Statistics for UQ and Time Series

Dr Adam Sykulski, Imperial College London

Outline

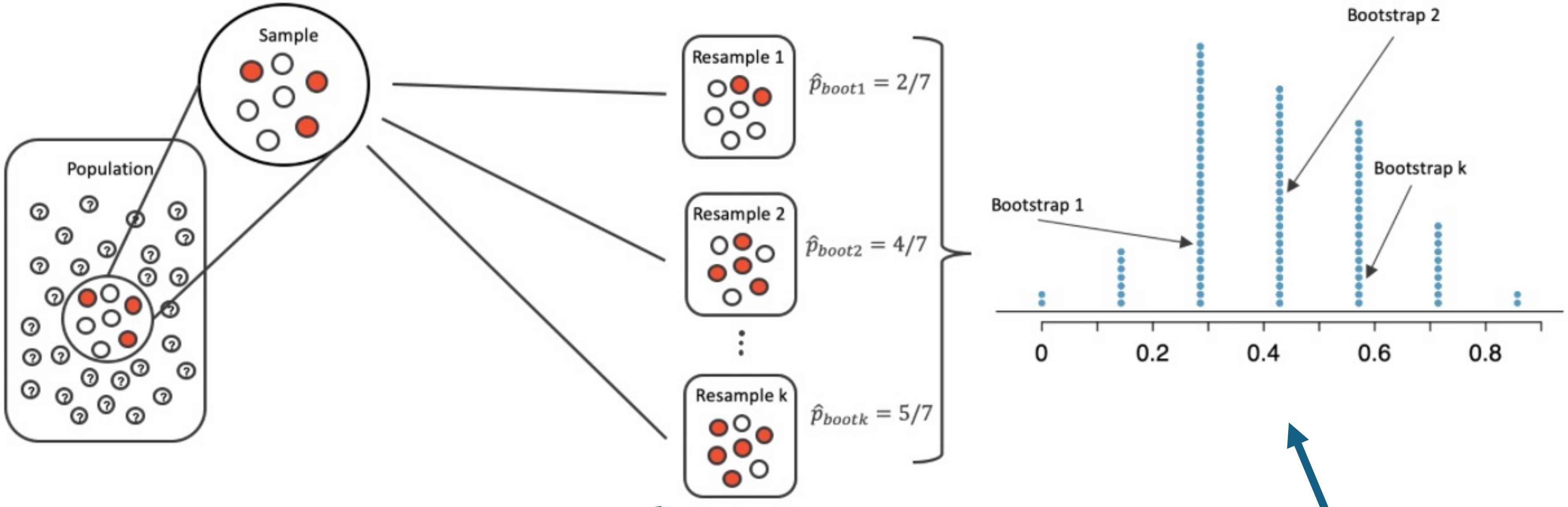
1. Estimating uncertainty and confidence intervals using the bootstrap (and/or the CLT)

2. Time Series and Spectral Analysis

3. Bonus section

Part I:
Estimating uncertainty and confidence intervals using the bootstrap (and/or the CLT)

What is the bootstrap?



Resample with replacement!

We calculate uncertainties from this distribution!

Don’t believe me? Let me show you in code!

https://github.com/AdamSykulski/OceanUQ

The bootstrap

The advantages

- Very simple method to understand and implement for calculating uncertainty
- Does not make distributional assumptions on the data or the distribution of the point estimate
- No need to collect more data!

The disadvantages

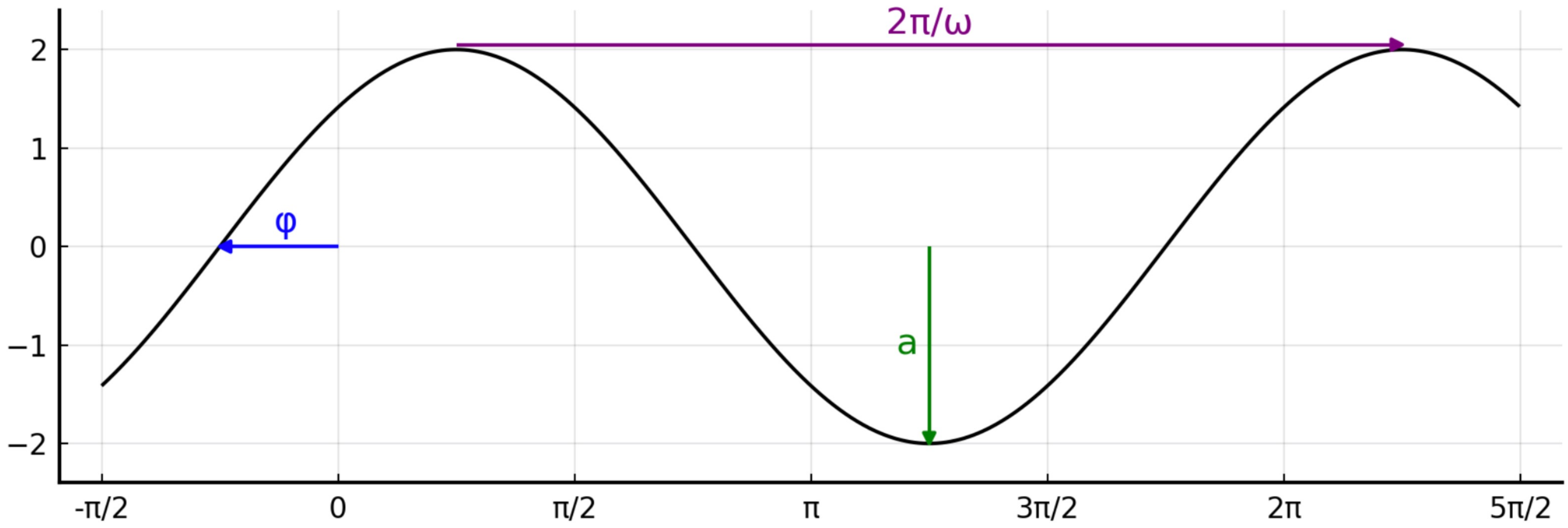
- Sample may not be representative
- As with all statistical methods, sensitive to small sample sizes
- Data might not be independent (e.g., a time series)
- All of the above can give misleading uncertainty measures!

Part II: Time Series and Spectral Analysis

Consider a basic sinusoid:

$$\eta(t) = a\sin(\omega t + \varphi)$$

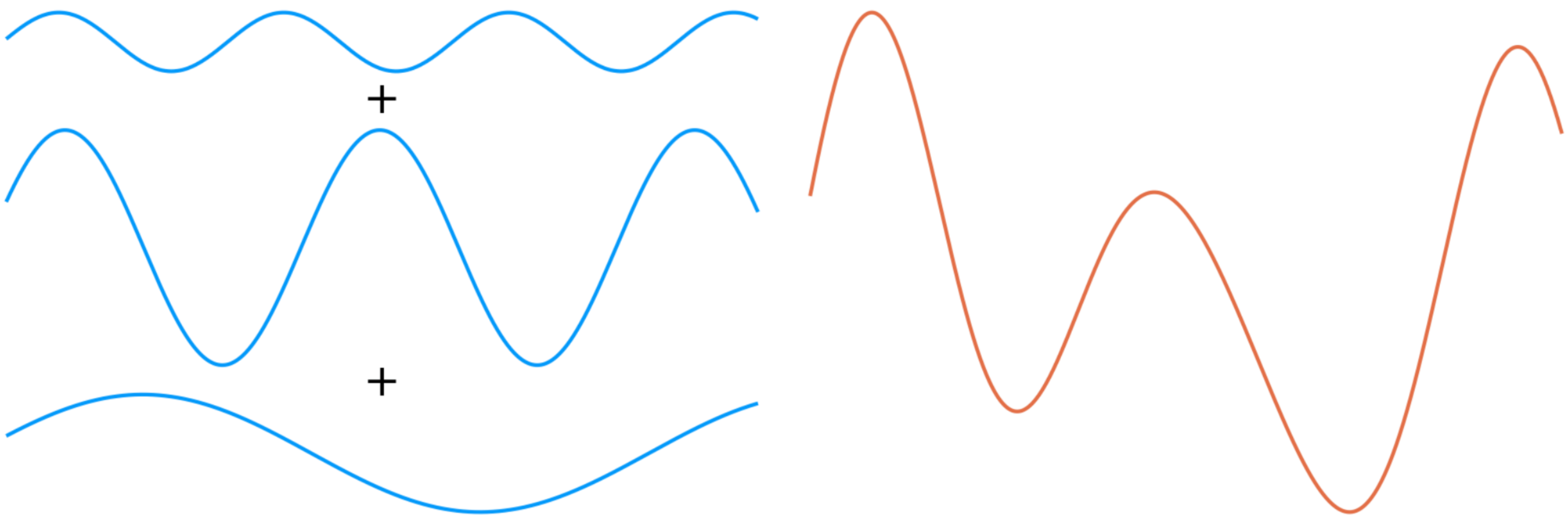
- a is the *amplitude*.
- φ is the *phase*.
- ω is the *angular frequency*.



We can represent more complicated functions as a sum of sinusoids:

$$\eta(t) = \sum_{\omega} a(\omega) \sin(\omega t + \varphi(\omega))$$

with different amplitudes and phases for each component

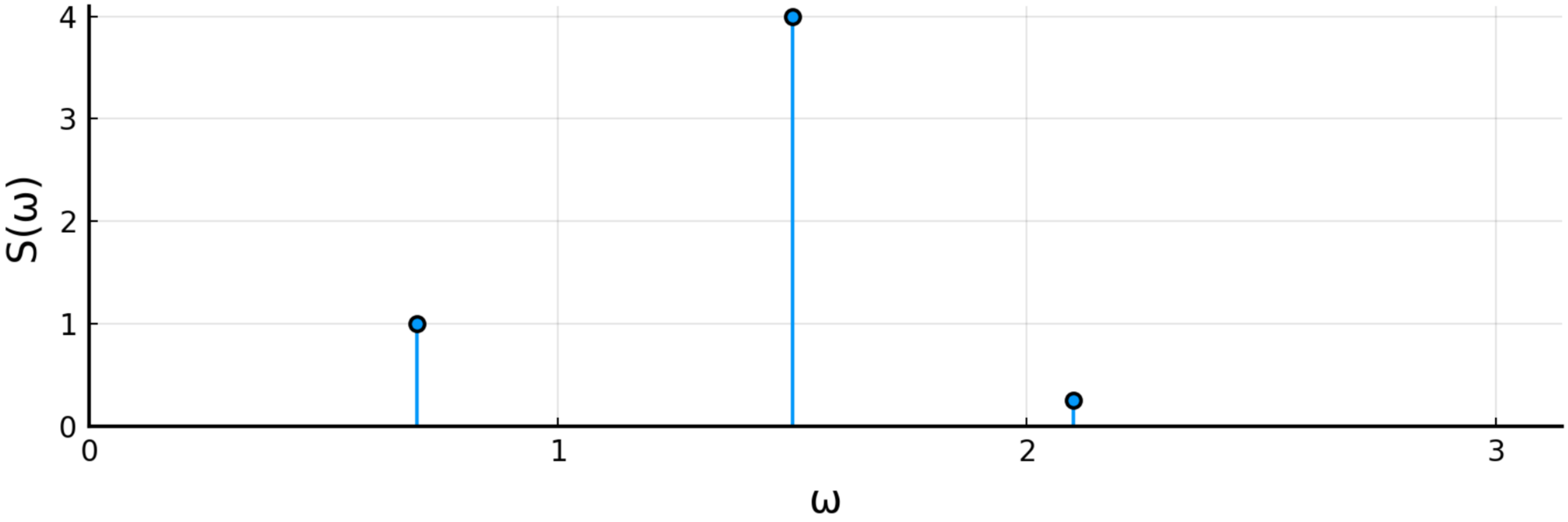


The power spectrum: discrete frequency

The power spectrum is then defined as

$$S(\omega) = |a(\omega)|^2$$

In our example:



We could keep going … but what if our time series process lives at all frequencies and is stochastic?

This is the field of spectral analysis!

To proceed we require some notation…

- $x(t)$: continuous real-valued stationary process, $t \in \mathbb{R}$
- x_t : discrete real-valued stationary process, $t \in \mathbb{Z}$
- ω : angular frequency, $\omega = 2\pi f$ (f is measured in hertz)
- τ : time-lag (positive or negative)
- $i = \sqrt{-1}$

To keep things tidy we will assume $x(t)$ (or x_t) is zero mean

The power spectral density

Fourier Transform: $f_x(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt, \quad \omega \in \mathbb{R}$

Inverse Fourier Transform: $x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_x(\omega)e^{i\omega t}d\omega, \quad t \in \mathbb{R}$

Power Spectral Density: $S_x(\omega) = \lim_{T\to\infty} \mathbb{E}\left(\frac{1}{2T}\left|\int_{-T}^{T} x(t)e^{-i\omega t}dt\right|^2\right)$

Relationship with autocovariance sequence $s_x(\tau) = \mathbb{E}[x(t)x(t-\tau)]$:

$$S_x(\omega) = \int_{-\infty}^{\infty} s_x(\tau)e^{-i\omega\tau}d\tau \iff s_x(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)e^{i\omega\tau}d\omega$$

Percival and Walden, Spectral Analysis for Univariate Time Series, page 65

Estimating from time series data

Theory: $S_x(\omega) = \lim_{T\to\infty} \mathbb{E}\left(\frac{1}{2T}\left|\int_{-T}^{T} x(t)e^{-i\omega t}dt\right|^2\right)$

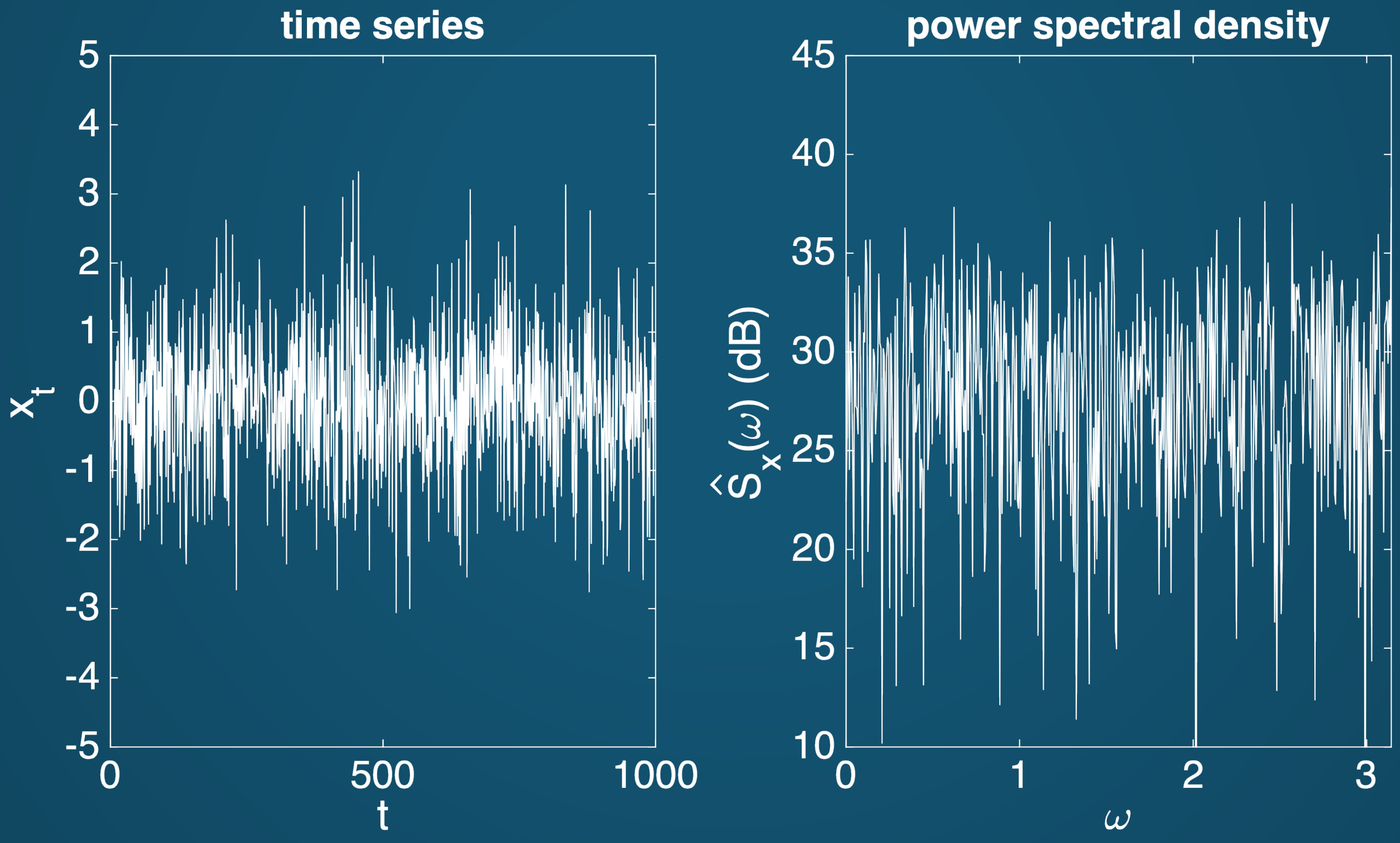
Practice: Observe some sample $X_1, \ldots, X_N$ at intervals Δ such that

$$\hat{S}(\omega) = \frac{\Delta}{N}\left|\sum_{t=1}^{N} X_t e^{-i\omega t}\right|^2$$

This is called the *periodogram* and is defined for $\omega \in [-\pi/\Delta, \pi/\Delta]$ where π/Δ is the *Nyquist frequency*.

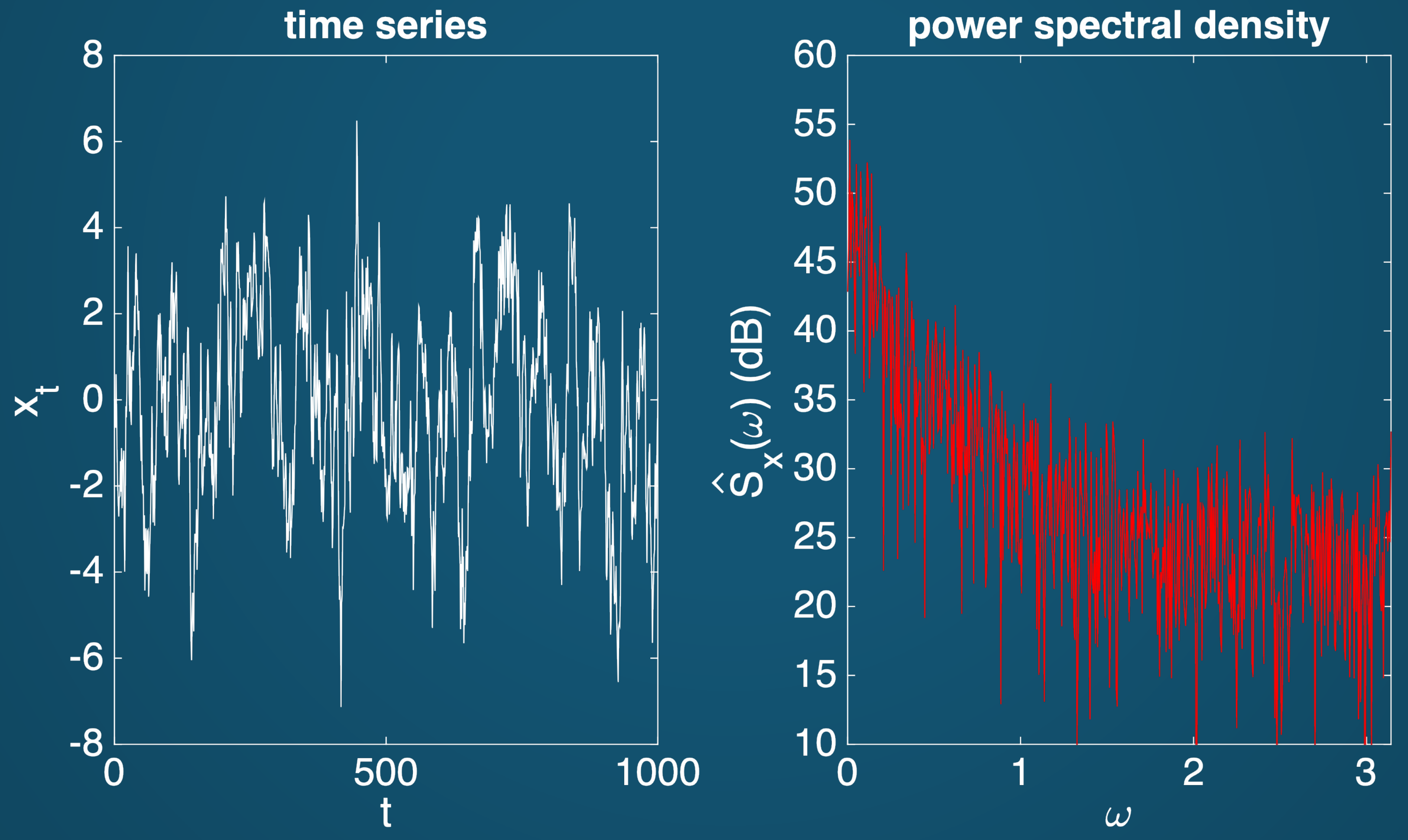
White noise process

$$x_t = \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$



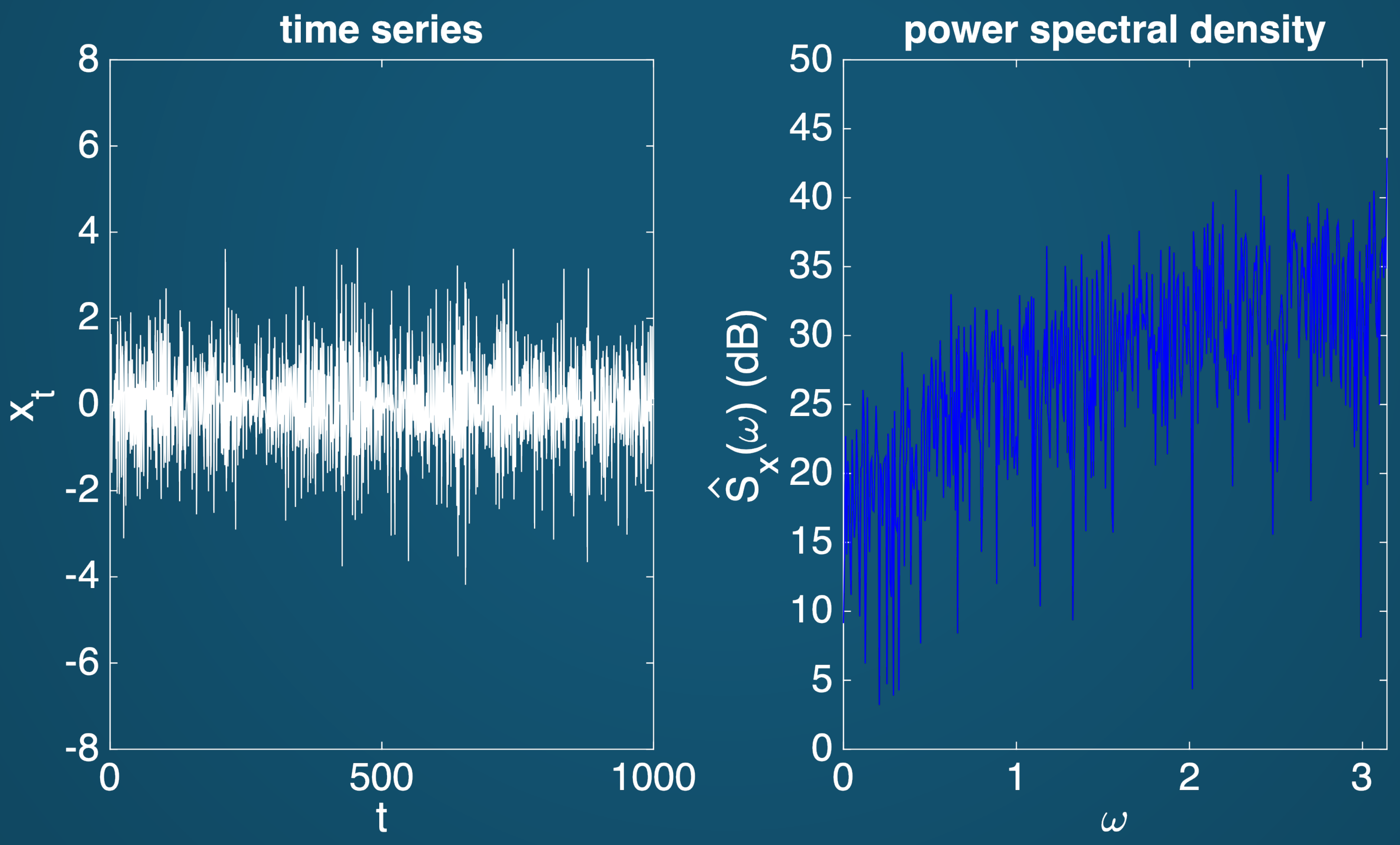
Auto-regressive process: AR(1)

$$x_t = 0.9x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$

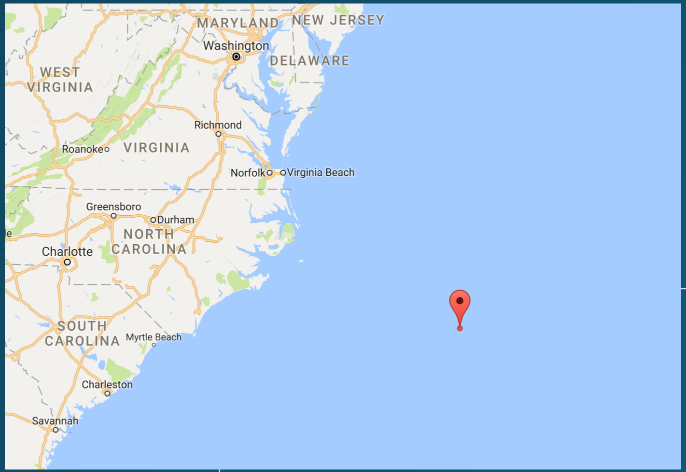


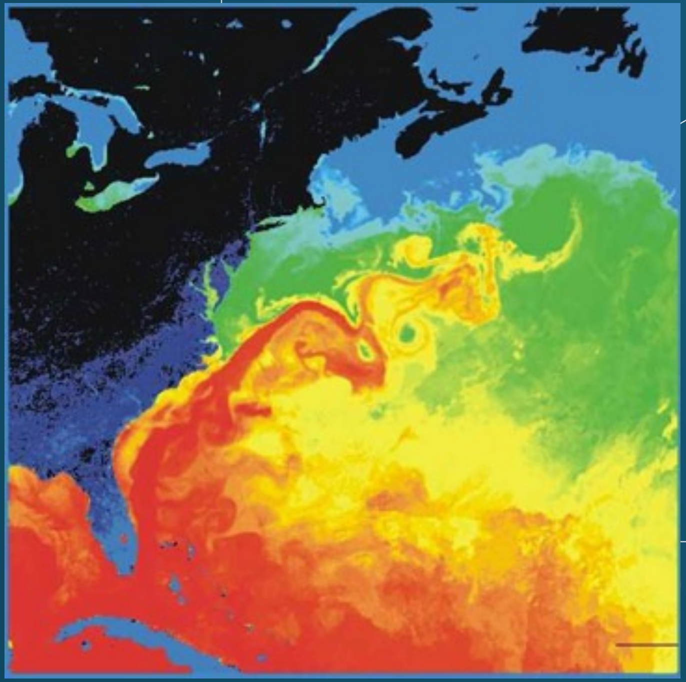
Moving average process: MA(1)

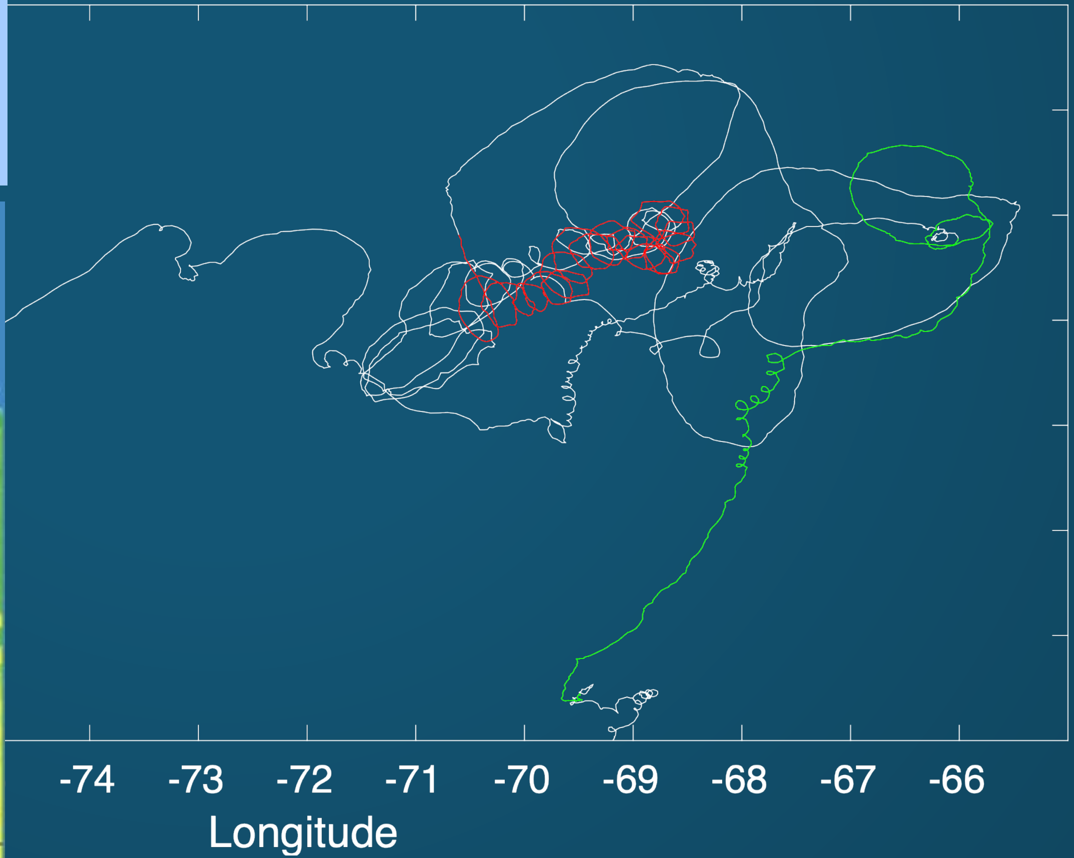
$$x_t = \varepsilon_t - 0.7\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$



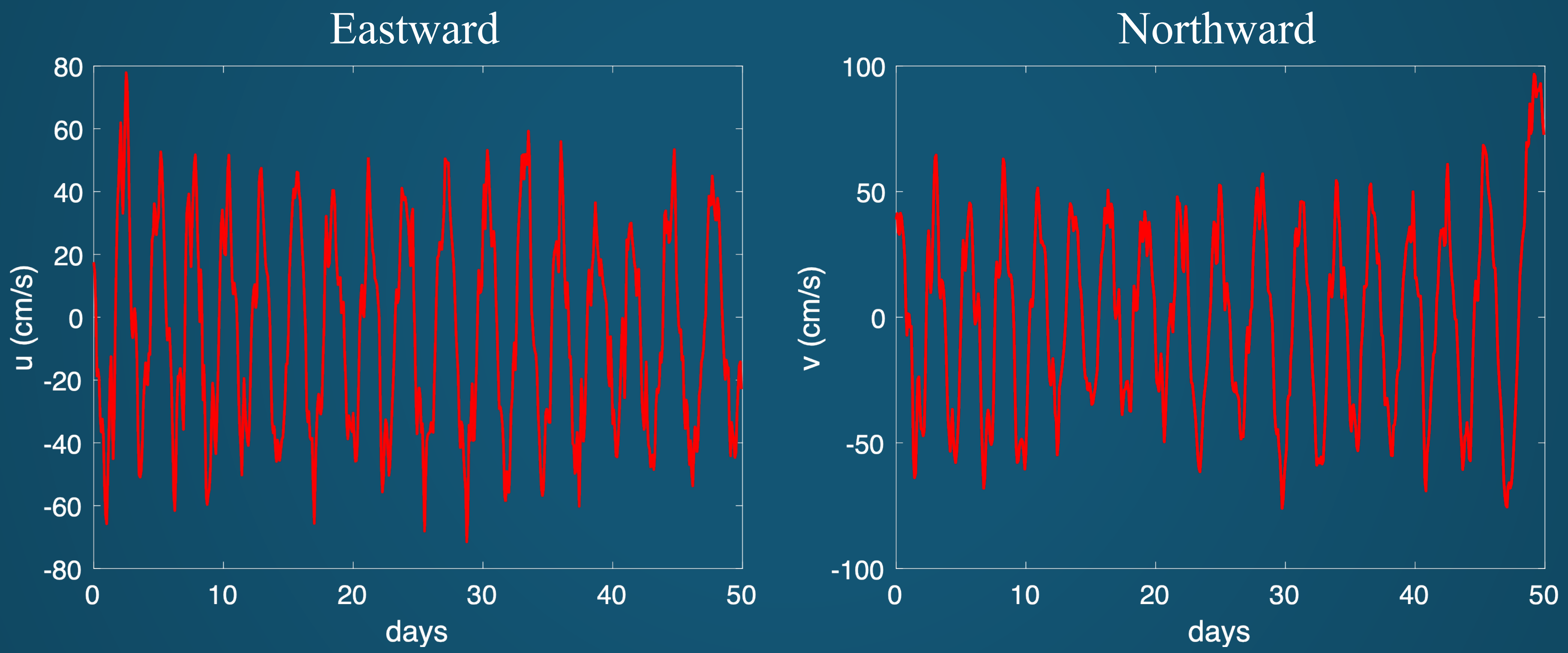
Oceanographic Drifter Data







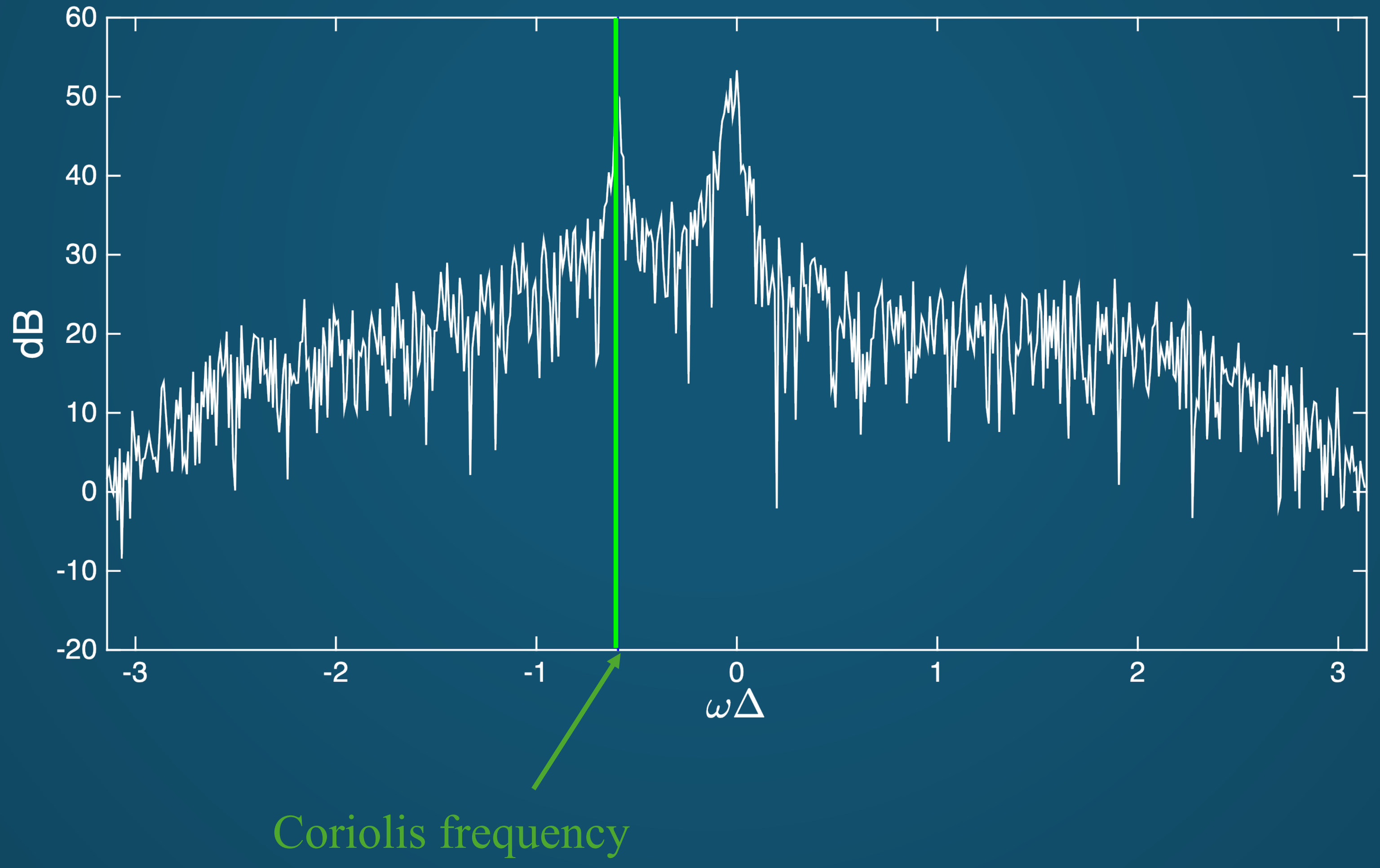
Drifter velocities for:



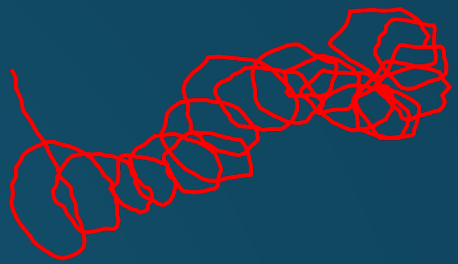
Represent particle velocities as complex-valued time series:

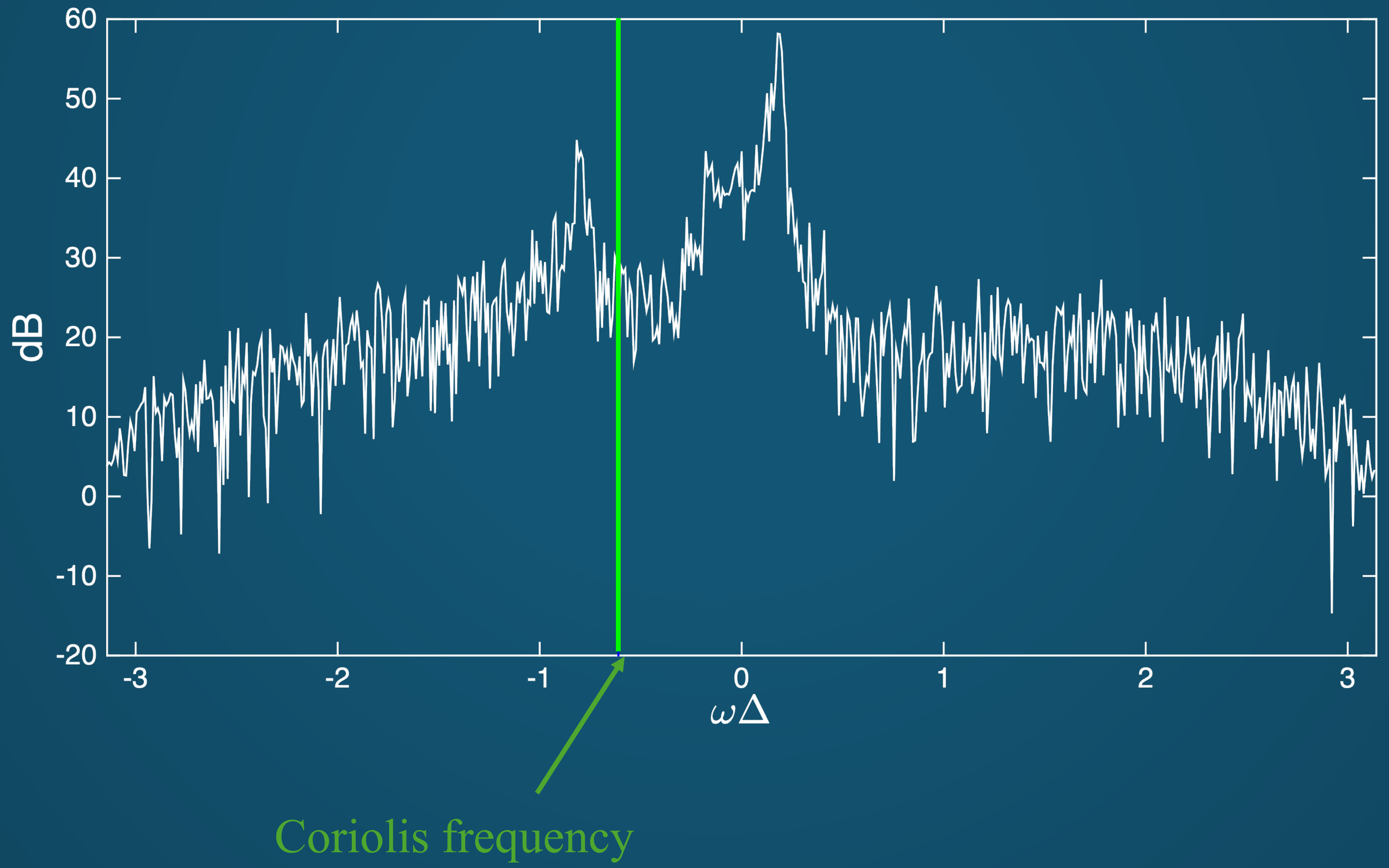
$$z_t = u_t + iv_t$$

Periodogram for velocities from:

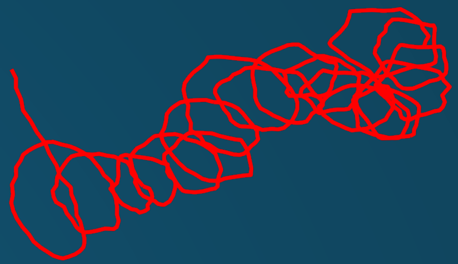


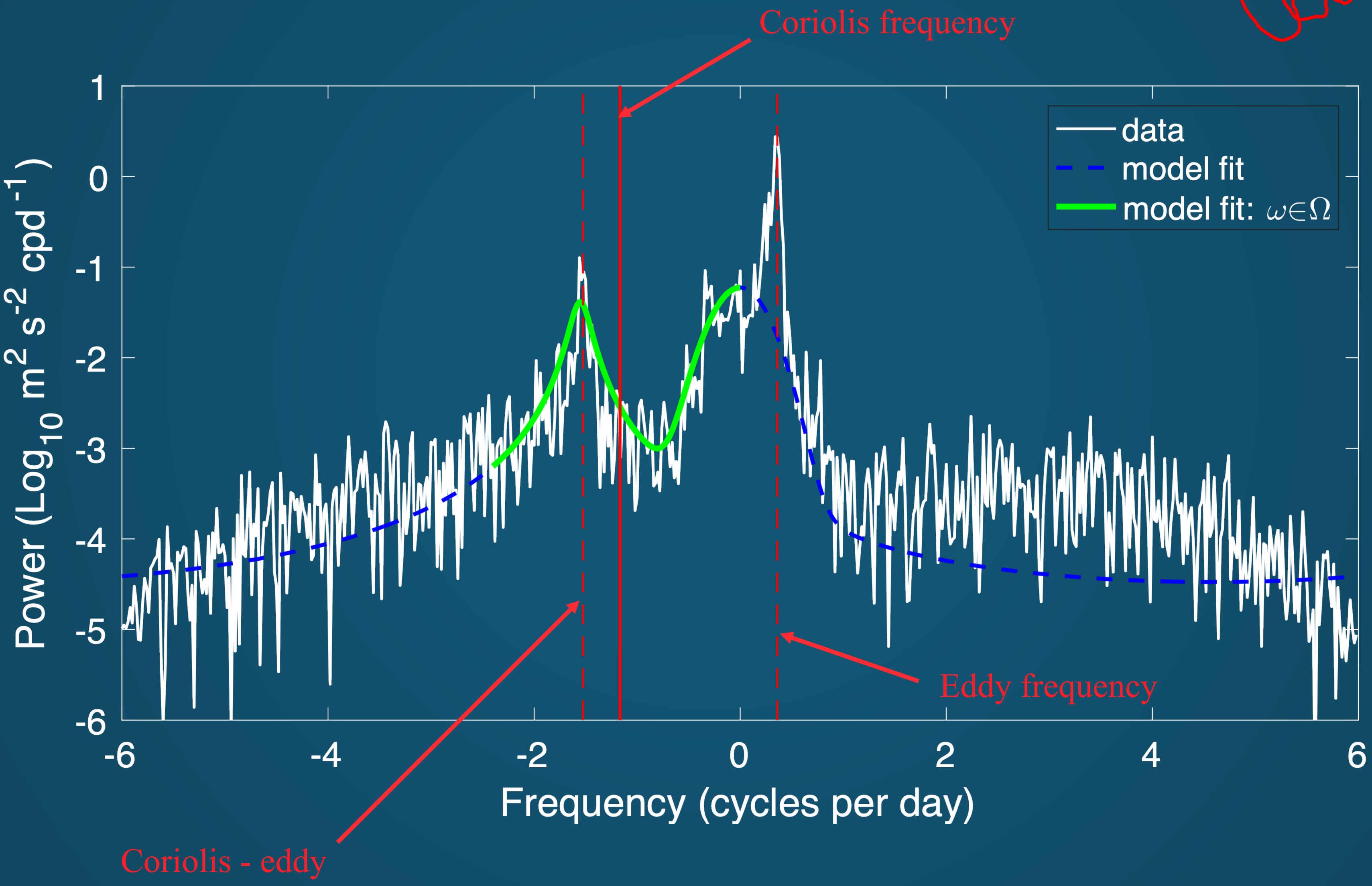
Periodogram for velocities from:





Periodogram for velocities from:





Sampling creates uncertainty

Theory: $S_x(\omega) = \lim_{T\to\infty} \mathbb{E}\left(\frac{1}{2T}\left|\int_{-T}^{T} x(t)e^{-i\omega t}dt\right|^2\right)$

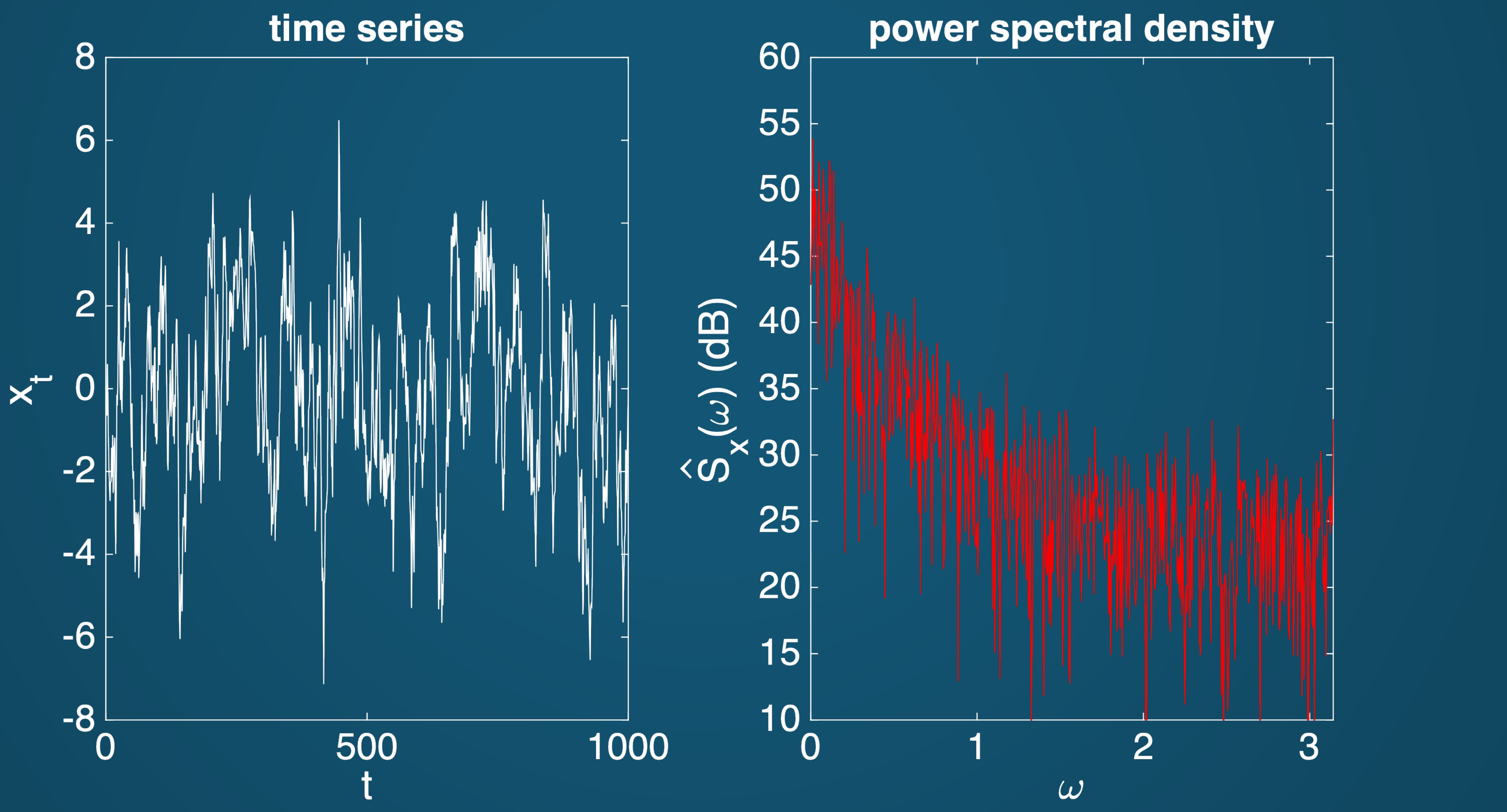
Practice: Observe some sample $X_1, \ldots, X_N$ at intervals Δ such that

$$\hat{S}(\omega) = \frac{\Delta}{N}\left|\sum_{t=1}^{N} X_t e^{-i\omega t}\right|^2$$

This is called the *periodogram* and is defined for $\omega \in [-\pi/\Delta, \pi/\Delta]$ where π/Δ is the *Nyquist frequency*.

Auto-regressive process: AR(1)

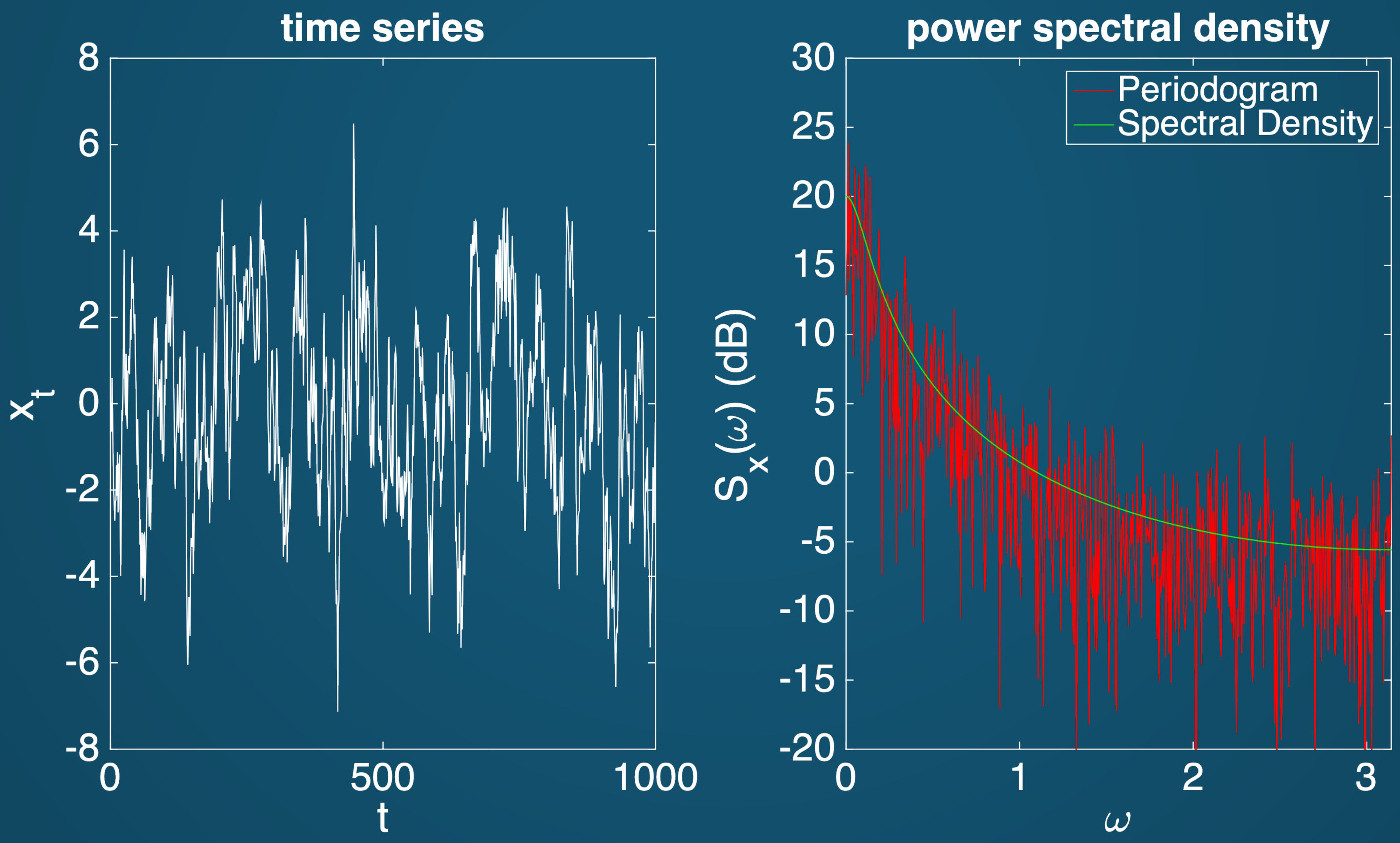
$$x_t = 0.9x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$



AR(1)

$$x_t = 0.9x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$

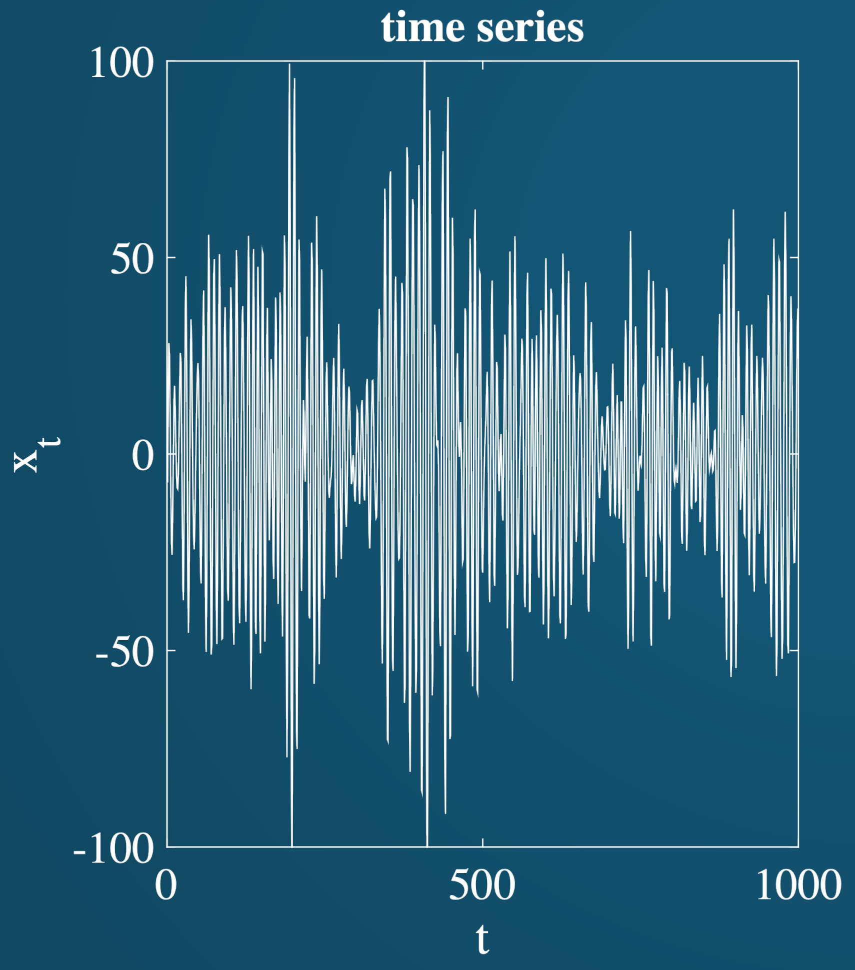
$$S_x(\omega) = \frac{\sigma_\varepsilon^2}{1 - 2\phi_1 \cos(\omega) + \phi_1^2}$$

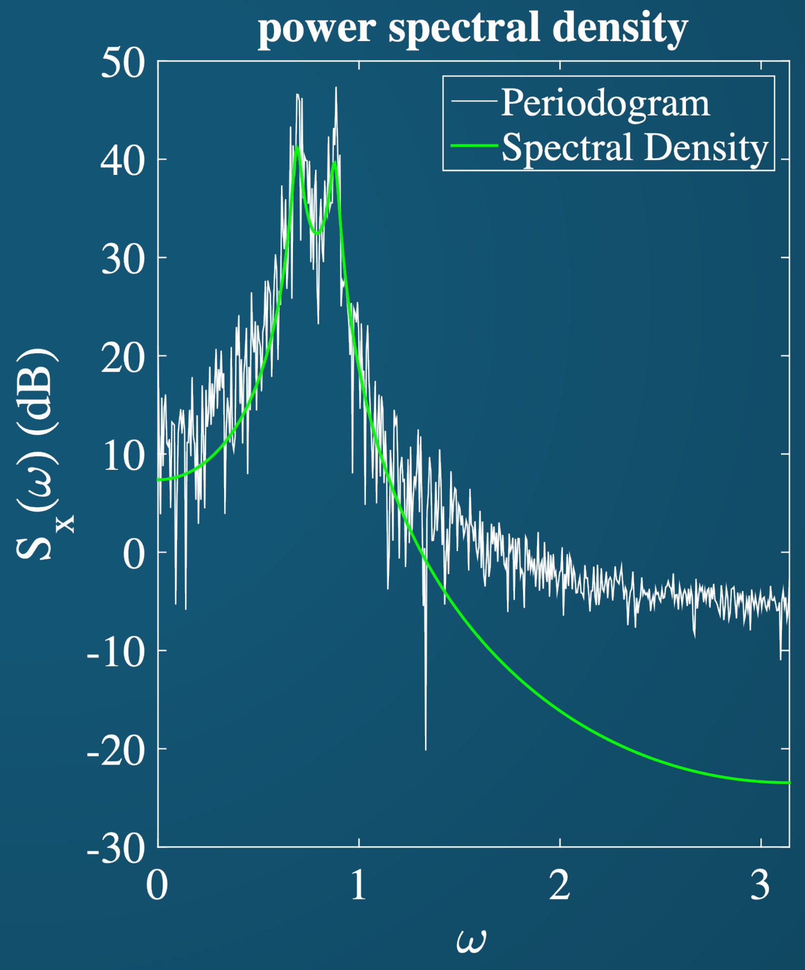


AR(4)

$$x_t = 2.7607x_{t-1} - 3.8106x_{t-2} + 2.6535x_{t-3} - 0.9238x_{t-4} + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1)$$

$$S_x(\omega) = \frac{\sigma_\epsilon^2}{|1 - \sum_{k=1}^{4} \phi_k e^{-ikw}|^2}$$





“More lives have been lost looking at the periodogram than by any other action involving time series”

John W. Tukey

Remember: sampling creates uncertainty

Theory: $S_x(\omega) = \lim_{T\to\infty} \mathbb{E}\left(\frac{1}{2T}\left|\int_{-T}^{T} x(t)e^{-i\omega t}dt\right|^2\right)$

Practice: $\hat{S}_X(\omega) = \left|\sum_{t=1}^{N} X_t e^{-i\omega t}\right|^2 / N$

Convolution: $\mathbb{E}\left\{\hat{S}_X(\omega)\right\} = \int_{-\pi}^{\pi} \mathcal{F}(\omega - \omega')S_x(\omega')d\omega'$

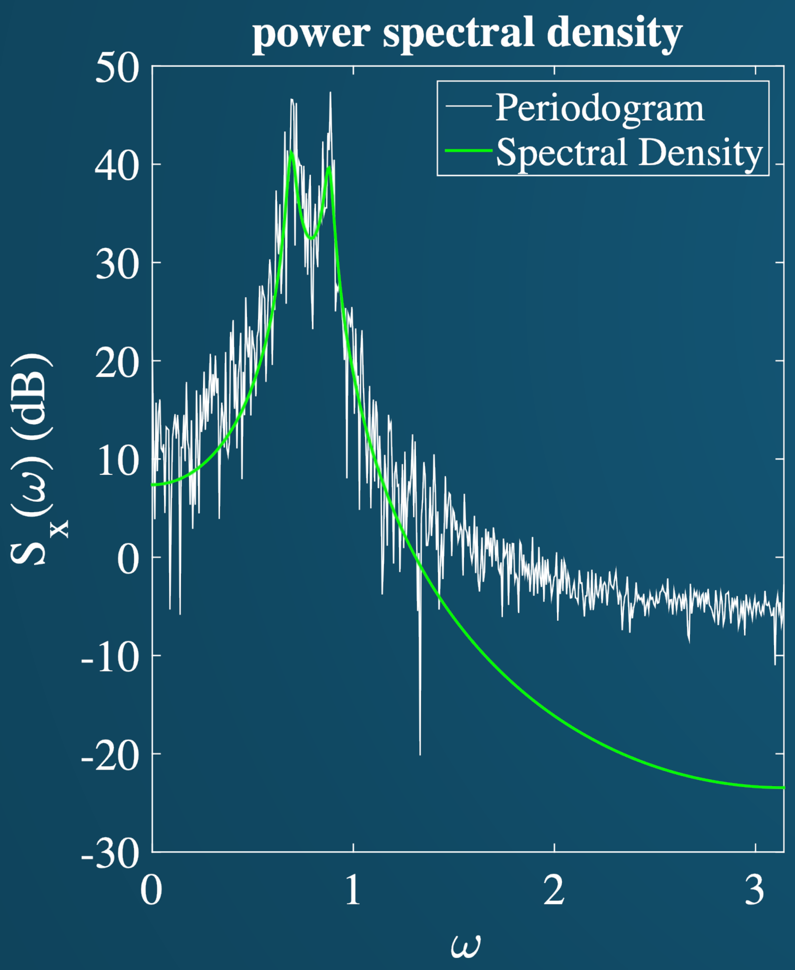
where $\mathcal{F}(\cdot)$ is the Fejér kernel: $\mathcal{F}(\omega) = \frac{1}{2\pi N}\frac{\sin^2(N\omega/2)}{\sin^2(\omega/2)}$

Percival and Walden, page 198

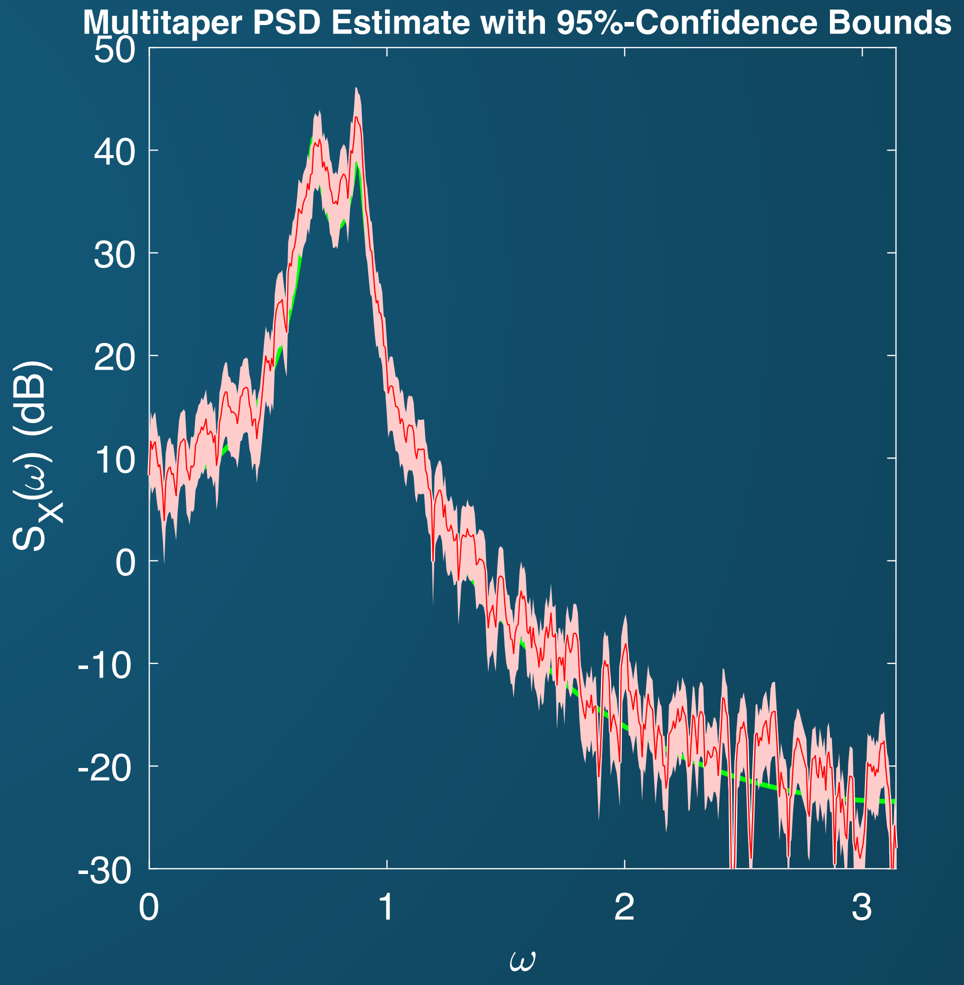
AR(4)

$$x_t = 2.7607x_{t-1} - 3.8106x_{t-2} + 2.6535x_{t-3} - 0.9238x_{t-4} + \varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1)$$

$$S_x(\omega) = \frac{\sigma_\epsilon^2}{|1 - \sum_{k=1}^{4} \phi_k e^{-ikw}|^2}$$



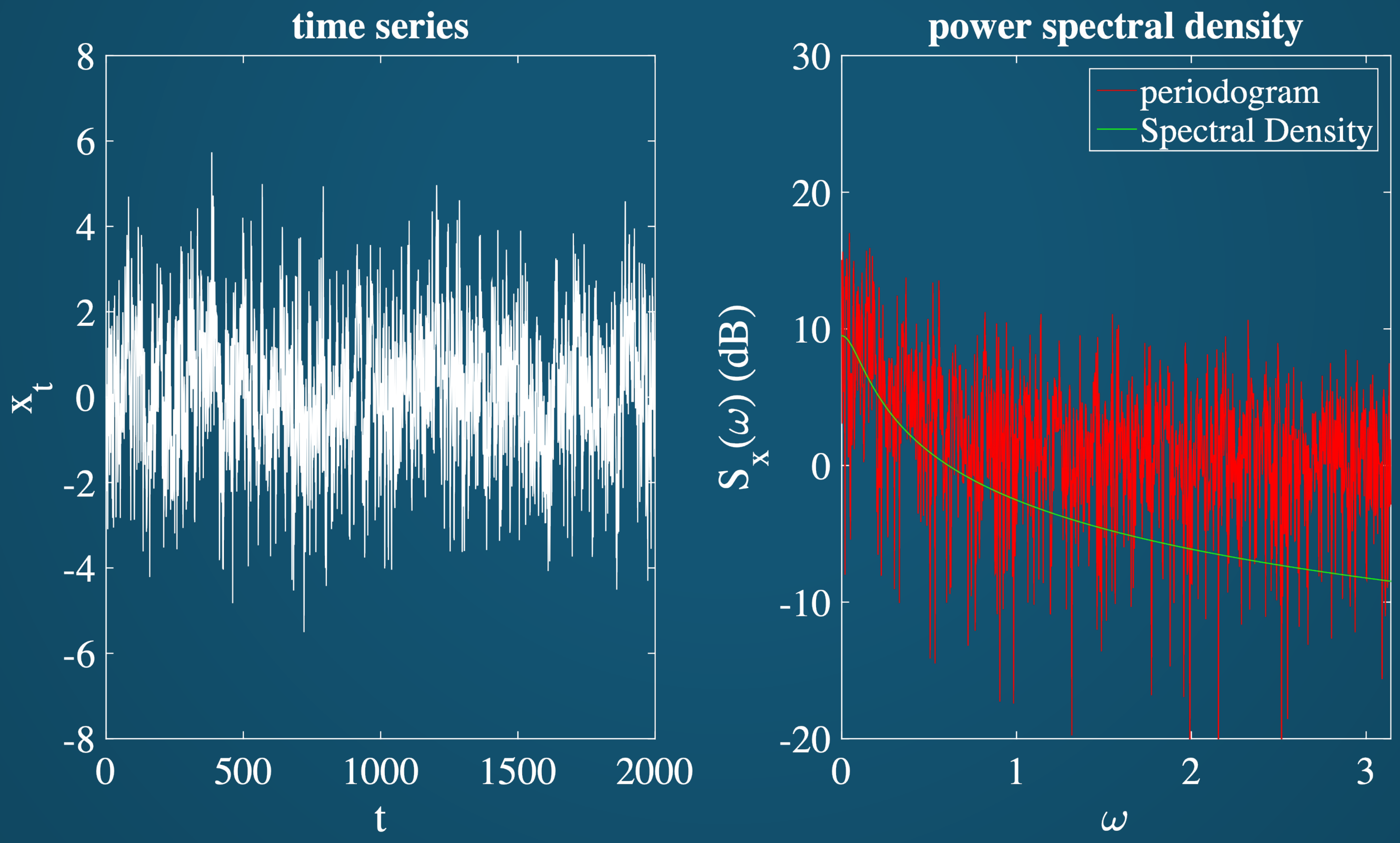
Percival and Walden page 354



Another good option is Welch's method, P+W page 412

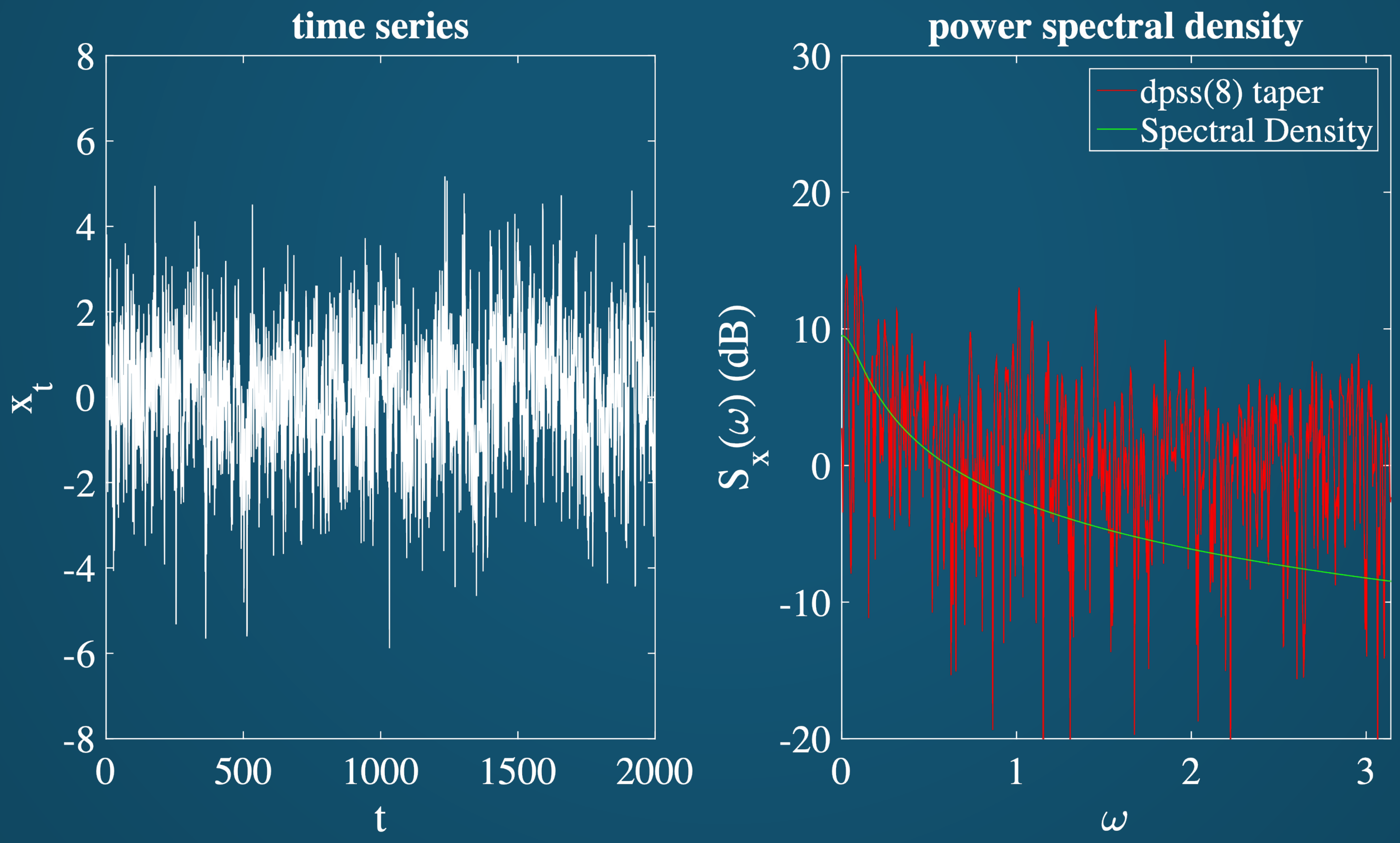
$$X(t) \sim \text{Matérn}(A = 1, \alpha = 0.6, h = 0.1)$$

$$S_x(\omega) = \frac{A^2}{(\omega^2 + h^2)^\alpha}$$

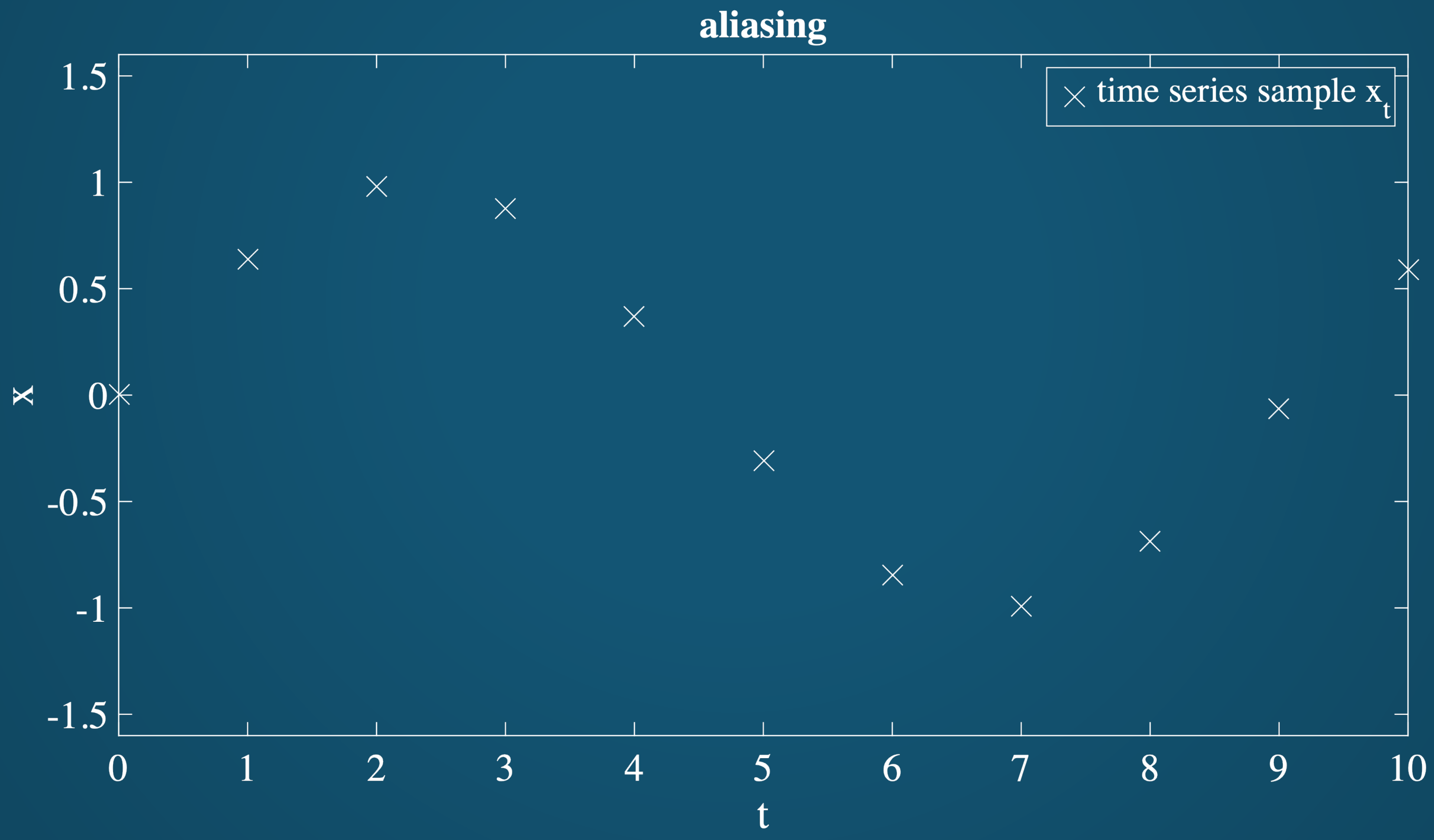


$$X(t) \sim \text{Matérn}(A = 1, \alpha = 0.6, h = 0.1)$$

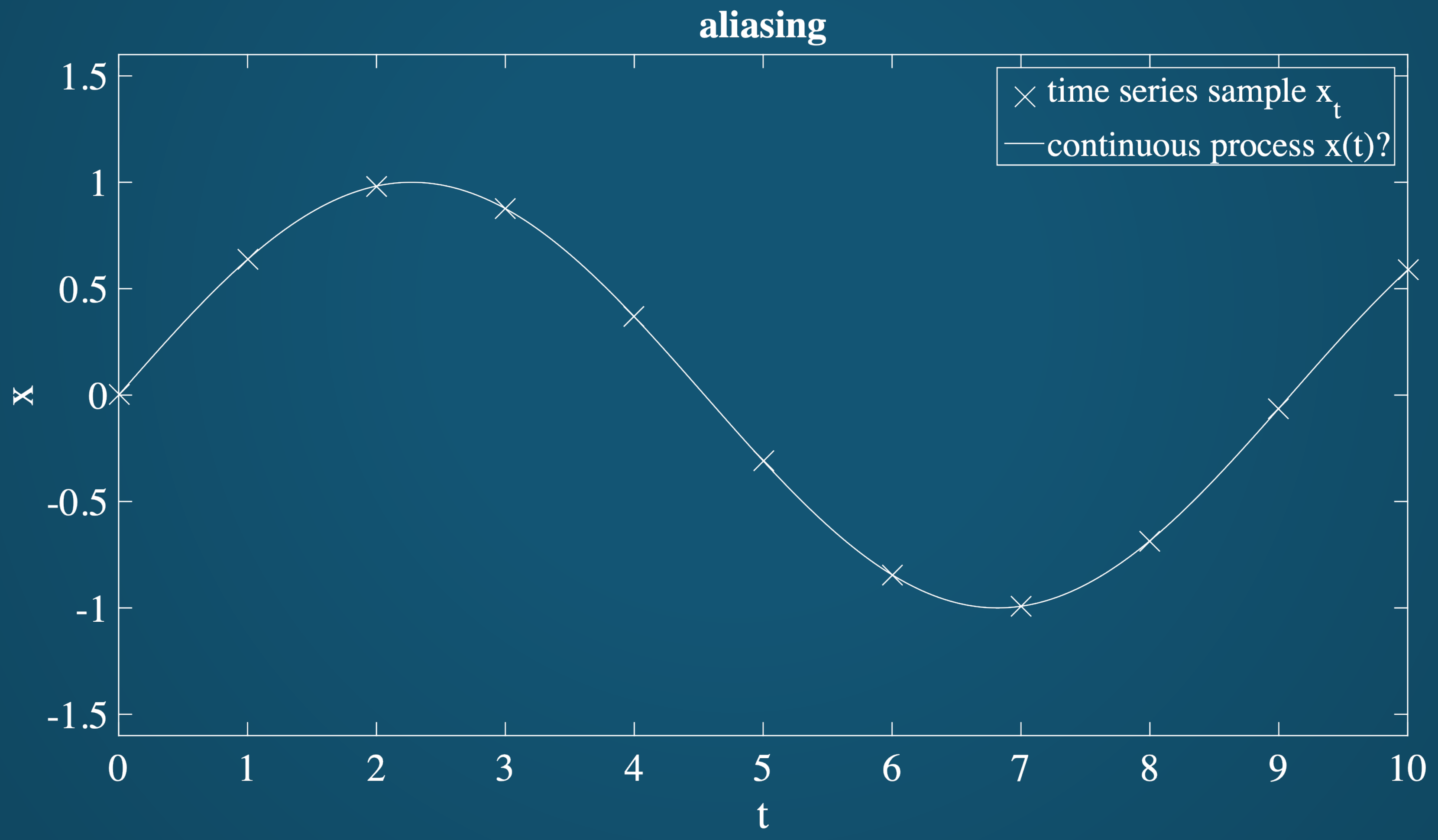
$$S_x(\omega) = \frac{A^2}{(\omega^2 + h^2)^\alpha}$$



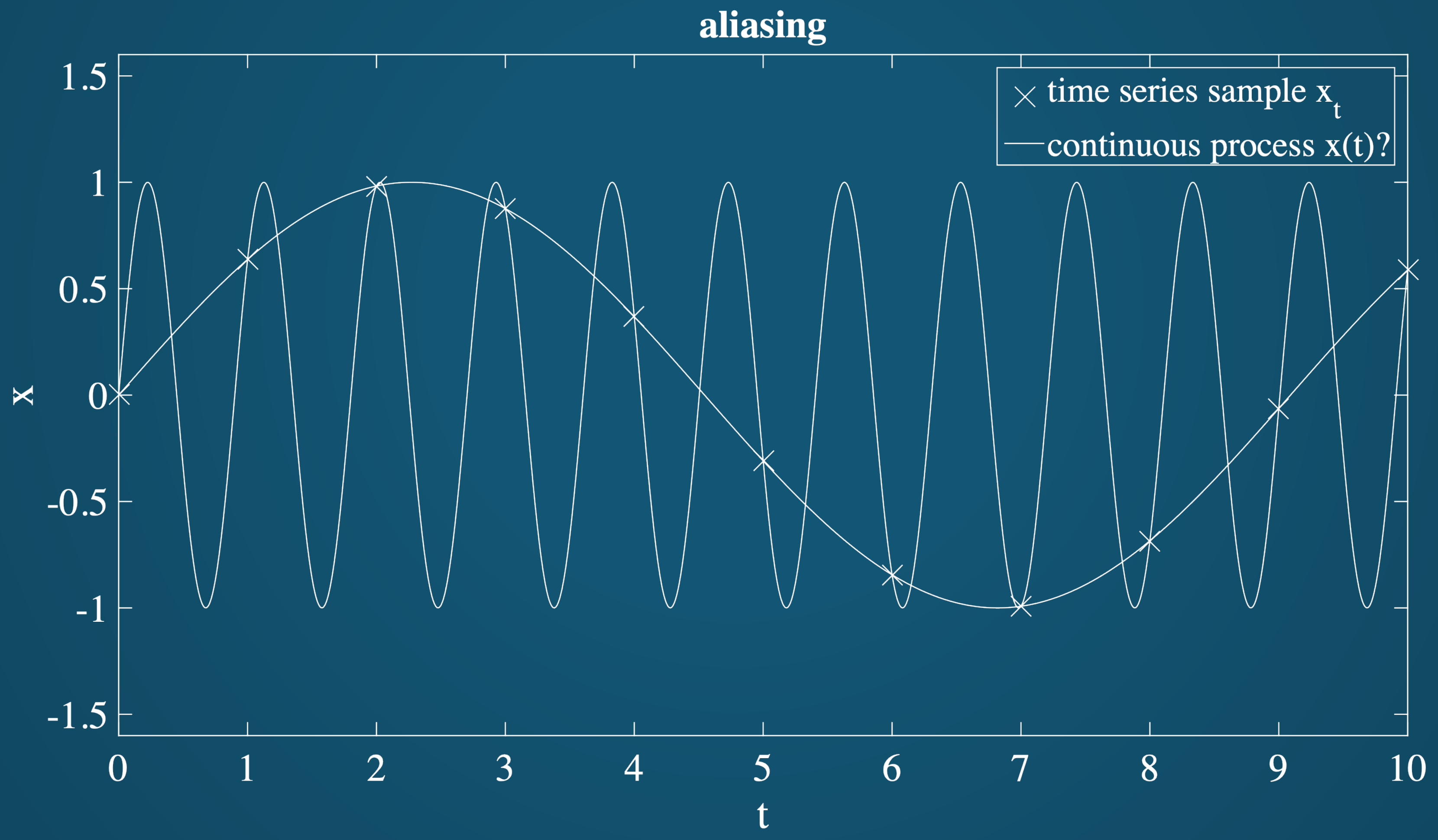
What's wrong this time? Sampling isn't (usually) continuous…



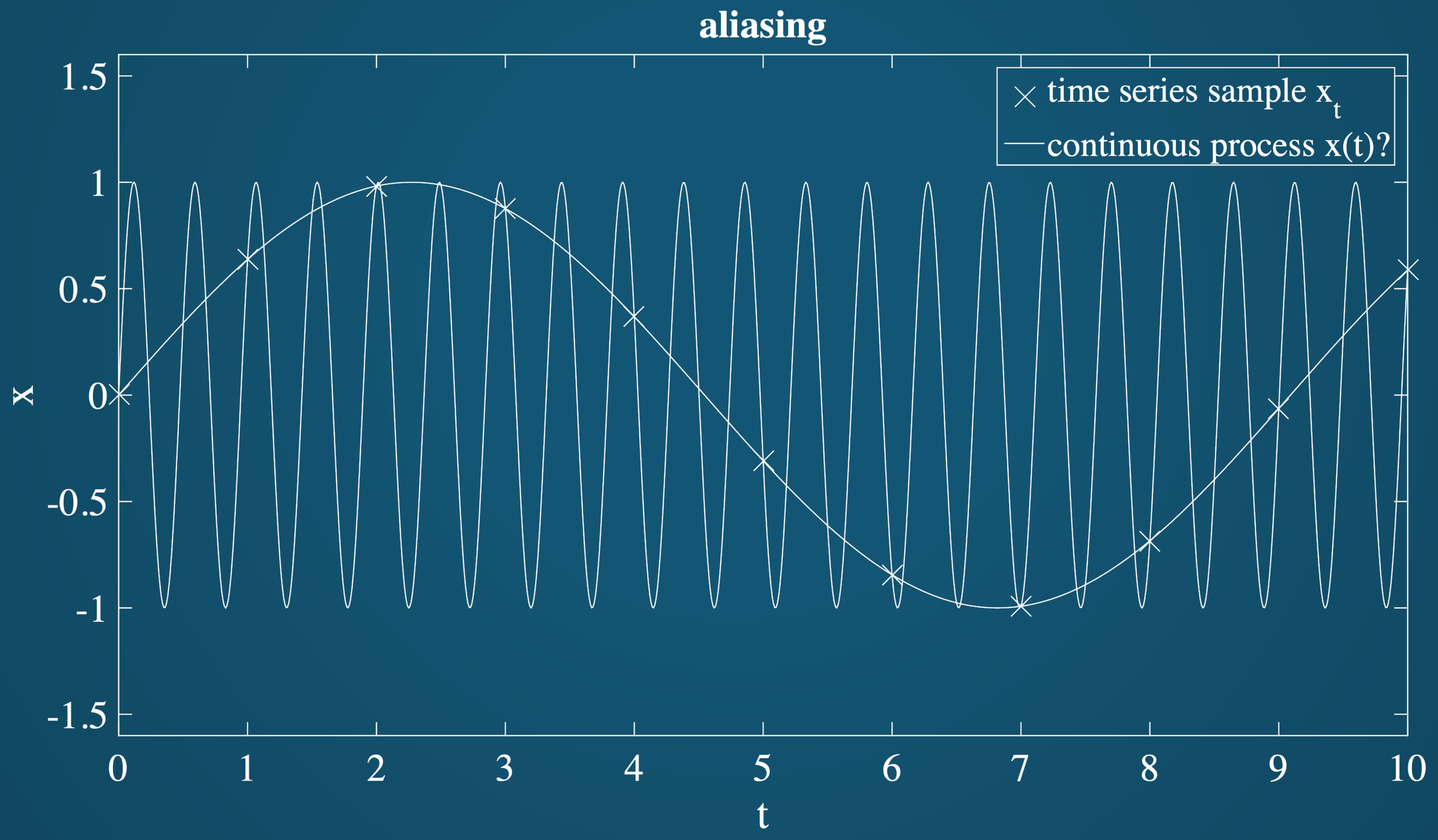
Percival and Walden, page 97

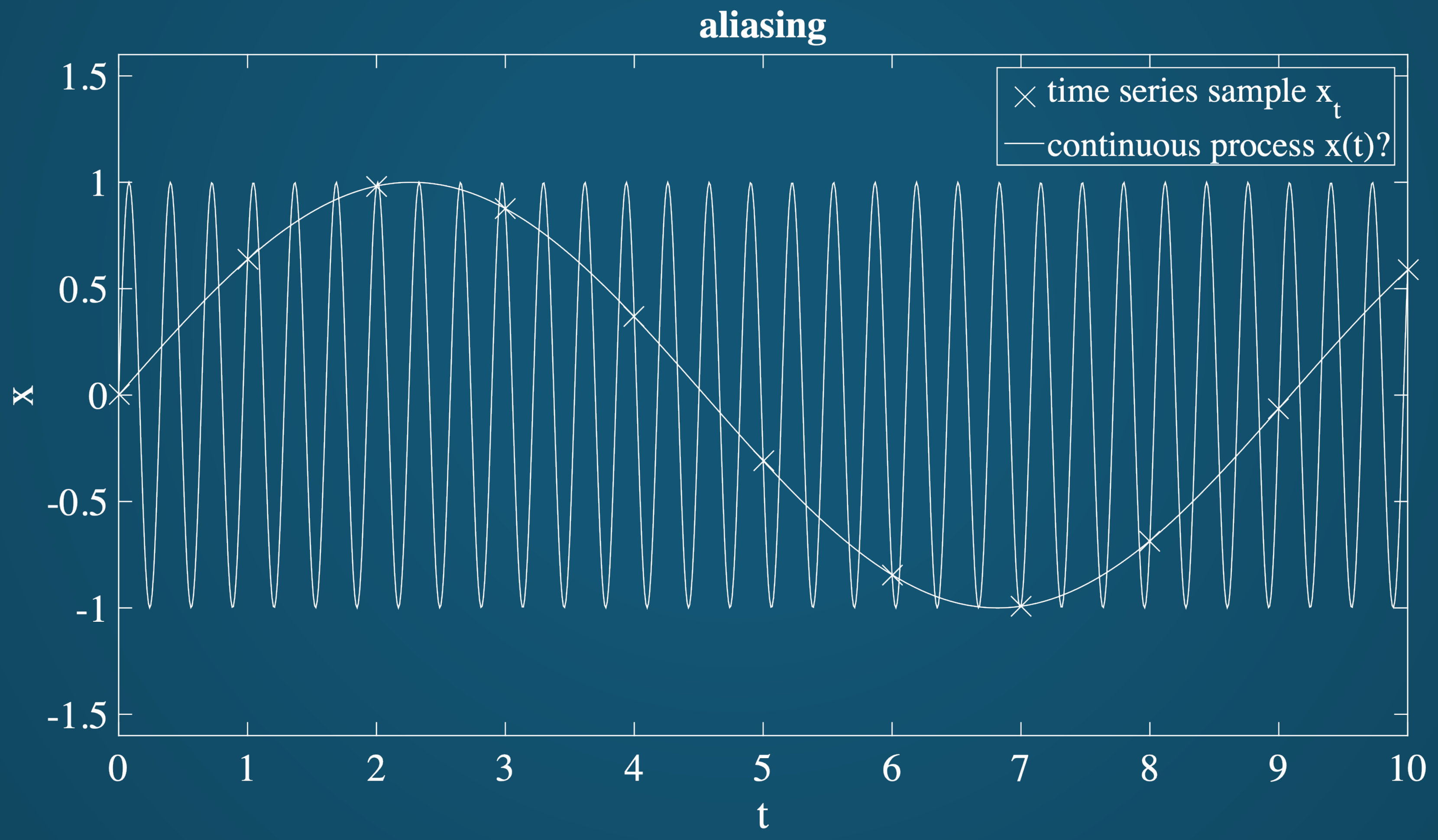


aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t

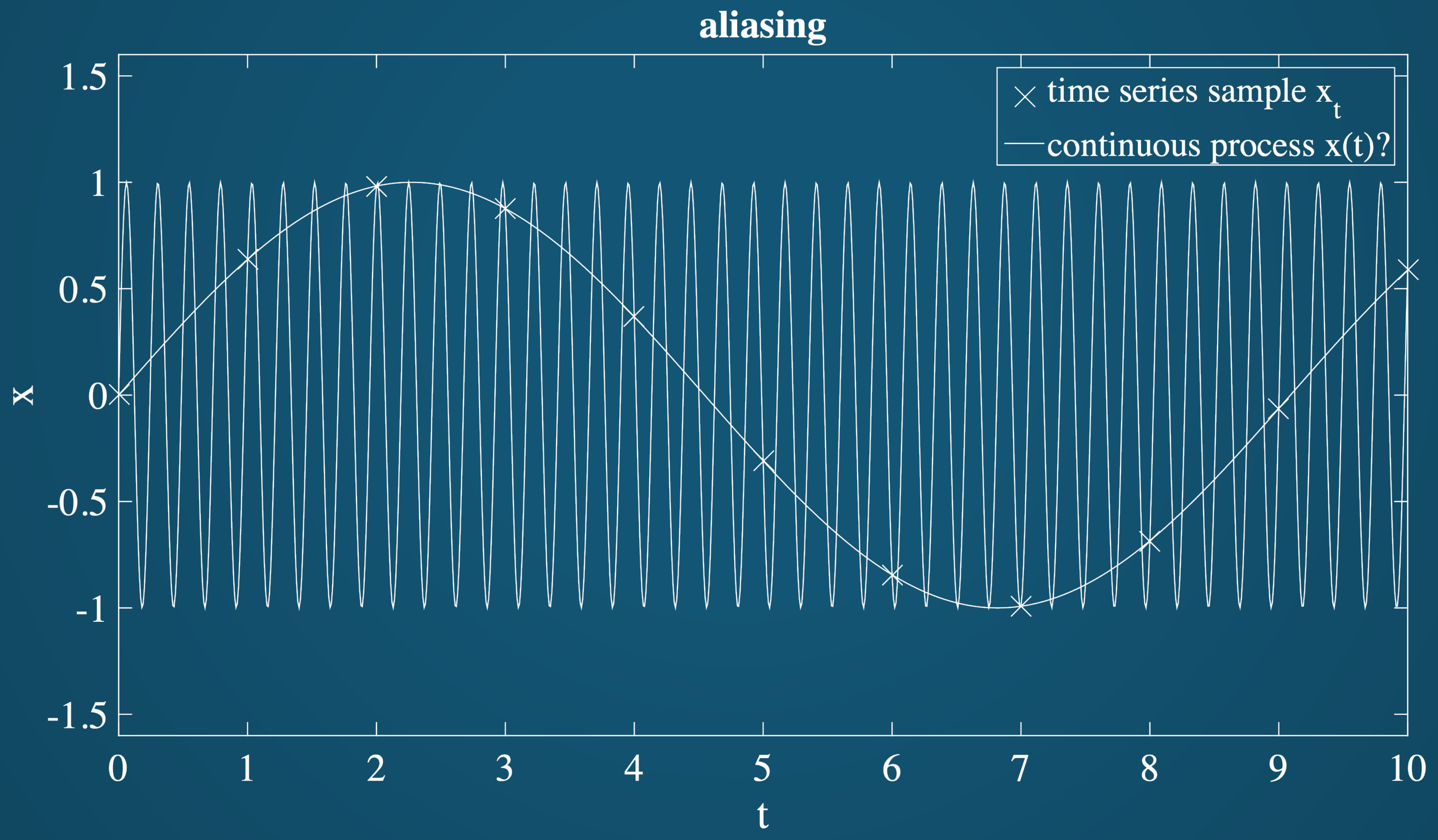


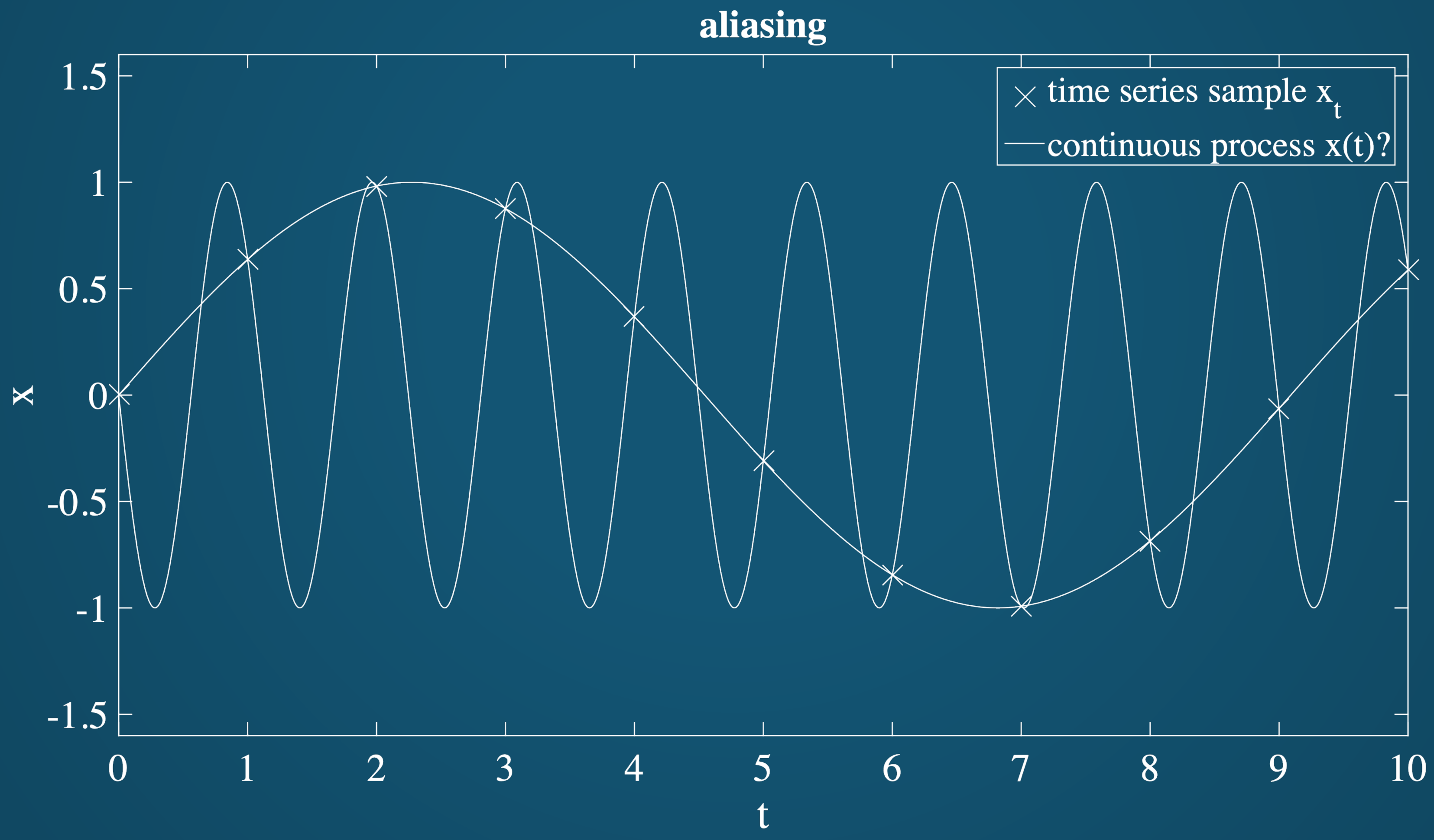
aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t

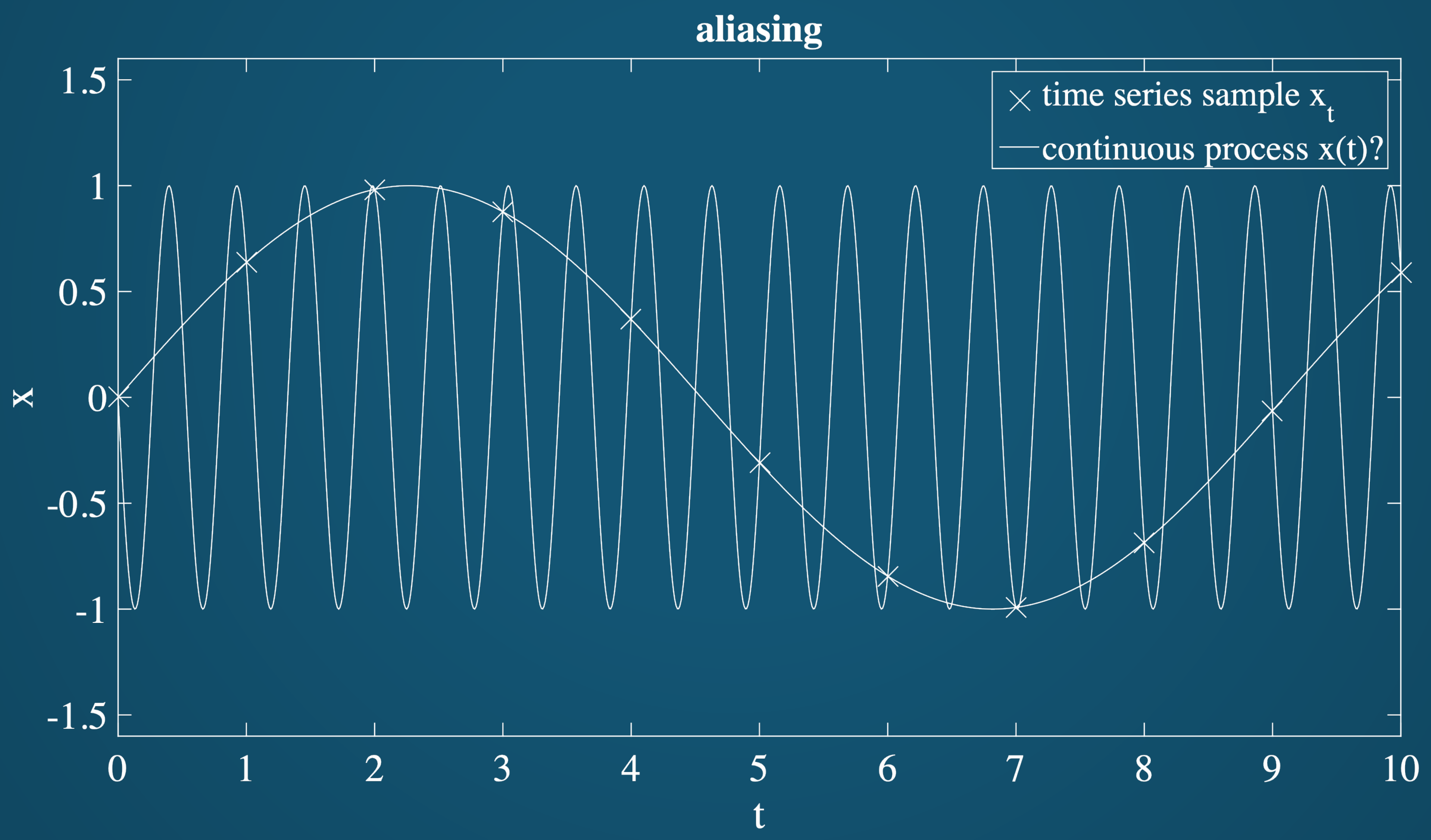

aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t



aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t


aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t


aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t

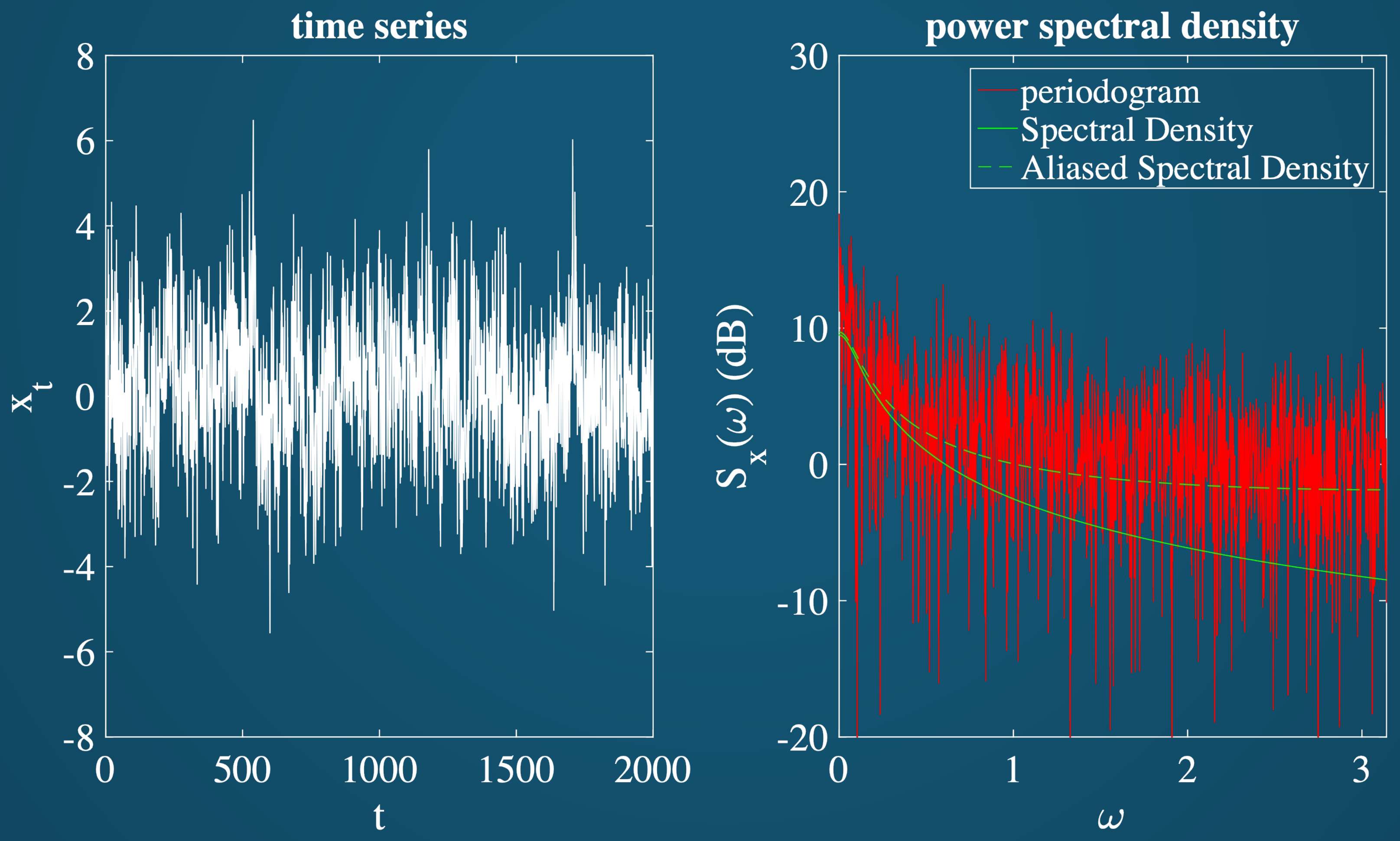

aliasing
time series sample x_t
continuous process x(t)?
1.5
1
0.5
0
-0.5
-1
-1.5
x
0
1
2
3
4
5
6
7
8
9
10
t

In 2-D the aliasing problem is also known as “the wagon-wheel effect”

Link to Video

$$X(t) \sim \text{Matérn}(A = 1, \alpha = 0.6, h = 0.1)$$

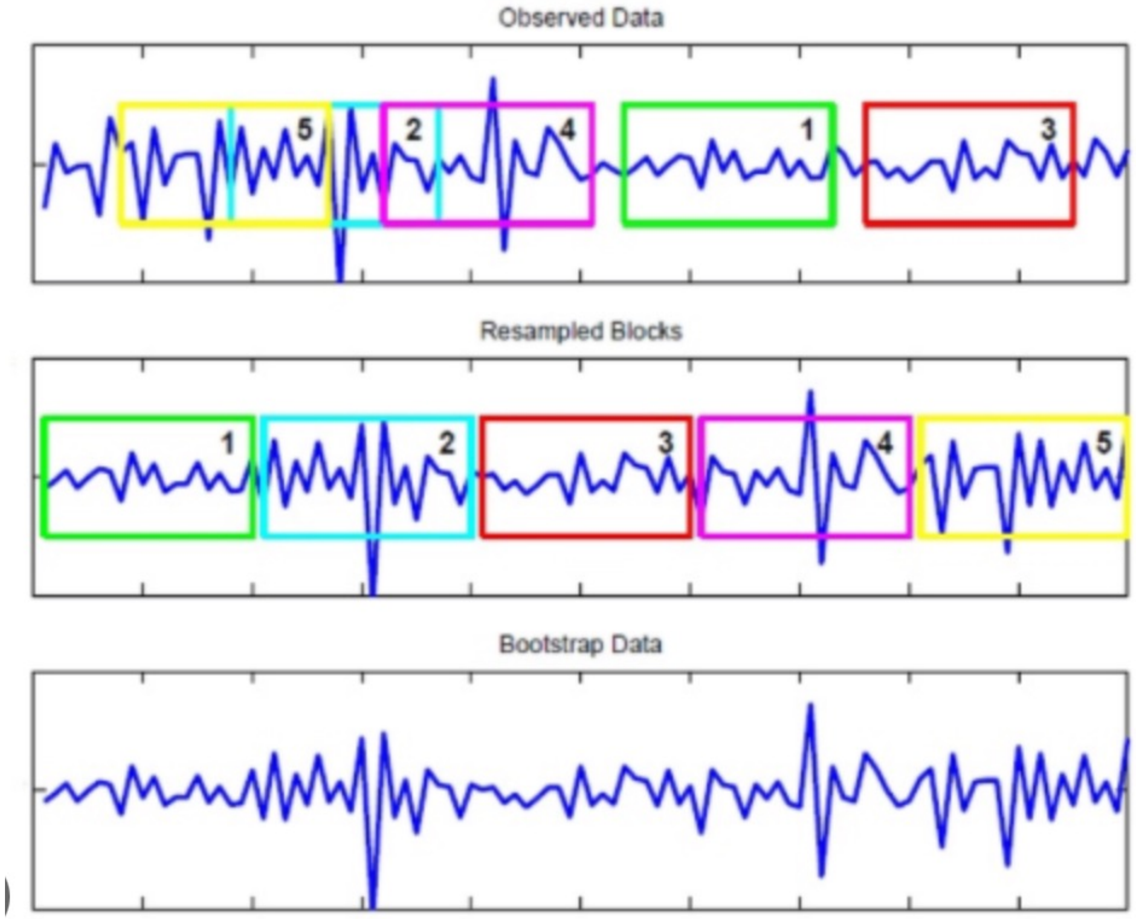
$$S_x(\omega) = \sum_{k=-\infty}^{\infty} \frac{A^2}{((\omega + 2\pi k)^2 + h^2)^\alpha}$$



Nonstationarity: Time-frequency "spectrograms" and Heisenberg-Gabor uncertainty

Bonus Section: How to bootstrap a time series!

Block bootstrap



Other approaches include:

- Parametric model fitting and then sampling from the model
- Spectral analysis methods which in essence sample from the spectrum and then Fourier transform back
- For existing Python code for the block bootstrap checkout the ARCH 6.3.0 package by Kevin Sheppard: https://doi.org/10.5281/zenodo.593254

Reference: Politis, D. N., & White, H. (2004). Automatic Block-Length Selection for the Dependent Bootstrap. Econometric Reviews, 23(1), 53–70.