UNIVERSIDAD ADOLFO IBÁÑEZ

Change Point Detection in Irregularly Sampled Time Series

AGN Light Curves Toy Example

Cinthya Leonor Vergara Silva - Student Talk

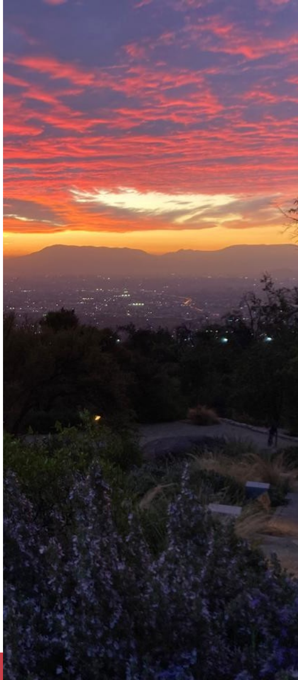


Table of contents

What Changepoint Detection is?

UNIVERSIDAD ADOLFO IBÁÑEZ

In general, changepoint detection techniques aim to identify the points in a stochastic process, typically a time series but it can be any process with ordered data, where a change occurs in the statistical properties of the data.

Can be found in a wide range of applications: quality control, finance, pricing, economics, medicine, social sciences, cybersecurity, physics, ocenoagrafy, proccess control, and more

It is also known as [**?**, **?**]:

- Change detection
- Breakpoints
- Segmentation
- Structural breaks
- Regime switching
- Detecting disorder

Note: A comprehensive review of changepoint detection techniques can be found in [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]

Why is interesting?

First, it is possible to identify 2 fundamental questions [**?**]:

- Have there been significant changes in my process over time? (how many? when?)
- Is something new or unusual?

As well as:

- What do we mean by a significant change?
- what type of change do we find?
- Are we sure it is a change or is it just noise?
- how confident are we about the change location?
- Why did it happen? was caused by external factors o is it part of the natural process?

Types of Changepoints

Example (Structural Breaks in Time Series)

1. **Mean structure changes:** Level shifts or trend changes in $\mathbb{E}[Y_t]$
2. **Variance changes:** Changes in $\mathrm{Var}(Y_t)$ (volatility shifts)
3. **Distributional changes:** Changes in the full distribution F_t
4. **Dependence structure changes:** Changes in autocorrelation, covariance, or spectral properties

img/change-point-types.png

General Design Structure for Parametric CPD

UNIVERSIDAD ADOLFO IBÁÑEZ

Typical methodological structure

1. A probabilistic (generative) model is specified for the series, allowing structural breaks in parameters.
2. The number and locations of changepoints, together with segment parameters, are estimated via a likelihood-based, penalized, or Bayesian criterion.
3. Since the resulting optimization is typically combinatorial, computational methods (dynamic programming, pruning, MCMC, recursive filtering) are employed.
4. Asymptotic and statistical properties of the estimators (consistency, convergence rates, identifiability, optimality) are established under suitable assumptions.

General Problem Formulation

Change Point Detection Problem [?]

Given time series $\{y_1, y_2, \ldots, y_T\}$, identify change points $\{\tau_1, \tau_2, \ldots, \tau_K\}$ where:

- Boundaries: $\tau_0 = 0 < \tau_1 < \tau_2 < \cdots < \tau_K < \tau_{K+1} = T$
- Segment $[\tau_i, \tau_{i+1})$ follows distribution P_i
- Distributions change: $P_i \neq P_{i+1}$ at each τ_i

$$\min_{\tau, K} \left[\sum_{i=0}^{K} \mathcal{C}(y_{\tau_i+1:\tau_{i+1}}) + \beta K \right]$$

- $\mathcal{C}(\cdot)$: segment cost (measures homogeneity)
- $\beta > 0$: penalty parameter (controls model complexity)

How many changes?

UNIVERSIDAD ADOLFO IBÁÑEZ

Cost Function Approach

Find change points by minimizing: $\sum_{i=1}^{K+1} C(y_{\tau_{i-1}+1:\tau_i}) + \beta K$ where:

- $C(\cdot)$: Cost function measuring segment homogeneity
- β: Penalty for number of change points
- K: Number of change points

Common Cost Functions

- **L2 norm**: $C = \sum ||y_t - \bar{y}||^2$ (mean changes)
- **L1 norm**: $C = \sum |y_t - \bar{y}|$ (robust to outliers)
- **Likelihood**: $C = -\sum \log f(y_t|\theta)$ (MLE approach)
- **Kernel (MMD-based)**: $C = \frac{1}{n^2} \sum_i \sum_j k(y_i, y_j) - \frac{2}{n} \sum_i k(y_i, \bar{y})$ (non-parametric, within-segment)

Popular Optimization Strategies

UNIVERSIDAD ADOLFO IBÁÑEZ

When K is Unknown

- **PELT** (Pruned Exact Linear Time): Optimal with linear complexity
- **Dynamic Programming**: Guaranteed optimal solution
- **Bottom-up**: Merge segments iteratively

When K is Known

- **Binary Segmentation**: Recursive splitting approach
- **Window-based**: Sliding window methods
- **Optimal Partitioning**: Direct optimization

Classical Approaches

Frequentist Methods [?]

Direct optimization: Minimize penalized cost function

- Choose $\mathcal{C}(\cdot)$ based on model assumptions (e.g., likelihood)
- Select β via validation or information criteria
- Point estimates for change point locations

Bayesian Methods [?]

Posterior inference: Maximum a posteriori (MAP) estimation

$$\arg\max_{\tau,K} p(\tau, K|y) = \arg\min_{\tau,K} [-\log p(y|\tau, K) - \log p(\tau, K)]$$

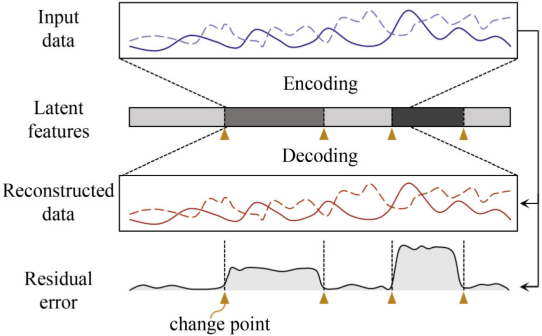
- Likelihood $p(y|\tau, K)$ defines cost function
- Prior $p(\tau, K) = p(K) \cdot p(\tau|K)$ provides regularization

Modern Approaches

Deep Learning-Based Methods

Data-driven detection: Learn mapping $f_\theta : \mathcal{Y} \to \mathcal{T}$ from data to change points

- **Supervised:** Train on labeled datasets with known change points
- **Unsupervised:** Learn representations that capture temporal structure
- Implicit cost function learned through neural network training



Unsupervised framework in deep learning-based CPD [?]

Methodological Evolution

UNIVERSIDAD ADOLFO IBÁÑEZ

Fundamental Trade-offs

1. **Bias-Variance:** Parametric power vs non-parametric robustness
2. **Computation-Accuracy:** Online speed vs offline optimality
3. **Sensitivity-Specificity:** Detection power vs false alarm control
4. **Local-Global:** Sequential methods vs batch optimization

1. **Parametric** $\rightarrow$ **Non-parametric**: Fewer distributional assumptions
2. **Offline** $\rightarrow$ **Online**: Real-time processing capabilities
3. **Low-dimensional** $\rightarrow$ **High-dimensional**: Scalability to $p \gg n$
4. **Univariate** $\rightarrow$ **Complex data**: Networks, functional, spatial
5. **Single type** $\rightarrow$ **Multiple types**: Simultaneous detection of different changes

What is Time Series Analysis?

UNIVERSIDAD ADOLFO IBÁÑEZ

Definition

Time series analysis involves studying data points collected or recorded at specific time intervals to identify patterns, trends, and anomalies.

- **Sequential Data**: Data points ordered by time: $\{y_1, y_2, \ldots, y_T\}$
- **Temporal Dependencies**: The current value is influenced by prior observations
- **Stationarity**: Whether statistical properties remain constant (stationary) or change over time (non-stationary)
- **Time Interval Regularity**: Data may be recorded at regular or irregular time intervals

Irregularly Sampled Time Series

UNIVERSIDAD ADOLFO IBÁÑEZ

The case of AGN light curves

An active galactic nucleus (AGN) is a region of space where a supermassive black hole is accreting matter.

- Astronomical surveys usually have irregular observation times.
- AGN variability is dominated by **stochastic red noise**, **irregular sampling**, **seasonal gaps** and often **low signal-to-noise** [**?**, **?**, **?**, **?**]

Example: An AGN toy example

ZTF19acnskyy - The awakening AGN

This is a **toy example** for changepoint detection. Many **free** light-curve and catalog data are available from astronomical repositories (ZTF, Gaia, LSST, IRSA, etc.) for similar analyses.

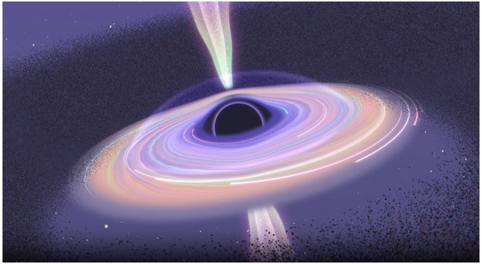


IMAGE: “Anatomy of an Active Galactic Nucleus” - NASA’s Goddard Space Flight Center [**?**]

- **Object**: SDSS J133519.91+072807.4 (aka ZTF19acnskyy)
- **Event**: 13 December 2019 - First detection of nucle
variability
- **Black Hole Mass**: $\sim 10^6\ M_\odot$ (SMBH)
- **Data Source**: ZTF (Zwicky Transient Facility)
- **Observations**: g-band light curve from IRSA (point count at run time; run `plots_astropy.py` to refresh)

Obtaining light curve data from surveys

UNIVERSIDAD ADOLFO IBÁÑEZ

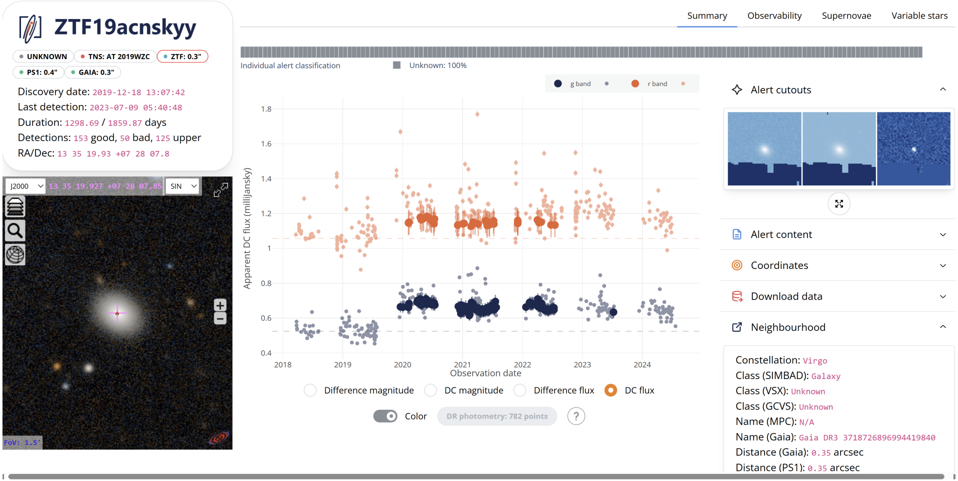
Survey APIs and the Astropy ecosystem

Light curves for AGN (e.g. ZTF, LSST) can be retrieved programmatically:

- **IRSA ZTF Lightcurve API**: https://irsa.ipac.caltech.edu/docs/program_interface/ztf_lightcurve_api.html (query by object ID or position).
- **astroquery** (Astropy ecosystem): `astroquery.ipac.irsa` for IRSA catalog and cone searches; suitable for scripting and reproducible workflows.
- **Astropy**: `astropy.timeseries.TimeSeries` to hold irregular observation times and fluxes. This toy example uses **PELT** (ruptures library) for changepoint detection (see below).

Many free resources exist in astronomical repositories (e.g. Rubin Science Platform, IRSA bulk and cloud); the present AGN toy example uses publicly available ZTF data.

AGN toy example: ZTF19acnskyy



ZTF g-band light curve. Data from public survey archives — Image ZTF Fink Portal https://ztf.fink-portal.org/ZTF19acnskyy. Run `plots_astropy.py` for current IRSA point count

ZTF19acnskyy Changepoints

img/change-in-mean.png

Change in Mean

img/change-in-variance.png

Change in Variance

img/change-in-mean-variance

Change in Mean & Variance

Using **PELT** (ruptures library): BIC-style penalty for variance and mean & variance panels; a lighter penalty for the mean-only panel so that mean shifts are detected. Three panels: change in mean (L2 cost), change in variance (normal cost on centered data), change in mean & variance (normal cost). Segment means (and spread) are shown. x-axis = MJD.

Piecewise Example Model (I)

Another technique is to fit a **piecewise** model: use the same parametric form in each segment between changepoints. Here we use a **binary segmentation (adaptive-pooled)** strategy: repeatedly apply AMOC (at most one change per segment), choosing the split that most reduces pooled MSE (error proportional to points in each segment). We take a **generic** linear model—trend plus one harmonic plus noise—and fit it separately in each regime:

$$y(t) = \alpha + \beta t + \gamma t^2 + a_1 \cos(2\pi f_1 t) + b_1 \sin(2\pi f_1 t) + \varepsilon$$

where:

- α = intercept; β, γ = linear and quadratic trend (to capture slow evolution)
- a_1, b_1 = harmonic coefficients; f_1 = fundamental frequency ($1/f_1 \approx 182.6$ days). We use $J = 1$ harmonic (AGN are not strictly periodic)
- ε = residual (e.g. measurement noise)

Piecewise Example Model (II)

Results show that it is possible to identify regime change and some periodicity.

But, AGN may have quasi-periodic oscillations in their lightcurves. Quasiperiodic oscillations occur mainly because of the dynamics of the accretion disk, and the relativistic effects or stochastic red noise.

img/piecewise-models.png

Looking for changes? - Final thoughts

UNIVERSIDAD ADOLFO IBÁÑEZ

Final thoughts

- Changepoint detection is a powerful tool to identify changes in time series data.
- Is used in a wide range of applications, including finance, economics, medicine, social sciences, linguistics, astronomy, oceanography, process control, and more.
- There's a lot of research in the field, and new methods are being developed all the time.

But! still there's a lot of challenges to overcome...

Challenges and Future Work

UNIVERSIDAD ADOLFO IBÁÑEZ

- **Data**: Irregular sampling, low signal-to-noise, seasonal gaps, graph data
- **Computational Complexity**: scalability to large datasets and algorithm complexity
- **Sensitivity-Specificity**: Detection power vs false alarm control
- **Local-Global**: Sequential methods vs batch optimization
- **Real-time Processing**: Online processing in real-time scenarios
- **Robustness**: Handling outliers, missing data, and non-stationarity
- **Statistical Guarantees**: consistency, power, robustness, scalability, computational efficiency.
- **Causality and Interpretability**: understanding the causal relationships between the change points and the underlying processes

Acknowledgments

UNIVERSIDAD ADOLFO IBÁÑEZ

- **Event funding:** This event was funded by the Center for Mathematical Modeling (CMM) through its ANID FB210005 Basal Project.
- **AI-assisted development:** AI-based tools (Cursor IDE) were used to assist with writing, generate code suggestions, and improve documentation. Methodological decisions, literature review, implementation choices, and final edits are the author's responsibility.
- **Changepoint methodology:** The use of PELT and the changepoint workflow was informed by the workshop *Introduction to Changepoint Analysis* [**?**] (R. Killick, CASI 2024; `changepoint` R package, PELT). The **binary segmentation** (adaptive-pooled) used in the piecewise model was developed by the author in the code, combining general ideas from the well-known literature.

I hope you enjoyed this talk!

UNIVERSIDAD ADOLFO IBÁÑEZ

Rubin First Look Gallery - Virgo Cluster “Cosmic Abundance” [?]

References I

- **Samaneh Aminikhanghahi and Diane J. Cook.**
  A survey of methods for time series change point detection.
  Knowledge and Information Systems, 51(2):339–367, 2017.

- **V Baseville, R Plante, T Deeming, and D Durand.**
  Change point analysis in astronomy.
  Astronomy and Astrophysics, 271:451, 1993.

- **B. E. Brodsky and B. S. Darkhovsky.**
  Nonparametric methods in change-point problems, volume 243 of *Mathematics and Its Applications*.
  Springer Netherlands, 1993.

References II

- **NASA's Goddard Space Flight Center, Krystofer Kim, Sophia Roberts, and Jeanette Kazmierczak.**
  Anatomy of an active galactic nucleus.
  https://svs.gsfc.nasa.gov/14922/, 2025.
  Animation released December 1, 2025. Animator: Krystofer Kim, Sophia Roberts (eMITS), Jeanette Kazmierczak. An artist's concept of an active galactic nucleus and its central black hole.

- **Muktesh Gupta, Rajesh Wadhvani, and Akhtar Rasool.**
  Comprehensive analysis of change-point dynamics detection in time series data: A review.
  Expert Systems with Applications, 248:123342, 2024.

References III

Lajos Horváth and Gregory Rice.
Change Point Analysis for Time Series.
Springer, 2024.

R. Killick and I. A. Eckley.
changepoint: An r package for changepoint analysis.
Journal of Statistical Software, 58(3), 2014.

R. Killick, P. Fearnhead, and I. A. Eckley.
Optimal detection of changepoints with a linear computational cost.
Journal of the American Statistical Association, 107(500):1590–1598, 2012.

References IV

Rebecca Killick.
Introduction to changepoint analysis.
CASI 2024 workshop; https://github.com/rkillick/intro-changepoint-course/blob/master/IntroCptWorkshop.Rmd, 2024.
R Markdown workshop: `changepoint`, `changepoint.np`, PELT, BinSeg.

Szymon Kozłowski.
Revisiting stochastic variability of agns with structure functions.
The Astrophysical Journal, 826(2):118, 2016.

Jackeline Moreno, Michael S Vogeley, Gordon T Richards, and Weixiang Yu.
Stochastic modeling handbook for optical agn variability.
Publications of the Astronomical Society of the Pacific, 131(1000):063001, 2019.

References V

- **NSF-DOE Vera C. Rubin Observatory/NOIRLab/SLAC/AURA.**
  Cosmic abundance – virgo cluster.
  https://rubinobservatory.org/gallery/collections/first-look-gallery/urcqf35h9h28ha8huluk8jb02j, 2025.
  Image, accessed June 5, 2025.

- **Laurent Oudre.**
  Intelligence artificielle & machine learning pour la modélisation de séries temporelles et de signaux. séance 8 : Détection de ruptures et d'anomalies.
  https://www.laurentoudre.fr/signalml.html, 2026.
  laurent.oudre@ens-paris-saclay.fr.

- **P. Sánchez-Sáez et al.**
  Sdss1335+0728: The awakening of a $\sim 10^6\, m_\odot$ black hole.
  Astronomy & Astrophysics, 688:A157, 2024.

References VI

- **Jeffrey D. Scargle, Jay P. Norris, Brad Jackson, and James Chiang.**
  Studies in astronomical time series analysis. VI. bayesian block representations.
  The Astrophysical Journal, 764(2):167, 2013.

- **Charles Truong, Laurent Oudre, and Nicolas Vayatis.**
  Selective review of offline change point detection methods.
  Signal Processing, 167:107299, 2020.

- **Haoyun Wang and Yao Xie.**
  Sequential change-point detection: Computation versus statistical performance.
  Wiley Interdisciplinary Reviews: Computational Statistics, 16(1):e1628, 2024.

References VII

Ruiyu Xu, Zheren Song, Jianguo Wu, Chao Wang, and Shiyu Zhou.
Change-point detection with deep learning: A review.
Frontiers of Engineering Management, 12(1):154–176, 2025.