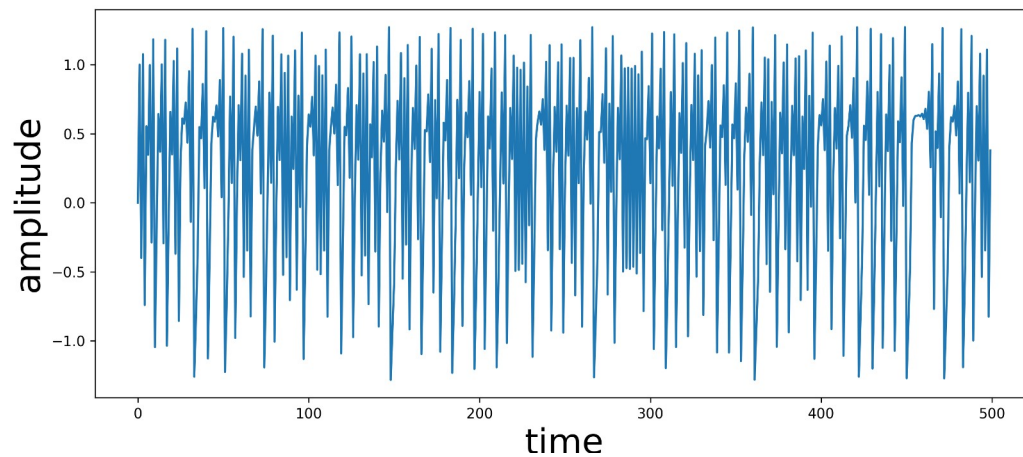
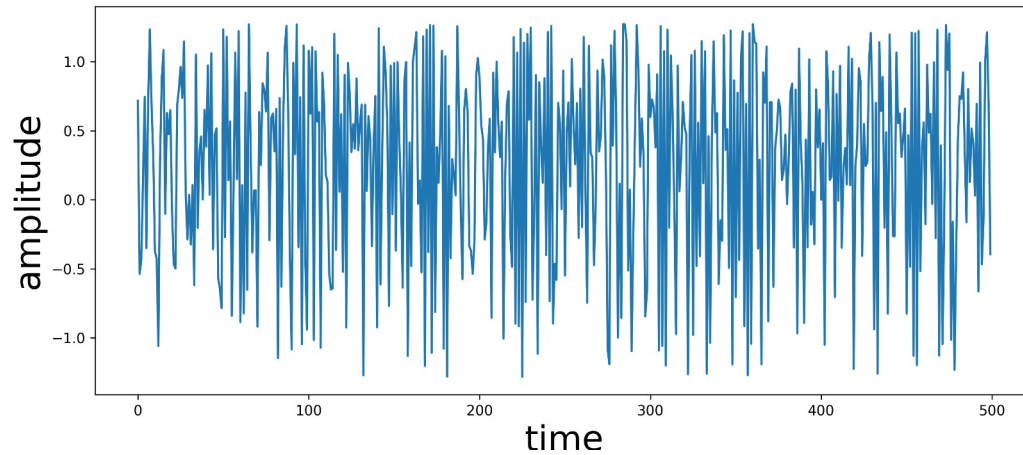


# Monte Carlo Simulation and Surrogates

Computational Physics Seminar on  
Analyzing Biomedical Signals

Konrad Beck

# Which Signal is Linear ?



# Motivation

- Nonlinearity can't be seen with naked eye
- Nonlinear measures can report structure in purely linear / stochastic data
- Before using nonlinear tools, check whether a simpler explanation is sufficient
- Setup Null Hypothesis
- Test Null by generating Monte Carlo surrogates

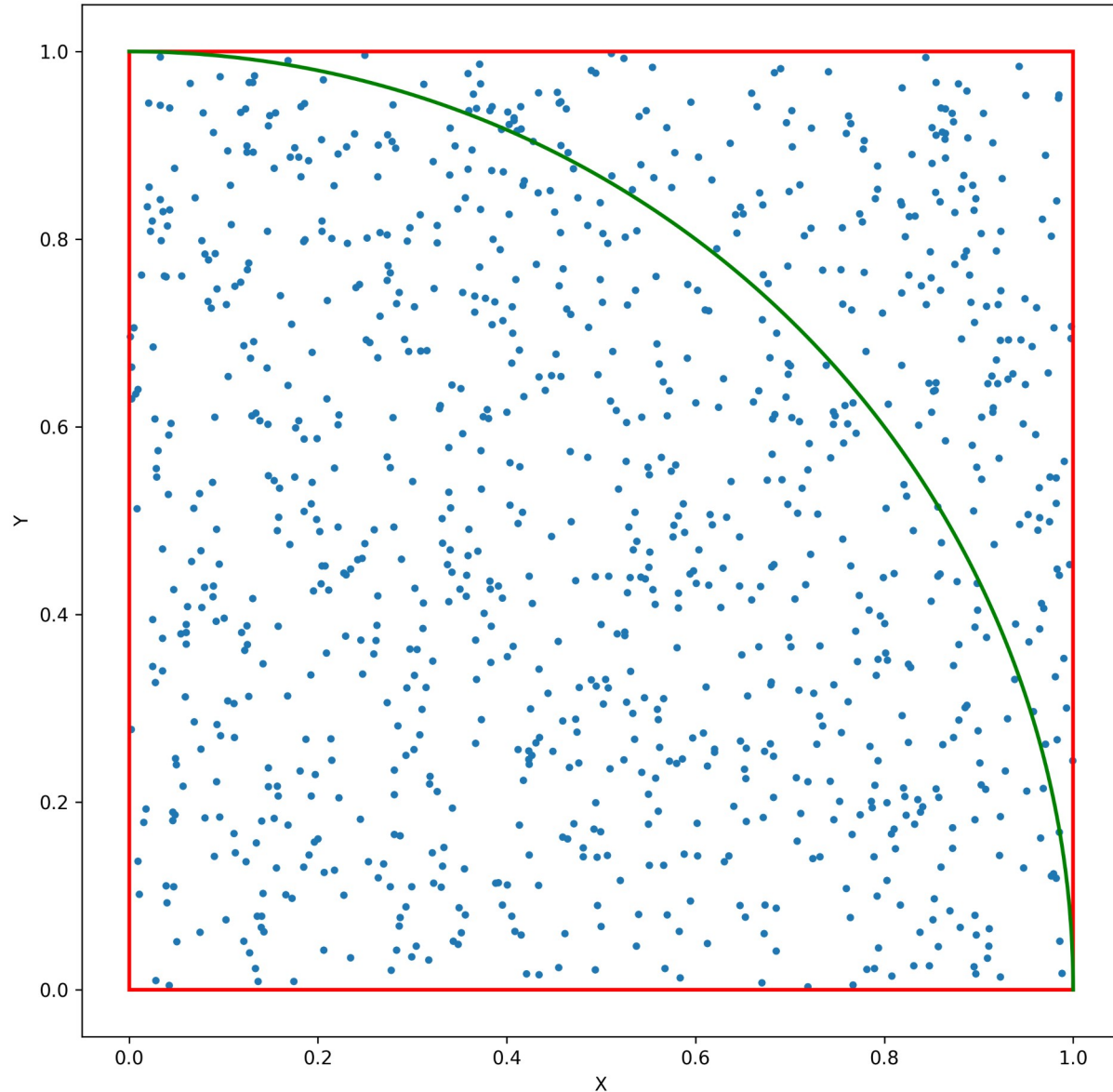
# Outline

- Motivation
- Monte Carlo Simulation / The Bootstrap Method
- Surrogate testing workflow
- Null Hypothesis
- Surrogate generation methods
- Test Statistics
- Examples
- Interpretation and Limits

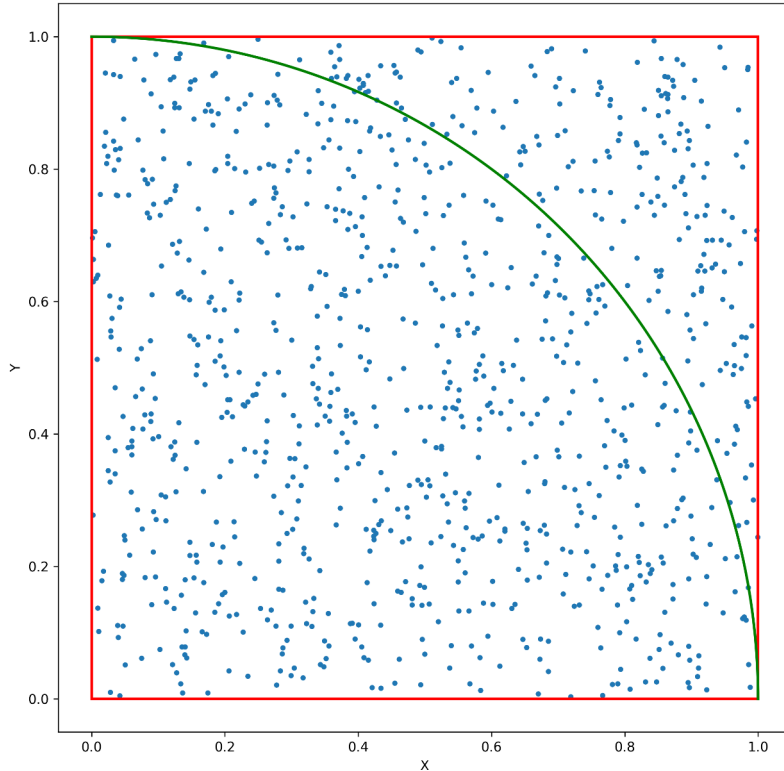
# Monte-Carlo-Simulation

- Use random sampling to approximate difficult or analytically unsolvable problems
- Accuracy improves with more samples
- Mostly used in optimization, numerical integration and sampling methods

# Monte-Carlo-Simulations



# Monte-Carlo-Simulations



$$\frac{\pi}{4} = \frac{A_{circle}}{A_{total}} \approx \frac{N_{circle}}{N_{total}}$$

for  $N_{total} = 10000$ :

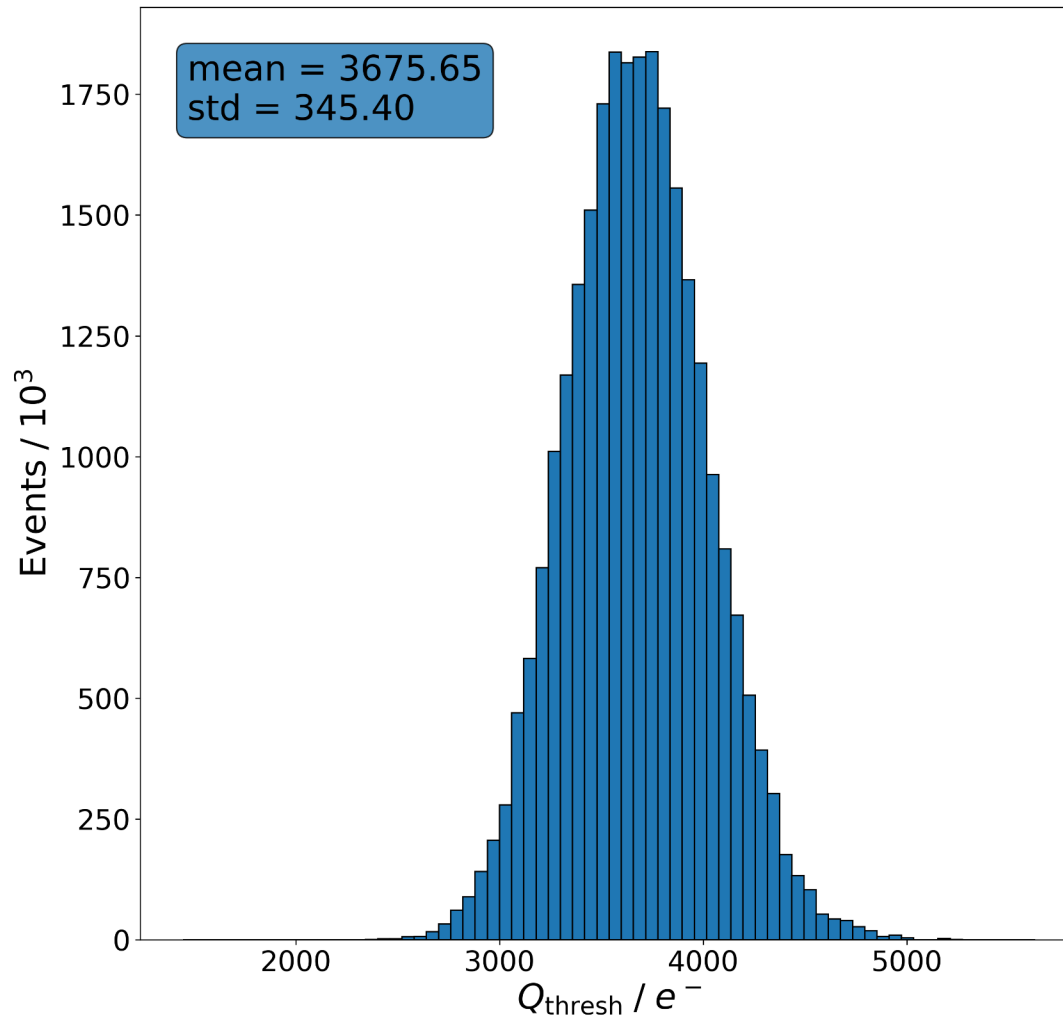
$$\pi \approx 3.1592$$

# The Bootstrap Method

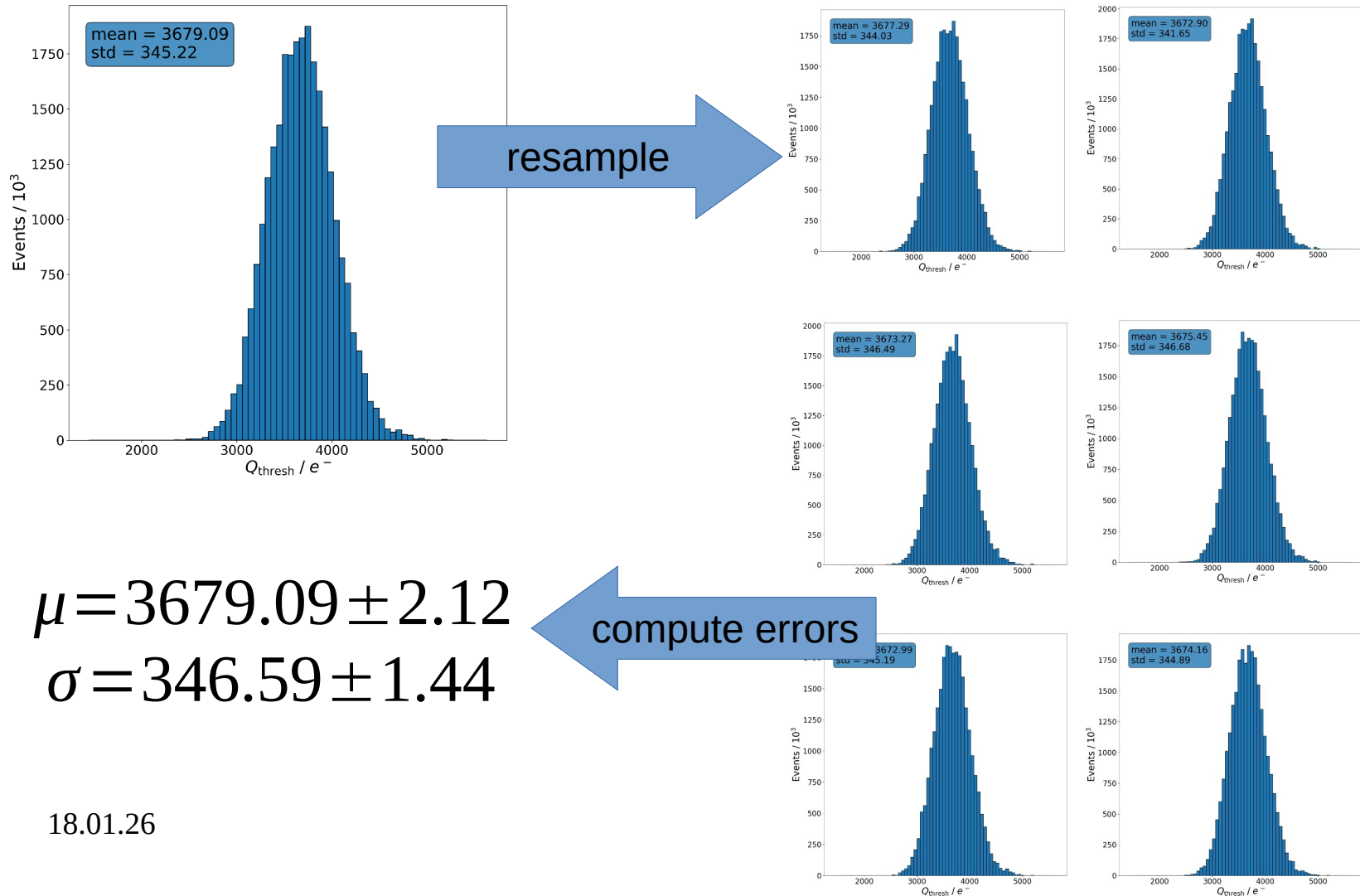
- Estimate statistical measures when true distribution is unknown
- Resample from observed data (with replacement) and compute statistics for new data set
- The resampled statistics form an empirical distribution
- Parametric Bootstrap: first fit model to data then resample from model
- Key Assumption: Measured distribution is representative of true distribution



# Bootstrap for Error Estimation



# Bootstrap for Error Estimation



# Null Hypothesis

- The Null Hypothesis is the simplest plausible explanation for the observed signal
- Specifies the class of processes the data could come from
- Used to generate reference data (surrogates) consistent with  $H_0$

# Surrogate Test Workflow

- Choose a null hypothesis  $H_0$
- Generate many surrogate time series consistent with  $H_0$
- Choose a discriminating statistic  $T$
- Compute  $T$  for the original data and for each surrogate
- Compare data vs surrogate distribution
- reject or do not reject  $H_0$

# Independently Distributed

- $H_0$ : the data are independent draws from a fixed distribution
- Create surrogates by randomly permuting the measured values
- Preserves the amplitude distribution exactly, removes all temporal correlations
- If the data differ significantly from the shuffles, reject independence

# Linear Gaussian Systems

- Null hypothesis: Gaussian linear stochastic process

$$s_n = \sum_{i=1}^M a_i s_{n-i} + \sum_{i=0}^N b_i \eta_{n-i}$$

- Option 1: Assume independence of nuisance parameters
- Option 2: Fit parameters of nuisance parameters → Bootstrap
-

# Fourier phase-randomized surrogates

- All structure lies in first and second order statistics
- Option 3: Generate surrogates by keeping Fourier amplitudes (DFT) and randomizing Fourier phases, then transforming back

$$|S_k|^2 = \left| \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_n \exp\left(\frac{i 2 \pi k n}{N}\right) \right|^2$$
$$\bar{s}_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{i \alpha_k} |S_k| \exp\left(\frac{-i 2 \pi k n}{N}\right)$$

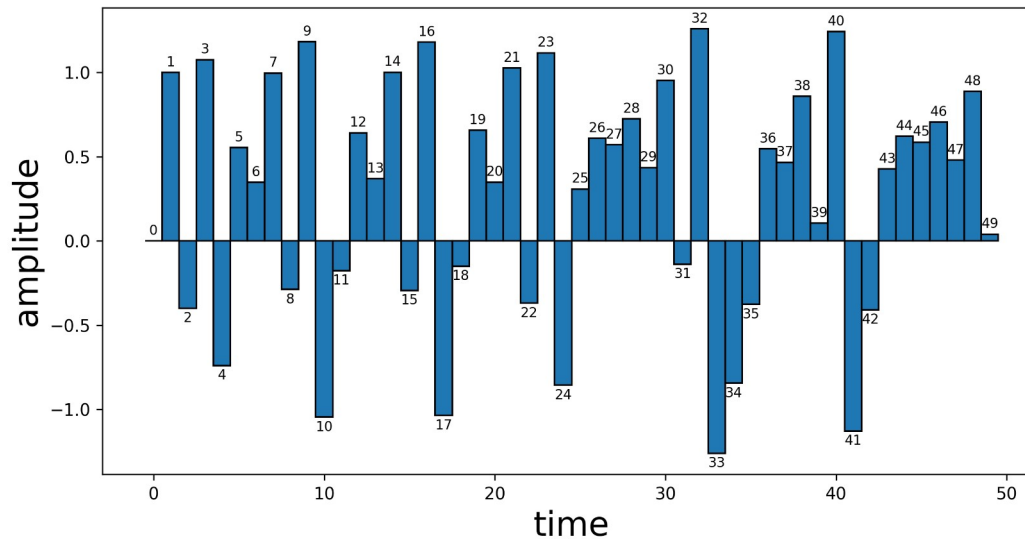
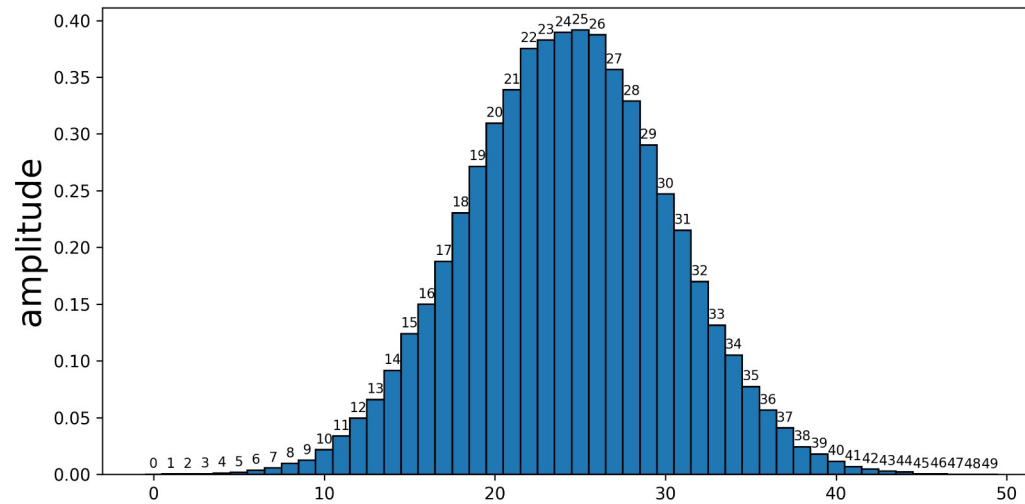
# Rescaled Linear Gaussian Processes

- Null: linear Gaussian process observed through an invertible static measurement function
- Idea: undo the unknown measurement distortion by rank-rescaling the data to a Gaussian distribution
- Nonlinearity lies in measurement function  $s$

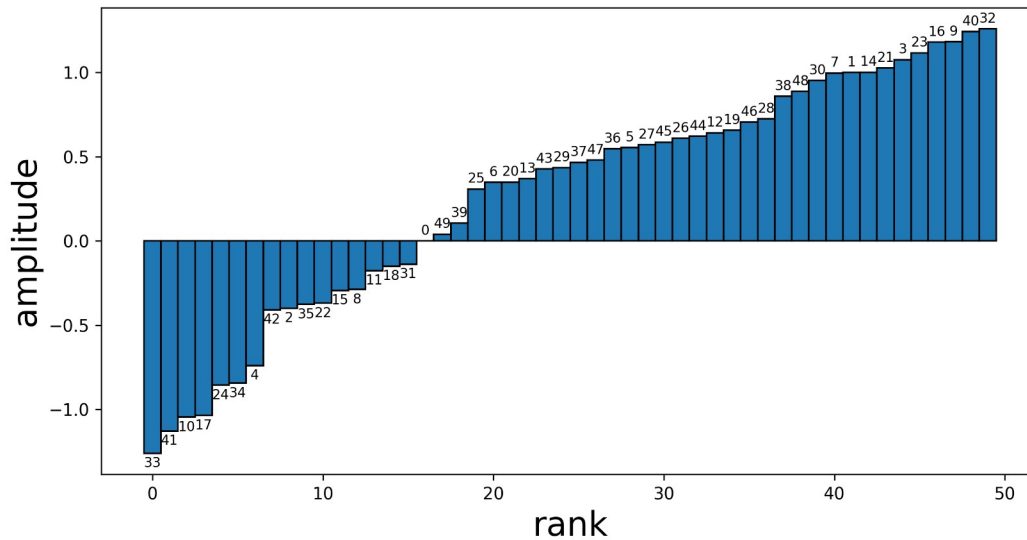
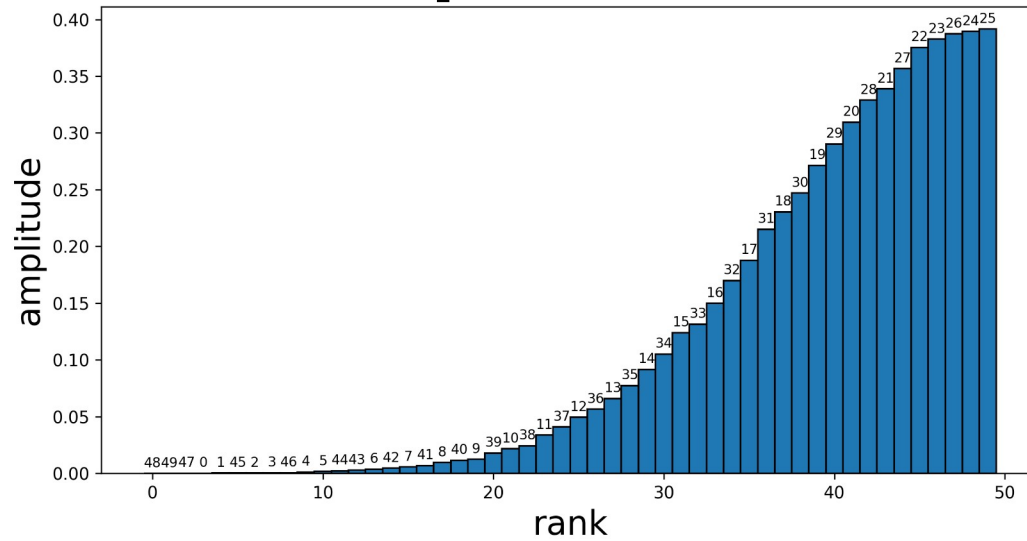
$$s_n = s(x_n) \quad x_n = \sum_{i=1}^M a_i x_{n-i} + \sum_{i=0}^N b_i \eta_{n-i}$$



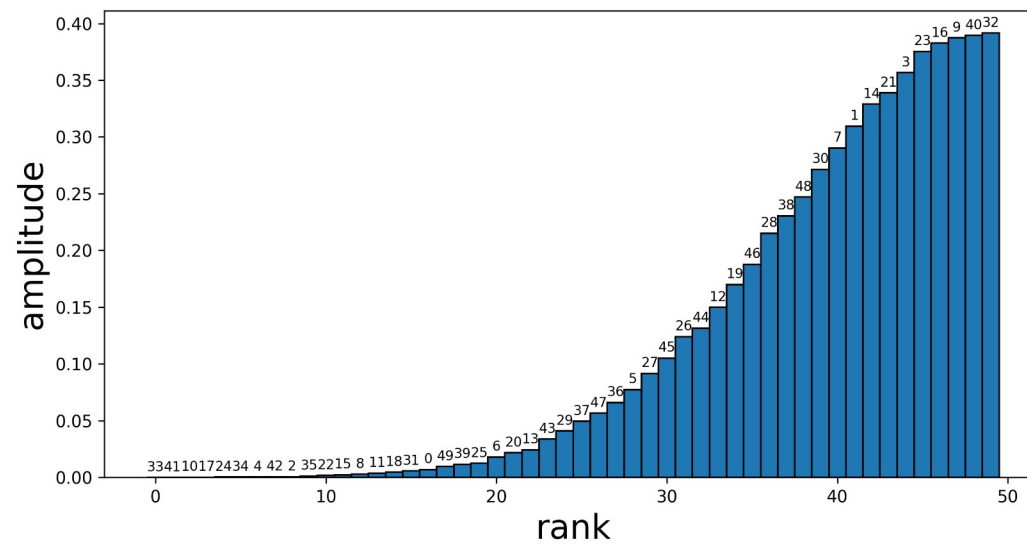
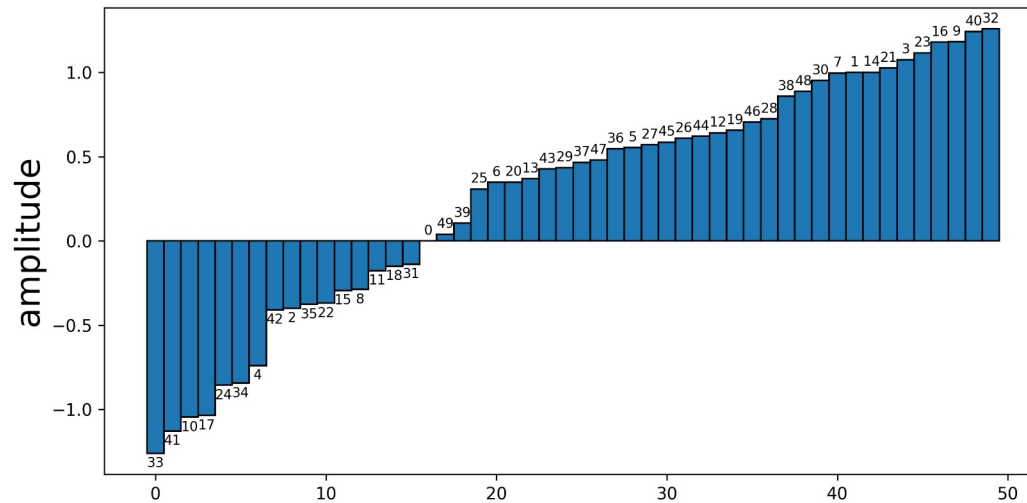
# Gaussianization



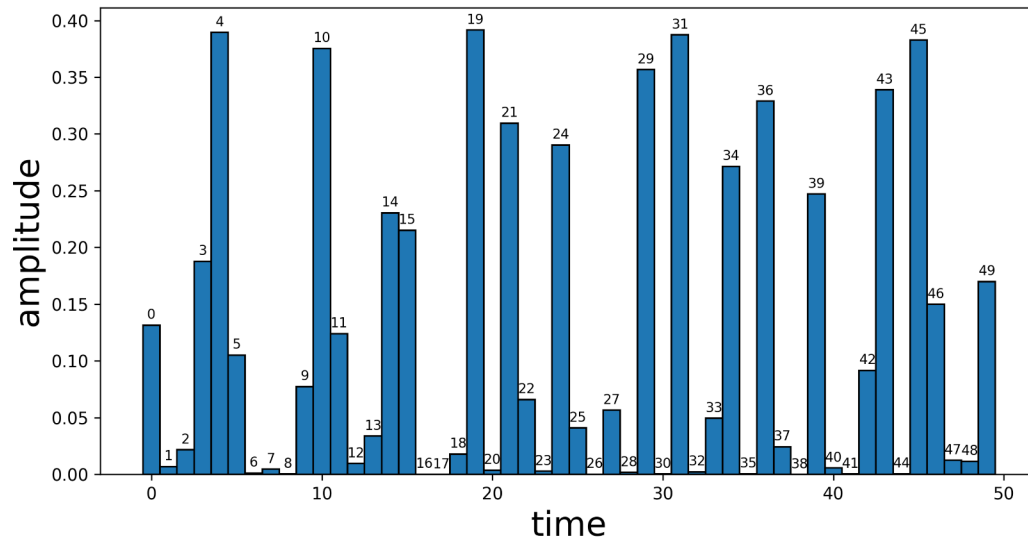
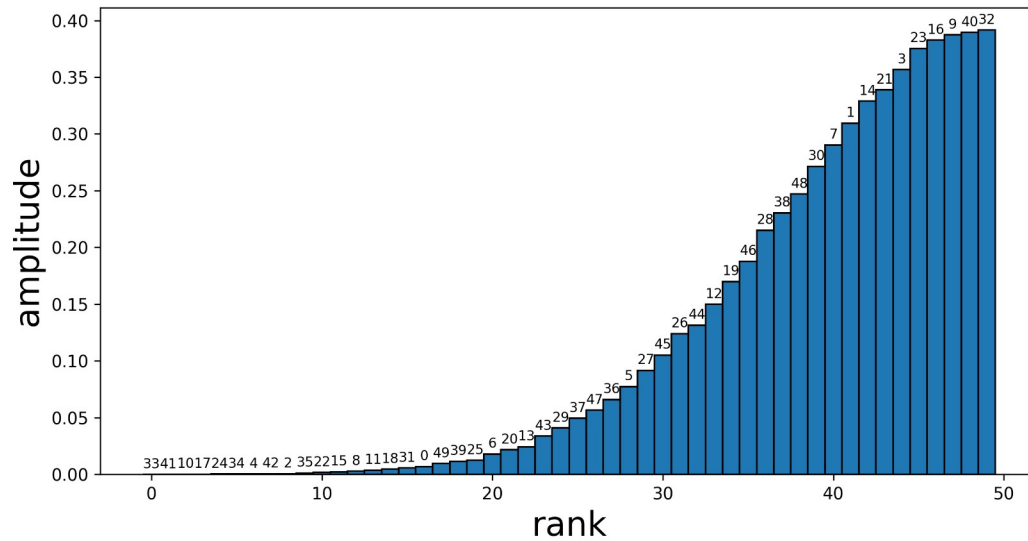
# Step 1: Sort



# Step 2: Rescale



# Step 3: Rearrange



# Amplitude Adjusted Fourier Transform Surrogates

- Gaussianize
- Randomize Fourier phases to keep only linear correlations while destroying nonlinear structure
- Rescale back by rank order to match the original amplitude distribution
- Works reliably mainly when the series is long, correlations are not too strong, and the measurement function is close to identity

# Iterative AAFT Surrogates

- AAFT can bias the spectrum
- IAAFT alternates two steps until convergence:
  - enforce the target Fourier amplitudes
  - enforce the target amplitude distribution by rank ordering
- Step 0: Generate a Surrogate using reshuffling or AAFT

$c_k \rightarrow$  sorted copy of data

$|S_K|^2 \rightarrow$  Fourier amplitudes of data

$\bar{r}_n^i \rightarrow$  sequence with correct amplitude distribution at iteration  $i$

$\bar{s}_n^i \rightarrow$  sequence with correct Fourier amplitudes at iteration  $i$

# IAAFT: Algorithm

Sep1: Compute DFT

$$\bar{R}_k^{(i)} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \bar{r}_n \exp\left(\frac{i 2 \pi k n}{N}\right)$$

Sep2: Compute the phases

$$\exp(i \psi_k^{(i)}) = \frac{\bar{R}_k^{(i)}}{|\bar{R}_k^{(i)}|}$$

Sep3: Force the correct Fourier Amplitudes and apply inverse DFT

$$\bar{s}_n^{(i)} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \exp(i \psi_k^{(i)}) |S_k| \exp\left(\frac{-i 2 \pi k n}{N}\right)$$

Sep4: Force the correct Amplitudes by rescaling)

$$\bar{r}_n^{(i+1)} = c_{rank}(\bar{s}_n^{(i)}).$$

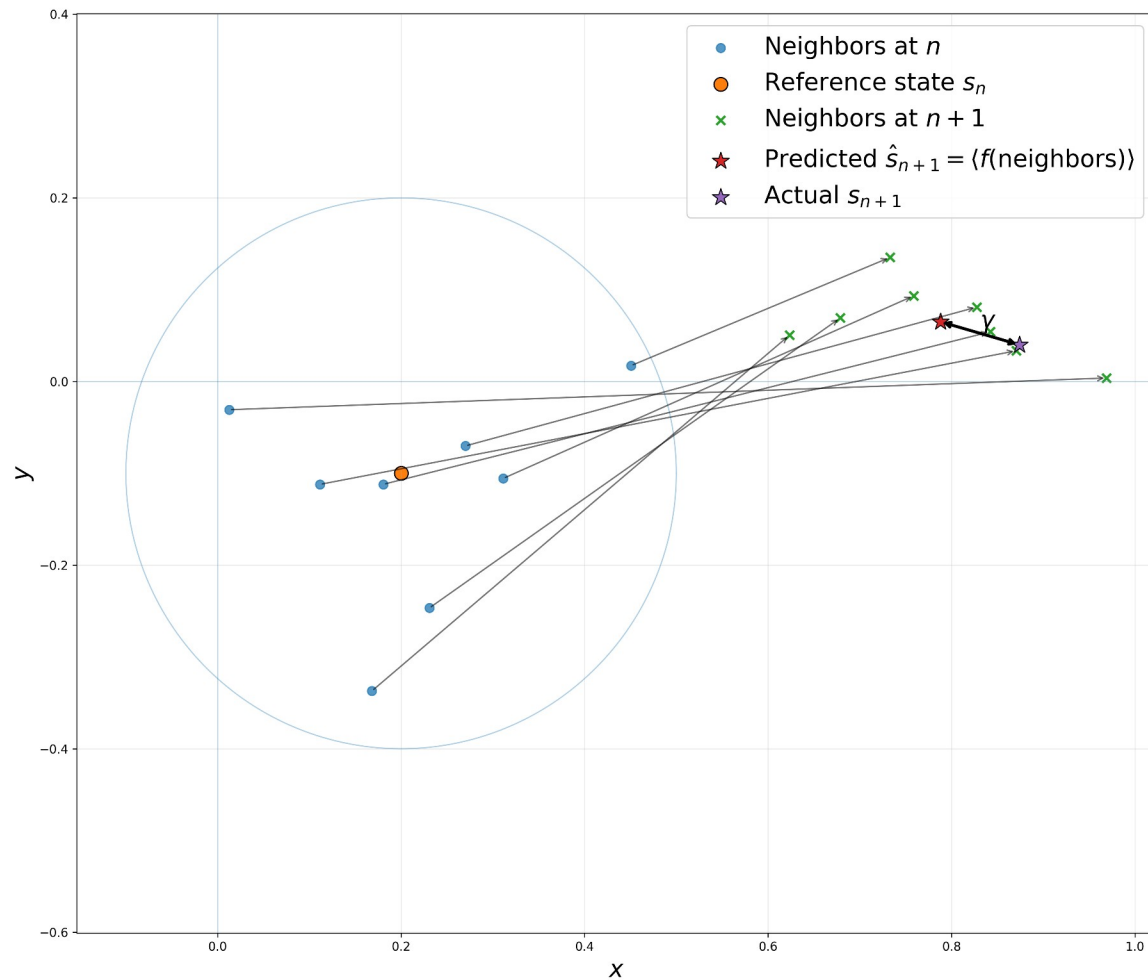
# Time Irreversibility

- Idea: compare the series to its time-reversed version
- Use a statistic based on increments to detect asymmetry between forward and backward time
- Normalize the statistic so results are comparable across signals
- Interpretation: significant deviation from 0 indicates time irreversibility

$$\phi^{rev}(\tau) = \frac{1}{N - \tau} \sum_{n=\tau+1}^N (s_n - s_{n-\tau})^3,$$



# Non Linear Prediction Error



# Non Linear Prediction Error

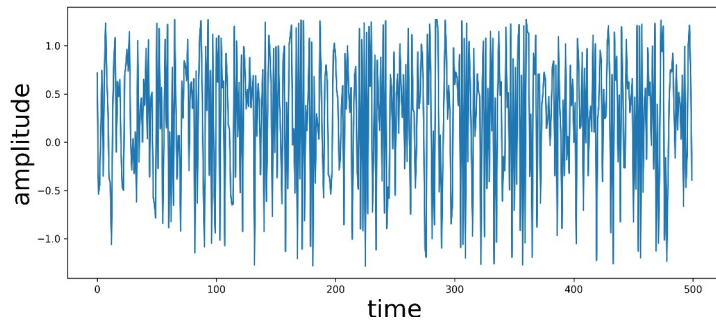
$$\gamma(m, \tau, \epsilon) = \left( \frac{1}{N} \sum [s_{n+1} - F(s_n)]^2 \right)^{\frac{1}{2}}$$

# The rank-order test

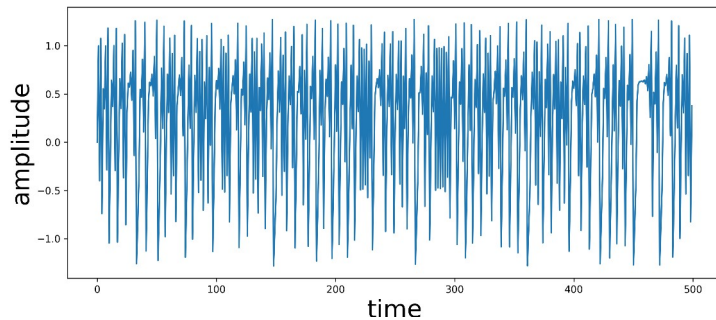
- Choose false-rejection probability
- Choose an integer  $K$
- Compute the amount of Surrogates  $M$  needed
- Rank the statistics of the surrogates
- Reject if the data's statistic is among the  $K$  most extreme

$$\text{one sided: } M = \frac{K}{\alpha} - 1, \quad \text{two sided: } M = 2 \frac{K}{\alpha} - 1$$

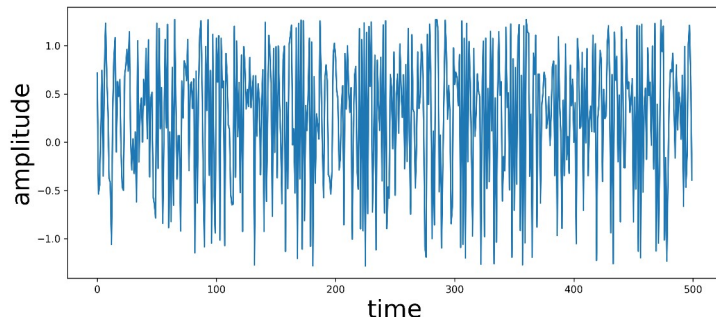
# Example 1



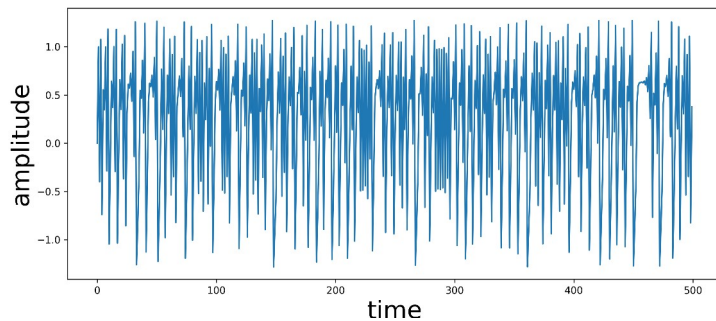
Null: linear Gaussian process  
observed through an invertible  
static measurement function



# Example 1

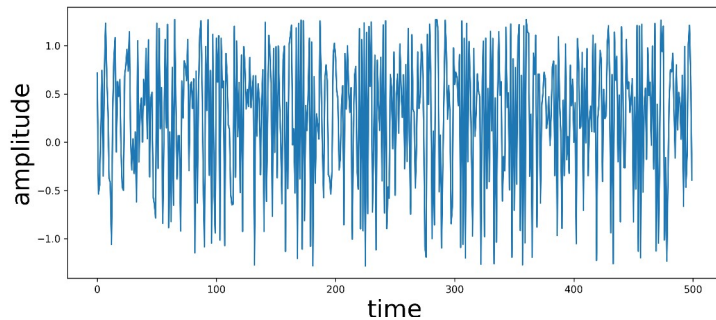


- Generated 99 surrogates for both datasets
- Choose parameters
- Computed Non Linear Prediction Error for all

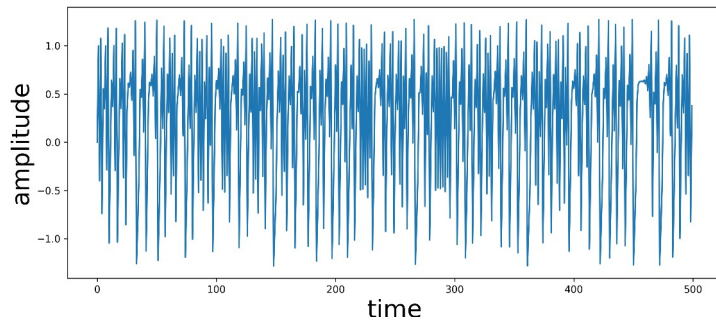


$$\begin{aligned}\alpha &= 0.01 \\ K &= 1 \\ M &= \frac{K}{\alpha} - 1 = 99\end{aligned}$$

# Example 1

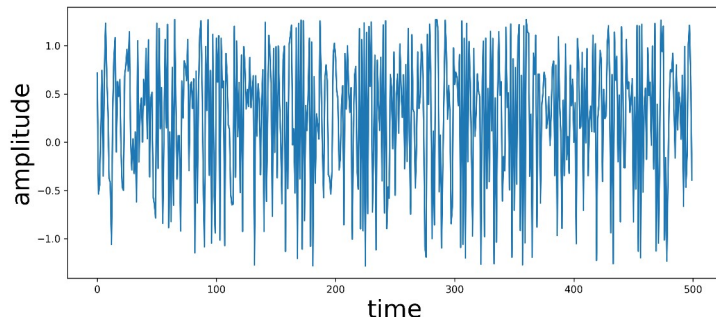


- Choose parameters
- Generated 99 surrogates for both datasets using IAAFT
- Computed Non Linear Prediction Error for all
- Rank Errors



$$\begin{aligned}\alpha &= 0.01 \\ K &= 1 \\ M &= \frac{K}{\alpha} - 1 = 99\end{aligned}$$

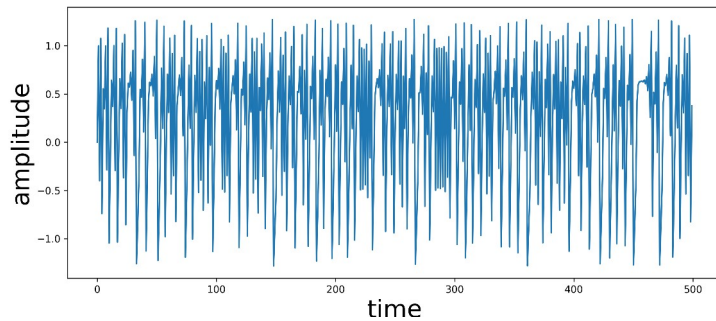
# Example 1



$$\gamma \approx 0.95$$



Not rejected

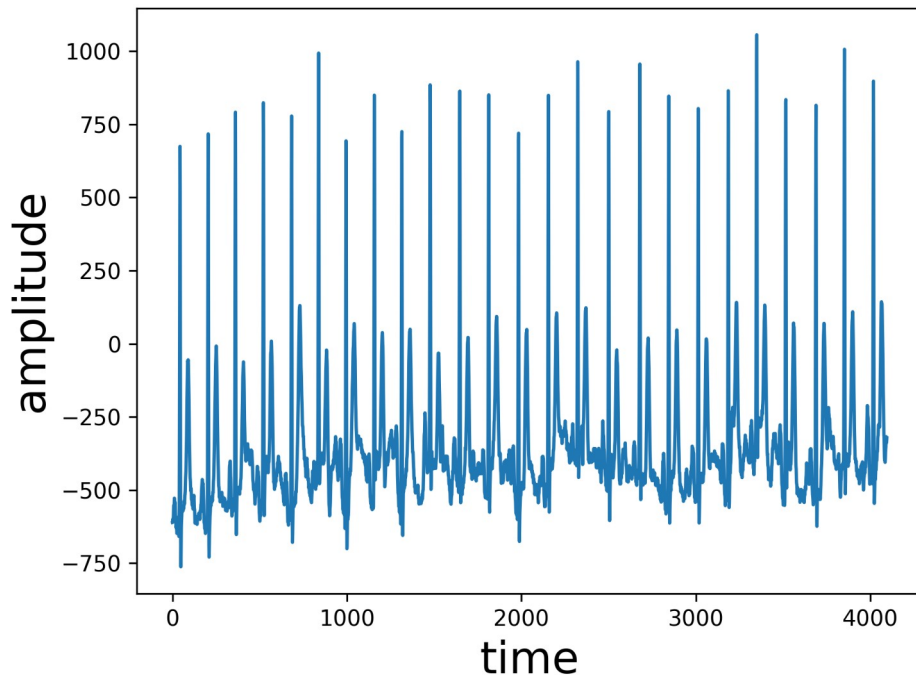


$$\gamma \approx 0.10$$



Rejected

# ECG Example



- Null: linear Gaussian process observed through an invertible static measurement function
- Choose parameters
- Generated M surrogates for both datasets using IAAFT
- Compute Time reversibility
- Rank Time irreversibility

original data:  $\phi^{rev}(\tau) \approx -1.017$

surrogates:  $\phi^{rev}(\tau) \approx 0.0027$

$$\alpha = 0.01$$

$$K = 1$$

$$M = 2 \frac{K}{\alpha} - 1 = 199$$



# What does it mean to reject?

- Reject  $H_0$  means the observed statistic is very unlikely if  $H_0$  were true (at level  $\alpha$ )
- Conclusion: data contains structure not explained by this specific  $H_0$
- Do not reject  $H_0$ : no evidence against  $H_0$  with this statistic and this amount of data
- Important: non-rejection does not prove  $H_0$  (could be low power or wrong statistic)

# Limits of Surrogate Testing

- Conclusions are only about the chosen null  $H_0$   
(wrong  $H_0$   $\rightarrow$  wrong conclusion)
- Artifacts can cause false rejections
- Surrogates may fail to match the intended constraints
- Finite data and noise reduce test power
- Multiple testing (many statistics/parameters) increases false positives unless corrected

