Debiasing Welch’s Method of Spectral Density Estimation

Lachlan Astfalck
School of Physics, Mathematics and Computing & Oceans Graduate School
The University of Western Australia

With contributions from Adam Sykulski, Ed Cripps, Andrew Zulberti, Aurelien Ponte, Michael Cutler and Paul Branson

TIDE

a.k.a. ARC ITRH for Transforming energy Infrastructure through Digital Engineering

Physical Oceanography

Hydrodynamics of Sea-surface Structures

Sea-bed Geotechnics

TIDE

a.k.a. ARC ITRH for Transforming energy Infrastructure through Digital Engineering

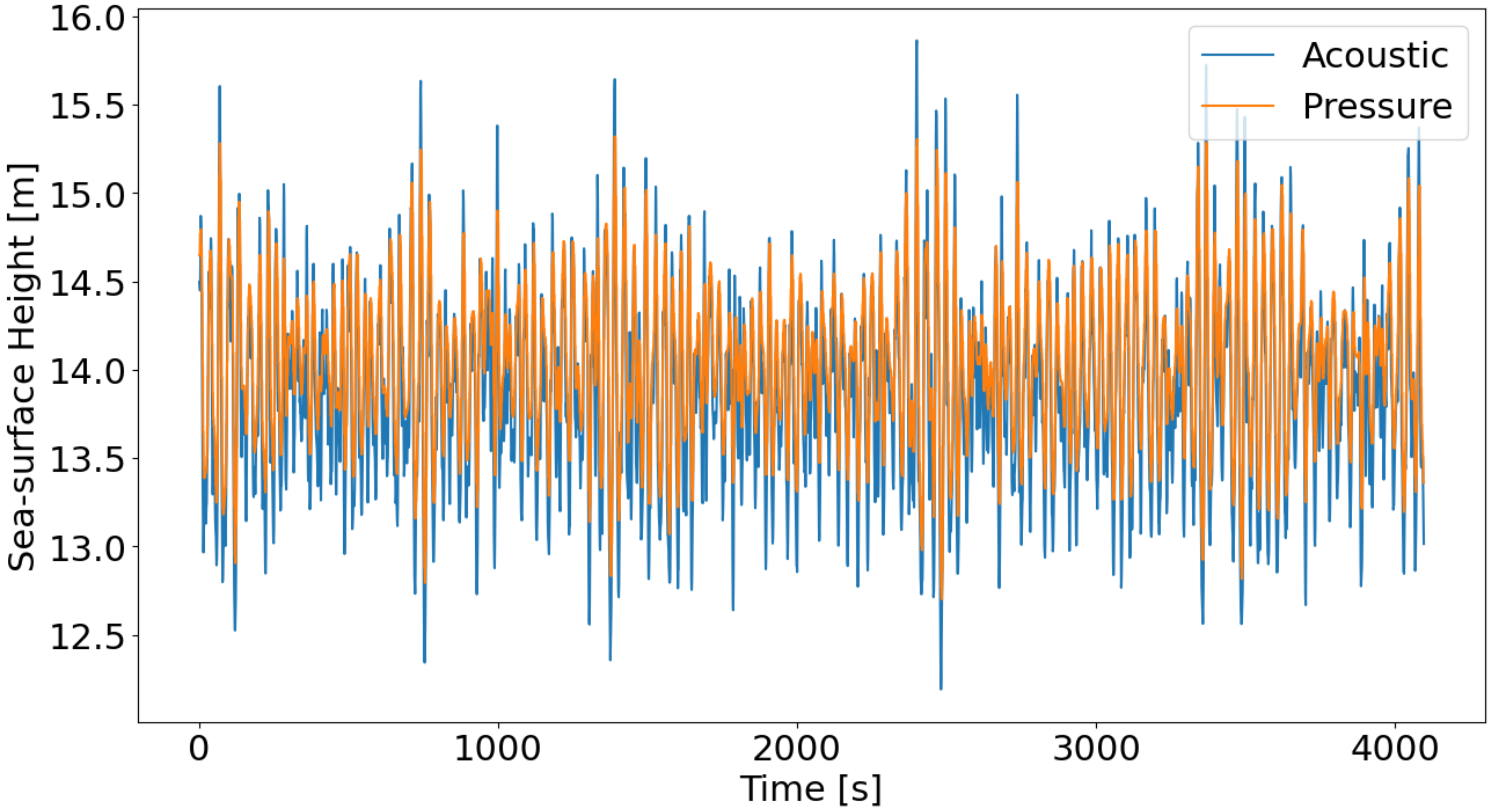
Physical Oceanography

Hydrodynamics of Sea-surface Structures

Sea-bed Geotechnics

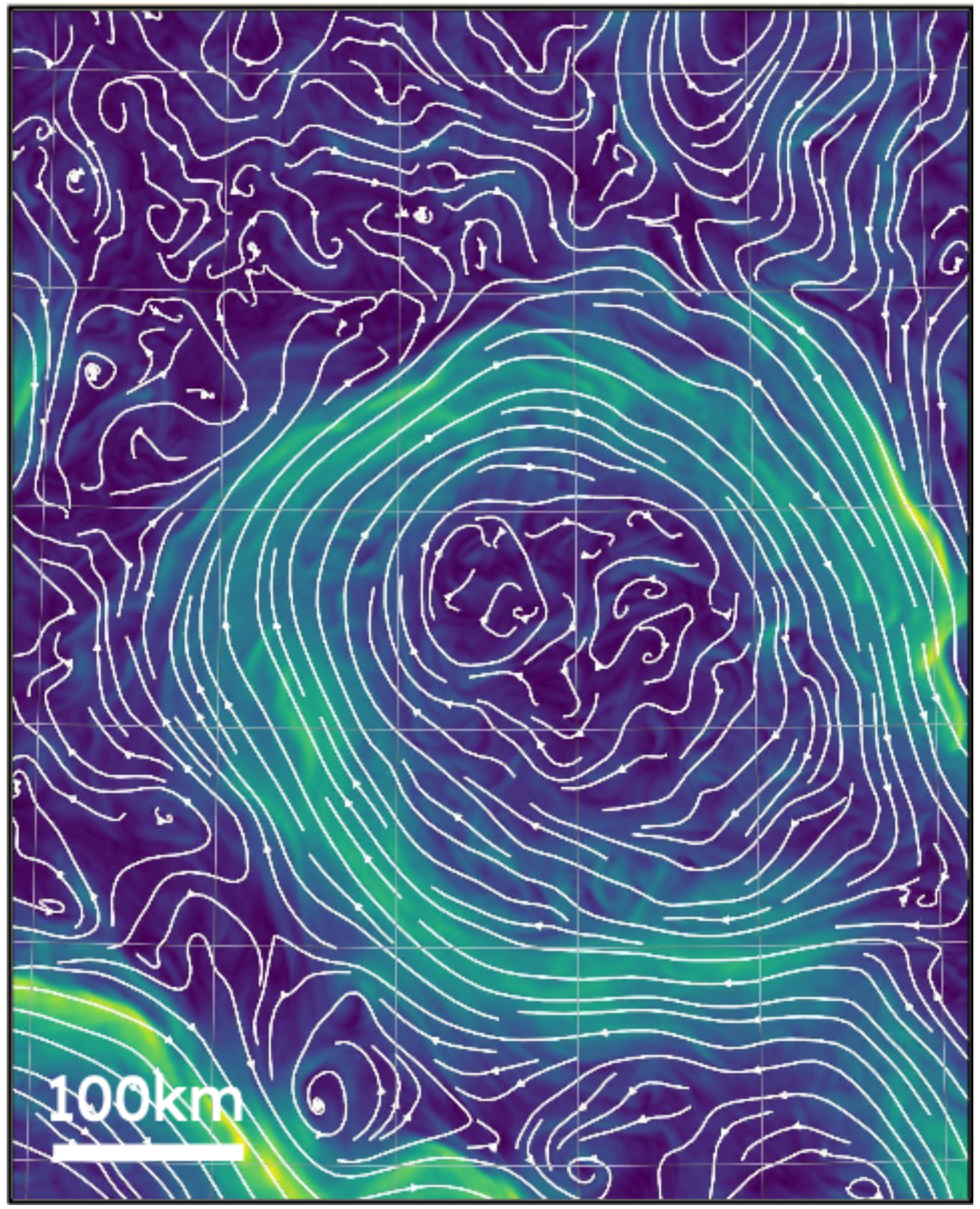
Data Science

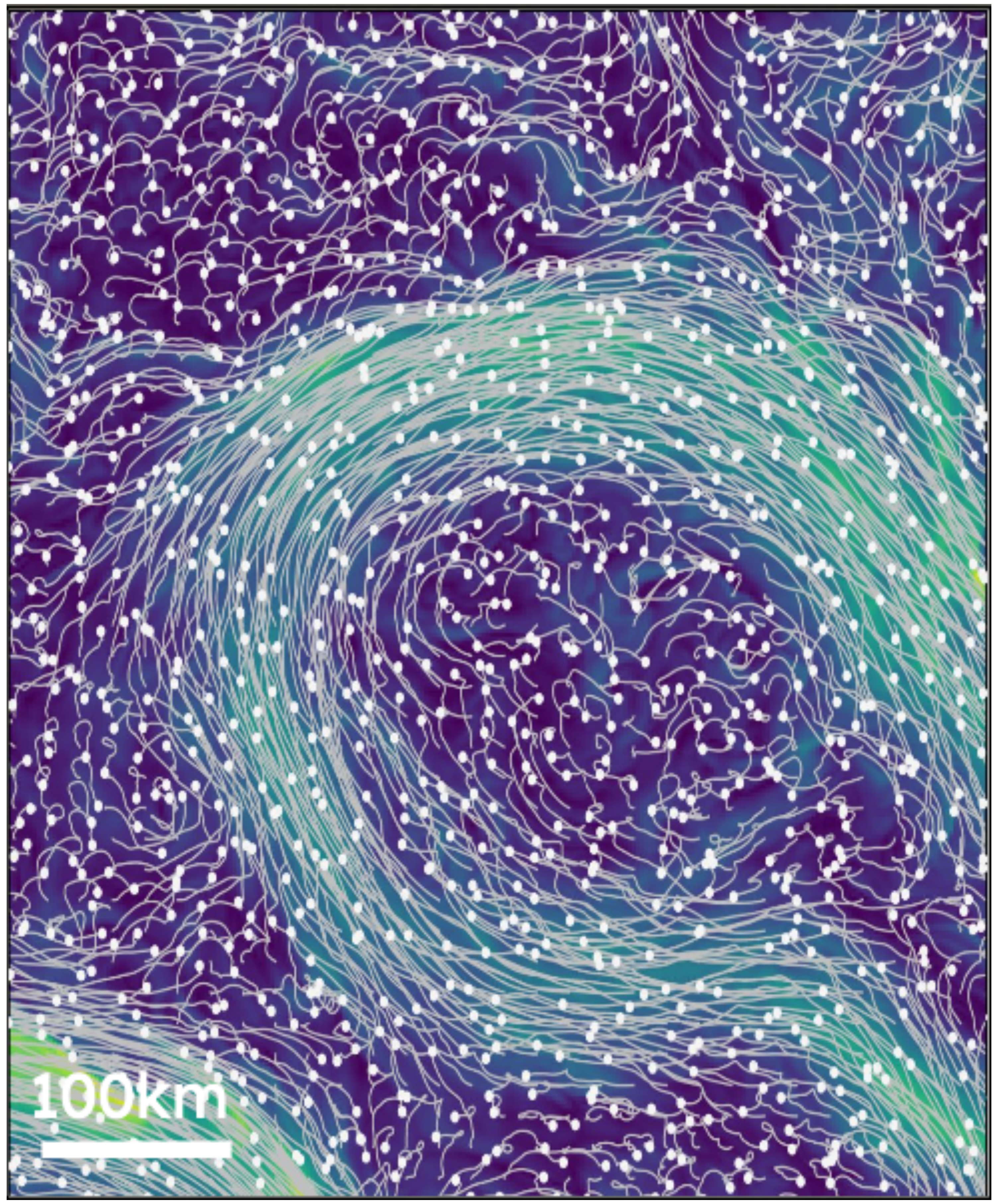
Coastal Wave Measurements



Ocean Beach sea-surface heights

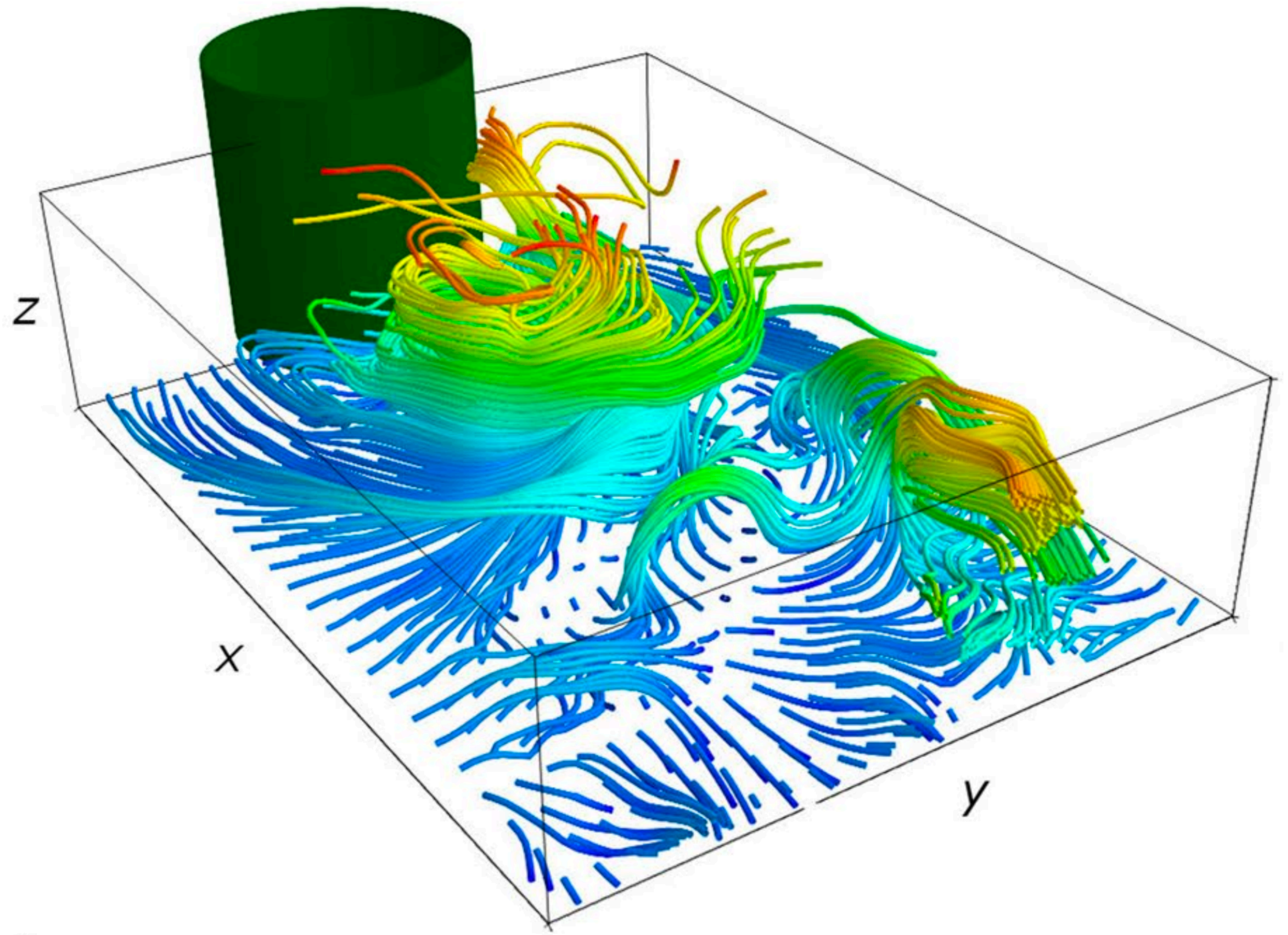
Complex-valued data





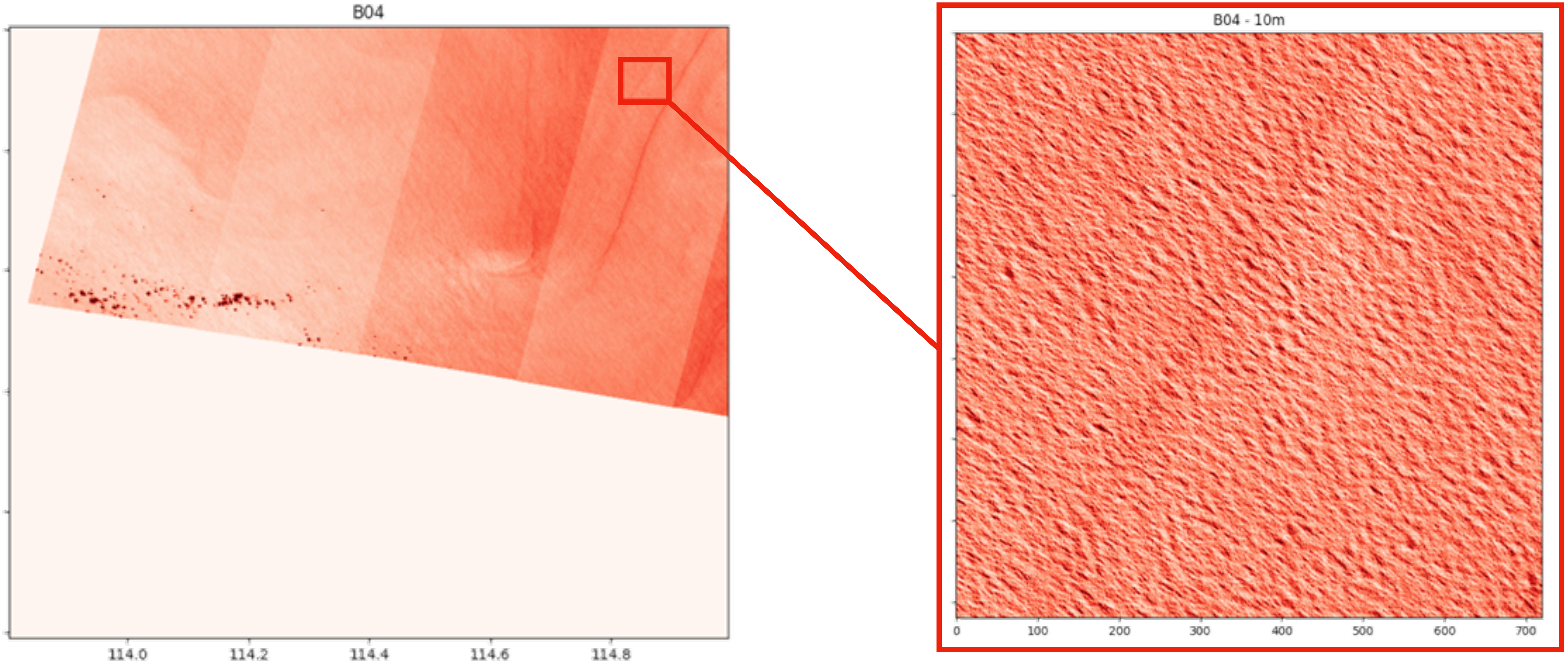
Model simulation of Lagrangian drifters

Multivariate data



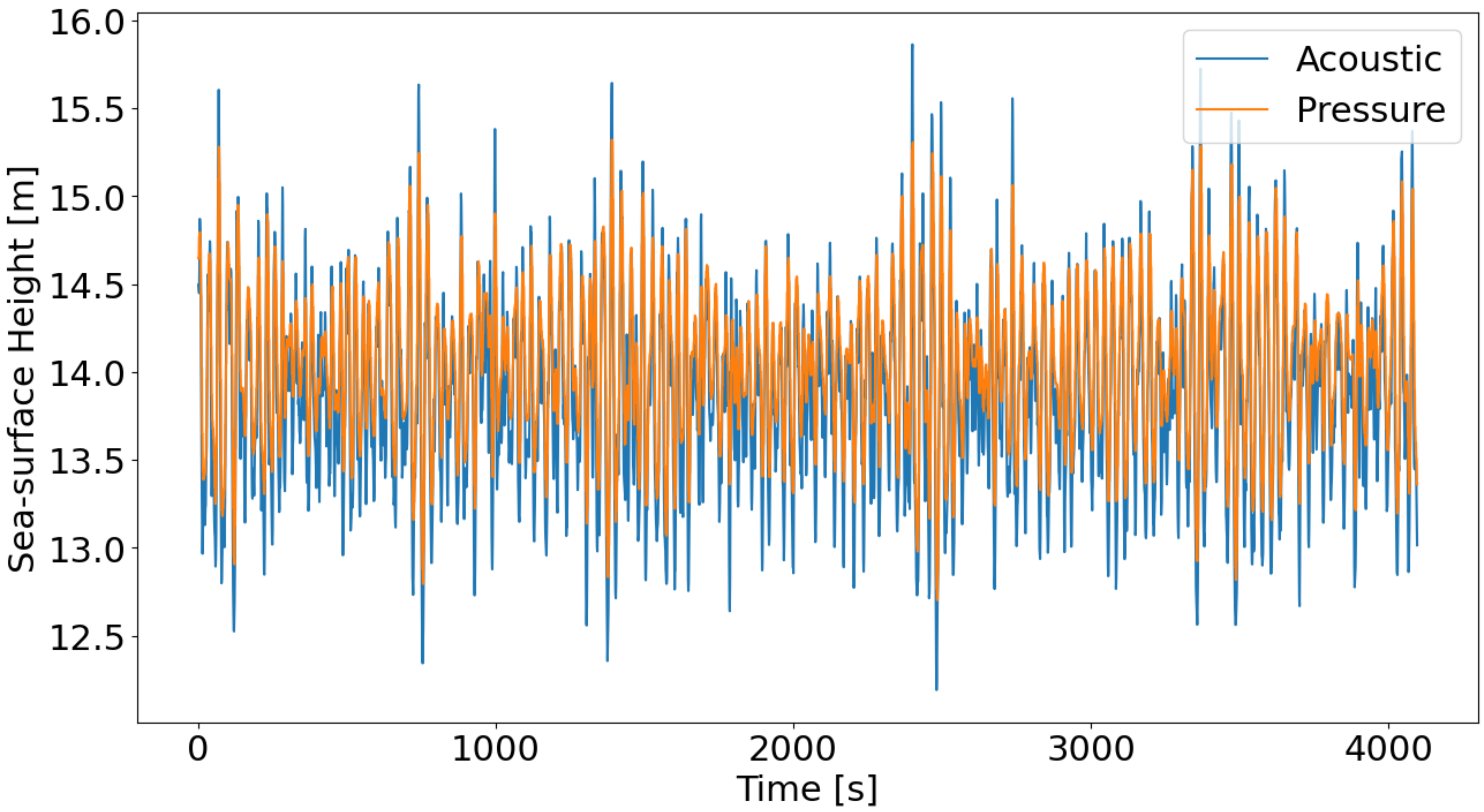
3D Shallow Island Wakes

Multidimensional data



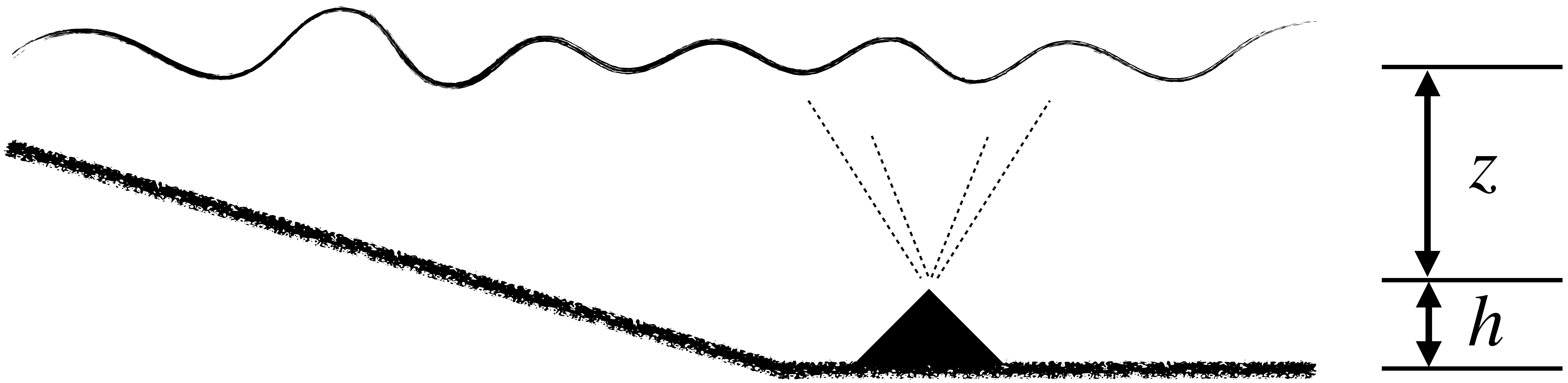
Sentinel-2 Sea Surface Imaging

Coastal Wave Measurements



Ocean Beach sea-surface heights

Coastal Wave Measurements



Pressure is attenuated as $K_p(k,z)^2 = \left(\frac{\cosh(kh+kz)}{\cosh(kh)}\right)^2$

$$\ldots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \ldots$$

$$\ldots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \cdots$$

- Assume we observe a real-valued stochastic process $\{X_t\}$ for $t \in \mathbb{Z}$, observed at the interval Δ.

$$\ldots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \cdots$$

- Assume we observe a real-valued stochastic process $\{X_t\}$ for $t \in \mathbb{Z}$, observed at the interval Δ.
- Assume $\mathrm{E}[X_t] = \text{constant}$ and Gaussian $\{X_t\}$

$$\ldots, x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}, \cdots$$

- Assume we observe a real-valued stochastic process $\{X_t\}$ for $t \in \mathbb{Z}$, observed at the interval Δ.
- Assume $\mathrm{E}[X_t] = \mathrm{constant}$ and Gaussian $\{X_t\}$
- If $\{X_t\}$ is second-order stationary we define the auto-covariance function $\gamma(\tau) = \mathrm{E}[X_t X_{t+\tau}]$

$$
\mathrm{X}_n \sim \mathrm{MN}\left(0, \begin{bmatrix} \gamma(0) & \dots & \gamma(n-1) \\ \vdots & \ddots & \vdots \\ \gamma(n-1) & \dots & \gamma(0) \end{bmatrix}\right)
$$

$$X_n \sim \mathrm{MN}\left(0, \begin{bmatrix} \gamma(0) & \dots & \gamma(n-1) \\ \vdots & \ddots & \vdots \\ \gamma(n-1) & \dots & \gamma(0) \end{bmatrix}\right)$$

- We can specify a parametric form for $\gamma(\tau)$ and fit via maximum likelihood.

$$X_n \sim \mathrm{MN}\left(0, \begin{bmatrix} \gamma(0) & \dots & \gamma(n-1) \\ \vdots & \ddots & \vdots \\ \gamma(n-1) & \dots & \gamma(0) \end{bmatrix}\right)$$

- We can specify a parametric form for $\gamma(\tau)$ and fit via maximum likelihood.
- Once we've fit a $\gamma(\tau)$ we can interpolate data, make predictions, and everything else a GP does.

$$X_n \sim MN\left(0, \begin{bmatrix} \gamma(0) & \dots & \gamma(n-1) \\ \vdots & \ddots & \vdots \\ \gamma(n-1) & \dots & \gamma(0) \end{bmatrix}\right)$$

- We can specify a parametric form for $\gamma(\tau)$ and fit via maximum likelihood.
- Once we've fit a $\gamma(\tau)$ we can interpolate data, make predictions, and everything else a GP does.
- $\gamma(\tau)$ must be positive semi-definite.

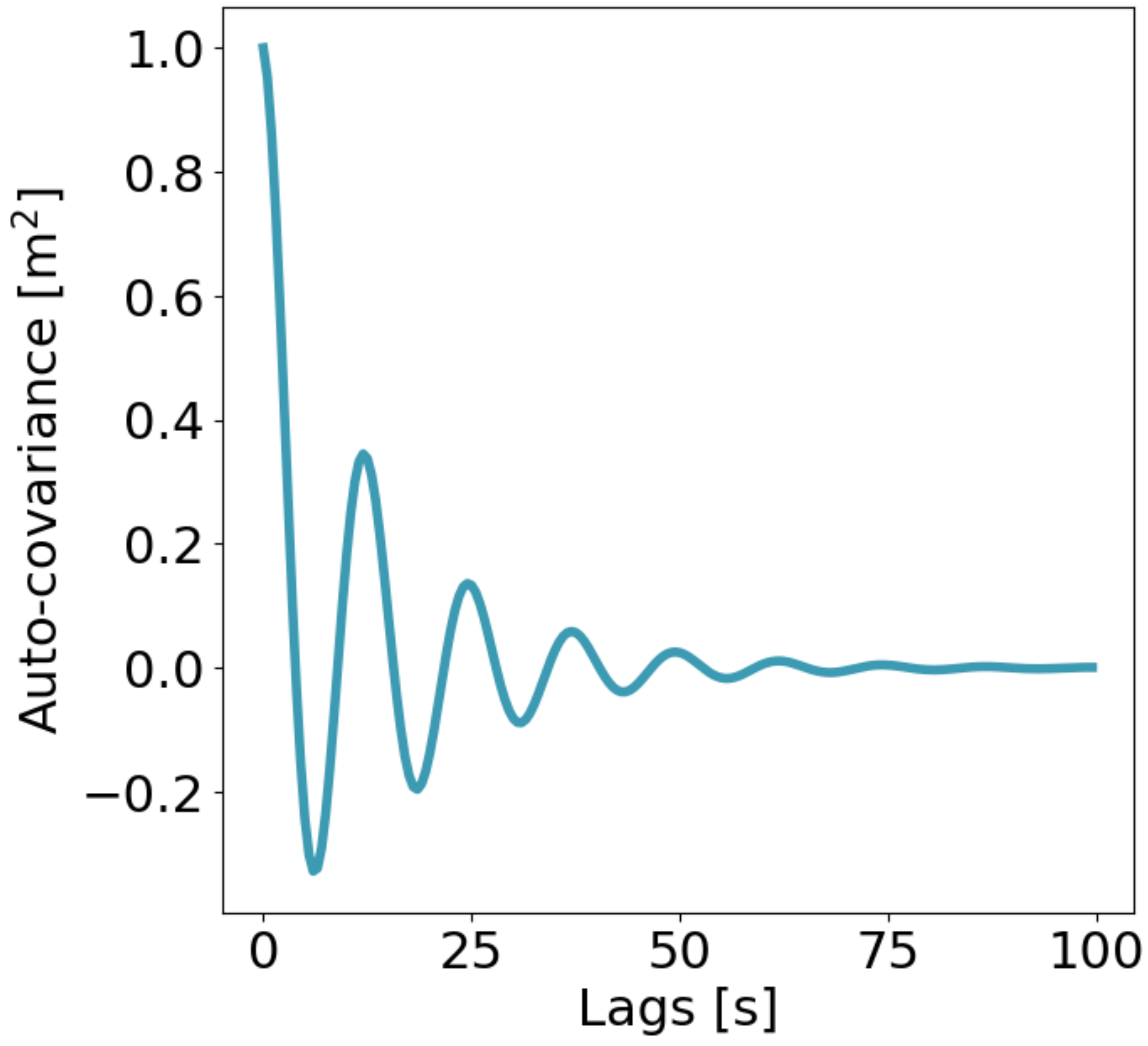
Bochner's Theorem

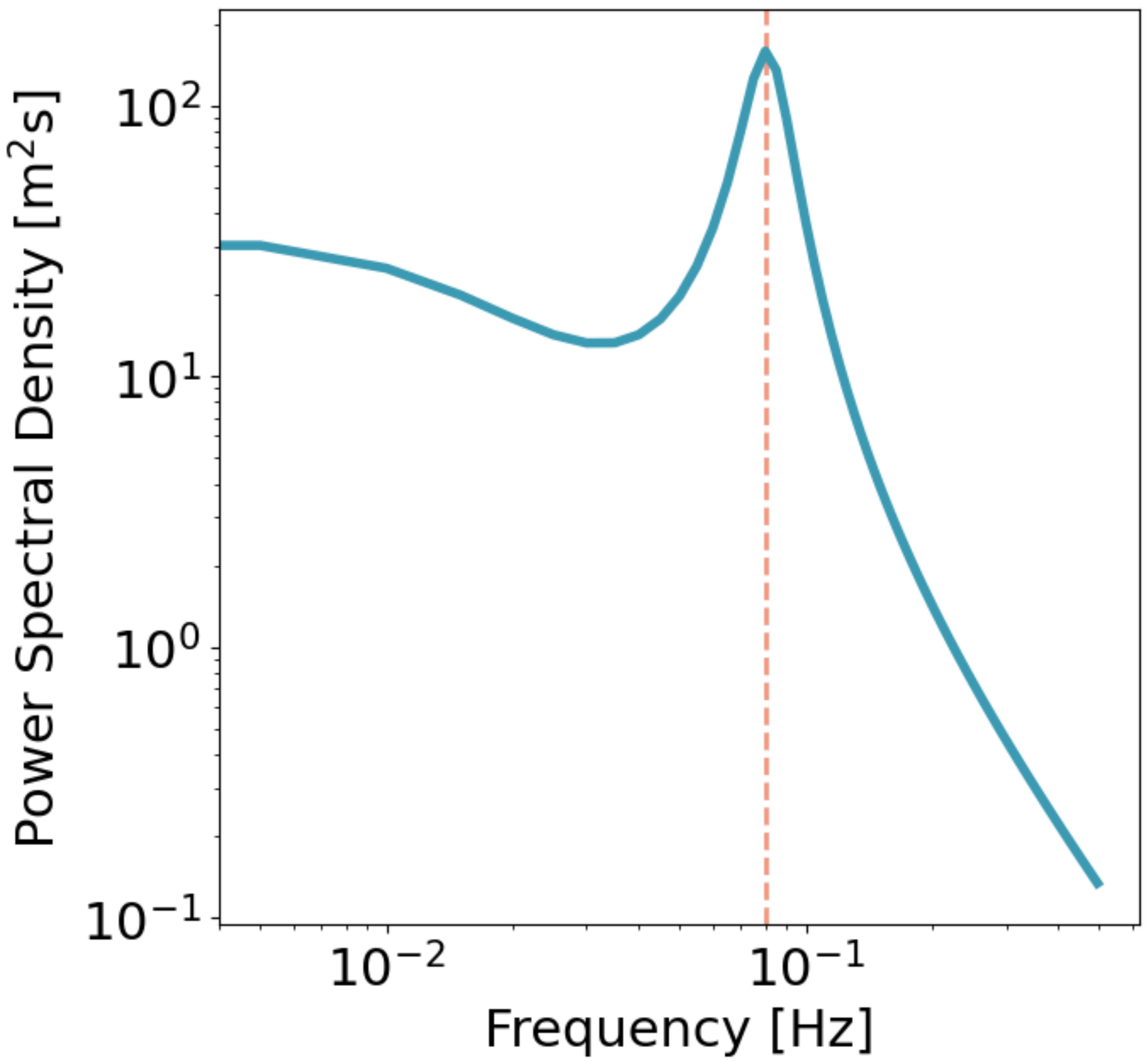
(aka Wiener-Khinchin's Theorem)

If $\gamma(\tau)$ is absolutely summable then there exists a power spectral density $f(\omega)$ that forms a Fourier pair with $\gamma(\tau)$

$$\gamma(\tau) = \int_{-1/2}^{1/2} f(\omega) e^{i\omega\tau}\, \mathrm{d}\omega, \qquad f(\omega) = \sum_{\tau=\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau}$$

The ACF vs the PSD





Discrete sampling and estimation

Empirically estimating the PSD

Discrete sampling and estimation

Empirically estimating the PSD

What's a good empirical estimate of $f(\omega) = \mathscr{F}\{\gamma(\tau)\}$?

Discrete sampling and estimation

Empirically estimating the PSD

What's a good empirical estimate of $f(\omega) = \mathcal{F}\{\gamma(\tau)\}$?

$$I_n(\omega) = \frac{1}{n}\left|\sum_{t=0}^{n-1} x_t e^{-i\omega\tau}\right|^2 = \sum_{\tau=-(n-1)}^{n-1} \hat{\gamma}_b(\tau) e^{-i\omega\tau}$$

This defines the ***periodogram***. This gives a strictly positive, although biased, estimate to $f(\omega)$.

Discrete sampling and estimation

Empirically estimating the PSD

What's a good empirical estimate of $f(\omega) = \mathcal{F}\{\gamma(\tau)\}$?

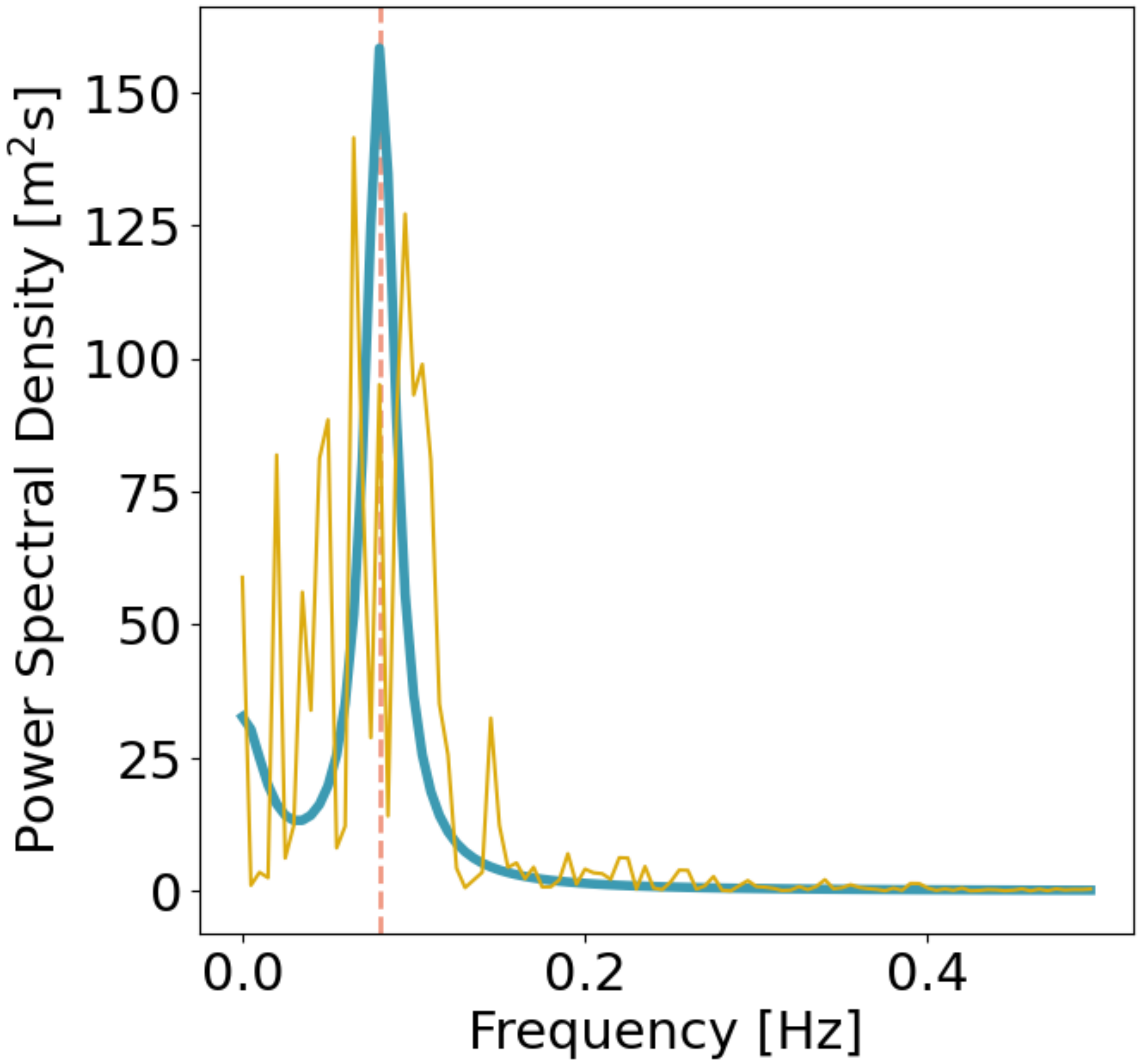
$$I_n(\omega) = \frac{1}{n}\left|\sum_{t=0}^{n-1} x_t e^{-i\omega\tau}\right|^2 = \sum_{\tau=-(n-1)}^{n-1} \hat{\gamma}_b(\tau)e^{-i\omega\tau}$$

This defines the ***periodogram***. This gives a strictly positive, although biased, estimate to $f(\omega)$.

$$\hat{\gamma}_b(\tau) = \frac{1}{n}\sum_{t=0}^{n-|\tau|-1} x_t x_{t+|\tau|} \longrightarrow \mathrm{E}[\hat{\gamma}_b(\tau)] = \left(1 - \frac{|\tau|}{n}\right)\gamma(\tau)$$

The periodogram is noisy and inconsistent

$$I_n(\omega) \sim \mathrm{E}[I_n(\omega)]\ \mathscr{X}_2^2$$

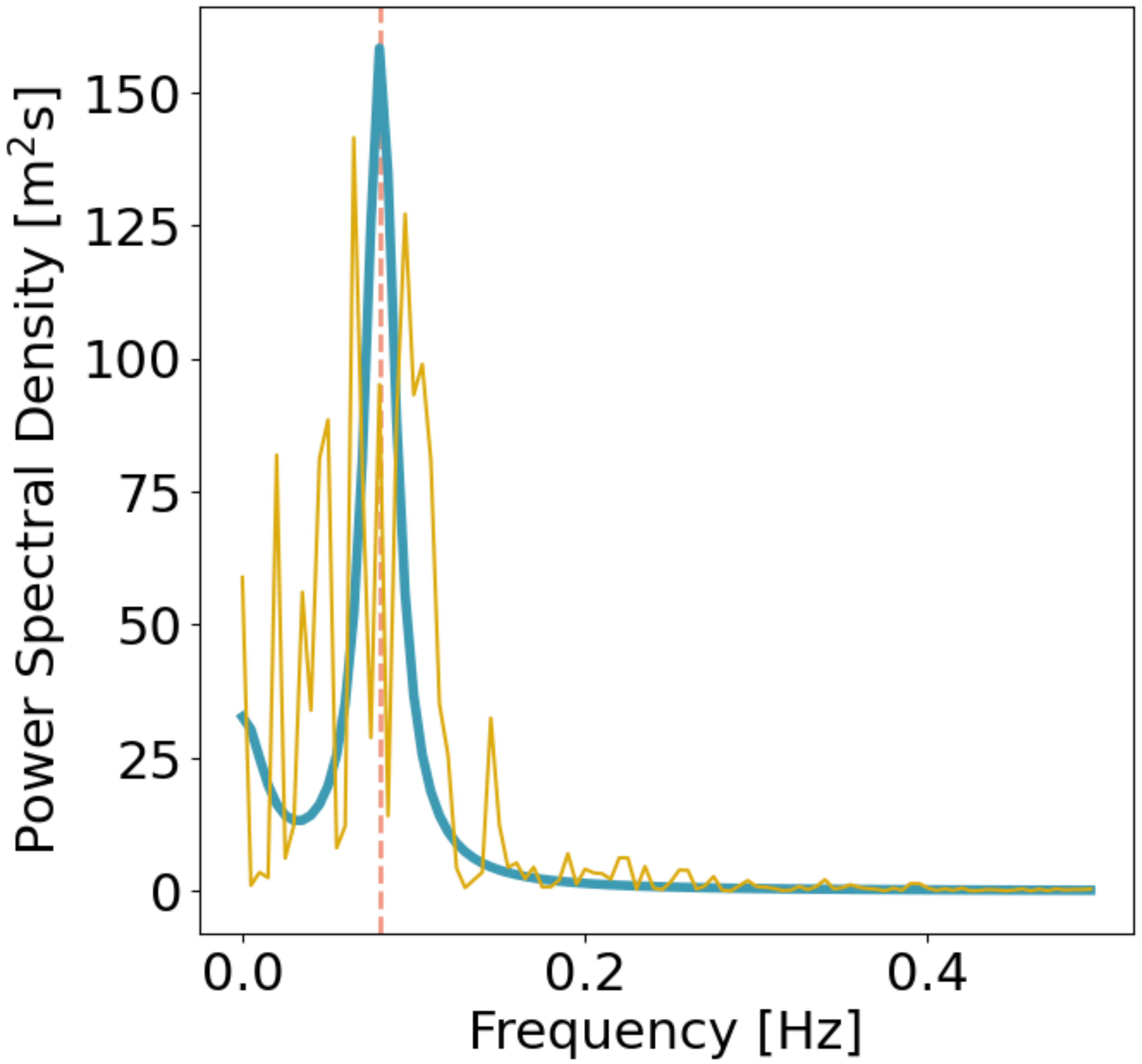


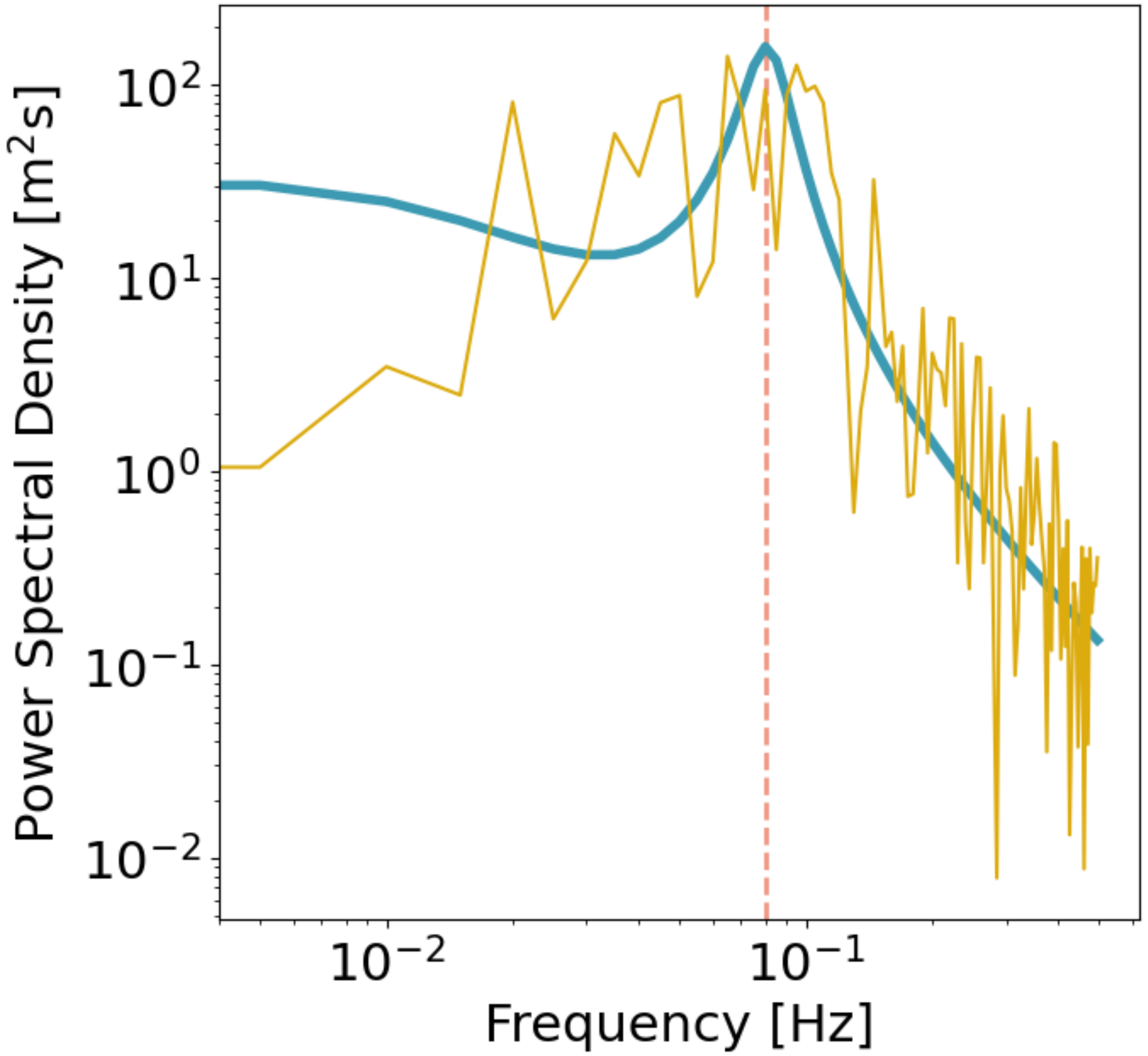
(approximately)

The periodogram is noisy and inconsistent

$$I_n(\omega) \sim \mathrm{E}[I_n(\omega)]\ \mathcal{X}_2^2$$

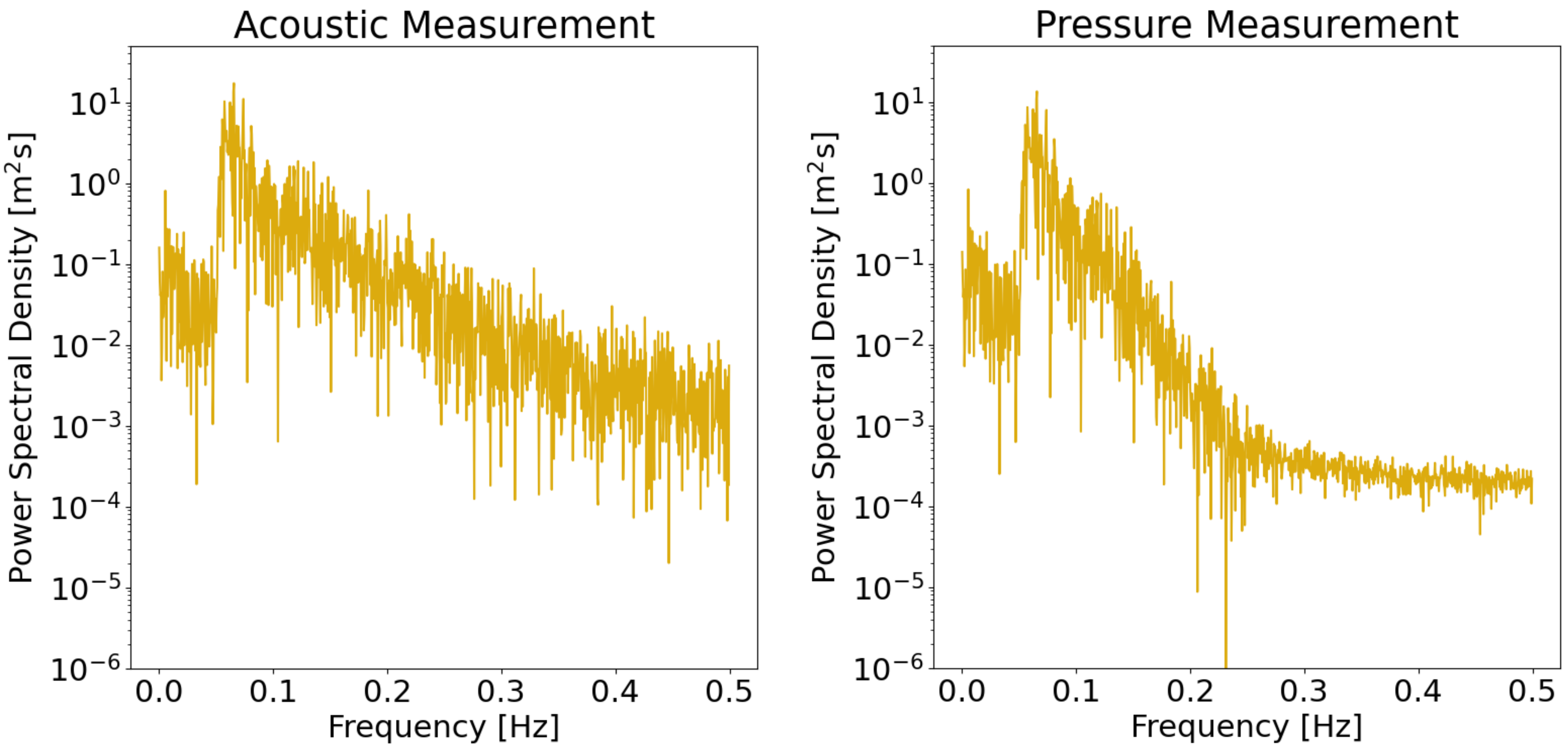
$$\log I_n(\omega) \sim \log \mathrm{E}[I_n(\omega)] + \log \mathcal{X}_2^2$$



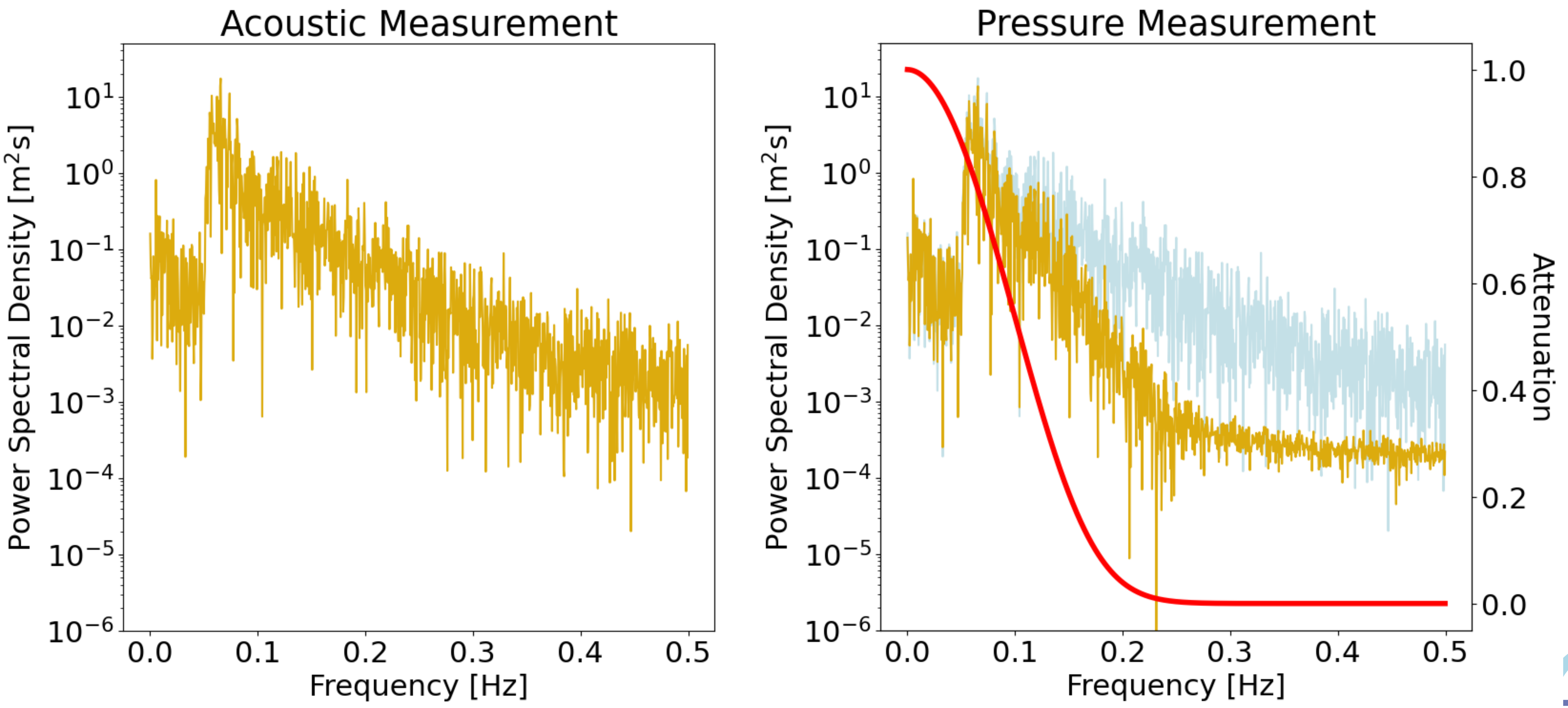


(approximately)

Log Periodograms of Wave Measurements



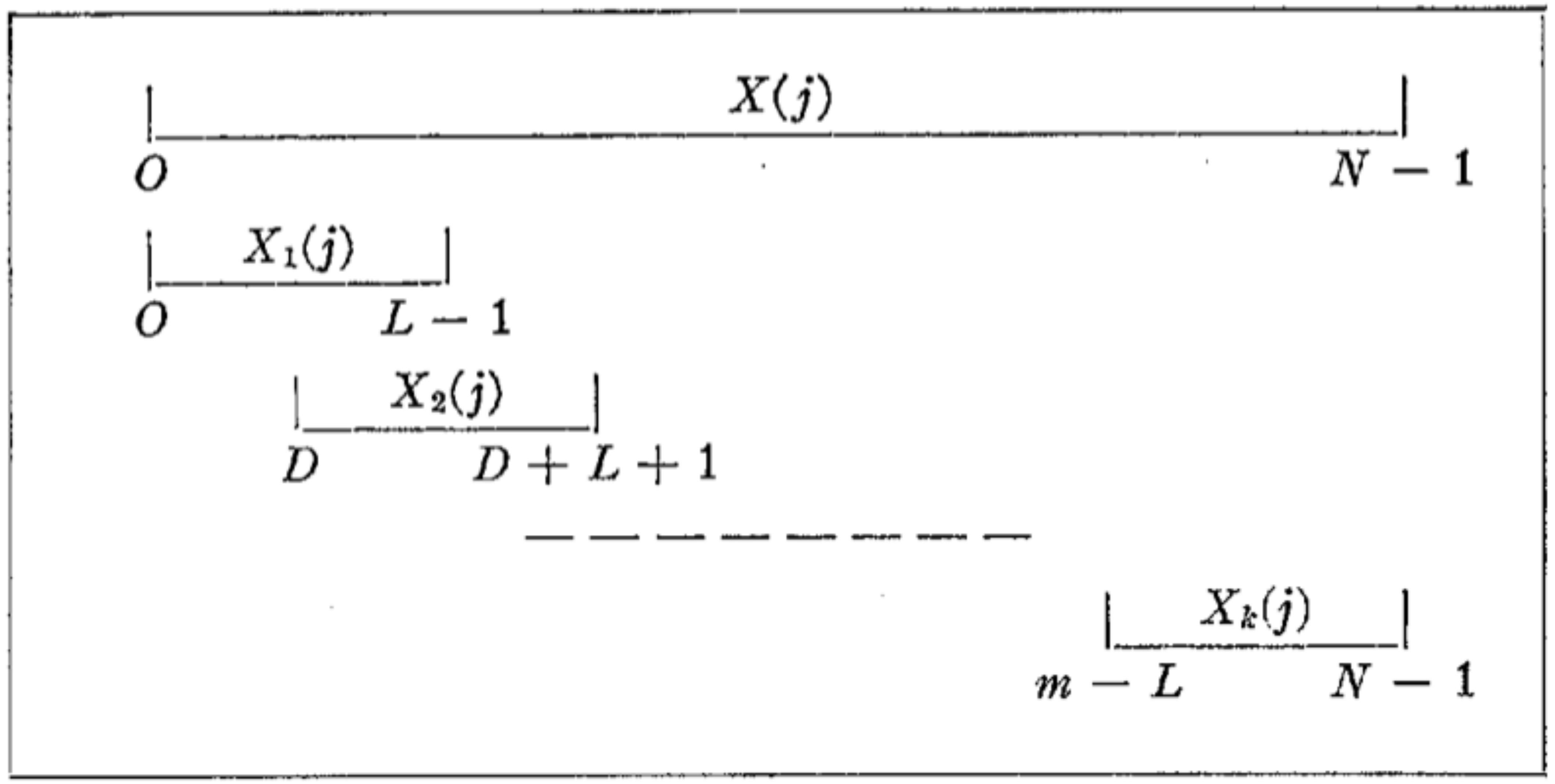
Log Periodograms of Wave Measurements



Welch's Estimate

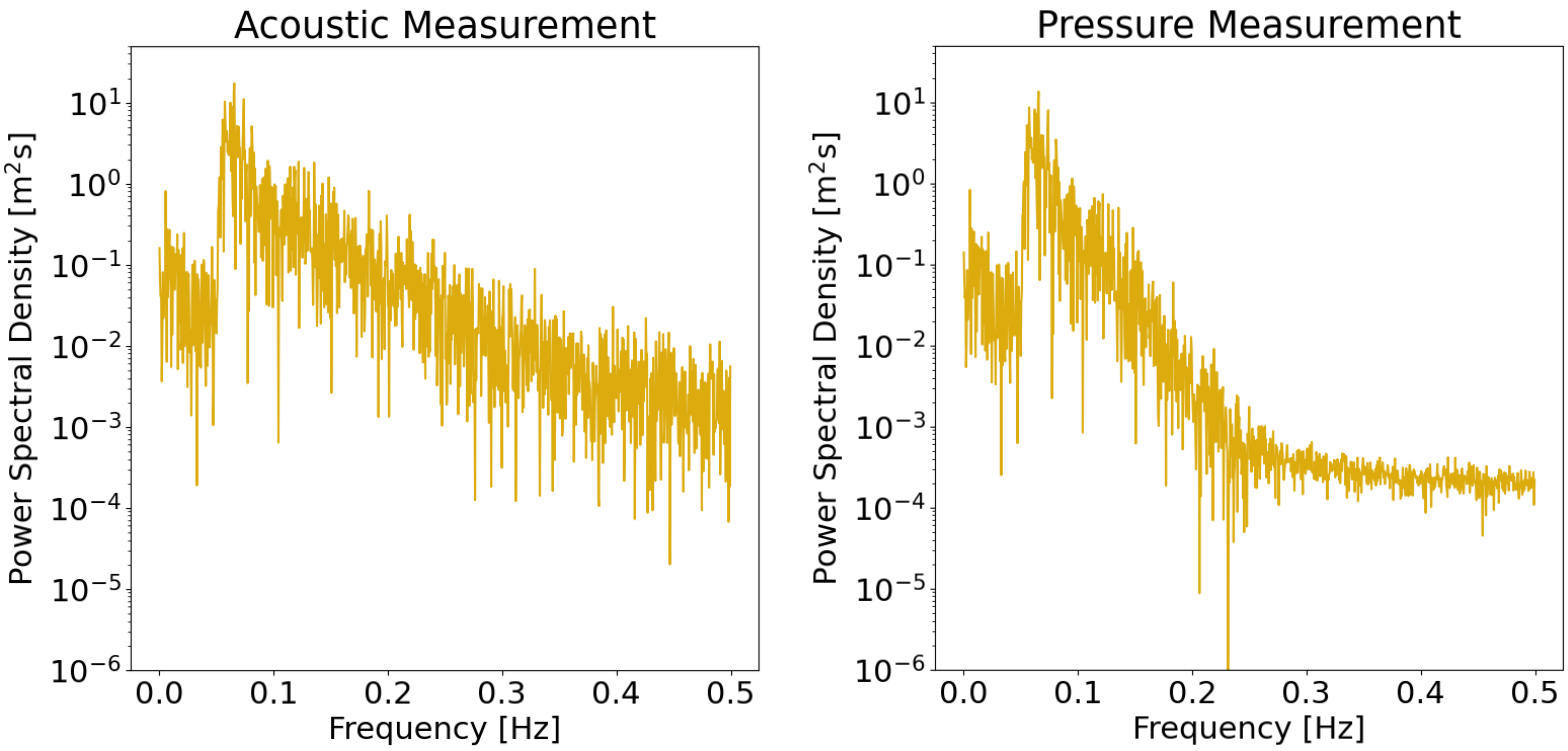
Welch's estimator partitions a time-series into m overlapping blocks of length L, calculates the taper periodograms of each block, $I_L^j(\omega; h)$, and is defined as

$$\bar{I}_L(\omega; h) = \frac{1}{m} \sum_{j=0}^{j-1} I_L^j(\omega; h)$$

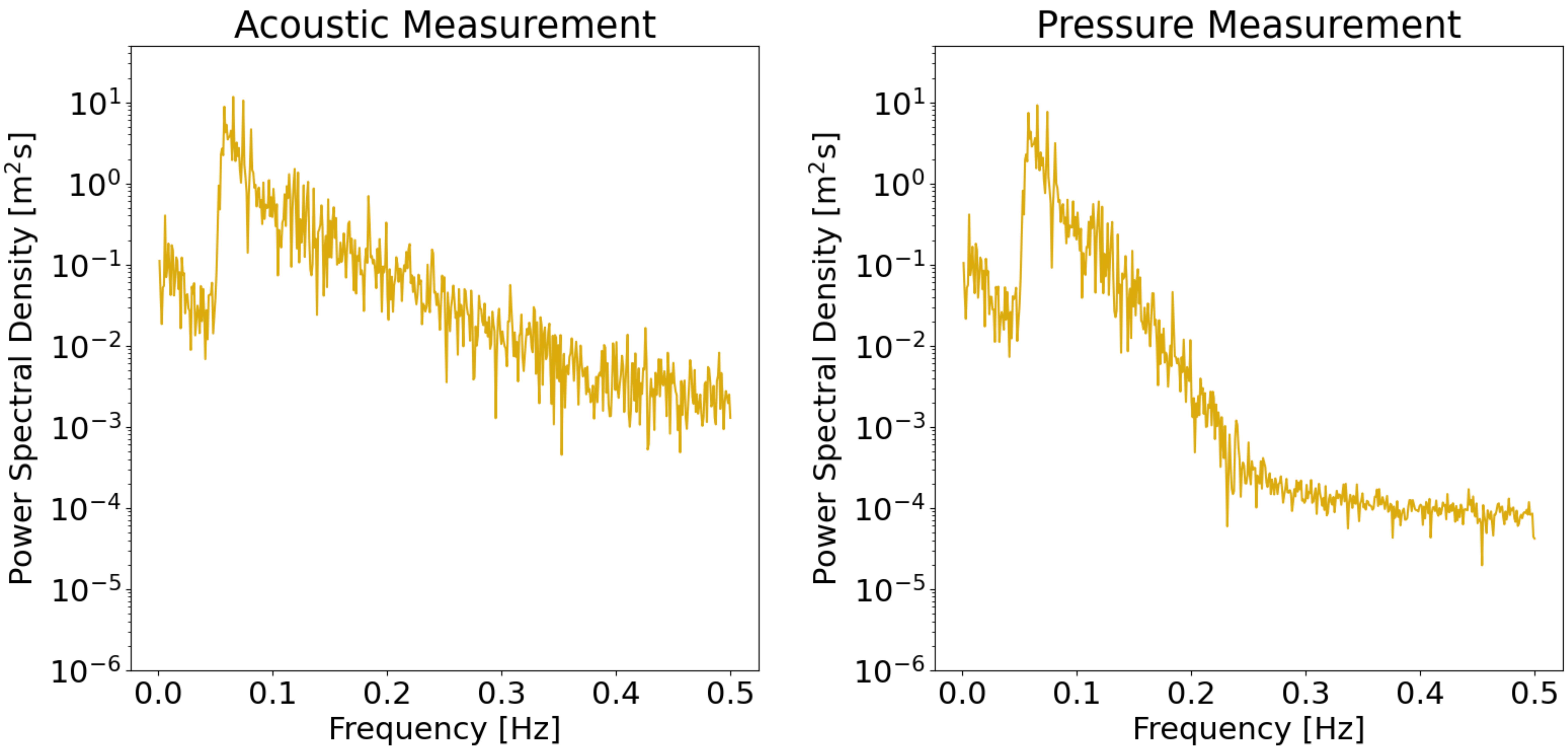


Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, *15*(2), 70-73.

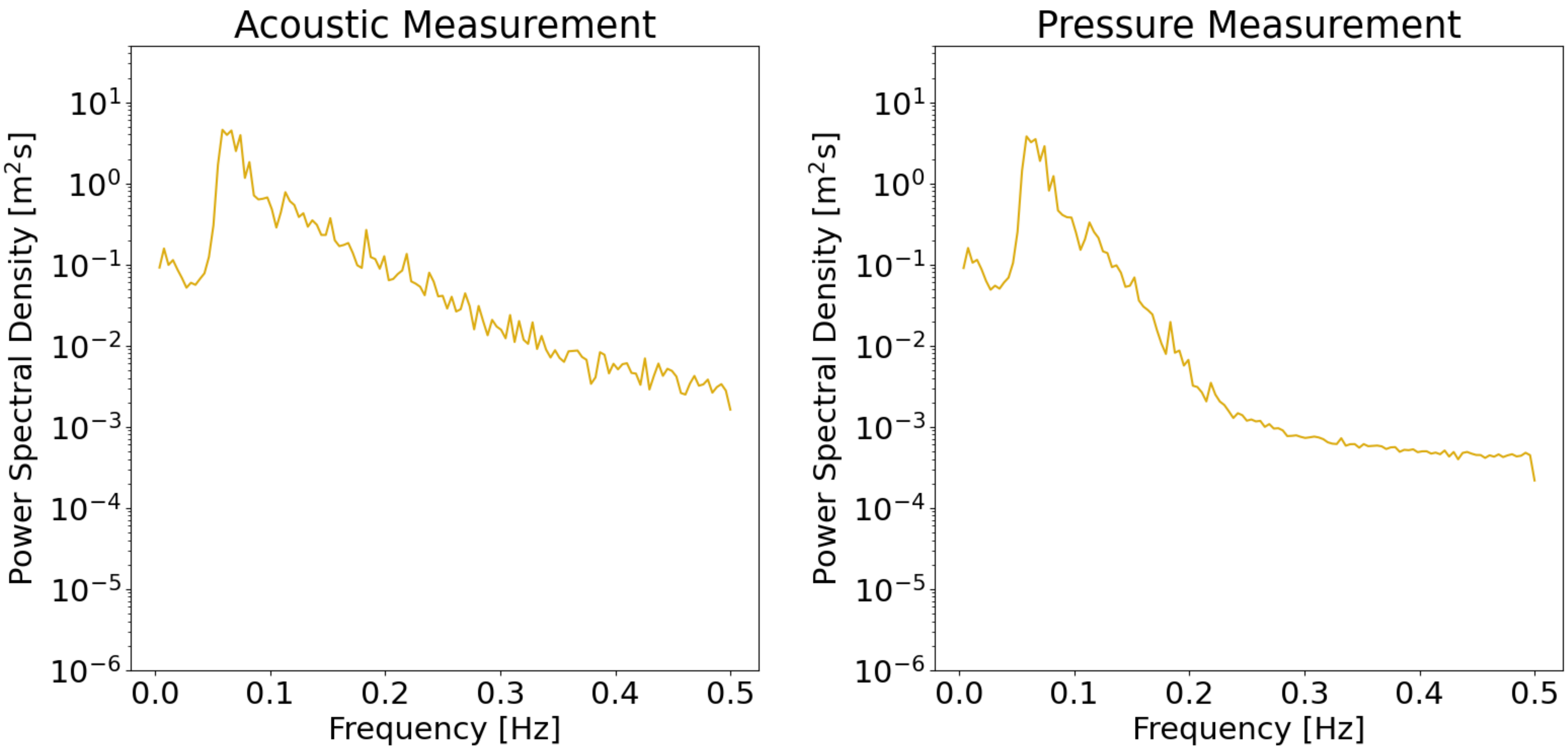
Welch’s Estimate of Wave Data



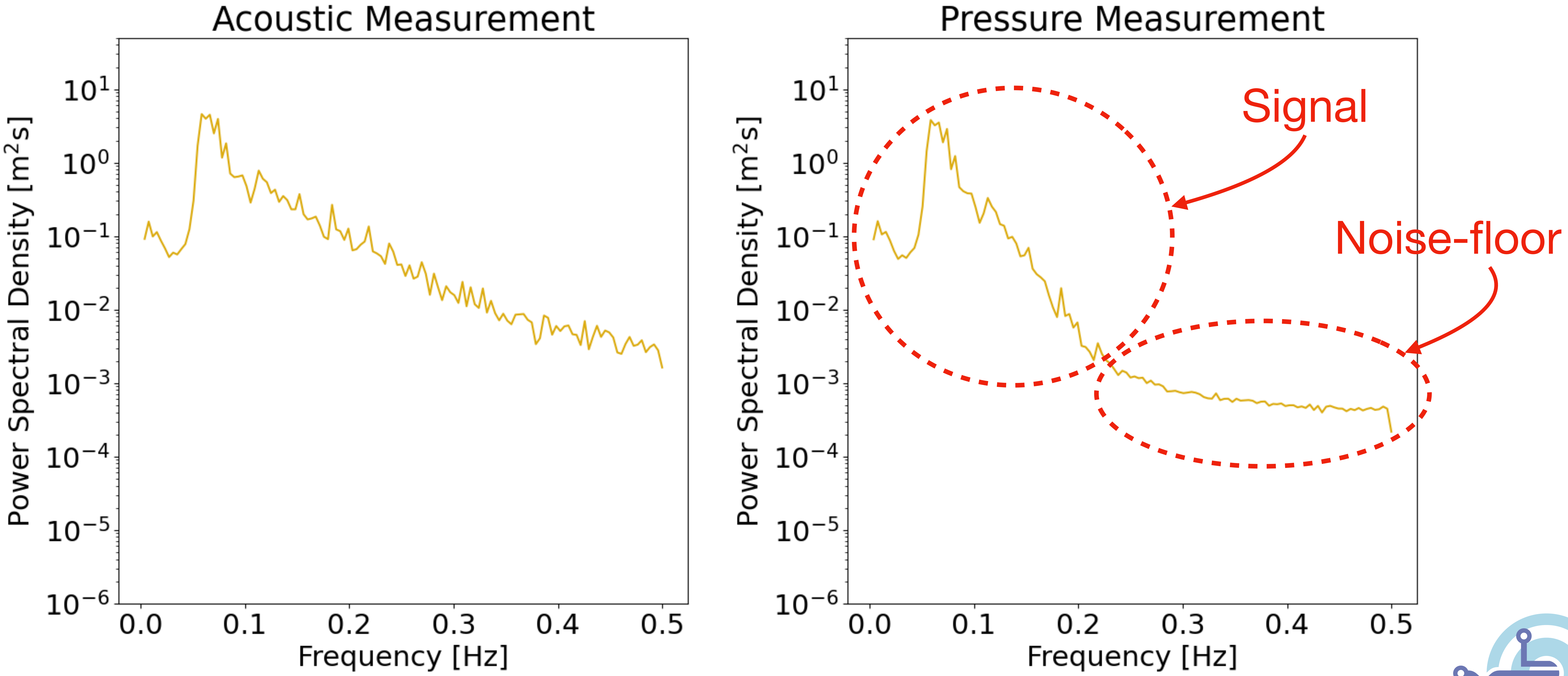
Welch’s Estimate of Wave Data



Welch's Estimate of Wave Data

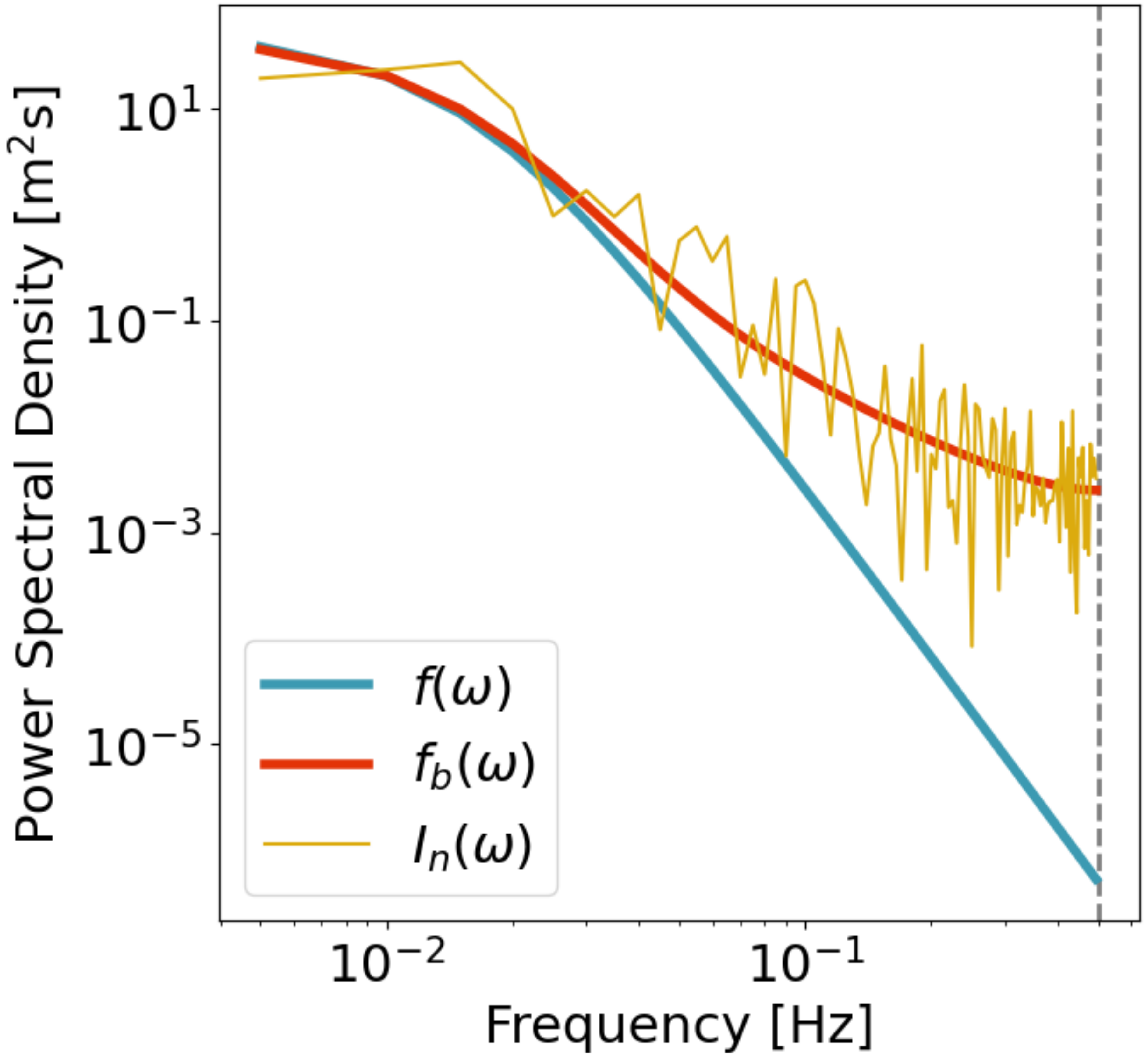


Welch's Estimate of Wave Data



The periodogram is blurred

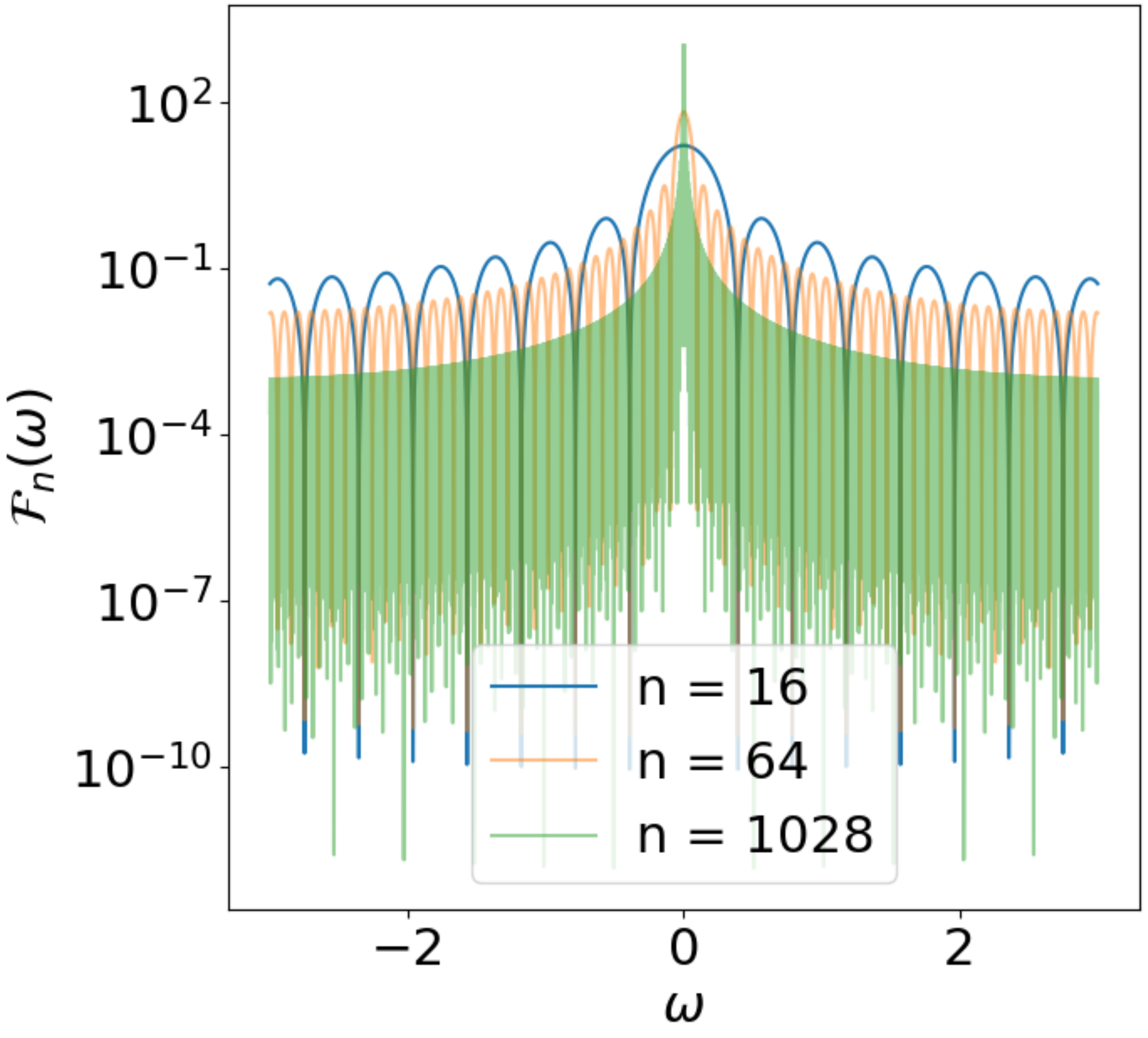
Seeing as the periodogram is defined from a biased estimate of the ACF, we may also suspect that $I_n(\omega)$ is also biased.



Fejer's kernel

Define Fejer's kernel

$$\mathscr{F}_n(\omega) = \frac{1}{n}\left(\frac{1-\cos(n\omega)}{1-\cos(\omega)}\right)$$
$$= \sum_{\tau=1-n}^{n-1}\left(1-\frac{|\tau|}{n}\right)e^{-i\omega\tau}$$



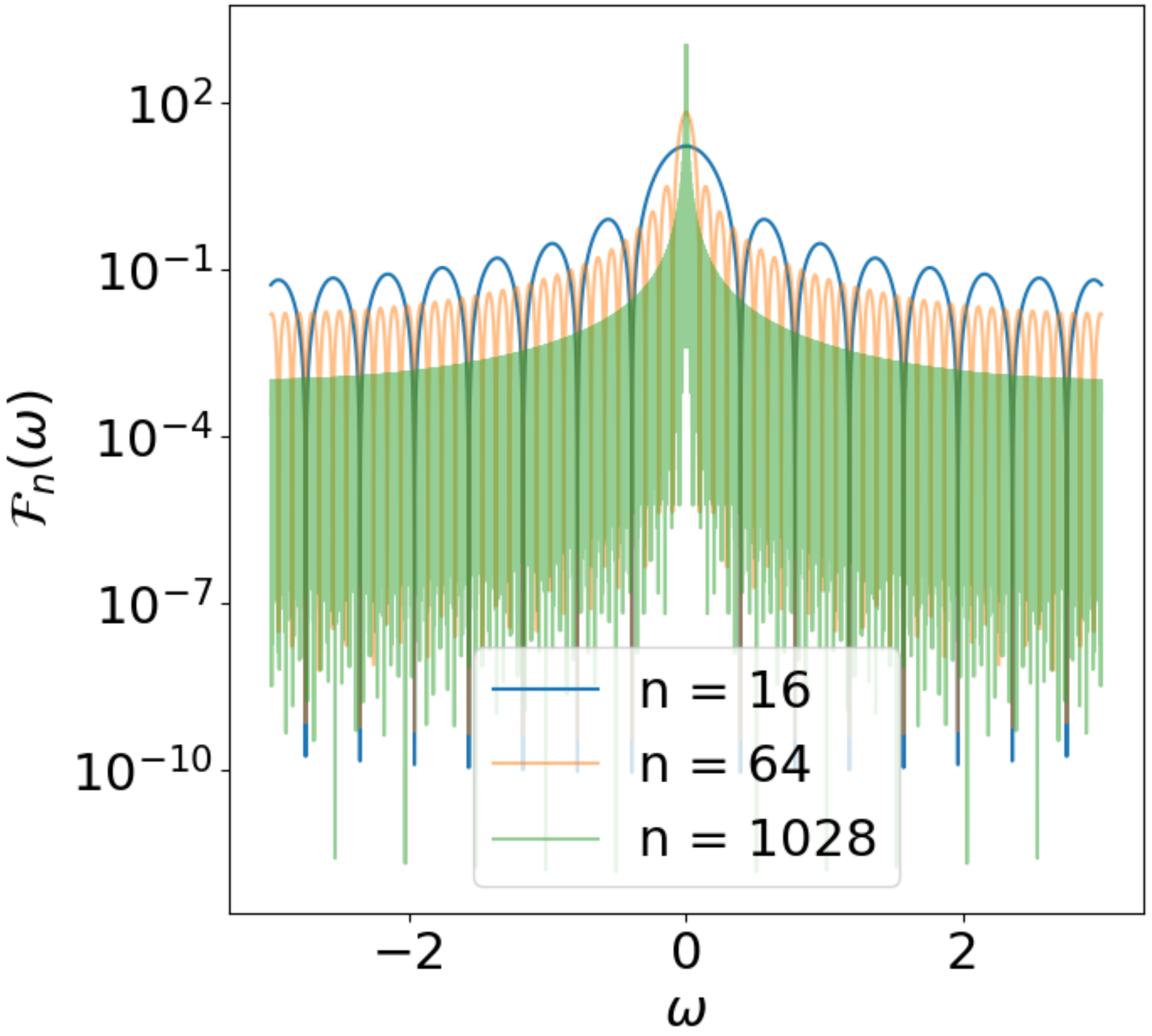
The blurred PSD is $\tilde{f}(\omega) = f(\omega) * \mathscr{F}_n(\omega)$

Fejer's kernel

Define Fejer's kernel

$$\mathcal{F}_n(\omega) = \frac{1}{n}\left(\frac{1-\cos(n\omega)}{1-\cos(\omega)}\right)$$
$$= \sum_{\tau=1-n}^{n-1}\left(1-\frac{|\tau|}{n}\right)e^{-i\omega\tau}$$

The blurred PSD is $\tilde{f}(\omega) = f(\omega) * \mathcal{F}_n(\omega)$



Bias decreases as n increases

What's wrong with Welch's estimate?

- Welch's estimate enforces consistency by partitioning and averaging
- As the m blocks increase the variance of our estimator decreases
- Also, as m increases, block length L decreases (and hence bias increases)

What's wrong with Welch's estimate?

- Welch's estimate enforces consistency by partitioning and averaging
- As the m blocks increase the variance of our estimator decreases
- Also, as m increases, block length L decreases (and hence bias increases)

We become increasingly more confident in an estimate that is increasingly more wrong

How do we manage spectral bias?

Parametric Estimation

To avoid expensive matrix inversions we can fit some $f(\omega;\theta)$ with the Whittle likelihood calculated in $\mathcal{O}(n\log n)$

$$l_W(\theta) = -\sum_{\omega\in\Omega_n}\left\{\log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)}\right\}$$

For a nice proof of this, see Kirch, C., et al. (2019). Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis.

Debiased parametric estimation

Either debias the periodogram…

Correct the bias in the data?

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Debiased parametric estimation

Either debias the periodogram…

Correct the bias in the data?

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Ways we could remove bias from the periodogram:

Debiased parametric estimation

Either debias the periodogram…

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Correct the bias in the data?

Ways we could remove bias from the periodogram:

- Tapering. But at the cost of introducing a different type of bias.

Debiased parametric estimation

Either debias the periodogram…

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Correct the bias in the data?

Ways we could remove bias from the periodogram:

- Tapering. But at the cost of introducing a different type of bias.
- Pre-whitening. Doesn’t always work, hard to tune.

Debiased parametric estimation

Either debias the periodogram…

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Correct the bias in the data?

Ways we could remove bias from the periodogram:

- Tapering. But at the cost of introducing a different type of bias.
- Pre-whitening. Doesn't always work, hard to tune.
- Collect more data at a higher sampling rate. Not really possible.

Biased parametric estimation

…or bias the spectral density

$$l_W(\theta) = -\sum_{\omega\in\Omega_n}\left\{\log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)}\right\}$$

Or match the model to the biased data?

Biased parametric estimation

…or bias the spectral density

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Or match the model to the biased data?

Call $\tilde{f}_n(\omega;\theta)$ the biased spectral density. We need it to incorporate the effects of blurring and aliasing.

Biased parametric estimation

…or bias the spectral density

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Or match the model to the biased data?

Call $\tilde{f}_n(\omega;\theta)$ the biased spectral density. We need it to incorporate the effects of blurring and aliasing.

Blurred spectrum: $f_b(\omega;\theta) = f(\omega;\theta) * \mathscr{F}_n(\omega)$

Biased parametric estimation

…or bias the spectral density

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log f(\omega;\theta) + \frac{I_n(\omega)}{f(\omega;\theta)} \right\}$$

Or match the model to the biased data?

Call $\tilde{f}_n(\omega;\theta)$ the biased spectral density. We need it to incorporate the effects of blurring and aliasing.

Blurred spectrum: $f_b(\omega;\theta) = f(\omega;\theta) * \mathscr{F}_n(\omega)$

Aliased spectrum: $f_a(\omega;\theta) = \sum_{k=-\infty}^{k=\infty} f(\omega + k)$

The debiased Whittle likelihood

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log \tilde{f}_n(\omega; \theta) + \frac{I_n(\omega)}{\tilde{f}_n(\omega; \theta)} \right\}$$

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., & Early, J. J. (2019). The debiased Whittle likelihood. *Biometrika*, *106*(2), 251-266.

The debiased Whittle likelihood

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log \tilde{f}_n(\omega; \theta) + \frac{I_n(\omega)}{\tilde{f}_n(\omega; \theta)} \right\}$$

Where we calculate $\tilde{f}_n(\omega; \theta)$ by

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., & Early, J. J. (2019). The debiased Whittle likelihood. *Biometrika*, *106*(2), 251-266.

The debiased Whittle likelihood

$$l_W(\theta) = - \sum_{\omega \in \Omega_n} \left\{ \log \tilde{f}_n(\omega; \theta) + \frac{I_n(\omega)}{\tilde{f}_n(\omega; \theta)} \right\}$$

Where we calculate $\tilde{f}_n(\omega; \theta)$ by

$$\tilde{f}_n(\omega; \theta) = 2 \times \mathrm{Re} \left\{ \sum_{\tau=0}^{n-1} \left(1 - \frac{\tau}{n}\right) \gamma(\tau; \theta) e^{-i\omega\tau} \right\} - \gamma(0; \theta)$$

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., & Early, J. J. (2019). The debiased Whittle likelihood. *Biometrika*, *106*(2), 251-266.

The debiased Whittle likelihood

$$l_W(\theta) = -\sum_{\omega \in \Omega_n} \left\{ \log \tilde{f}_n(\omega; \theta) + \frac{I_n(\omega)}{\tilde{f}_n(\omega; \theta)} \right\}$$

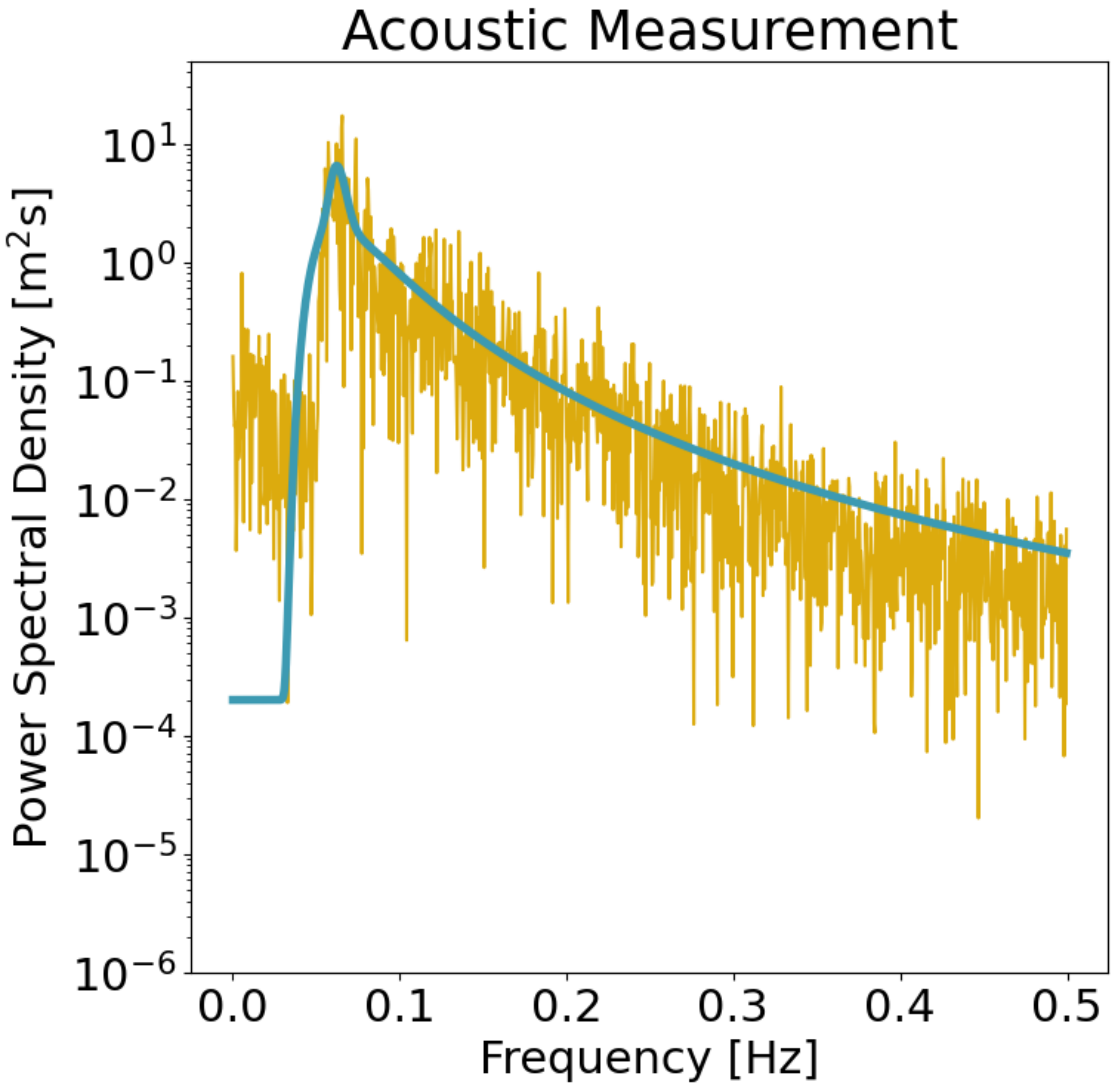
Where we calculate $\tilde{f}_n(\omega; \theta)$ by

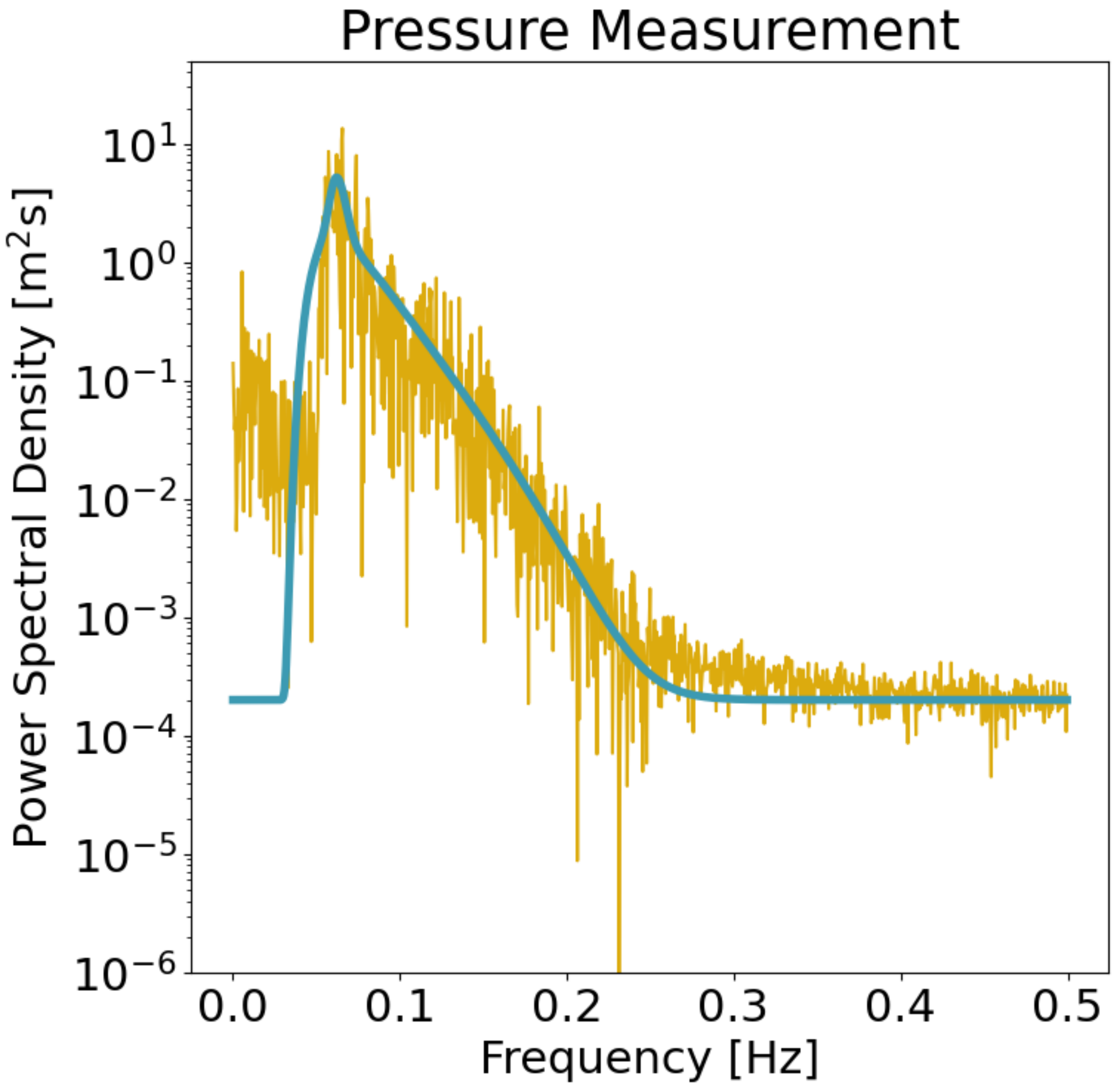
$$\tilde{f}_n(\omega; \theta) = 2 \times \mathrm{Re} \left\{ \sum_{\tau=0}^{n-1} \left(1 - \frac{\tau}{n}\right) \gamma(\tau; \theta) e^{-i\omega\tau} \right\} - \gamma(0; \theta)$$

We can show that $\mathrm{E}[I_n(\omega)] = \tilde{f}_n(\omega; \theta)$

Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., & Early, J. J. (2019). The debiased Whittle likelihood. *Biometrika*, *106*(2), 251-266.

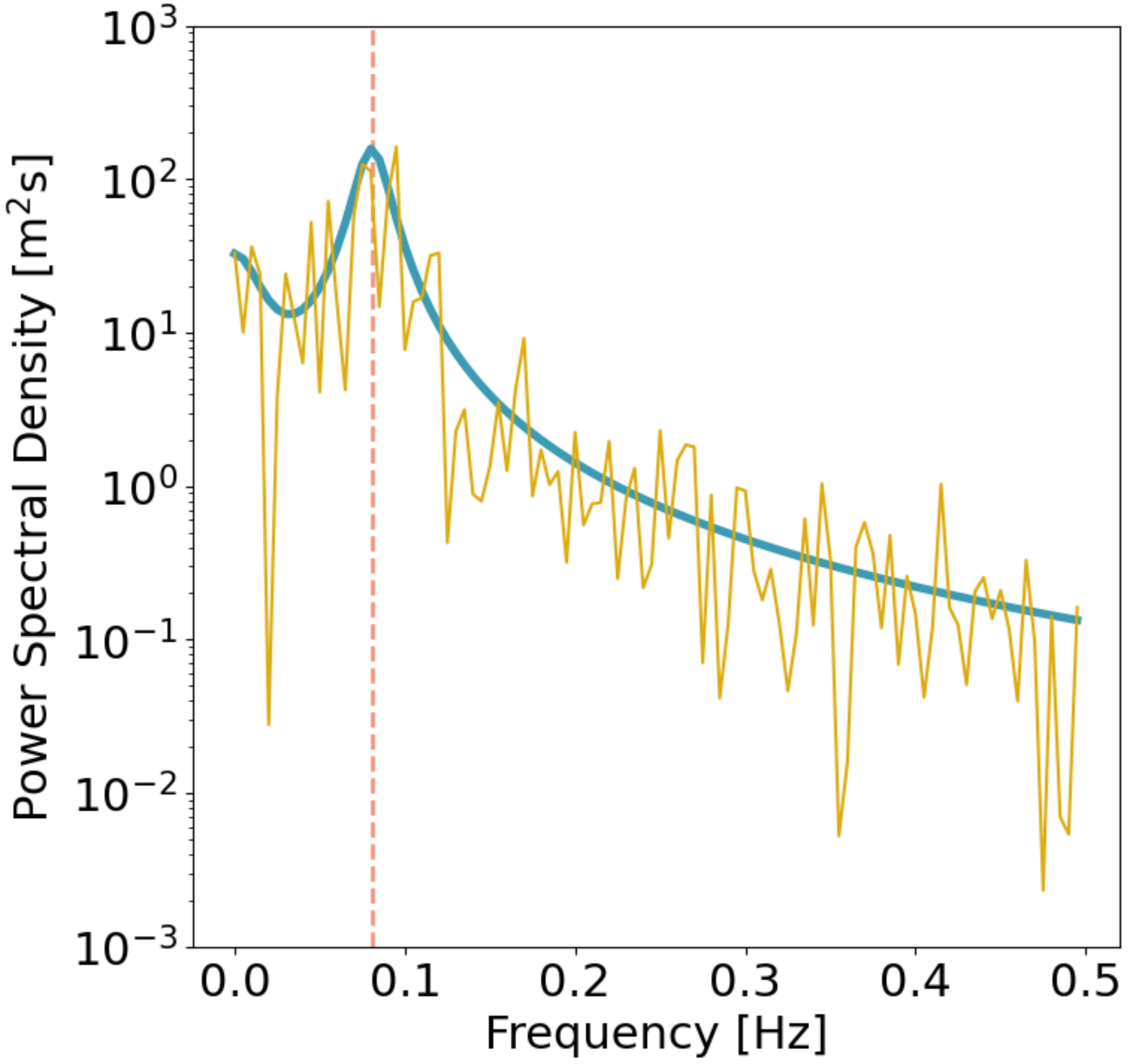
Assuming a model for the waves





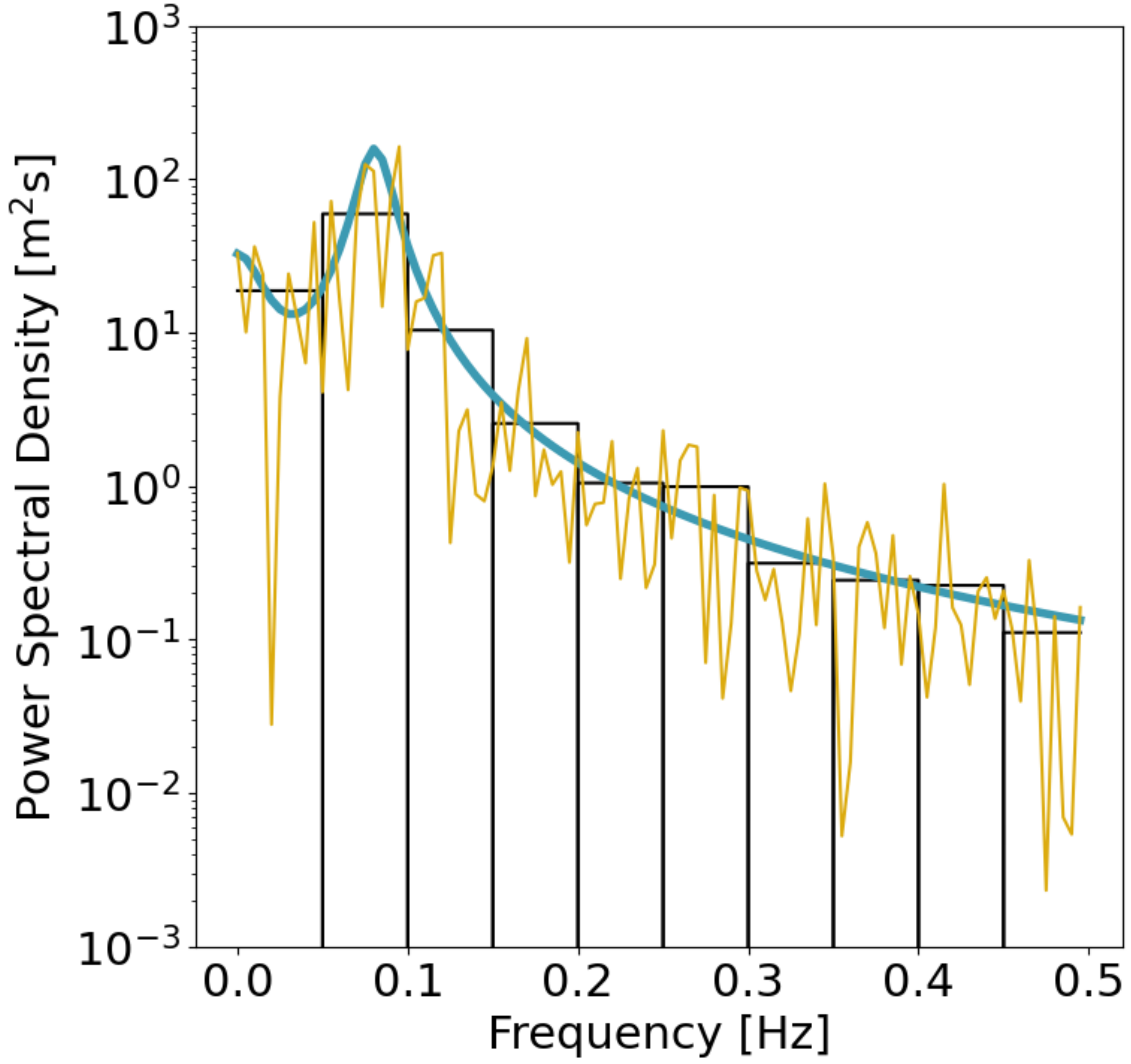
Debiasing Welch's estimate

Riemann approximation to the PSD



Debiasing Welch's estimate

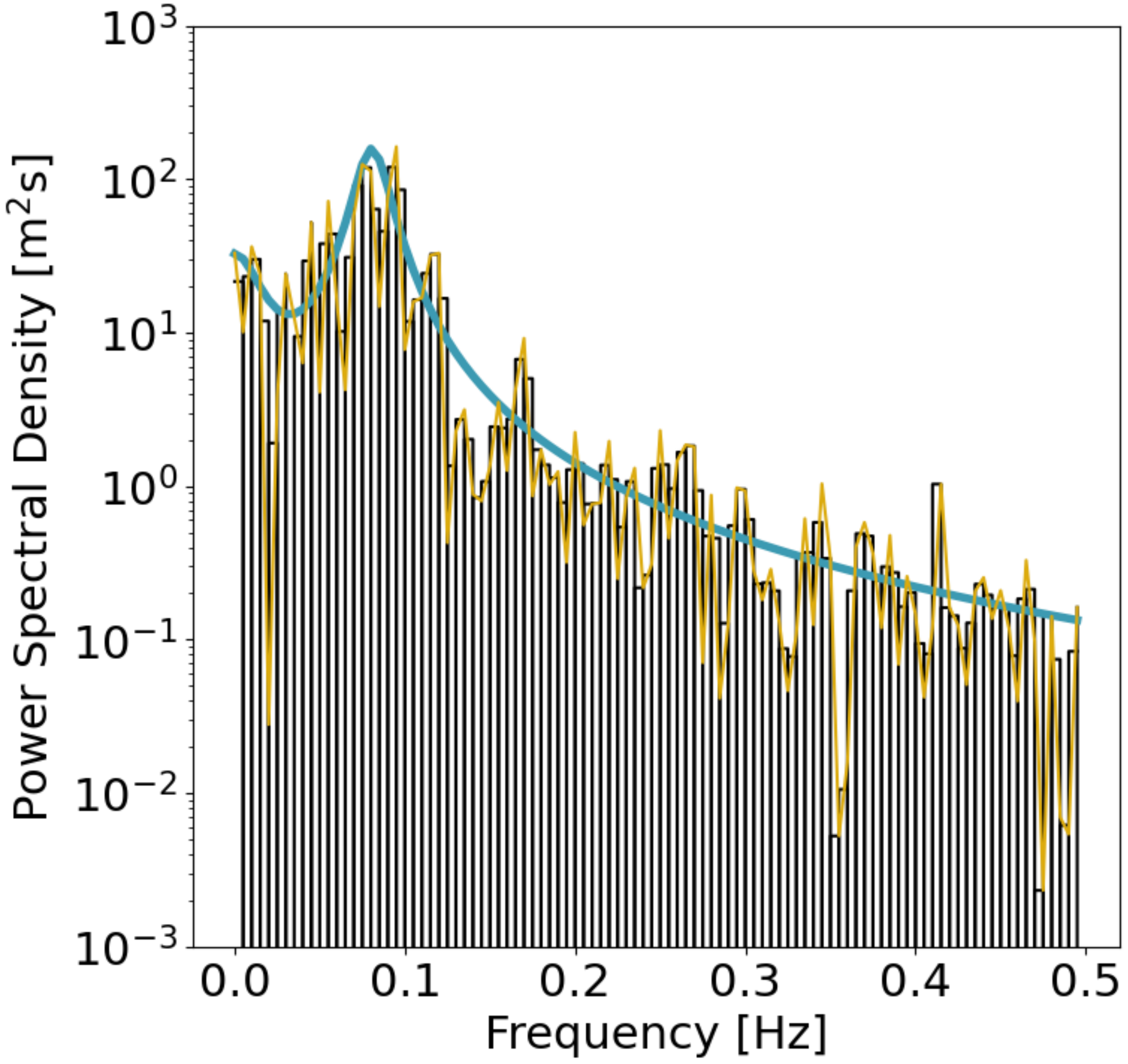
Riemann approximation to the PSD



- We can think about a periodogram as a Riemann approximation to the true biased PSD

Debiasing Welch's estimate

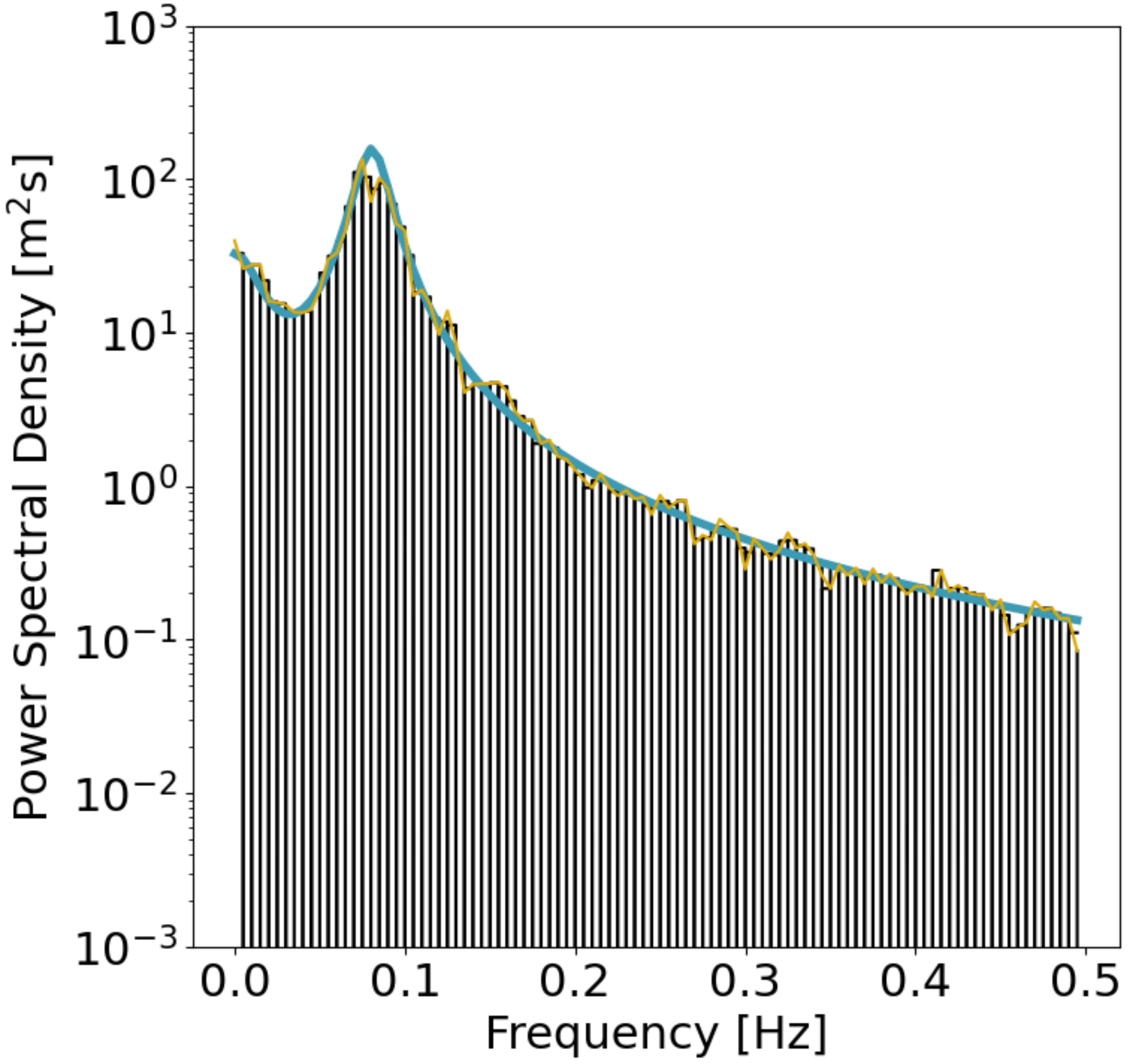
Riemann approximation to the PSD



- We can think about a periodogram as a Riemann approximation to the true biased PSD
- As we increase the resolution of the bases we converge on the true integral

Debiasing Welch's estimate

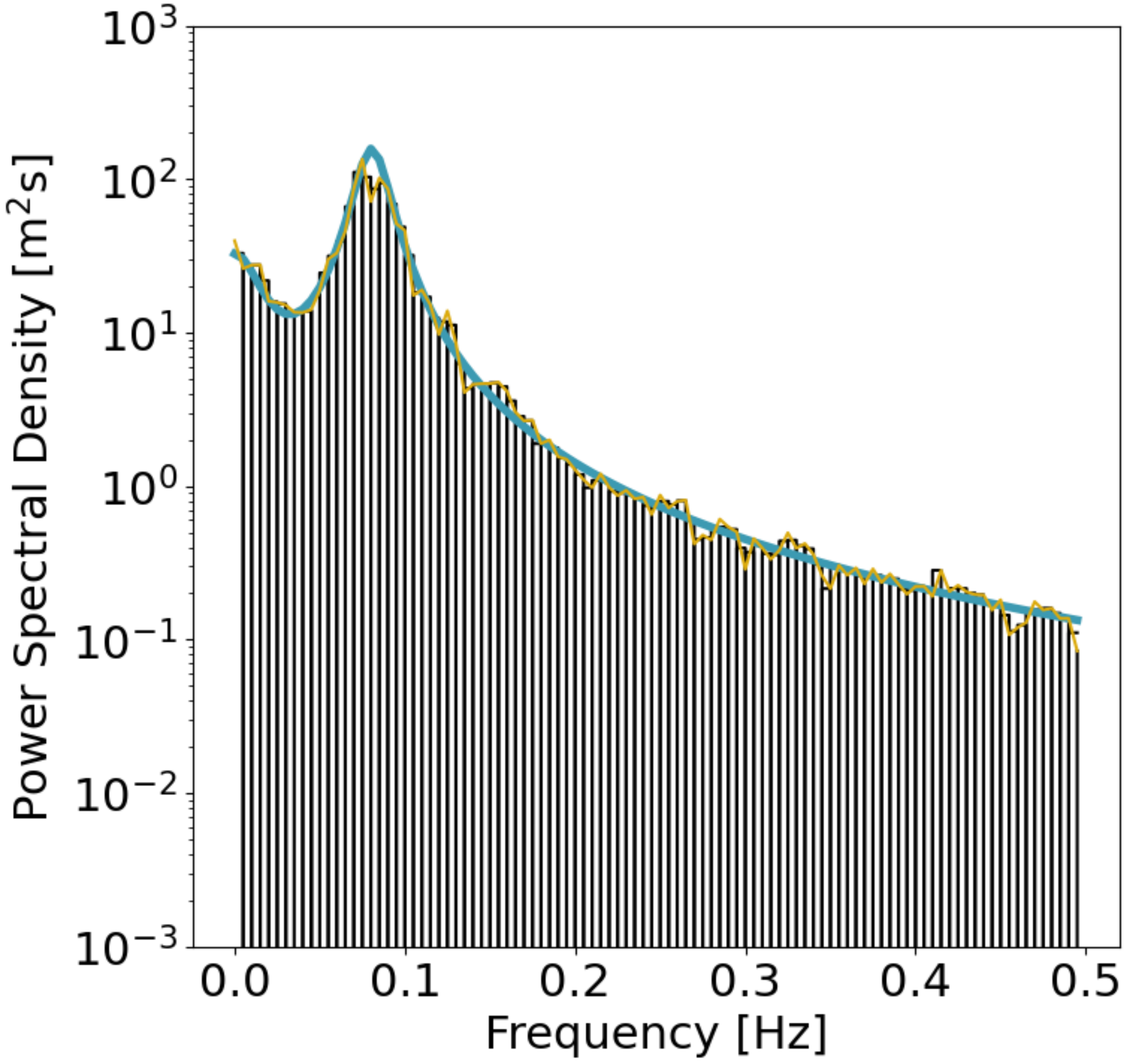
Riemann approximation to the PSD



- We can think about a periodogram as a Riemann approximation to the true biased PSD
- As we increase the resolution of the bases we converge on the true integral
- As we increase m we converge on the true biased PSD

Debiasing Welch's estimate

Riemann approximation to the PSD



We model our spectral density with the rectanglular basis

$$f(\omega) = \sum_i a_i B_i(\omega)$$

and solve similar to the parametric case.

(De)biased semi-parametric inference

Tobar, F. (2019). Band-limited Gaussian processes: The sinc kernel. *Advances in Neural Information Processing Systems*, *32*.

(De)biased semi-parametric inference

The corresponding ACF to each basis is given by

$$\rho_i(\tau) = \int_{-1/2}^{1/2} B_i(\omega)\, e^{i\omega\tau}\, \mathrm{d}\omega = \frac{\operatorname{sinc}(\tau\delta)\ \cos(\omega_i\tau)}{\delta}$$

Tobar, F. (2019). Band-limited Gaussian processes: The sinc kernel. *Advances in Neural Information Processing Systems*, *32*.

(De)biased semi-parametric inference

The corresponding ACF to each basis is given by

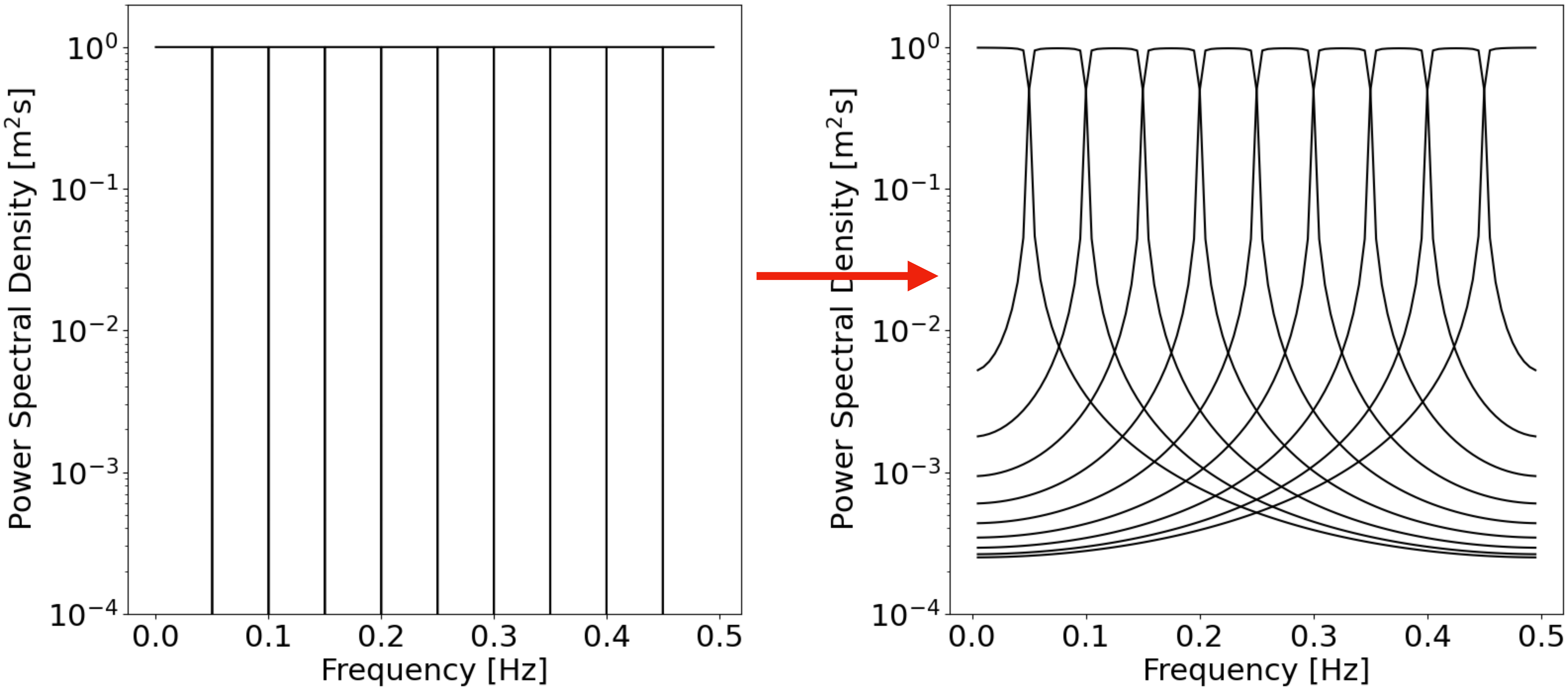
$$\rho_i(\tau) = \int_{-1/2}^{1/2} B_i(\omega)\, e^{i\omega\tau}\, \mathrm{d}\omega = \frac{\mathrm{sinc}(\tau\delta)\ \cos(\omega_i\tau)}{\delta}$$

The biased basis $\tilde{B}(\omega)$ is calculated similar to before

$$\tilde{B}_i(\omega) = 2 \times \mathrm{Re}\left\{ \sum_{\tau=0}^{n-1} \left(1 - \frac{\tau}{n}\right) \rho_i(\tau) e^{-i\omega\tau} \right\} - \phi_i(0)$$

Tobar, F. (2019). Band-limited Gaussian processes: The sinc kernel. *Advances in Neural Information Processing Systems*, *32*.

The biased bases



Computing the debiased Welch estimator

We are required to fit a large number of basis to data that we've established is non-Gaussian, so how?

Once we've established strong mixing we appeal to the central limit theorem and treat the data as Gaussian for a big enough m

$$\hat{\vartheta} = \arg\min_{\vartheta} \left\{ \text{var}\left[\bar{I}_L(\omega)\right]^{-1} \left(\bar{I}_L(\omega) - \vartheta \tilde{B}(\omega)\right)^2 \right\}$$

This term here is a problem as it's dependent on $f(\omega)$

Mathematical Intricacies

The main mathematical results of this work establish two results:

1. $\lim_{L \to \infty} \mathrm{var}[\bar{I}_L(\omega)] = c \ \mathrm{var}[I_L(\omega)]$, for c constant over ω
2. $\mathrm{var}[I_L(\omega)] = \bar{I}_L(\omega)^2 + \mathcal{O}_p\left(\frac{1}{m} + \frac{\log L}{L}\right)$

Mathematical Intricacies

The main mathematical results of this work establish two results:

1. $\lim_{L\to\infty} \mathrm{var}[\bar{I}_L(\omega)] = c\ \mathrm{var}[I_L(\omega)]$, for c constant over ω
2. $\mathrm{var}[I_L(\omega)] = \bar{I}_L(\omega)^2 + \mathcal{O}_p\left(\frac{1}{m} + \frac{\log L}{L}\right)$

$$\hat{\vartheta} = \arg\min_{\vartheta} \left\{ \mathrm{var}\left[\bar{I}_L(\omega)\right]^{-1} \left(\bar{I}_L(\omega) - \vartheta \check{B}(\omega)\right)^2 \right\}$$

Mathematical Intricacies

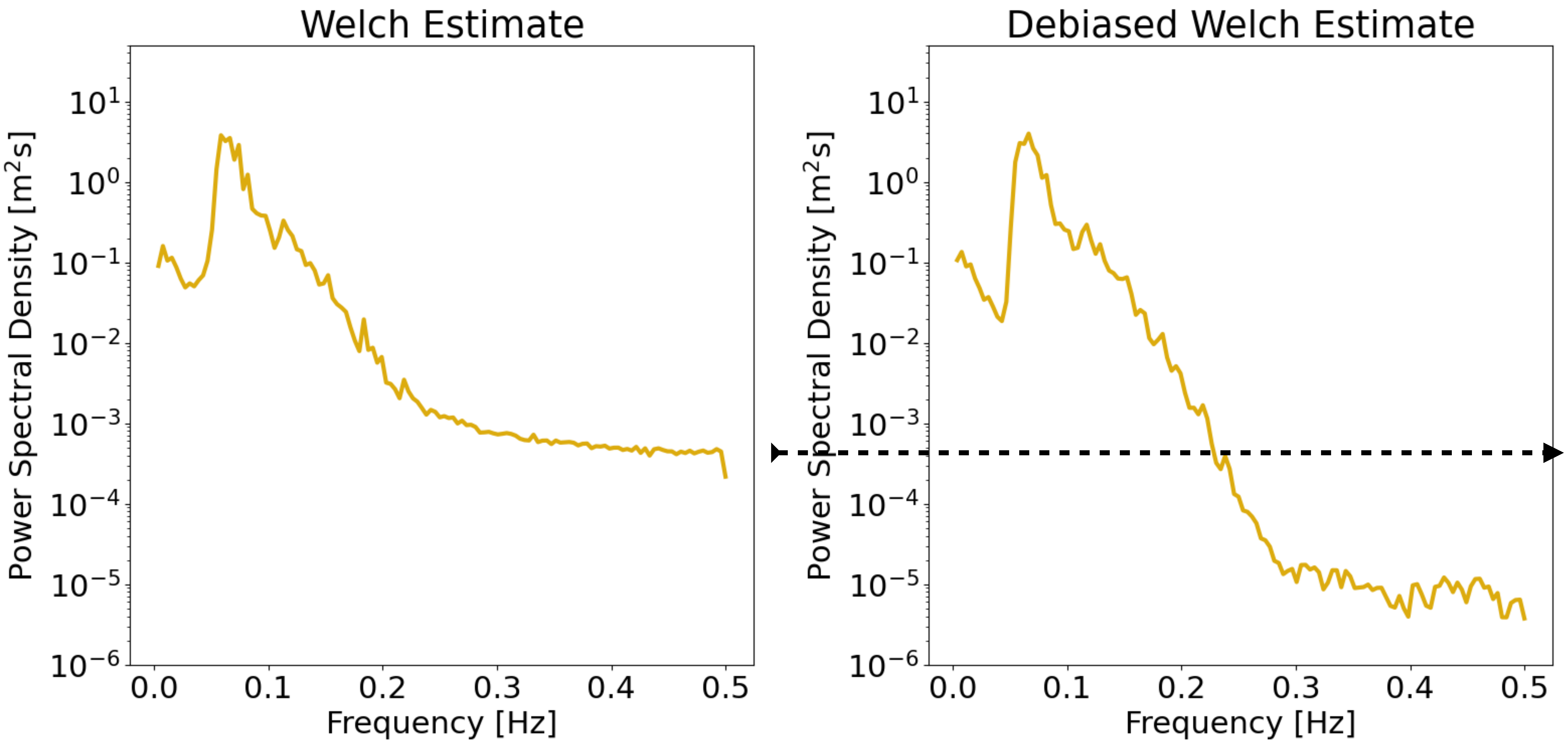
The main mathematical results of this work establish two results:

1. $\lim_{L\to\infty} \mathrm{var}[\bar{I}_L(\omega)] = c\ \mathrm{var}[I_L(\omega)]$, for c constant over ω
2. $\mathrm{var}[I_L(\omega)] = \bar{I}_L(\omega)^2 + \mathcal{O}_p\left(\frac{1}{m} + \frac{\log L}{L}\right)$

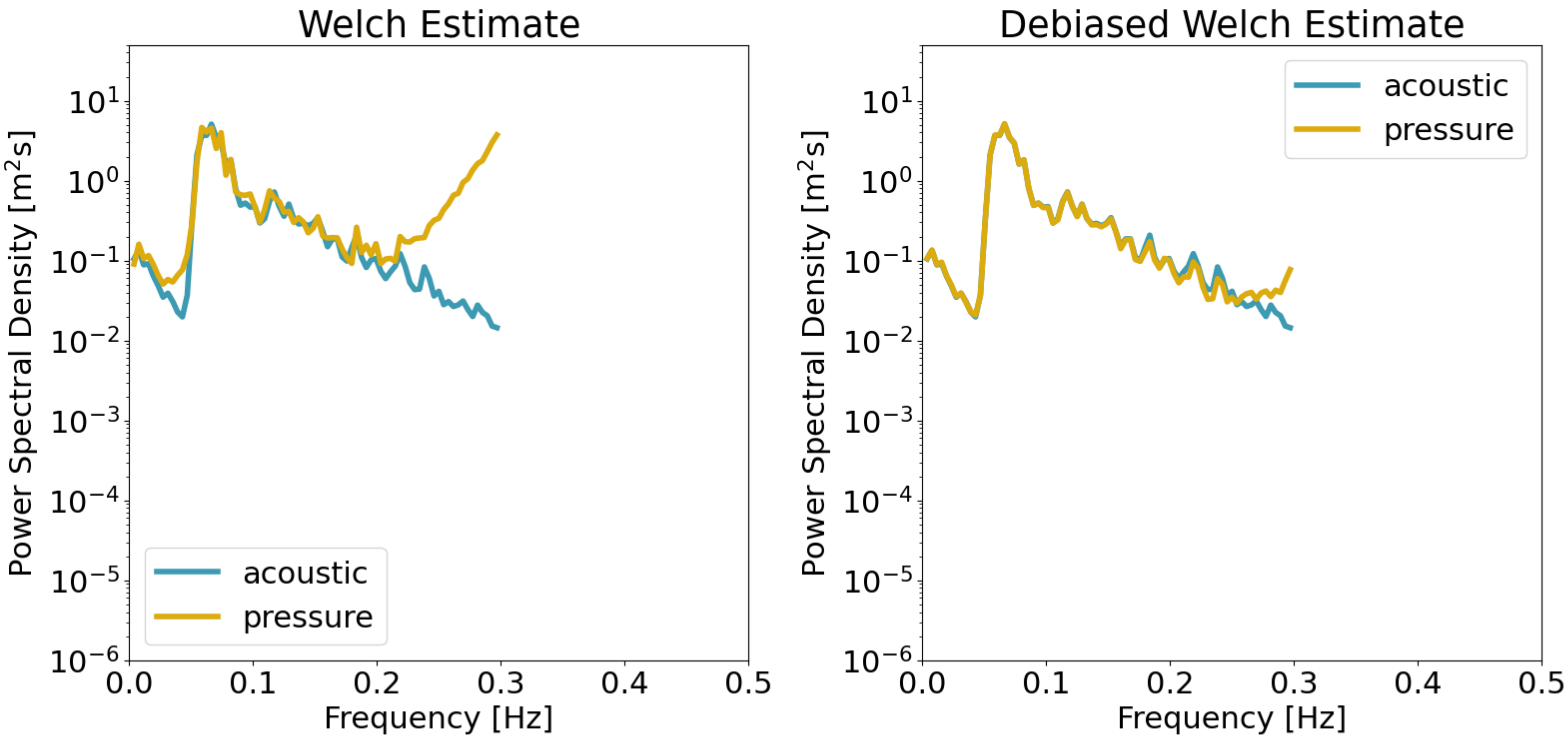
$$\hat{\vartheta} = \arg\min_{\vartheta} \left\{ \left[\bar{I}_L(\omega)\right]^{-2} \left(\bar{I}_L(\omega) - \vartheta \check{B}(\omega)\right)^2 \right\}$$

This solution is analytical!

Welch estimates

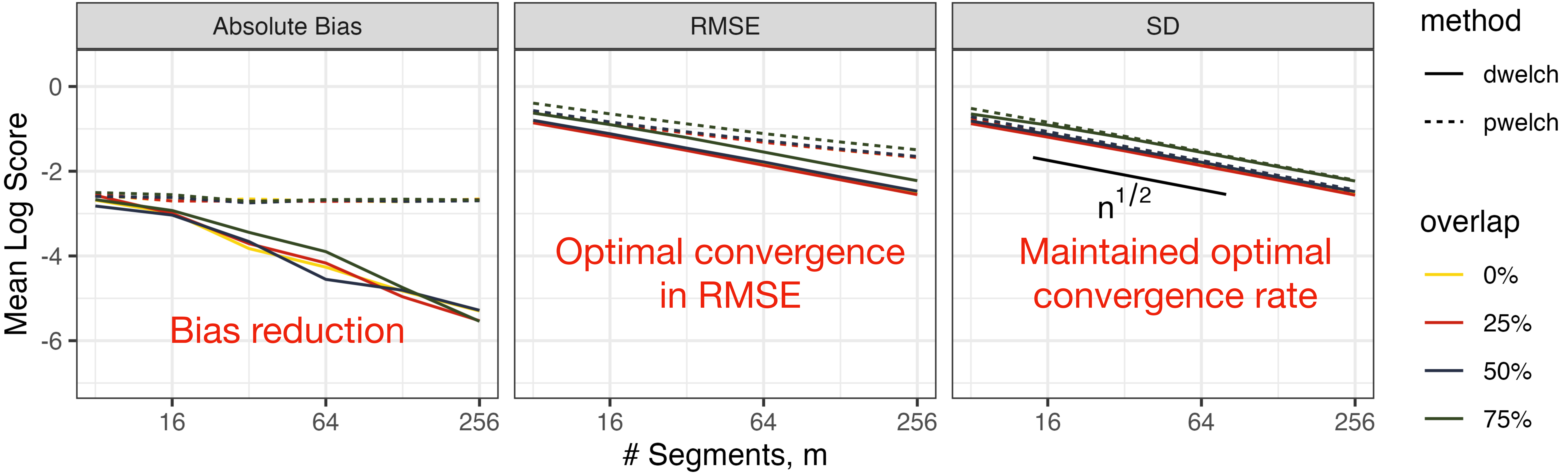


Reversing Attenuation



Performance over repeated simulations

Percival and Walden's AR(4) model



$$\text{MSE}\left[I_n(\omega)\right] = \left(\text{Bias}\left[I_n(\omega)\right]\right)^2 + \text{Var}\left[I_n(\omega)\right]$$

Some code in development

Non-Australians should look up
Gary Moorcroft Mark of the Year

GPTIDE

SPE
CCY

github.com/TIDE-ITRH

Some code in development

Non-Australians should look up
Gary Moorcroft Mark of the Year

GPTIDE

SPE
CCY

github.com/TIDE-ITRH

Thank you, and check out this research

- Kirch, C., Edwards, M. C., Meier, A., & Meyer, R. (2019). Beyond Whittle: Nonparametric correction of a parametric likelihood with a focus on Bayesian time series analysis.
- Lilly, J. M., Sykulski, A. M., Early, J. J., & Olhede, S. C. (2017). Fractional Brownian motion, the Matérn process, and stochastic modeling of turbulent dispersion. Nonlinear Processes in Geophysics, 24(3), 481-514.
- Percival, D. B., & Walden, A. T. (2020). Spectral analysis for univariate time series (Vol. 51). Cambridge University Press.
- Shumway, R. H., & Stoffer, D. S. (2000). Time series analysis and its applications (Vol. 3). New York: springer.
- Sykulski, A. M., Olhede, S. C., Guillaumin, A. P., Lilly, J. M., & Early, J. J. (2019). The debiased Whittle likelihood. Biometrika, 106(2), 251-266.
- Sykulski, A. M., Olhede, S. C., Lilly, J. M., & Danioux, E. (2016). Lagrangian time series models for ocean surface drifter trajectories. Journal of the Royal Statistical Society Series C: Applied Statistics, 65(1), 29-50.
- Tobar, F. (2019). Band-limited Gaussian processes: The sinc kernel. Advances in Neural Information Processing Systems, 32.
- Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics, 15(2), 70-73.

lachlan.astfalck@uwa.edu.au