AUTOCORRELATION, PARTIAL AUTOCORRELATION, AND THE SPECTRUM

James Doss-Gollin

CEVE 543

2026-09-02

OBJECTIVES

By the end of today, you will be able to:

1. Define the autocorrelation and partial autocorrelation functions
2. Explain how the spectrum relates to the autocorrelation function
3. Identify red noise, white noise, and a periodic component from a spectrum
4. Read a spectrogram

Autocorrelation

Partial autocorrelation

Second order

The spectrum

Some examples

Real data

DEFINITION

The **autocorrelation at lag** h is the correlation between X_t and X_{t-h}:

$$\rho_h = \frac{\text{Cov}(X_t, X_{t-h})}{\text{Var}(X_t)}$$

Assuming stationarity, ρ_h depends on h alone.

$\rho_0 = 1$, and $\rho_h \in [-1, 1]$.

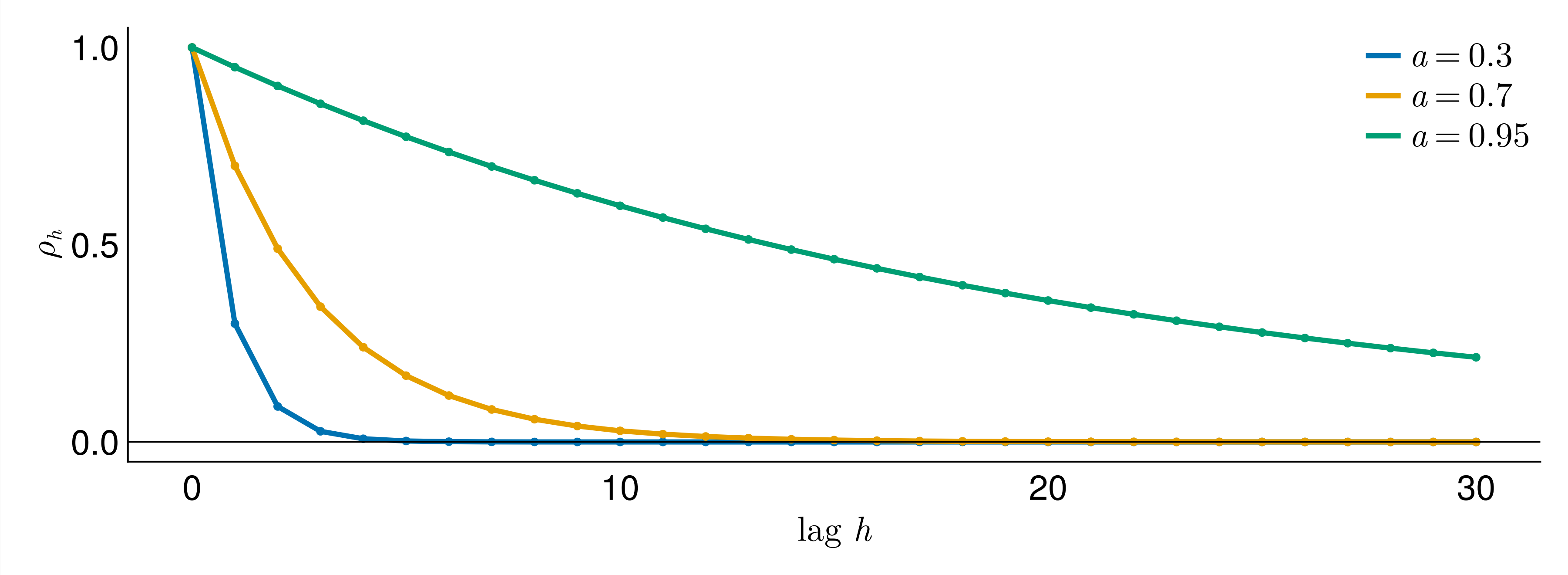
The AR(1) Model

$$x_t = a\,x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \qquad |a| < 1$$

The autocorrelation is

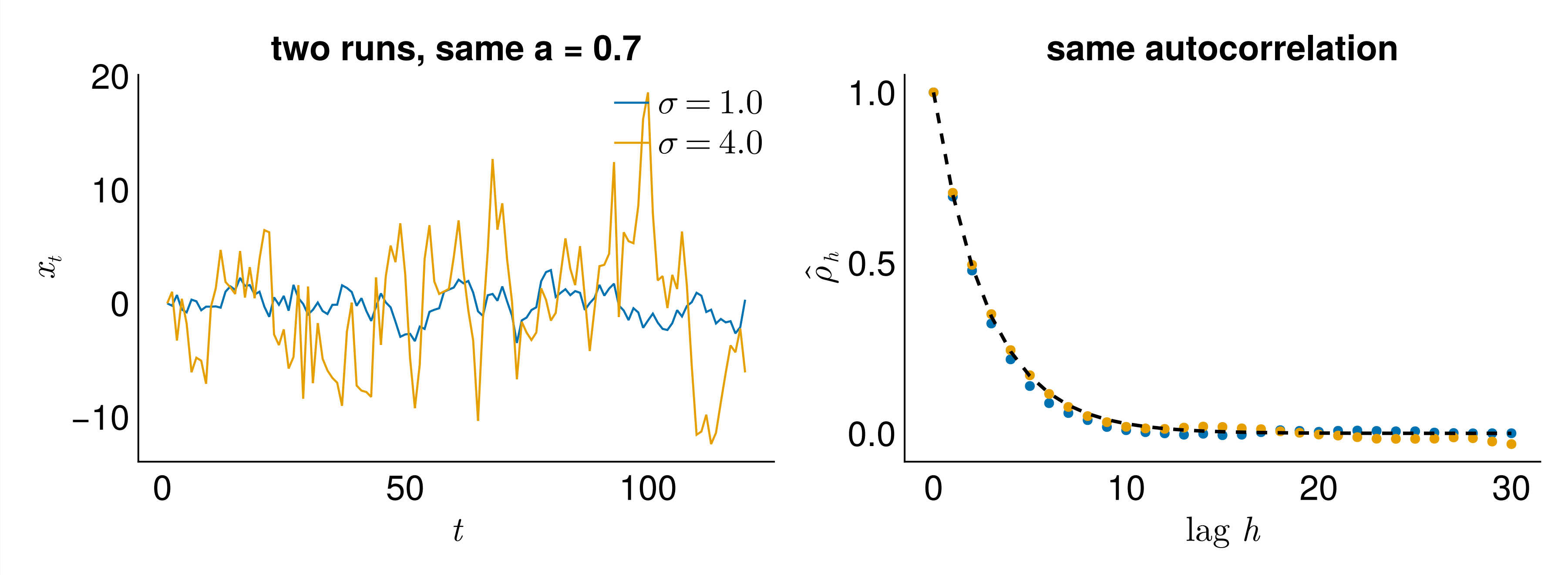
$$\rho_h = a^{|h|}$$

HOW THE DECAY DEPENDS ON a




```
a = 0.30: rho_10 = 0.0000
a = 0.70: rho_10 = 0.0282
a = 0.95: rho_10 = 0.5987
```

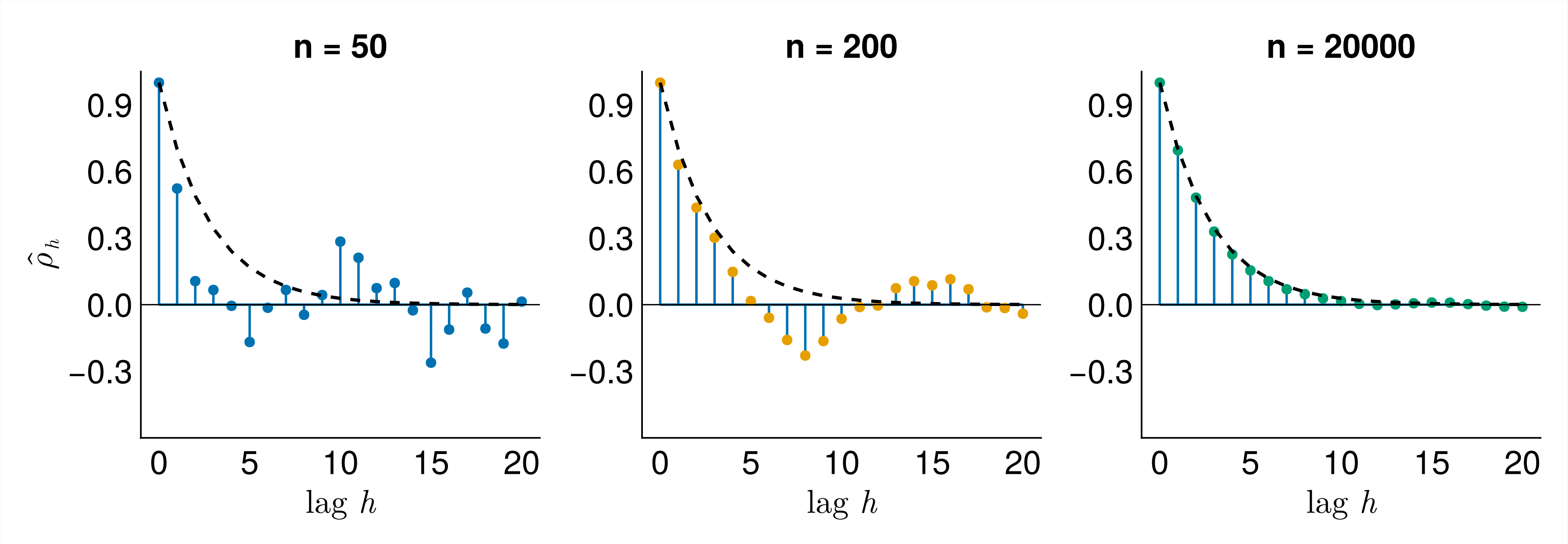
σ DOES NOT AFFECT THE AUTOCORRELATION



Correlation is scale-free, so a shows up in ρ_h; σ does not.

WHAT A SHORT RECORD CAN TELL YOU

The formula $\rho_h = a^{|h|}$ describes the model. An estimate $\hat{\rho}_h$ from n observations is subject to *sampling variability*:



Autocorrelation

Partial autocorrelation

Second order

The spectrum

Some examples

Real data

DEFINITION

An AR(1) has $\rho_2 = a^2 \neq 0$, but x_{t-2} does not appear in the model. The lag-2 correlation exists because x_{t-2} affects x_{t-1}, which affects x_t.

The **partial autocorrelation at lag** h removes the contribution of intervening lags:

$$\phi_h = \text{coefficient on } \; x_{t-h} \; \text{ in a regression of } \; x_t \; \text{ on } \; x_{t-1}, \ldots, x_{t-h}$$

PACF of an AR(1)

$\phi_1 = a$, because regressing x_t on x_{t-1} alone recovers the AR(1) coefficient.

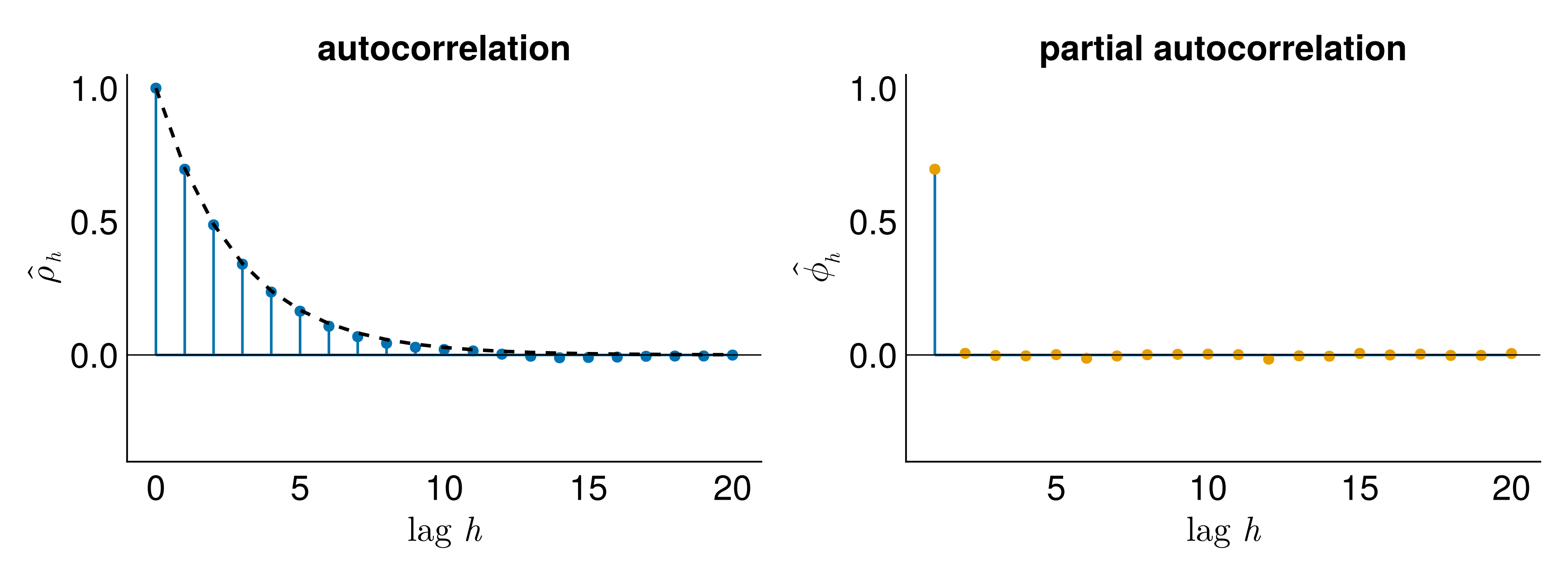
Adding x_{t-2} to the regression adds no predictive power, so $\phi_h = 0$ for $h > 1$.

$$\phi_1 = a, \qquad \phi_h = 0 \text{ for } h > 1$$

For an AR(p), $\phi_h = 0$ for $h > p$.

The autocorrelation decays gradually; the partial autocorrelation cuts off, and the cutoff identifies the order.

AR(1), BOTH DIAGNOSTICS



Autocorrelation

Partial autocorrelation

Second order

The spectrum

Some examples

Real data

WHAT DOES A SECOND LAG BUY?

An AR(1) can only decay: today's value drifts back toward zero, and the rate depends on a.

Adding a second lag lets the process look back two steps:

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \varepsilon_t$$

The character of the solution depends on a_1 and a_2:

- **Decay** (both roots real and inside the unit circle)
- **Oscillation** (complex roots inside the unit circle)
- **Blow up** (a root outside the unit circle, the model is unstable)

The roots of $z^2 - a_1 z - a_2 = 0$ are complex when $a_1^2 + 4a_2 < 0$.

Our AR(2) example

Take $a_1 = 1.2$, $a_2 = -0.7$.

$$a_1^2 + 4a_2 = 1.44 - 2.8 = -1.36 < 0$$

Complex roots, so it oscillates: damped, and never repeating exactly.

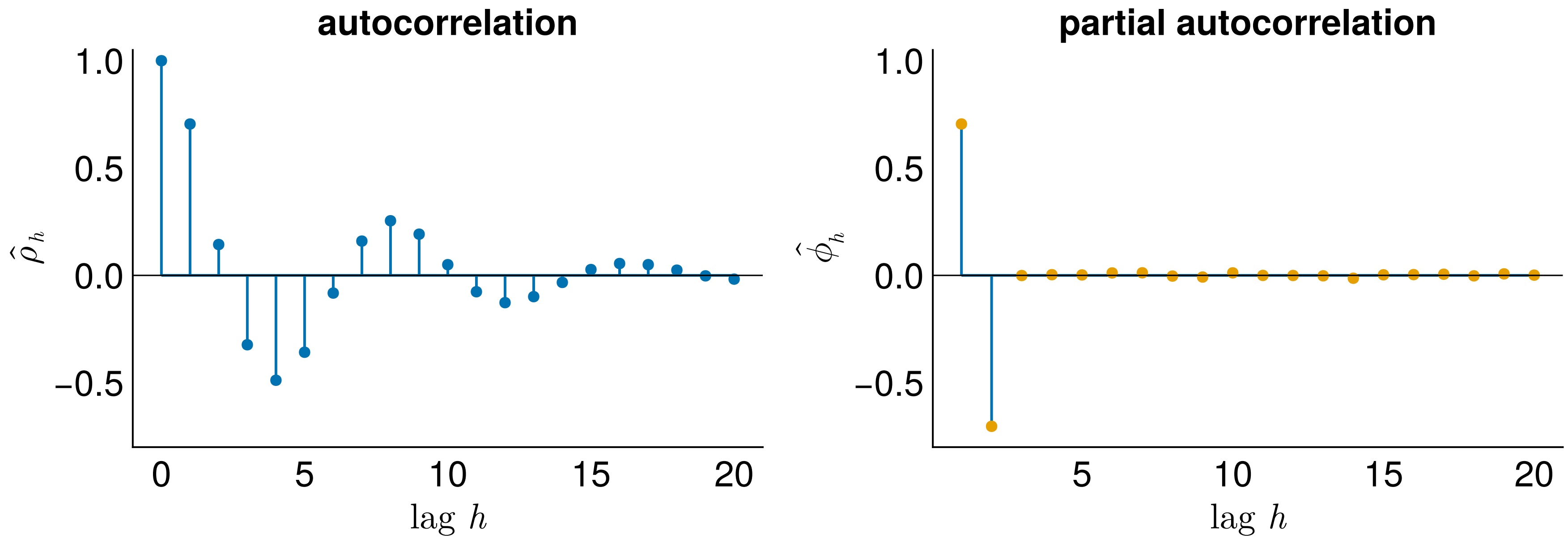
AR(2) Partial Autocorrelation

$$\phi_1 = \frac{a_1}{1 - a_2}, \qquad \phi_2 = a_2, \qquad \phi_h = 0 \text{ for } h > 2$$

For our $a_1 = 1.2$, $a_2 = -0.7$:

```
φ₁ = 0.706, φ₂ = -0.7
```

AR(2), BOTH DIAGNOSTICS



Autocorrelation

Partial autocorrelation

Second order

The spectrum

Some examples

Real data

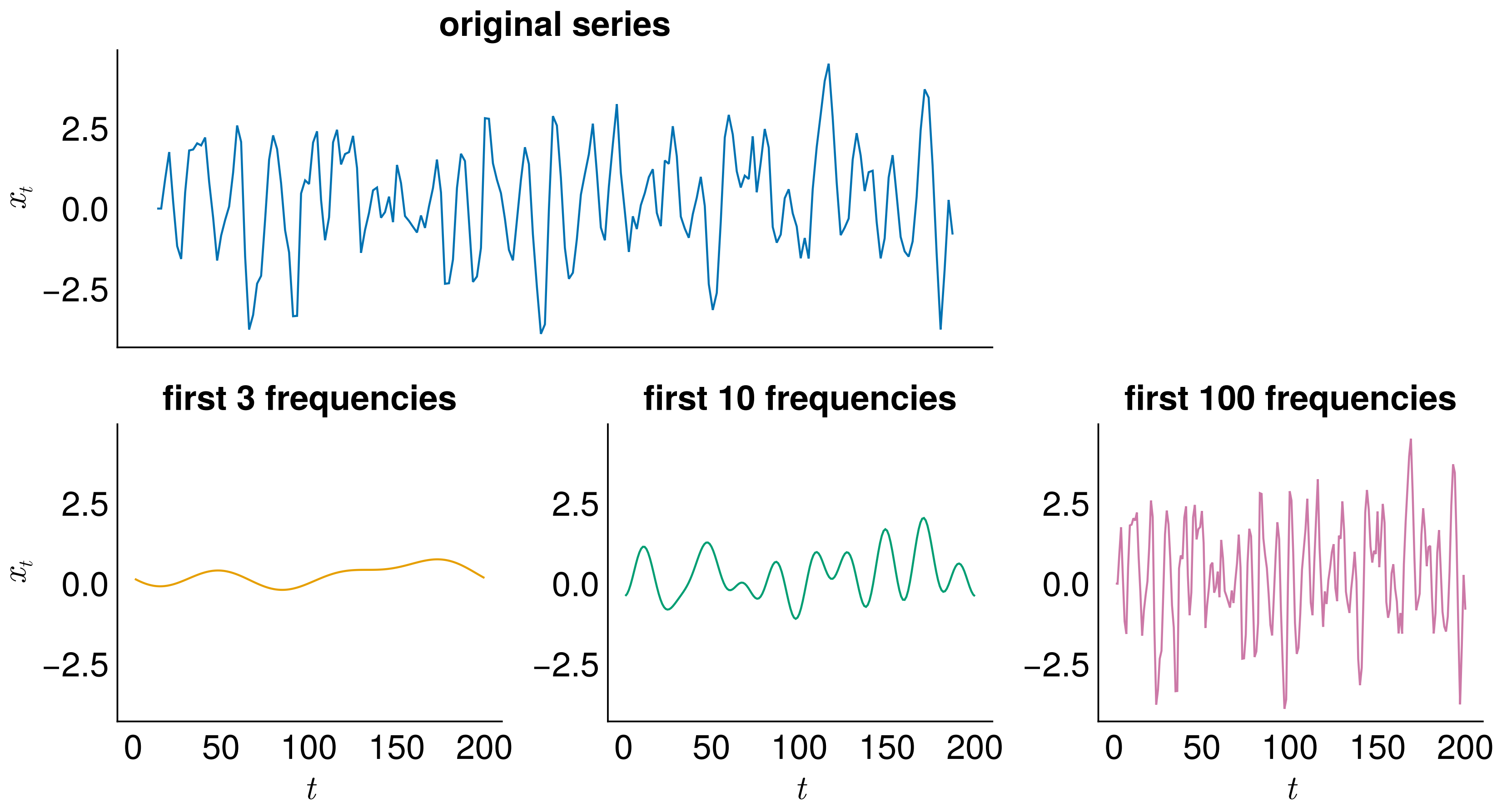
ANY TIME SERIES IS A SUM OF SINE WAVES

A zero-mean time series with n points can be written exactly as

$$x_t = \sum_{k=1}^{n/2} [A_k \cos(2\pi f_k t) + B_k \sin(2\pi f_k t)]$$

at frequencies $f_k = k/n$. Each frequency contributes some amplitude; the spectrum measures how much variance each one carries.

RECONSTRUCTING FROM FREQUENCIES



RECONSTRUCTING FROM FREQUENCIES

Adding more frequencies recovers more of the original signal. All $n/2$ frequencies recover it exactly.

VARIANCE BY FREQUENCY

- The **spectrum** $S(f)$ is the variance of the record at frequency f.
- Frequency is in cycles per time step, so $f = 1/12$ is one cycle every 12 steps.
- Frequencies run from 0 to $1/2$ (the Nyquist frequency).

Wiener-Khinchin theorem

The spectrum is the Fourier transform of the autocorrelation function:

$$S(f) = \sum_{h=-\infty}^{\infty} \rho_h \, e^{-2\pi i f h}$$

For supplementary notes, read Kantz (2004) §2.3, printed pp. 18-23.

WHITE NOISE

A **white noise** sequence is a set of independent, identically distributed random variables, each with mean zero and variance σ^2.

$\rho_h = 0$ for all $h \neq 0$: because the draws are independent, there is no autocorrelation at any lag.

Since $S(f)$ is the Fourier transform of ρ_h, and ρ_h is zero everywhere except $h = 0$:

$$S(f) = \sigma^2 \qquad \text{for every} \quad f$$

A flat spectrum: equal power at every frequency (by analogy with white light). Adding white noise to a signal adds the same amount of variance at every frequency.

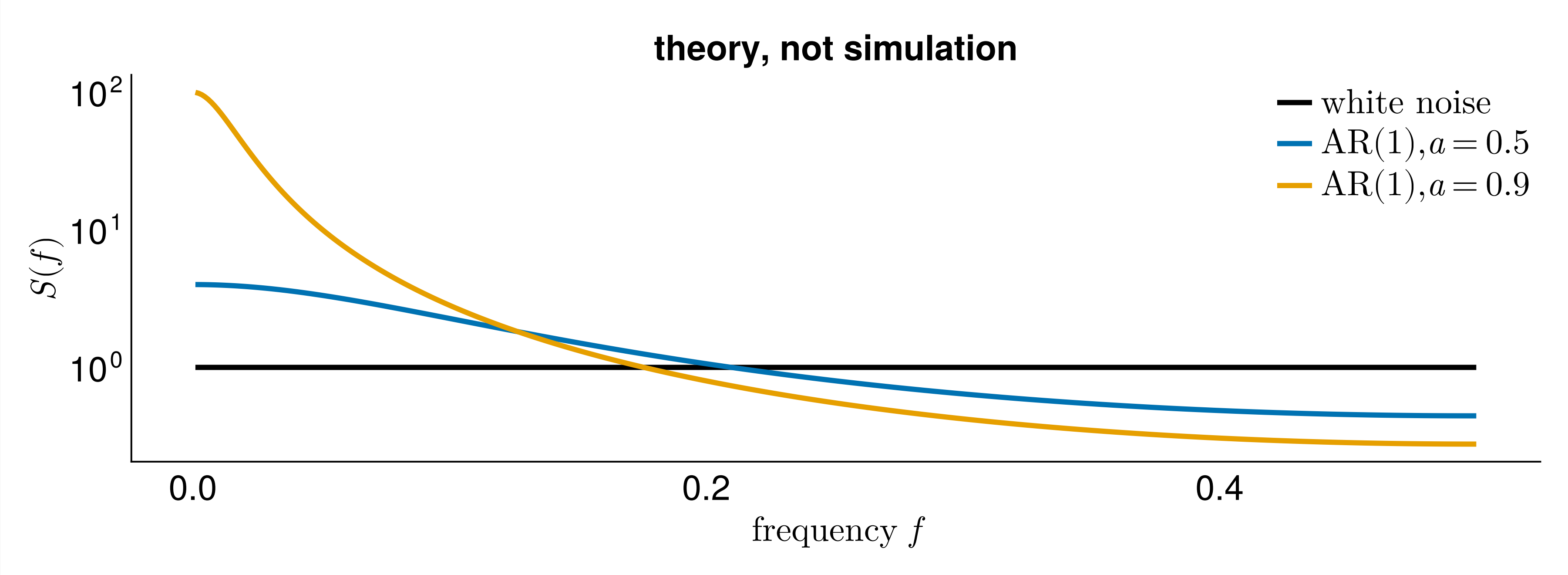
AR(1) SPECTRUM: RED NOISE

$$S(f) = \frac{\sigma^2}{\left|1 - a\,e^{-2\pi i f}\right|^2} = \frac{\sigma^2}{1 - 2a\cos 2\pi f + a^2}$$

At $f = 0$: denominator is $(1 - a)^2$, power is largest. At $f = 1/2$: denominator is $(1 + a)^2$, power is smallest.

For $a > 0$, power decreases with frequency. Called **red noise** because red is the low-frequency end of visible light.

TWO THEORETICAL SPECTRA



- Autocorrelation
- Partial autocorrelation
- Second order
- The spectrum
- **Some examples**
- Real data

FIVE TEST SIGNALS

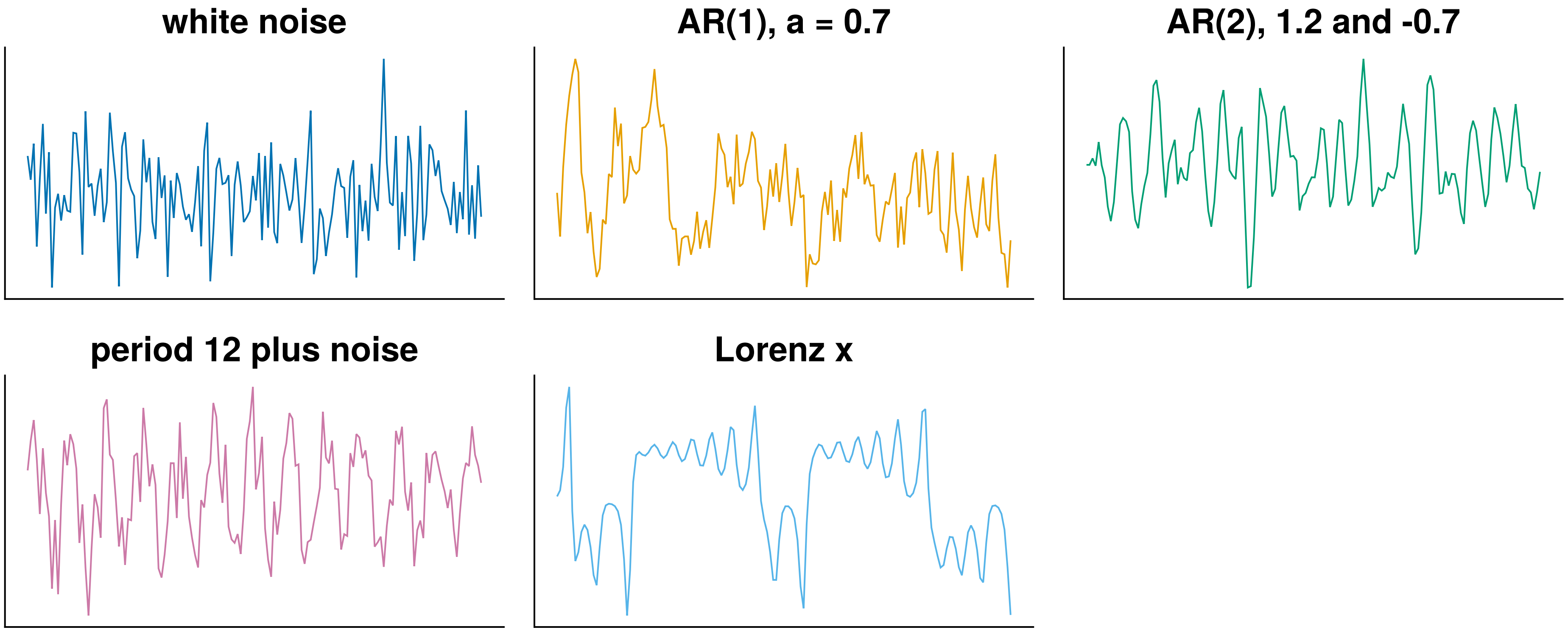
```julia
n = 20_000

white_noise = randn(n)
ar1 = simulate_ar([0.7], n)
ar2 = simulate_ar([1.2, -0.7], n)
seasonal = [sin(2π * t / 12) + 0.5 * randn() for t in 1:n]

function lorenz_x(n; dt=0.01, subsample=10)
    total = n * subsample
    x, y, z = zeros(total), zeros(total), zeros(total)
    x[1], y[1], z[1] = 1.0, 1.0, 1.0
    for i in 1:(total - 1)
        x[i + 1] = x[i] + dt * 10.0 * (y[i] - x[i])
        y[i + 1] = y[i] + dt * (x[i] * (28.0 - z[i]) - y[i])
        z[i + 1] = z[i] + dt * (x[i] * y[i] - (8 / 3) * z[i])
    end
```

FIVE TEST SIGNALS

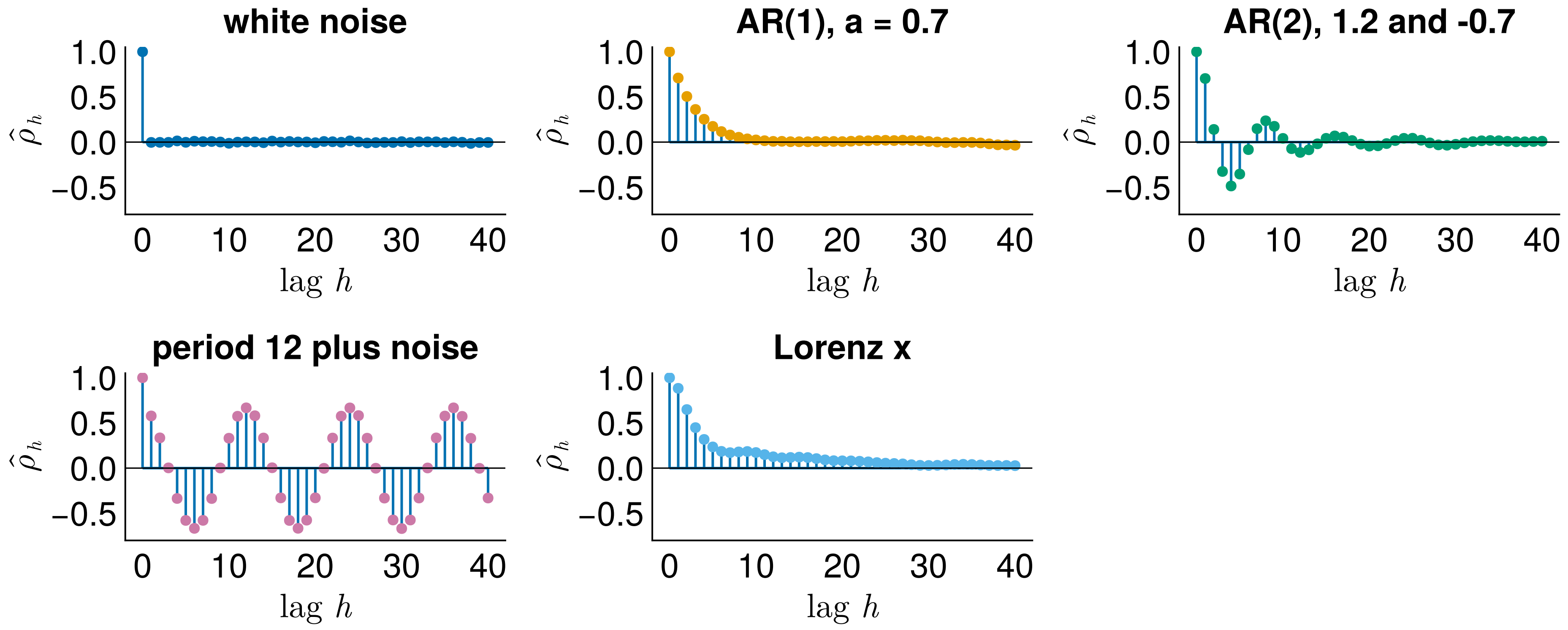
```
    return x[1:subsample:end]
end

lorenz = lorenz_x(n)
```

The Lorenz system from Monday is included as a deterministic comparison (no noise term).

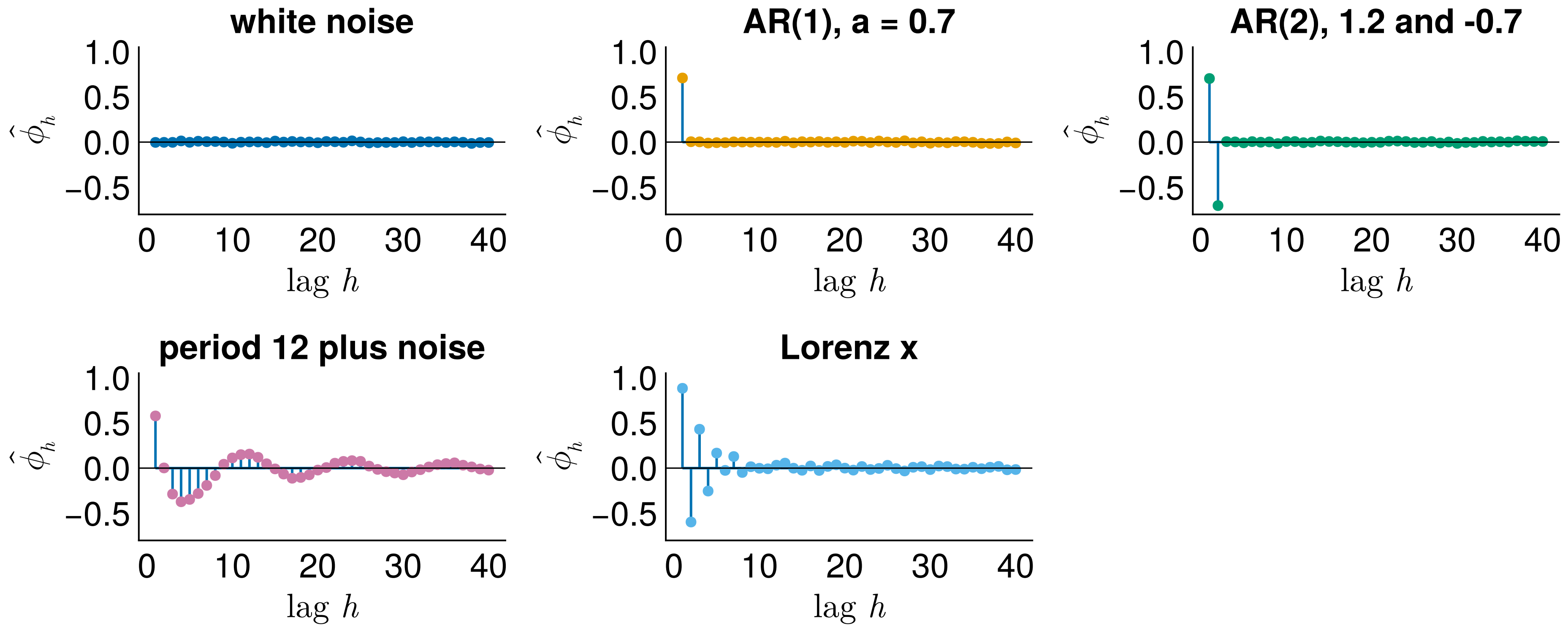
WHAT THEY LOOK LIKE



AUTOCORRELATION



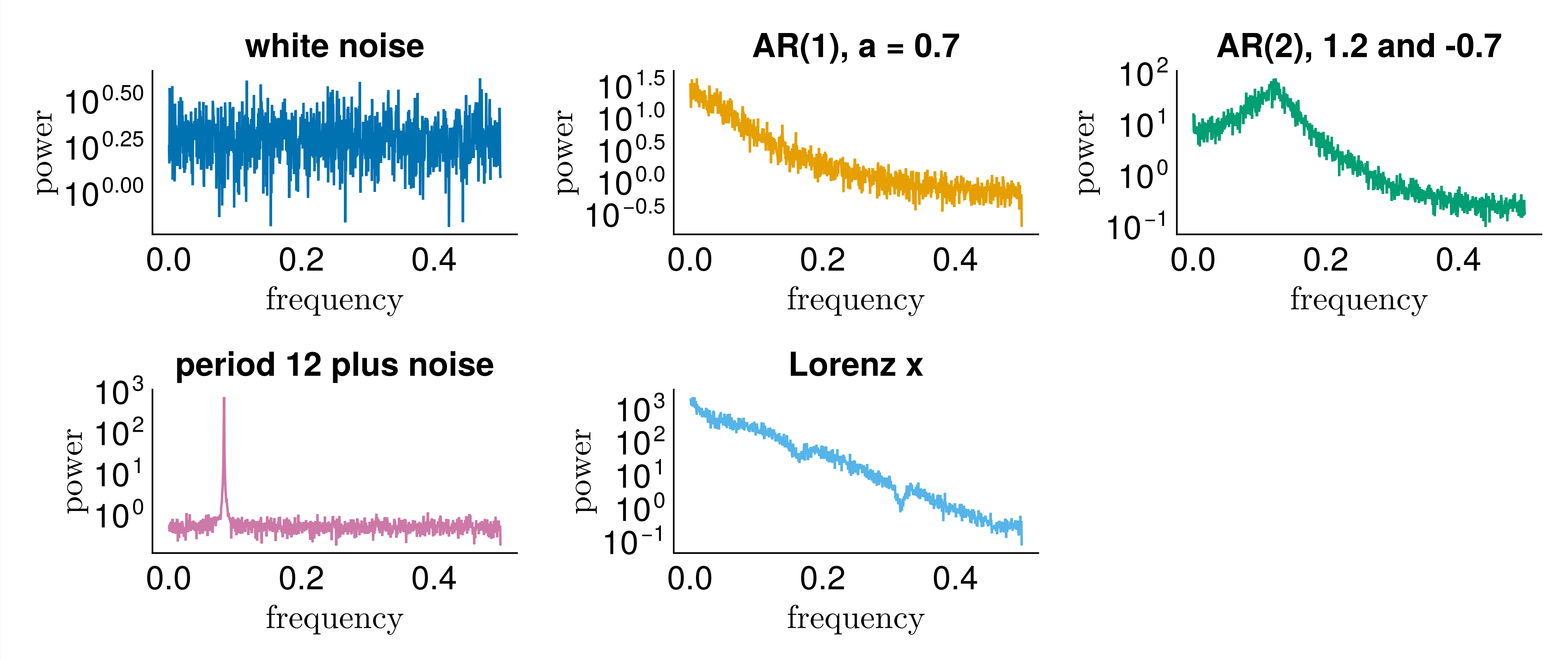
PARTIAL AUTOCORRELATION



Partial Autocorrelation

The Lorenz x has alternating partial autocorrelations that decay slowly, unlike any finite-order AR model.

SPECTRA


white noise
AR(1), a = 0.7
AR(2), 1.2 and -0.7
period 12 plus noise
Lorenz x
power
frequency
$10^{0.50}$
$10^{0.25}$
$10^{0.00}$
$10^{1.5}$
$10^{1.0}$
$10^{0.5}$
$10^{0.0}$
$10^{-0.5}$
10^{2}
10^{1}
10^{0}
10^{-1}
10^{3}
0.0
0.2
0.4

WHY THE AR(2) PEAK IS WHERE IT IS

The AR(2) spectrum is $S(f) = \sigma^2 / |1 - a_1 e^{-2\pi i f} - a_2 e^{-4\pi i f}|^2$.

The denominator expands to

$$|1 - a_1 e^{-2\pi i f} - a_2 e^{-4\pi i f}|^2 = (1 - a_1 \cos 2\pi f - a_2 \cos 4\pi f)^2 + (a_1 \sin 2\pi f + a_2 \sin 4\pi f)^2$$

The peak sits at the frequency where this denominator is smallest. Setting the derivative to zero and using $\cos 4\pi f = 2\cos^2 2\pi f - 1$:

$$f_{\text{peak}} = \frac{1}{2\pi} \arccos\left(\frac{-a_1(1 - a_2)}{4a_2}\right)$$

```
f_peak = 0.12 (period = 8.33 time steps)
```

Autocorrelation

Partial autocorrelation

Second order

The spectrum

Some examples

Real data

THE AMERICAN RIVER AT FOLSOM, CALIFORNIA

The American River basin, draining the western Sierra Nevada

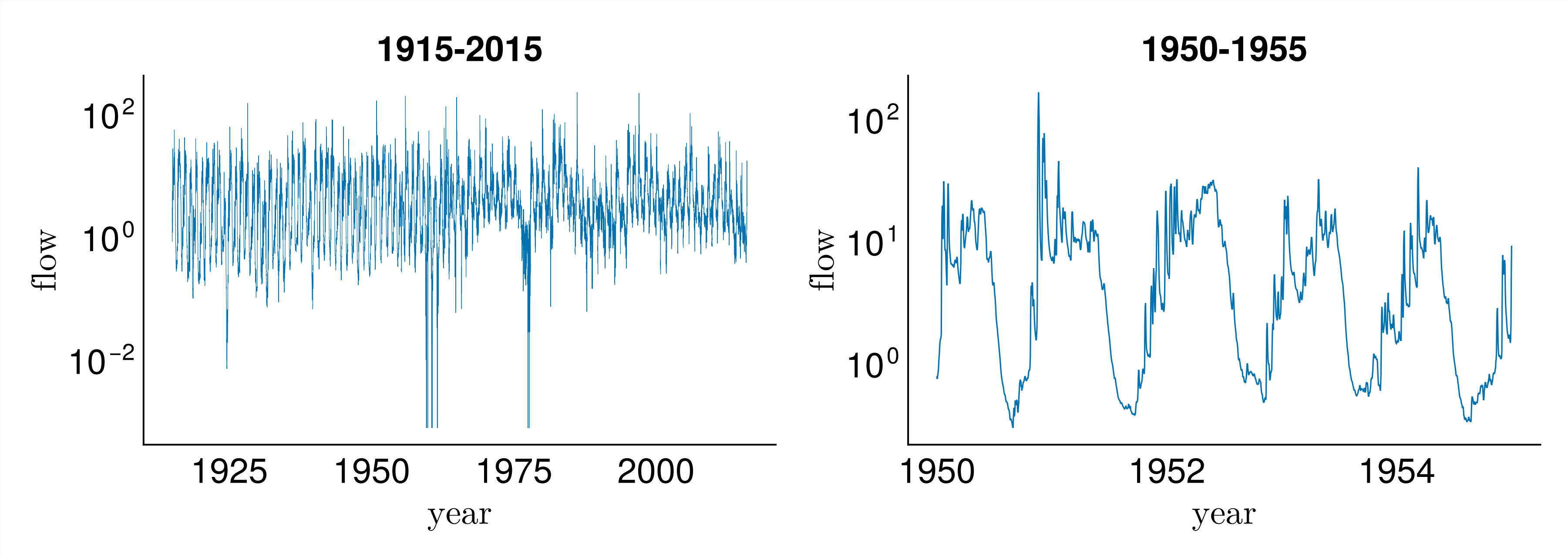
Daily flow, 1915 to 2015, 36,888 days.

Snowmelt-driven, so expect a strong seasonal cycle.

WHAT ONE WET YEAR DID

K Street in Sacramento during the Great Flood of 1862, with people rowing boats between the storefronts. Lithograph published by A. Rosenfield, San Francisco; public domain.

THE RECORD



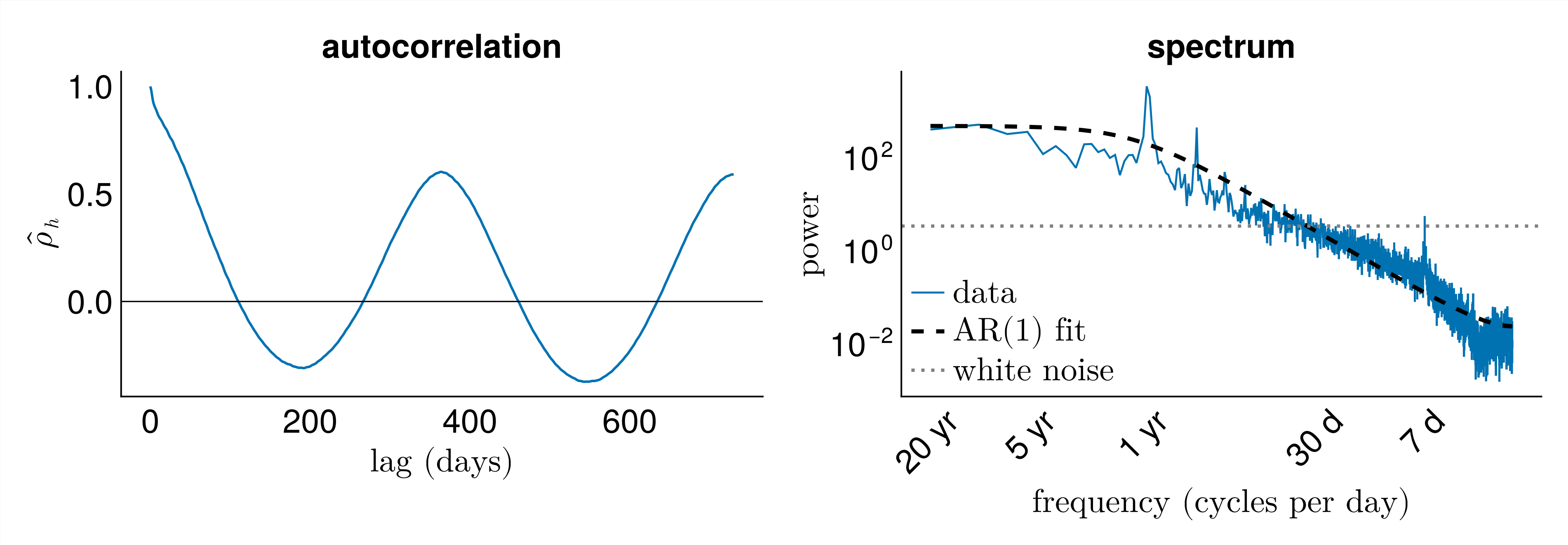
The raw flow spans four orders of magnitude, so all analysis uses $\log$ flow.

SKETCH IT

Draw the autocorrelation out to two years, and the spectrum.

Where do the peaks go?

AMERICAN RIVER AUTOCORRELATION AND SPECTRUM



Both axes are logarithmic, the convention for a spectrum spanning decades of frequency and power.

AMERICAN RIVER AUTOCORRELATION AND SPECTRUM

Nothing stands above the red-noise fit at two to seven years, where the El Niño-Southern Oscillation (ENSO) would sit.

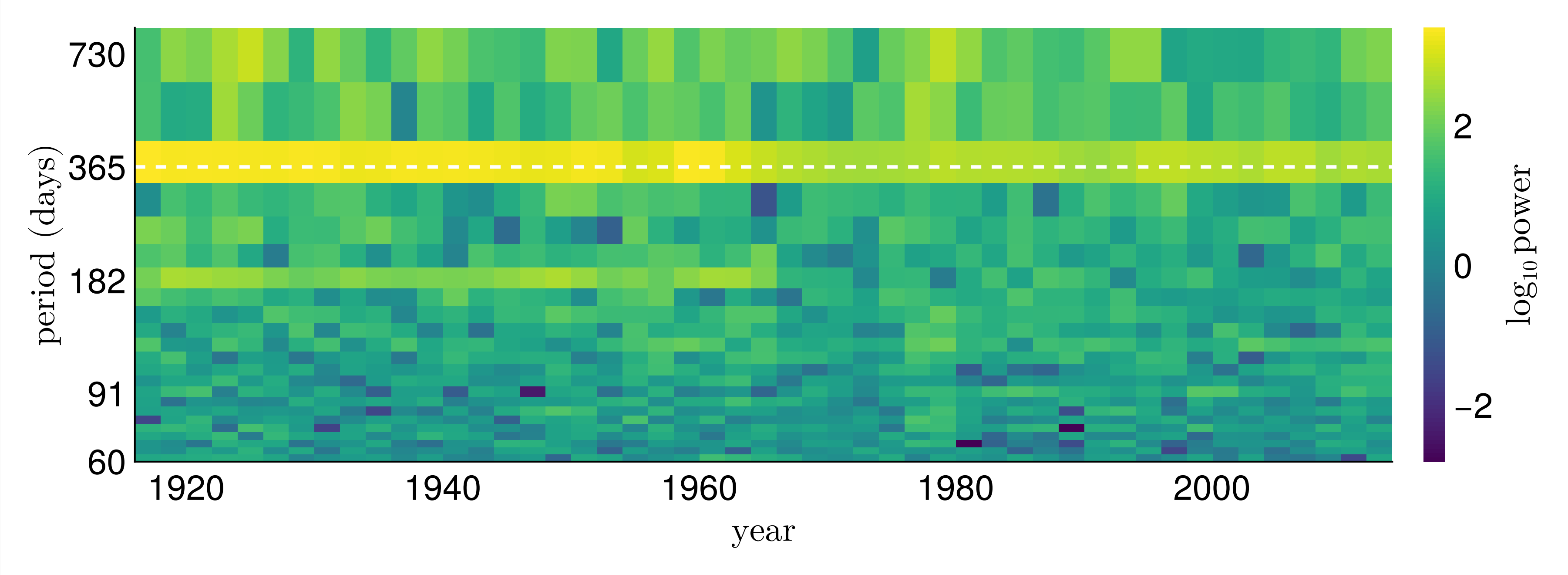
SPECTROGRAM

A **spectrogram** computes a spectrum in each of several overlapping windows of m points:

$$S(f,t) = \left| \sum_{s=1}^{m} w_s \, x_{t+s} \, e^{-2\pi i f s} \right|^2$$

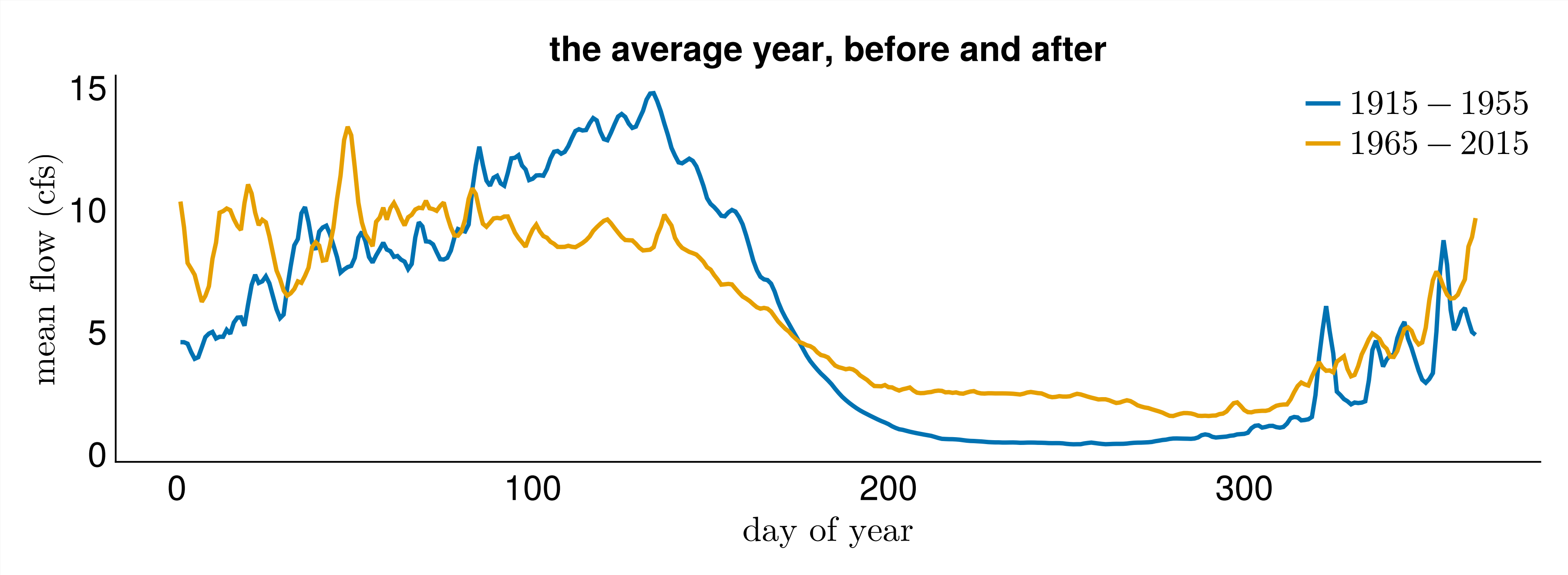
The frequency resolution is $1/m$, so a longer window separates nearby frequencies but blurs their timing.

AMERICAN RIVER SPECTROGRAM



The frequency grid is linear in f but the y-axis is $\log$ period, so long periods get coarser blocks.

THE HALF-YEAR BAND FADES




```
1915-1955: peak on day 134, second harmonic is 0.32 of the first
1965-2015: peak on day 48, second harmonic is 0.05 of the first
```

FRIDAY

Lab 2 runs a two-variable model of ENSO with fixed parameters and no trend, then samples 50-year windows from it.