

# **Spectrum and Autocovariance**



# Three Important Points

1. For any dataset you analyze, you should know the sample interval  $\Delta$ , number of points  $N$ , and duration  $T = N\Delta$  in physical meaningful units that you can relate to.
2. From these, you should know the Rayleigh frequency  $f^{\mathcal{R}} = 1/T$  and the Nyquist frequency  $f^{\mathcal{N}} = 1/(2\Delta)$ .
3. Be aware that there are two conventions for frequency, radian  $\omega$  and cyclic  $f$ , and note with care which you are using. Recall  $\omega = 2\pi f$ , as can be seen from  $\cos(\omega t)$  and  $\cos(2\pi f t)$ .



# The Basic Idea

The basic idea of spectral analysis is to break apart the variability of a time series into contributions from different frequencies.

Generally, this is done in the context of the assumption that the time series is a sample of some type of stochastic process. Thus, we need to review ideas such as the autocovariance function.

As will be seen, the autocovariance function and the spectrum are two equivalent quantities containing the essential second-order statistical information of a stochastic signal.

However, the spectrum tends to be more useful in practice, as well as easier to estimate.



# A Conceptual Model

We believe that our time series  $z_n$  is a *sample* of a real-world time series  $z(t)$  that, for simplicity, we presume to go on forever.

We can usefully imagine  $z(t)$  to be some type of random process.

This means that, while we *happened* to observe one series of values, we could just as readily have observed another series.

The abstract set of all possible times series that we *could* have observed is called the ensemble.

The easiest way to understand the ensemble is with respect to repeated runs of a numerical model.

Each run, or *realization*, of a numerical model (of the ocean, atmosphere, or climate) will tend to be different from the all others even if all model parameters remain unchanged.

This is because of the turbulent nature of the system, generating different results from the same initial conditions.



# Ensemble Averaging

Thus we believe that it is useful to think of there being a random process  $z(t)$  underlying our observations  $z_n$ .

In order to correctly describe the *statistical* behavior of this random process, we need a type of averaging called *ensemble averaging*, also known as the *expected value*.

The expected value, or ensemble average, is denoted “ $E\{\}$ ”, and represents an average over all the time series we *might* have observed, that is, over the ensemble.

The expected value or ensemble mean of  $z(t)$  is

$$\eta(t) \equiv E\{z(t)\}$$

which importantly can be a function of time! For example, in a greenhouse gas simulation, the globally averaged temperature, averaged over the model ensemble, will tend to increase.



# Types of Averaging

It is important to understand that the ensemble average is a very different kind of average from what we may be used to.

The ensemble average is, normally, not something you compute. Rather, it is a *theoretical tool* employed to help describe reality.

This average is needed because we have *assumed*, as a conceptual model, that what we have observed is some type of random process.

The more familiar type of mean that we forms from observed values of a discrete-time process  $z_n$  or a continuous-time process  $z(t)$

$$\bar{z}_n \equiv \frac{1}{N} \sum_{n=0}^{N-1} z_n, \quad \bar{z} \equiv \frac{1}{T} \int_{-T/2}^{T/2} z(t) dt$$

are called the *sample means*. These are not functions of time!



# The Autocovariance

In describing the statistical behavior of the random process  $z(t)$ , we seek to quantify not only its mean or *expected value* but also its typical spread about the mean. The *variance* is defined as

$$\sigma^2(t) \equiv \mathbb{E}\{|z(t) - \eta(t)|^2\}, \quad \eta(t) \equiv \mathbb{E}\{z(t)\}$$

which gives the expected squared deviation from the mean.

Again, unlike the *sample variance* of the observe time series  $z_n$ , the variance of  $z(t)$  involves an average over *realizations* and not over time. It is not something we can normally compute directly.

Similarly, the autocovariance function of  $z(t)$  is defined as

$$R(t, \tau) \equiv \mathbb{E}\{[z(t + \tau) - \eta(t + \tau)] [z(t) - \eta(t)]^*\}$$

The conjugation is there in case  $z(t)$  is complex-valued. If it is real-valued, the conjugation does nothing and can be ignored.

✧ Note, importantly, that at zero time offset we have  $R(t, 0) = \sigma^2(t)$ . 7 / 35

# Stationarity

The autocovariance is, in general, a function of *two* time variables:

$$R(t, \tau) \equiv \text{E}\{[z(t + \tau) - \eta(t + \tau)] [z(t) - \eta(t)]^*\}.$$

The first,  $t$ , indicates where we are in time. The second,  $\tau$ , is an *offset* that we use to compare what is happening now, at time  $t$ , with what will happen some time later.

A crucial simplification happens if our time series is such that its expected value and autocovariance function *do not change* with time  $t$ . In this case,  $\eta \equiv \text{E}\{z(t)\}$  and the autocovariance becomes

$$R(\tau) \equiv \text{E}\{[z(t + \tau) - \eta] [z(t) - \eta]^*\}$$

and the process is then said to be *second-order stationary* (because the autocovariance is a second-order statistical quantity).

It is common, and helpful, to *assume* stationarity even it is not strictly true. Or, rather, we decide we will treat our time series *as if it were* a sample of a stationary process. We do this henceforth.





# The Spectrum

The Fourier transform of the autocovariance function, denoted  $S(\omega)$  and called the spectrum, gives the Fourier coefficients decomposing the autocovariance into contributions from different frequencies,

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega, \quad S(\omega) \equiv \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau$$

where for the moment we take the right equation as a *definition*.

Since the spectrum is a Fourier transform pair with the autocovariance function, it contains equivalent information.

But the true autocovariance function is not observable unless we have (i) infinite time and (ii) access to an abstract set of other universes where things might have happened differently!

Spectral analysis is about trying to *estimate* the “true” spectrum  $S(\omega)$  from  $z_n$ , a finite sample of one realization of  $z(t)$ .



# Relationship to Variance

The spectrum has a fundamental relationship to the variance.

For a second-order stationary process, the mean, variance, autocovariance function are all independent of global time  $t$ :

$$\eta \equiv \mathbf{E}\{z(t)\}$$

$$\sigma^2 \equiv \mathbf{E}\{|z(t) - \eta|^2\}$$

$$R(\tau) \equiv \mathbf{E}\{[z(t + \tau) - \eta] [z(t) - \eta]^*\}$$

so that  $R(0) = \sigma^2$ , the autocovariance at zero time offset gives the variance (as we saw in the nonstationary case). But this means

$$\sigma^2 = R(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega 0} S(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega$$

so that, importantly, the spectrum can be said to *partition the variance* into contributions from different frequencies.



# The Spectrum (Take One)

The spectrum is for the moment *defined* as the Fourier transform of the autocovariance function

$$R(\tau) \equiv \mathbf{E}\{[z(t + \tau) - \eta] [z(t) - \eta]^*\}$$

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega, \quad S(\omega) \equiv \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau$$

Note that (iv) implies as a consequence the important formula

$$\sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega$$

which states that the spectrum partitions the variance across frequency.



# Spectral Estimates

Recall from yesterday the Fourier pair relationships:

$$z_n = \frac{1}{N} \sum_{m=0}^{N-1} Z_m e^{i2\pi mn/N}, \quad Z_m \equiv \sum_{n=0}^{N-1} z_n e^{-i2\pi mn/N}.$$

It can be shown (but not now) that the spectrum can be estimated simply by taking the modulus squared of the Fourier coefficients,

$$\hat{S}_m \equiv \frac{1}{N} |Z_m|^2, \quad n = 0, 1, 2, \dots, (N-1).$$

This quantity is known as the *periodogram*.

As we shall see, the periodogram is *not* the spectrum! It is an *estimate* of the spectrum—and generally speaking, a very poor one.

It is also said to be the *naive* spectral estimate, meaning it is the spectral estimate that you get if you don't know that there is something better.



# Three Problems

The periodogram estimate of the spectrum suffers from three problems:

1. *Aliasing*. This is due to the fact that a continuous-time process is *sampled*,  $z_n \equiv z(n\Delta)$ .
2. *Blurring*. Values of the spectrum bleed out to affect other frequencies. We will understand this in a later lecture.
3. *Variance*. This is due to the randomness of the underlying process  $z(t)$ .

The problem of aliasing one is more or less stuck with, but the other two problems can be greatly improved by a more sophisticated choice of spectral estimate.

To minimize aliasing, you should make sure your experiment has an adequate sample rate!

Blurring is a particularly difficult problem because it doesn't reduce with averaging. This problem is also called *bias* or *leakage*.



# Aliasing

Imagine that our time series extends at discrete times into the infinite past and future, that is, we have  $z_n = z(n\Delta)$  for *all* positive and negative integers  $n$ , but not in between those times.

In this case, it turns out that the spectrum can still be considered a function of continuous frequency  $\omega$ , but *only* up to the Nyquist at  $\omega = \pm\pi/\Delta$ . (Recall that  $f = 1/(2\Delta)$  is the same as  $\omega = \pi/\Delta$ .)

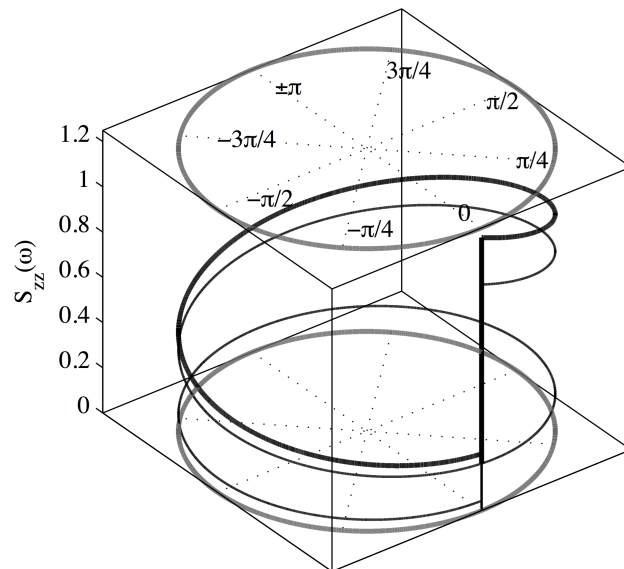
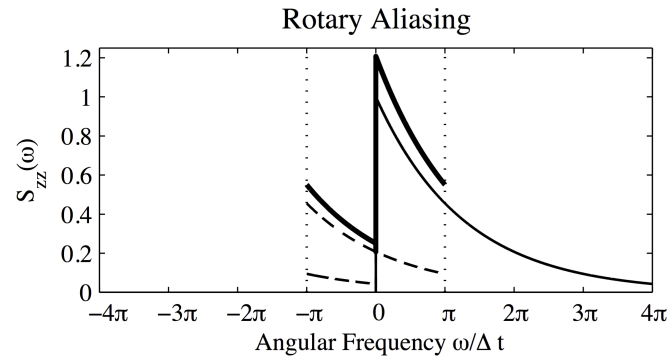
It can be shown (but not right now) that the spectrum then becomes

$$\sum_{p=-\infty}^{\infty} S\left(\omega + p\frac{2\pi}{\Delta}\right) \equiv S^{\Delta}(\omega)$$

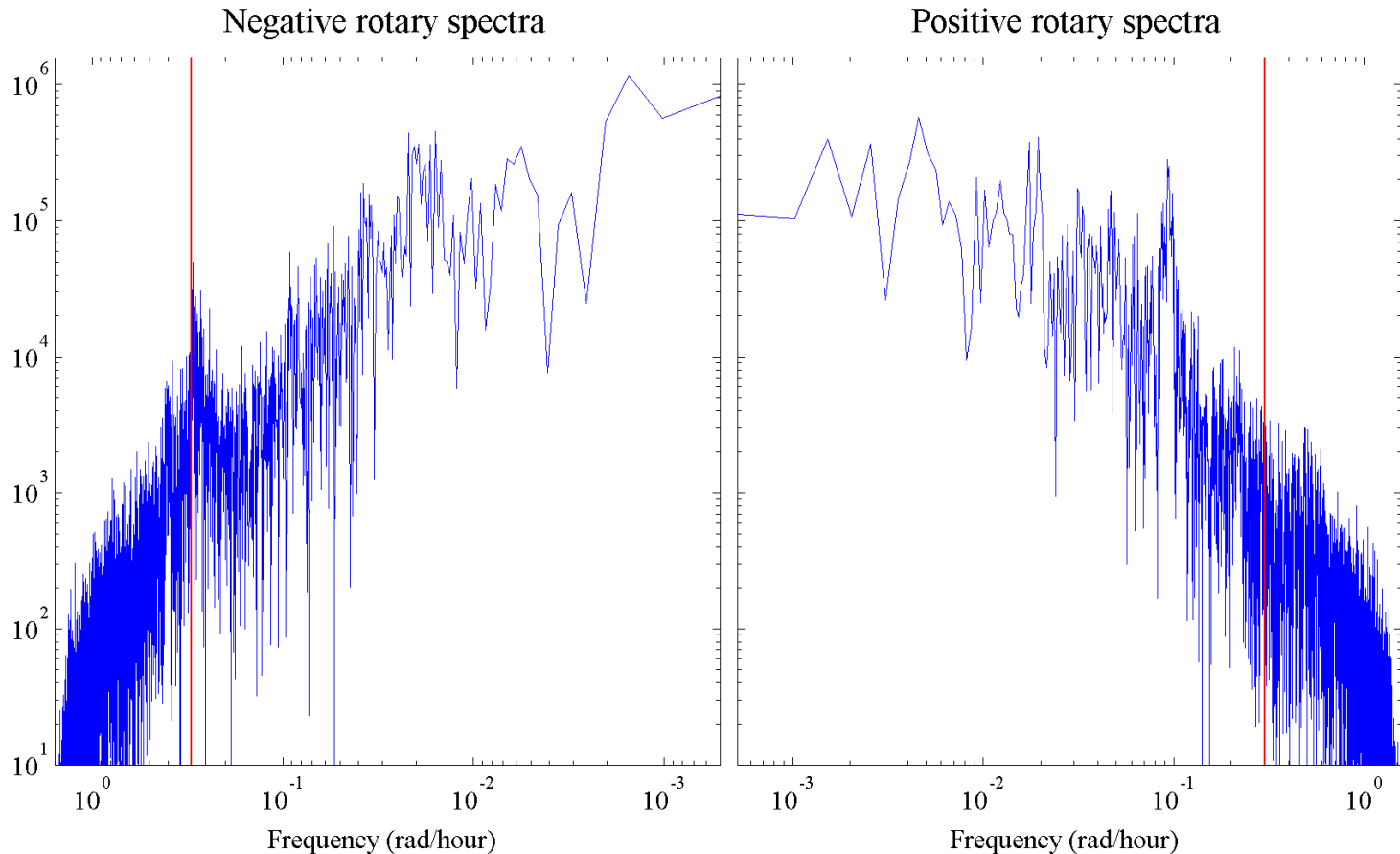
which states that all frequencies higher or lower than  $\omega = \pm\pi/\Delta$  are *wrapped around* to the next resolved frequency.



# Aliasing Illustration



# Variance Illustration

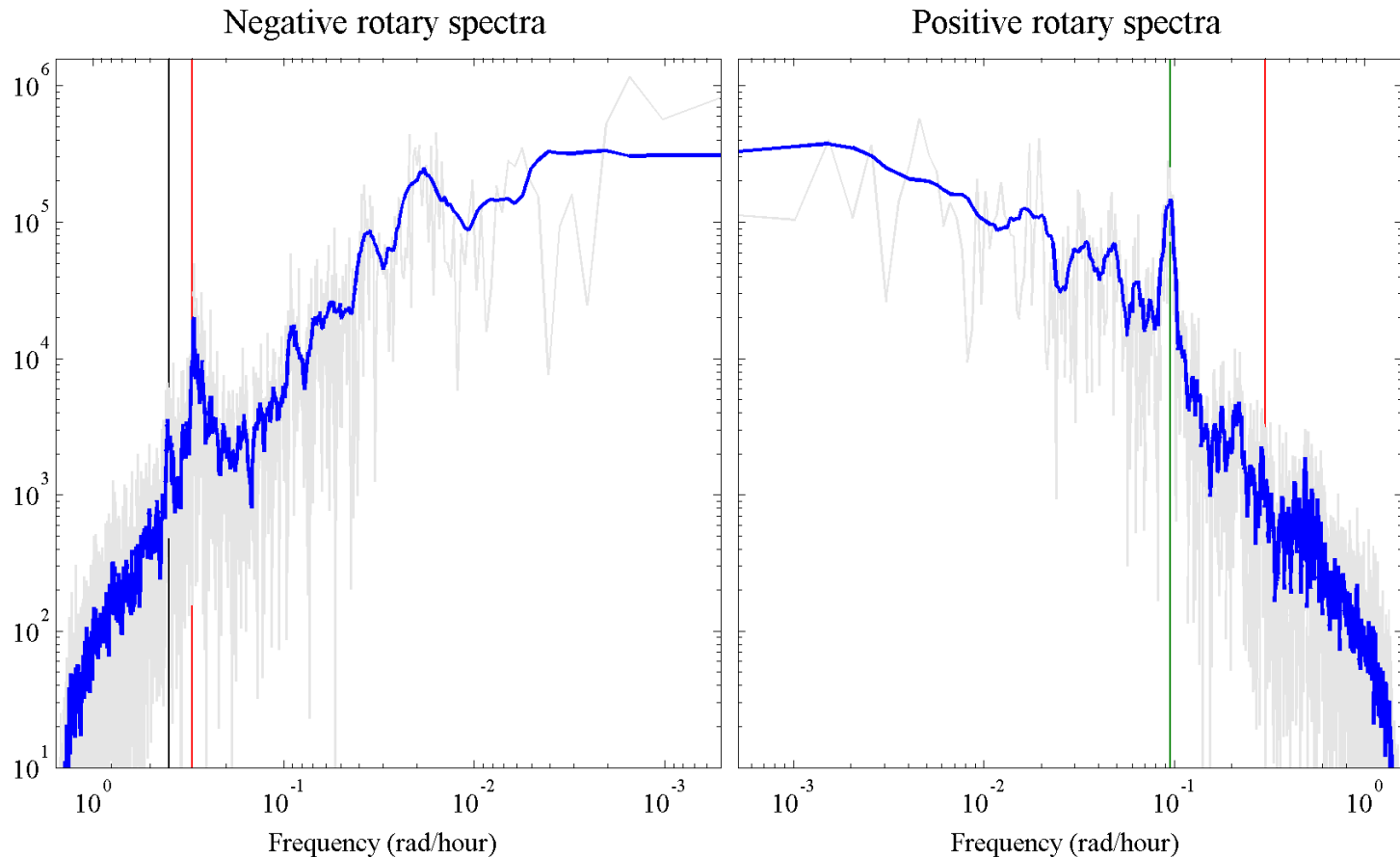


The main problem here is variance of the periodogram spectral estimate. Most meaningful structure in the spectrum is obscured.





# Reducing Variance

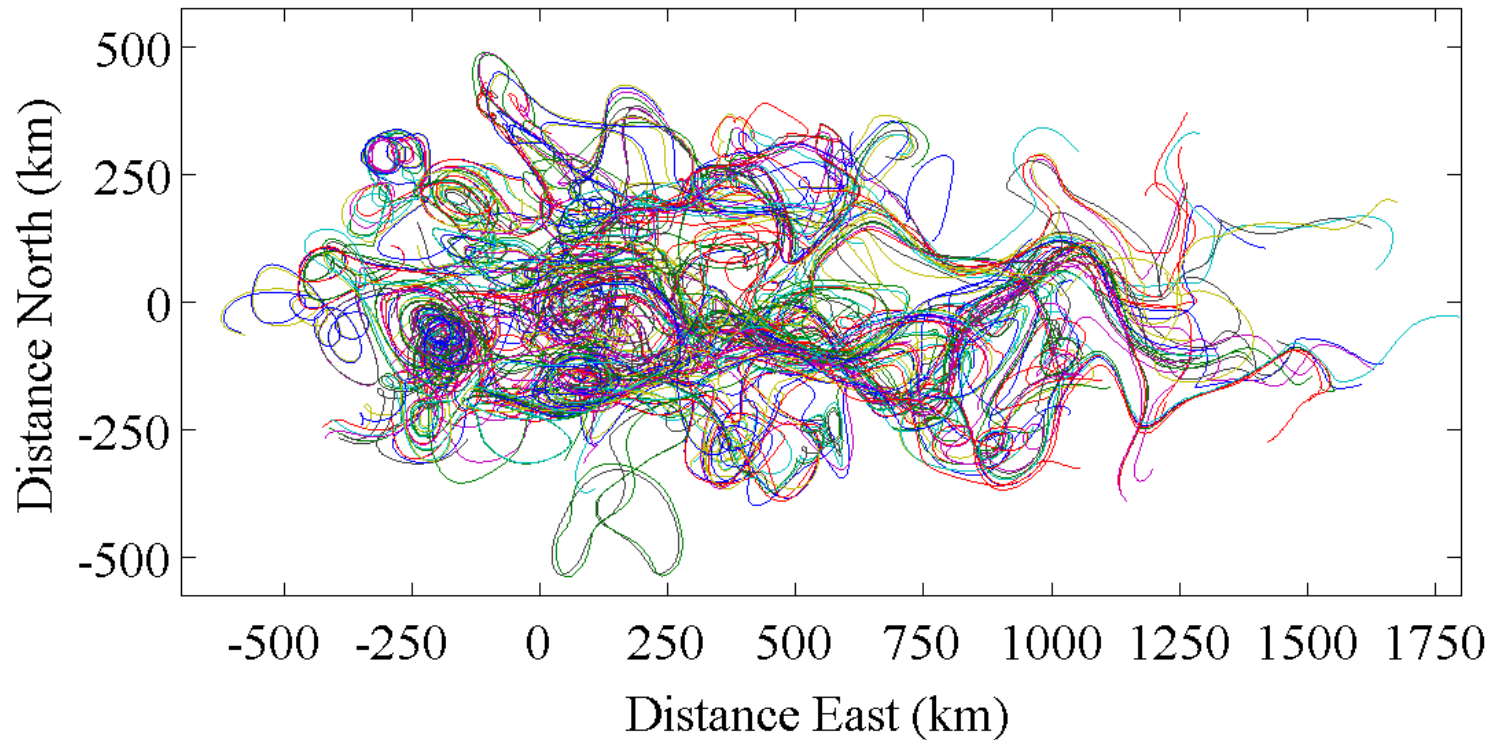


Variance can be mitigated by a spectral estimate that incorporates some type of *frequency-domain averaging*, as is the case here.



# Model Trajectories

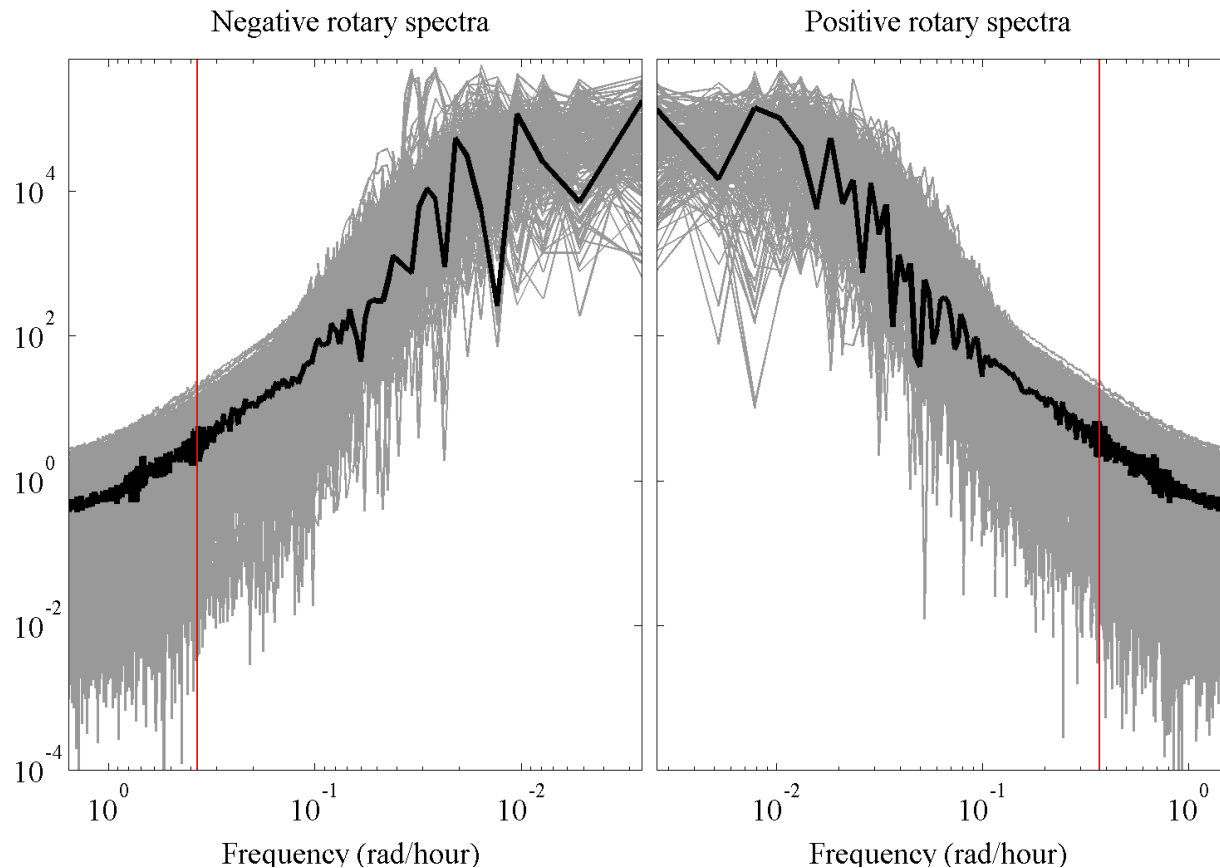
Lagrangian Trajectories in a Model of Baroclinic Instability



Particle trajectories in a model of baroclinic turbulence on a  $\beta$  plane, from a simulation by E. Danioux.



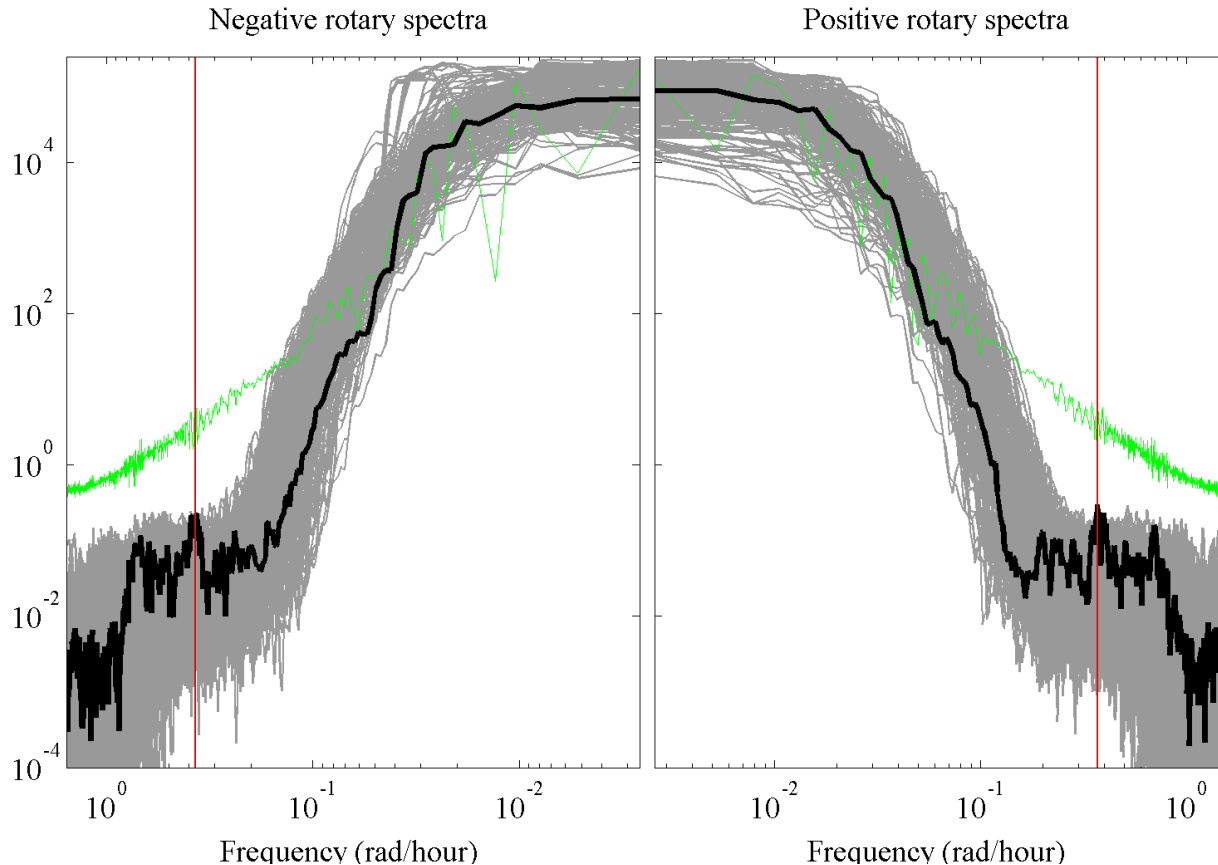
# The Periodogram



Rotary spectra estimated using the periodogram. Red line is a characteristic frequency that we can ignore.



# Blurring Illustration



Periodogram vs. adaptive multitaper estimate. We will see that the periodogram is dominated by an artifact called *spectral blurring*.



# The Spectrum (Take Two)

We have said that the spectrum is the Fourier transform of the autocovariance. But what is the spectrum itself?

We know the spectrum is somehow about squared frequency-domain coefficients, and we have said  $S_m \equiv \frac{1}{N} |Z_m|^2$  is an *estimate* of the spectrum called the periodogram.



# The Spectrum (Take Two)

We have said that the spectrum is the Fourier transform of the autocovariance. But what is the spectrum itself?

We know the spectrum is somehow about squared frequency-domain coefficients, and we have said  $S_m \equiv \frac{1}{N} |Z_m|^2$  is an *estimate* of the spectrum called the periodogram.

We would like to just write  $z(t)$  in terms of its Fourier transform

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega) e^{i\omega t} d\omega, \quad Z(\omega) \equiv \int_{-\infty}^{\infty} z(t) e^{-i\omega t} dt$$

and then define the spectrum as  $S(\omega) \equiv E \left\{ |Z(\omega)|^2 \right\}$ .

The spectrum is indeed, essentially, the expected squared magnitude of a Fourier-domain coefficient.

✧ But  $S(\omega) \equiv E \left\{ |Z(\omega)|^2 \right\}$  is not quite right because the above equations have a mathematical difficulty that we must circumvent

# Defining the Spectrum

Earlier in this lecture we defined the spectrum as the Fourier transform of the autocovariance function.

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega, \quad S(\omega) \equiv \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau$$

This is unsatisfactory because we know that the spectrum is related to the square of the Fourier transform of a time series.

However, we intentionally defined it in this way in order to sidestep a technical difficulty, which we now address.



# Continuous Time Fourier

For processes defined on continuous time  $z(t)$ , we need to use the continuous-time, continuous frequency Fourier transform equations

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega) e^{i\omega t} d\omega, \quad Z(\omega) \equiv \int_{-\infty}^{\infty} z(t) e^{-i\omega t} dt$$

These differ from the discrete Fourier equations because *both*  $z(t)$  and  $Z(\omega)$  have values everywhere, not only at discrete intervals and perhaps not only over a certain time or frequency interval.

For these equations to be valid, it is necessary for  $z(t)$  and  $Z(\omega)$  to satisfy some conditions. If both functions are absolutely integrable,

$$\int_{-\infty}^{\infty} |z(t)| dt < \infty, \quad \int_{-\infty}^{\infty} |Z(\omega)| d\omega < \infty$$

then it can be shown that the Fourier transform equations hold.





# A Wrinkle in Time Domain

When dealing with time series  $z(t)$  that we consider as stochastic processes, we run into a wrinkle when we try to represent them in the Fourier domain.

Our time series  $z(t)$  will be absolutely integrable if it is, say, a Gaussian shape, or a boxcar, or in general any time-localized signal with bounded (non-infinite) amplitude.

This is a big problem with stochastic processes! We want to think of them as having a Fourier transform, but most processes of interest—such as white noise—will not be absolutely integrable.

This means the Fourier transform

$$Z(\omega) \neq \int_{-\infty}^{\infty} z(t) e^{-i\omega t} dt$$

is not even valid when  $z(t)$  is a stochastic process, because the integral cannot be associated with a specific finite value.



# A Wrinkle in Time Domain

If you remember from day one, we said that we believed the observed data represents *samples* of some real-world process,

$$z_n \equiv z(\Delta n), \quad n = 0, 1, \dots, N - 1$$

where we can conceptualize  $z(t)$  as a stochastic process that goes on forever in both directions.



# A Wrinkle in Time Domain

If you remember from day one, we said that we believed the observed data represents *samples* of some real-world process,

$$z_n \equiv z(\Delta n), \quad n = 0, 1, \dots, N - 1$$

where we can conceptualize  $z(t)$  as a stochastic process that goes on forever in both directions.

But because they are not absolutely integrable, **stochastic processes do not have Fourier transforms in the usual sense!**



# Unwrinkle the Wrinkle

Loosely speaking, the problem is that  $Z(\omega)$  would need to be infinite.

There is a way around this, in which we replace the usual inverse Fourier transform with a relative:

(deterministic)

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega) e^{i\omega t} d\omega$$

(stochastic)

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

The crucial equation on the right is called the *Cramér spectral representation theorem*.

Note a seemingly small change:  $Z(\omega)d\omega$  is replaced with  $dZ(\omega)$ .

This lets us think of the stochastic process  $z(t)$  as being, *in a generalized sense*, the Fourier transform of a frequency-domain process  $dZ(\omega)$ , known as the *increment process*.



# The Cramér Theorem

Let's compare the two equations

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega) e^{i\omega t} d\omega, \quad z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

The equation on the left involves deterministic quantities, whereas the equation on the right involves stochastic quantities.

More importantly, however, the very definition of an integral has changed! On the left is the usual Riemann integral defined as the limit of a sum as the step size goes to zero. On the right we have something somewhere different called a Riemann–Stieltjes integral involving a different limiting process.

The essential idea is that we group the potentially non-finite value of  $Z(\omega)$  together with an infinitesimal  $d\omega$  in the integrand. This combination makes the integral manageable again.



# The Cramér Theorem

Fortunately, we rarely (probably never) need to worry about the technical details of the integration.

The main point is that there is the *idea* of an inverse Fourier transform that works for stochastic processes, and that is called the Cramér spectral representation theorem.

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

Like it relative for deterministic processes, this states again that  $z(t)$  can be built up from different oscillatory components.



# The Cramér Theorem

Fortunately, we rarely (probably never) need to worry about the technical details of the integration.

The main point is that there is the *idea* of an inverse Fourier transform that works for stochastic processes, and that is called the Cramér spectral representation theorem.

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

Like it relative for deterministic processes, this states again that  $z(t)$  can be built up from different oscillatory components.

Note, for stochastic processes, we don't have a forward Fourier transform equation in the usual sense!

$$dZ(\omega) \neq \int_{-\infty}^{\infty} e^{-i\omega t} z(t) dt$$



# Stationarity Implications

For the proof, we substitute the Cramér spectral representation theorem for  $z(t)$  in the definition of the autocovariance, leading to

$$\begin{aligned} R(t, \tau) &\equiv \mathbb{E}\{z(t + \tau) z^*(t)\} \\ &= \mathbb{E} \left\{ \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega(t+\tau)} dZ(\omega) \right] \left[ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\nu t} dZ^*(\nu) \right] \right\} \\ &= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(\omega-\nu)t+i\omega\tau} \mathbb{E} \{dZ(\omega) dZ^*(\nu)\} . \end{aligned}$$

In general, the right hand side is a function of both  $t$  and  $\tau$ .

If we want to enforce the condition that the time series is stationary, we need  $R(\tau)$ . This *requires*  $\mathbb{E} \{dZ(\omega) dZ^*(\nu)\}$  to vanish or  $\omega \neq \nu$ !

This fact means *we must have*, for some function  $S(\omega)$

$$S(\omega)\delta(\omega - \nu) d\omega d\nu \equiv \frac{1}{2\pi} \mathbb{E} \{dZ(\omega) dZ^*(\nu)\} .$$





# Stationarity Implications

At the moment,  $S(\omega)$  is simply the coefficient of the delta function that we need to guarantee stationarity

$$S(\omega)\delta(\omega - \nu) d\omega d\nu \equiv \frac{1}{2\pi} \mathbb{E} \{dZ(\omega)dZ^*(\nu)\}.$$

Inserting this into the expression for the autocovariance, we find

$$\begin{aligned} R(t, \tau) &\equiv \mathbb{E}\{z(t + \tau) z^*(t)\} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(\omega - \nu)t + i\omega\tau} S(\omega)\delta(\omega - \nu) d\omega d\nu \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega \end{aligned}$$

using the fundamental property of a delta function. Thus we see that this choice has eliminated time on the right-hand side, and thus now the autocovariance is now  $R(\tau)$ .



# Defining the Spectrum

For a stationary random process  $z(t)$ , we can now re-define the spectrum as

$$S(\omega)\delta(\omega - \nu) d\omega d\nu \equiv \frac{1}{2\pi} \mathbb{E} \{dZ(\omega)dZ^*(\nu)\}$$

directly in terms of  $dZ(\omega)$ , the frequency-domain increment process generating  $z(t)$  via the Cramér spectral representation theorem.

This equation admittedly looks bizarre at first, with a delta function as a part of the definition, as well as a  $d\omega$  that we normally only see inside of an integral.

However, its intuitive content is that for  $\omega = \nu$ , we have the value of  $S(\omega)$  essentially being set to  $\frac{1}{2\pi} \mathbb{E} \left\{ |dZ(\omega)|^2 \right\}$ , plus extra maths.

Now we begin to see why the discrete Fourier spectrum can be estimated by the periodogram,  $\hat{S}_m \equiv \frac{1}{N} |Z_m|^2$ .



# Spectrum and Autocovariance

It follows from this new definition of the spectrum that

$$S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau, \quad R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega$$

where the previous “ $\equiv$ ” sign on the left is now an equals sign.

This is another key foundation of spectral analysis. If one defines the autocovariance  $R(\tau)$  and spectrum  $S(\omega)$ , respectively, as

$$R(\tau) \equiv \mathbb{E}\{z(t + \tau) z^*(t)\}, \quad S(\omega)\delta(\omega - \nu) d\omega d\nu \equiv \frac{1}{2\pi} \mathbb{E}\{dZ(\omega)dZ^*(\nu)\}$$

then **the spectrum and the autocovariance form a Fourier transform pair.**

We can write this in a shorthand notation as

$$R(\tau) \iff S(\omega).$$



# Summary, Part I

(i) The sampling model

$$z_n \equiv z(\Delta n), \quad n = 0, 1, \dots, N-1$$

(ii) Euler's formula

$$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t)$$

(iii) The discrete Fourier transform

$$z_n = \frac{1}{N} \sum_{m=0}^{N-1} Z_m e^{i2\pi nm/N}, \quad Z_m \equiv \sum_{n=0}^{N-1} z_n e^{-i2\pi nm/N}$$

(iv) The continuous Fourier transform

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega) e^{i\omega t} d\omega, \quad Z(\omega) \equiv \int_{-\infty}^{\infty} z(t) e^{-i\omega t} dt$$



# Summary, Part II

(v) The wavelet transform

$$w(t, s) = \int_{-\infty}^{\infty} \frac{1}{s} \psi^* \left( \frac{t - \tau}{s} \right) z(\tau) d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Psi^*(s\omega) Z(\omega) e^{i\omega t} d\omega$$

(vi) The Cramér spectral representation of a stochastic process

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

(vii) The spectrum and autocovariance are a Fourier transform pair

$$R(\tau) \equiv \mathbb{E}\{[z(t + \tau) - \eta] [z(t) - \eta]^*\}$$

$$S(\omega) \delta(\omega - \nu) d\omega d\nu \equiv \frac{1}{2\pi} \mathbb{E} \{dZ(\omega) dZ^*(\nu)\}$$

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega\tau} S(\omega) d\omega, \quad S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau$$



# Conclusion

The spectrum is the fundamental second-order description of a stochastic time series.

Loosely speaking, it's the expected value of the square of the Fourier coefficients. But defining this precisely is a bit of mathematical challenge.

There are three challenges in estimating it, the details of which we'll understand shortly.

