Analysing Ocean Drifter Data in R

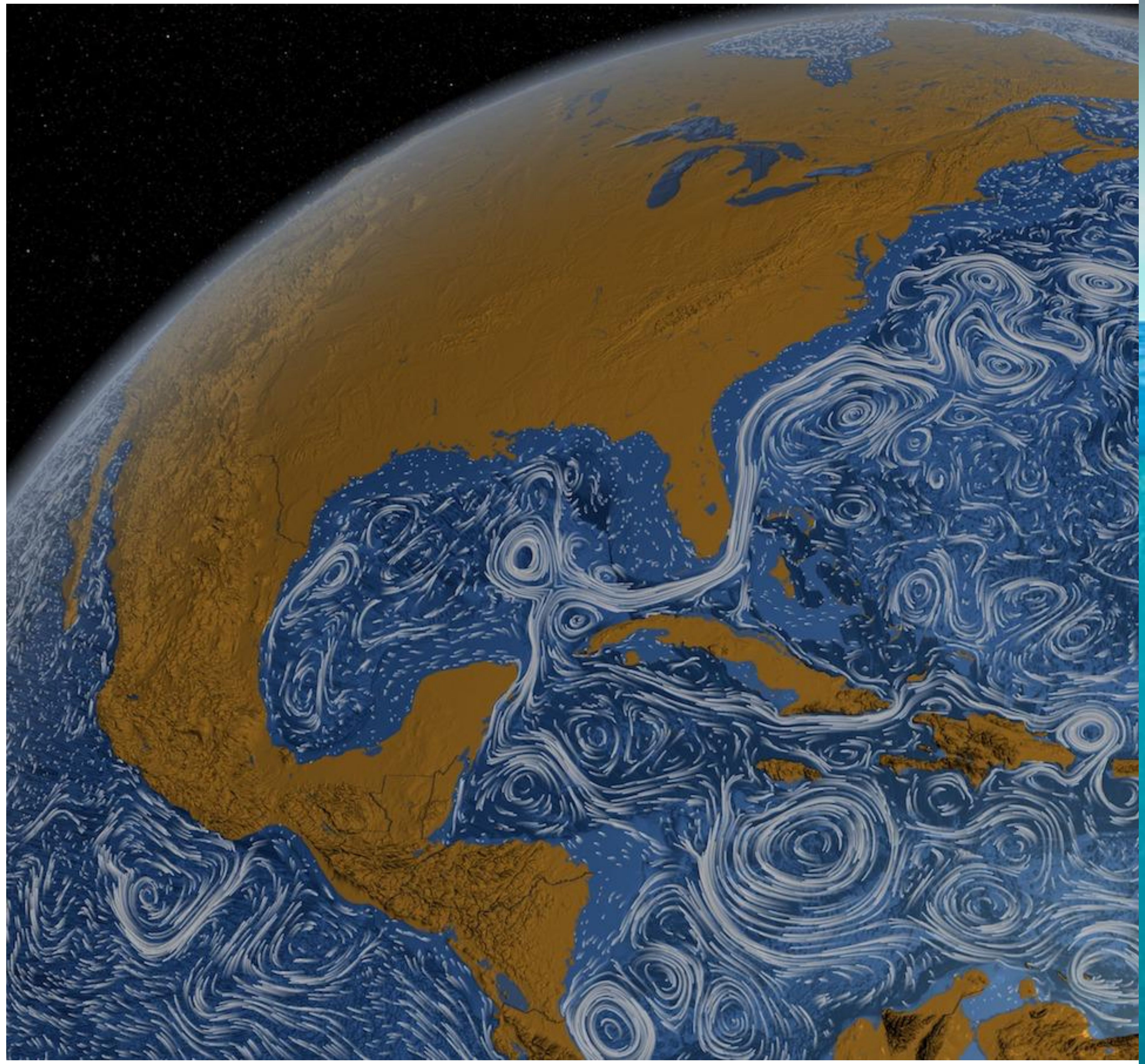
Adam M. Sykulski
Senior Lecturer in Statistics
Imperial College London
adam.sykulski@imperial.ac.uk

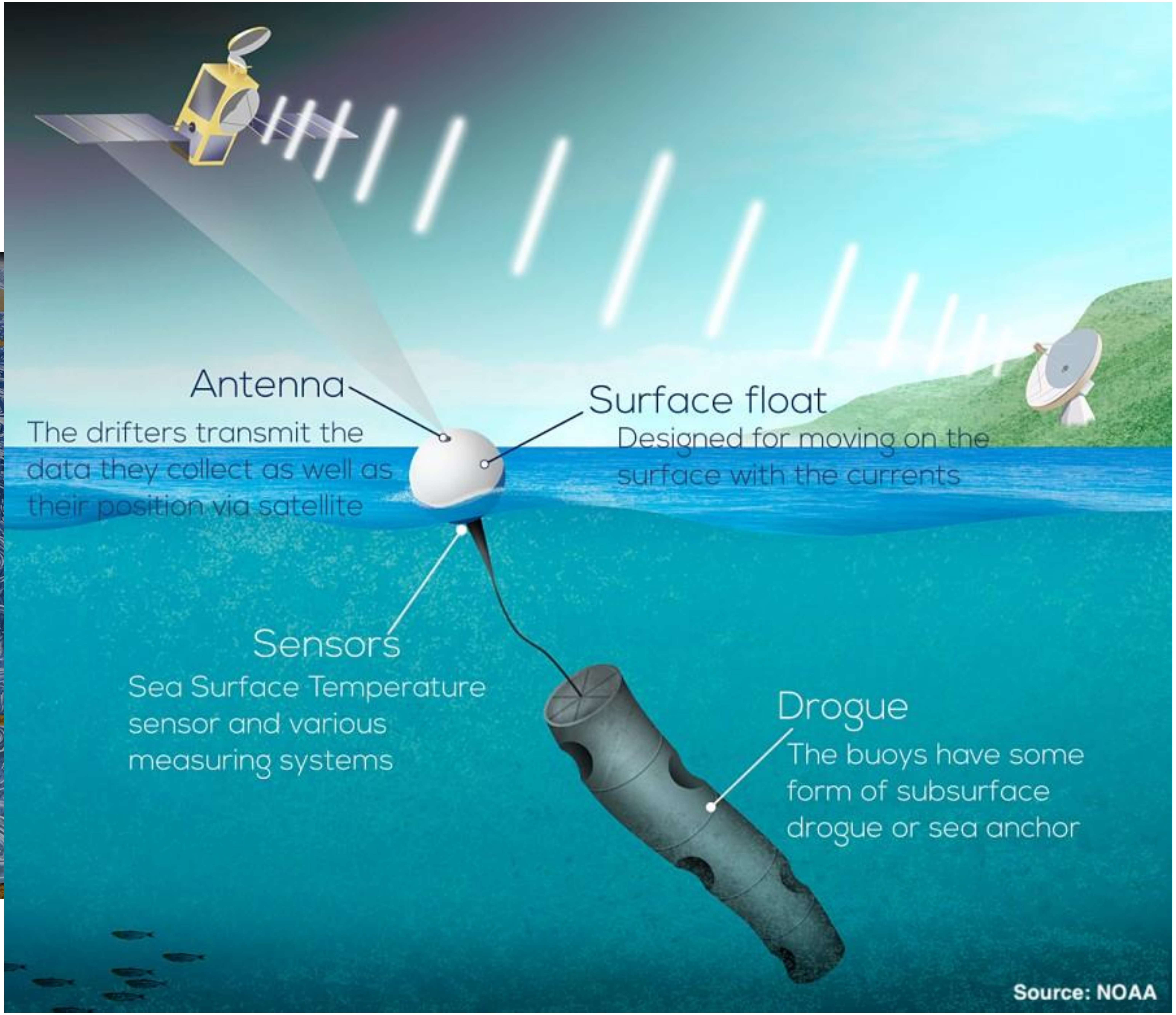
The goal(s) of this session

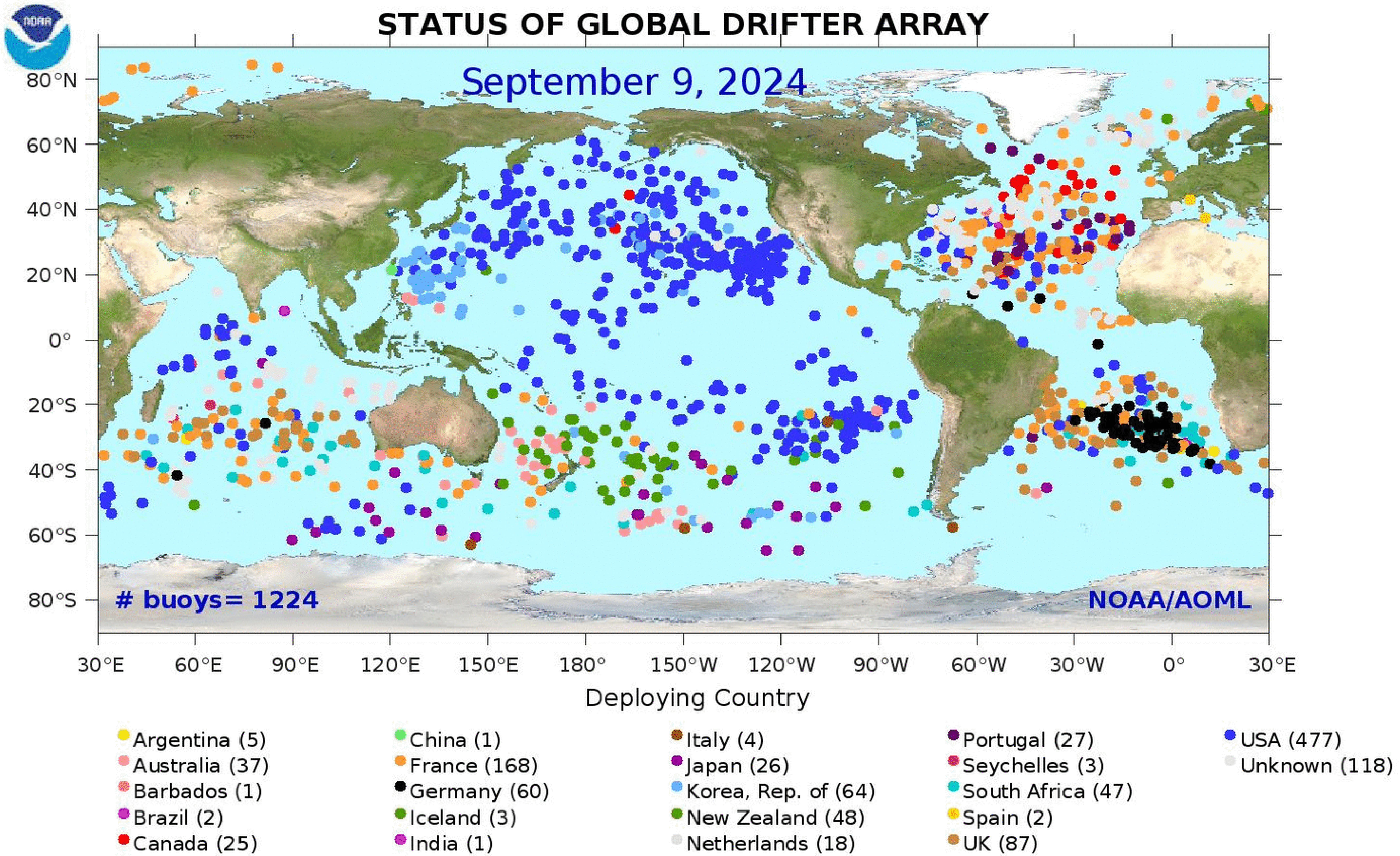
- Learning about ocean drifter data, what it is and how to model and analyse it
- Learning how to do statistical analysis and visualisations in R

- Bonus: working in small groups, learning to communicate, code efficiently and use resources!
- Next Wednesday (10.30am), prepare one slide showing your favourite visualisation from today's programming session! Email it to me by next Tuesday evening (adam.sykulski@imperial.ac.uk)

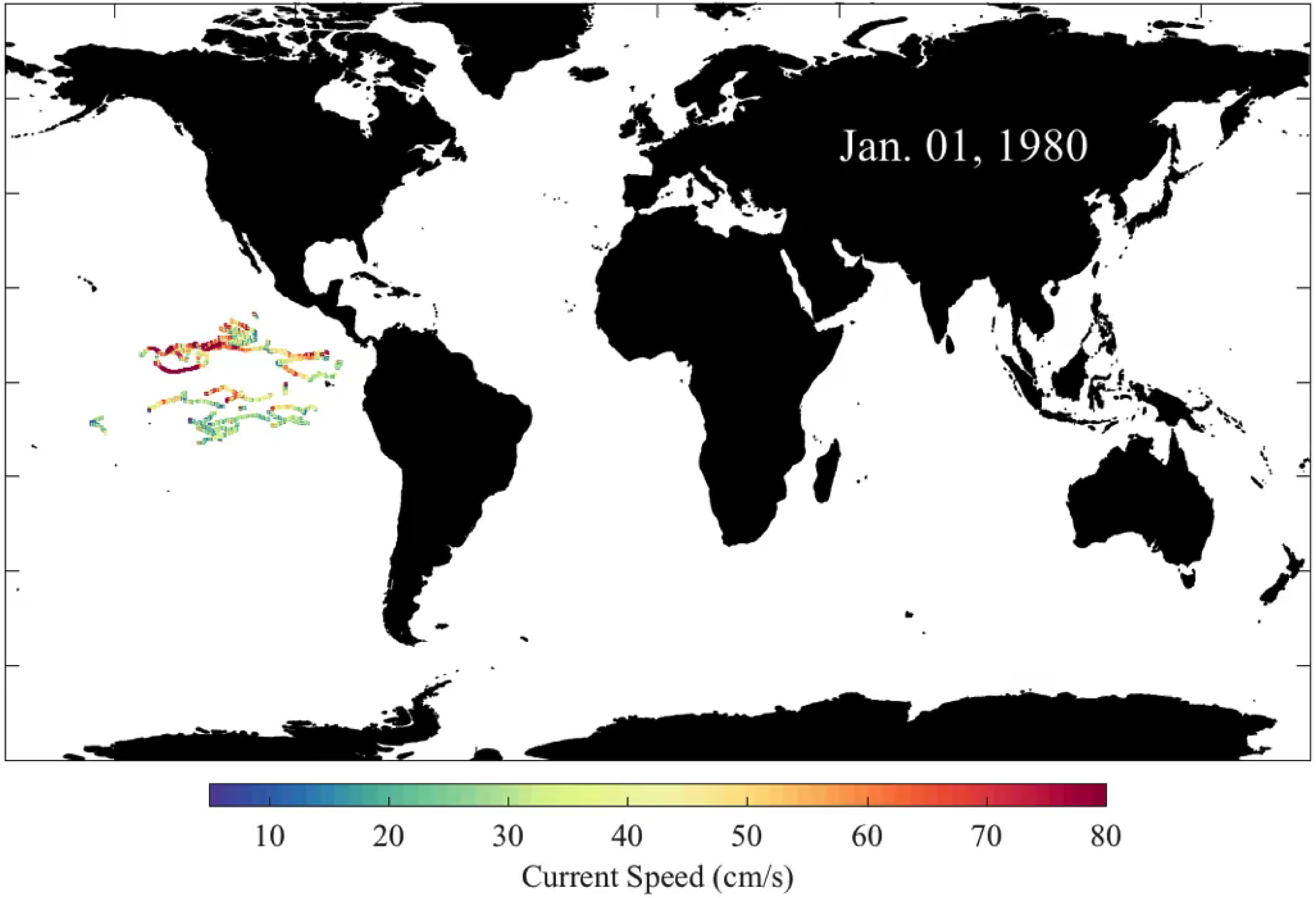
Monitoring our Oceans



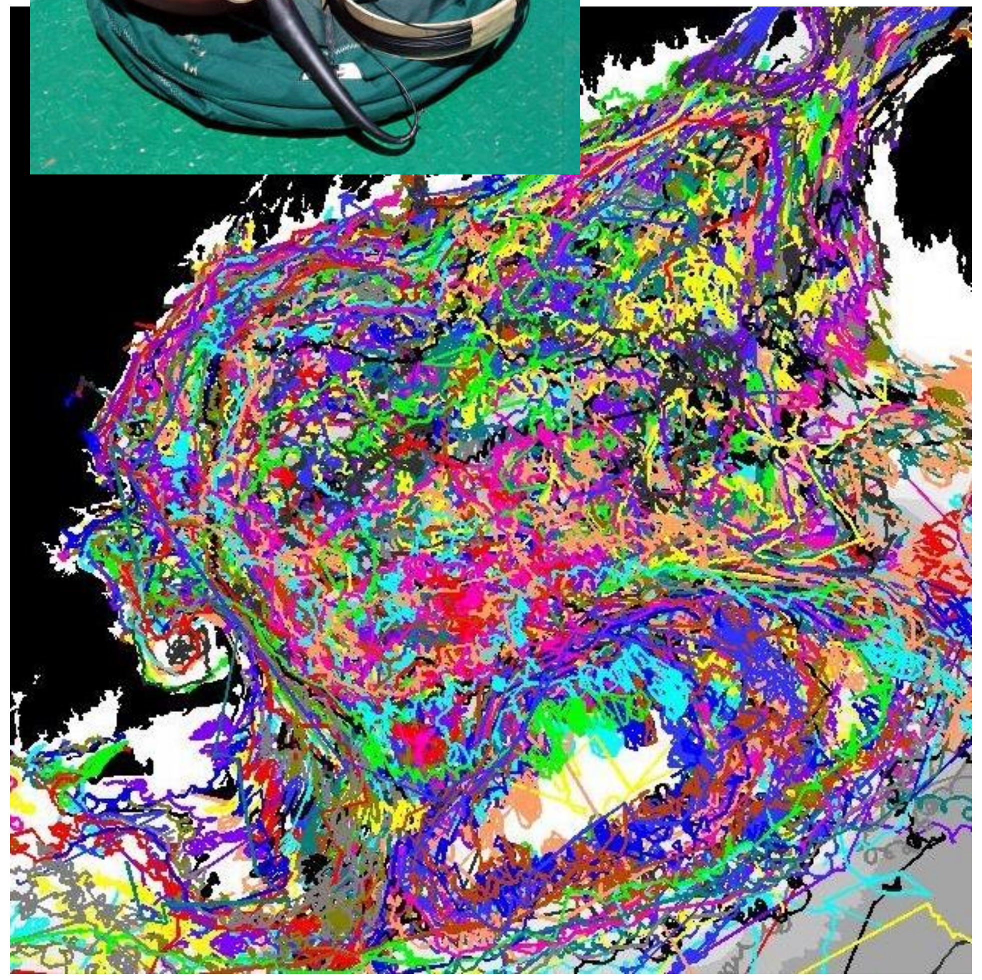


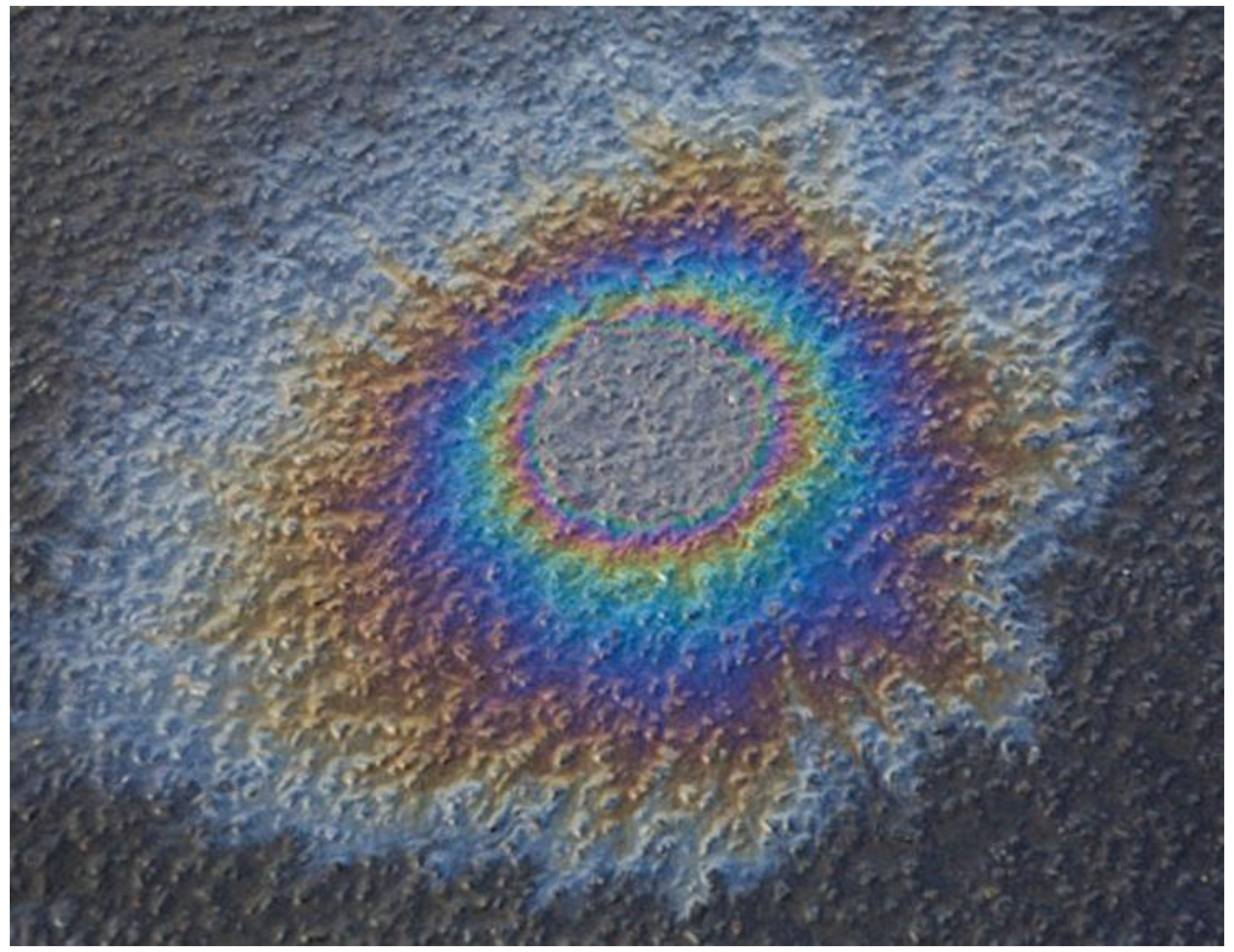

STATUS OF GLOBAL DRIFTER ARRAY
September 9, 2024
buoys= 1224
NOAA/AOML
80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
80°S
30°E
60°E
90°E
120°E
150°E
180°
150°W
120°W
90°W
60°W
30°W
0°
30°E
Deploying Country
Argentina (5)
Australia (37)
Barbados (1)
Brazil (2)
Canada (25)
China (1)
France (168)
Germany (60)
Iceland (3)
India (1)
Italy (4)
Japan (26)
Korea, Rep. of (64)
New Zealand (48)
Netherlands (18)
Portugal (27)
Seychelles (3)
South Africa (47)
Spain (2)
UK (87)
USA (477)
Unknown (118)

Global Drifter Program

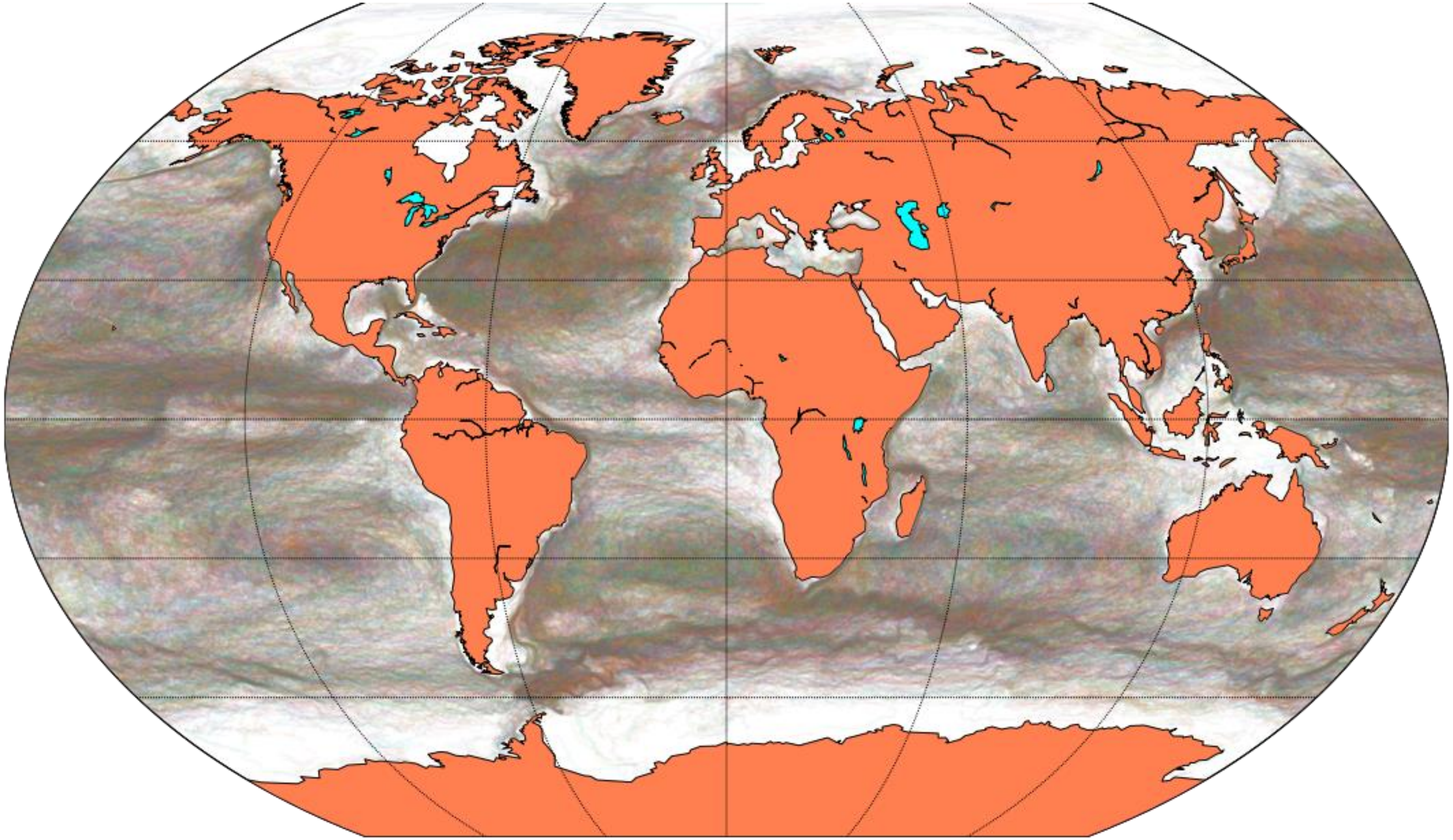


Wait…what is this good for?

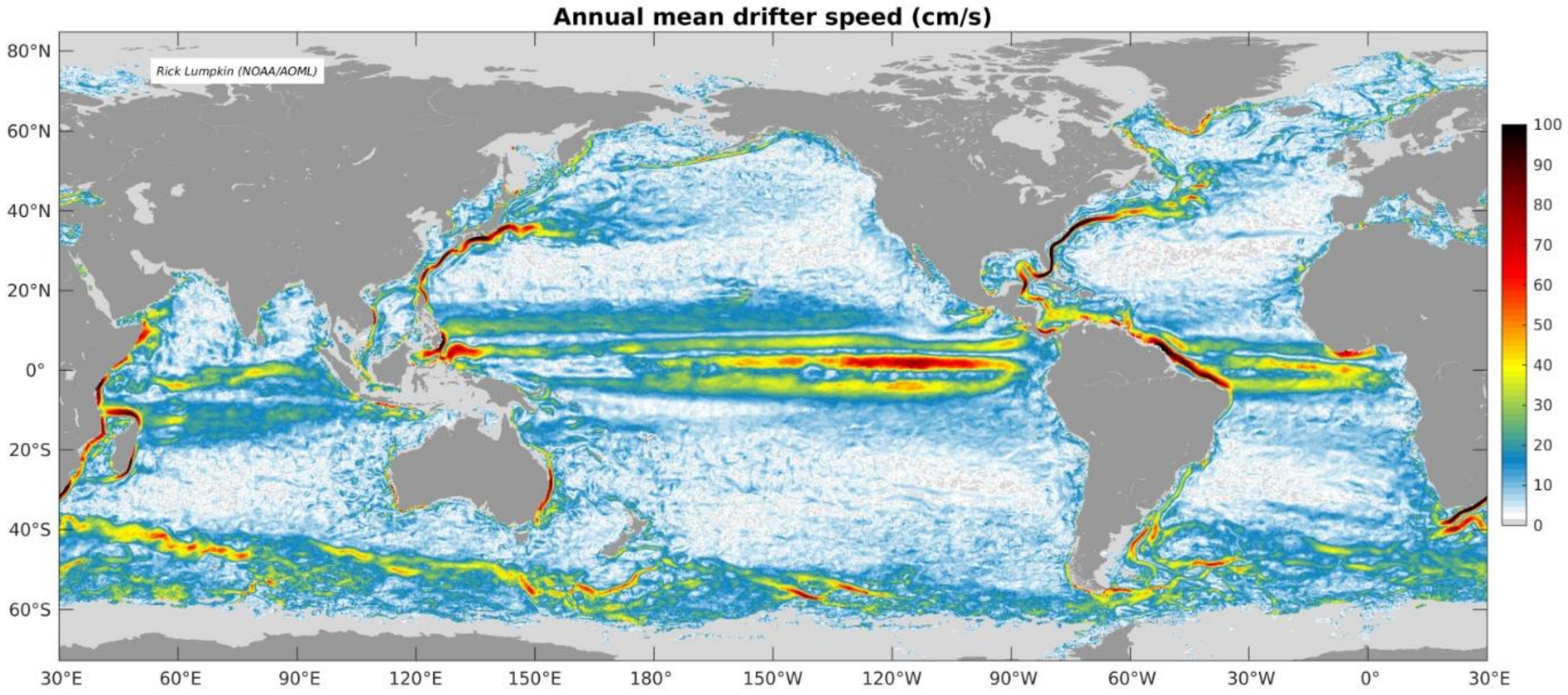




Spaghetti Plot – "data dump"

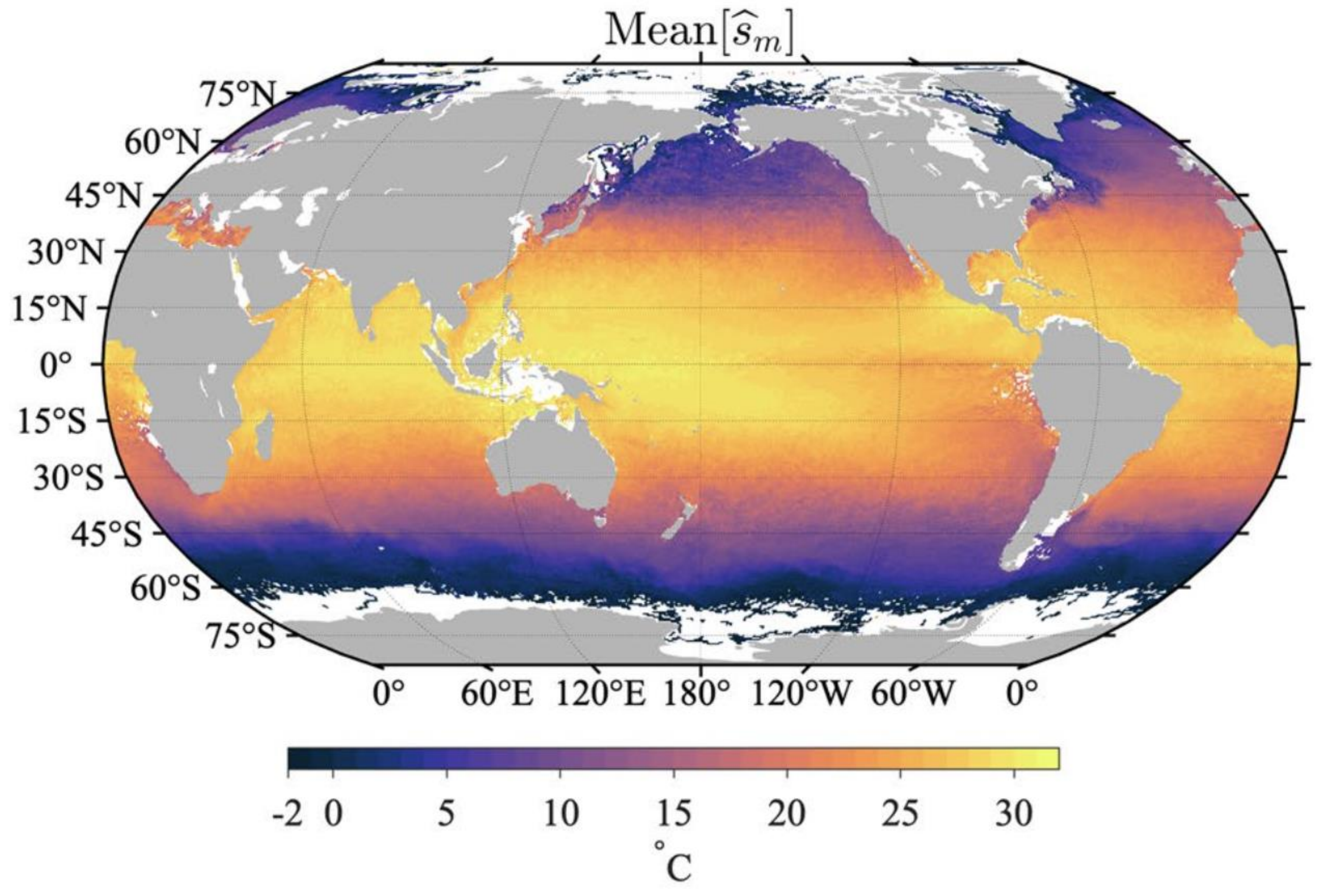


First step: Averaging the data

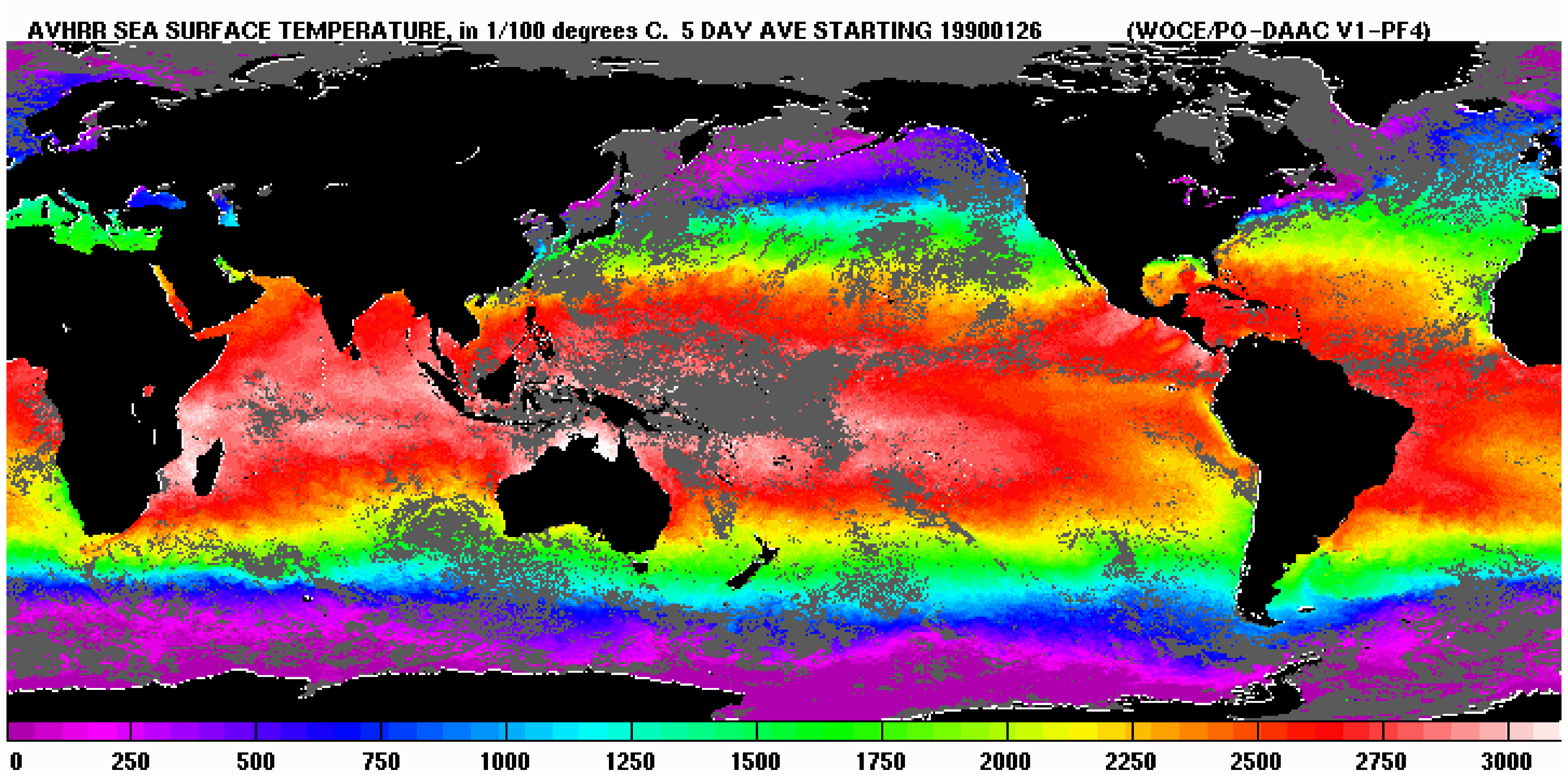


Activity 1: Drifters vs Satellites

Sea Surface Temperature



Drifter SST



Satellite SST

Download data and R scripts from:
github.com/AdamSykulski/OceanDrifterDataInR.git

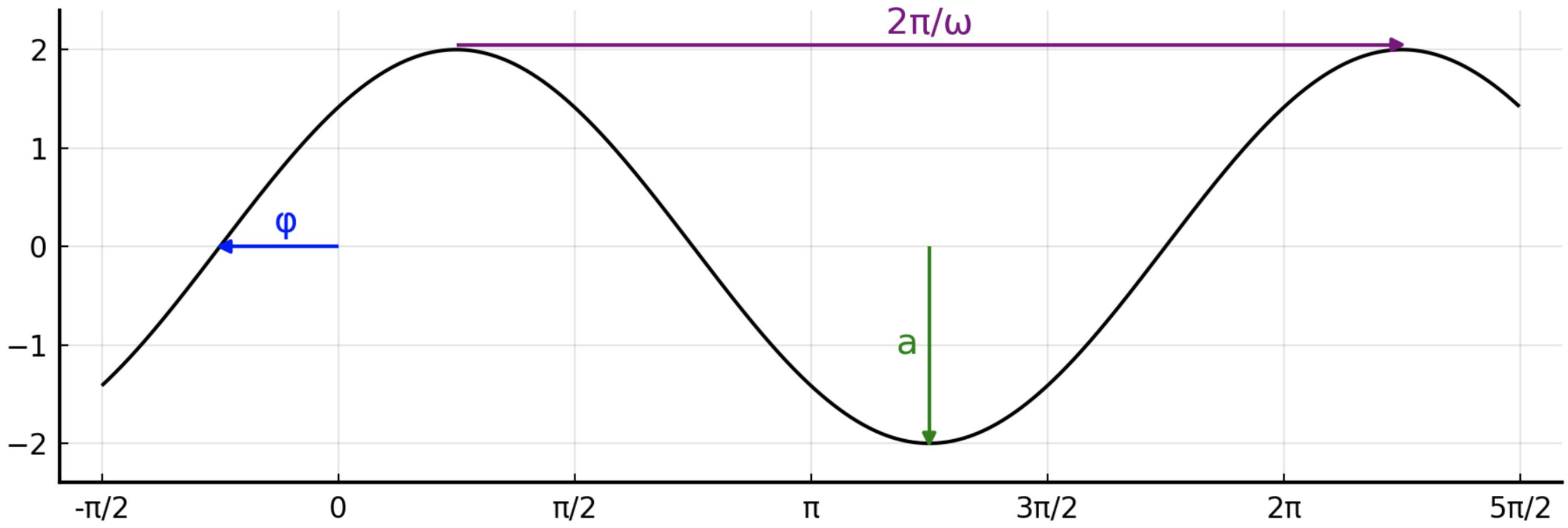
And then head over to RStudio for Activity 1!

Activity 2: Spectral Analysis of Time Series

Consider a basic sinusoid:

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

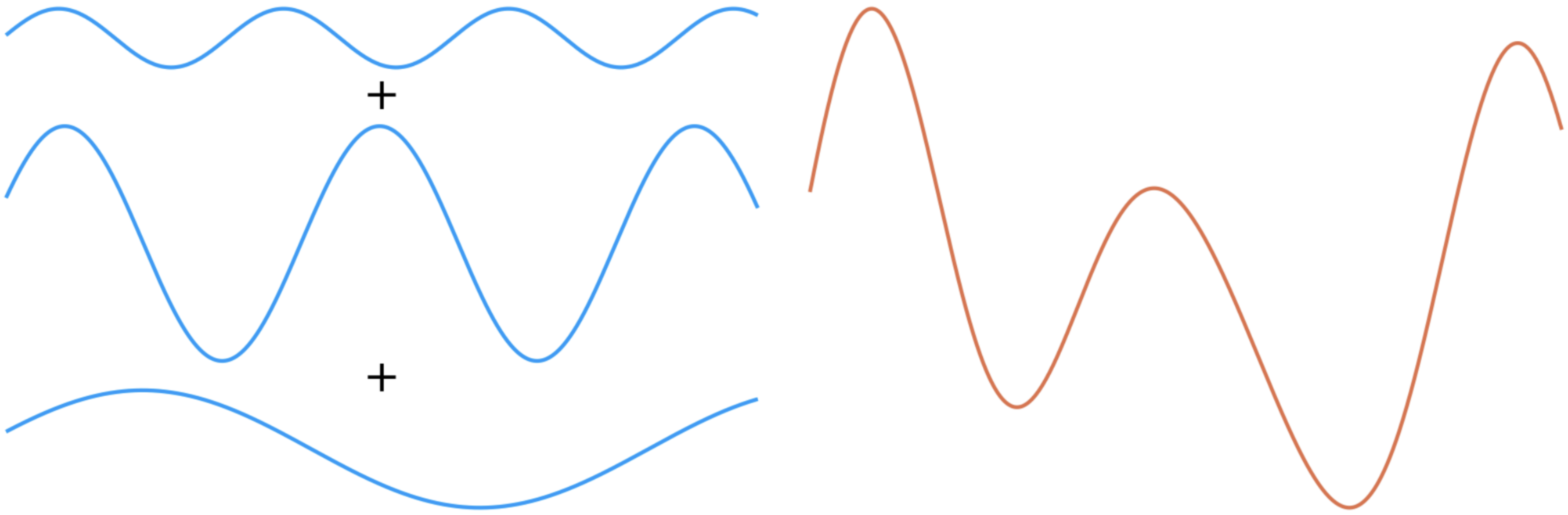
- a is the *amplitude*.
- φ is the *phase*.
- ω is the *angular frequency*.



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

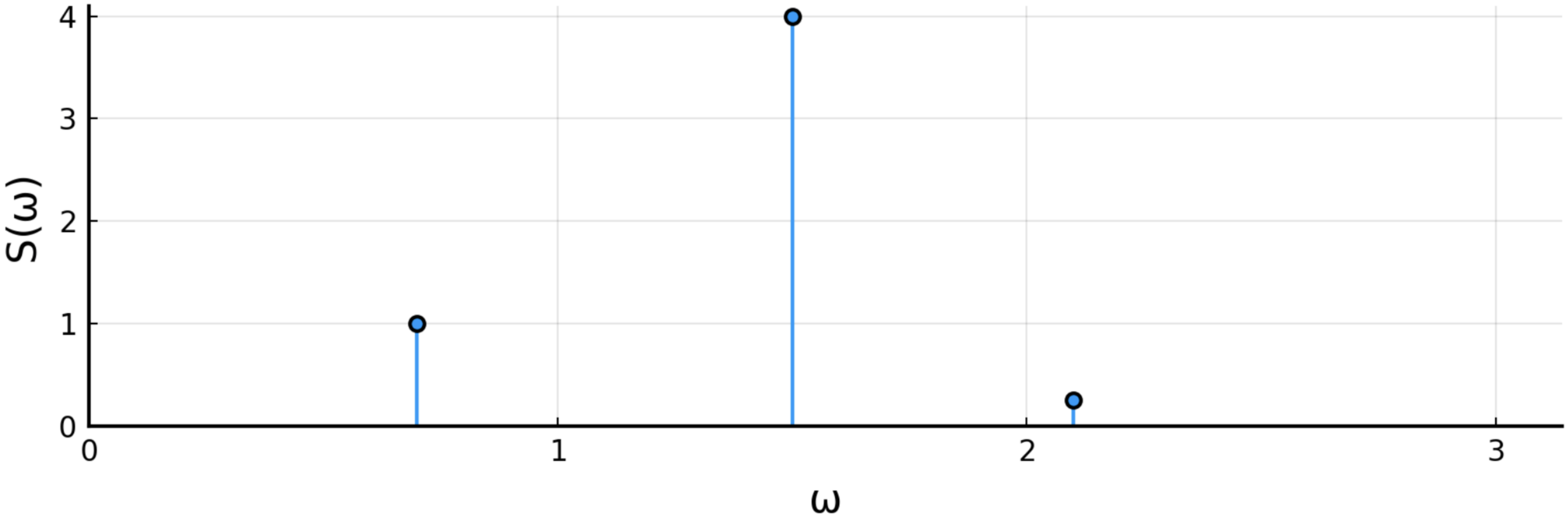


The power spectrum: discrete frequency

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 the power spectral density

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

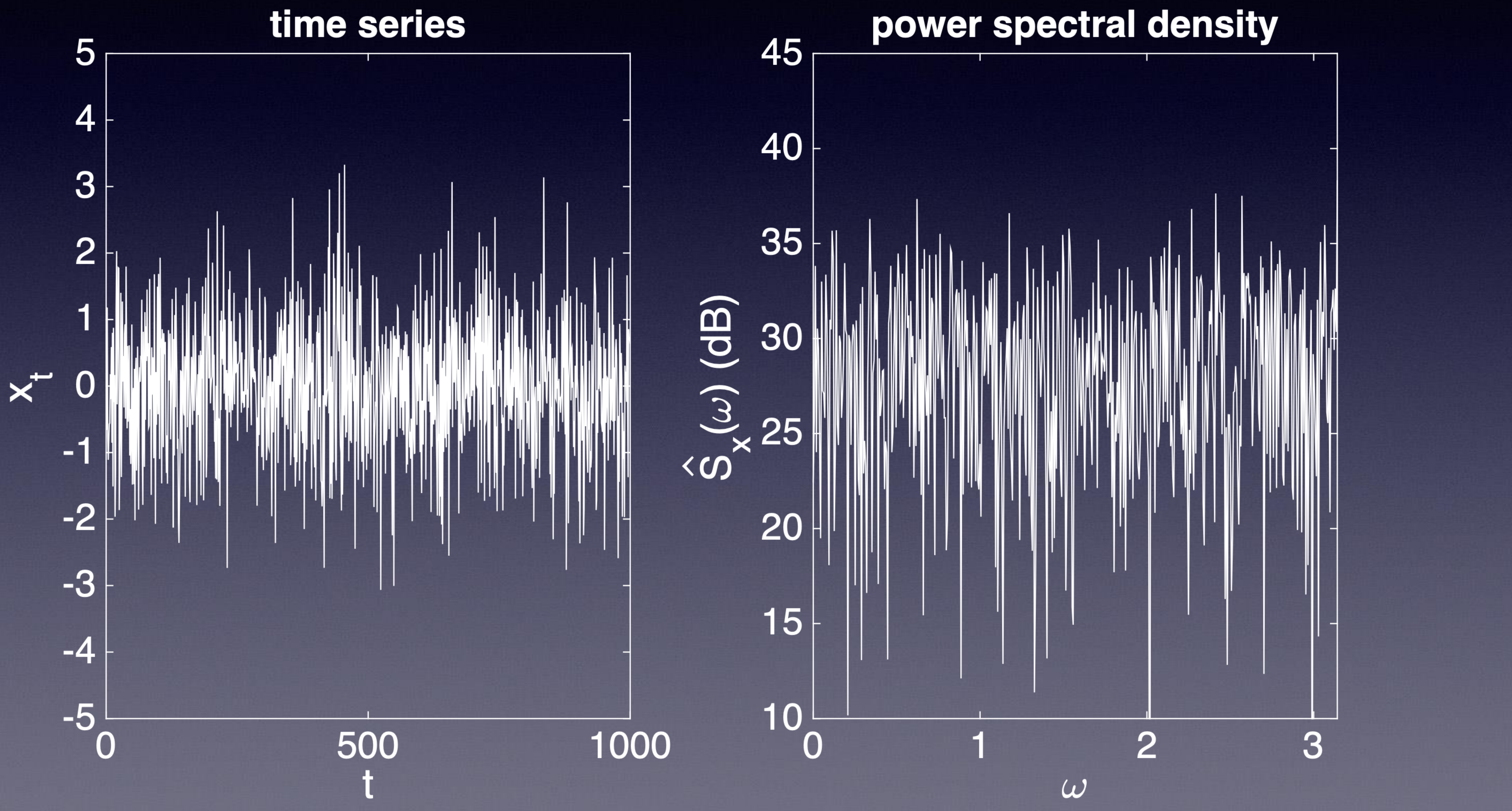
Practice: Observe some sample $X_1, \ldots, X_N$ at intervals Δ such that

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

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

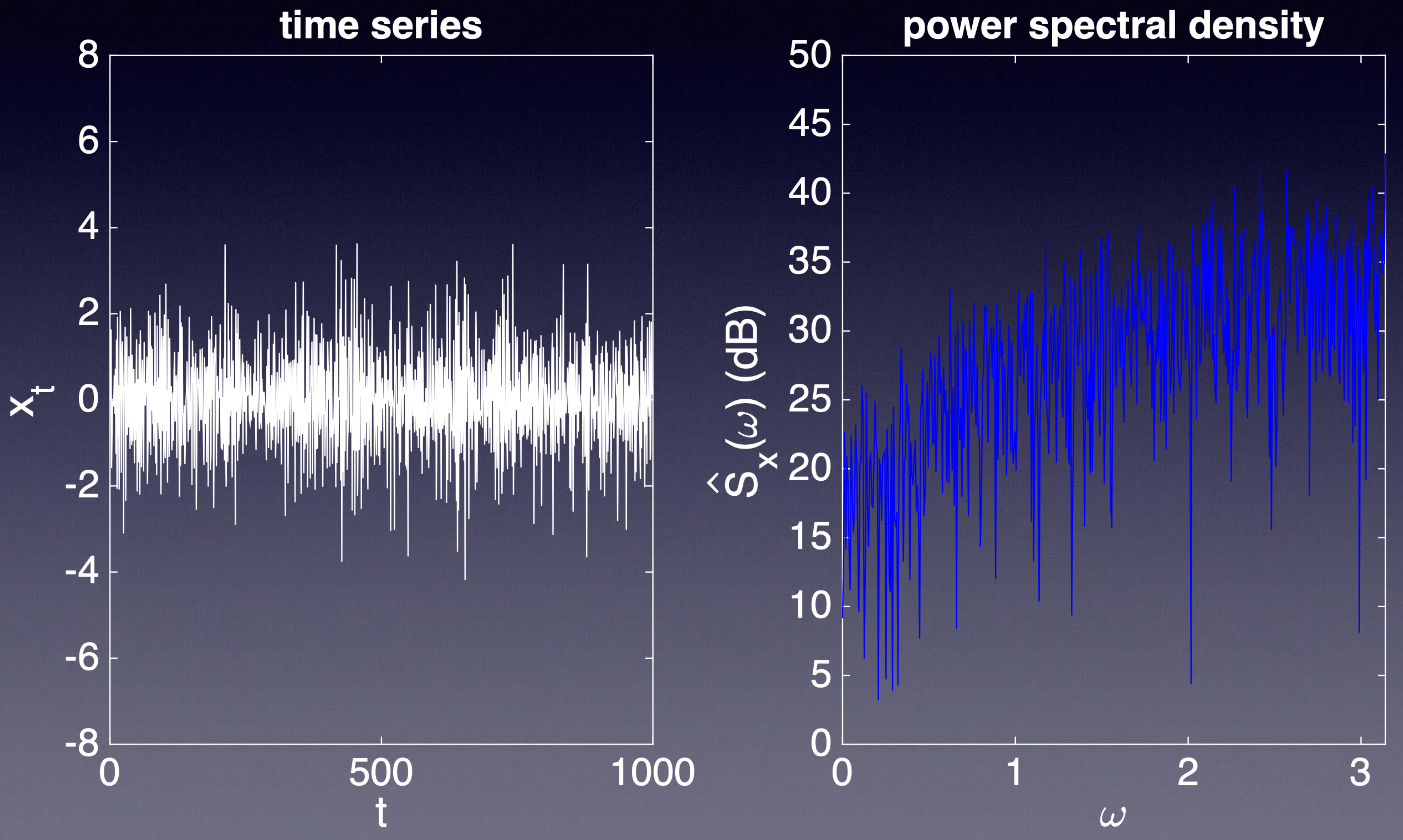
White noise process

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



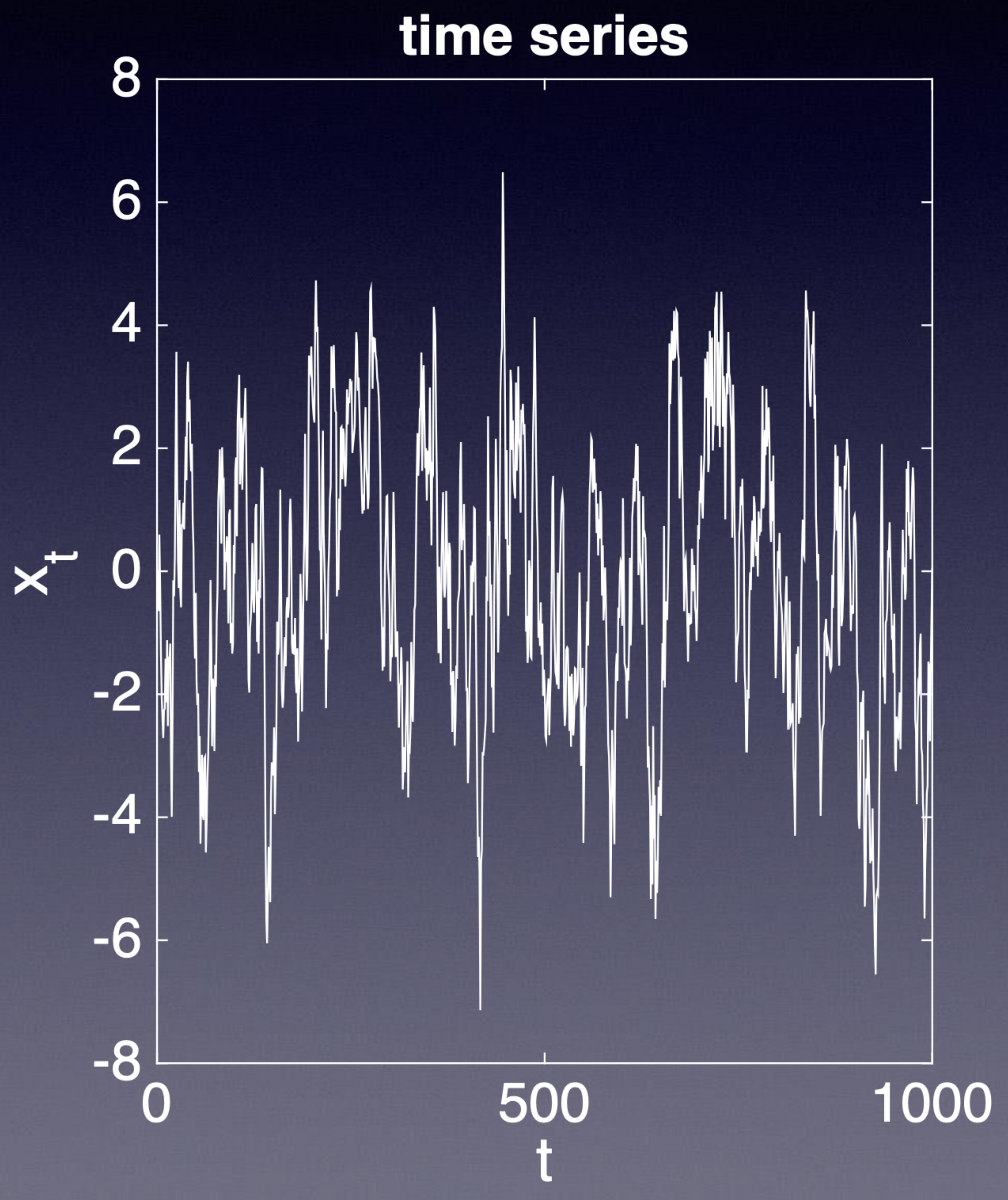
Moving average process: MA(1)

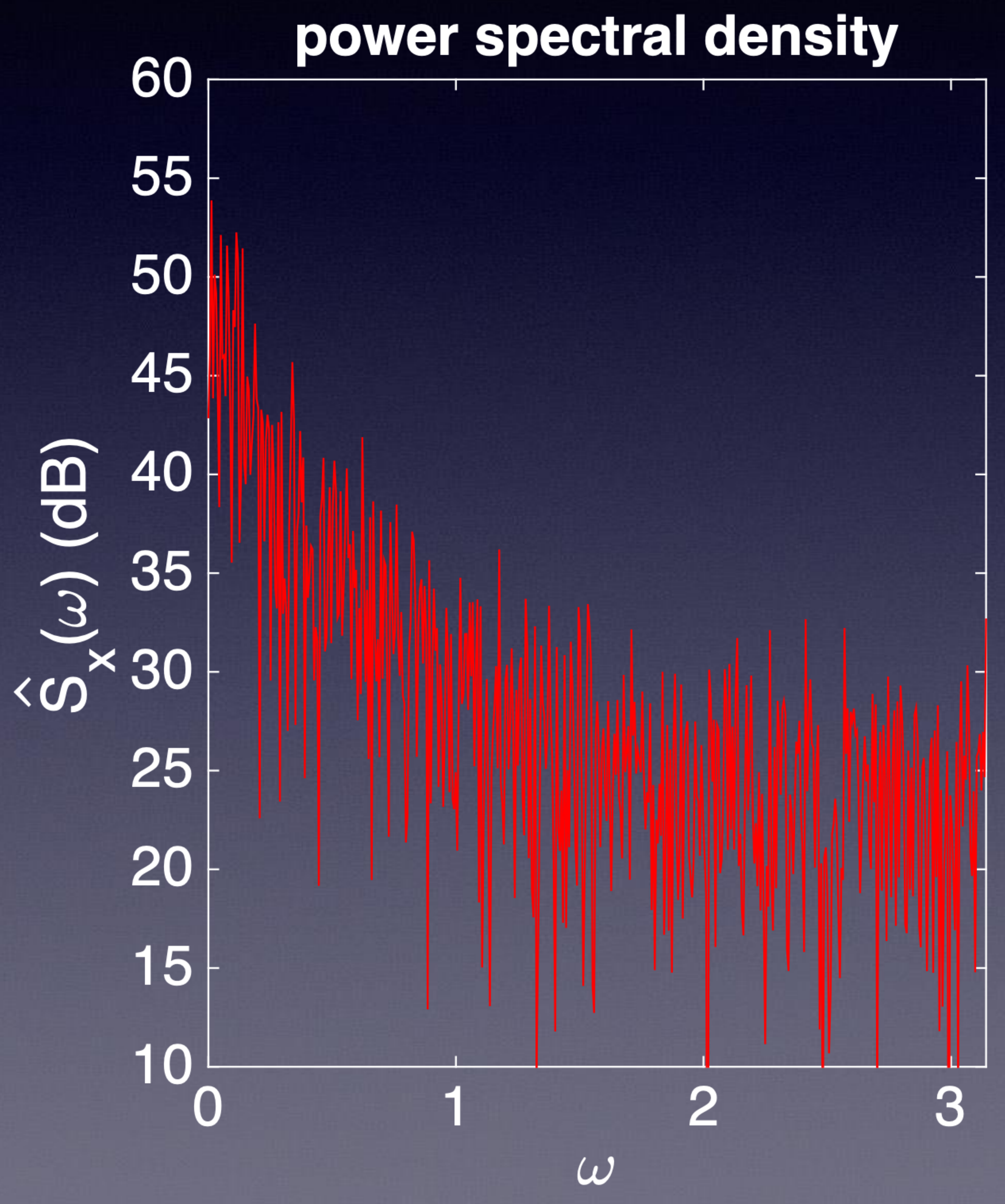
$$x_t = \varepsilon_t - 0.7\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{N}(0,1)$$



Auto-regressive process: AR(1)

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

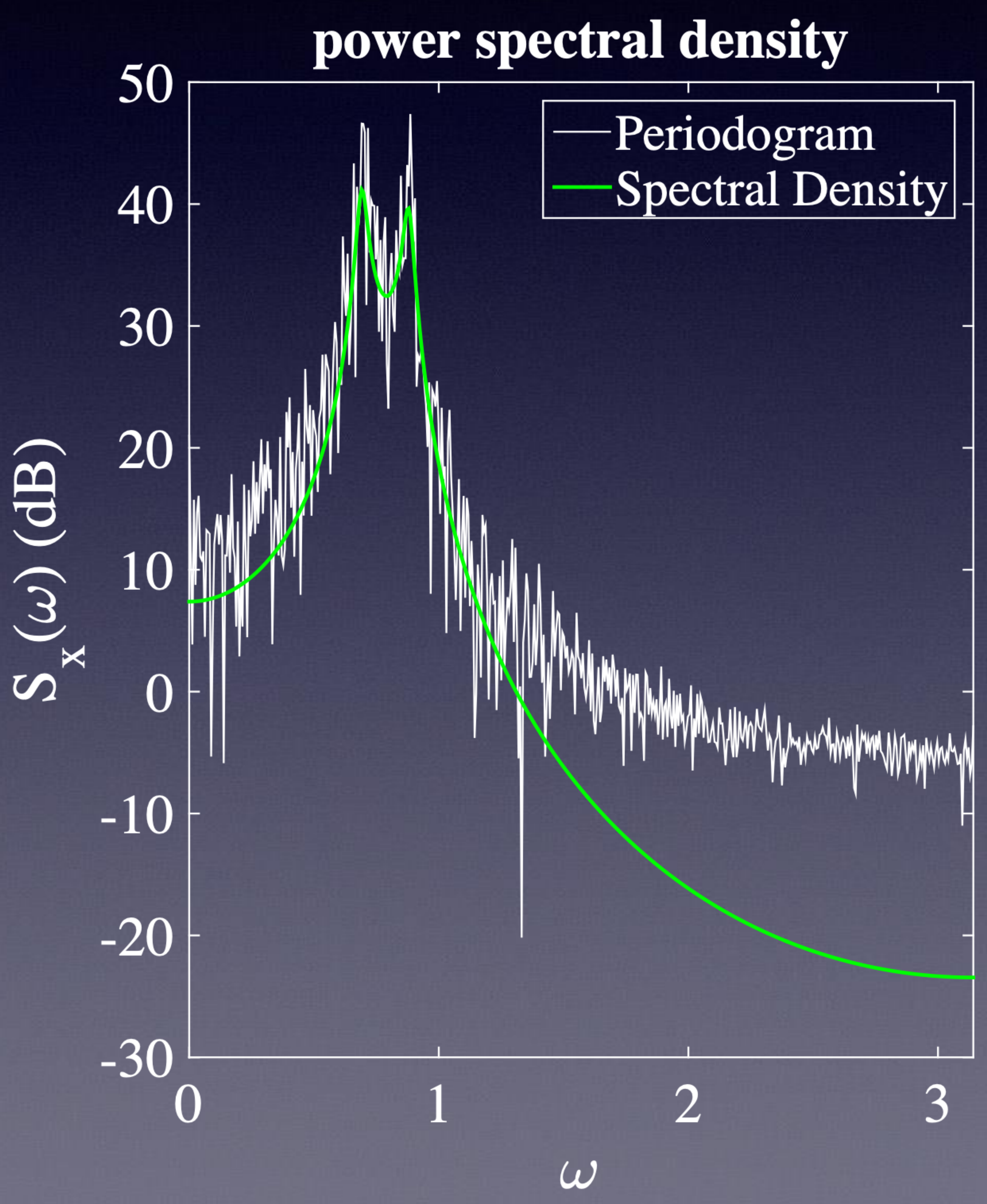




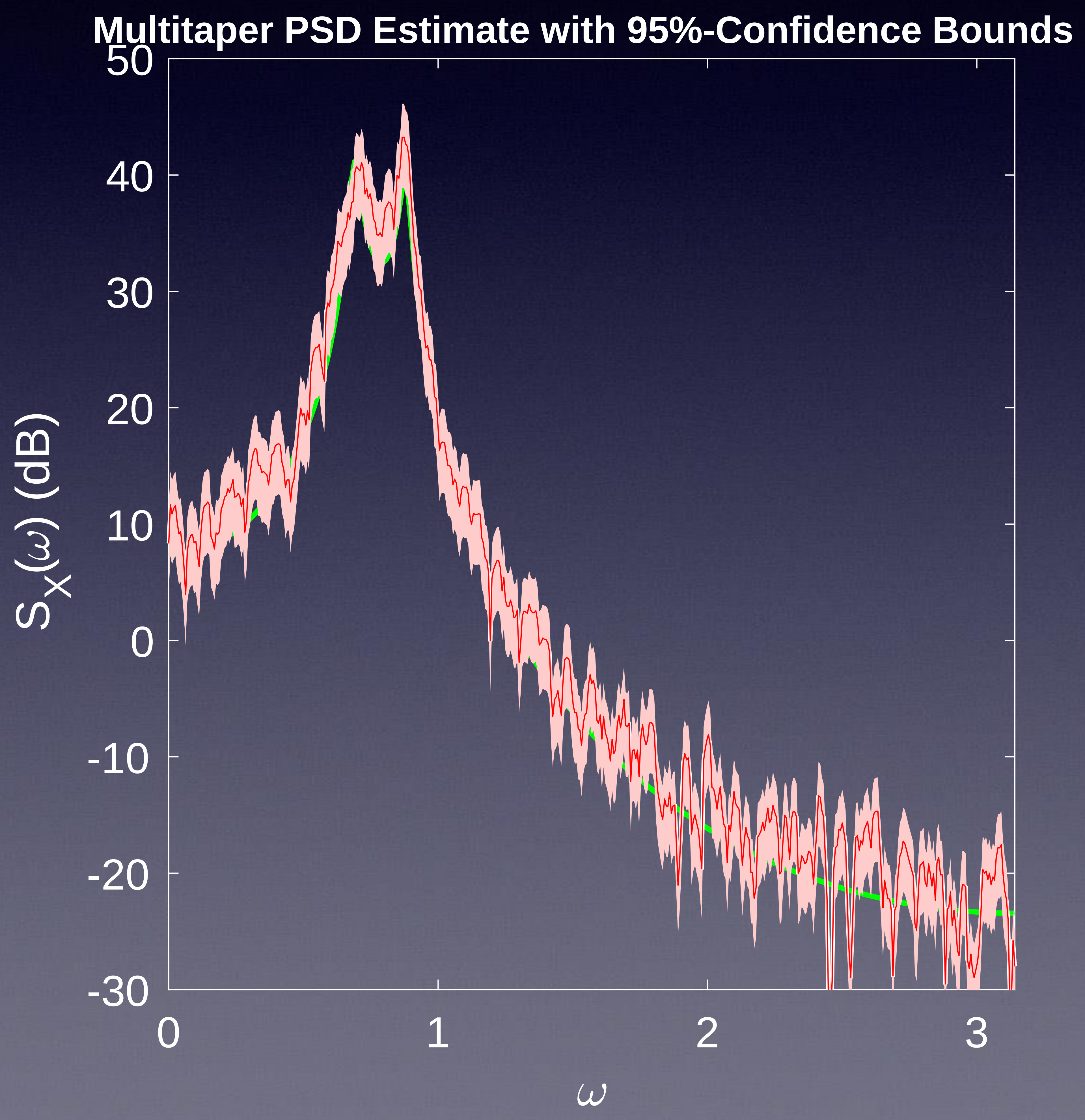
AR(4)

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

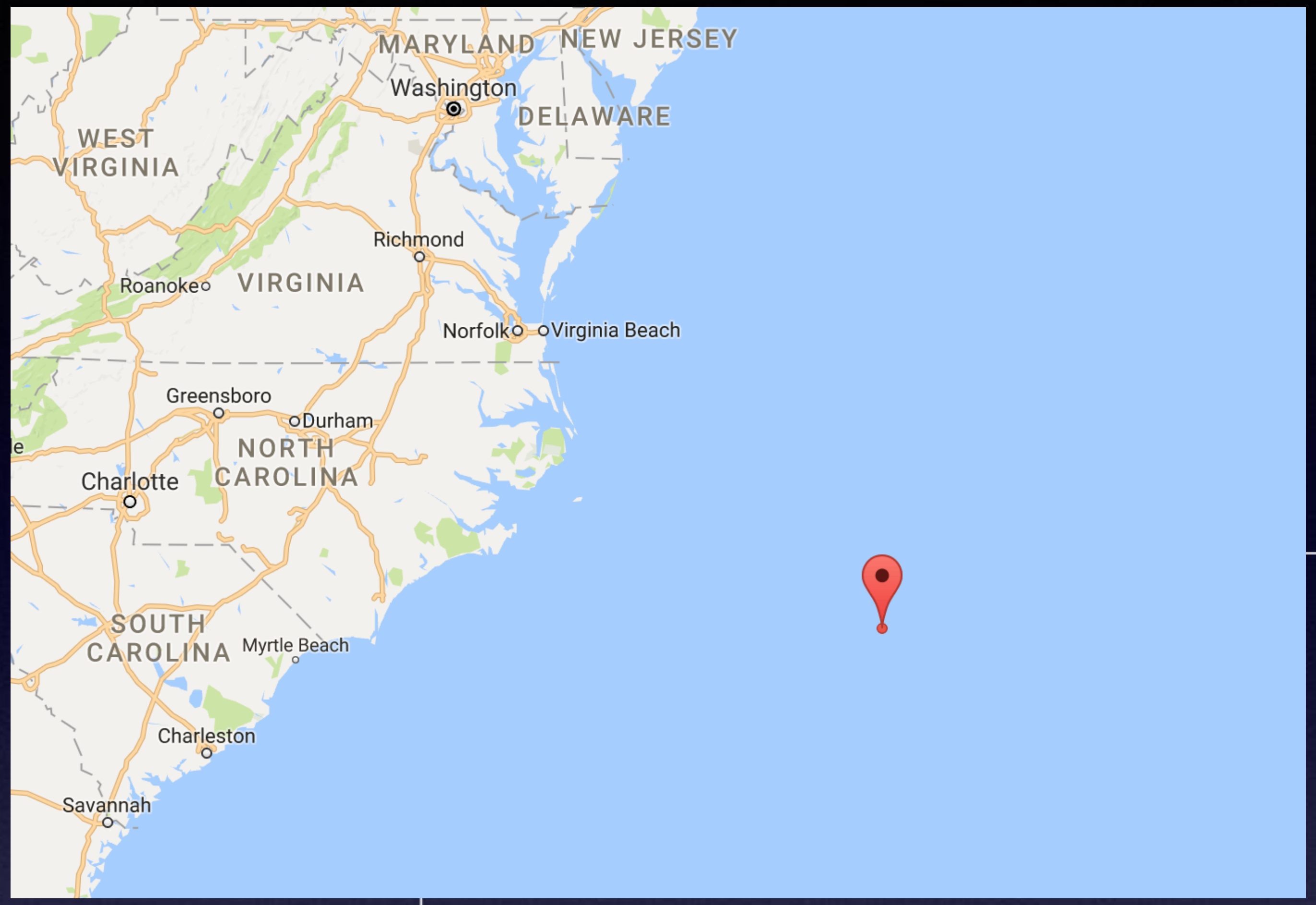
$$S_x(\omega) = \frac{\sigma_\epsilon^2}{|1 - \sum_{k=1}^{4} \phi_k e^{-ikw}|^2}$$

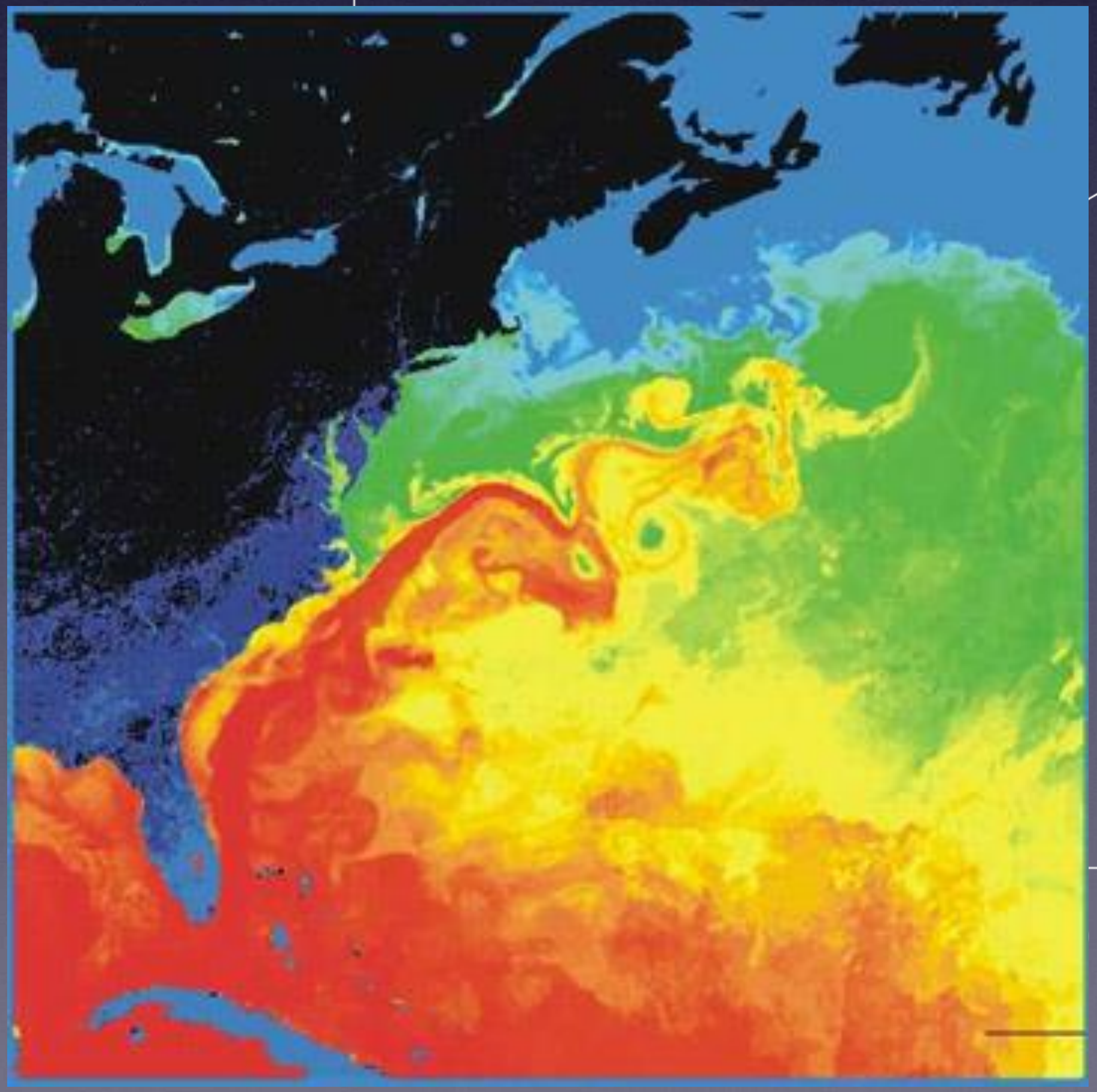


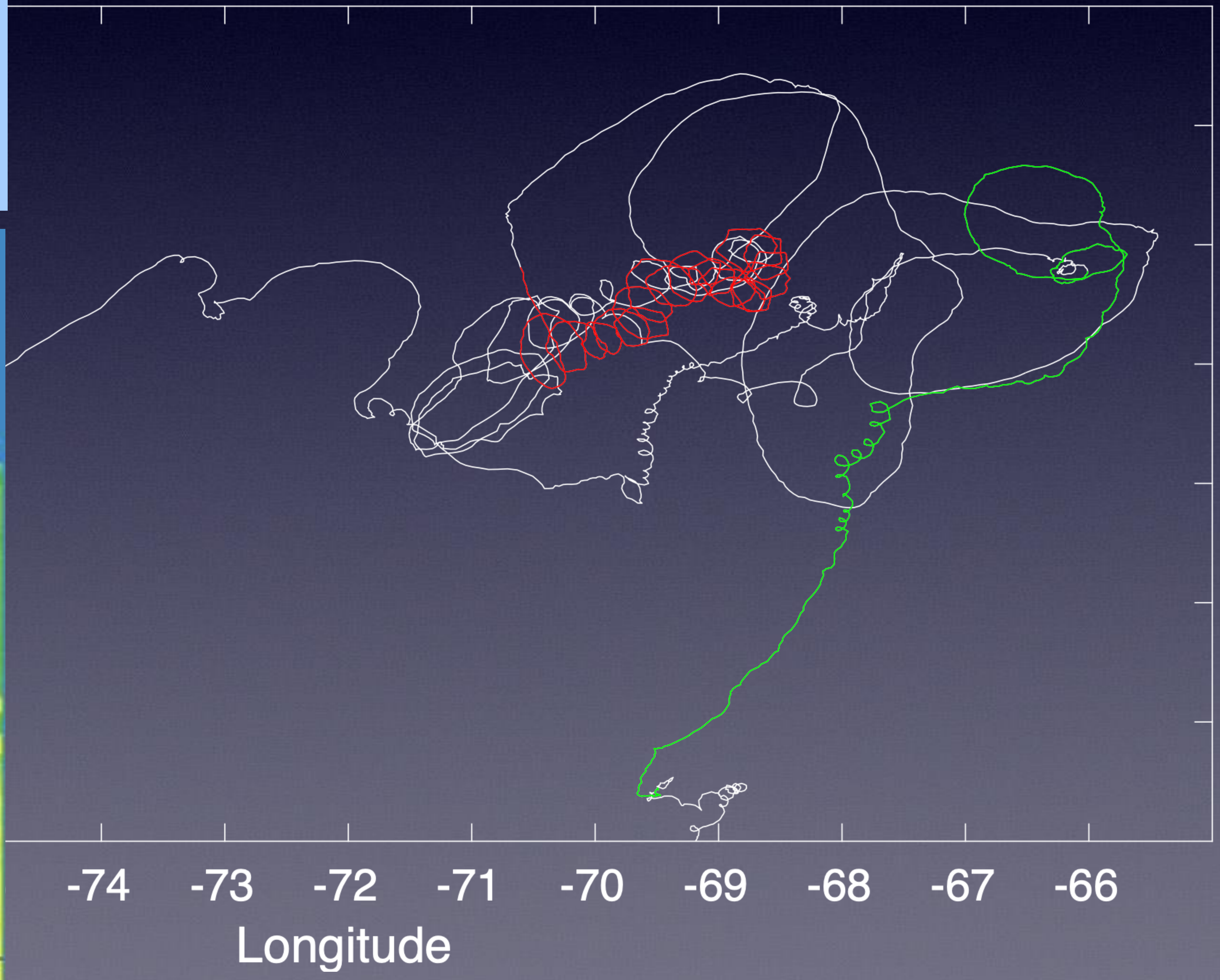
tapering



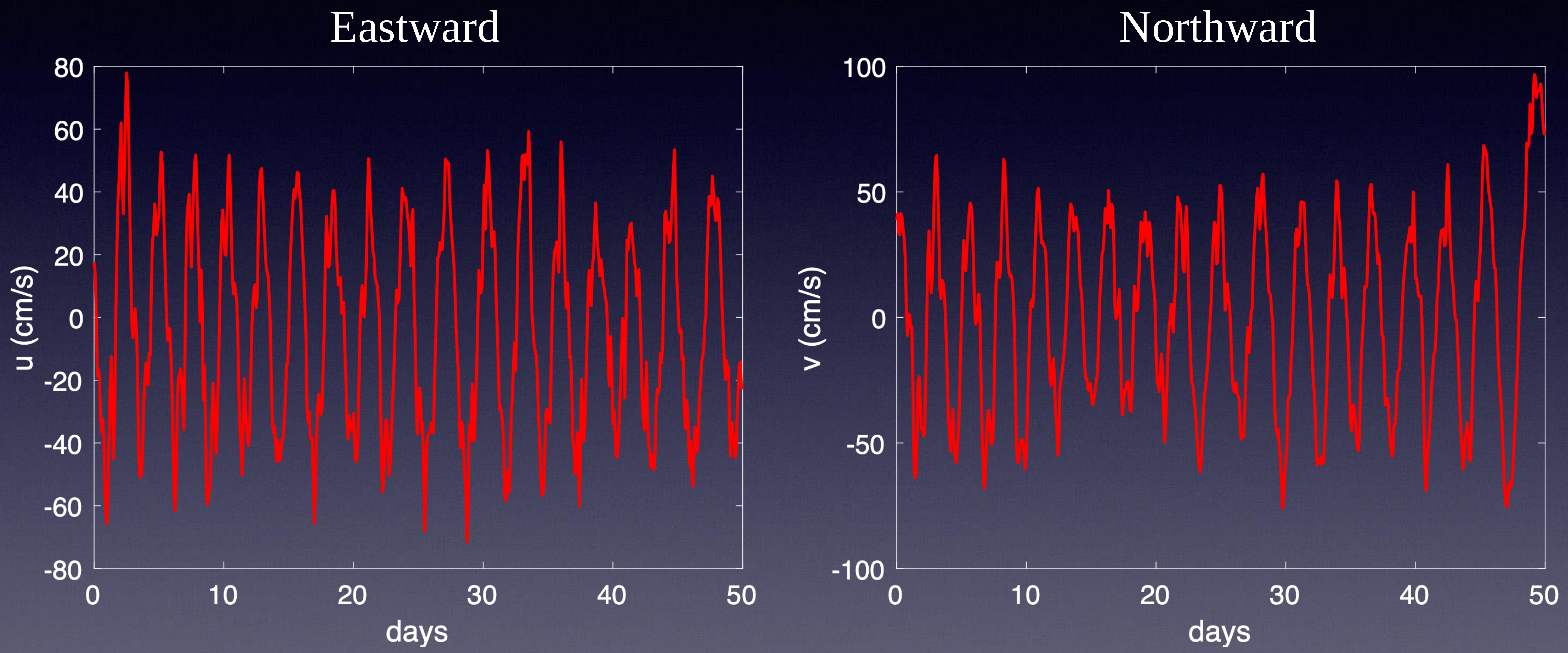
Oceanographic Drifter Data







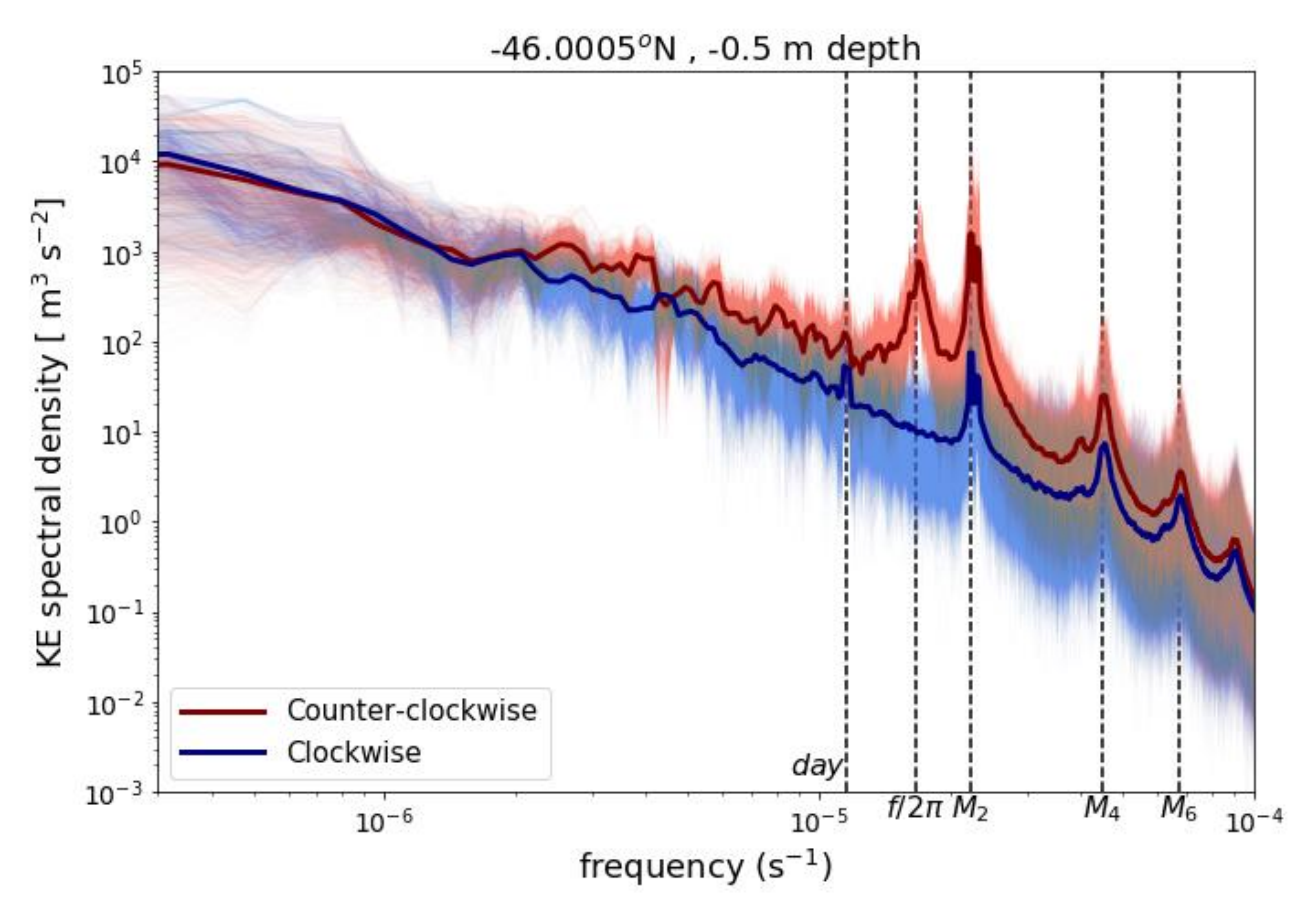
Drifter velocities for:



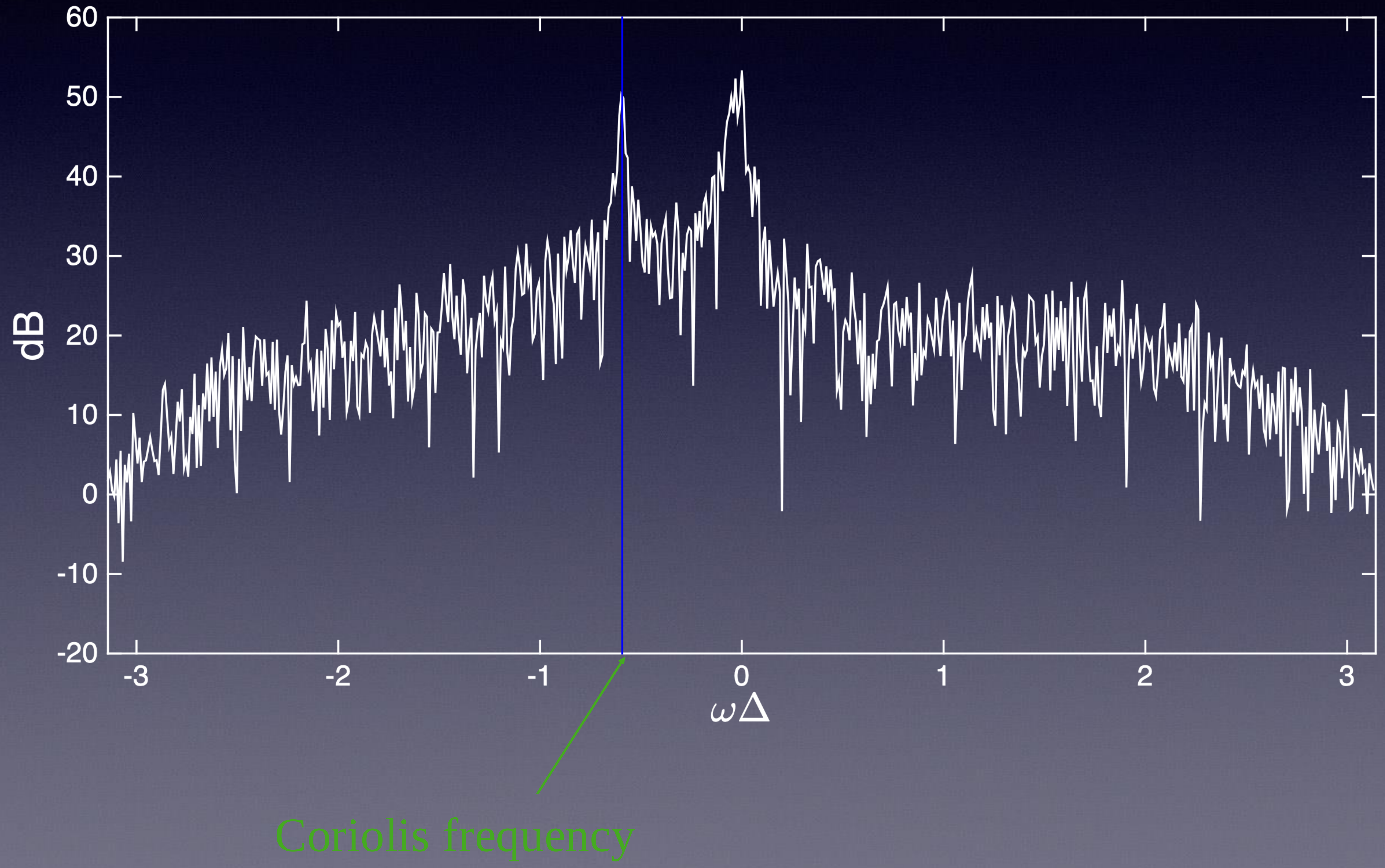
Represent particle velocities as complex-valued time series:

$$z_t = u_t + iv_t$$

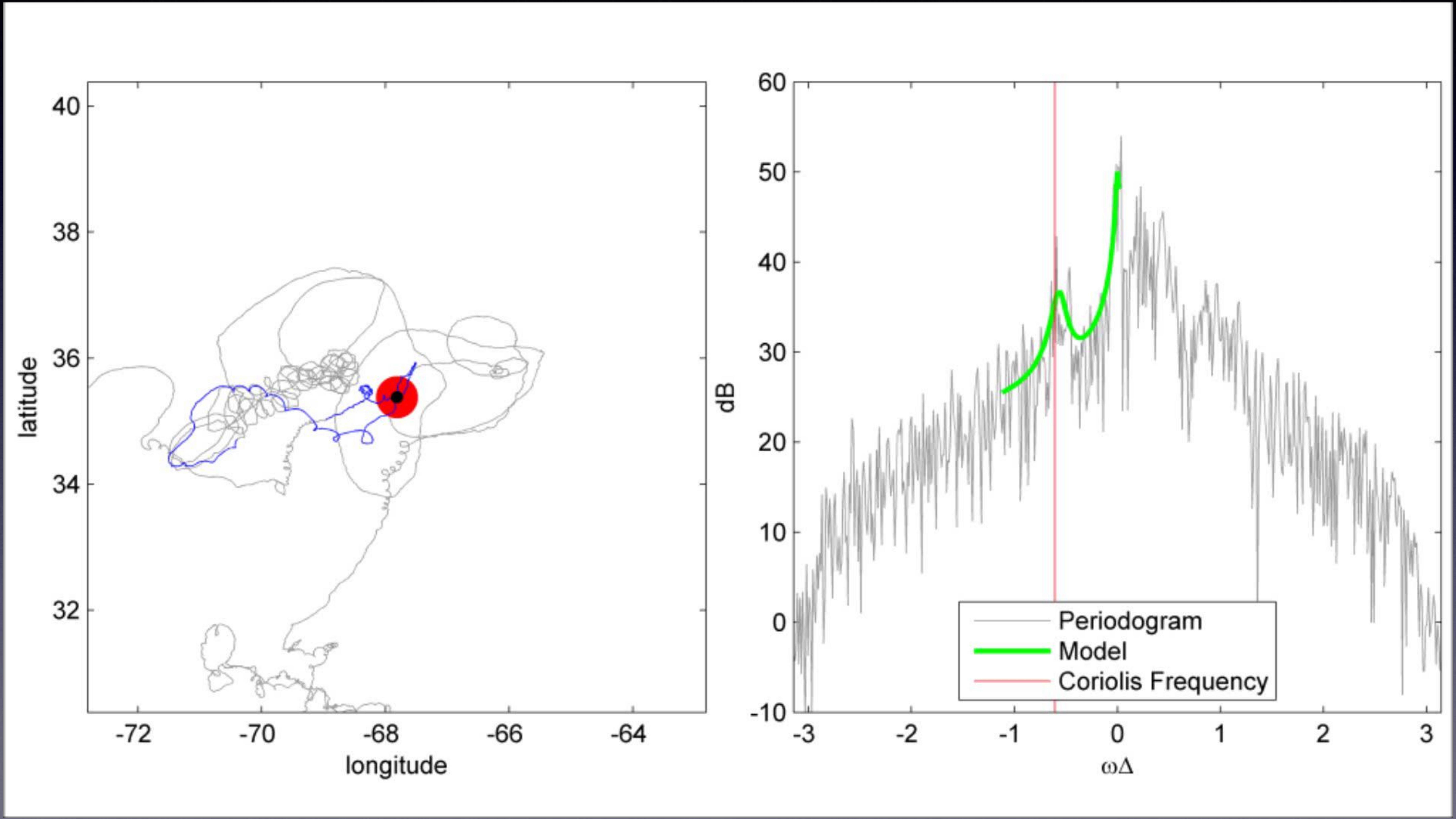
Power spectral density for velocities from:



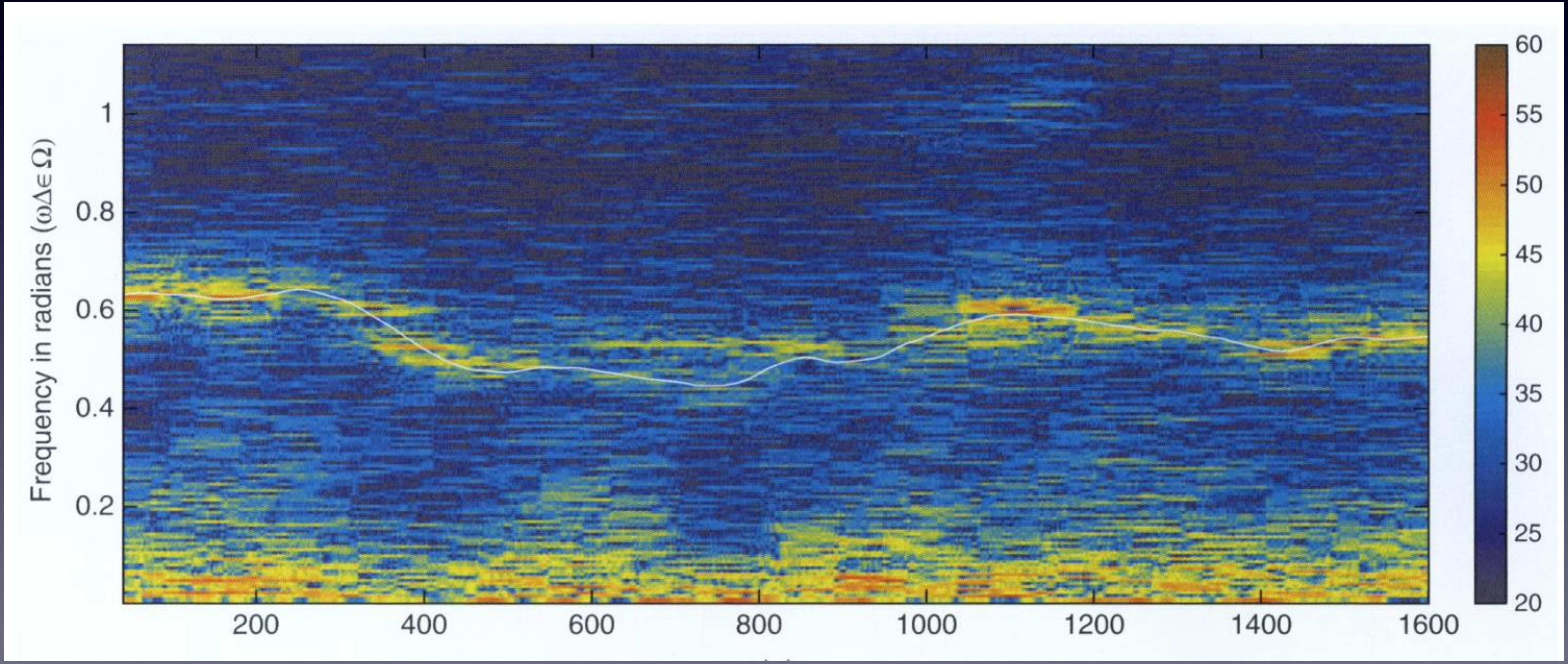
Power spectral density for velocities from:



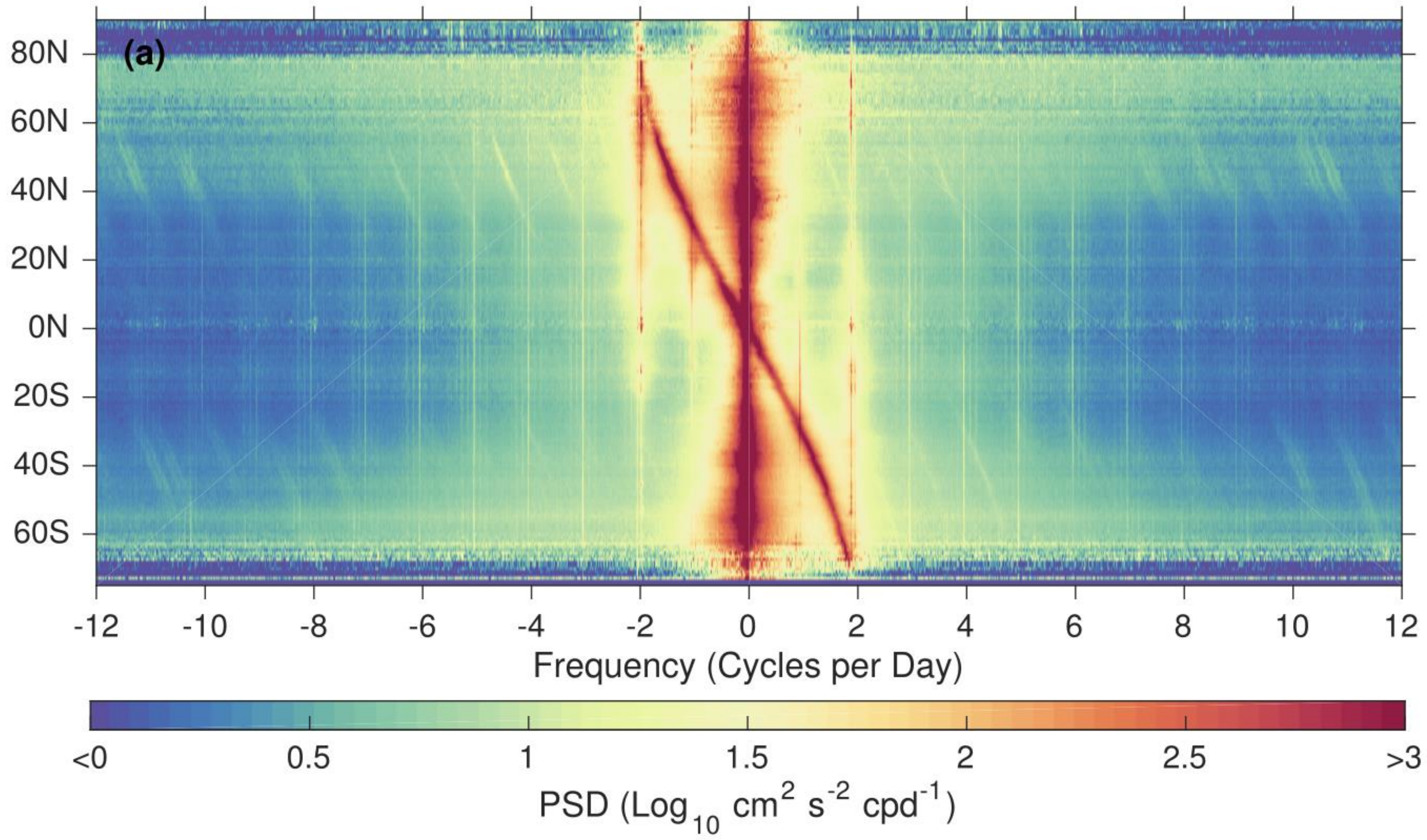
Locally stationary modelling

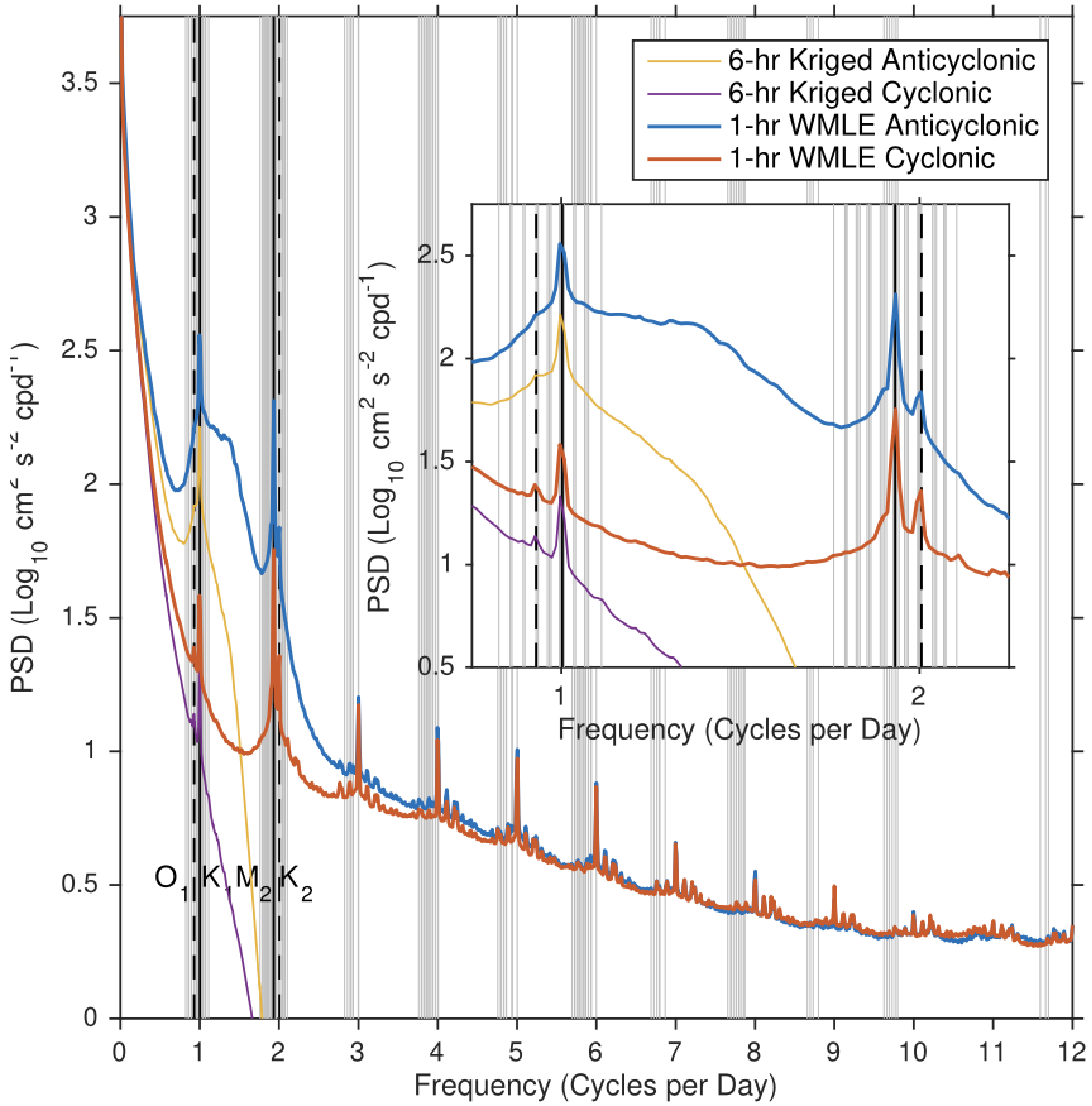


Nonstationarity: Time-frequency "spectrograms" and Heisenberg-Gabor uncertainty



Peaks in the power spectral density…





Head over to RStudio for Activity 2!