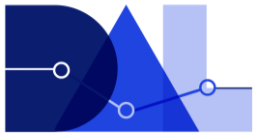
CEFET/RJ



Harbinger

A Unified Framework for Event Detection in Time Series

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
https://eic.cefet-rj.br/~eogasawara

Harbinger: A Unified Framework for Event Detection in Time Series

Harbinger

Harbinger is a framework for event detection in time series. It provides an integrated environment for anomaly detection, change point detection, and motif discovery. Harbinger offers a broad range of methods and functions for plotting and evaluating detected events.

For anomaly detection, methods are based on: - Machine learning model deviation: Conv1D, ELM, MLP, LSTM, Random Regression Forest, and SVM - Classification models: Decision Tree, KNN, MLP, Naive Bayes, Random Forest, and SVM - Clustering: k-means and DTW - Statistical techniques: ARIMA, FBIAD, GARCH

For change point detection, Harbinger includes: - Linear regression, ARIMA, ETS, and GARCH-based approaches - Classic methods such as AMOC, ChowTest, Binary Segmentation (BinSeg), GFT, and PELT

For motif discovery, it provides: - Methods based on Hashing and Matrix Profile

Harbinger also supports **multivariate time series analysis** and **event evaluation** using both traditional and soft computing metrics.

The architecture of Harbinger is based on **Experiment Lines** and is built on top of the DAL Toolbox. This design makes it easy to extend and integrate new methods into the framework.

https://cran.r-project.org/web/packages/harbinger

Why a unified framework?

- Various methods exist, but they are focused on specific types
- The appropriate choice depends on the nature of the time series
- Real-world applications require systematic comparison, combination of detectors, and rigorous evaluation

Harbinger: Um framework para integração e análise de métodos de detecção de eventos em séries temporais*

Rebeca Salles[1], Luciana Escobar[1], Lais Baroni[1], Roccio Zorrilla[1], Artur Ziviani[2], Vinicius Kreischer[1], Flavia C. Delicato[2], Paulo F. Pires[1], Luciano Maia[3], Rafaelli Coutinho[1], Laura Assis[1], Eduardo Ogasawara[1]

[1]CEFET/RJ - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

[2]LNCC - Laboratório Nacional de Computação Científica

[3]UFF - Universidade Federal Fluminense

[4]Petrobras

{rebecca.salles,luciana.vignoli,lais.baroni}@eic.cefet-rj.br

{ziviani,roccio.zorrilla,vinicius.kreischer}@lncc.br

{fdelicato,paulo.pires}@ic.uff.br, luciano.maia@petrobras.com.br

{rafaelli.coutinho,laura.assis}@cefet-rj.br, eogasawara@ieee.org

Abstract. *When analyzing time series, it is possible to observe significant changes in the behavior of its observations that frequently characterize the occurrence of events. Events may appear as anomalies, change points, or frequent patterns. In literature, there are several methods for event detection. However, the search for a suitable method for a time series is not a simple task, especially considering that the nature of the events is often not known. In this context, this work presents Harbinger, a framework for integration and analysis of event detection methods. Harbinger was evaluated with synthetic and real data, where it was possible to verify that its functionalities foster the selection of methods and the understanding of detected events.*

Resumo. *Ao analisar séries temporais é possível observar mudanças significativas no comportamento das observações que frequentemente caracterizam a ocorrência de eventos. Eventos se apresentam como anomalias, pontos de mudança, ou padrões frequentes. Na literatura existem diversos métodos para detecção de eventos. Entretanto, a busca por um método adequado para uma série temporal não é uma tarefa simples, principalmente considerando-se que a natureza dos eventos muitas vezes não é conhecida. Neste contexto, este trabalho apresenta Harbinger, um framework para integração e análise de métodos de detecção de eventos. O Harbinger foi avaliado em dados sintéticos e reais, onde foi possível constatar que suas funcionalidades promovem a seleção de métodos e a compreensão dos eventos detectados.*

1. Introdução

Na análise de séries temporais, frequentemente é possível observar a ocorrência de uma mudança significativa em seu comportamento em um certo ponto ou intervalo de tempo.

*Os autores agradecem a Petrobras, FAPERJ, CAPES (cód. 001) e CNPq pelo financiamento do projeto.

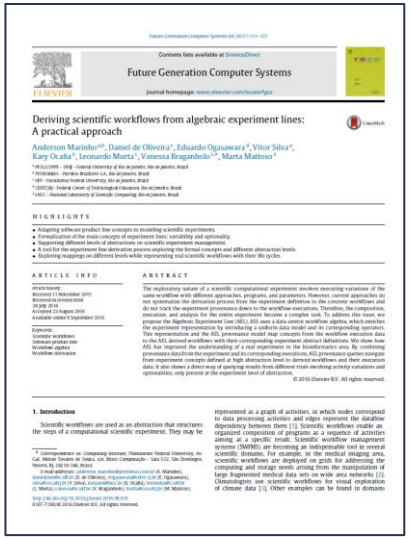
	Anomalies	**Change point**	**Motifs**
Methods	50	10	9

Features	
Metrics	Hard (traditional) and Fuzzy
Models	Ensemble and Ensemble Fuzzy
Thresholds customization	Deviation, Filters, and Candidate selection
Visualization	Time Series, Comparison, Residuals, and thresholds

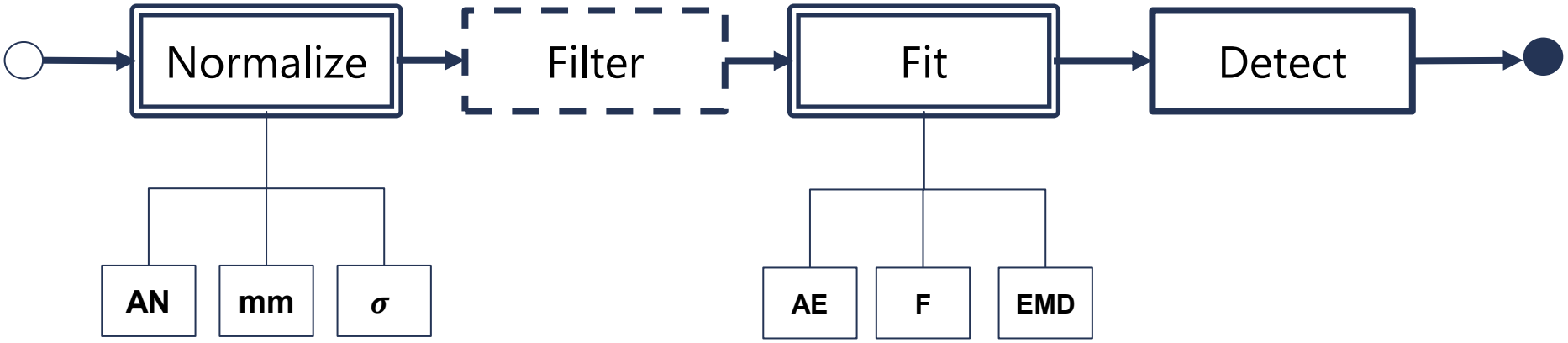
R. Salles, L. Escobar, L. Baroni, R. Zorrilla, A. Ziviani, V. Kreischer, F. Delicato, P. F. Pires, L. Maia, R. Coutinho, L. Assis, e E. Ogasawara, "Harbinger: Um framework para integração e análise de métodos de detecção de eventos em séries temporais", em *Anais do Simpósio Brasileiro de Banco de Dados (SBBD)*, SBC, set. 2020, p. 73–84. doi: 10.5753/sbbd.2020.13626.

Modular architecture

- Detection, Evaluation, Combination, Comparison
- Each module can be used alone or integrated
- Rigid interface (based on workflow algebra and experiment lines)
- Built on top of DAL Toolbox
  - https://cran.r-project.org/web/packages/daltoolbox
  - Inspired by Scikit-learn: fit() and detect() functions

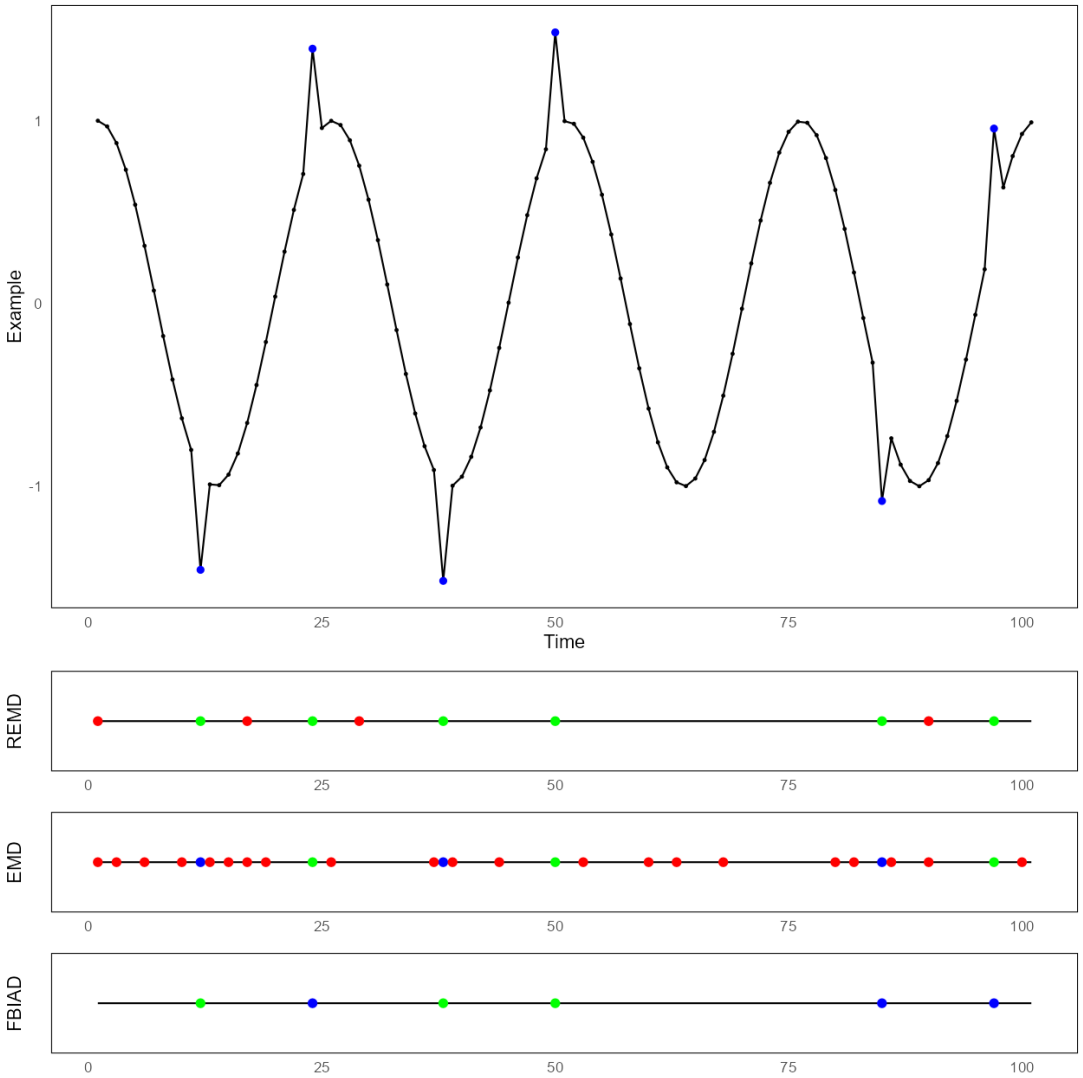

Future Generation Computer Systems

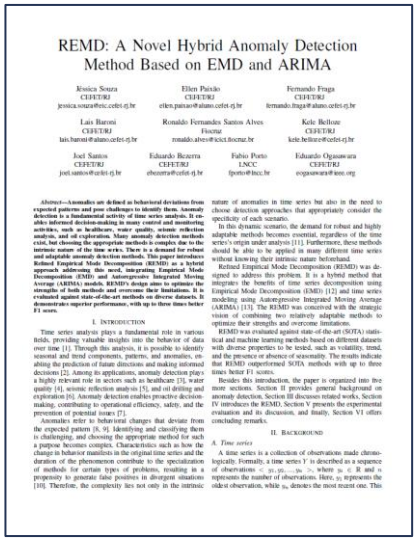
Deriving scientific workflows from algebraic experiment lines: A practical approach



A. Marinho, D. de Oliveira, E. Ogasawara, V. Silva, K. Ocaña, L. Murta, V. Braganholo, e M. Mattoso, "Deriving scientific workflows from algebraic experiment lines: A practical approach", *Future Generation Computer Systems*, vol. 68. p. 111–127, 2017. doi: 10.1016/j.future.2016.08.016.

Harbinger allows the evaluation and benchmark of detectors





REMD: A Novel Hybrid Anomaly Detection Method Based on EMD and ARIMA

Jéssica Souza, CEFET/RJ, jessica.souza@eic.cefet-rj.br
Ellen Paixão, CEFET/RJ, ellen.paixao@aluno.cefet-rj.br
Fernando Fraga, CEFET/RJ, fernando.fraga@aluno.cefet-rj.br
Laís Baroni, CEFET/RJ, lais.baroni@aluno.cefet-rj.br
Ronaldo Fernandes Santos Alves, Fiocruz, ronaldo.alves@ticict.fiocruz.br
Kele Belloze, CEFET/RJ, kele.belloze@cefet-rj.br
Joel Santos, CEFET/RJ, joel.santos@cefet-rj.br
Eduardo Bezerra, CEFET/RJ, ebezerra@cefet-rj.br
Fabio Porto, LNCC, fporto@lncc.br
Eduardo Ogasawara, CEFET/RJ, eogasawara@ieee.org

Abstract—Anomalies are defined as behavioral deviations from expected patterns and pose challenges to identify them. Anomaly detection refers to the identification of time series segments able to support decision-making in various critical application activities, such as healthcare, water quality, seismic reflection analysis, and oil exploration. Many anomaly detection methods exist, but choosing the appropriate methods is complex due to the intrinsic nature of the time series. There is a demand for robust and adaptable anomaly detection methods. This paper introduces Refined Empirical Mode Decomposition (REMD) as a hybrid approach addressing this need, integrating Empirical Mode Decomposition (EMD) and Autoregressive Integrated Moving Average (ARIMA) models. REMD's design aims to optimize the strengths of both methods and overcome their limitations. It is evaluated against state-of-the-art methods on diverse datasets. It demonstrates superior performance, with up to three times better F1 scores.

I. Introduction

Time series analysis plays a fundamental role in various fields, providing valuable insights into the behavior of data over time [1]. Through this analysis, it is possible to identify seasonal and trend components, patterns, and anomalies, enabling the prediction of future directions and making informed decisions [2]. Among its applications, anomaly detection plays a highly relevant role in sectors such as healthcare [3], water quality [4], seismic reflection analysis [5], and oil drilling and exploration [6]. Anomaly detection enables proactive decision-making, contributing to operational efficiency, safety, and the prevention of potential issues [7].

Anomalies refer to behavioral changes that deviate from the expected pattern [8, 9]. Identifying and classifying them is challenging, and choosing the appropriate method for such a purpose becomes complex. Characteristics such as how the change in behavior manifests in the original time series and the duration of the phenomenon contribute to the specialization of methods for certain types of problems, resulting in a propensity to generate false positives in divergent situations [10]. Therefore, the complexity lies not only in the intrinsic nature of anomalies in time series but also in the need to choose detection approaches that appropriately consider the specificity of each scenario.

In this dynamic scenario, the demand for robust and highly adaptable methods becomes essential, regardless of the time series's origin under analysis [11]. Furthermore, these methods should be able to be applied in many different time series without knowing their intrinsic nature beforehand.

Refined Empirical Mode Decomposition (REMD) was designed to address this problem. It is a hybrid method that integrates the benefits of time series decomposition using Empirical Mode Decomposition (EMD) [12] and time series modeling using Autoregressive Integrated Moving Average (ARIMA) [13]. The REMD was conceived with the strategic vision of combining two relatively adaptable methods to optimize their strengths and overcome limitations.

REMD was evaluated against state-of-the-art (SOTA) statistical and machine learning methods based on different datasets with diverse properties to be tested, such as volatility, trend, and the presence or absence of seasonality. The results indicate that REMD outperformed SOTA methods with up to three times better F1 scores.

Besides this introduction, the paper is organized into five more sections. Section II provides general background on anomaly detection. Section III discusses related works. Section IV introduces the REMD. Section V presents the experimental evaluation and its discussion, and finally, Section VI offers concluding remarks.

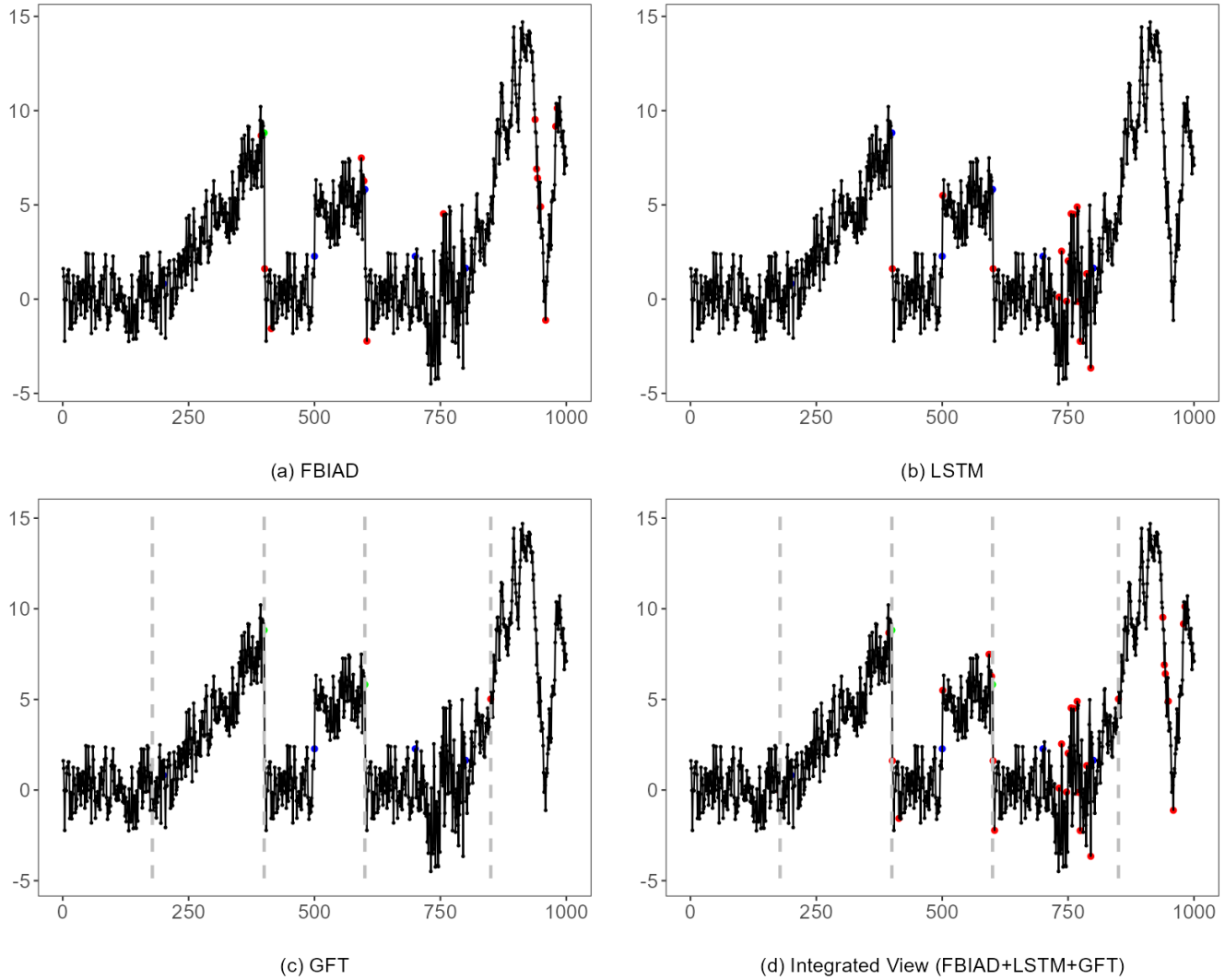
II. Background

A. Time series

A time series is a collection of observations made chronologically. Formally, a time series Y is described as a sequence of observations $< y_1, y_2, ..., y_n >$, where $y_i \in \mathbb{R}$ and n represents the number of observations. Here, y_1 represents the oldest observation, while y_n denotes the most recent one. This

J. Souza, E. Paixão, F. Fraga, L. Baroni, R. F. S. Alves, K. Belloze, J. Dos Santos, E. Bezerra, F. Porto, e E. Ogasawara, "REMD: A Novel Hybrid Anomaly Detection Method Based on EMD and ARIMA", *Proceedings of the International Joint Conference on Neural Networks*. 2024. doi: 10.1109/IJCNN60899.2024.10651192.

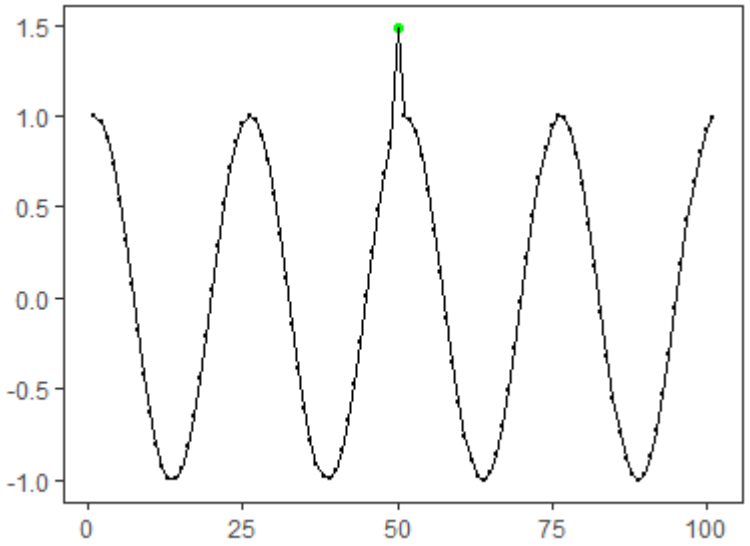
Combined visualization of detectors in a time series



E. Ogasawara *et al.*, "harbinger: A Unified Time Series Event Detection Framework".
Disponível em: https://cran.r-project.org/web/packages/harbinger/index.html

Simple anomaly detection example

- The Harbinger comes with a few example series
- Labeled time series
- Plotting functions
- Many detectors, such as ARIMA



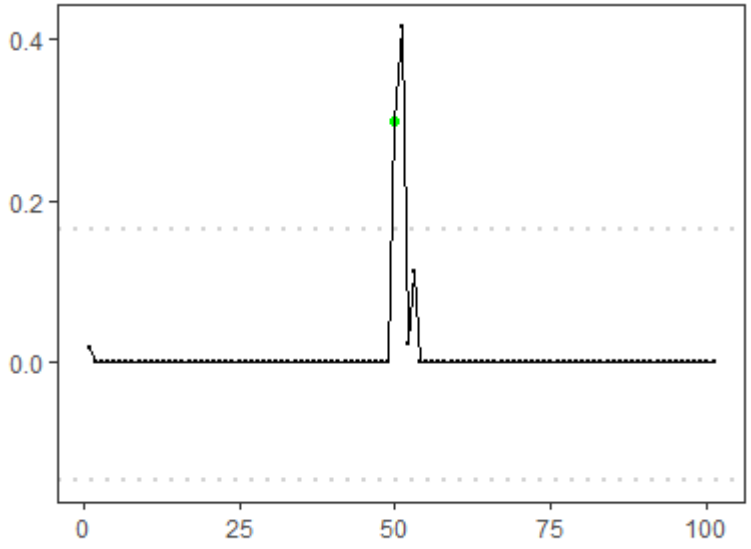

```r
library(daltoolbox)

library(harbinger)

data(examples_anomalies)
dataset <- examples_anomalies$simple
model <- hanr_arima()
# fitting the model
model <- fit(model, dataset$serie)
# making detections
detection <- detect(model, dataset$serie)
# plotting the results
har_plot(model(), dataset$serie, detection)
```

Understanding the detection

- Exploring detections
- Visualizing the residuals
- Evaluating the results




```
# filtering detected events
print(detection |>
        dplyr::filter(event==TRUE))
```

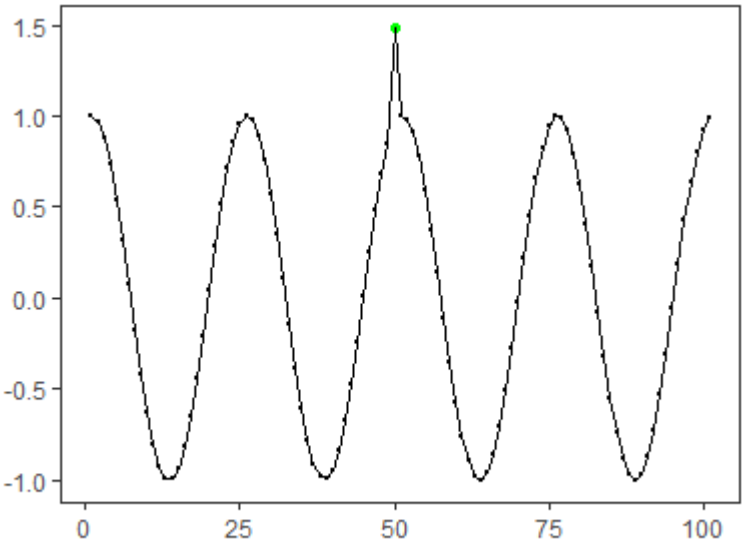
```
##   idx event    type
## 1  50  TRUE anomaly
```

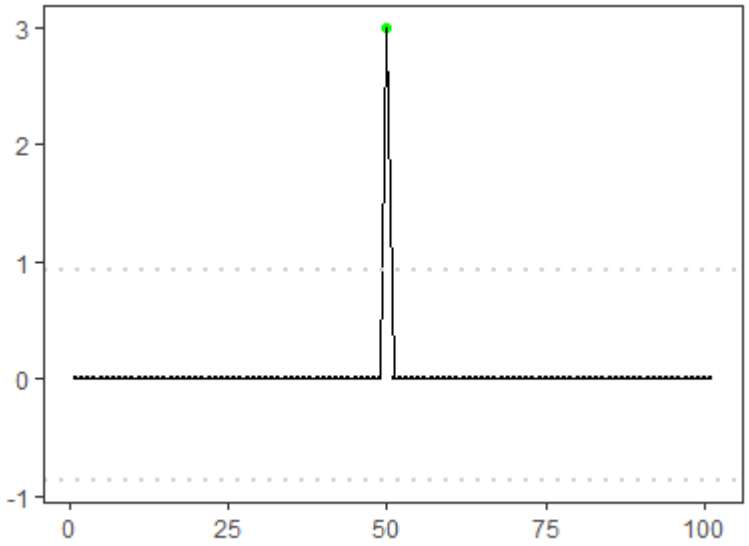
```
# understating detected events
har_plot(model, attr(detection, "res"), detection,
        dataset$event, yline = attr(detection, "threshold"))
# evaluating detected events
evaluation <- evaluate(model, detection$event, dataset$event)
print(evaluation$confMatrix)
```

```
##           event
## detection TRUE  FALSE
## TRUE      1     0
## FALSE     0     100
```

Ensemble Models

```
model <- har_ensemble(hanr_fbiad(), hanr_arima(), hanr_emd())
model <- fit(model, dataset$serie)
detection <- detect(model, dataset$serie)
har_plot(model, dataset$serie, detection, dataset$event)
har_plot(model, attr(detection, "res"), detection,
    dataset$event, yline = attr(detection, "threshold"))
```

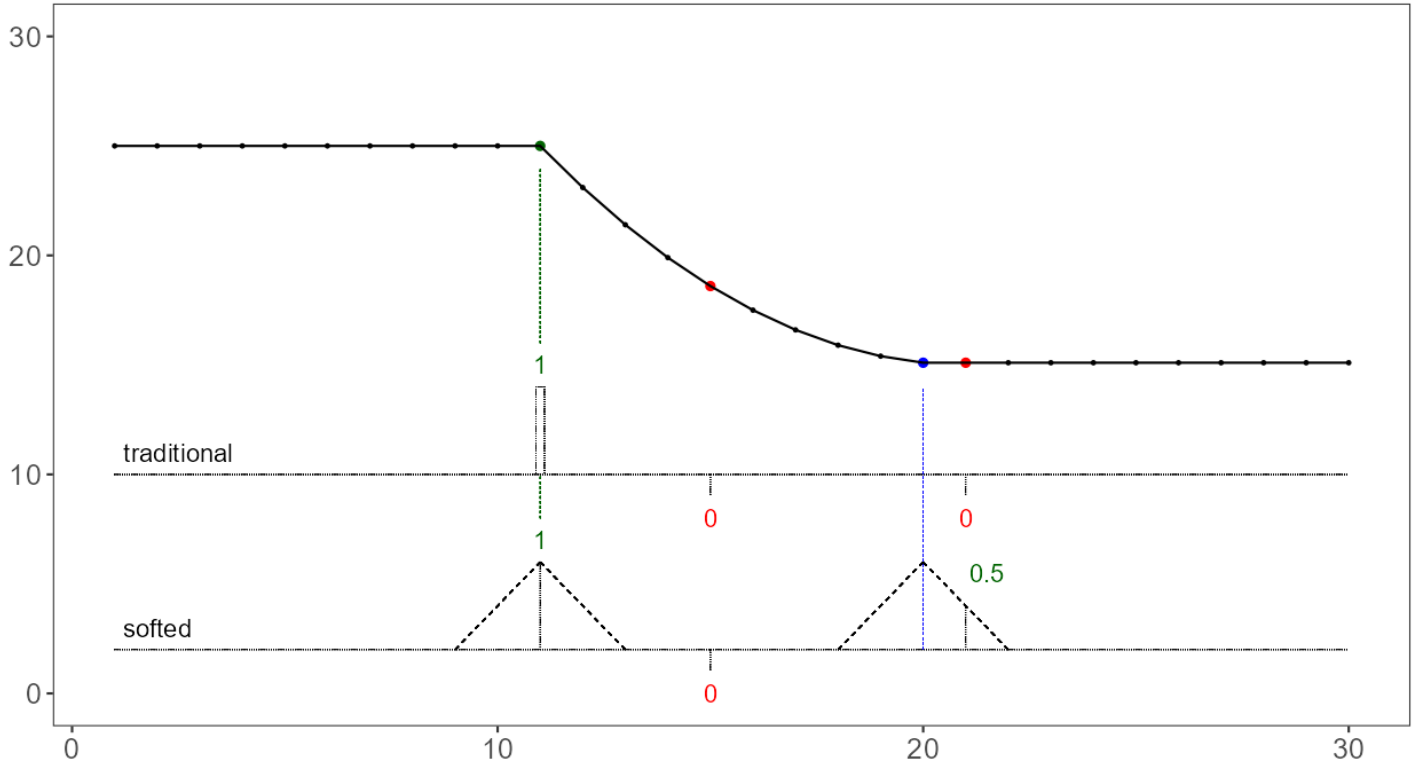




Harbinger also allows combining multiple detectors in ensemble models, extending detection robustness

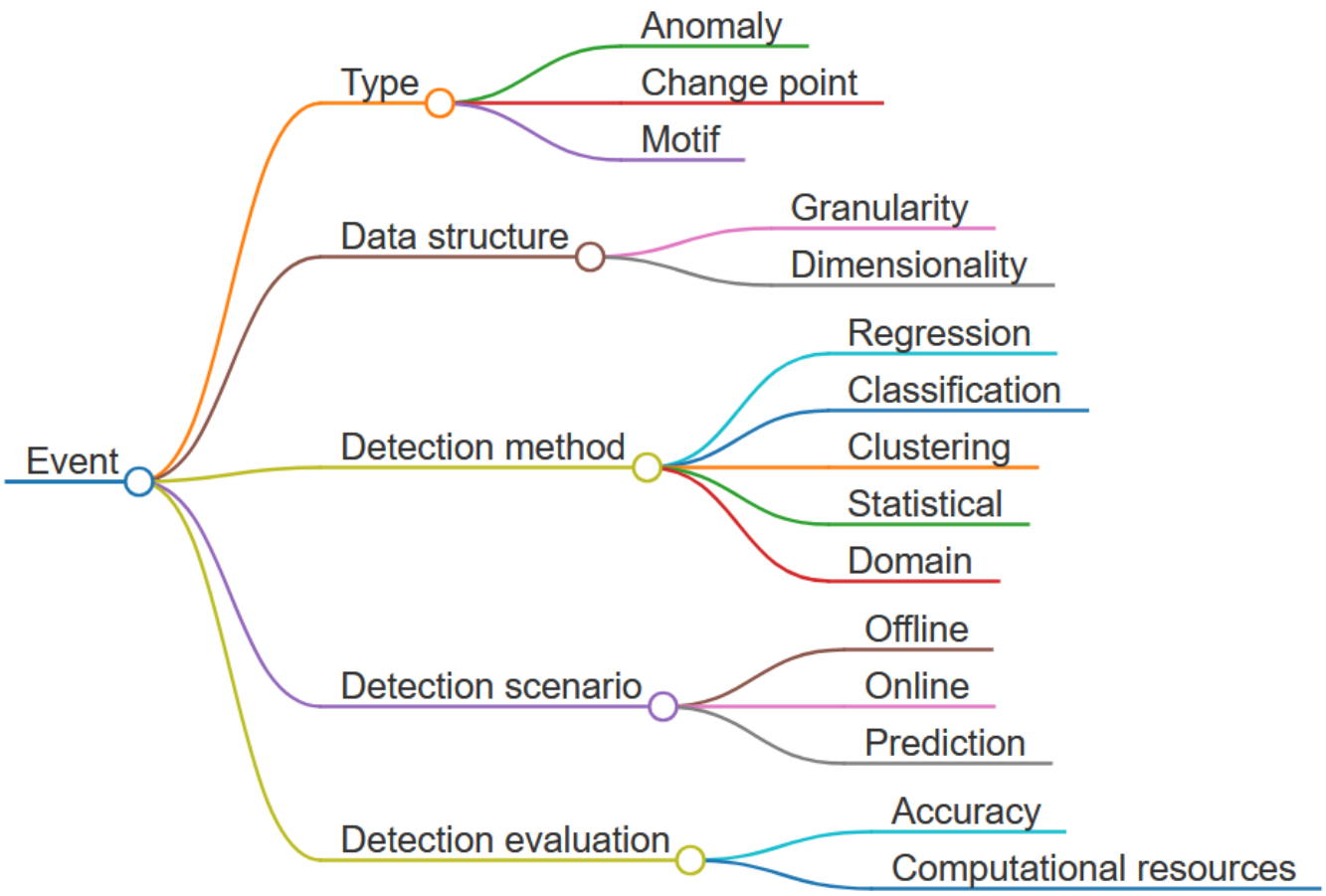
Fuzzyfing events

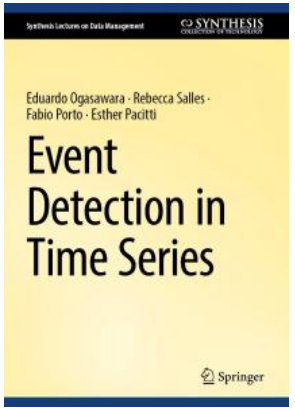
- Unlike traditional metrics, SoftED tolerates minor temporal misalignments but imposes stronger penalties on fragmented or inconsistent detection patterns



R. Salles, J. Lima, M. Reis, R. Coutinho, E. Pacitti, F. Masseglia, R. Akbarinia, C. Chen, J. Garibaldi, F. Porto, e E. Ogasawara, "SoftED: Metrics for soft evaluation of time series event detection", *Computers and Industrial Engineering*, vol. 198. 2024. doi: 10.1016/j.cie.2024.110728.

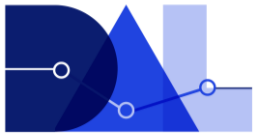
Taxonomy of Event Detection in Time Series





https://eic.cefet-rj.br/~eogasawara/en/event-detection-in-time-series

CEFET/RJ



Harbinger

A Unified Framework for Event Detection in Time Series

Eduardo Ogasawara
eduardo.ogasawara@cefet-rj.br
https://eic.cefet-rj.br/~eogasawara