Potsdam Institute for
Climate Impact Research

PIK

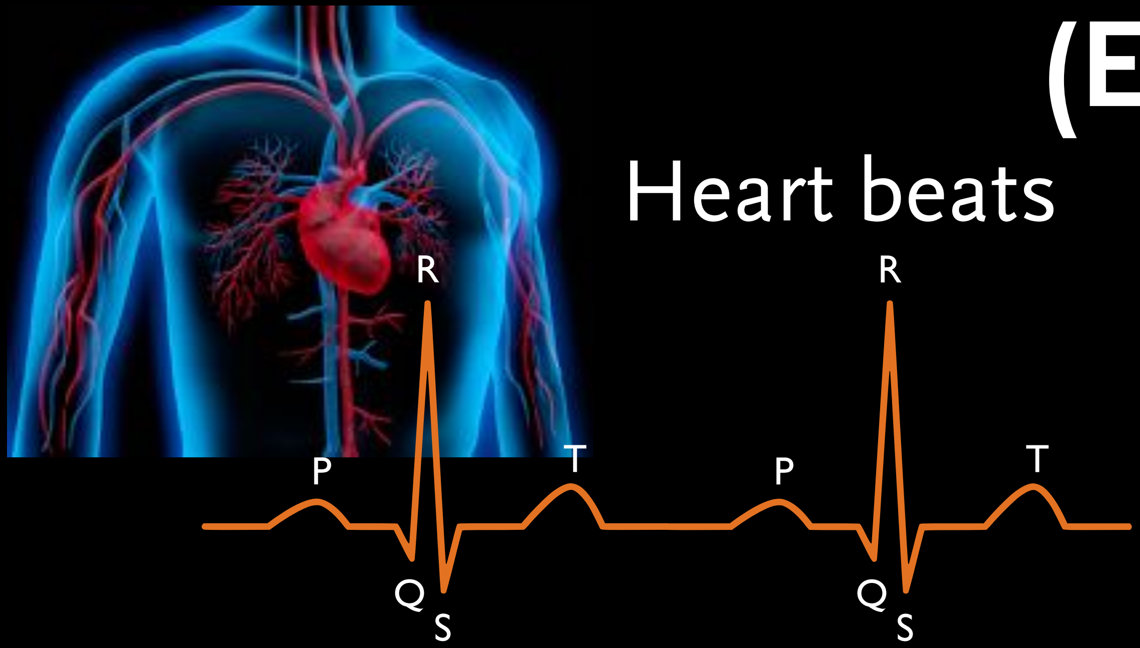
Norbert Marwan

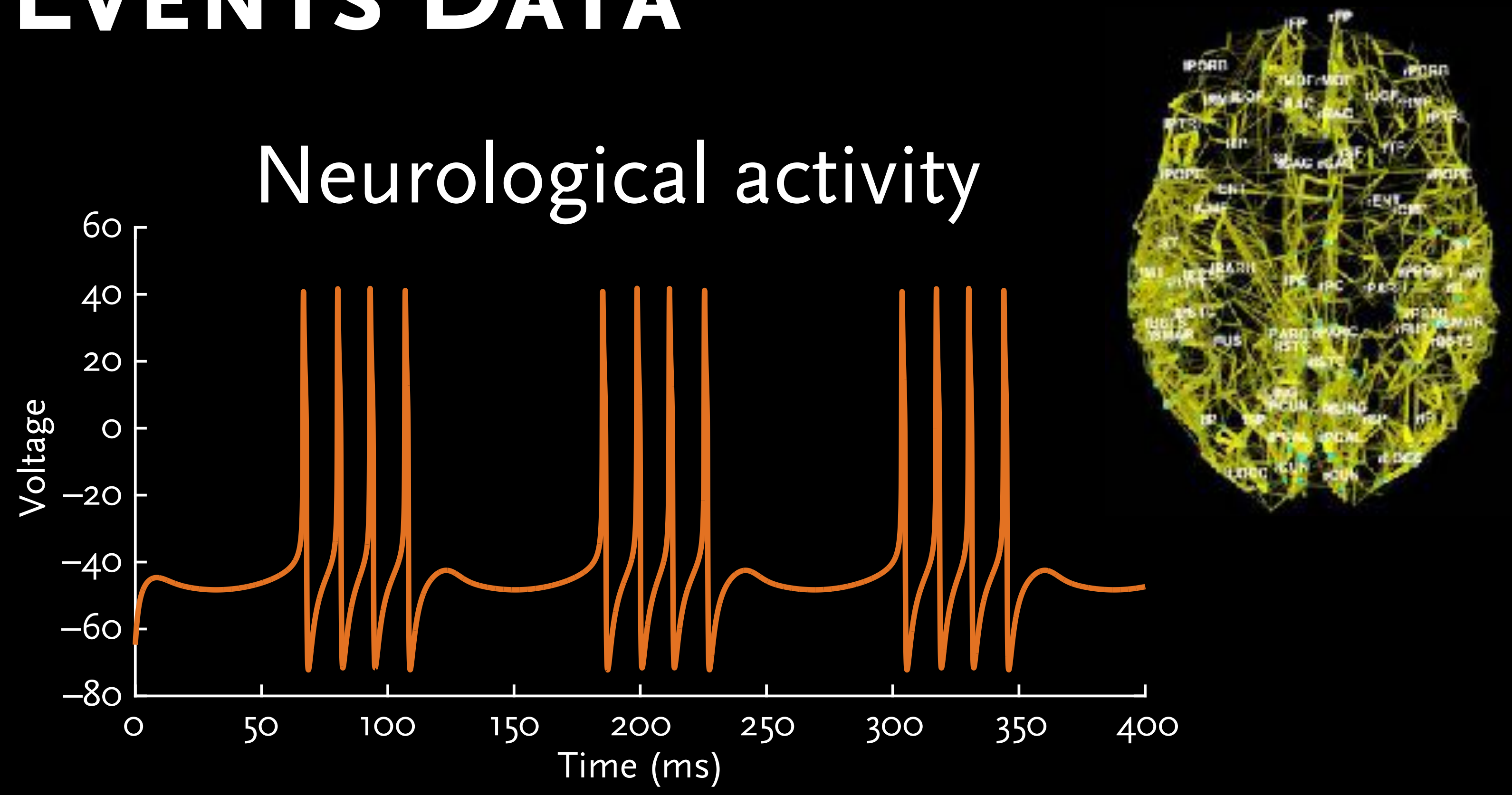
Tobias Braun

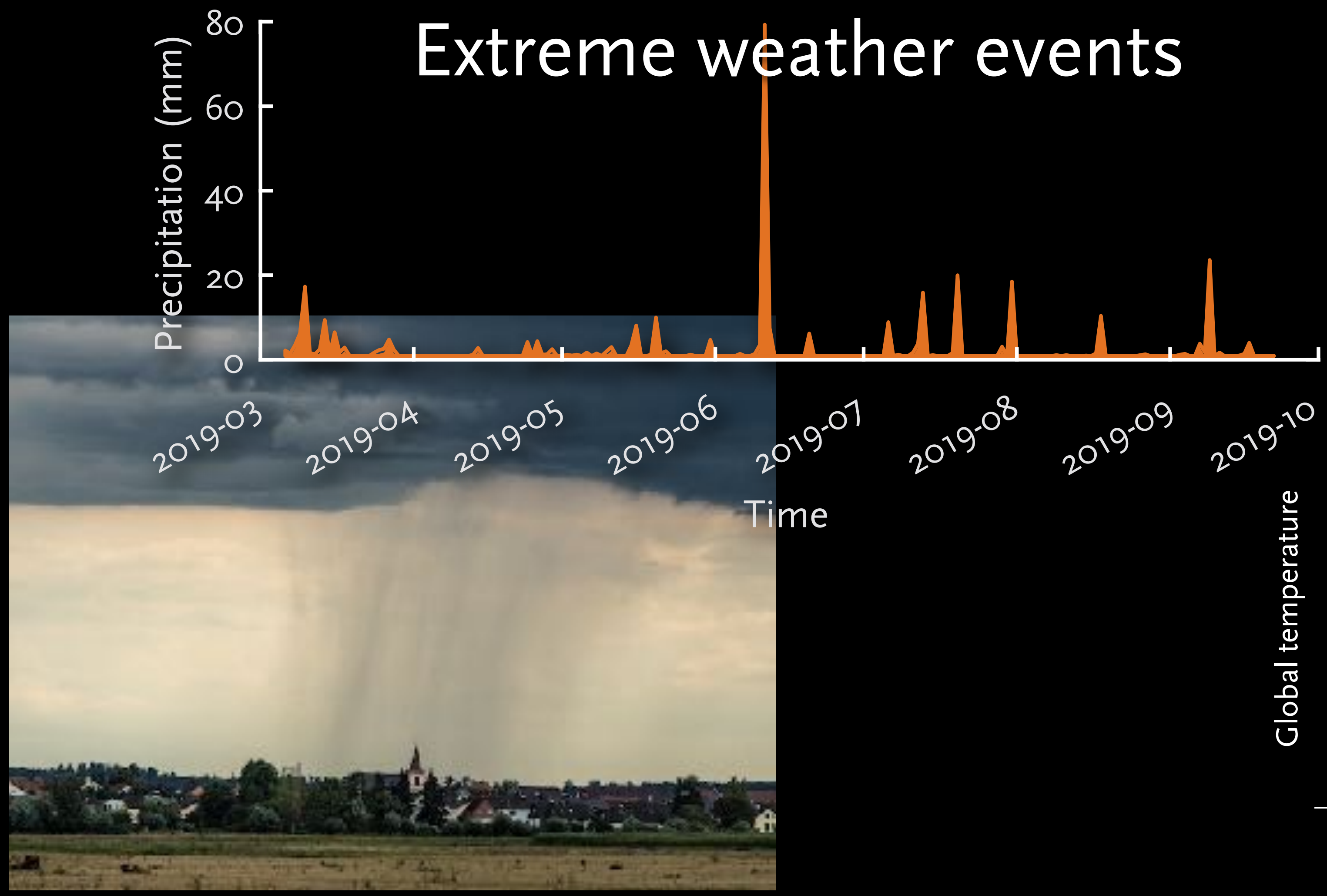
Power Spectrum Estimation for (Extreme) Events Data

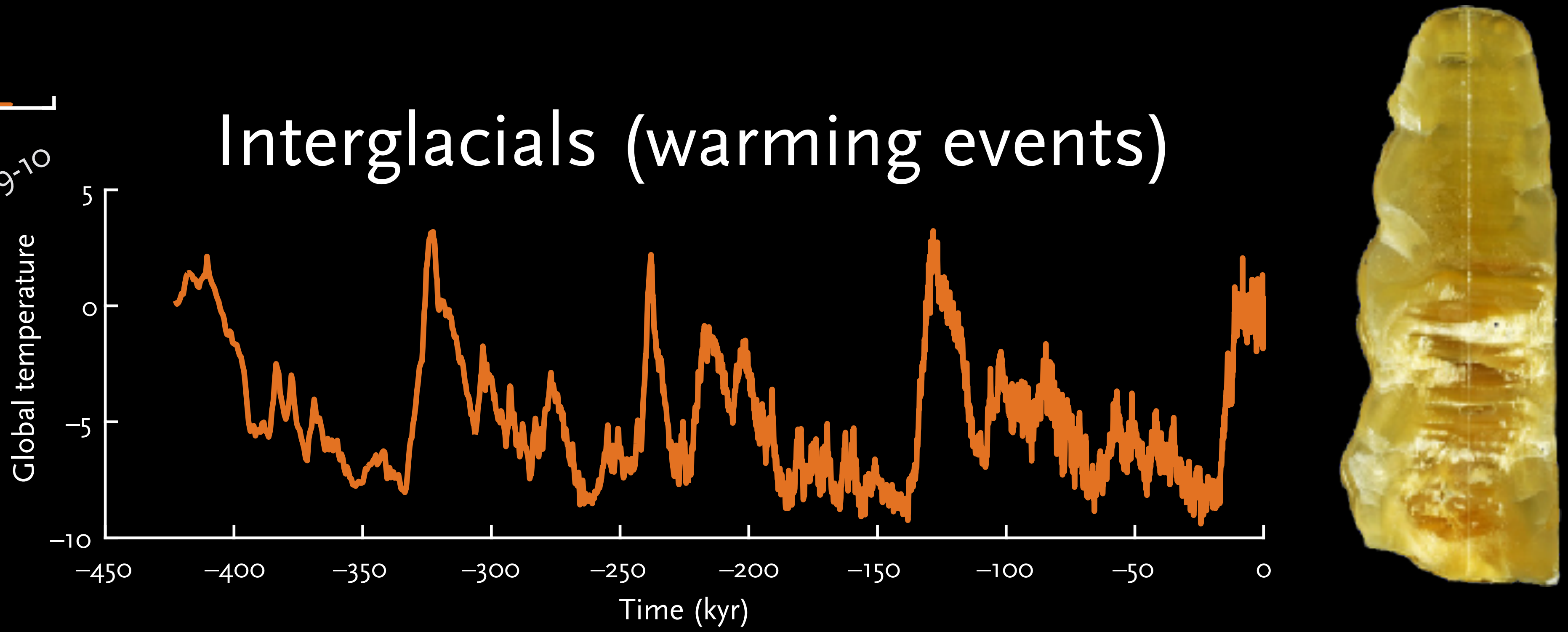
Leibniz-Gemeinschaft

(Extreme) Events Data

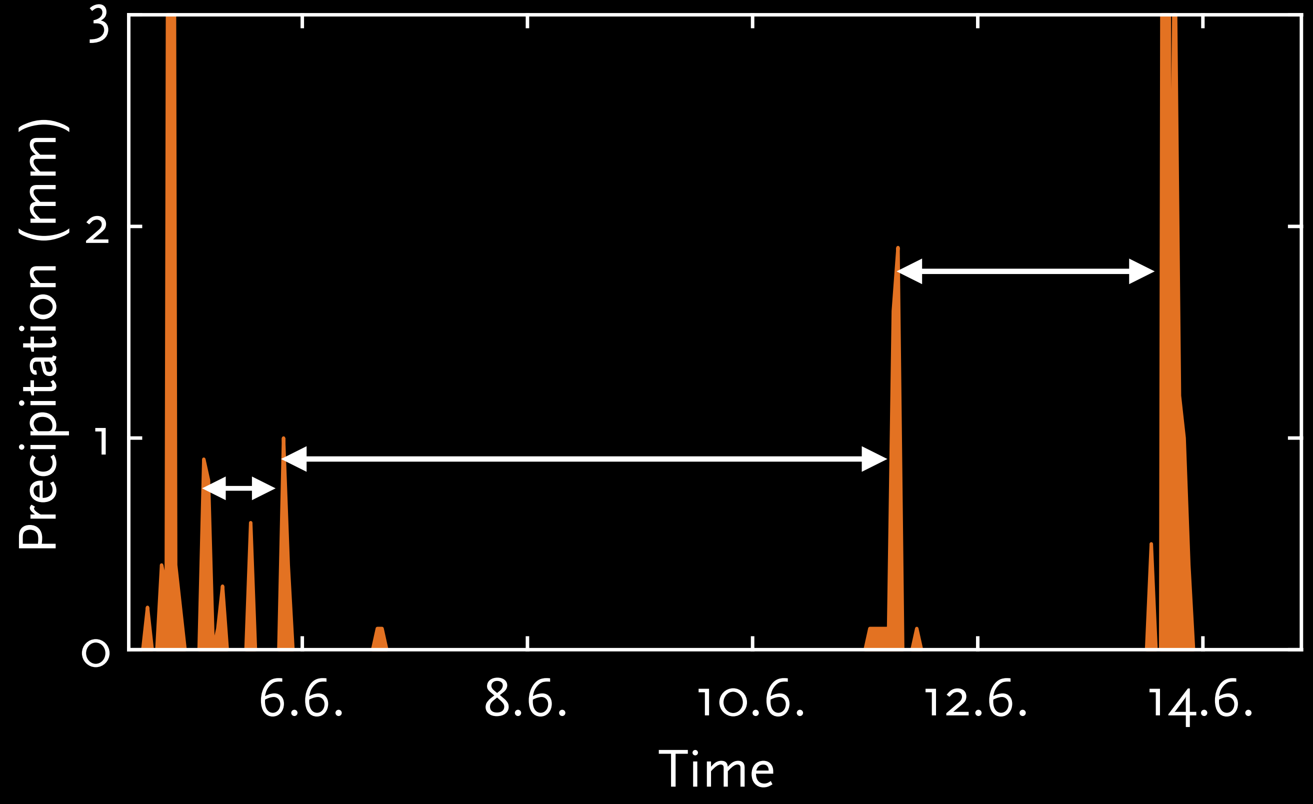


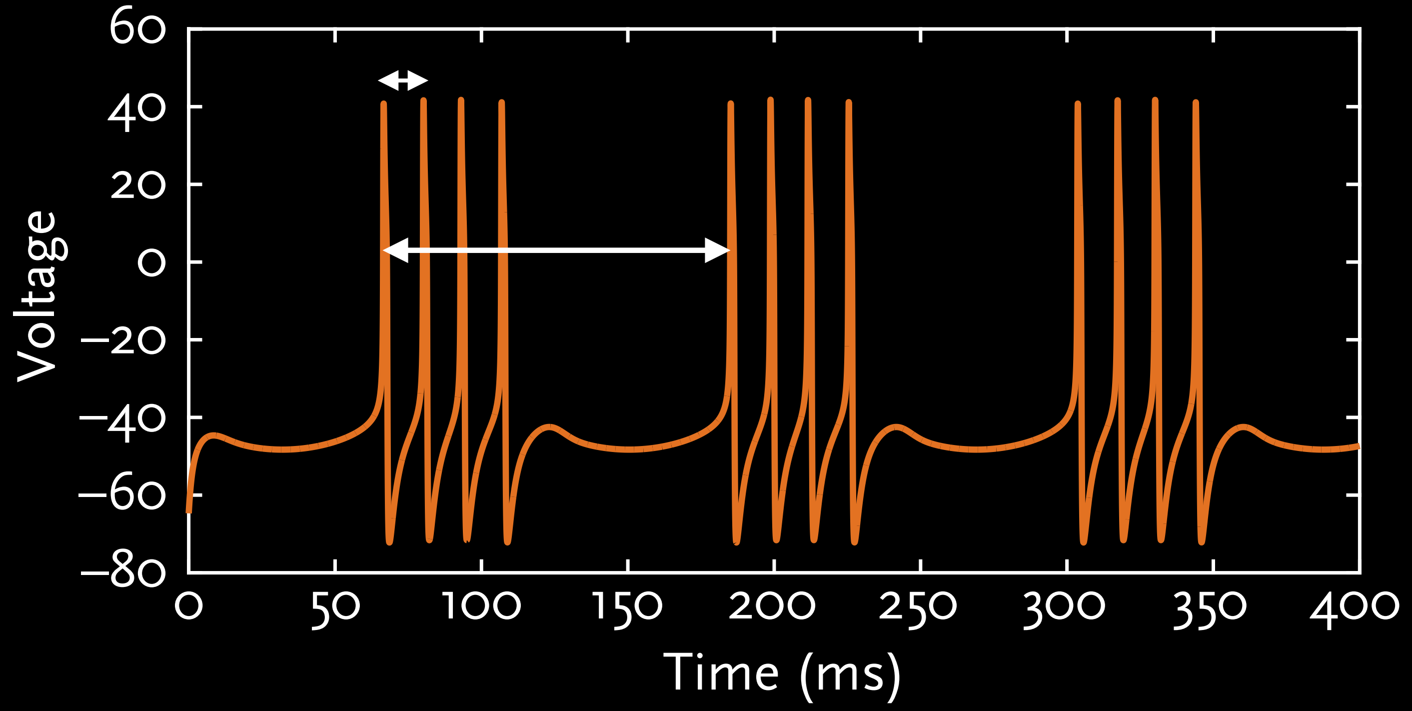






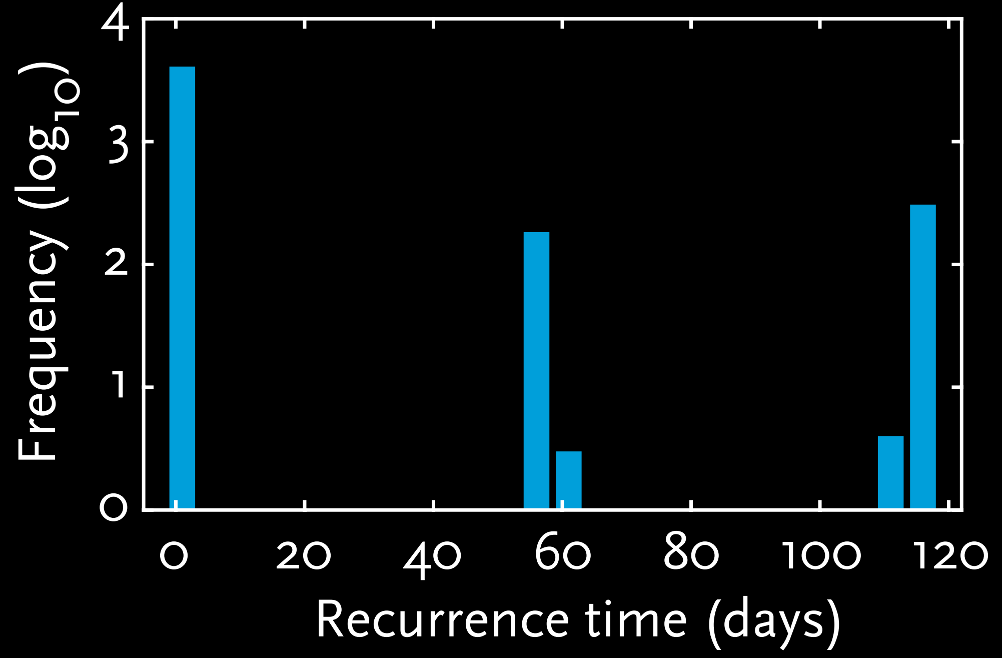
Time Scales of Event Occurrences?



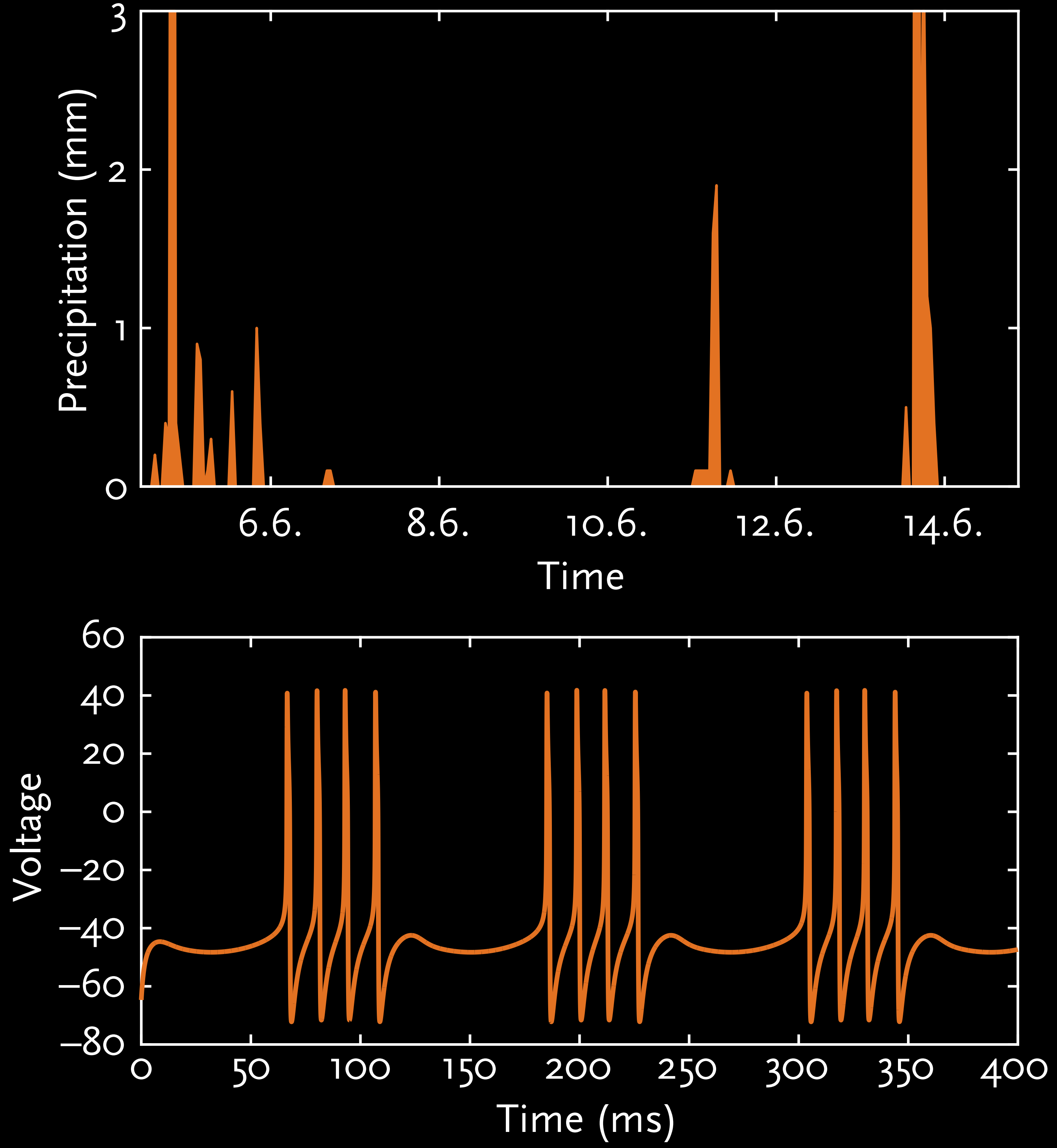


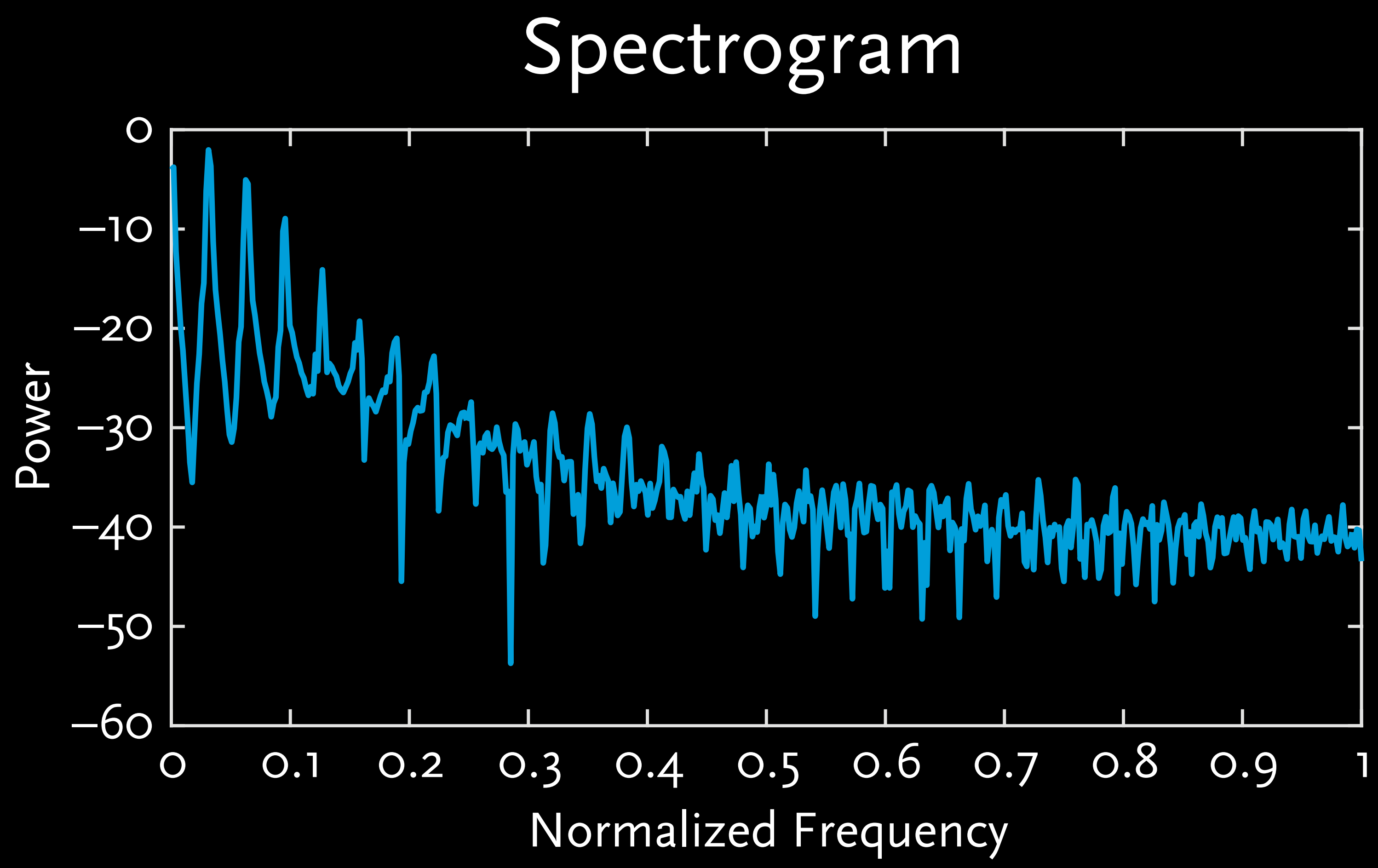
- Interspike intervals (recurrence times)

➡ No unique selection
Discretisation bias
Neglecting amplitude



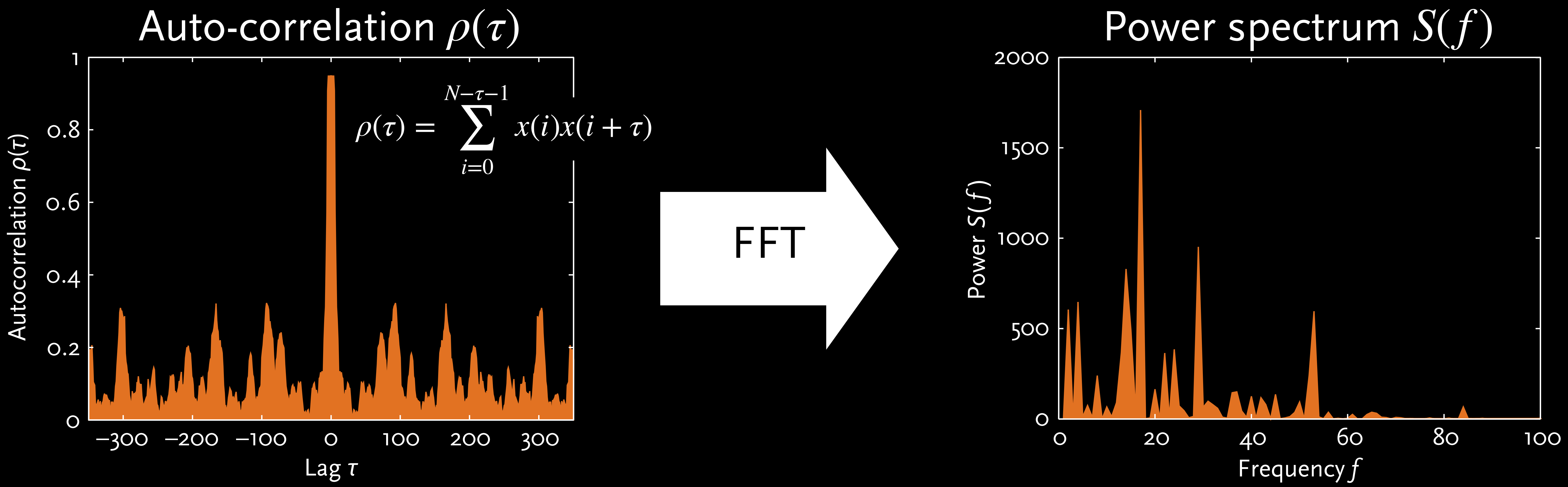
Time Scales of Event Occurrences?





- Simple, no/ few parameters, objective

Power Spectrum from Auto-Correlation

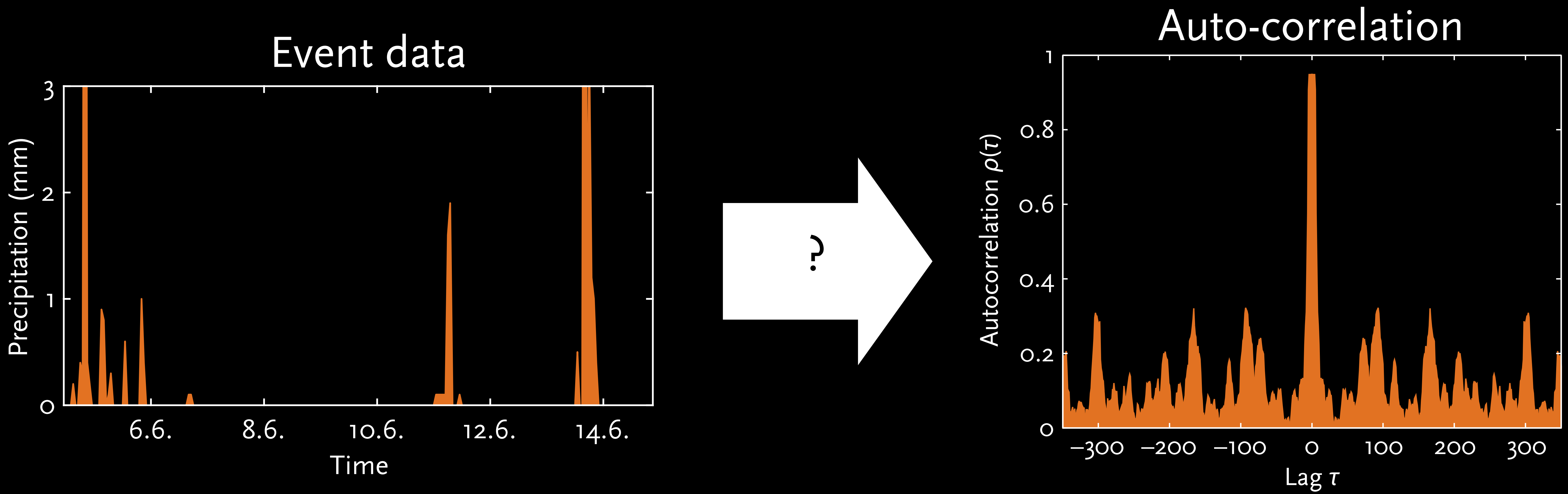


$$\rho(\tau) = \sum_{i=0}^{N-\tau-1} x(i)x(i+\tau)$$

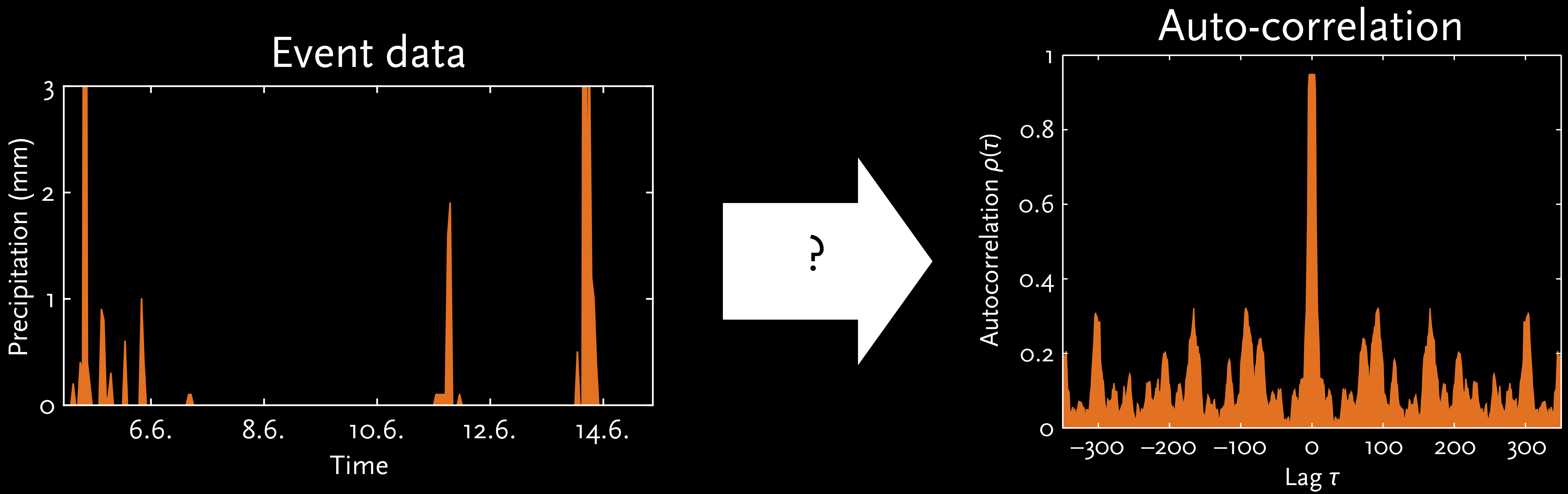
Wiener Khinchin theorem: $S(f) = \sum_{\tau=-\infty}^{\infty} \rho(\tau)e^{-i2\pi\tau f}$

Einstein, Archives sciences phys. 37, 1914; Wiener, Acta Math. 55, 1930; Khintchine, Math. Annalen 109, 1934

Auto-Correlation of Event Series



Auto-Correlation of Event Series

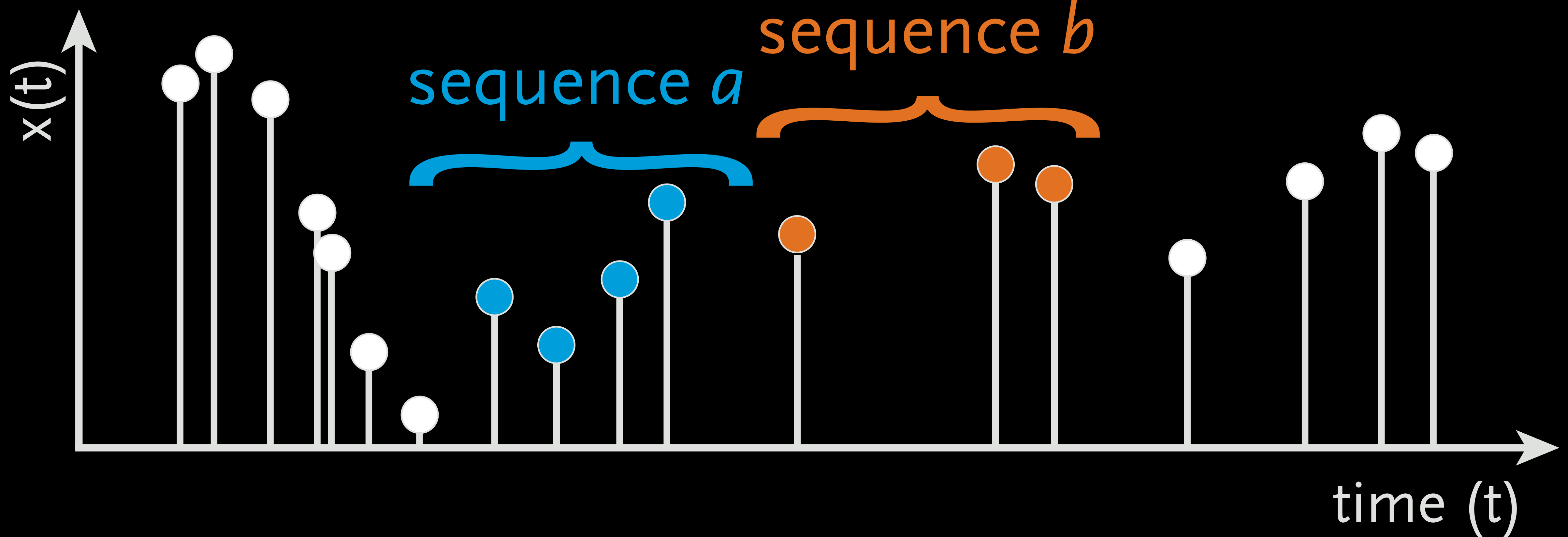


Similarity measure for event data $\rho\left(x(t), x(t+\tau)\right)$

(e.g., Levenshtein metric, event synchronisation, ...)

Edit Distance*

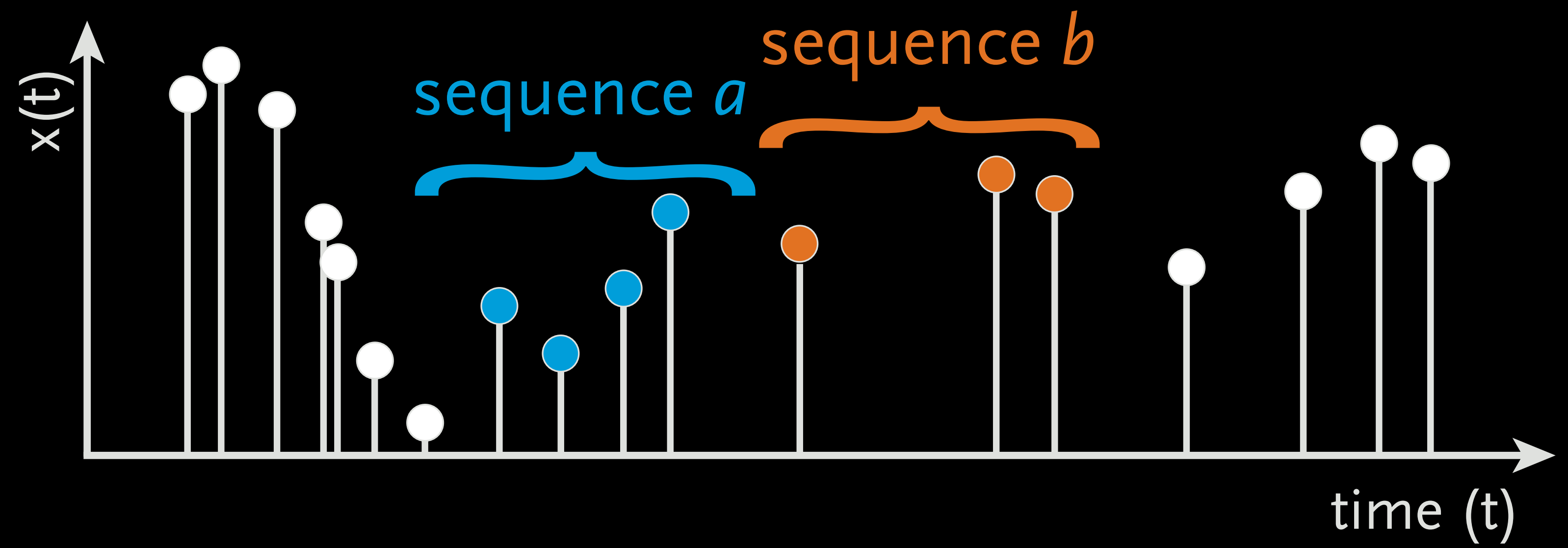
- Distance d = minimize the cost to transform sequence a to sequence b



Victor & Purpura, Network 8, 1997; Suzuki et al., Int. J. Bif. Chaos. 20, 2010

* modified Levenshtein distance

Edit Distance



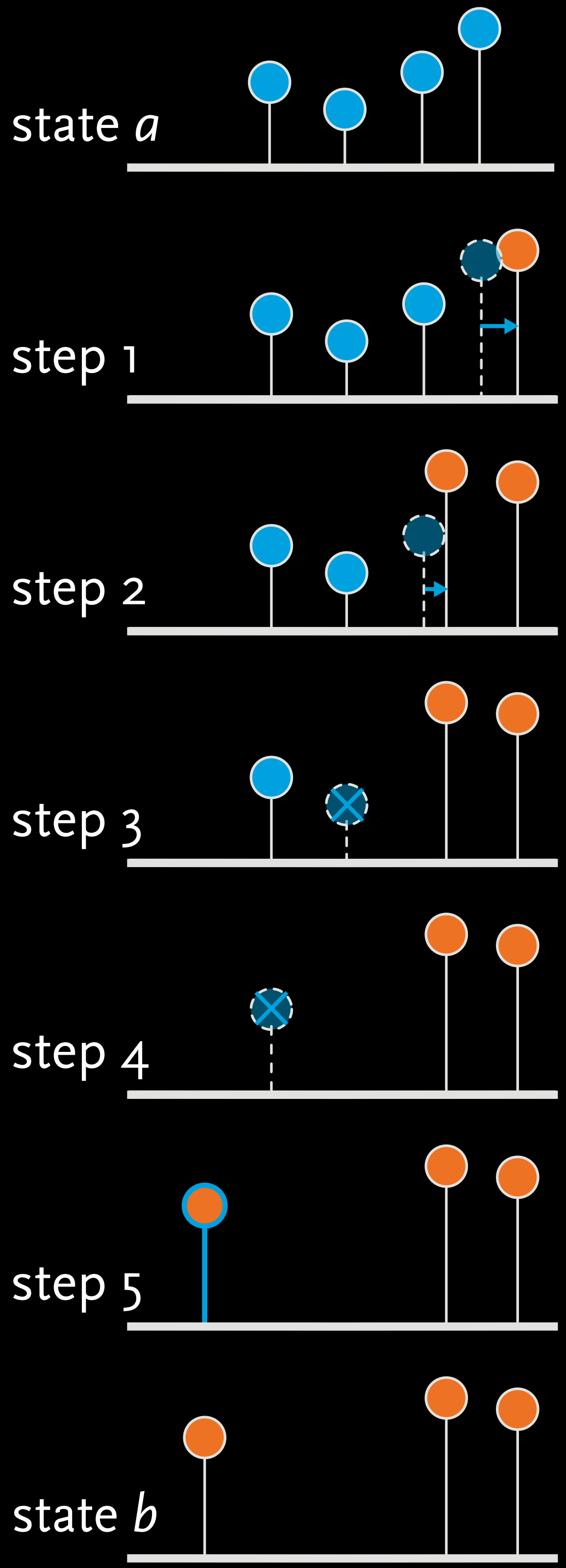
4 Operations:

(1) shifting (cost λ_0)

(2) adding (cost λ_s)

(3) deleting (cost λ_s)

(4) amplitude modification (cost λ_k)



Initial state
$d = 0$

Shifting + amplitude modif.
$+\lambda_0 \, |t_a(4) - t_b(3)| + \lambda_k \, |L_a(4) - L_b(3)|$

Shifting + amplitude modif.
$+\lambda_0 \, |t_a(3) - t_b(2)| + \lambda_k \, |L_a(3) - L_b(2)|$

Deleting
$+\lambda_s$

Deleting
$+\lambda_s$

Adding
$+\lambda_s$

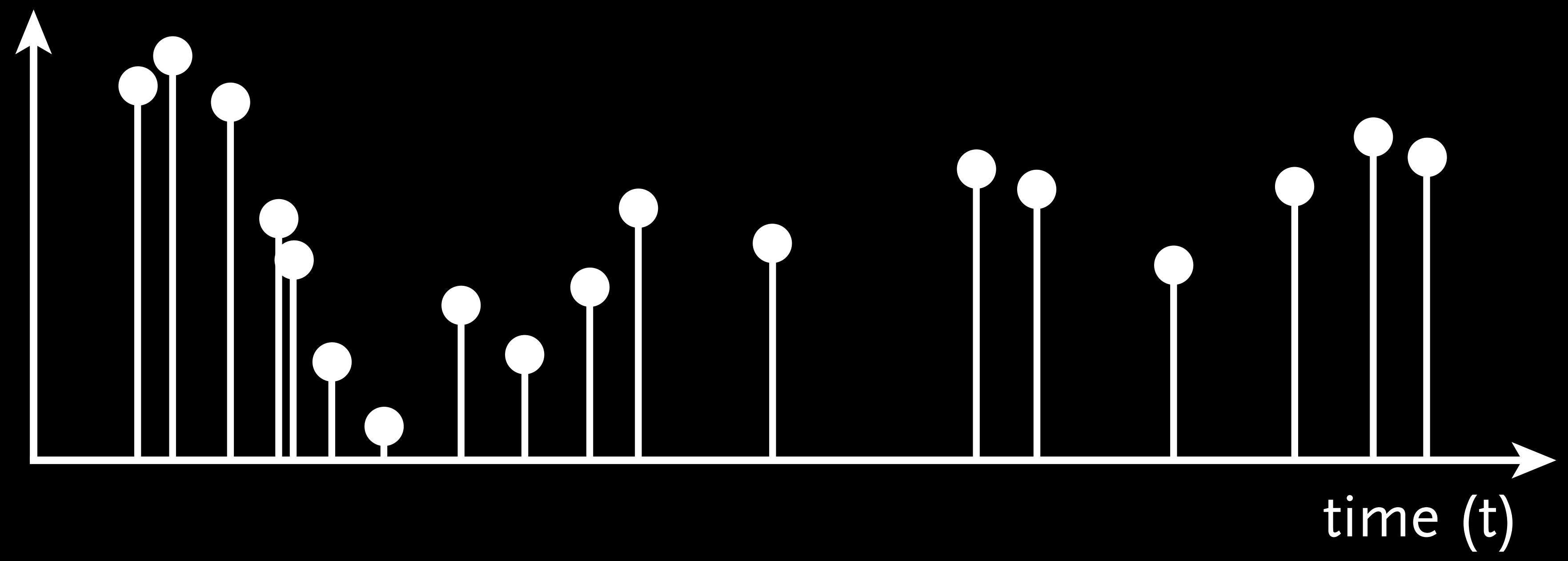
Final state
$= d(a,b) \to d(\vec{x}_i, \vec{x}_j)$

Suzuki et al., Int. J. Bif. Chaos. 20, 2010

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathcal{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta \in \mathcal{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \to d\left(\mathcal{S}, \mathcal{S}(\tau)\right)$$

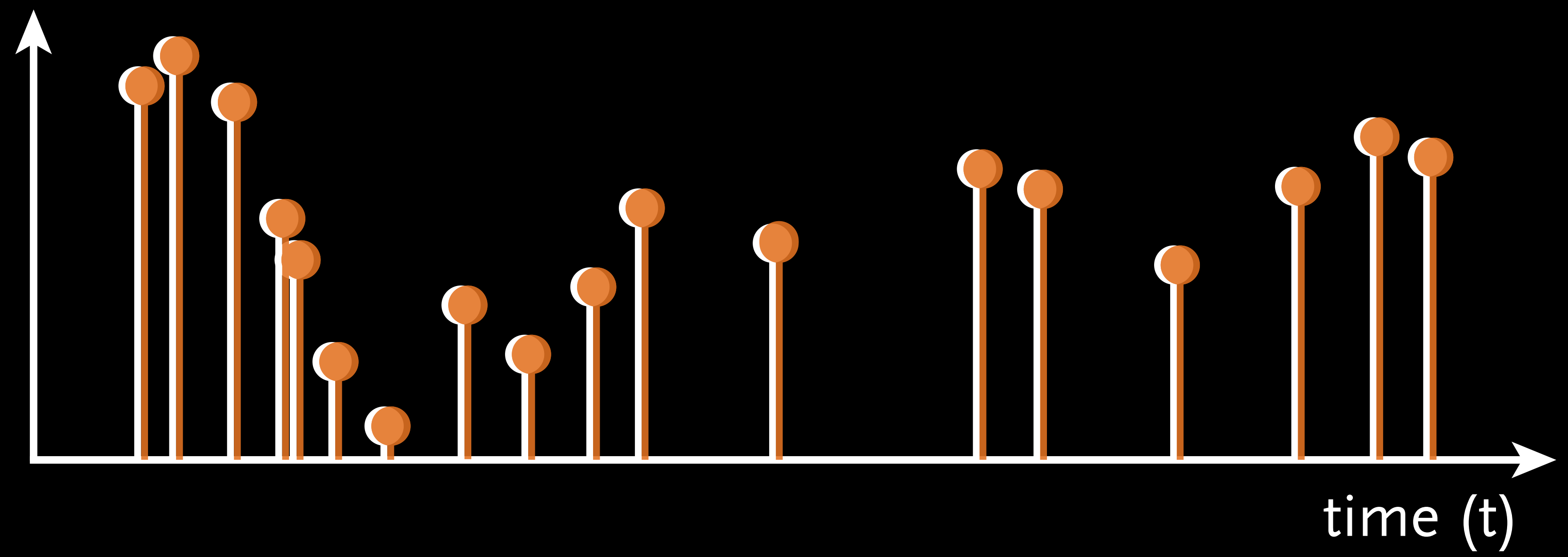


Marwan & Braun, Chaos 33, 2023

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathcal{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta \in \mathcal{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \rightarrow d\left(\mathcal{S}, \underline{\mathcal{S}(\tau)}\right)$$

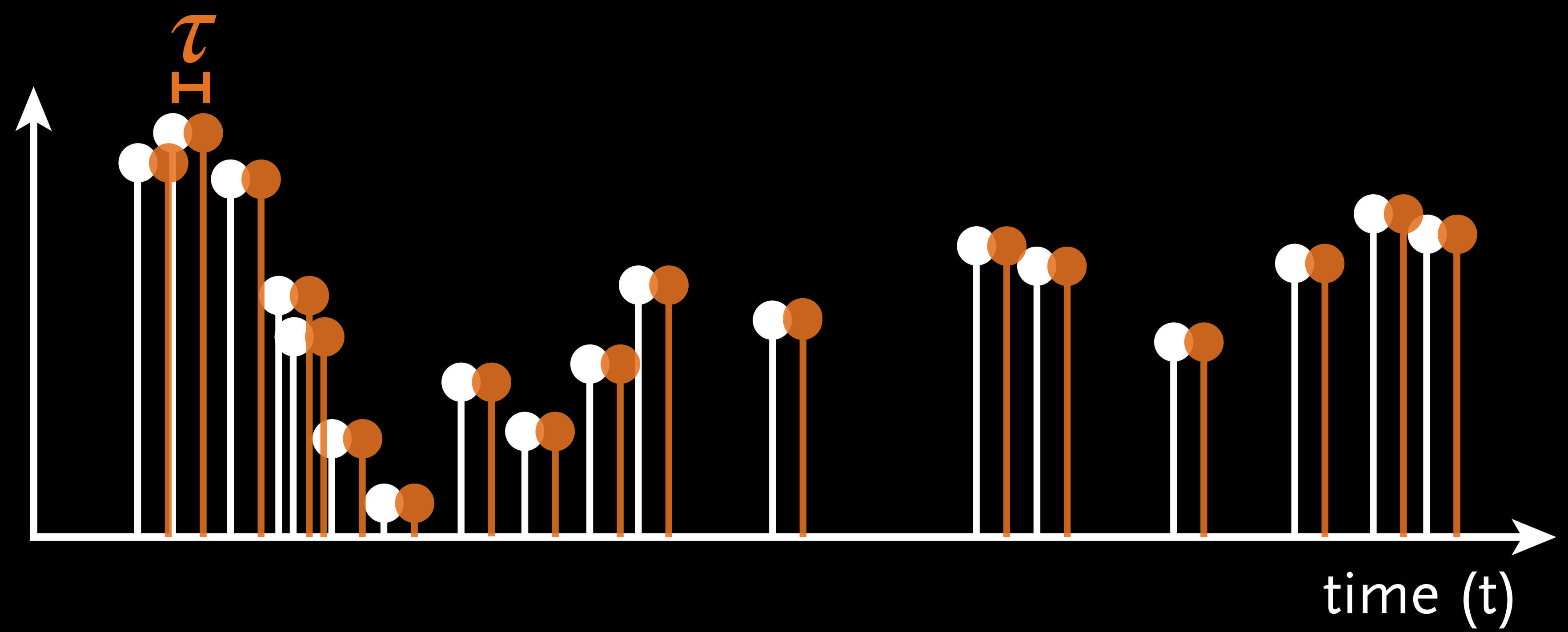


Marwan & Braun, Chaos 33, 2023

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathcal{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta \in \mathcal{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \to d\left(\mathcal{S}, \underline{\mathcal{S}(\tau)}\right)$$

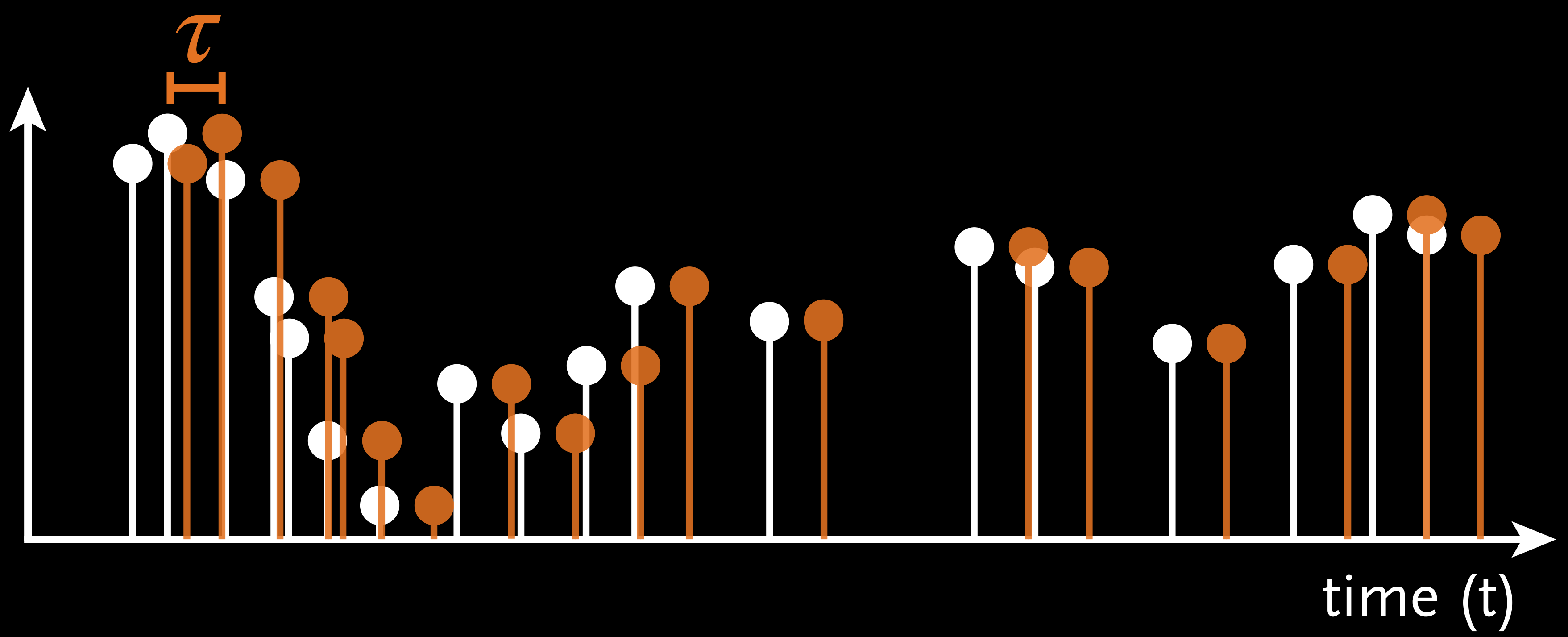


Marwan & Braun, Chaos 33, 2023

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathscr{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta \in \mathscr{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \rightarrow d\left(\mathcal{S}, \underline{\mathcal{S}(\tau)}\right)$$

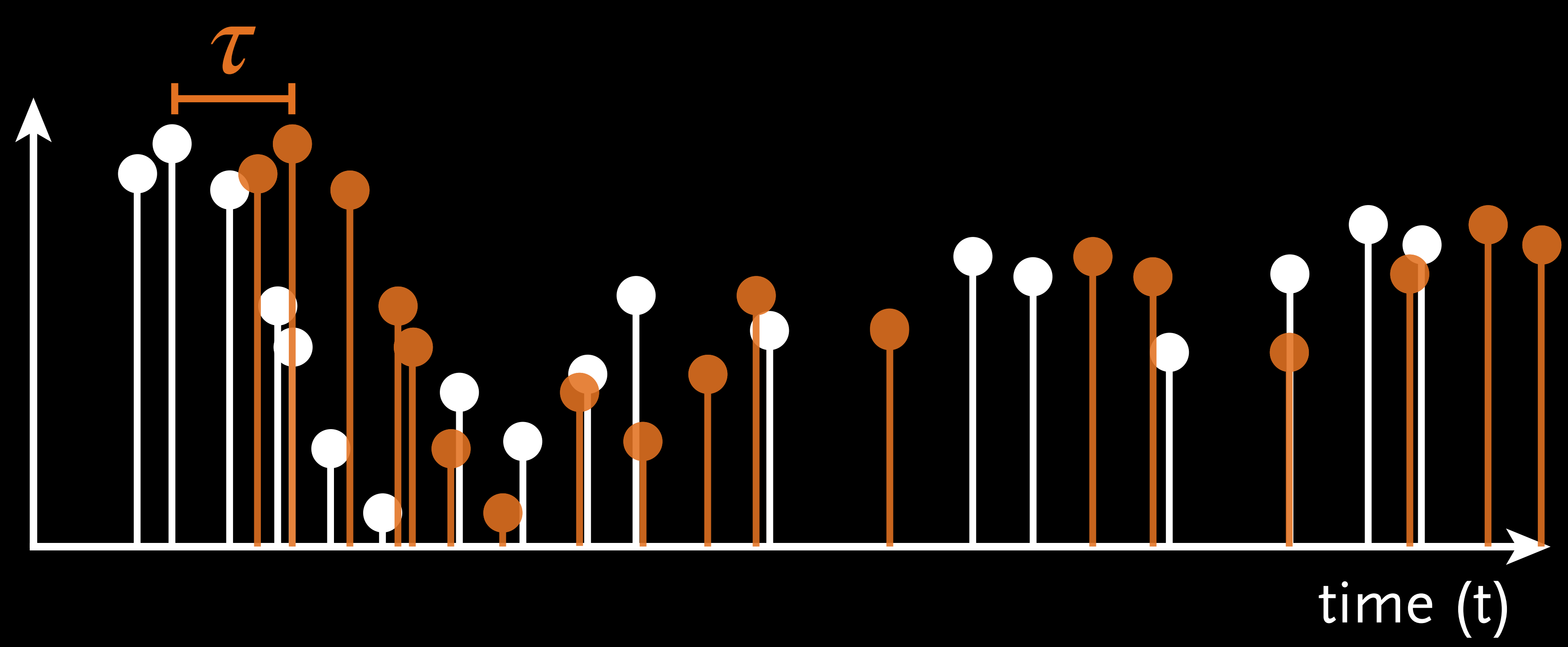


Marwan & Braun, Chaos 33, 2023

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathscr{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta\in\mathscr{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \rightarrow d\left(\mathcal{S}, \underline{\mathcal{S}(\tau)}\right)$$



Marwan & Braun, Chaos 33, 2023

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathcal{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta \in \mathcal{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \to d\left(\mathcal{S}, \mathcal{S}(\tau)\right)$$

$d\left(\mathcal{S}, \mathcal{S}(\tau)\right)$ – distance: $d = 0$: identical $d \gg 0$: different

$\rho\left(\mathcal{S}, \mathcal{S}(\tau)\right)$ – correlation: $\rho = 0$: different $\rho = 1$: identical

$$\rho\left(\mathcal{S}, \mathcal{S}(\tau)\right) \to 1 - \tilde{d}\left(\mathcal{S}, \mathcal{S}(\tau)\right)$$

$$\tilde{d} = \frac{d}{\max(d)}$$

Marwan & Braun, Chaos 33, 2023

Edit Distance as Auto-covariance

Edit distance (for binary events):

$$d(\mathcal{S}_a, \mathcal{S}_b) = \min \left\{ \underbrace{\Lambda_S \left(N_a + N_b - 2|\mathcal{C}| \right)}_{\text{adding and deleting}} + \sum_{\alpha,\beta \in \mathcal{C}} \underbrace{\Lambda_0 \left\| t_\alpha^{(a)} - t_\beta^{(b)} \right\|}_{\text{shifting}} \right\} \to d\left(\mathcal{S}, \mathcal{S}(\tau)\right)$$

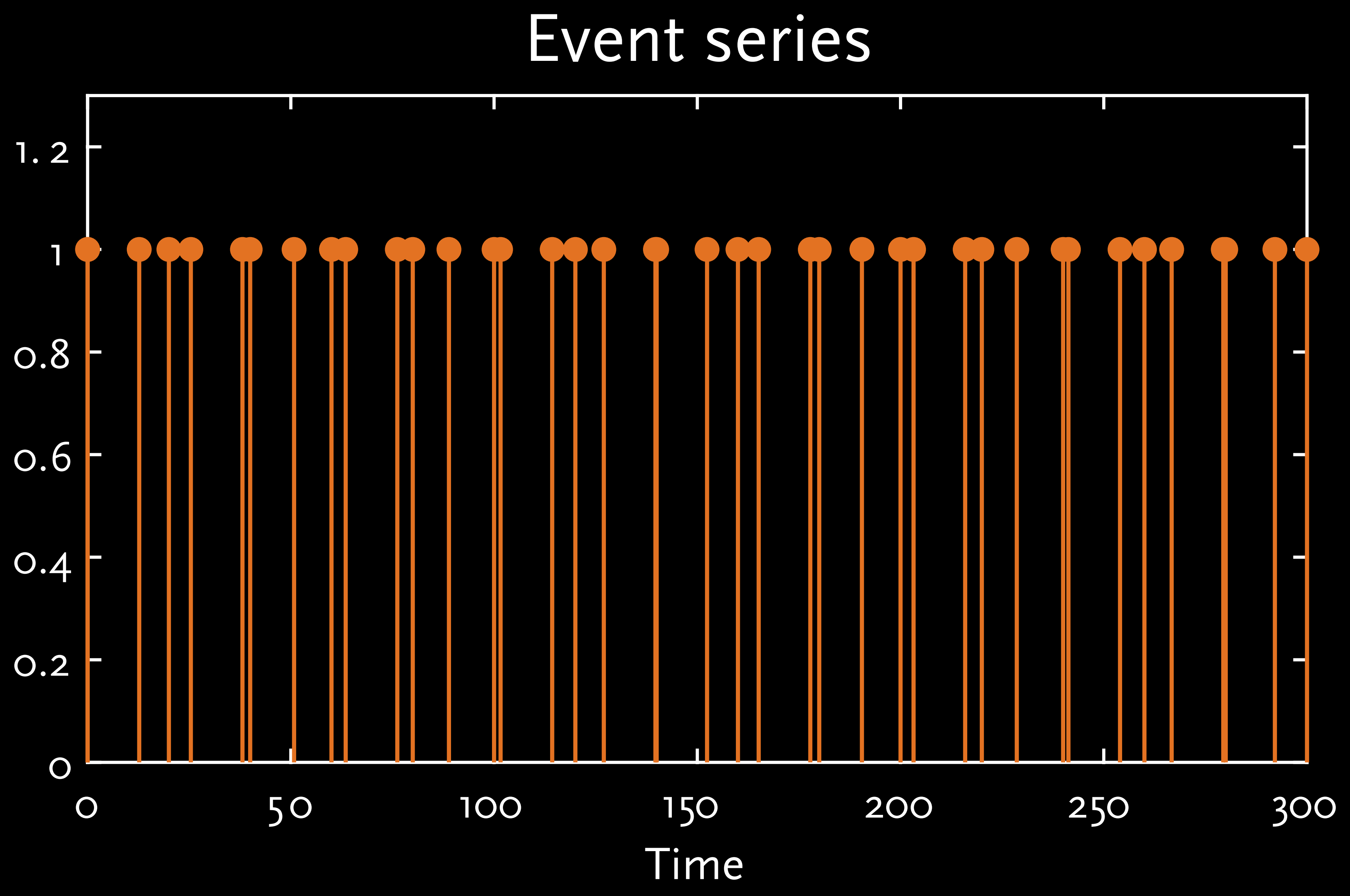
Combine with Wiener Khinchin theorem:

$$S_{\mathcal{S}}^{\text{edit}}(f) = \sum_{\tau=-\infty}^{\infty} \frac{\left(1 - \tilde{d}\left(\mathcal{S}, \mathcal{S}(\tau)\right)\right) - \left\langle 1 - \tilde{d}\left(\mathcal{S}, \mathcal{S}(\tau)\right) \right\rangle}{\text{std}\left(1 - \tilde{d}\left(\mathcal{S}, \mathcal{S}(\tau)\right)\right)} \mathrm{e}^{-j2\pi f\tau}$$

Marwan & Braun, Chaos 33, 2023

Edit Distance-based Spectrum

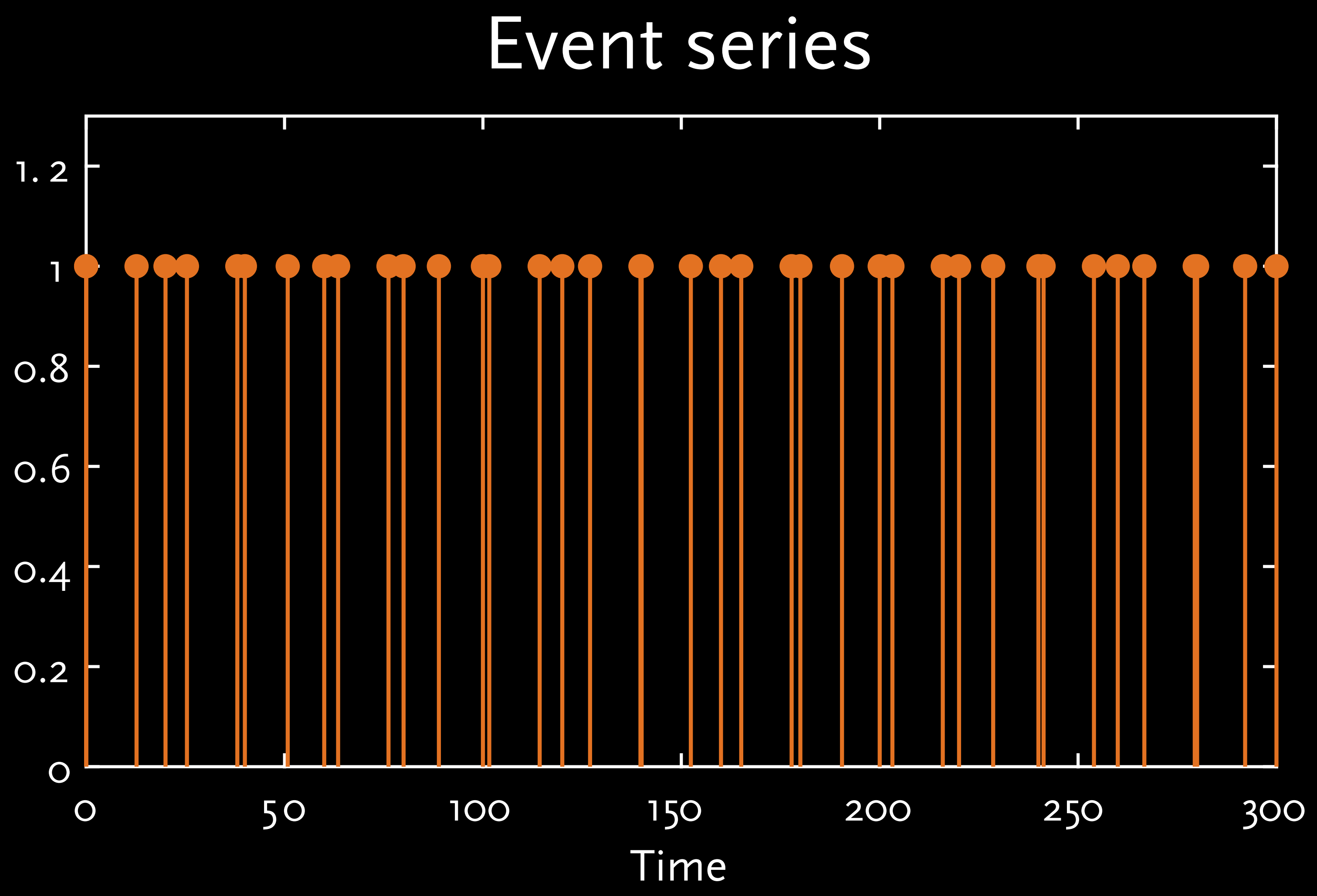
Event series with two frequencies ($\omega_1 = 1/20$, $\omega_2 = 1/12.7$):

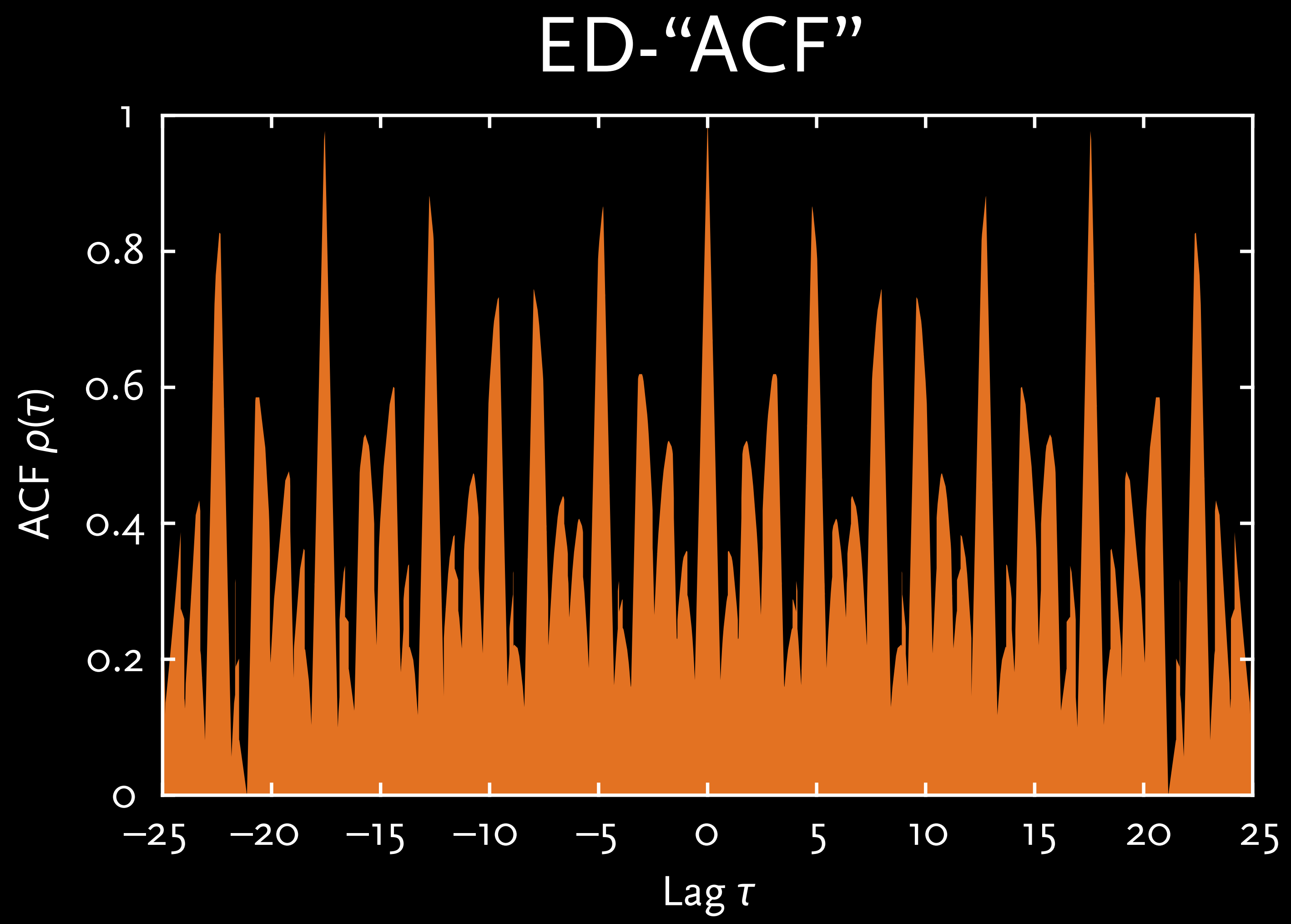


Marwan & Braun, Chaos 33, 2023

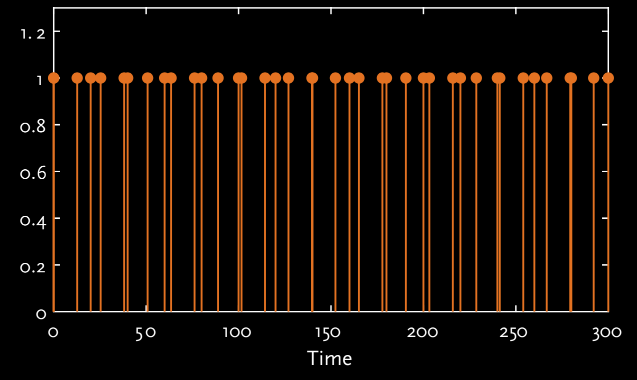
Edit Distance-based Spectrum

Event series with two frequencies ($\omega_1 = 1/20$, $\omega_2 = 1/12.7$):



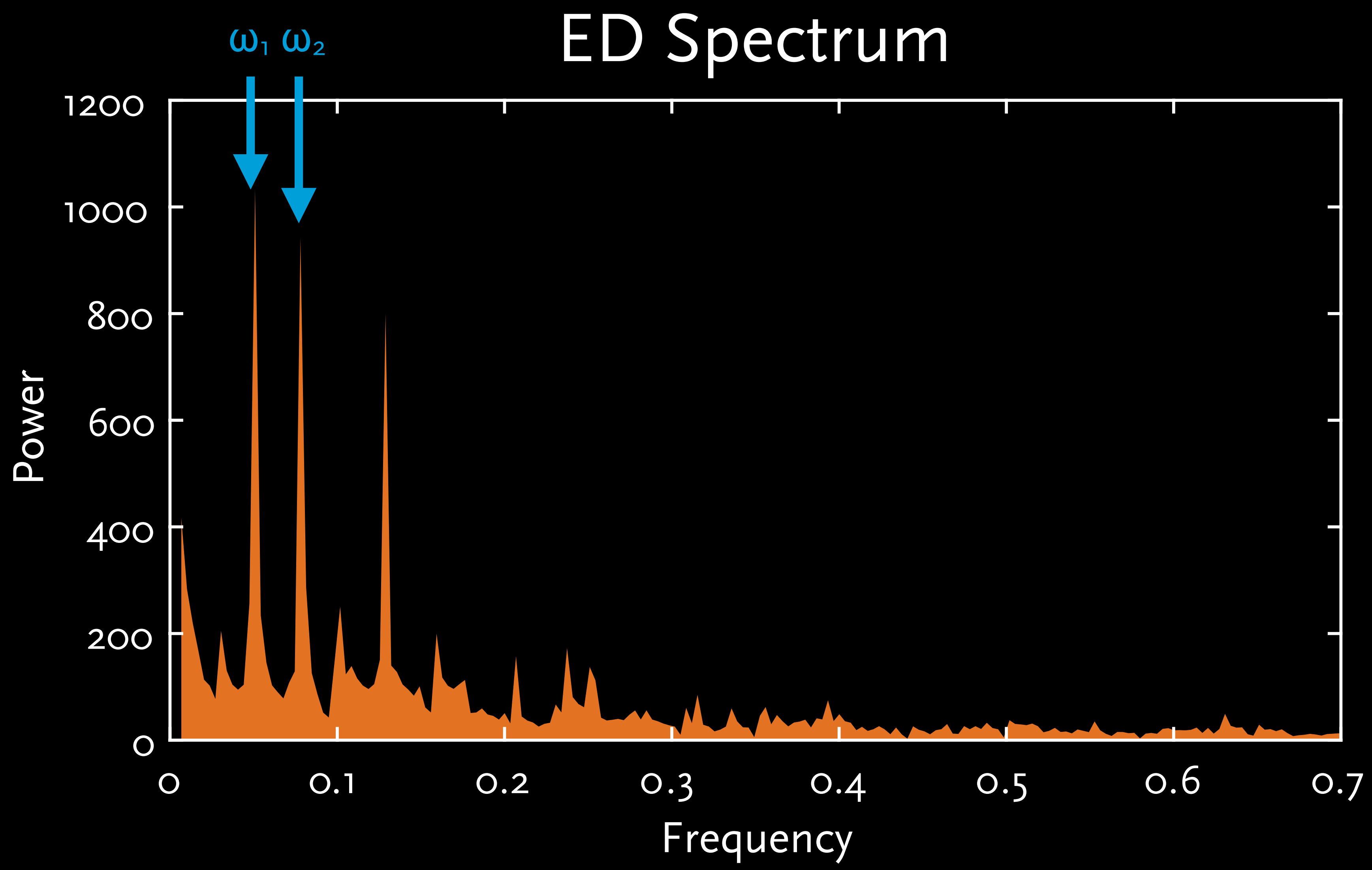


Marwan & Braun, Chaos 33, 2023

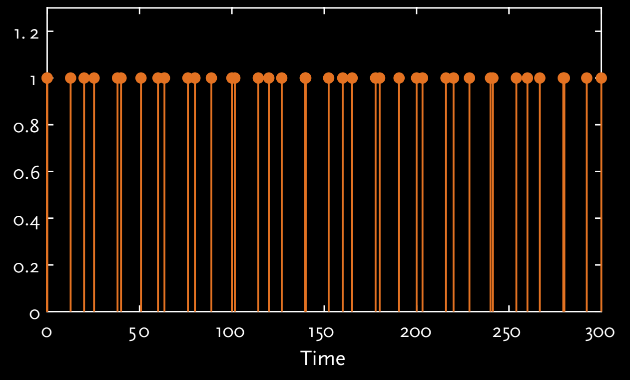


Edit Distance-based Spectrum

Event series with two frequencies ($\omega_1 = 1/20$, $\omega_2 = 1/12.7$):



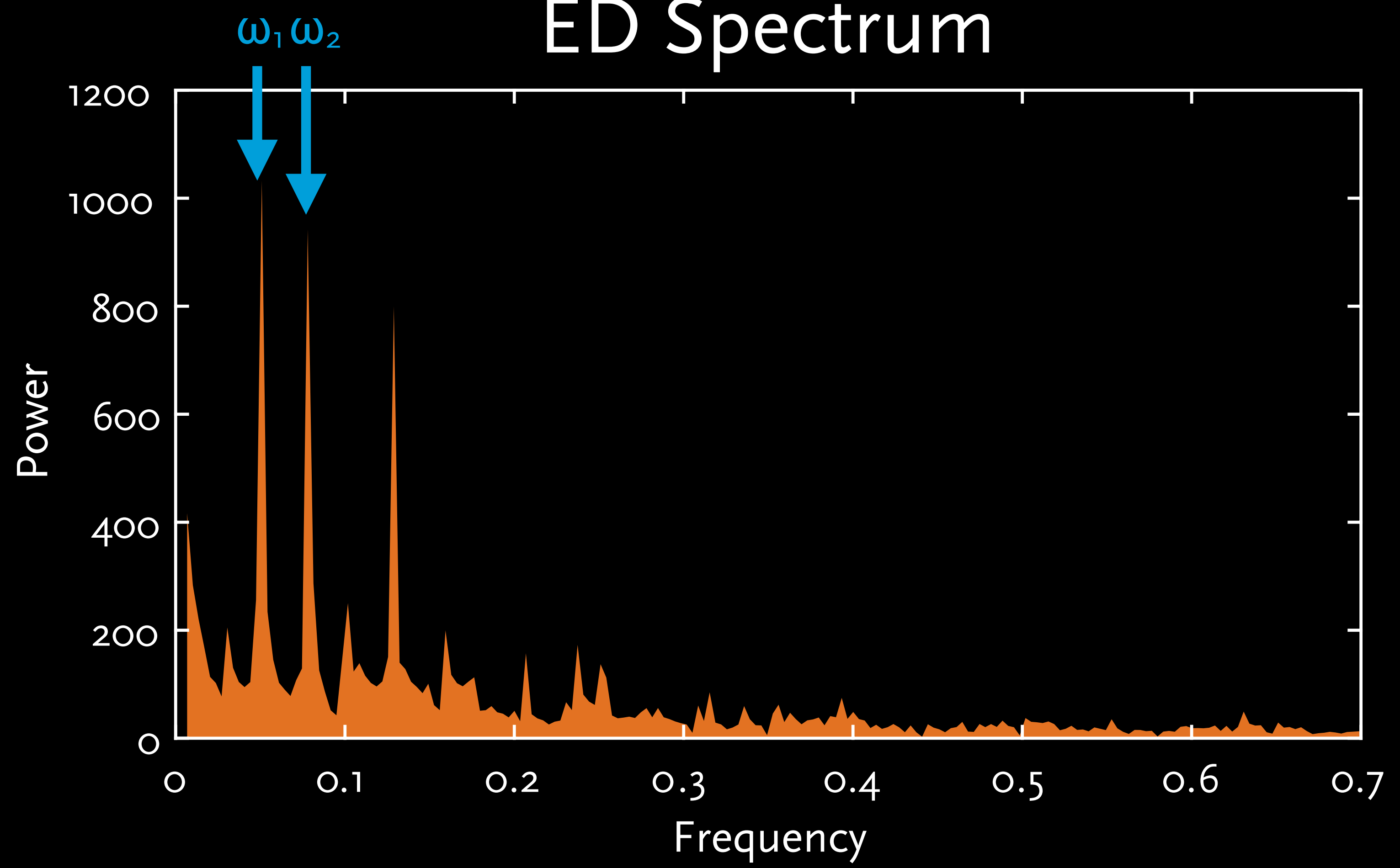
Marwan & Braun, Chaos 33, 2023



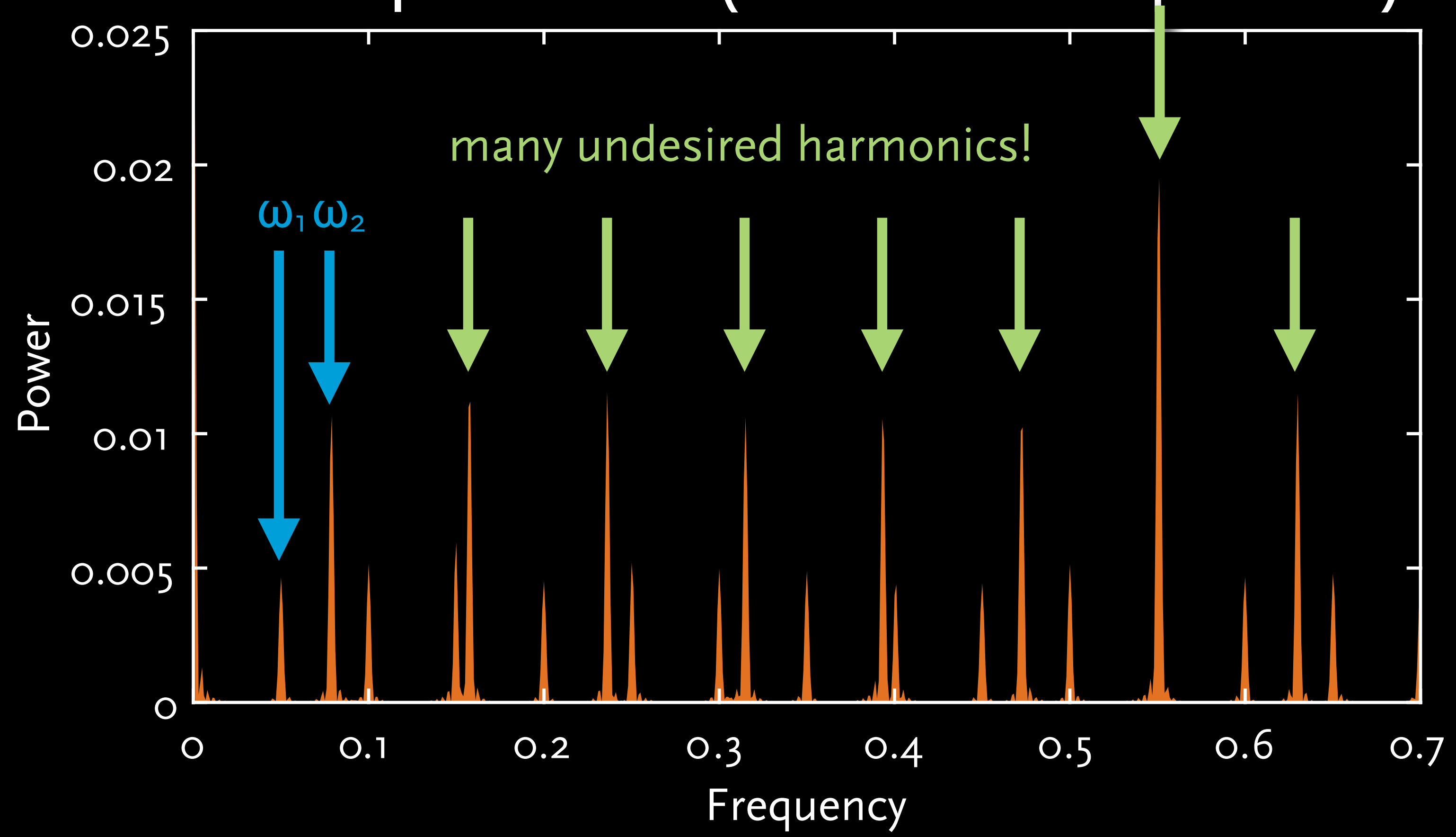
Edit Distance-based Spectrum

Event series with two frequencies ($\omega_1 = 1/20$, $\omega_2 = 1/12.7$):

ED Spectrum

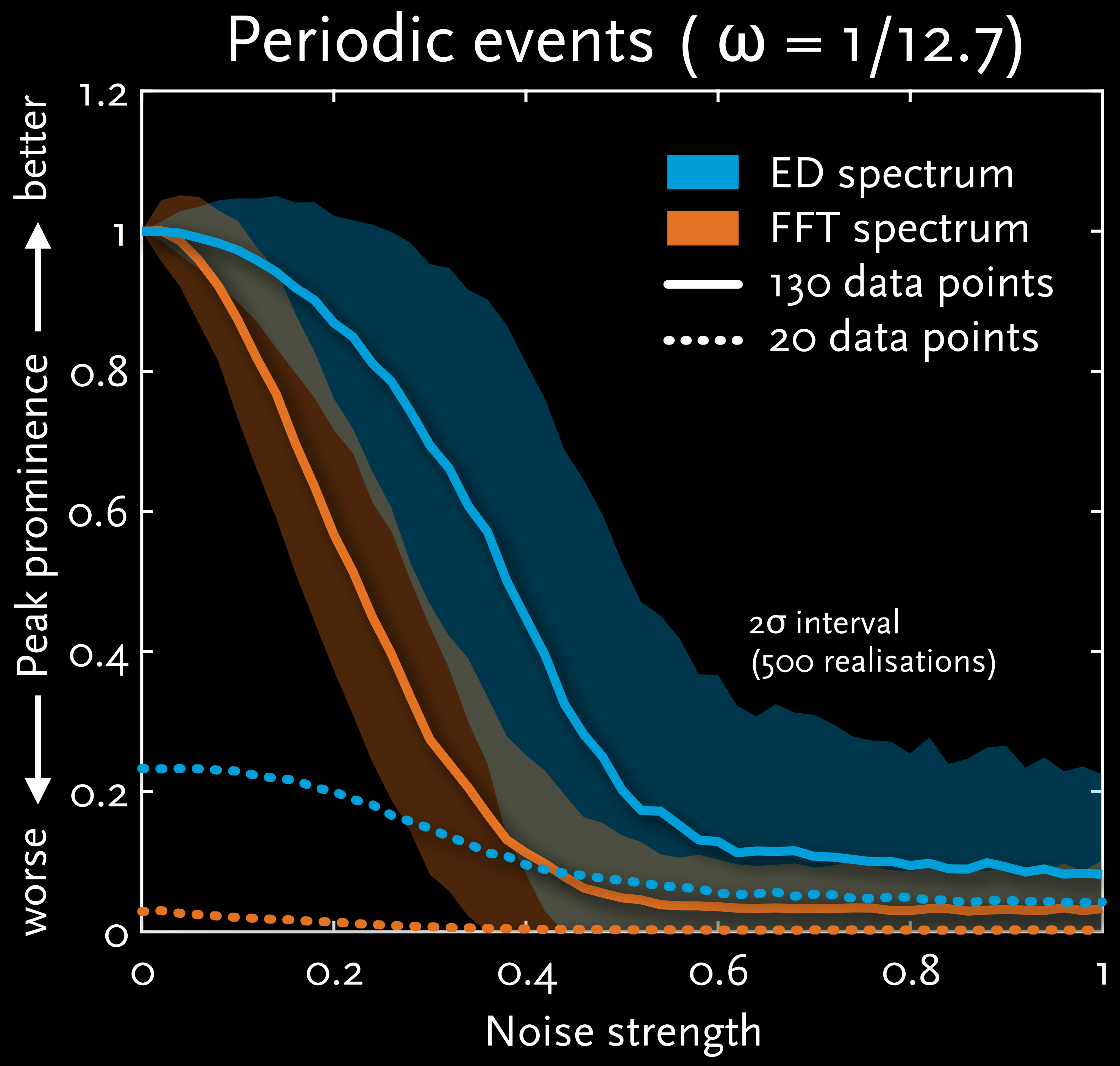


FFT Spectrum (of event sequence)



Marwan & Braun, Chaos 33, 2023

Influence of Length and Noise

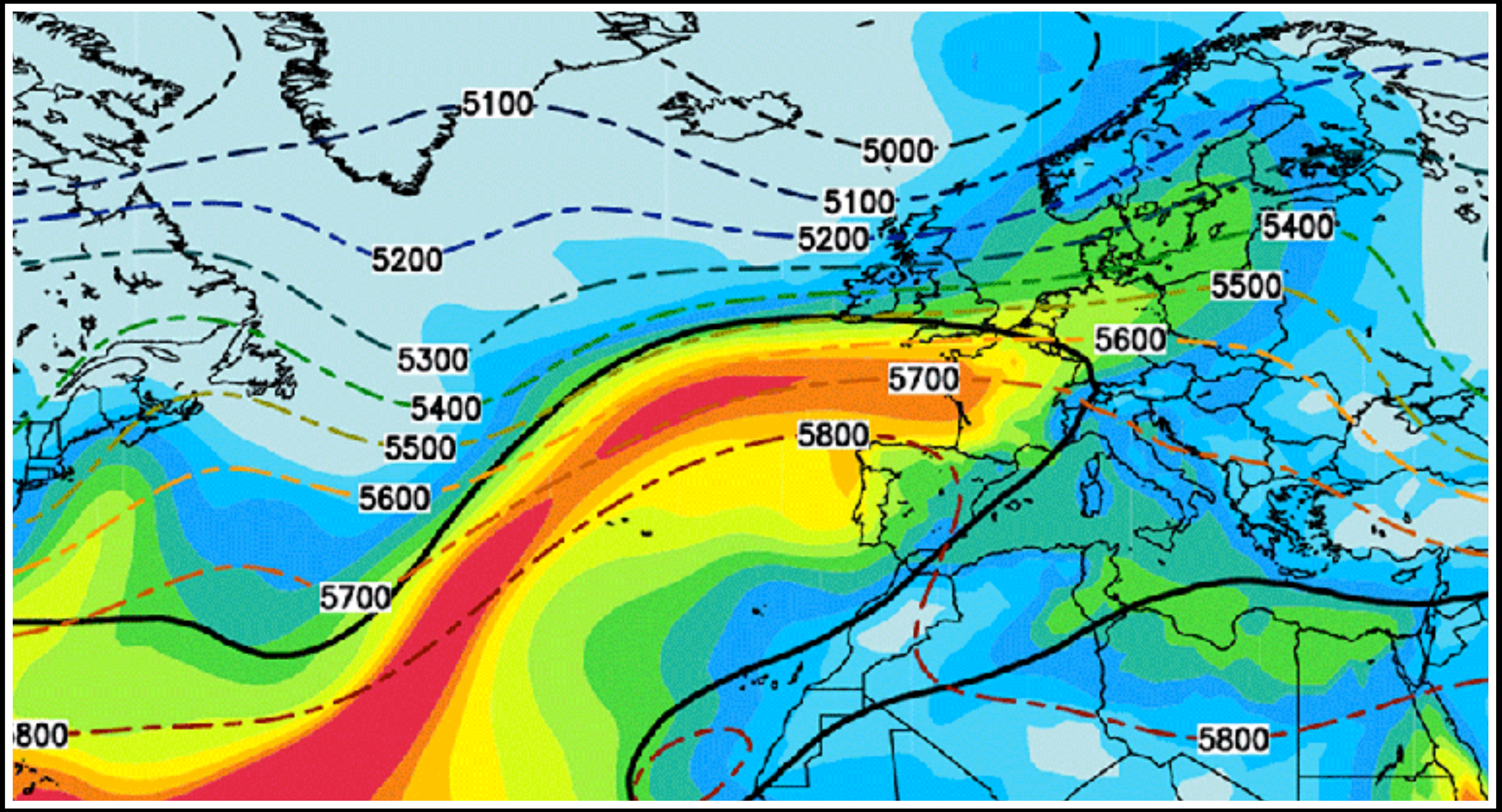


- Peak prominence: difference between peak and its baseline
- ED spectrum:
  - less sensitive to noise
  - more robust even for shorter time series

Marwan & Braun, Chaos 33, 2023

Atmospheric Rivers

Integrated water vapor transport (IVT)



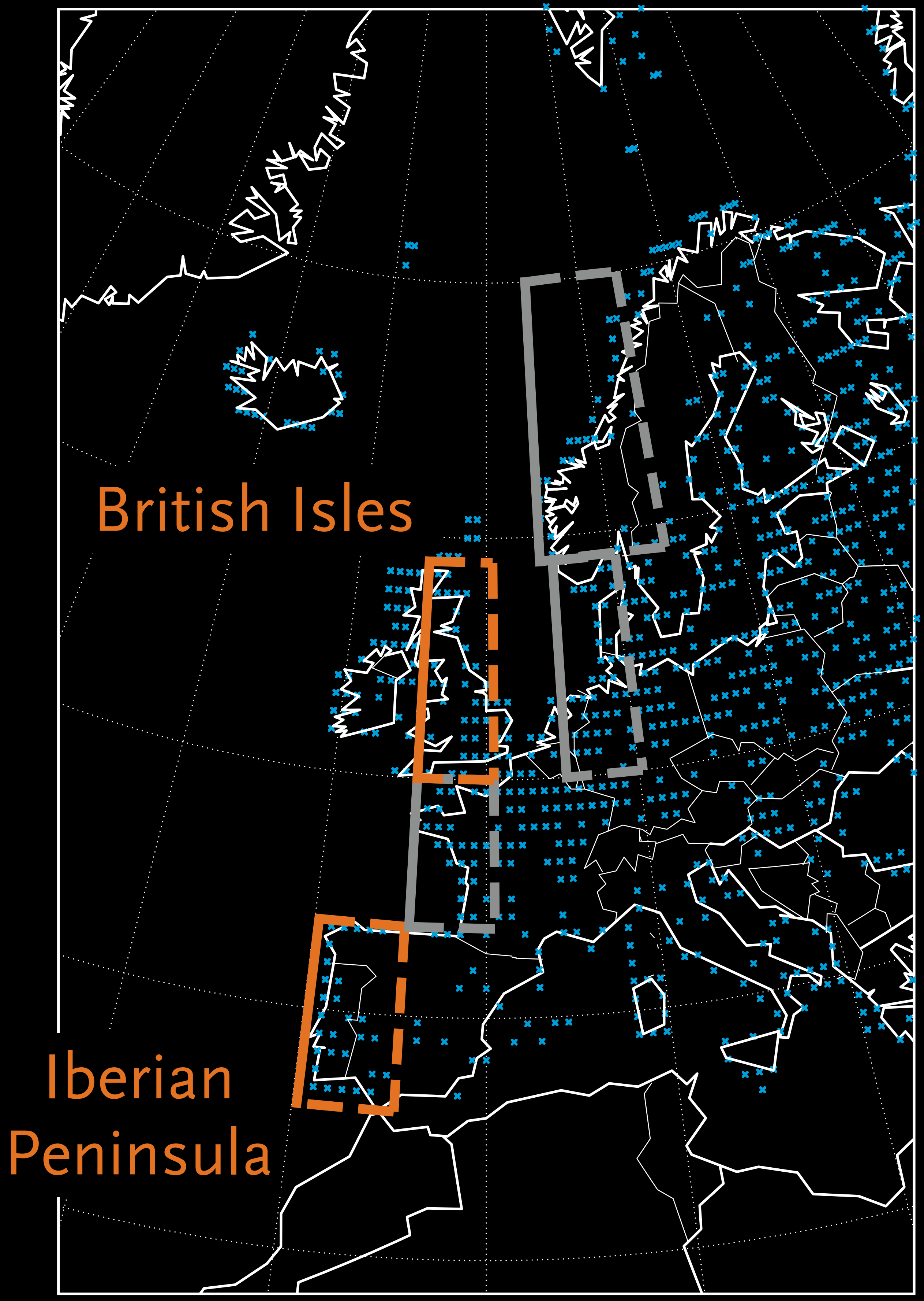
Dec 19, 1993

- Important source of water supply
- Extreme rainfall
- Flash floods
- Landslides

Ionita et al., HESS 24, 2020

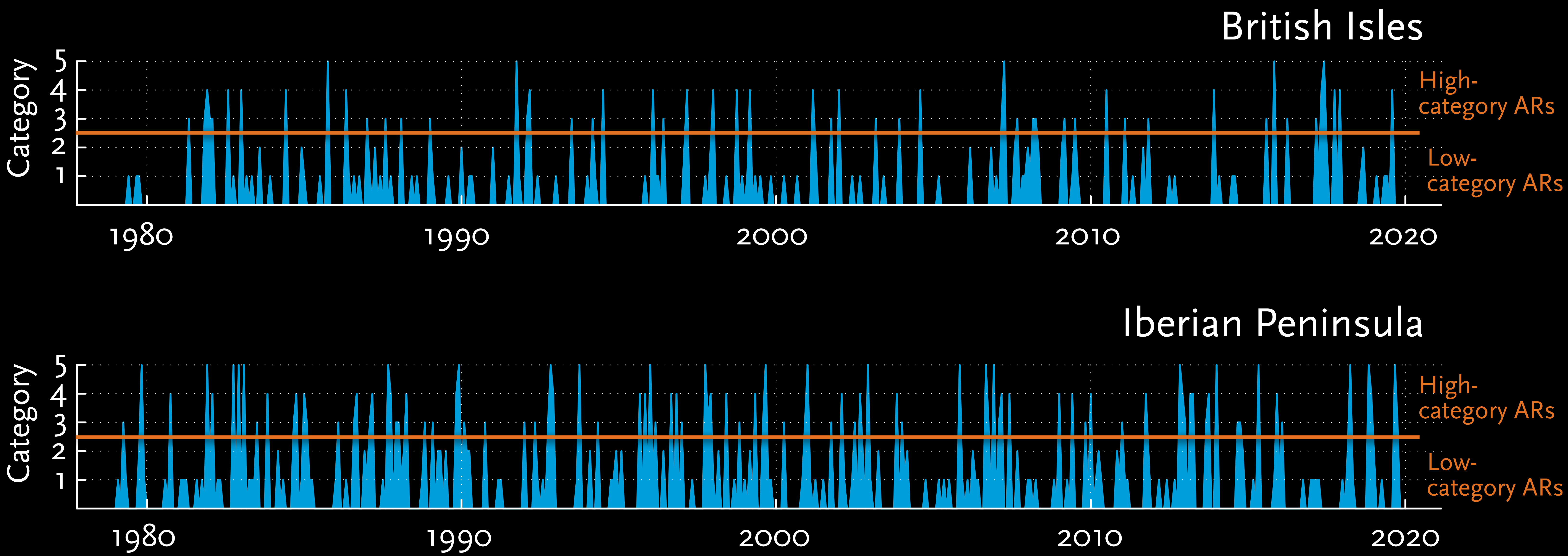
Atmospheric Rivers

- Landfalling atmospheric rivers
- Regional differences?



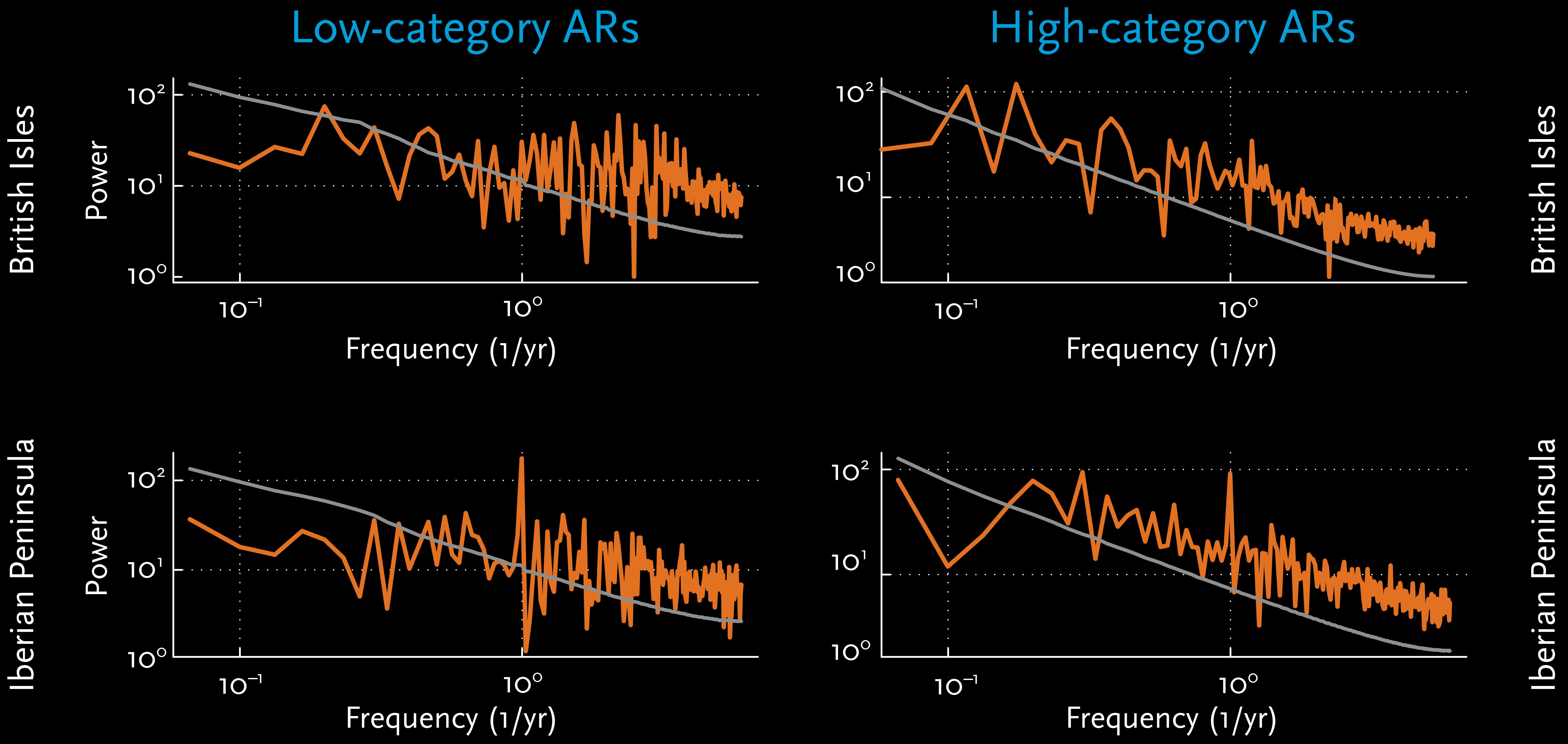
Marwan & Braun, Chaos 33, 2023

Atmospheric Rivers



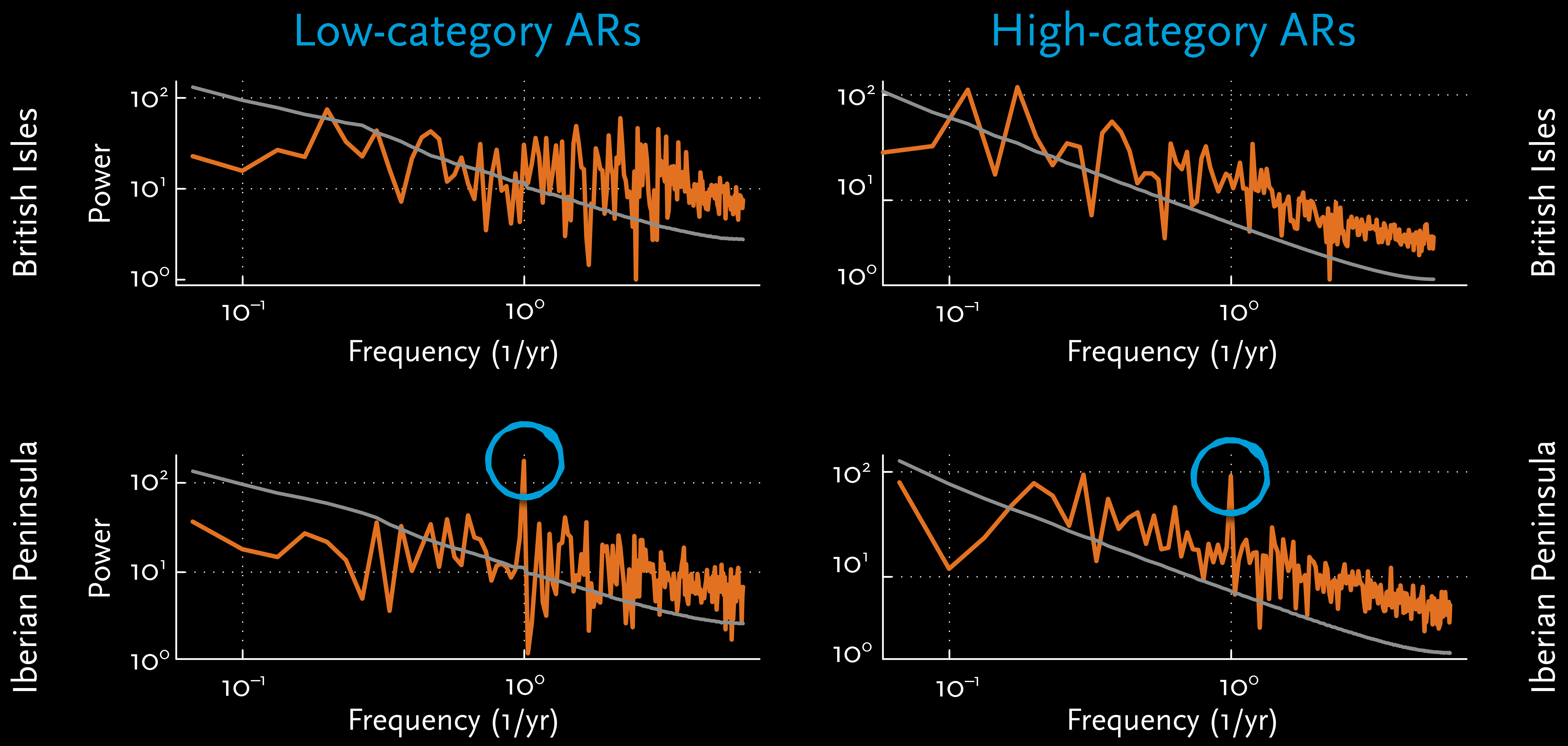
Marwan & Braun, Chaos 33, 2023

Powerspectra Atmospheric Rivers



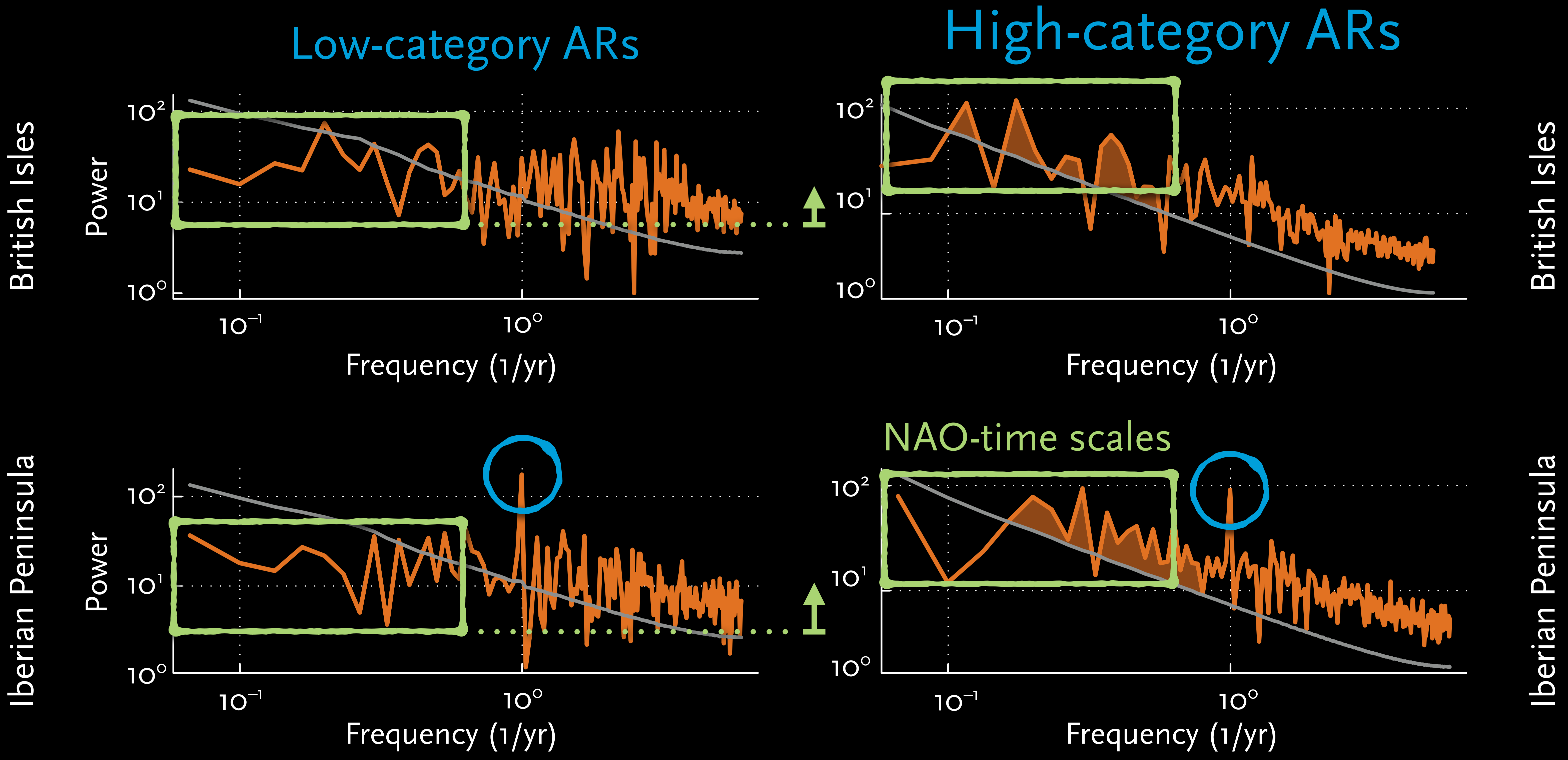
Marwan & Braun, Chaos 33, 2023

Powerspectra Atmospheric Rivers



Marwan & Braun, Chaos 33, 2023

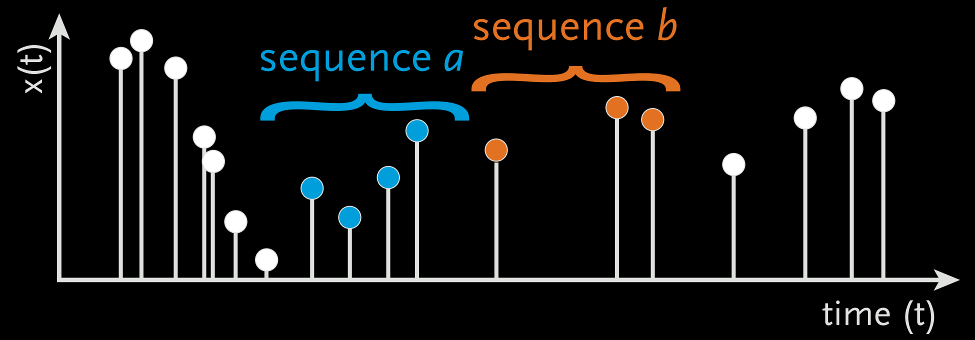
Powerspectra Atmospheric Rivers

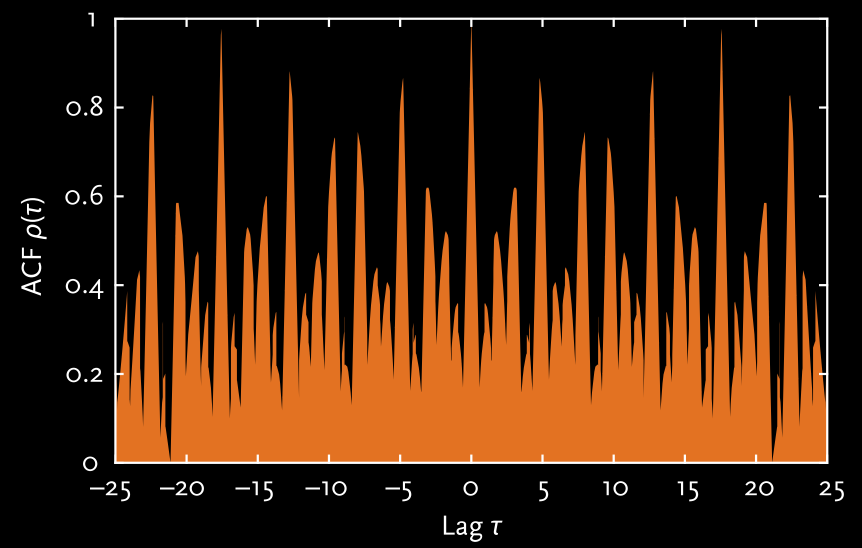


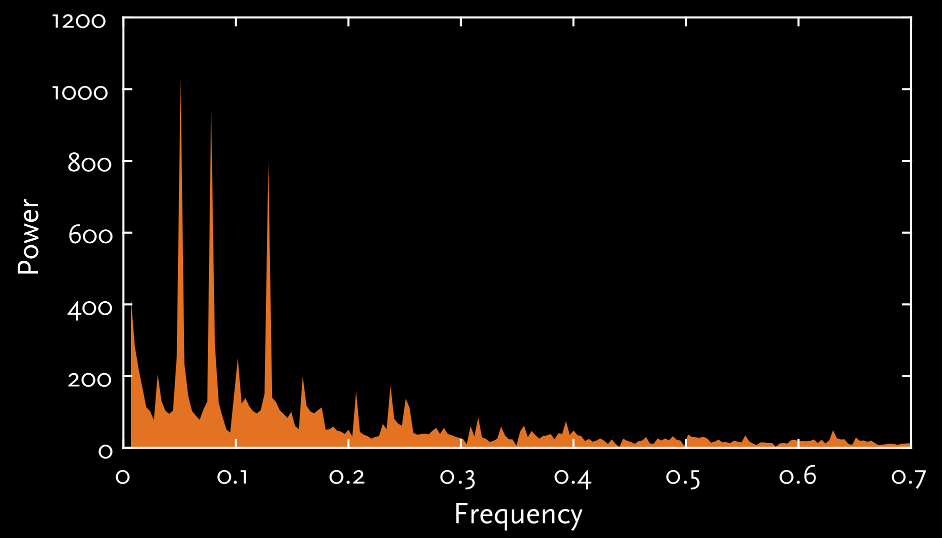
Marwan & Braun, Chaos 33, 2023

Open Questions

- Non-stationary event distribution, clustered events
- Anti-correlation in event series
- Effect of normalisation in ED-ACF
- Including amplitude variability
- Alternative metrics (ARI-SPIKE, Needleman-Wunsch distance, LCSS)
- Harmonics



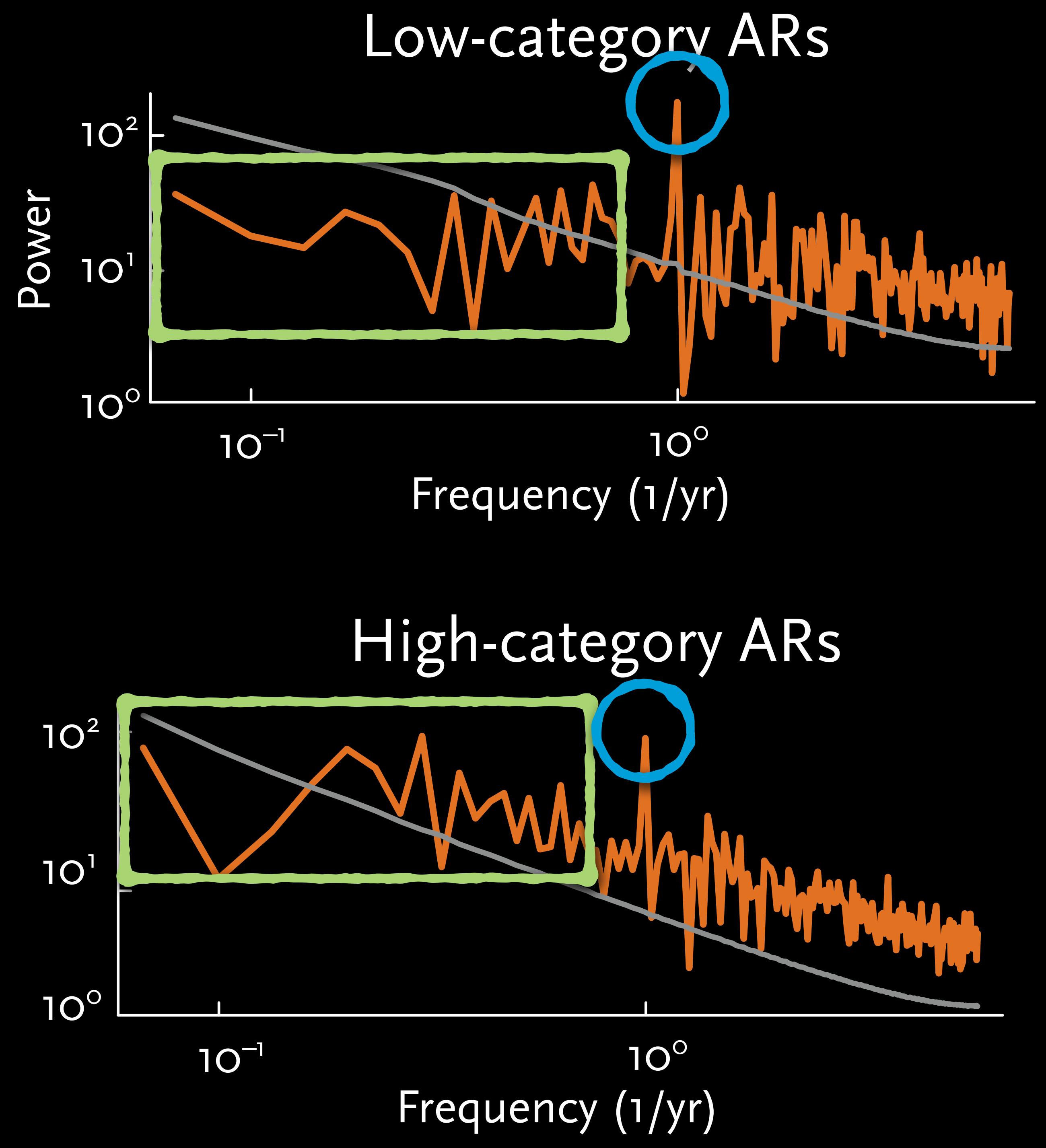




Marwan, Front. Appl. Math. Stat. 9, 2023

Take Home Message

- Simple power spectrum estimation for event data
- Atmospheric rivers in Europe:
  - Clear seasonal cycle, except British isles
  - High spectral power multiannual/ decadal time-scales for high-cat. ARs
- Julia script: doi:10.5281/zenodo.7555049



Marwan & Braun, Chaos 33, 2023

Chaos ARTICLE scitation.org/journal/cha

Power spectral estimate for discrete data

Cite as: Chaos **33**, 053118 (2023); doi: 10.1063/5.0143224
Submitted: 20 January 2023 · Accepted: 3 April 2023 ·
Published Online: 25 May 2023

View Online Export Citation CrossMark

Norbert Marwan[1,2,a] and Tobias Braun[1]

AFFILIATIONS

[1] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Telegrafenberg A31, 14473 Potsdam, Germany
[2] University of Potsdam, Institute of Geoscience, Karl-Liebknecht-Straße 32, 14476 Potsdam, Germany

Note: This paper is part of the Focus Issue on Ordinal Methods: Concepts, Applications, New Developments and Challenges.
[a] **Author to whom correspondence should be addressed:** marwan@pik-potsdam.de

ABSTRACT

The identification of cycles in periodic signals is a ubiquitous problem in time series analysis. Many real-world datasets only record a signal as a series of discrete events or symbols. In some cases, only a sequence of (non-equidistant) times can be assessed. Many of these signals are furthermore corrupted by noise and offer a limited number of samples, e.g., cardiac signals, astronomical light curves, stock market data, or extreme weather events. We propose a novel method that provides a power spectral estimate for discrete data. The edit distance is a distance measure that allows us to quantify similarities between non-equidistant event sequences of unequal lengths. However, its potential to quantify the frequency content of discrete signals has so far remained unexplored. We define a measure of serial dependence based on the edit distance, which can be transformed into a power spectral estimate (EDSPEC), analogous to the Wiener–Khinchin theorem for continuous signals. The proposed method is applied to a variety of discrete paradigmatic signals representing random, correlated, chaotic, and periodic occurrences of events. It is effective at detecting periodic cycles even in the presence of noise and for short event series. Finally, we apply the EDSPEC method to a novel catalog of European atmospheric rivers (ARs). ARs are narrow filaments of extensive water vapor transport in the lower troposphere and can cause hazardous extreme precipitation events. Using the EDSPEC method, we conduct the first spectral analysis of European ARs, uncovering seasonal and multi-annual cycles along different spatial domains. The proposed method opens new research avenues in studying of periodic discrete signals in complex real-world systems.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0143224

Many dynamical processes exhibit characteristic periodic time scales, which can be assessed by power spectral density estimation. This is the usual and standard procedure of time series analysis, where the data are, in general, equally sampled and follows a more or less continuous nature. However, in specific applications, only events are observable, which could be irregularly occurring activ-

be denoted by a set of ordered pairs $\{(t_i, x_i)\}$ of time t_i with $t_{i+1} > t_i$ and corresponding data value x_i; and with sampling index i. PSD estimation becomes even more challenging, if instead only the time points of the events are available (i.e., in contrast to the time series definition above, we have only a set of time points $\{t_i\}$ which indicate the presence of an event; time series analysis tools usually handle

14 August 2023 10:06:51