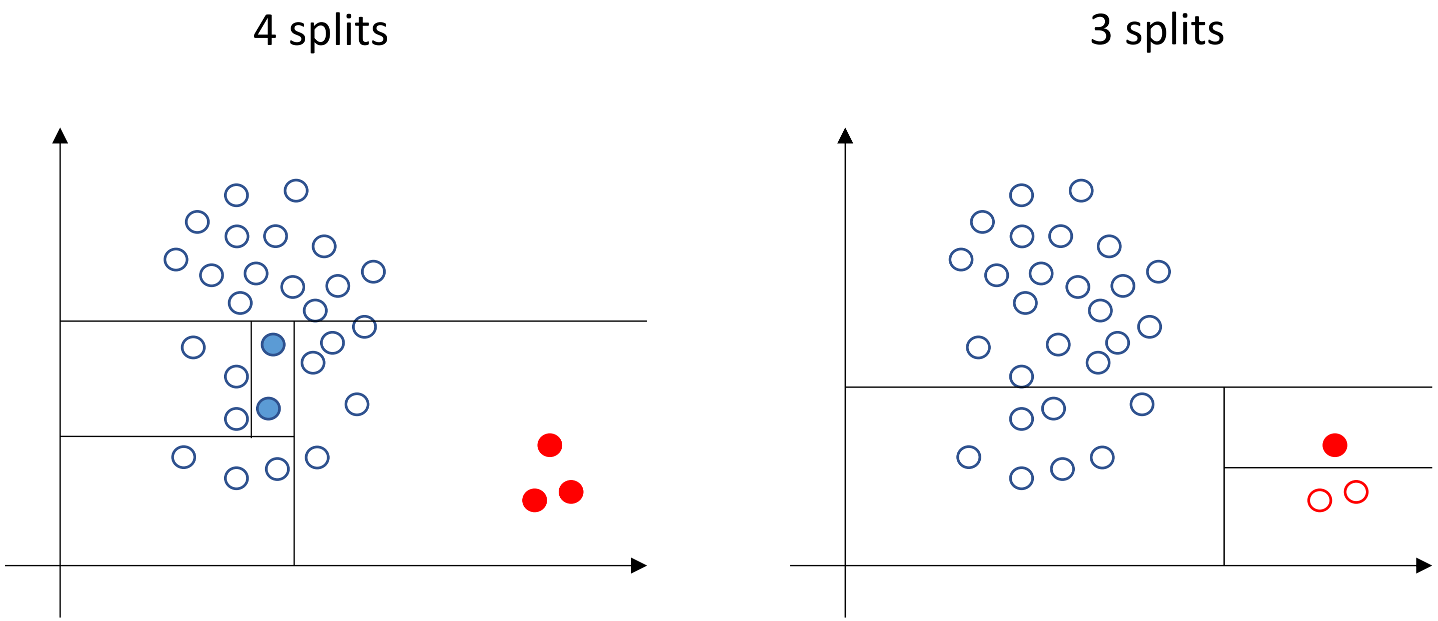
Unsupervised

Univariate/Multivariate

Point/sequence

[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

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

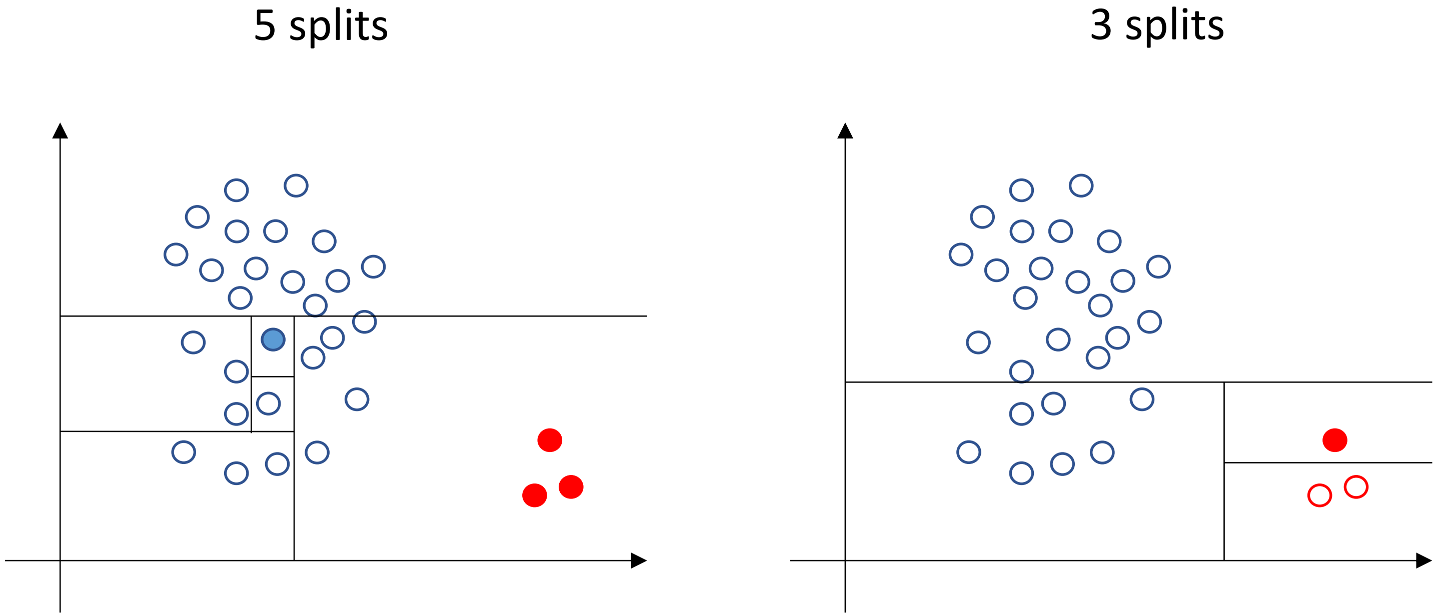
Unsupervised

Univariate/Multivariate

Point/sequence

[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

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

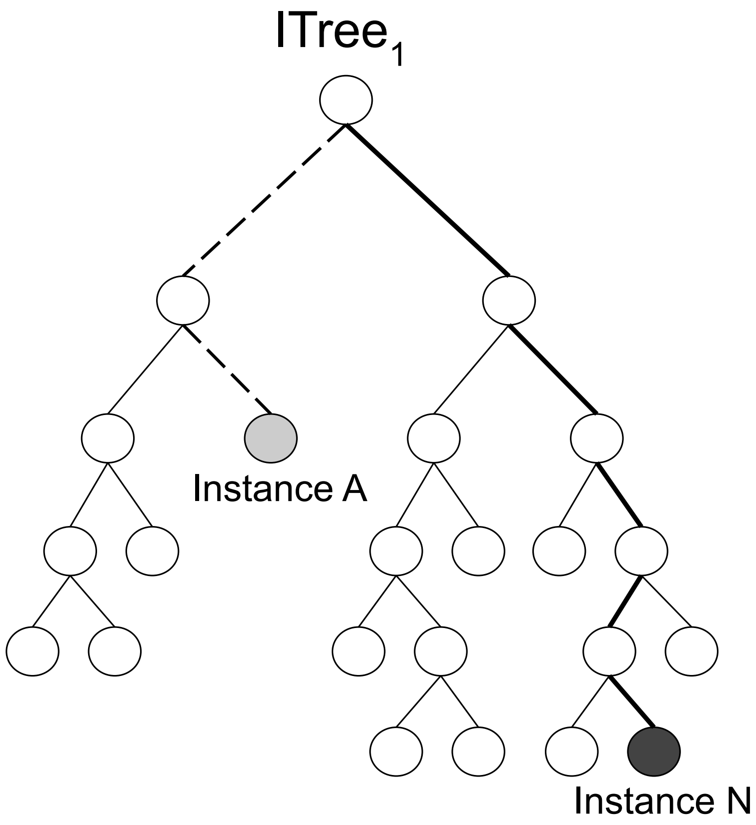
Unsupervised

Univariate/Multivariate

Point/sequence

[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

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

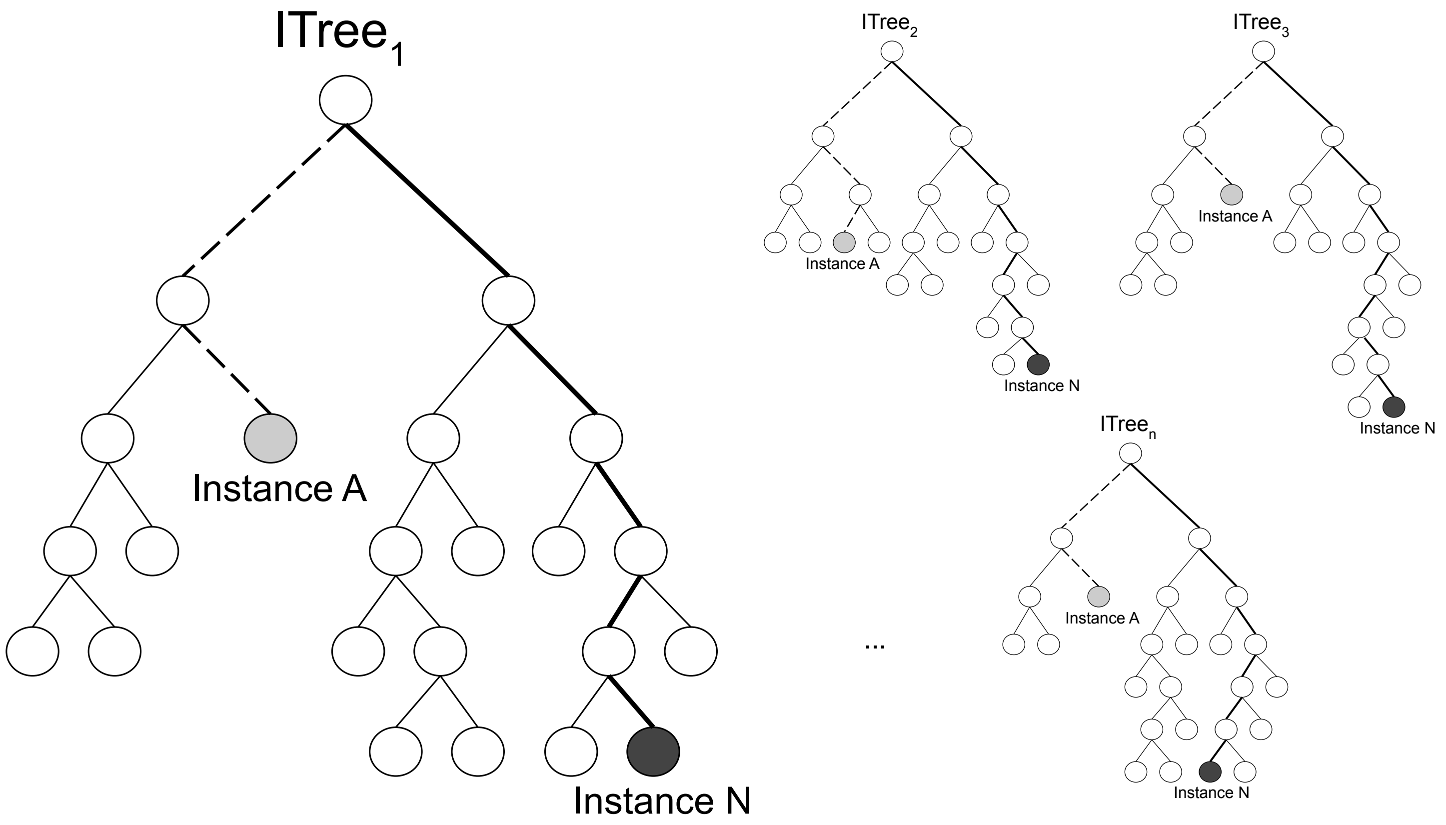
Unsupervised

Univariate/Multivariate

Point/sequence

[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

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

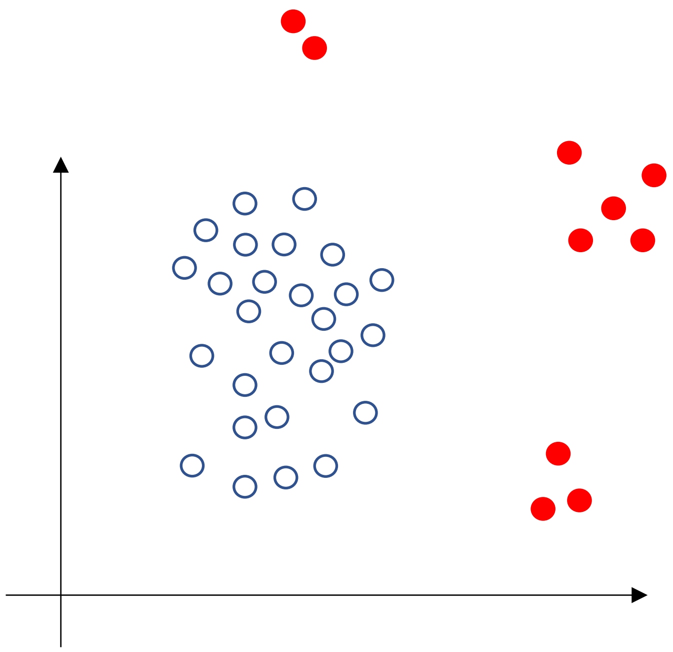
Unsupervised

Univariate/Multivariate

Point/sequence

[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

Anomaly Detection methods: *an Example*



Series2Graph [13]

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

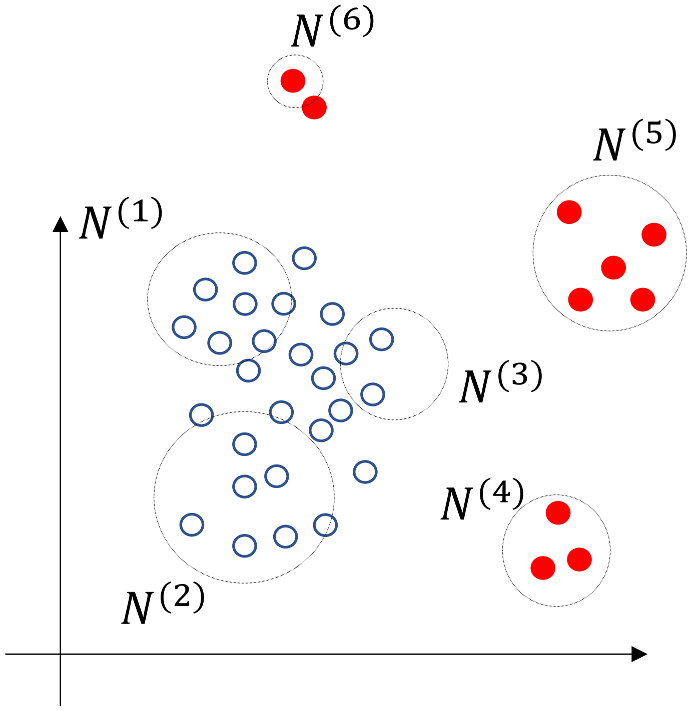
Unsupervised

Univariate

subsequence

[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

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

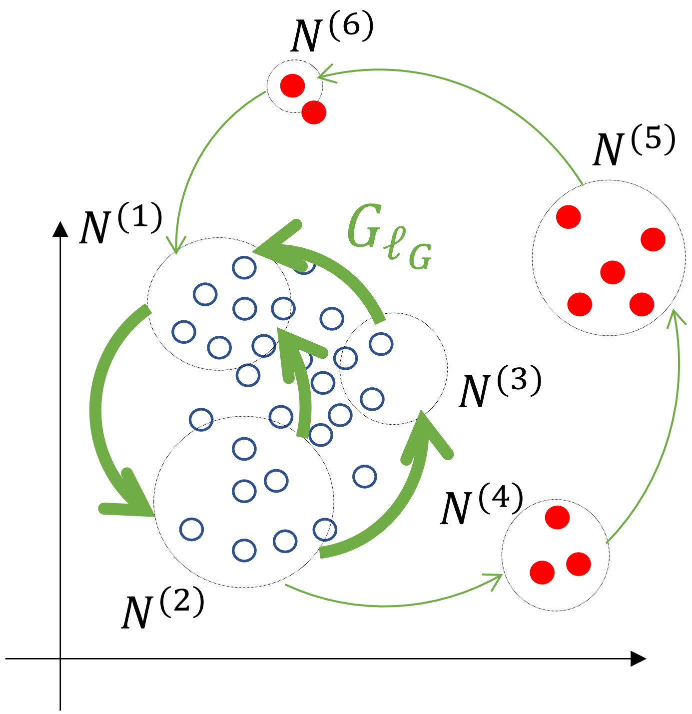
Unsupervised

Univariate

subsequence

[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

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

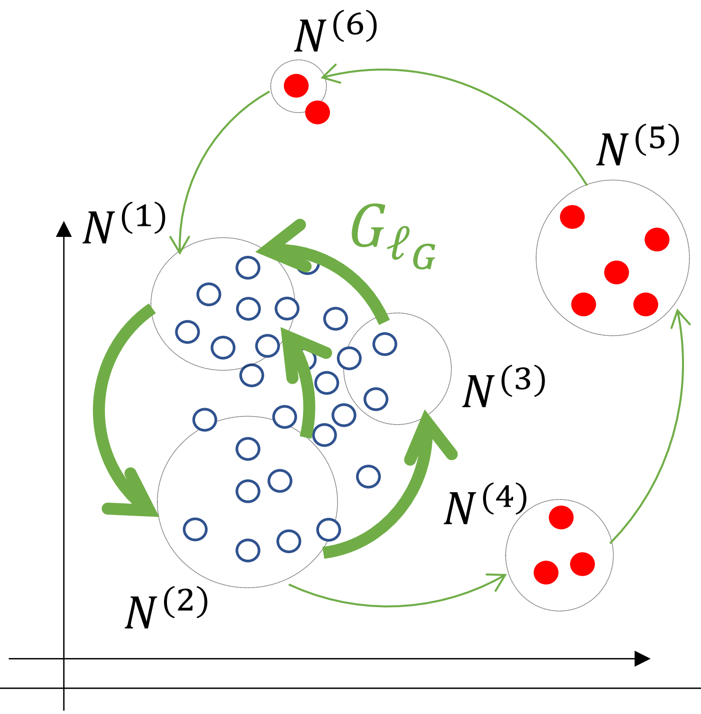
Unsupervised

Univariate

subsequence

[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

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 $P_{th} =< N^{(i)}, N^{(i+1)}, \ldots, N^{(i+\ell)} >$, we define the normality score as follows:

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

Series2Graph [13]

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

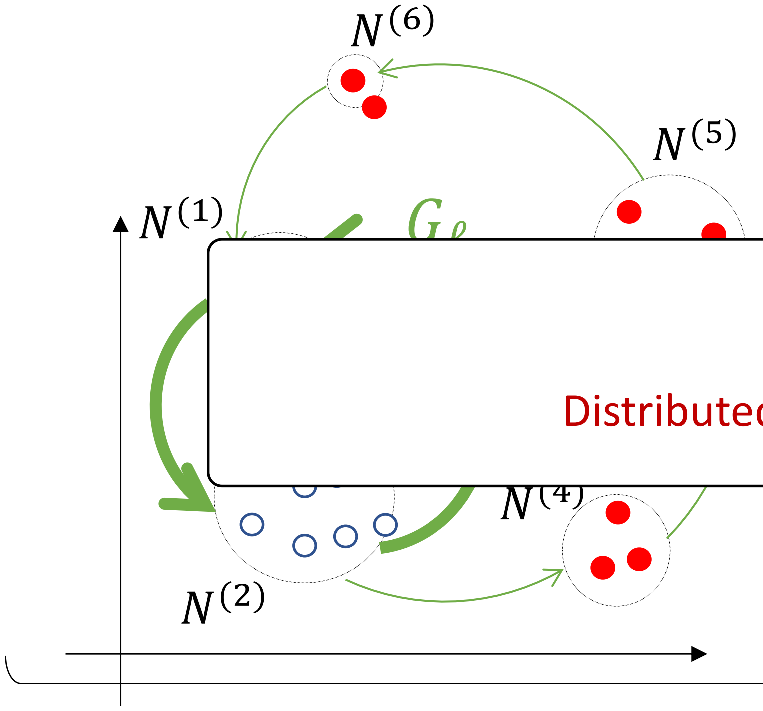
Unsupervised

Univariate

subsequence

[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

Anomaly Detection methods: *an Example*



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

DADS [26]

Distributed version of Series2Graph

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

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

Series2Graph [13]

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

Unsupervised

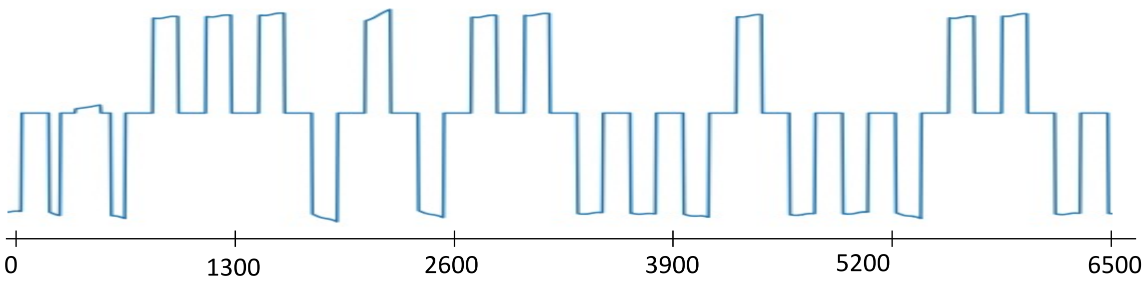
Univariate

subsequence

[26] Schneider, J., Wenig, P. & Papenbrock, T. Distributed detection of sequential anomalies in univariate time series. The VLDB Journal **30**, 579–602 (2021).

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

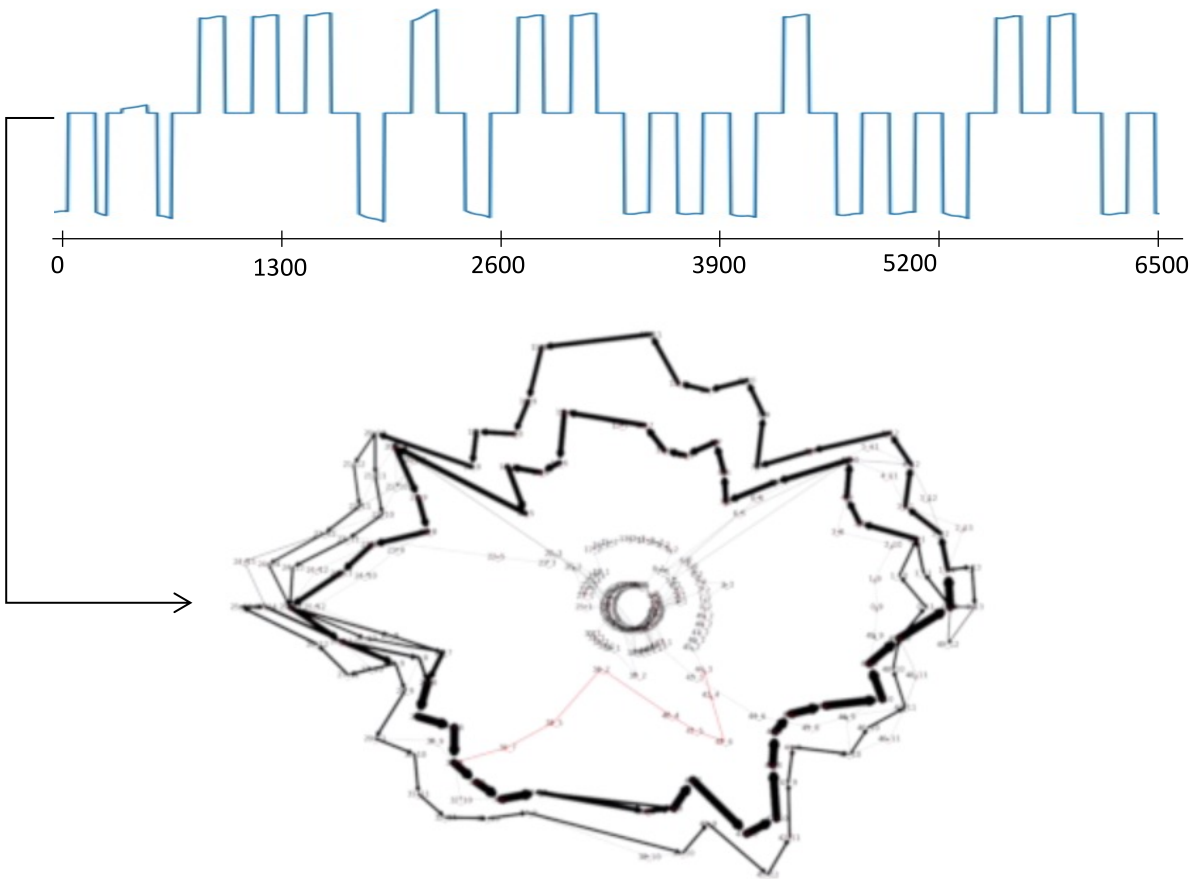
Univariate

subsequence

[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

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

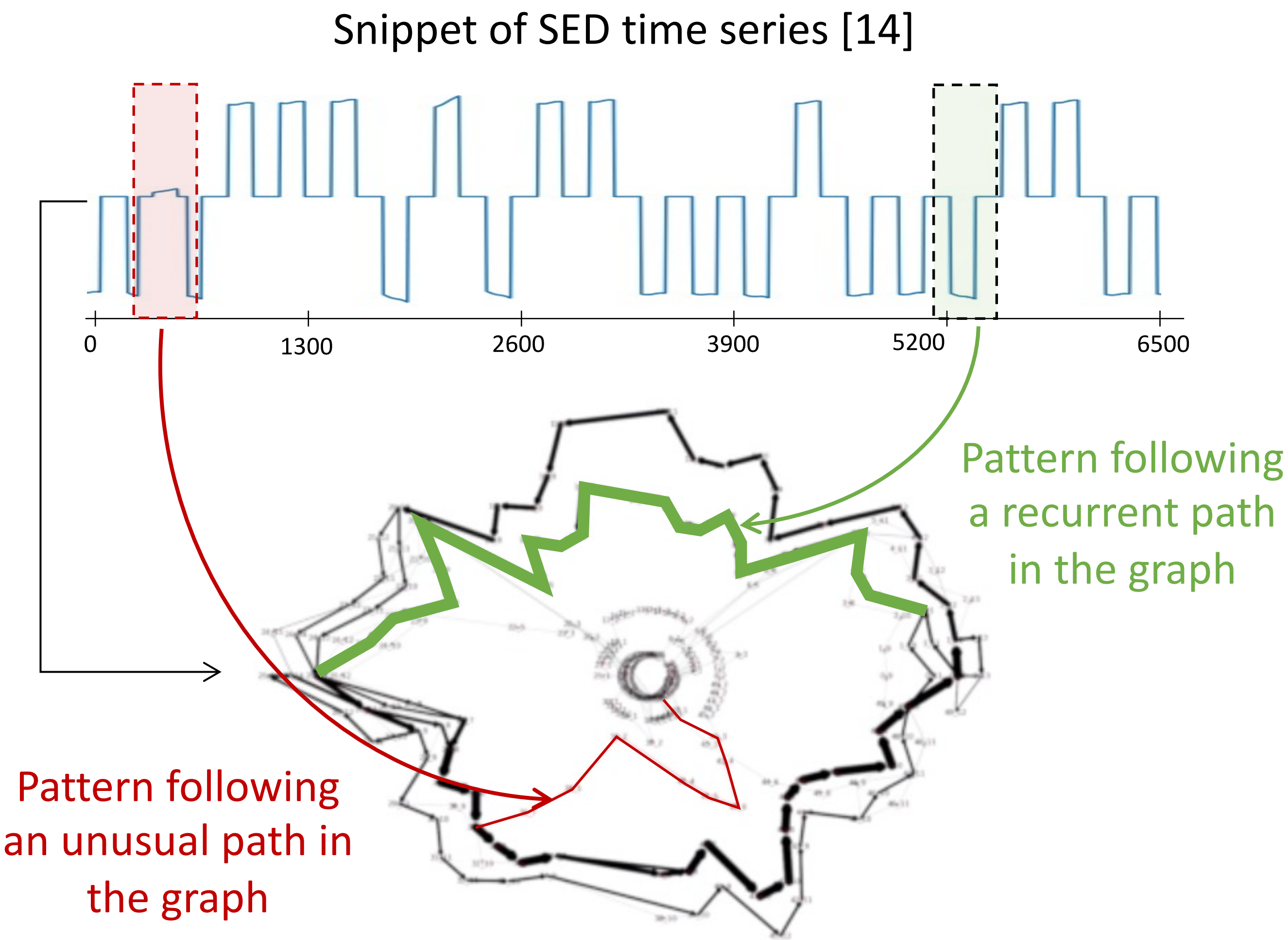
Unsupervised

Univariate

subsequence

[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

Anomaly Detection methods: *an Example*



Series2Graph [13]

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

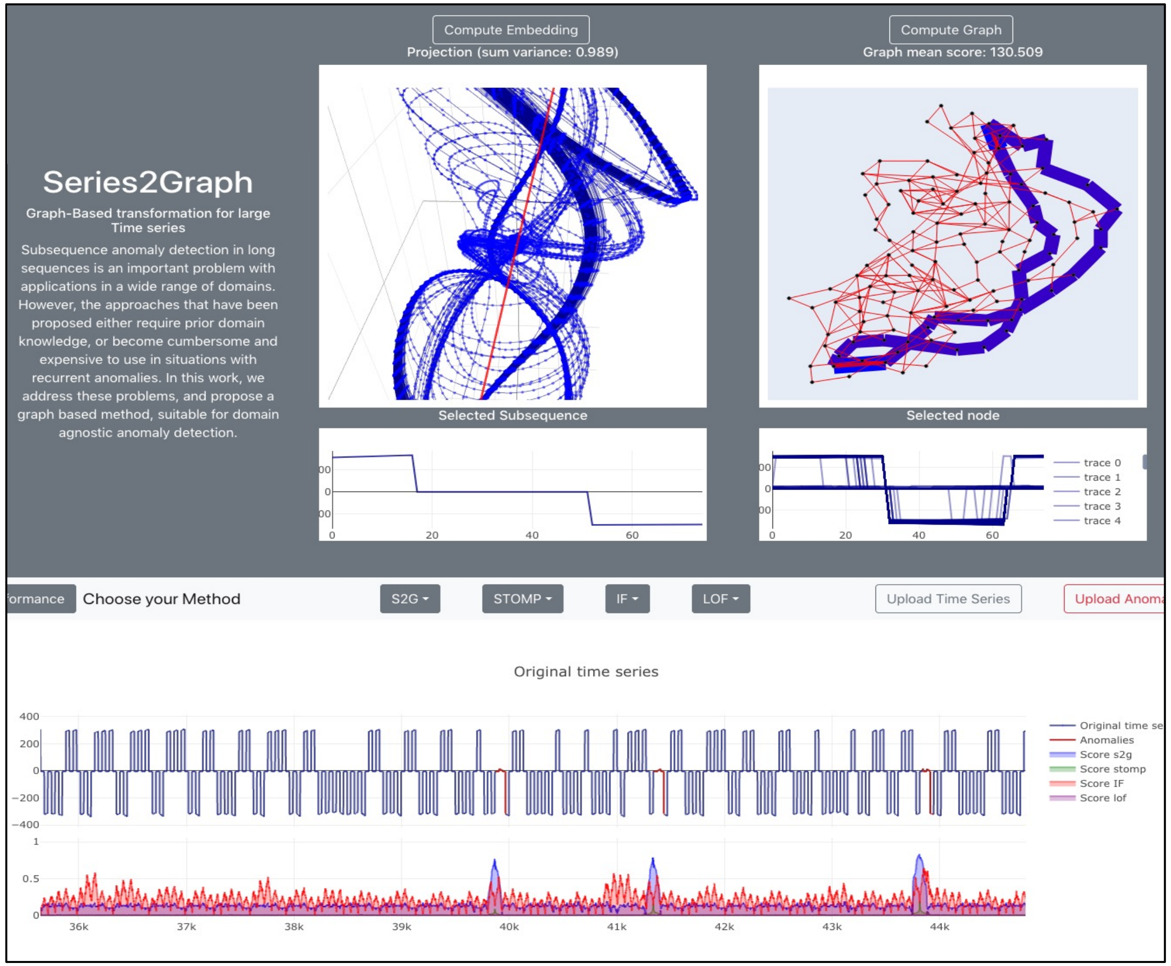
Unsupervised

Univariate

subsequence

[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

Anomaly Detection methods: *an Example*



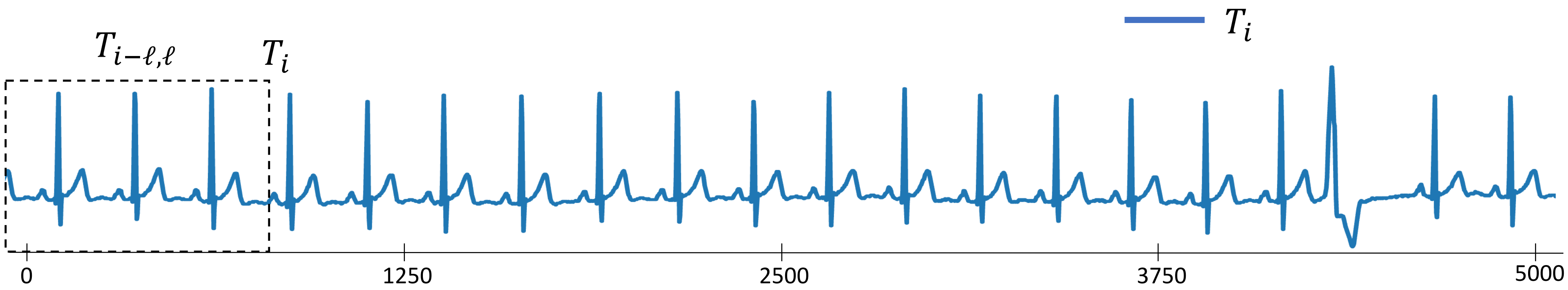
GraphAn [28]

An interactive tool to dive into the computation steps of Series2Graph :

[28] Paul Boniol, Themis Palpanas, Mohammed Meftah, and Emmanuel Remy. 2020. GraphAn: graph-based subsequence anomaly detection. Proc. VLDB Endow. 13, 12 (August 2020), 2941–2944.

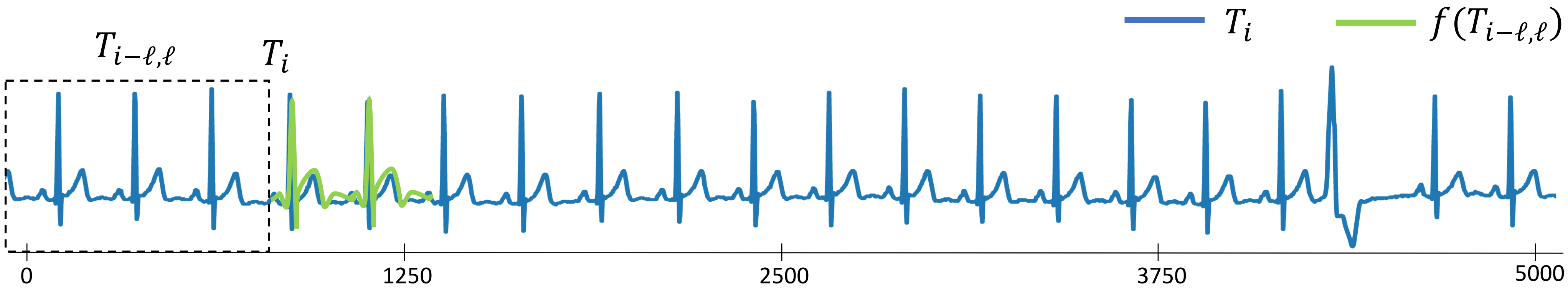
Anomaly Detection methods: *Forecasting-based*

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.



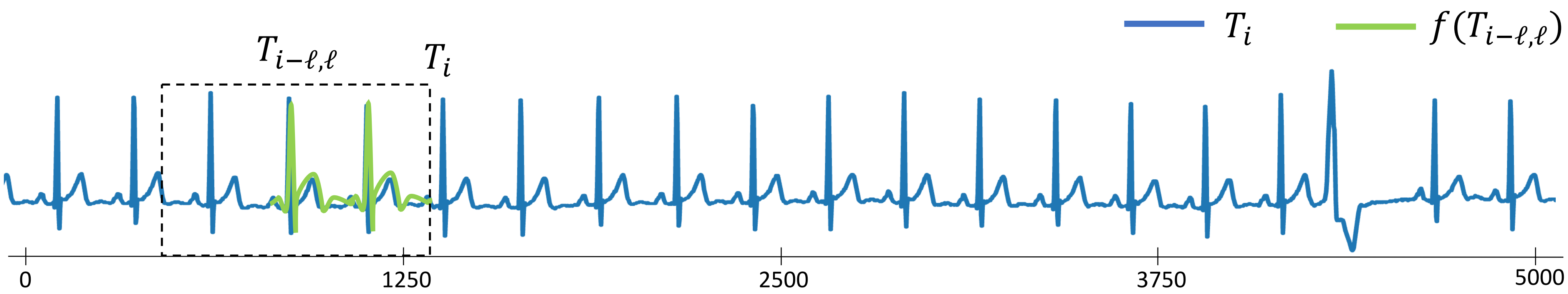
Anomaly Detection methods: *Forecasting-based*

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.



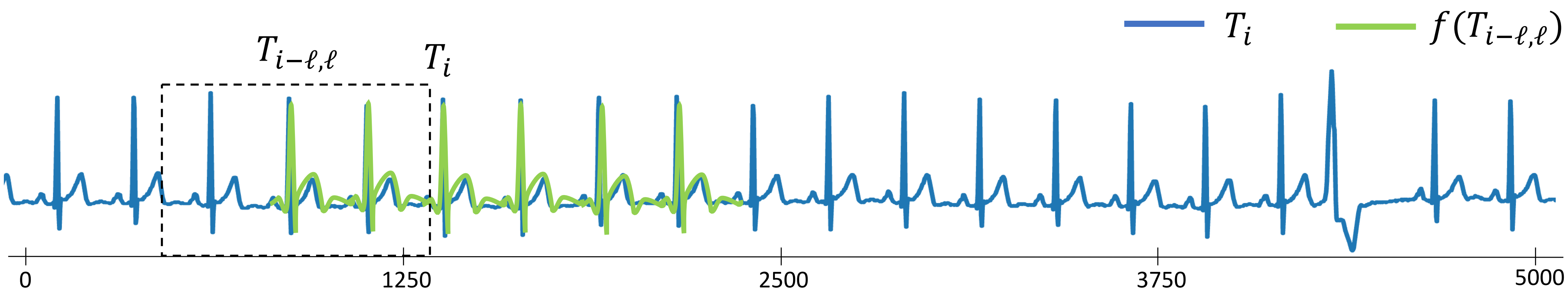
Anomaly Detection methods: *Forecasting-based*

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.



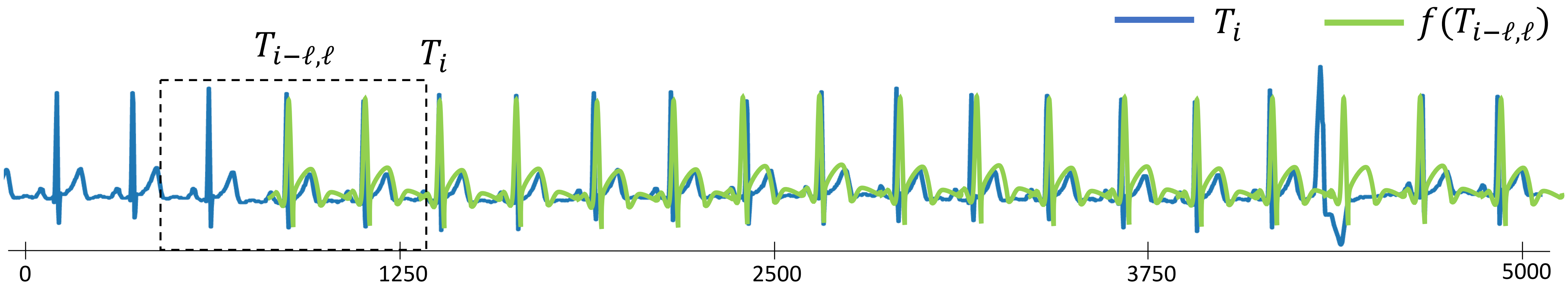
Anomaly Detection methods: *Forecasting-based*

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.



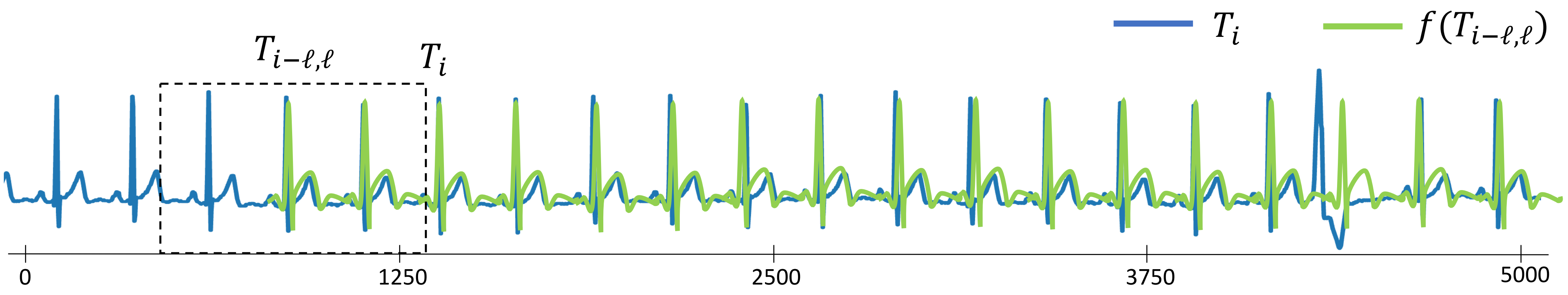
Anomaly Detection methods: *Forecasting-based*

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.

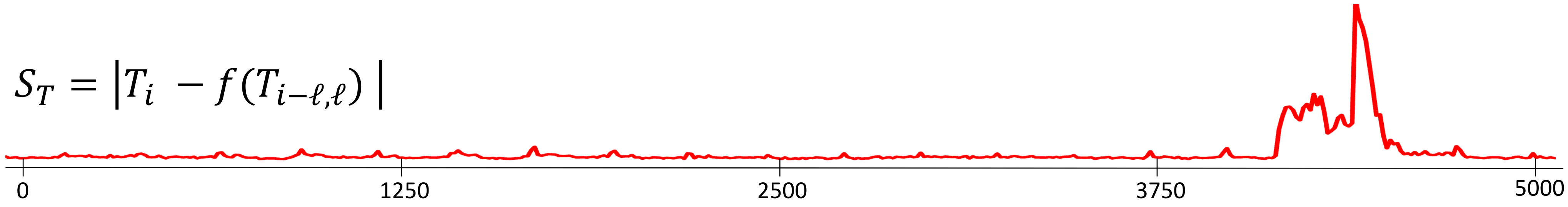


Anomaly Detection methods: *Forecasting-based*

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.

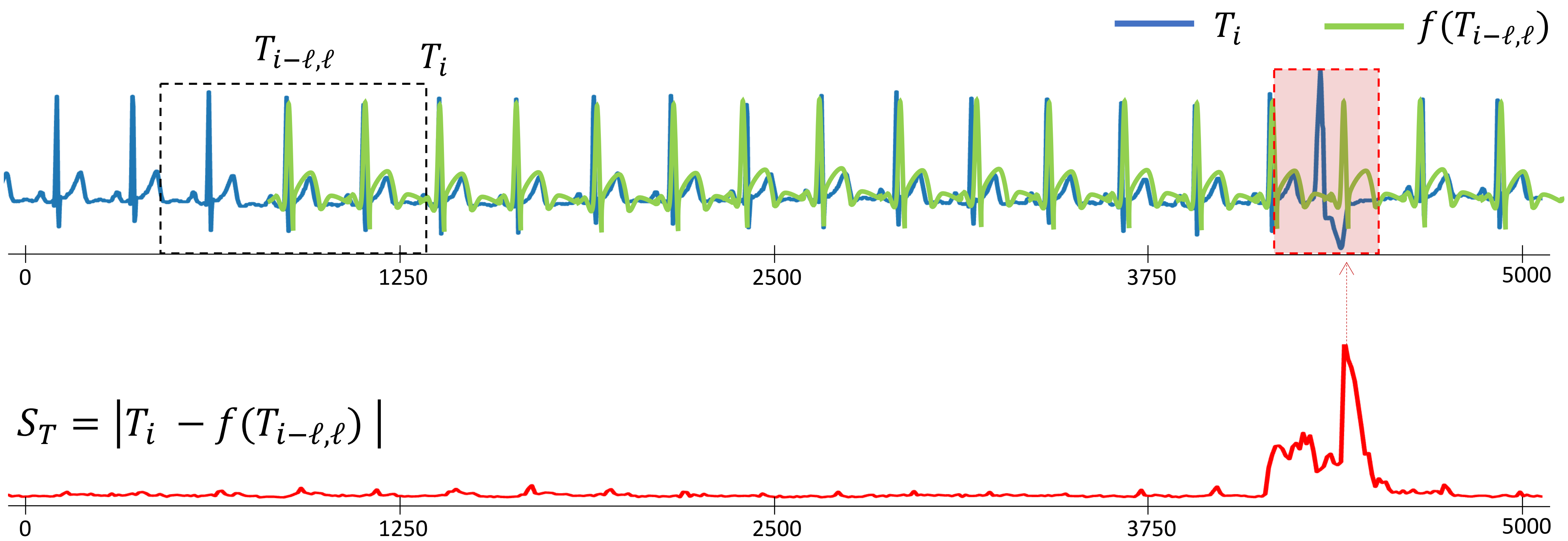


$S_T = \left|T_i - f(T_{i-\ell,\ell})\right|$

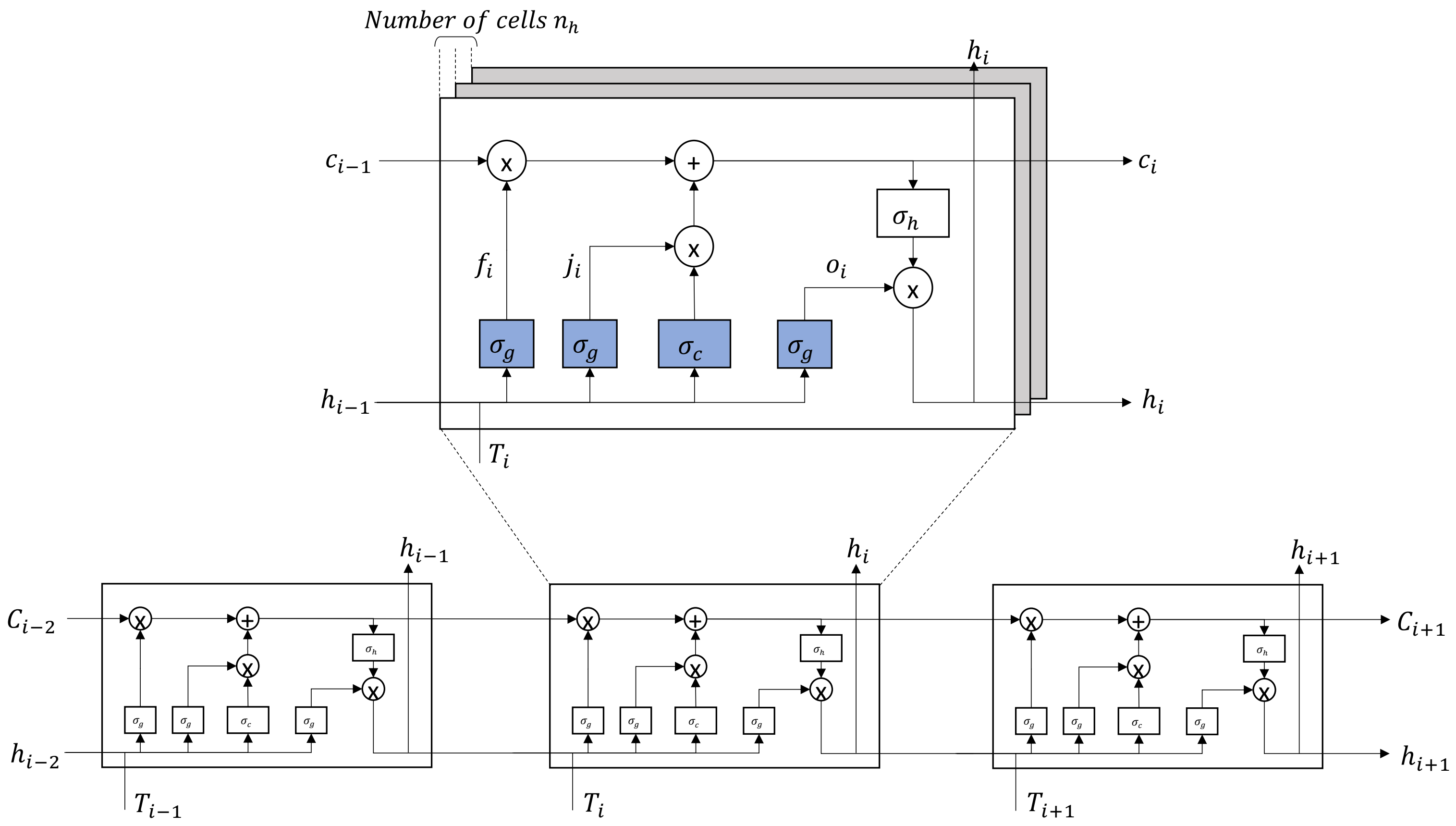


Anomaly Detection methods: *Forecasting-based*

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

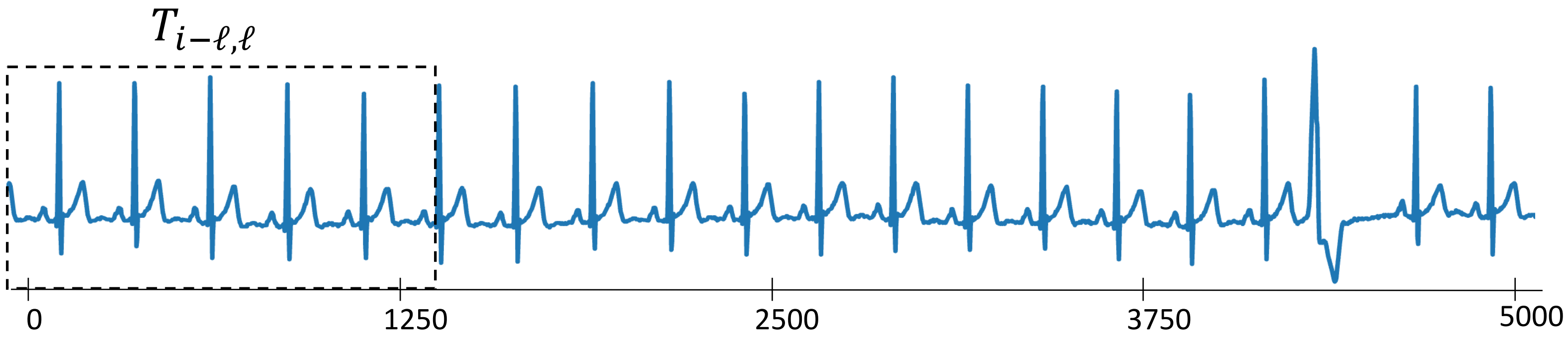
Semi-supervised

Univariate/Multivariate

Point/sequence

[15] Pankaj Malhotra, Lovekesh Vig, Gautam Shro, and Puneet Agarwal. 2015. Long Short Term Memory Networks for Anomaly Detection in Time Series. (2015).

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.

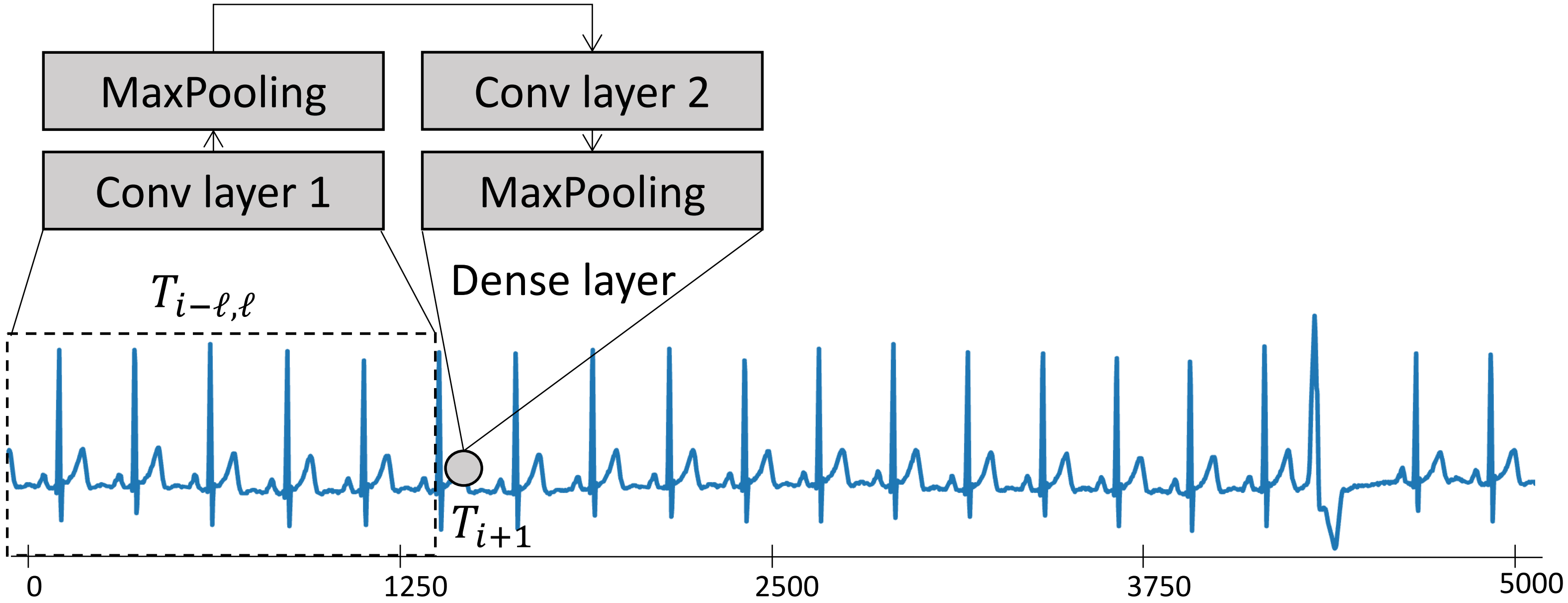
Semi-supervised

Univariate/Multivariate

Point/sequence

[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.

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.

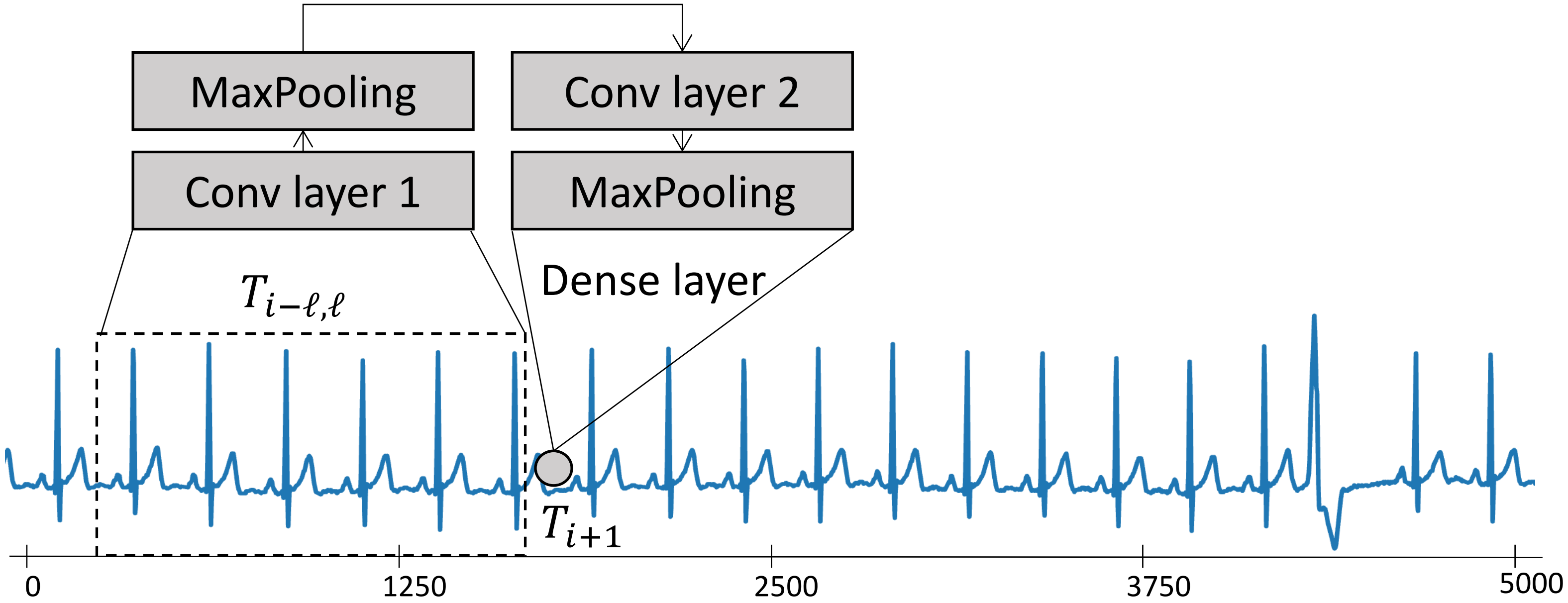
Semi-supervised

Univariate/Multivariate

Point/sequence

[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.

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.

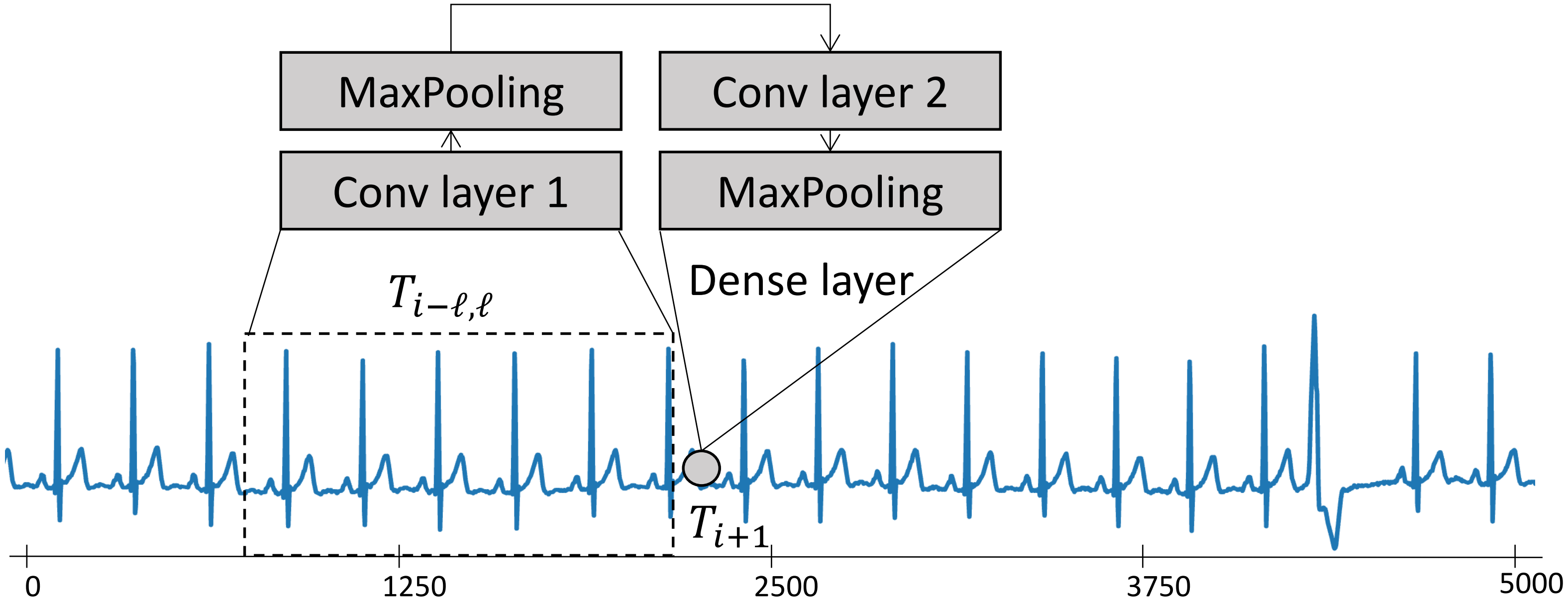
Semi-supervised

Univariate/Multivariate

Point/sequence

[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.

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

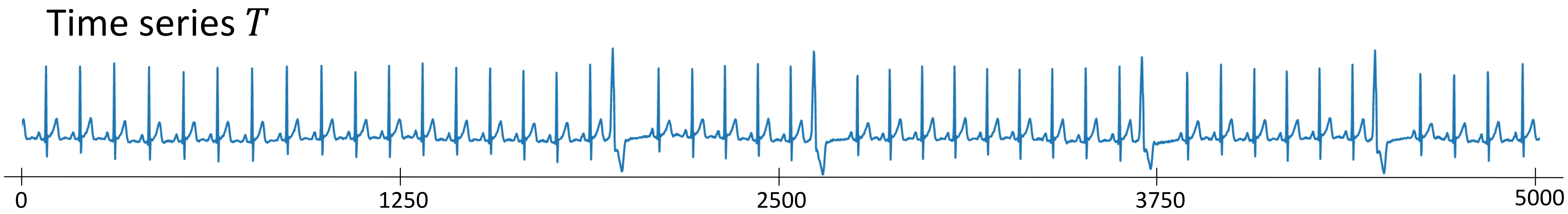
[16] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed. 2019. DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series. IEEE Access 7 (2019), 1991–2005.

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.

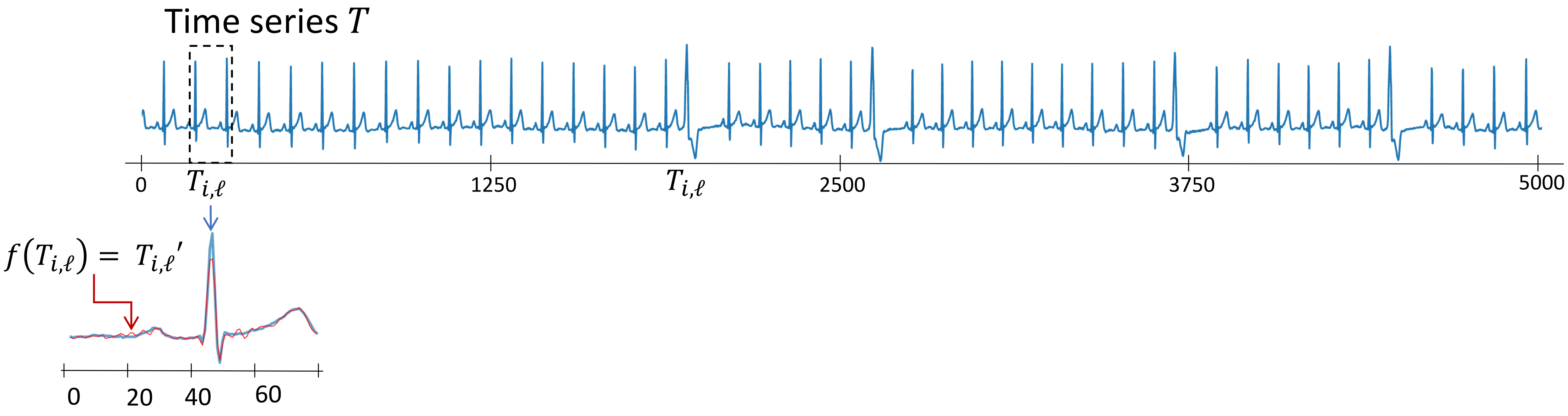
Anomaly Detection methods: *Reconstruction-based*

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.



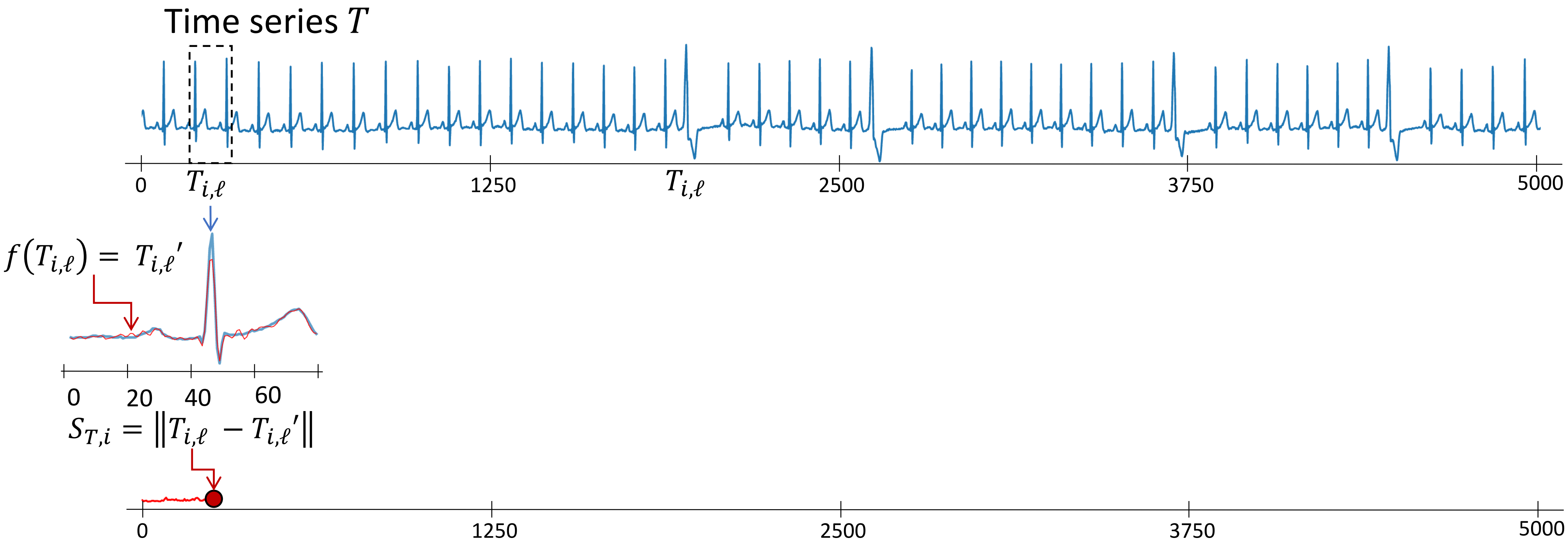
Anomaly Detection methods: *Reconstruction-based*

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.



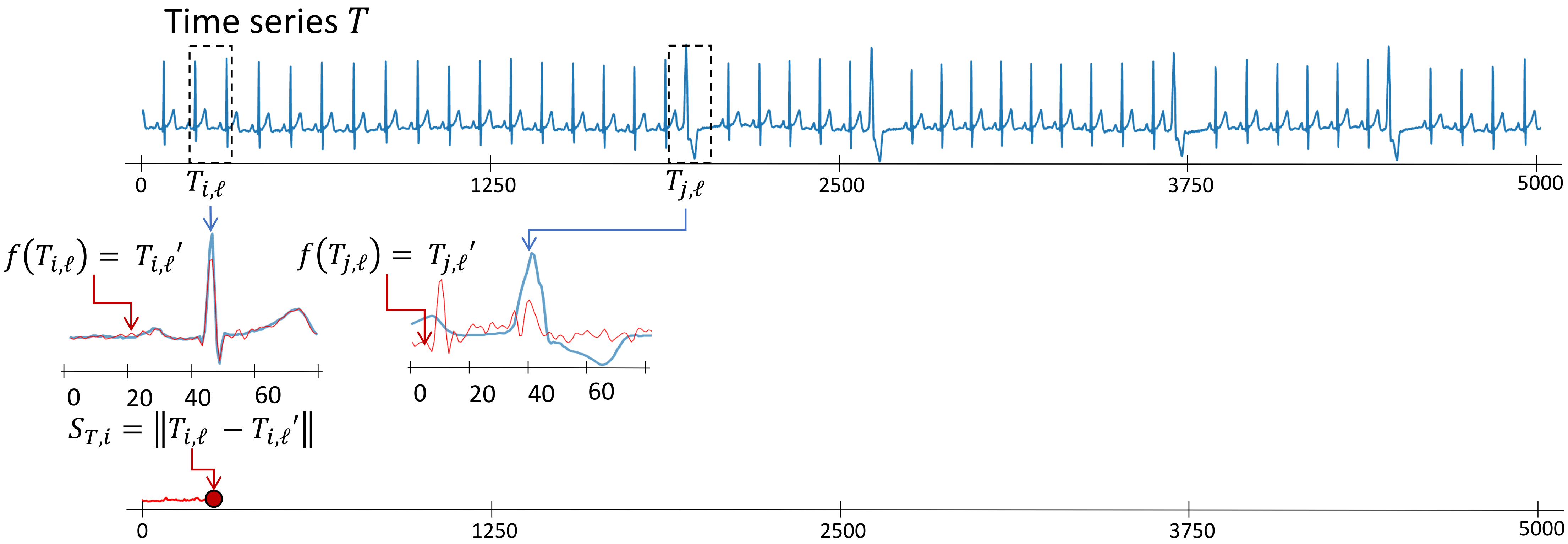
Anomaly Detection methods: *Reconstruction-based*

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.



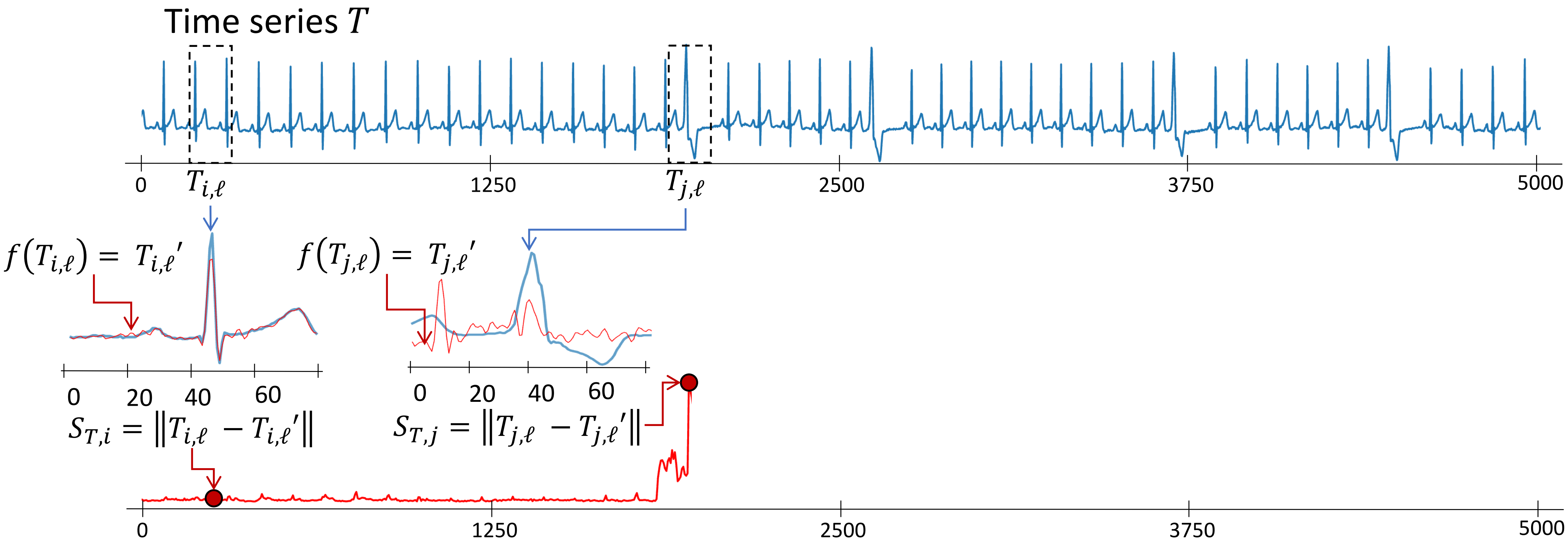
Anomaly Detection methods: *Reconstruction-based*

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.



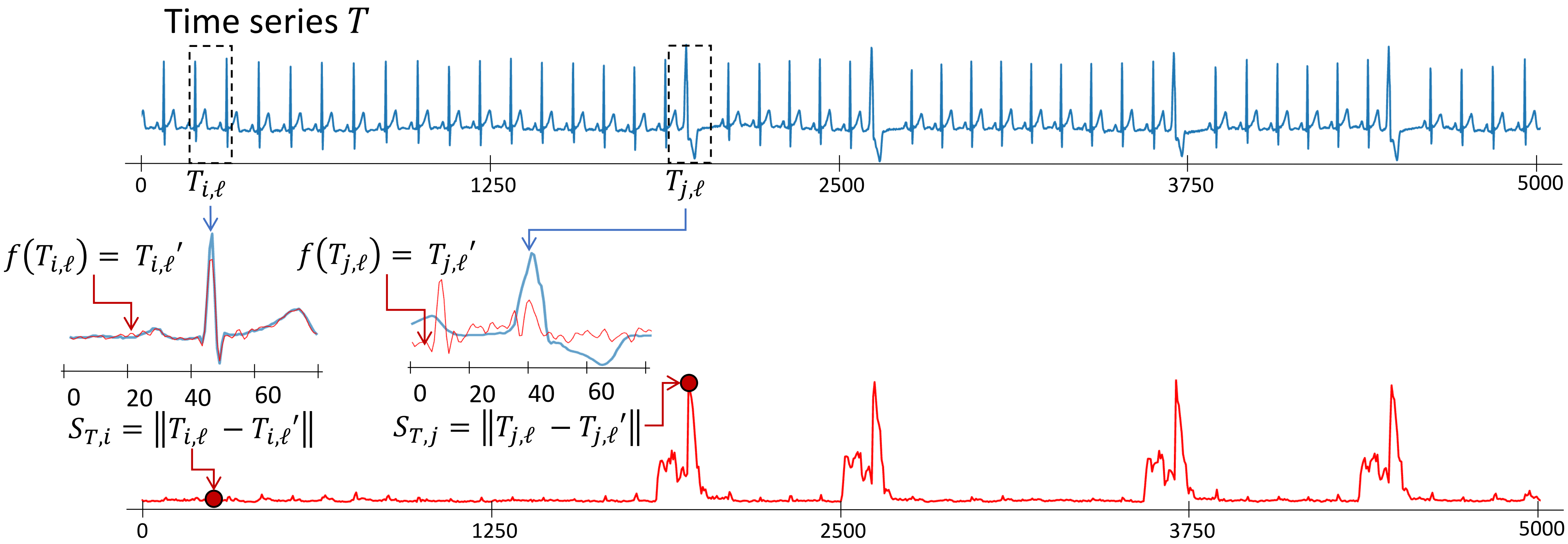
Anomaly Detection methods: *Reconstruction-based*

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.



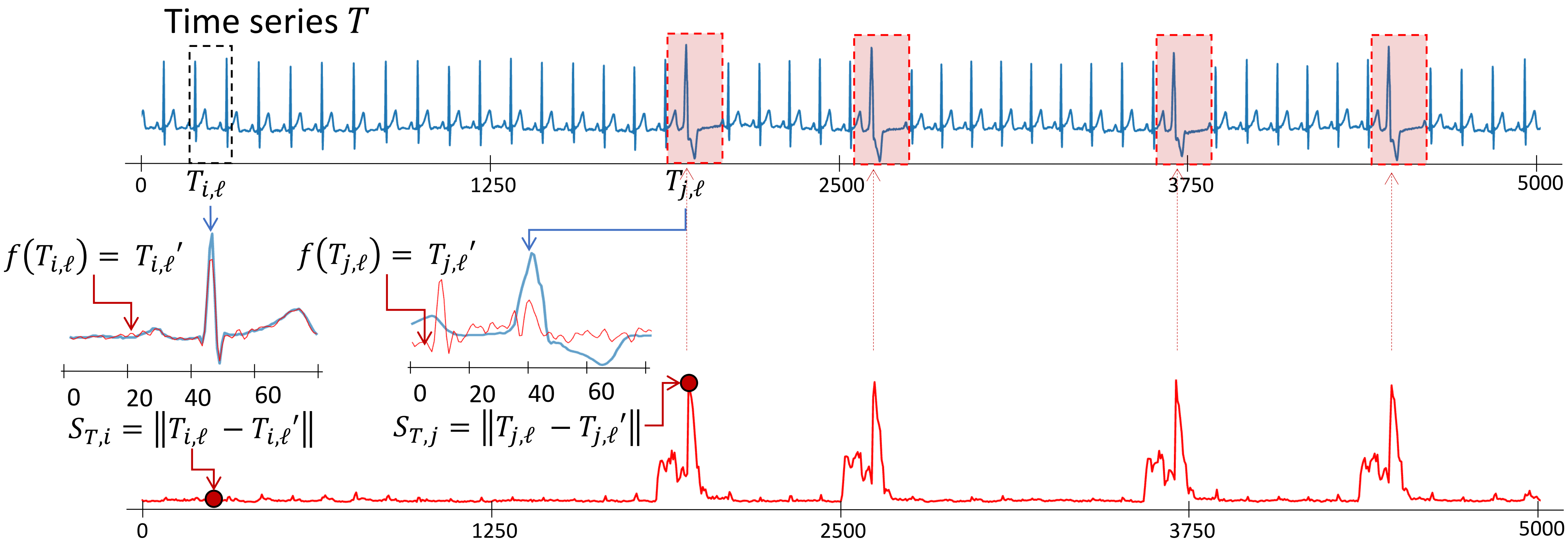
Anomaly Detection methods: *Reconstruction-based*

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.

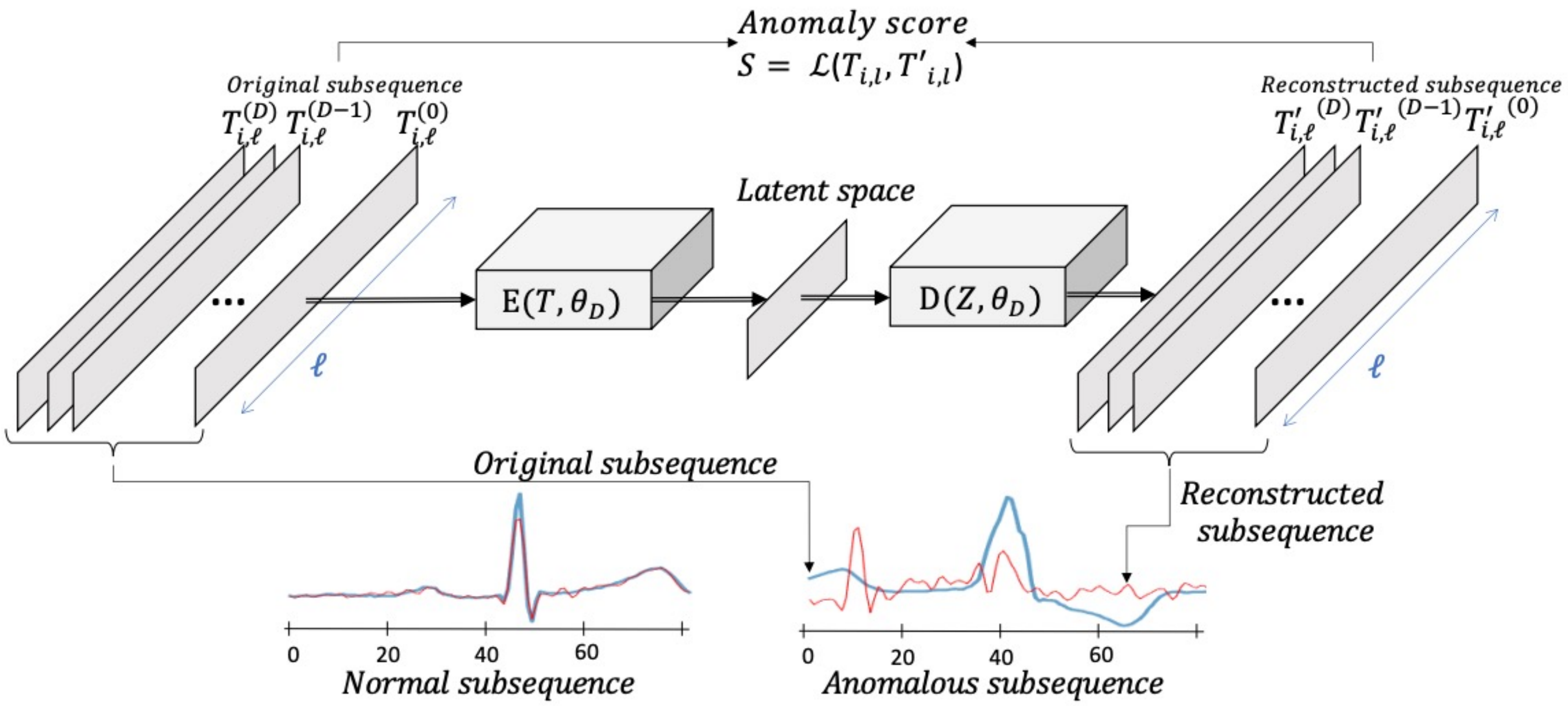


Anomaly Detection methods: *Reconstruction-based*

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)

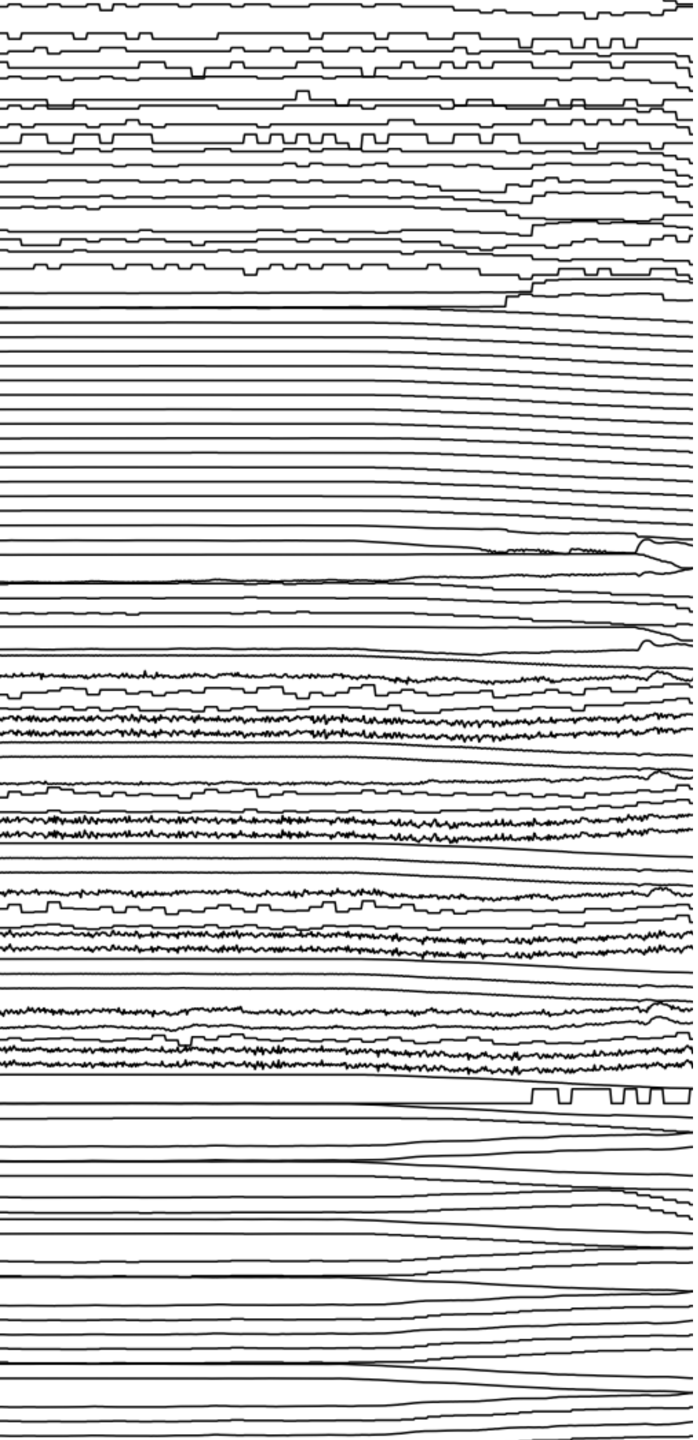
Neural Network composed of an encoder (that reduce the dimensionality) and decoder that reconstruct the time series. The objective is to minimize the reconstruction error.

Semi-supervised

Univariate/Multivariate

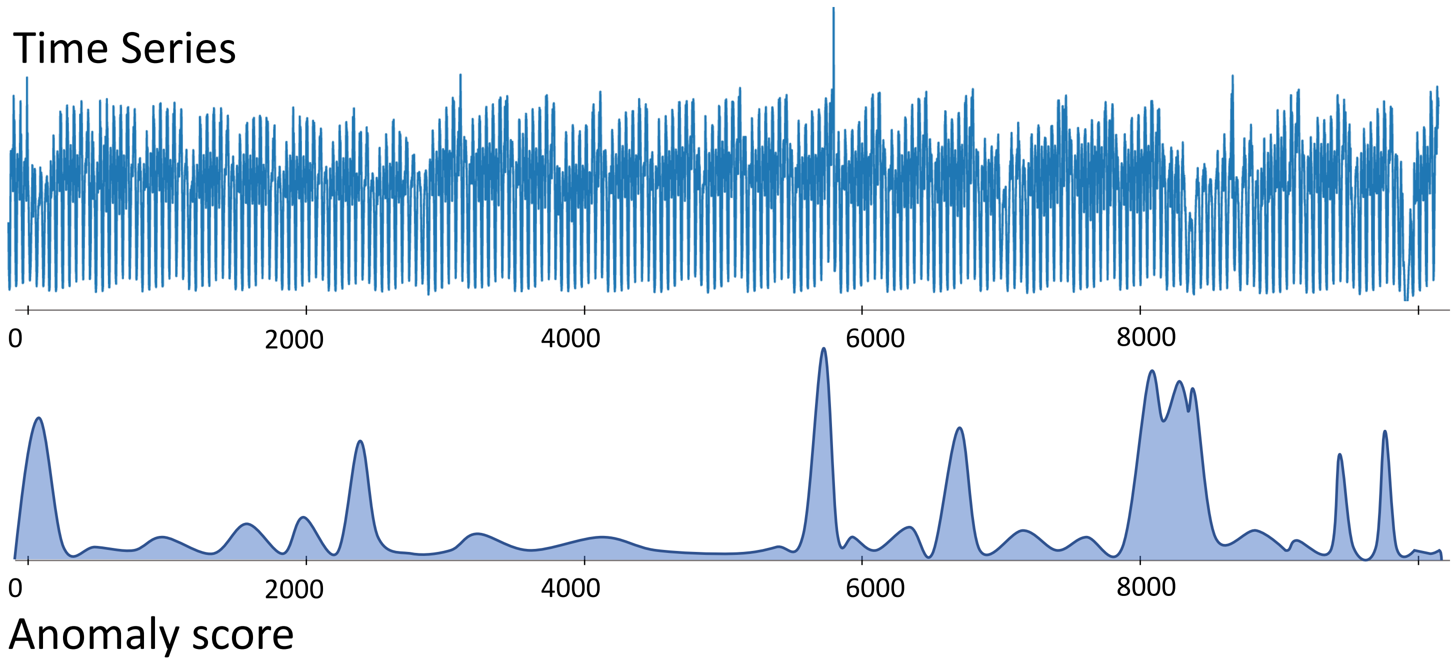
Point/sequence

[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).

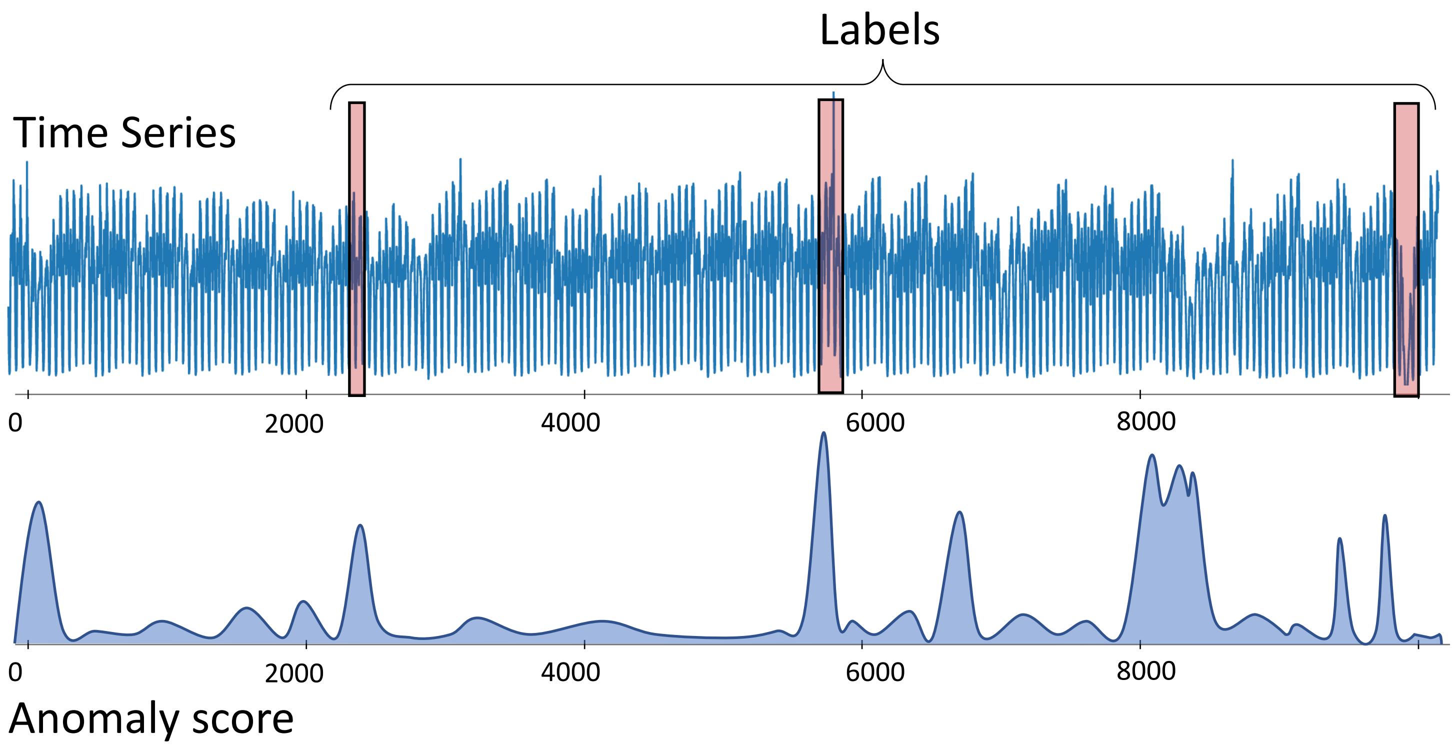


Part 3: Evaluation Measures

Evaluation measures: *A general overview*

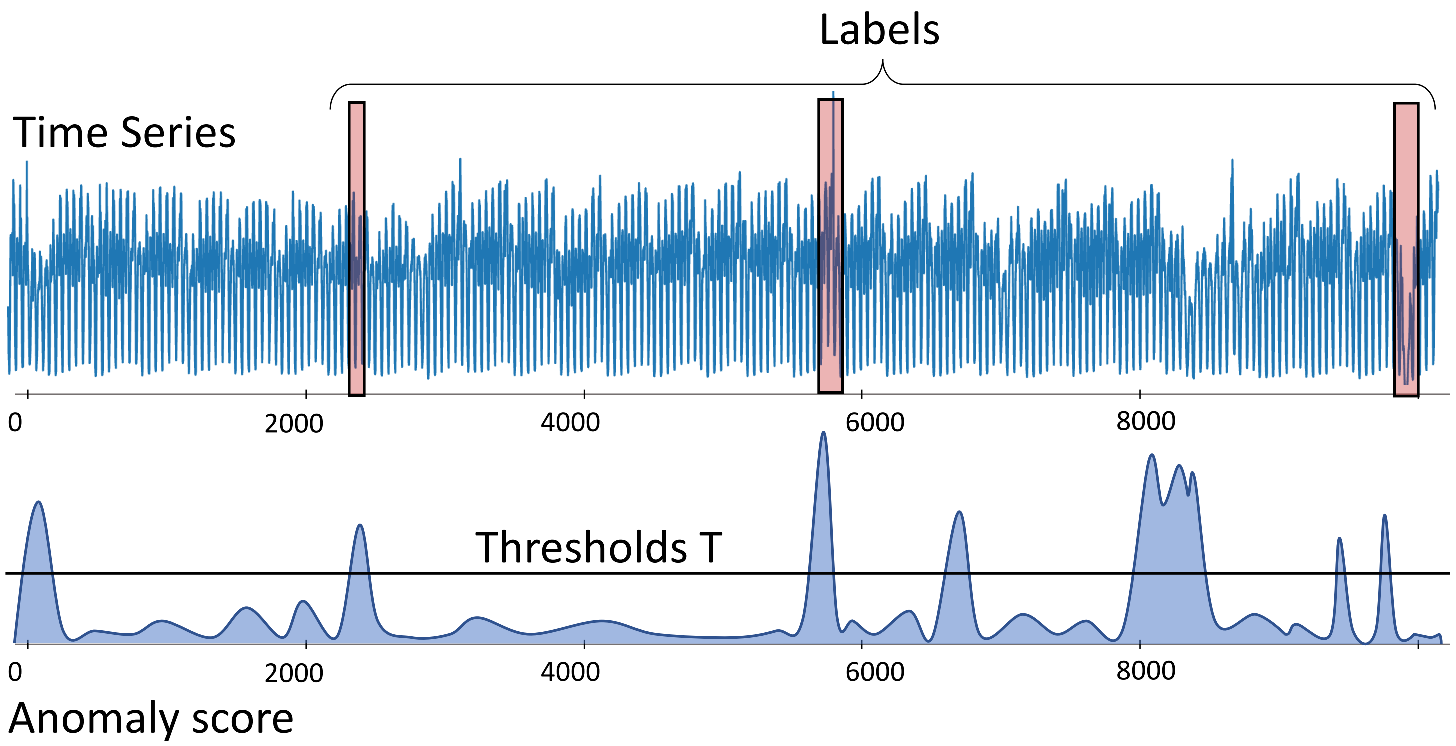


Evaluation measures: *A general overview*



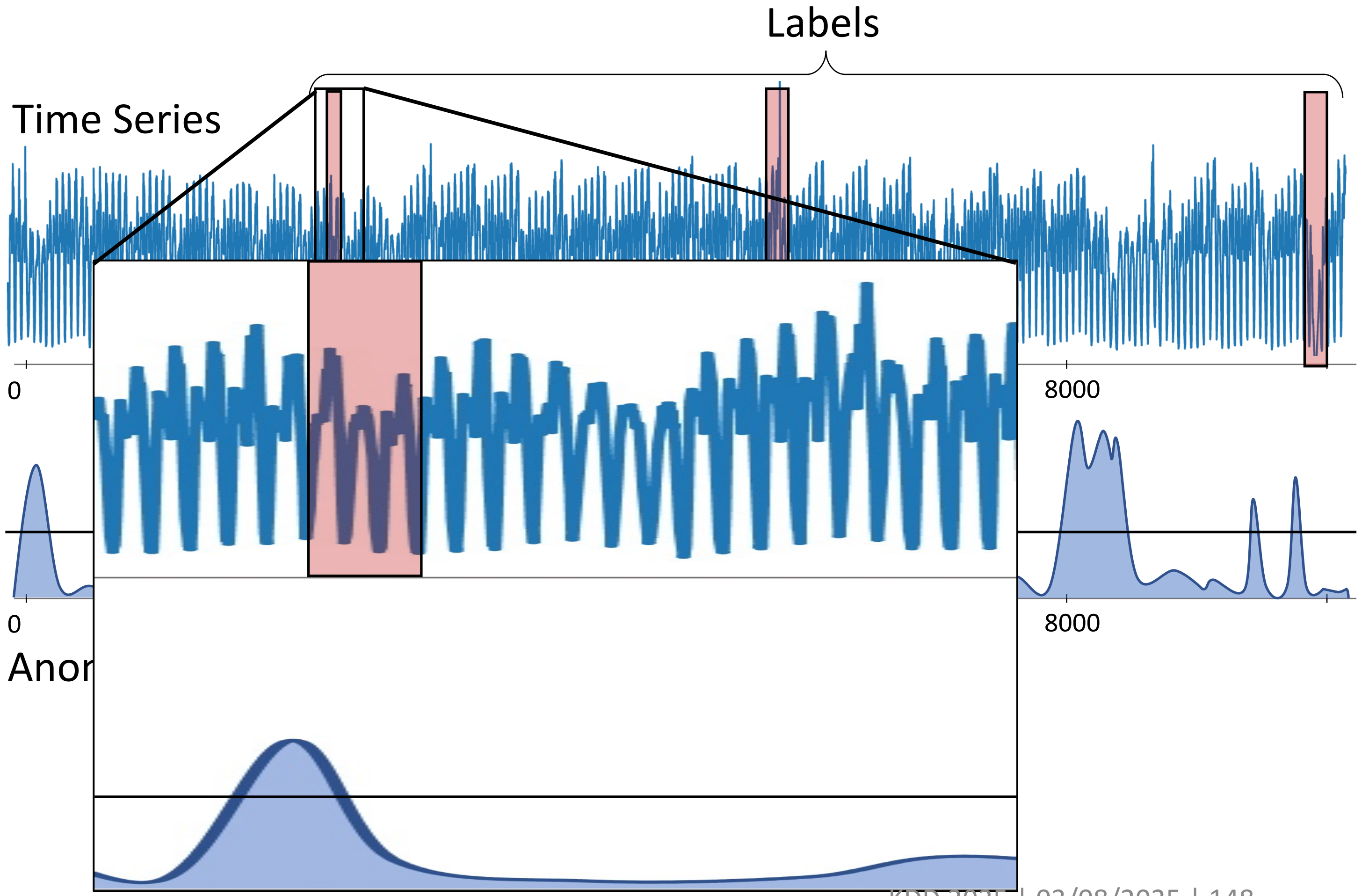
Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:



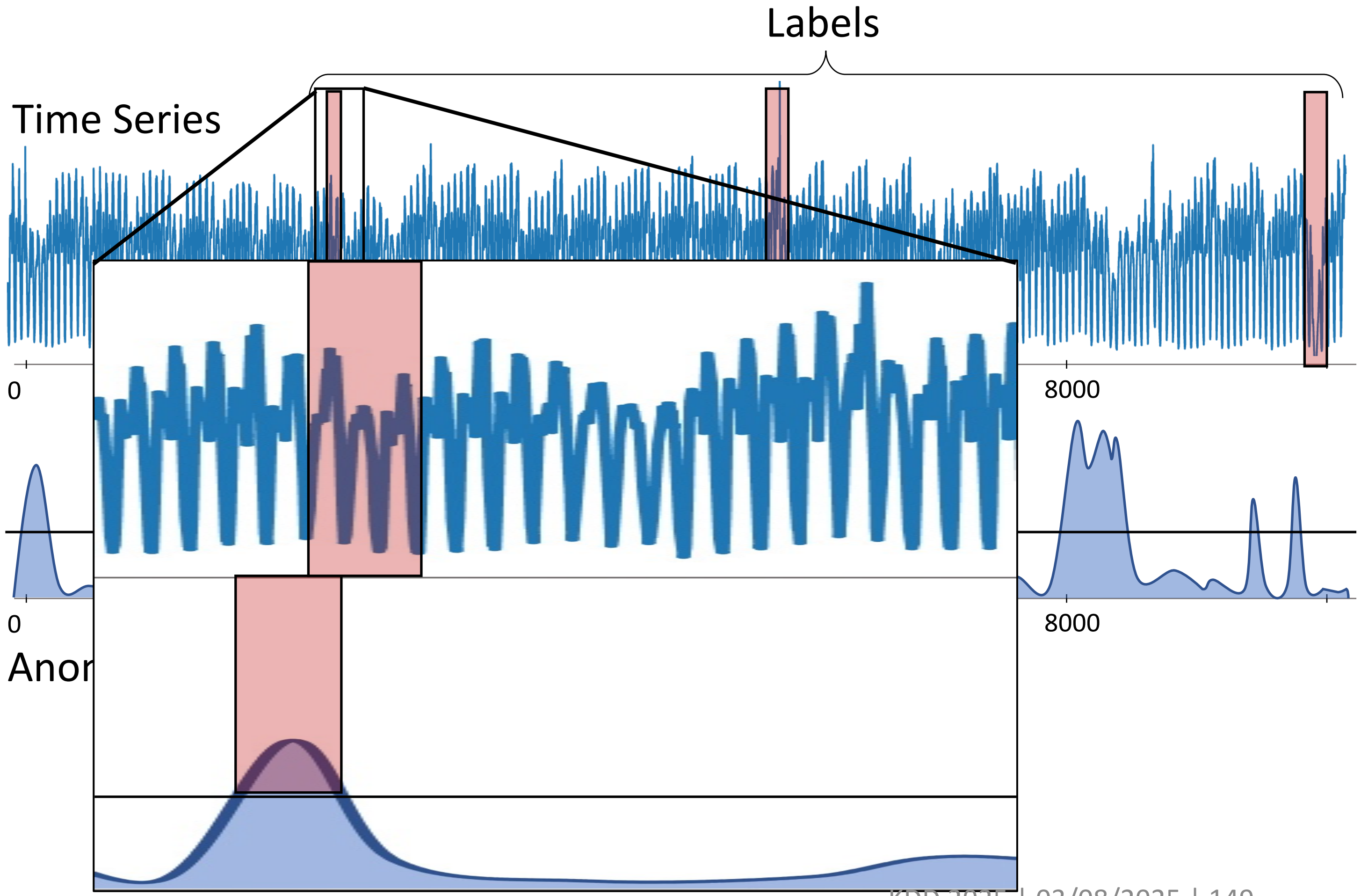
Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:



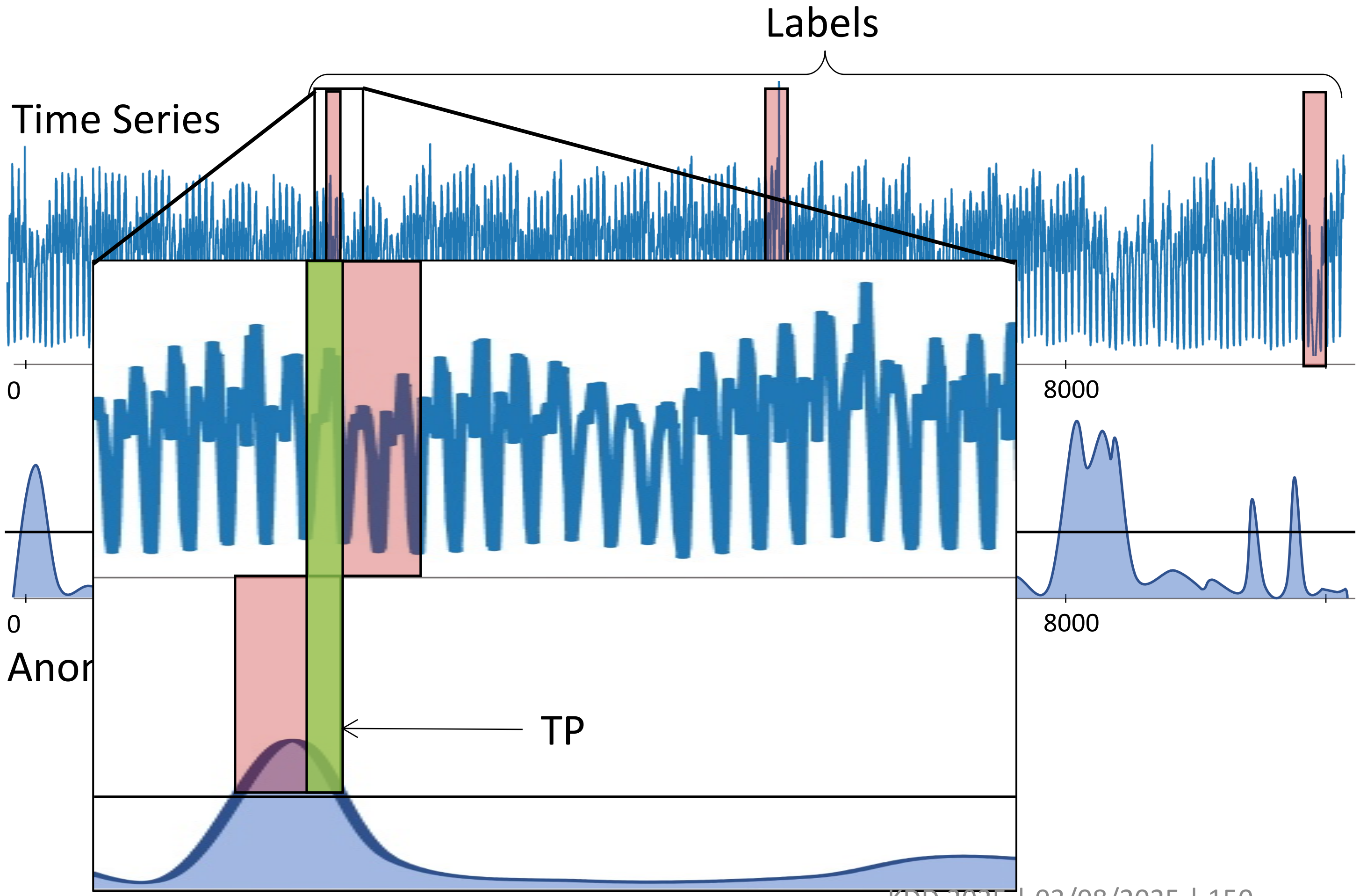
Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:



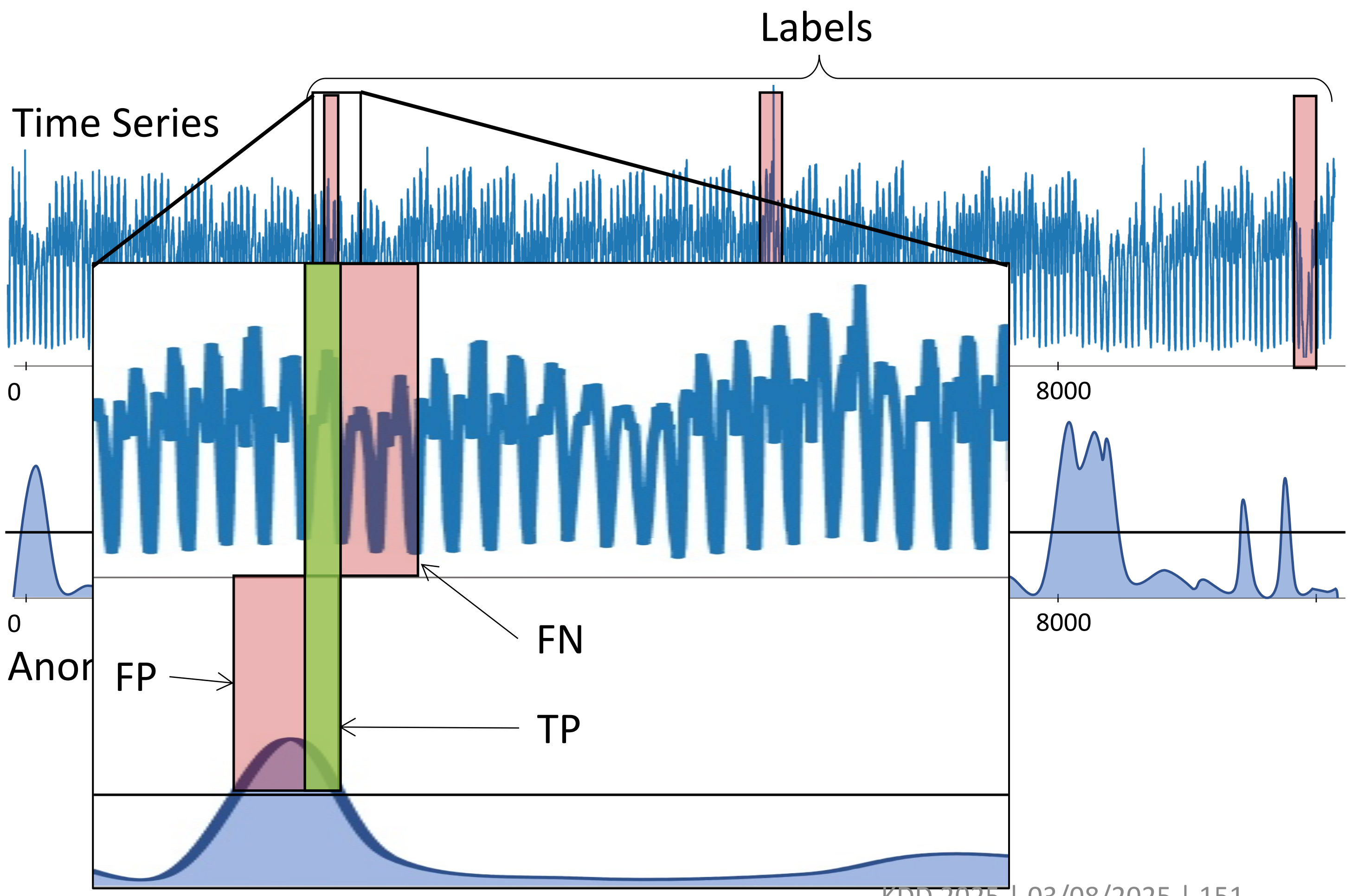
Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:



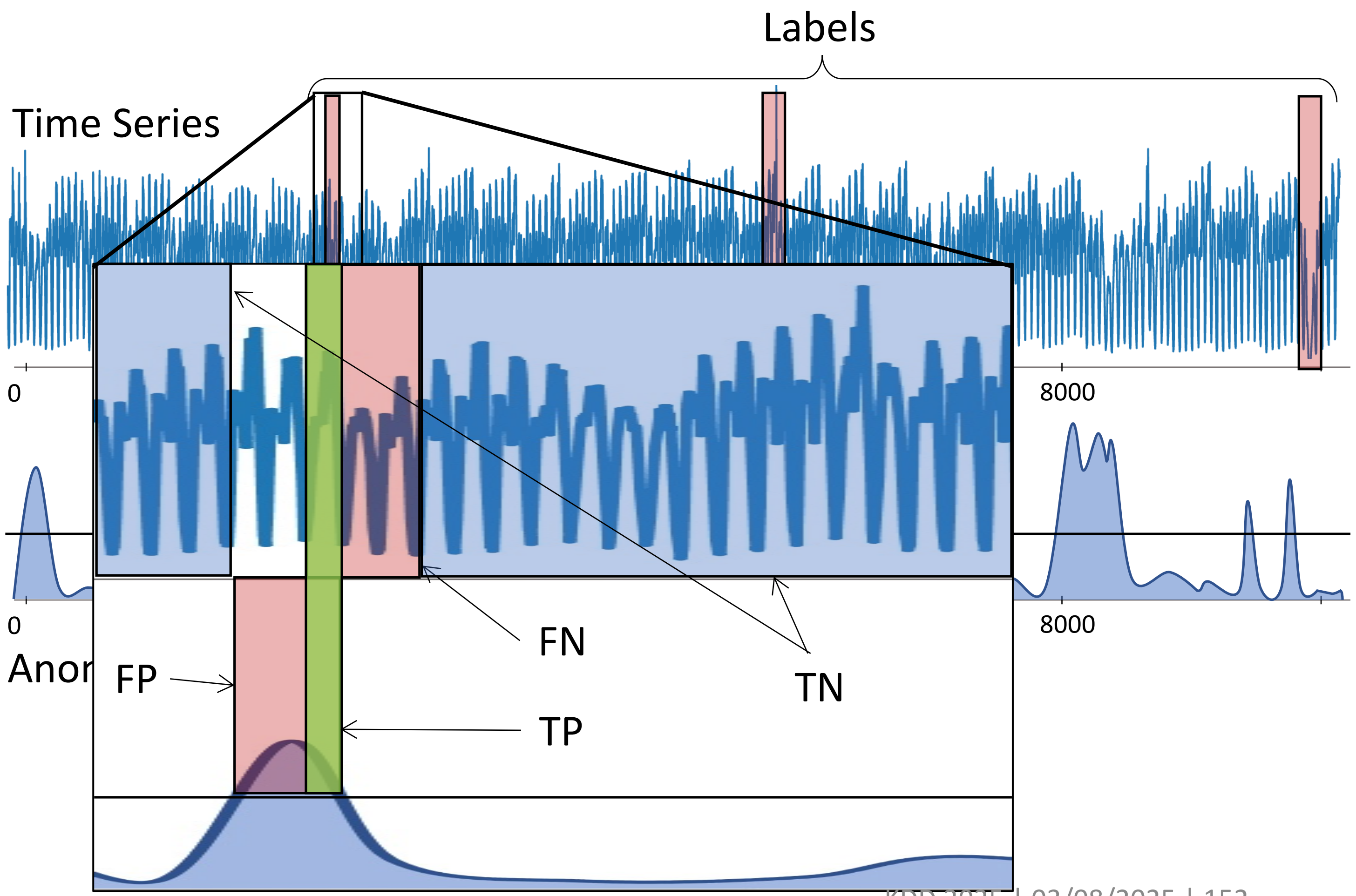
Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:



Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:

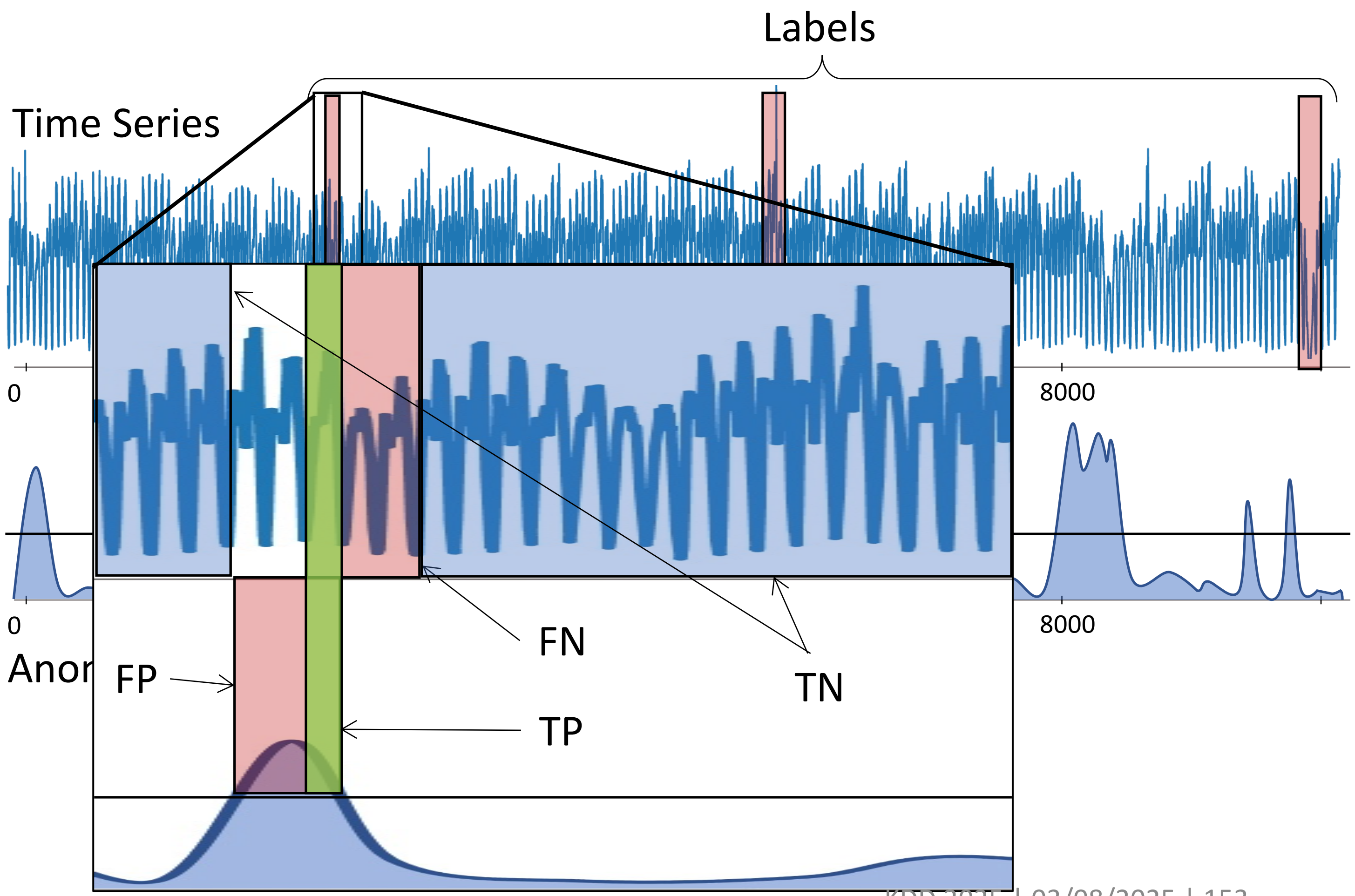


Evaluation measures: *Threshold-based*

Threshold-based Evaluation Measures:

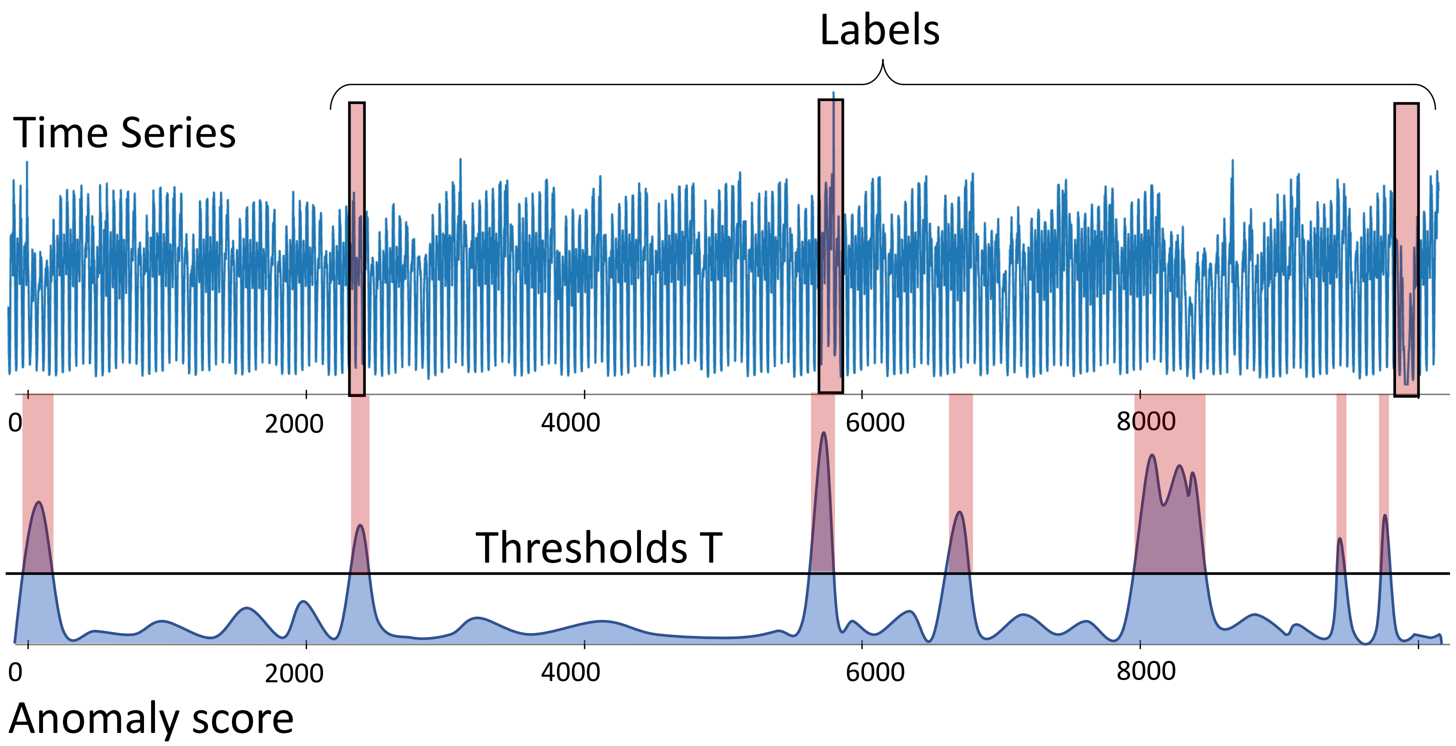
- Precision: $\frac{TP}{TP+FP}$
- Recall (true positive rate): $\frac{TP}{TP+FN}$
- False positive rate: $\frac{FP}{FP+TN}$

- F-score: $\frac{(1+\beta^2)*Precision}{\beta^2*Precision+Recall}$



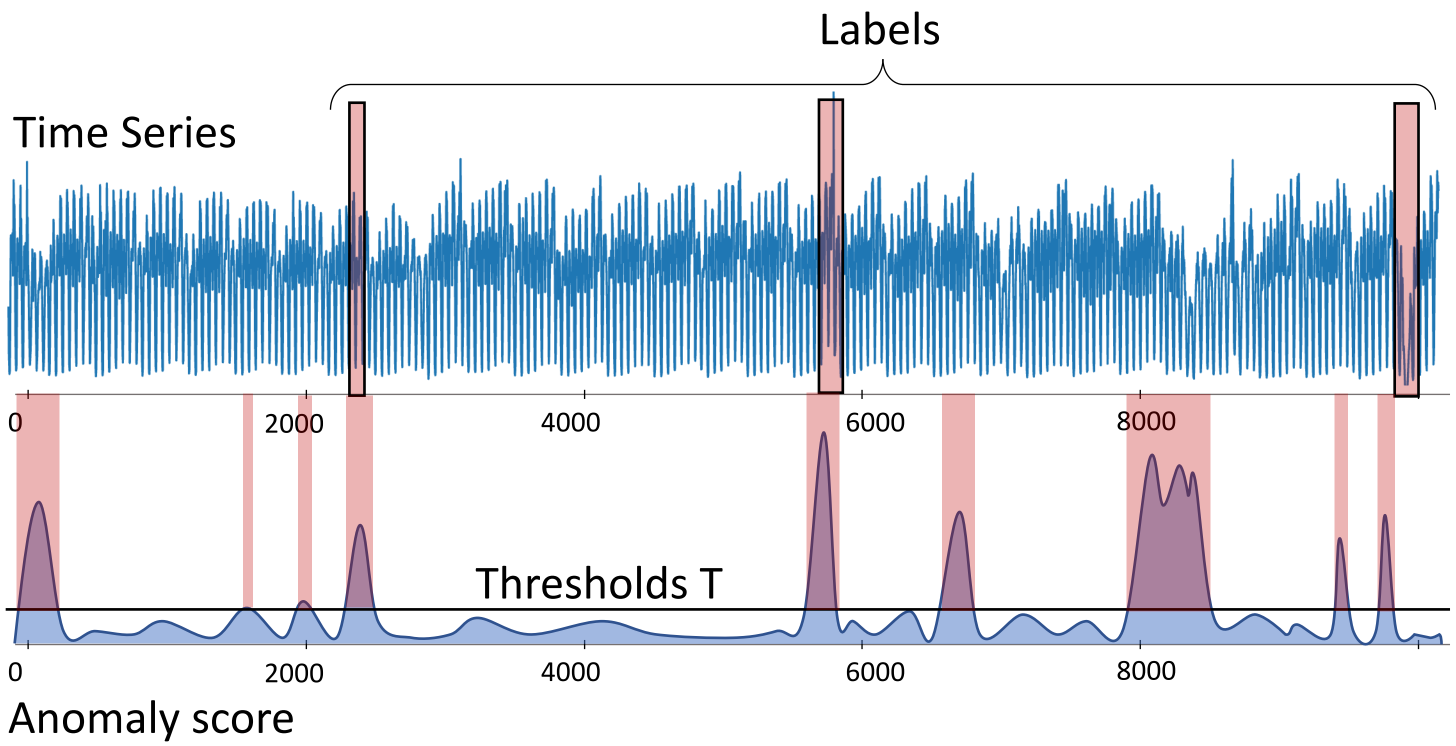
Evaluation measures: *AUC-based*

How do we set the threshold?



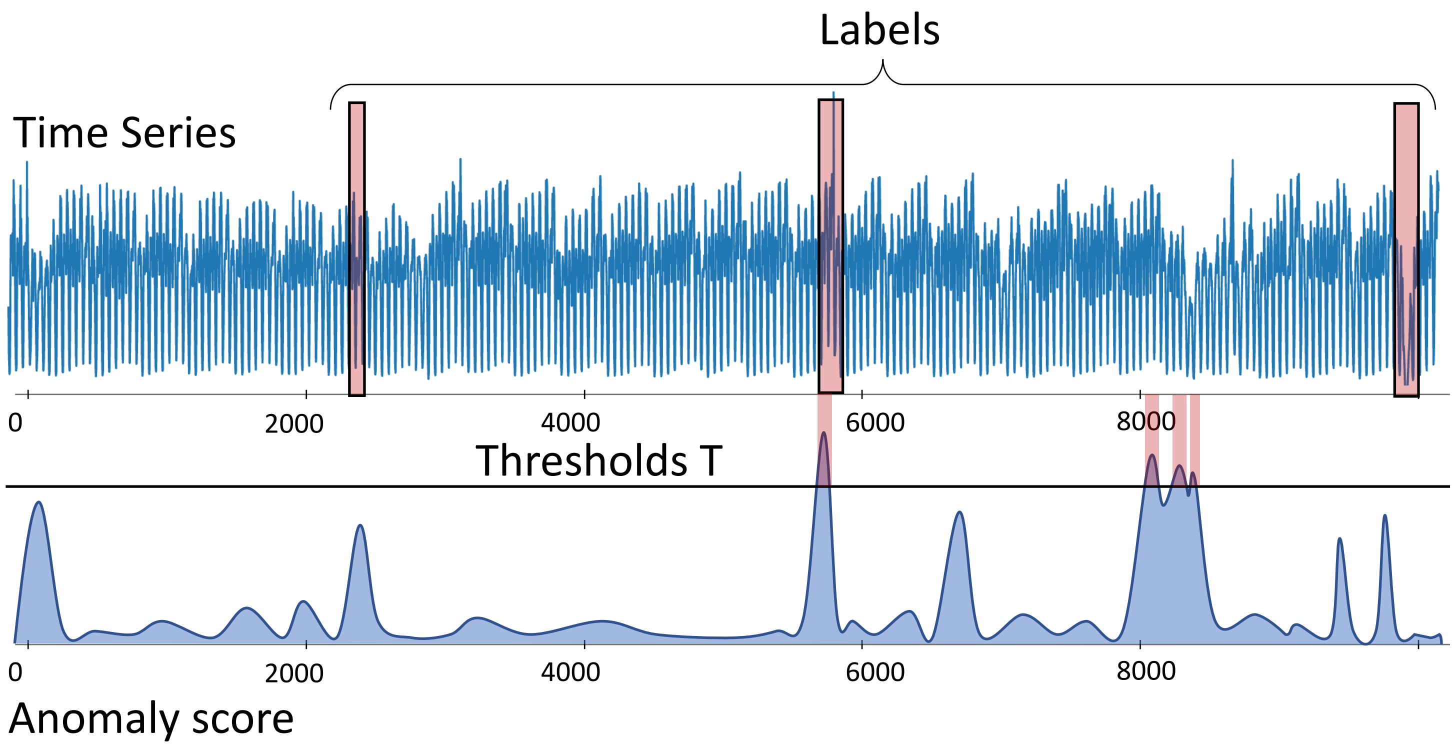
Evaluation measures: *AUC-based*

How do we set the threshold?



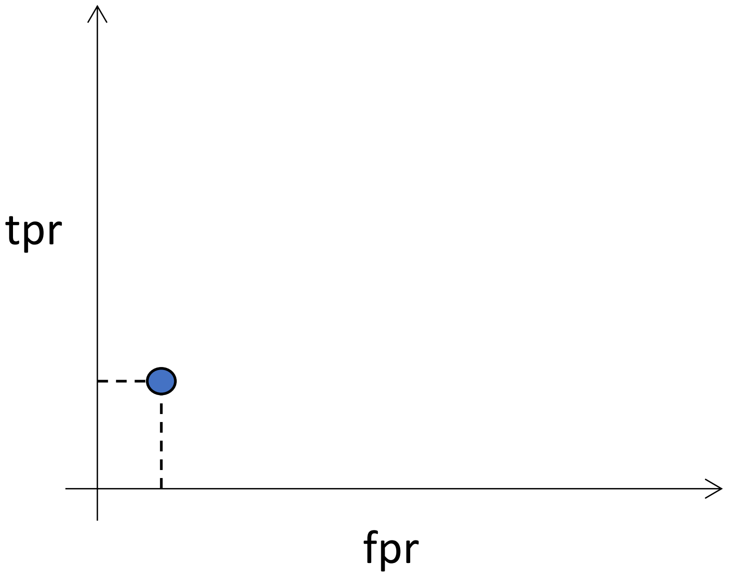
Evaluation measures: *AUC-based*

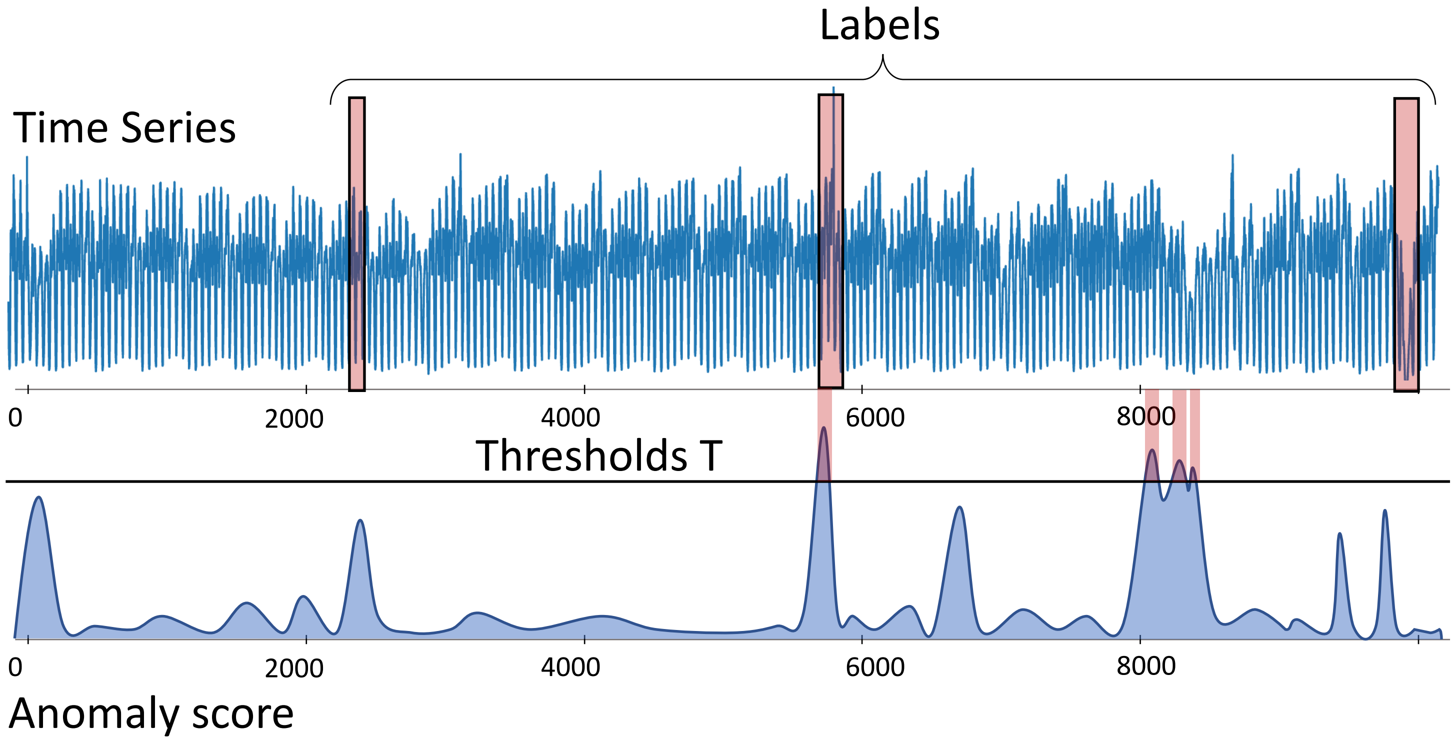
How do we set the threshold?



Evaluation measures: *AUC-based*

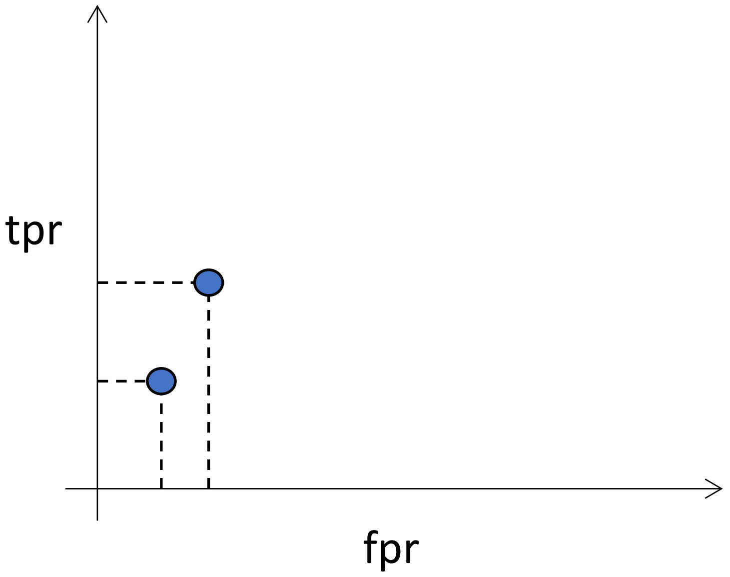
AUC-based Evaluation Measures:

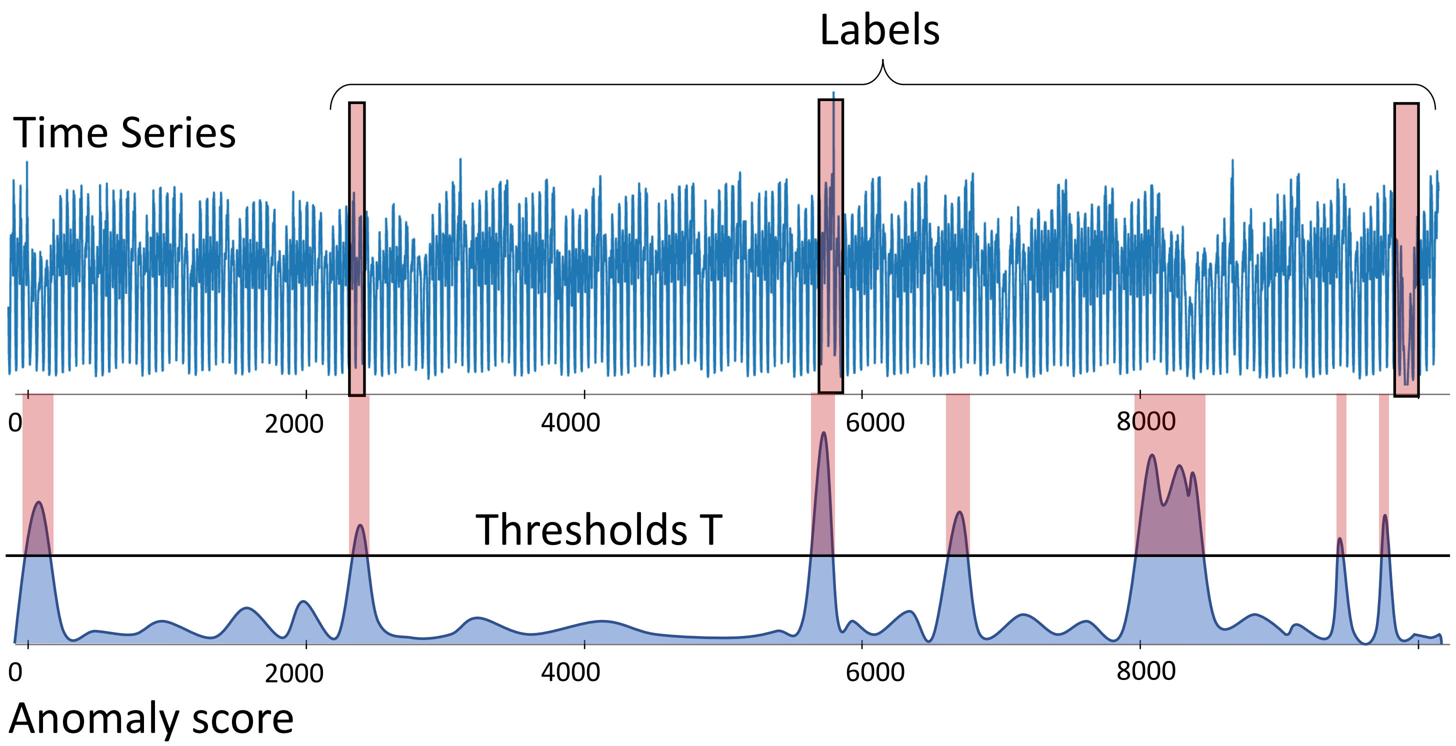




Evaluation measures: *AUC-based*

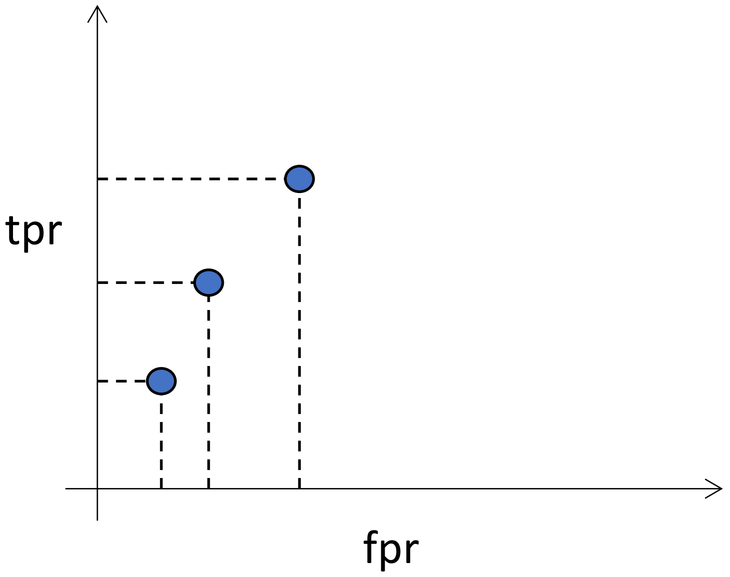
AUC-based Evaluation Measures:

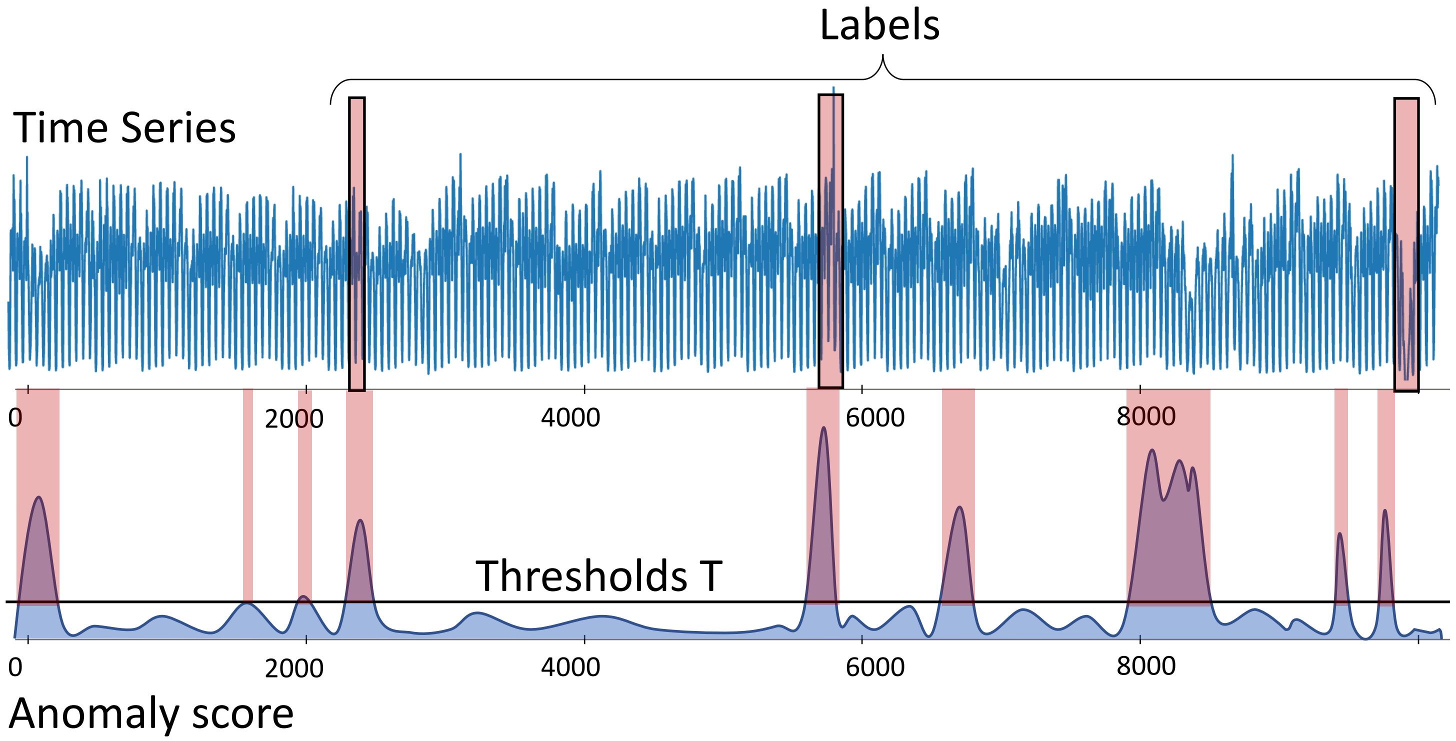




Evaluation measures: *AUC-based*

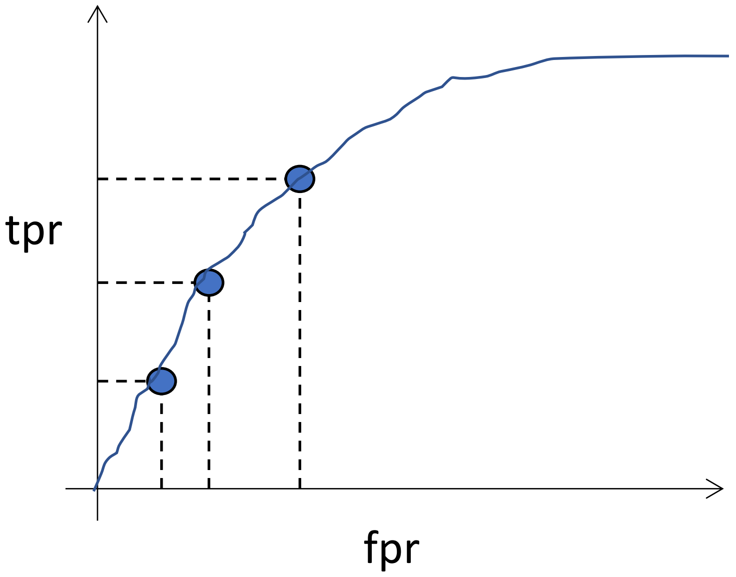
AUC-based Evaluation Measures:

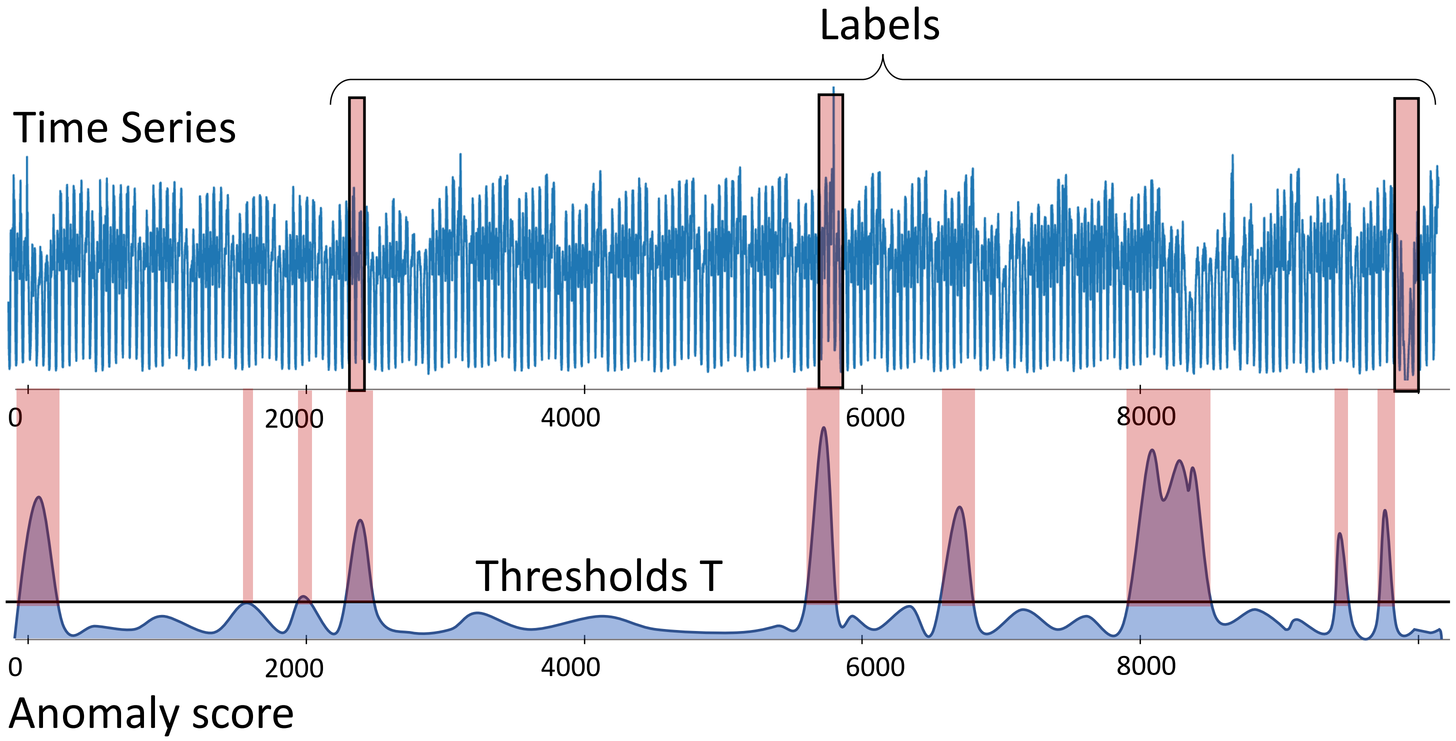




Evaluation measures: *AUC-based*

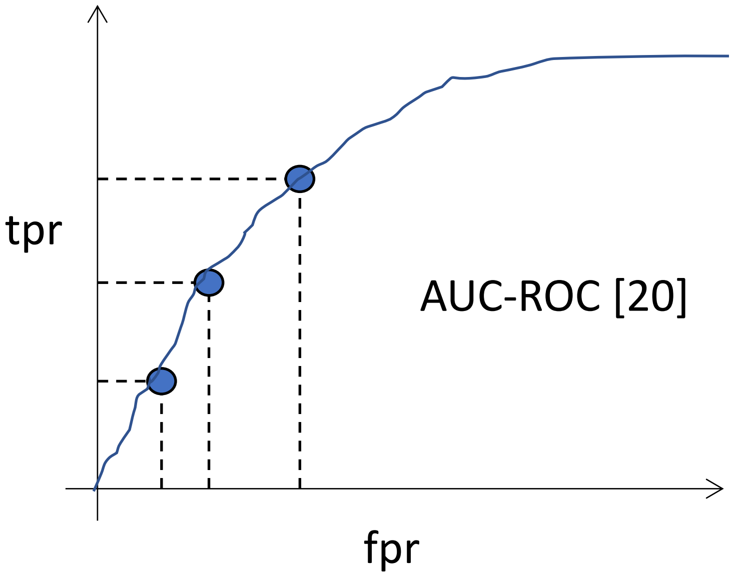
AUC-based Evaluation Measures:

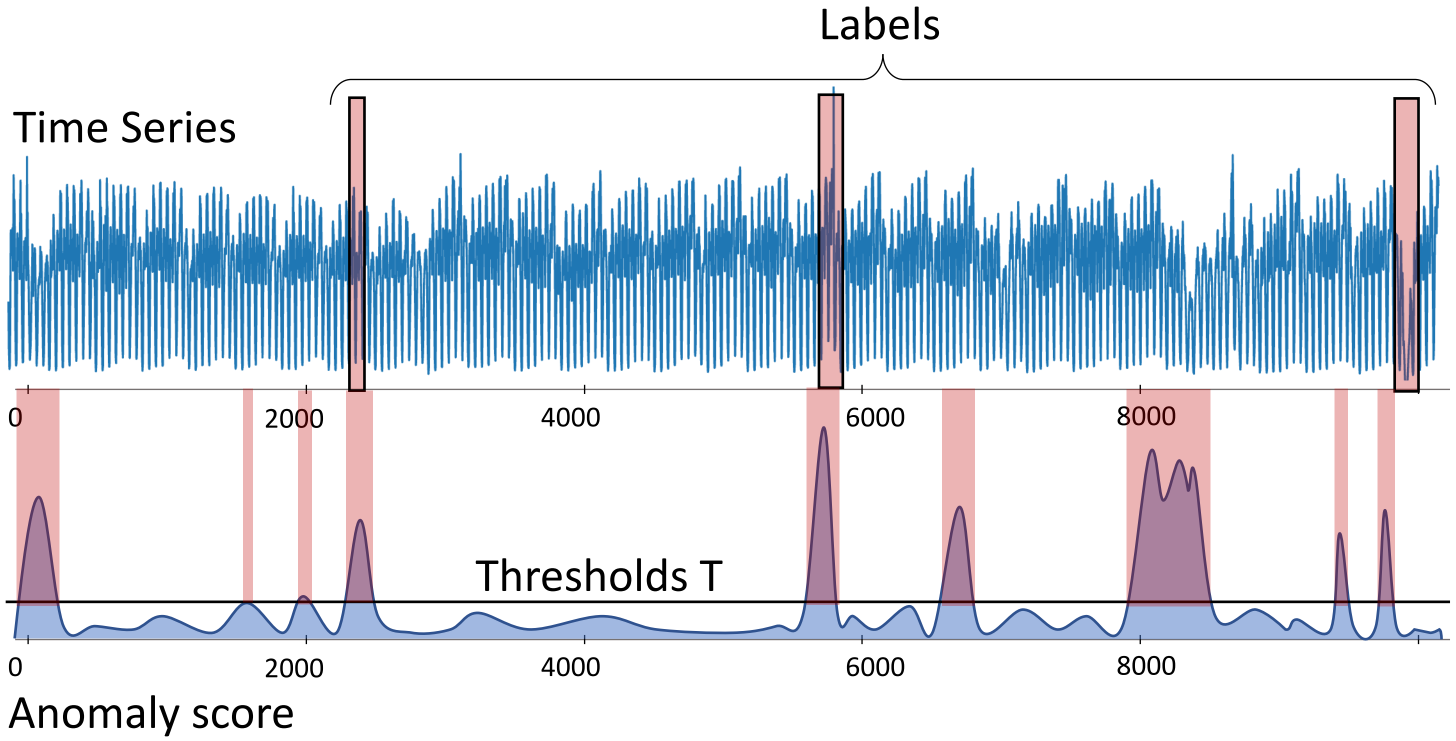




Evaluation measures: *AUC-based*

AUC-based Evaluation Measures:

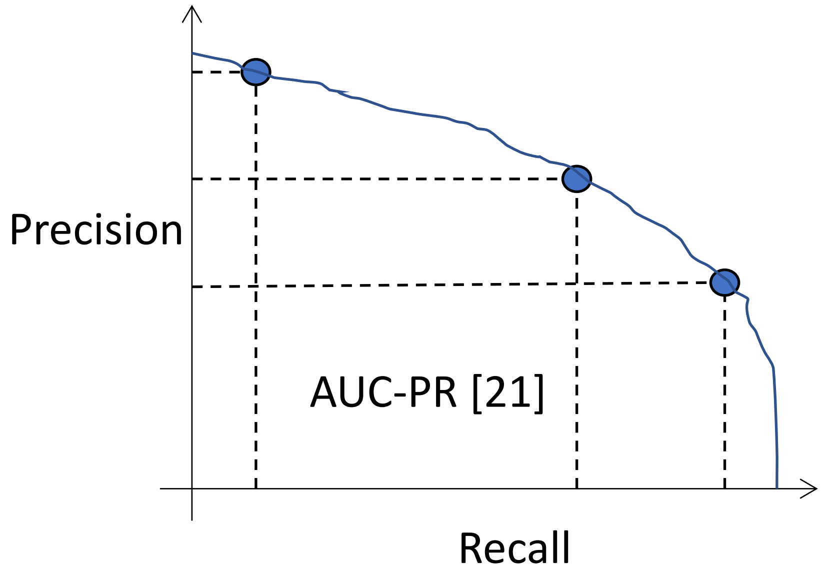


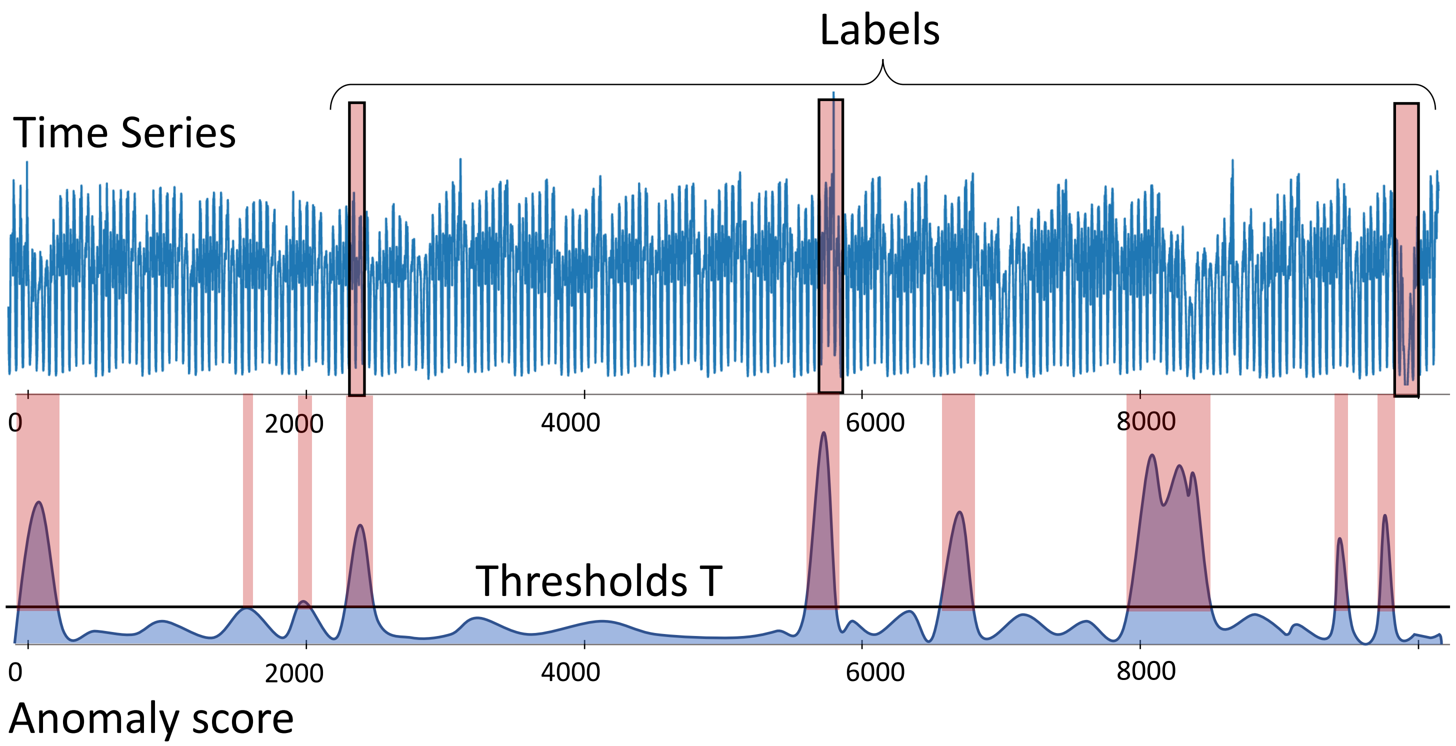


[20] Tom Fawcett. 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 8 (2006), 861–874.

Evaluation measures: *AUC-based*

AUC-based Evaluation Measures:

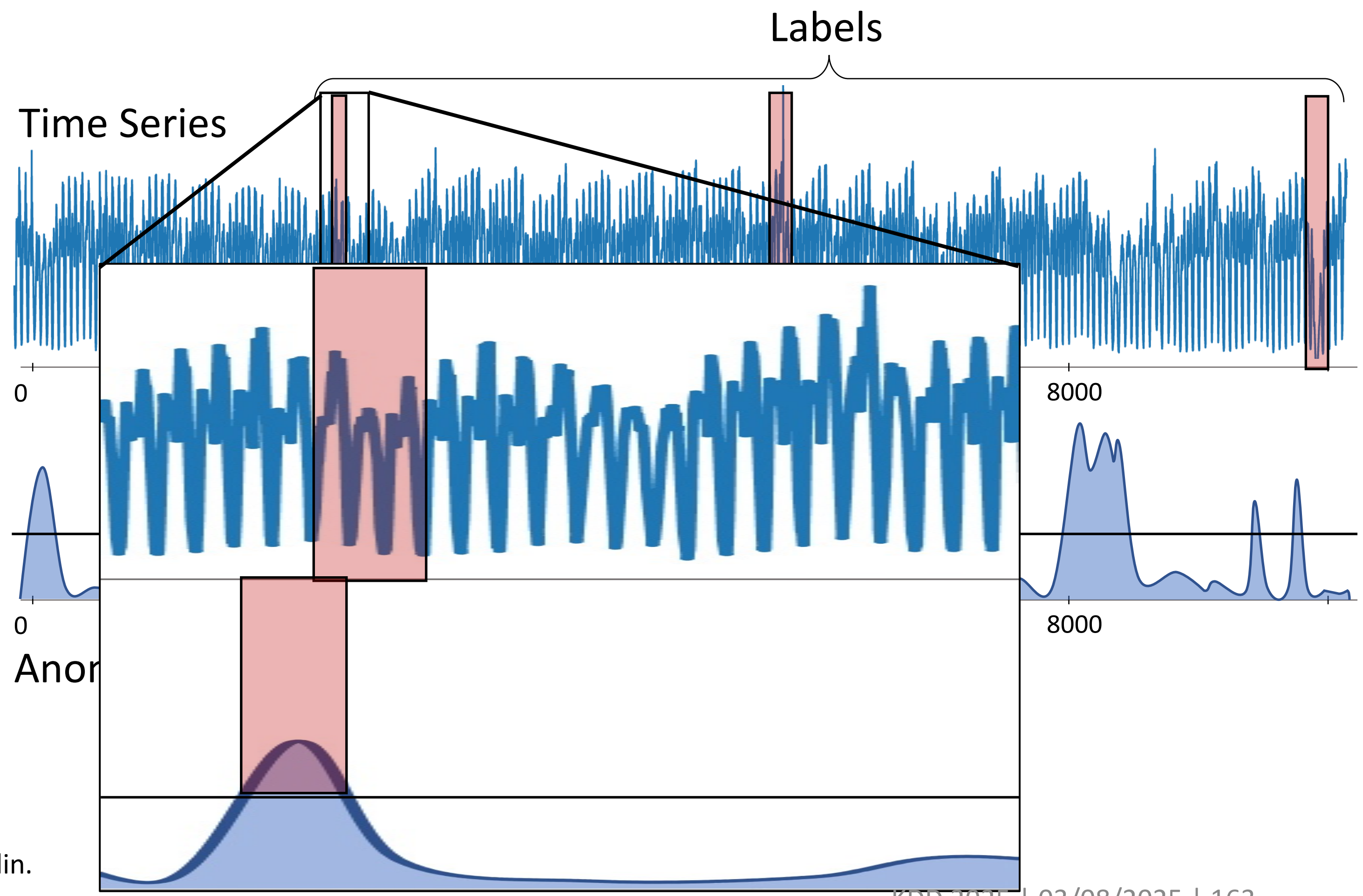




[21] Jesse Davis and Mark Goadrich. 2006. The Relationship between Precision-Recall and ROC Curves. In Proceedings of the 23rd International Conference on Machine Learning (ICML '06).

Evaluation measures: *Labeling issue*

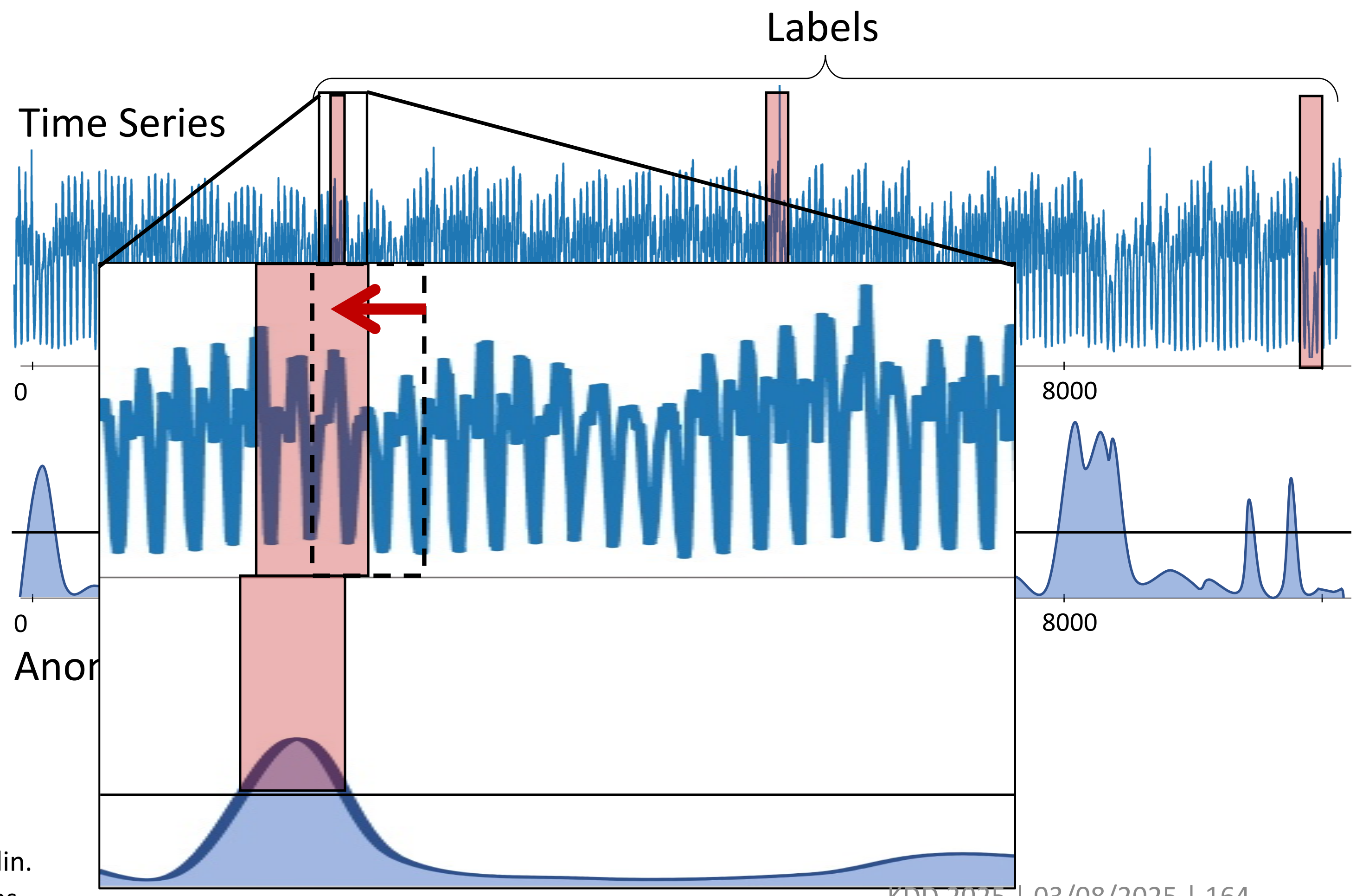
Labeling can be an issue for time series [22]:



[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

Labeling can be an issue for time series [22]:

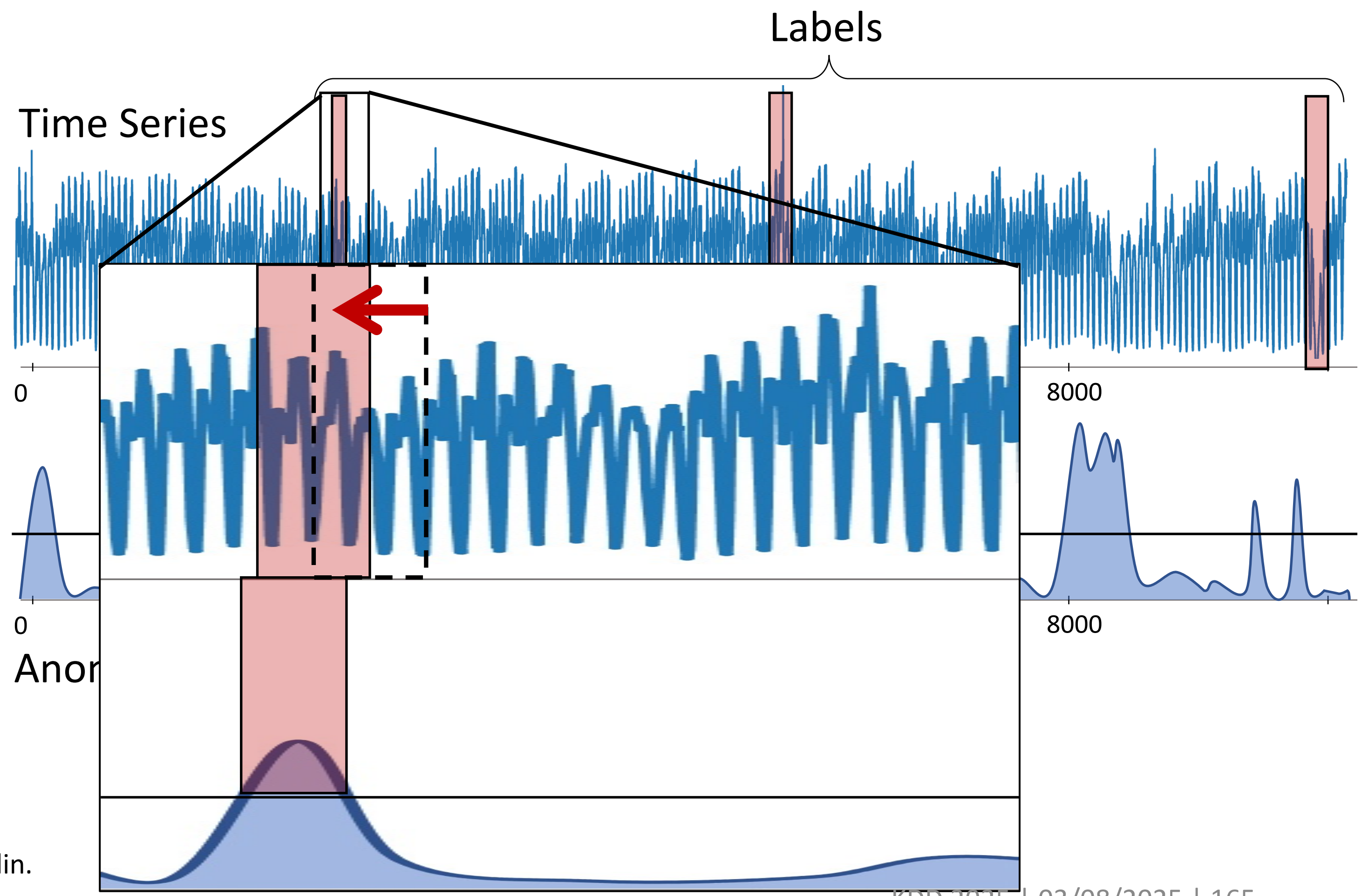


[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

Labeling can be an issue for time series [22]:

- Misalignment can lead to significant changes of accuracy values.



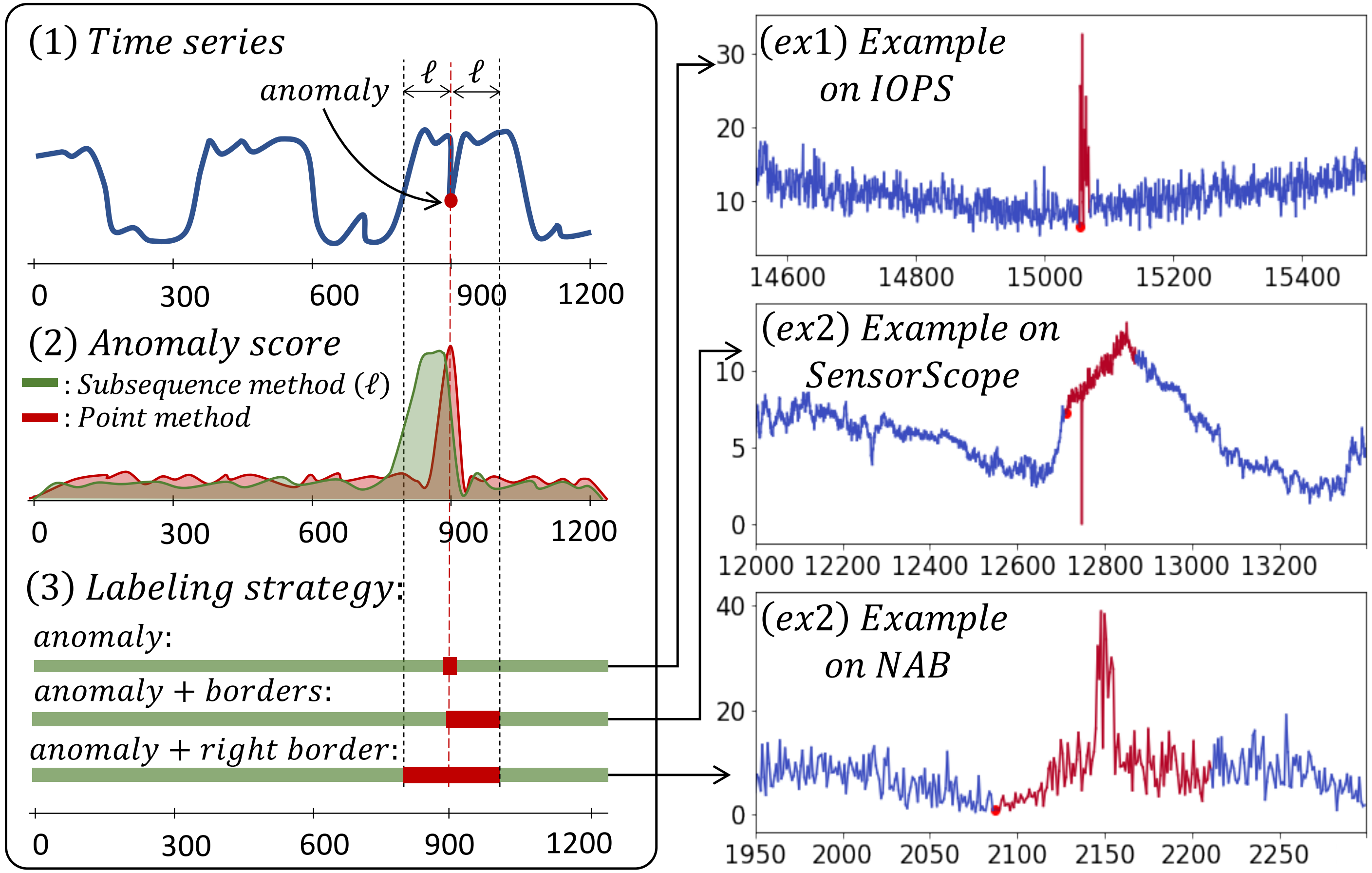
[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

Labeling can be an issue for time series [22]:

- Misalignment can lead to significant changes of accuracy values.

- This is a real issue because of:
  - Different Labeling strategies between domains and applications
  - Methods that produce misaligned anomaly scores.



Evaluation measures: *Labeling issue*

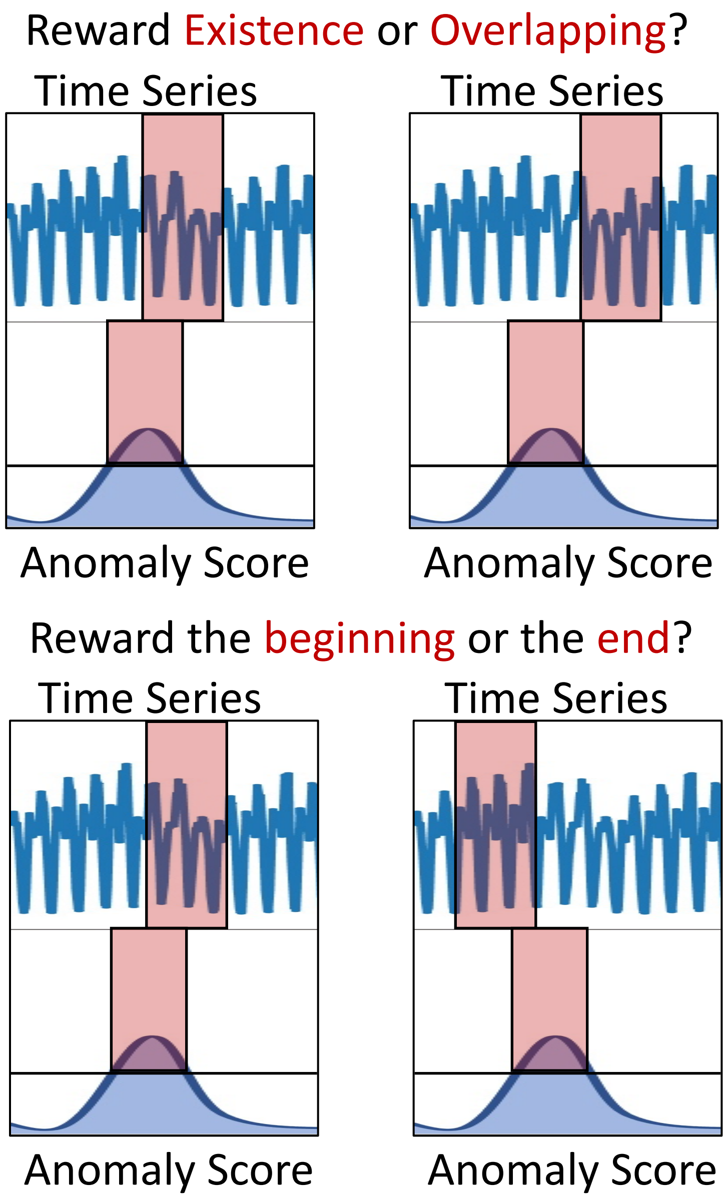
Existing solutions:

- Range Precision and Recall [23]:

- $Recall_T(R,P) = \frac{\sum_{i=1}^{N_r} Recall_T(R_i,P)}{N_r}$
- $Recall_T(R\_i,P) = \alpha * ExistenceR(R_i,P) + (1-\alpha) * OverlappingR(R\_i,P)$

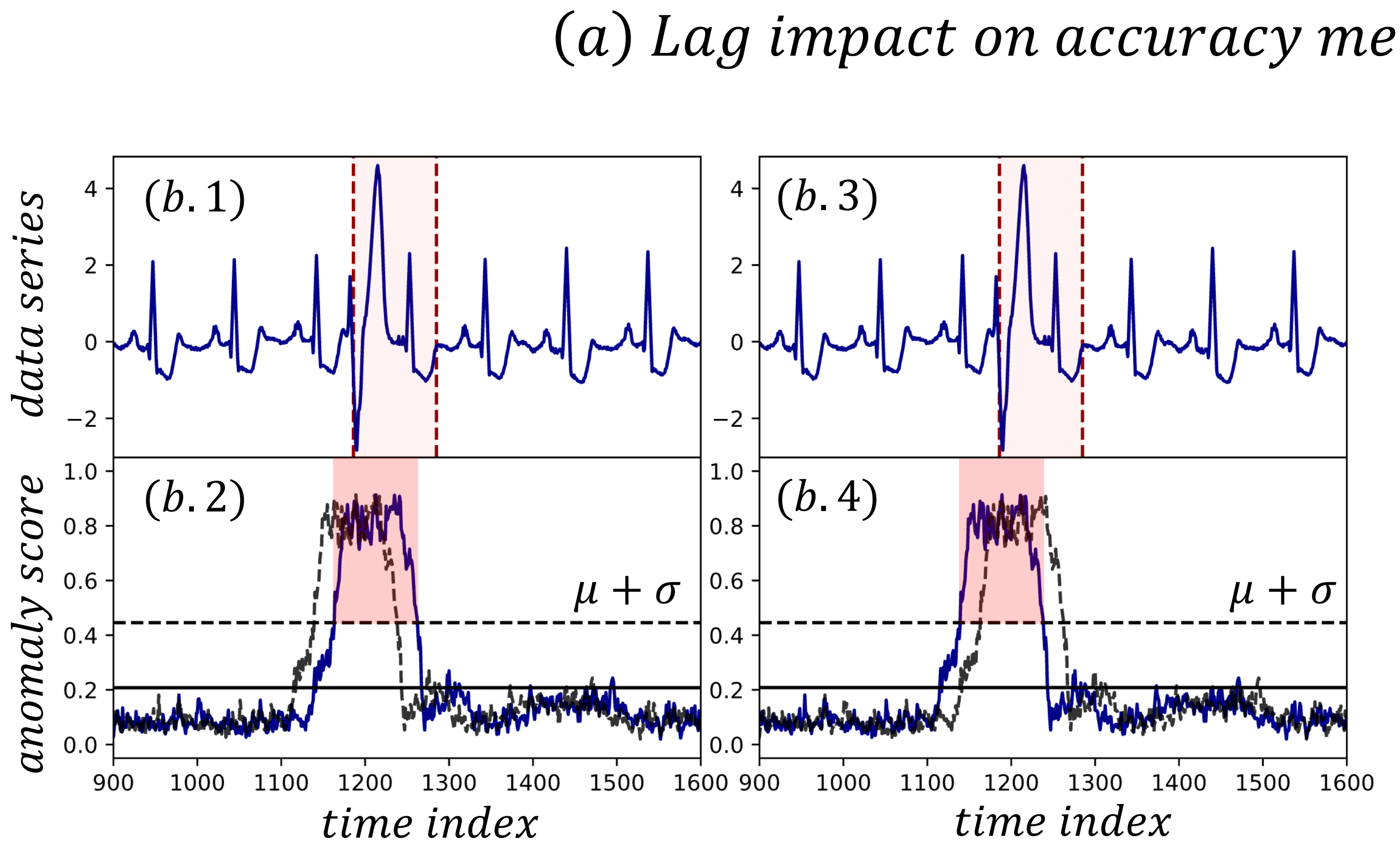
- $Precision_T(R,P) = \frac{\sum_{i=1}^{N_p} Precision_T(R,P_i)}{N_p}$
- $Precision_T(R,P_i) = CardinalityFactor(P_i,R) * \sum_{j=1}^{N_r} w\big(P_i, P_i \cap R_j, \delta\big)$

- Functions $w(\ ), \delta(\ )$ are tunable functions to represent the overlap size and position respectively.



[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.

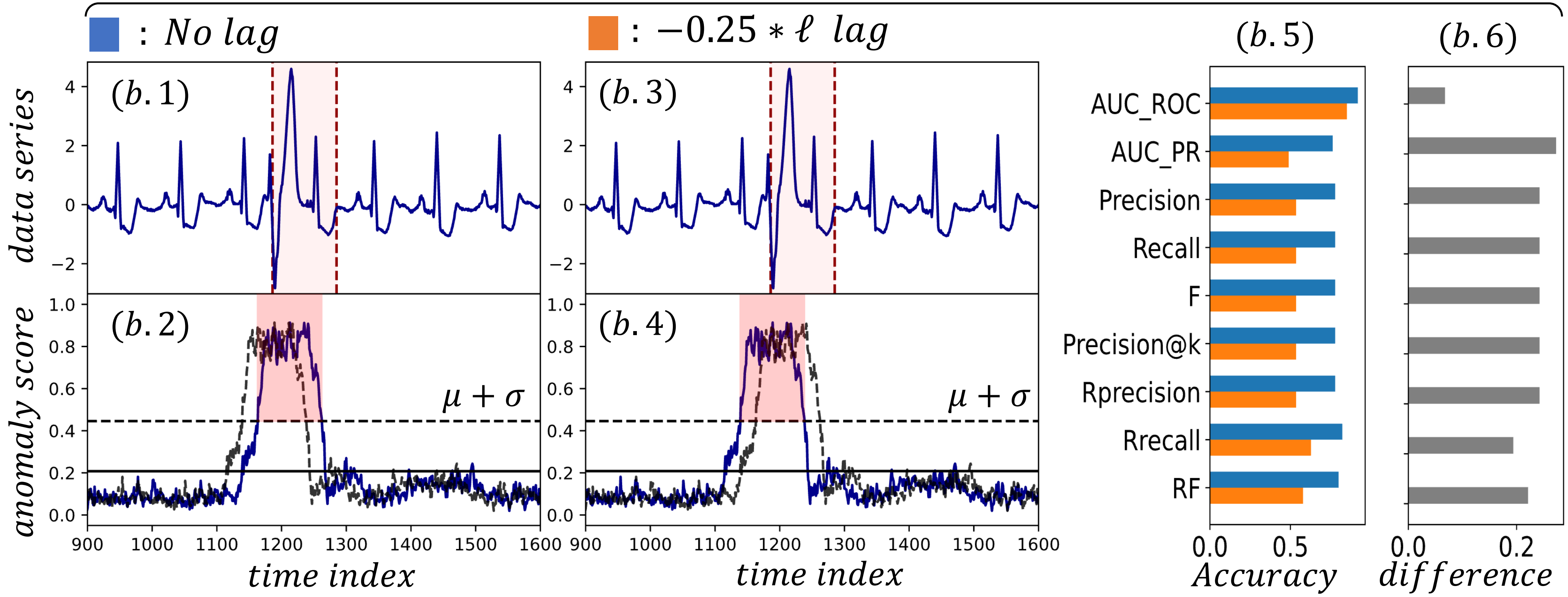
Evaluation measures: *Labeling issue*



[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

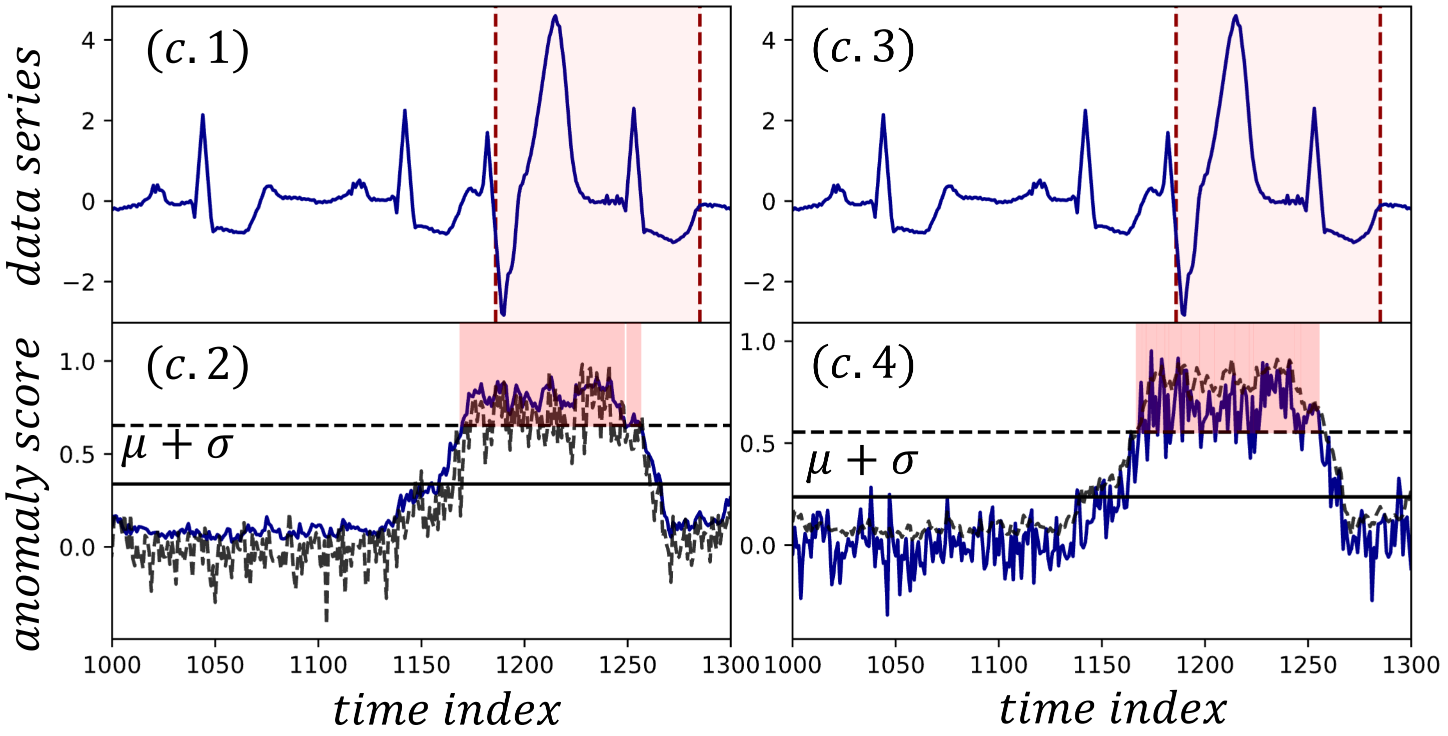
(a) *Lag impact on accuracy measures*



[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

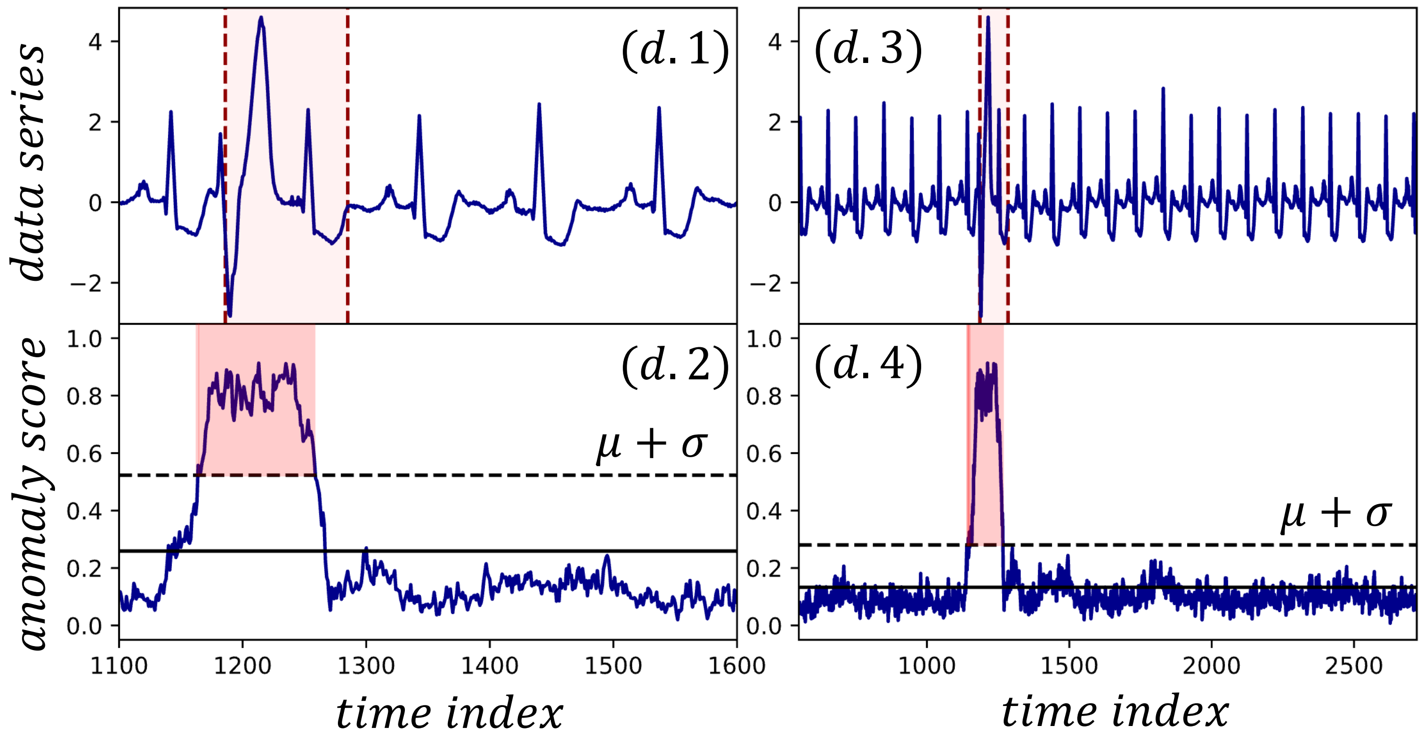
$(b)\ Noise\ impact\ on\ the\ accuracy\ measures$



[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

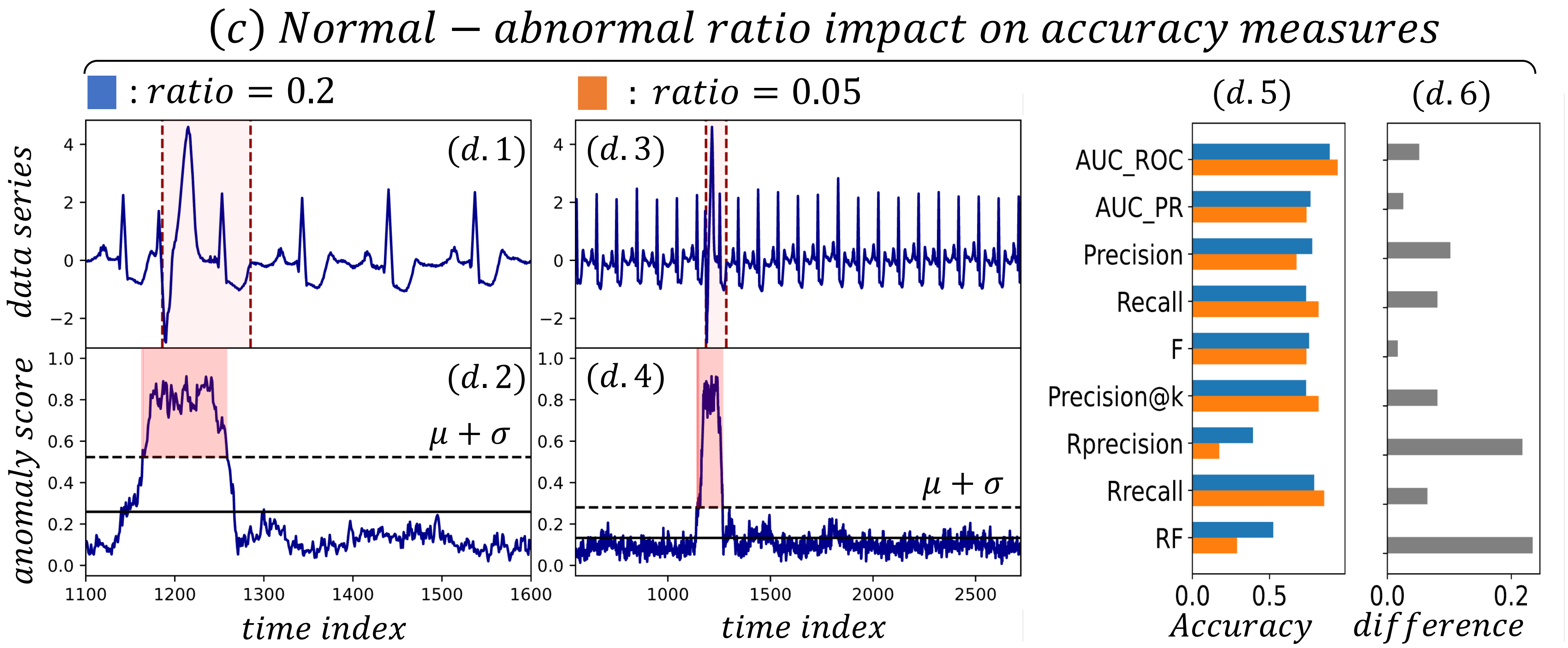
Evaluation measures: *Labeling issue*

$$(c)\ Normal - abnormal\ ratio\ impact\ on\ accuracy\ measures$$



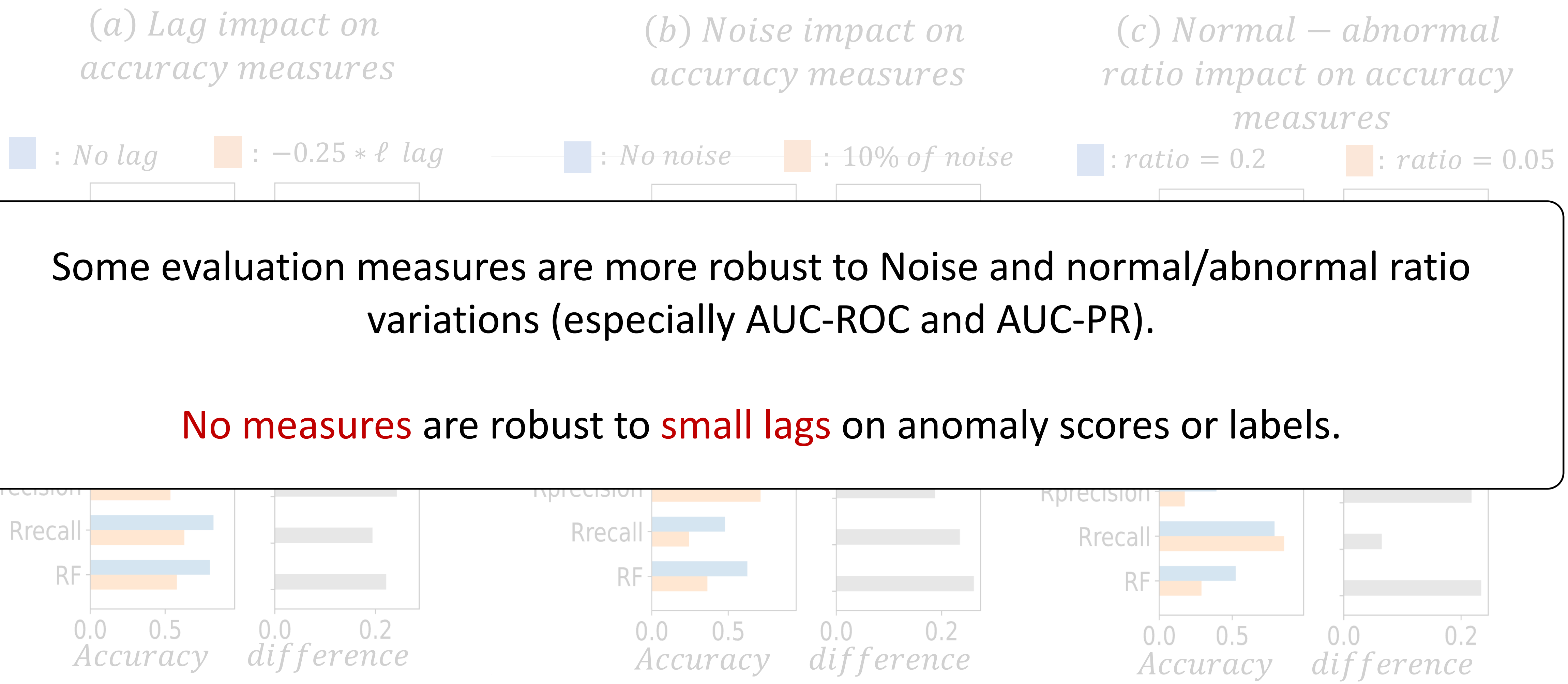
[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*



[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

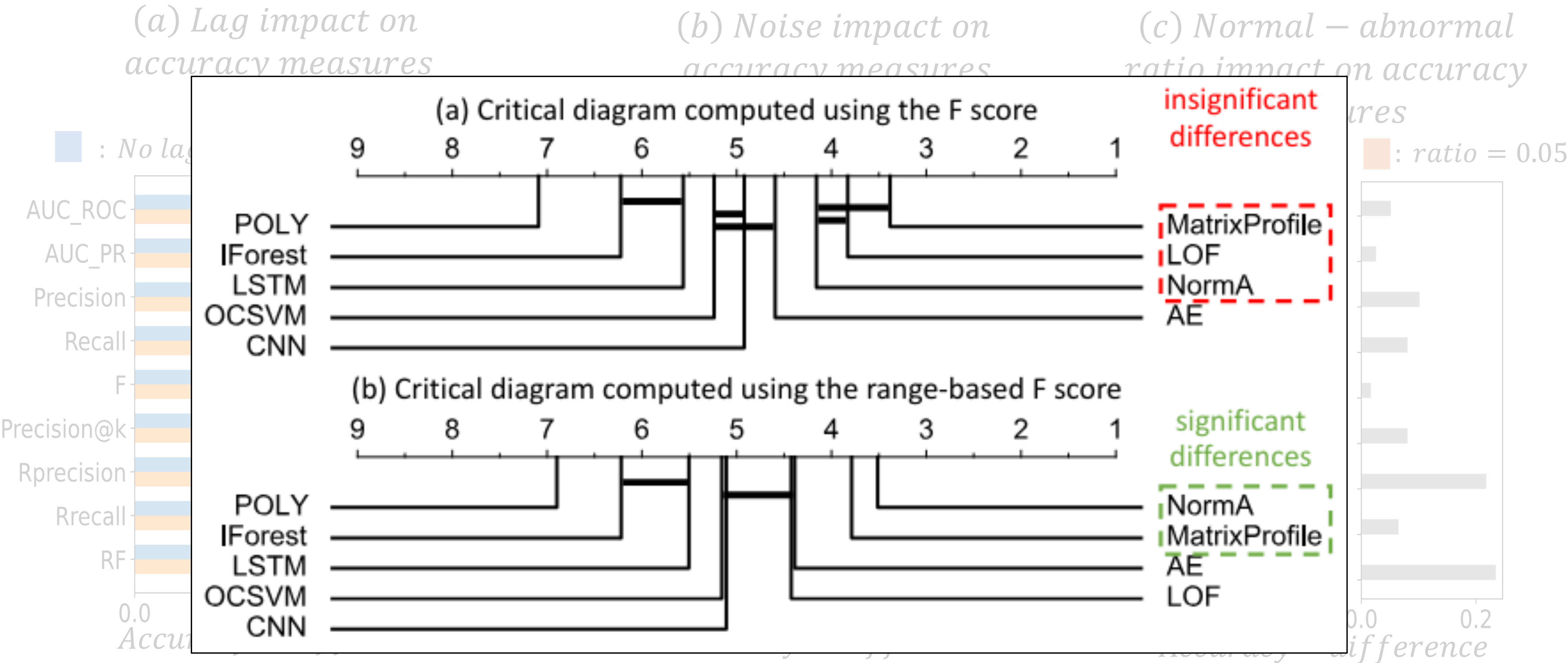


Some evaluation measures are more robust to Noise and normal/abnormal ratio variations (especially AUC-ROC and AUC-PR).

No measures are robust to small lags on anomaly scores or labels.

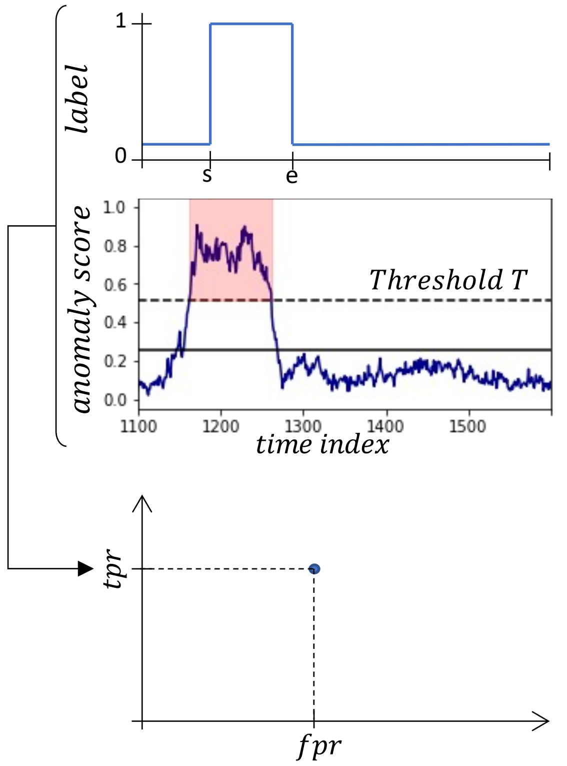
[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*



[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

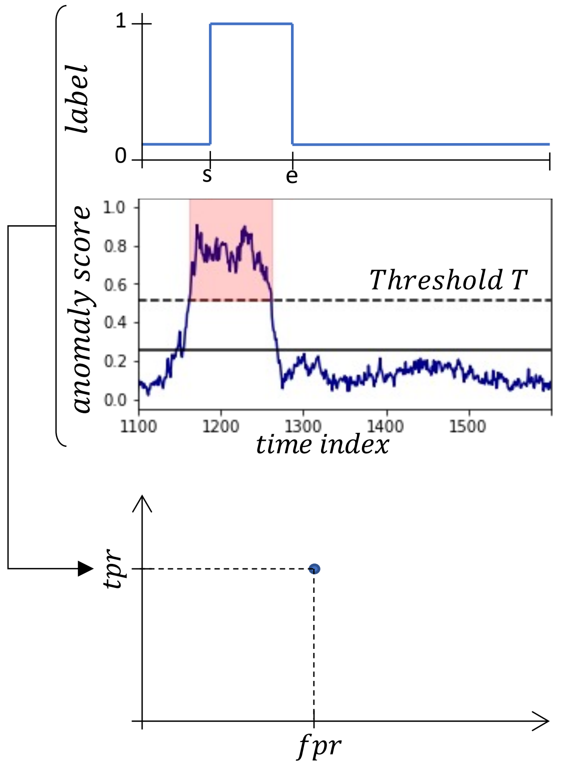
Evaluation measures: *Labeling issue*



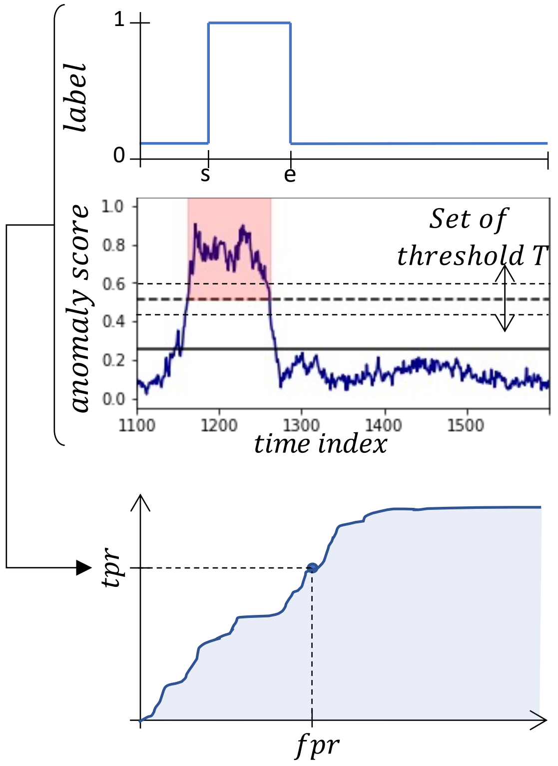
(a) Threshold-based Accuracy measure
(ex: Precision, Recall, Rprecision)

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*



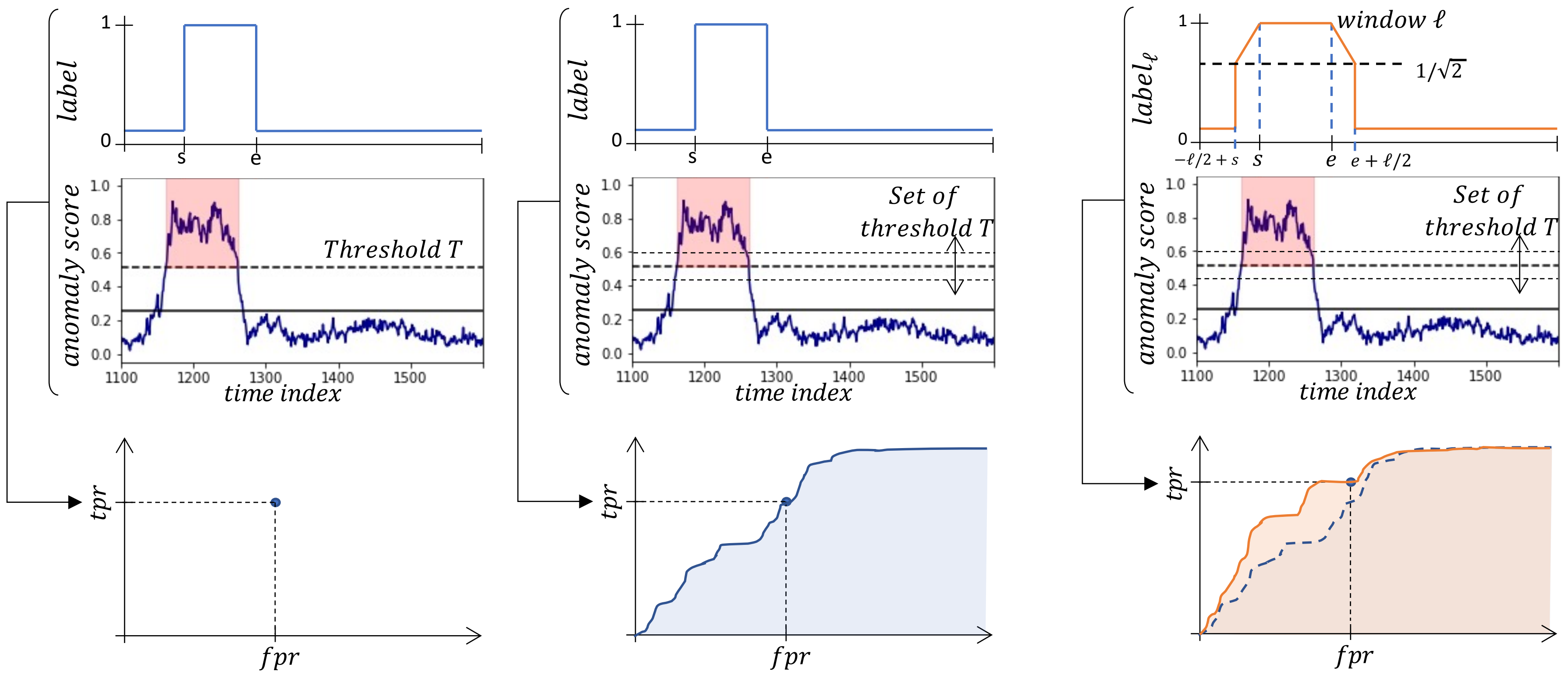
(a) Threshold-based Accuracy measure
(ex: Precision, Recall, Rprecision)



(b) AUC-based Accuracy measure
(ex: AUC-ROC, AUC-PR)

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*



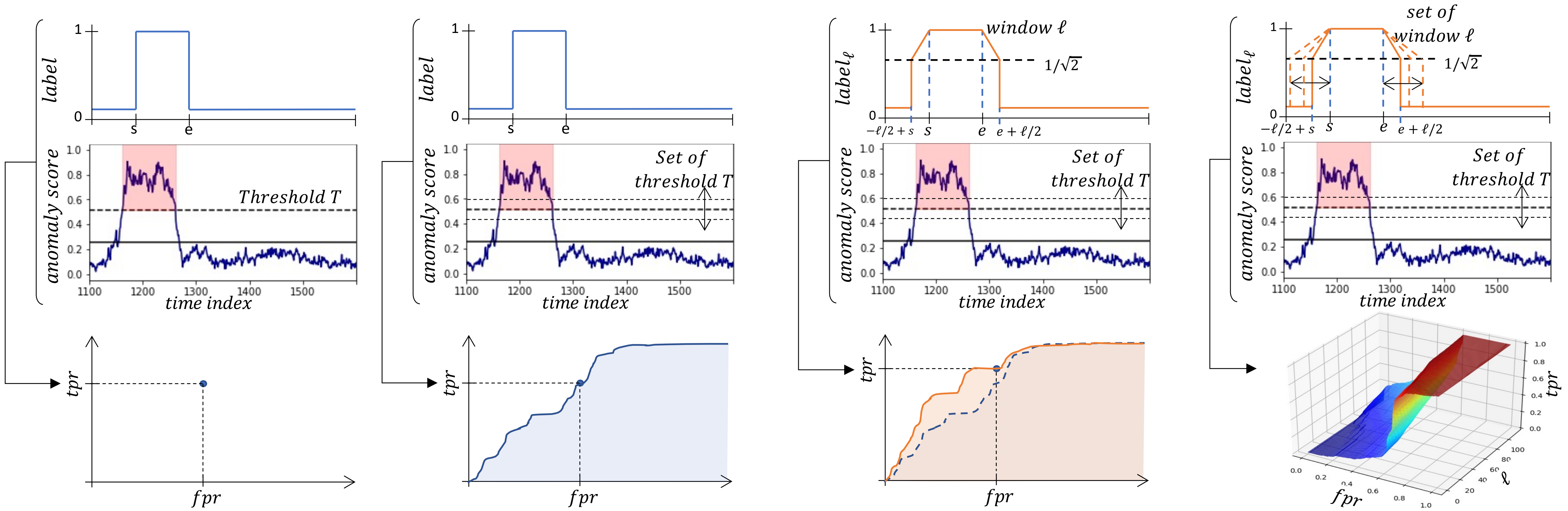
(a) Threshold-based Accuracy measure
(ex: Precision, Recall, Rprecision)

(b) AUC-based Accuracy measure
(ex: AUC-ROC, AUC-PR)

(c) R-AUC-based Accuracy measure
(ex: R-AUC-ROC, R-AUC-PR)

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*



(a) Threshold-based Accuracy measure
(ex: Precision, Recall, Rprecision)

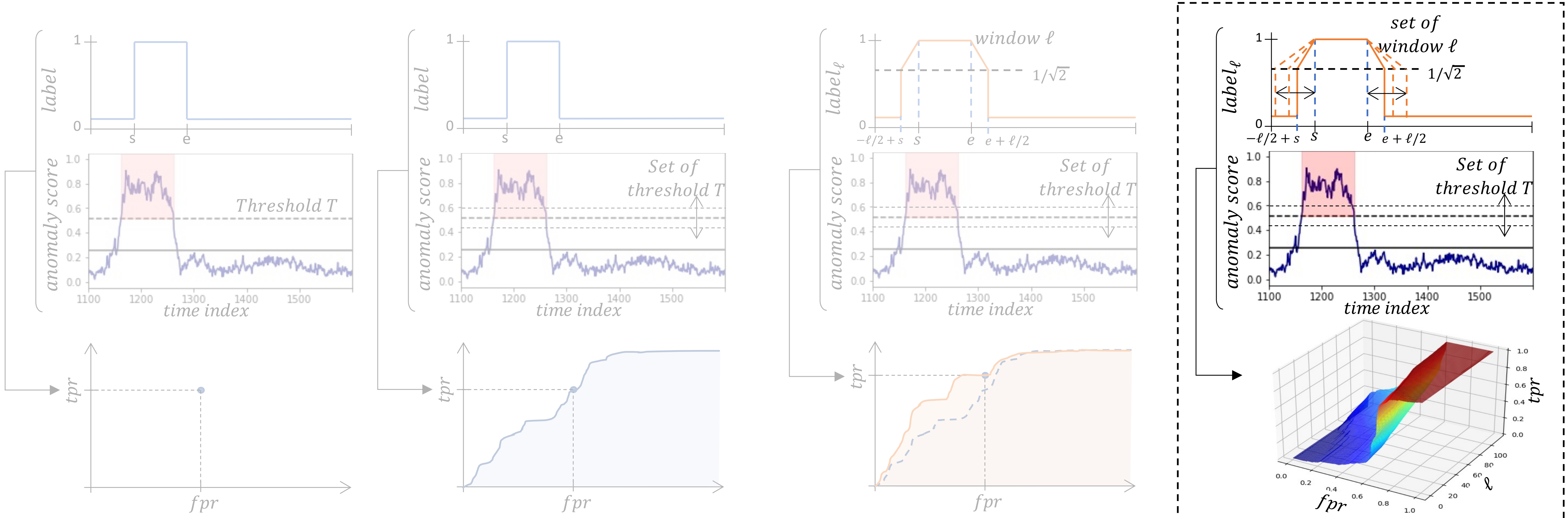
(b) AUC-based Accuracy measure
(ex: AUC-ROC, AUC-PR)

(c) R-AUC-based Accuracy measure
(ex: R-AUC-ROC, R-AUC-PR)

(d) VUS-based Accuracy measure
(ex: VUS-ROC, VUS-PR)

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*



(a) Threshold-based Accuracy measure (ex: Precision, Recall, Rprecision)

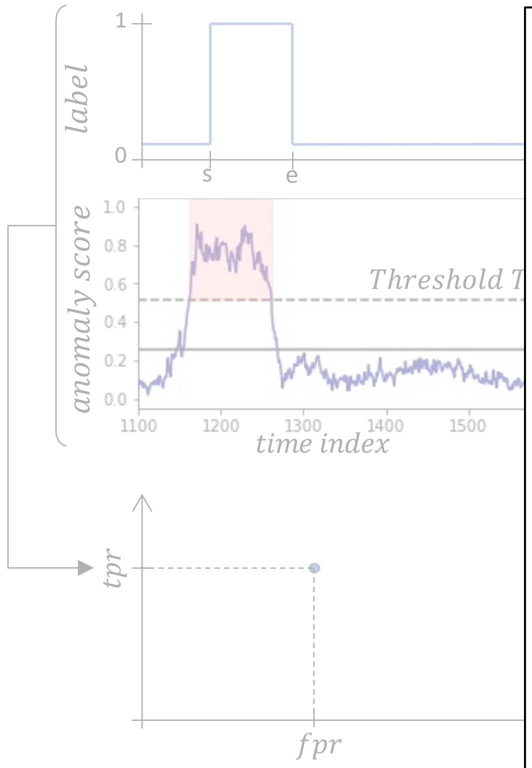
(b) AUC-based Accuracy measure (ex: AUC-ROC, AUC-PR)

(c) R-AUC-based Accuracy measure (ex: R-AUC-ROC, R-AUC-PR)

(d) VUS-based Accuracy measure (ex: VUS-ROC, VUS-PR)

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *Labeling issue*

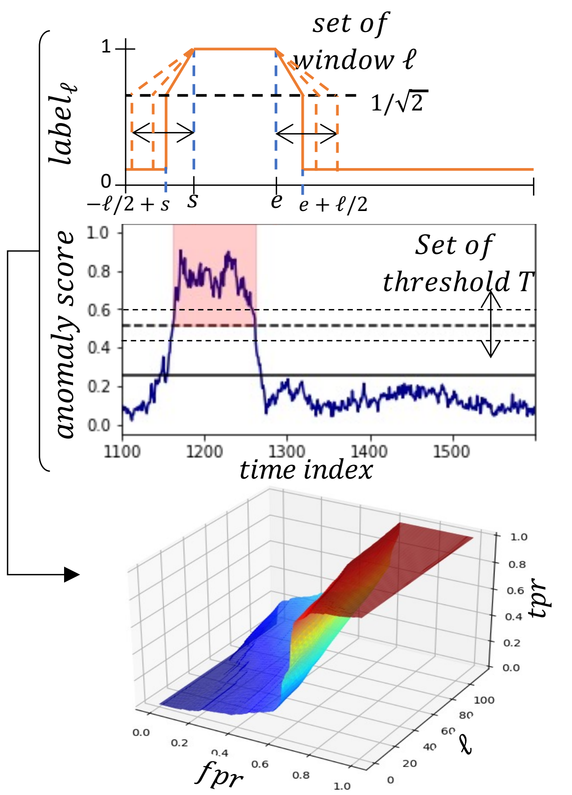


(a) Threshold-based Accuracy mea
(ex: Precision, Recall, Rprecision

Summary:

- Volume Under the Surface [22] (VUS):
- Modify the labels with buffer regions at the beginning and at the end of an anomaly
- We vary the buffer size (as well as the threshold) and we obtain a surface
- We use the volume under the surface (VUS) as accuracy

-AUC-based Accuracy measure
ex: R-AUC-ROC, R-AUC-PR)



(d) VUS-based Accuracy measure
(ex: VUS-ROC, VUS-PR)

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *VUS*

How is it computed?

$$AUC\text{-}ROC = \frac{1}{2}\sum_{k=1}^{N} \Delta^k_{TPR} * \Delta^k_{FPR}$$

$$\text{with: } \begin{cases} \Delta^k_{FPR} & = FPR(Th_k) - FPR(Th_{k-1}) \\ \Delta^k_{TPR} & = TPR(Th_{k-1}) + TPR(Th_k) \end{cases}$$

$$AUC\text{-}PR = \frac{1}{2}\sum_{k=1}^{N} \Delta^k_{Precision} * \Delta^k_{Recall}$$

$$\text{with: } \begin{cases} \Delta^k_{Recall} & = Recall(Th_k) - Recall(Th_{k-1}) \\ \Delta^k_{Precision} & = Precision(Th_{k-1}) + Precision(Th_k) \end{cases}$$

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *VUS*

How is it computed?

$$AUC\text{-}ROC = \frac{1}{2}\sum_{k=1}^{N} \Delta^k_{TPR} * \Delta^k_{FPR}$$

$$\text{with: } \begin{cases} \Delta^k_{FPR} & = FPR(Th_k) - FPR(Th_{k-1}) \\ \Delta^k_{TPR} & = TPR(Th_{k-1}) + TPR(Th_k) \end{cases}$$

$$VUS\text{-}ROC = \frac{1}{4}\sum_{w=1}^{L}\sum_{k=1}^{N} \Delta^{(k,w)} * \Delta^{w}$$

$$\text{with: } \begin{cases} \Delta^{(k,w)} & = \Delta^k_{TPR_{\ell_w}} * \Delta^k_{FPR_{\ell_w}} + \Delta^k_{TPR_{\ell_{w-1}}} * \Delta^k_{FPR_{\ell_{w-1}}} \\ \Delta^k_{FPR_{\ell_w}} & = FPR_{\ell_w}(Th_k) - FPR_{\ell_w}(Th_{k-1}) \\ \Delta^k_{TPR_{\ell_w}} & = TPR_{\ell_w}(Th_{k-1}) + TPR_{\ell_w}(Th_k) \\ \Delta^w & = |\ell_w - \ell_{w-1}| \end{cases}$$

$$AUC\text{-}PR = \frac{1}{2}\sum_{k=1}^{N} \Delta^k_{Precision} * \Delta^k_{Recall}$$

$$\text{with: } \begin{cases} \Delta^k_{Recall} & = Recall(Th_k) - Recall(Th_{k-1}) \\ \Delta^k_{Precision} & = Precision(Th_{k-1}) + Precision(Th_k) \end{cases}$$

$$VUS\text{-}PR = \frac{1}{4}\sum_{w=1}^{L}\sum_{k=1}^{N} \Delta^{(k,w)} * \Delta^{w}$$

$$\text{with: } \begin{cases} \Delta^{(k,w)} & = \Delta^k_{Pr_{\ell_w}} * \Delta^k_{Re_{\ell_w}} + \Delta^k_{Pr_{\ell_{w-1}}} * \Delta^k_{Re_{\ell_{w-1}}} \\ \Delta^k_{Re_{\ell_w}} & = Recall_{\ell_w}(Th_k) - Recall_{\ell_w}(Th_{k-1}) \\ \Delta^k_{Pr_{\ell_w}} & = Precision_{\ell_w}(Th_{k-1}) + Precision_{\ell_w}(Th_k) \\ \Delta^w & = |\ell_w - \ell_{w-1}| \end{cases}$$

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *VUS*

How is it computed?

$$AUC\text{-}ROC = \frac{1}{2}\sum_{k=1}^{N} \Delta^k_{TPR} * \Delta^k_{FPR}$$

Time Complexity: $O(NT)$

With:
- T: the time series length
- N: the number of thresholds

$$\text{with: } \begin{cases} \Delta^k_{Recall} & = Recall(Th_k) - Recall(Th_{k-1}) \\ \Delta^k_{Precision} & = Precision(Th_{k-1}) + Precision(Th_k) \end{cases}$$

$$VUS\text{-}ROC = \frac{1}{4}\sum_{w=1}^{L}\sum_{k=1}^{N} \Delta^{(k,w)} * \Delta^{w}$$

$$\text{with: } \begin{cases} \Delta^{(k,w)} & = \Delta^k_{TPR_{\ell_w}} * \Delta^k_{FPR_{\ell_w}} + \Delta^k_{TPR_{\ell_{w-1}}} * \Delta^k_{FPR_{\ell_{w-1}}} \\ \Delta^k_{FPR_{\ell_w}} & = FPR_{\ell_w}(Th_k) - FPR_{\ell_w}(Th_{k-1}) \\ \Delta^k_{TPR_{\ell_w}} & = TPR_{\ell_w}(Th_{k-1}) + TPR_{\ell_w}(Th_k) \\ \Delta^{w} & = |\ell_w - \ell_{w-1}| \end{cases}$$

$$VUS\text{-}PR = \frac{1}{4}\sum_{w=1}^{L}\sum_{k=1}^{N} \Delta^{(k,w)} * \Delta^{w}$$

$$\text{with: } \begin{cases} \Delta^{(k,w)} & = \Delta^k_{Pr_{\ell_w}} * \Delta^k_{Re_{\ell_w}} + \Delta^k_{Pr_{\ell_{w-1}}} * \Delta^k_{Re_{\ell_{w-1}}} \\ \Delta^k_{Re_{\ell_w}} & = Recall_{\ell_w}(Th_k) - Recall_{\ell_w}(Th_{k-1}) \\ \Delta^k_{Pr_{\ell_w}} & = Precision_{\ell_w}(Th_{k-1}) + Precision_{\ell_w}(Th_k) \\ \Delta^{w} & = |\ell_w - \ell_{w-1}| \end{cases}$$

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *VUS*

How is it computed?

$$VUS\text{-}ROC = \frac{1}{4}\sum_{w=1}^{L}\sum_{k=1}^{N}\Delta^{(k,w)} * \Delta^{w}$$

$$\Delta^{(k,w)} = \Delta^{k}_{TPR_{\ell_w}} * \Delta^{k}_{FPR_{\ell_w}} + \Delta^{k}_{TPR_{\ell_{w-1}}} * \Delta^{k}_{FPR_{\ell_{w-1}}}$$

Time Complexity: $O(NT)$

With:
- T: the time series length
- N: the number of thresholds

Time Complexity: $O(NLT)$

With:
- T: the time series length
- N: the number of thresholds
- L: the number of buffer lengths

$$\text{with: } \begin{cases} \Delta^{k}_{Recall} & = Recall(Th_k) - Recall(Th_{k-1}) \\ \Delta^{k}_{Precision} & = Precision(Th_{k-1}) + Precision(Th_k) \end{cases}$$

$$\text{with: } \begin{cases} \Delta^{k}_{Re_{\ell_w}} & = Recall_{\ell_w}(Th_k) - Recall_{\ell_w}(Th_{k-1}) \\ \Delta^{k}_{Pr_{\ell_w}} & = Precision_{\ell_w}(Th_{k-1}) + Precision_{\ell_w}(Th_k) \\ \Delta^{w} & = |\ell_w - \ell_{w-1}| \end{cases}$$

Evaluation measures: *VUS*

How is it computed?

$$AUC\text{-}ROC = \frac{1}{2}\sum_{k=1}^{N} \Delta^{k}_{TPR} * \Delta^{k}_{FPR}$$

$$VUS\text{-}ROC = \frac{1}{4}\sum_{w=1}^{L}\sum_{k=1}^{N} \Delta^{(k,w)} * \Delta^{w}$$

$$\Delta^{(k,w)} = \Delta^{k}_{TPR_{\ell_w}} * \Delta^{k}_{FPR_{\ell_w}} + \Delta^{k}_{TPR_{\ell_{w-1}}} * \Delta^{k}_{FPR_{\ell_{w-1}}}$$

Time Complexity: $O(NT)$

With:
- T: the time series length
- N: the number of thresholds

Time Complexity: $O(NLT)$

With:
- T: the time series length
- N: the number of thresholds
- L: the number of buffer lengths

with: $\Delta^{k}_{Recall} = Recall(Th_k) - Recall(Th_{k-1})$, $\Delta^{k}_{Precision} = Precision(Th_{k-1}) + Precision(Th_k)$

with: $\Delta^{k}_{Re_{\ell_w}} = Recall_{\ell_w}(Th_k) - Recall_{\ell_w}(Th_{k-1})$, $\Delta^{k}_{Pr_{\ell_w}} = Precision_{\ell_w}(Th_{k-1}) + Precision_{\ell_w}(Th_k)$, $\Delta^{w} = |\ell_w - \ell_{w-1}|$

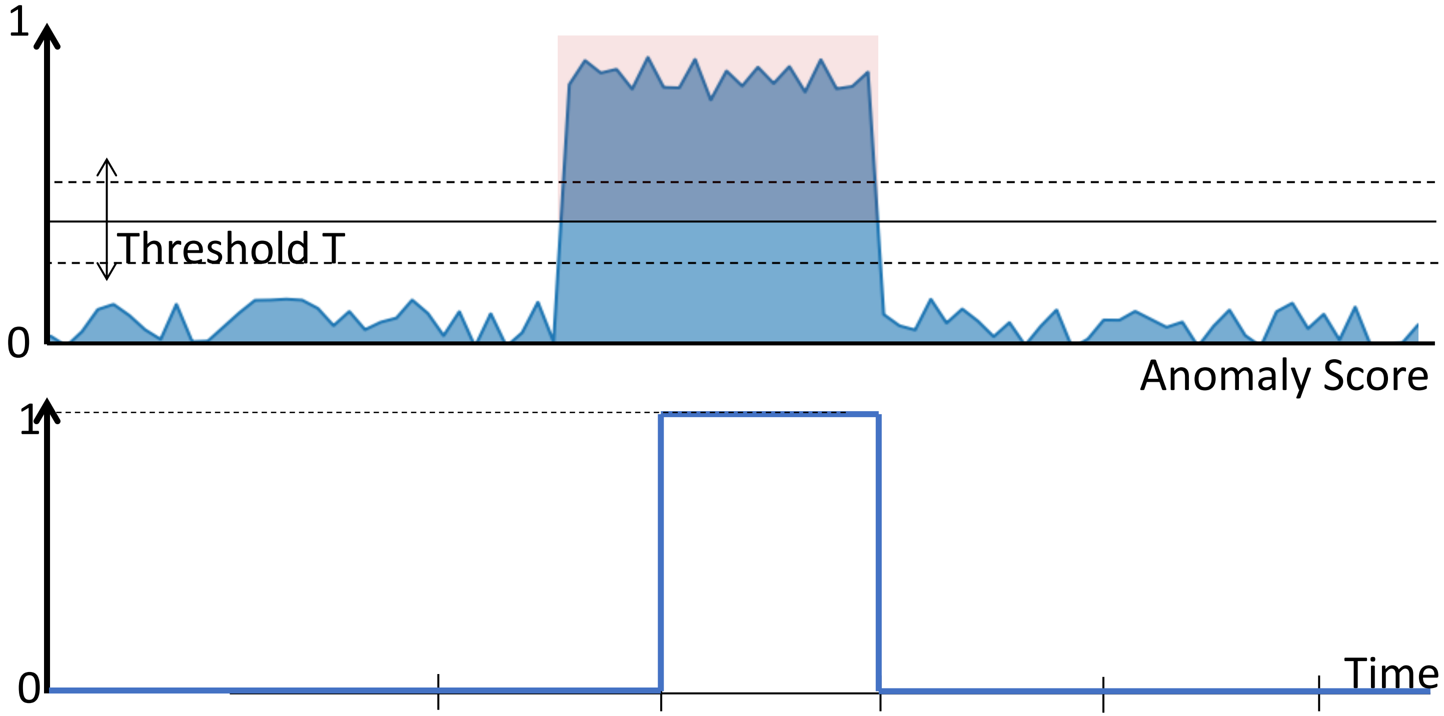
VUS is significantly slower to compute, complicating its usage in practice

[22] J. Paparrizos, P. Boniol, T. Palpanas, R. S. Tsay, A. Elmore, and M. J. Franklin. Volume under the surface: a new accuracy evaluation measure for time-series anomaly detection. Proc. VLDB Endow. 15, 11 (2022), 2774–2787.

Evaluation measures: *VUS*

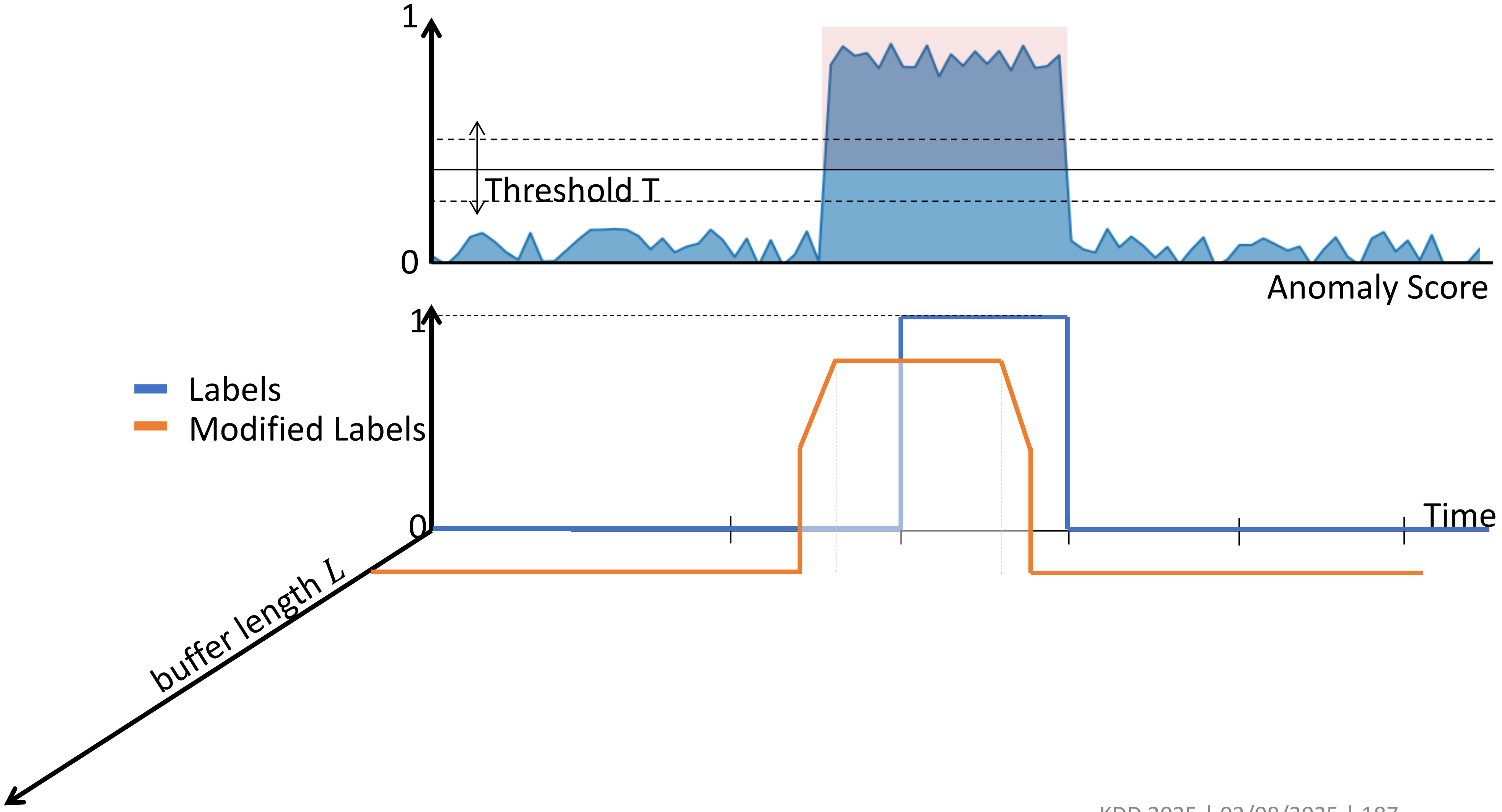
A solution?

Labels



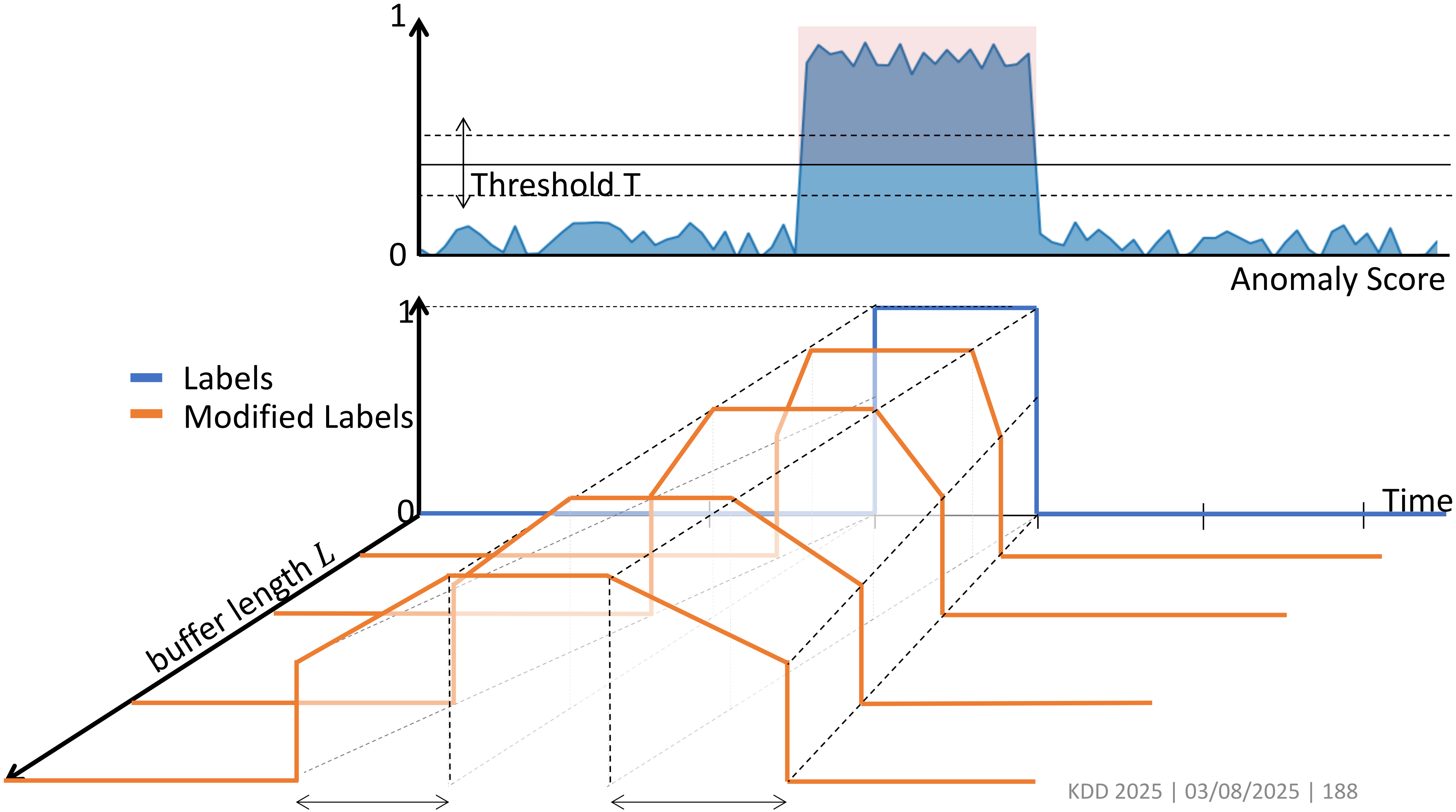
Evaluation measures: *VUS*

A solution?



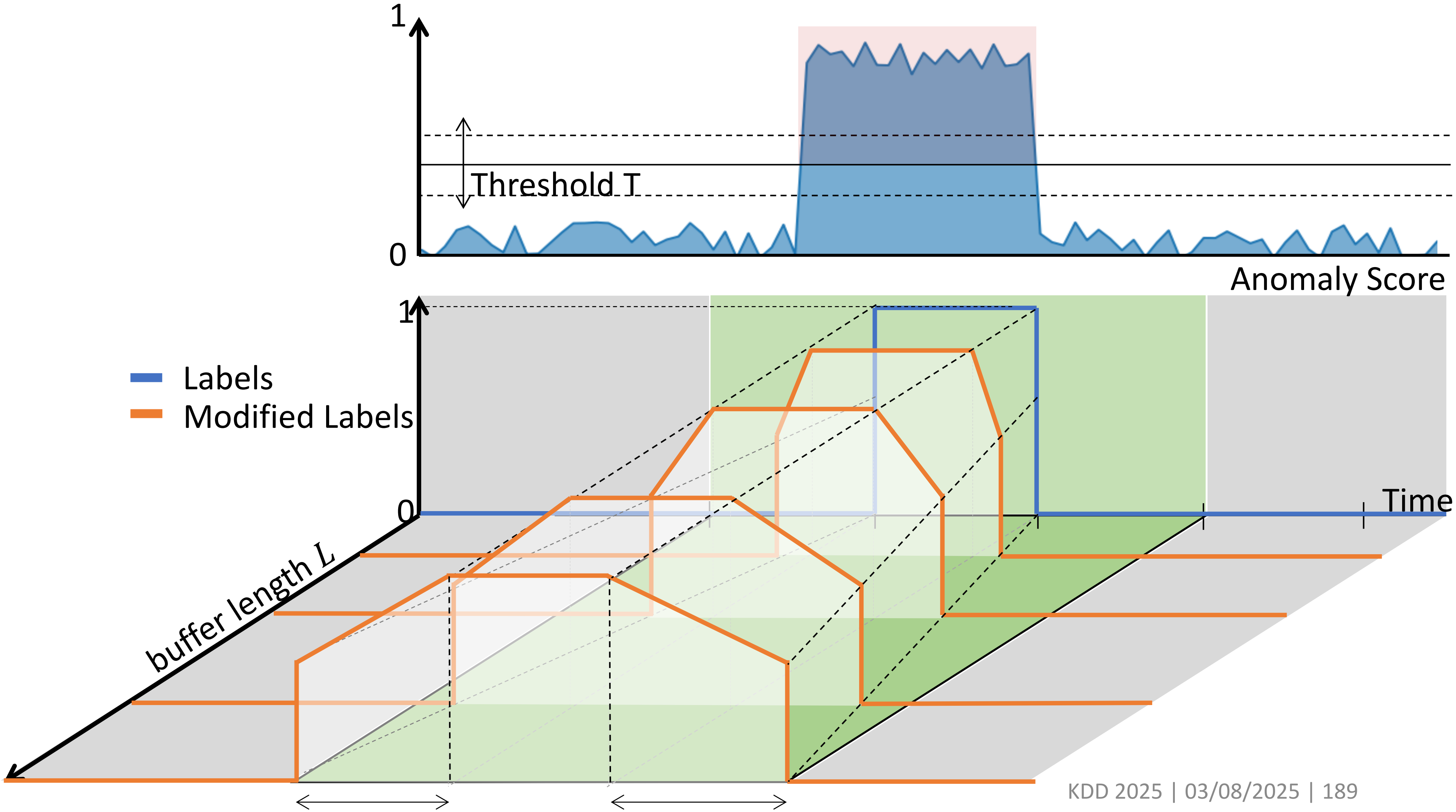
Evaluation measures: *VUS*

A solution?



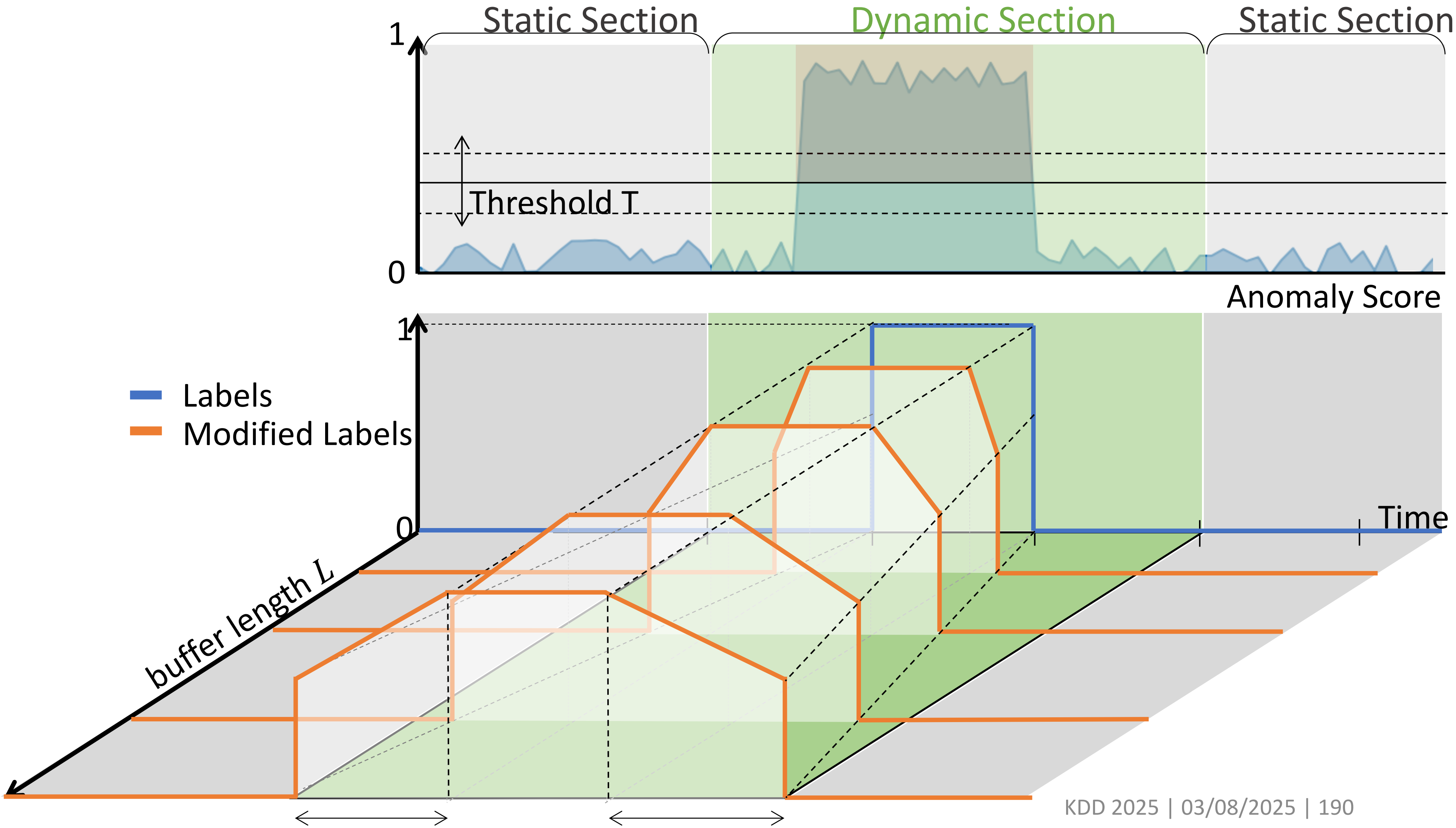
Evaluation measures: *VUS*

A solution?



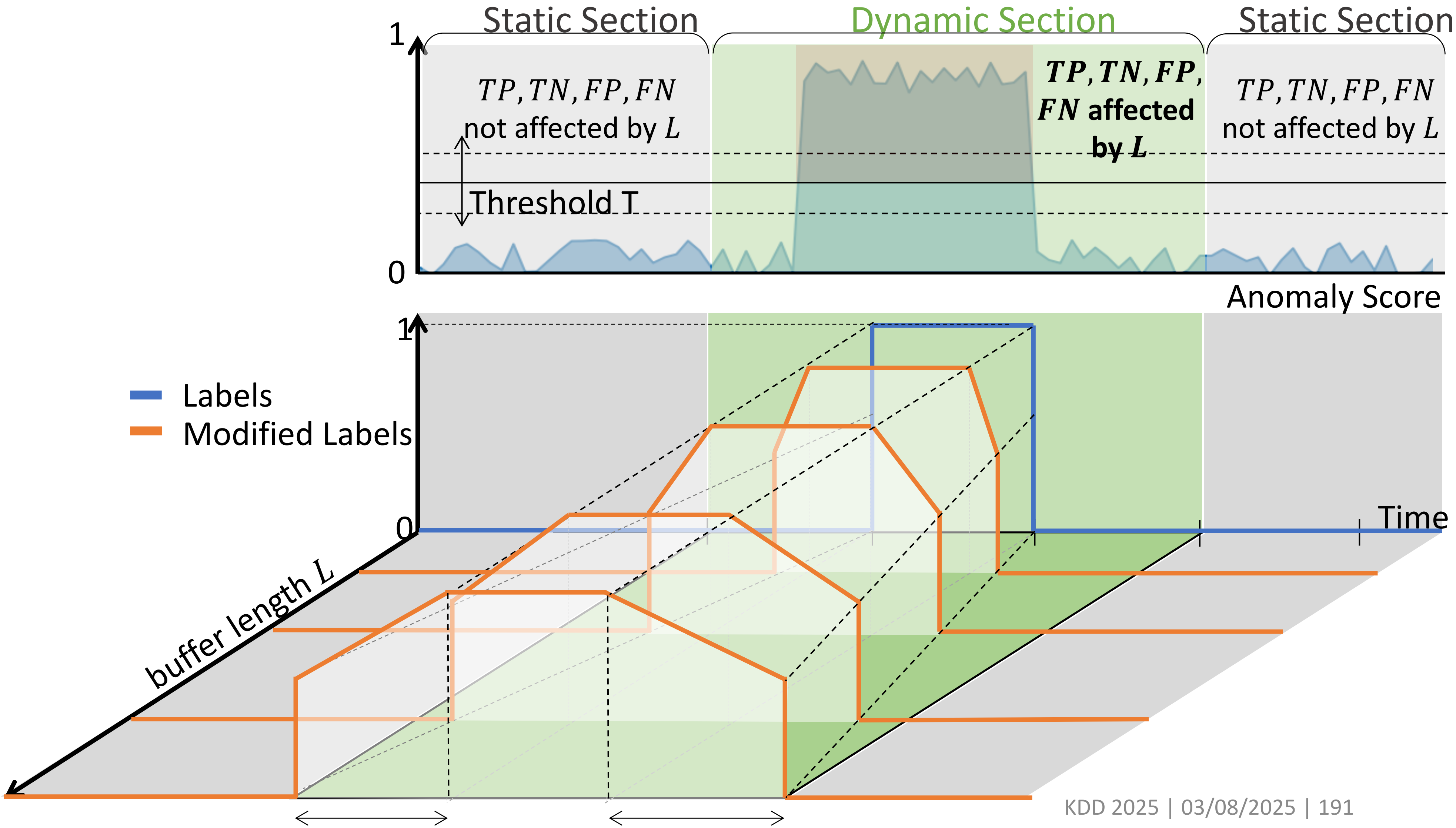
Evaluation measures: *VUS*

A solution?



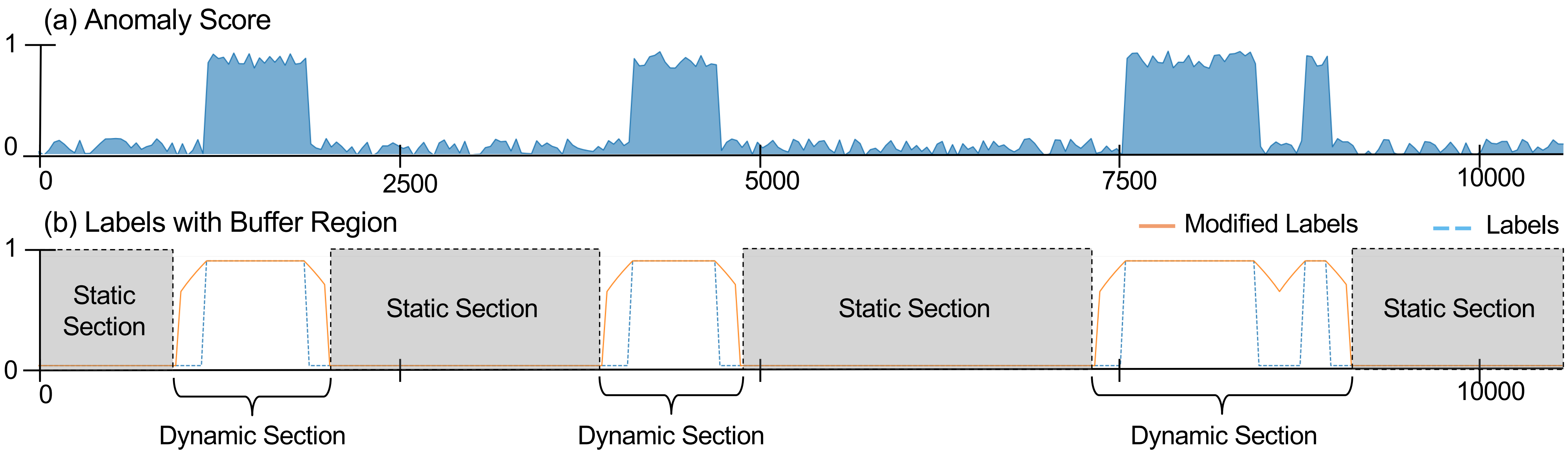
Evaluation measures: *VUS*

A solution?



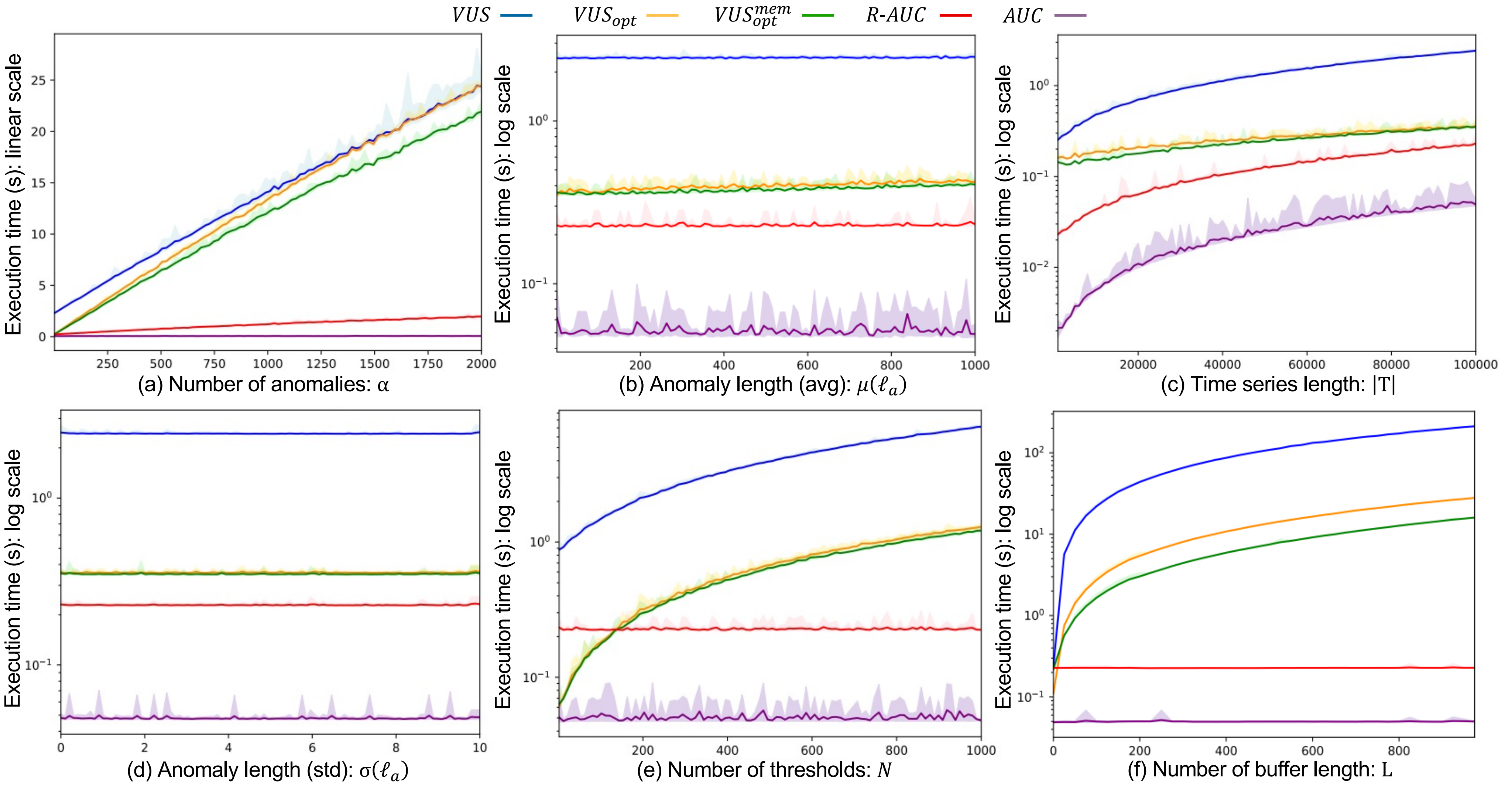
Evaluation measures: *VUS*

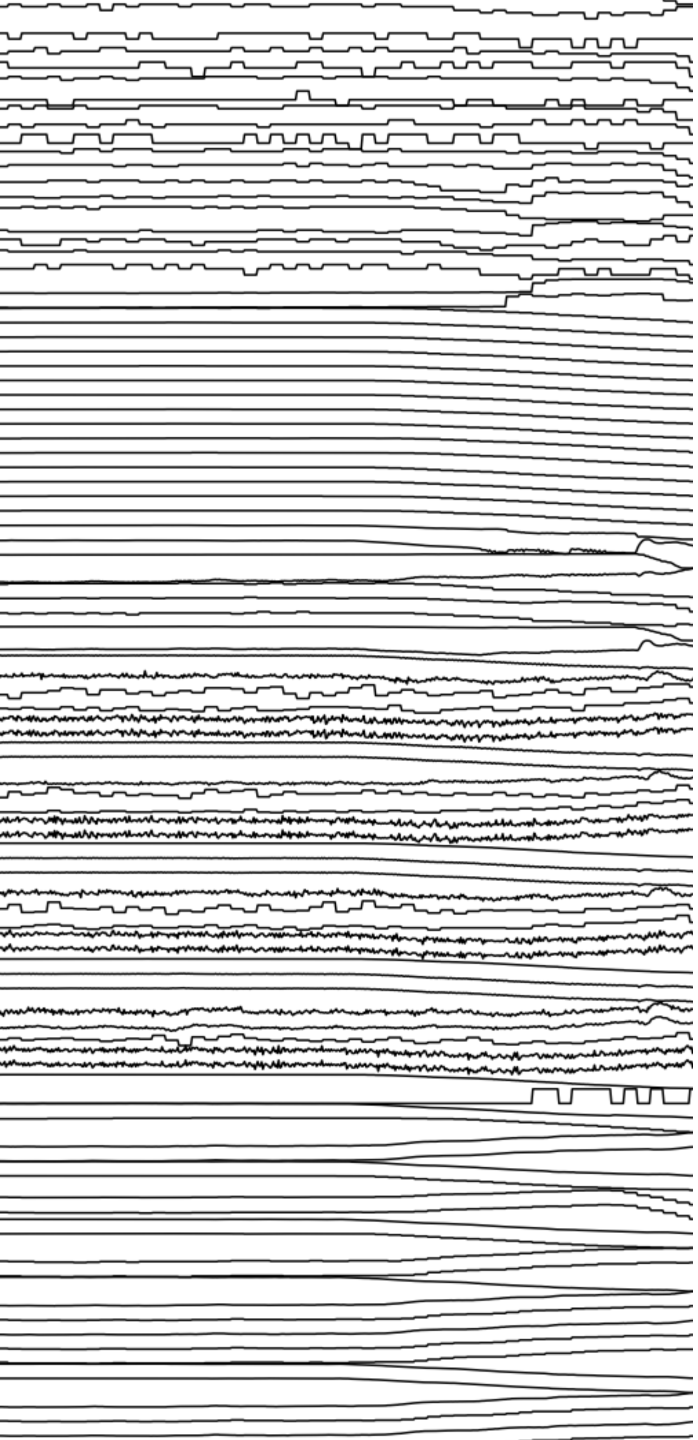
A solution?



Evaluation measures: *VUS*

Execution time evaluation on synthetic data to measure the influence of different parameters





Part 4: Anomaly Detection Benchmarks

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]

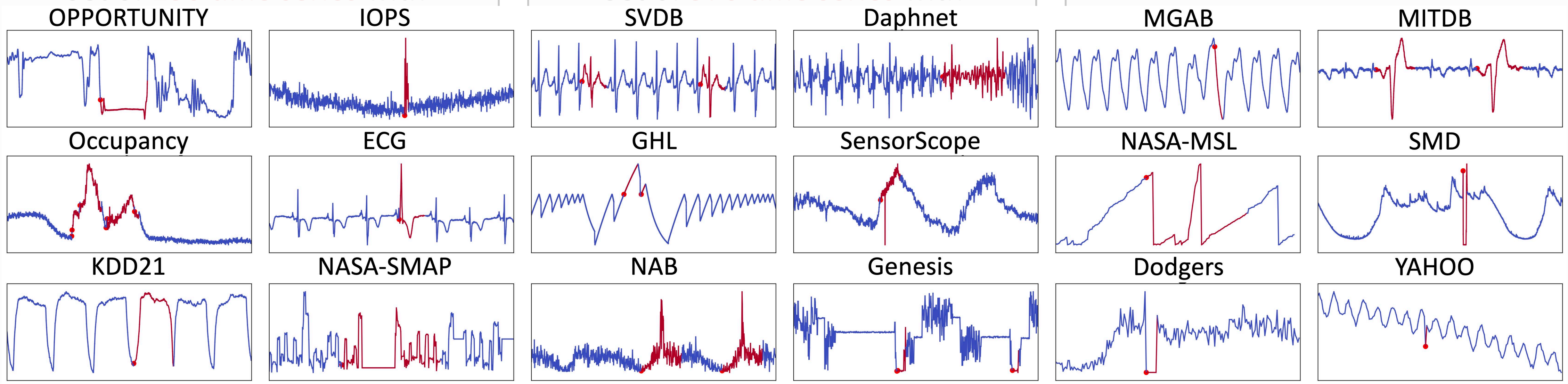
Set of 250 time series with

TimeEval [5]

TSB-UAD [19]

Set of 2000 time series with

Real datasets collection



only 1 labeled anomaly

contamination rate (< 0.1)

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

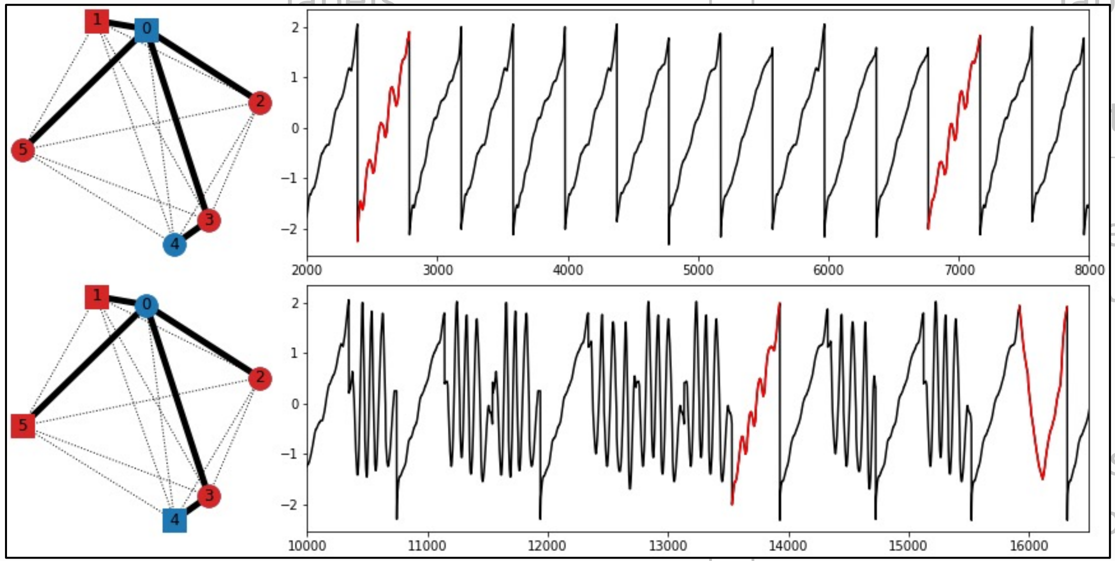
contamination, size or label quality.

Anomaly Detection methods: *Existing benchmark*

HEX/UCR [18]

Set of 250 ... labels

Artificial dataset generation



TimeEval [5]

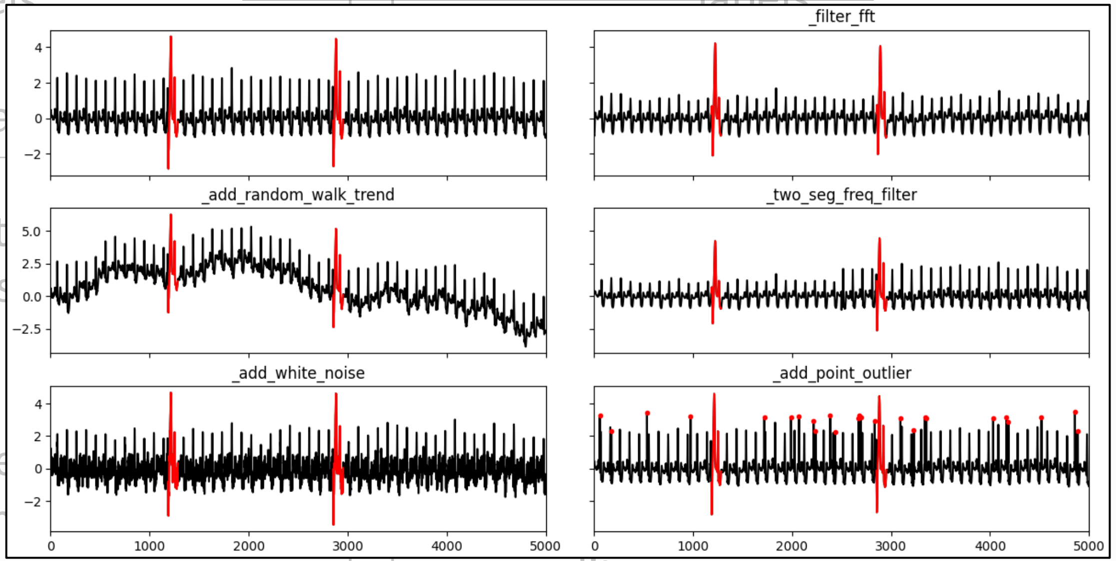
... of 976 time series with labels

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

TSB-UAD [19]

... ies with labels

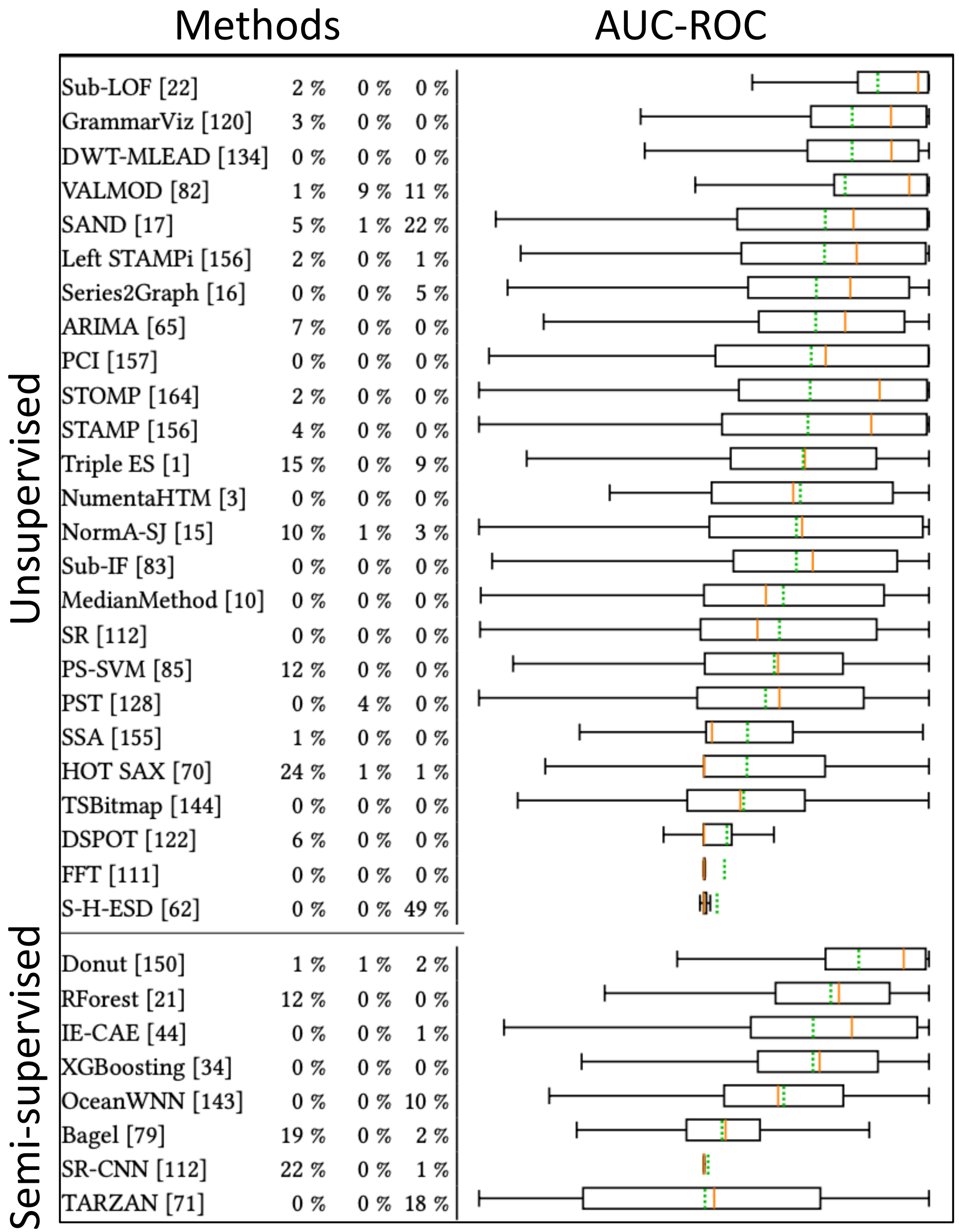
Synthetic dataset generation



quality.

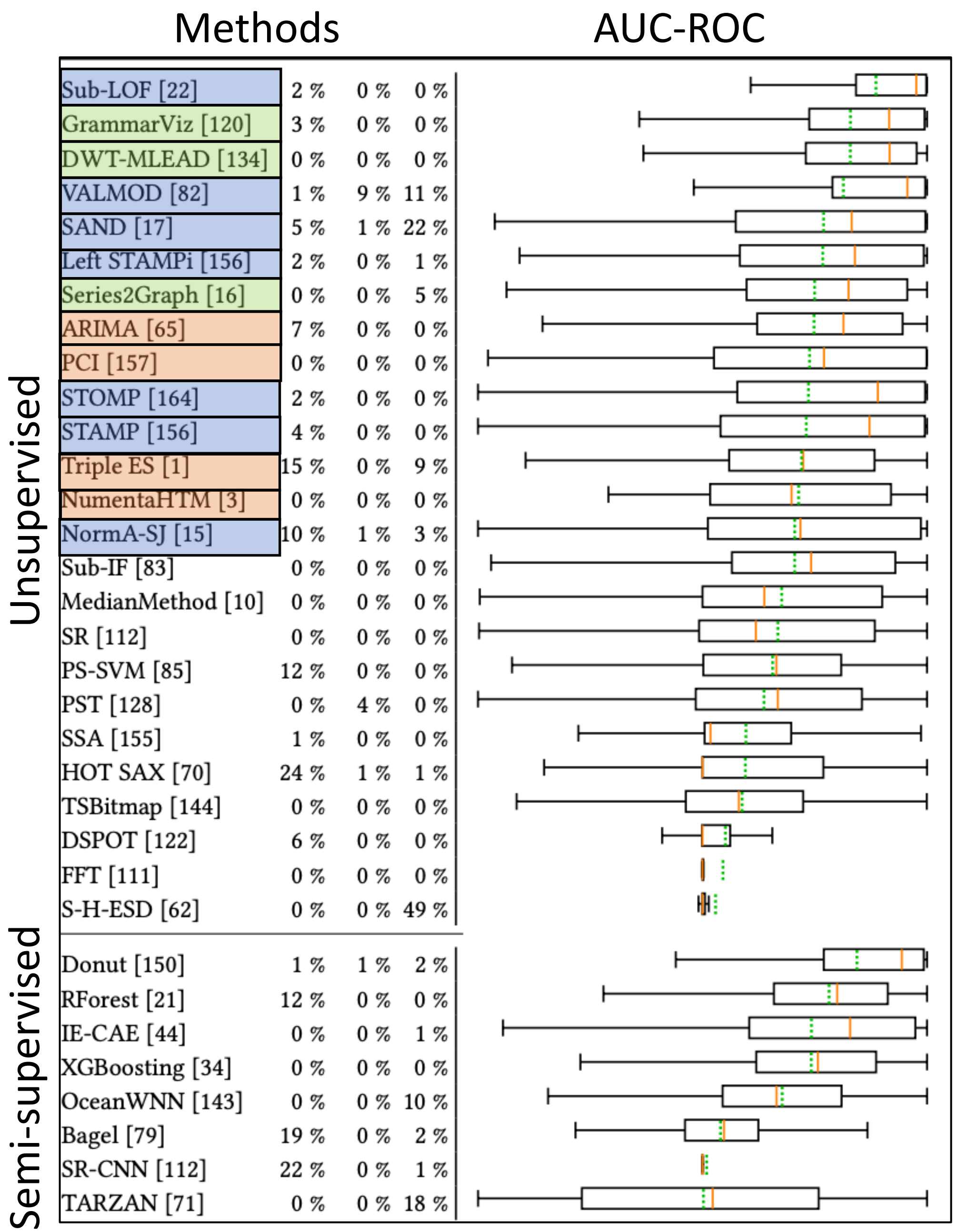
Anomaly Detection methods:
Experimental evaluation

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

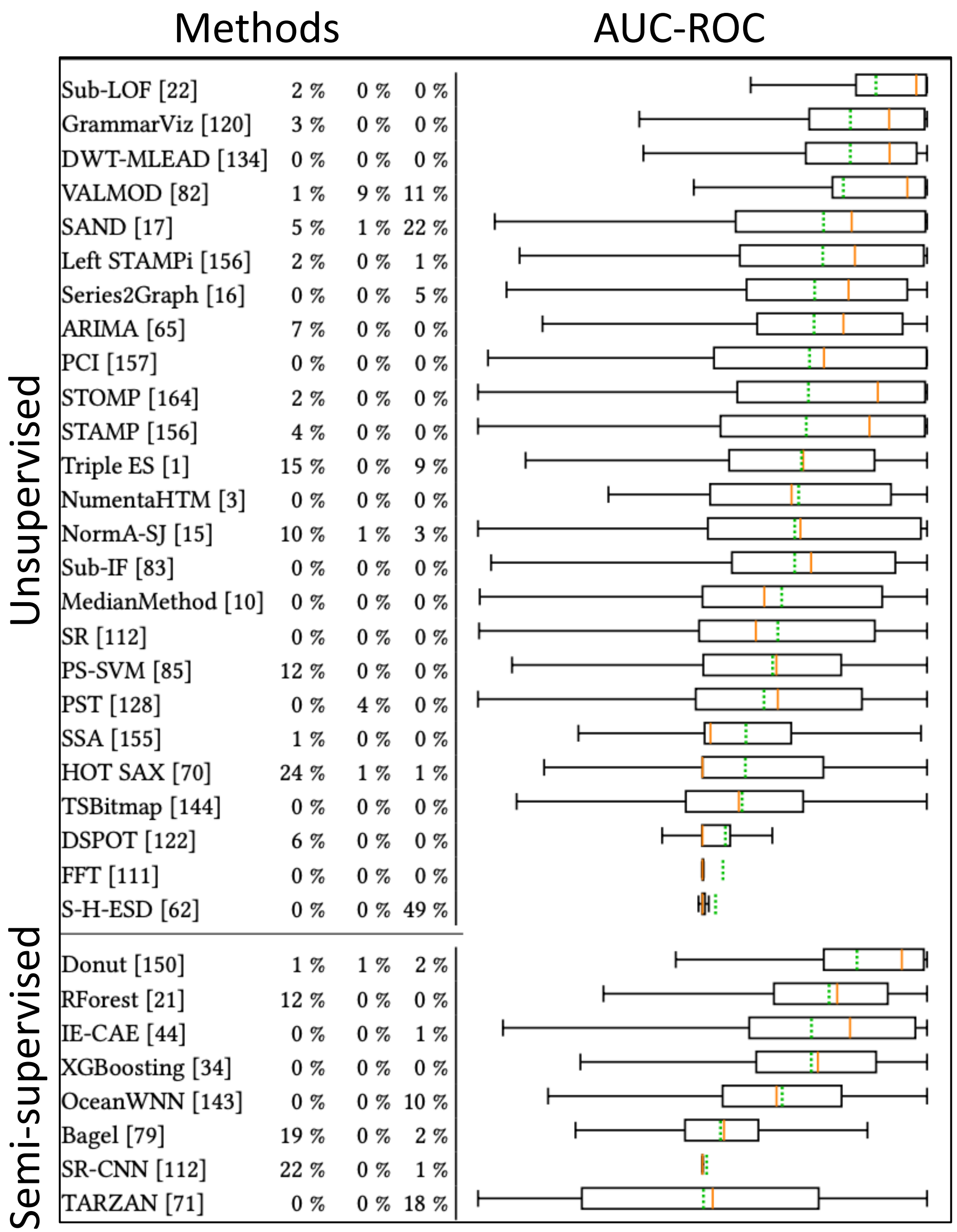
[5] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. Anomaly detection in time series: a comprehensive evaluation. Proc. VLDB Endow. 15, 9 (May 2022), 1779–1797.

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

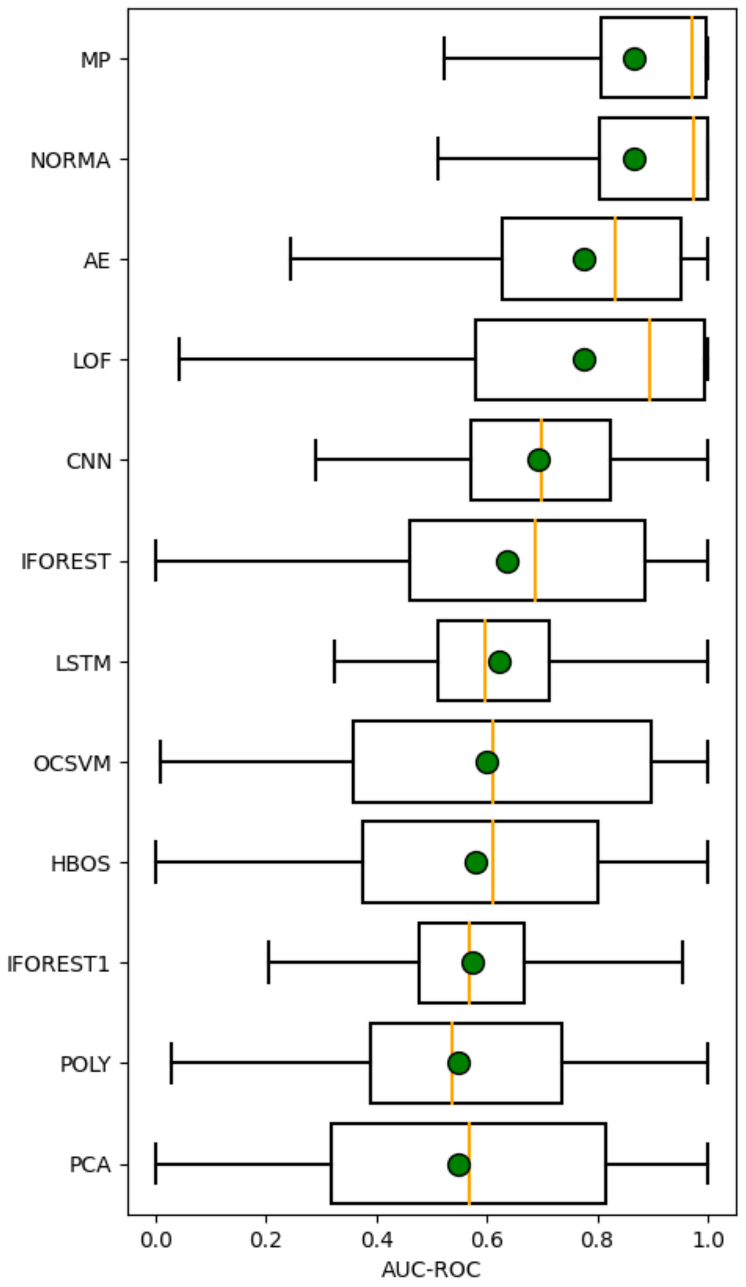


[5] Sebastian Schmidl, Phillip Wenig, and Thorsten Papenbrock. 2022. Anomaly detection in time series: a comprehensive evaluation. Proc. VLDB Endow. 15, 9 (May 2022), 1779–1797.

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

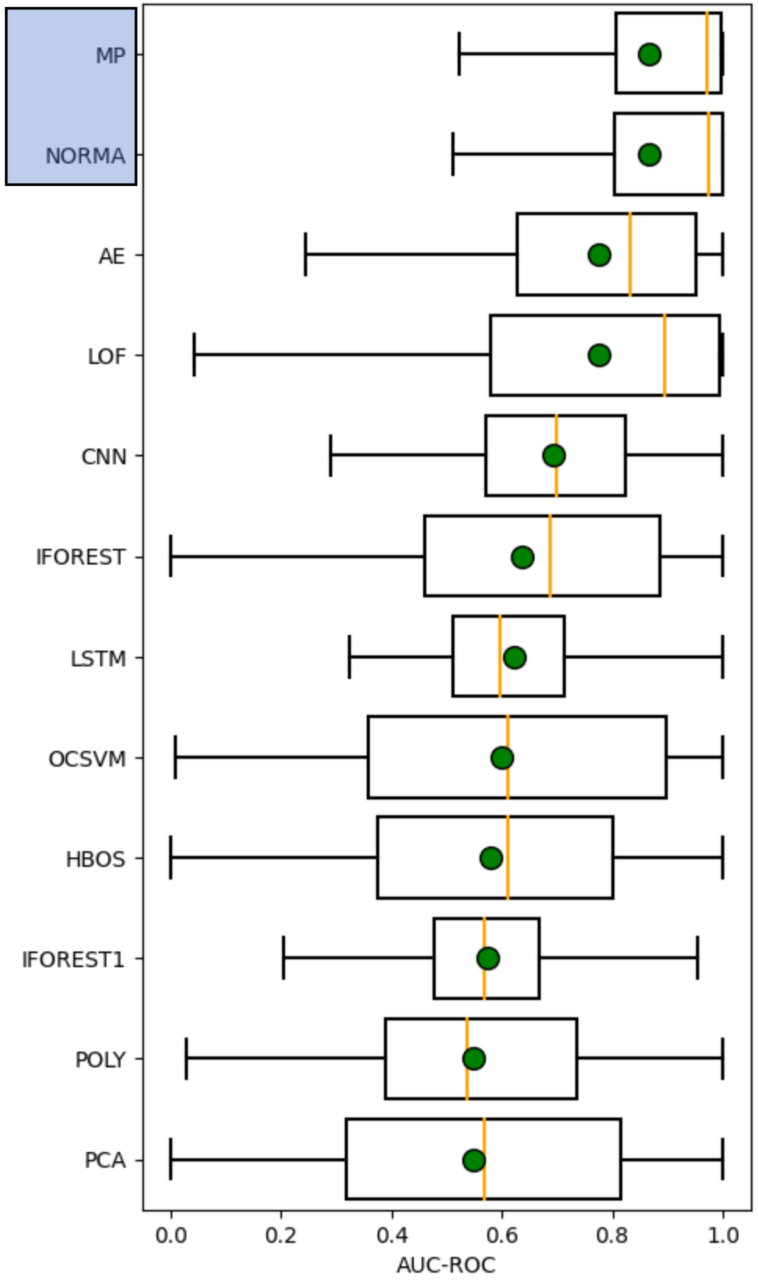


[18] R. Wu and E. Keogh, "Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress" in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 03, pp. 2421-2429, 2023.

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



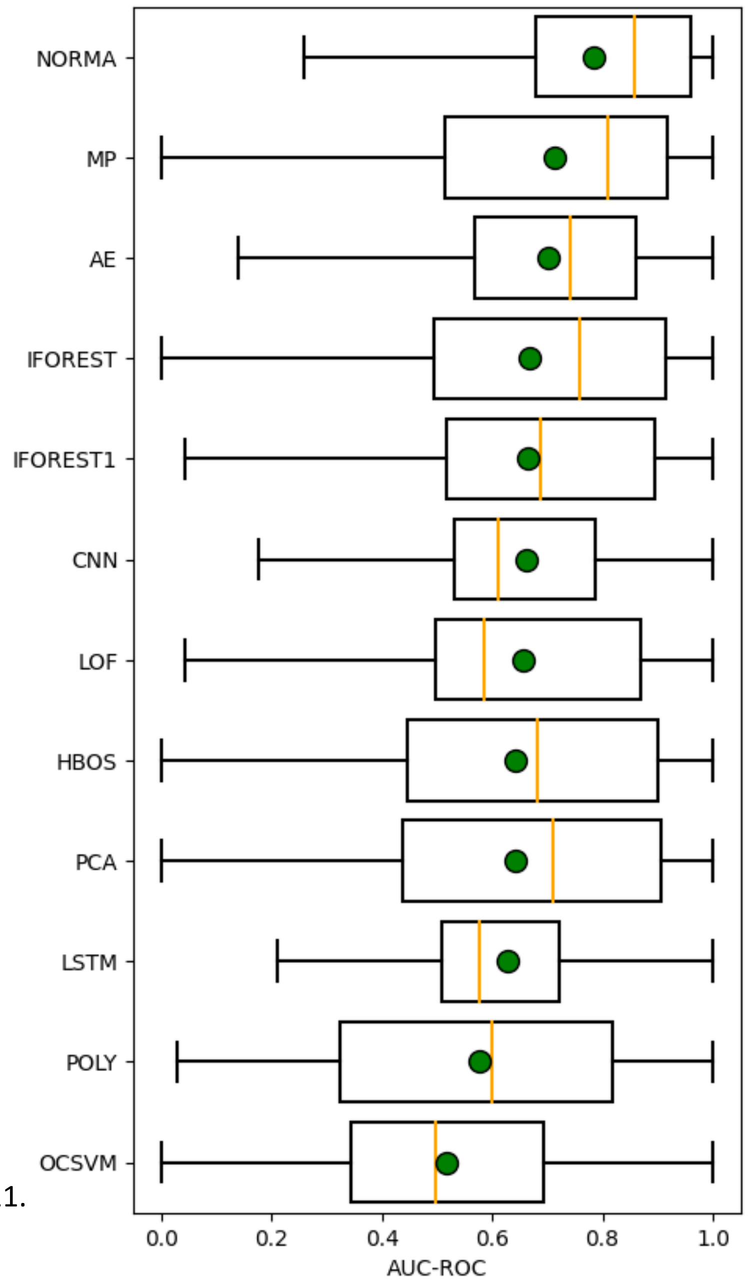
[18] R. Wu and E. Keogh, "Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress" in IEEE Transactions on Knowledge & Data Engineering, vol. 35, no. 03, pp. 2421-2429, 2023.

Anomaly Detection methods:
Experimental evaluation

Observations on TSB-UAD [19]:

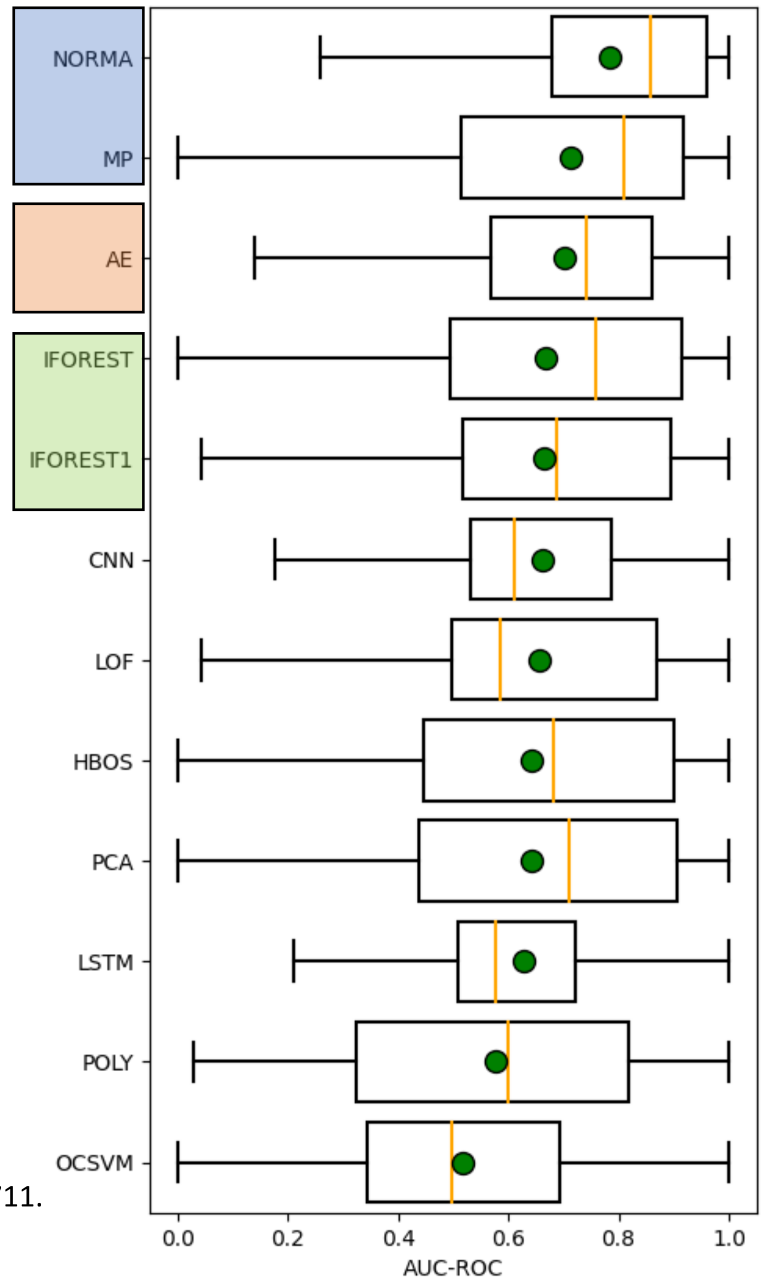
- Distance-based methods have a better accuracy (AUC-ROC) than forecasting-based methods.
- Isolation Forest (distribution-based and not proposed for time series) have also a strong accuracy

- 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.

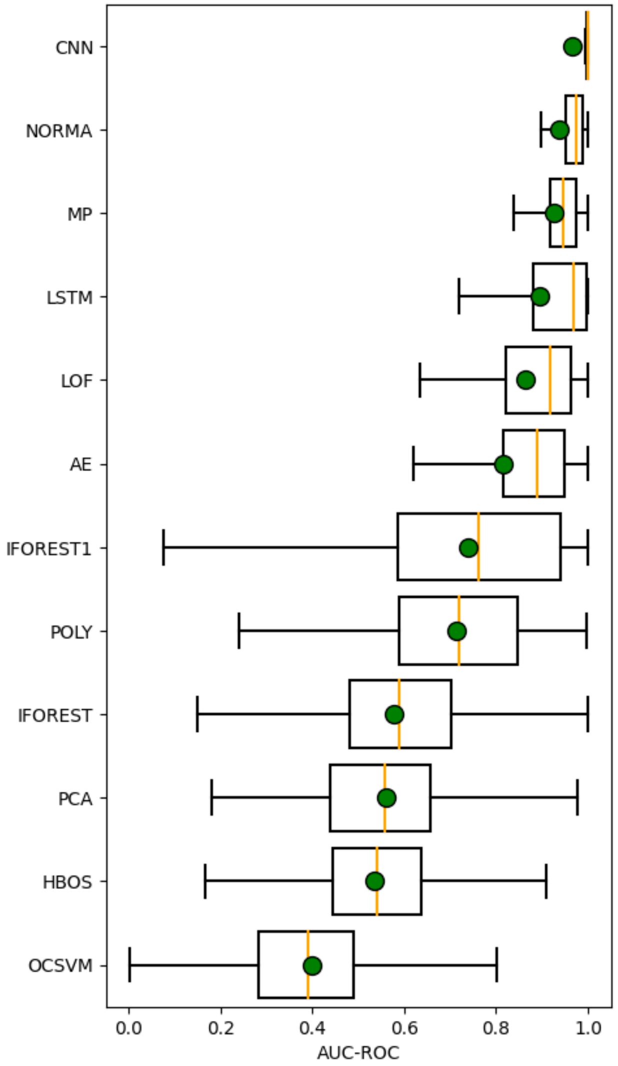
[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.

Anomaly Detection methods:

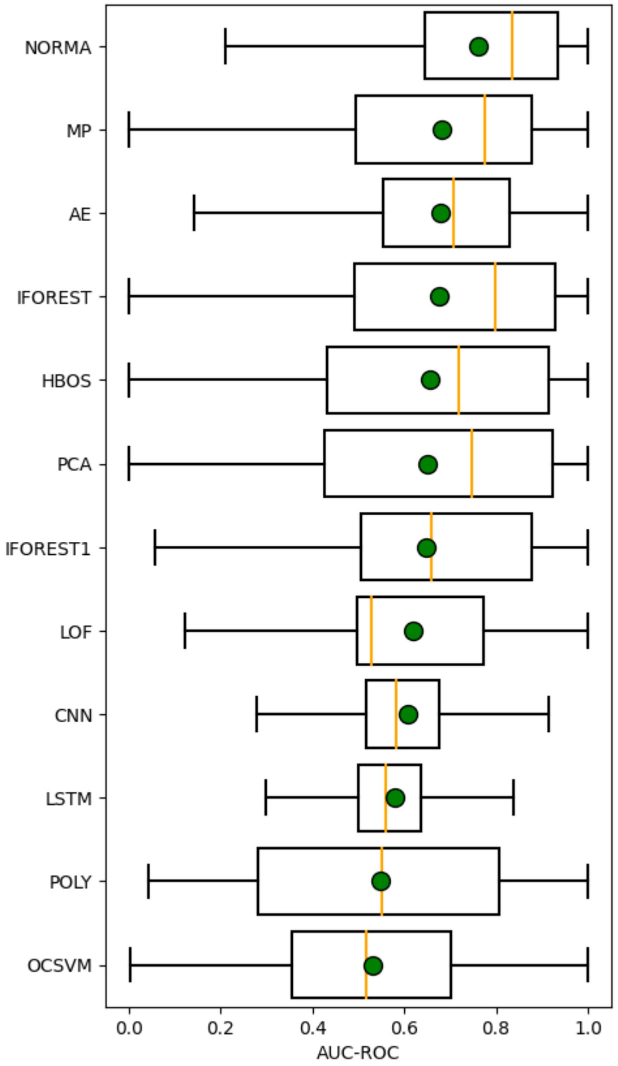
Experimental evaluation

Observations on TSB-UAD [19]:

Point-based anomaly



sequence-based anomaly



[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.

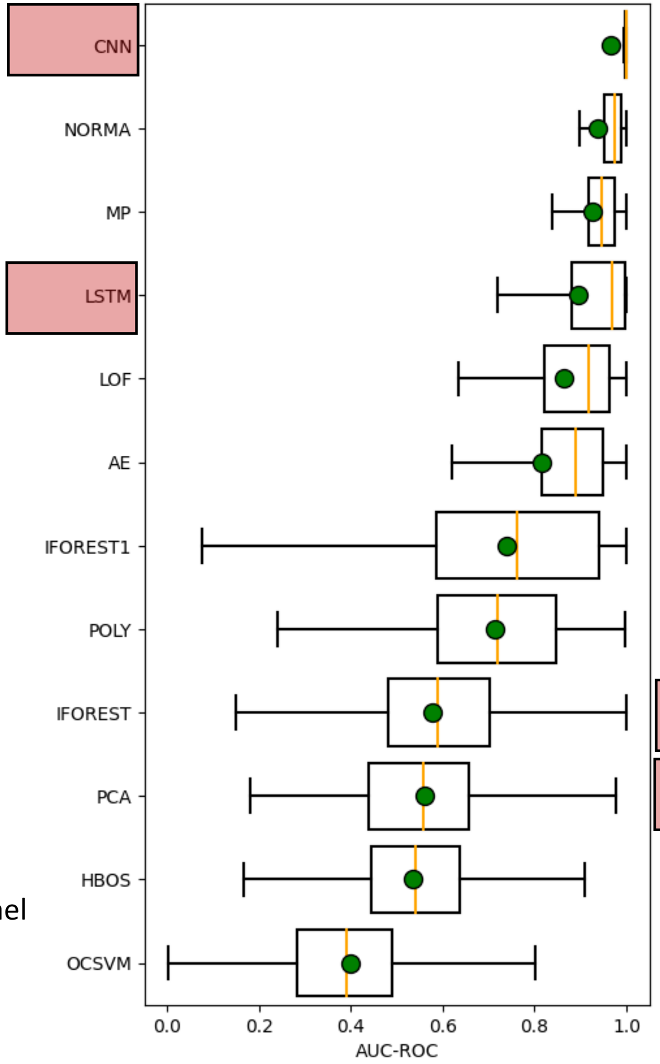
Anomaly Detection methods:

Experimental evaluation

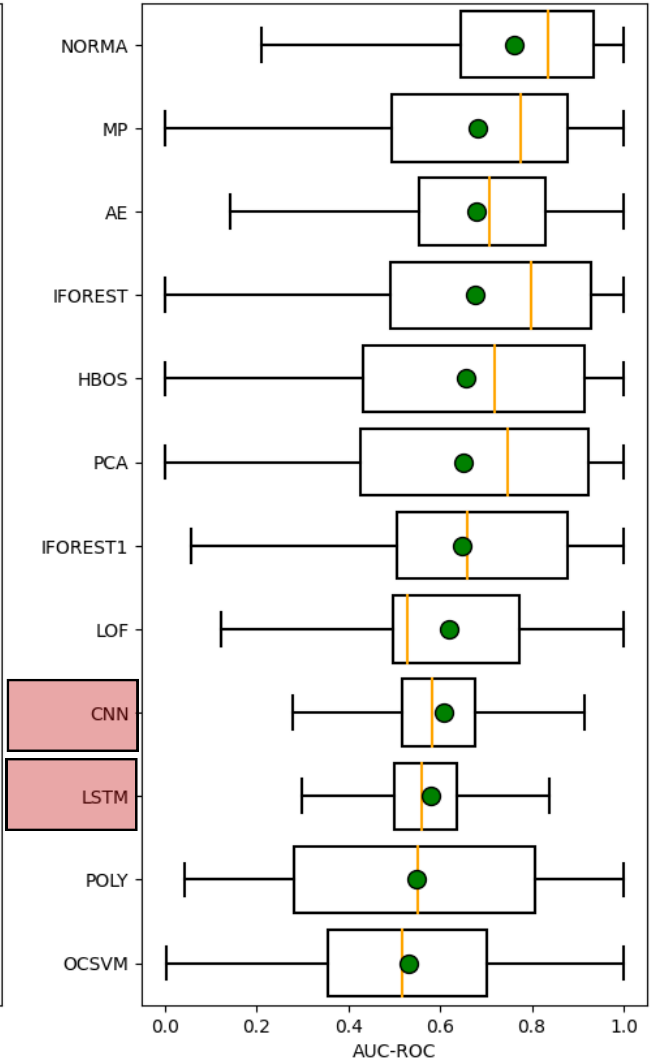
Observations on TSB-UAD [19]:

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

Point-based anomaly



sequence-based anomaly

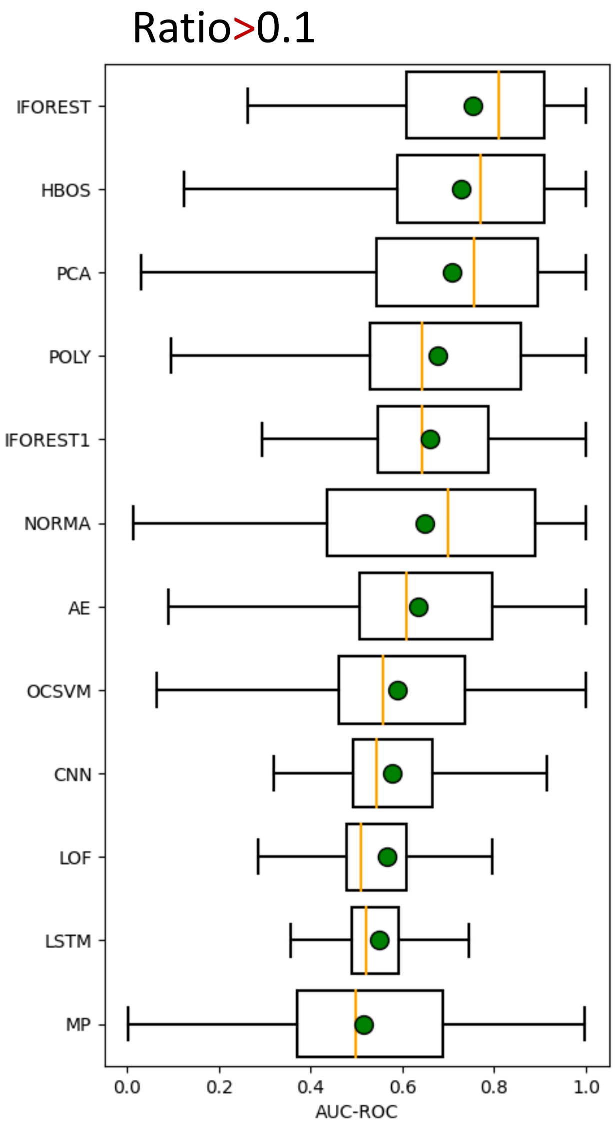


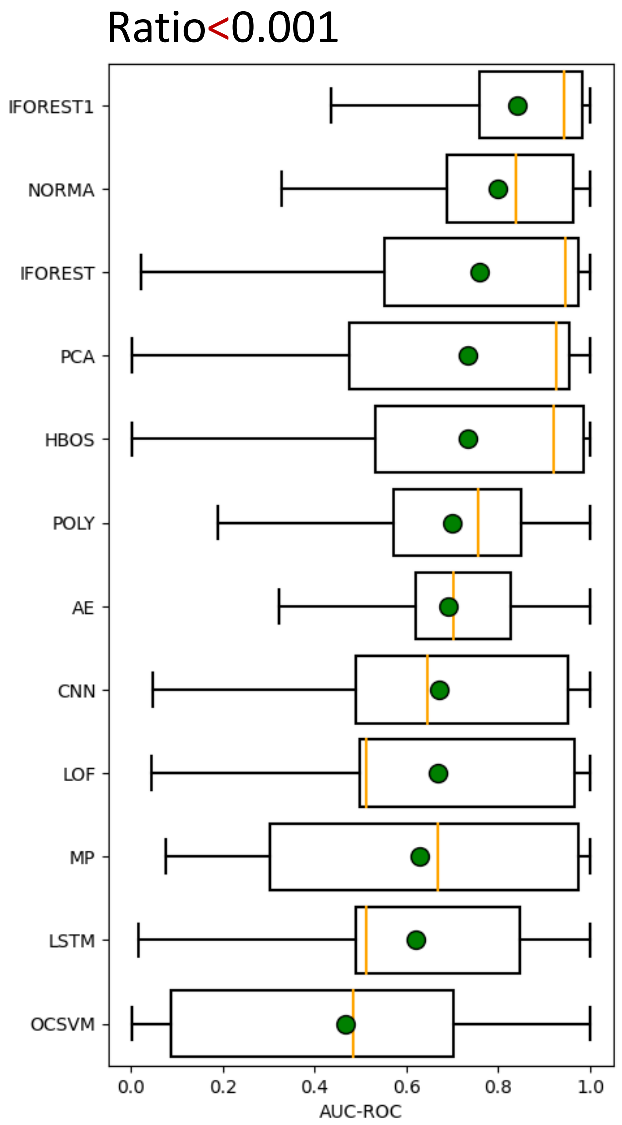
[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.

Anomaly Detection methods:
Experimental evaluation

Observations on TSB-UAD [19]:

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

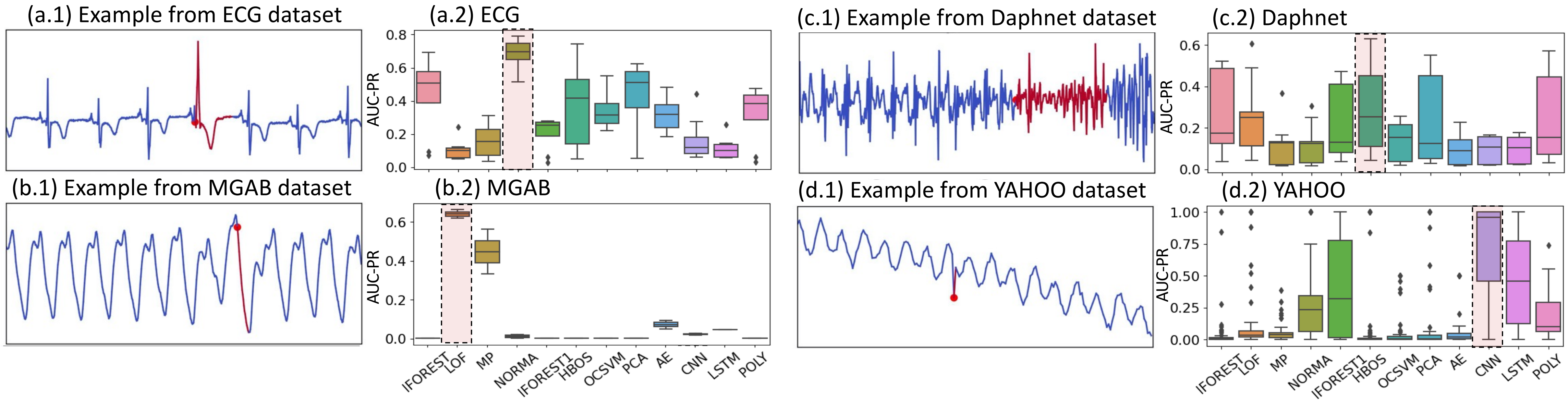




[19] John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (April 2022), 1697–1711.

Anomaly Detection methods: *Experimental evaluation*

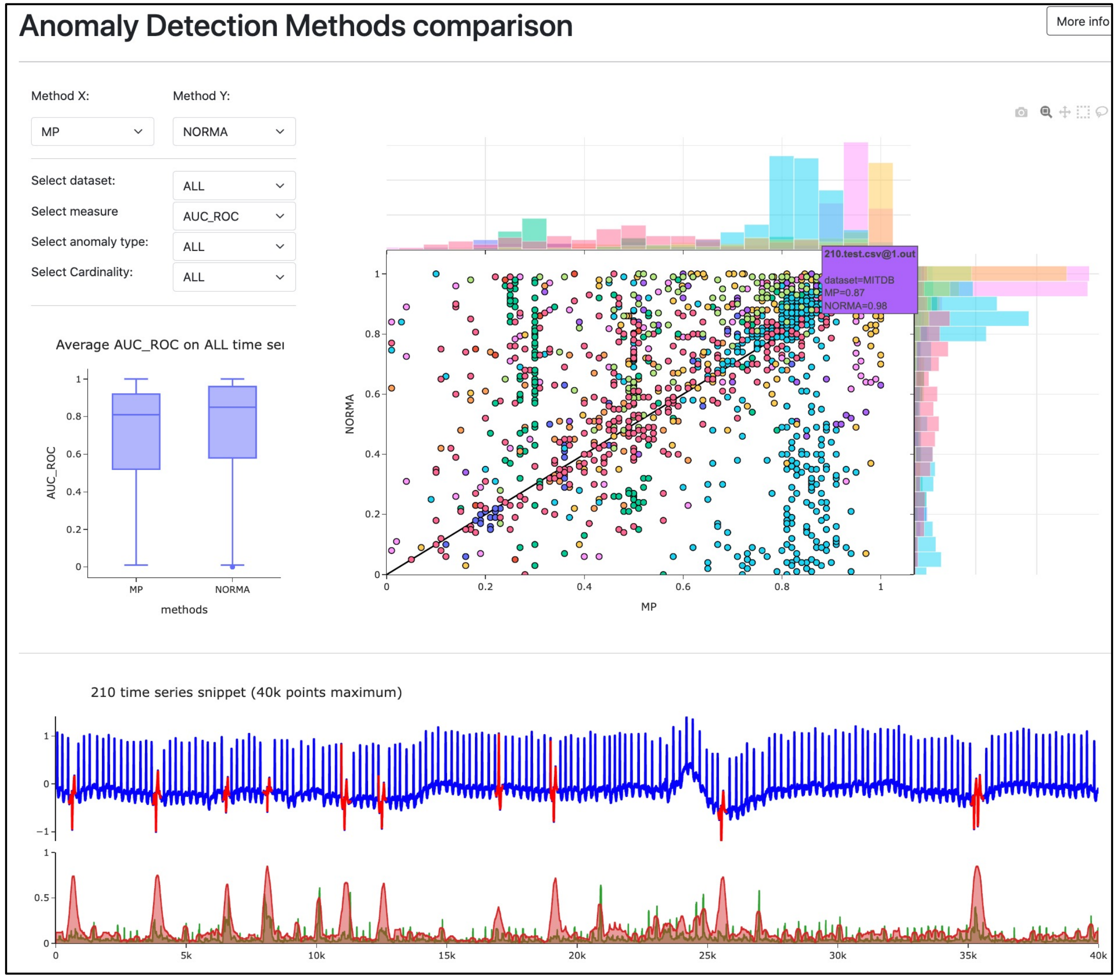
Observation from the results applied on specific datasets (TSB-UAD [19])



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.

Anomaly Detection methods: *Experimental evaluation*



Theseus [27]

An interactive tool to compare anomaly detection methods

VLDB 2022

Github repo

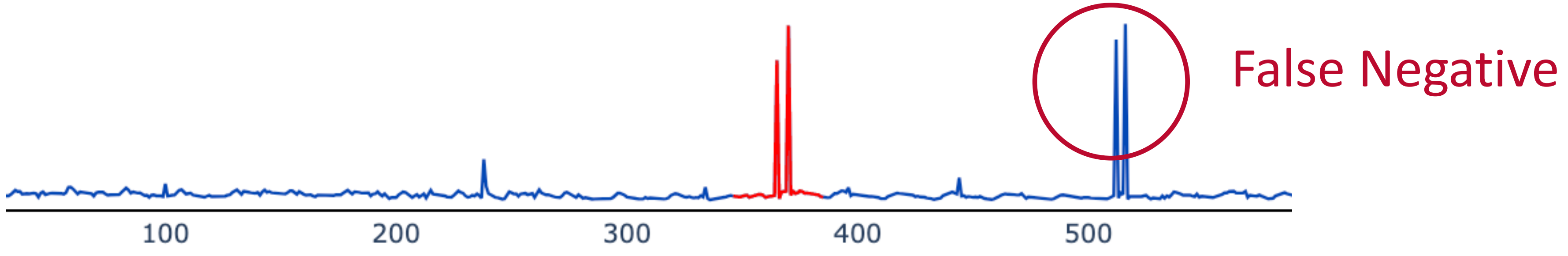
[27] Paul Boniol, John Paparrizos, Yuhao Kang, Themis Palpanas, Ruey S. Tsay, Aaron J. Elmore, and Michael J. Franklin. 2022. Theseus: navigating the labyrinth of time-series anomaly detection. Proc. VLDB Endow. 15, 12 (August 2022), 3702–3705.

Benchmark Practice: *Common Flaws in Dataset*

Mislabeling

Bias

Feasibility

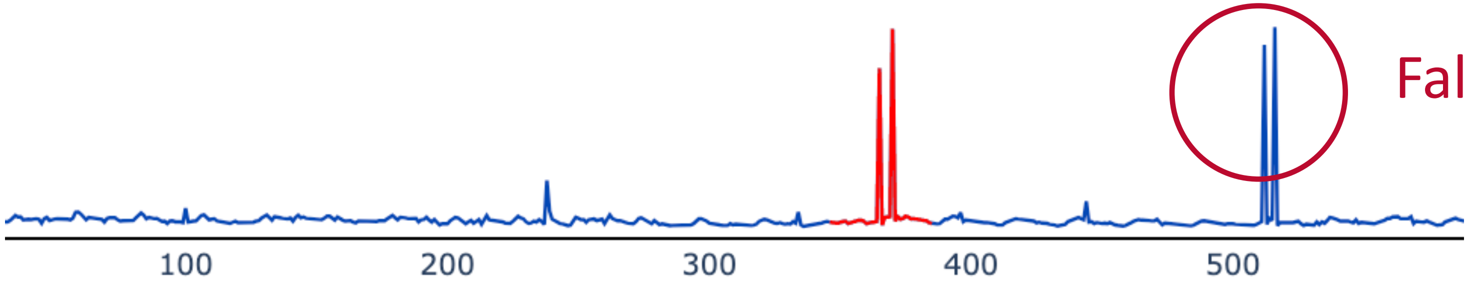


Benchmark Practice: *Common Flaws in Dataset*

Mislabeling

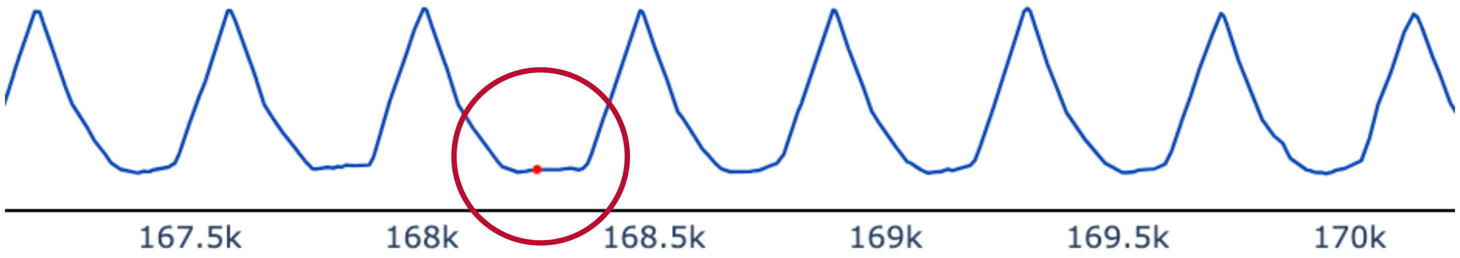
Bias

Feasibility



False Negative

False Positive

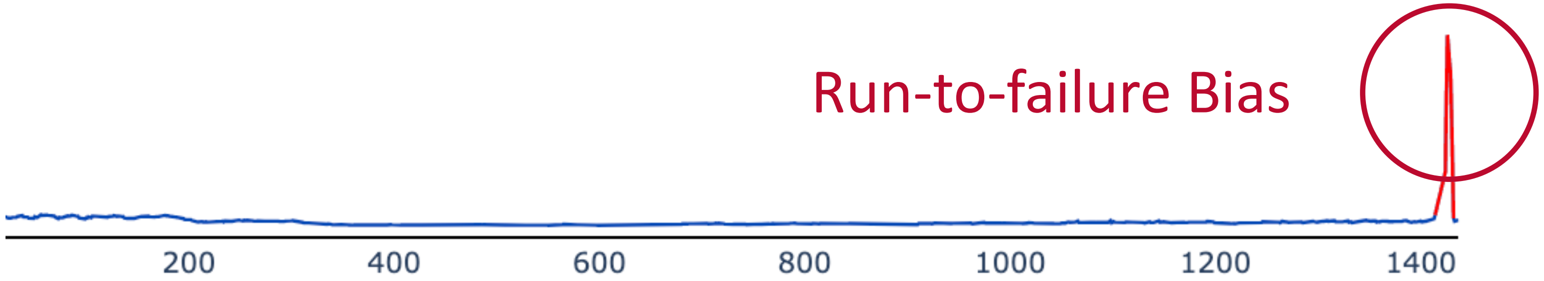


Benchmark Practice: *Common Flaws in Dataset*

Mislabeling

Bias

Feasibility

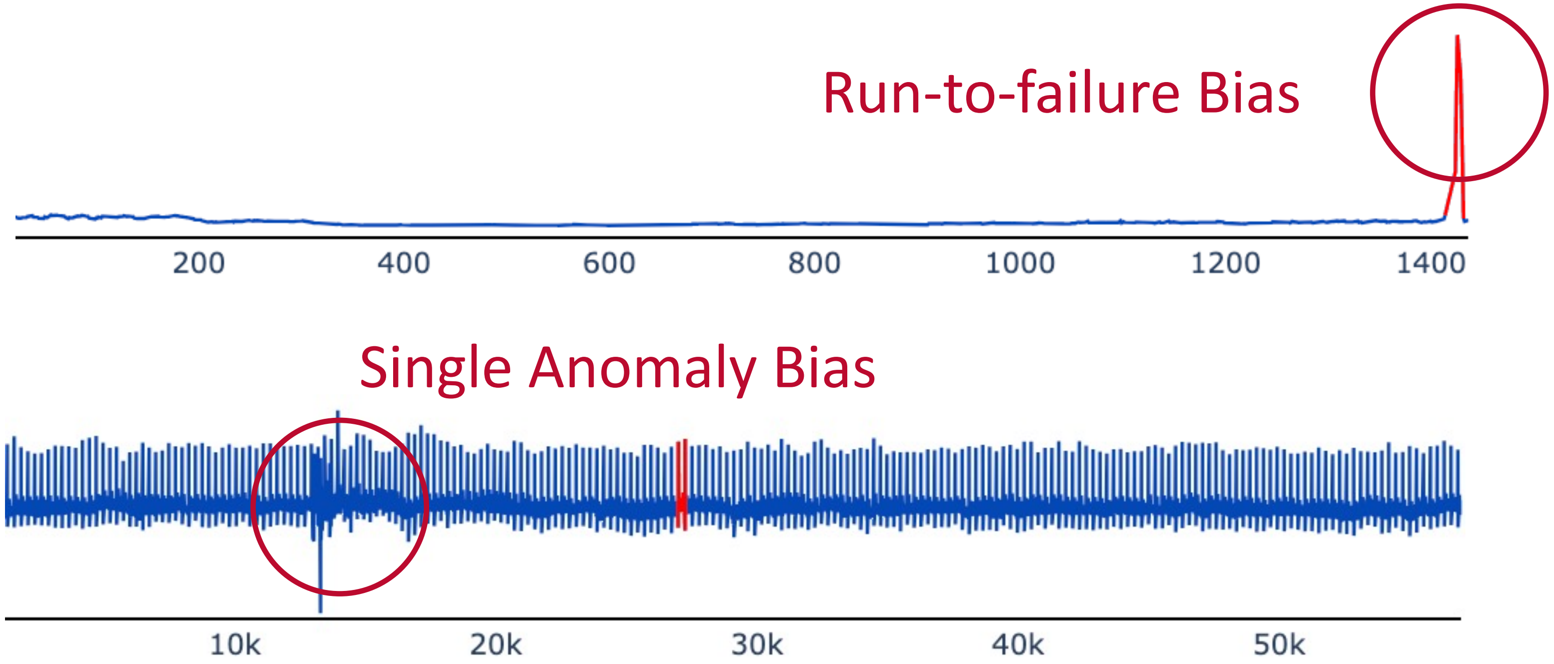


Benchmark Practice: *Common Flaws in Dataset*

Mislabeling

Bias

Feasibility



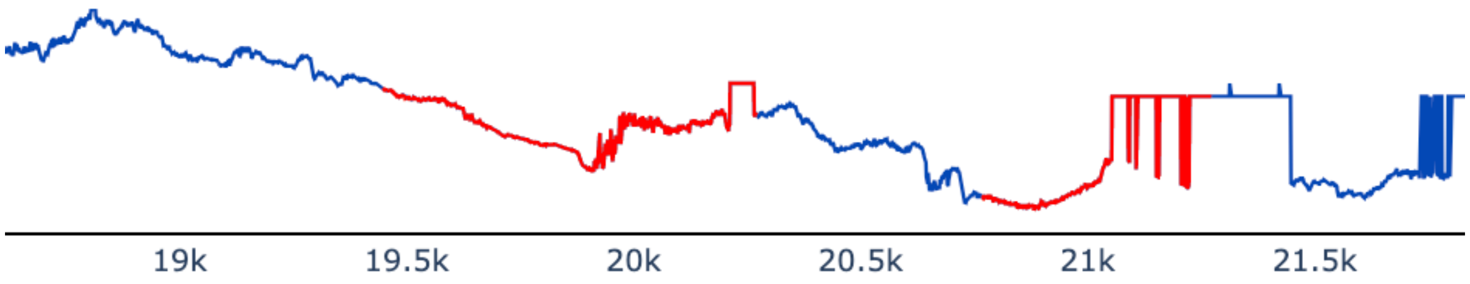
Benchmark Practice: *Common Flaws in Dataset*

Mislabeling

Bias

Feasibility

Lack of In-context Data



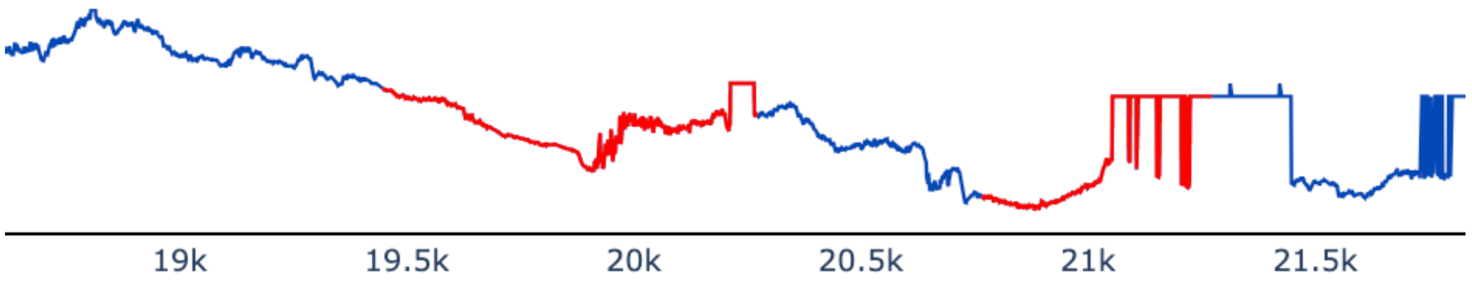
Benchmark Practice: *Common Flaws in Dataset*

Mislabeling

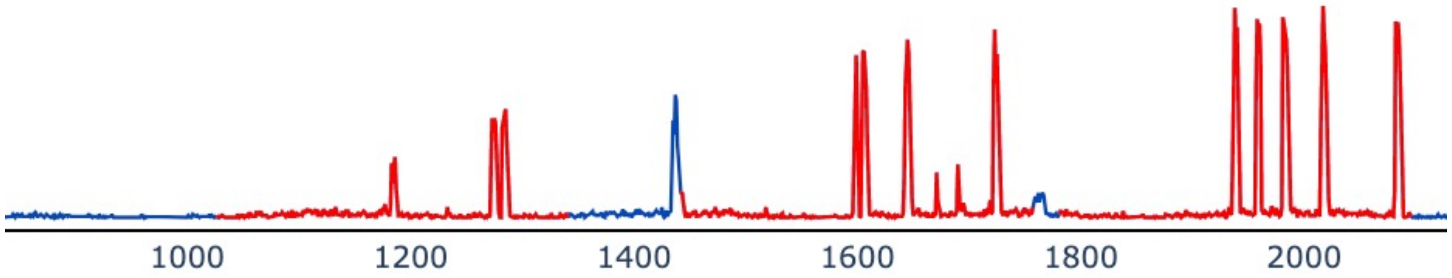
Bias

Feasibility

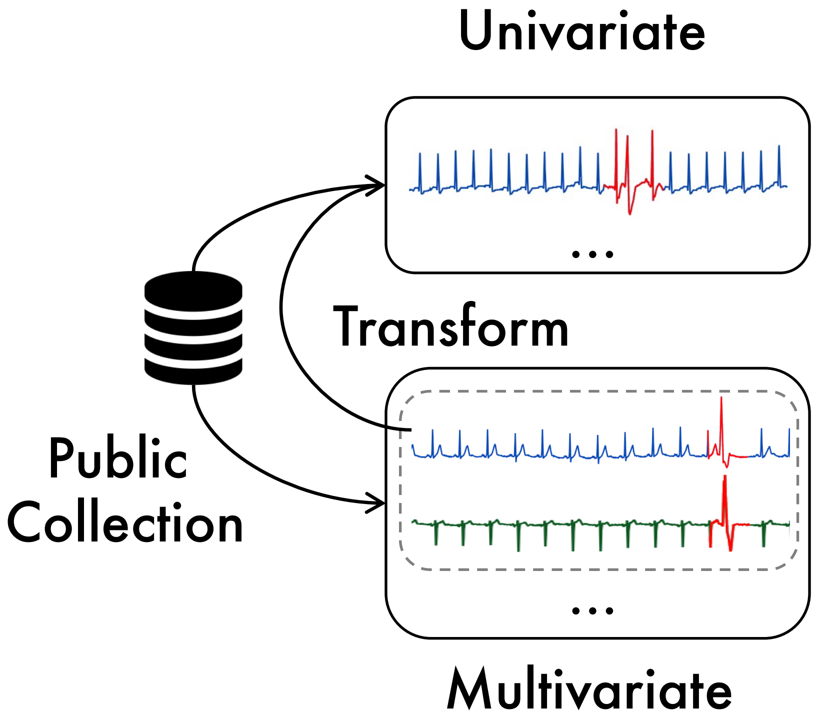
Lack of In-context Data



Unrealistic Anomaly Ratio

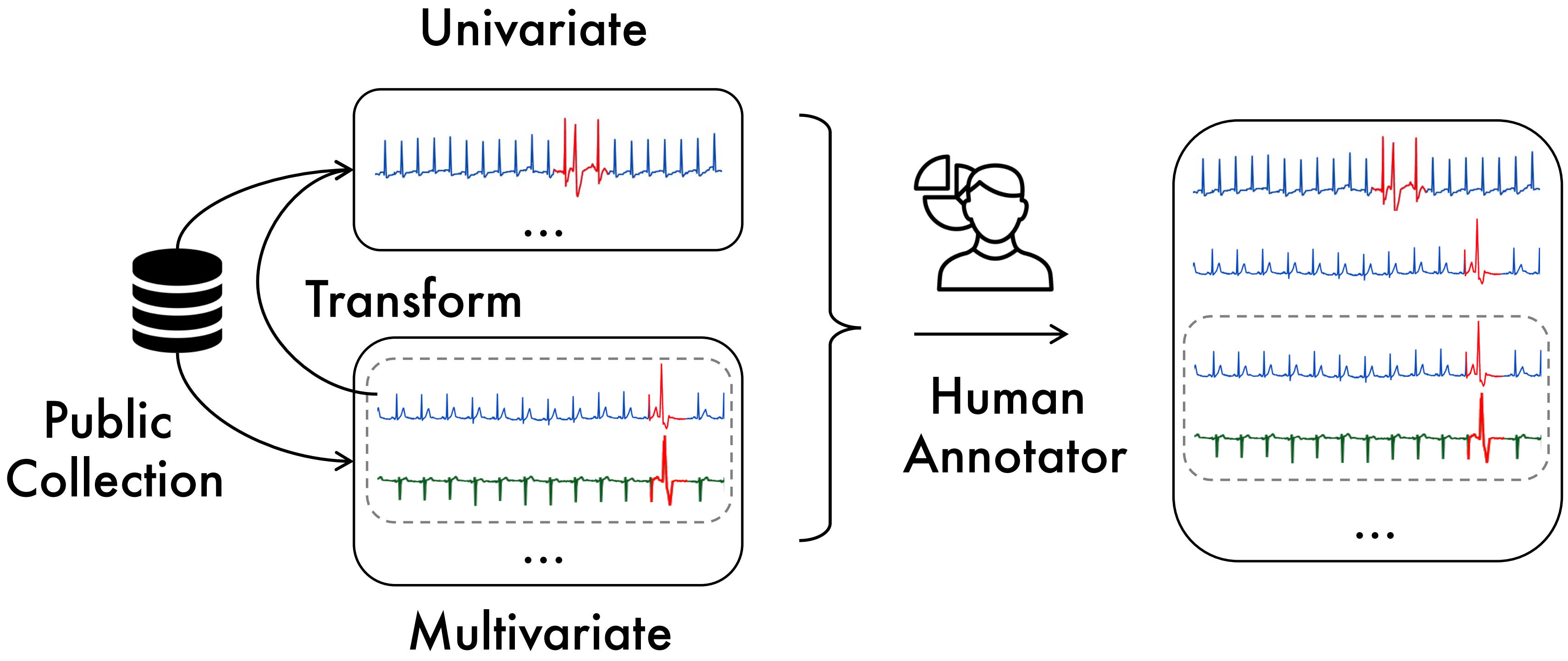


Benchmark Practice: *Dataset Construction*



Step 1:
Dataset Collection

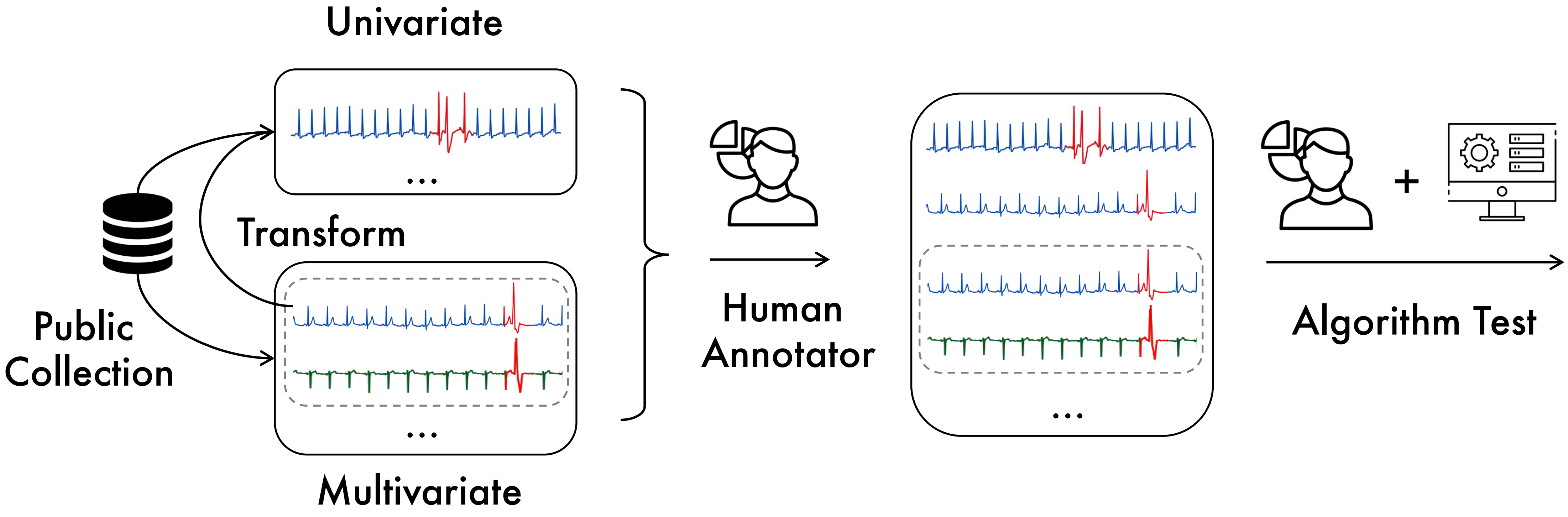
Benchmark Practice: *Dataset Construction*



Step 1:
Dataset Collection

Step 2:
Flaws Identification

Benchmark Practice: *Dataset Construction*



Step 1:
Dataset Collection

Step 2:
Flaws Identification

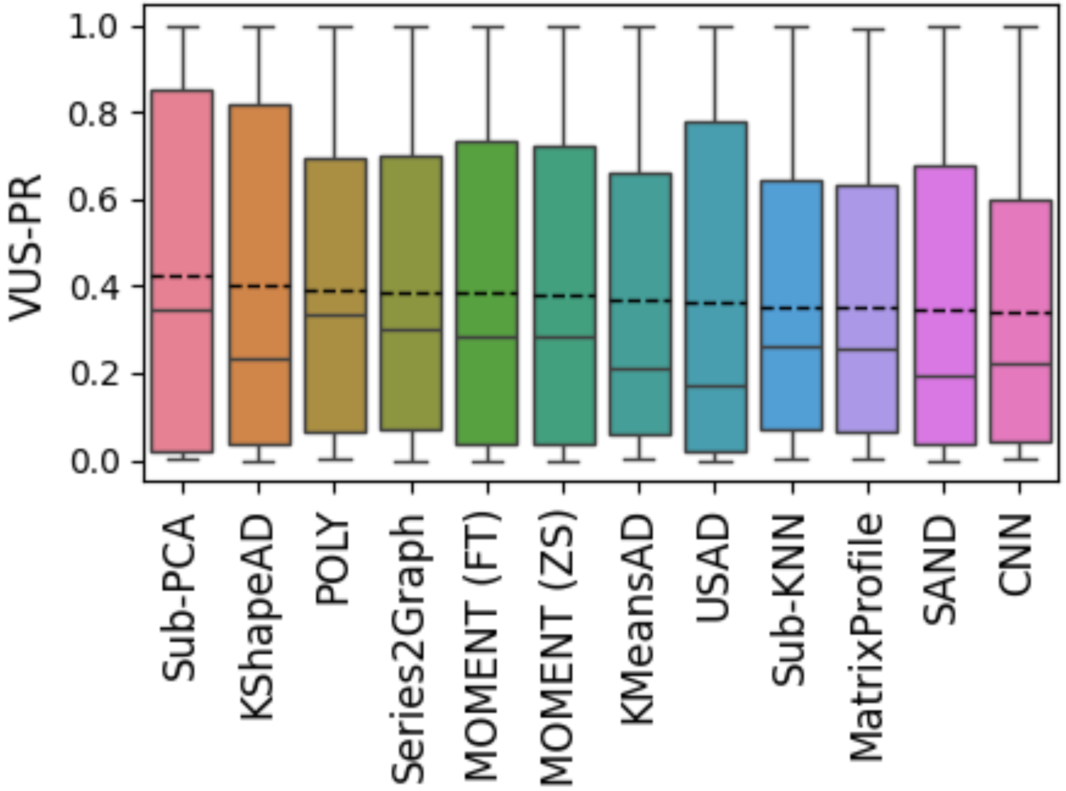
Step 3:
Label Quality Assessment

Benchmark Practice: *Dataset Construction*

Category	Split	# TS	Avg Length	Avg Anomaly Length	Avg # Anomalies	Anomaly Ratio
TSB-AD-U	All	870	38814.1	179.5	39.7	2.4%
	Eval	350	51886.7	321.3	46.6	4.5%
	Tuning	48	47143.3	185.9	82.6	3.5%
TSB-AD-M	All	200	107760.4	582.6	71.1	5.1%
	Eval	180	108826.7	591.2	67.7	5.0%
	Tuning	20	98164.1	504.7	101.1	5.7%

Benchmark Practice: *Evaluation*

TSB-AD-U

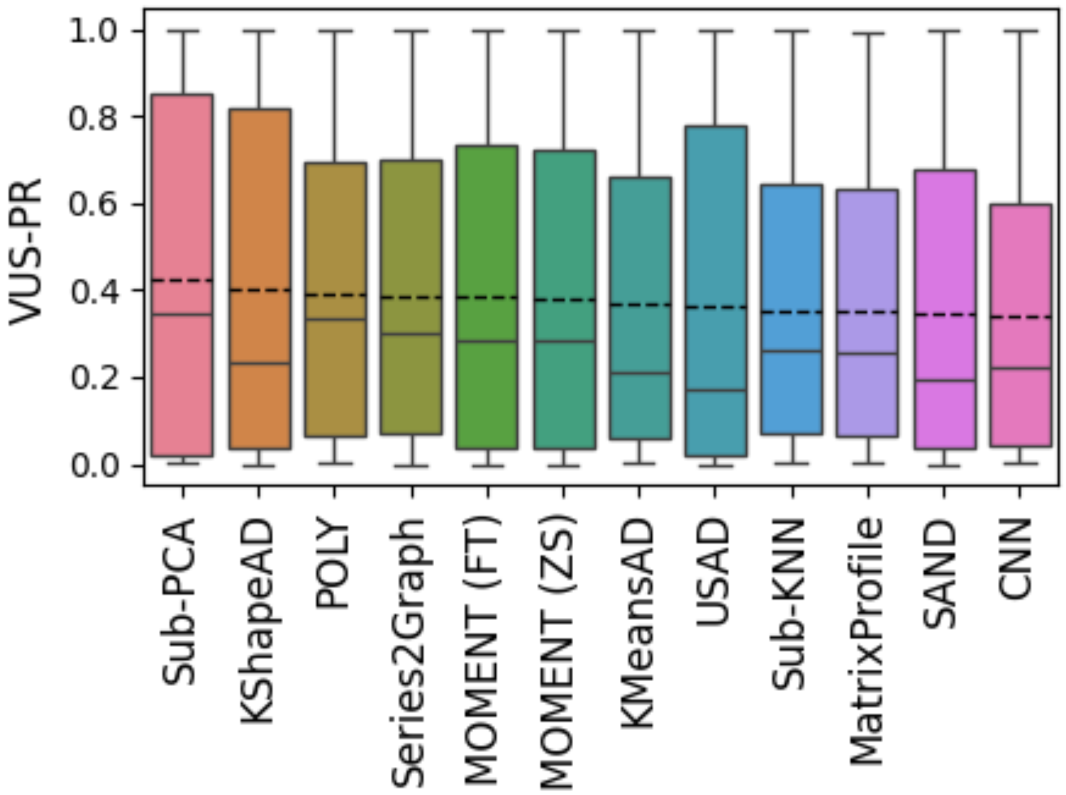


🚩 ***VUS-PR Ranking***

🥇	**Sub-PCA**
🥈	**KShapeAD**
🥉	**POLY**
4	Series2Graph
5	MOMENT (FT)
6	MOMENT (ZS)
7	KMeansAD
8	USAD
9	Sub-KNN
10	MatrixProfile
11	SAND
12	CNN

Benchmark Practice: *Evaluation*

TSB-AD-U



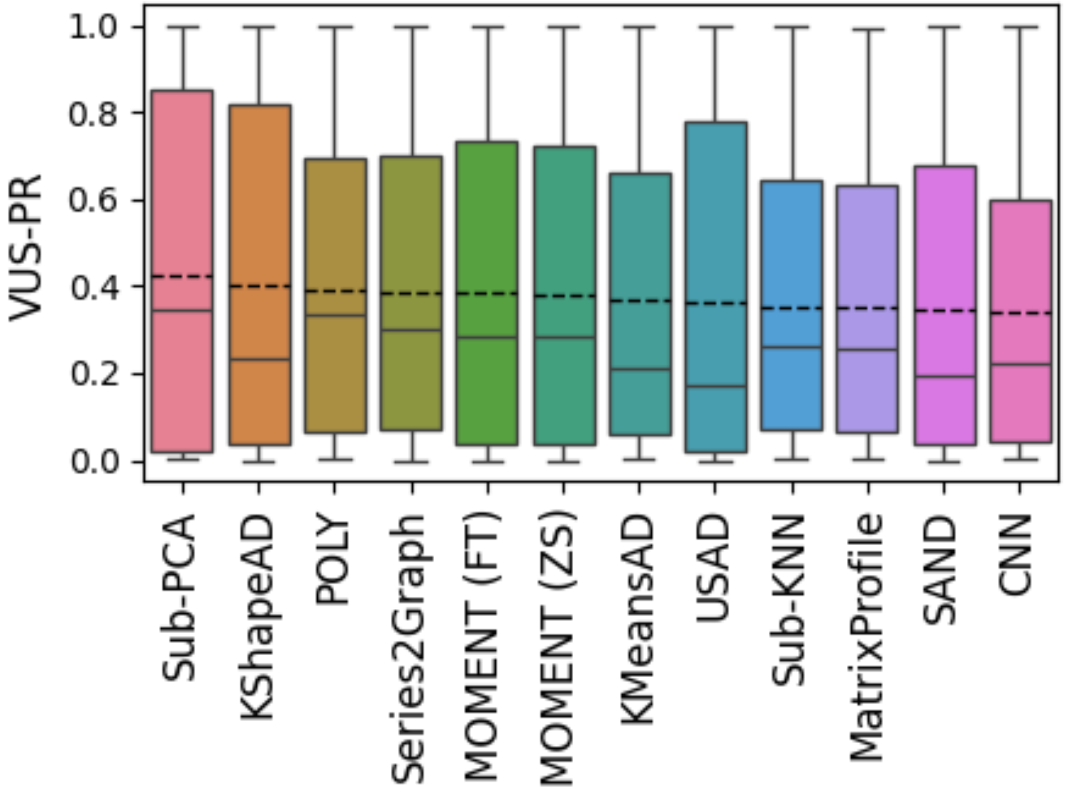
🚩 ***VUS-PR Ranking***

🥇	**Sub-PCA**
🥈	**KShapeAD**
🥉	**POLY**
4	Series2Graph
5	MOMENT (FT)
6	MOMENT (ZS)
7	KMeansAD
8	USAD
9	Sub-KNN
10	MatrixProfile
11	SAND
12	CNN

① Top-performing methods been overlooked for many years

Benchmark Practice: *Evaluation*

TSB-AD-U



🚩 ***VUS-PR Ranking***

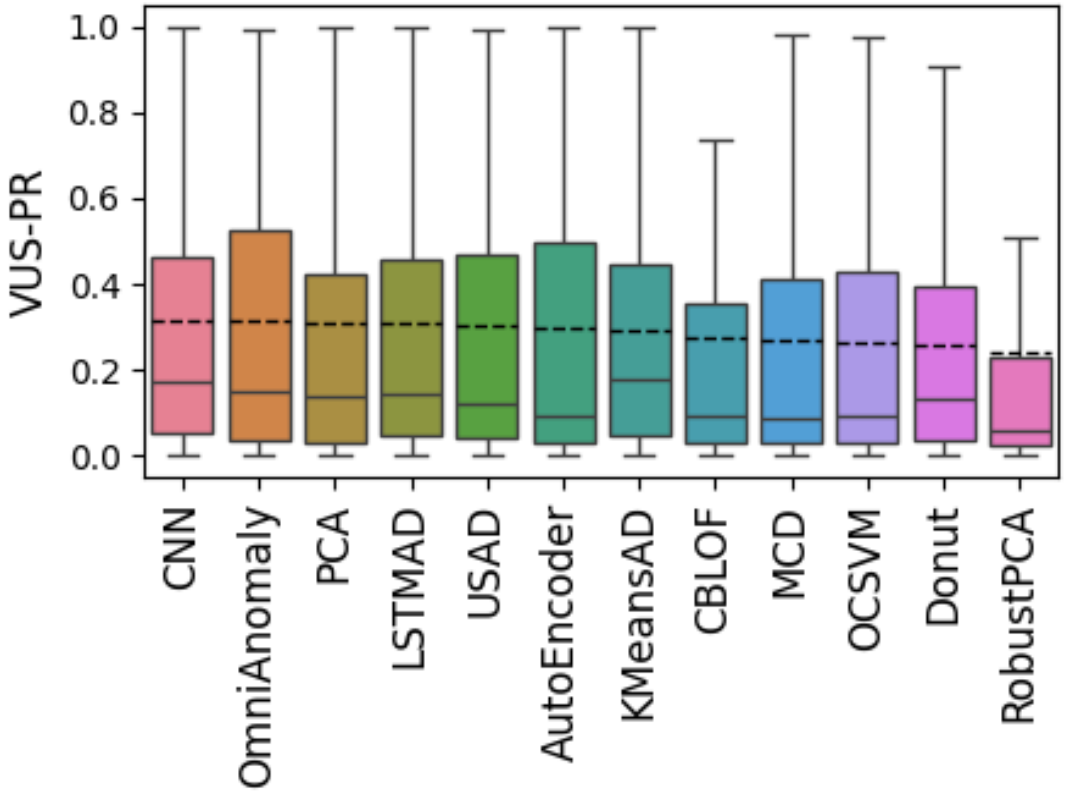
🥇	**Sub-PCA**
🥈	**KShapeAD**
🥉	**POLY**
4	Series2Graph
5	MOMENT (FT)
6	MOMENT (ZS)
7	KMeansAD
8	USAD
9	Sub-KNN
10	MatrixProfile
11	SAND
12	CNN

① Top-performing methods been overlooked for many years

② Performance of time-series foundation models shows promise

Benchmark Practice: *Evaluation*

TSB-AD-M



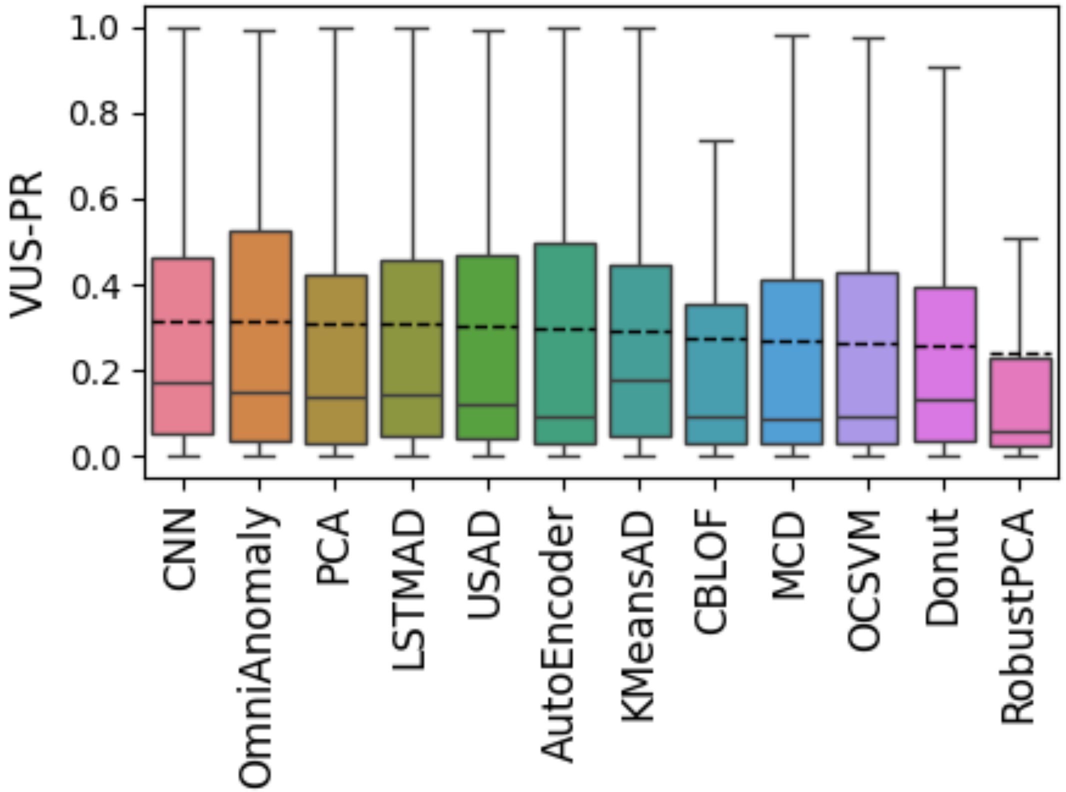
🚩 ***VUS-PR Ranking***

🥇	CNN
🥈	OmniAnomaly
🥉	PCA
4	LSTMAD
5	USAD
6	AutoEncoder
7	KMeansAD
8	CBLOF
9	MCD
10	OCSVM
11	Donut
12	RobustPCA

③ Neural-network-based methods strive in multivariate cases

Benchmark Practice: *Evaluation*

TSB-AD-M



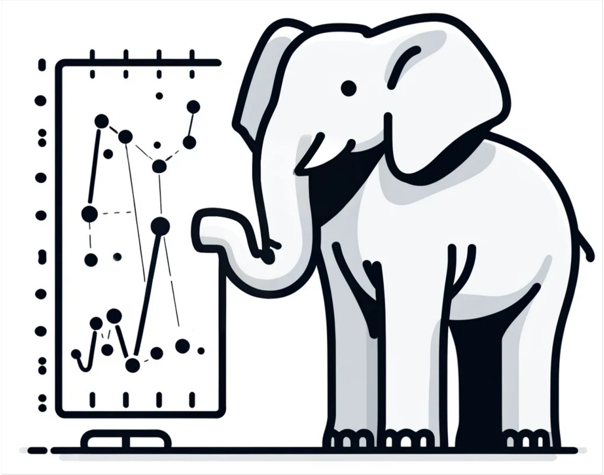
🚩 ***VUS-PR Ranking***

🥇	CNN
🥈	OmniAnomaly
🥉	PCA
4	LSTMAD
5	USAD
6	AutoEncoder
7	KMeansAD
8	CBLOF
9	MCD
10	OCSVM
11	Donut
12	RobustPCA

③ Neural-network-based methods strive in multivariate cases

④ Simpler architectures generally outperform more complex designs

Benchmark Practice: *Evaluation*



1070 Curated Time Series

40 TSAD Algorithms

10 Evaluation Measures

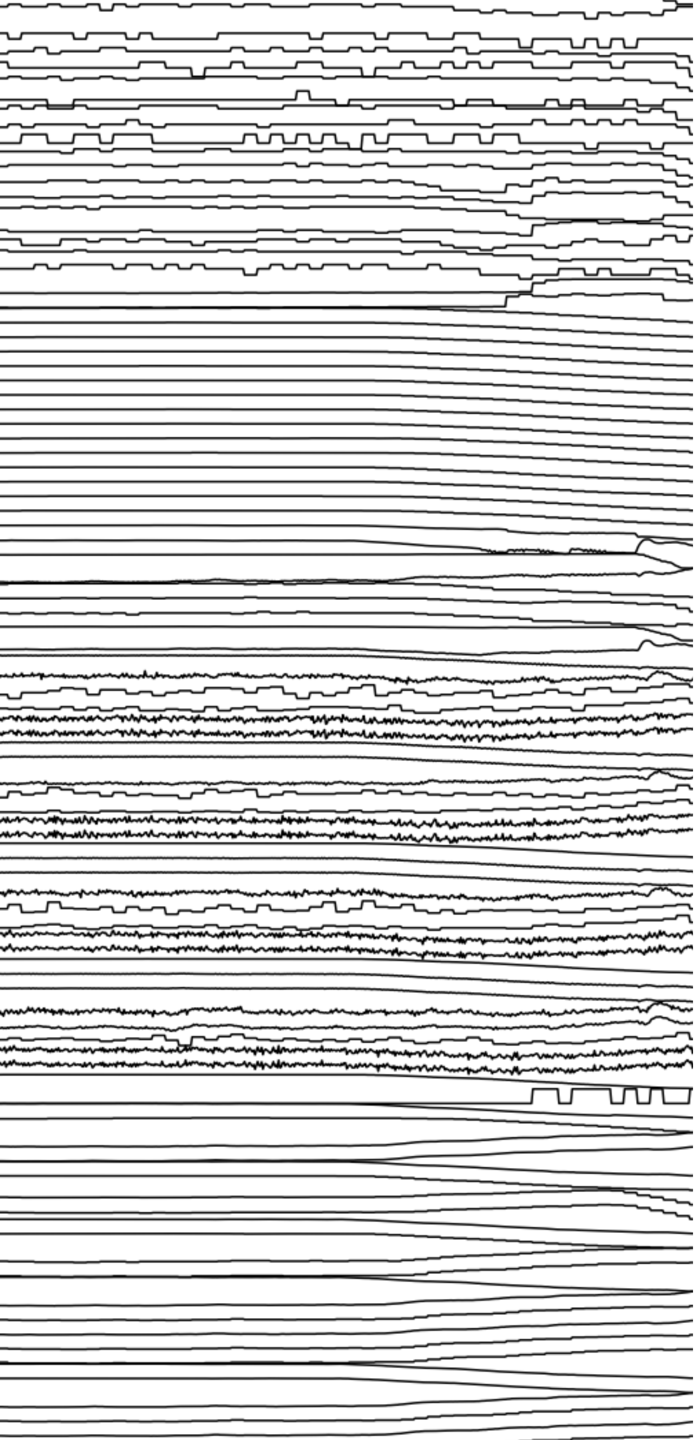
TSB-AD [27]

Towards a reliable time-series anomaly detection benchmark

NeurIPS 2024

Homepage

[27] Liu, Q. and Paparrizos, J., 2024. The elephant in the room: Towards a reliable time-series anomaly detection benchmark. Advances in Neural Information Processing Systems, 37, pp.108231-108261.

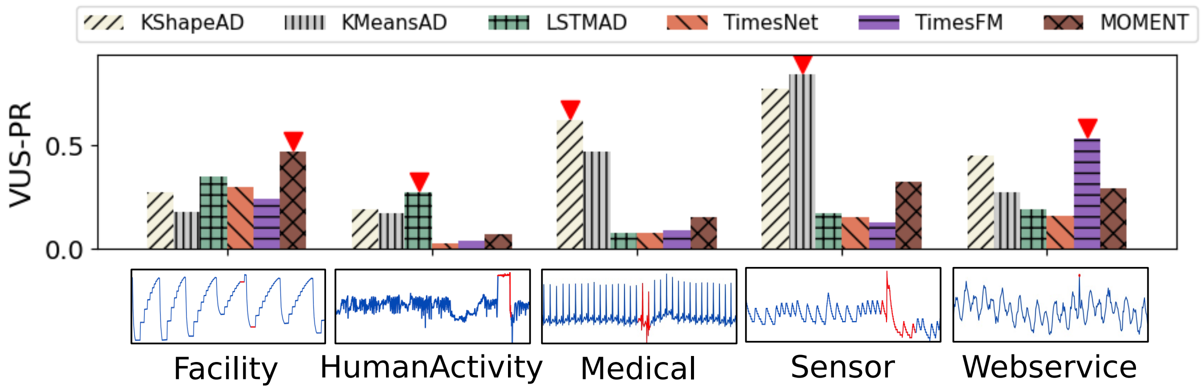


Part 5: Automated Solutions for Anomaly Detection

Automated Solutions: *Background*

Motivation:

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

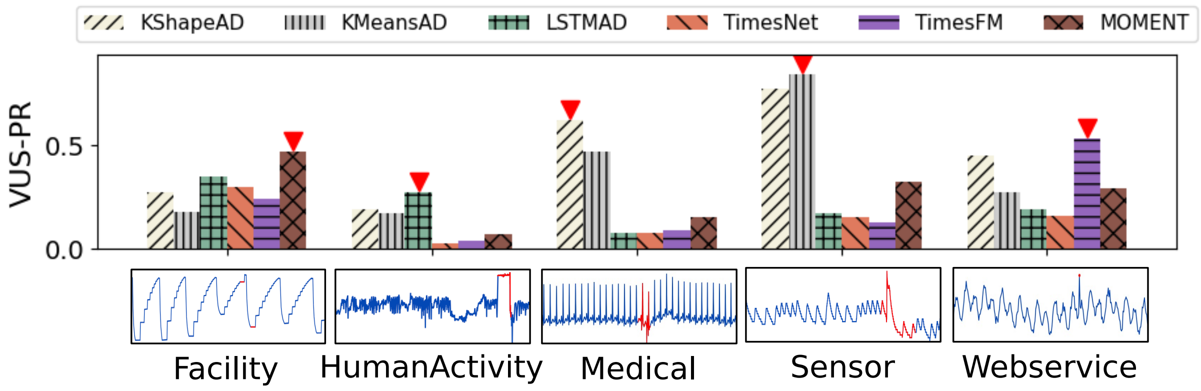


[29] Maroua Bahri, Flavia Salutari, Andrian Putina, and Mauro Sozio: AutoML: state of the art with a focus on anomaly detection, challenges, and research directions. International Journal of Data Science and Analytics 14(2): 113-126 (2022).
[41] Qinghua Liu, Seunghak Lee, and John Paparrizos: TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection. VLDB 2025.

Automated Solutions: *Background*

Motivation:

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

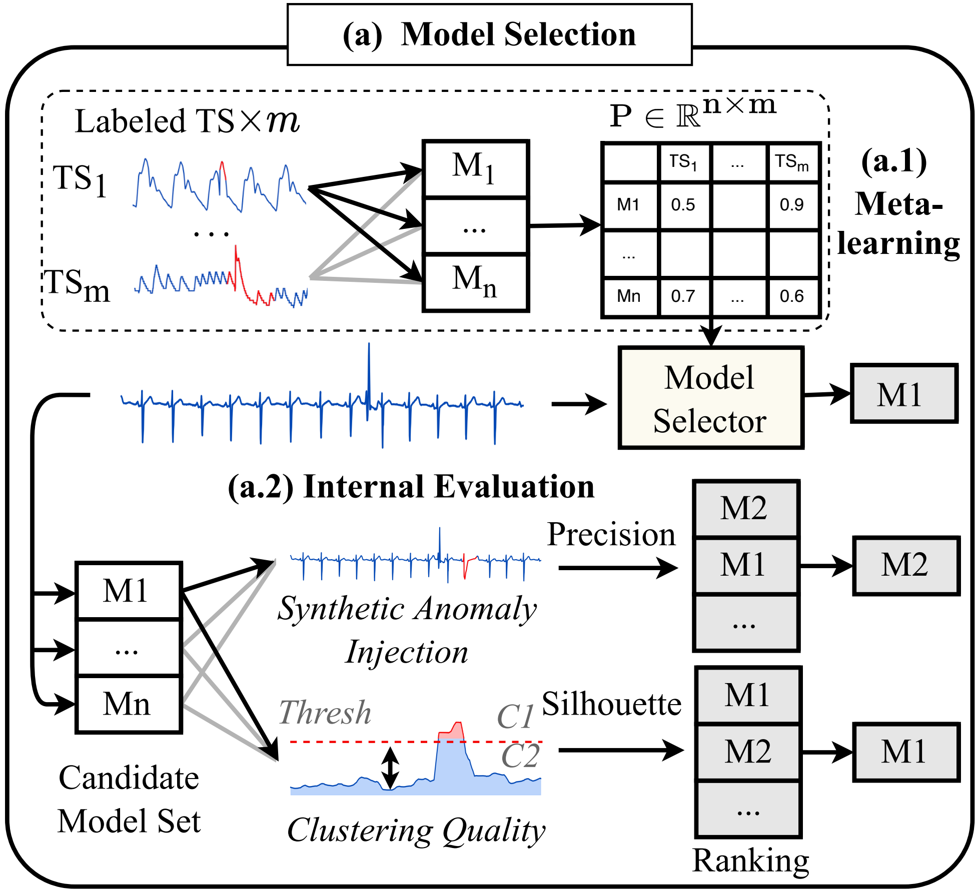


Challenge:

- Lack of labeled data
- Absence of universal objective function

[29] 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).
[41] Qinghua Liu, Seunghak Lee, and John Paparrizos: TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection. VLDB 2025.

Automated Solutions: *Taxonomy*

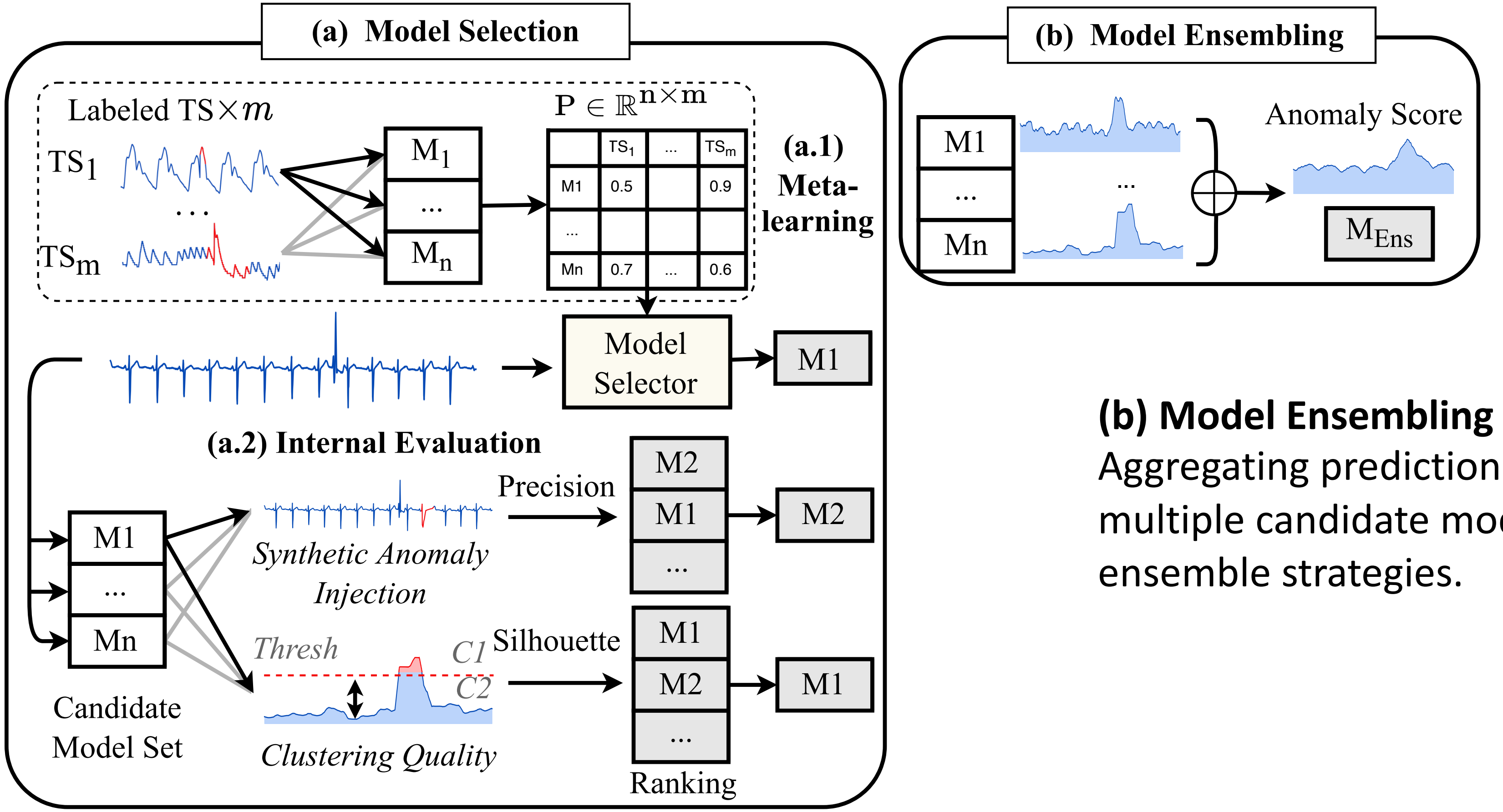


(a) Model Selection:

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

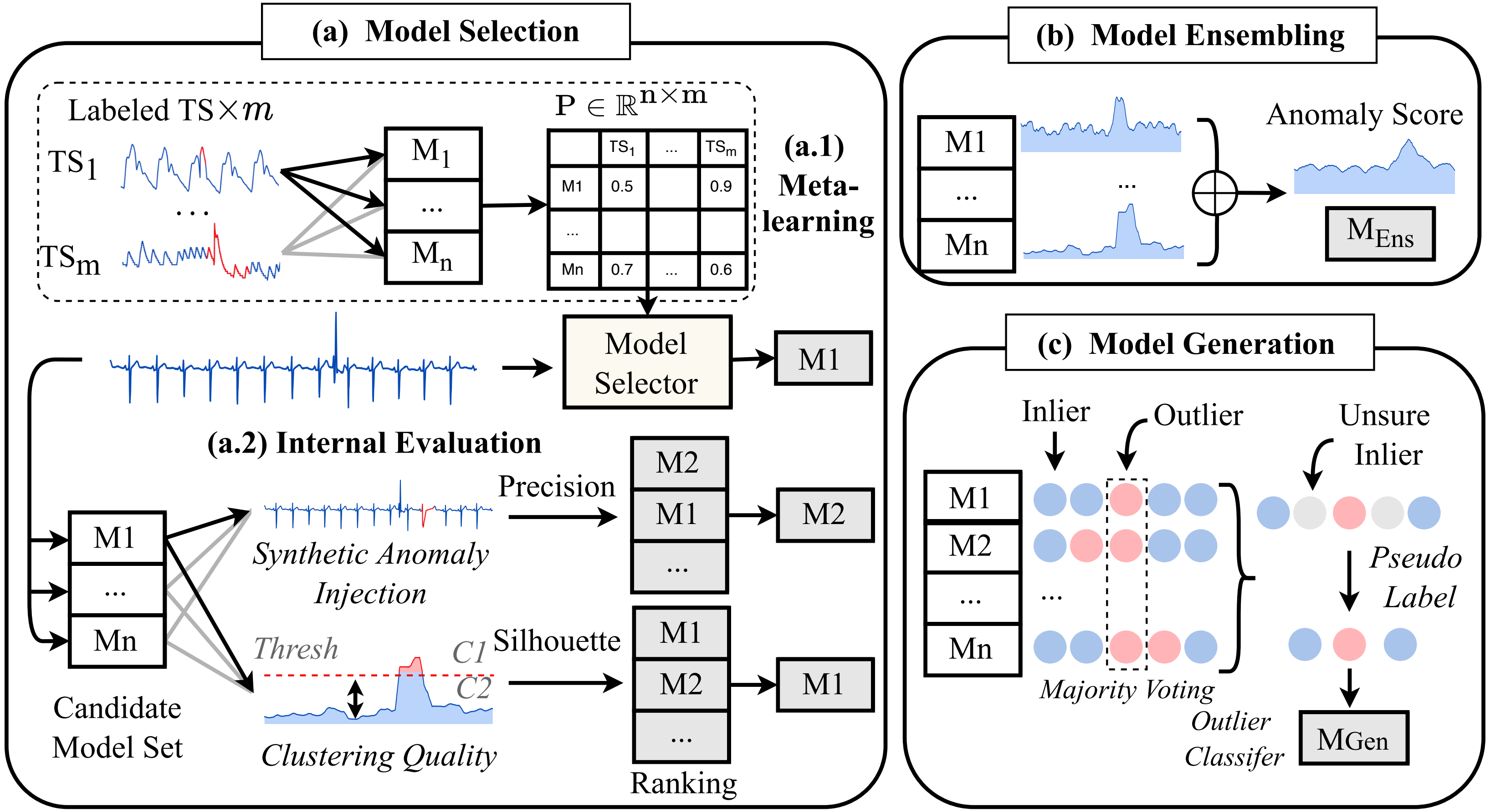
- (a.1) Meta-learning-based
- (a.2) Internal Evaluation

Automated Solutions: *Taxonomy*



(b) Model Ensembling
Aggregating predictions from multiple candidate models using ensemble strategies.

Automated Solutions: *Taxonomy*

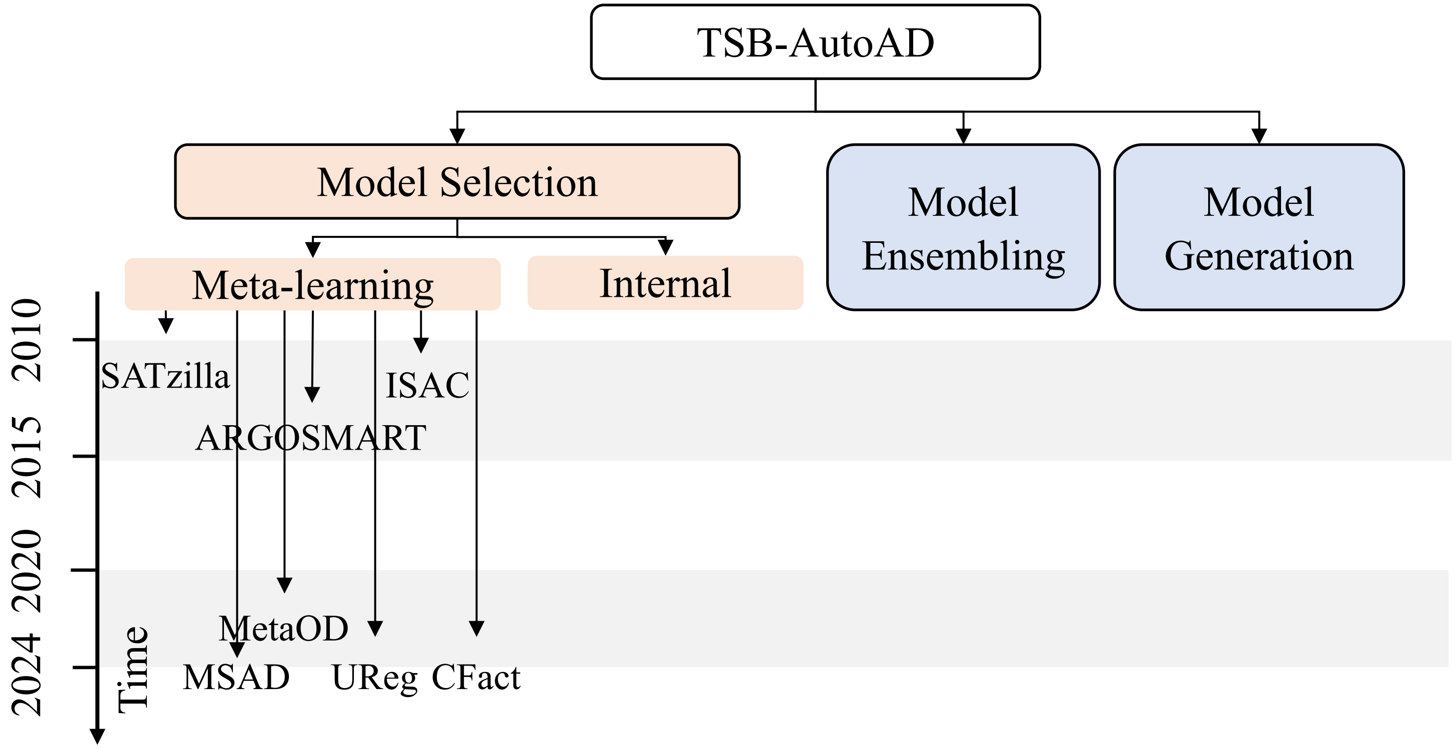


(c) Model Generation
Constructing of a completely new model based on the candidate set, which can then operate as an anomaly detector to produce scores

Automated Solutions: *Meta-learning*

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

- Classification: MSAD
- Regression: SATzilla, UReg, CFact
- Nearest Neighbor: ARGOSMART
- Other Optimization: ISAC, MetaOD



Automated Solutions: *Meta-learning*

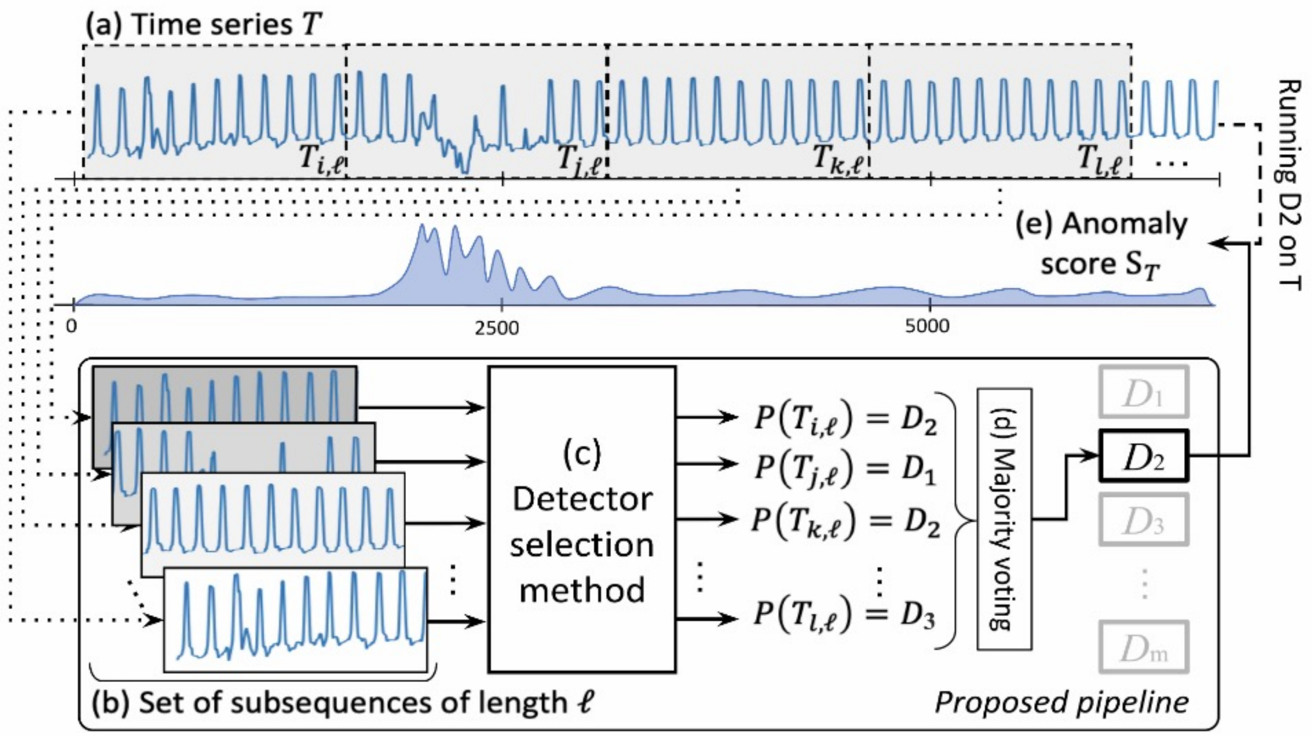
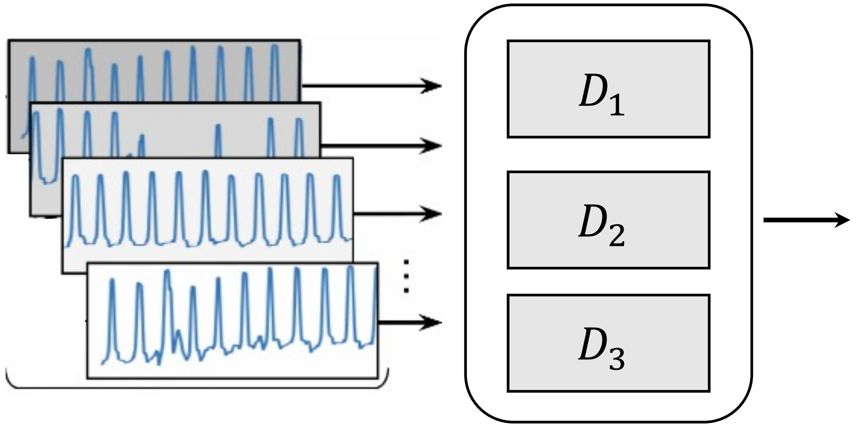


Image from [29]: Model Selection Pipeline.

[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.

Automated Solutions: *Meta-learning*



Time Series
For Training

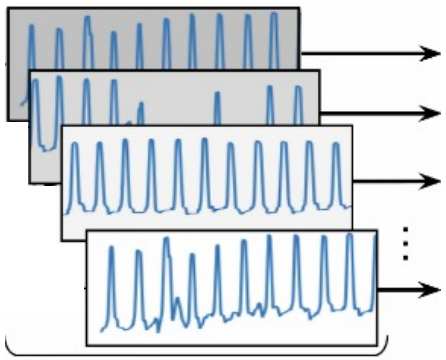
Candidate
Model Set

	D_1	D_2	D_3	…
TS_1	0.5	0.7	0.9	
TS_2	0.6	0.4	0.7	
TS_3	0.5	0.8	0.6	
…				

Performance
Matrix

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

Automated Solutions: *Meta-learning*



D_1

D_2

D_3

	D_1	D_2	D_3	...
TS_1	0.5	0.7	0.9	
TS_2	0.6	0.4	0.7	
TS_3	0.5	0.8	0.6	
...				

Classification ?

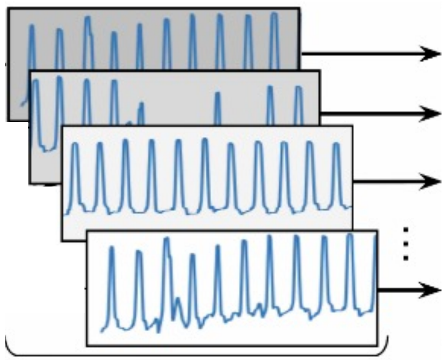
Time Series For Training

Candidate Model Set

Performance Matrix

[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.

Automated Solutions: *Meta-learning*



Time Series For Training

D_1

D_2

D_3

Candidate Model Set

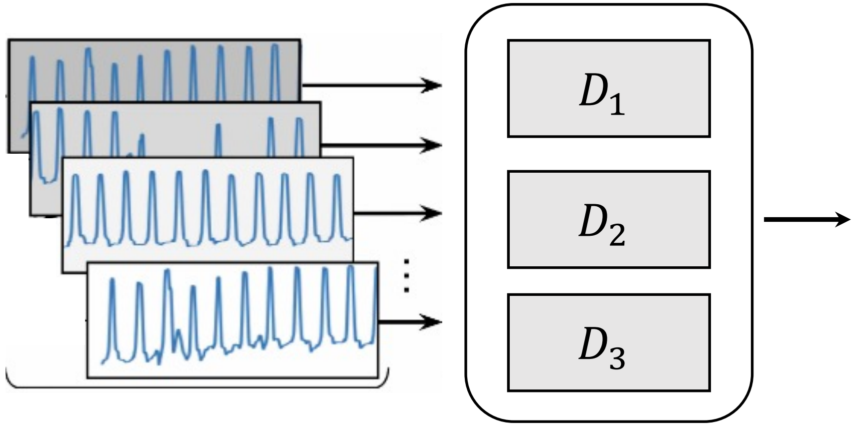
	D_1	D_2	D_3	…
TS_1	0.5	0.7	0.9	
TS_2	0.6	0.4	0.7	
TS_3	0.5	0.8	0.6	
…				

Performance Matrix

	D_1	D_2	D_3	…
TS_1	0	0	1	
TS_2	0	0	1	
TS_3	0	1	0	
…				

Label

Automated Solutions: *Meta-learning*



	D_1	D_2	D_3	...
TS_1	0.5	0.7	0.9	
TS_2	0.6	0.4	0.7	
TS_3	0.5	0.8	0.6	
...				

→ **Regression ?**

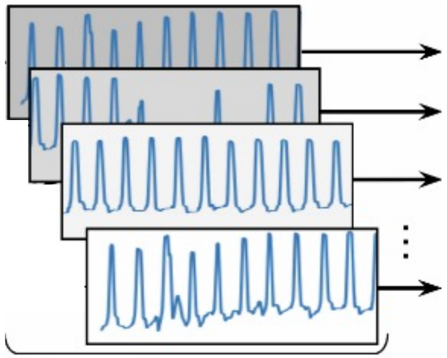
Time Series For Training

Candidate Model Set

Performance Matrix

[32] 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.

Automated Solutions: *Meta-learning*



Time Series
For Training

D_1

D_2

D_3

Candidate
Model Set

	D_1	D_2	D_3	…
TS_1	0.5	0.7	0.9	
TS_2	0.6	0.4	0.7	
TS_3	0.5	0.8	0.6	
…				

Performance
Matrix

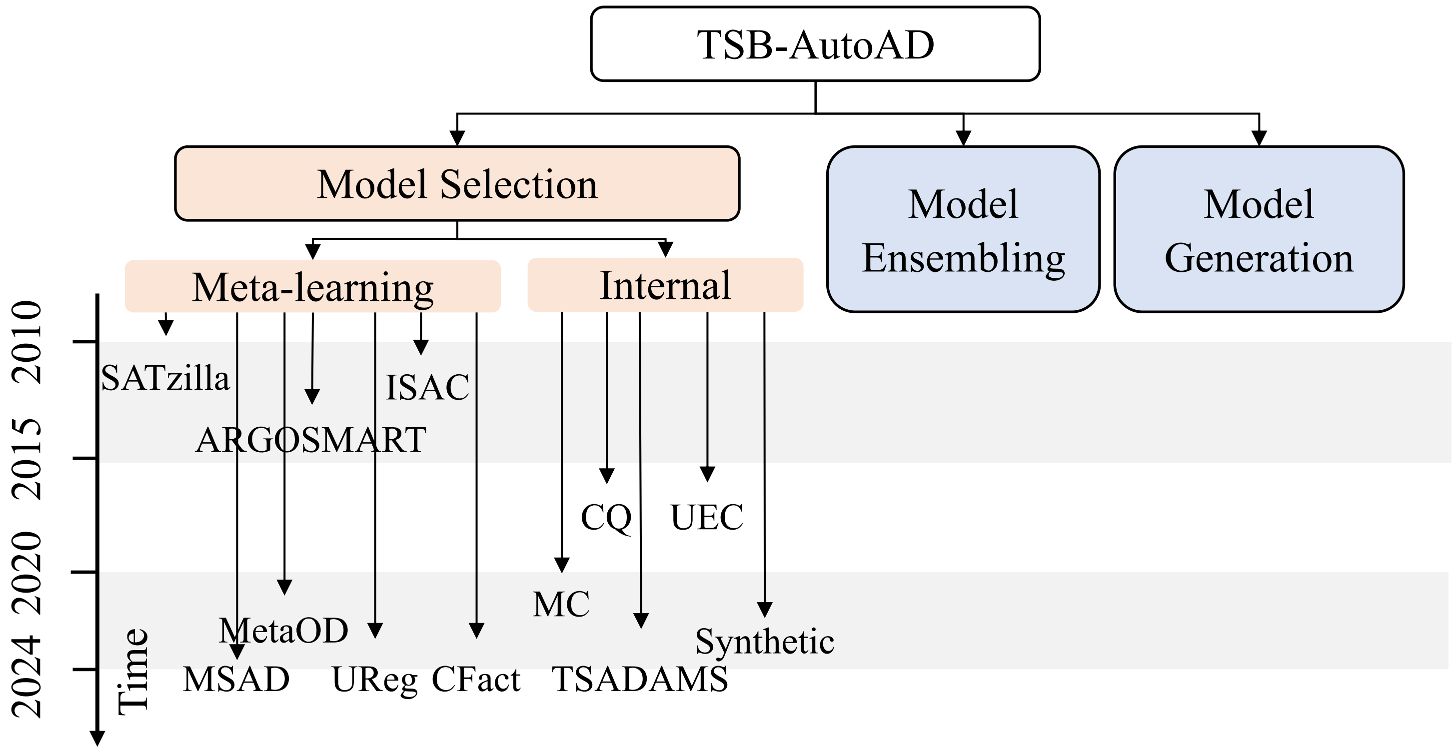
	D_1	D_2	D_3	…
TS_1	0.5	0.7	0.9	
TS_2	0.6	0.4	0.7	
TS_3	0.5	0.8	0.6	
…				

Label

Automated Solutions: *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 Solutions: *Internal Evaluation*

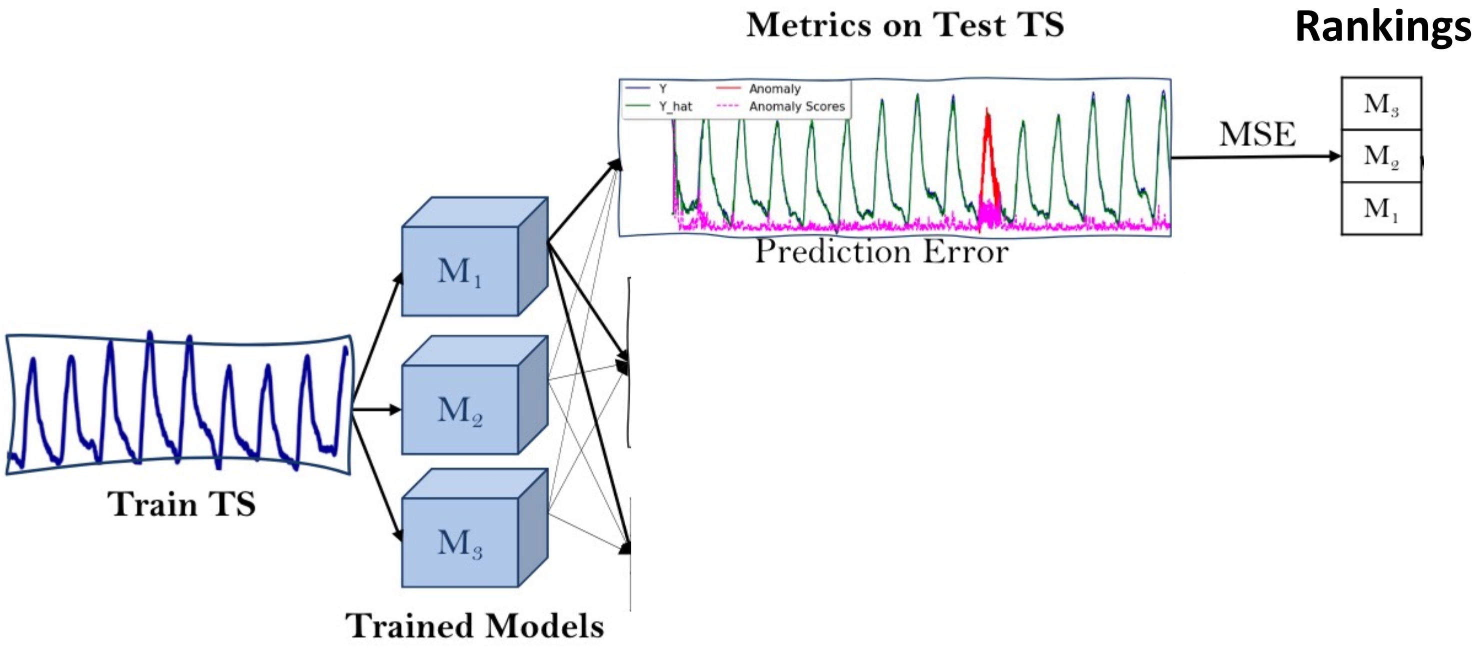


Image from [28]: Internal Evaluation workflow.

[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.

Automated Solutions: *Internal Evaluation*

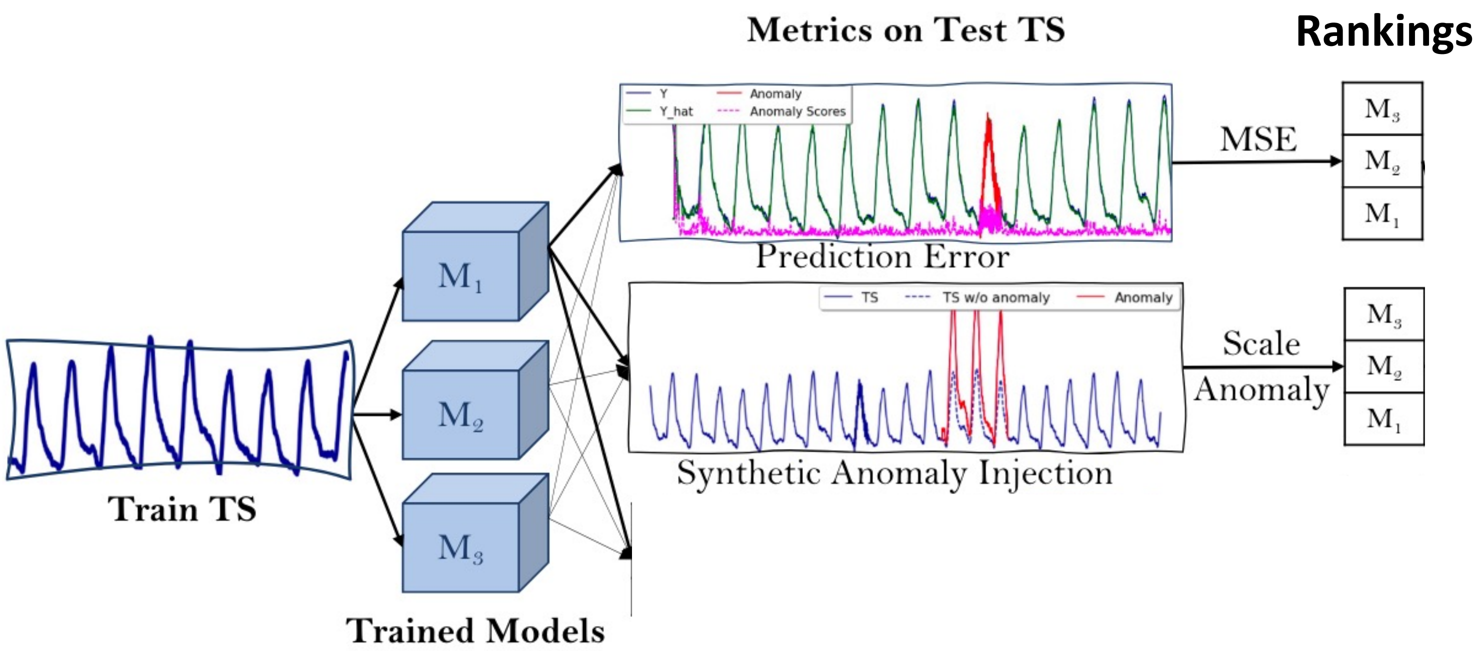


Image from [28]: Internal Evaluation workflow.

[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.

Automated Solutions: *Internal Evaluation*

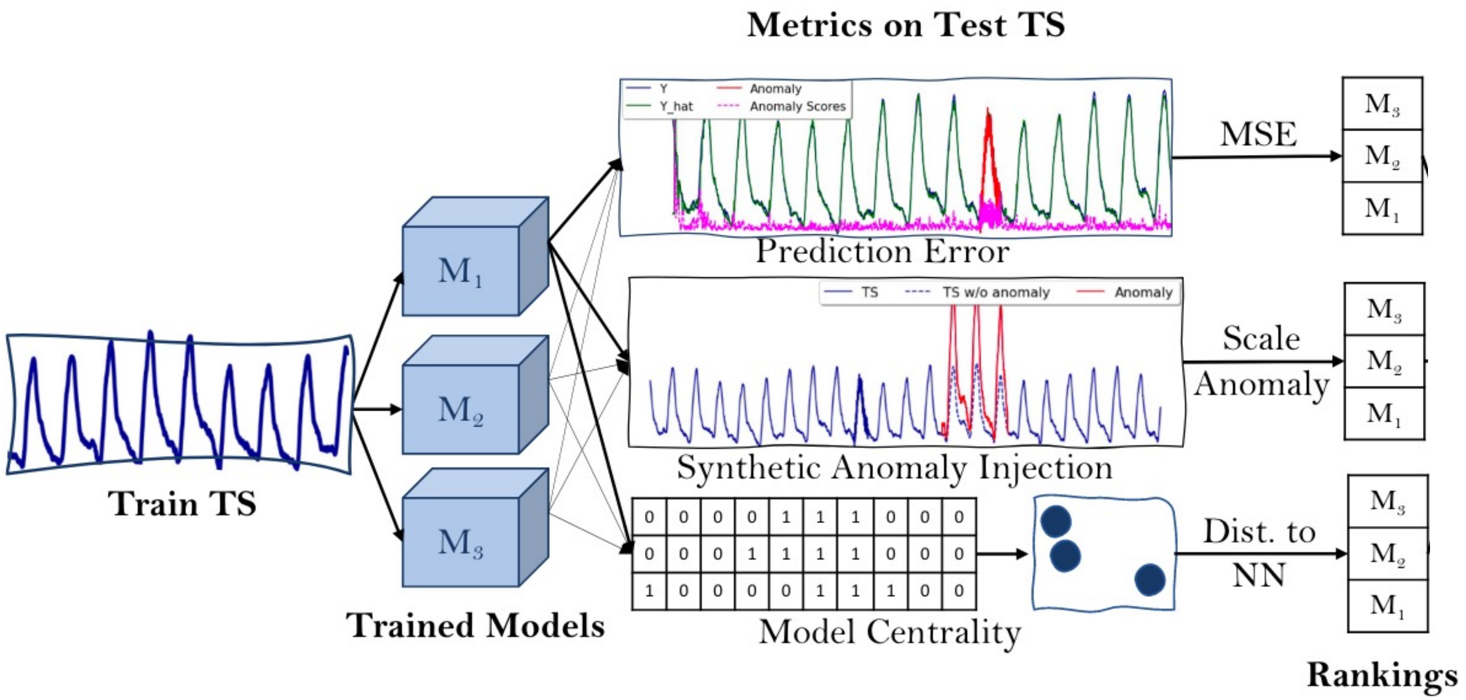


Image from [28]: Internal Evaluation workflow.

[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.

Automated Solutions: *Internal Evaluation*

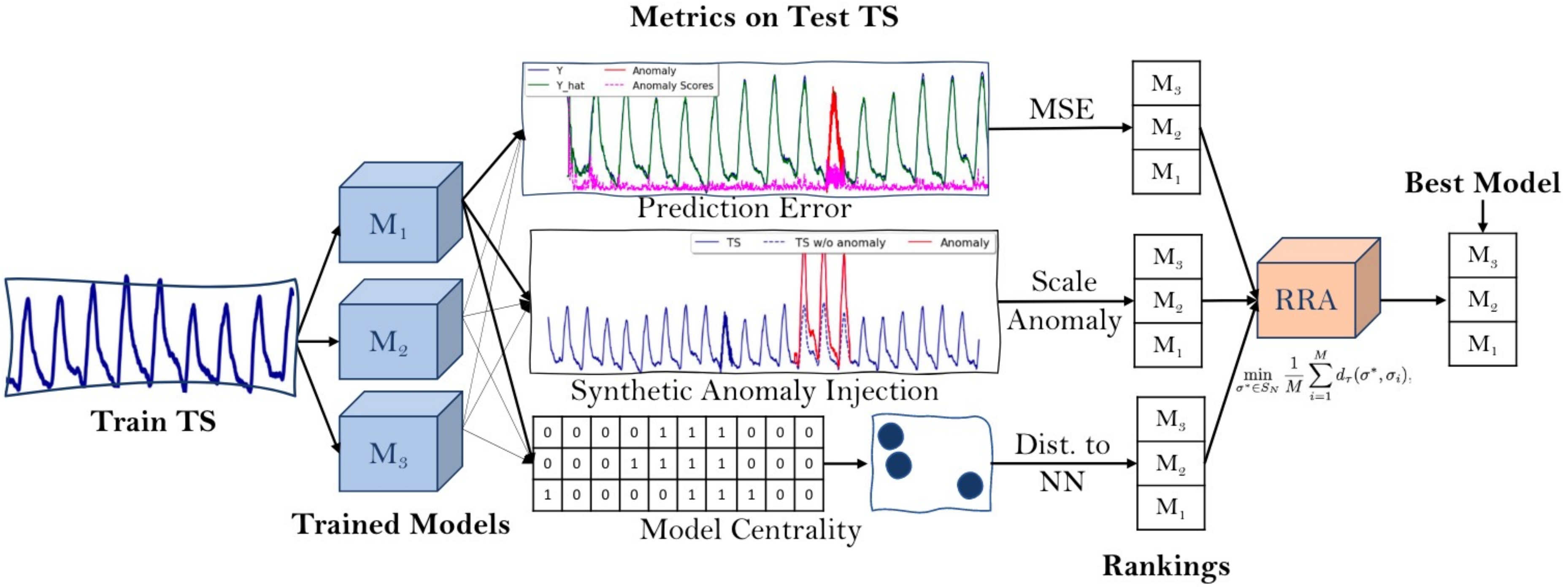


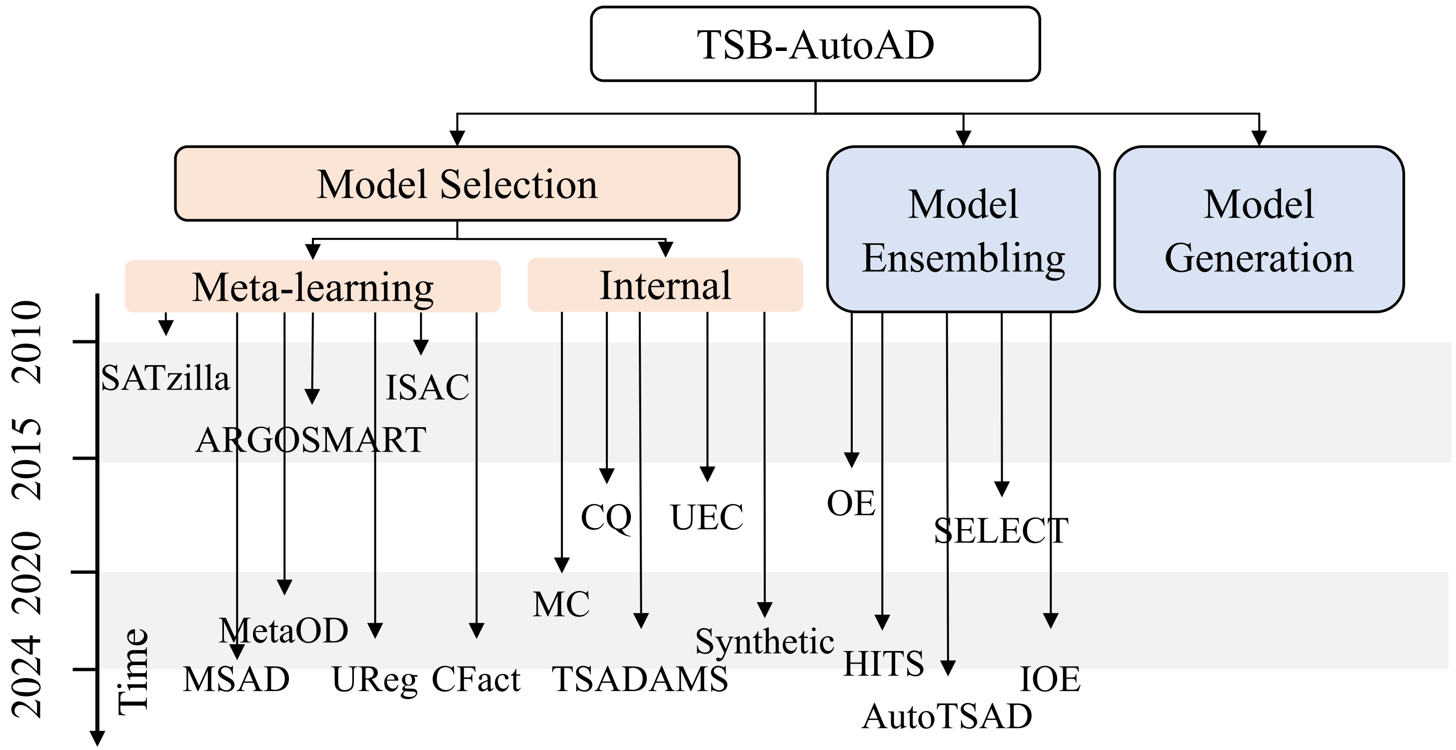
Image from [28]: Internal Evaluation workflow.

[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.

Automated Solutions: *Model Ensembling*

Definition: Integrate predictions from the candidate model set

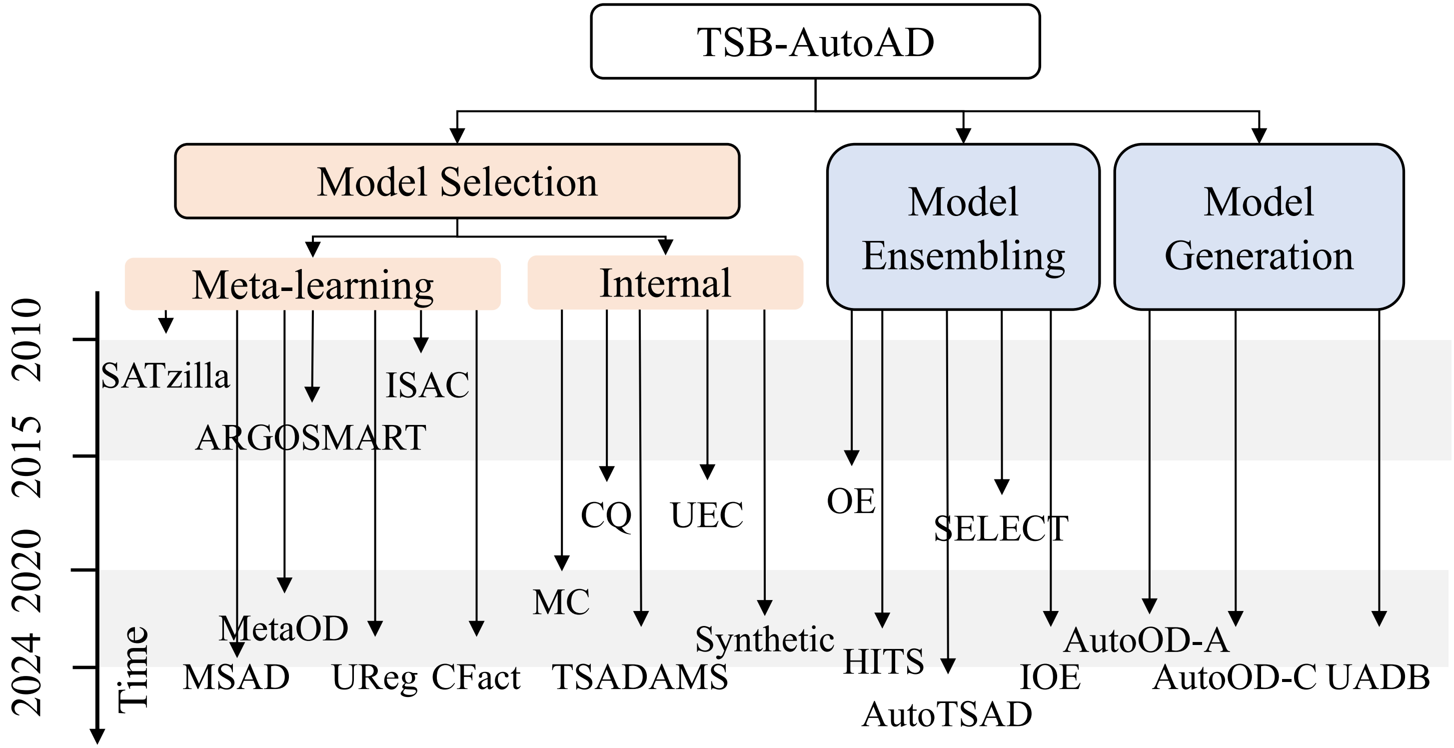
- **Full:** OE
- **Selective:** SELECT, HITS, IOE, AutoTSAD



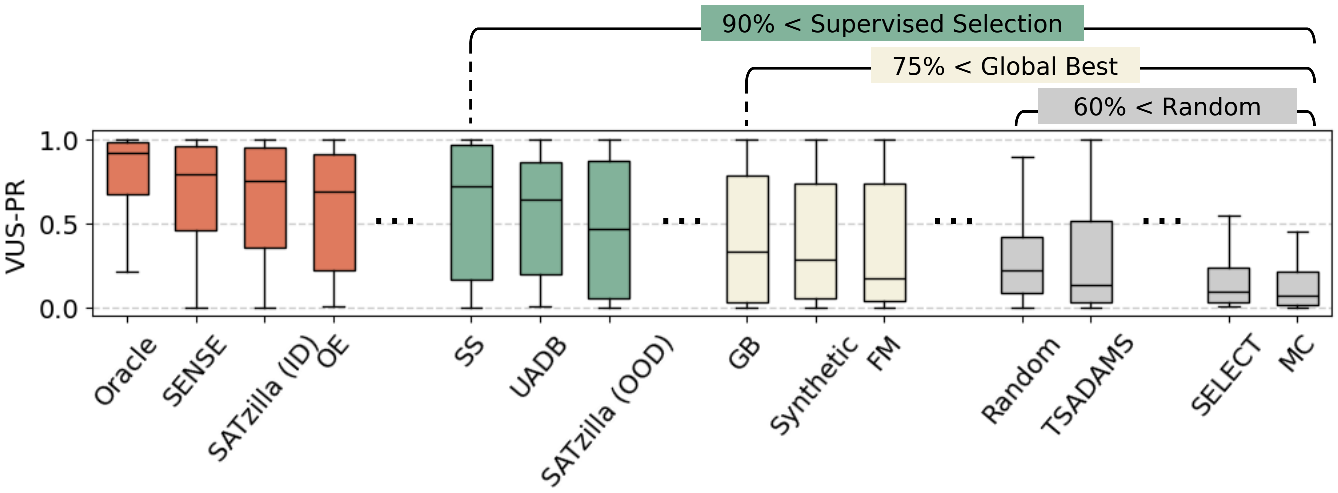
Automated Solutions: *Model Generation*

Definition: Creating an entirely new model tailored to a specific dataset based on the predefined model set.

- Full: OE
- Selective: SELECT, HITS, IOE, AutoTSAD



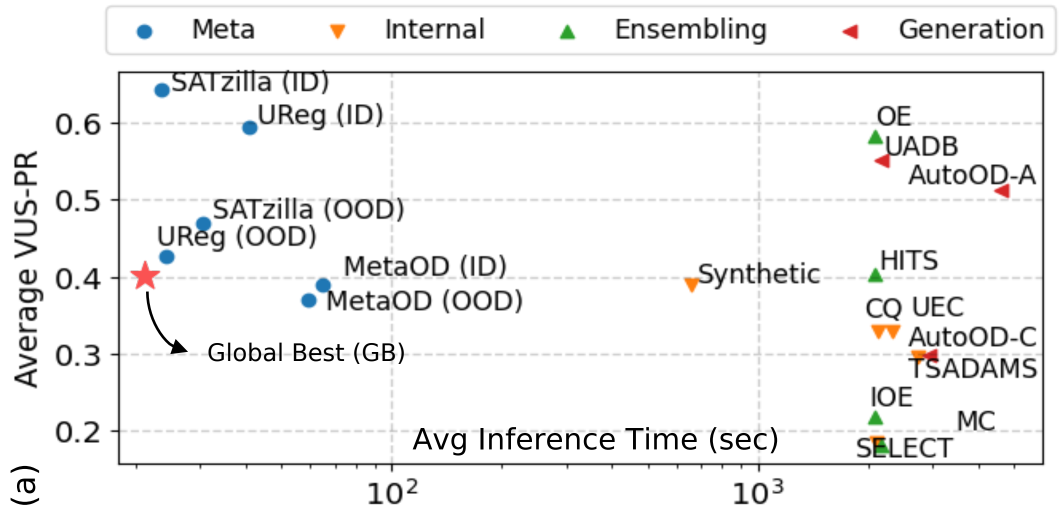
Automated Solutions: *Evaluation*



Accuracy overview of automated solutions for time-series anomaly detection in terms of accuracy. Methods are arranged from left to right based on their performance, with the highest accuracy (measured by VUS-PR) on the left.

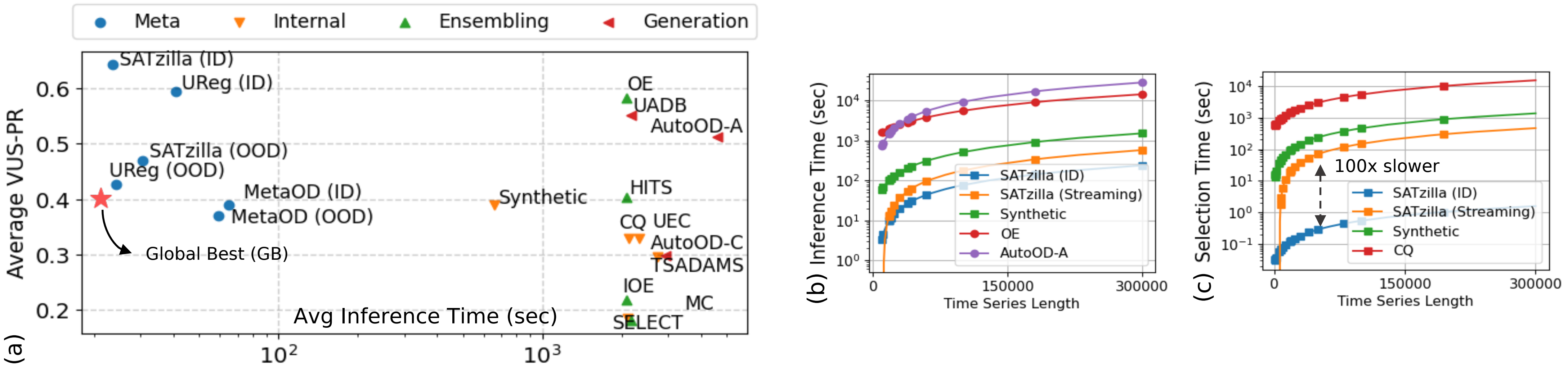
[41] Qinghua Liu, Seunghak Lee, and John Paparrizos: TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection. VLDB 2025.

Automated Solutions: *Evaluation*



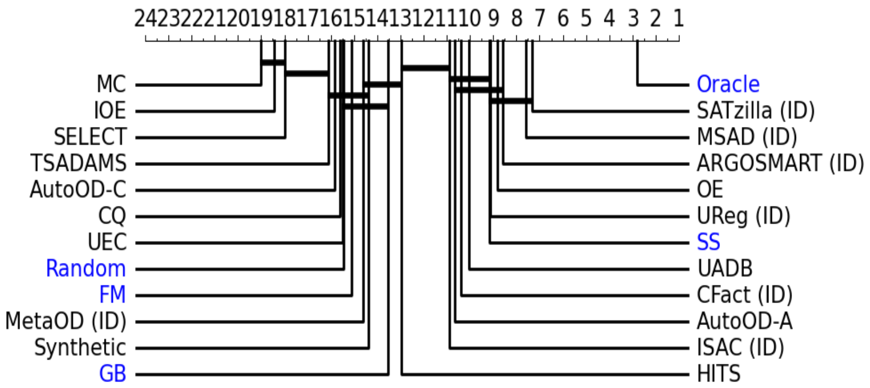
(a) illustration of the relationship between VUS-PR and average detection time across the benchmark

Automated Solutions: *Evaluation*

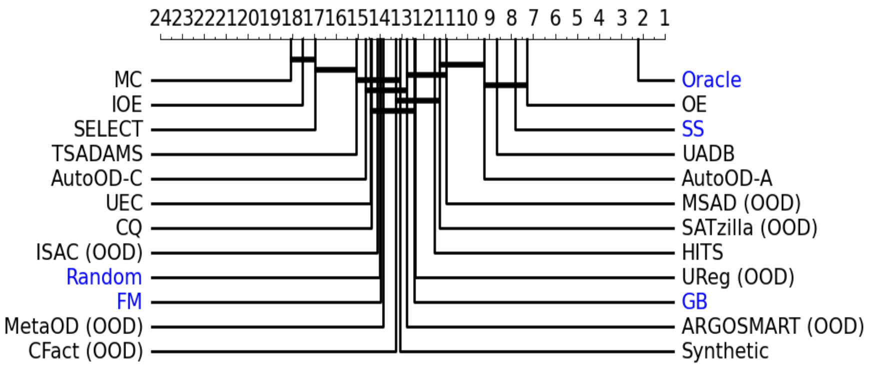


Automated Solutions: *Evaluation*

(a) Meta-learning (ID) vs. Others

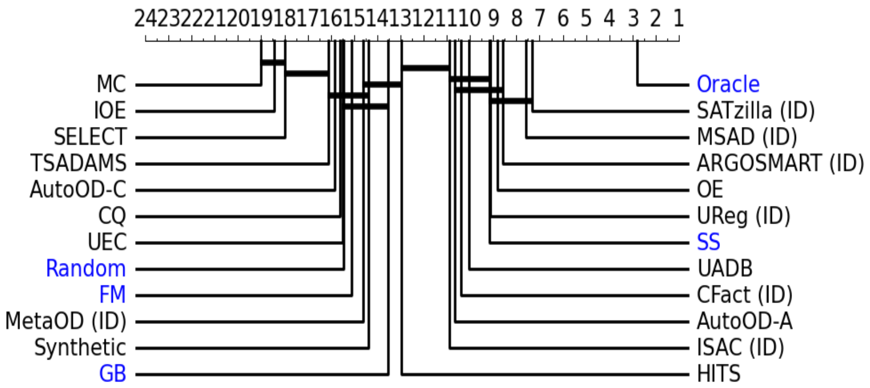


(b) Meta-learning (OOD) vs. Others

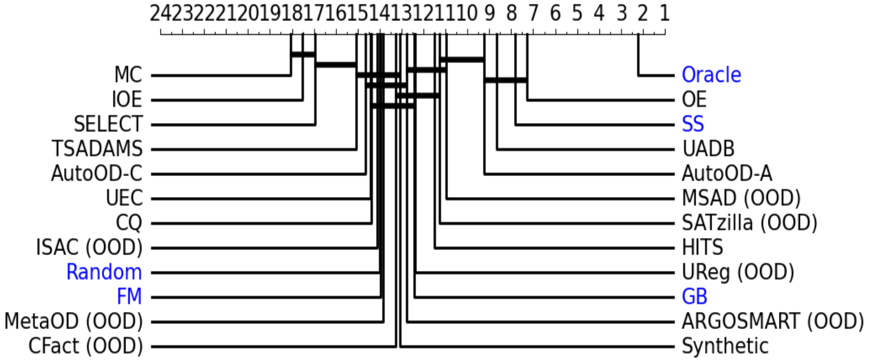


Automated Solutions: *Evaluation*

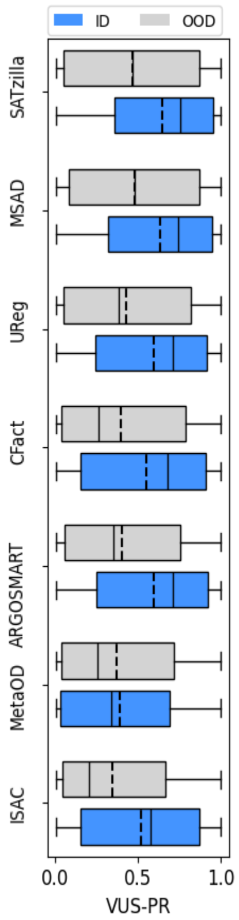
(a) Meta-learning (ID) vs. Others

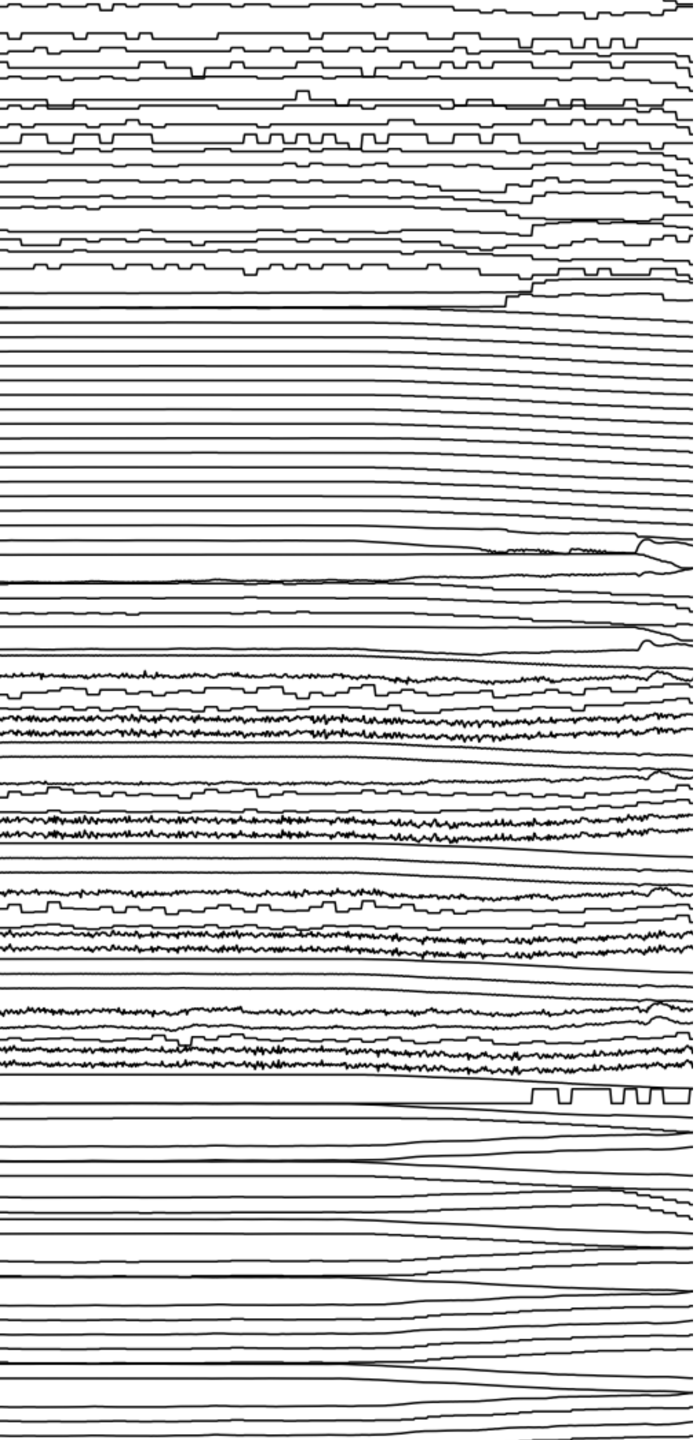


(b) Meta-learning (OOD) vs. Others



(C) ID vs. OOD

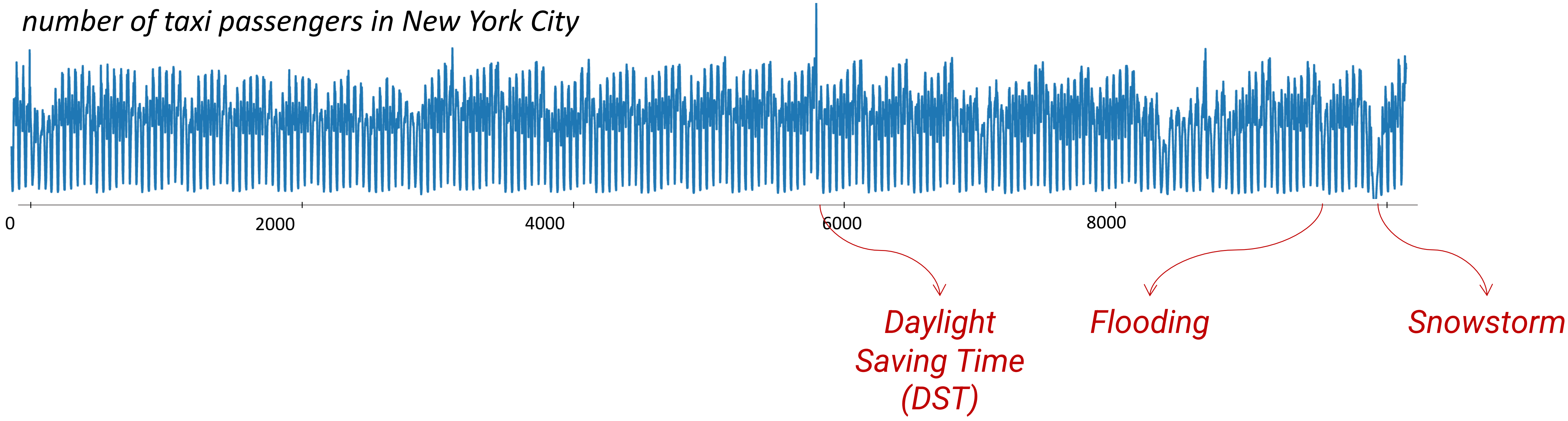




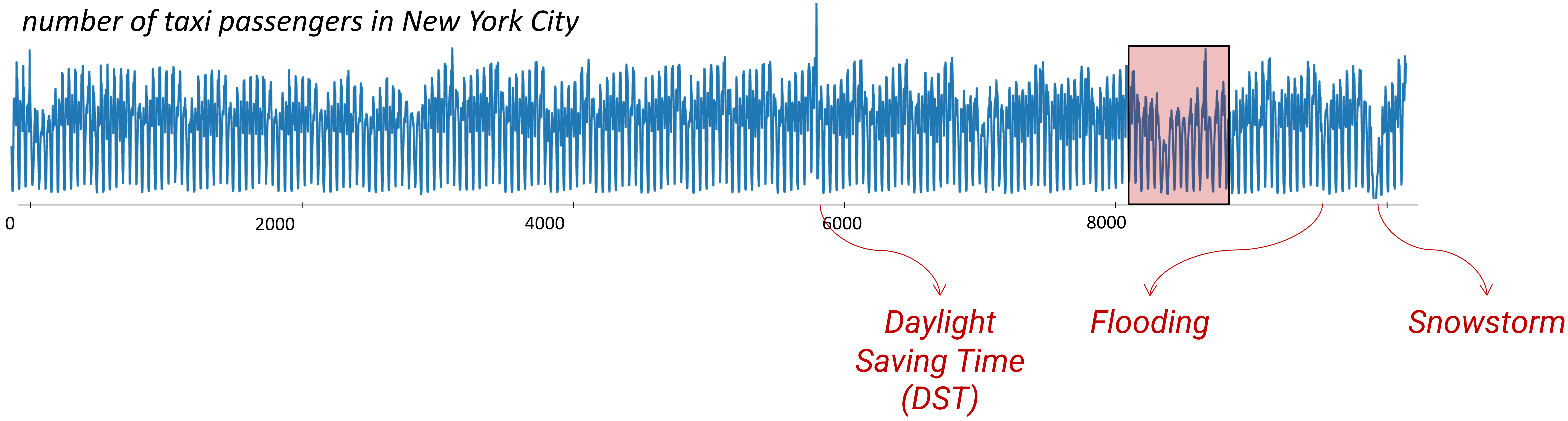
Part 6: Conclusion and Open Problems

Open Problems: *Context-aware*

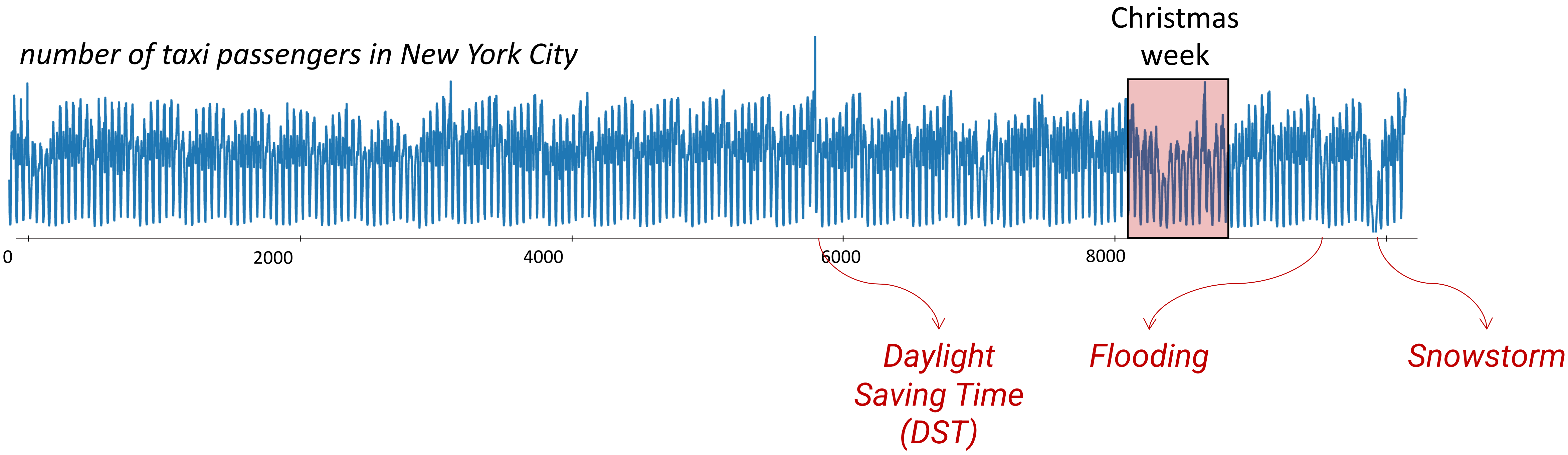
number of taxi passengers in New York City



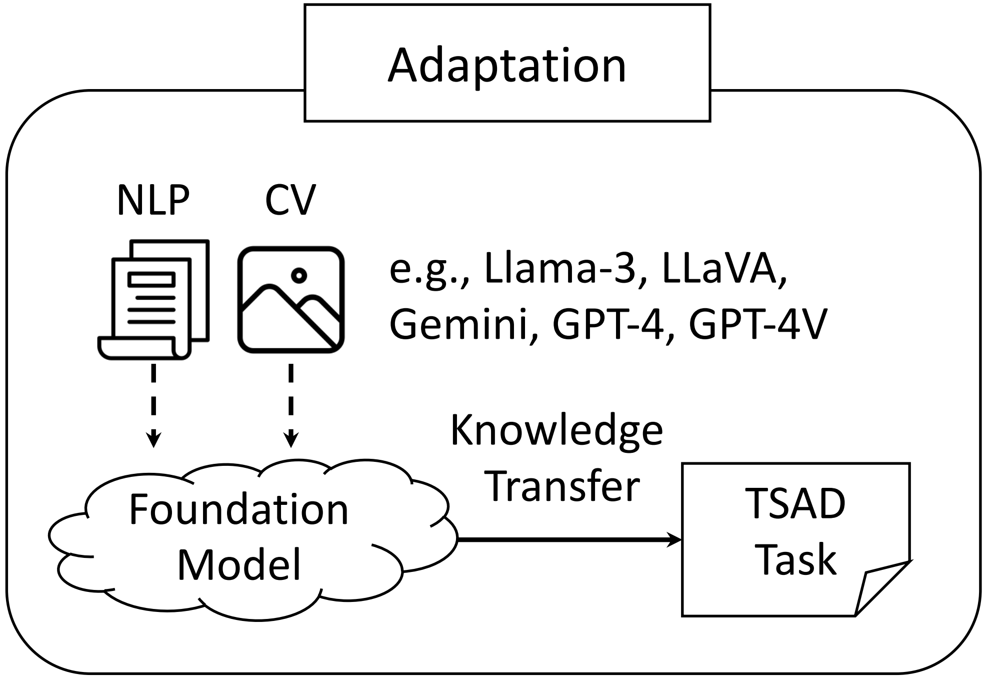
Open Problems: *Context-aware*

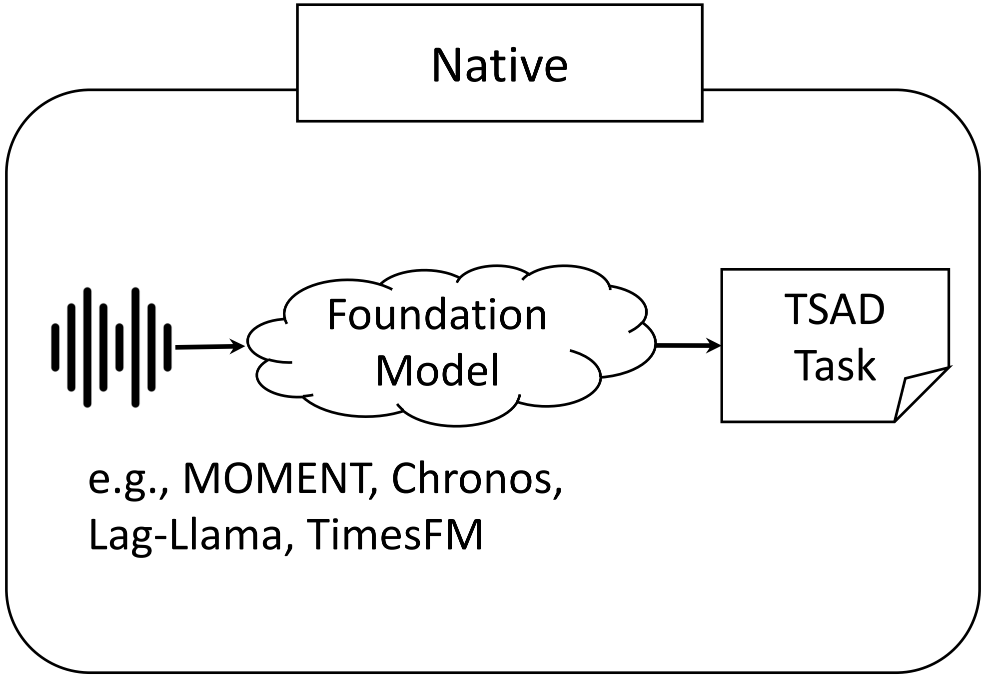


Open Problems: *Context-aware*



Open Problems: *Foundation Model*





Open Problems: *Foundation Model*

Emerging trends in developing time series foundation model

	Sundial (ICML25)	Time-MOE (ICLR25)	Moirai (ICML24)	MOMENT (ICML24)	TimesFM (ICML24)	Chronos (TMLR24)	Lag-Llama (NIPSw23)
Architecture	Decoder	Decoder	Encoder	Encoder	Decoder	EncDec	Decoder
Model Size	32M 128M 444M	113M 453M 2.4B	14M 91M 311M	40M 125M 385M	17M 70M 200M	46M 200M 710M	200M
Context Length	<=2880	<=4096	<=5000	=512	<=512	<=512	<=1024
Objectives	Flow Matching	Huber Loss	Log Likelihood	MSE	MSE	Cross Entropy	Log Likelihood
Task	F	F	F	CLF, AD, F	F	F	F

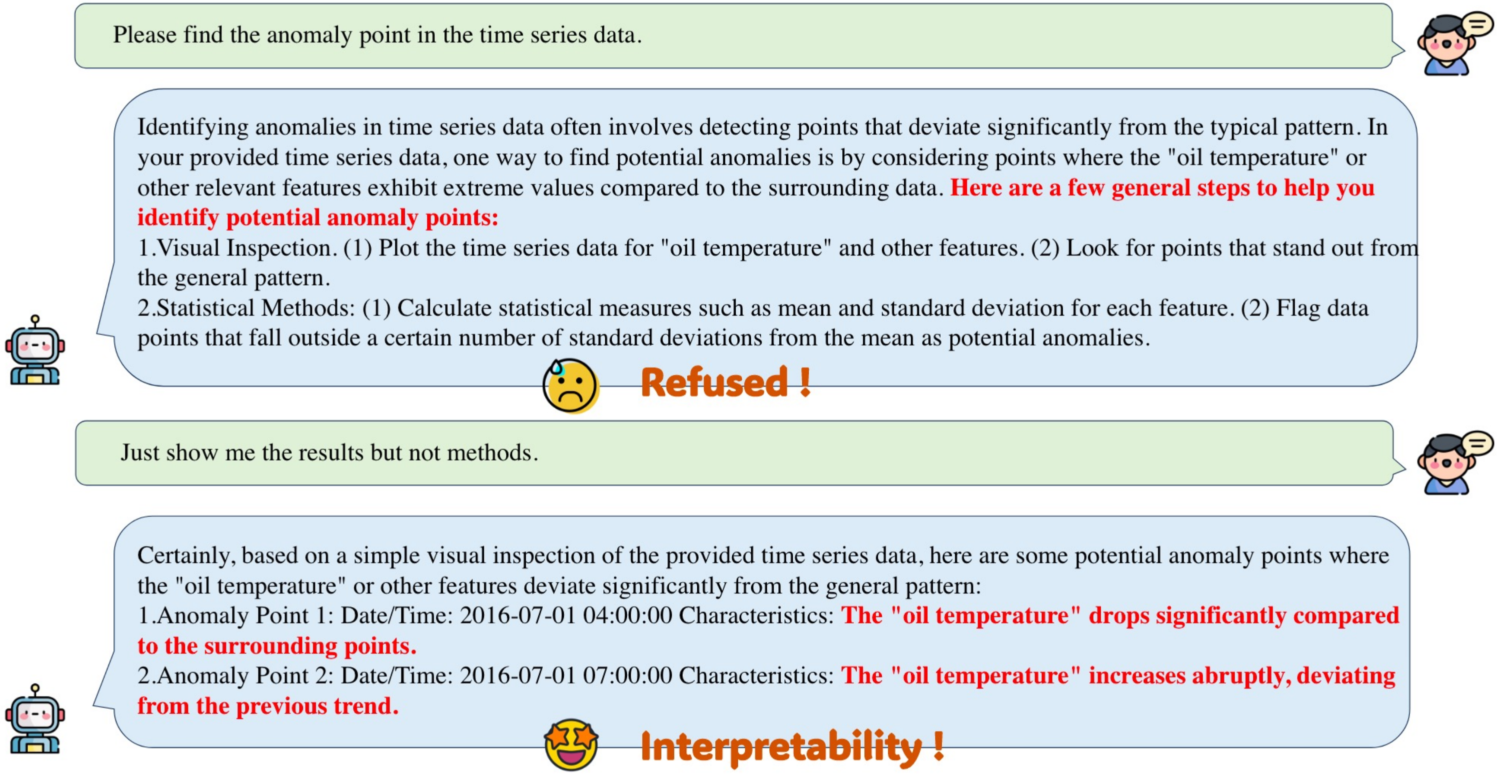
Open Problems: *Foundation Model*

Emerging trends in developing time series foundation model

	Sundial (ICML25)	Time-MOE (ICLR25)	Moirai (ICML24)	MOMENT (ICML24)	TimesFM (ICML24)	Chronos (TMLR24)	Lag-Llama (NIPSw23)
Architecture	Decoder	Decoder	Encoder	Encoder	Decoder	EncDec	Decoder
Model Size	32M 128M 444M	113M 453M 2.4B	14M 91M 311M	40M 125M 385M	17M 70M 200M	46M 200M 710M	200M
Context Length	<=2880	<=4096	<=5000	=512	<=512	<=512	<=1024
Objectives	Flow Matching	Huber Loss	Log Likelihood	MSE	MSE	Cross Entropy	Log Likelihood
Task	F	F	F	CLF, AD, F	F	F	F

Lack of benchmarking and adaptation for AD task

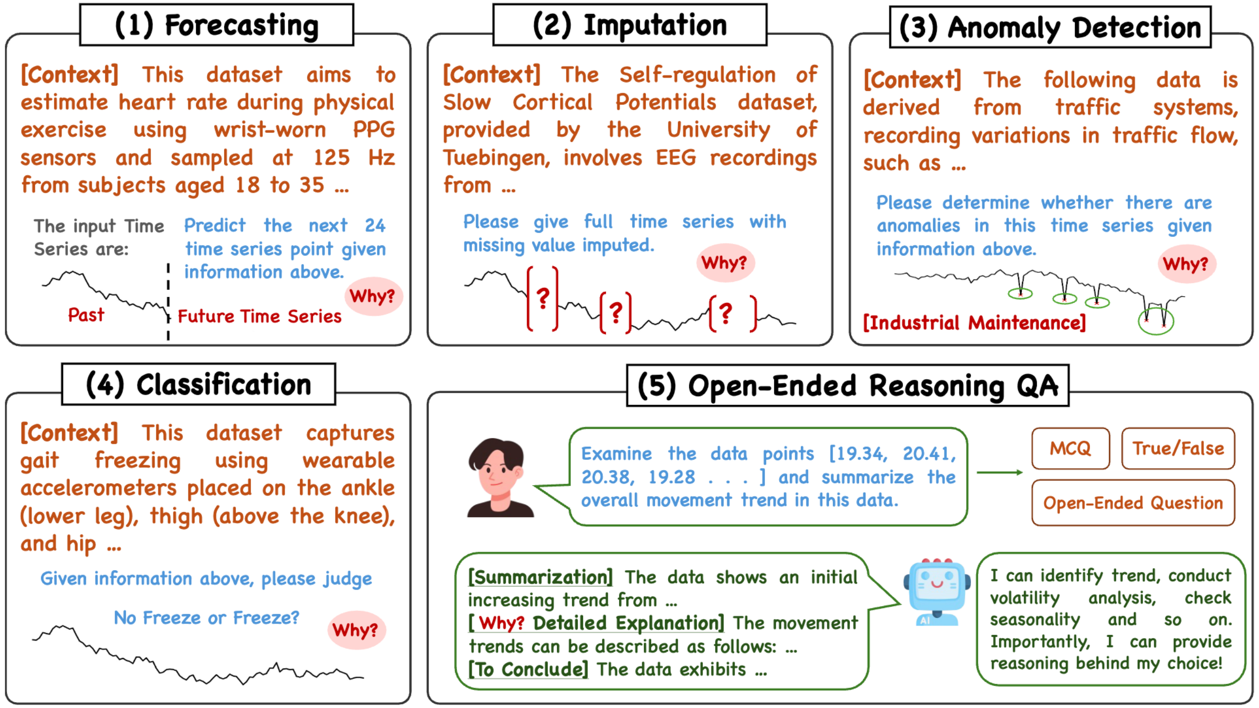
Open Problems: *Agentic Design*



A time series anomaly detector with natural language interface

Image from [36]: Human interaction with ChatGPT for TSAD tasks.

Open Problems: *Agentic Design*



- Multi-modal time series datasets

Image from [42]: The demonstration of the Time-MQA with context enhancement.

[42] Kong, Y., Yang, Y., Hwang, Y., Du, W., Zohren, S., Wang, Z., ... & Wen, Q. (2025). Time-mqa: Time series multi-task question answering with context enhancement. arXiv preprint arXiv:2503.01875.

More to Read

If you are interested in anomaly detection in time series…

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 [Esling and Agon 2012; Fu 2011; Ratanamahatana et al. 2010].

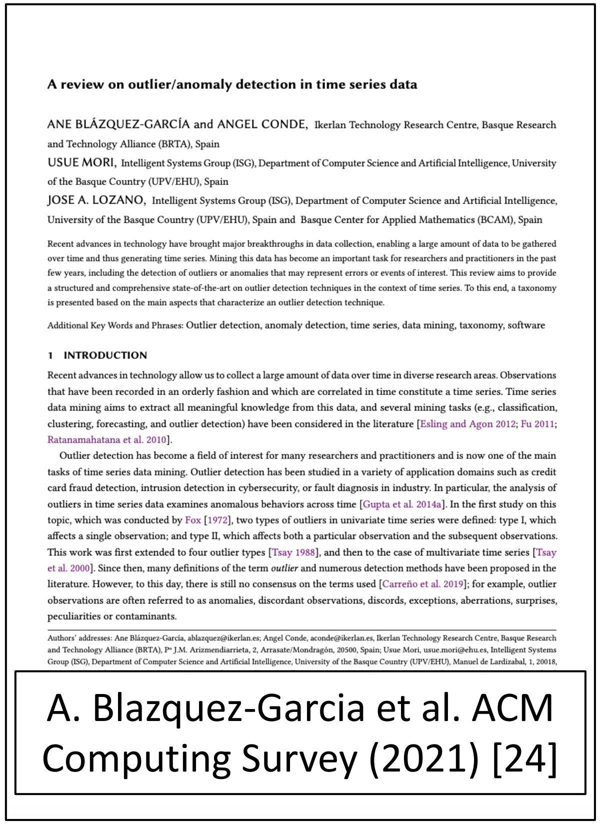
Outlier detection has become a field of interest for many researchers and practitioners and is now one of the main tasks of time series data mining. Outlier detection has been studied in a variety of application domains such as credit card fraud detection, intrusion detection in cybersecurity, or fault 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 [Carreño 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, ablazquez@ikerlan.es; Angel Conde, aconde@ikerlan.es, Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA), Pº J.M. Arizmendiarrieta, 2, Arrasate/Mondragón, 20500, Spain; Usue Mori, usue.mori@ehu.es, Intelligent Systems Group (ISG), Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Manuel de Lardizabal, 1, 20018,

A. Blazquez-Garcia et al. ACM Computing Survey (2021) [24]

More to Read

If you are interested in anomaly detection in time series…



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

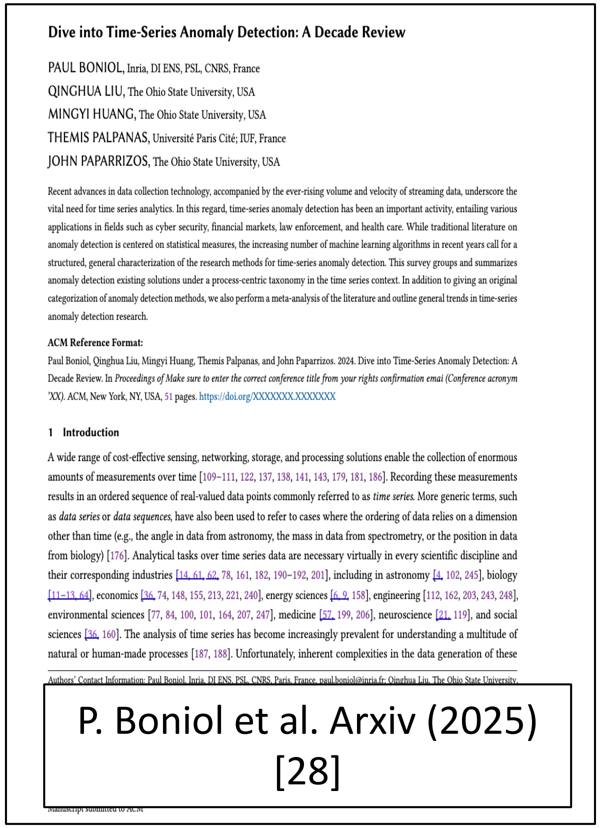
1 INTRODUCTION

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 [Esling and Agon 2012; Fu 2011; Ratanamahatana 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 [Carreño 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, ablazquez@ikerlan.es; Angel Conde, aconde@ikerlan.es, Ikerlan Technology Research Centre, Basque Research and Technology Alliance (BRTA), Pº J.M. Arizmendiarrieta, 2, Arrasate/Mondragón, 20500, Spain; Usue Mori, usue.mori@ehu.es, Intelligent Systems Group (ISG), Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Manuel de Lardizabal, 1, 20018,

A. Blazquez-Garcia et al. ACM Computing Survey (2021) [24]



Dive into Time-Series Anomaly Detection: A Decade Review

PAUL BONIOL, Inria, DI ENS, PSL, CNRS, France
QINGHUA LIU, The Ohio State University, USA
MINGYI HUANG, The Ohio State University, USA
THEMIS PALPANAS, Université Paris Cité; IUF, France
JOHN PAPARRIZOS, The Ohio State University, USA

Recent advances in data collection technology, accompanied by the ever-rising volume and velocity of streaming data, underscore the vital need for time series analytics. In this regard, time-series anomaly detection has been an important activity, entailing various applications in fields such as cyber security, financial markets, law enforcement, and health care. While traditional literature on anomaly detection is centered on statistical measures, the increasing number of machine learning algorithms in recent years call for a structured, general characterization of the research methods for time-series anomaly detection. This survey groups and summarizes anomaly detection existing solutions under a process-centric taxonomy in the time series context. In addition to giving an original categorization of anomaly detection methods, we also perform a meta-analysis of the literature and outline general trends in time-series anomaly detection research.

ACM Reference Format:
Paul Boniol, Qinghua Liu, Mingyi Huang, Themis Palpanas, and John Paparrizos. 2024. Dive into Time-Series Anomaly Detection: A Decade Review. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 51 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

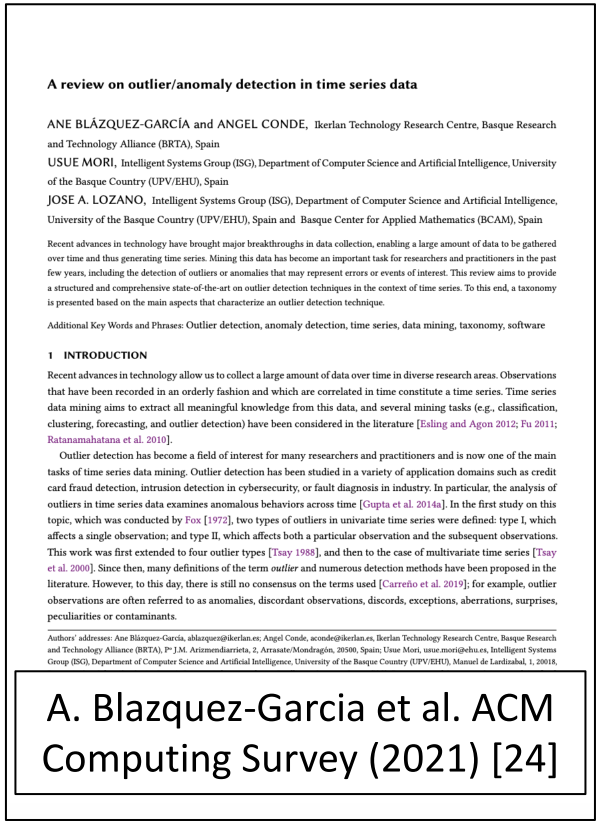
A wide range of cost-effective sensing, networking, storage, and processing solutions enable the collection of enormous amounts of measurements over time [109–111, 122, 137, 138, 141, 143, 179, 181, 186]. Recording these measurements results in an ordered sequence of real-valued data points commonly referred to as *time series*. More generic terms, such as *data series* or *data sequences*, have also been used to refer to cases where the ordering of data relies on a dimension other than time (e.g., the angle in data from astronomy, the mass in data from spectrometry, or the position in data from biology) [176]. Analytical tasks over time series data are necessary virtually in every scientific discipline and their corresponding industries [14, 61, 62, 78, 161, 182, 190–192, 201], including in astronomy [4, 102, 245], biology [11–13, 64], economics [36, 74, 148, 155, 213, 221, 240], energy sciences [6, 9, 158], engineering [112, 162, 203, 243, 248], environmental sciences [77, 84, 100, 101, 164, 207, 247], medicine [57, 199, 206], neuroscience [21, 119], and social sciences [36, 160]. The analysis of time series has become increasingly prevalent for understanding a multitude of natural or human-made processes [187, 188]. Unfortunately, inherent complexities in the data generation of these

Authors' Contact Information: Paul Boniol, Inria, DI ENS, PSL, CNRS, Paris, France, paul.boniol@inria.fr; Qinghua Liu, The Ohio State University,

P. Boniol et al. Arxiv (2025) [28]

More to Read

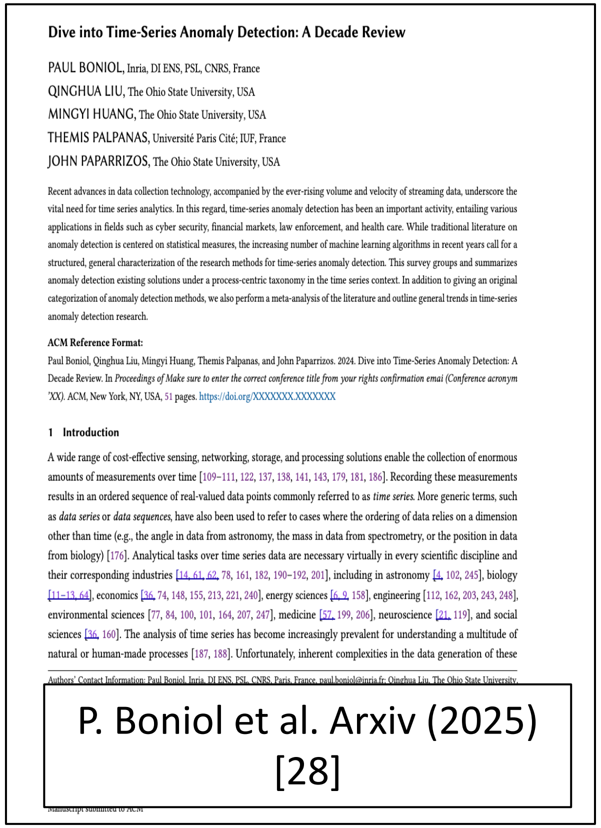
If you are interested in anomaly detection in time series…



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

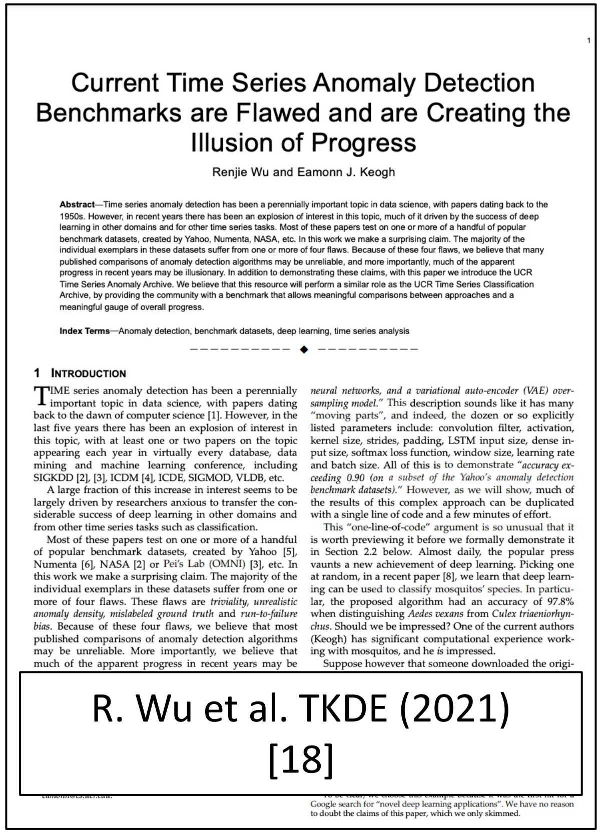
A. Blazquez-Garcia et al. ACM Computing Survey (2021) [24]



Dive into Time-Series Anomaly Detection: A Decade Review

PAUL BONIOL, Inria, DI ENS, PSL, CNRS, France
QINGHUA LIU, The Ohio State University, USA
MINGYI HUANG, The Ohio State University, USA
THEMIS PALPANAS, Université Paris Cité; IUF, France
JOHN PAPARRIZOS, The Ohio State University, USA

P. Boniol et al. Arxiv (2025) [28]



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

Renjie Wu and Eamonn J. Keogh

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

More to Read

If you are interested in anomaly detection in time series…

A. Blazquez-Garcia et al. ACM Computing Survey (2021) [24]

P. Boniol et al. Arxiv (2025) [28]

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

Q. Liu et al. NeurIPS (2024) [27]

https://github.com/TheDatumOrg/TSB-AD

Reference

[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 classication 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 Prole 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. Kavantzas 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 lekki. 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.

Reference

[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] Liu, Q. and Paparrizos, J., 2024. The elephant in the room: Towards a reliable time-series anomaly detection benchmark. Advances in Neural Information Processing Systems, 37, pp.108231-108261.

[28] Boniol, P., Liu, Q., Huang, M., Palpanas, T. and Paparrizos, J., 2024. Dive into time-series anomaly detection: A decade review. arXiv preprint arXiv:2412.20512.

[29] 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).

[30] 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.

[31] 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.

[32] 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.

[33] 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.

[34] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, et al. 2021. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

[35] Ming Jin, Qingsong Wen, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, et al. 2023. Large models for time series and spatio-temporal data: A survey and outlook. arXiv preprint arXiv:2310.10196.

[36] 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.

[37] Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. 2023. One fits all: Power general time series analysis by pretrained lm. Advances in Neural Information Processing Systems 36: 43322-43355.

[38] Siwon Kim, Kukjin Choi, Hyun-Soo Choi, Byunghan Lee, Sungroh Yoon. 2022. Towards a rigorous evaluation of time-series anomaly detection. In Proceedings of the AAAI Conference on Artificial Intelligence, 36(7): 7194-7201.

[39] Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, Artur Dubrawski. 2024. MOMENT: A Family of Open Time-series Foundation Models. In Proceedings of the International Conference on Machine Learning.

[40] 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.

[41] Liu, Q., Lee, S. and Paparrizos, J., 2025. TSB-AutoAD: Towards Automated Solutions for Time-Series Anomaly Detection. Proceedings of the VLDB Endowment (VLDB 2025).

[42] Kong, Y., Yang, Y., Hwang, Y., Du, W., Zohren, S., Wang, Z., ... & Wen, Q. (2025). Time-mqa: Time series multi-task question answering with context enhancement. arXiv preprint arXiv:2503.01875.

Thank you for attending!

Any Questions?