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: Mo, 4-5pm, Room 608

TAs
Caroline Haimerl, Artie Shen
Office hours: Thu, 11am, TBD

Couse page: https://github.com/savinteachingorg/pTSAfall2019
Piazza: https://piazza.com/nyu/fall2019/dsga3001001/home

[Quick feedback much appreciated. (anonymous)]

Course logistics: grading

5 problem sets 25%
Primarily derivations, 2w for each

Lab work 20%
Coding: **python**, weekly, ideally finished during lab

Midterm 25%
Oct 28th, ARIMA+ LDS+HMMs

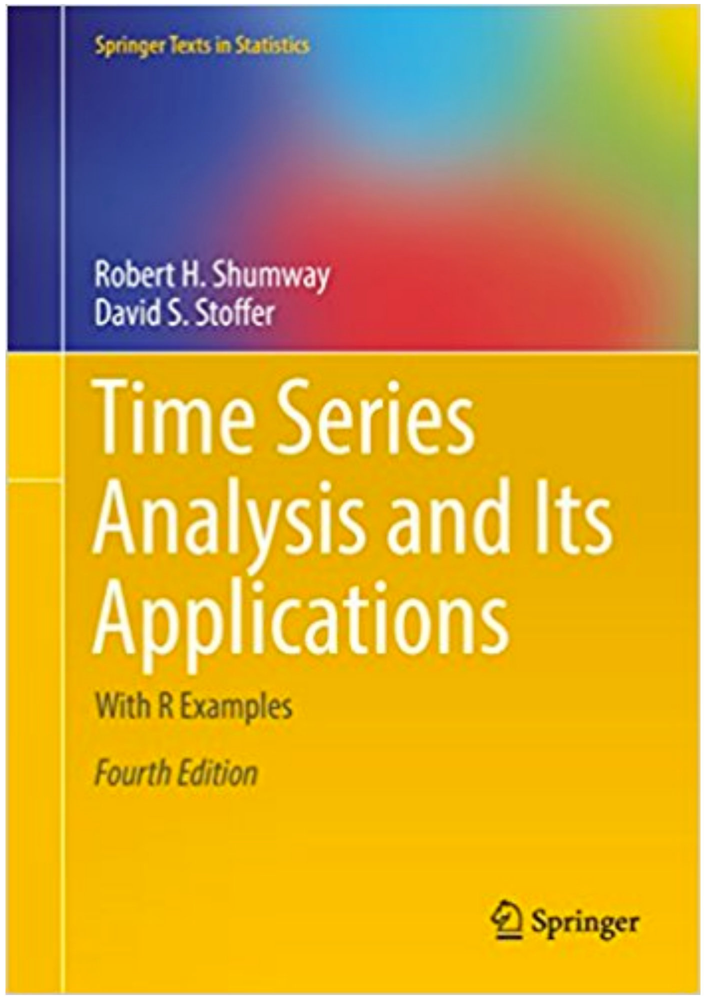
Project 25%
Groups of 2-3, topic of choice
Project proposal due **Oct. 7th**

Participation 5%
Class discussions, office hours, piazza

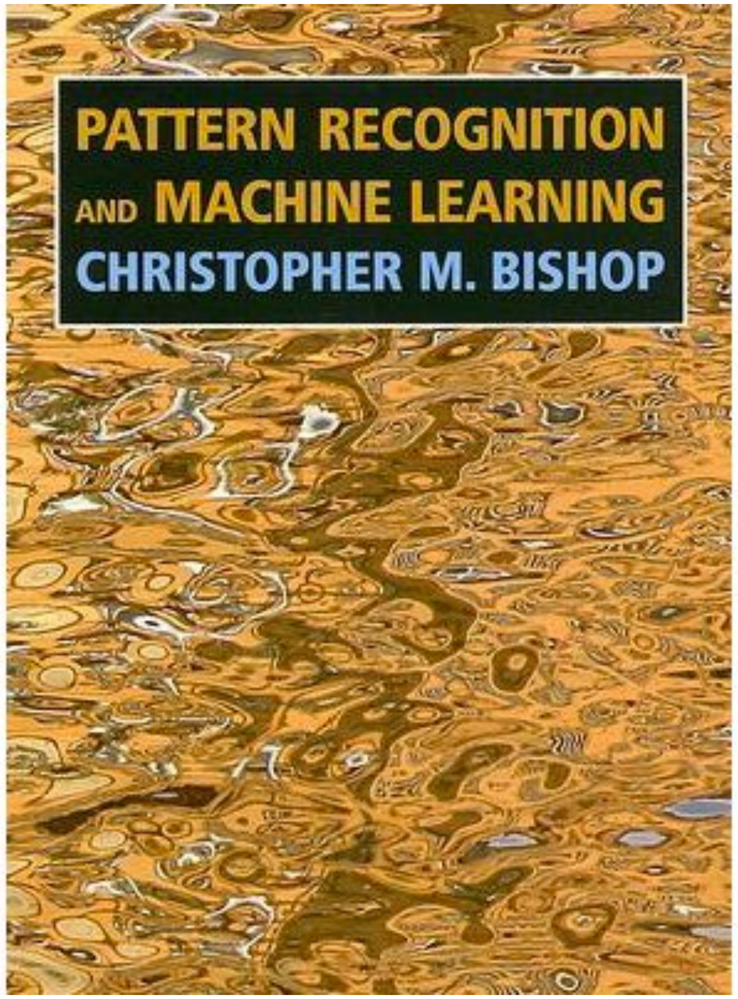
No official prerequisites, but if you haven't seen much **probability** and **linear algebra** recently, a quick refresher is strongly encouraged

Bibliography

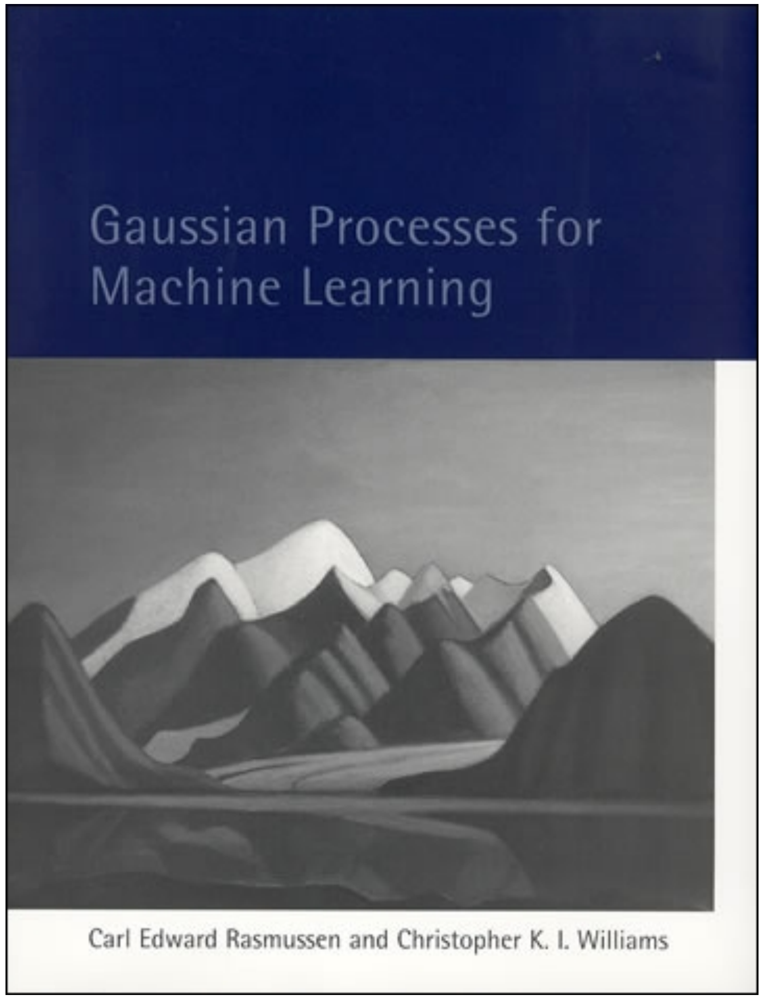
No course textbook, handouts for each section. Lectures based on:



Intro, AR(I)MA, Spectral methods



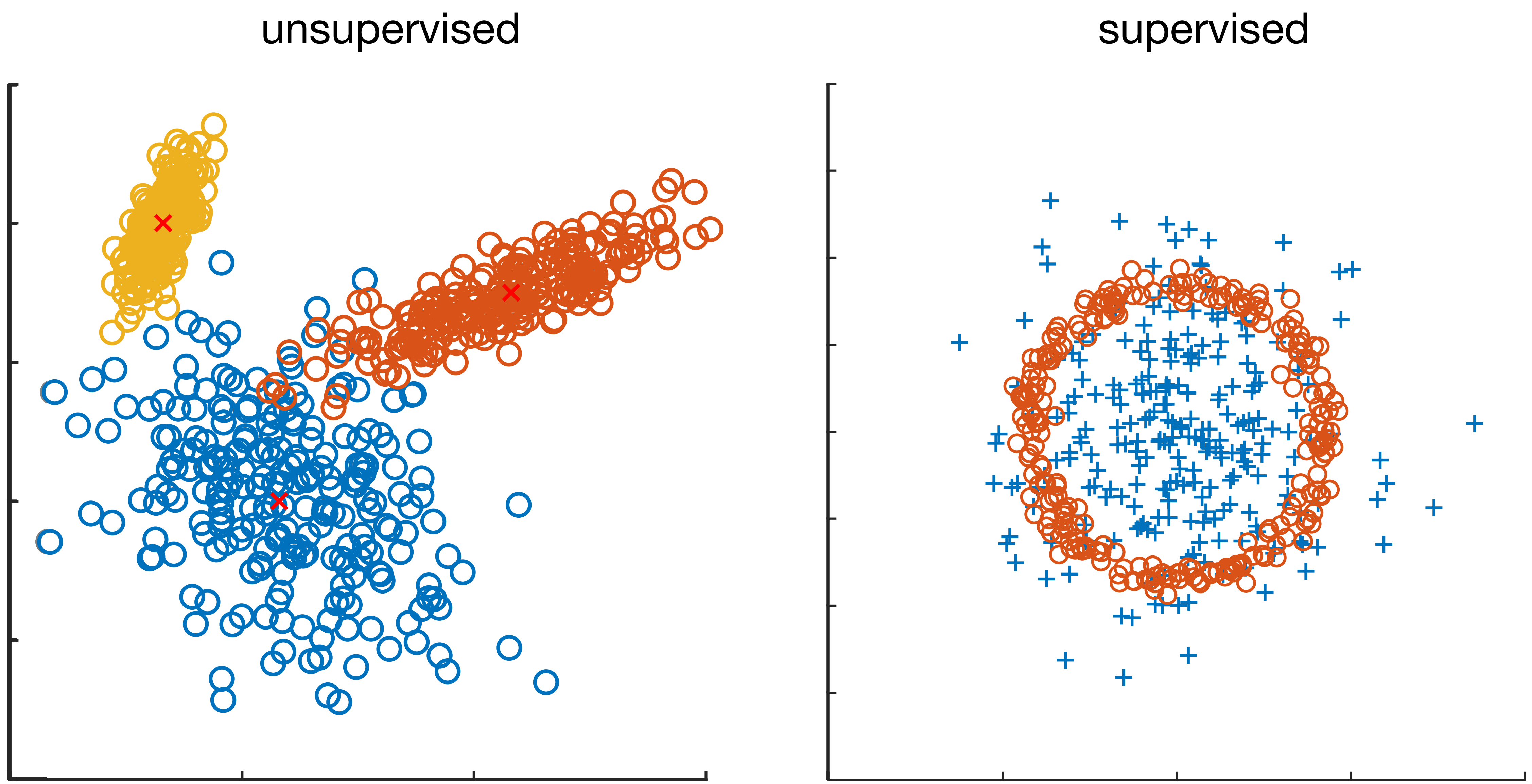
Latent space models



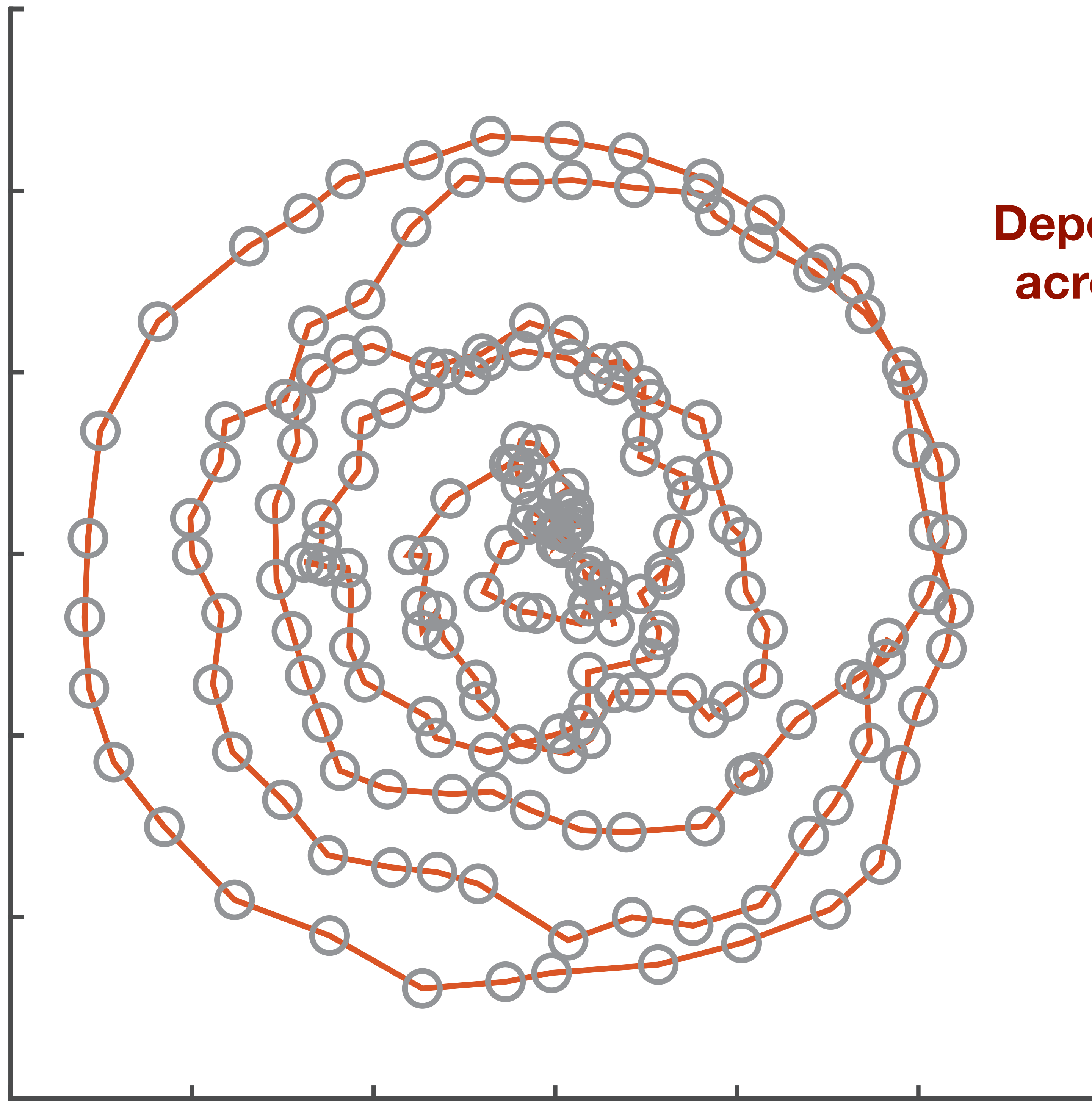
GP

What is probabilistic
time series analysis?

All ML: finding structure and using it to make predictions



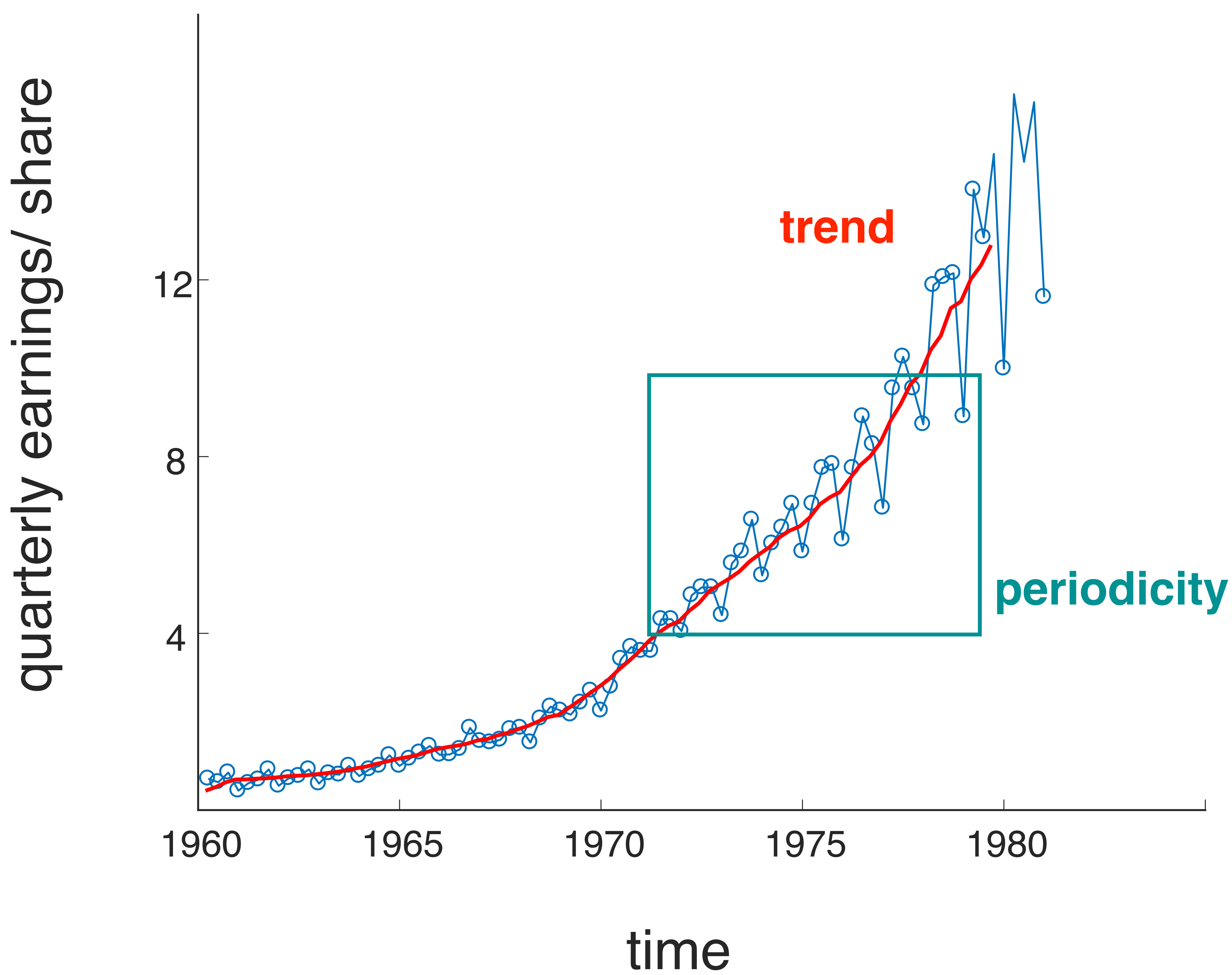
TSA: finding structure and using it to make predictions in <u>sequential data</u>



Dependencies across time!

Some concrete examples

Task: predict future earnings, interpret data structure

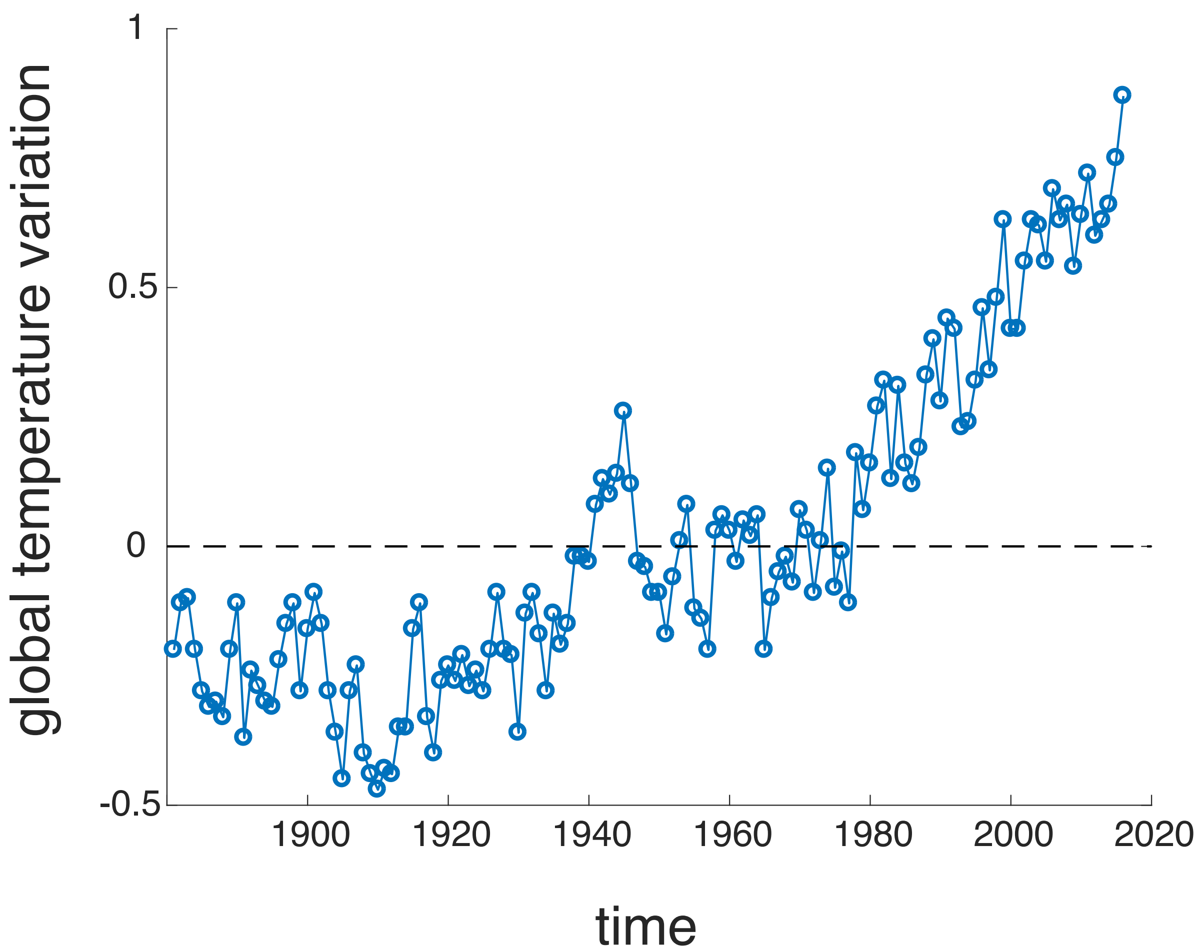


Johnson & Johnson

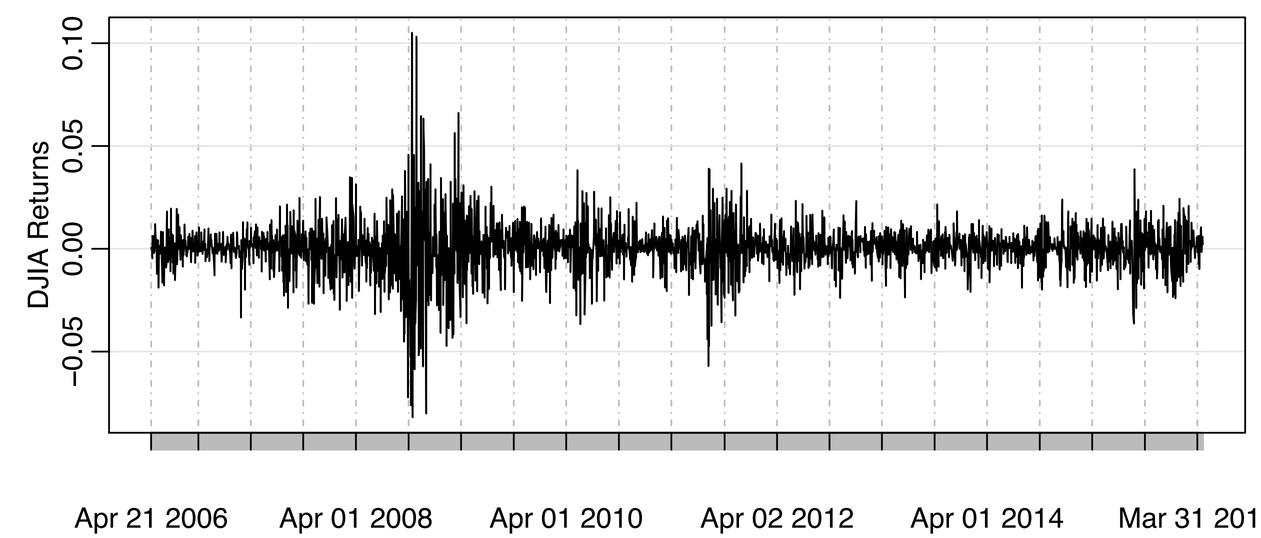
Task: identifying trends, do things change systematically over time?

Task: detect high volatility periods

Global warming

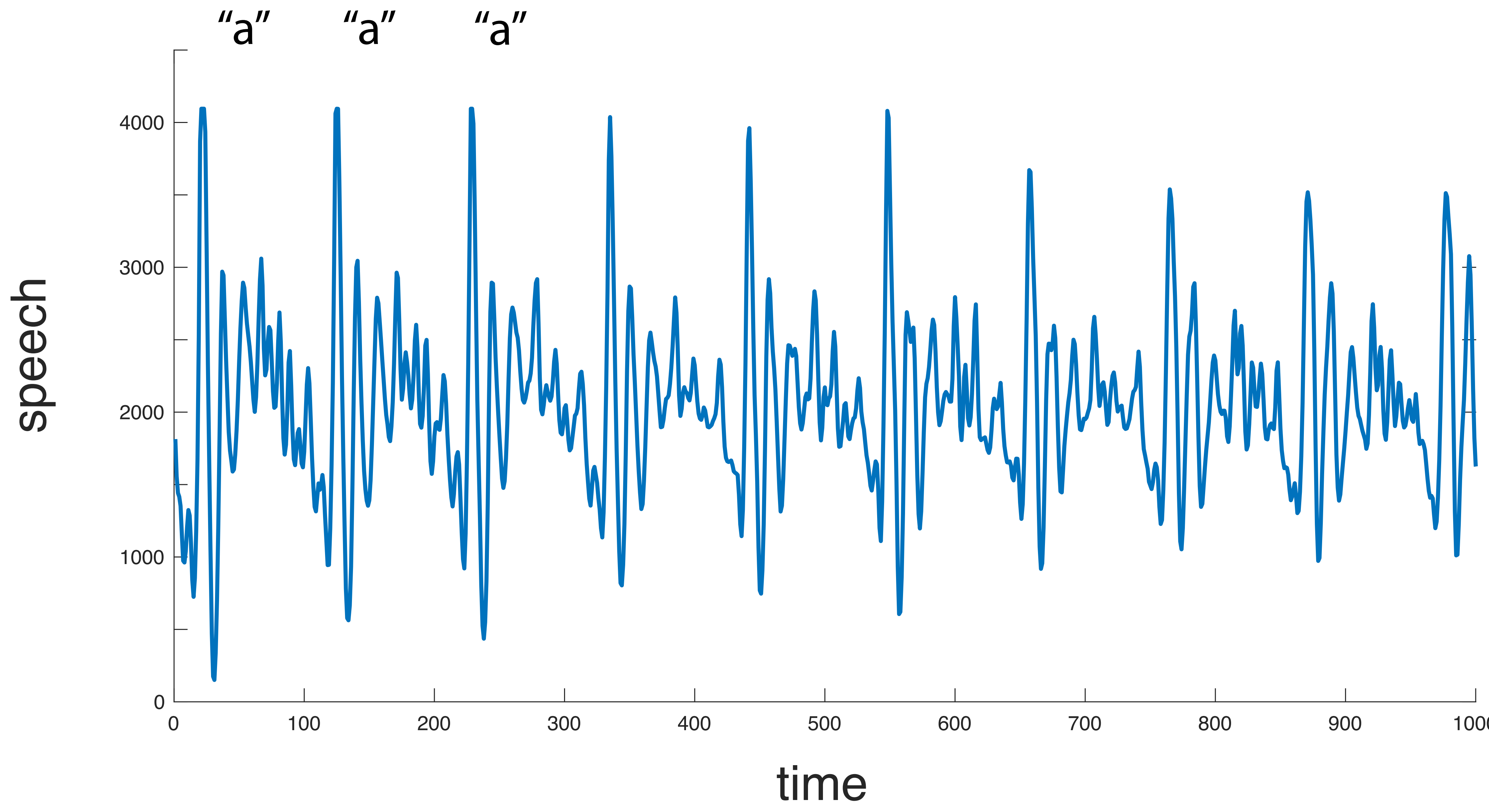


Financial crisis

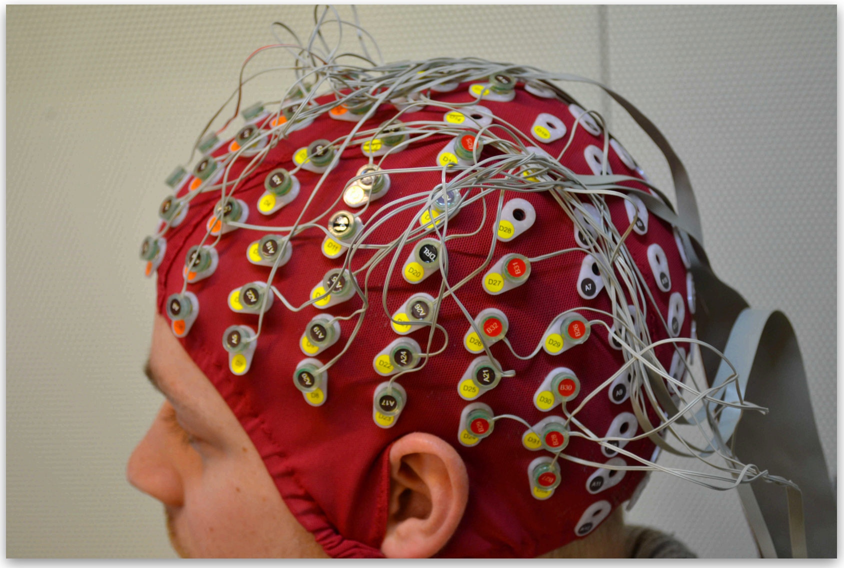


DJIA: *Dow Jones Industrial Average*

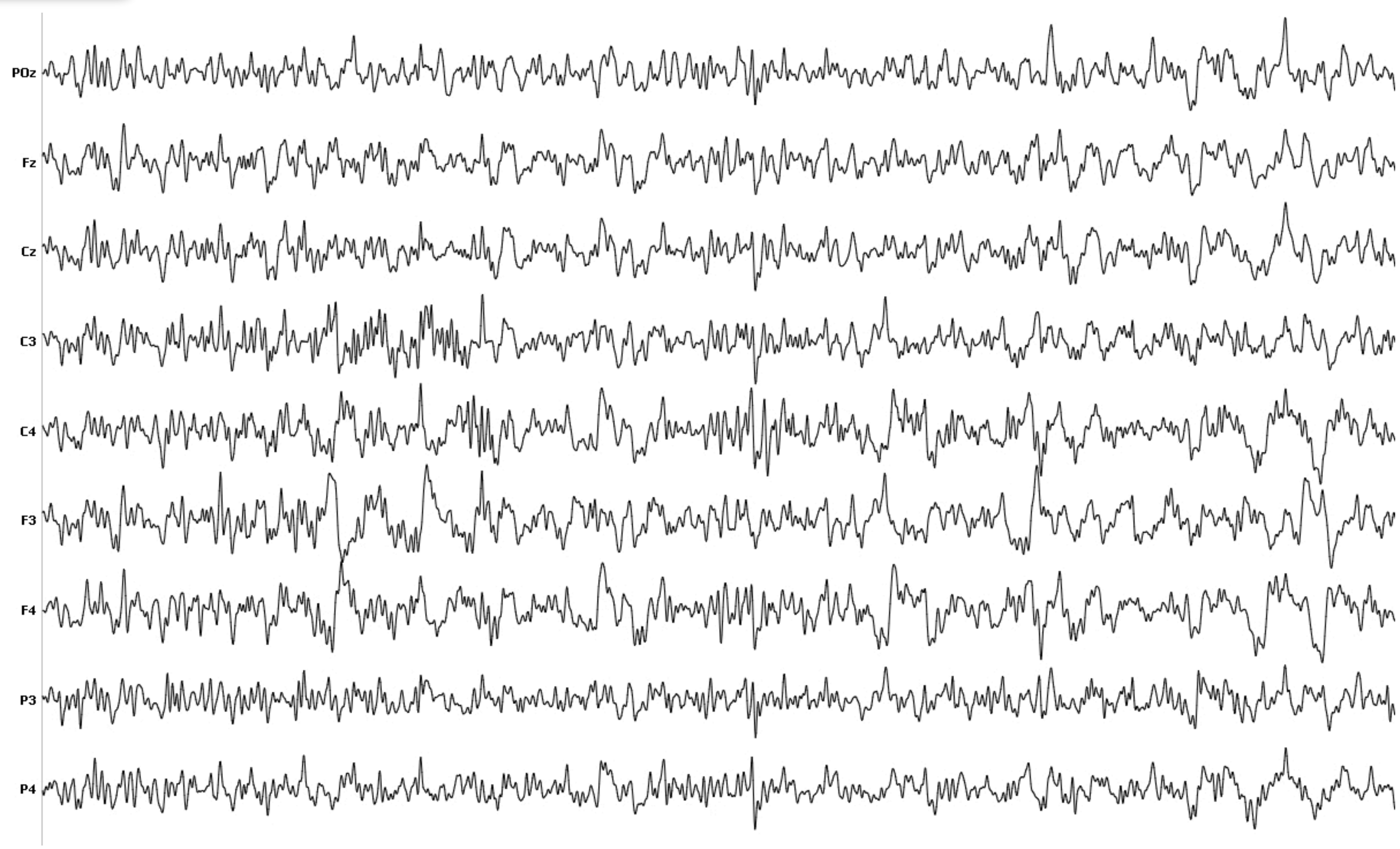
Task: infer discrete latent structure from analog signals



Voice recognition

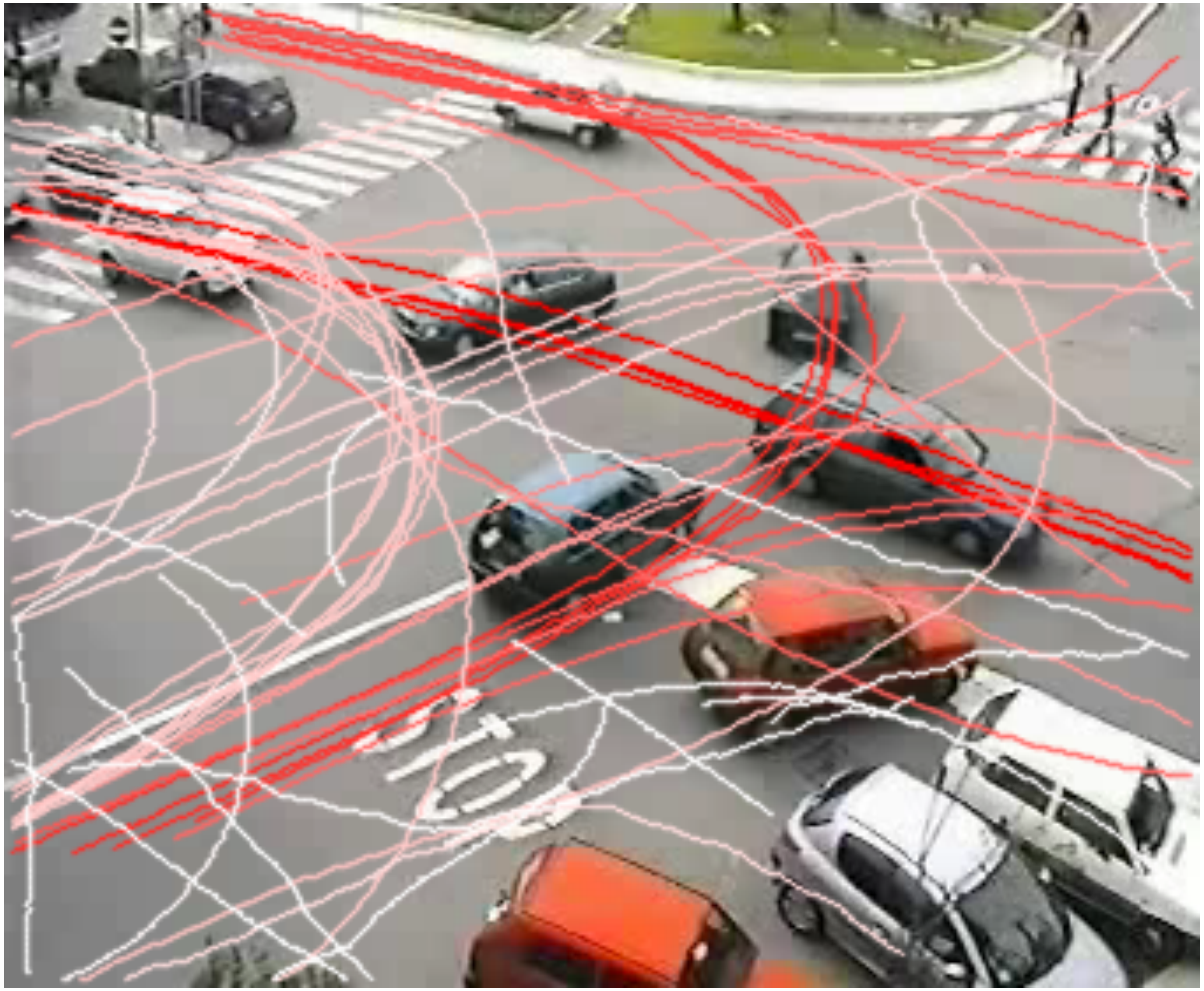


Multivariate time series: EEG



Task: decoding
Brain-computer interfaces

Self driving cars

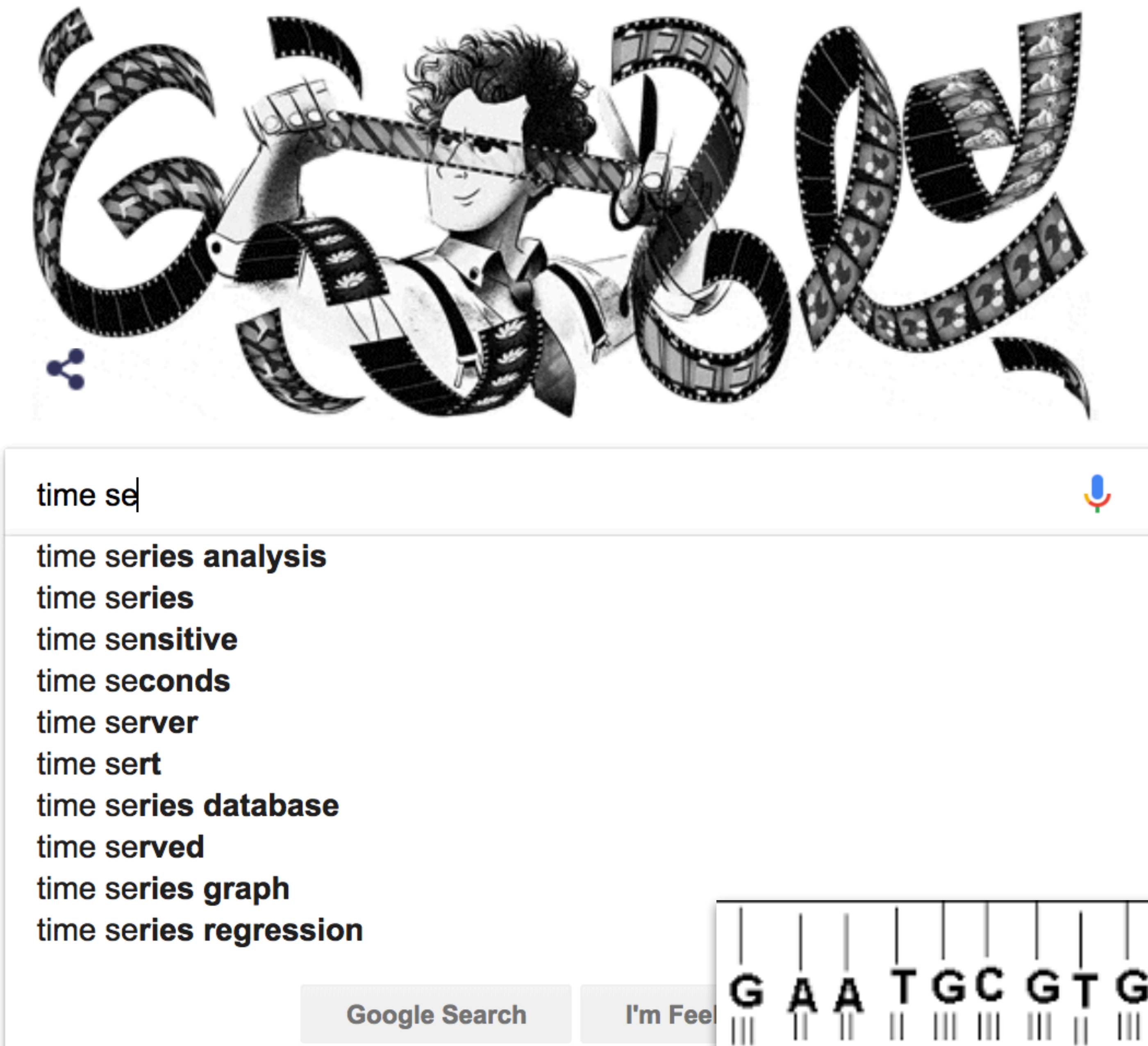


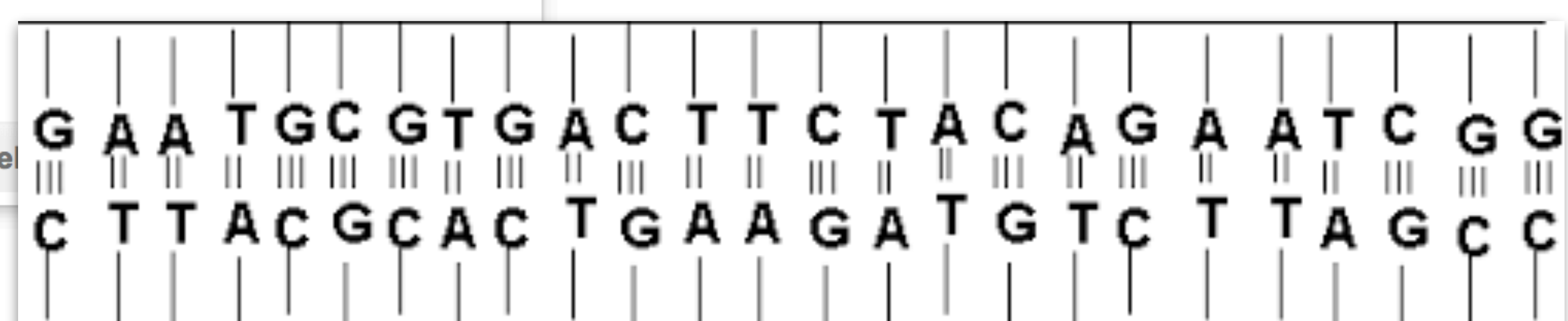
Tasks: simulation, control

We need uncertainty representation for
optimal decision making, risk minimization

From Saunier & Sayed 2008

Intro: what is a time series?

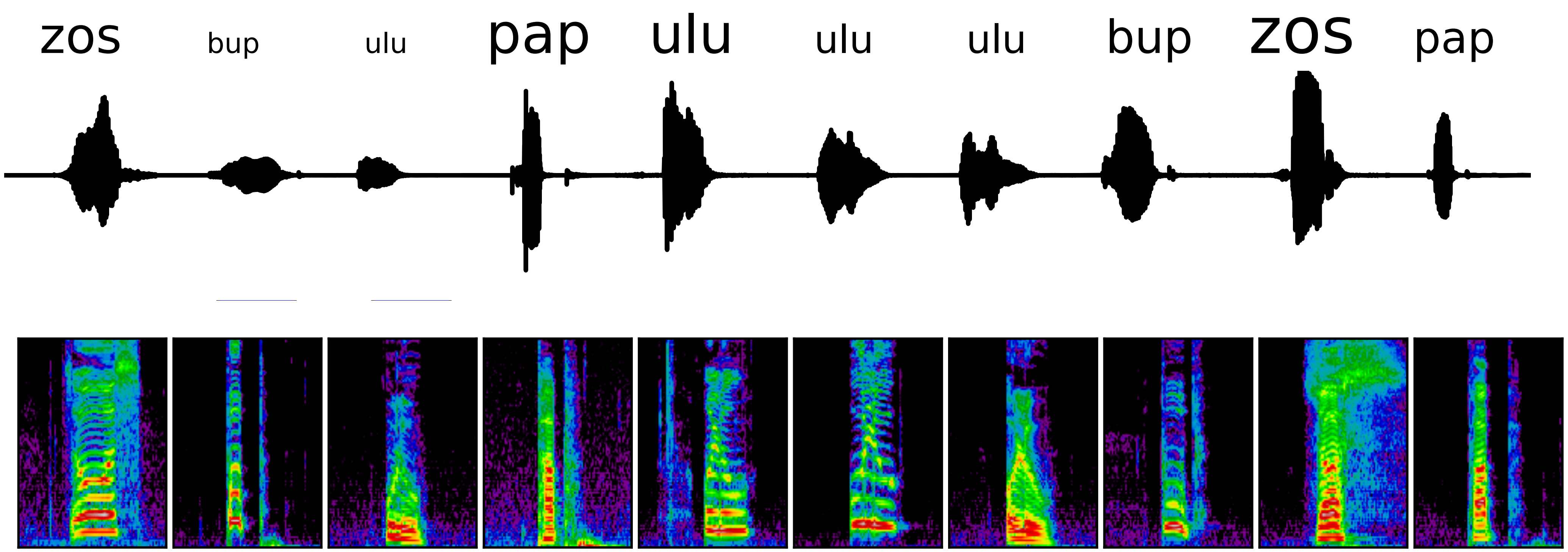




Language!
NLP
Machine translation

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

Intro: what is a time series?

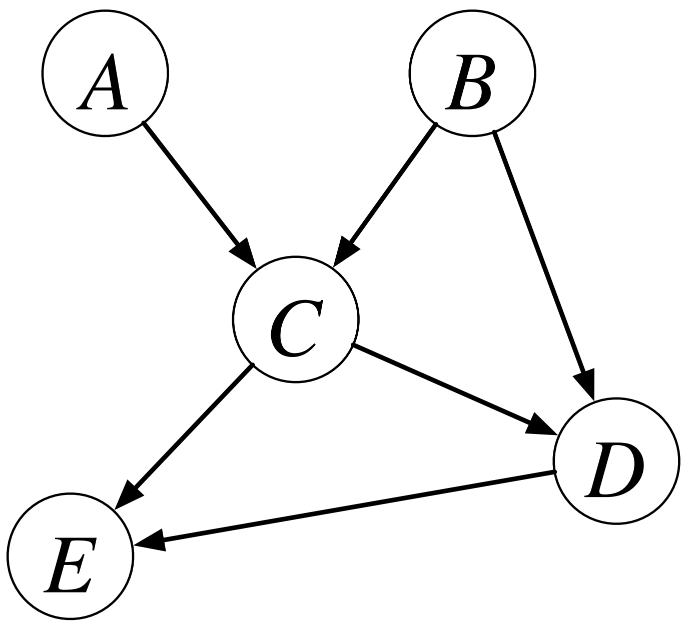


Model data in the frequency domain

e.g. automated speech recognition

Tasks: identify latent structure, denoising…

Why **probabilistic** time series analysis



1. Define <u>generative models</u> capturing relevant statistical structure in the data
(vs. recognition models)

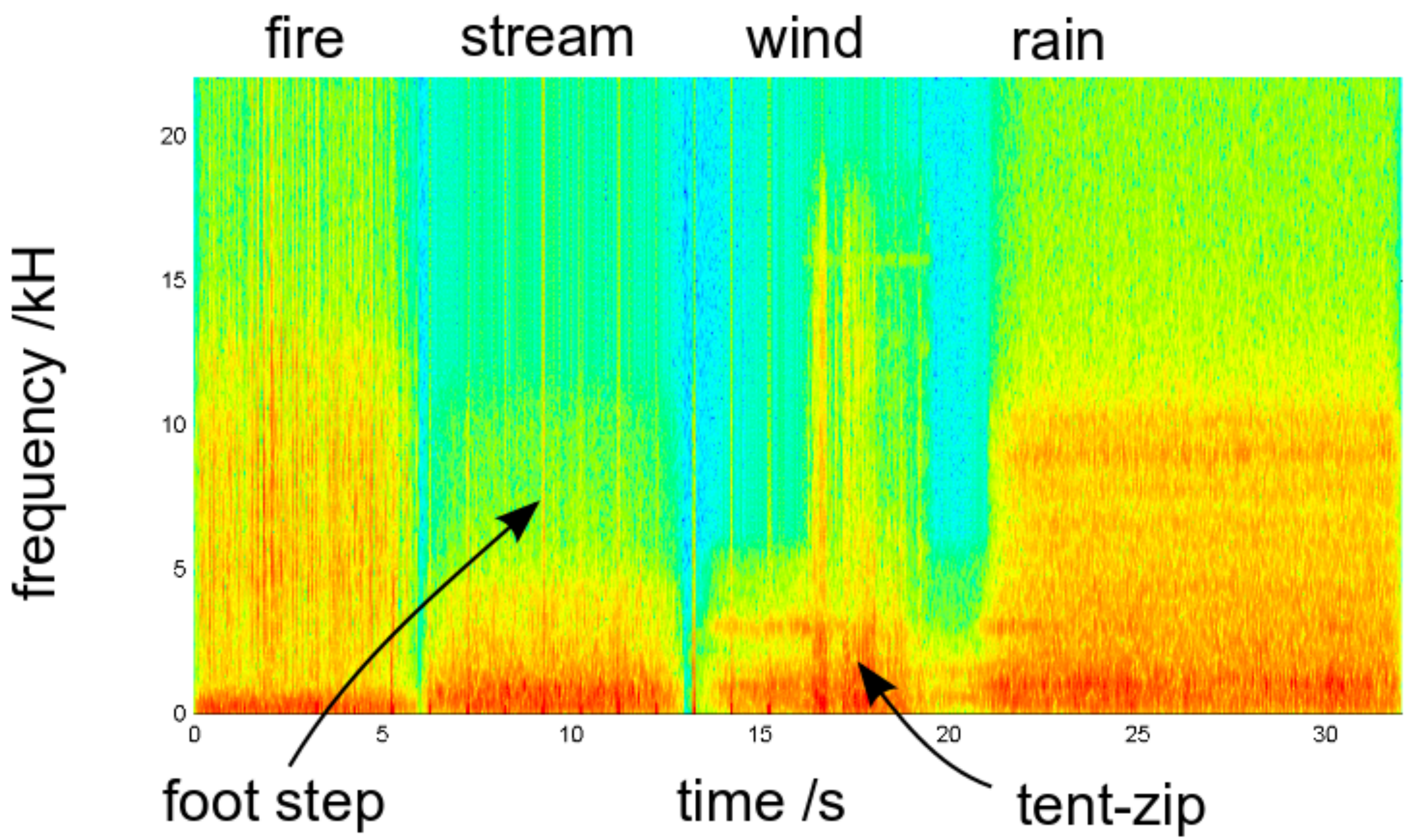
Depiction**: graphical model**

$$p(A, B, C, D, E) = p(A)p(B)p(C|A, B)p(D|B, C)p(E|C, D)$$

2. Fit models to data

3. Make predictions

4. Generate artificial data with same statistics



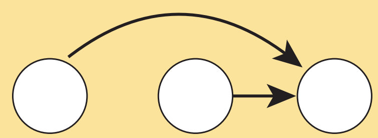
Overview of the course

TIME

Lagged relationships

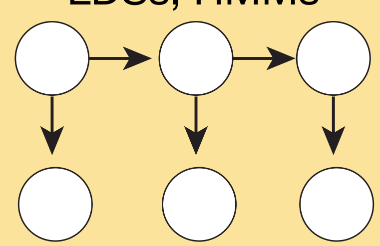
Model temporal dependencies directly

AR+ friends



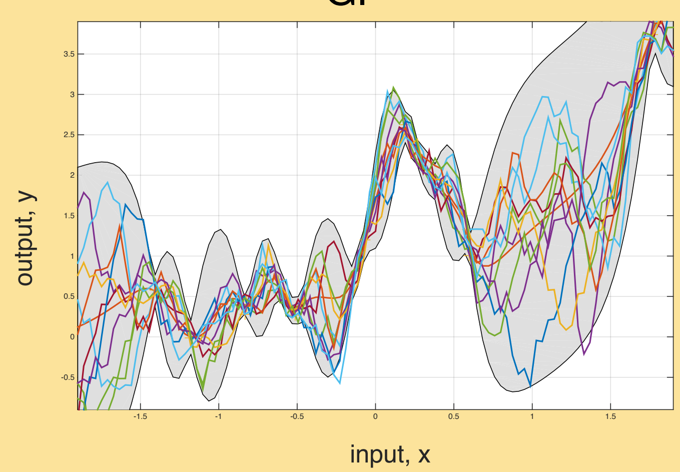
Latent structure

LDSs, HMMs



Distribution function

GP



FREQUENCY

Periodic structure

Stationary processes seasonality/periodicity

Probabilistic spectral analysis

Mixing it up

Non stationary spectral structure (?)

Guest lecture: example state of the art

Intro: what is a time series?

Definitions. Basic statistics

What is a time series?

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

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

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

"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*:

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

These **dependencies** are also the main problem: they make the math hard.

***Intractable in general, we limit ourselves to more structured classes of distributions*

Basic statistics of a time series

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

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

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

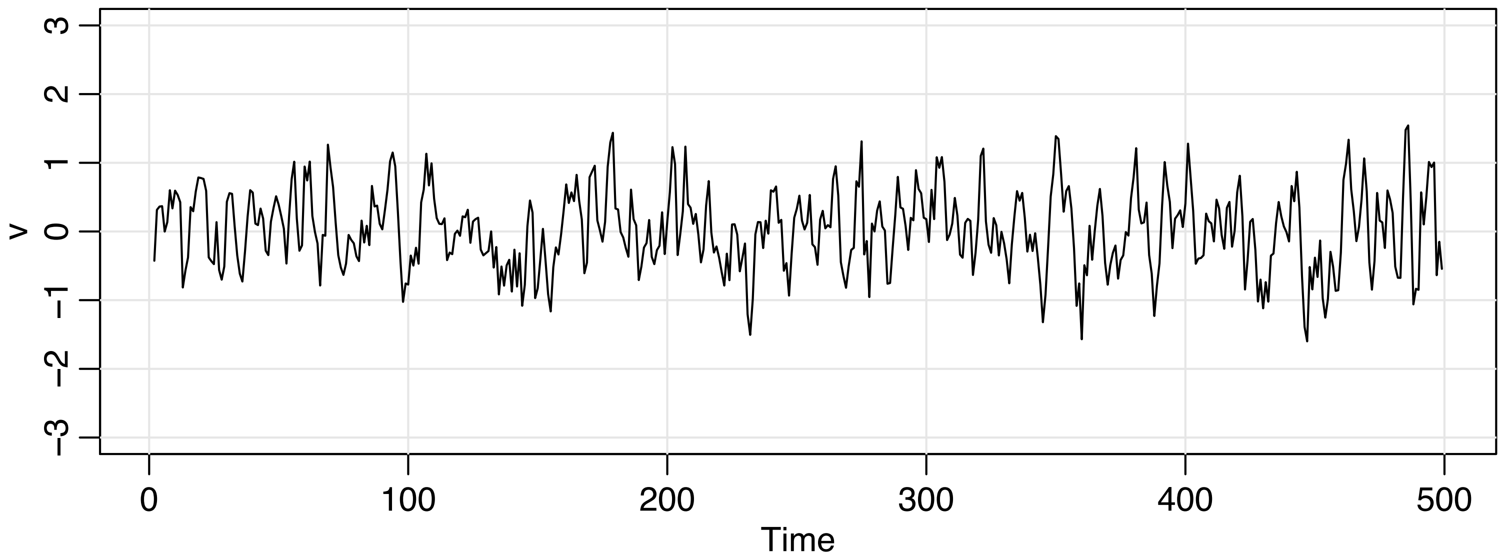
$$-1 \leq \rho_X(t,u) \leq 1$$

Cross-Covariance $$R_{X,Y}(t,u) = \mathrm{cov}\left(X_t, Y_u\right)$$

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.



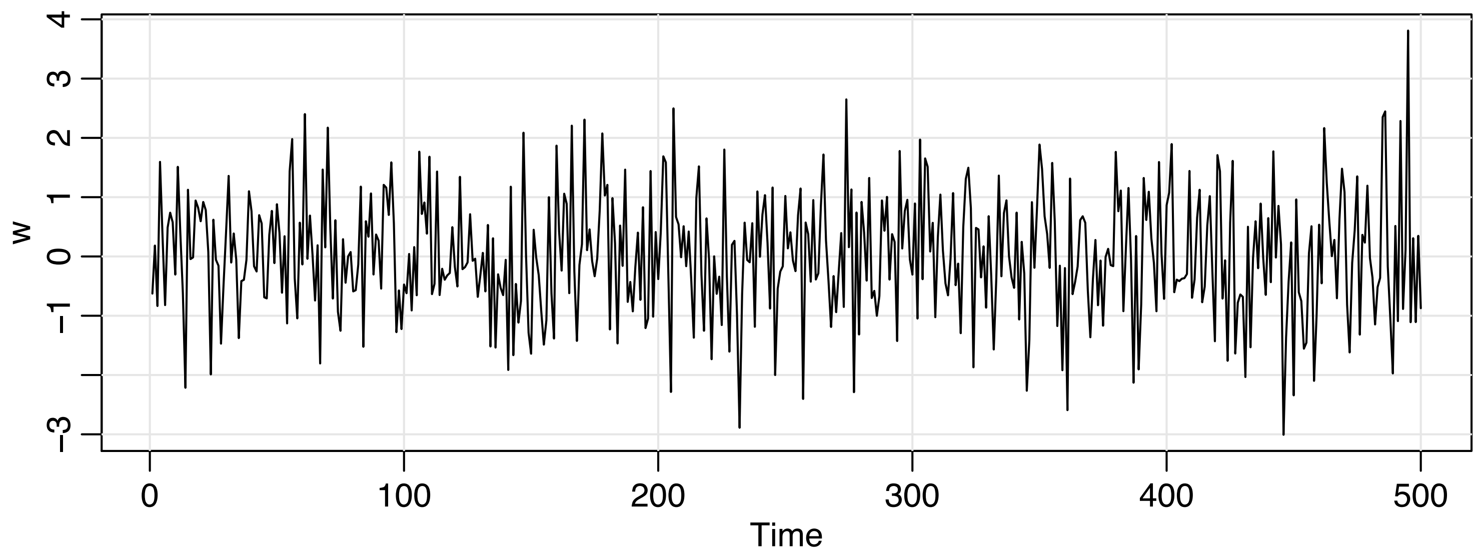
Trivially, white noise has

$$\mu_X(t) = 0$$

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

Example stochastic processes

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



filtered white noise

$$\mu_V(t) = 0$$

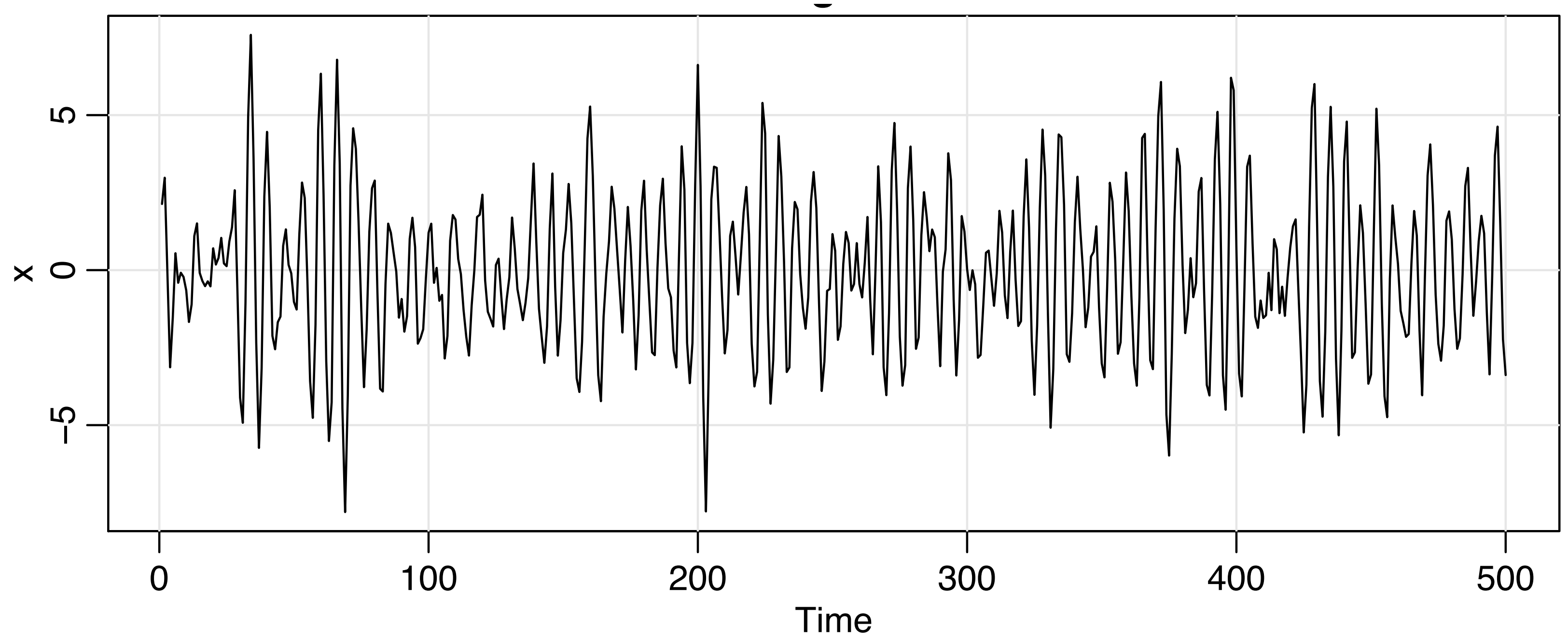
$$R_V(t,u) = \begin{cases} 1/3\,\sigma^2 & , t = u \\ 2/9\,\sigma^2 & , |t-u| = 1 \\ 1/9\,\sigma^2 & , |t-u| = 2 \\ 0, & |t-u| > 2 \end{cases}$$

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

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

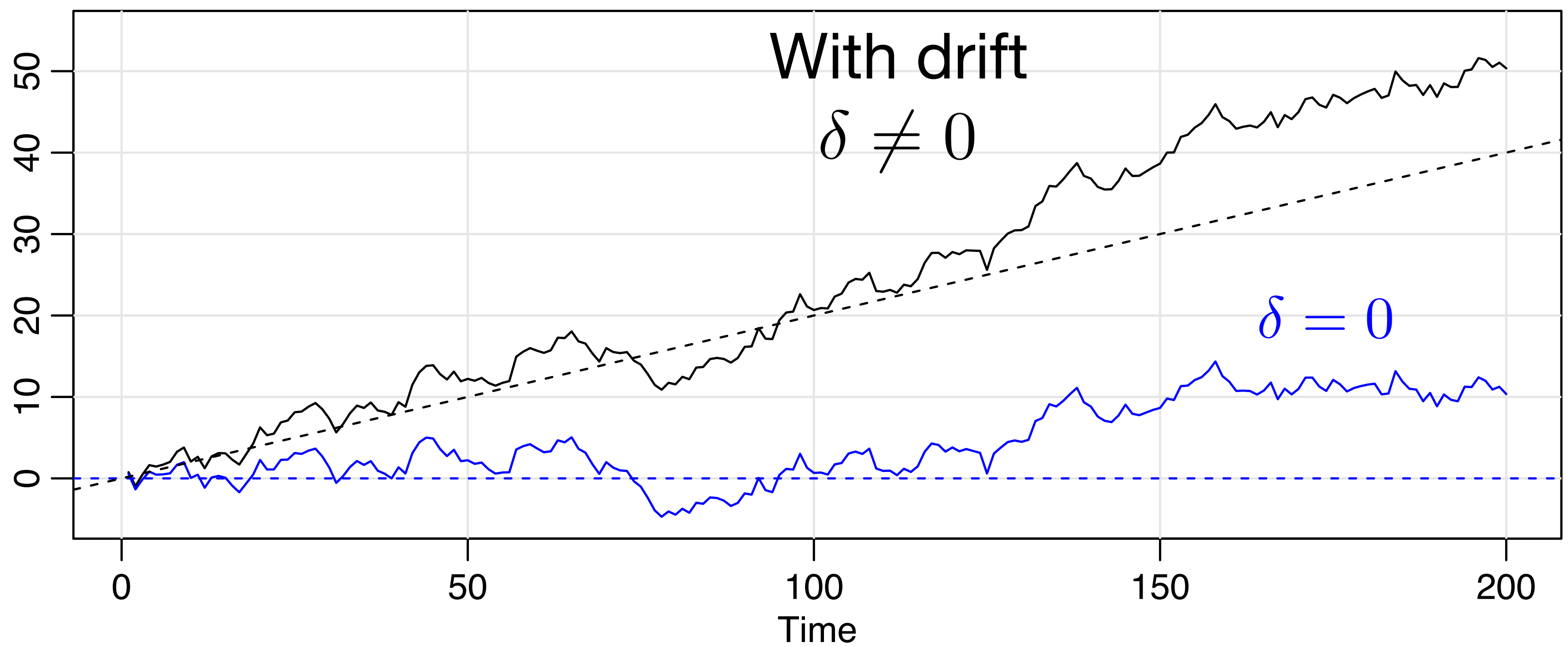
Autoregressive process (AR)

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



These are simple examples of ARIMA models —discussed in L2 .

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



If we unfold recursion:

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

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

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

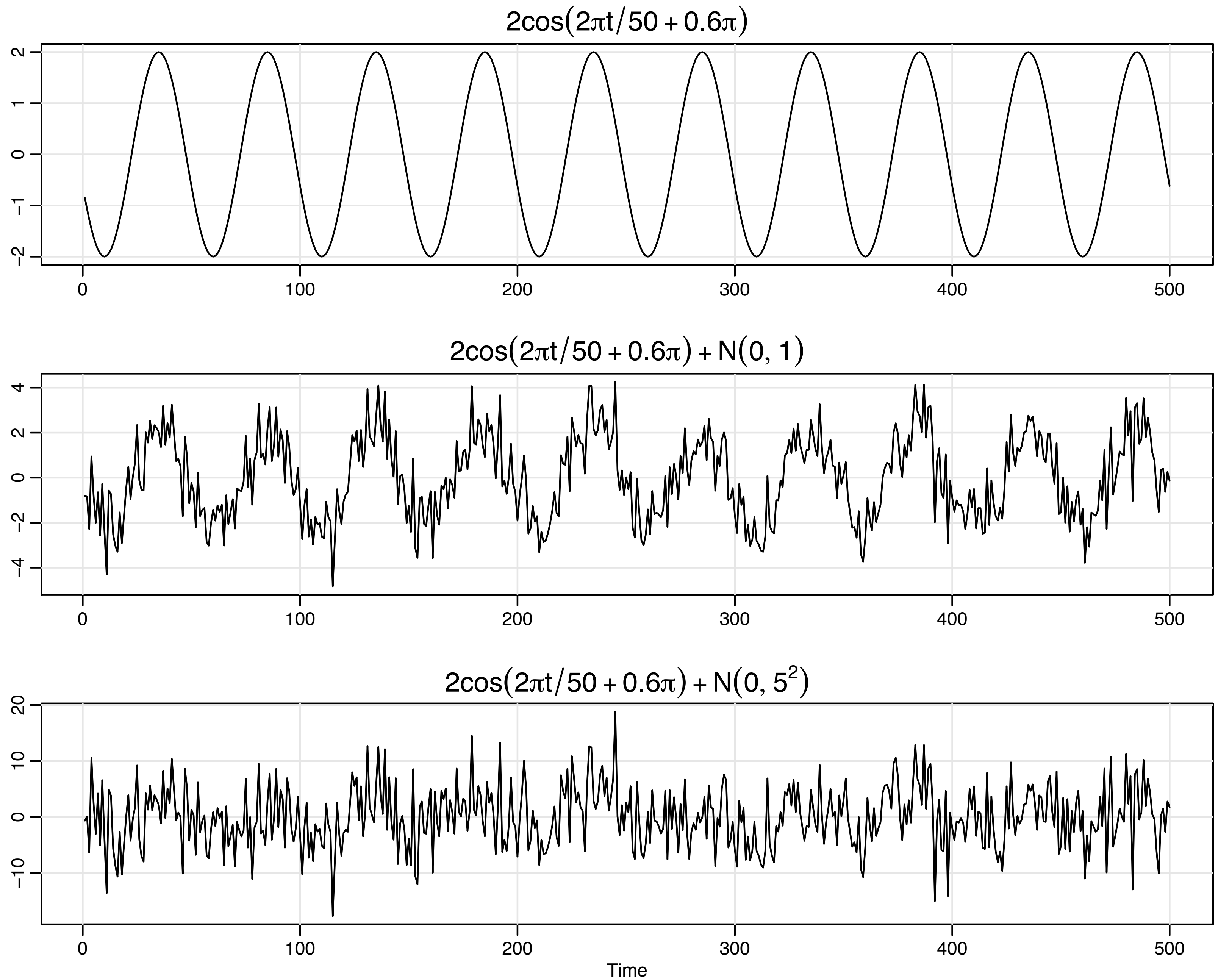
Cross-Covariance $$R_{X,Y}(t,u) = \text{cov}\left(X_t, Y_u\right)$$

Cross-Correlation Function (ACF) $$\rho_{X,Y}(t,u) = \frac{R_{X,Y}(t,u)}{\sqrt{R_{X,Y}(t,t),R_{X,Y}(u,u)}}$$

E.g. $$x_t = w_t + w_{t-1} \quad \text{and} \quad y_t = w_t - w_{t-1},$$

$$\rho_{xy}(h) = \begin{cases} 0 & h = 0, \\ 1/2 & h = 1, \\ -1/2 & h = -1, \\ 0 & |h| \geq 2. \end{cases}$$

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

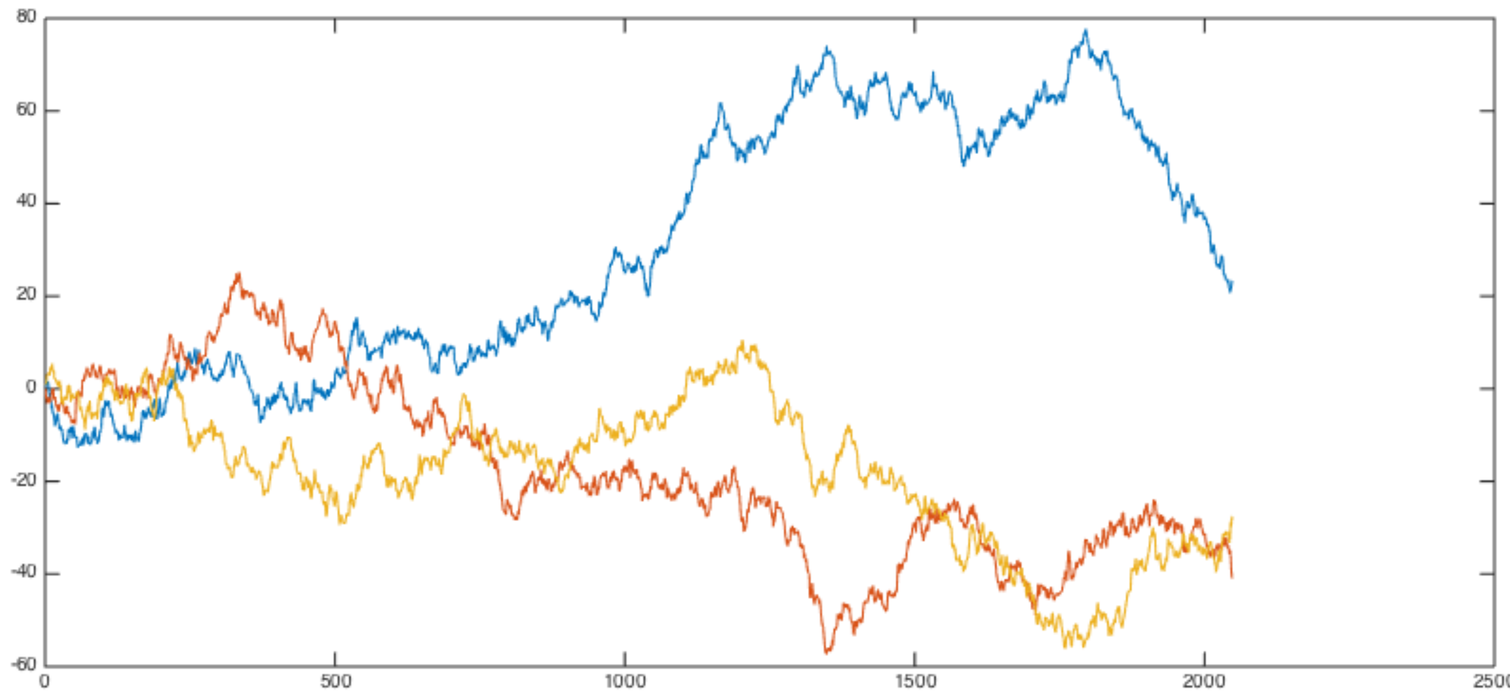
Present depends on past but not on future

It's often a natural assumption for real world data

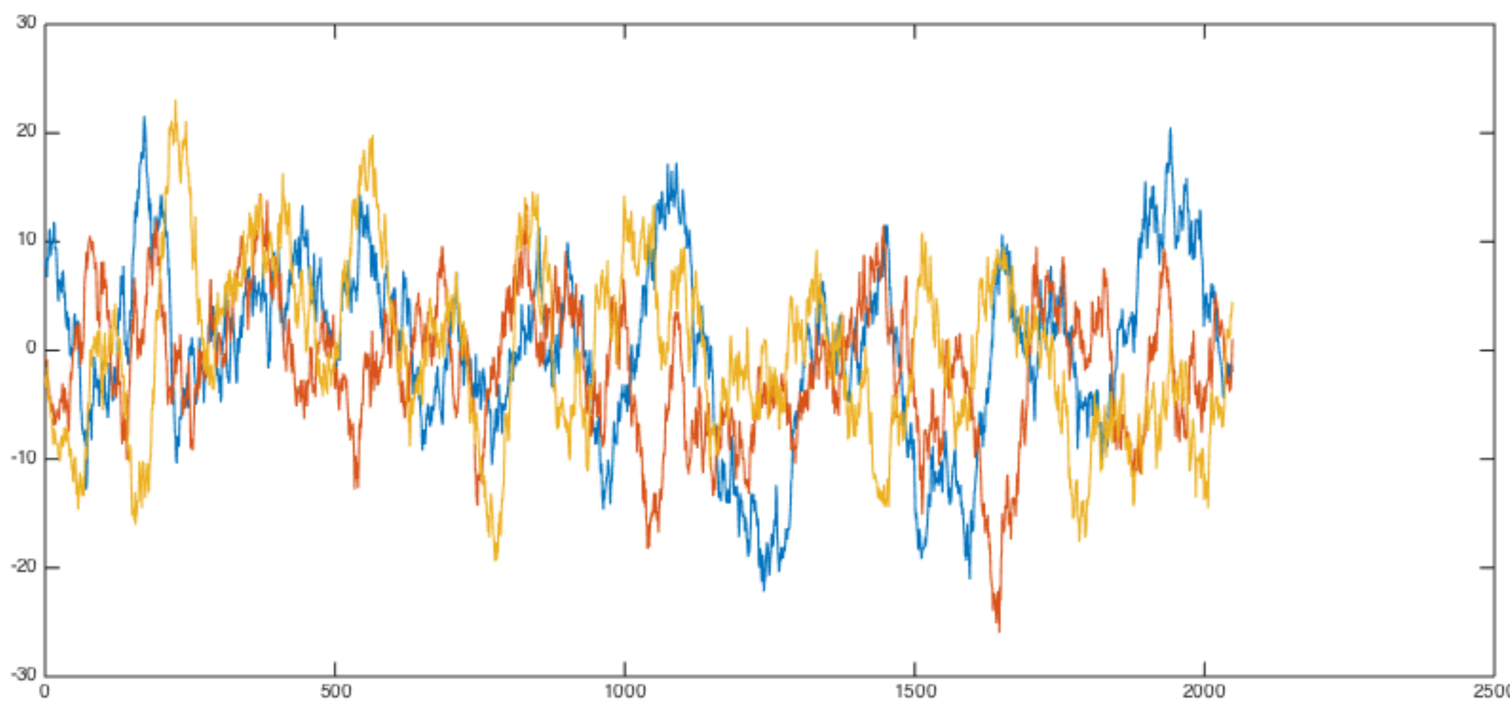
Not necessarily true for all sequential data (e.g. DNA)

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)

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

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

NEXT WEEK: AR(I)MA