JUNE 30 - JULY 5, 2024 YOKOHAMA, JAPAN

IEEE WCCI 2024

CEFET/RJ

REMD: A NEW HYBRID METHOD FOR ANOMALY DETECTION IN TIME SERIES

Jessica Souza, Ellen Paixão, Fernando Fraga, Lais Baroni, Ronaldo Alves, Kele Belloze, Joel dos Santos, Eduardo Bezerra, Fabio Porto, Eduardo Ogasawara

LNCC

FIOCRUZ

Eduardo Ogasawara
eogasawara@ieee.org
https://eic.cefet-rj.br/~eogasawara

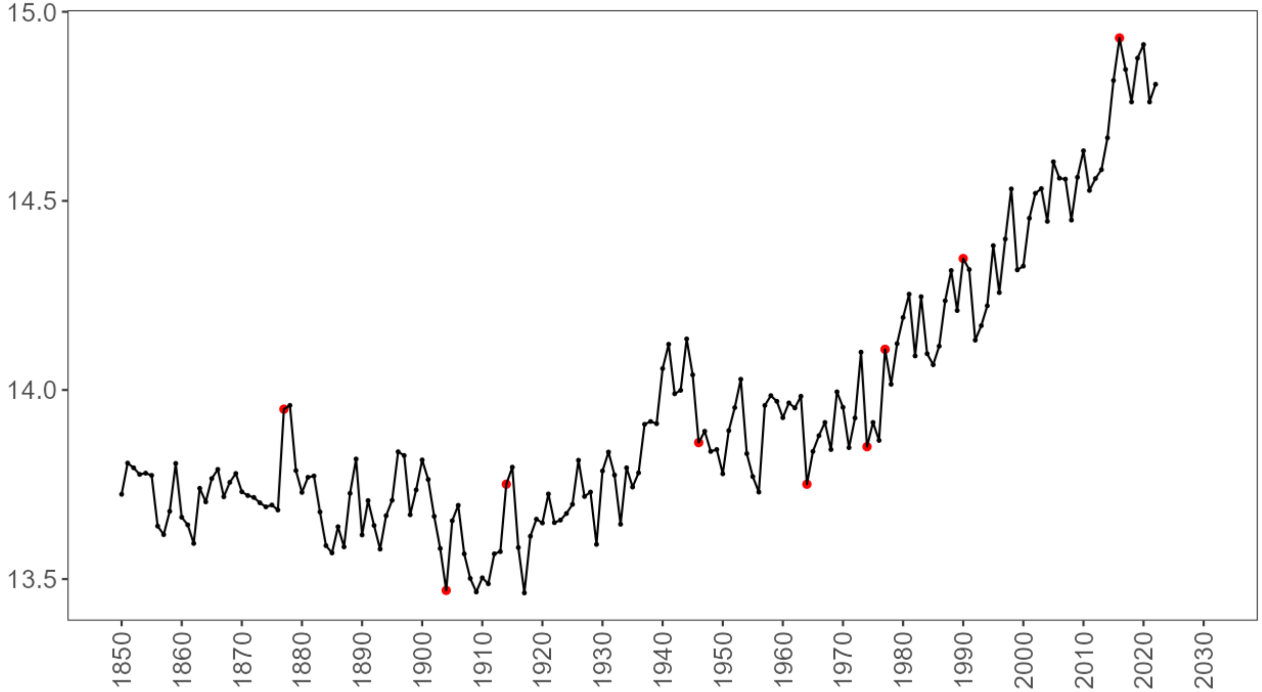
Road map

- Context
- Background
- REMD
- Results
- Conclusion

ChatGPT. (2024). Illustration of a data-centric AI for time series event detection

Anomalies

- Anomalies are observations that do not conform to the typical ones at the time series
  - Observations seem not to be derived by the time series process
- Let X be a time series $< x_1, \dots, x_n >$
  - $A(X) = \{i\} \mid x_i \notin [Q1(X) - 1.5IQR(X), Q3 + 1.5IQR(X)]$



[1] M. Olteanu, F. Rossi, and F. Yger, "Meta-survey on outlier and anomaly detection," Neurocomputing, vol. 555. 2023. doi: 10.1016/j.neucom.2023.126634.

Anomaly Detection

- Process of identifying anomalies
- Important for monitoring and surveillance
  - Industry, seismic, oil exploration, epidemiology, climate
- There are many anomaly detection methods
- Currently, detectors tend to be specialized for certain domains or types of anomalies

ChatGPT. (2024). Illustration of a time series event detection

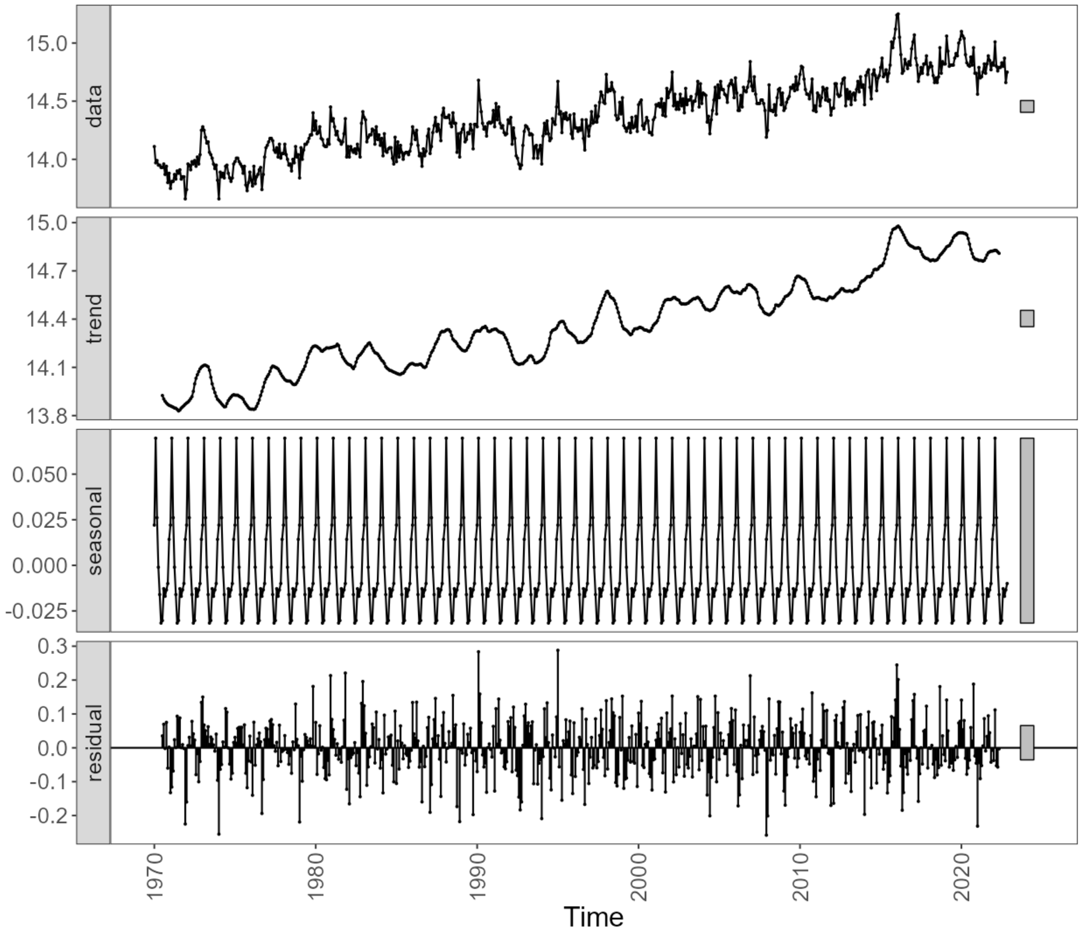
Problem definition

- Provide more adaptive/generalizable anomaly detection methods regardless of time series properties

Background

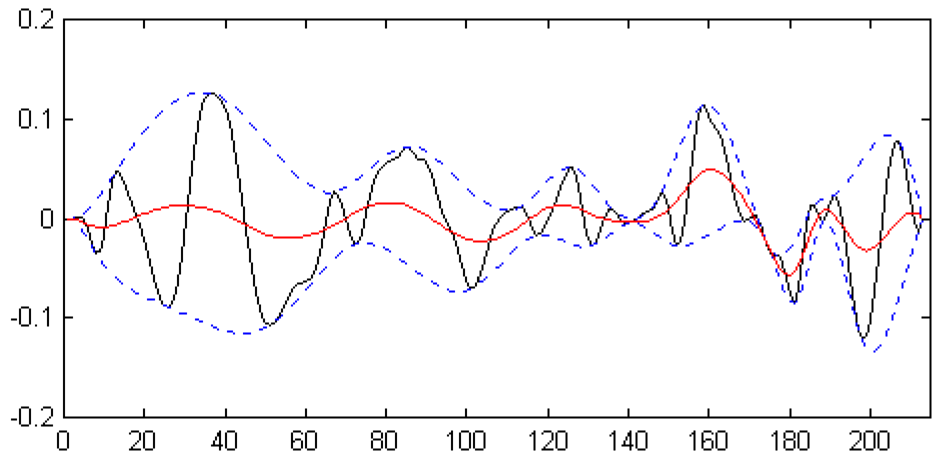
Time Series Decomposition

- Time series can be decomposed into trend, seasonality, and noise
  - $x_t = \beta_t + \pi_t + \omega_t$



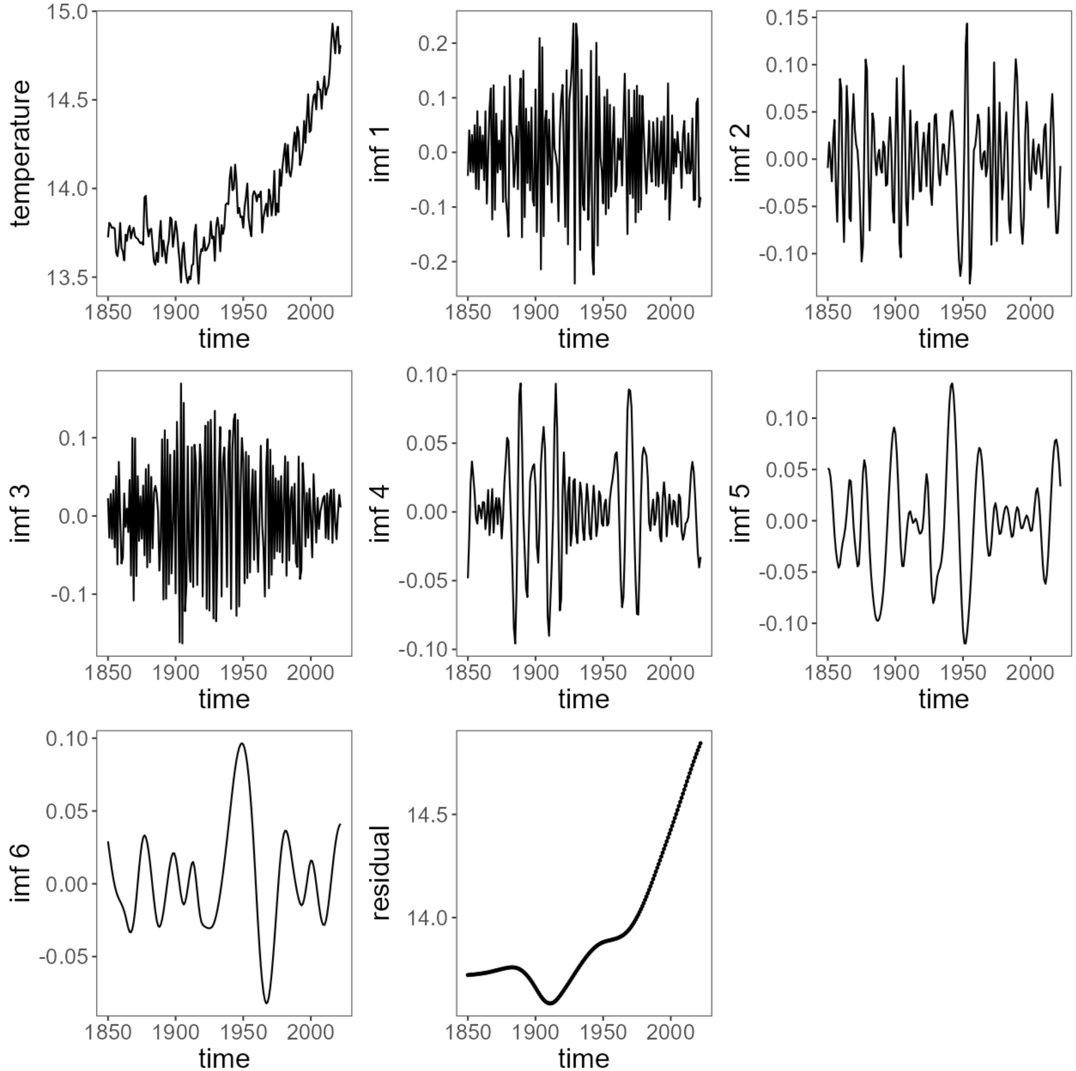
Empirical Mode Decomposition (EMD)

- EMD is a technique for decomposing non-linear and non-stationary series into Intrinsic Mode Functions (IMFs)
- IMF is a decomposed and oscillatory time series
  - The number of crossings between extrema maximums and minimums must be equal to zero or differ at most by one
  - At any point, the mean value of the envelope defined by local maxima and local minima is zero



[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Snin, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis,” Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971. pp. 903–995, 1998. doi: 10.1098/rspa.1998.0193.

EMD Example



[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Snin, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis,” Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971. pp. 903–995, 1998. doi: 10.1098/rspa.1998.0193.

EMD as an anomaly detection method

- Apply EMD
- Analyze outliers in IMF_1

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015.

ARIMA

- ARIMA is a statistical model for time series analysis and forecasting
  - combines autoregressive, moving averages, and differencing
- ARIMA (p, d, q)
  - p = Order of the autoregressive term
  - d = Order of the differencing term
  - q = Order of the moving average term
- $\theta(B)(1-B)^d x_t = \phi(B)\,\omega_t$

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015.

ARIMA is usually used for anomaly detection

- Model a time series using ARIMA (perhaps using auto-arima)
- Compute the residual
- Analyze outliers in the residual

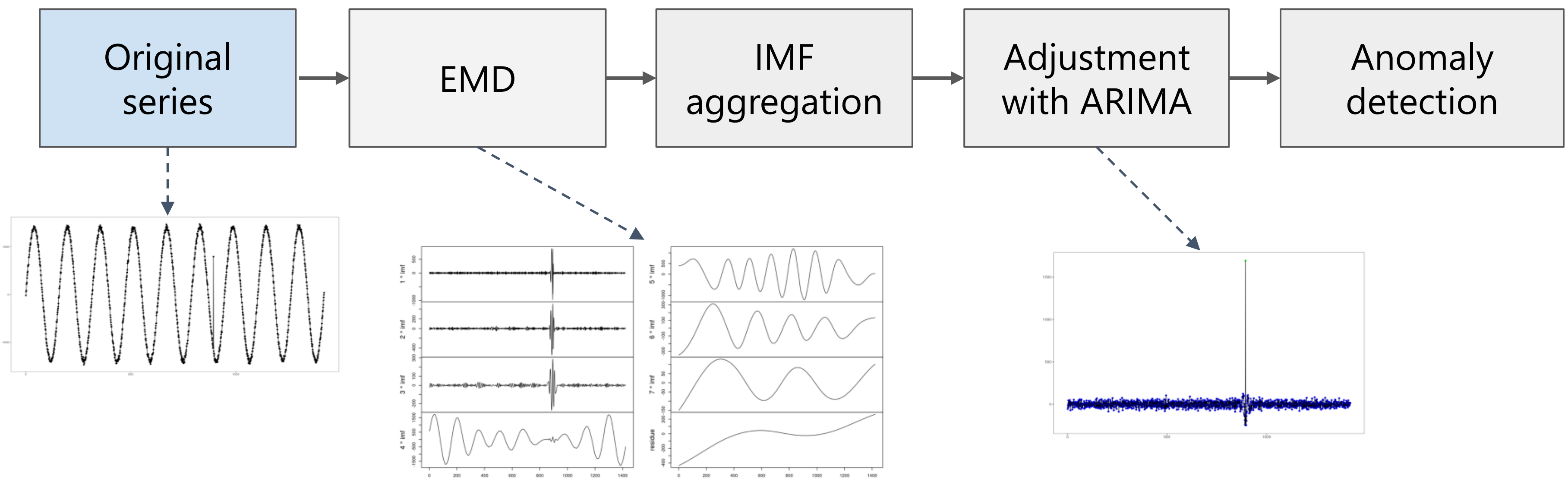
[1] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," Journal of Statistical Software, vol. 27, no. 3. pp. 1–22, 2008. doi: 10.18637/jss.v027.i03.

REMD: Refined Empirical Mode Decomposition

REMD

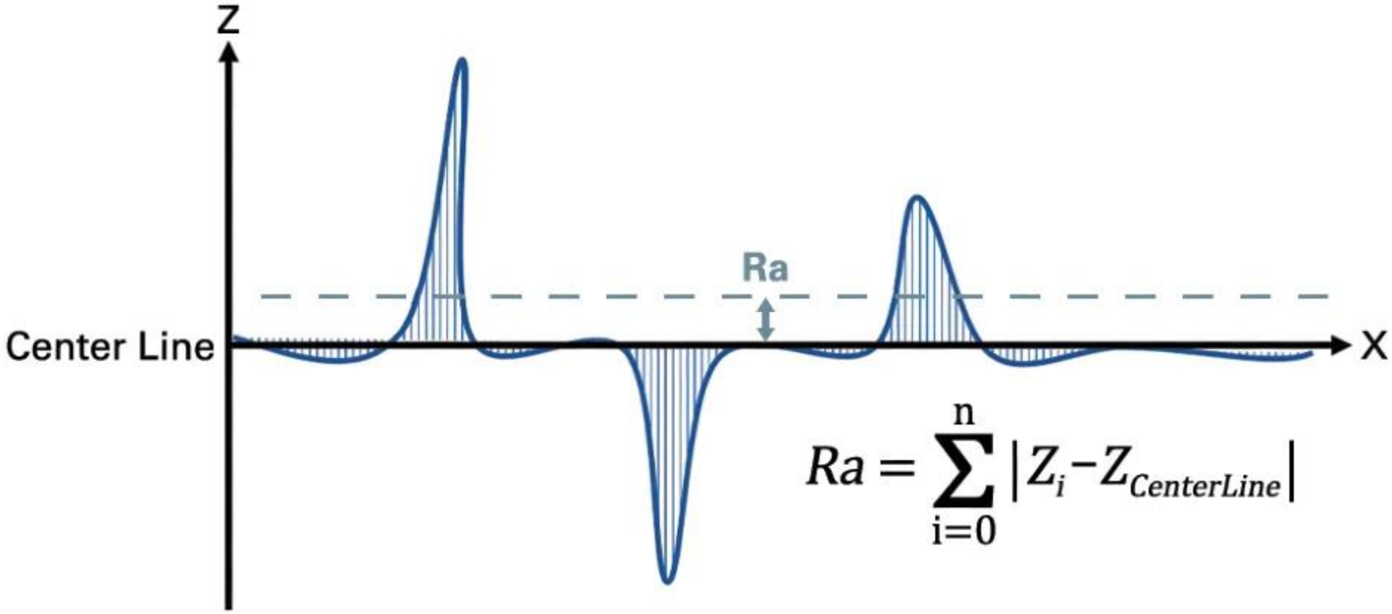
- REMD: A hybrid method consisting of four steps
  - EMD decomposition
  - IMF aggregation
  - ARIMA adjustment of aggregated IMF
  - Anomaly detection: analysis of distribution residuals


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



IMF aggregation

- Roughness in a time series refers to the irregularity or variability in values over time. It measures how "rough" or "smooth" a series is, indicating oscillation or stability in different periods



[1] T. R. Thomas, “Characterization of surface roughness,” Precision Engineering, vol. 3, no. 2. pp. 97–104, 1981. doi: 10.1016/0141-6359(81)90043-X.

REMD Algorithm

function REMD(Y) ▷ Initialization
    $IMFs, \omega \leftarrow \text{EMD}(Y)$ ▷ Algorithm 1
    $H, L \leftarrow \text{FrequencySeparation}(IMFs, \omega)$
    $\hat{H} \leftarrow \text{ModelAdjustment}(H)$
    $anomalies \leftarrow \text{ResidualAnalysis}(H, \hat{H})$
    return $anomalies$
function FrequencySeparation($IMFs$, ω)
    $RAs \leftarrow \text{Roughness}(IMFs)$
    $m \leftarrow \text{MaxCurvature}(RAs)$
    $H \leftarrow \{IMFs_1, \ldots, IMFs_m\}$
    $L \leftarrow \{IMFs_{m+1}, \ldots, IMFs_{|IMFs|}, \omega\}$
    return H, L
function ModelAdjustment(H)
    $model \leftarrow \text{FitARIMA}(H)$
    $\hat{H} \leftarrow \text{PredictARIMA}(H)$
    return $\hat{H}$
function ResidualAnalysis($H, \hat{H}$)
    $\omega \leftarrow H - \hat{H}$
    $p_i = \frac{|1-\omega_i|}{\max \omega}$ ▷ Equation 10
    Let P be the set of p_i
    $anomalies \leftarrow A(P)$ ▷ Equation 3
    return $anomalies$

Harbinger: Framework for Time Series Event Detection

- Holistic view of the problem
  - Anomalies
  - Change points
  - Motif discovery
- Properties
  - Uniform Data Model
  - Rigid interface (algebraic)
  - Expansible
  - Based on experimental line
- Inspiration from Sci-Kit Learn
  - Fit()
  - Detection()
- More than 50 event detectors
- R Package available at CRAN


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

Harbinger

Harbinger

Stars 14 downloads 608/month

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

In the anomaly classes, methods are based on machine learning model deviation (Conv1D, ELM, MLP, LSTM, Random Regression Forest, SVM), machine learning classification model (Decision Tree, KNN, MLP, Naive Bayes, Random Forest, SVM), clustering (kmeans and DTW) and statistical methods (ARIMA, FBIAD, GARCH).

In the change points classes, methods are based on linear regression, ARIMA, ETS, GARCH, AMOC, ChowTest, BinSeg, GFT, PELT.

In the motifs classes, methods are based on Hash and Matrix Profile. There are specific methods for multivariate series. The evaluation of detections includes both traditional and soft computing.

Harbinger architecture is based on Experiment Lines and is built on top of the DAL Toolbox. Such an organization makes it easy to customize and add novel methods to the framew

Installation

The latest version of Harbinger at CRAN is available at: https://

You can install the stable version of Harbinger from CRAN with

```
install.packages("harbinger")
```

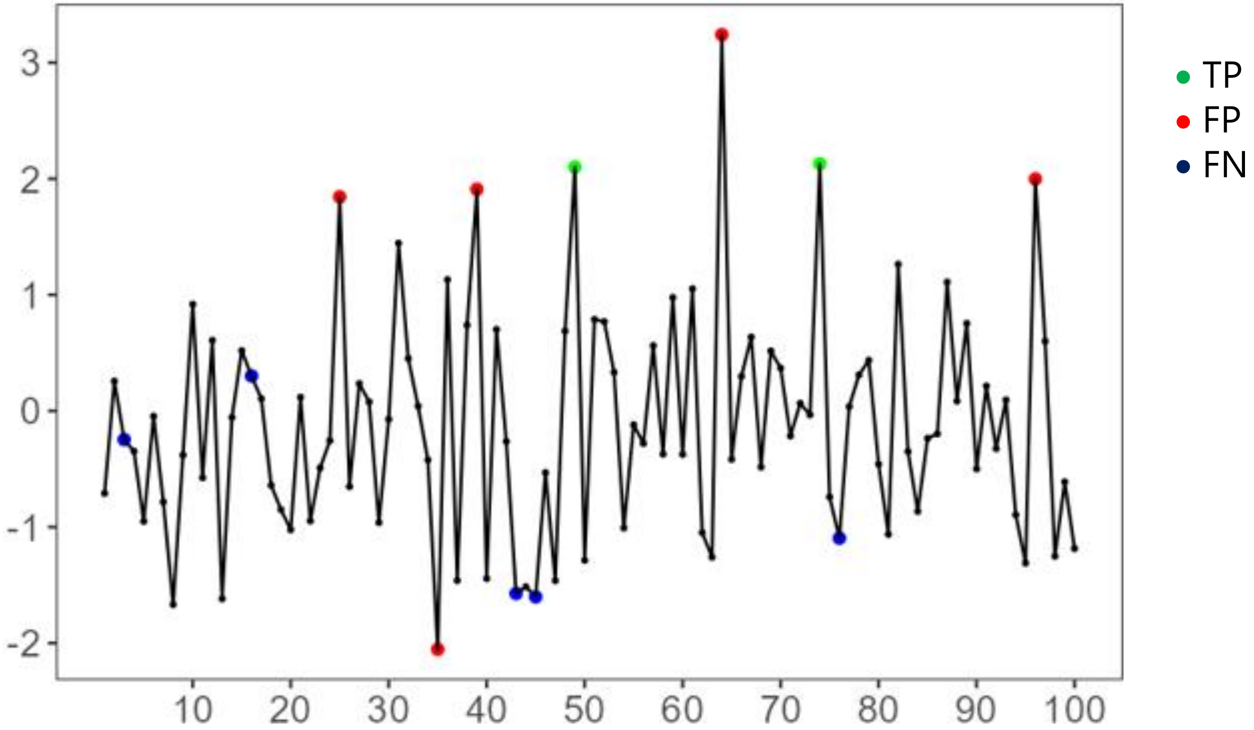
You can install the development version of Harbinger from GitH

```
# install.packages("devtools")
library(devtools)
```

Results

Basic metrics for event detection

- precision $= \frac{TP}{TP+FP}$
- recall $= \frac{TP}{TP+FN}$
- $F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

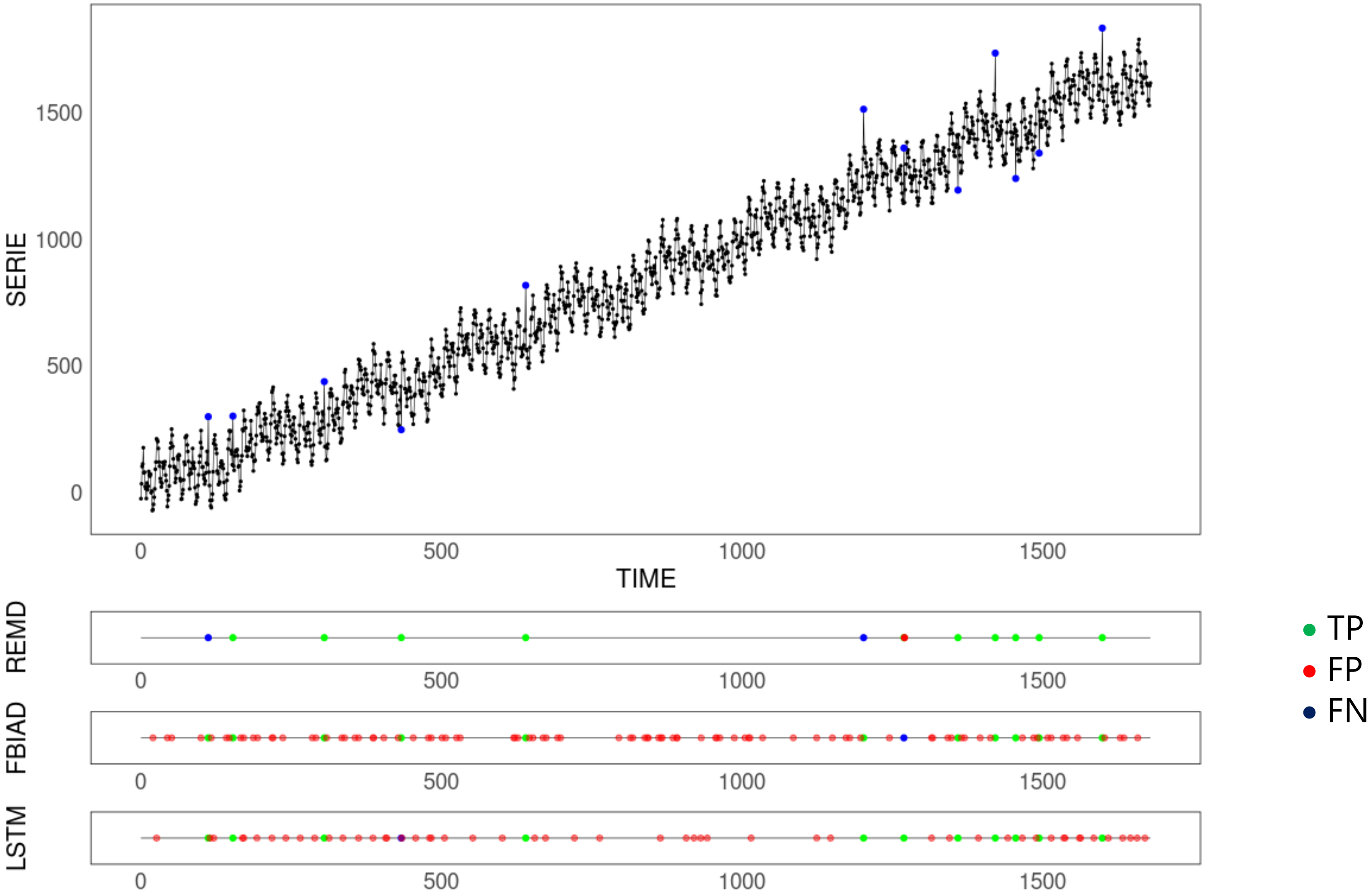


Results

- Datasets: Yahoo, Numenta, and Gecco
- REMD presents a much better performance than the second-placed method
  - EMD-based method, when we use F1 as the main selection criterion

Method	Precision	Recall	F1
REMD	**0.684**	0.386	**0.448**
EMD	0.243	0.408	0.207
FBIAD	0.066	0.528	0.085
ARIMA	0.045	**0.556**	0.067
LSTM	0.041	0.534	0.062
ELM	0.041	0.517	0.063
Conv1D	0.036	0.519	0.055
SVM	0.030	0.542	0.049

Inspecting performance comparison



[1] J. Souza, E. Paixão, F. Fraga, L. Baroni, R. F. S. Alves, K. Belloze, J. Santos, E. Bezerra, F. Porto, and E. Ogasawara, “REMD: A Novel Hybrid Anomaly Detection Method Based on EMD and ARIMA,” Proceedings of the International Joint Conference on Neural Networks, vol. 2024-July. pp. 1–8, 2024.

Conclusion

- This paper introduces a novel anomaly detection method named Refined Empirical Mode Decomposition (REMD)
  - The method goes through four phases: EMD decomposition, frequency separation using the roughness curve, ARIMA model fitting, and outlier detection
- We conducted an extensive experimental evaluation of REMD on three diverse datasets
  - Yahoo Labs, Numenta Anomaly Benchmark, and GECCO 2018 Challenge
- REMD outperformed many methods
  - Statistical EMD, FBIAD and ARIMA
  - Machine learning approaches such as LSTM, ELM, Conv1D, and SVM

JUNE 30 - JULY 5, 2024 YOKOHAMA, JAPAN

IEEE WCCI 2024

CEFET/RJ

REMD: A NEW HYBRID METHOD FOR ANOMALY DETECTION IN TIME SERIES

Jessica Souza, Ellen Paixão, Fernando Fraga, Lais Baroni, Ronaldo Alves,
Kele Belloze, Joel dos Santos, Eduardo Bezerra, Fabio Porto, Eduardo Ogasawara

LNCC FIOCRUZ

Eduardo Ogasawara
eogasawara@ieee.org
https://eic.cefet-rj.br/~eogasawara