DS-GA 3001.001
Probabilistic time series analysis
Lecture 1
Logistics. Introduction to time series

Instructor: Cristina Savin

NYU, CNS & CDS

Course logistics

Instructor
Cristina Savin, csavin@nyu.edu
Office hours: Tue, 4-5pm, Room 608

TA
Tim Kunisky, dk3105@nyu.edu
Office hours: Wed 6:10-7:10pm, Room TBD

Couse page: https://github.com/charlieblue17/probabilisticTSAFall2018/blob/master/README.md
Piazza: https://piazza.com/nyu/fall2018/dsga3001001/home

Course logistics: grading

5 problem sets 40%
A mix of derivations + coding (python/matlab)

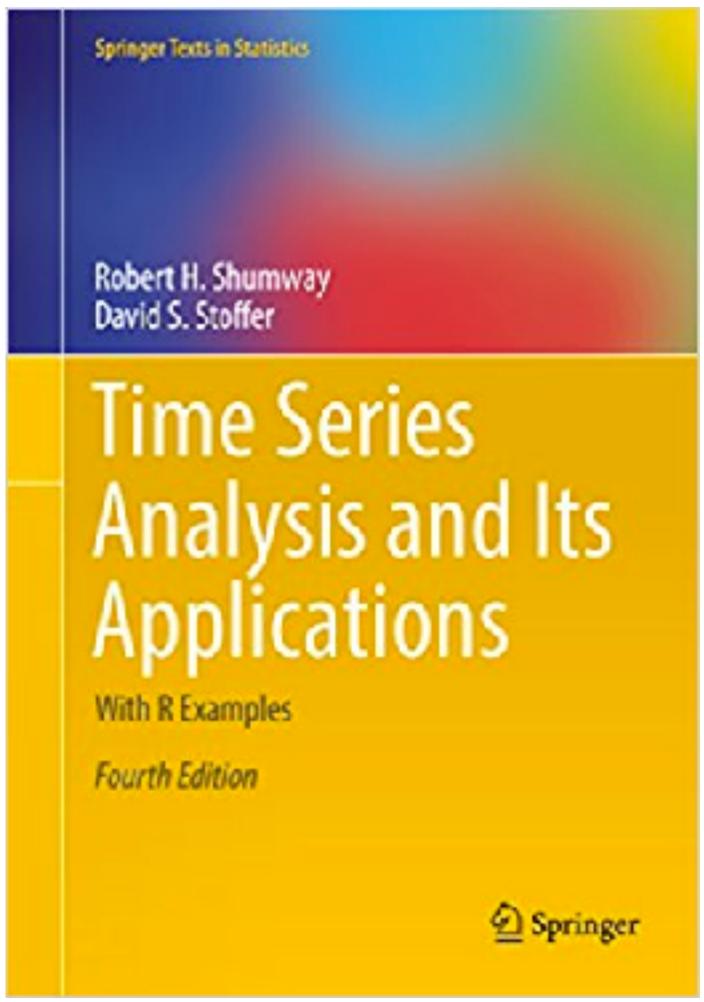
Midterm 25%
March 6th, Covers AR+ latent state models

Project 25%
Groups of 2-3, topic of choice
Project proposal due **Sept. 23rd**

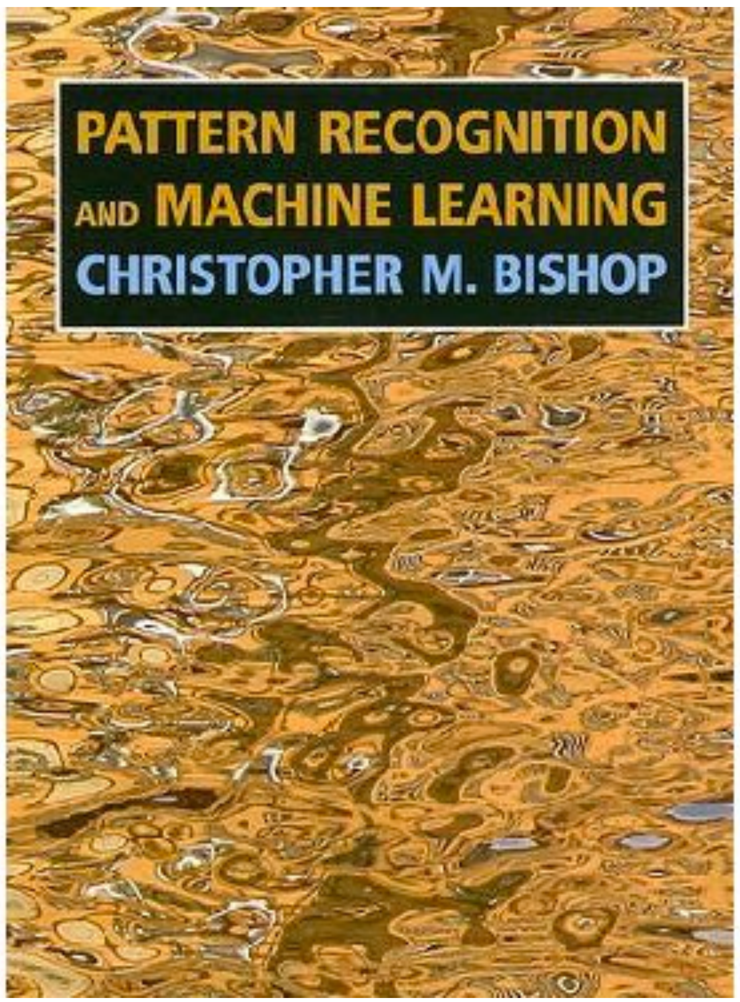
Participation 10%
Lab work, lecture discussions, piazza

Bibliography

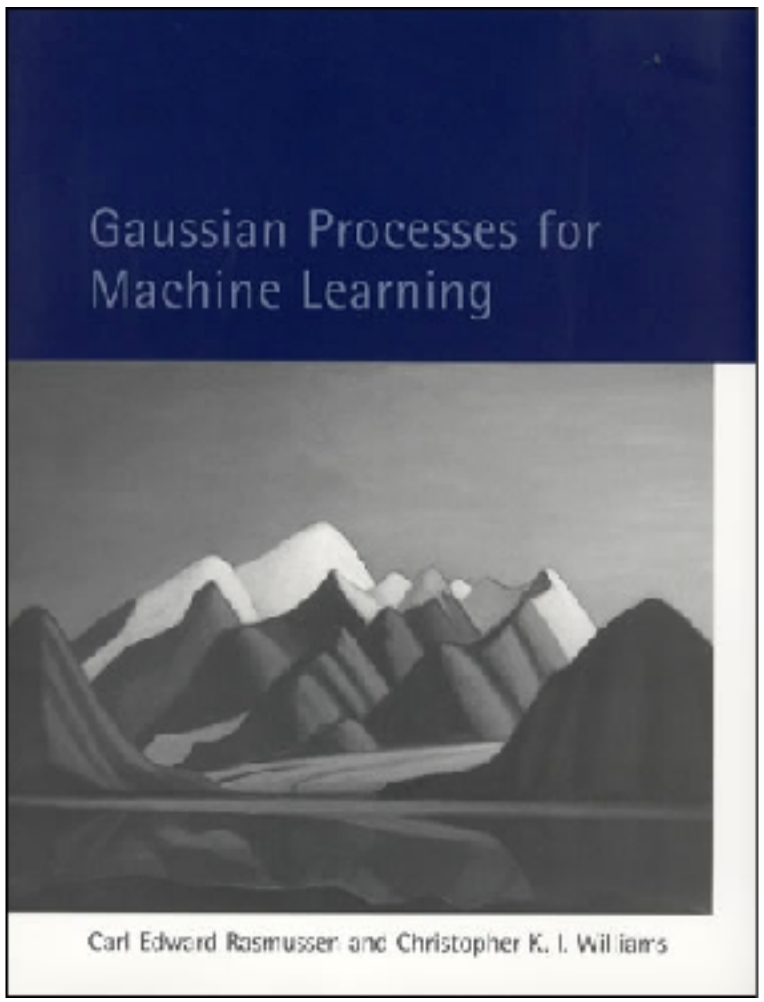
No course textbook, one handout for each section. Lectures based on:



Intro, AR(I)MA, Spectral methods



Latent space models

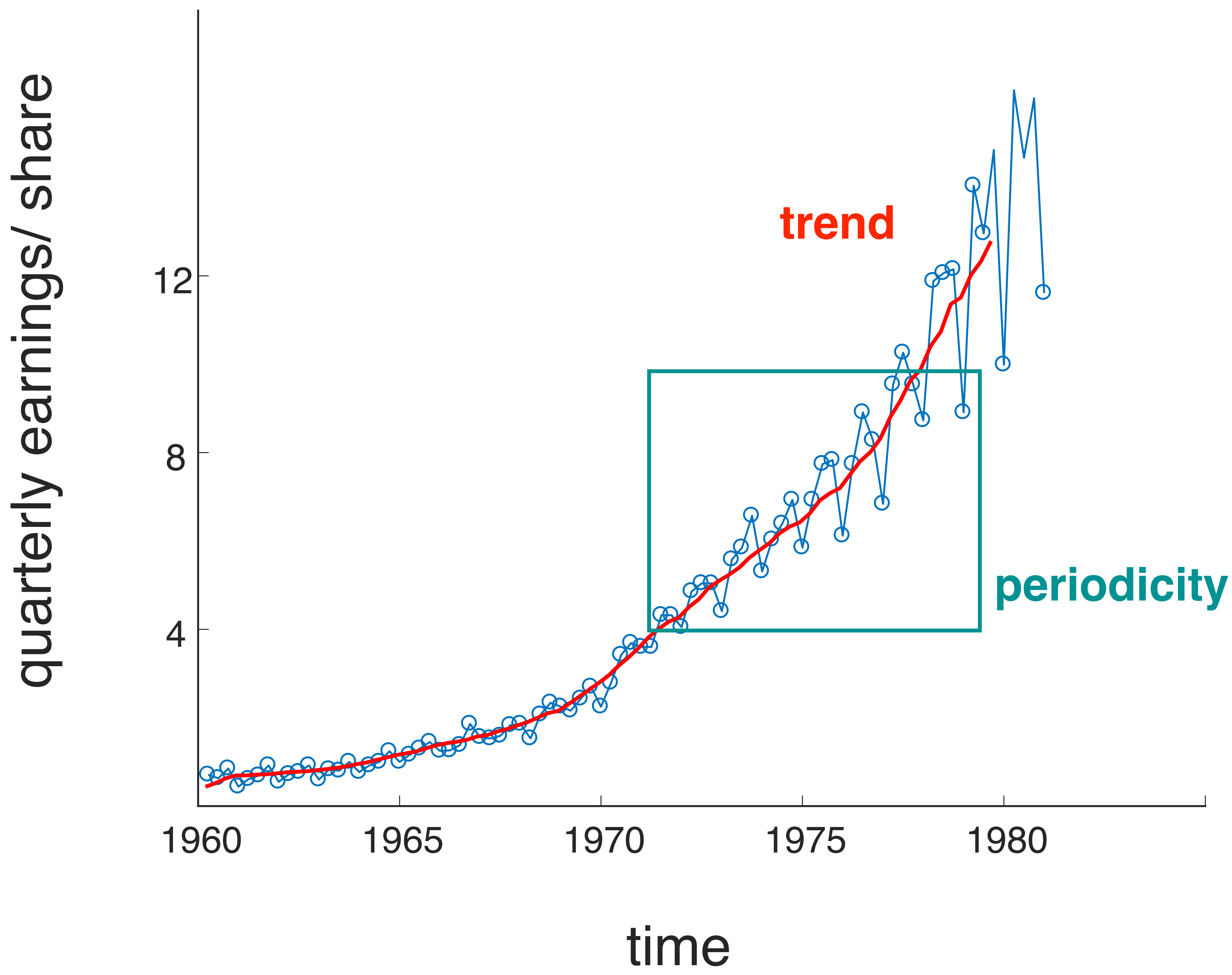


GP

Intro: what is a time series?

Some examples

Task: predict future earnings, interpret data structure

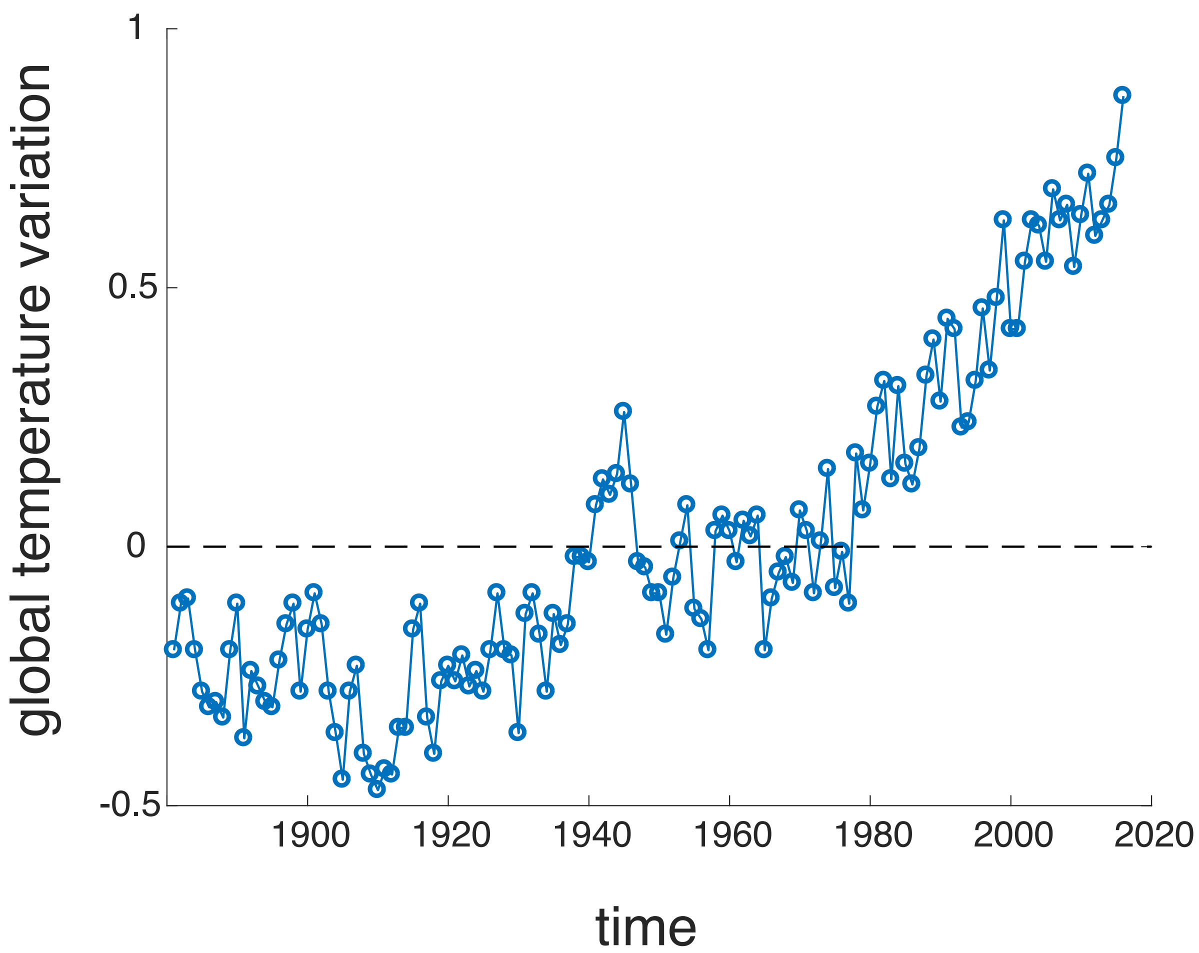


Johnson & Johnson

Intro: what is a time series?

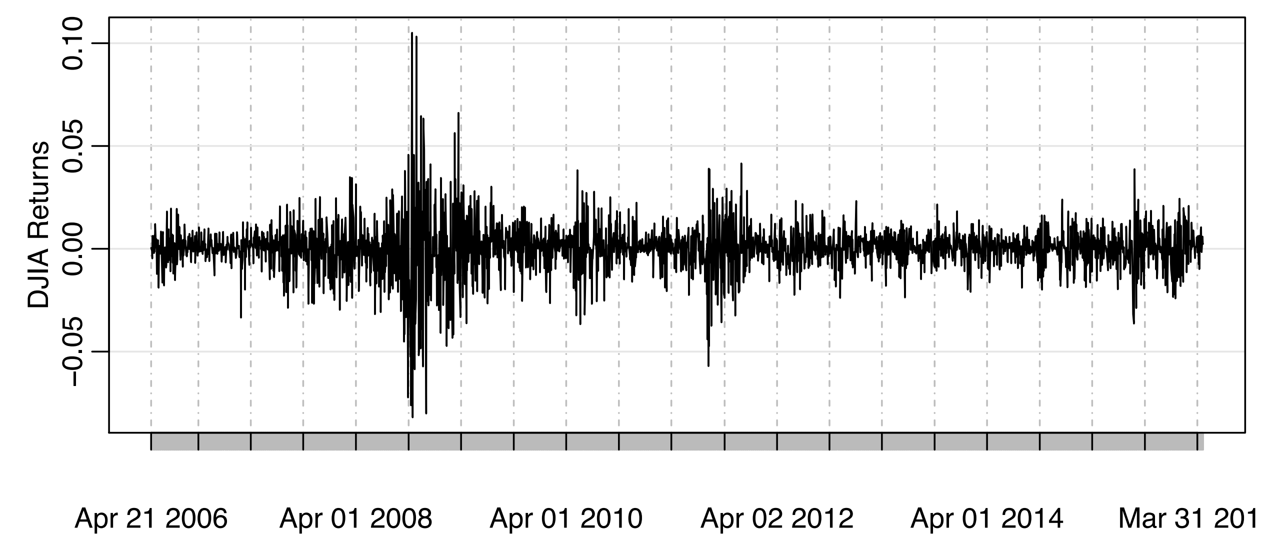
Task: do things change systematically over time?

Global warming



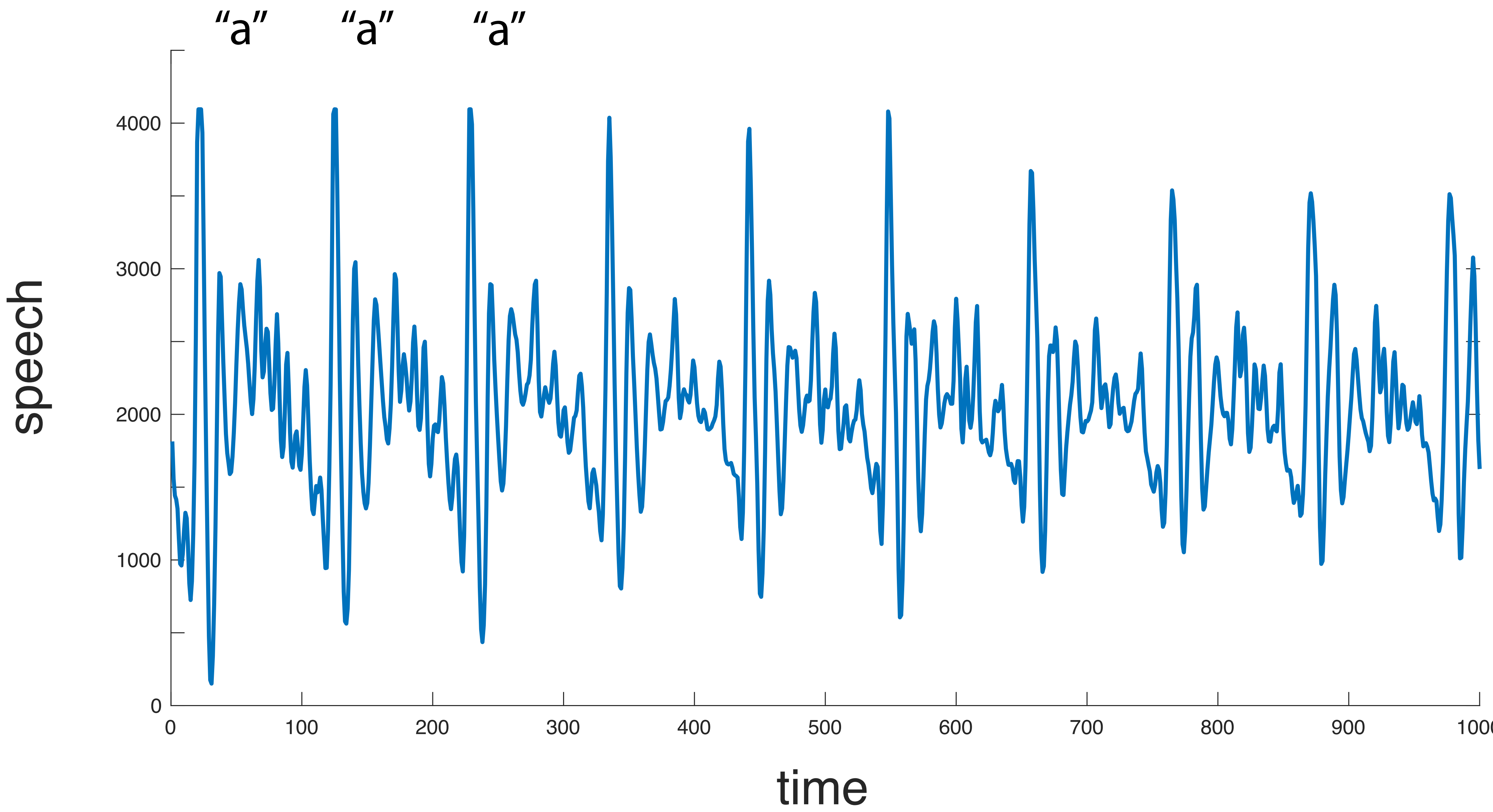
Task: detect high volatility periods

Financial crisis



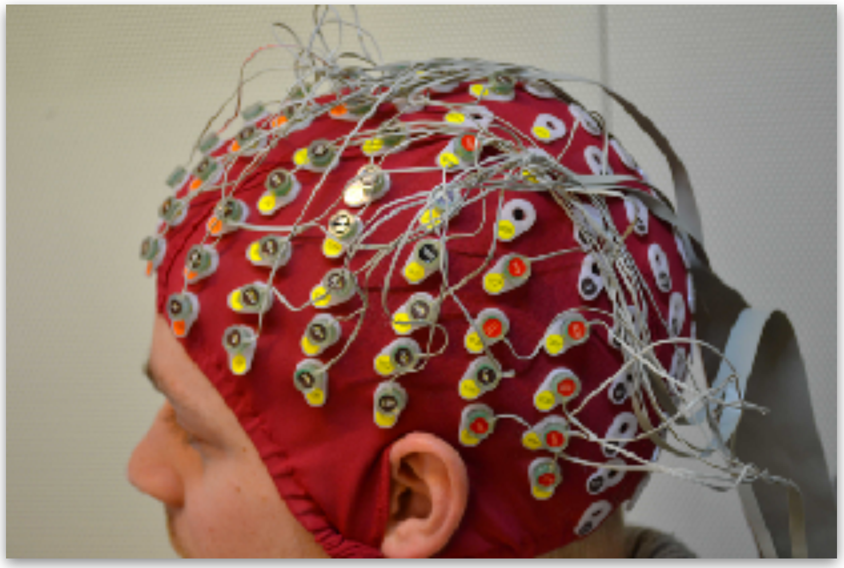
Intro: what is a time series?

Task: infer discrete latent structure from analog signals

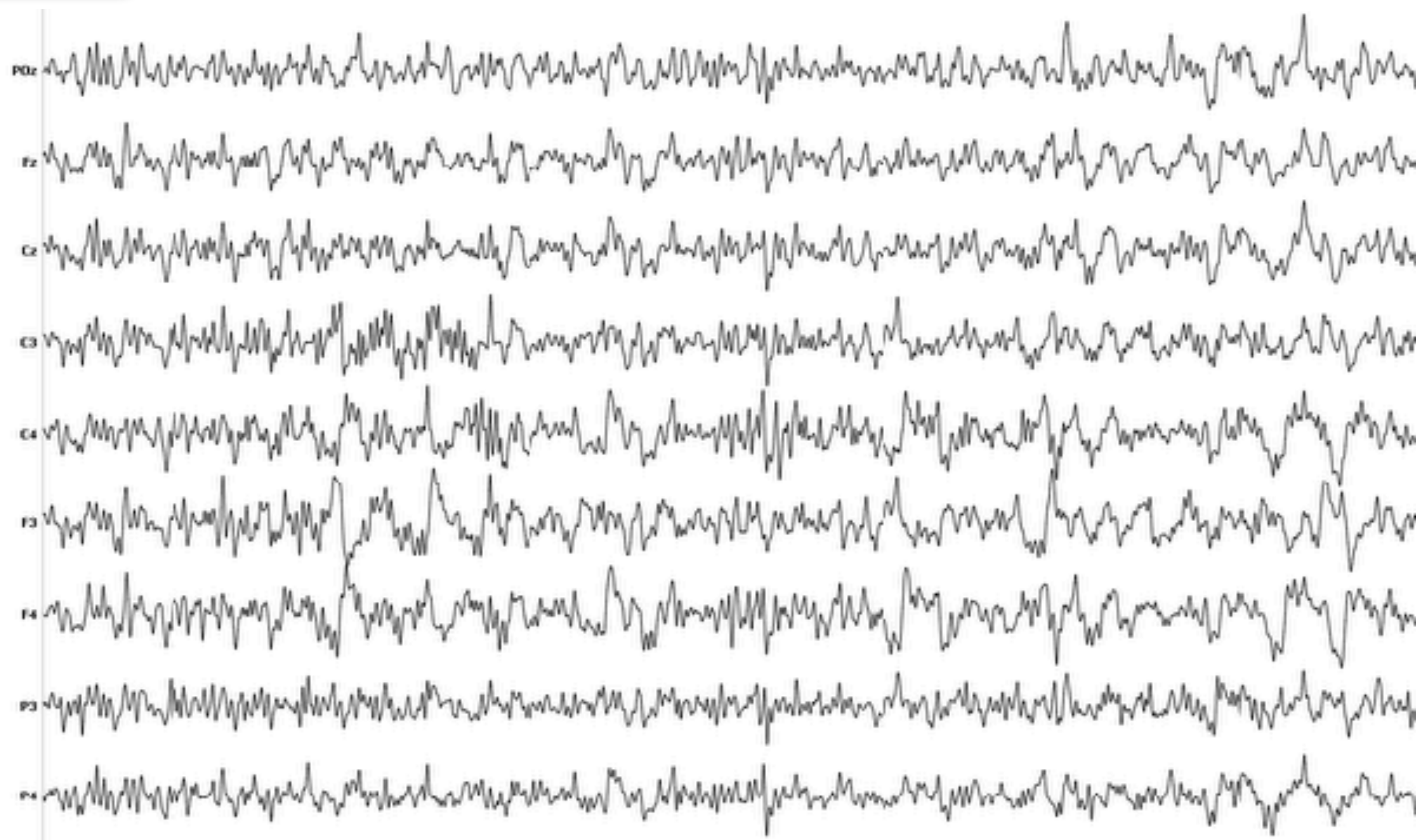


Voice recognition

Intro: what is a time series?

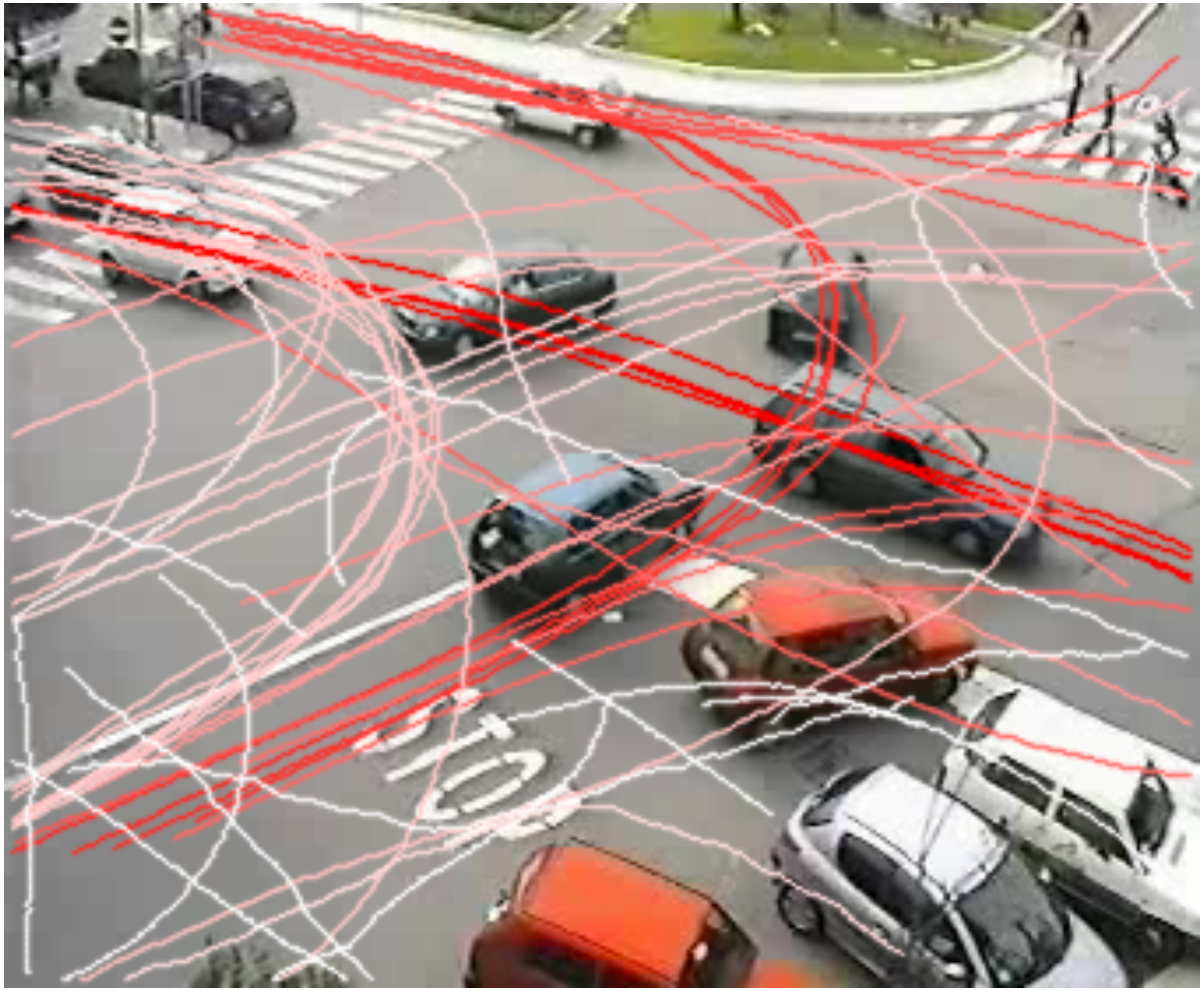


Multivariate time series: EEG



Task: decoding
Brain-computer interfaces

Self driving cars



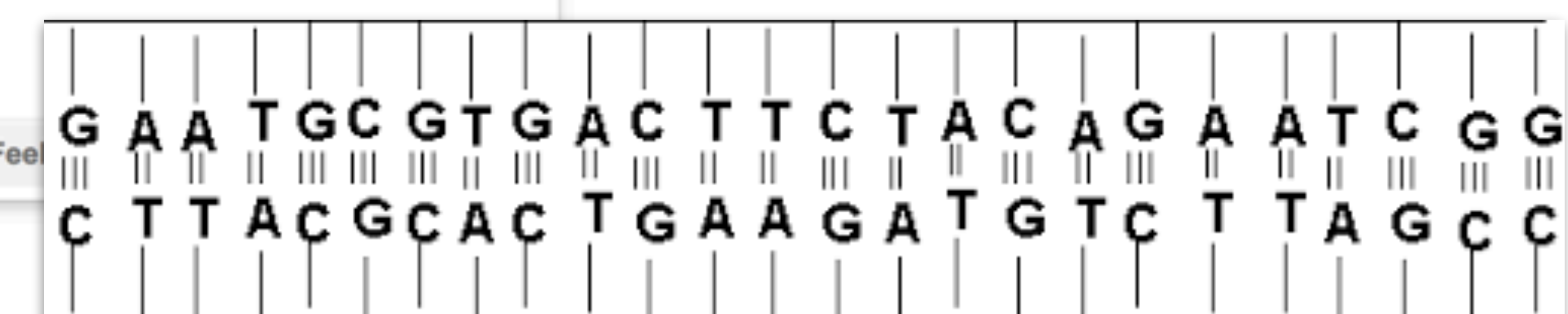
Tasks: simulation, control

We need uncertainty representation for optimal decision making, risk minimization

From Saunier & Sayed 2008

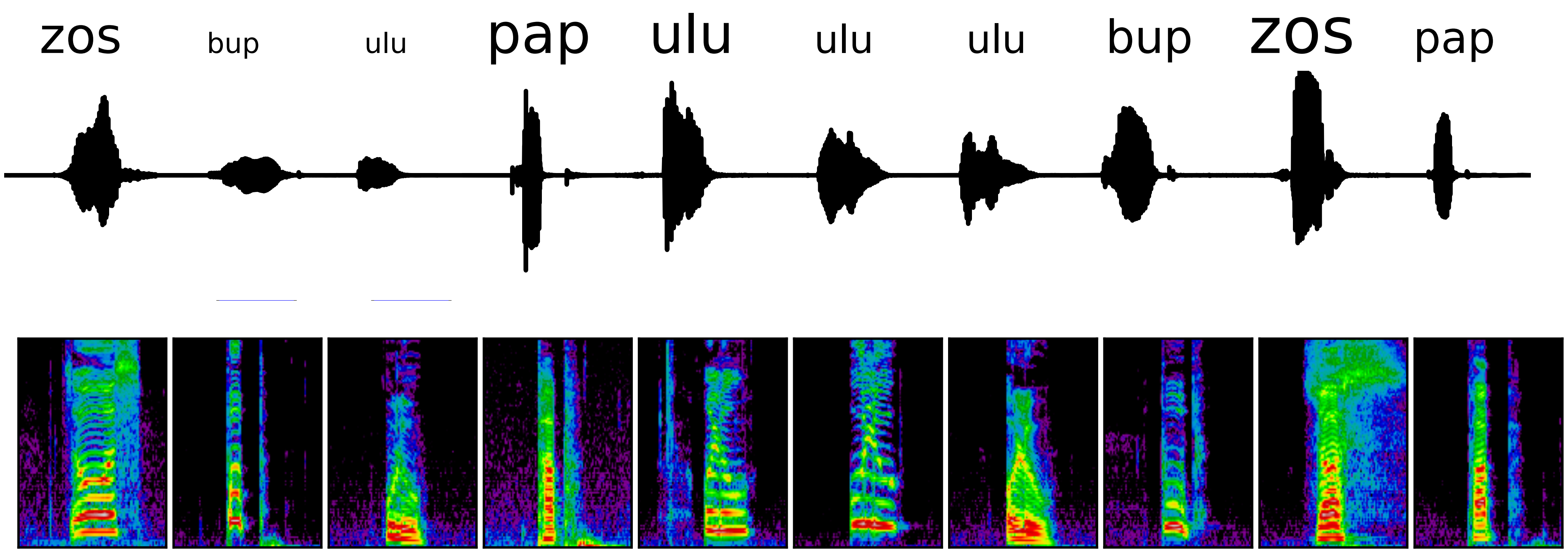
Intro: what is a time series?

Language!
NLP
Machine translation



More general sequential structure:
e.g. sequence of nucleotide base pairs in DNA

Intro: what is a time series?



Model data in the frequency domain

e.g. automated speech recognition

Tasks: identify latent structure, denoising…

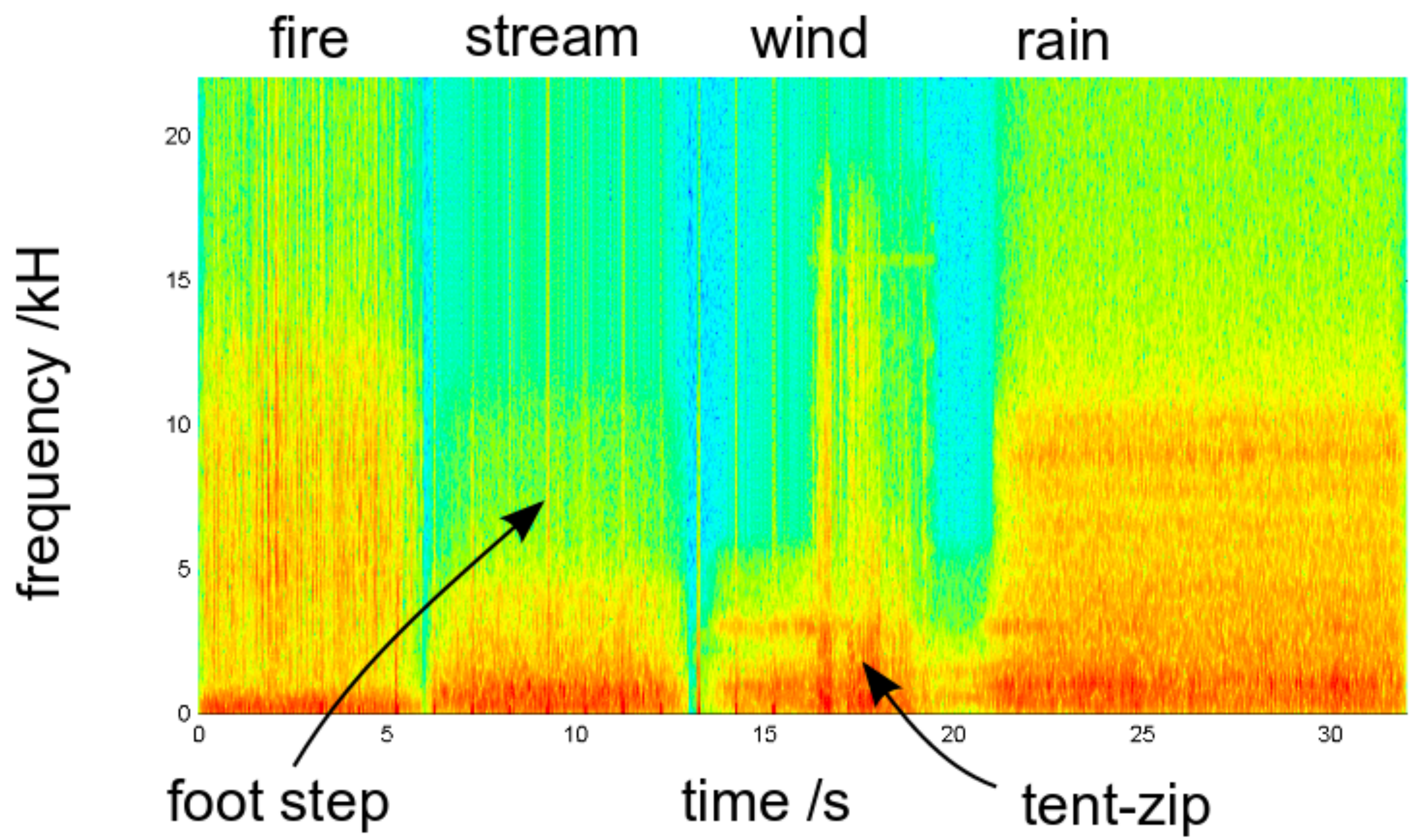
Why **probabilistic** time series analysis

1. Define generative models capturing the relevant statistical structure in the data

2. Fit such models to data

3. Make predictions

4. Generate artificial data with same statistics


fire
stream
wind
rain
frequency /kH
20
15
10
5
0
0
5
10
15
20
25
30
time /s
foot step
tent-zip

[redacted]

Overview of course

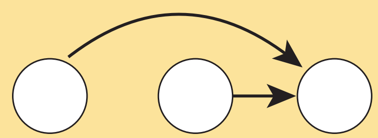
TIME

Lagged relationships

Next Lecture

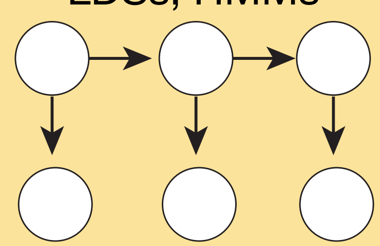
Model temporal dependencies directly

AR+ friends



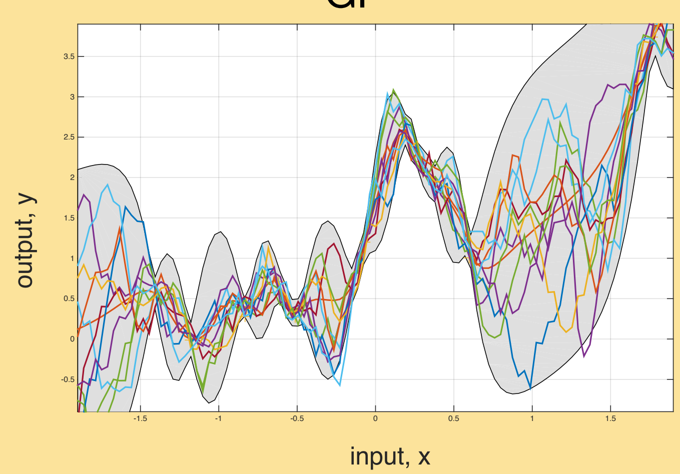
Latent structure

LDSs, HMMs



Distribution function

GP



FREQUENCY

Stationary processes seasonality/periodicity

Probabilistic spectral analysis

Intro: what is a time series?

Definitions. Basic statistics

What is a time series?

Formally, a collection of random variables indexed by time, t*

$$\{X_1, X_2, \ldots, X_t \ldots\}$$

"stochastic process"

data = "realisation"

Unlike the traditional case, NOT I.I.D. !!!

These **dependencies** are the main point; it's what makes prediction possible.

Fully specified by joint*:

$$\mathrm{P}(X_1 \leq x_1, \ldots, X_t \leq x_t \ldots)$$

**Usually discrete time (digital data collection), but continuous time can be convenient in some cases*

***Intractable in general, usually we limit ourselves to simpler statistics*

Basic statistics of a time series

Mean

$$\mu_X(t) = \mathbb{E}(X_t)$$

Covariance

$$R_X(t,u) = \text{cov}(X_t, X_u)$$

Auto-Correlation Function (ACF)

$$\rho_X(t,u) = \frac{R_X(t,u)}{\sqrt{R_X(t,t), R_X(u,u)}}$$

measures linear predicability of X_t from X_s

$$-1 \leq \rho_X(t,u) \leq 1$$

Cross-Covariance

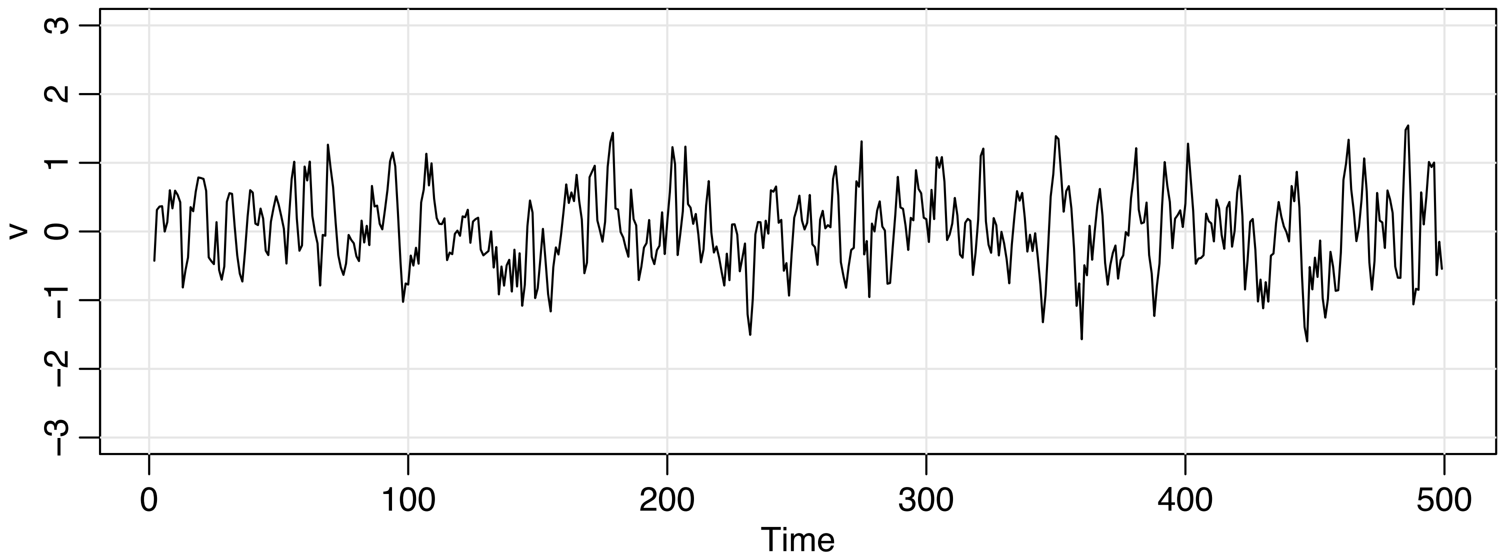
$$R_{X,Y}(t,u) = \text{cov}\left(X_t, Y_u\right)$$

Cross-Correlation Function (ACF)

$$\rho_{X,Y}(t,u) = \frac{R_{X,Y}(t,u)}{\sqrt{R_{X,Y}(t,t), R_{X,Y}(u,u)}}$$

Example stochastic processes

White noise $X_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d.



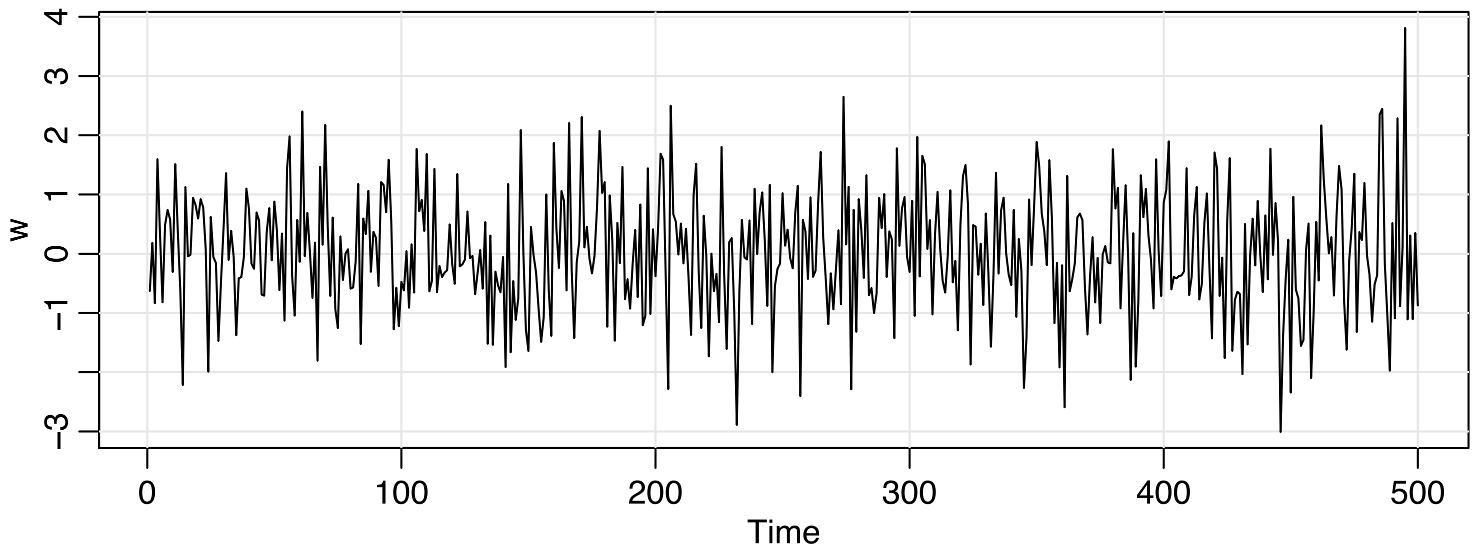
Trivially, white noise has

$$\mu_X(t) = 0$$

$$R_X(t, u) = \begin{cases} \sigma^2, t = u \\ 0, t \neq u \end{cases}$$

Example stochastic processes

Moving average (MA) $v_t = \frac{1}{3}\left(w_{t-1} + w_t + w_{t+1}\right)$ filtered white noise



$$\mu_V(t) = 0$$

$$R_V(t,u) = \begin{cases} 1/3\,\sigma^2 & ,\, t = u \\ 2/9\,\sigma^2 & ,\, |t-u| = 1 \\ 1/9\,\sigma^2 & ,\, |t-u| = 2 \\ 0, & |t-u| > 2 \end{cases}$$

***Useful: Cov. of linear combinations**

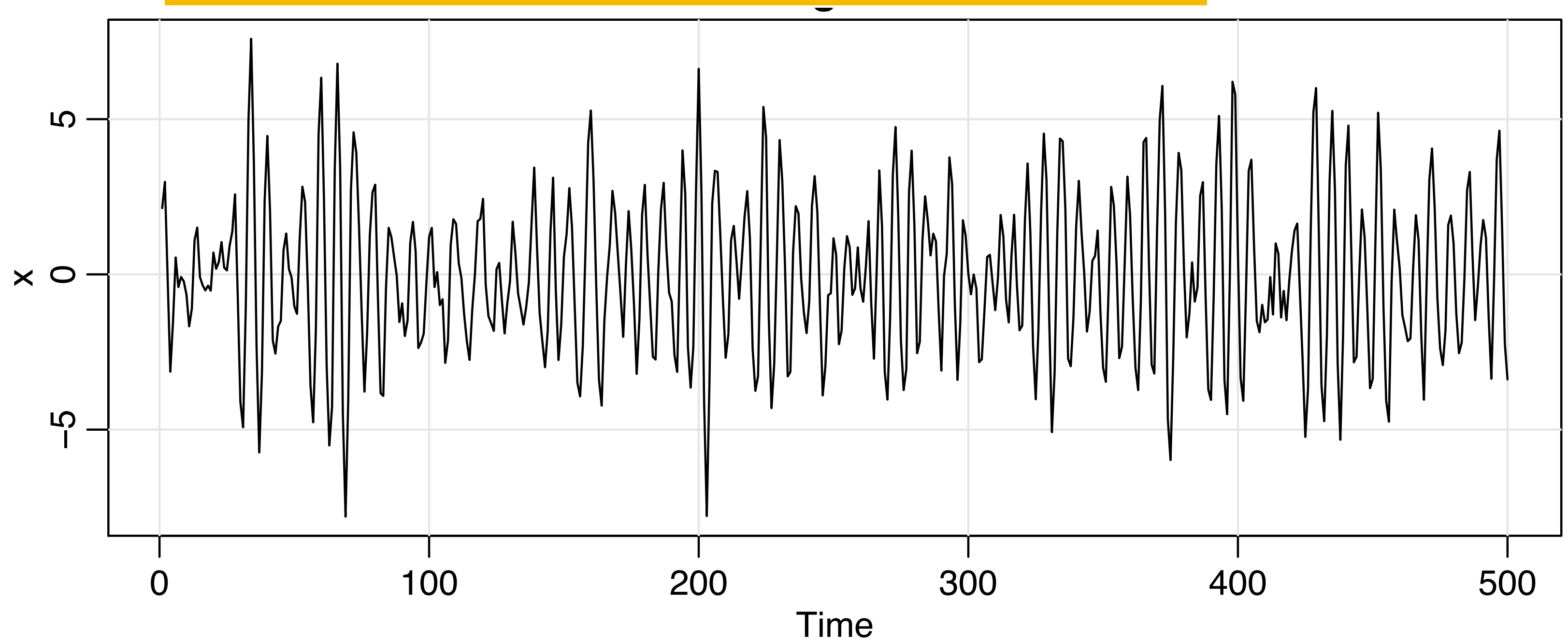
$$U = \sum_i a_i X_i$$

$$V = \sum_i b_i Y_i$$

$$\mathrm{cov}(V,U) = \sum_{i,j} a_i b_j \mathrm{cov}(X_i, Y_j)$$

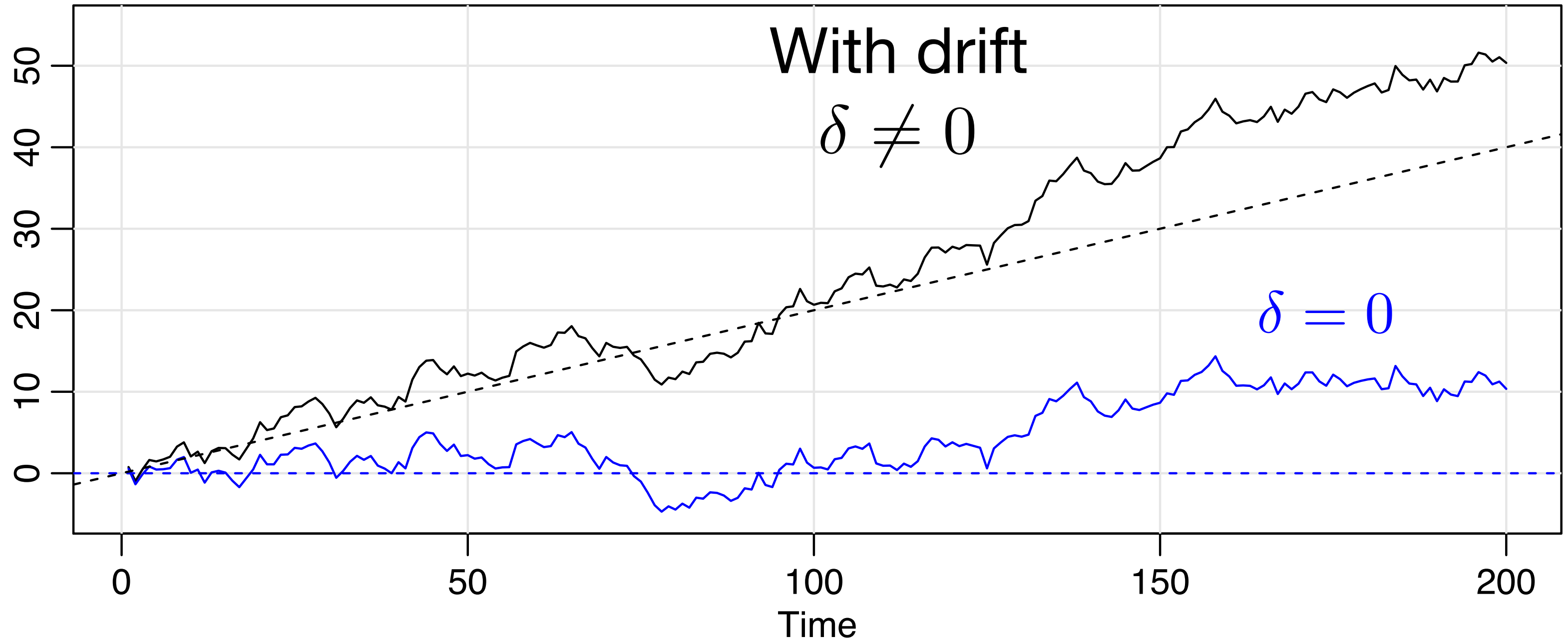
Autoregressive process (AR)

$$x_t = x_{t-1} - 0.9x_{t-2} + w_t$$



More on these next week.

Random walk $x_t = \delta + x_{t-1} + w_t$



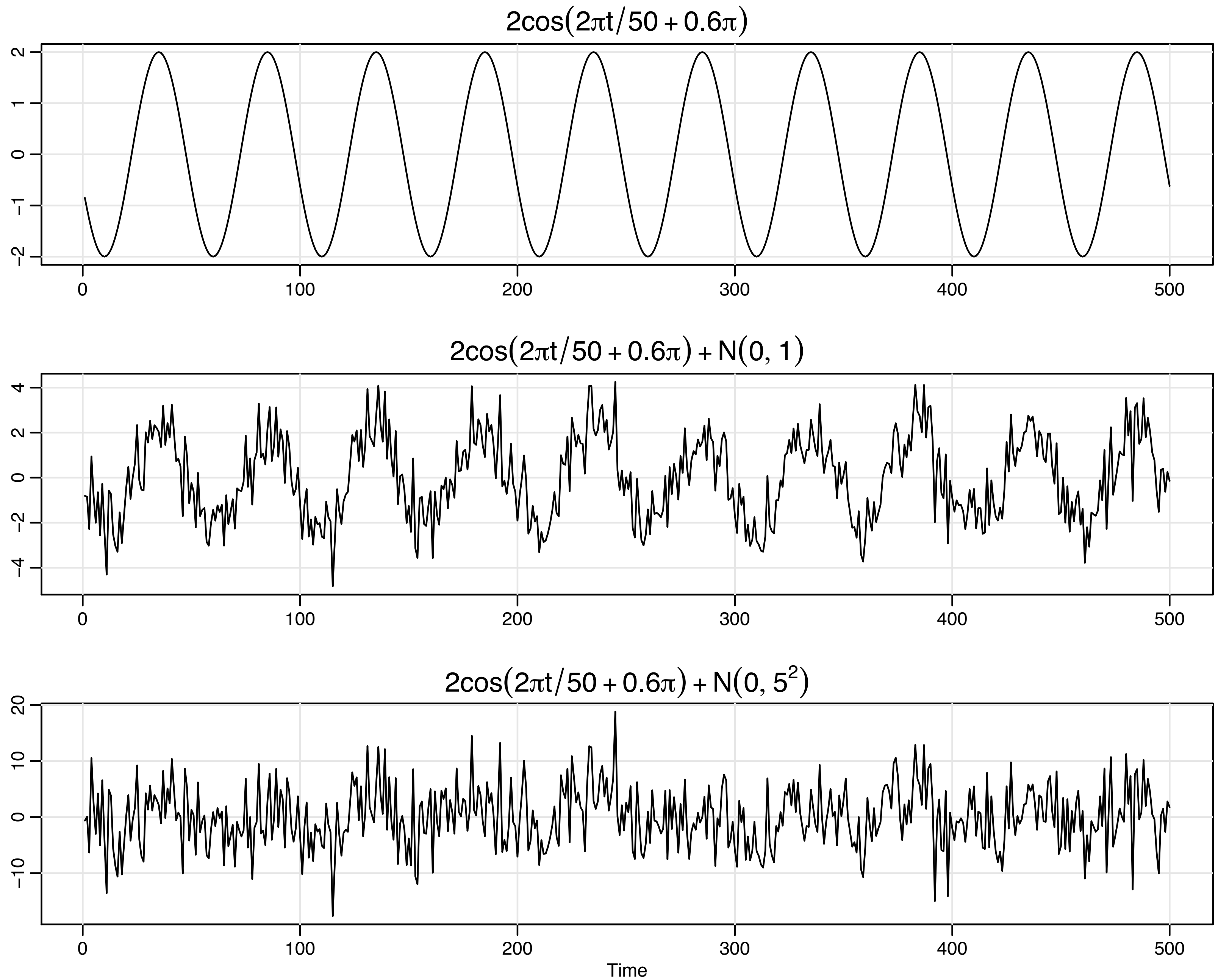
If we unfold recursion:

$$x_t = t\delta + \sum_{i \leq t} w_i$$

$$\mu_X(t) = t\delta$$

$$R_X(t, u) = \min(t, u)\sigma^2$$

Signal vs noise



denoising in frequency domain,
inferring latent structure in temporal domain

Basic statistics of a time series

"Strong stationarity"

$$\{X_t, ..., X_{t+K}\}$$
$$\{X_{t+h}, ..., X_{t+h+K}\}$$

Identically distributed subsets
for all t,h,K

Consequences:

For single variables (K=0) this implies same marginals everywhere

$\mathrm{P}(X_t < x) = \mathrm{P}(X_{t+h} < x)$ for all t,h, and so $\mu_X(t) = \mathrm{cte}$

For single variables (K=1) this implies same pairwise dependencies

Basic statistics of a time series

“(Weak) stationarity”

$$\mu_X(t) = \text{const.}$$
$$R_X(\Delta t) = \text{cov}(X_t, X_{t+\Delta t})$$

+finite variance

Example: moving averages

A strongly stationary process with finite variance is weakly stationary

Converse is more complicated:
a gaussian weakly stationary process is strictly stationary

*Note: change in notation, for stationary processes R_x has a single argument

Basic statistics of a time series

“Trend stationarity”

$$\mu_X(t) \neq \text{const.}$$

$$R_X(\Delta t) = \text{cov}(X_t, X_{t+\Delta t})$$

This means that data can be partitioned into a
time-dependent term + zero-mean stationary process

e.g. sigmoid + white noise

Final note: random walks are non-stationary

Basic statistics of a time series

Linear process

A general version of filtered white noise

$$x_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \qquad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Causality

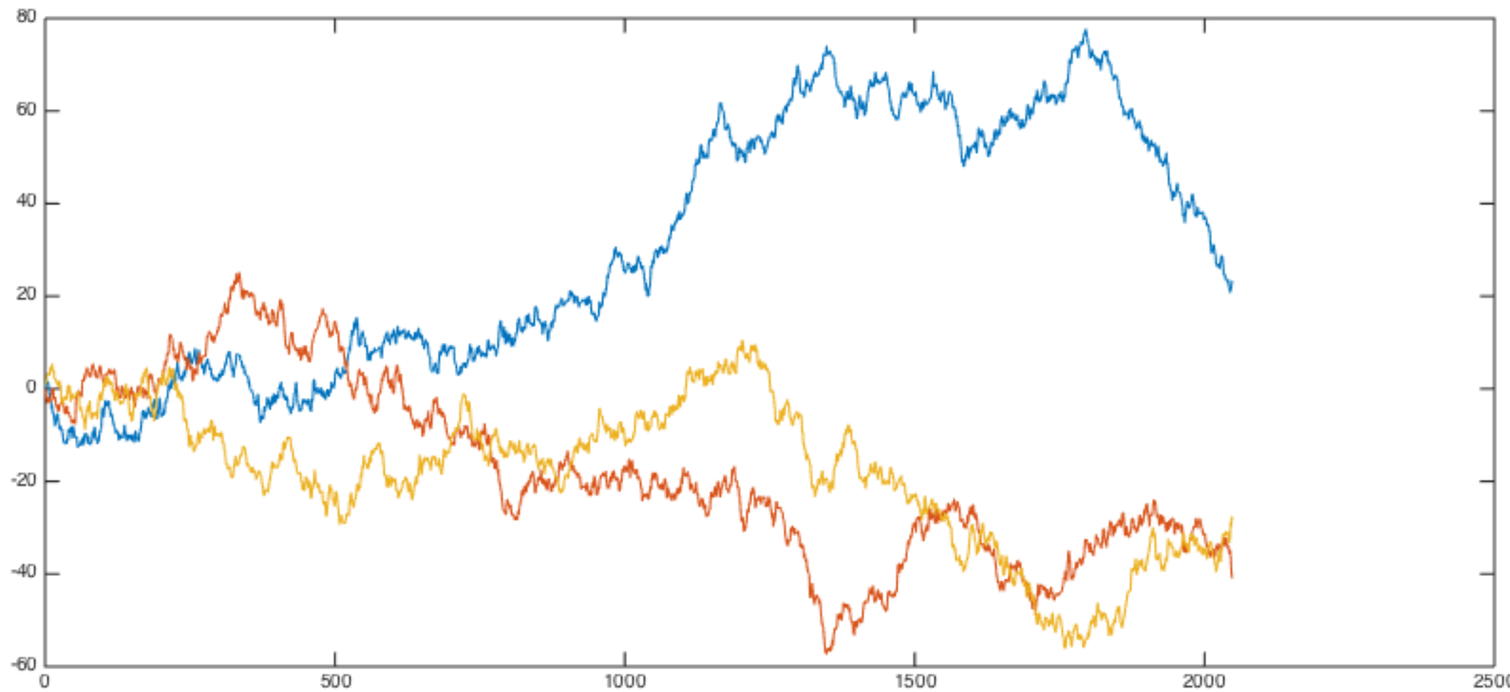
Present depends on past but not on future

It’s often a natural assumption for real world data

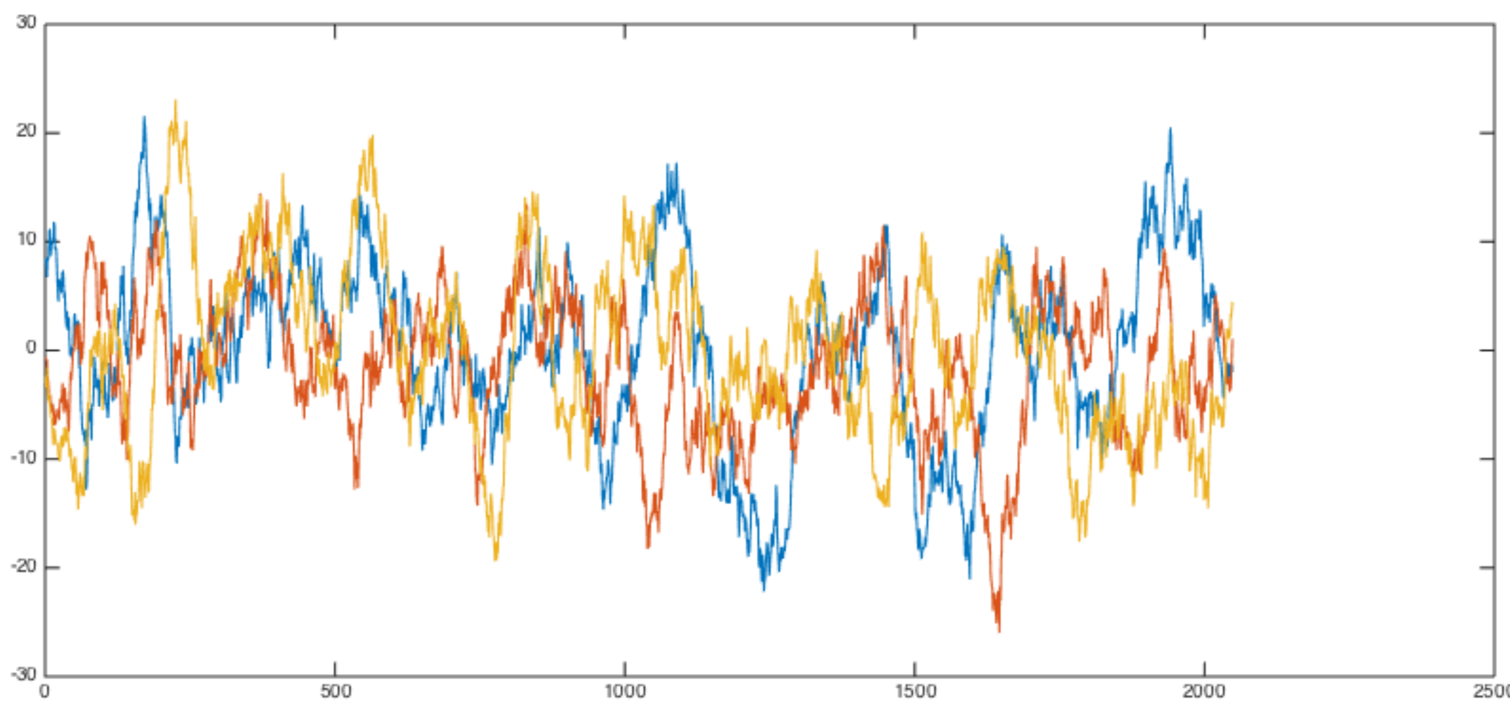
Not necessarily true for all sequential data (e.g. DNA)

Causal linear process:

$\psi_j = 0$ for future components (j<0)



Random walk
(non stationary)



stationary

Simpler structure,
Easier to estimate

Empirical measurements (stationary process)

$$\boxed{\hat{\mu}_x = \frac{1}{T} \sum_t x_t}$$

$$\boxed{\hat{R}_x(\Delta t) = \mathrm{cov}(x_t, x_{t+\Delta t})}$$