# Lecture 4: Physiological Sensor Data

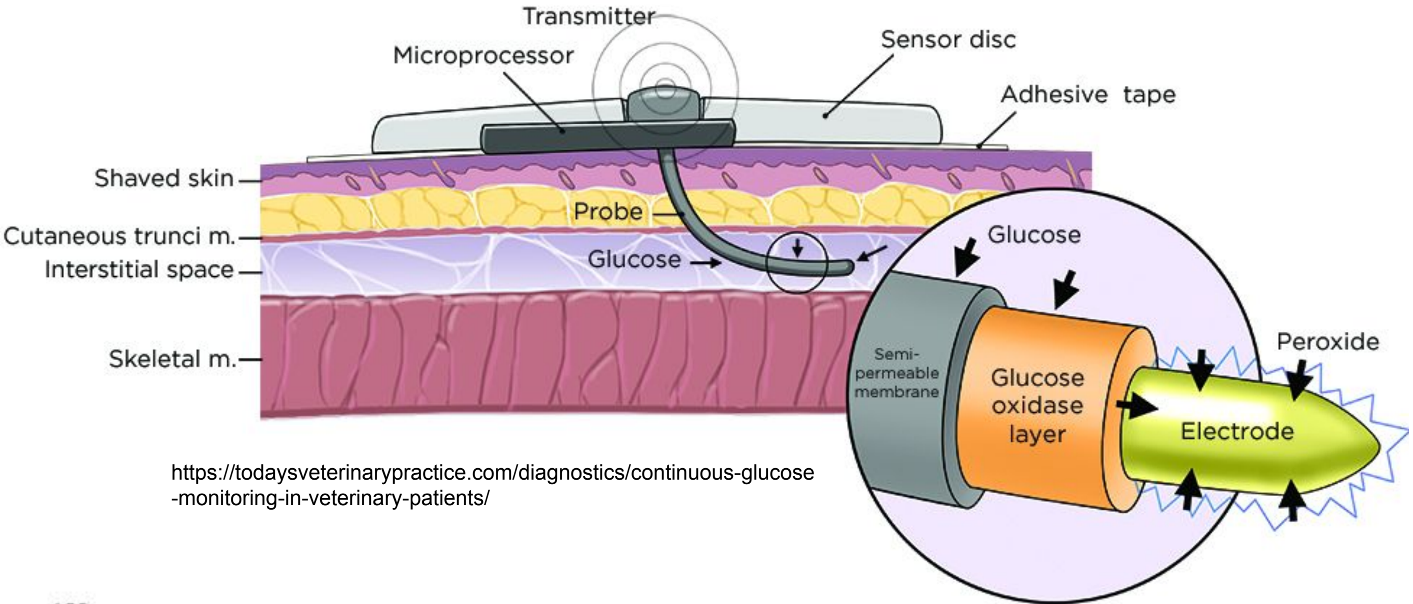
CSCI6XXX/CHE6XXX/CSCI4XXX
(CSCI6093)

Finlay Maguire (finlay.maguire@dal.ca)

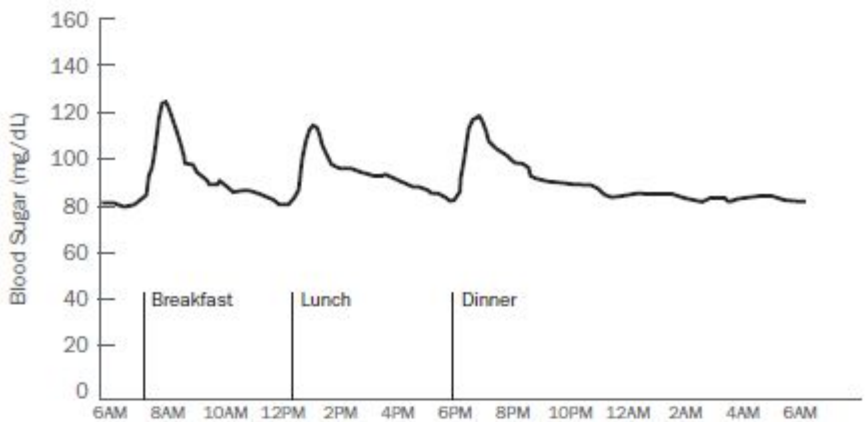
# Learning Overview

- Types of medical sensor data
- Time-domain approaches: detrending/regression models
- Alternative decomposition: frequency/time-frequency
- State-space approaches: hidden markov models
- Handling data from multiple sensors
- General purpose Bayesian approaches: Gaussian Processes
- Cough-detection example
- Segmentation of heartbeats example
- Seizure prediction example

# Physiological sensors (typically) capture data over time




https://todaysveterinarypractice.com/diagnostics/continuous-glucose-monitoring-in-veterinary-patients/

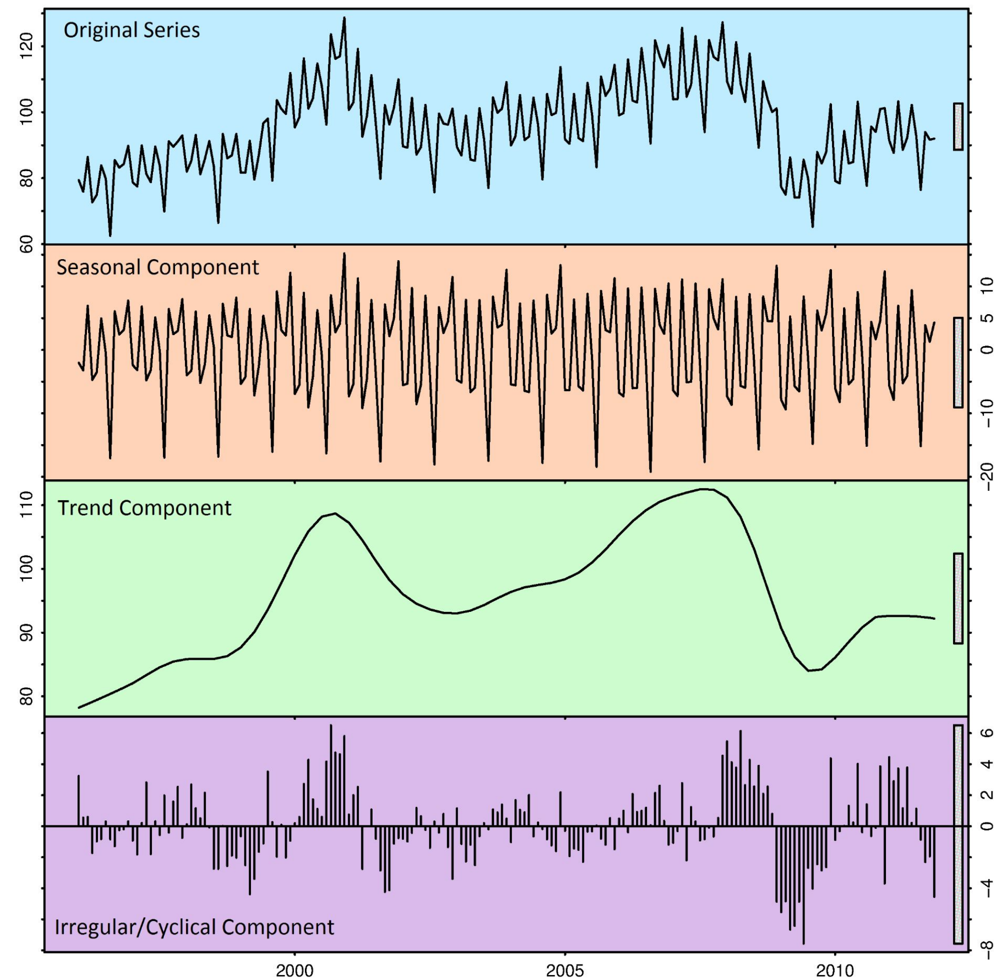



https://www.diabetesdaily.com/learn-about-diabetes/understanding-blood-sugars/is-my-blood-sugar-normal/

- Time-series: set of values ordered by time
- Single object observed over time vs cross-section of multiple objects on common time axis
- Simple 1-dimensional variable: continuous glucose monitoring

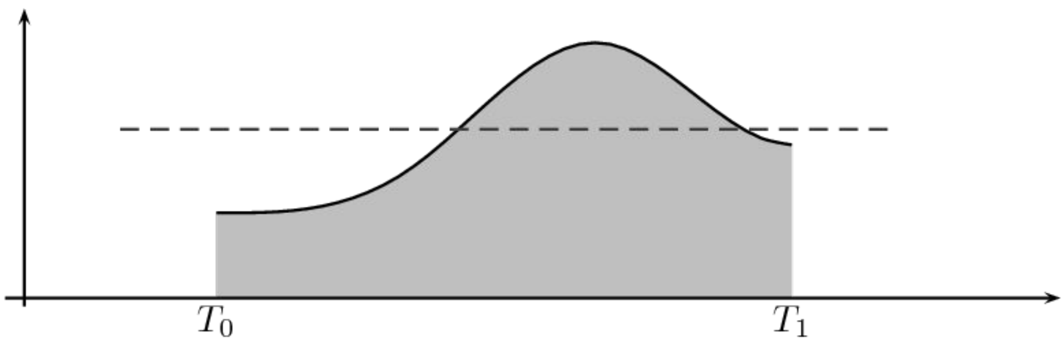
# Time-series often have multiple components

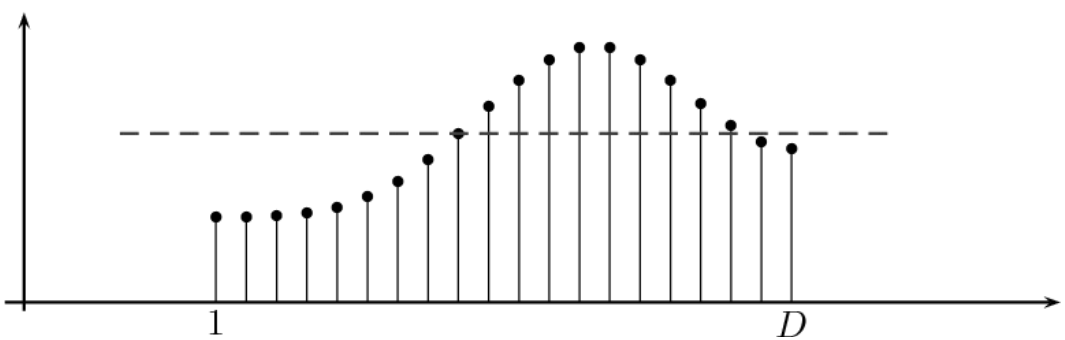
- Short period fluctuations: seasonality
- Long period fluctuations: cycles
- Long-term directionality: trend
- Noise: stochasticity
- Correlation in successive times: autocorrelation
- Distribution changes over time: stationarity/non-stationarity



# Time can be a discrete or continuous value

- Continuous time: defined at every real value of ***T***
- Discrete time: defined at discrete intervals of ***T***
- Real-world data is continuous but sampled during collection as discrete data (*sampling rate*)
- Conversion to some fidelity is possible
- Impacts analysis methods



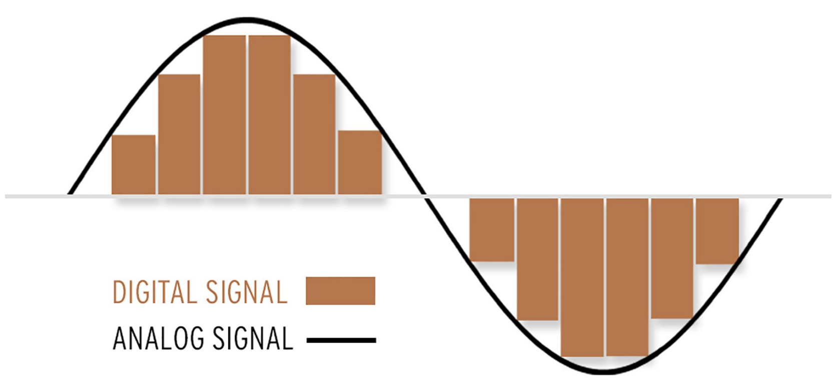




https://www.sp4comm.org/webversion/livre.html

# Physiological sensors capture signal data

- "Signal" is a broadly defined term
- For today: signals are analogue or digital representations of analogue physical quantities.
- Typically electrical representations created by a transducer
- Data encoded in voltage, current and/or frequency
- Medicine: often directly capturing bioelectrical signals

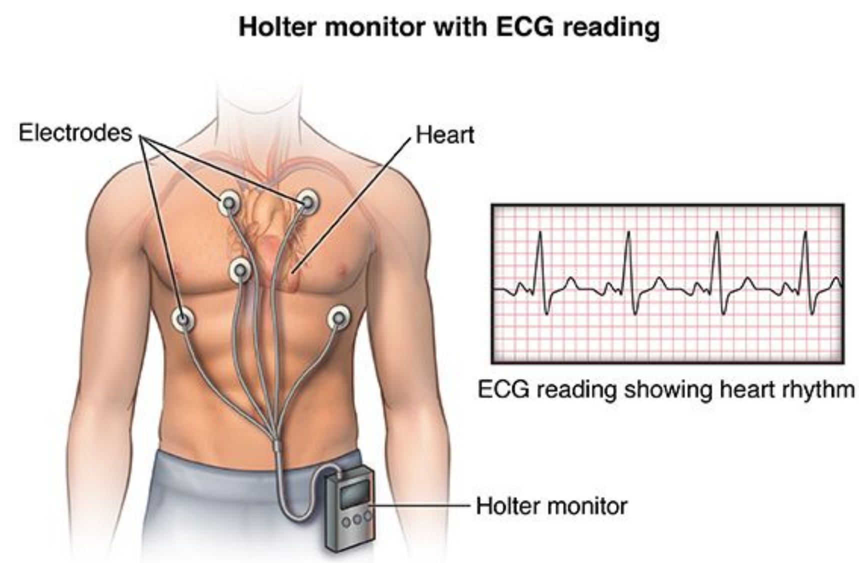

https://mynewmicrophone.com/how-do-microphones-work-a-helpful-illustrated-guide/




https://www.klipsch.ca/blog/digital-vs-analog-audio

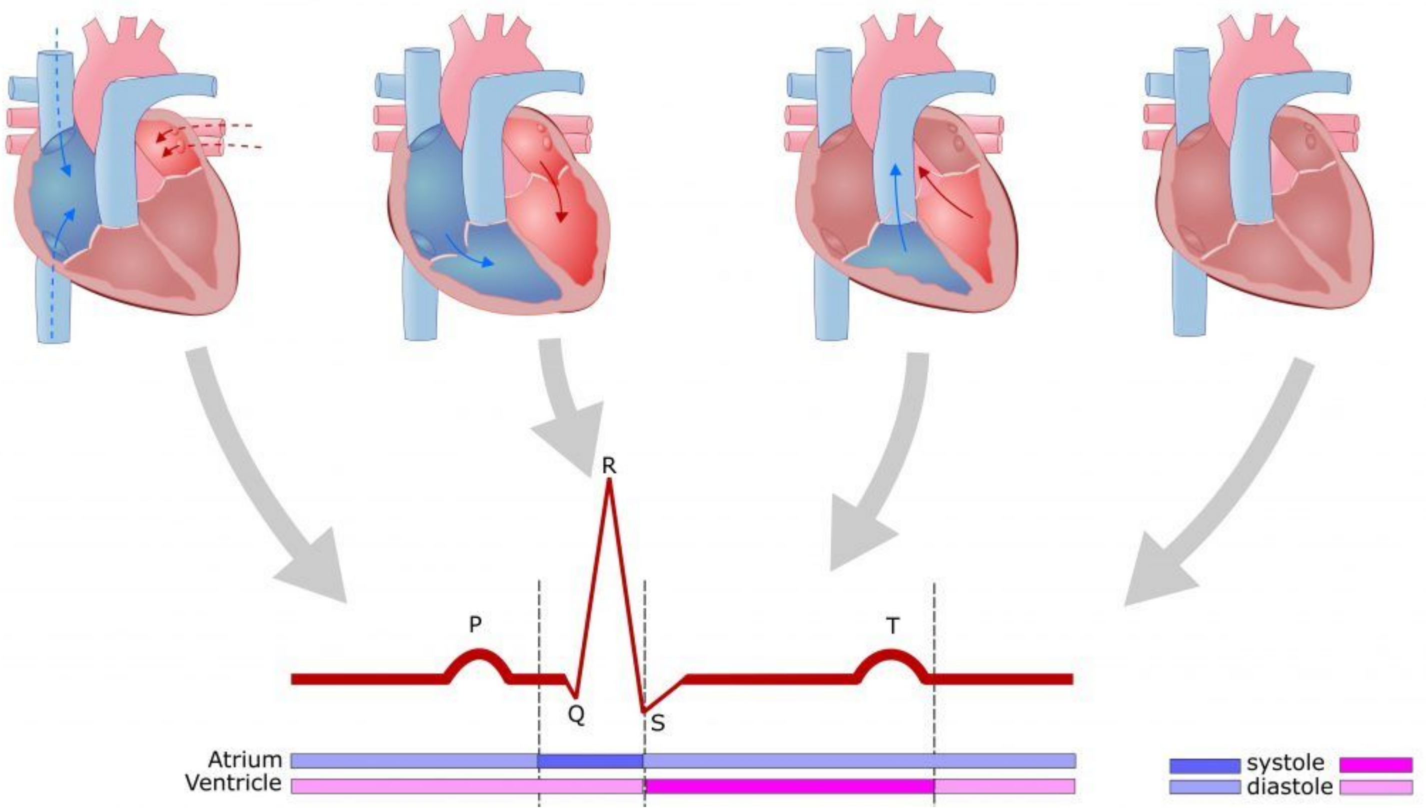
# **E**lectro**c**ardio**g**ram (ECG): cardiac electrical signals

- Recording of heart's electrical activity using multiple electrodes
- Signal from cardiac muscle {de,}polarisation during systole (contraction) and diastole (relaxation).
- Changes in pattern indicate abnormalities (e.g., rhythm disturbances, coronary blood flow, electrolyte disturbances)
- **Heart Rate**: number of cycles within period (bpm)
- **Inter-Beat Interval**: time between cycles (ms)
- **Heart Rate Variability**: variation in IBI




https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/holter-monitor

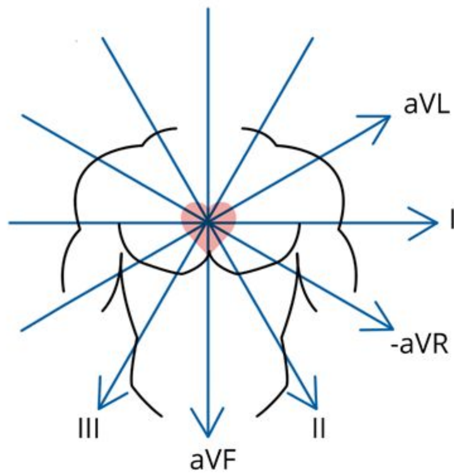
# <u>E</u>lectro<u>c</u>ardio<u>g</u>ram (ECG): inscrutable defined components

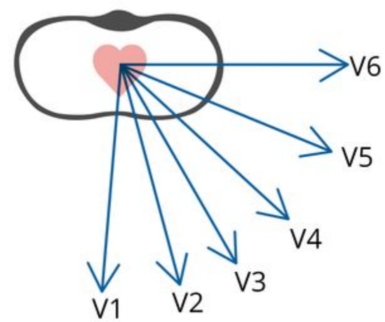


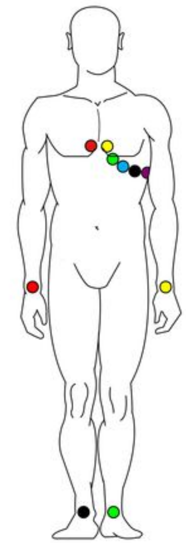

**P-wave:** depolarisation of atria -> atrial systole
**PR-interval**
**QRS complex:** atrial diastole -> depolarisation of ventricles -> ventricular systole
**ST segment**
**T-wave:** repolarization of ventricles -> ventricular diastole
**TP segment**

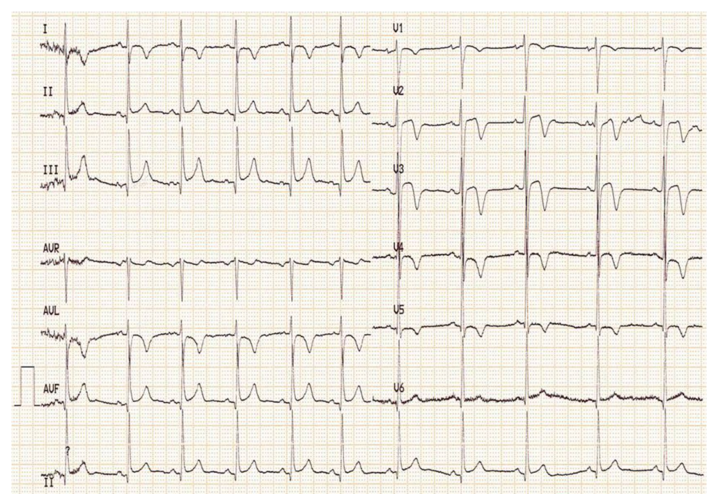
# Multiple sensors recording same signal improves data

- By capturing a signal from multiple sensors we get a lot more information
- 12-lead ECGs given spatial resolution on cardiac abnormalities
- Increases analytical complexity (e.g., handling inter-channel covariance + autocorrelation)



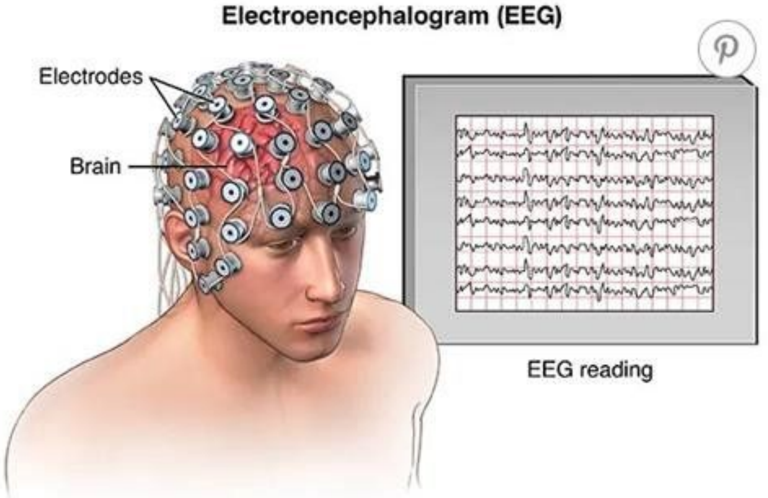



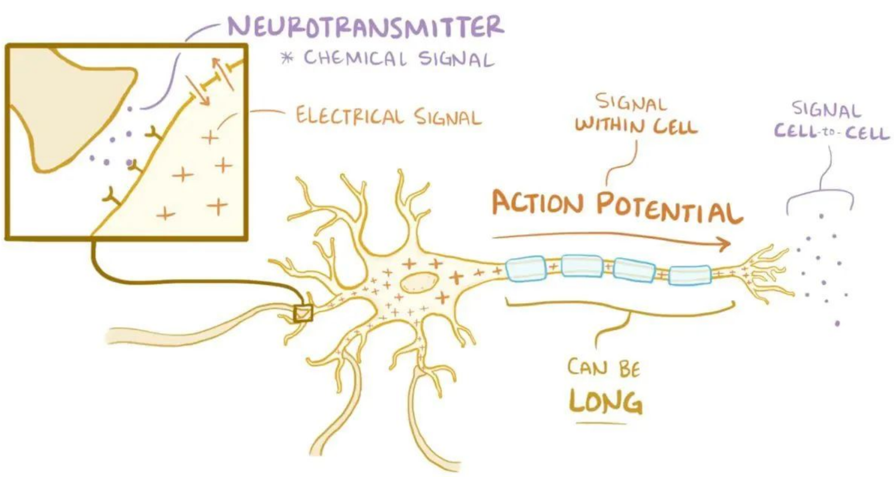


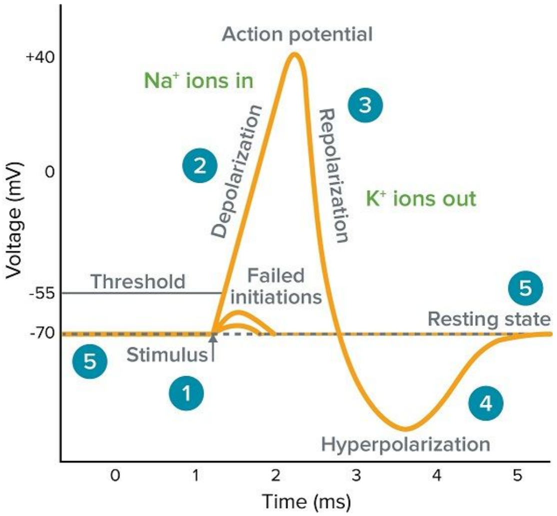

Schreibgeschwindigkeit 25mm/sec

# **E**lectro**e**ncephalo**g**ram (EEG): many channels

- Electrogram of macroscopic brain activity measured from scalp (or intracranially).
- Signal from sets of neuron action potentials (ion-gated membrane de/repolarisation)
- Different electrode layouts/types impact signal resolution

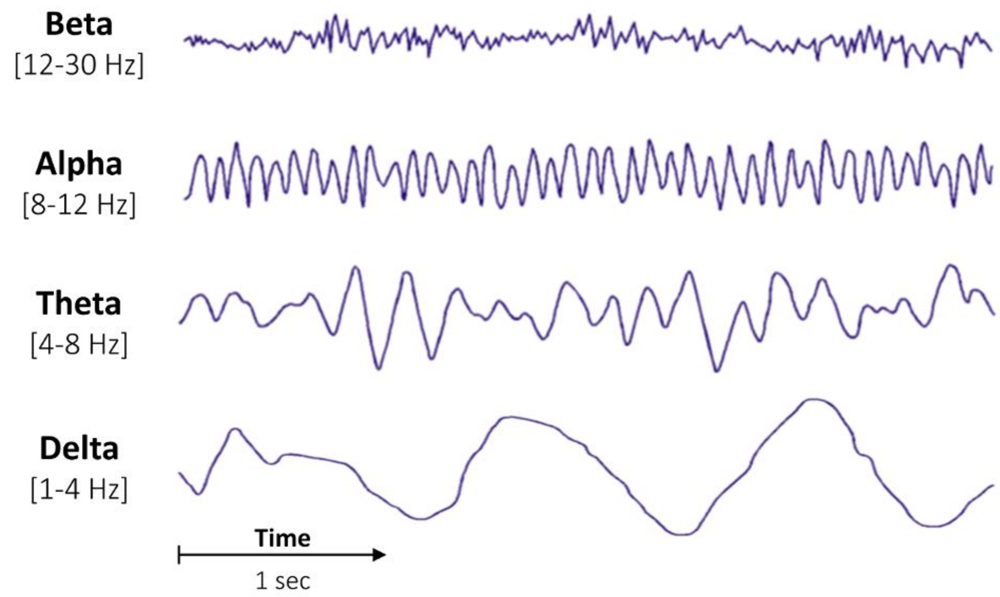


# **E**lectro**e**ncephalo**g**ram (EEG): defined frequency bands

- EEG signals get divided into defined frequency bands
- Different brain activity typified by band of majority of activity (e.g., Delta -> Deep Sleep).
- Patterns in EEG can diagnose neurological disorders (including sleep disorders and epilepsy)



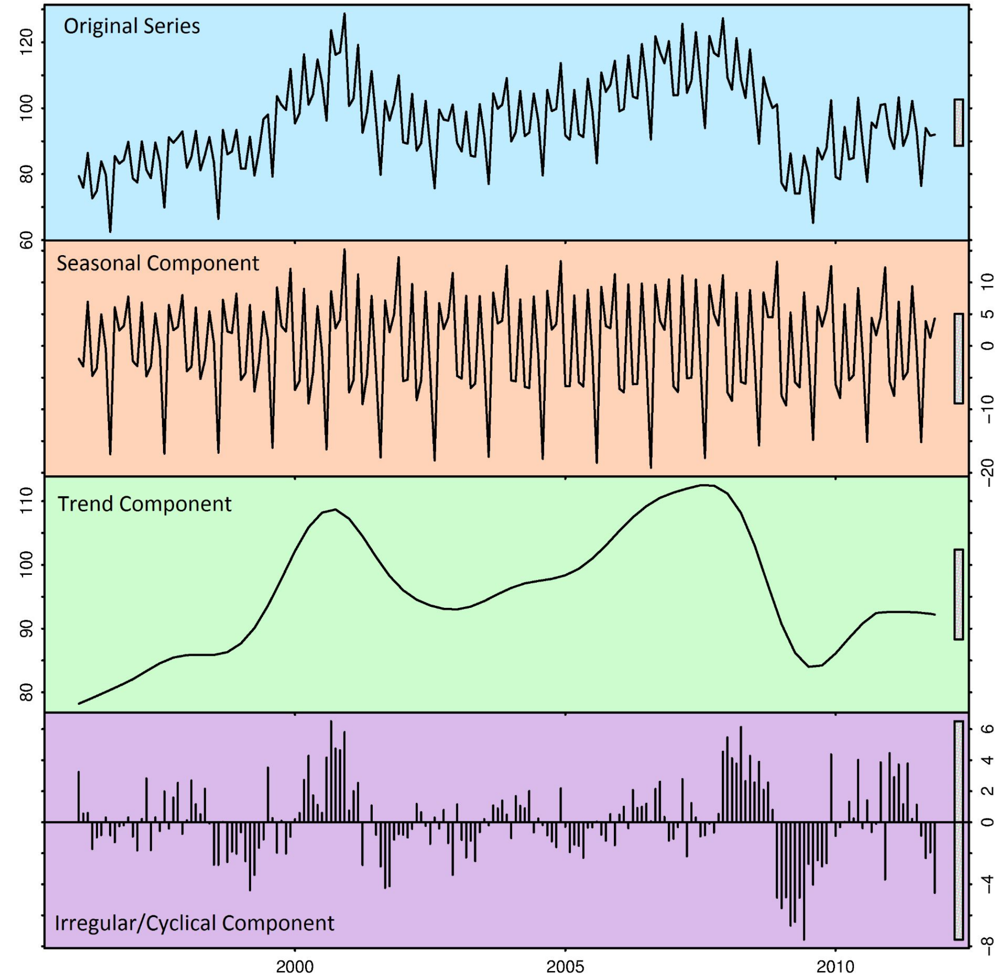
# How do we analyse sensor data?

# Approaches for sensor data

- Moment (time domain) representation
  - Considering the statistical properties of the input data jointly over time
- Spectral (frequency domain) representation
  - Analysing the frequency-space representation
- Path (state space) representation
  - Describe the system as a dynamic system over time
- Change representation: systems of differential equations
  - Not going to discuss these but very common classical statistic/applied maths approach to sensor data.
  - Stochastic (SDEs) or deterministic (ODEs)

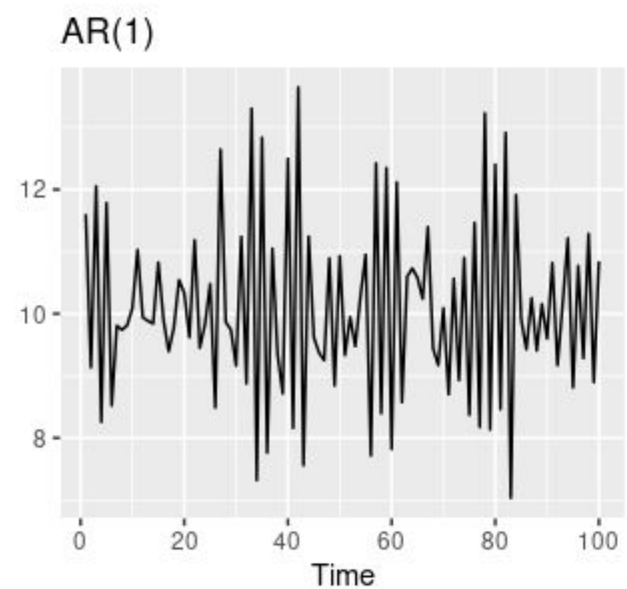
# Time Domain: Decomposition

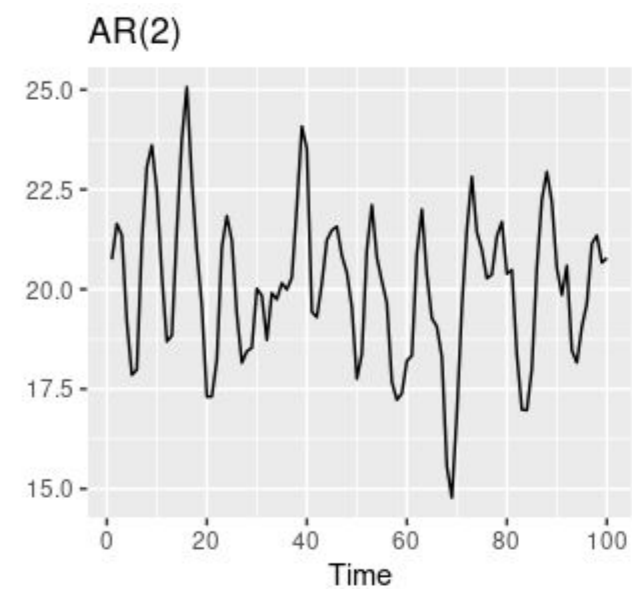
- Decomposition enables measuring strength of trend and seasonality
- Estimate trend/cycle using moving average
- Moving average: smooth series using average over window (size = order)
- Detrend series: signal - moving average
- Moving average of detrended data: seasonality
- Multiplicative decomposition (divide rather than subtract)
- More advanced modern decomposition methods (STL/SEATS/X11)



# Time Domain: Differencing and AutoRegressive models

- Differencing: computed differences between consecutive observations
- AutoRegression: Predict value at time t based on linear combination of past values of variable: $y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + \varepsilon_t$
- Order of model is number of lagged values used
- $\Theta_1 = 0$ represents white noise
- $\Theta_1 = 1$ represents a random walk (with or without constant drift)



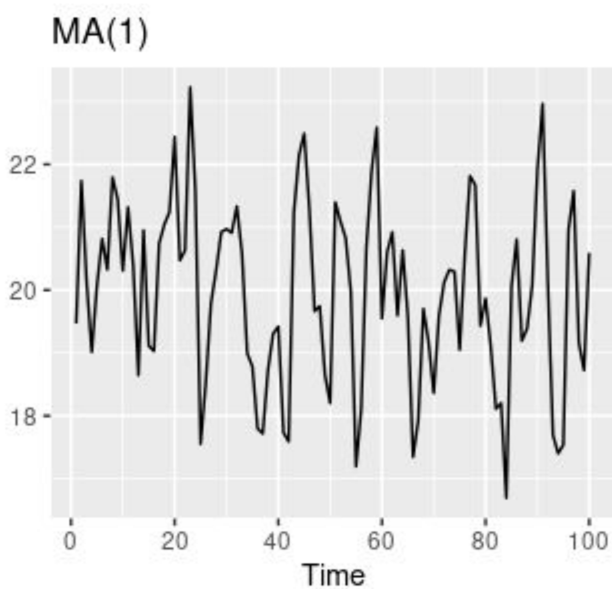



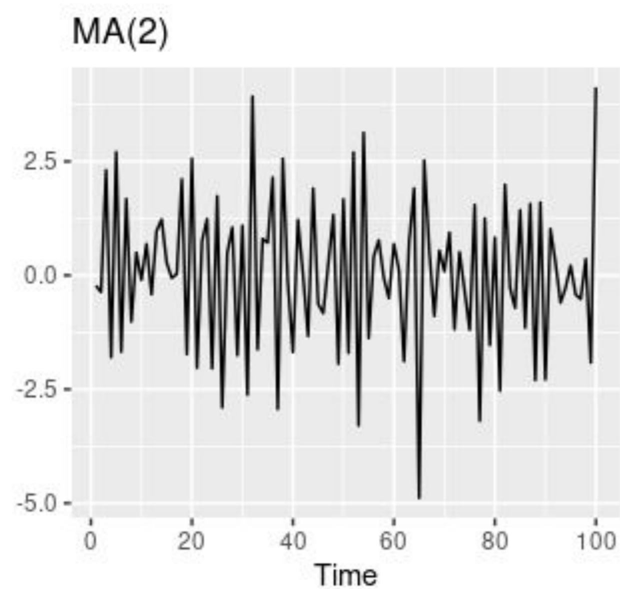
# Time Domain: Moving Average models

- Instead of past values predict using past errors:

  $y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_p\varepsilon_{t-q}$

- For stationary data AR(p) = MA(inf)
- Not to be confused with moving average smoothing used in decomposition



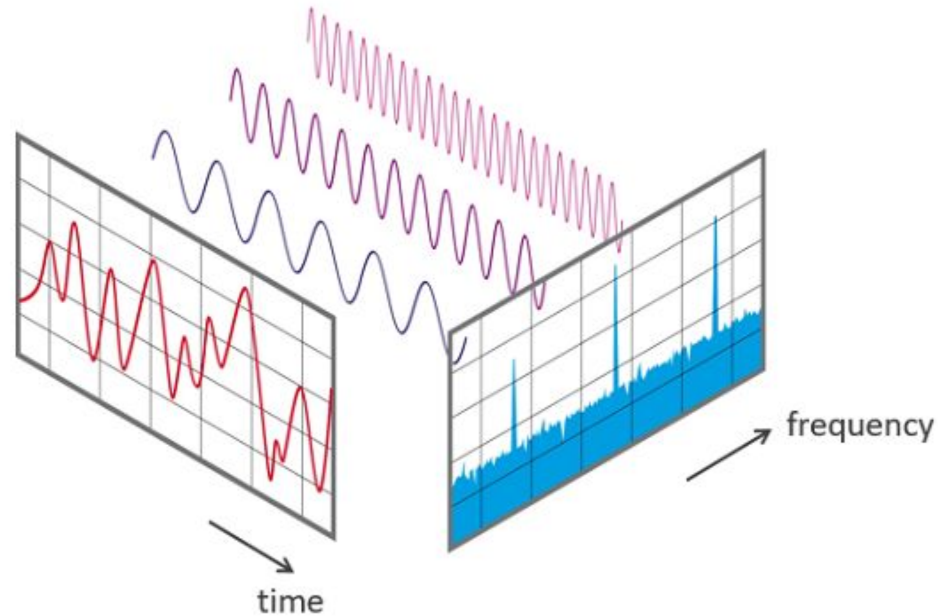



# ARIMA: Combining Differencing, AR and MA models

- AutoRegressive Integrated Moving Average: Predict differenced value of y (y’) using lagged values and errors
- $y'_t=\theta_1 y'_{t-1}+\cdots+\theta_p y'_{t-p}+\theta_1 \varepsilon_{t-1}+\cdots+\theta_q \varepsilon_{t-q}+\varepsilon_t$
- ARIMA(p,d,q): p = order of AR, d = differencing degree, q= order of MA
- Requires MLE / Information Criterion to fit orders
- Core of gold-standard time-series regression/forecasting method
- More advanced methods:
  - Vector Autoregression (VAR): enables feedback between forecasted variable and predictors (more realistic for real-world data)
  - Feed lagged values (or error) into ML model e.g., neural network with or without convolutions

# Frequency/Spectral Domain

- Signal composed multiple frequencies (e.g., EEG power bands)
- Can greatly simplify analysis (offers simple decomposition)
- Feeds into many useful mathematical tools (resonance, harmonics, power spectral densities, eigenvalues, …)
- Several different ways of converting time-domain to frequency-domain
- Laplace and fourier methods are most common

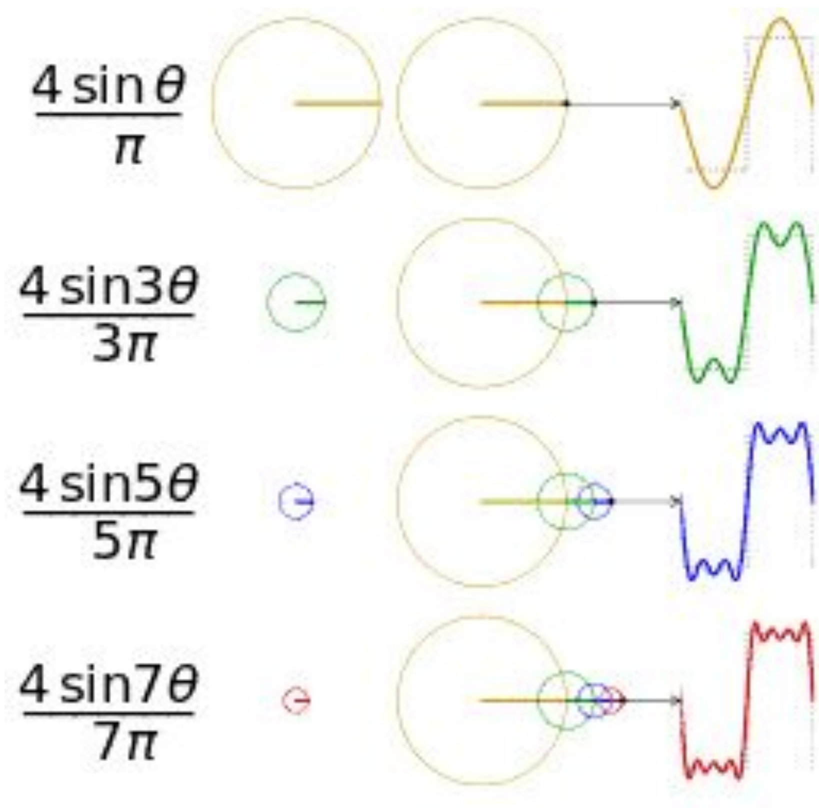


# Frequency Domain: Fourier Transform

- Fourier Transform: Time -> Frequency
- Inverse Fourier Transform: Frequency -> Time
- Decompose signal into series of angular components

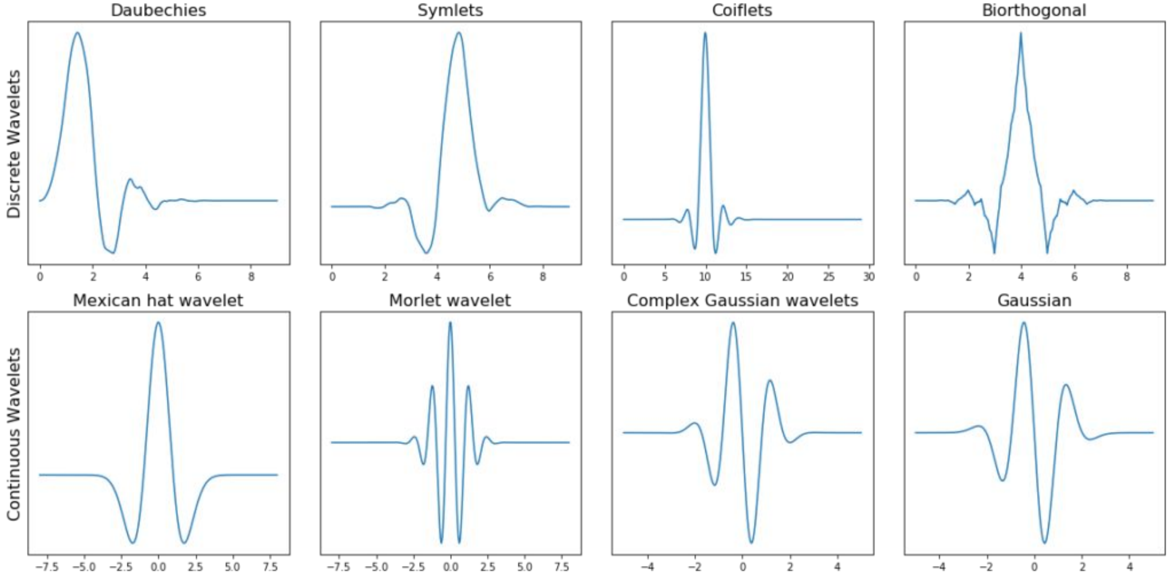
  $x(t) = a \sin(\omega t + \varphi) = a \sin(2\pi f\, t + \varphi)$

- Location (frequency) and height (amplitude) of frequency spectra peaks (among other statistical summaries of spectral space) can be used as input for whatever model you want.



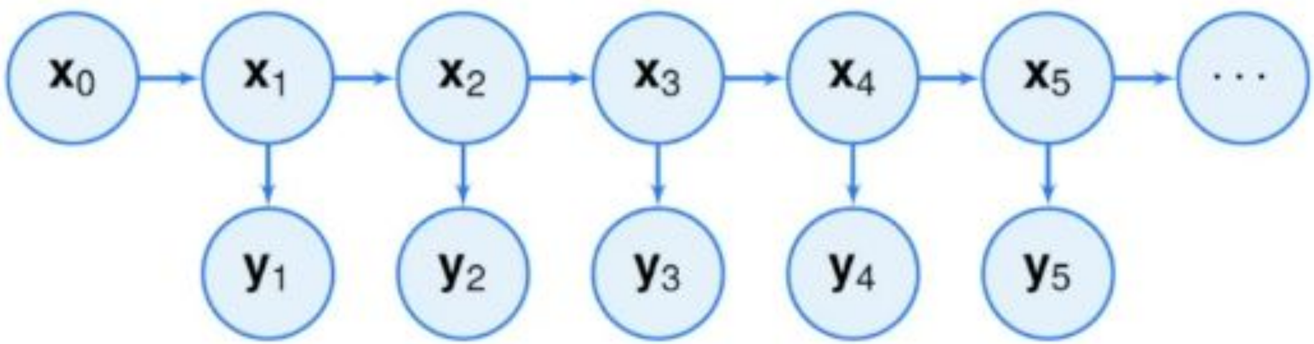
# Time-Frequency Domain: Wavelet Transforms

- Fourier transform has great frequency resolution but no time resolution
- Wavelet allows retaining frequency and time resolution: capture dynamic frequency spectra within signal
- Convolve signal with variety of waves (wavelets) with scale (frequency) and location (time) properties
- Wavelet can be learnt a la convolutional kernels



# State-space models: Hidden Markov Models

- Data is represented resulting from a series of hidden states
- Model describes movements between hidden states
- Observed values are derived from hidden states
- Markov property: only previous state(s) matter
- More naturally discrete time (but continuous time possible)
- Well suited to classification/detection



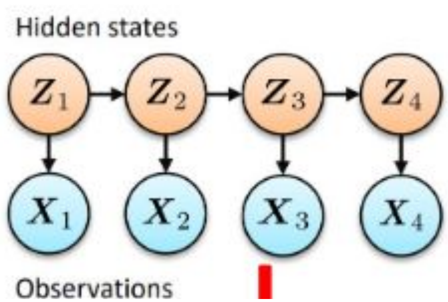

- A canonical state space model:

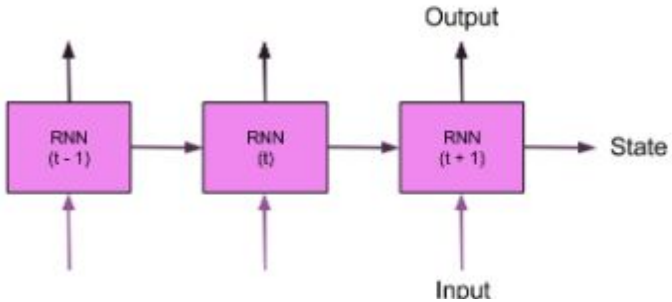
$$\begin{aligned} \text{Dynamics:} \quad & \mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_k), \quad && \mathbf{q}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_k), \\ \text{Measurement:} \quad & \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k), \quad && \mathbf{r}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k) \end{aligned}$$

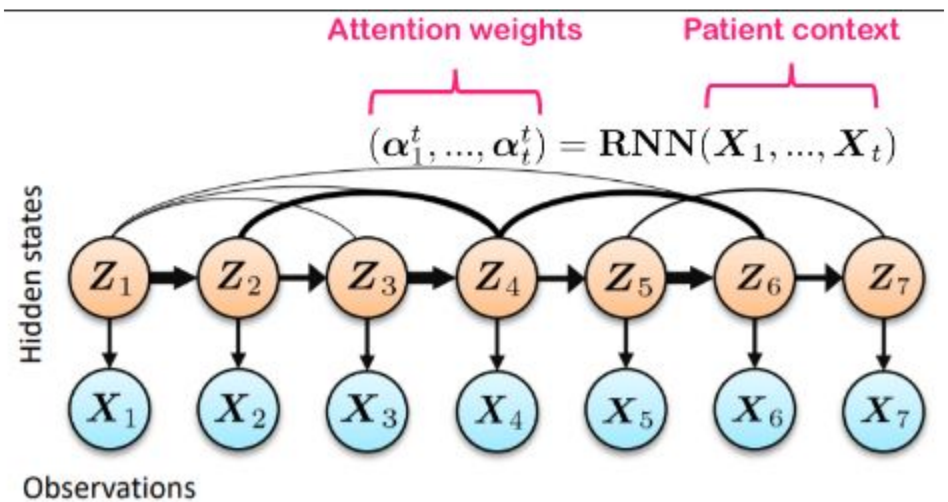
# Going beyond HMMs

- RNNs can act like HMMs more complex dependencies
- Alternative state space models: best of both worlds
- Attention mechanism similar to soft/variable-order HMMs

**Maintain probabilistic structure of HMMs** **But use RNNs to model state dynamics**



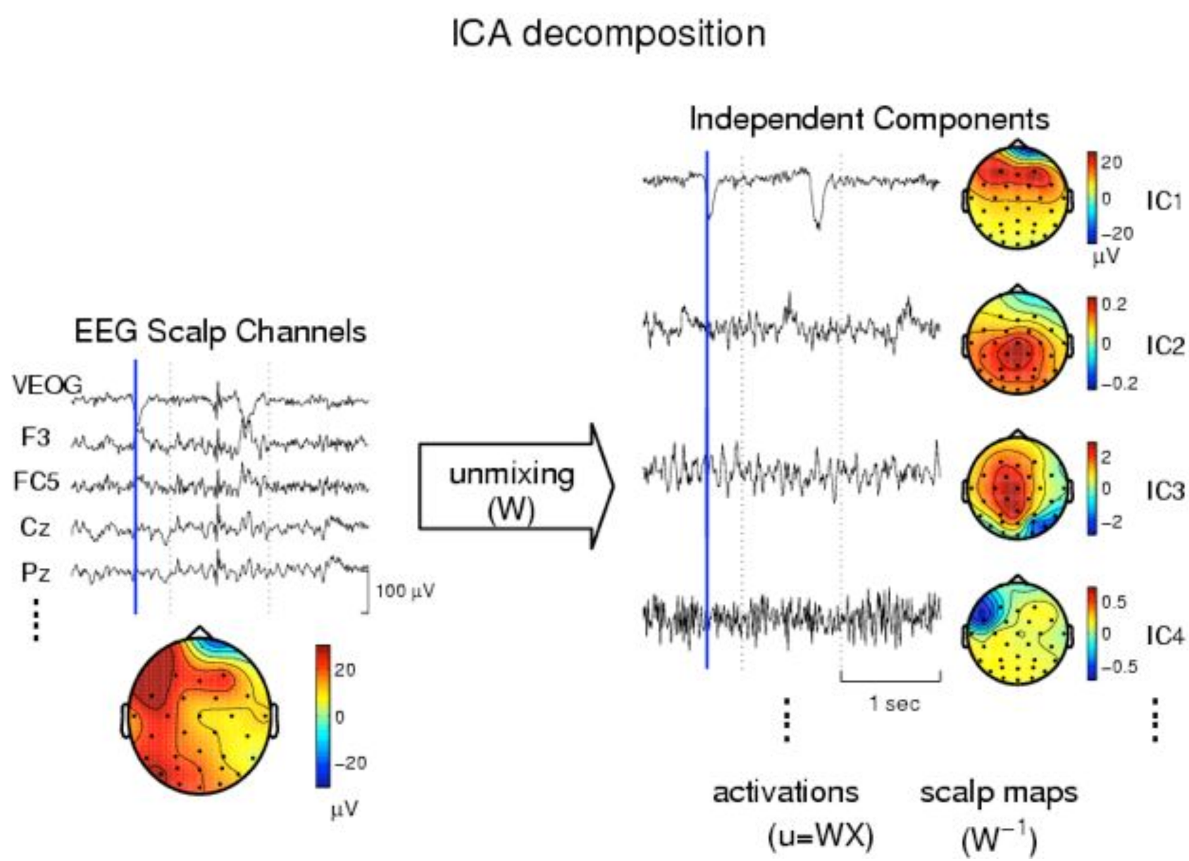




https://www.vanderschaar-lab.com/papers/2022_AAAI_tutorial_web.pdf

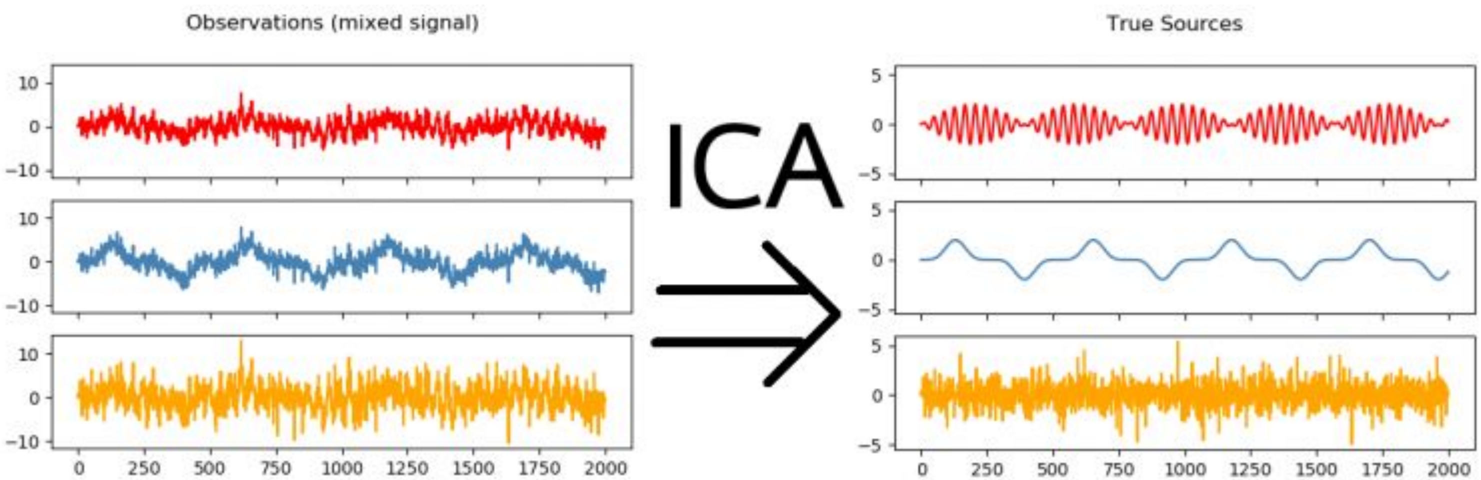
# Decomposing data from multiple sensors

- Measured medical phenomena are often a mixture of signals from different sources
- Multiple sensors = each captures those sources (or a subset)
- Same source through different sensors will have different characteristics (amplitude, lag) due to sensor location



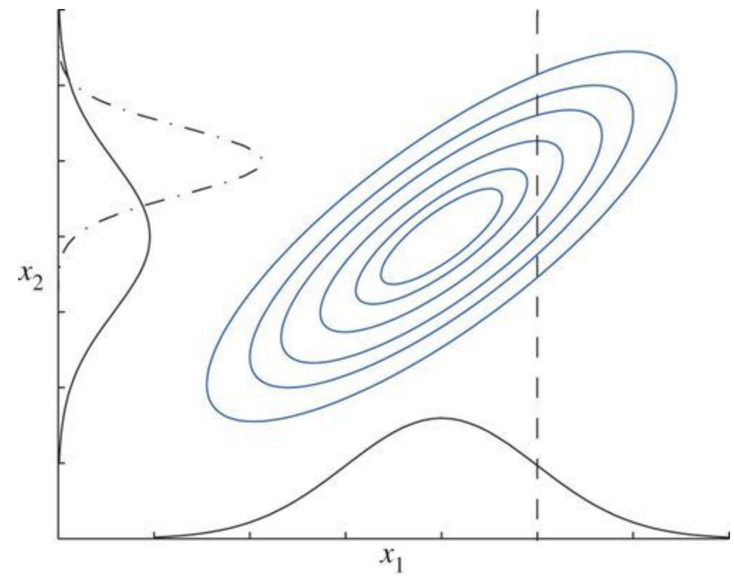
# Decomposing data from multiple sensors: Independent Component Analysis

- Decompose signal into linear mixture of independent sources
- Part of most EEG analysis/processing workflows
- Whiten data (remove correlations between channels: cholesky decomposition with covariation matrix)
- Identify gaussian components of whitened matrix

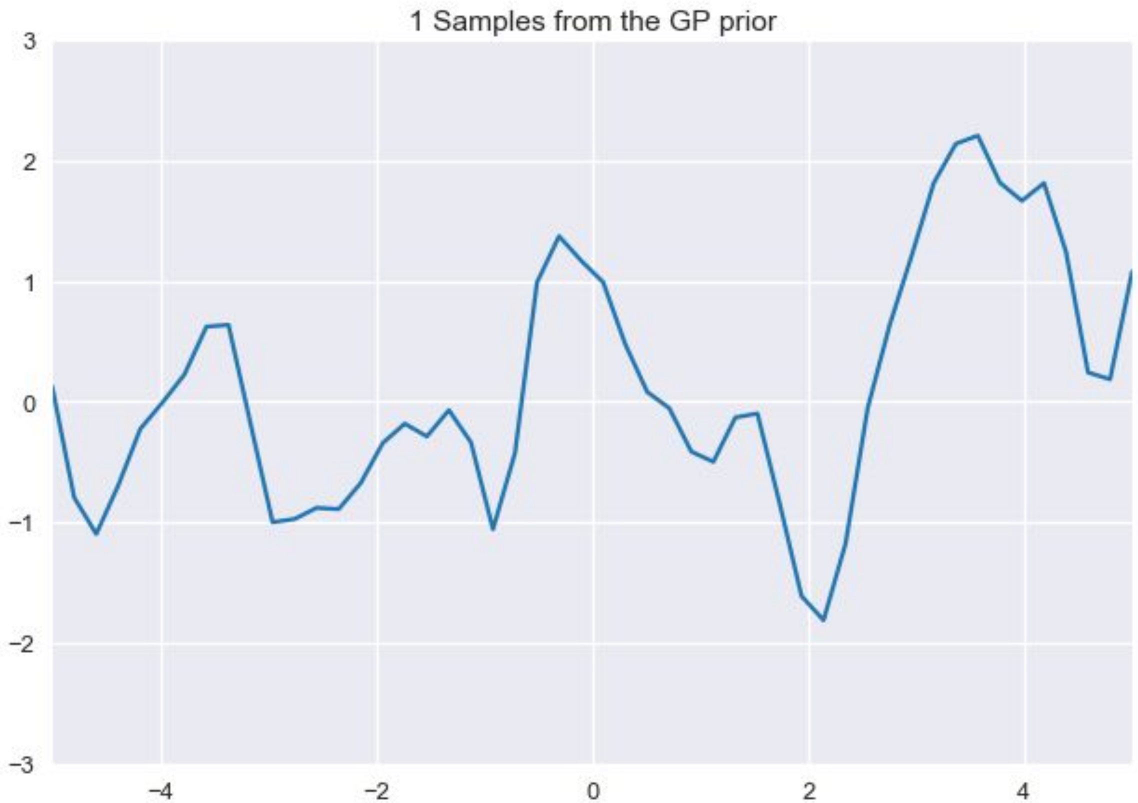


# Gaussian Processes: non-parametric models with infinite parameters

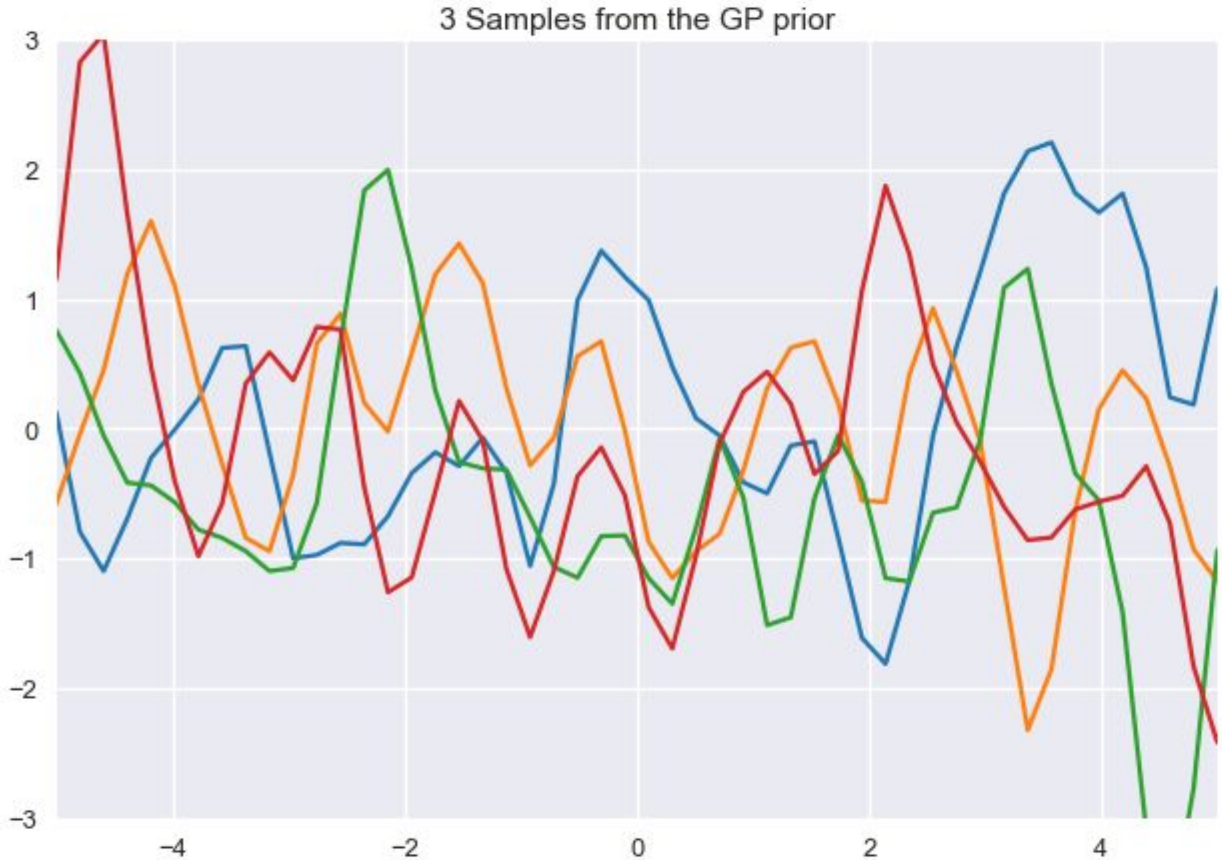
- Bayesian linear regression: find distribution over the parameters consistent with observed data
- Gaussian process: find distribution over all possible functions that are consistent with observed data
- Defined by covariance kernel between functions (draws from multivariate gaussian)
- Can capture time, frequency, and state-space models
- Yet another “should be an entire graduate course”



# Gaussian Process Prior

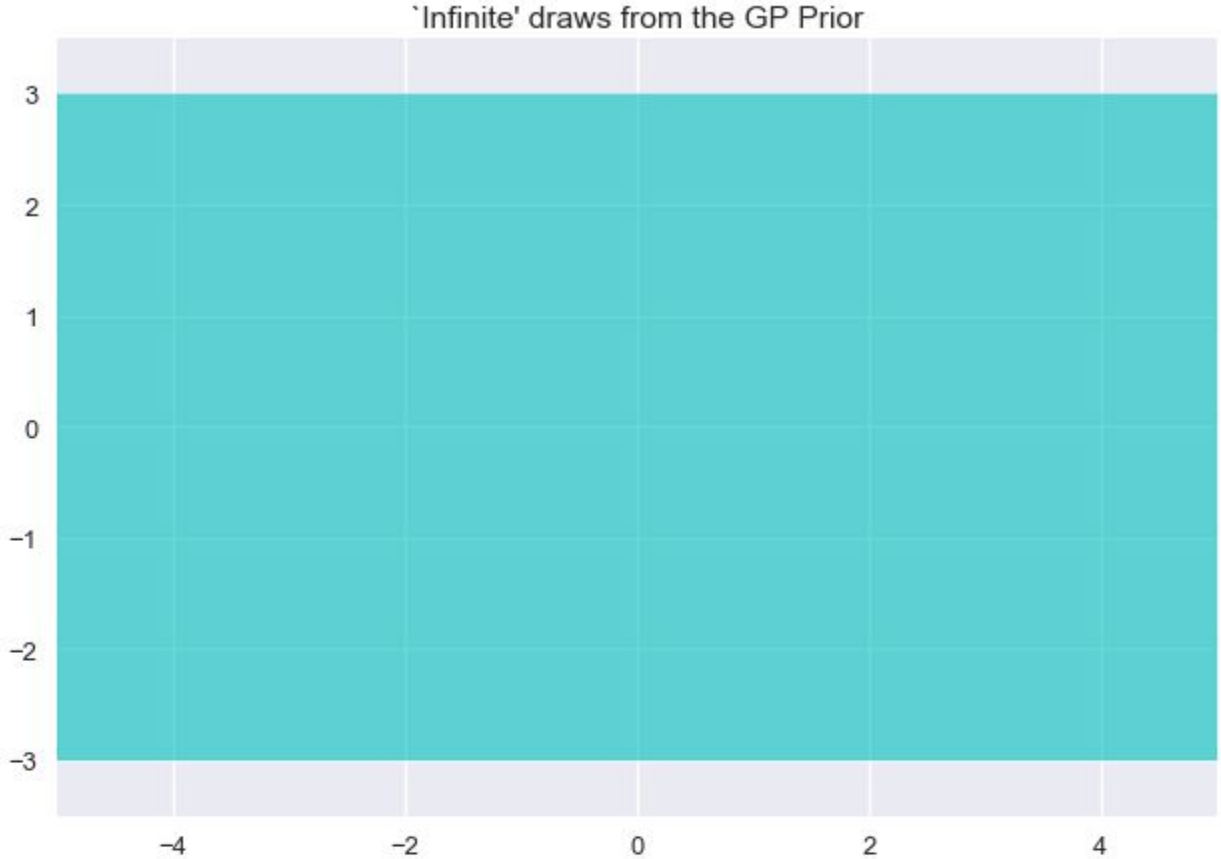


# Gaussian Process Prior

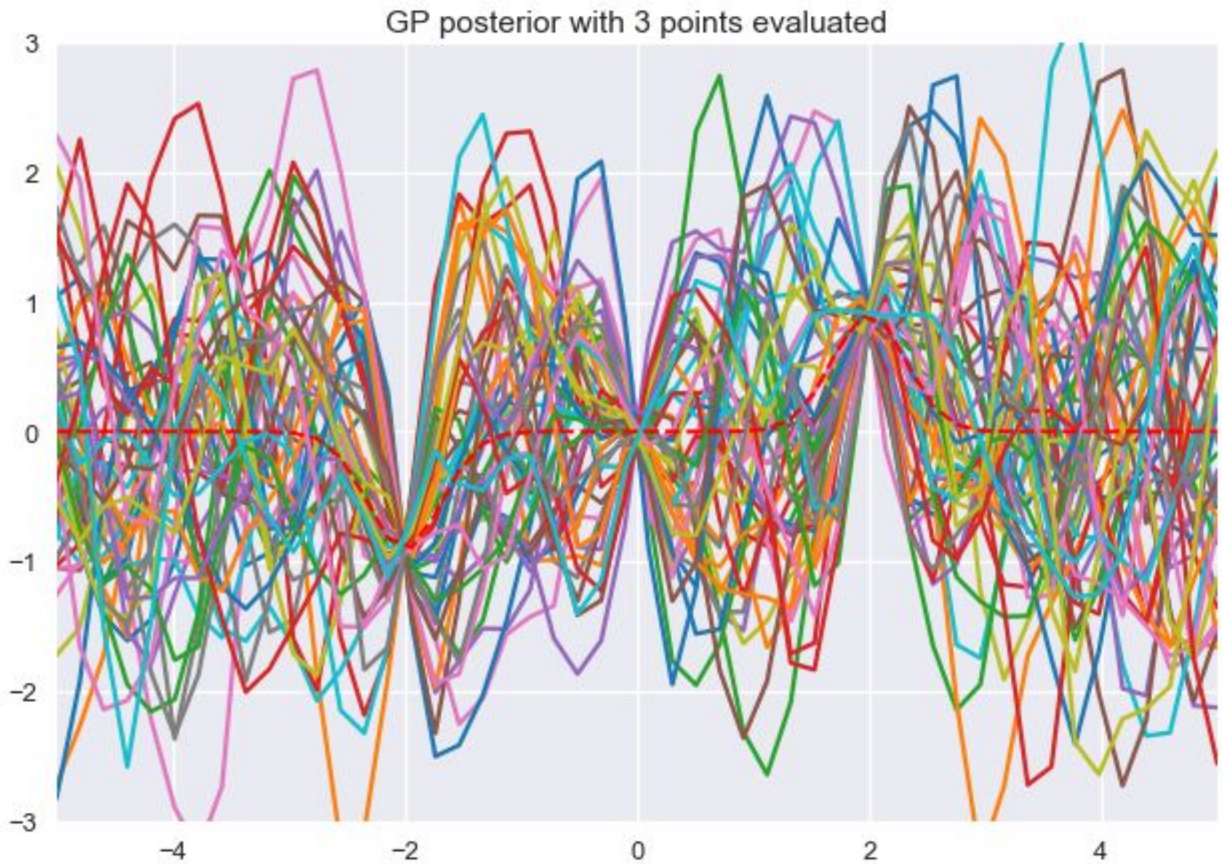


# Gaussian Process Prior

# Gaussian Process Prior

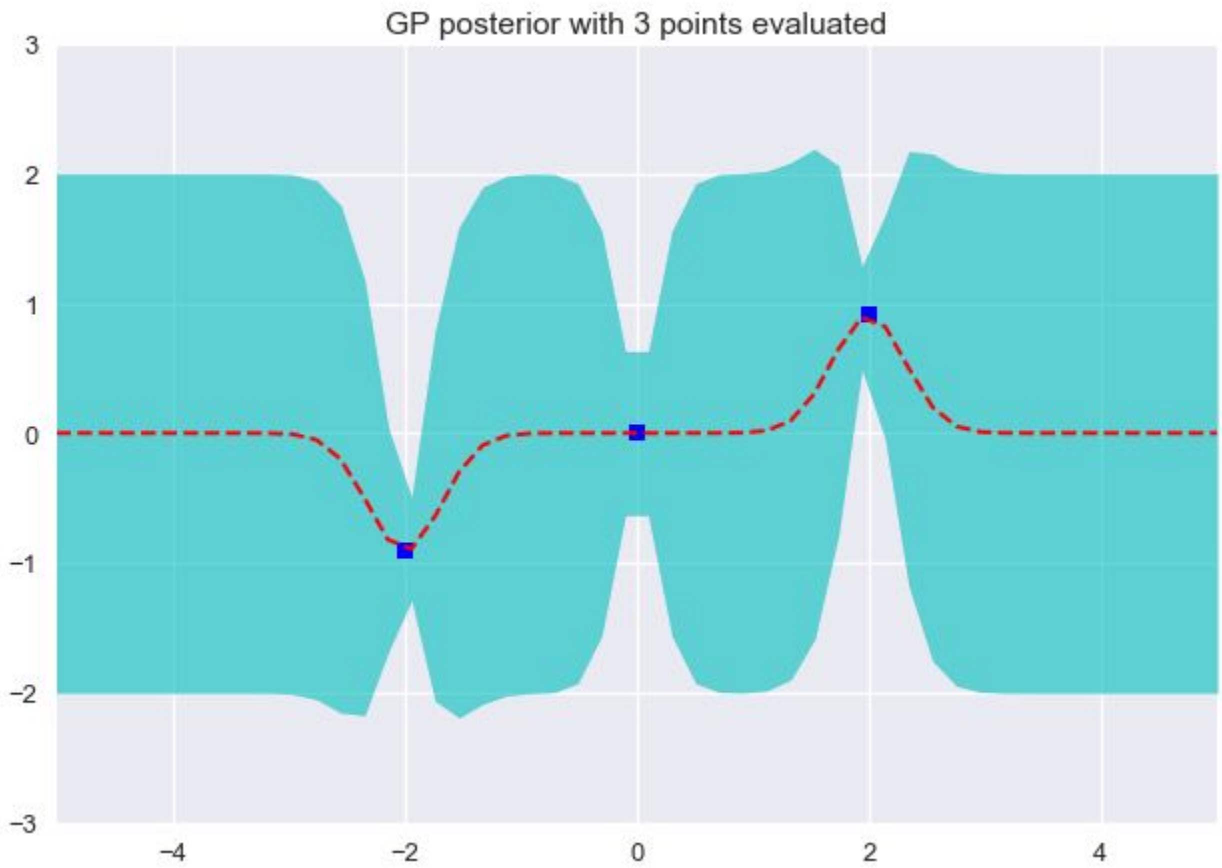


# Constrain prior based on observed data: posterior

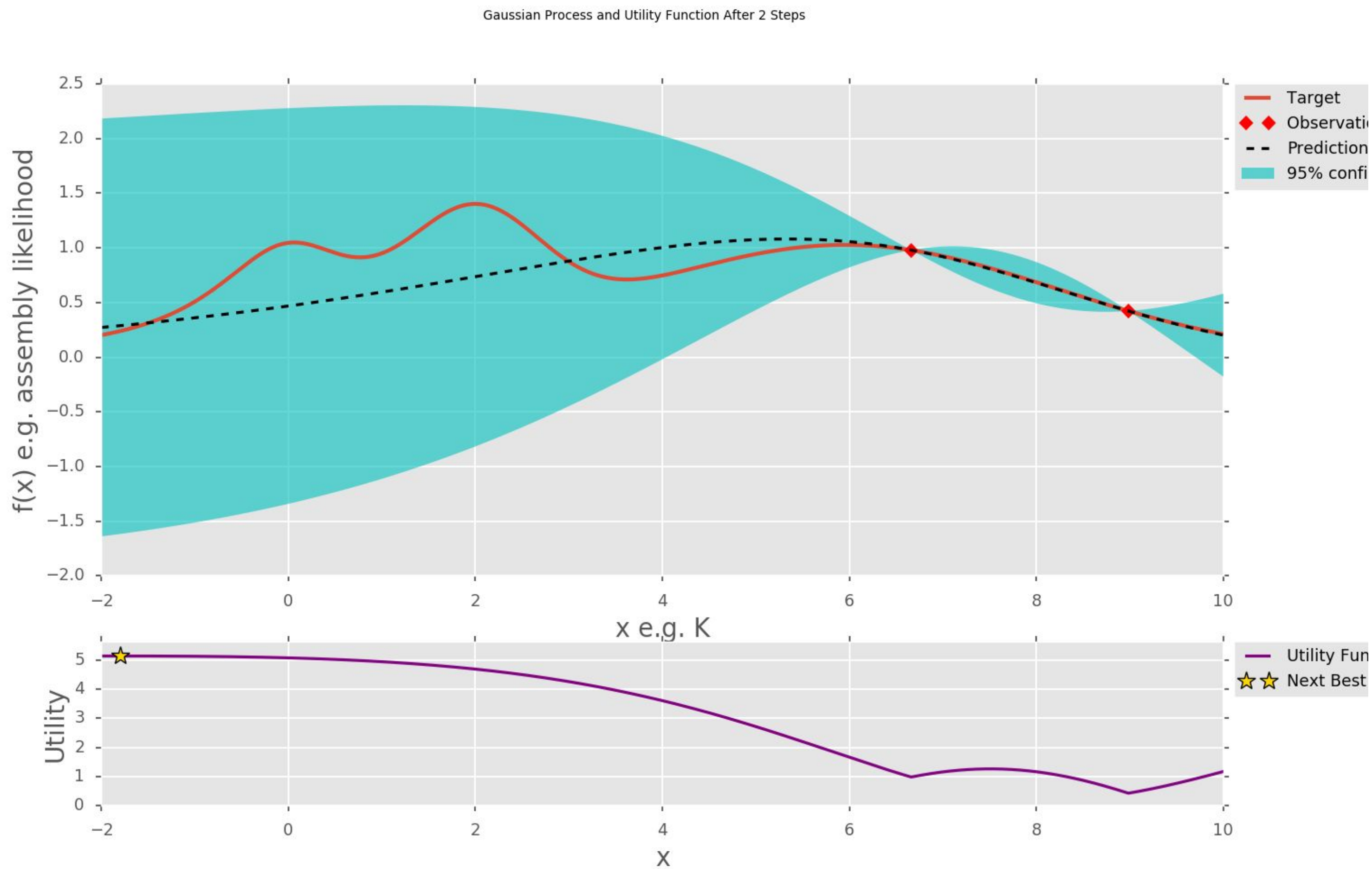


# Constrain prior based on observed data: posterior

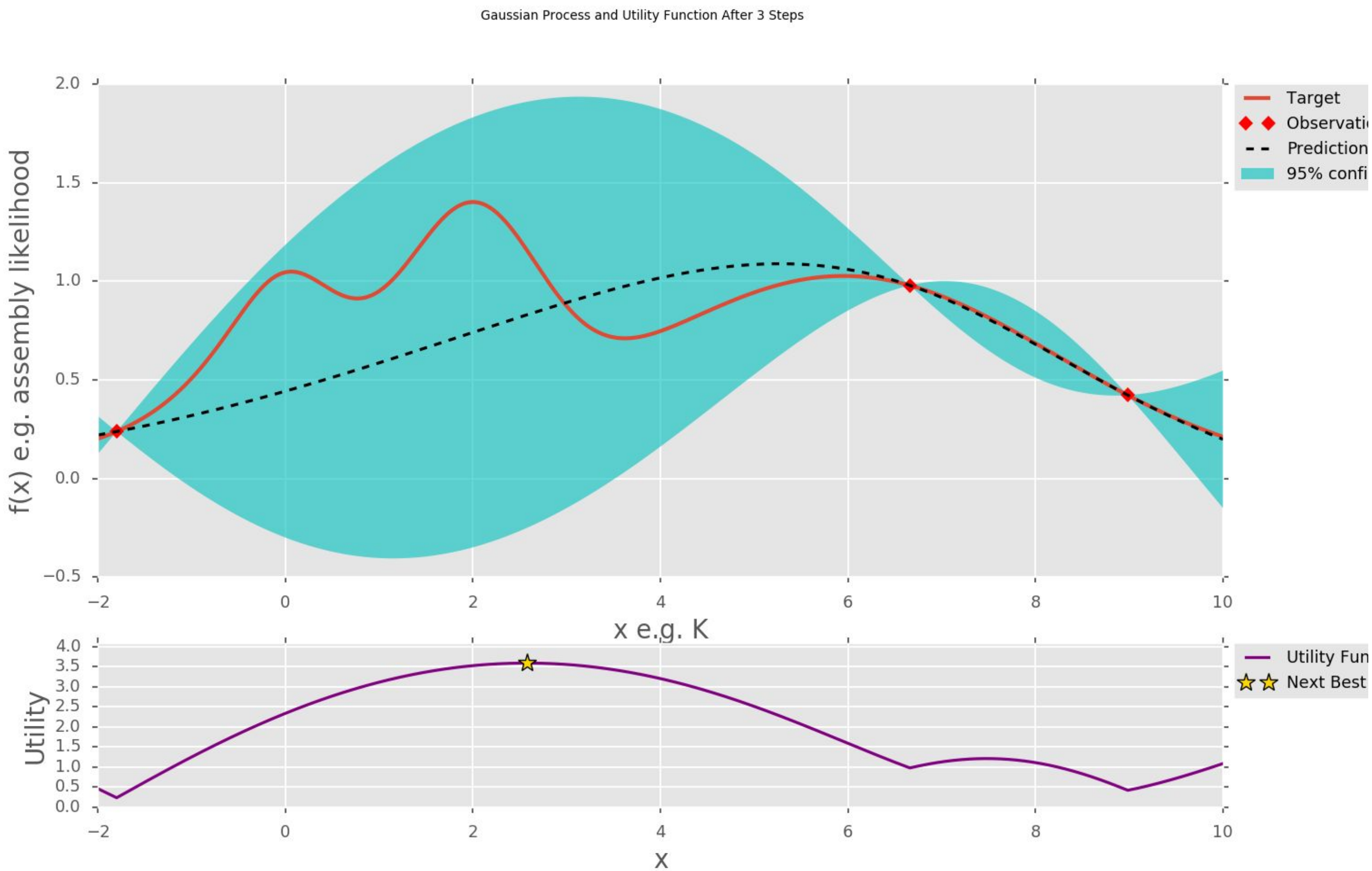
# Constrain prior based on observed data: posterior



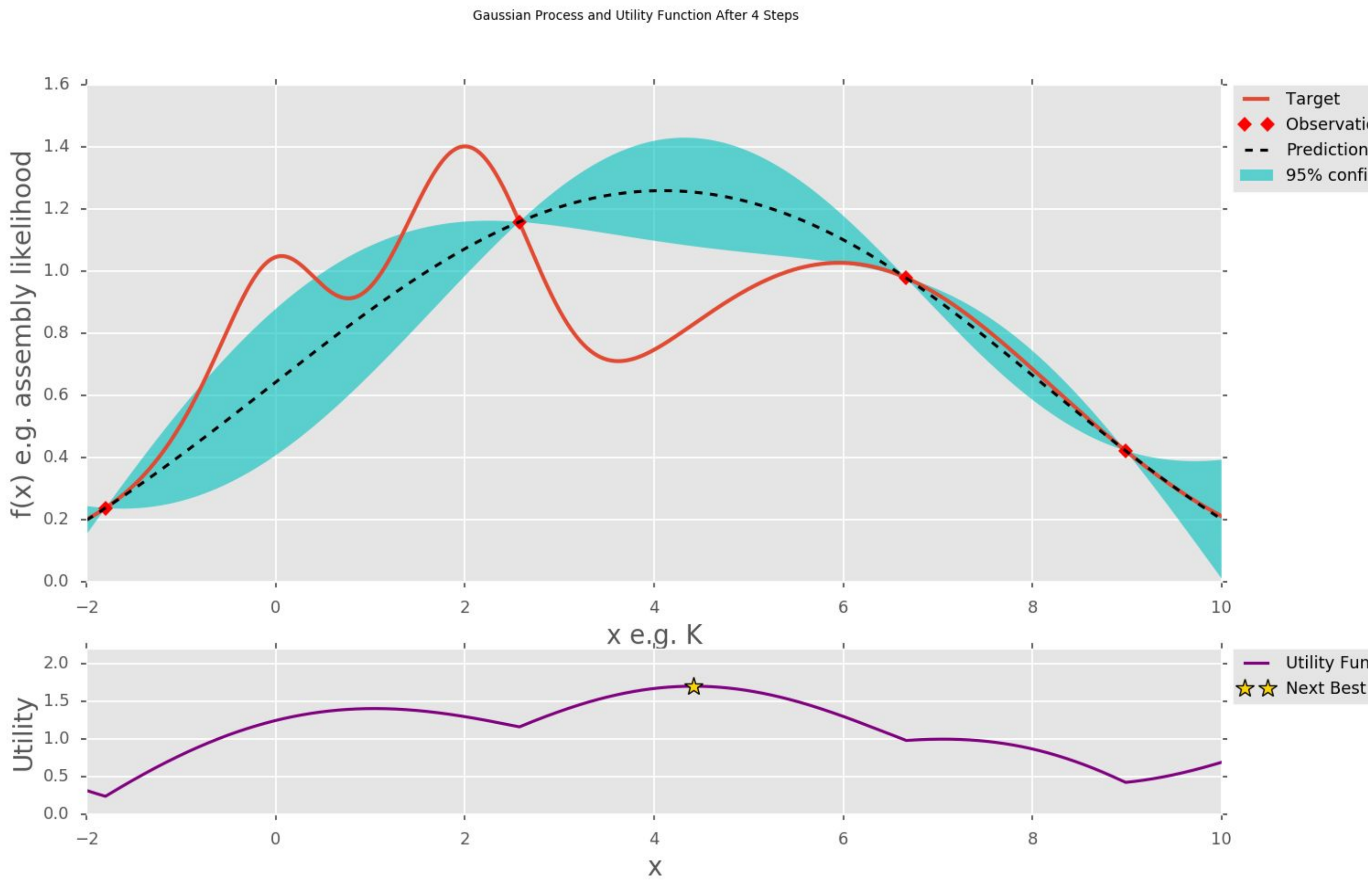
# Probabilistic numerics: data efficient framework



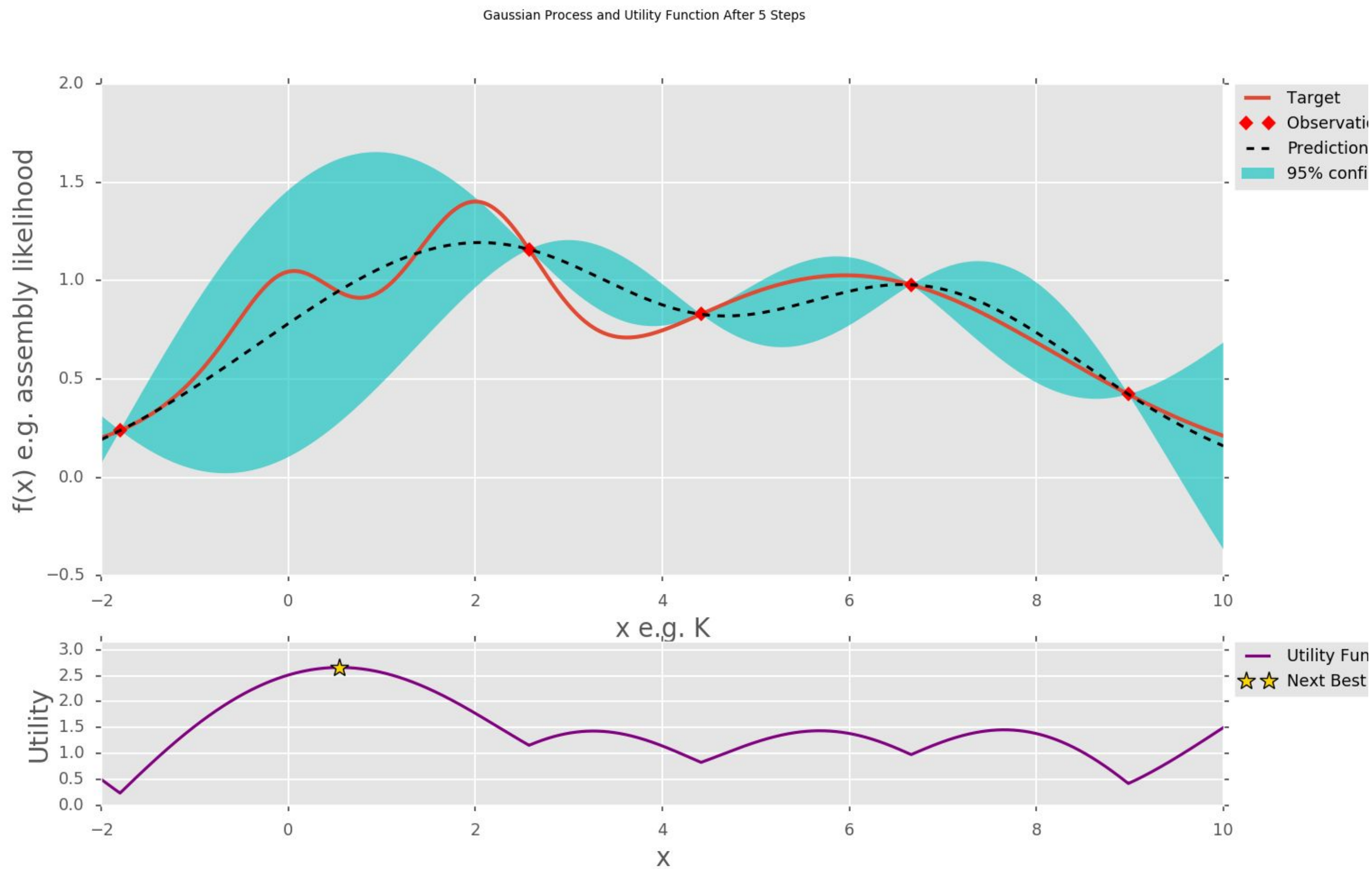
# Probabilistic numerics: data efficient framework



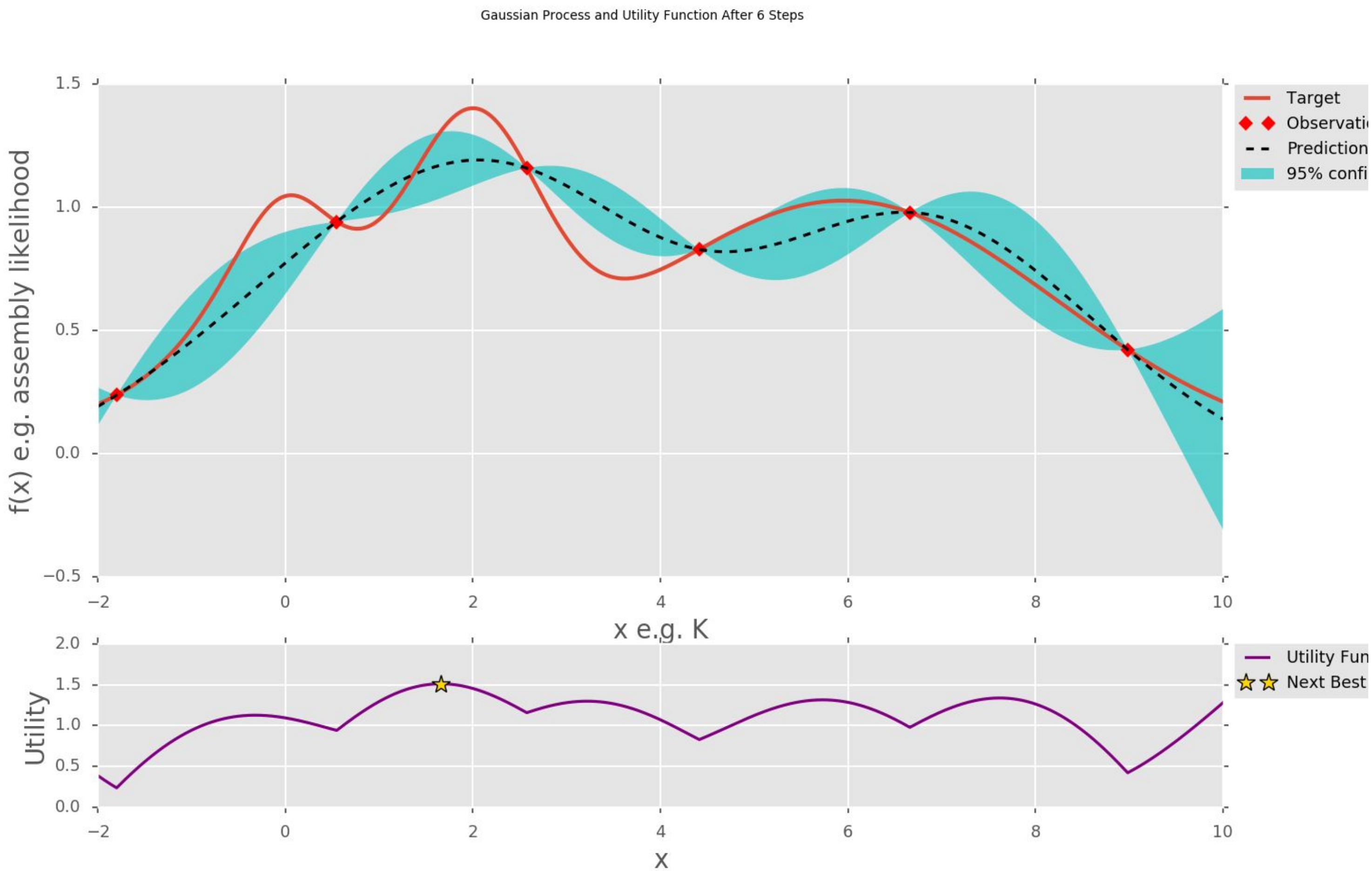
# Probabilistic numerics: data efficient framework



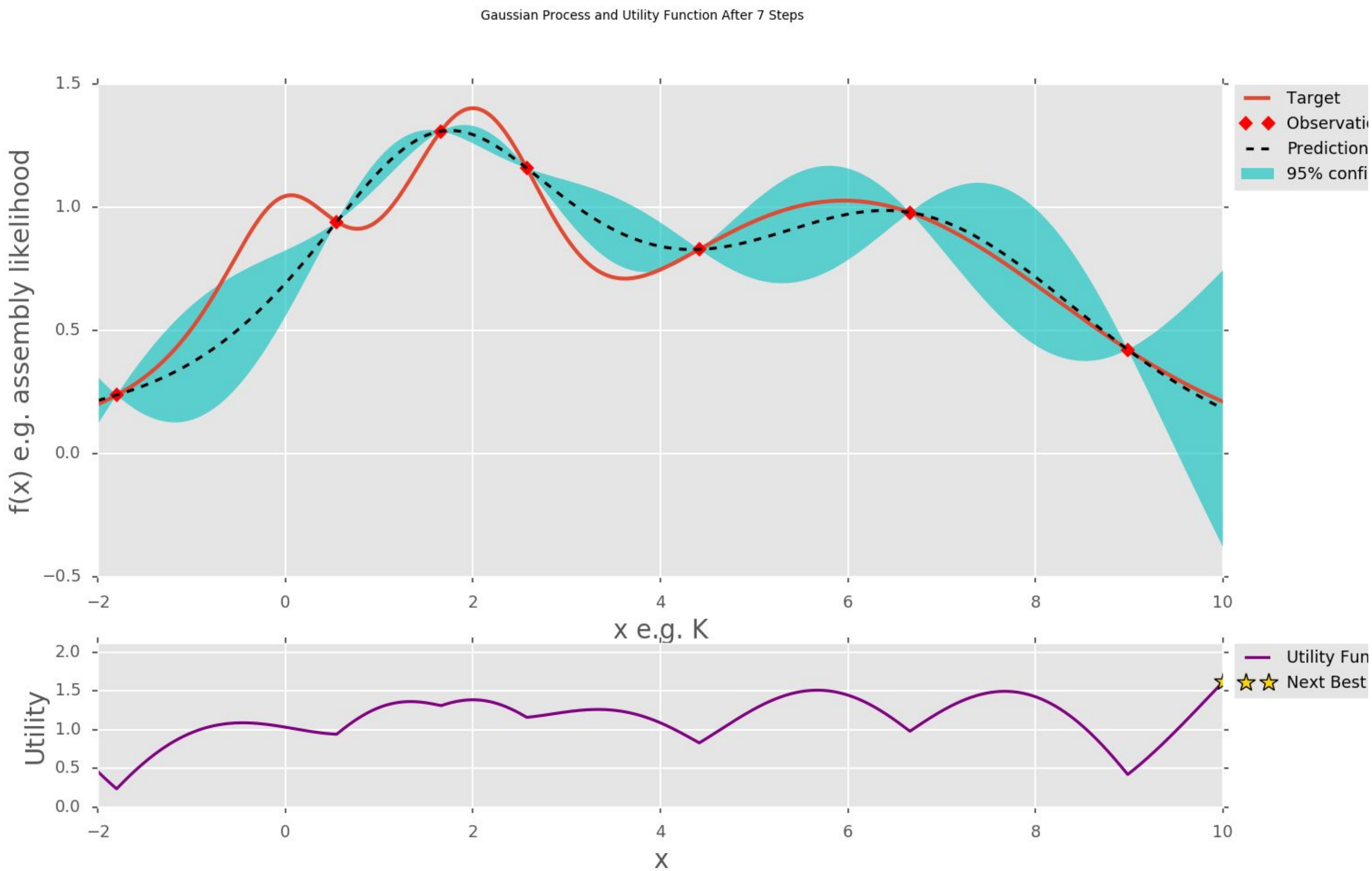
# Probabilistic numerics: data efficient framework



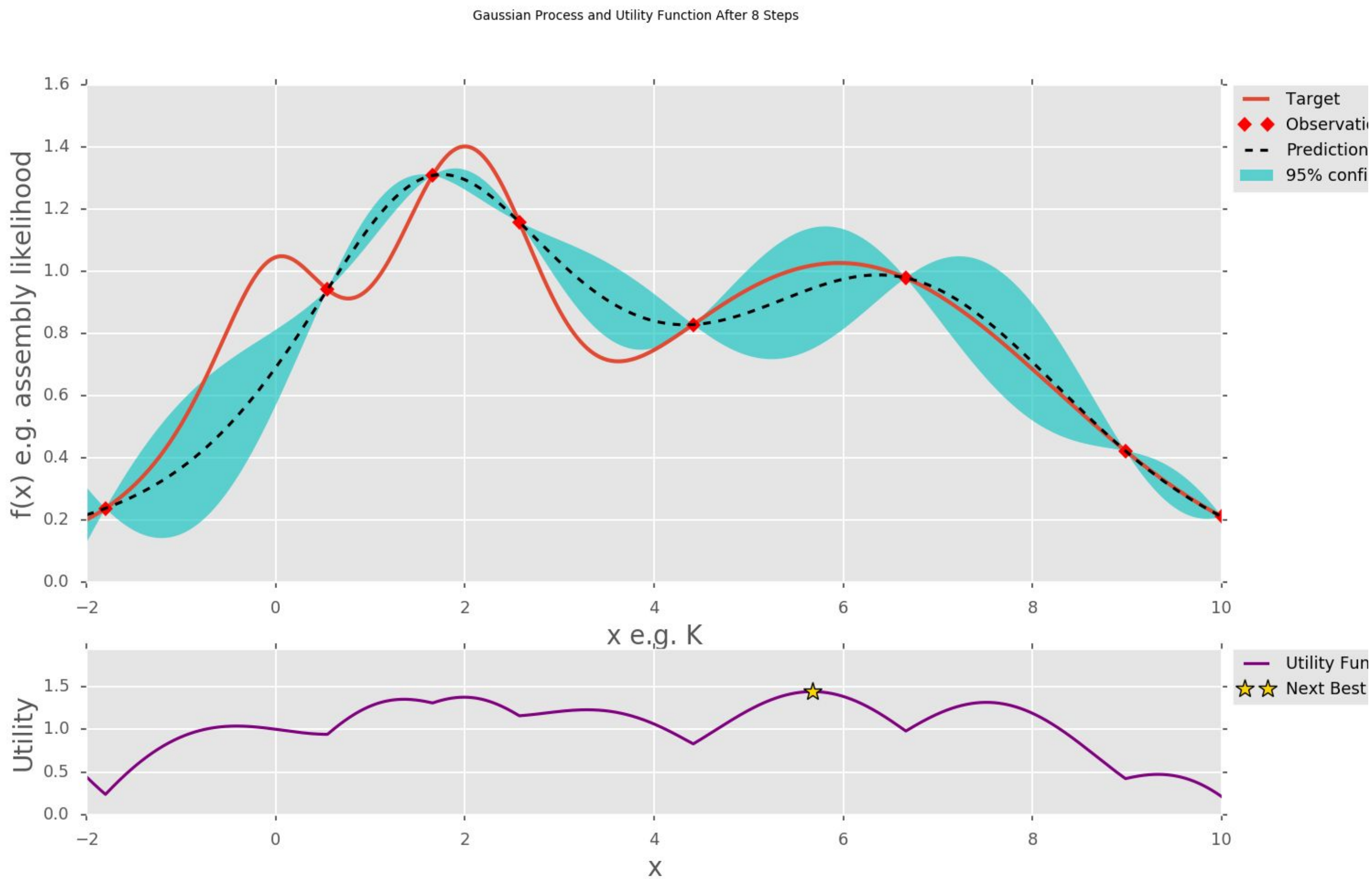
# Probabilistic numerics: data efficient framework



# Probabilistic numerics: data efficient framework



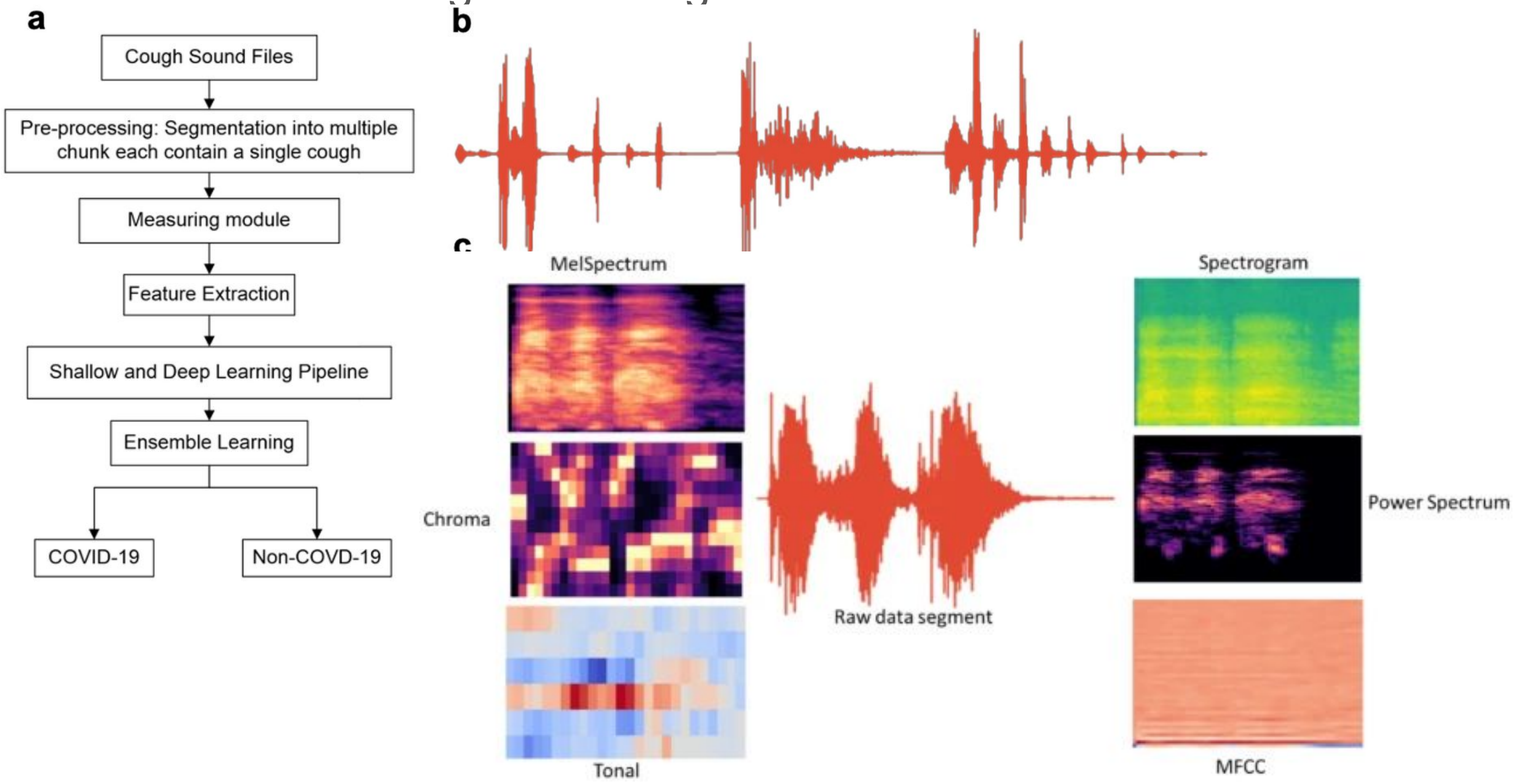
# Probabilistic numerics: data efficient framework



# Let’s discuss some case studies

# Cough detection/classification

- Quick/Non-invasive screening of COVID-19 could protect health resources/optimise non-pharmaceutical interventions
- Crowdsourced cough recordings

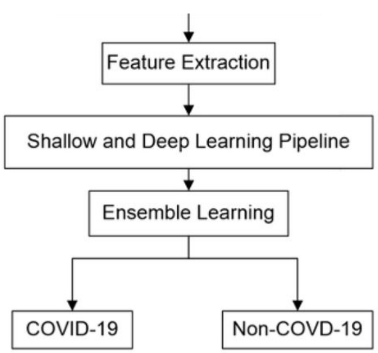


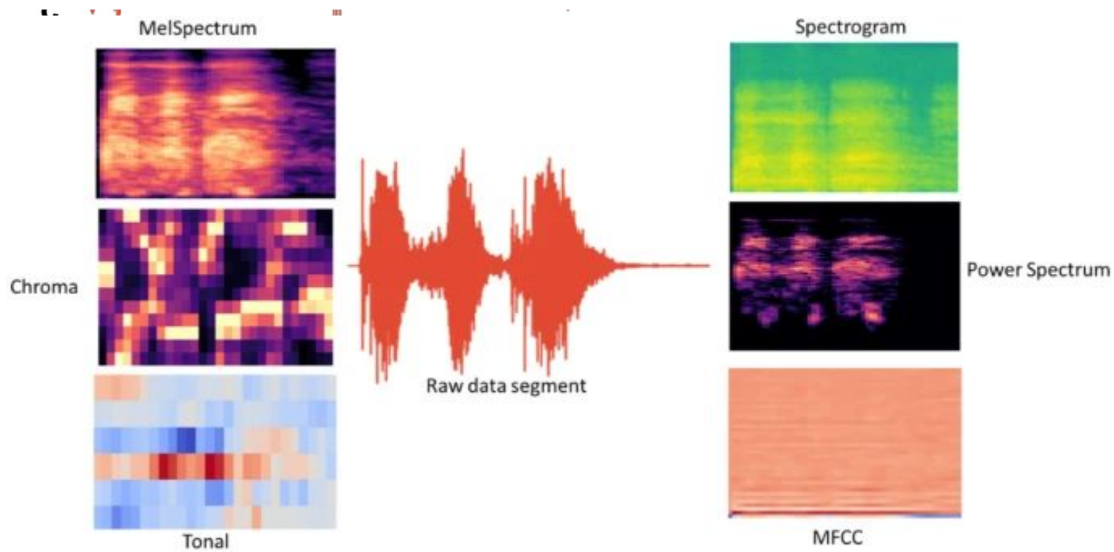

10.1038/s41598-021-95042-2

# Cough detection/classification

| Feature/classifier | NB | | KNN | | LogitReg | | RF | | SGD | | XGB | | SVM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pre | NPV | Pre | NPV | Pre | NPV | Pre | NPV | Pre | NPV | Pre | NPV | Pre | NPV |
| Chroma | 0.52 | 0.50 | 0.53 | 0.53 | 0.54 | 0.54 | 0.53 | 0.53 | 0.55 | 0.54 | 0.51 | 0.51 | 0.54 | 0.55 |
| MelSpectrum | 0.68 | 0.55 | 0.63 | 0.63 | 0.64 | 0.64 | 0.63 | 0.61 | 0.58 | 0.59 | 0.61 | 0.62 | 0.64 | 0.63 |
| MFCC | 0.55 | 0.64 | 0.60 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.59 | 0.61 | 0.63 | 0.68 | 0.64 |
| PowerSpec | 0.54 | 0.58 | 0.60 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.57 | 0.61 | 0.61 | 0.63 | 0.63 |
| RAW | 0.59 | 0.53 | 0.60 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0.52 | 0.56 | 0.58 | 0.61 | 0.59 |
| Spec | 0.56 | 0.57 | 0.65 | 0.66 | 0.63 | 0.66 | 0.68 | 0.68 | 0.60 | 0.62 | 0.65 | 0.65 | 0.73 | 0.68 |
| Tonal | 0.53 | 0.63 | 0.53 | 0.55 | 0.55 | 0.55 | 0.56 | 0.56 | 0.51 | 0.51 | 0.57 | 0.53 | 0.53 | 0.54 |

CNN/VGGish transfer: Precision ~0.65



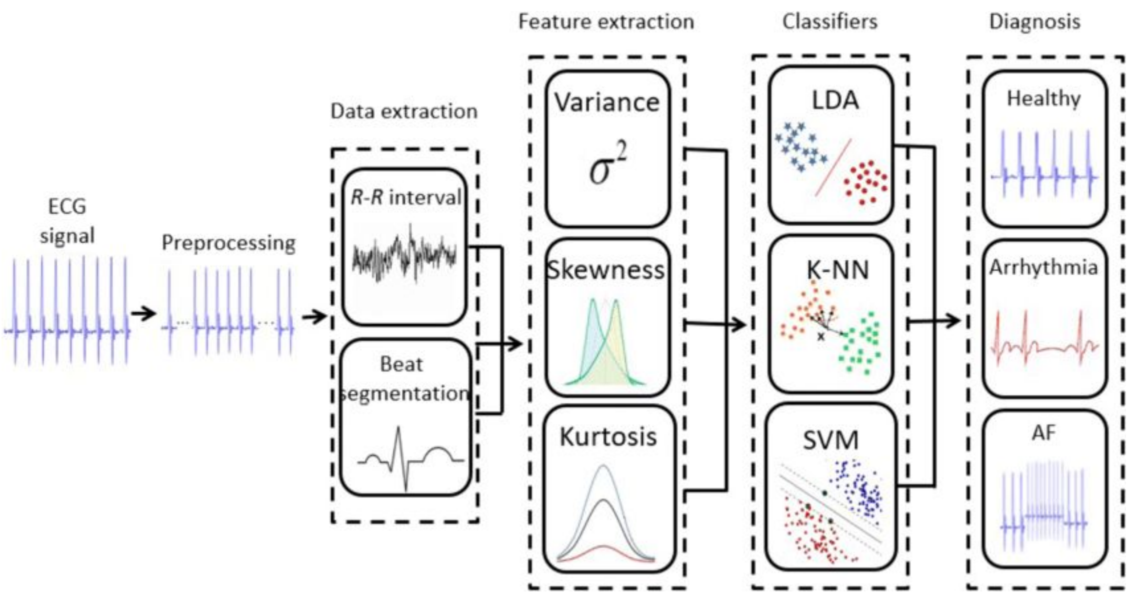

10.1038/s41598-021-95042-2

# Electrocardiogram (ECG)

Many useful and useless analyses of ECGs:

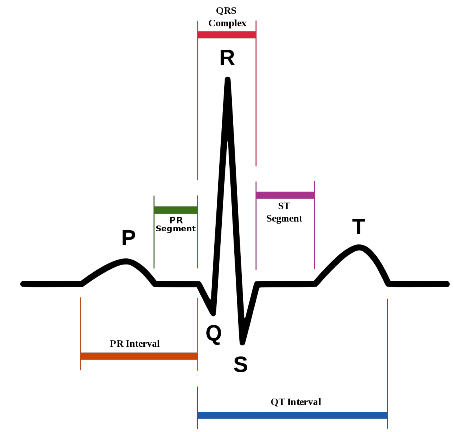
- Predict age and sex
- Detect anaemia (>90% accuracy with demographic data)
- Predict likelihood of low ejection fraction
- Automated detection of amyloid heart/cardiomyopathy/mitral valve prolapse
- Predict 1-year mortality (AUC > 0.85)

Methods generally require some form of ECG segmentation:

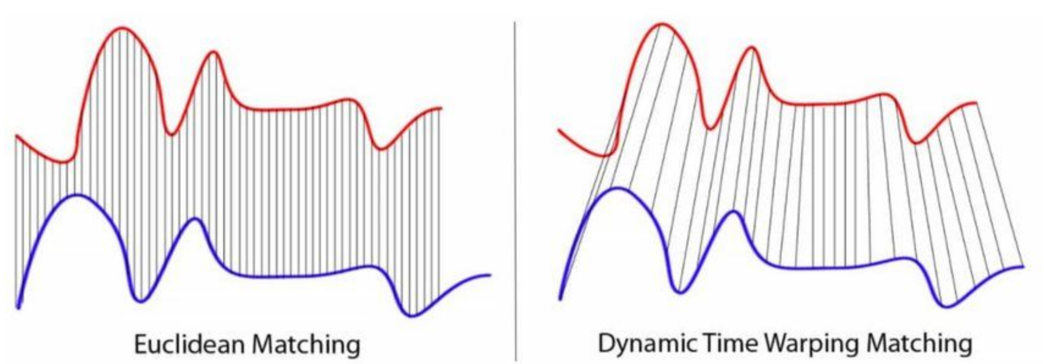


# Dynamic Time-Warping beat segmentation

- Clean ECG: just identify highest peaks but ECG is often noisy
- Know what a heartbeat looks like: align to ECG
- Often detecting arrhythmias/abnormal heartbeats: may not align
- Allow time to be “fuzzy” in alignment: dynamic time warping



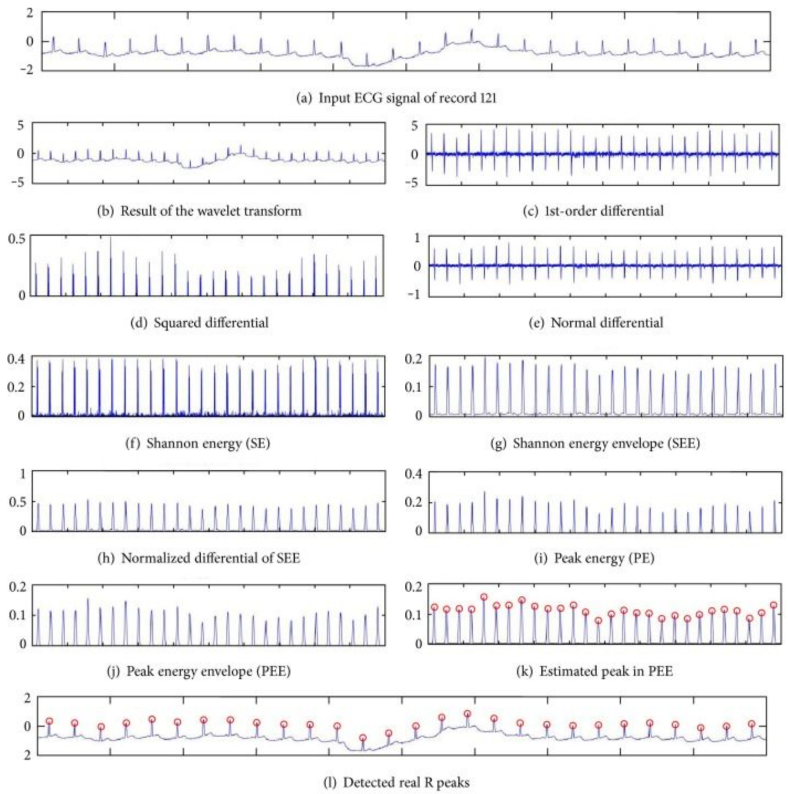


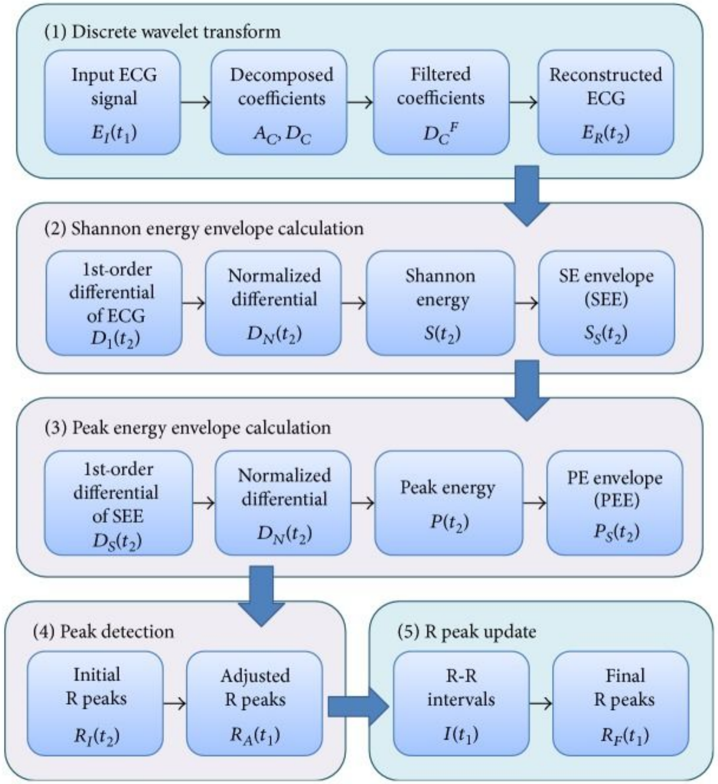


# Wavelet Transform Beat Segmentation

- Wavelet transforms can make R-peaks very clear even in noisy data
- Hand-crafted features can then be extracted

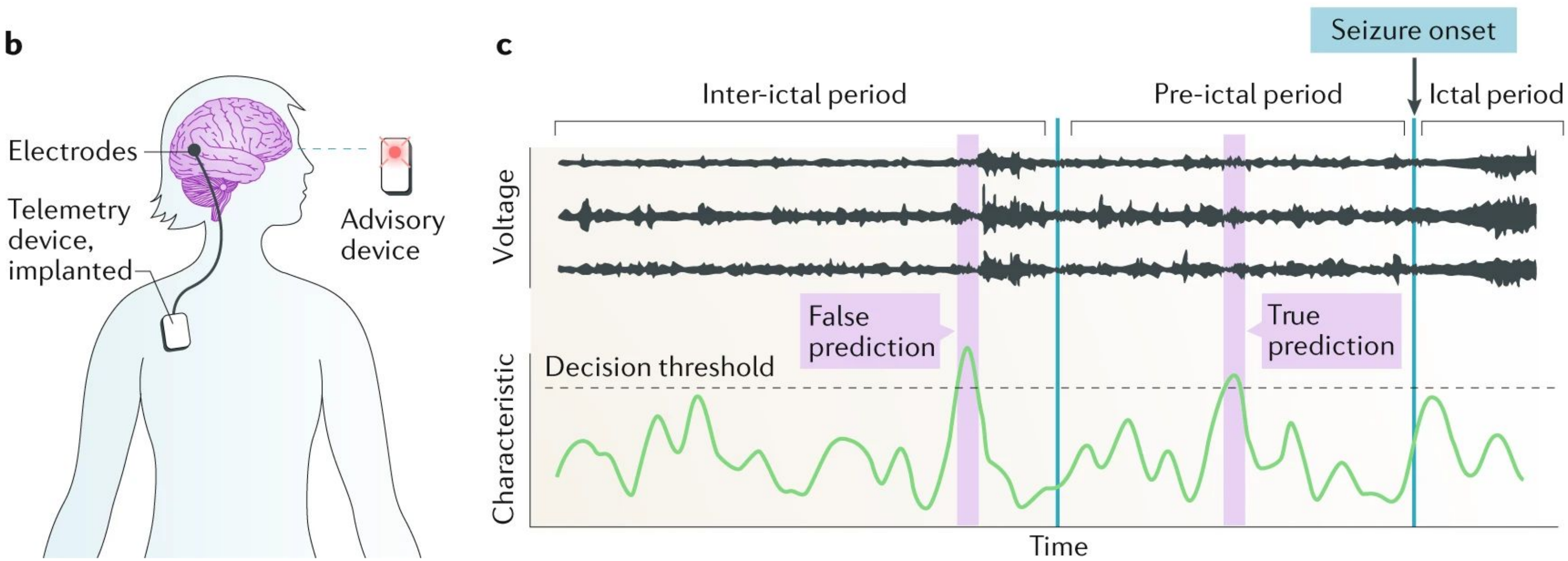
- Alternatively deep models can be used to learn wavelets and segmentations (unsupervised or supervised)



# Predicting epileptic seizures

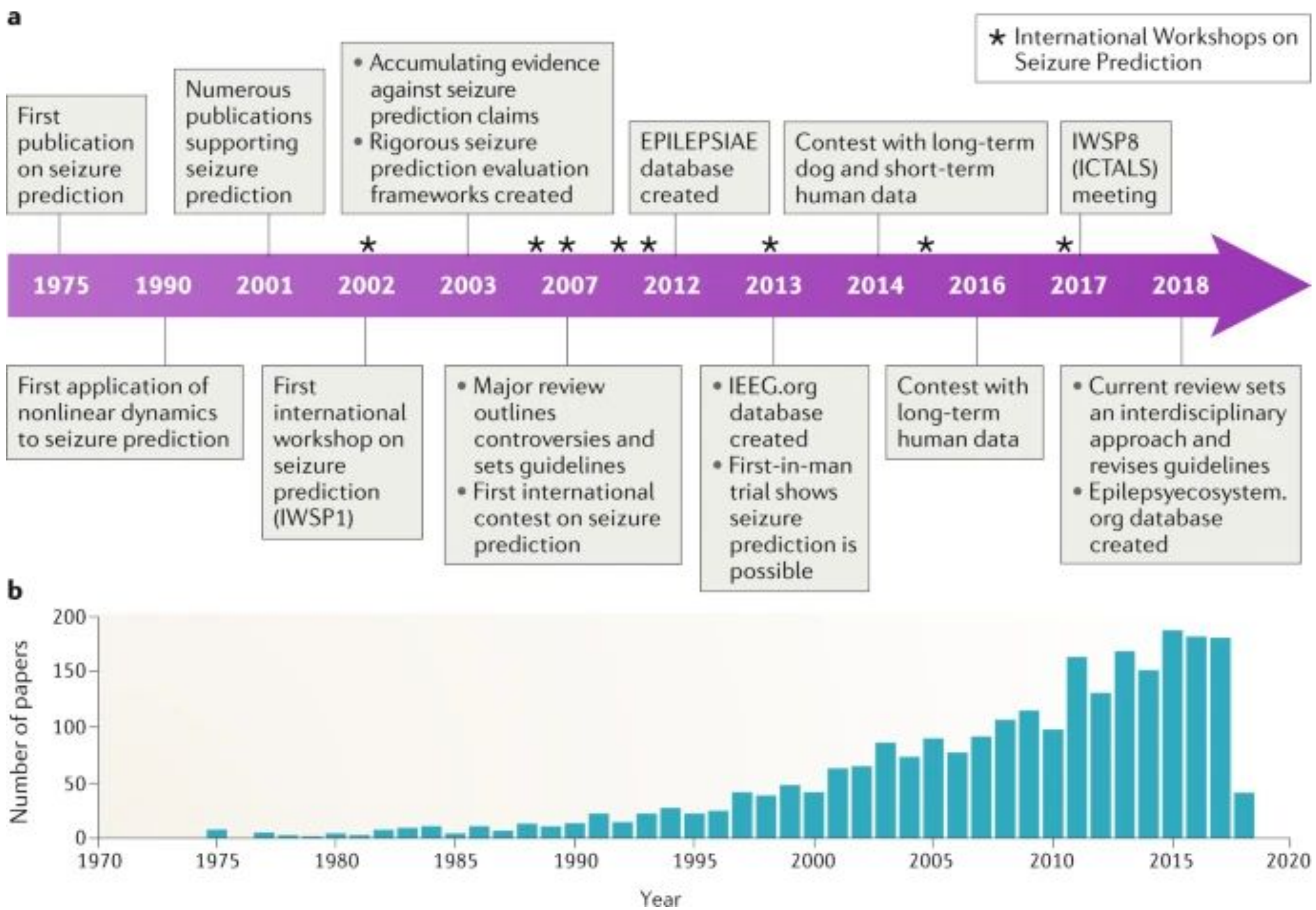
- Epilepsy has a global prevalence of 1% (80 million)
- 30% of cases not treatable with anti-epiletic medication  (2.4 million)
- Unpredictability of seizures is major source of mortality and morbidity
- Permanent intracranial EEGs now possible




10.1038/s41582-018-0055-2

# Lots of research

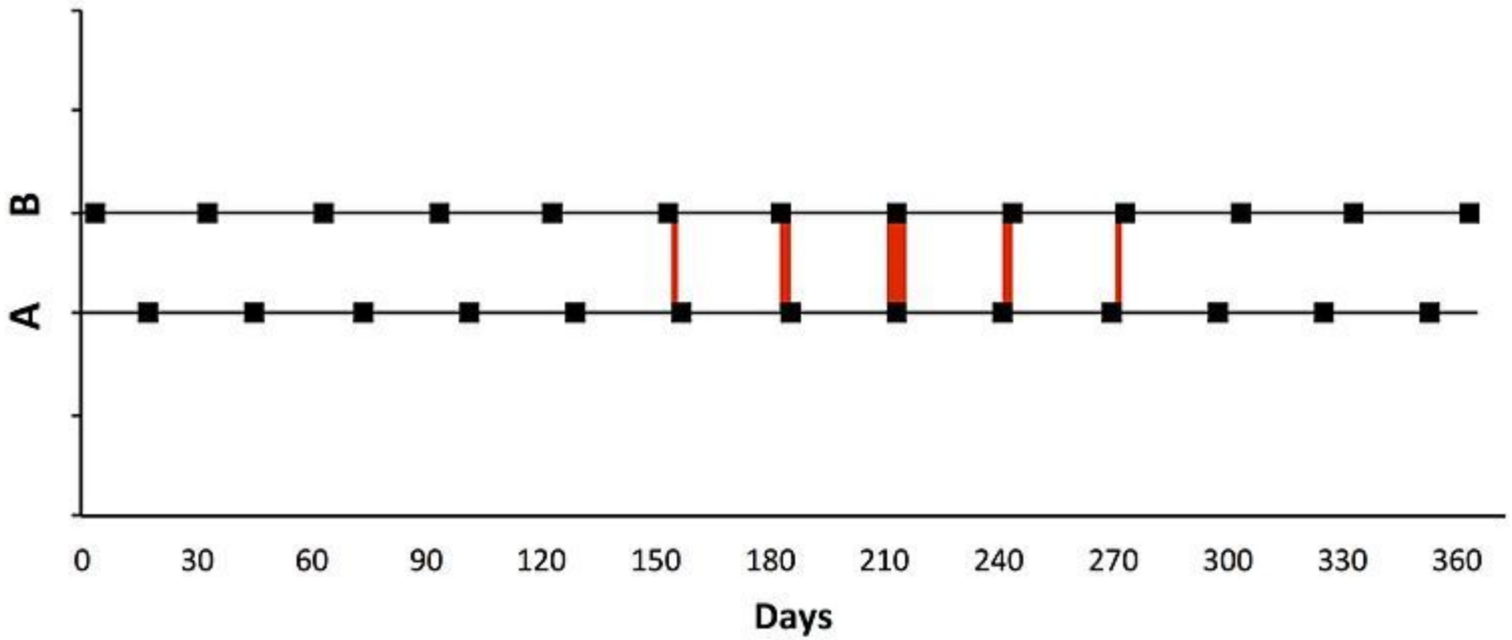
- 2007 review: insufficient evidence that seizures can be predicted




10.1038/s41582-018-0055-2

# Nulls for periodic signal can be challenging

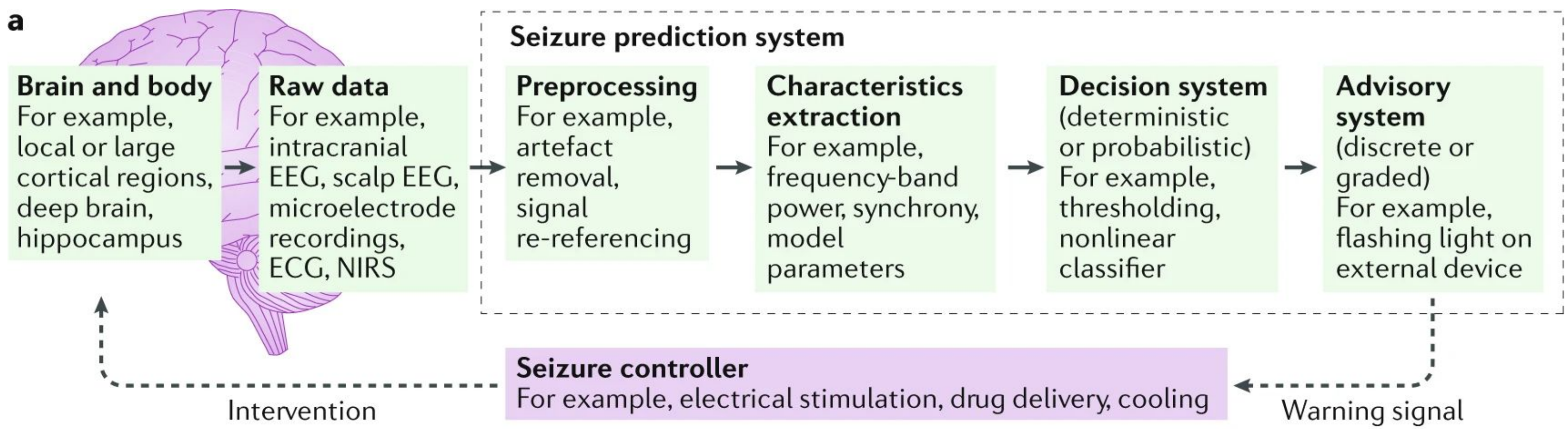
- Randomly shuffle seizure onset times => no pre-2007 actually worked




https://en.wikipedia.org/wiki/File:Yang_and_Schank_2006_converging_diverging_cycles2.jpg

# Most take a similar approach

- Inherently unbalanced data (seizures are rare compared to interictal EEG)
- Non-continuous datasets (i.e., inter-ictal and pre-ictal chunks) can make task easier than reality
- Scoring predictions is challenging (prediction window / time to seizure onset)
- Suitable baseline performance metric (random prediction: diurnal?)
- Inter-person variance can be large (electrode placement, cycles)



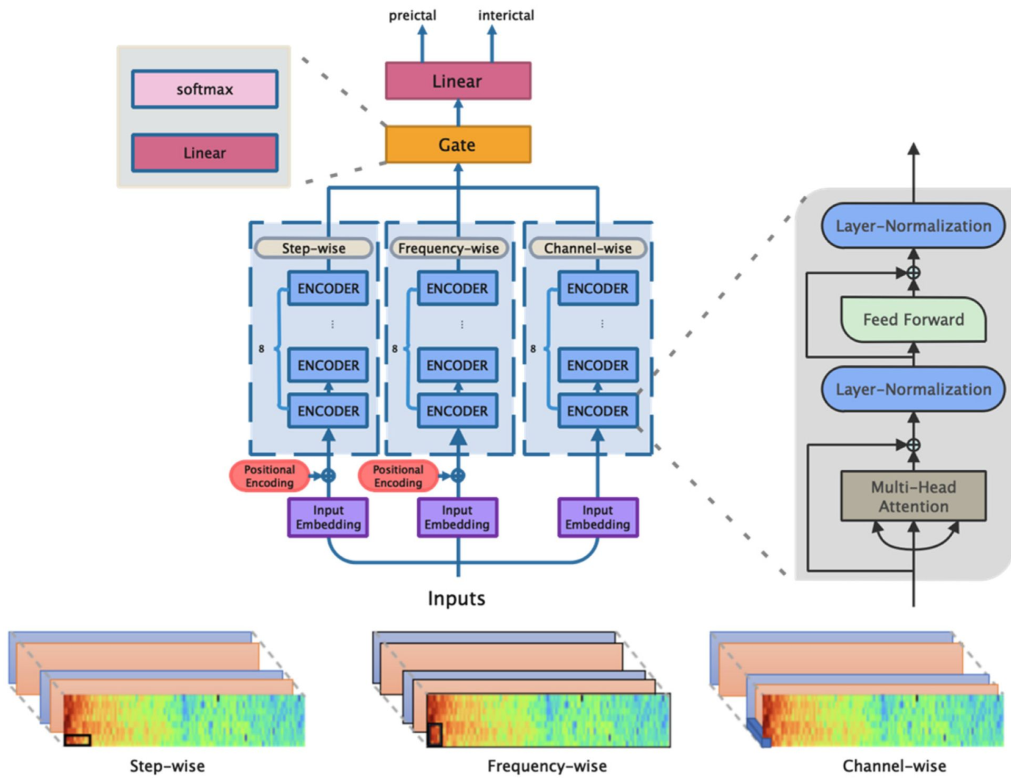
# American Epilepsy Society Seizure Prediction Challenge

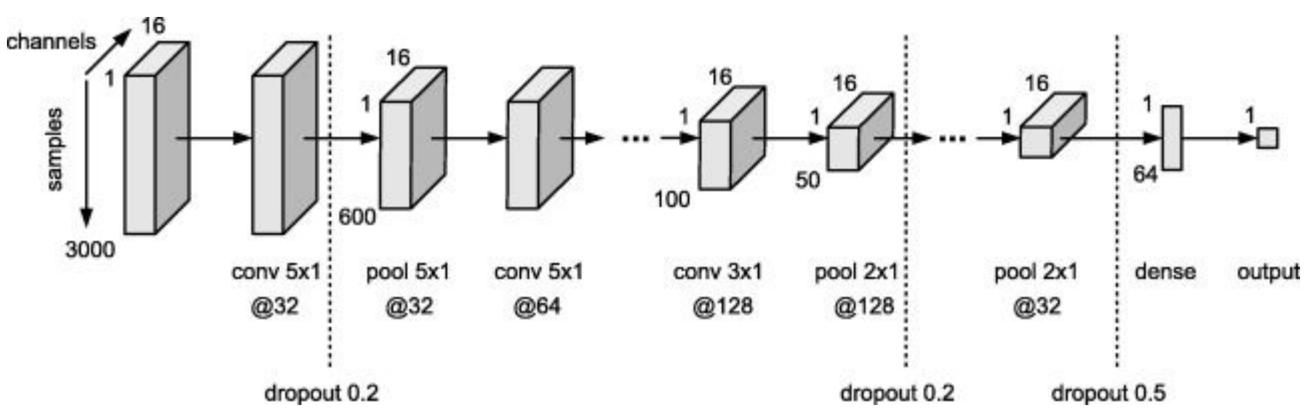
- Winning entry: well-crafted features for GLMs (averages with RF)
- General approach: bunch of features into large ensemble models
- 2014 (still relatively early days for CNNs being used outside of images)

| | Dog_1 | Dog_2 | Dog_3 | Dog_4 | Dog_5 | Patien | Patien | all |
|---|---|---|---|---|---|---|---|---|
| SVC_ica_psd_logfBB_AND_ica_xcorr-tpeak | 0.8175 | 0.9253 | 0.8129 | 0.7371 | 0.8617 | 0.8346 | 0.7526 | 0.8493 |
| SVC_ica_ilingam-causalindex_AND_ica_PSDlogfcorrcoef | 0.8125 | 0.9904 | 0.7356 | 0.6645 | 0.8540 | 0.9355 | 0.4759 | 0.8362 |
| SVC_ica_PSDlogfcorrcoef_AND_ica_pwling5 | 0.8101 | 0.8916 | 0.7657 | 0.6663 | 0.9073 | 0.8236 | 0.5182 | 0.8280 |
| SVC_ica_cov_AND_ica_lmom-3 | 0.8081 | 0.9383 | 0.8109 | 0.6689 | 0.9018 | 0.7532 | 0.6210 | 0.8458 |
| SVC_ica_corrcoefeig_AND_ica_PSDlogfcorrcoef | 0.8071 | 0.9858 | 0.7791 | 0.6792 | 0.8970 | 0.9208 | 0.4920 | 0.8456 |
| SVC_ica_psd_logfBB_AND_ica_PSDlogfcorrcoef | 0.8063 | 0.9856 | 0.8033 | 0.7442 | 0.8477 | 0.9002 | 0.7618 | 0.8618 |
| SVC_ica_ilingam-causalorder_AND_ica_psd_logfBB | 0.8059 | 0.9685 | 0.8166 | 0.7623 | 0.8398 | 0.8660 | 0.8117 | 0.8596 |
| SVC_ica_ilingam-causalindex_AND_ica_psd_logfBB | 0.8049 | 0.9783 | 0.8133 | 0.7610 | 0.8429 | 0.8653 | 0.8020 | 0.8619 |
| SVC_ica_lmom-3_AND_ica_PSDlogfcorrcoef | 0.8011 | 0.9816 | 0.7282 | 0.6717 | 0.8756 | 0.9547 | 0.4971 | 0.8400 |
| SVC_ica_lmom-2_AND_ica_psd_logfBB | 0.8008 | 0.9755 | 0.8358 | 0.7575 | 0.8591 | 0.8486 | 0.8169 | 0.8643 |
| SVC_ica_ampcorrcoef-alpha-eig_AND_ica_pib_ratioBB | 0.8004 | 0.9546 | 0.8612 | 0.7408 | 0.8666 | 0.8825 | 0.7204 | 0.8584 |
| SVC_ica_pib_ratioBB_AND_ica_pwling5 | 0.7991 | 0.8737 | 0.8724 | 0.7281 | 0.8607 | 0.8163 | 0.5582 | 0.8353 |
| SVC_ica_gcaus_AND_ica_pib_ratioBB | 0.7973 | 0.9770 | 0.8548 | 0.7212 | 0.8296 | 0.8927 | 0.7014 | 0.8566 |
| SVC_ica_lmom-4_AND_ica_psd_logfBB | 0.7962 | 0.9711 | 0.8367 | 0.7613 | 0.8446 | 0.8588 | 0.8186 | 0.8613 |
| SVC_ica_ampcorrcoef-high_gamma_AND_ica_phase-beta-sync | 0.7921 | 0.9450 | 0.7474 | 0.6594 | 0.9699 | 0.9032 | 0.5327 | 0.8439 |
| SVC_ica_ampcorrcoef-low_gamma_AND_ica_psd_logfBB | 0.7899 | 0.9449 | 0.8367 | 0.7463 | 0.8714 | 0.8434 | 0.7935 | 0.8506 |
| SVC_ica_ampcorrcoef-alpha-eig_AND_ica_phase-beta-sync | 0.7889 | 0.9789 | 0.7460 | 0.7239 | 0.9605 | 0.9157 | 0.5430 | 0.8559 |
| SVC_ica_ampcorrcoef-high_gamma-eig_AND_ica_corrcoef | 0.7873 | 0.9320 | 0.8261 | 0.6221 | 0.9360 | 0.7966 | 0.6627 | 0.8563 |
| SVC_ica_PSDlogfcorrcoef_AND_ica_xcorr-ypeak | 0.7869 | 0.9826 | 0.7565 | 0.6327 | 0.9433 | 0.8982 | 0.7139 | 0.8641 |
| SVC_ica_psd_logf_AND_ica_PSDlogfcorrcoef | 0.7866 | 0.9816 | 0.8291 | 0.7369 | 0.8993 | 0.8998 | 0.7395 | 0.8725 |
| SVC_ica_phase-beta-sync_AND_ica_pib | 0.7770 | 0.9806 | 0.8304 | 0.6789 | 0.9556 | 0.9166 | 0.6674 | 0.8626 |
| SVC_ica_ampcorrcoef-beta_AND_ica_phase-beta-sync | 0.7737 | 0.9440 | 0.7624 | 0.7355 | 0.9632 | 0.8947 | 0.5755 | |

# Modern approaches

- Deep neural networks (static or dynamic input)
- Learnt representations of EEGs (wavelet, kernels, attention, embeddings)
- Still suffer from inter-person variance (and relatively rarity of seizures): individualised tuning
- Specificity still challenging



# High variance clinical trials: implementation science is key

- 3-100% accuracy across >=3 seizures across individuals
- Seizures are non-random (short and long-term temporal dependence)
- Diving into why they don't haven't worked:
  - Individual seizure frequency
  - long-term temporal variations in seizure frequency
  - multimodal distributions of seizure duration and inter-ictal intervals

Lessons learnt:

- EEGs give poor mechanistic insight
- Emerging ideas about how seizures work: excitation/inhibition imbalance vs aberrant behaviour emerging from network parameters
- Implementation science is often more important than underlying ML

# Learning Overview

- Types of medical sensor data
- Time-domain approaches: detrending/regression models
- Alternative decomposition: frequency/time-frequency
- State-space approaches: hidden markov models
- Handling data from multiple sensors
- General purpose Bayesian approaches: Gaussian Processes
- Cough-detection example
- Segmentation of heartbeats example
- Seizure prediction example