Detecting Subsequence Anomalies in Time Series

Sebastian Schmidl
sebastian.schmidl@hpi.de

Hasso Plattner Institute (HPI)
University of Potsdam
Germany

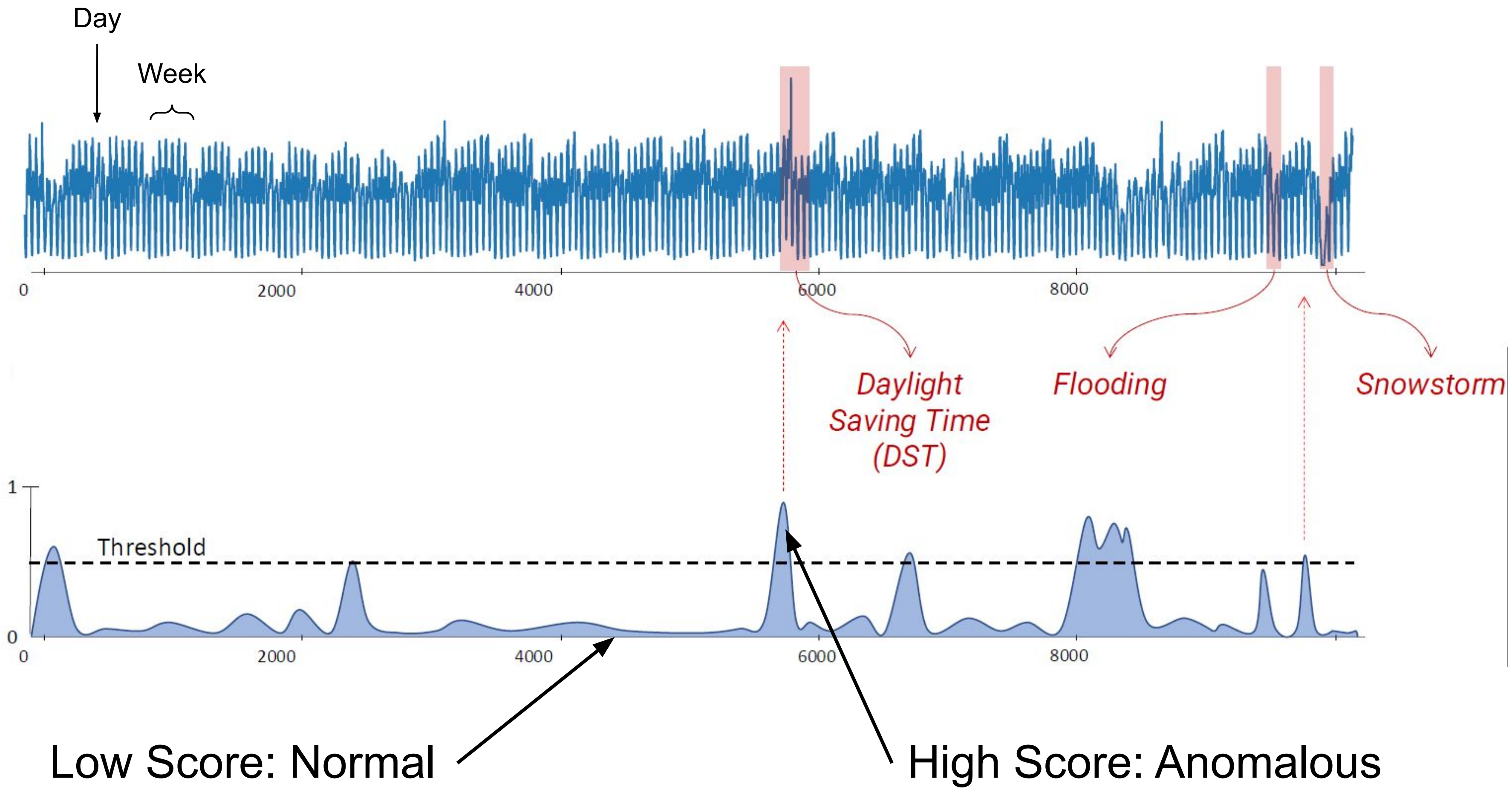
Phillip Wenig
phillip.wenig@hpi.de

Hasso Plattner Institute (HPI)
University of Potsdam
Germany

Thorsten Papenbrock
papenbrock@uni-marburg.de

Big Data Analytics
Philllipps-University of Marburg
Germany

Introduction



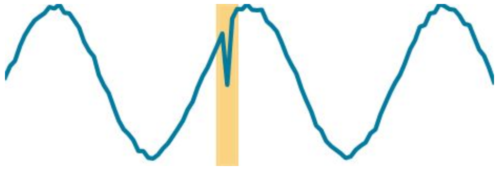
" An anomaly is a point or a sequence of points that deviate w.r.t. some measure or model from the regular patterns of the time series. "

For this tutorial:
A potentially undesired rare point or rare sequence of points of a given length.

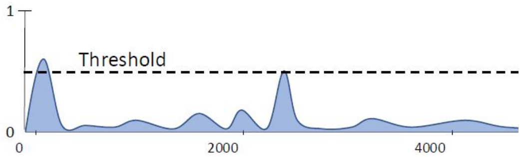
Image from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

Outline

1. Time Series Anomalies



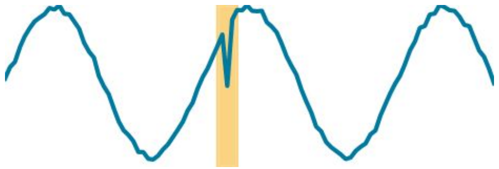
2. Anomaly Detection Methods



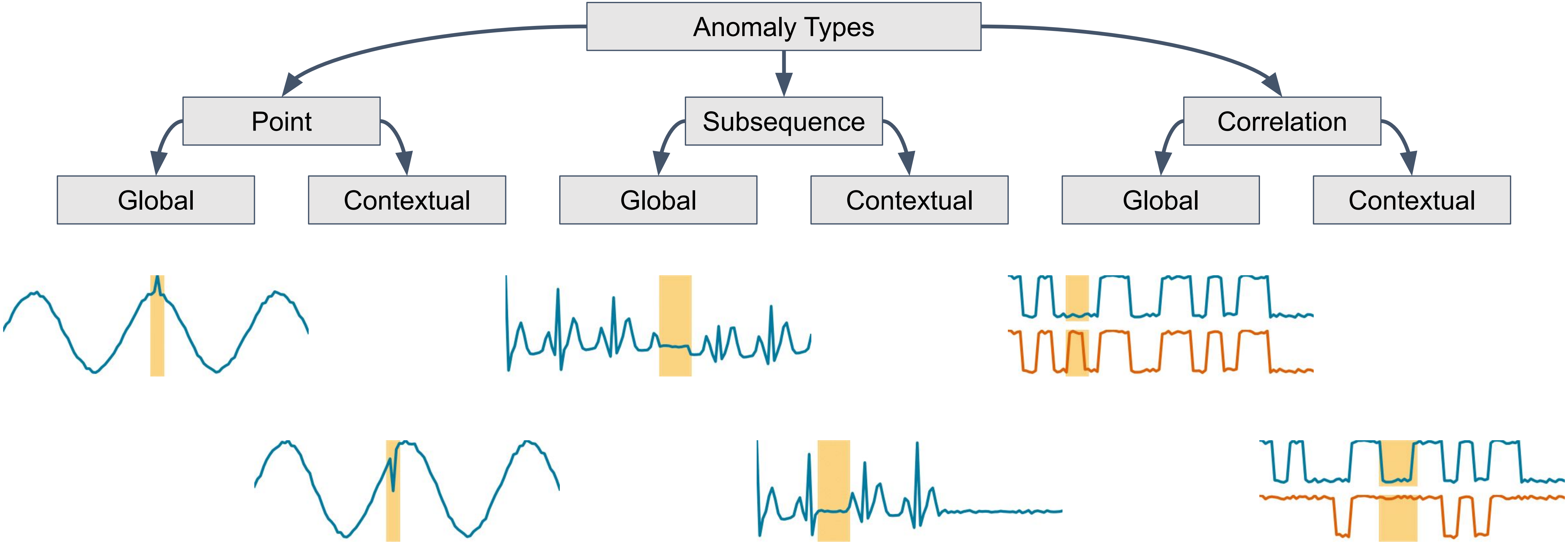
3. TSAD Benchmarks & Datasets

4. Outlook

Time Series Anomalies

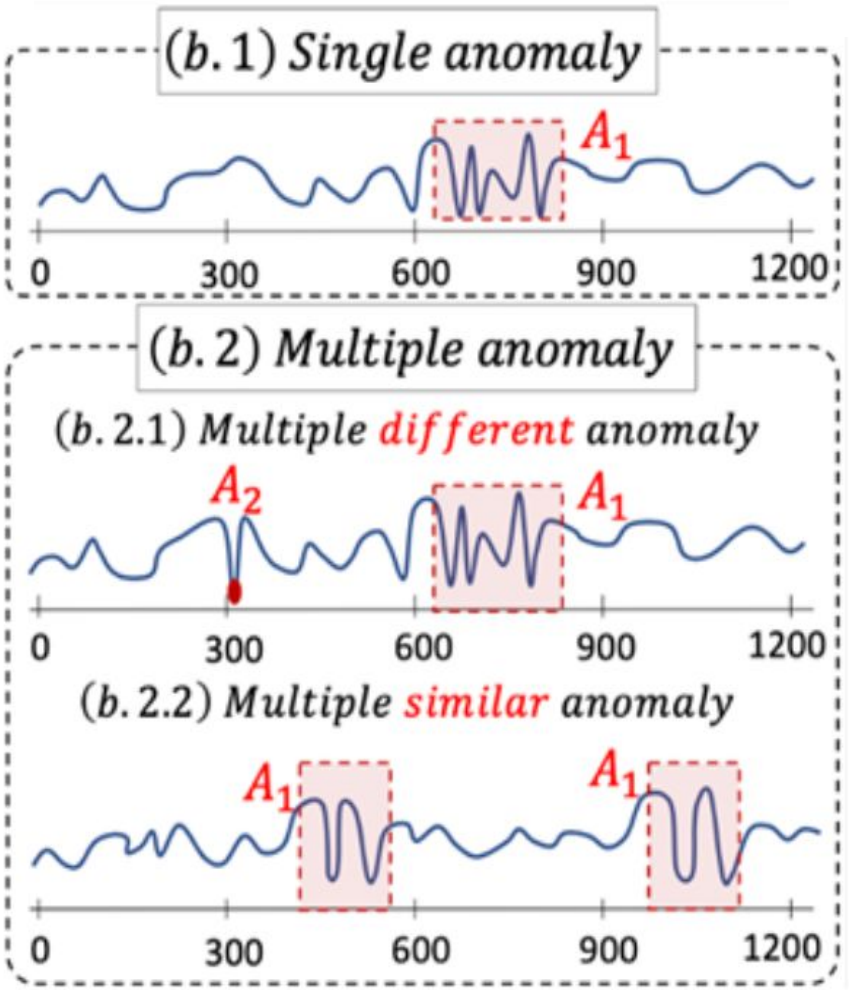


Anomaly Taxonomy



Phillip Wenig, Sebastian Schmidl, Thorsten Papenbrock: Anomaly Detectors for Multivariate Time Series: The Proof of the Pudding is in the Eating. Proceedings of the International Conference on Data Engineering Workshops (ICDEW), 2024. DOI:10.1109/ICDEW61823.2024.00018

Anomaly Taxonomy



Anomaly Arity

- for multivariate time series
- how many channels are affected

n-ary anomaly:

anomaly affects *n* channels/dimensions of the time series

Image from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

Real-world examples

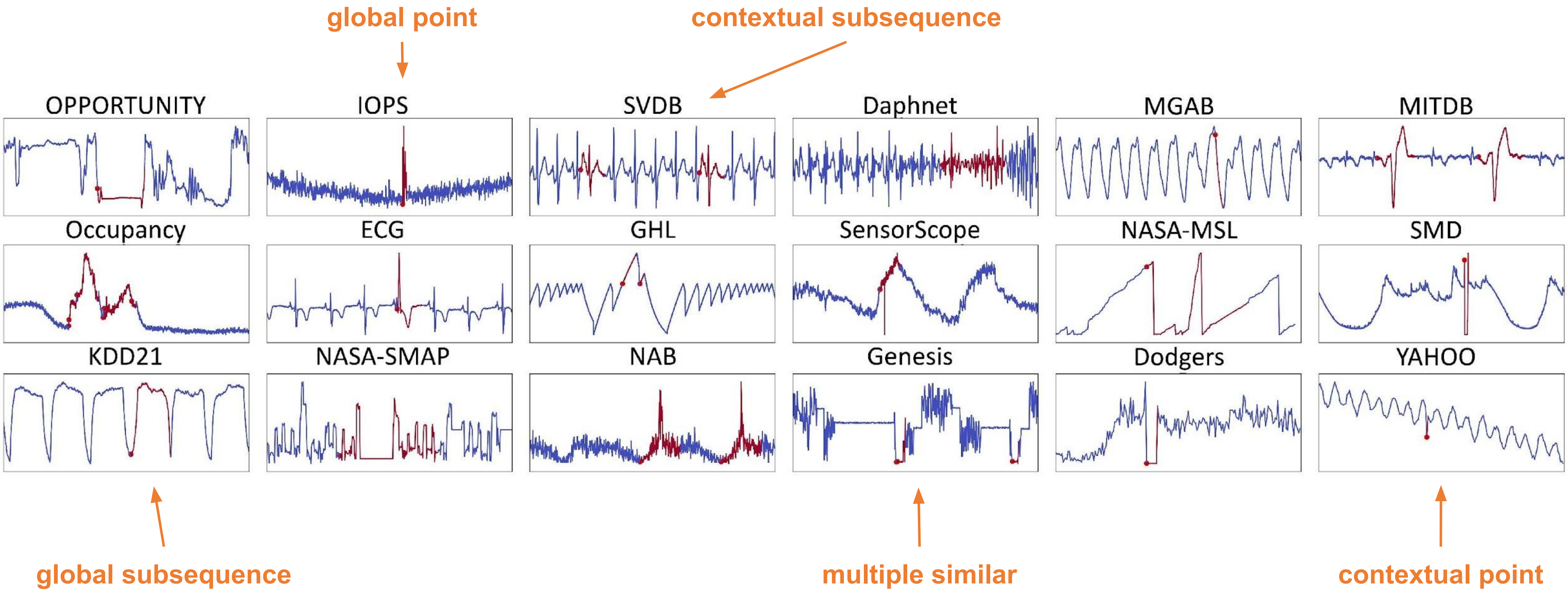
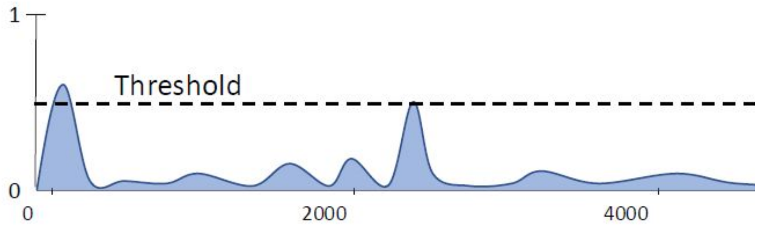
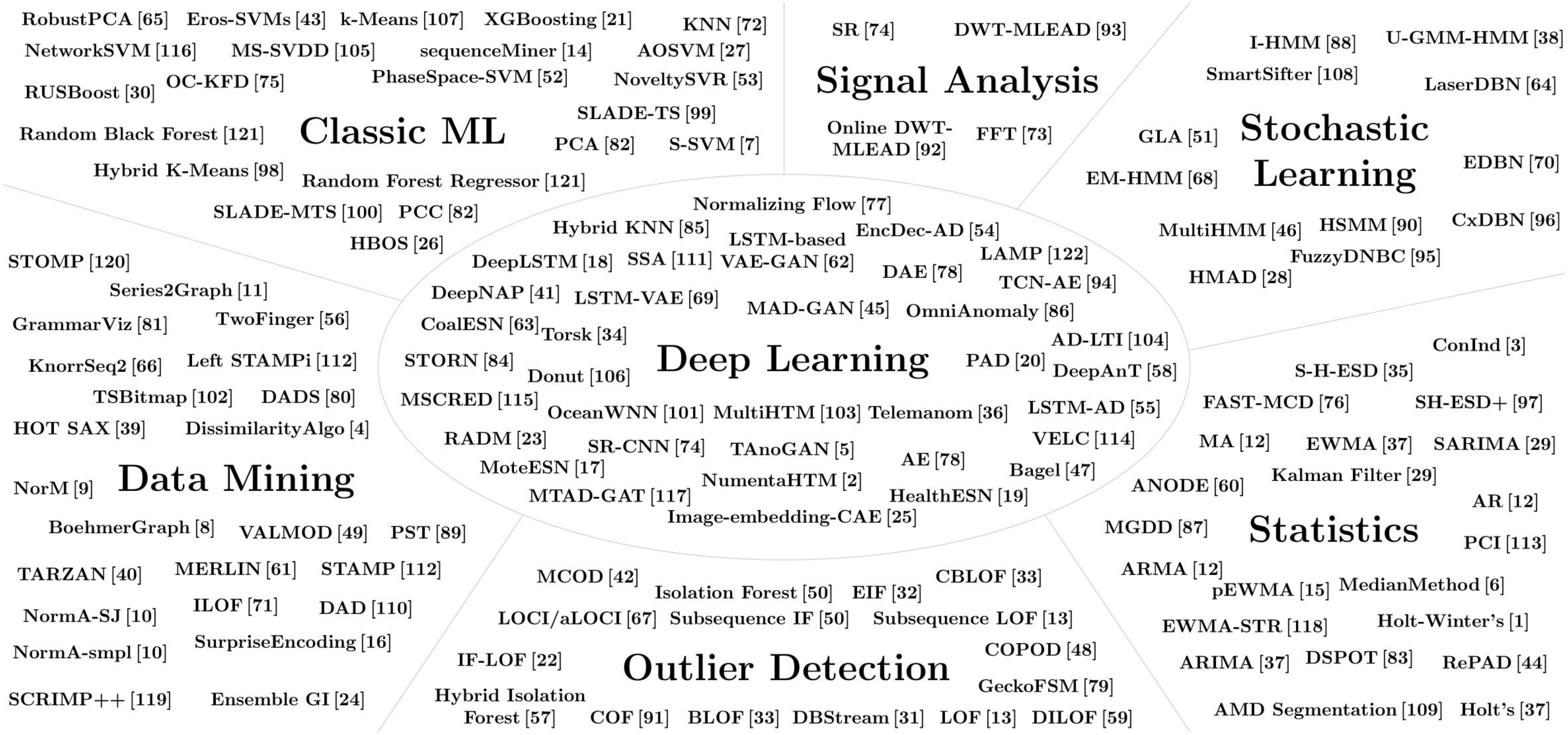


Image from TSB-UAD repository.

Anomaly Detection Methods

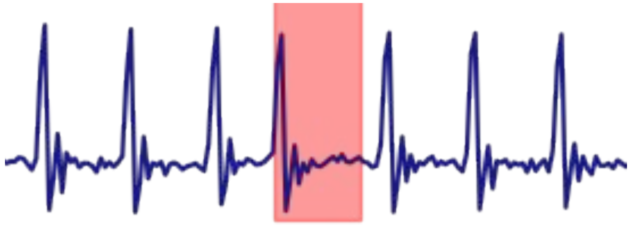
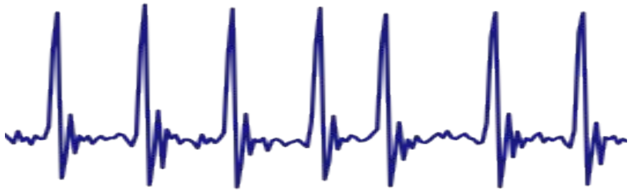
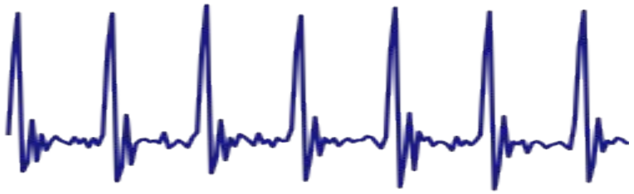


Anomaly Detection Methods – by Research Domains



Sebastian Schmidl, Phillip Wenig, Thorsten Papenbrock: Anomaly Detection in Time Series: A Comprehensive Evaluation. PVLDB 9:(15), 2022. DOI:10.14778/3538598.3538602

Anomaly Detection Methods – by Learning Type

	Training	**Testing**
Unsupervised (Type I [a])	∅	
Supervised (Type II [a])		
Semi-Supervised (Type III [a])		

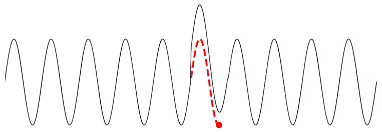
[a] Victoria J. Hodge and Jim Austin. 2004. A Survey of Outlier Detection Methodologies. Artificial Intelligence Review, 22, 2, 85–126. doi: 10.1007/s10462-004-4304-y

Anomaly Detection Methods – by Family

- Forecasting
- Reconstruction
- Encoding
- Distance
- Distribution
- Isolation Tree

Anomaly Detection Methods – by Family

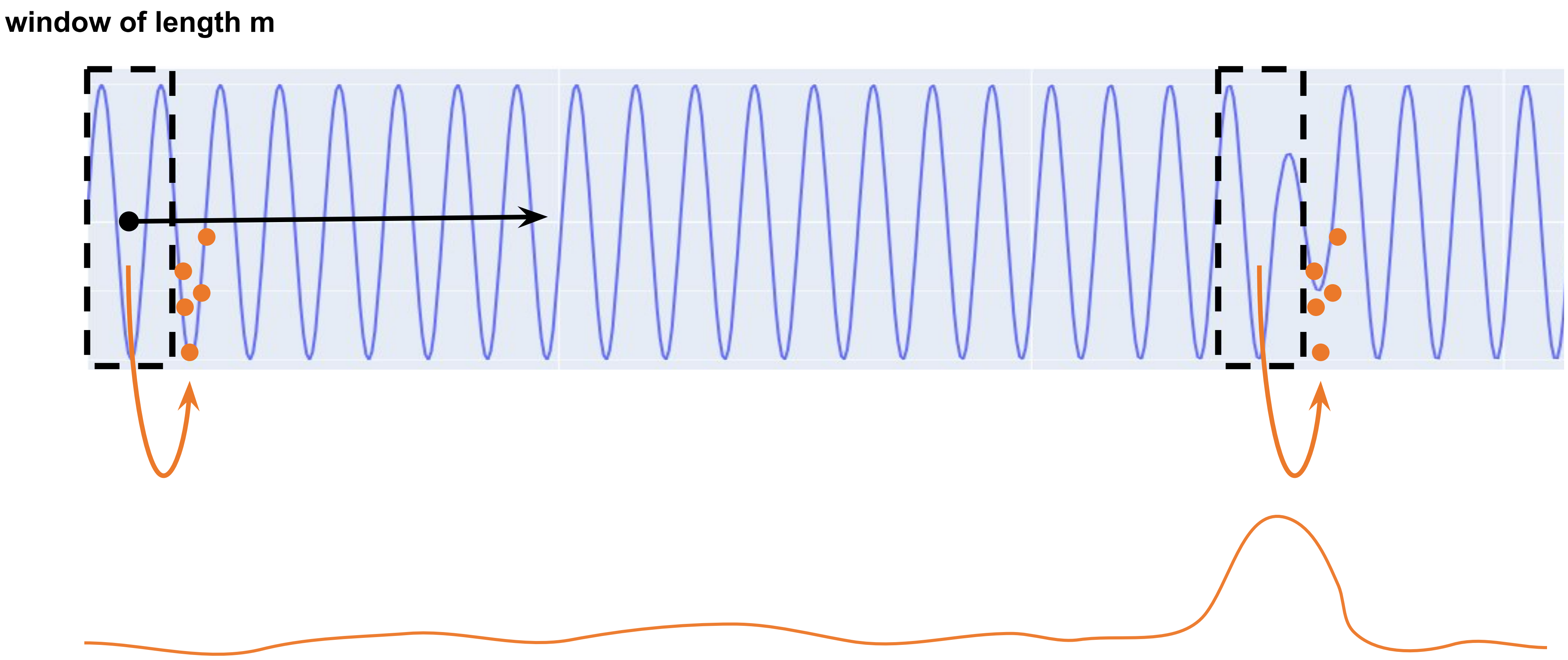
Forecasting



ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

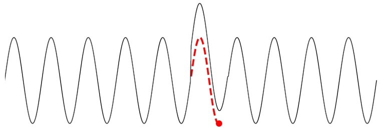
Forecasting Methods

- Methods try to **forecast the next point or sequence of points** based on previous window
- Anomaly score = **prediction error**



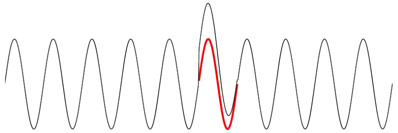
Anomaly Detection Methods – by Family

Forecasting



ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

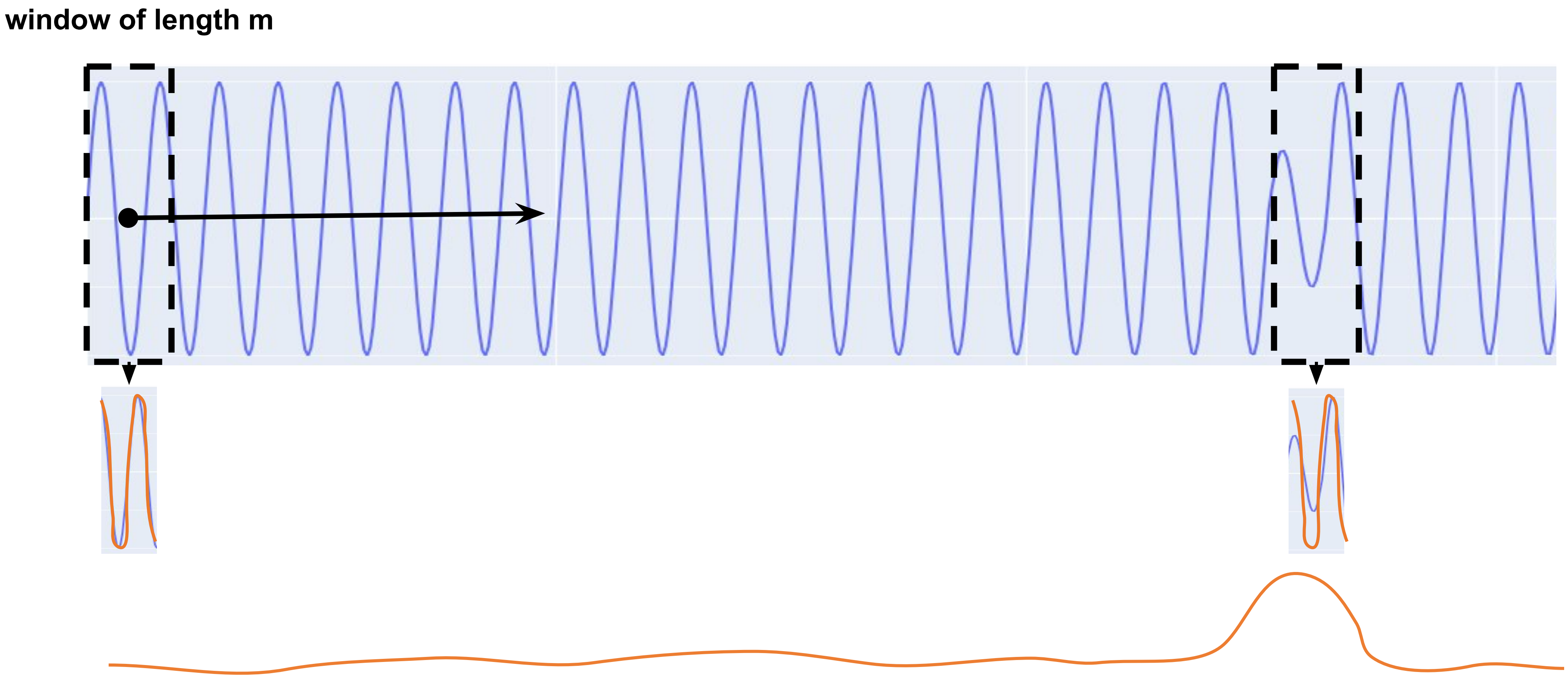
Reconstruction



AE, Bagel, DAE, Donut, EncDec-AD, FFT, IE-CAE, LSTM-VAE, MSCRED, OmniAnomaly, PCI, PCC, RobustPCA, SR, SR-CNN, TAnoGan

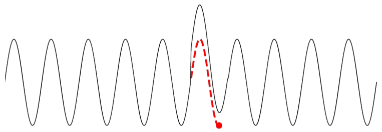
Reconstruction Methods

- Training: Methods build normal model by encoding normal subsequences in latent space
- Testing: Methods try to **reconstruct subsequences from latent space**
- Anomaly score = **reconstruction error**



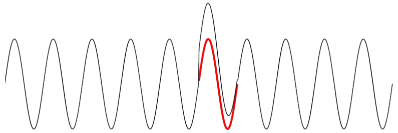
Anomaly Detection Methods – by Family

Forecasting



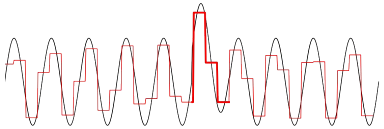
ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

Reconstruction



AE, Bagel, DAE, Donut, EncDec-AD, FFT, IE-CAE, LSTM-VAE, MSCRED, OmniAnomaly, PCI, PCC, RobustPCA, SR, SR-CNN, TAnoGan

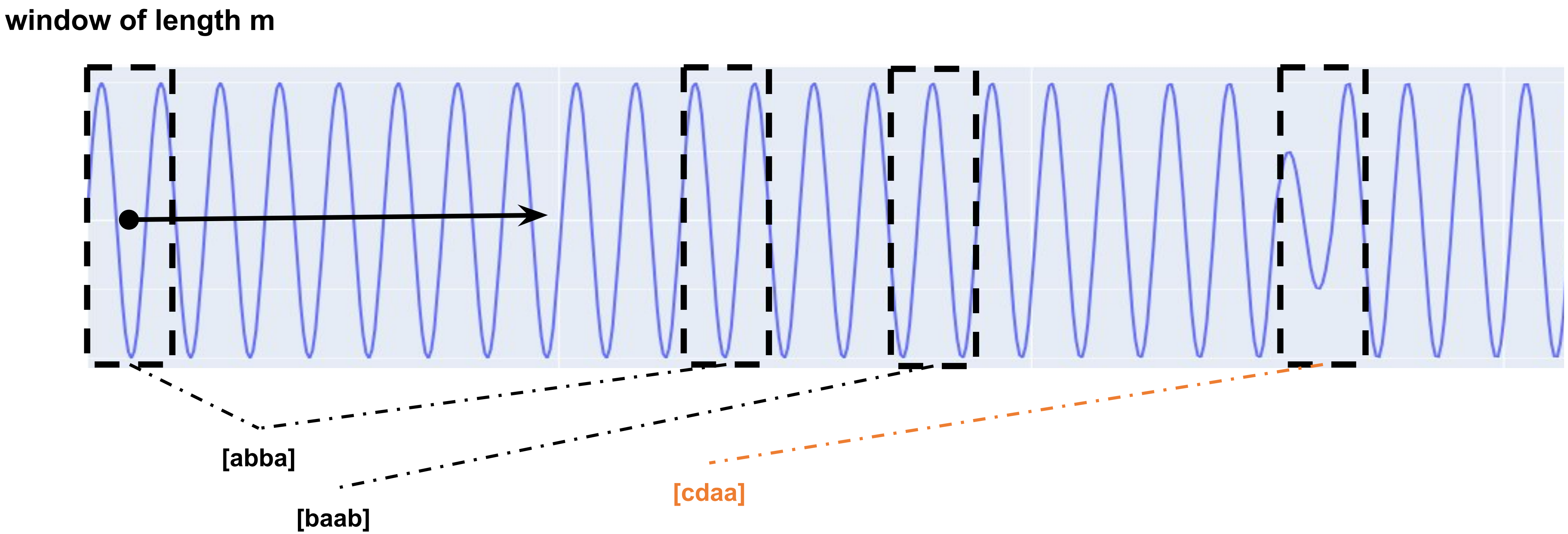
Encoding



Ensemble GI, GrammarViz, LaserDBN, MultiHMM, PST, Series2Graph, TARZAN, TSBitmap

Encoding Methods

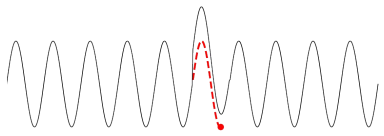
- Methods **encode subsequences in latent space** and identify anomalies in this space
- Anomaly score = **depends on representation**



GrammarViz: Coverage of grammar rules
Series2Graph: Edge weight
TsBitmap: Bitmap similarity
…

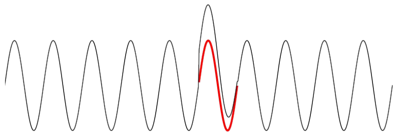
Anomaly Detection Methods – by Family

Forecasting



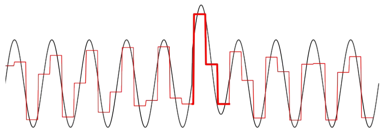
ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

Reconstruction



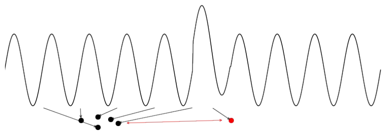
AE, Bagel, DAE, Donut, EncDec-AD, FFT, IE-CAE, LSTM-VAE, MSCRED, OmniAnomaly, PCI, PCC, RobustPCA, SR, SR-CNN, TAnoGan

Encoding



Ensemble GI, GrammarViz, LaserDBN, MultiHMM, PST, Series2Graph, TARZAN, TSBitmap

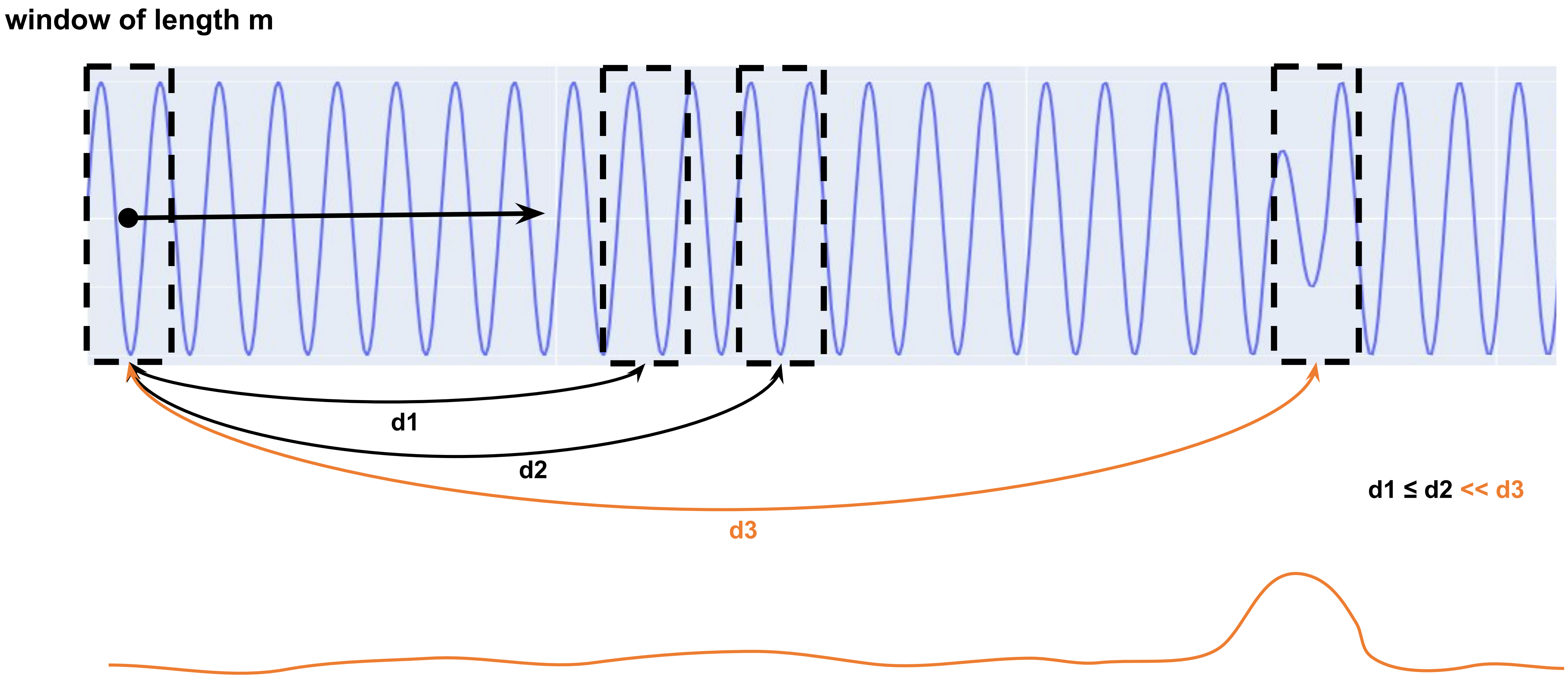
Distance



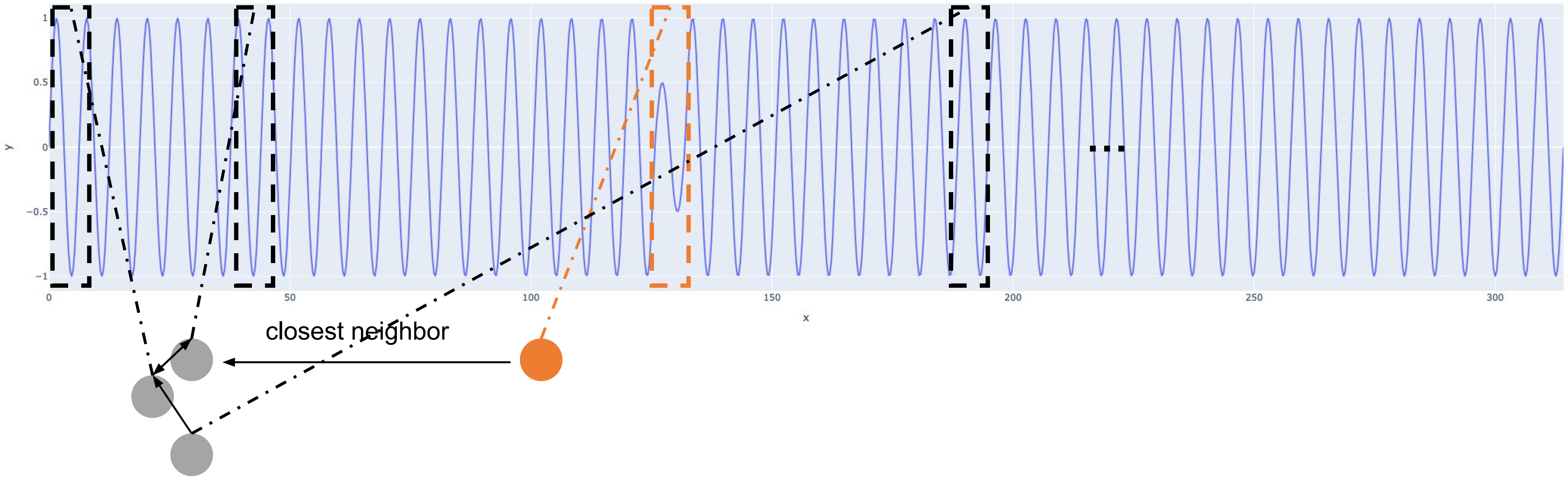
CBLOF, COF, DBStream, HOT SAX, Hybrid KNN, k-Means, KNN, LOF, NormA-SJ, PS-SVM, SAND, SSA, STAMP, STOMP, Sub-LOF, VALMOD, Left STAMPi

Distance Methods

- Methods compute distances between points/subsequences (or groups of subsequences)
- Anomaly score = **distance value**

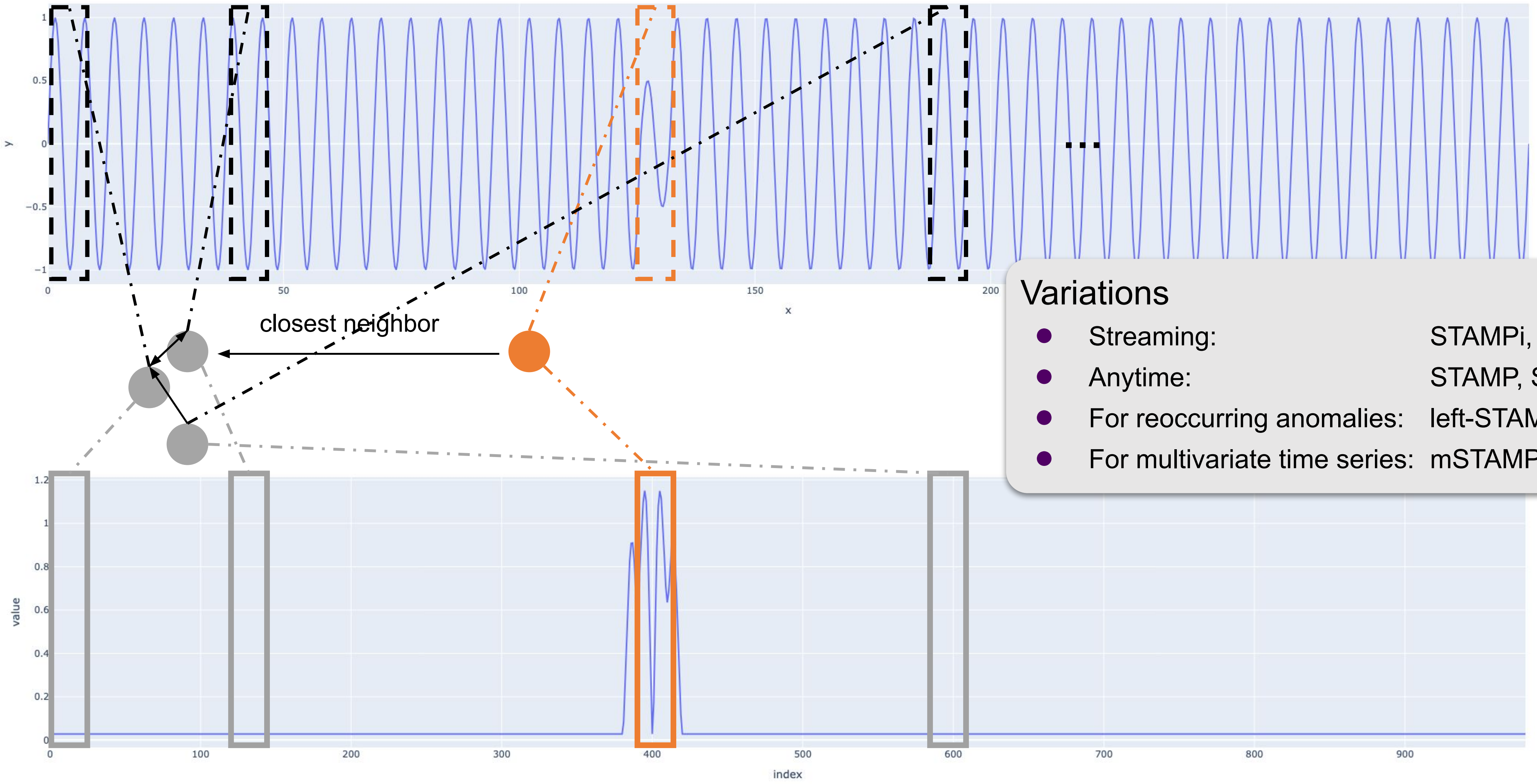


Distance Methods Example: MatrixProfile



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.

Distance Methods Example: MatrixProfile



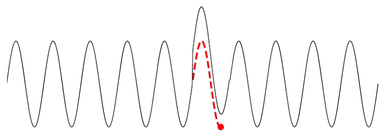
Variations

- Streaming: STAMPi, DAMP
- Anytime: STAMP, STOMP
- For reoccurring anomalies: left-STAMP
- For multivariate time series: mSTAMP

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.

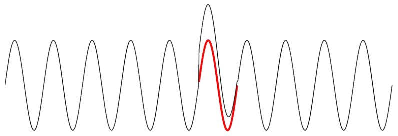
Anomaly Detection Methods – by Family

Forecasting



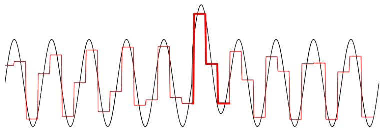
ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

Reconstruction



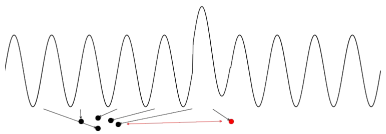
AE, Bagel, DAE, Donut, EncDec-AD, FFT, IE-CAE, LSTM-VAE, MSCRED, OmniAnomaly, PCI, PCC, RobustPCA, SR, SR-CNN, TAnoGan

Encoding



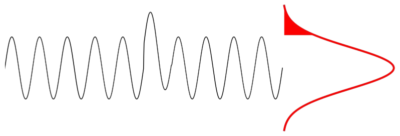
Ensemble GI, GrammarViz, LaserDBN, MultiHMM, PST, Series2Graph, TARZAN, TSBitmap

Distance



CBLOF, COF, DBStream, HOT SAX, Hybrid KNN, k-Means, KNN, LOF, NormA-SJ, PS-SVM, SAND, SSA, STAMP, STOMP, Sub-LOF, VALMOD, Left STAMPi

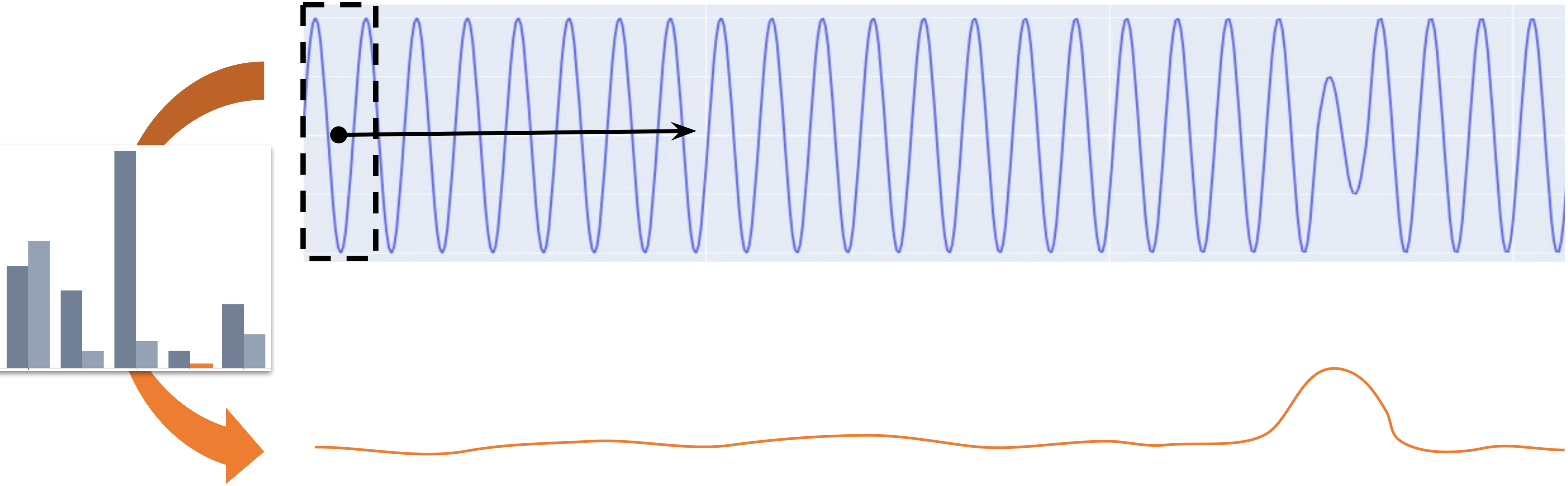
Distribution



COPOD, DWT-MLEAD, Fast-MCD, HBOS, NF, S-H-ESD, DSPOT, Sub-Fast-MCD

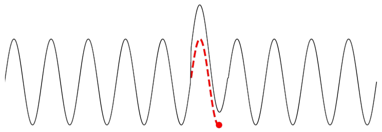
Distribution Methods

- Methods **estimate the distribution** of the data or fit a distribution model to the data (frequency instead of distance)
- Anomaly score = **probabilities / likelihoods**



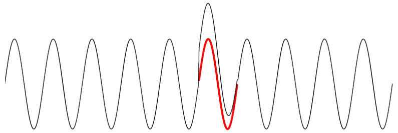
Anomaly Detection Methods – by Family

Forecasting



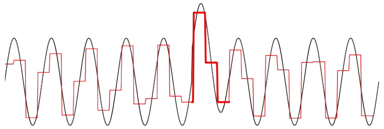
ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

Reconstruction



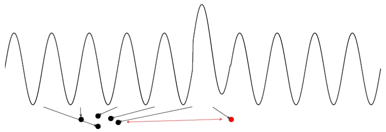
AE, Bagel, DAE, Donut, EncDec-AD, FFT, IE-CAE, LSTM-VAE, MSCRED, OmniAnomaly, PCI, PCC, RobustPCA, SR, SR-CNN, TAnoGan

Encoding



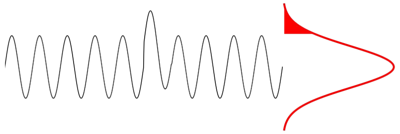
Ensemble GI, GrammarViz, LaserDBN, MultiHMM, PST, Series2Graph, TARZAN, TSBitmap

Distance



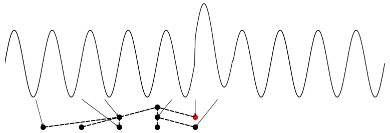
CBLOF, COF, DBStream, HOT SAX, Hybrid KNN, k-Means, KNN, LOF, NormA-SJ, PS-SVM, SAND, SSA, STAMP, STOMP, Sub-LOF, VALMOD, Left STAMPi

Distribution



COPOD, DWT-MLEAD, Fast-MCD, HBOS, NF, S-H-ESD, DSPOT, Sub-Fast-MCD

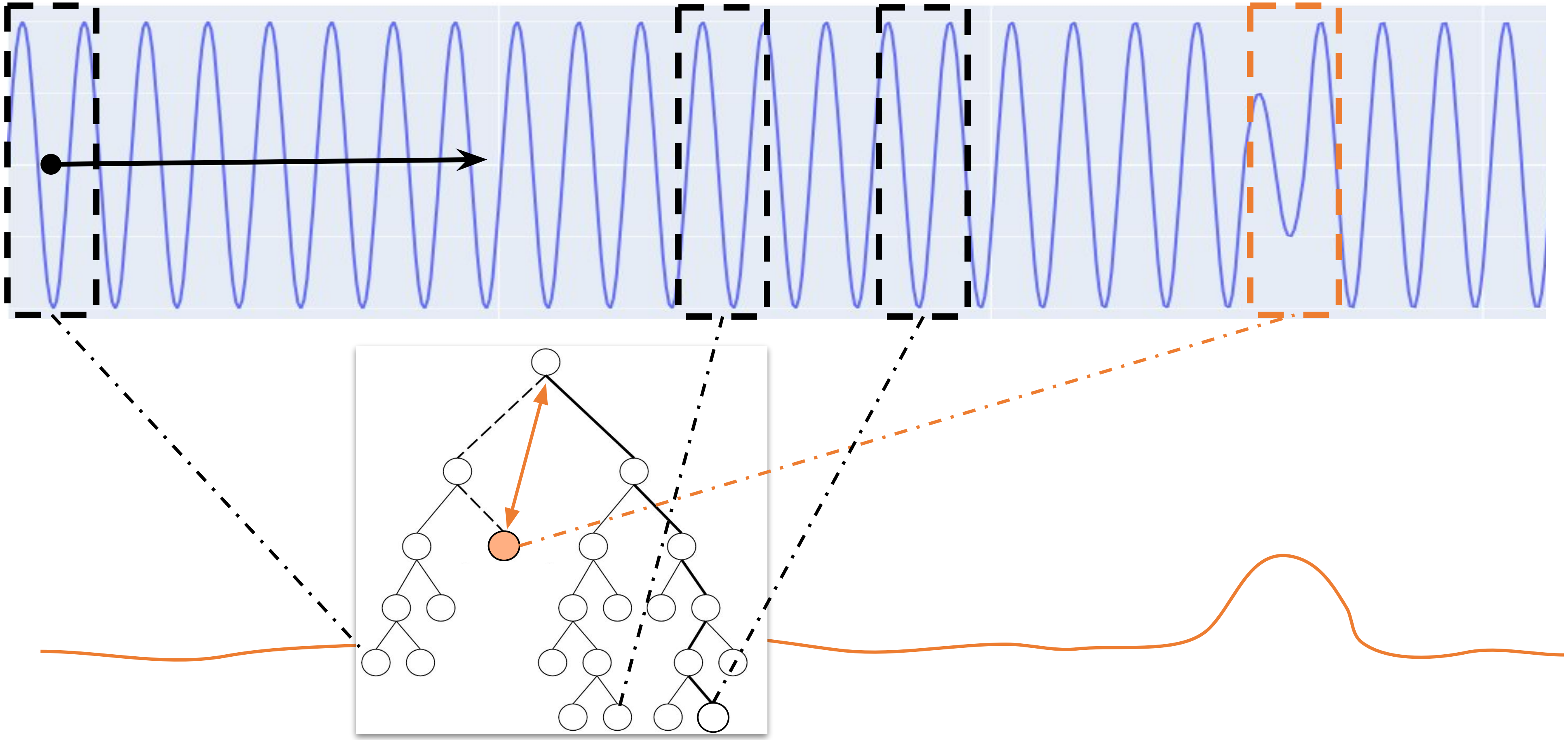
Isolation Tree



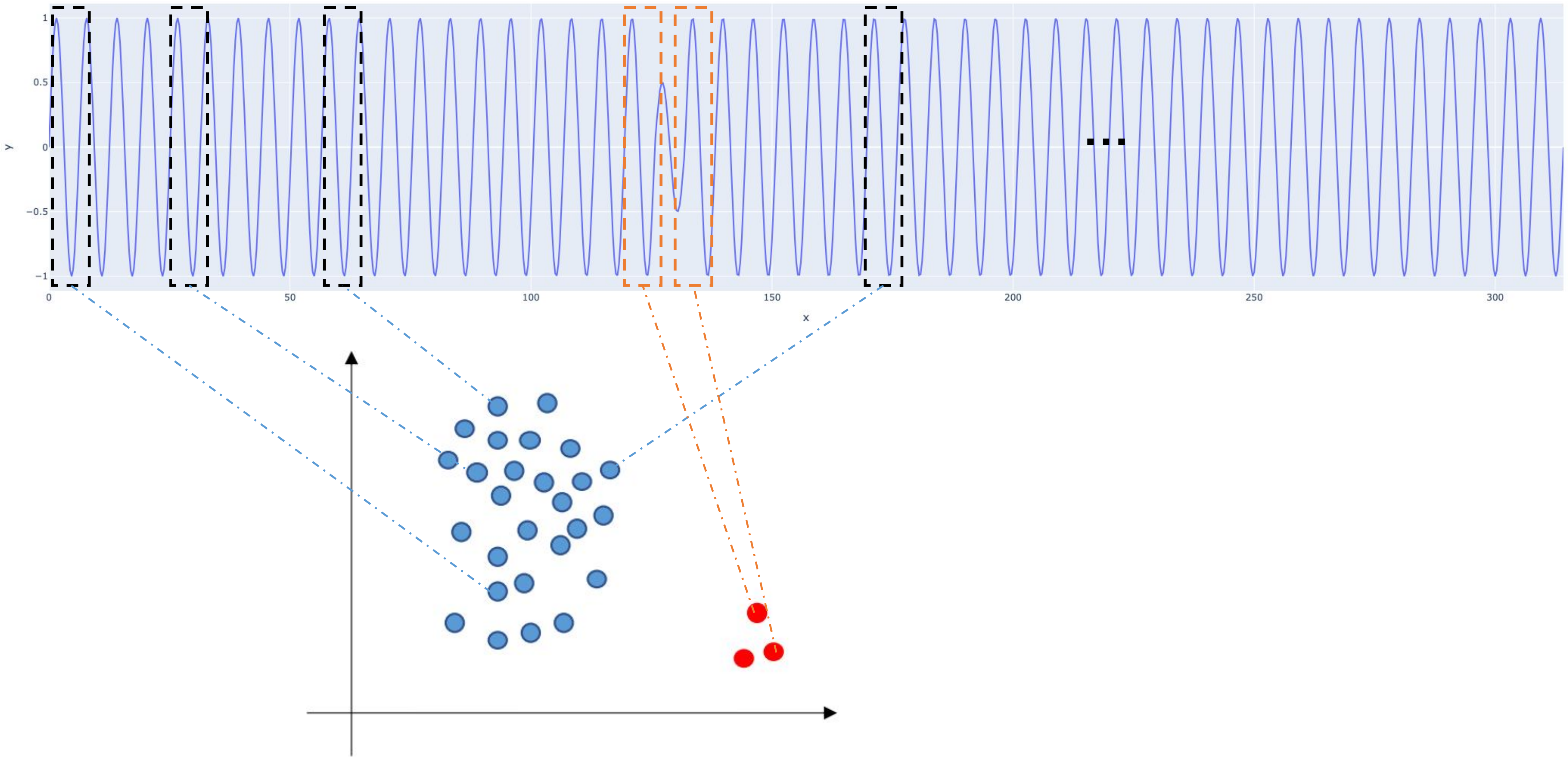
EIF, HIF, IF-LOF, iForest, Sub-IF

Isolation Tree Methods

- Methods build an **ensemble of random trees that partition the samples** (points or subsequences)
- Anomaly score = **average reciprocal path length**

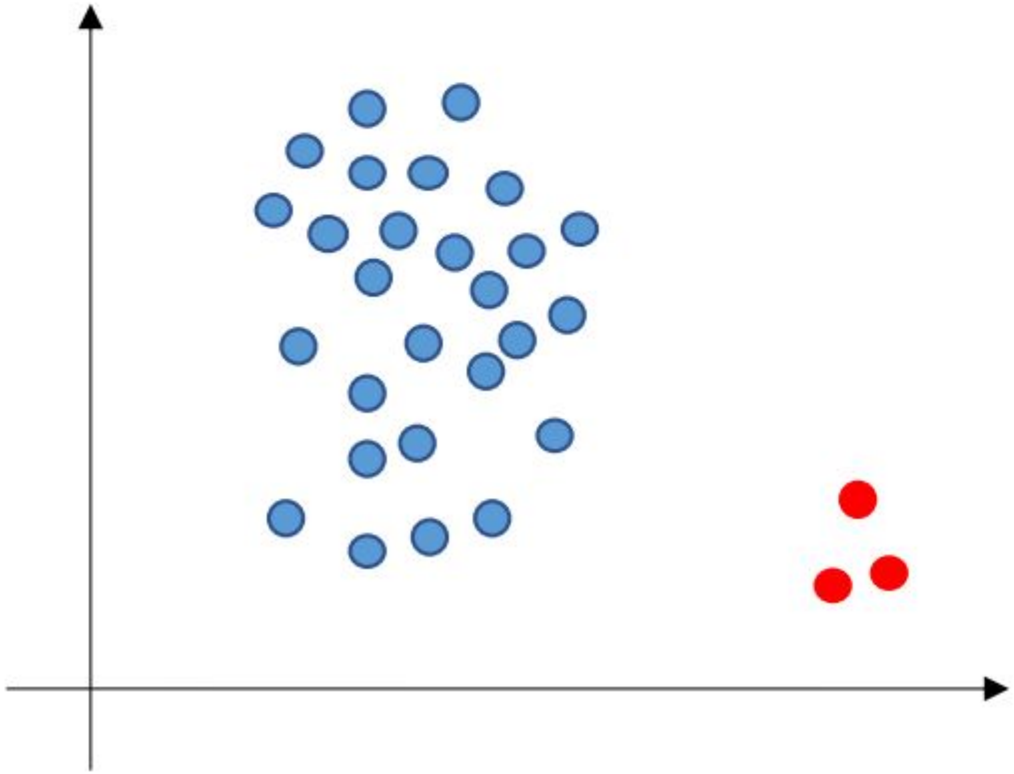


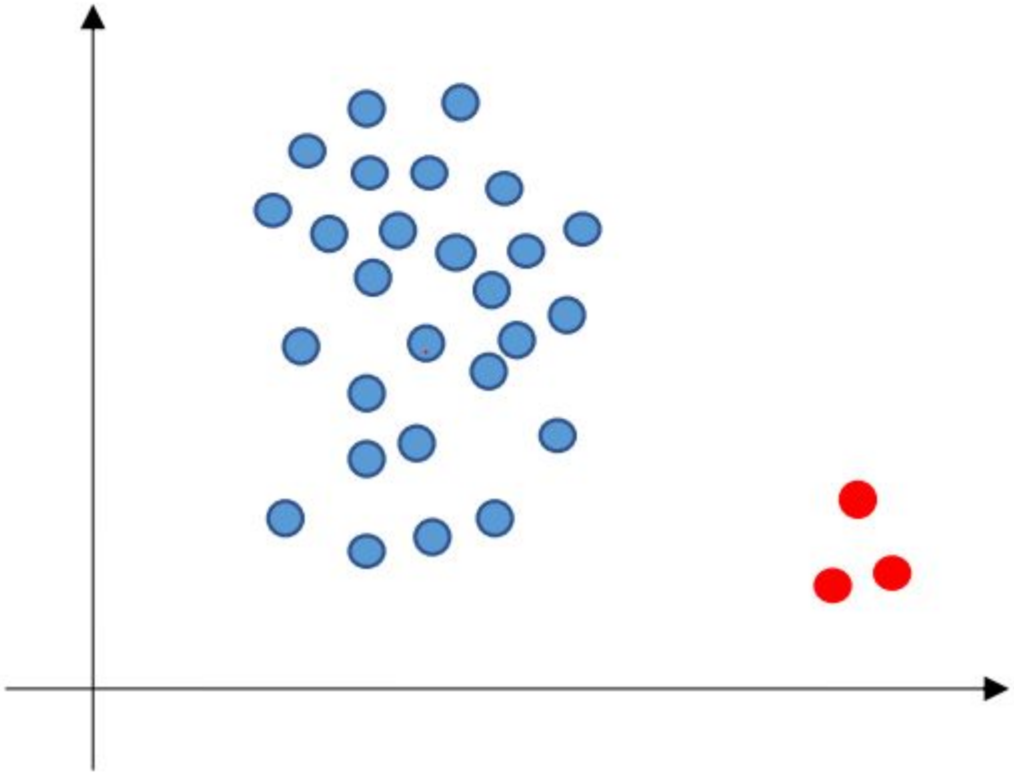
Isolation Tree Methods Example: Isolation Forest



Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

Isolation Tree Methods Example: Isolation Forest

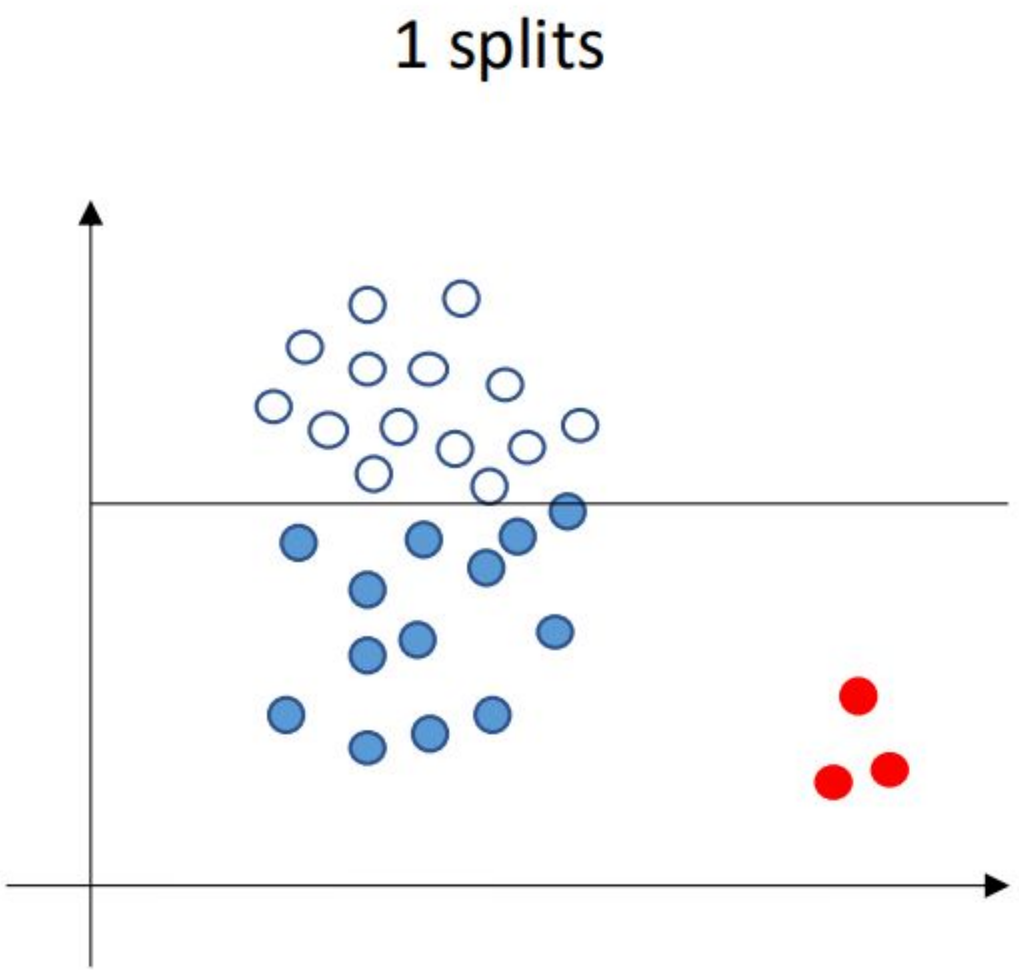


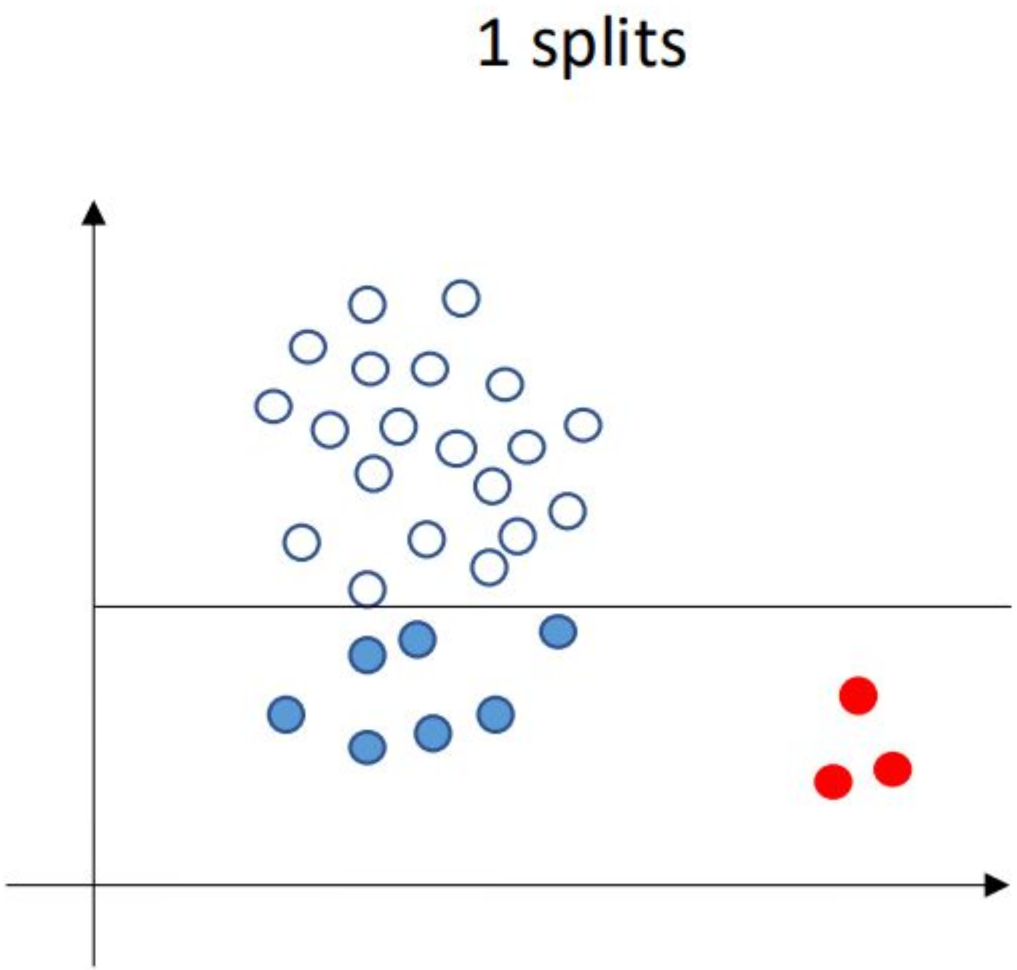


Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

Screenshot from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

Isolation Tree Methods Example: Isolation Forest

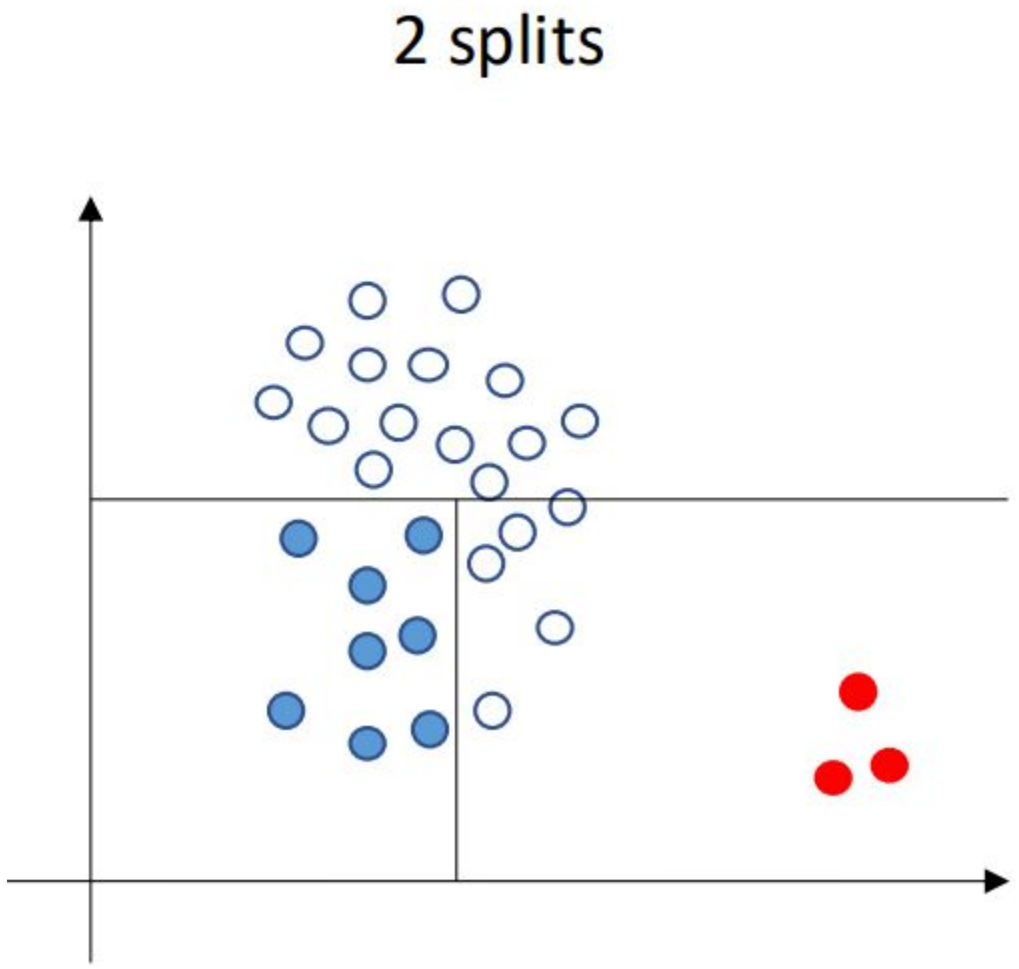


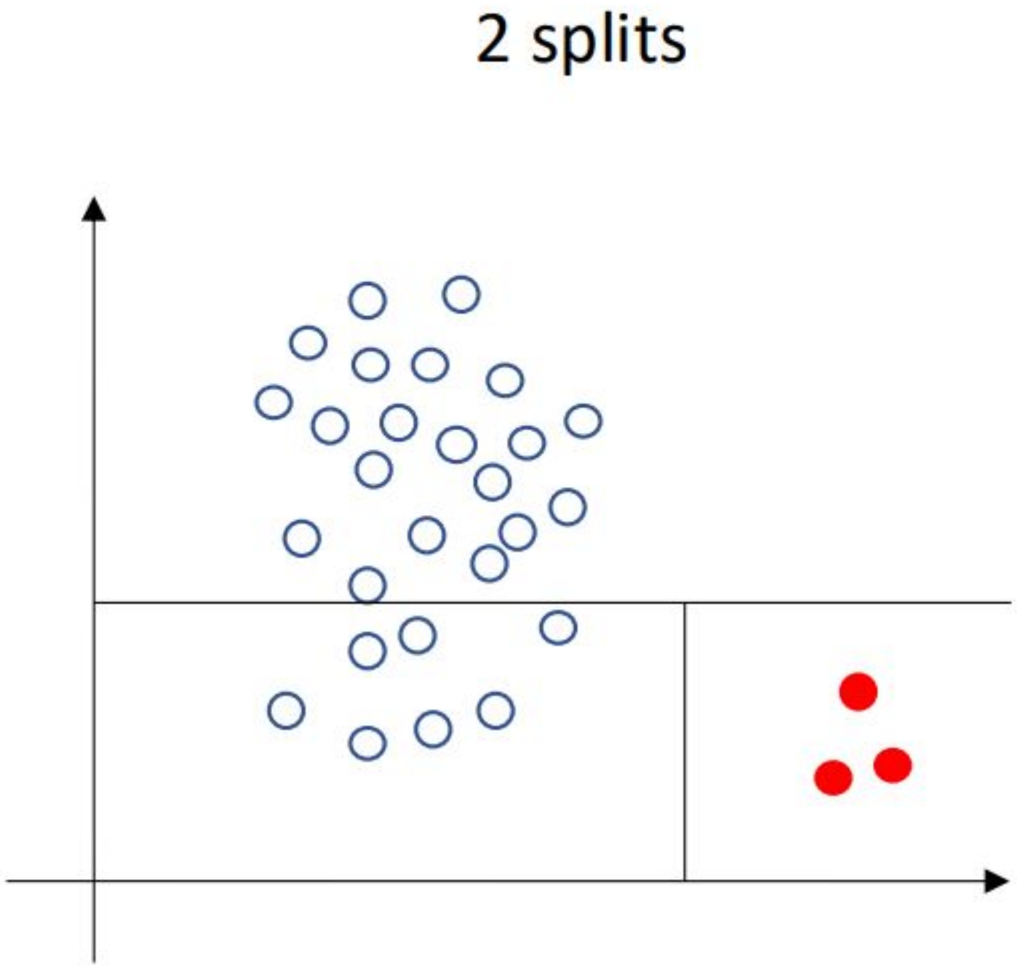


Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

Screenshot from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

Isolation Tree Methods Example: Isolation Forest



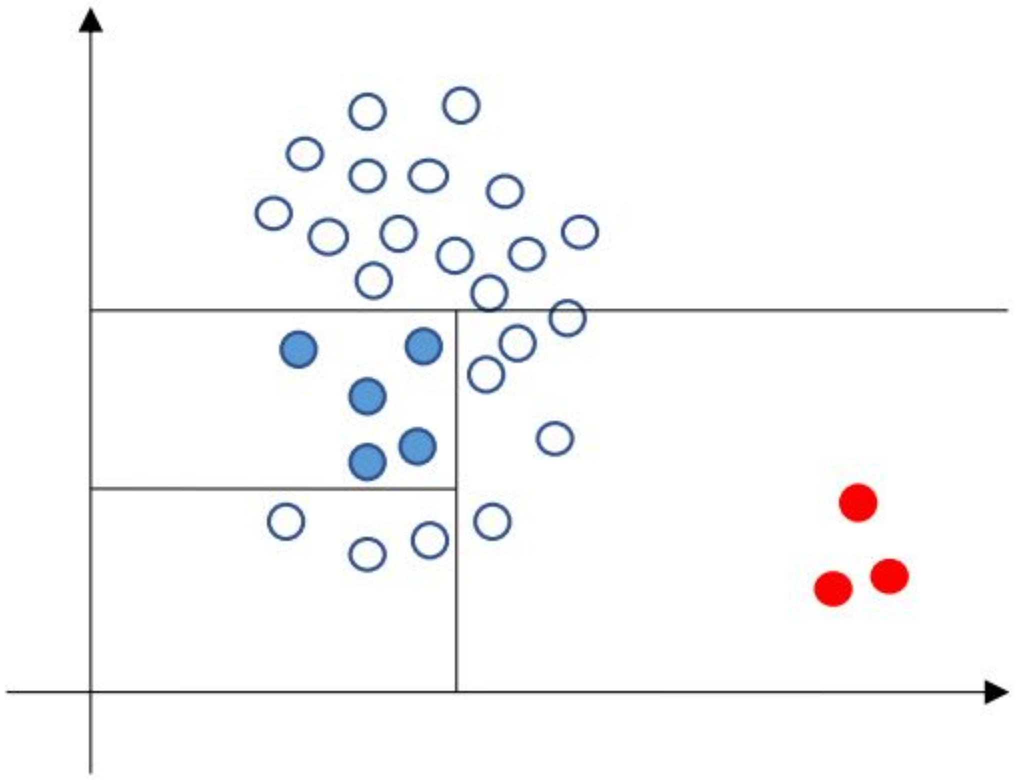


Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

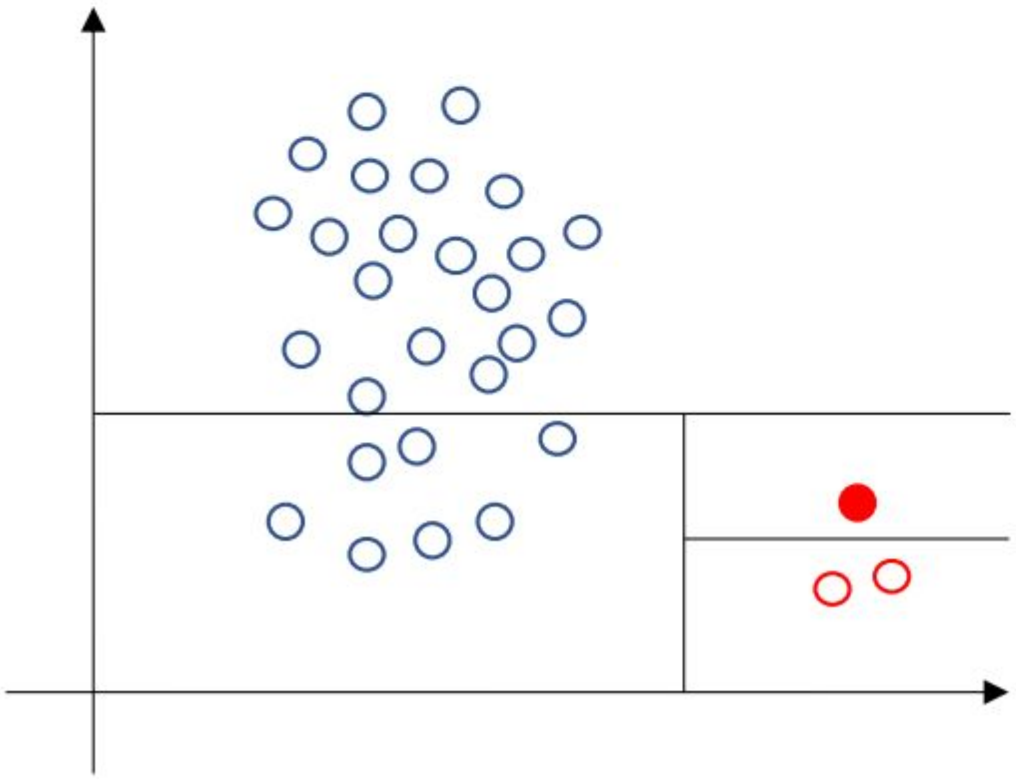
Screenshot from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

Isolation Tree Methods Example: Isolation Forest

3 splits



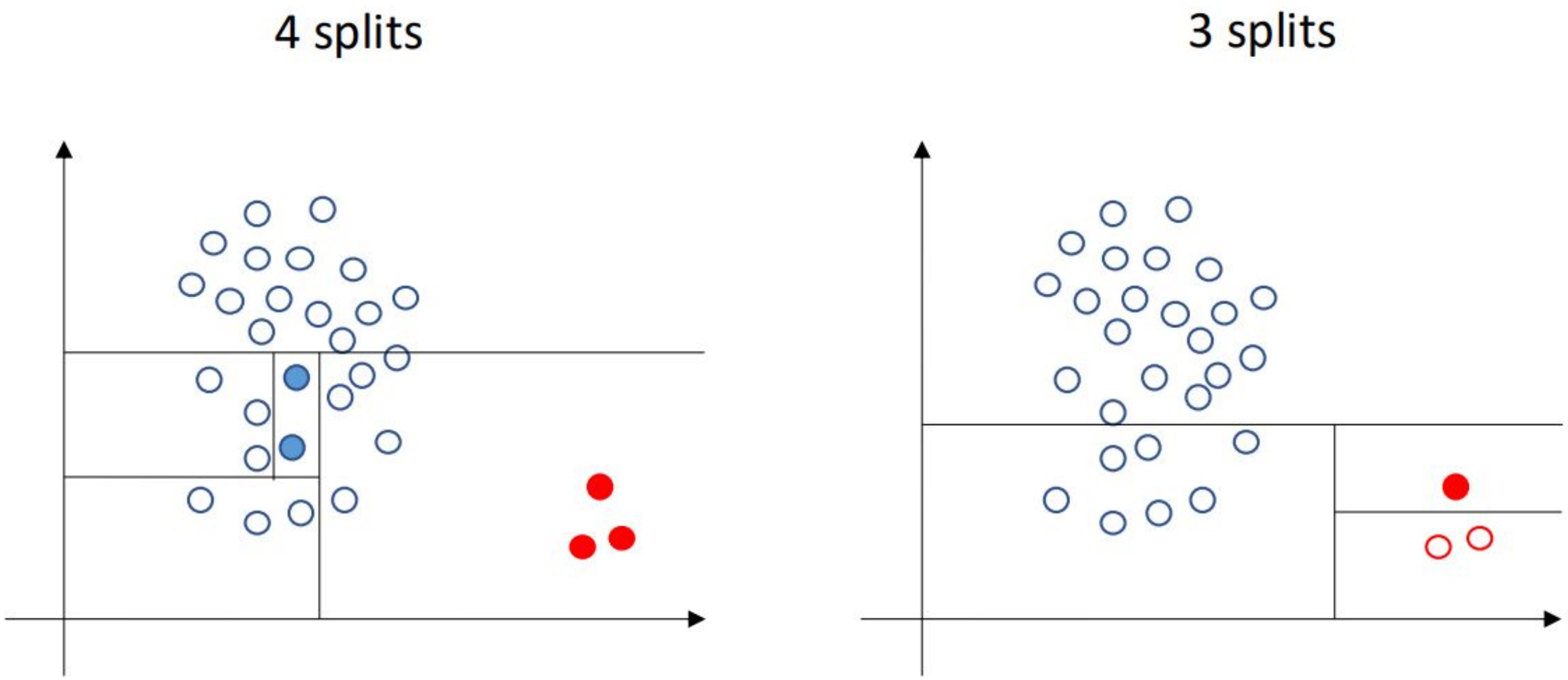
3 splits



Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

Screenshot from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

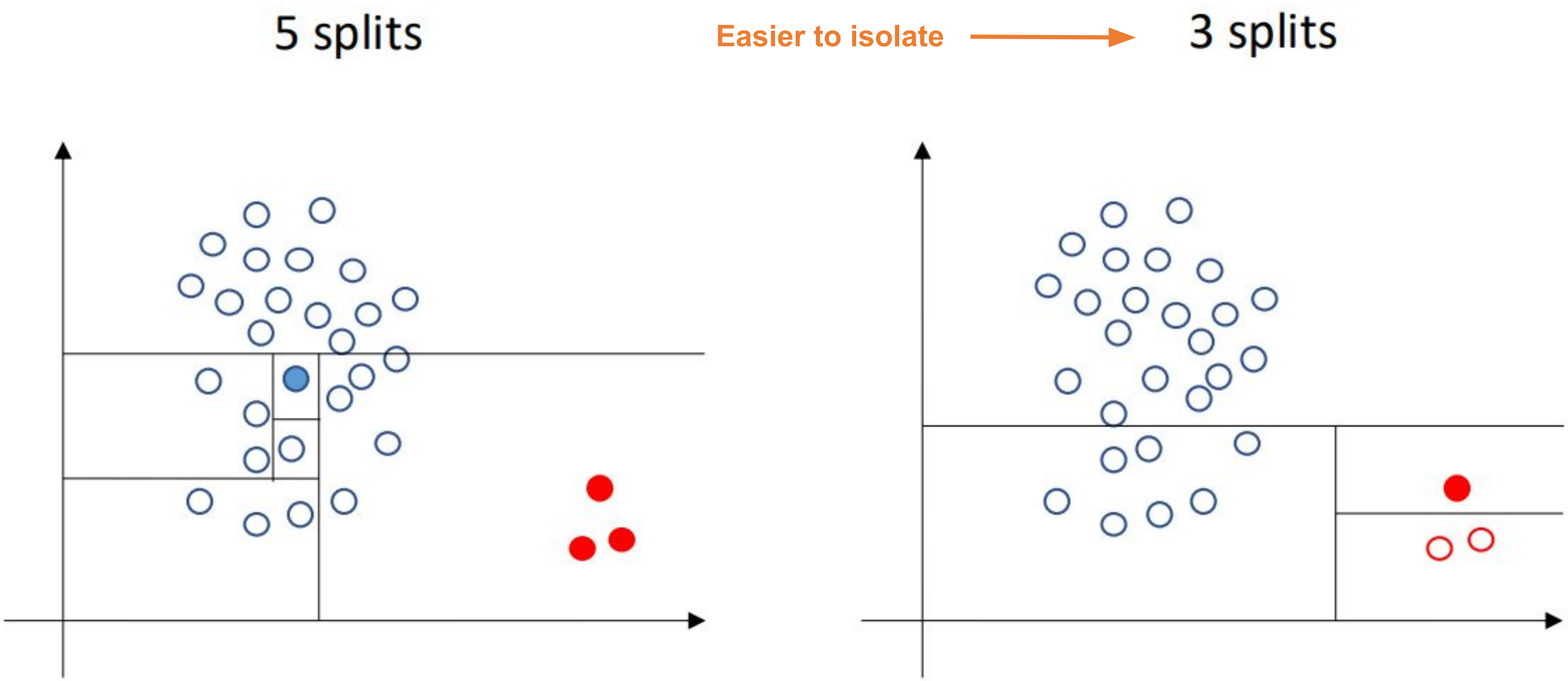
Isolation Tree Methods Example: Isolation Forest



Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

Screenshot from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

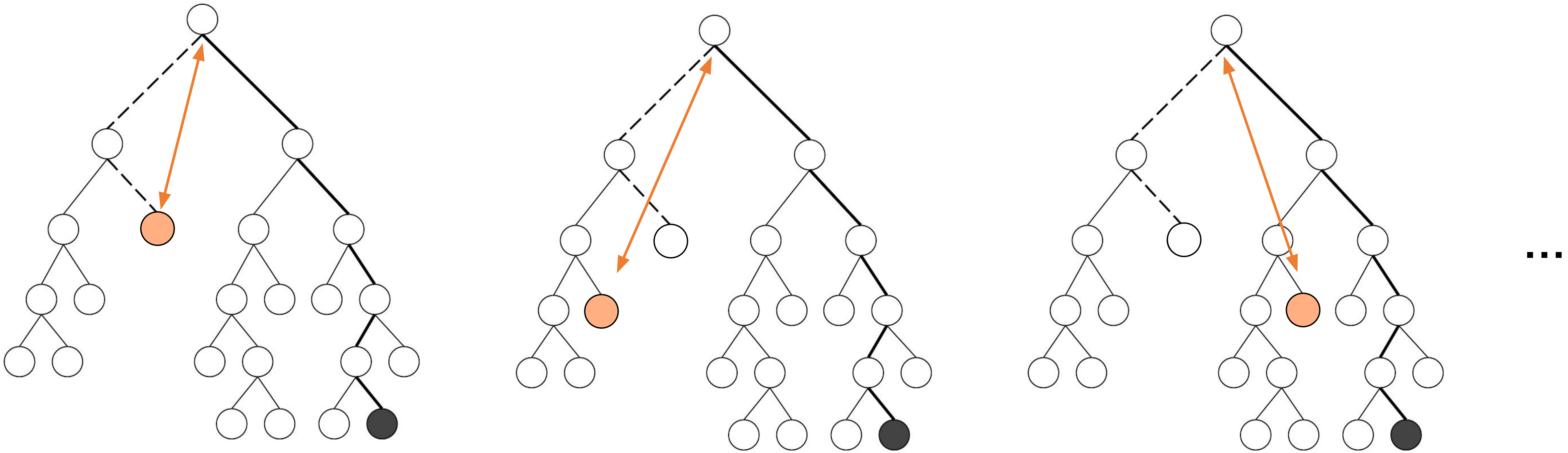
Isolation Tree Methods Example: Isolation Forest



Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

Screenshot from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

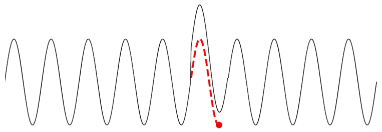
Isolation Tree Methods Example: Isolation Forest



Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2008. Isolation Forest. In Proceedings of the International Conference on Data Mining (ICDM), 413–422. doi: 10.1109/ICDM.2008.17.

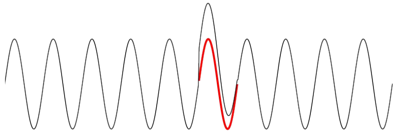
Anomaly Detection Methods in aeon

Forecasting



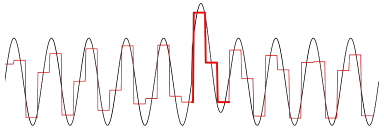
ARIMA, DeepAnT, DeepNAP, HealthESN, LSTM-AD, MedianMethod, MTAD-GAT, NumentaHTM, NoveltySVR, OceanWNN, RBForest, RForest, SARIMA, Telemanom, Torsk, Triple ES, XGBoosting

Reconstruction



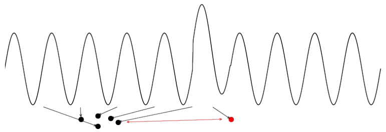
AE, Bagel, DAE, Donut, EncDec-AD, FFT, IE-CAE, LSTM-VAE, MSCRED, OmniAnomaly, PCI, **PCC**, RobustPCA, SR, SR-CNN, TAnoGan

Encoding



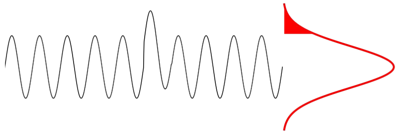
Ensemble GI, GrammarViz, LaserDBN, MultiHMM, PST, Series2Graph, TARZAN, TSBitmap

Distance



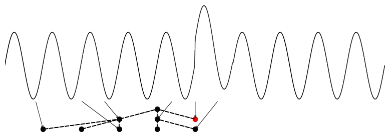
CBLOF, **COF**, DBStream, HOT SAX, Hybrid KNN, **k-Means**, **KNN**, **LOF**, NormA-SJ, PS-SVM, SAND, SSA, **STAMP**, STOMP, **Sub-LOF**, VALMOD, Left STAMPi

Distribution



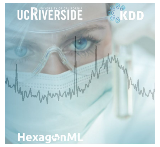
COPOD, **DWT-MLEAD**, Fast-MCD, **HBOS**, NF, S-H-ESD, DSPOT, Sub-Fast-MCD

Isolation Tree



EIF, HIF, IF-LOF, **iForest**, **Sub-IF**

TSAD Benchmark & Datasets



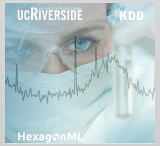
Benchmarks & Datasets

Numenta

NAB

58 time series

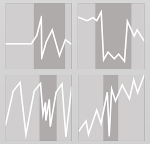
- Univariate
- Real-world and synthetic
- Streaming
- Own scoring mechanism
- Various domains



HEX/UCR

250 time series

- Univariate
- Manually checked labels
- Only 1 annotated anomaly per time series
- Various domains



TSB-UAD

2 000 time series

- Univariate
- No filtering
- Proposes data transformations to introduce new anomalies in existing time series
- Includes time series generated from classification datasets

TimeEval

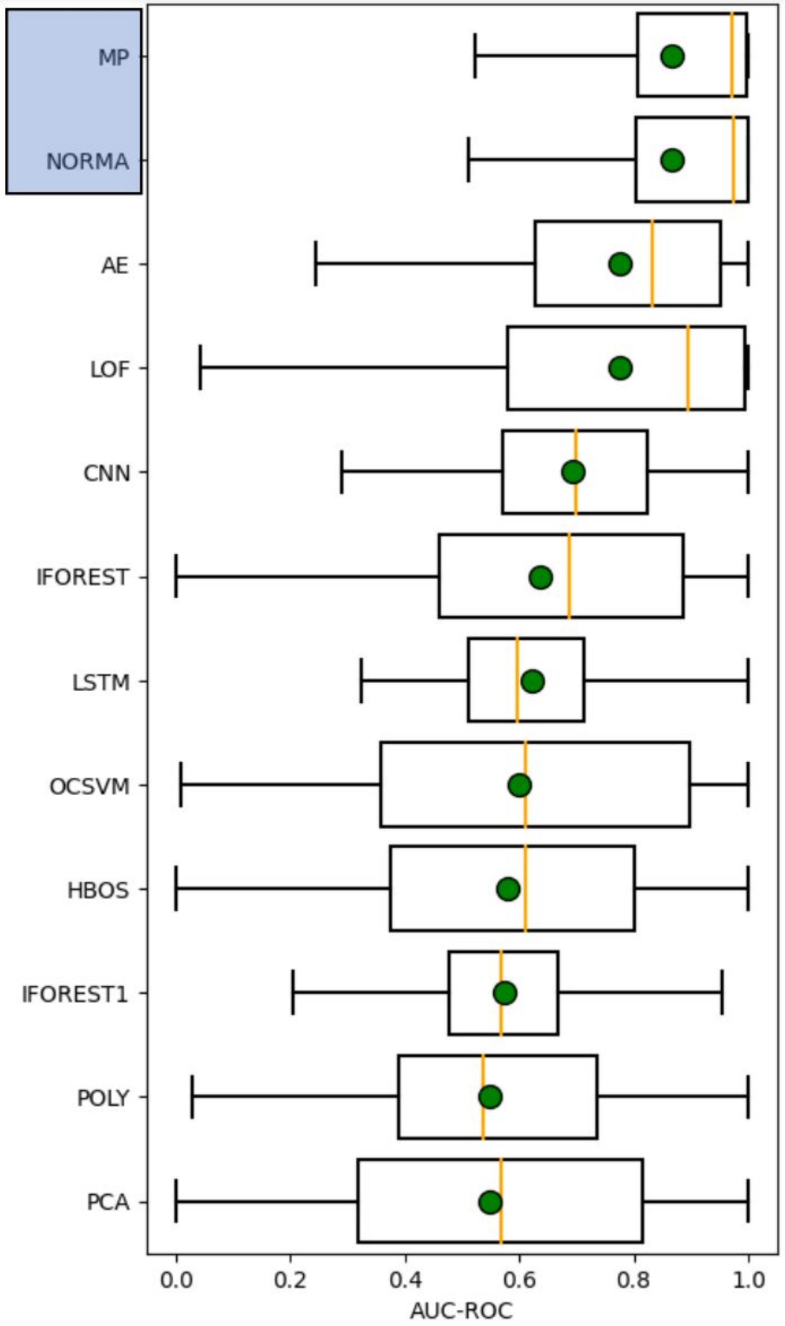
976 time series

- Uni- & multivariate
- Only time series with contamination <0.1, at least one method above 0.8 ROC-AUC and at least one anomaly
- Synthetic time series generator GutenTAG for parameter tuning

All datasets are available in aeon!

HEX/UCR Insights

- Distance methods have better accuracy than forecasting or reconstruction methods



Renjie Wu and Eamonn J. Keogh. 2020. Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress. IEEE Transactions on Knwoledge & Data Engineering, vol. 35, no. 04, 2023.

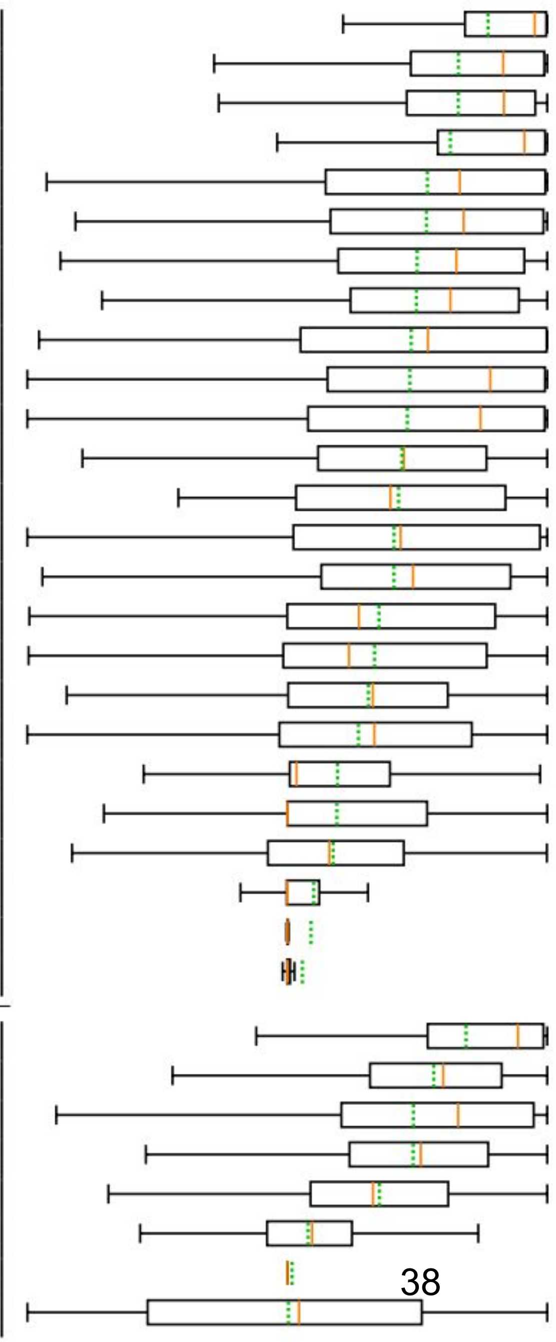
Image from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

TimeEval Insights

- ● Distance methods have better accuracy than ★ forecasting or ■ reconstruction methods
- Semi-supervised are NOT outperforming unsupervised
- **There is no overall winner!**

Dim.	Learn.	Algorithm	TL	OOM	ERR
UNIVARIATE	UNSUPERVISED	● Sub-LOF [22]	2 %	0 %	0 %
		▼ GrammarViz [120]	3 %	0 %	0 %
		▲ DWT-MLEAD [134]	0 %	0 %	0 %
		● VALMOD [82]	1 %	9 %	11 %
		● SAND [17]	5 %	1 %	22 %
		● Left STAMPi [156]	2 %	0 %	1 %
		▼ Series2Graph [16]	0 %	0 %	5 %
		★ ARIMA [65]	7 %	0 %	0 %
		■ PCI [157]	0 %	0 %	0 %
		● STOMP [164]	2 %	0 %	0 %
		● STAMP [156]	4 %	0 %	0 %
		★ Triple ES [1]	15 %	0 %	9 %
		★ NumentaHTM [3]	0 %	0 %	0 %
		● NormA-SJ [15]	10 %	1 %	3 %
		✚ Sub-IF [83]	0 %	0 %	0 %
		★ MedianMethod [10]	0 %	0 %	0 %
		■ SR [112]	0 %	0 %	0 %
		● PS-SVM [85]	12 %	0 %	0 %
		✚ PST [128]	0 %	4 %	0 %
		● SSA [155]	1 %	0 %	0 %
		● HOT SAX [70]	24 %	1 %	1 %
		▼ TSBitmap [144]	0 %	0 %	0 %
		▲ DSPOT [122]	6 %	0 %	0 %
		■ FFT [111]	0 %	0 %	0 %
		▲ S-H-ESD [62]	0 %	0 %	49 %
	SEMI-SUPERVISED	■ Donut [150]	1 %	1 %	2 %
		★ RForest [21]	12 %	0 %	0 %
		■ IE-CAE [44]	0 %	0 %	1 %
		★ XGBoosting [34]	0 %	0 %	0 %
		★ OceanWNN [143]	0 %	0 %	10 %
		■ Bagel [79]	19 %	0 %	2 %
		■ SR-CNN [112]	22 %	0 %	1 %
		▼ TARZAN [71]	0 %	0 %	18 %

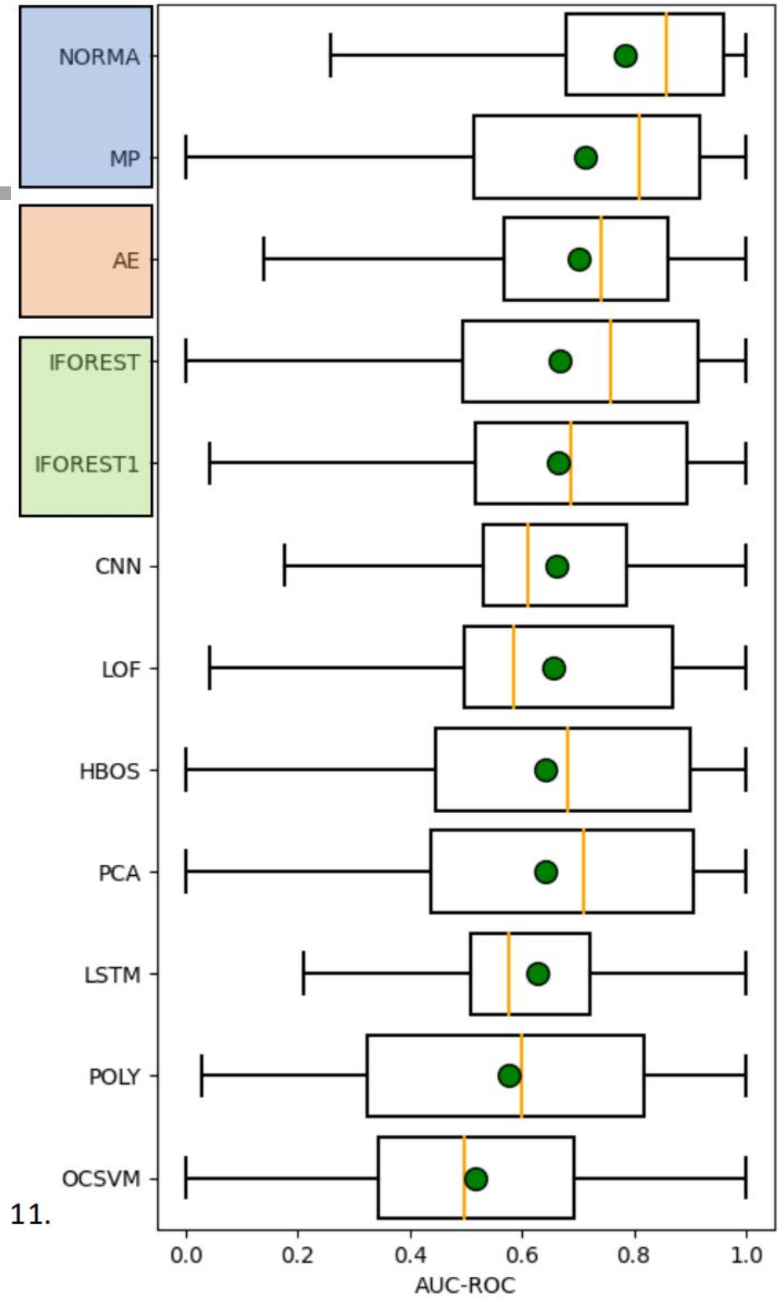
AUC-ROC all datasets



Sebastian Schmidl, Phillip Wenig, Thorsten Papenbrock: Anomaly Detection in Time Series: A Comprehensive Evaluation. PVLDB 9:(15), 2022. DOI:10.14778/3538598.3538602

TSB-UAD Insights

- Distance methods have better accuracy than forecasting methods
- Isolation Forest is very competitive

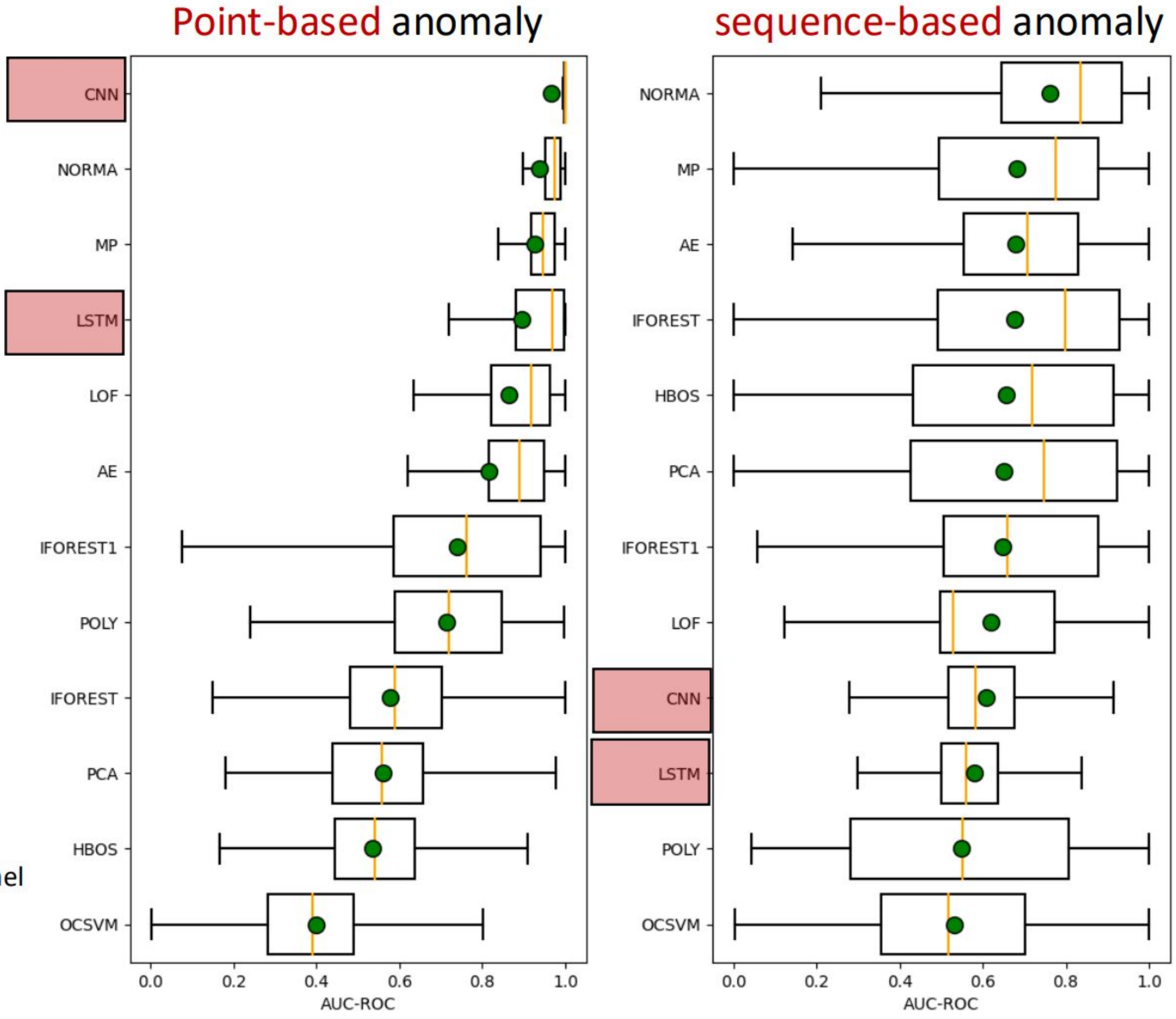


John Paparrizos, Yuhao Kang, Paul Boniol, Ruey S. Tsay, Themis Palpanas, and Michael J. Franklin. 2022. TSB-UAD: an end-to-end benchmark suite for univariate time-series anomaly detection. Proc. VLDB Endow. 15, 8 (2022).

Image from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

TSB-UAD Insights

- Distance methods have better accuracy than forecasting methods
- Isolation Forest is very competitive
- Forecasting methods are good at detecting point anomalies, but bad at detecting subsequence anomalies
- **There is no overall winner!**



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 (2022).

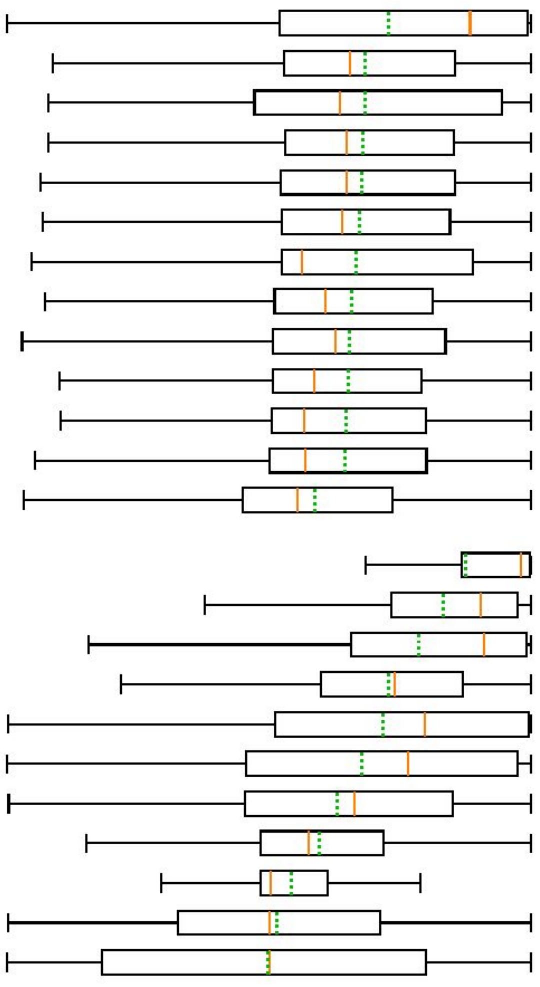
Image from Boniol et al. Time-Series Anomaly Detection: Overview and New Trends. VLDB 2024

Multivariate: TimeEval Insights

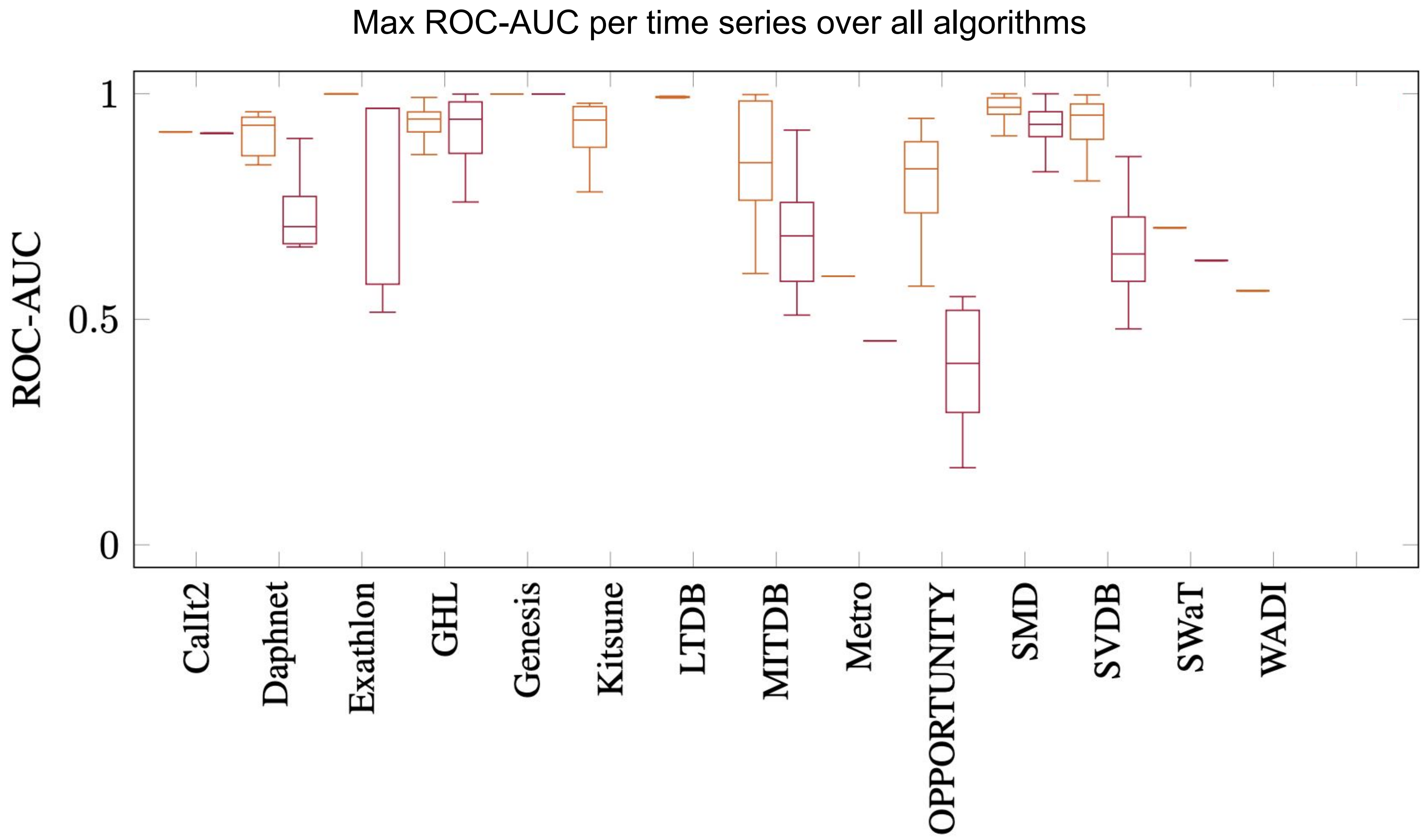
- k-Means and Telemanom stand out
- Overall bad accuracy
- Semi-supervised often exceed resource limits, but are outperforming unsupervised slightly
- **There is no overall winner!**

Dim.	Learn.	Algorithm	TL	OOM	ERR
MULTIVARIATE	UNSUPERVISED	k-Means [151]	0 %	4 %	1 %
		KNN [110]	0 %	0 %	0 %
		Torsk [60]	7 %	0 %	0 %
		EIF [58]	0 %	0 %	0 %
		iForest [83]	0 %	0 %	0 %
		HBOS [47]	0 %	0 %	0 %
		DBStream [55]	0 %	2 %	78 %
		CBLOF [59]	0 %	0 %	0 %
		COPOD [80]	0 %	0 %	0 %
		IF-LOF [36]	0 %	0 %	3 %
		LOF [22]	0 %	0 %	0 %
		COF [130]	0 %	24 %	0 %
		PCC [121]	0 %	0 %	0 %
	SEMI-SUPERVISED	LSTM-AD [89]	14 %	50 %	3 %
		HealthESN [32]	66 %	0 %	0 %
		Telemanom [64]	0 %	0 %	0 %
		RBForest [165]	6 %	5 %	0 %
		EncDec-AD [88]	55 %	26 %	3 %
		DeepAnT [94]	0 %	0 %	5 %
		OmniAnomaly [125]	0 %	4 %	2 %
		LaserDBN [100]	4 %	0 %	7 %
		RobustPCA [101]	0 %	0 %	0 %
		TAnoGan [8]	65 %	0 %	1 %
		Hybrid KNN [124]	0 %	1 %	1 %

AUC-ROC all datasets

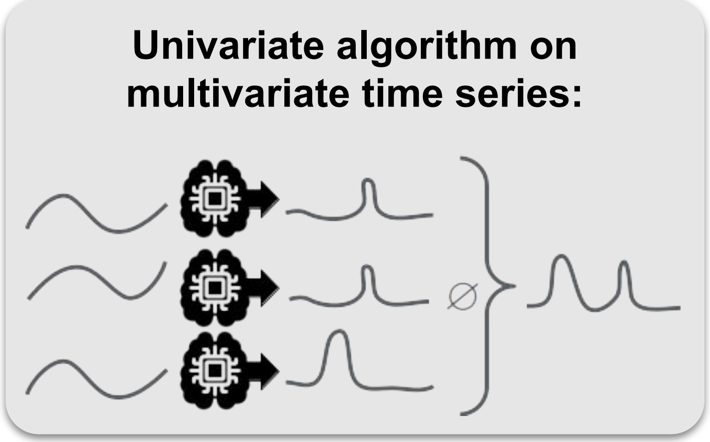


Multivariate: Datasets are too easy



Univariate

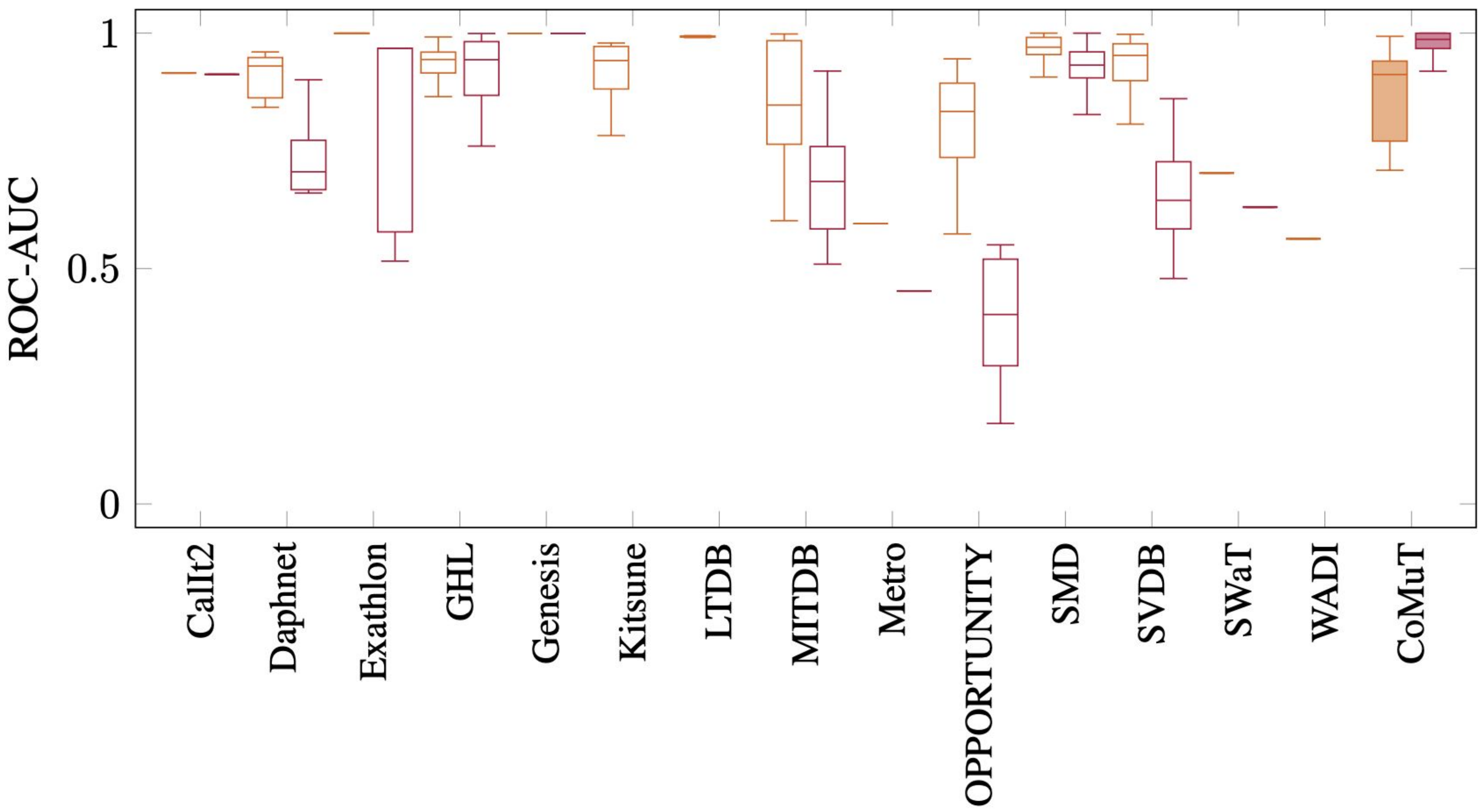
Multivariate



Phillip Wenig, Sebastian Schmidl, Thorsten Papenbrock: Anomaly Detectors for Multivariate Time Series: The Proof of the Pudding is in the Eating. Proceedings of the International Conference on Data Engineering Workshops (ICDEW), 2024. DOI:10.1109/ICDEW61823.2024.00018

Multivariate: Datasets are too easy

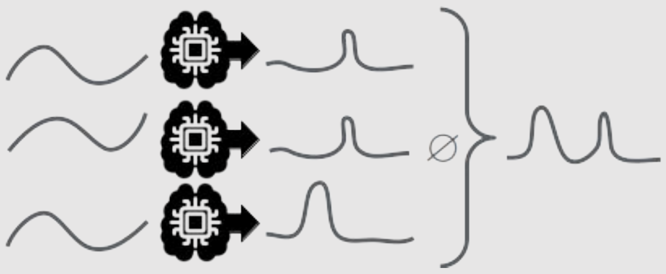
Max ROC-AUC per time series over all algorithms



Univariate

Multivariate

Univariate algorithm on multivariate time series:

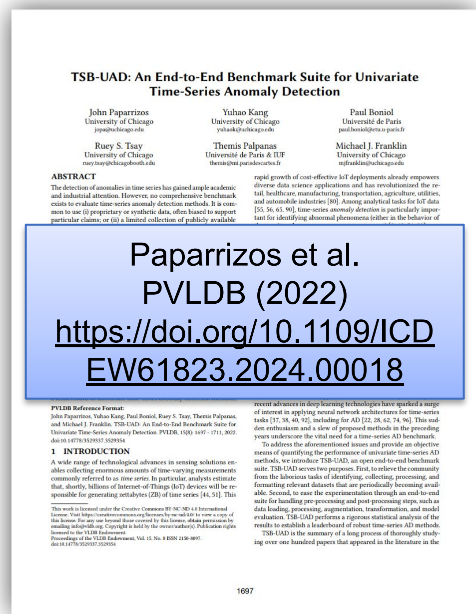


Phillip Wenig, Sebastian Schmidl, Thorsten Papenbrock: Anomaly Detectors for Multivariate Time Series: The Proof of the Pudding is in the Eating. Proceedings of the International Conference on Data Engineering Workshops (ICDEW), 2024. DOI:10.1109/ICDEW61823.2024.00018

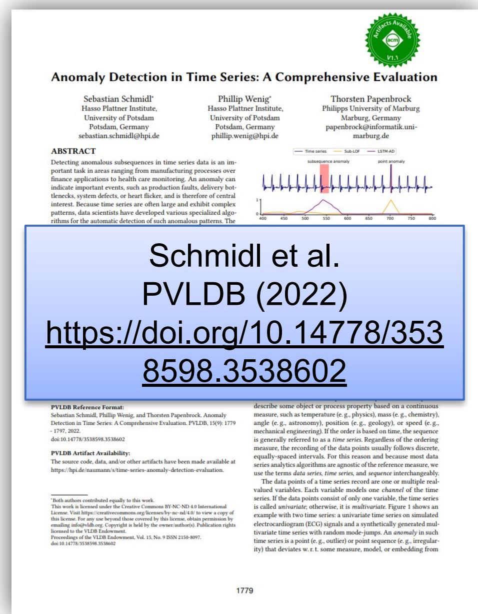
Outlook

Recommended Reads

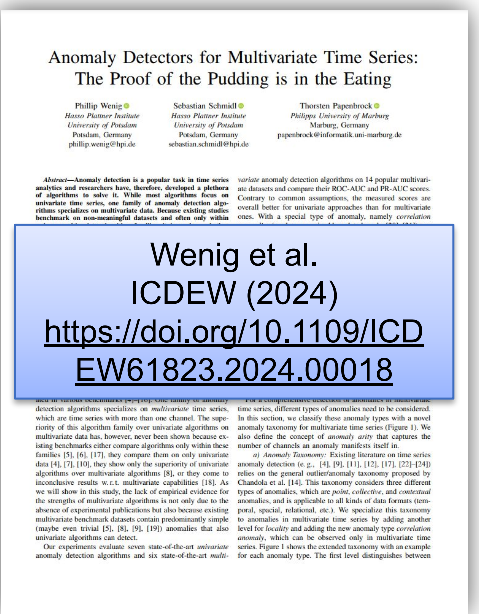
Evaluation and Survey Papers



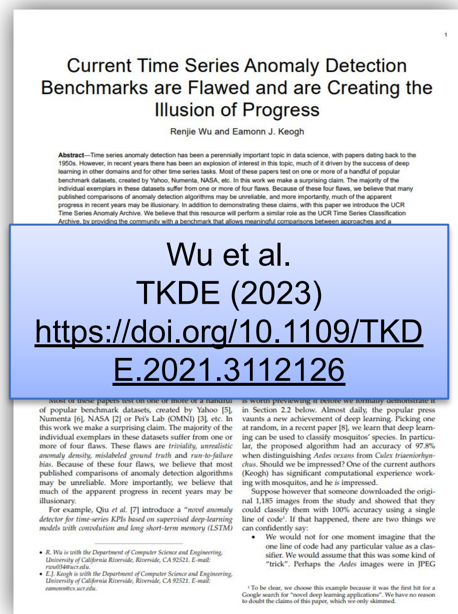
Paparrizos et al.
PVLDB (2022)
https://doi.org/10.1109/ICDEW61823.2024.00018



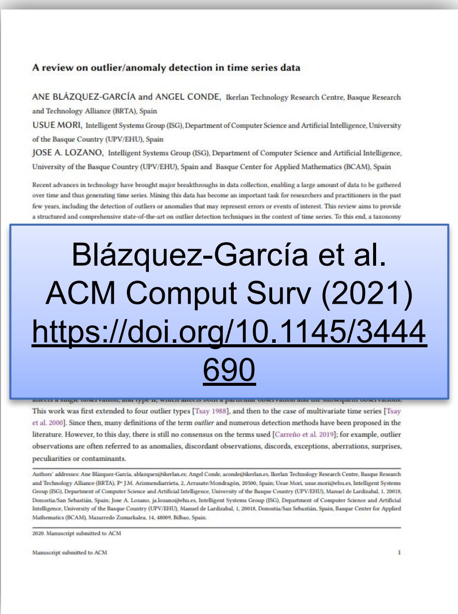
Schmidl et al.
PVLDB (2022)
https://doi.org/10.14778/3538598.3538602



Wenig et al.
ICDEW (2024)
https://doi.org/10.1109/ICDEW61823.2024.00018



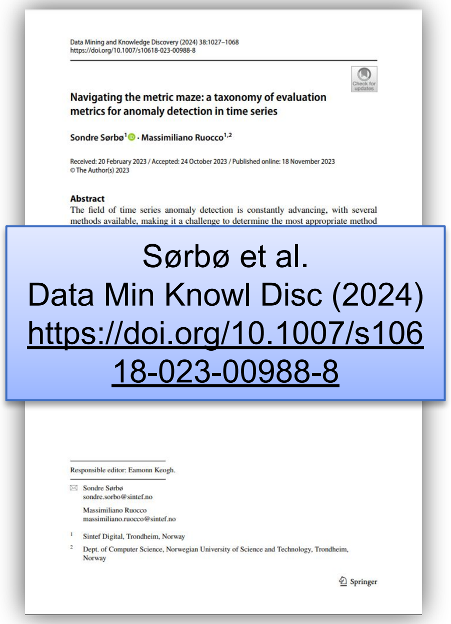
Wu et al.
TKDE (2023)
https://doi.org/10.1109/TKDE.2021.3112126



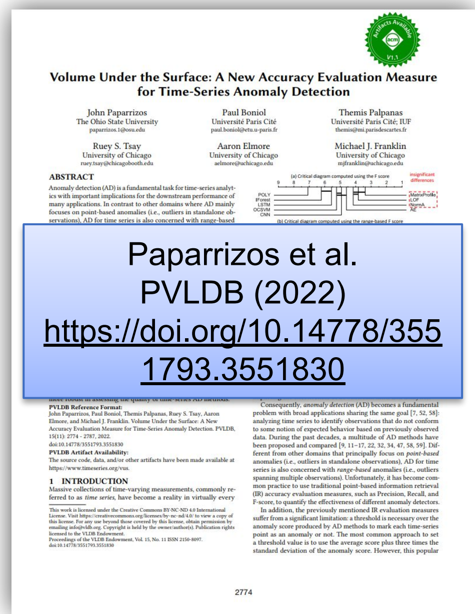
Blázquez-García et al.
ACM Comput Surv (2021)
https://doi.org/10.1145/3444690

Recommended Reads

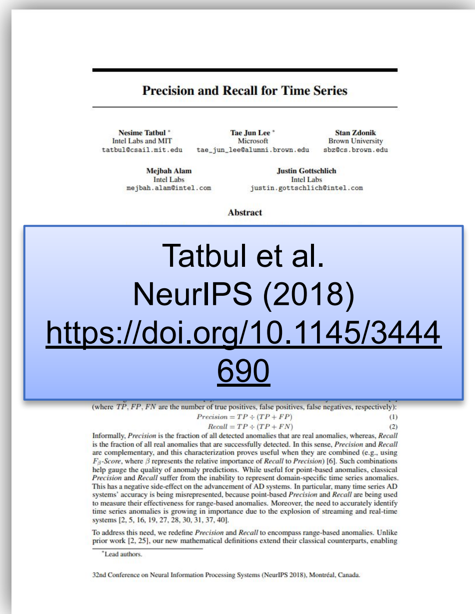
Metric Papers



Sørbø et al.
Data Min Knowl Disc (2024)
https://doi.org/10.1007/s10618-023-00988-8



Paparrizos et al.
PVLDB (2022)
https://doi.org/10.14778/3551793.3551830



Tatbul et al.
NeurIPS (2018)
https://doi.org/10.1145/3444690

Huet et al.
KDD (2022)
https://doi.org/10.1145/3534678.3539339

Future Work

For Research

- Automatic method selection and ensembling methods
  - No one-size-fits-all solution, but already 100s of methods
- More and better benchmarks
  - Existing benchmark datasets have flaws
  - Existing evaluations are biased
- TS Foundation Models
- Explainability / Precursor Detection / Anomaly Classification

In aeon

- Add more detectors (especially deep learning)
- Add missing metrics

Example code

See part6_anomaly_detection.ipynb

Thank you!

NEXT: Deep Learning by Ali Ismail-Fawaz

Datasets repositories