

# An Interactive Dive into Time-Series Anomaly Detection

Paul Boniol

Inria, ENS, PSL University, CNRS  
boniol.paul@gmail.com

John Paparrizos

The Ohio State University  
paparrizos.1@osu.edu

Themis Palpanas

Université Paris Cité & IUF  
themis@mi.parisdescartes.fr

**Abstract**—Anomaly detection is an important problem in data analytics with applications in many domains. In recent years, there has been an increasing interest in anomaly detection tasks applied to time series. In this tutorial, we take a holistic view of anomaly detection in time series, starting from the core definitions and taxonomies related to time series and anomaly types, to an extensive description of the anomaly detection methods proposed by different communities in the literature. We explore the literature and the proposed methods by demonstrating systems that help users understand the core computational steps of some methods and navigate benchmark results. Finally, we describe the problem of model selection for anomaly detection and discuss recent experimental results.

## I. INTRODUCTION

A wide range of sensing, networking, and processing solutions enable the collection of large amounts of data over time [1], [2], [3]. The latter results in ordered sequences of real-valued data points commonly referred to as *time series*. Analytical tasks over time series data are becoming increasingly important in virtually every domain [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], including astronomy [20], energy [21], environmental [22], and social [23] sciences.

Anomaly detection has received ample academic and industrial attention [24], [25]. Moreover, as illustrated in Figure 1, anomaly detection applied to time series (compared to other data types) is attracting more interest lately. As commonly defined in the literature [26], *anomalies* refer to data points (single points or group of points) that do not conform to some notion of normality or an expected behavior based on previously observed data. In practice, anomalies can correspond to erroneous data (e.g., broken sensors) or data of interest (e.g., anomalous behavior of the measured system) [27]. Detecting such cases is crucial for many applications [28].

In recent years, time-series anomaly detection is becoming increasingly popular. In particular, multiple surveys, benchmarks, and experimental studies summarize and analyze the state-of-the-art methods [29], [30], [31], [32], [33], [34], exploring different aspects of the problem. Based on these recent works, this tutorial provides a comprehensive view of the anomaly detection task in time series.

We start from definitions for time series, anomalies, and method categories to an interactive dive into demonstration systems. Our goal is four-fold: (i) introduce the motivation

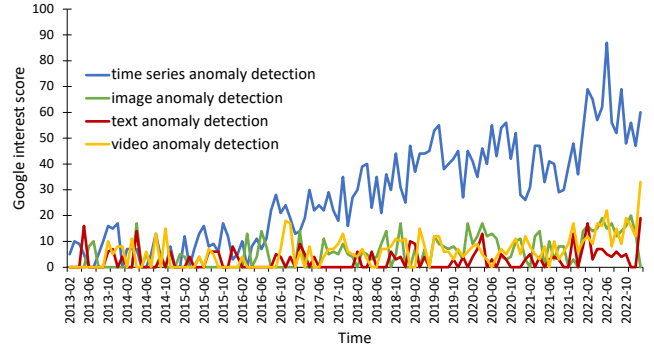


Fig. 1. Evolution of the Google interest score for anomaly detection over time series, images, text, and video.

related to the anomaly detection task in time series by describing a taxonomy of time series and anomaly types usually considered in the literature; (ii) describe recently proposed benchmarks and experimental evaluations and discuss the challenges and problems in the evaluation itself; (iii) demonstrate interactive systems for exploring a methods' computation steps and experimental evaluation; and (iv) discuss the challenges and opportunities that (automatic) model selection brings to time series anomaly detection.

**[Relation to Previous Tutorials]** Recent tutorials related to anomaly detection have focused (i) on specific types of methods (such as deep-learning methods [35]) ignoring several of the state-of-the-art methods, (ii) on specific types of temporal data (such as spatiotemporal data [36]), or (iii) on a much more general topic (only briefly discussing anomaly detection as a subpart of time series analysis [37]). An earlier version of this tutorial [38] focused on a description of the state-of-the-art time series anomaly detection methods, offering details on the way they operate. In contrast, this tutorial briefly presents a taxonomy of the methods proposed in the literature, and then focuses on the evaluation of methods, on prototype systems that enable an interactive inspection of methods and their inner-workings, and on the new opportunities that automatic model selection brings to the field. To our knowledge, this is the first tutorial that (i) discusses anomaly detection through the lens of interactivity and demonstration systems, and (ii) addresses model selection opportunities for the specific case of time series anomaly detection.

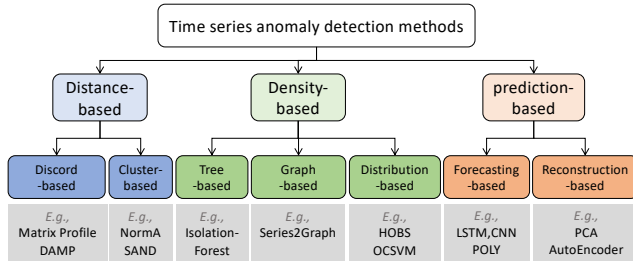


Fig. 2. Taxonomy of anomaly detection methods.

**[Audience and Expected Background]** This tutorial is for researchers and data analysts and will focus on recent advances in time series anomaly detection. The tutorial aims to initiate new collaborations between members of the data management community and data science practitioners in various domains and increase the interest in anomaly detection for time series. We will present material that includes the necessary background to follow the entire presentation and technical details of the current state-of-the-art solutions for anomaly detection. We will also discuss the drawbacks of existing solutions and the open problems. Both experts and newcomers in the area will be able to follow the material and benefit from the tutorial.

## II. TUTORIAL SCOPE

In this **90 minutes lecture-style tutorial**, we will go through the problem of anomaly detection in time series, starting from fundamental definitions of time series and anomaly to open problems and opportunities that model selection techniques can bring both in terms of accuracy and performance.

### A. Introduction, Motivation and Foundations

We will start by discussing examples of scientific and industrial applications that rely on time-series anomaly detection.

**[Type of Time Series]** We first introduce the different types of time series. Specifically, we define a time series as an ordered sequence of real values on one (for *univariate* time series) or multiple dimensions (for *multivariate* time series). Moreover, we define *static* and *streaming* time series as sequences with a fixed length or continuously arriving subsequences. Finally, the normal behavior (i.e., subsequences representing the normal and recurrent behavior) might change over time. In this case, we differentiate *single* and *multiple normalities* time series.

**[Type of Anomalies]** We then introduce the different types of anomalies. The two first categories, *point* and *contextual* anomalies, refer to data points deviating remarkably from the rest of the data globally or given a specific context, respectively. With *collective* anomalies, we refer to sequences of points that do not repeat a typical (previously observed) pattern. Then, on top of these categories, their combination also matters. First, we need to differentiate time series containing *single* anomalies from time series containing **multiple** anomalies (with *similar* or *different* anomalies). For all these definitions and taxonomies, we will provide explicit examples.

### B. Taxonomy of Anomaly Detection Methods

We will then dive into anomaly detection methods proposed in the literature. As many papers appear every year proposing new methods for anomaly detection in time series based on different applications, it is beyond our scope to cover all proposed methods extensively here. In this tutorial, we will briefly summarize popular categories of methods, and we refer the attendees to three recent survey papers for detailed coverage of methods [29], [39], [40].

**[Classification by Inputs]** We will first mention the three categories of methods based on the external knowledge provided to them. First, *unsupervised* methods take time series as input and are not provided by any other information. Then, *semi-supervised* methods take as input time series without any anomalies and are trained on normal data only. Finally, *supervised* methods take as input separately both normal and abnormal data. Thus, the model is trained to discriminate the anomalies from the normality.

**[Classification by methodologies]** Then, we will describe the following categories of methods (cf. Figure 2). First, *distance-based* approaches analyze subsequences by utilizing distances to a given model to detect anomalies. For instance, discord-based approaches use the nearest neighbor distances among subsequences [41], [42], [43], [44], [45], [46], [47], [48]. As another example, recent methods in this category first cluster data to obtain the normal behavior and compute the distance to the clusters to detect anomalies [49], [50], [51]. Second, *density-based* methods focus on detecting recurring or isolated behaviors by evaluating the density of the points or subsequences space [52], [53], [54]. Third, forecasting-based methods, such as recurrent or convolutional [55] neural network-based [56], have been proposed for this task. Such methods use the past values as input, predict the following, and use the forecasting error as an anomaly score. Last, reconstruction-based methods, such as the AutoEncoder-based approach [57], are trained to reconstruct the time series and use the reconstruction error as an anomaly score. We group these two aforementioned categories into *prediction-based methods*.

### C. Evaluating Anomaly Detection Methods

After describing the methods proposed in the literature, we will focus on how to evaluate them. The choice of benchmarks and accuracy measures may significantly bias the evaluation.

**[Evaluation Measures]** Thus, we will start by describing the evaluation measures. Briefly, we will first discuss traditional measures, such as Precision, Recall, and F-score, that assess the methods by assuming each time-series point can be marked as an anomaly or not (e.g., by a threshold on an anomaly score). We will then discuss range-based variants [63] that aim to overcome shortcomings of traditional measures when evaluating time series containing subsequence anomalies. We will discuss AUC (Area Under the Curve) measures that, contrary to the previous measures, eliminate the need to define a threshold. We will finally discuss VUS [32] (Volume Under the Surface) measures, which provide more robustness.

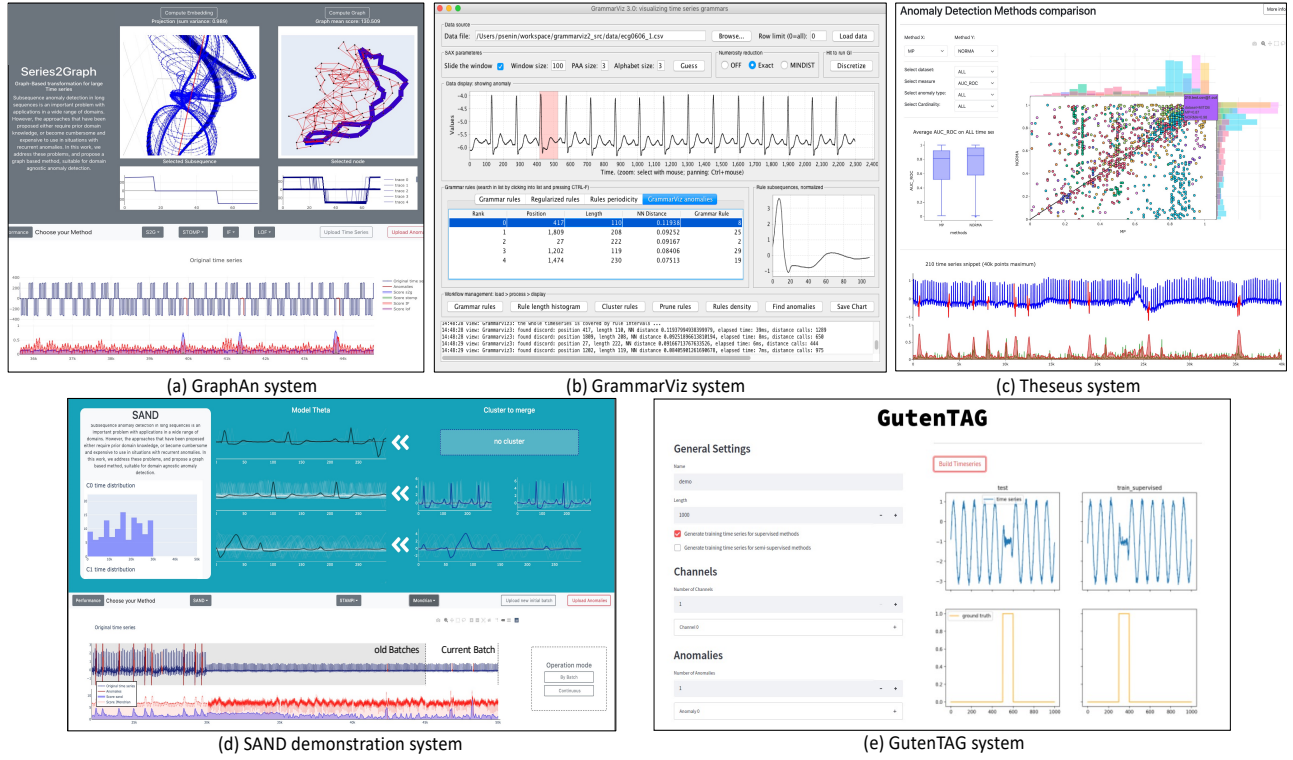


Fig. 3. Example of interactive systems that (a,b,d) allow the user to dive into computational steps [58], [59], [60], or (c,e) experimental results [61], [62].

**[Experimental Evaluations]** Then, we will discuss recent benchmarks proposed for anomaly detection in time series task [33], [31], [32], [64], [30]. Such benchmarks provide an extensive collection of time series from various domains and evaluate multiple methods belonging to the aforementioned categories. In this tutorial, we will discuss the results and conclusions of these recent benchmarks, as well as the criticisms that have been expressed about the characteristics and suitability of some datasets for this task [65], [30].

#### D. New Perspectives for Anomaly Detection

In the last part of the tutorial, we will discuss the challenges and perspectives for improving interpretability and accuracy.

**[Interpretability for the User]** In contrast to the previous edition that focuses only on a lecture-style description of the anomaly detection methods [38], this tutorial employs interactive exploration to assist in understanding the concepts. Such interactive exploration is important for users to overcome the lack of interpretability that anomaly detection methods can have [56], [57]. This is becoming possible by recently proposed prototype systems [59], [60], which enable users to interactively explore methods and their inner-workings, and thus, better understand the different computation steps. In this tutorial, we will discuss and demonstrate recently proposed systems for anomaly detection [58], [60], [59], as well as systems that allow the user to navigate and explore large experimental evaluation studies [62], [61]. Figure 3 illustrates examples of these systems. We will use such systems during

the tutorial, in order to explain interactively the anomaly detection methods introduced in Section II-B, and the experimental conclusions described in Section II-C.

**[Model Selection for Time Series Anomaly Detection]** Then, recent benchmarks and evaluation studies (described in the previous section) demonstrated that no overall best anomaly detection methods exist when applied on very heterogeneous time series (i.e., coming from very different domains). This tutorial will discuss possible solutions to overcome this accuracy limitation. We will first discuss ensembling solutions [66], but also model selection methods that select the best anomaly detection method to run based on time series characteristics. We will first describe recent research works related to AutoML (Automatic Machine Learning) for the general case of anomaly detection [67], and also for time series [68]. We will finally discuss the different strategies to apply model selection methods for the specific case of time series anomaly detection.

#### E. Challenges and Conclusions

We will conclude this tutorial by summarizing the main insights obtained from recent benchmarks on the performances anomaly detection methods [30] and the proper evaluation measures [32]. Finally, we will summarize the new challenges that model selection can bring to the field and elaborate on research directions that could solve these new open problems.

### III. PRESENTERS

**Paul Boniol** is a researcher at Inria, member of the VALDA project-team. Previously, he worked at ENS Paris-Saclay (Centre Borelli), Université Paris Cité, EDF Research lab, and Ecole Polytechnique (LIX). His research interests lie between data analytics, machine learning, and time-series analysis. His Ph.D. dissertation focused on subsequence anomaly detection and time-series classification. His work has been published in the top data management and analytics venues.

**John Paparrizos** is an assistant professor at The Ohio State University, leading The DATUM Lab. His research focuses on adaptive solutions for data-intensive and machine-learning applications. His doctoral work was recognized at the 2019 ACM SIGKDD Doctoral Dissertation Award competition. He has also received the inaugural ACM SIGMOD Research Highlight Award, a NetApp Faculty Award, and the 2023 IEEE TCDE Rising Star Award. His ideas have been widely adopted across scientific areas, Fortune 100-500 companies (e.g., Exelon and Nokia), and organizations such as ESA.

**Themis Palpanas** is an elected Senior Member of the French University Institute (IUF), and Distinguished Professor of computer science at the University of Paris (France). He is the author of 14 patents, has received 3 best paper awards and the IBM SUR award, has been Program Chair for VLDB 2025 and IEEE BigData 2023, General Chair for VLDB 2013, and has served Editor in Chief for BDR. He has been working in the fields of Data Series Management and Analytics for more than 15 years, and has developed several of the state of the art techniques. He has delivered 19 tutorials in top conferences.

### REFERENCES

- [1] T. Palpanas and V. Beckmann, "Report on the first and second interdisciplinary time series analysis workshop (ITISA)," *SIGMOD Rec.*, vol. 48, no. 3, pp. 36–40, 2019.
- [2] J. Paparrizos, C. Liu, B. Barbarioli, J. Hwang, I. Edian, A. J. Elmore, M. J. Franklin, and S. Krishnan, "Vergedb: A database for iot analytics on edge devices," in *CIDR*, 2021.
- [3] S. Liu, T. Mangla, T. Shaowang, J. Zhao, J. Paparrizos, S. Krishnan, and N. Feamster, "Amir: Active multimodal interaction recognition from video and network traffic in connected environments," *IMWUT*, vol. 7, no. 1, pp. 1–26, 2023.
- [4] K. McKeown, H. Daume III, S. Chaturvedi, J. Paparrizos, K. Thadani, P. Barrio, O. Biran, S. Bothe, M. Collins, K. R. Fleischmann *et al.*, "Predicting the impact of scientific concepts using full-text features," *Journal of the Association for Information Science and Technology*, vol. 67, no. 11, pp. 2684–2696, 2016.
- [5] J. Paparrizos, R. W. White, and E. Horvitz, "Screening for pancreatic adenocarcinoma using signals from web search logs: Feasibility study and results," *Journal of oncology practice*, vol. 12, no. 8, pp. 737–744, 2016.
- [6] R. Goel, S. Soni, N. Goyal, J. Paparrizos, H. Wallach, F. Diaz, and J. Eisenstein, "The social dynamics of language change in online networks," in *Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I* 8. Springer, 2016, pp. 41–57.
- [7] M. Bariya, A. von Meier, J. Paparrizos, and M. J. Franklin, "k-shapestream: Probabilistic streaming clustering for electric grid events," in *2021 IEEE Madrid PowerTech*. IEEE, 2021, pp. 1–6.
- [8] A. Dziedzic, J. Paparrizos, S. Krishnan, A. Elmore, and M. Franklin, "Band-limited training and inference for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019, pp. 1745–1754.
- [9] C. Liu, H. Jiang, J. Paparrizos, and A. J. Elmore, "Decomposed bounded floats for fast compression and queries," *Proc. VLDB Endow.*, vol. 14, no. 11, pp. 2586–2598, 2021.
- [10] H. Jiang, C. Liu, Q. Jin, J. Paparrizos, and A. J. Elmore, "Pids: attribute decomposition for improved compression and query performance in columnar storage," *Proc. VLDB Endow.*, vol. 13, no. 6, pp. 925–938, 2020.
- [11] H. Jiang, C. Liu, J. Paparrizos, A. A. Chien, J. Ma, and A. J. Elmore, "Good to the last bit: Data-driven encoding with codecdb," in *SIGMOD*, 2021, pp. 843–856.
- [12] J. Paparrizos, C. Liu, A. J. Elmore, and M. J. Franklin, "Debunking four long-standing misconceptions of time-series distance measures," in *SIGMOD*, 2020, pp. 1887–1905.
- [13] J. Paparrizos and M. J. Franklin, "Grail: efficient time-series representation learning," *Proc. VLDB Endow.*, vol. 12, no. 11, 2019.
- [14] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *SIGMOD*, 2015, pp. 1855–1870.
- [15] J. Paparrizos, I. Edian, C. Liu, A. J. Elmore, and M. J. Franklin, "Fast adaptive similarity search through variance-aware quantization," in *ICDE*. IEEE, 2022, pp. 2969–2983.
- [16] J. Paparrizos and L. Gravano, "Fast and accurate time-series clustering," *TODS*, vol. 42, no. 2, 2017.
- [17] I. Paparrizos, "Fast, scalable, and accurate algorithms for time-series analysis," Ph.D. dissertation, Columbia University, 2018.
- [18] J. Paparrizos, K. Wu, A. Elmore, C. Faloutsos, and M. J. Franklin, "Accelerating similarity search for elastic measures: A study and new generalization of lower bounding distances," *Proceedings of the VLDB Endowment*, vol. 16, no. 8, pp. 2019–2032, 2023.
- [19] J. Paparrizos, C. Liu, A. J. Elmore, and M. J. Franklin, "Querying time-series data: A comprehensive comparison of distance measures," *Data Engineering*, p. 69.
- [20] P. Huijse, P. A. Estevez, P. Protopapas, J. C. Principe, and P. Zegers, "Computational intelligence challenges and applications on large-scale astronomical time series databases," *IEEE Computational Intelligence Magazine*, vol. 9, no. 3, pp. 27–39, 2014.
- [21] M. Bach-Andersen, B. Rømer-Odgaard, and O. Winther, "Flexible non-linear predictive models for large-scale wind turbine diagnostics," *Wind Energy*, vol. 20, no. 5, pp. 753–764, 2017.
- [22] S. Goddard, S. K. Harms, S. E. Reichenbach, T. Tadesse, and W. J. Waltman, "Geospatial decision support for drought risk management," *Communications of the ACM*, vol. 46, no. 1, pp. 35–37, 2003.
- [23] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 2016.
- [24] E. Page, "On problems in which a change in a parameter occurs at an unknown point," *Biometrika*, vol. 44, no. 1/2, pp. 248–252, 1957.
- [25] A. J. Fox, "Outliers in time series," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 3, pp. 350–363, 1972.
- [26] D. M. Hawkins, *Identification of outliers*, 1980, vol. 11.
- [27] C. C. Aggarwal, "An introduction to outlier analysis," in *Outlier analysis*, 2017, pp. 1–34.
- [28] M. Hadjem, F. Naït-Abdesselam, and A. A. Khokhar, "St-segment and t-wave anomalies prediction in an ECG data using rusboost," in *Healthcom*, 2016.
- [29] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM CSUR*, vol. 54, no. 3, 2021.
- [30] J. Paparrizos, Y. Kang, P. Boniol, R. S. Tsay, T. Palpanas, and M. J. Franklin, "Tsb-uad: An end-to-end benchmark suite for univariate time-series anomaly detection," *Proc. VLDB Endow.*, vol. 15, no. 8, 2022.
- [31] S. Schmidl, P. Wenig, and T. Papenbrock, "Anomaly detection in time series: A comprehensive evaluation," *Proc. VLDB Endow.*, vol. 15, no. 9, p. 1779–1797, 2022.
- [32] 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.*, vol. 15, no. 11, 2022.
- [33] V. Jacob, F. Song, A. Stiegler, B. Rad, Y. Diao, and N. Tatbul, "Exathlon: A benchmark for explainable anomaly detection over time series," *Proc. VLDB Endow.*, vol. 14, no. 11, p. 2613–2626, oct 2021.
- [34] E. Sylligardos, P. Boniol, J. Paparrizos, P. Trahanias, and T. Palpanas, "Choose wisely: An extensive evaluation of model selection for anomaly detection in time series," *Proceedings of the VLDB Endowment*, vol. 16, no. 11, pp. 3418–3432, 2023.
- [35] R. Chalapathy, N. L. D. Khoa, and S. Chawla, "Robust deep learning methods for anomaly detection," in *SIGKDD*, ser. KDD '20, 2020.

- [36] G. Yang, N. Kulkarni, P. Dua, D. Khullar, and A. A. Chirayath, "Anomaly detection for spatiotemporal data in action," in *KDD*, 2022.
- [37] Q. Wen, L. Yang, T. Zhou, and L. Sun, "Robust time series analysis and applications: An industrial perspective," in *SIGKDD*, 2022.
- [38] P. Boniol, J. Paparrizos, and T. Palpanas, "New trends in time-series anomaly detection," in *EDBT*, 2023.
- [39] M. Braei and S. Wagner, "Anomaly detection in univariate time-series: A survey on the state-of-the-art," *arXiv:2004.00433*, 2020.
- [40] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: A survey," *TKDE*, vol. 26, no. 9, pp. 2250–2267, 2013.
- [41] C. M. Yeh, Y. Zhu, L. Ulanova, N. Begum, Y. Ding, H. A. Dau, D. F. Silva, A. Mueen, and E. J. Keogh, "Matrix profile I: all pairs similarity joins for time series," in *ICDM*, 2016.
- [42] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein, "Time series anomaly discovery with grammar-based compression," in *EDBT*, 2015.
- [43] E. Keogh, S. Lonardi, C. A. Ratanamahatana, L. Wei, S.-H. Lee, and J. Handley, "Compression-based data mining of sequential data," *DAMI*, 2007.
- [44] Y. Liu, X. Chen, and F. Wang, "Efficient Detection of Discords for Time Series Stream," *Advances in Data and Web Management*, 2009.
- [45] A. W. Fu, O. T. Leung, E. J. Keogh, and J. Lin, "Finding time series discords based on haar transform," in *ADMA*, 2006.
- [46] Y. Bu, O. T. Leung, A. W. Fu, E. J. Keogh, J. Pei, and S. Meshkin, "WAT: finding top-k discords in time series database," in *SIAM*, 2007.
- [47] M. Linardi, Y. Zhu, T. Palpanas, and E. J. Keogh, "Matrix Profile Goes MAD: Variable-Length Motif And Discord Discovery in Data Series," in *DAMI*, 2020.
- [48] Y. Lu, R. Wu, A. Mueen, M. A. Zuluaga, and E. J. Keogh, "DAMP: accurate time series anomaly detection on trillions of datapoints and ultra-fast arriving data streams," *Data Min. Knowl. Discov.*, vol. 37, no. 2, pp. 627–669, 2023. [Online]. Available: <https://doi.org/10.1007/s10618-022-00911-7>
- [49] P. Boniol, M. Linardi, F. Roncallo, T. Palpanas, M. Meftah, and E. Remy, "Unsupervised and scalable subsequence anomaly detection in large data series," *The VLDB Journal*, vol. 30, no. 6, 2021.
- [50] P. Boniol, J. Paparrizos, T. Palpanas, and M. J. Franklin, "Sand: streaming subsequence anomaly detection," *Proc. VLDB Endow*, vol. 14, no. 10, pp. 1717–1729, 2021.
- [51] P. Boniol and T. Palpanas, "Series2graph: Graph-based subsequence anomaly detection for time series," *Proc. VLDB Endow*, vol. 13, no. 11, 2020.
- [52] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *2008 Eighth IEEE International Conference on Data Mining*, Dec. 2008, pp. 413–422, iSSN: 2374-8486.
- [53] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: Identifying density-based local outliers," in *SIGMOD*, 2000.
- [54] H. Ma, B. Ghogh, M. N. Samad, D. Zheng, and M. Crowley, "Isolation mondrian forest for batch and online anomaly detection," 2020.
- [55] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "Deepant: A deep learning approach for unsupervised anomaly detection in time series," *IEEE Access*, vol. 7, pp. 1991–2005, 2019.
- [56] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," 2015.
- [57] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *MLSDA*, 2014.
- [58] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, S. Frankenstein, and M. Lerner, "Grammarviz 2.0: A tool for grammar-based pattern discovery in time series," in *ECML/PKDD*, 2014.
- [59] P. Boniol, T. Palpanas, M. Meftah, and E. Remy, "Graphan: Graph-based subsequence anomaly detection," *Proc. VLDB Endow.*, vol. 13, no. 12, pp. 2941–2944, 2020.
- [60] P. Boniol, J. Paparrizos, T. Palpanas, and M. J. Franklin, "Sand in action: subsequence anomaly detection for streams," *Proc. VLDB Endow.*, vol. 14, no. 12, 2021.
- [61] P. Boniol, J. Paparrizos, Y. Kang, T. Palpanas, R. S. Tsay, A. J. Elmore, and M. J. Franklin, "Theseus: navigating the labyrinth of time-series anomaly detection," *Proc. VLDB Endow*, vol. 15, no. 12, pp. 3702–3705, 2022.
- [62] P. Wenig, S. Schmidl, and T. Papenbrock, "Timeeval: A benchmarking toolkit for time series anomaly detection algorithms," *Proc. VLDB Endow.*, vol. 15, no. 12, p. 3678–3681, aug 2022.
- [63] N. Tatbul, T. J. Lee, S. Zdonik, M. Alam, and J. Gottschlich, "Precision and Recall for Time Series," in *NeurIPS*, vol. 31, 2018.
- [64] E. Keogh, T. Dutta Roy, U. Naik, and A. Agrawal, "Multi-dataset Time-Series Anomaly Detection Competition 2021," <https://compete.hexagon-ml.com/practice/competition/39/>, 2021.
- [65] R. Wu and E. J. Keogh, "Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 3, pp. 2421–2429, 2023. [Online]. Available: <https://doi.org/10.1109/TKDE.2021.3112126>
- [66] C. C. Aggarwal and S. Sathe, "Theoretical foundations and algorithms for outlier ensembles," *SIGKDD Explor. Newsl.*, vol. 17, no. 1, p. 24–47, sep 2015.
- [67] Y. Zhao, R. Rossi, and L. Akoglu, "Automatic unsupervised outlier model selection," in *NeurIPS*, vol. 34, 2021, pp. 4489–4502.
- [68] Y. Ying, J. Duan, C. Wang, Y. Wang, C. Huang, and B. Xu, "Automated model selection for time-series anomaly detection," 2020.