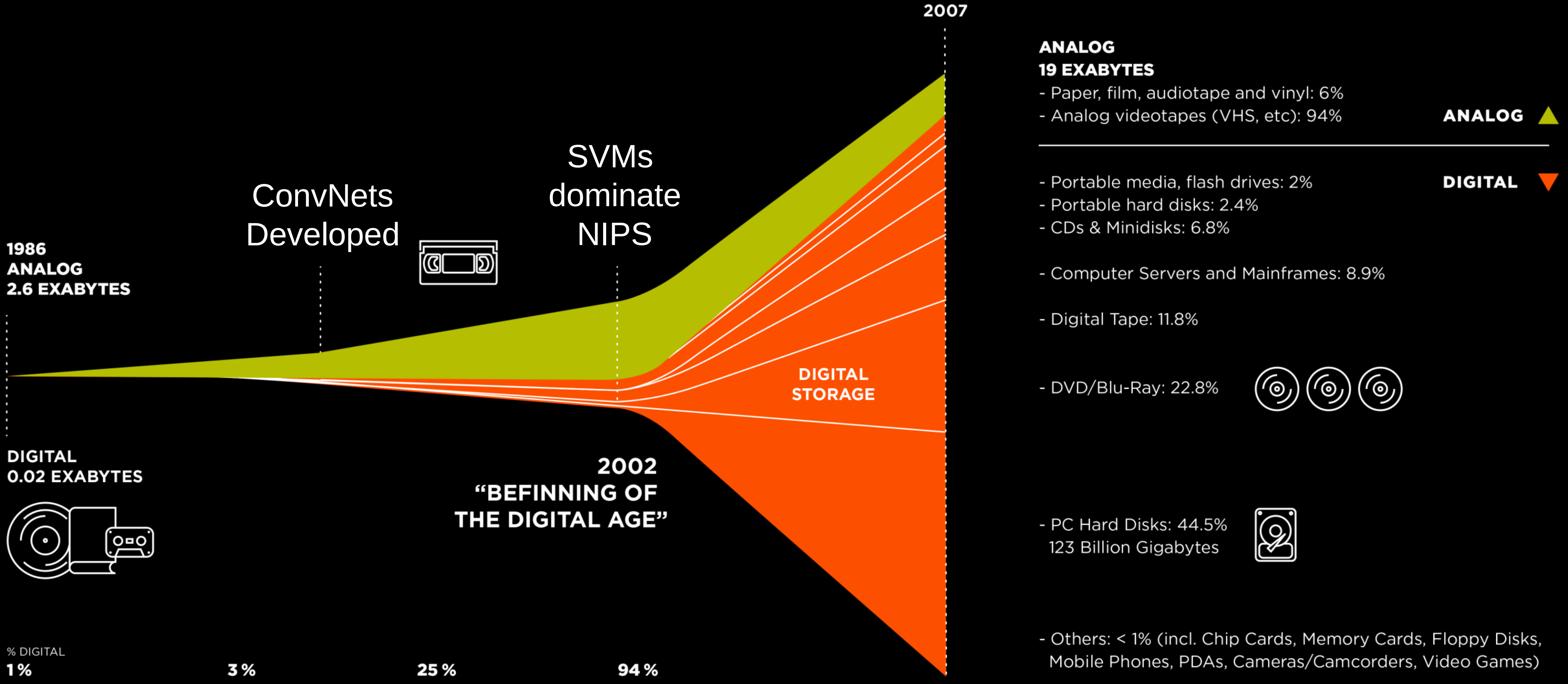
Gaussian Processes for Machine Learning

NEIL LAWRENCE

UNIVERSITY OF SHEFFIELD

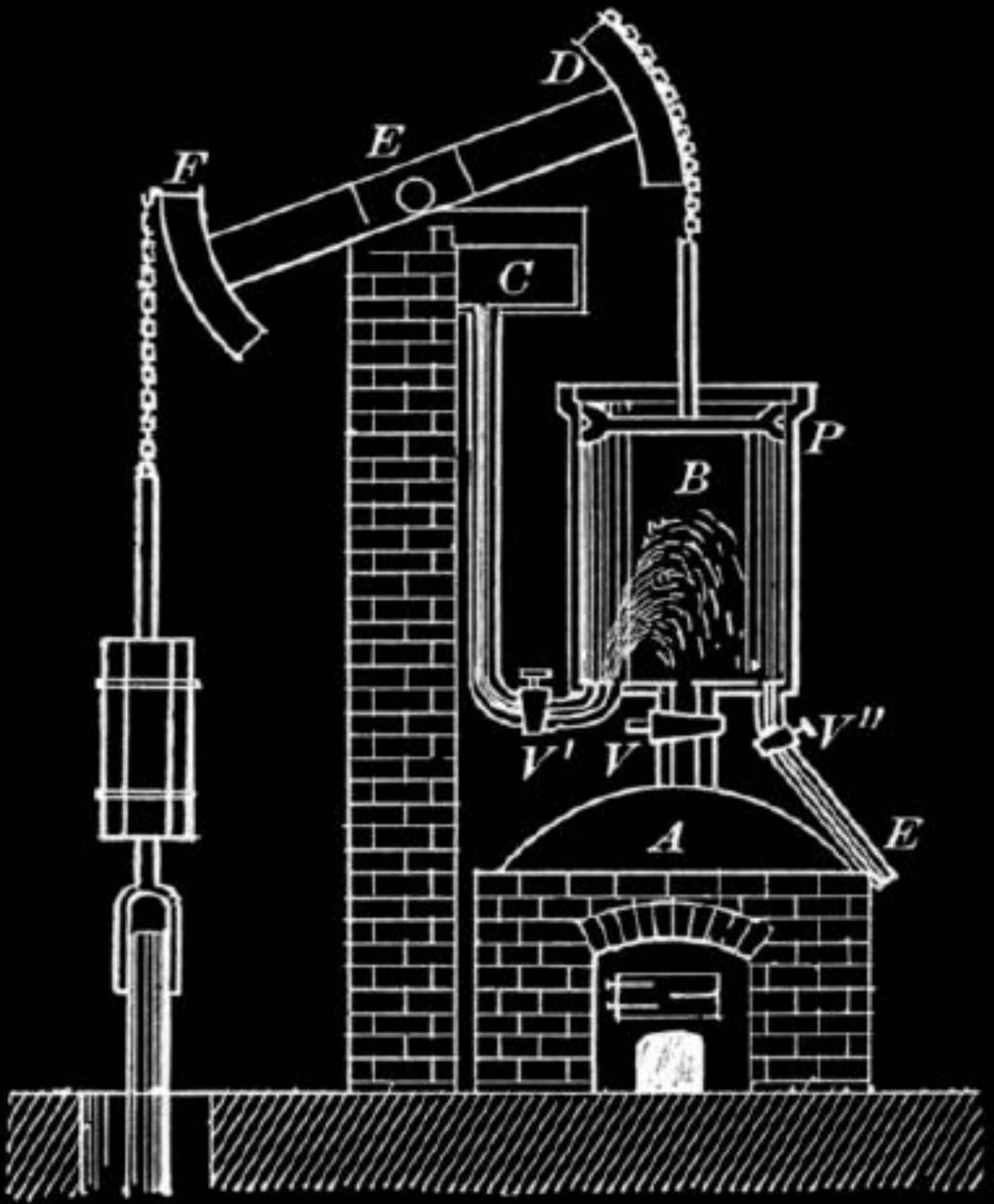
@lawrennd

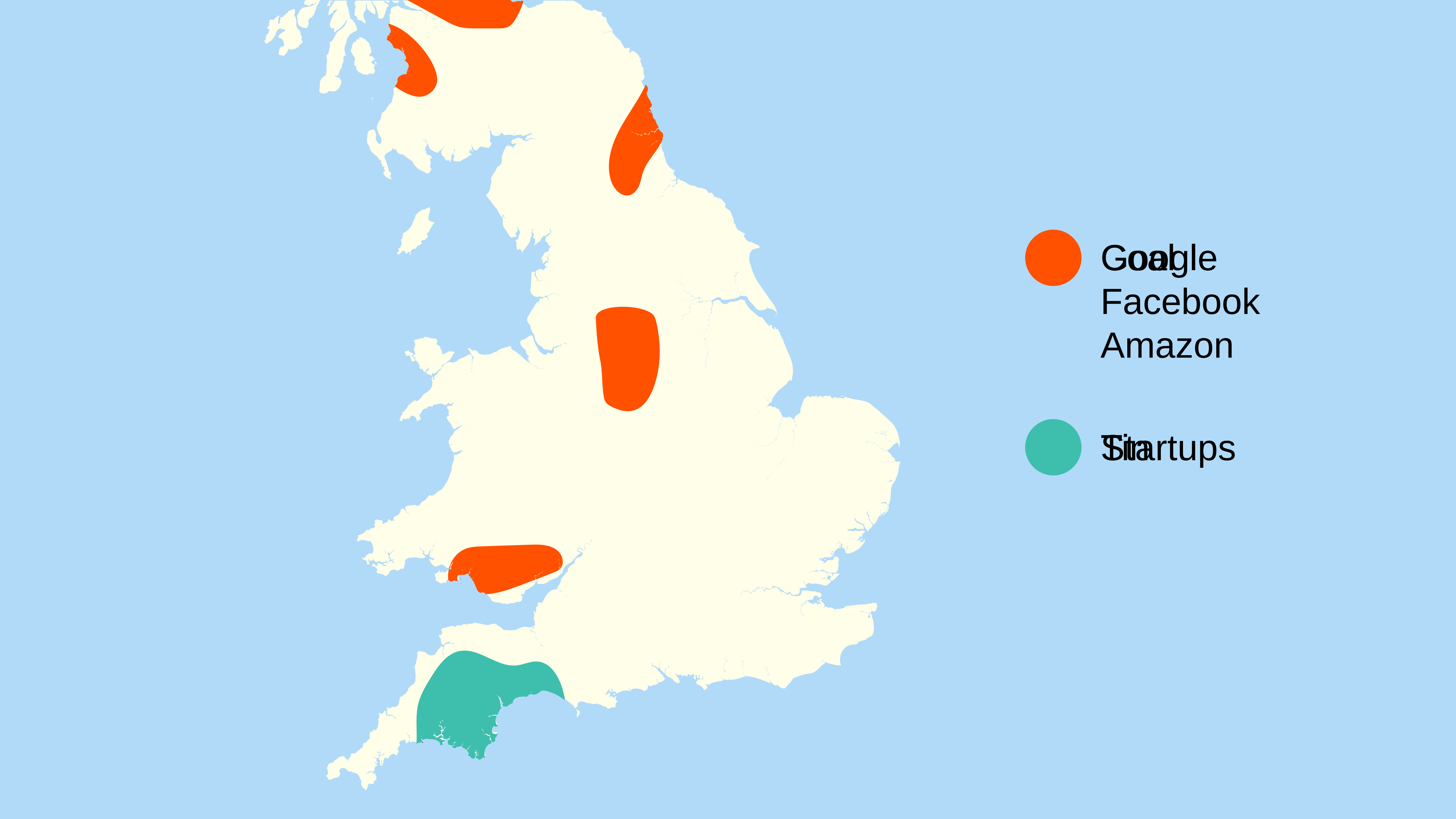
GLOBAL INFORMATION STORAGE CAPACITY
IN OPTIMALLY COMPRESSED BYTES

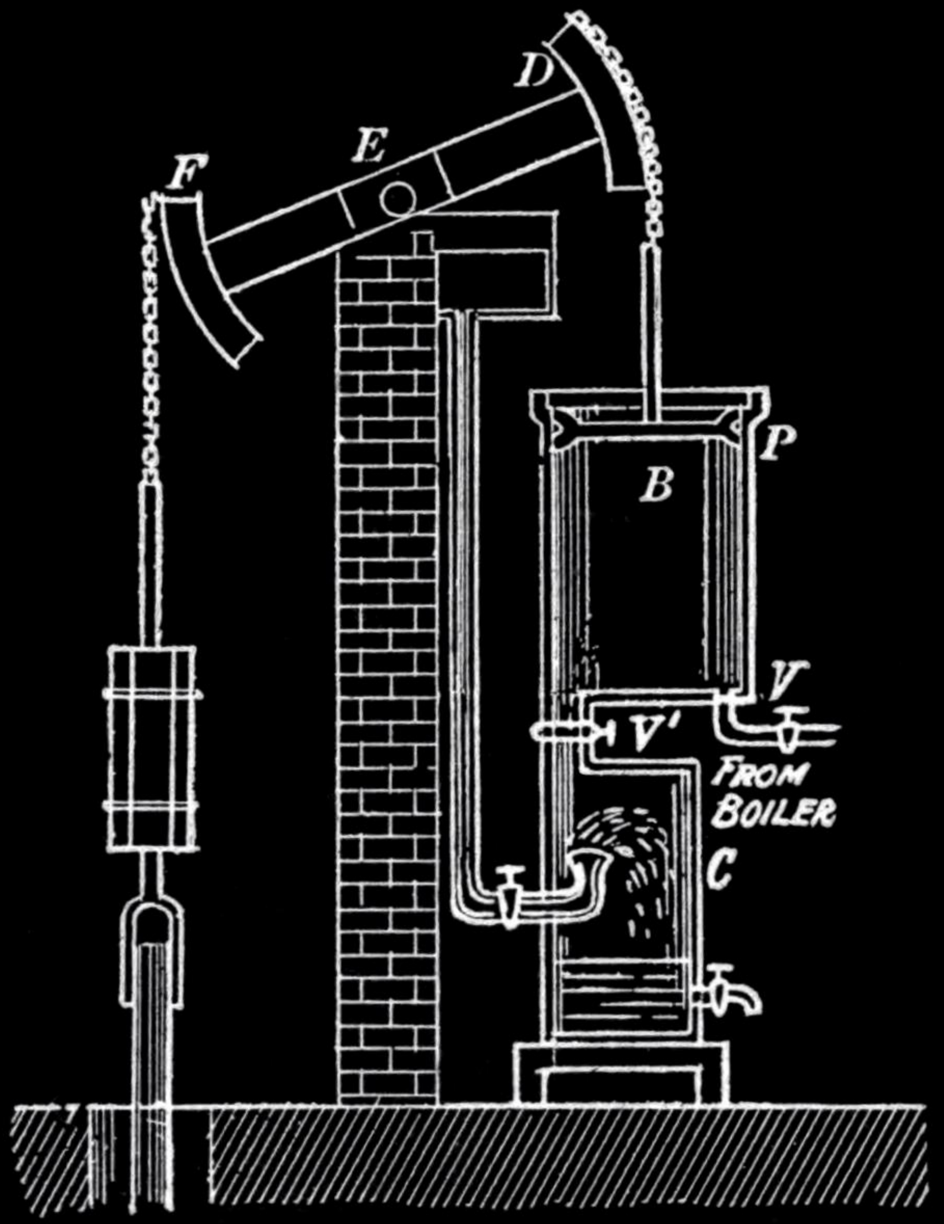


Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, andCompute Information. Science, 332 (6025), 60-65. martinhilbert.net/worldinfocapacity.html

DIGITAL
280 EXABYTES


D
E
F
C
P
B
V'
V
V''
A
E


Google
Facebook
Amazon
Startups


D
E
F
P
B
V
V'
FROM BOILER
C

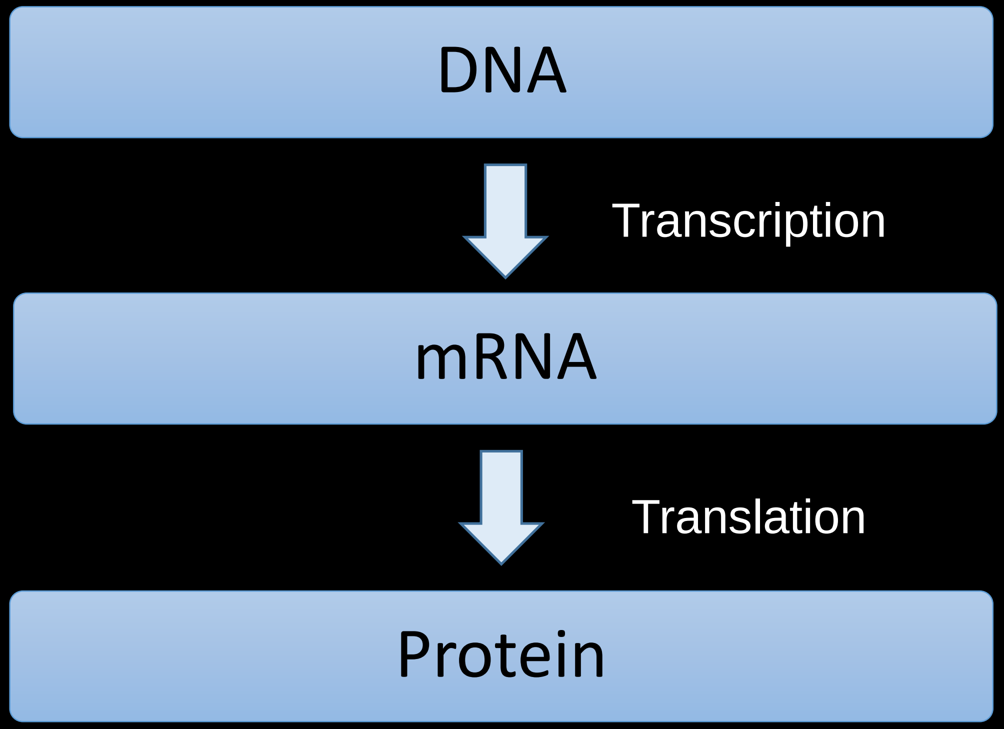
The Data are Not Enough

- Four pillars:
  - Deterministic/Stochastic
  - Mechanistic/Emipirical

- **Goal:** *model complex phenomena over time*
- **Problem:**
  - *Mechanistic* models are often inaccurate
  - Data is often not rich enough for an *empirical* approach

- Question 1: How do we combine *inaccurate physical model* with machine learning?

Central Dogma



Decision: Transcription Factors

Measured using Microarray since 1998 — mRNA

Translation

Difficult to measure — TF Protein

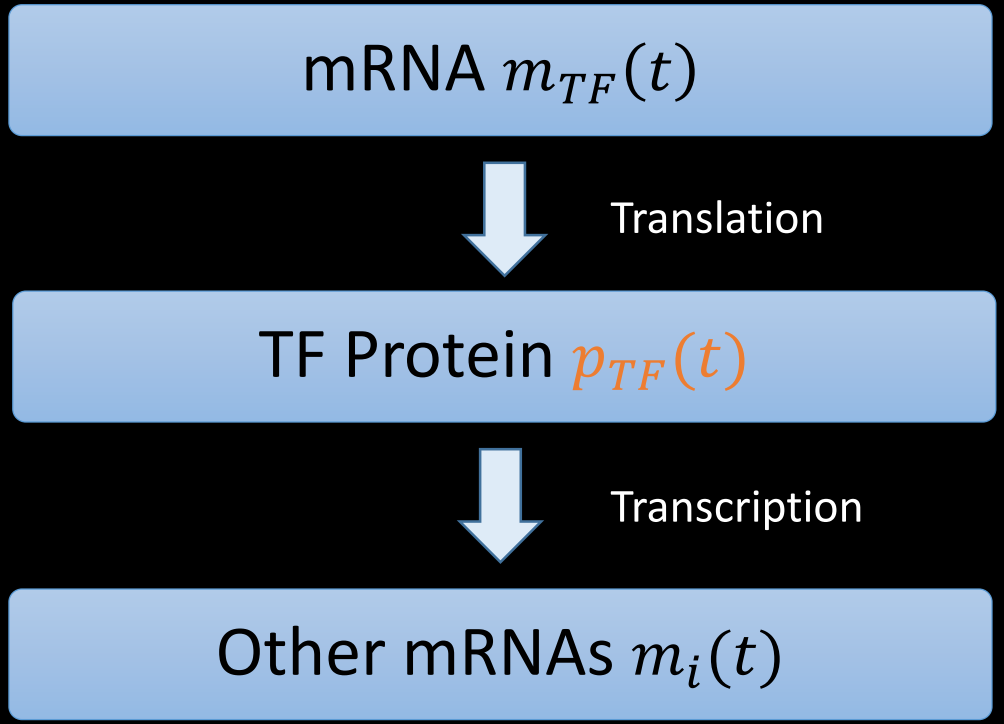
Transcription

Measured using Microarray since 1998 — Other mRNAs

Mechanistic Model

$$\frac{\mathrm{d}p_{TF}(t)}{\mathrm{d}t} = s_f m_{TF}(t) - d_f p_{TF}(t)$$

$$\frac{\mathrm{d}m_i(t)}{\mathrm{d}t} = s_i p_{TF}(t) - d_i m_i(t)$$



Need to Model $p_{TF}(t)$

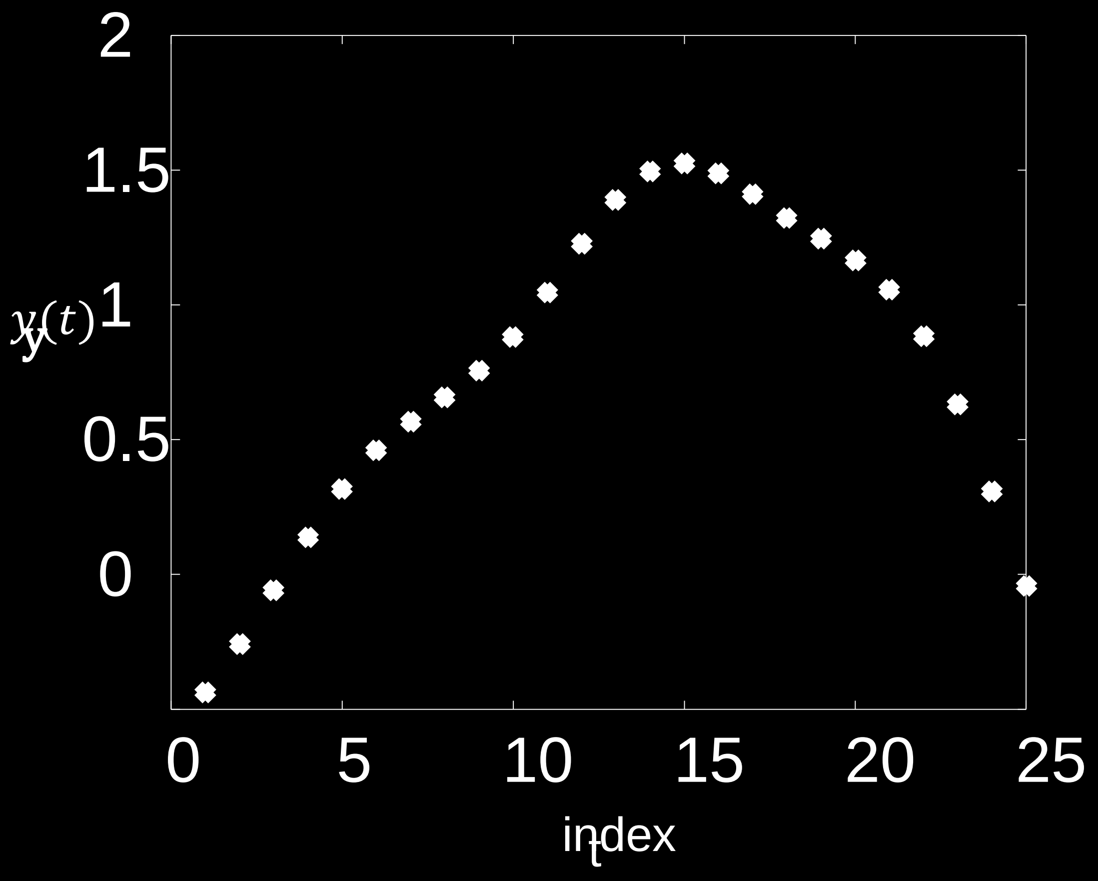
- Gaussian process: a *probabilistic* model for functions.
- Formally known as a *stochastic process*.
- Multivariate Gaussian is normally defined by a *mean vector*, $\boldsymbol{\mu}$, and a *covariance matrix*, $\mathbf{C}$.

$$y \sim N(\boldsymbol{\mu}, \mathbf{C})$$

- Gaussian process defined by a *mean function*, $\mu(t)$, and a *covariance function*, $c(t, t')$.

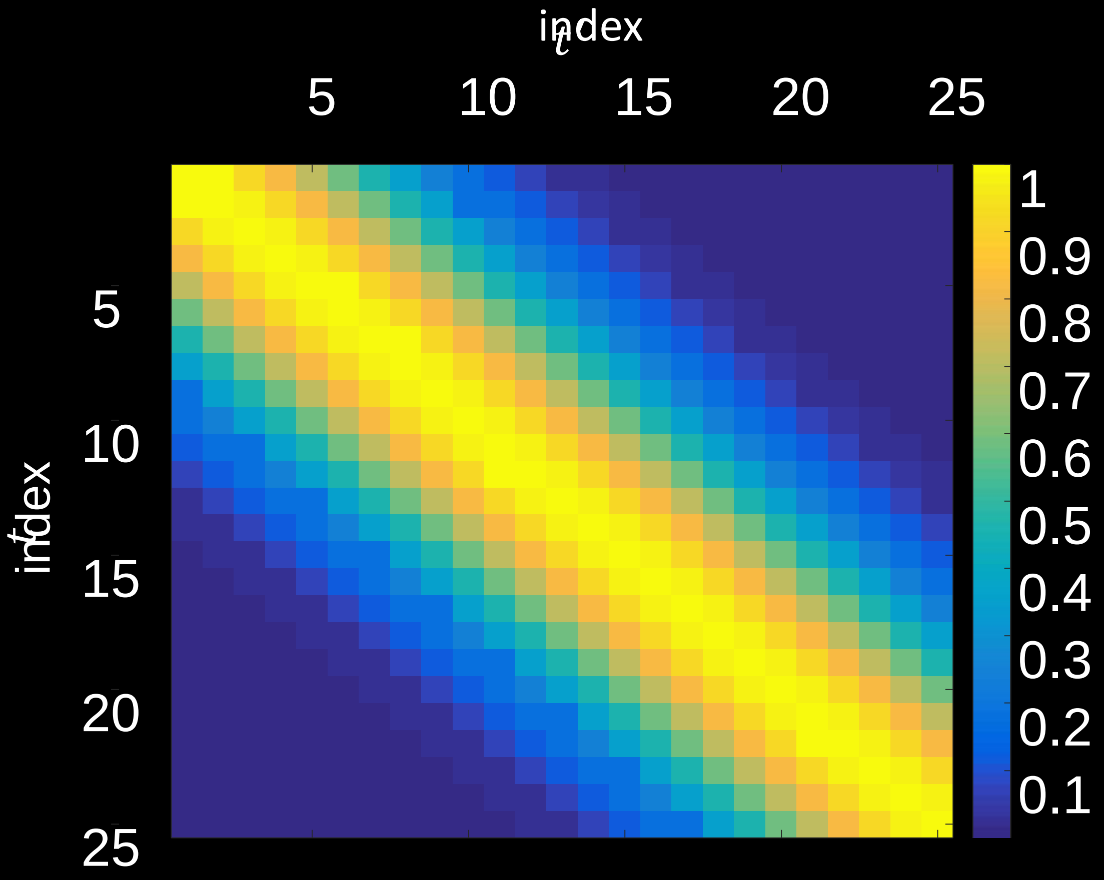
$$y(t) \sim N(\mu(t), c(t, t'))$$

Zero Mean Gaussian Samples
Zero Mean Gaussian Process Sample



samples from Gaussian

samples from Gaussian process



covariance $\mathbf{C}$

covariance function $c(t, t')$

Gaussian Processes

x_1, y_1 $\quad$ $p(\mathbf{f}|\mathbf{x})$ $\quad$ x_2, y_2

$p(y_1|f_1)$ $\quad$ $p(\mathbf{f}|\mathbf{y},\mathbf{x})$ $\quad$ $p(y_2|f_2)$

Results

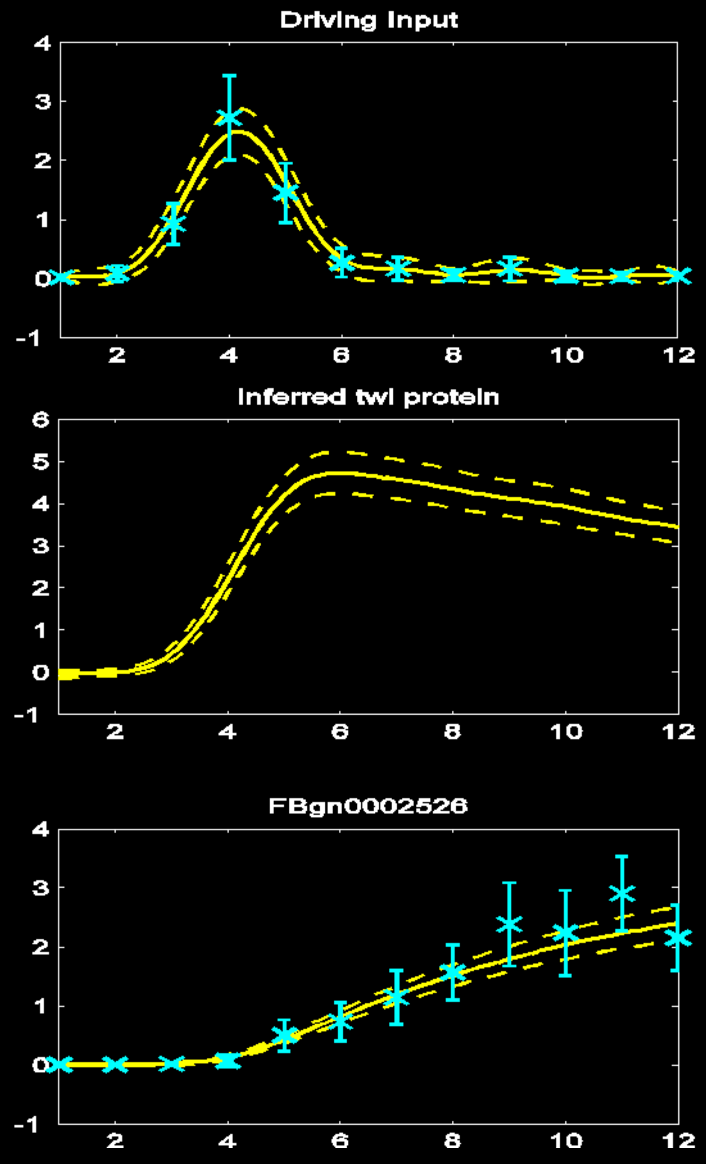
$$\frac{\mathrm{d}p_{TF}(t)}{\mathrm{d}t} = s_f m_{TF}(t) - d_f p_{TF}(t)$$

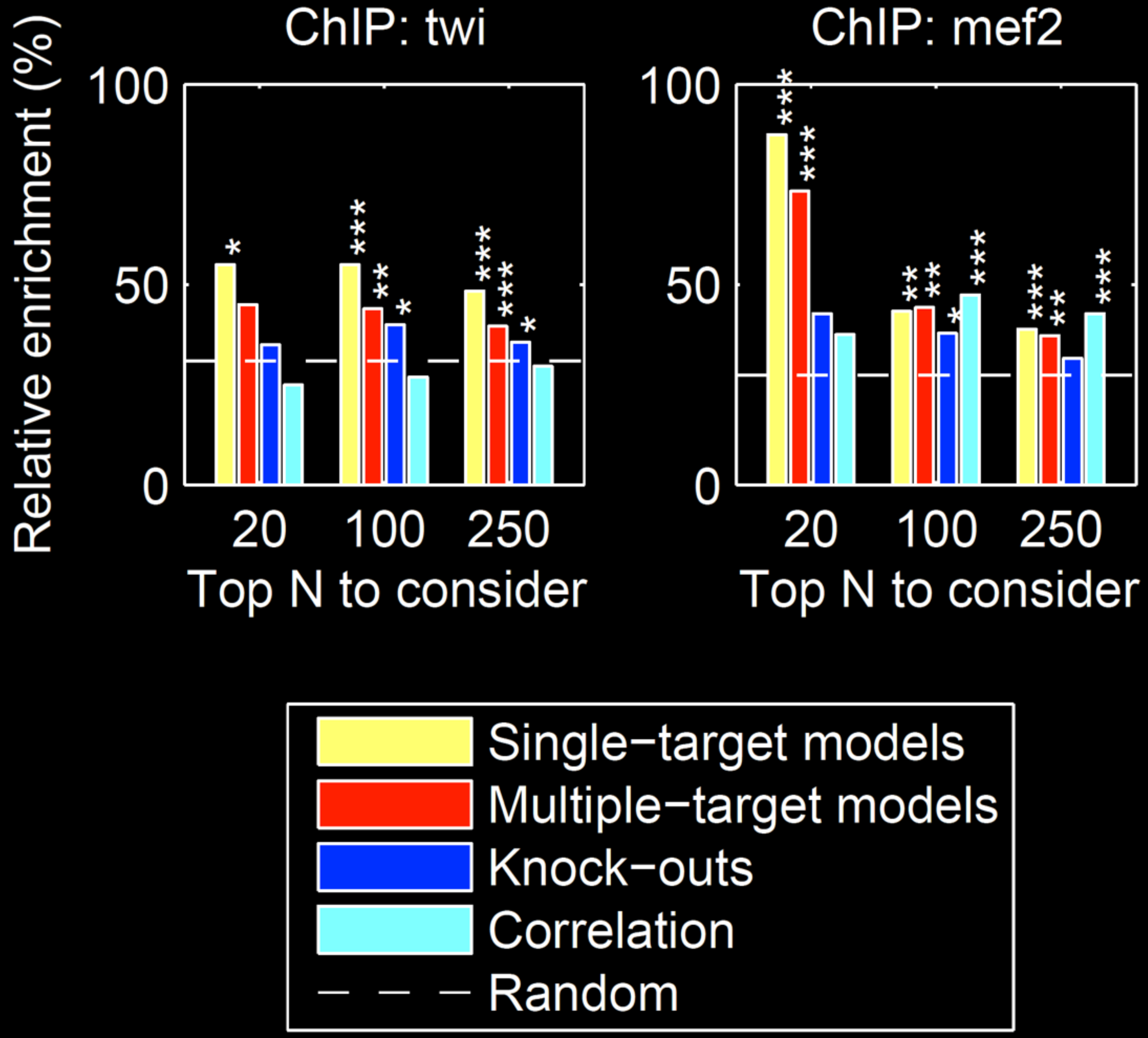
$$\frac{\mathrm{d}m_i(t)}{\mathrm{d}t} = s_i p_{TF}(t) - d_i m_i(t)$$

$m_{TF}(t)$

$p_{TF}(t)$

$m_i(t)$




ChIP: twi
ChIP: mef2
Relative enrichment (%)
100
50
0
20
100
250
Top N to consider
Single−target models
Multiple−target models
Knock−outs
Correlation
Random

MATLAB Demo

- demo_2016_04_28_amazon.m

Further Challenge

- This model inter-relates different functions with mechanistic understanding.

- What if you need to inter-relate across different modalities of data at different scales.

- *E.g.* biopsy images + genetic test + mammogram for breast cancer diagnostics.

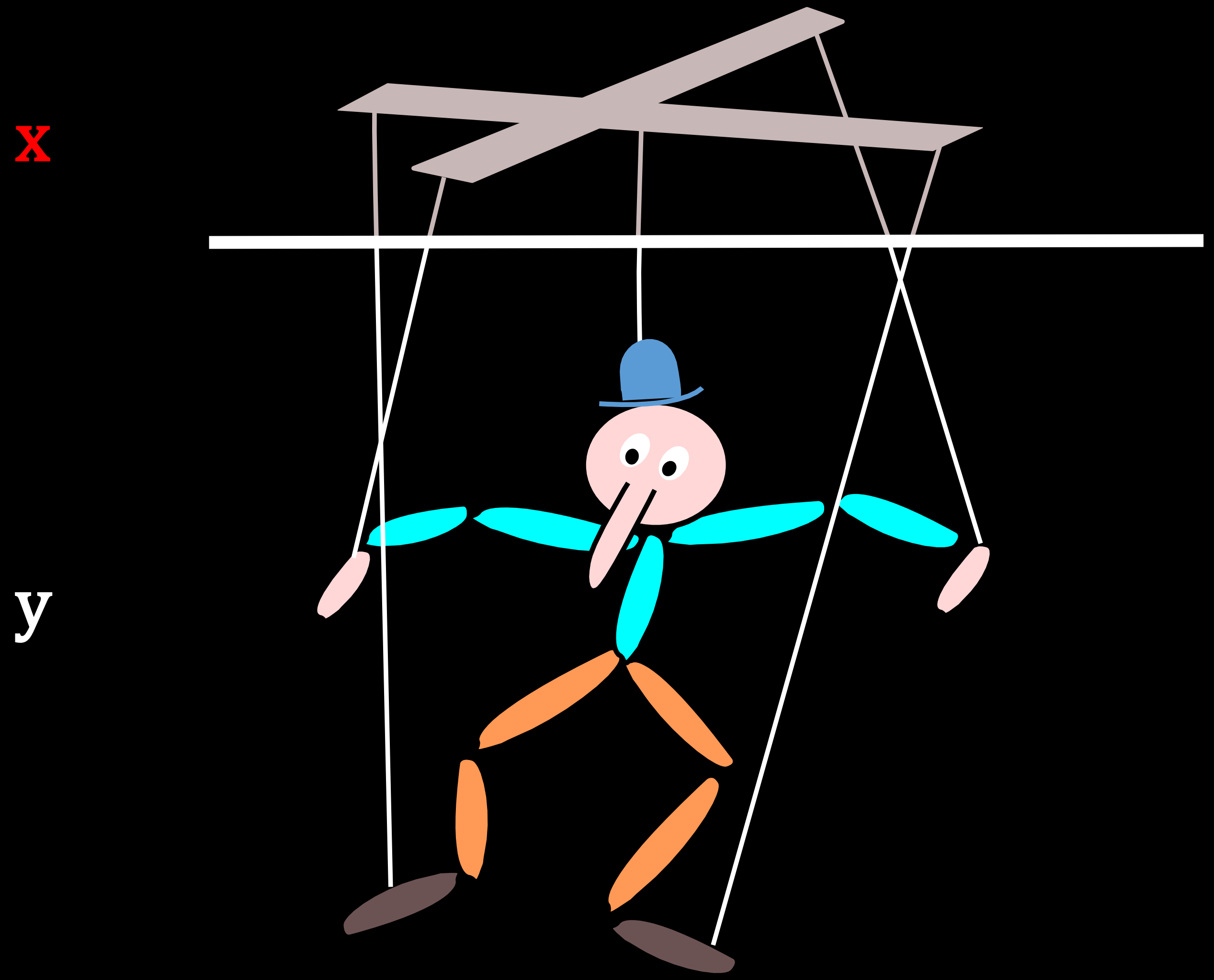
The Data are Not Enough

- Four pillars:
  - Deterministic/Stochastic
  - Mechanistic/Empirical

- **Goal:** *model complex phenomena over time*
- **Problem:**
  - *Mechanistic* models are often inaccurate
  - Data is often not rich enough for an *empirical* approach

- Question 2: How do we formulate the right representations to integrate different data modalities?

Classical Latent Variables



Classical Treatment

- Assume *a priori* that

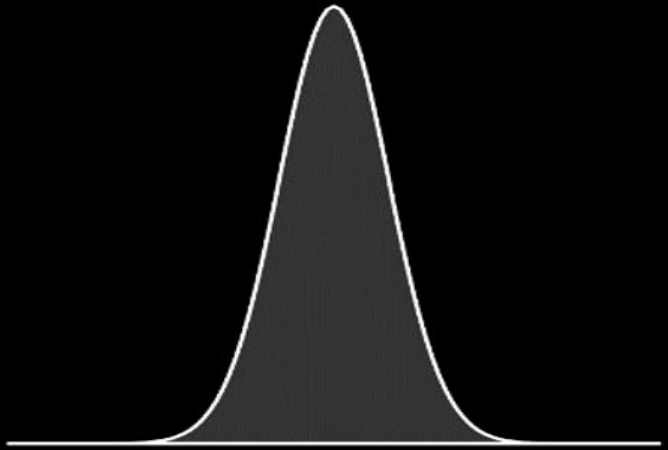
$$\mathrm{x} \sim N(0, \mathrm{I})$$

- Relate linearly to $\mathbf{y}$

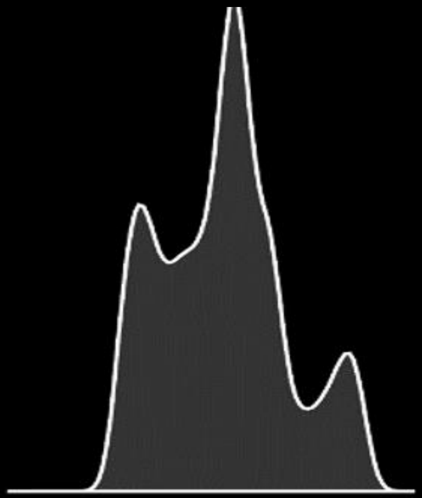
$$\mathbf{y} = \mathbf{W}\mathbf{x} + \boldsymbol{\epsilon}$$

- Framework covers many classical models PCA, Factor Analysis, ICA

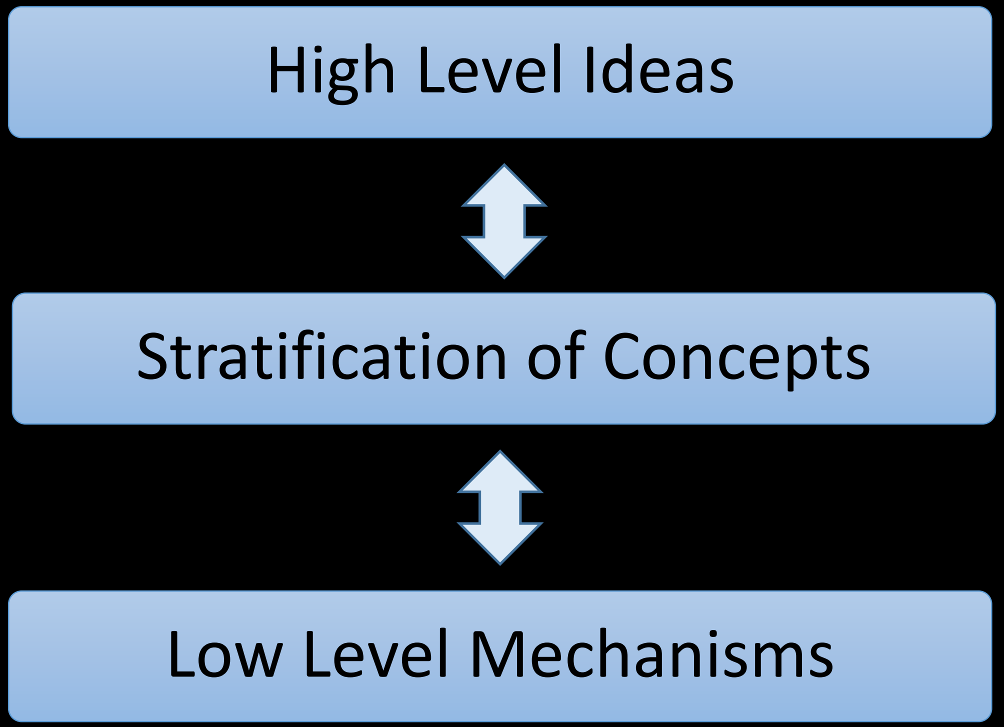
Render Gaussian Non Gaussian



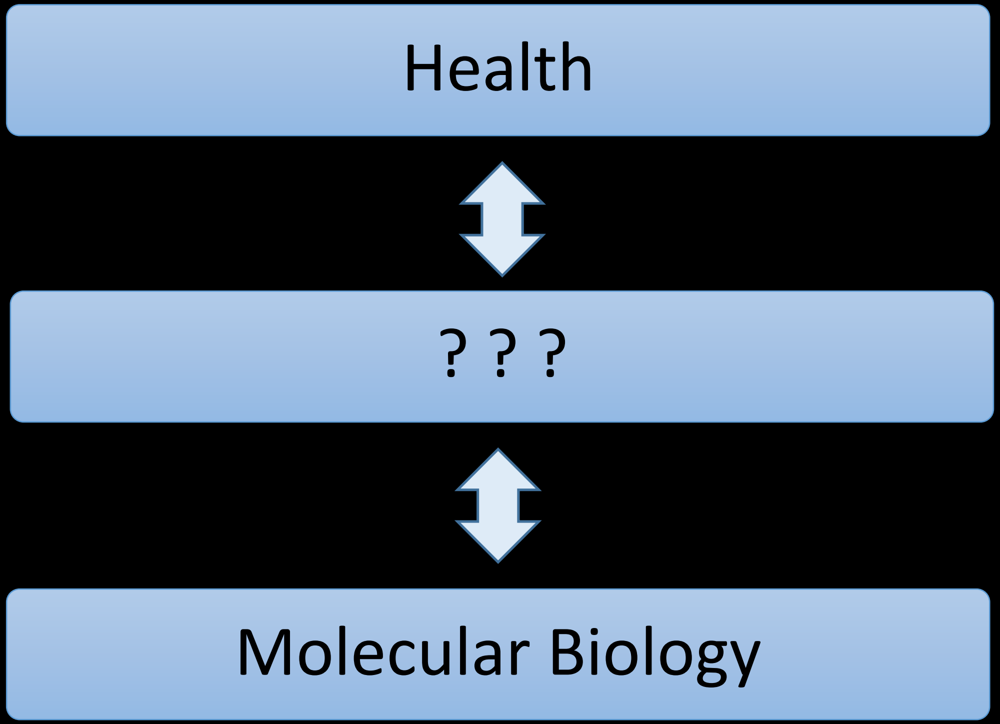
$$y = f(x)$$



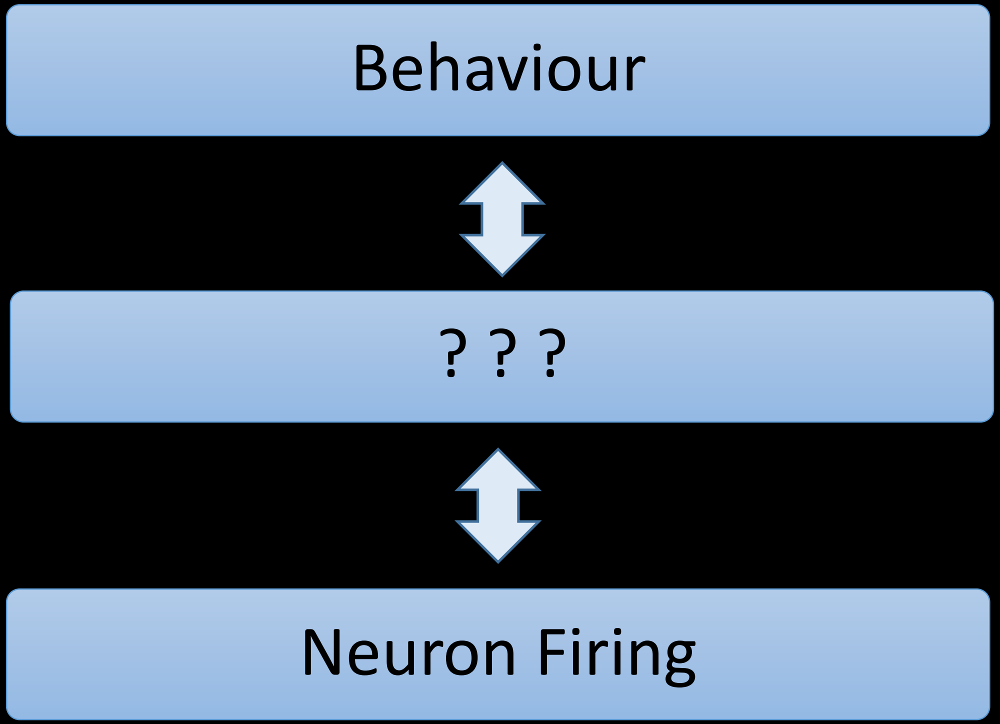
Use Abstraction for Complex Systems



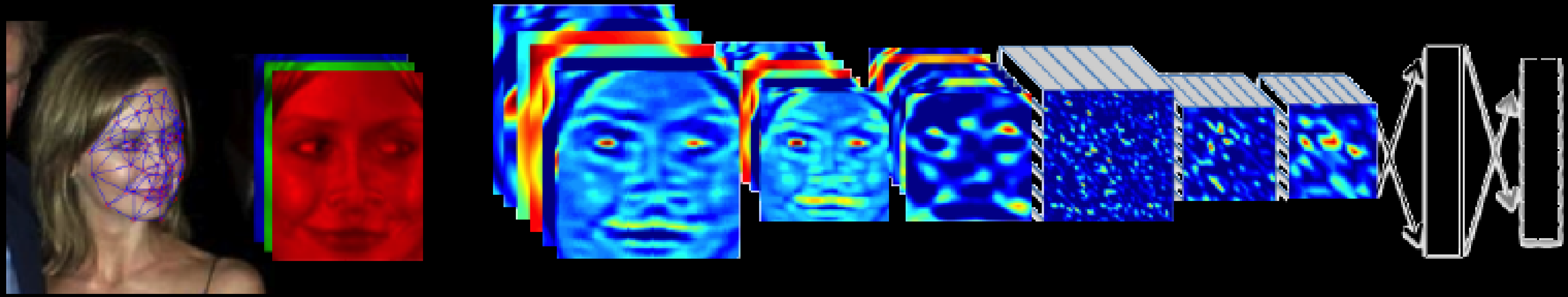
Biology and Health



Neuroscience



$$\mathrm{g}(x)$$



$\mathrm{f}_1(x)$ $\mathrm{f}_2(\cdot)$ $\mathrm{f}_3(\cdot)$ $\mathrm{f}_4(\cdot)$ $\mathrm{f}_5(\cdot)$ $\mathrm{f}_6(\cdot)$ $\mathrm{f}_7(\cdot)$ $\mathrm{f}_8(\cdot)$ $\mathrm{f}_9(\cdot)$

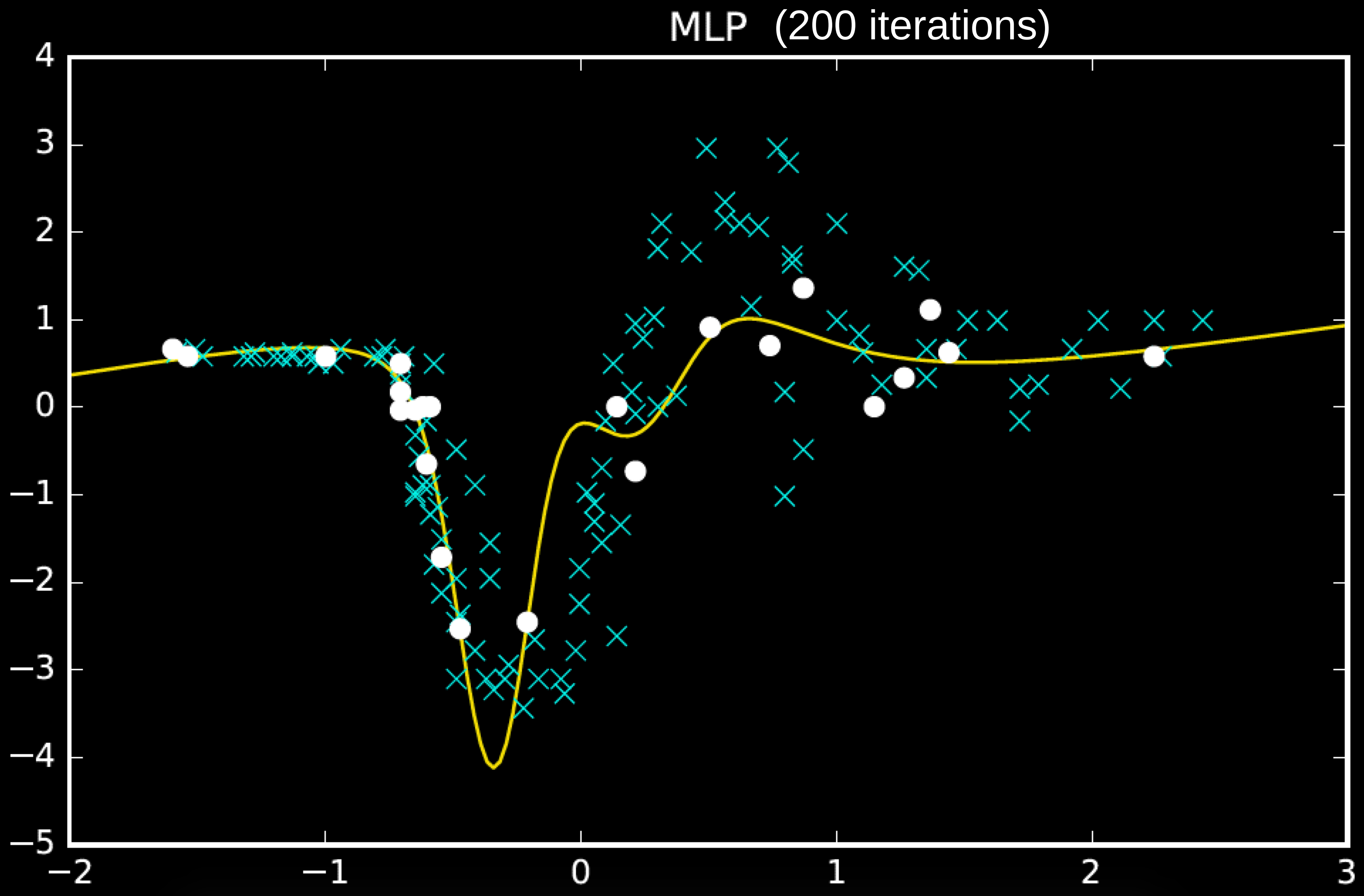
$$\mathrm{g}(x) = \mathrm{f}_9\left(\mathrm{f}_8\left(\mathrm{f}_7(\mathrm{f}_6(\cdots))\right)\right)$$

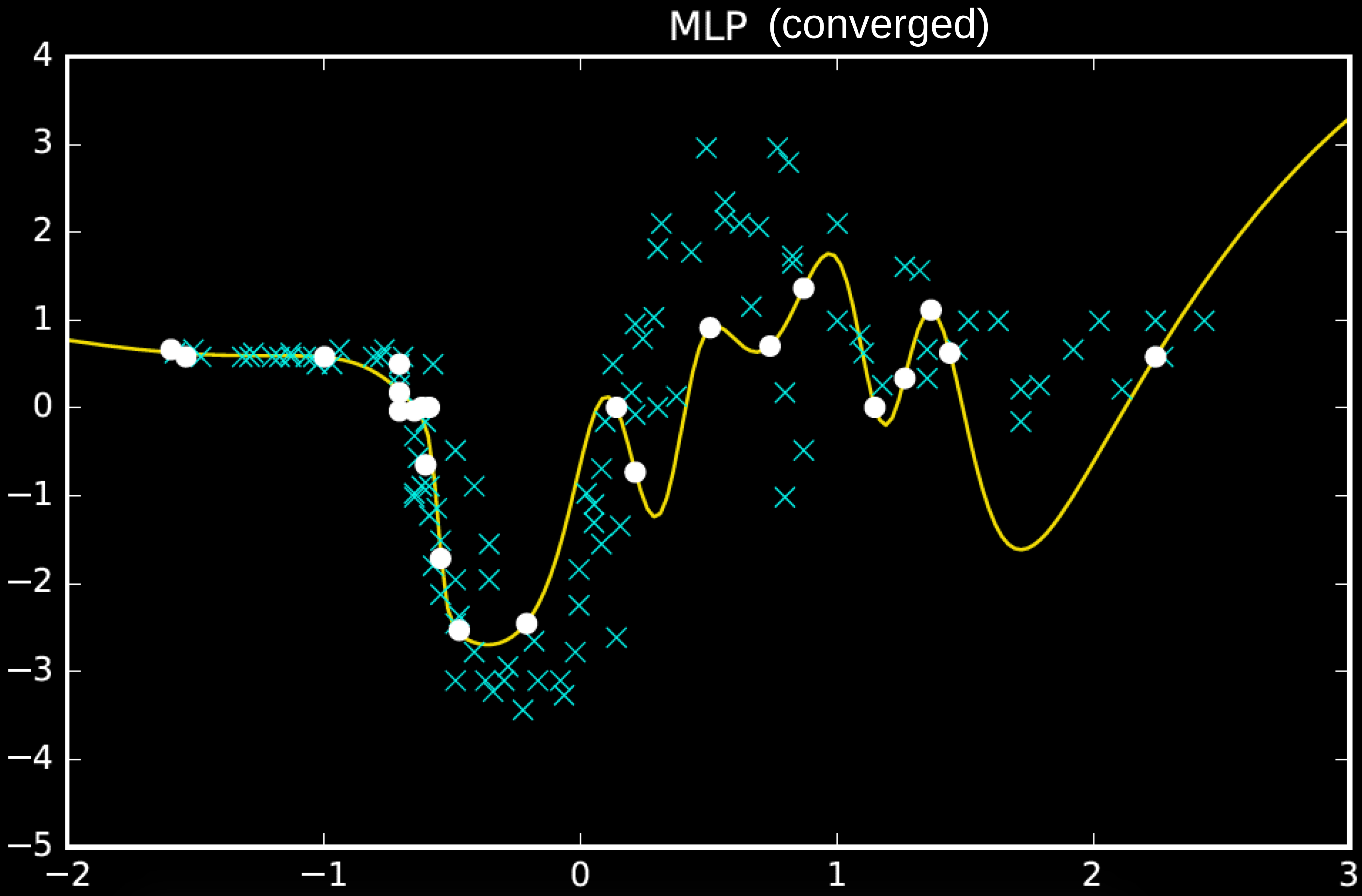
Stochastic Process Composition

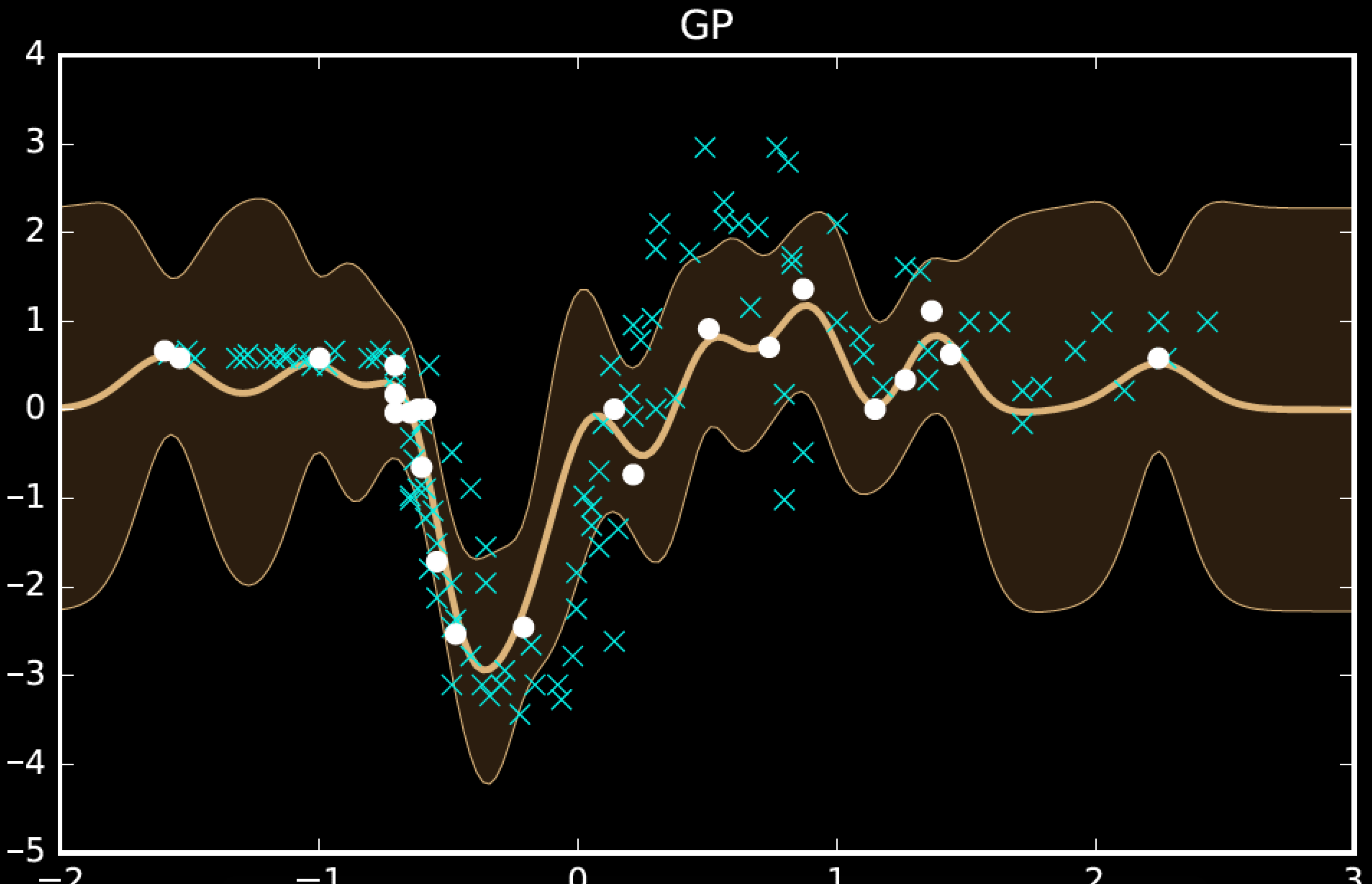
- A new approach to forming stochastic processes
- Mathematical composition:

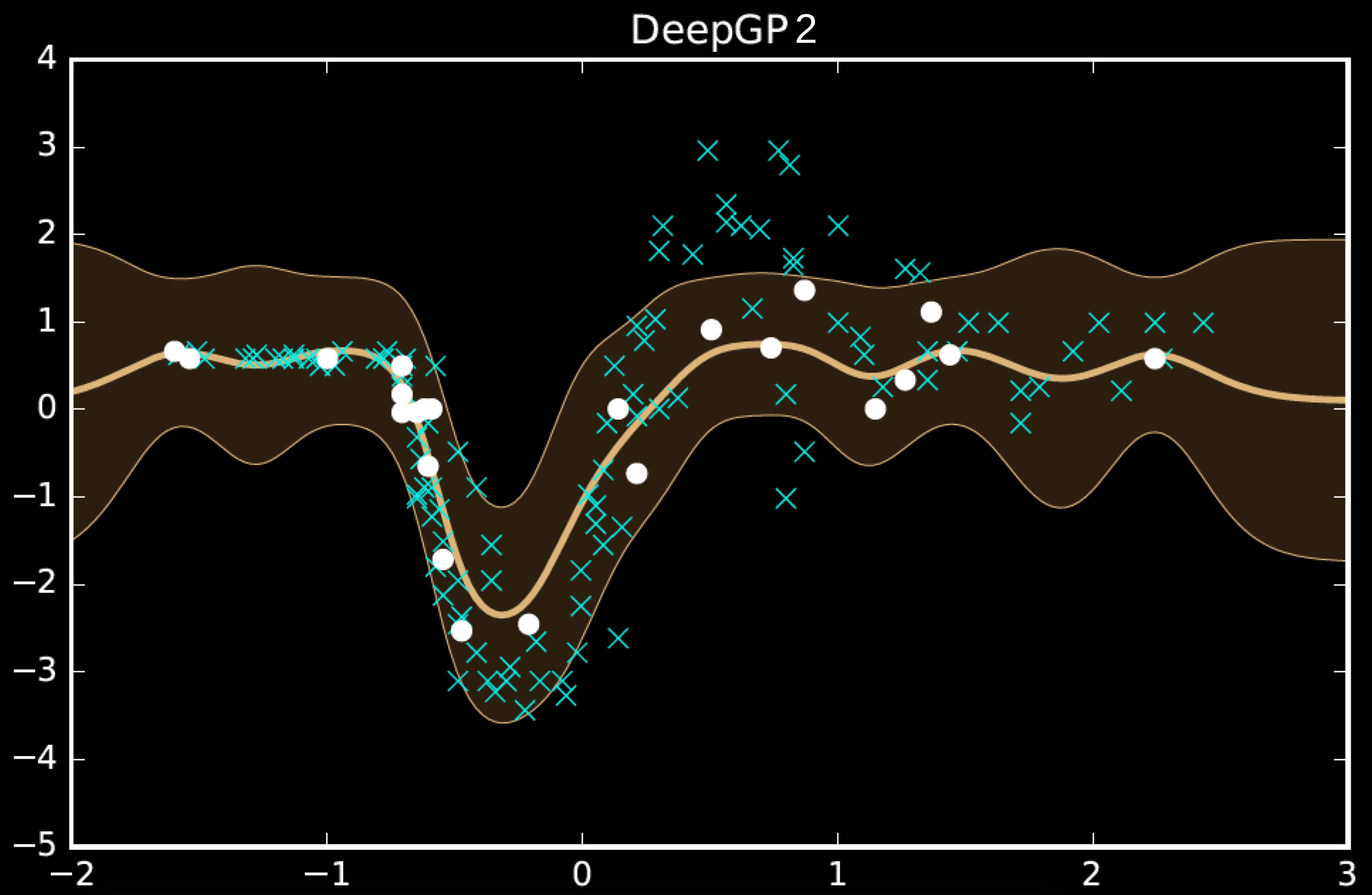
$$y(x) = f_1\left(f_2\left(f_3(x)\right)\right)$$

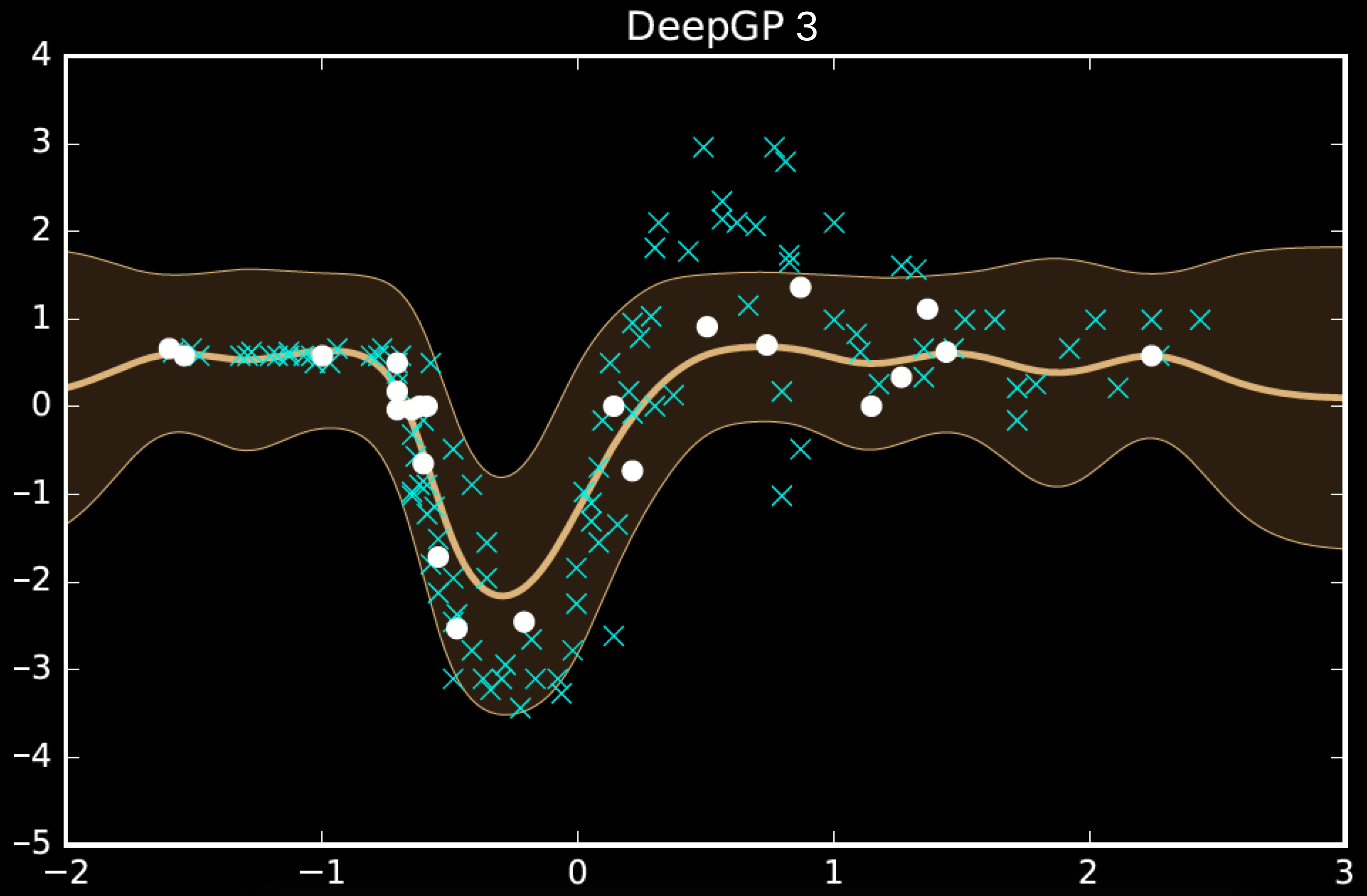
- Properties of resulting process highly non-Gaussian
- Allows for hierarchical structured form of model.
- Learning in models of this type has become known as: **deep learning**.


MLP (200 iterations)
4
3
2
1
0
−1
−2
−3
−4
−5
−2
−1
0
1
2
3


MLP (converged)
4
3
2
1
0
−1
−2
−3
−4
−5
−2
−1
0
1
2
3


GP
4
3
2
1
0
−1
−2
−3
−4
−5
−2
−1
0
1
2
3


DeepGP 2
4
3
2
1
0
−1
−2
−3
−4
−5
−2
−1
0
1
2
3



DeepGP 3
4
3
2
1
0
−1
−2
−3
−4
−5
−2
−1
0
1
2
3

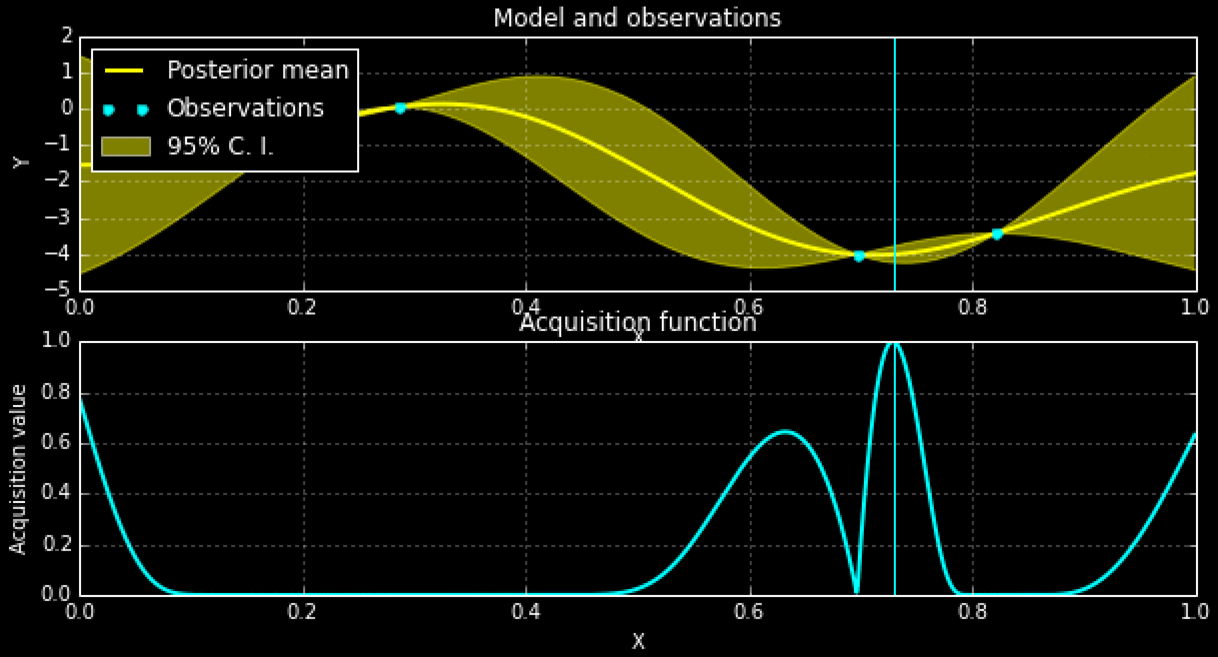
model	MSE (train)	MSE (test)
mlp (200 iters)	108.5	1185.1
mlp (converged)	24.0	1338.2
gp	59.2	1095.4
deep gp (2)	146.2	833.7
deep gp (3)	182.5	843.6

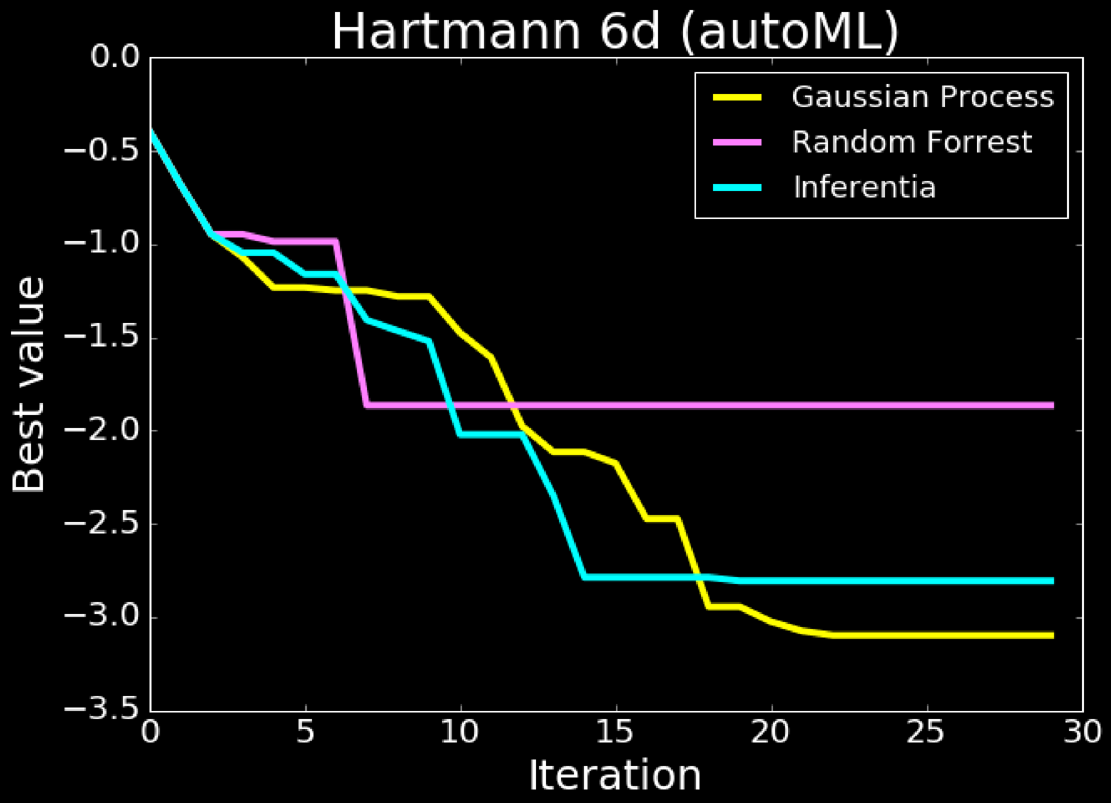
One hundred hidden nodes, one hundred inducing points

Regression

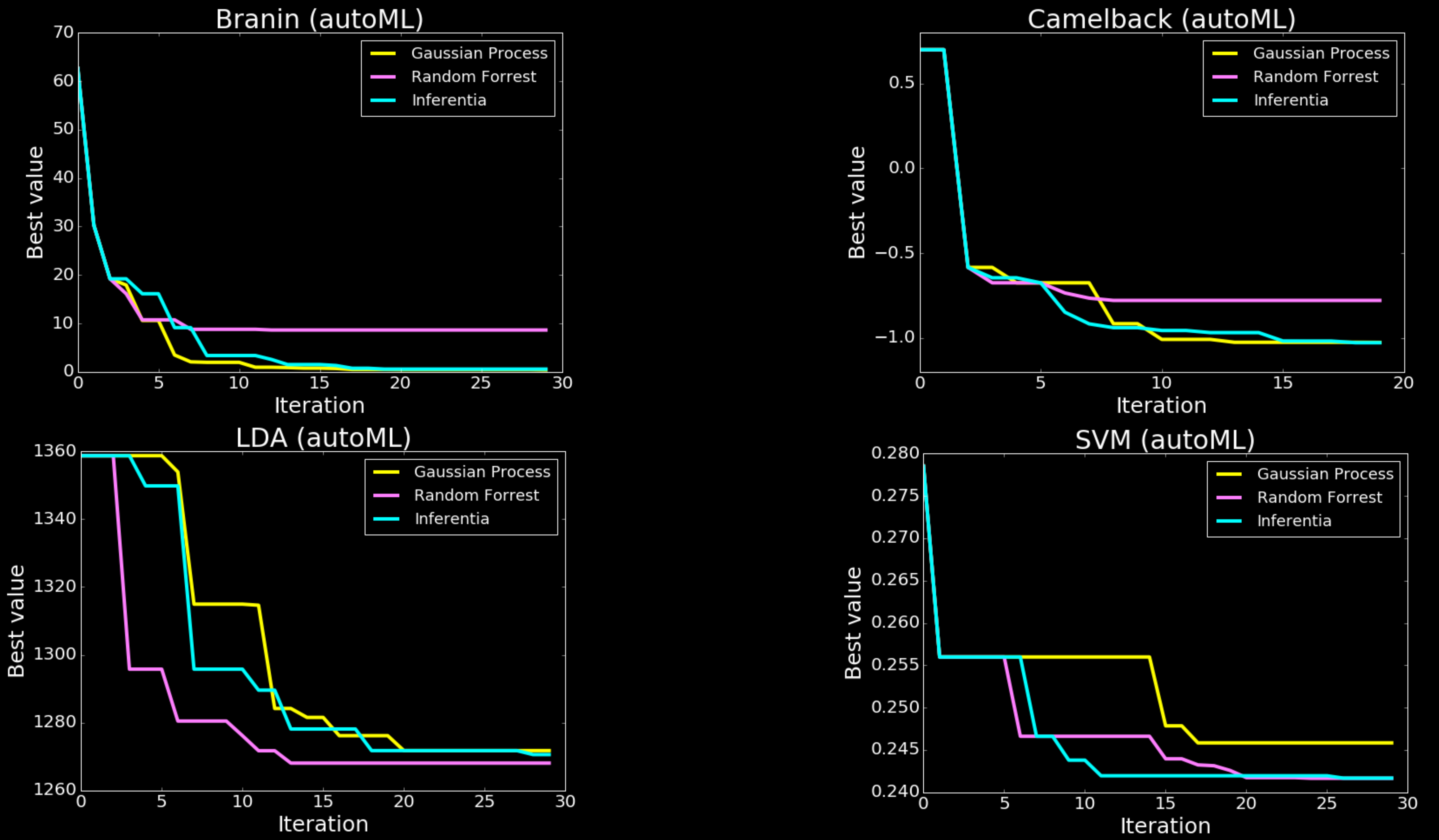
data set	n	p	GP	Sparse GP	Deep GP
housing	506	13	2.78±0.54	2.77±0.60	2.69±0.49
redwine	588	11	0.72±0.06	0.62±0.04	0.62±0.04
energy1	768	8	0.48±0.07	0.50±0.07	0.49±0.07
energy2	768	8	0.59±0.08	1.66±0.21	1.39±0.49
concrete	1030	8	5.26±0.67	5.81±0.62	5.66±0.62

Bayesian Optimization



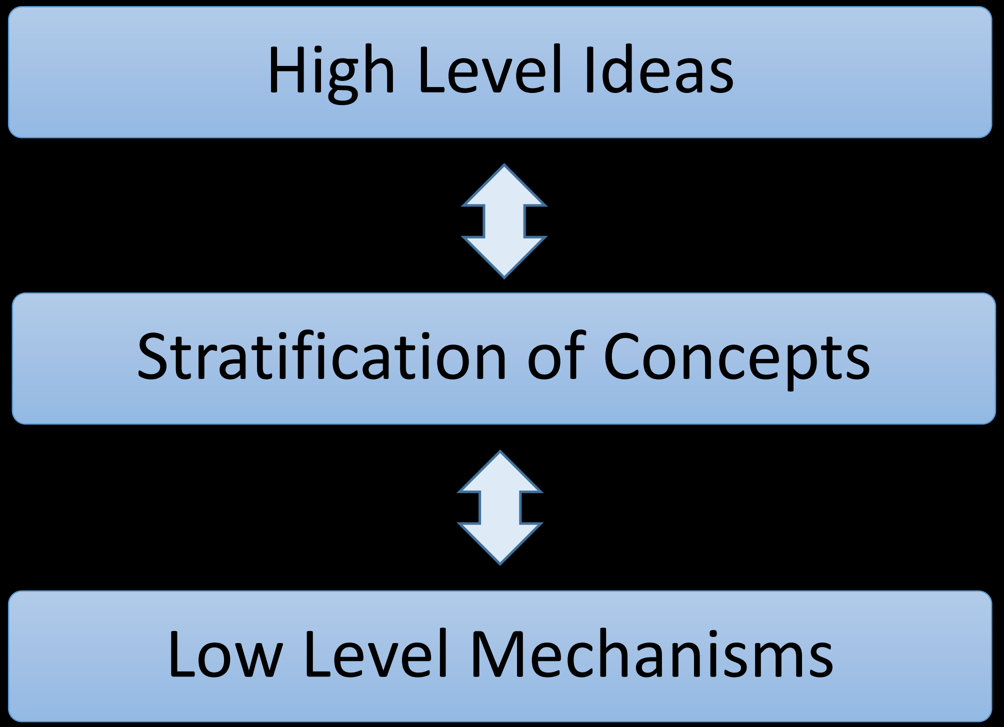


- Check http://sheffieldml.github.io/GPyOpt/

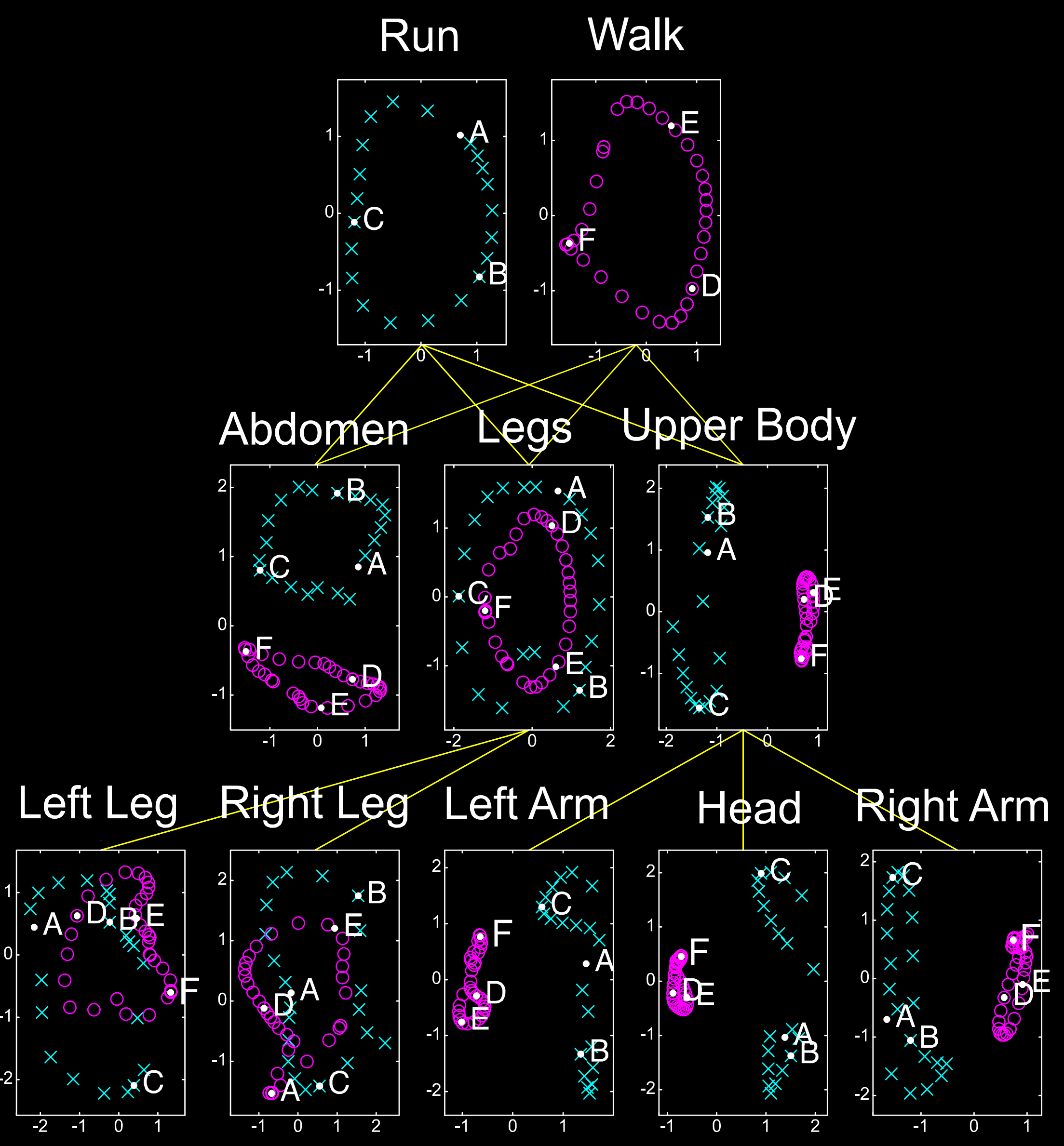


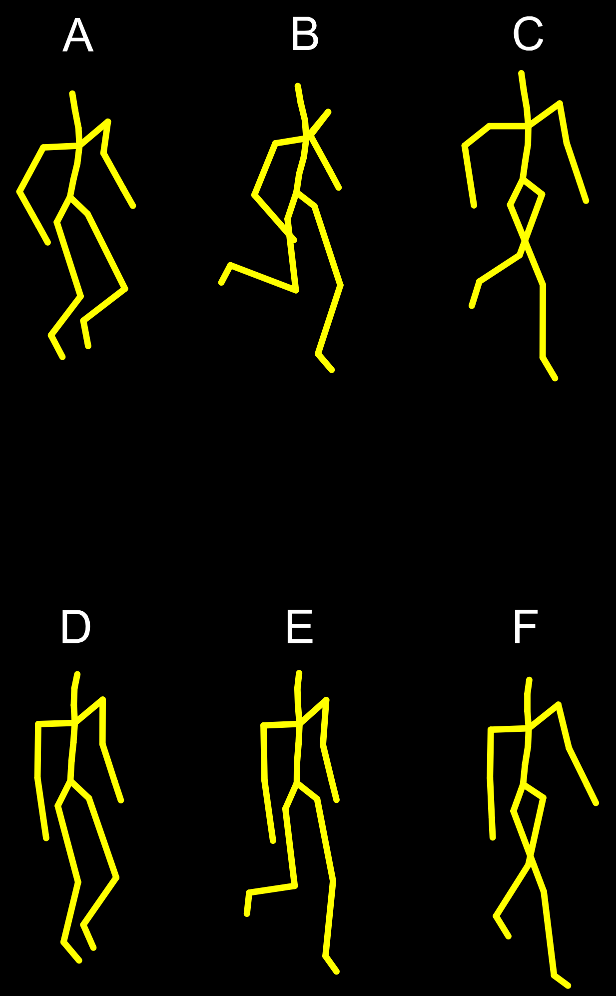
Branin (autoML)
Camelback (autoML)
LDA (autoML)
SVM (autoML)
Gaussian Process
Random Forrest
Inferentia
Best value
Iteration

Use Abstraction for Complex Systems



Example: Motion Capture Modelling

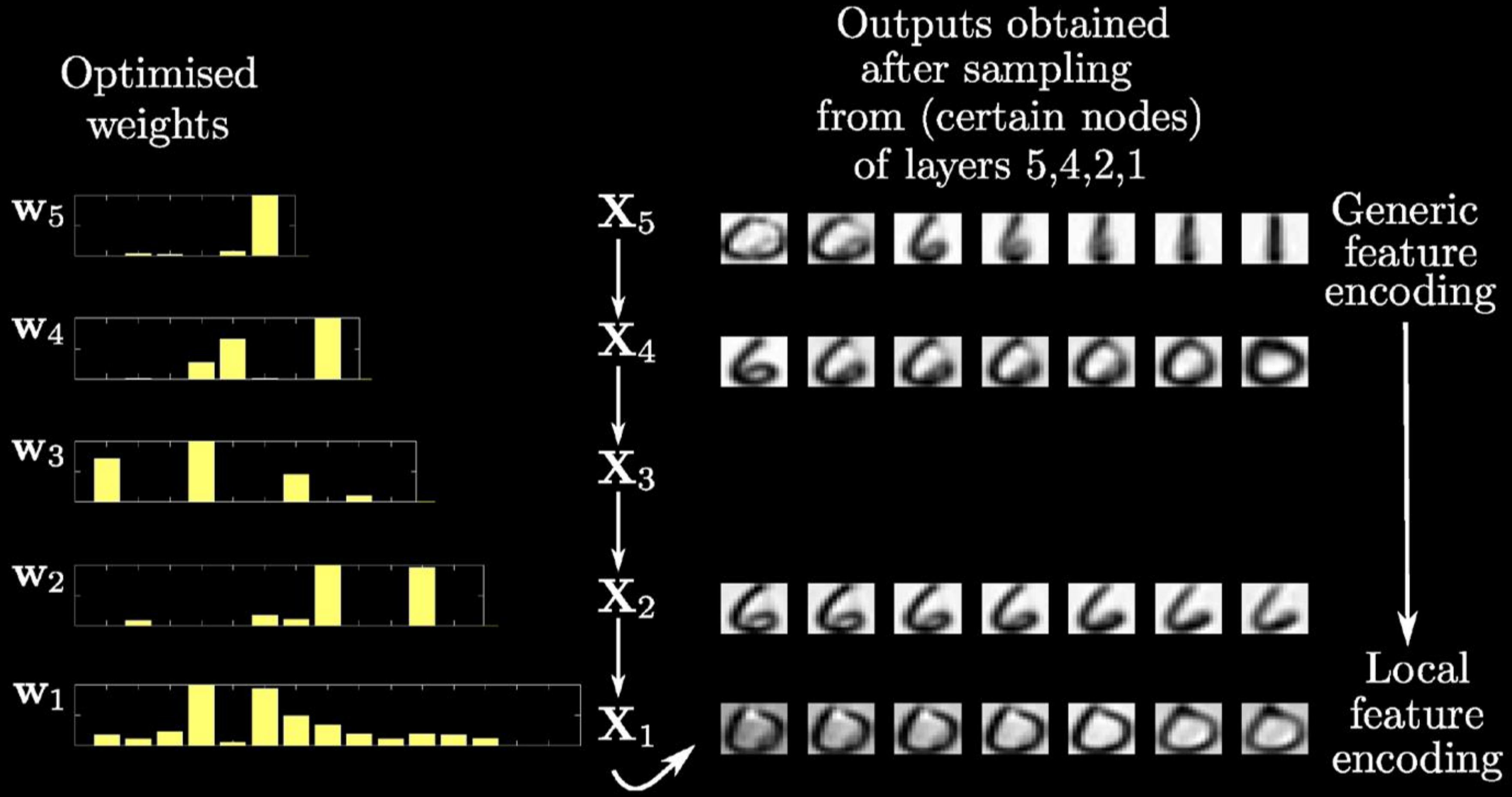




MATLAB Demo

- demo_2016_04_28_amazon.m

Modelling Digits



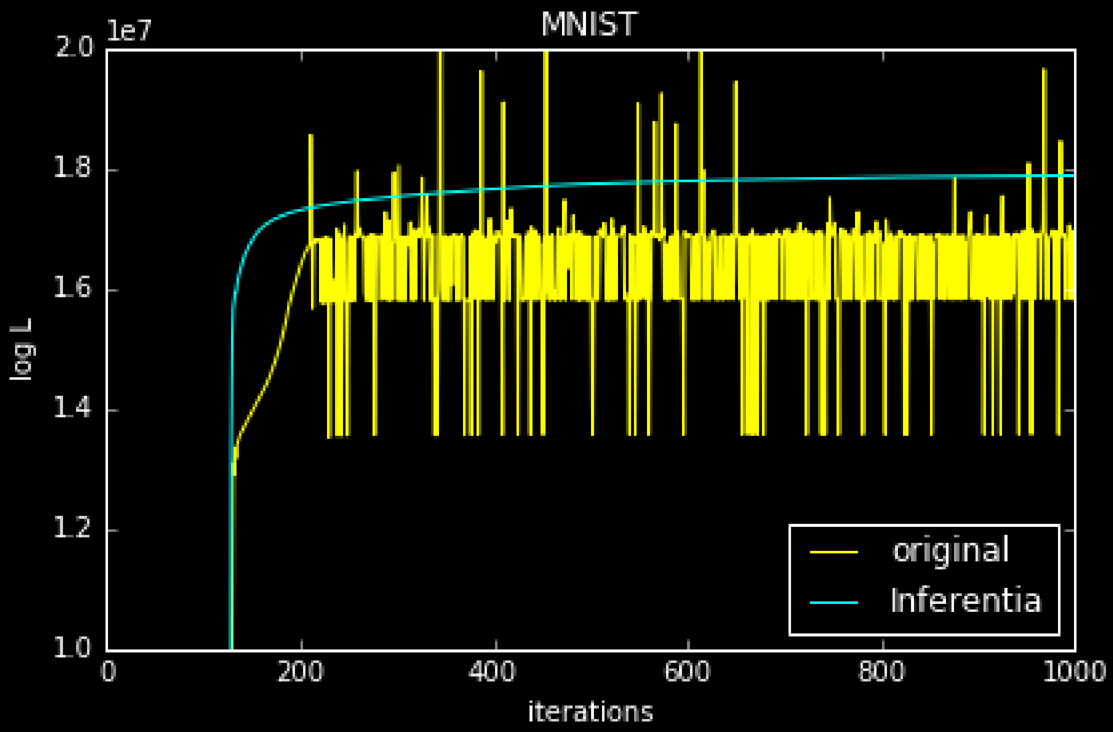
MATLAB Demo

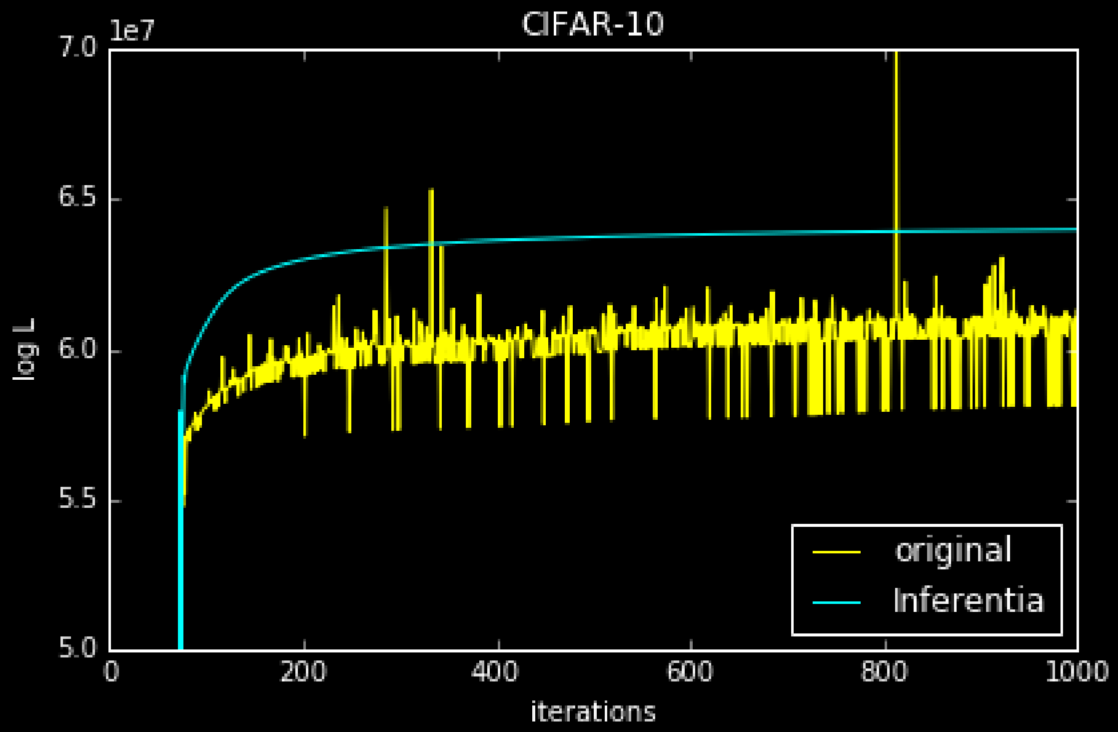
- demo_2016_04_28_amazon.m

Inferentia

Challenging Uncertainty

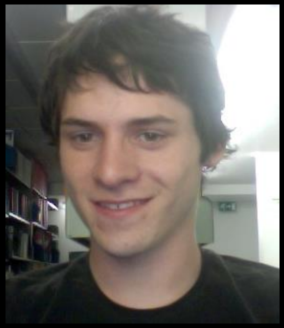
Numerical Issues



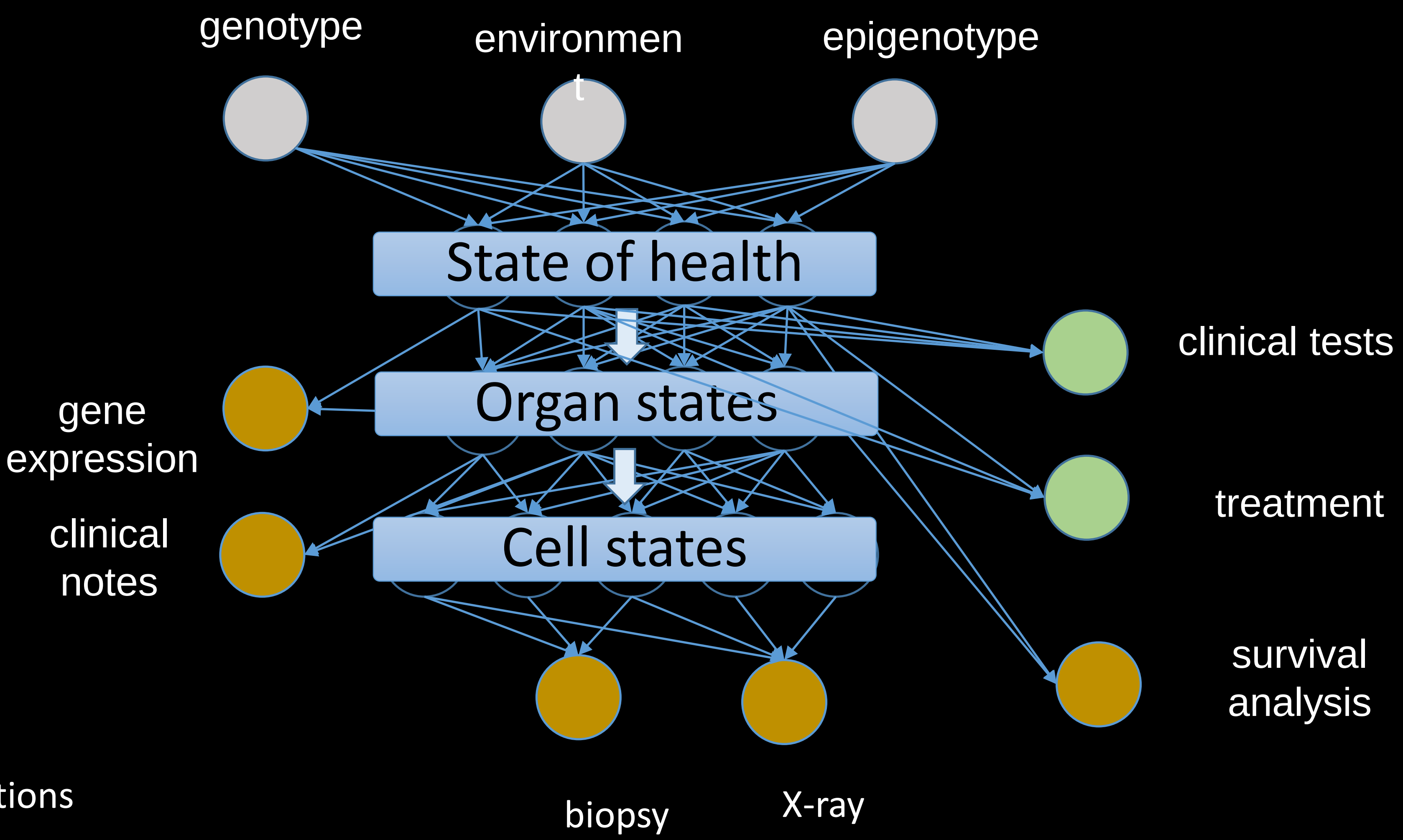


Health





- Complex system
- Scarce data
- Different modalities
- Poor understanding of mechanism
- Large scale



PLoS Comp Bio, Nature Communications

To Find Out More

- Gaussian Process Summer School
  - 12th-15th September 2016 in Sheffield
  - This year in parallel with/themed as a UQ orientated school (co-organisation with Rich Wilkinson)
  - Occurring alongside ENBIS Meeting
  - http://gpss.cc/

Future

- Methodology
  - Deep GPs (also current)
  - Latent Force Models (current but dormant)
  - Latent Action Models and Stochastic Optimal Control (new)
  - Probabilistic Geometries (starting)

- Exemplar Applications
  - Health and Biology (existing)
  - Developing world (existing)
  - Robotics at different scales (starting)
  - Perception: vision (dormant) haptic (new)

Summary

- Complex systems:
  - ‘big data’ is too ‘small’.
    - The data are not enough.

  - Need data efficient methods
    - http://www.theguardian.com/media-network/2016/jan/28/google-ai-go-grandmaster-real-winner-deepmind

- Solutions:
  - Hybrid mechanistic-empirical models

  - Structured models for automated data assimilation

Thank you

Neil Lawrence
http://inverseprobability.com
@lawrennd

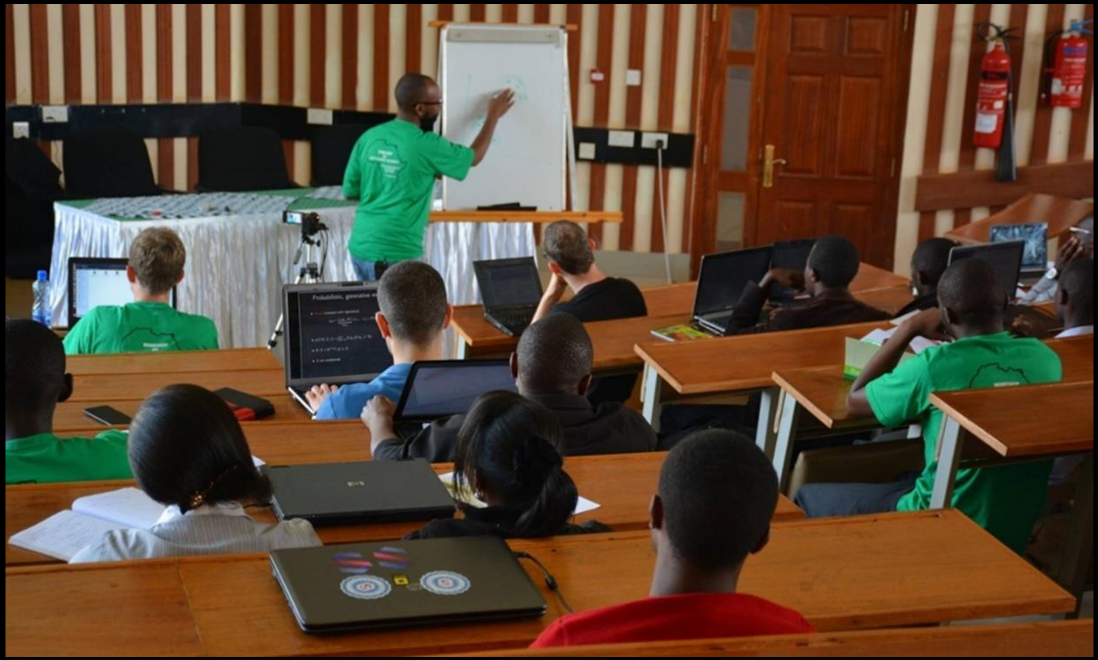
The Digital Oligarchy

- Response to concentration of power with data
- CitizenMe
  - London based start up
  - User-centric data modelling
- New challenges in ML
  - Integration of ML, systems, cryptography.

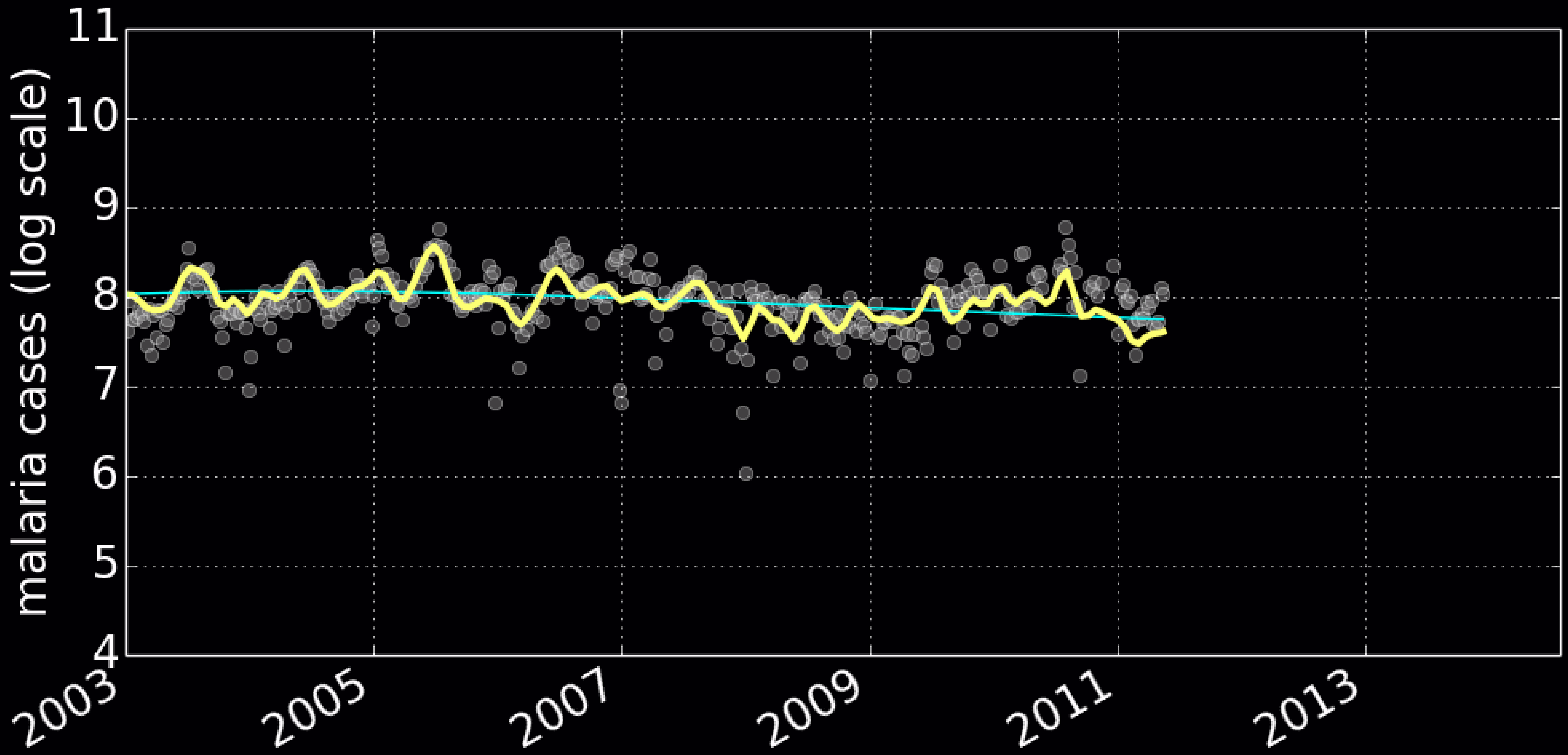
Open Data Science and Africa

Challenge

- “Whole pipeline challenge”
- Make software available
- Teach summer schools
- Support local meetings
    - Publicity in the Guardian
- Opportunities to deploy pipeline solution

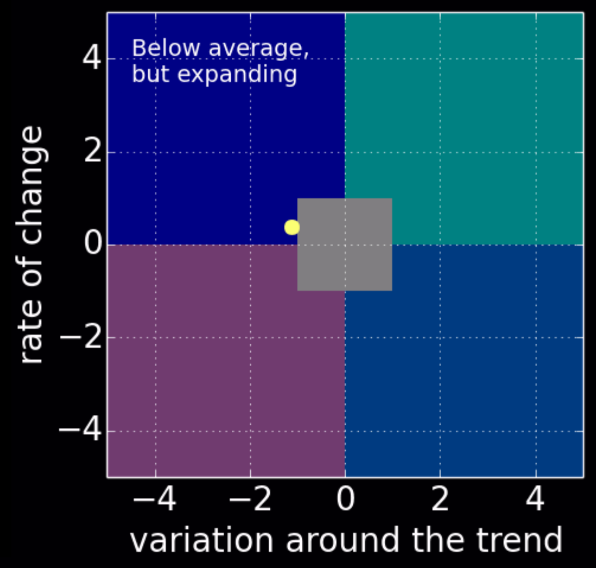


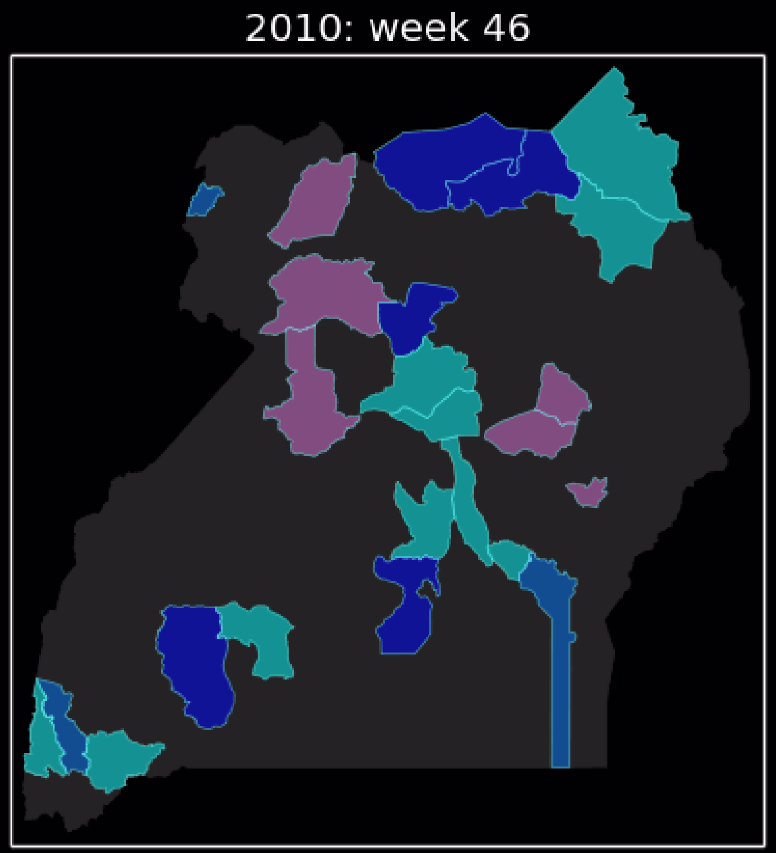
Disease Incidence for Malaria

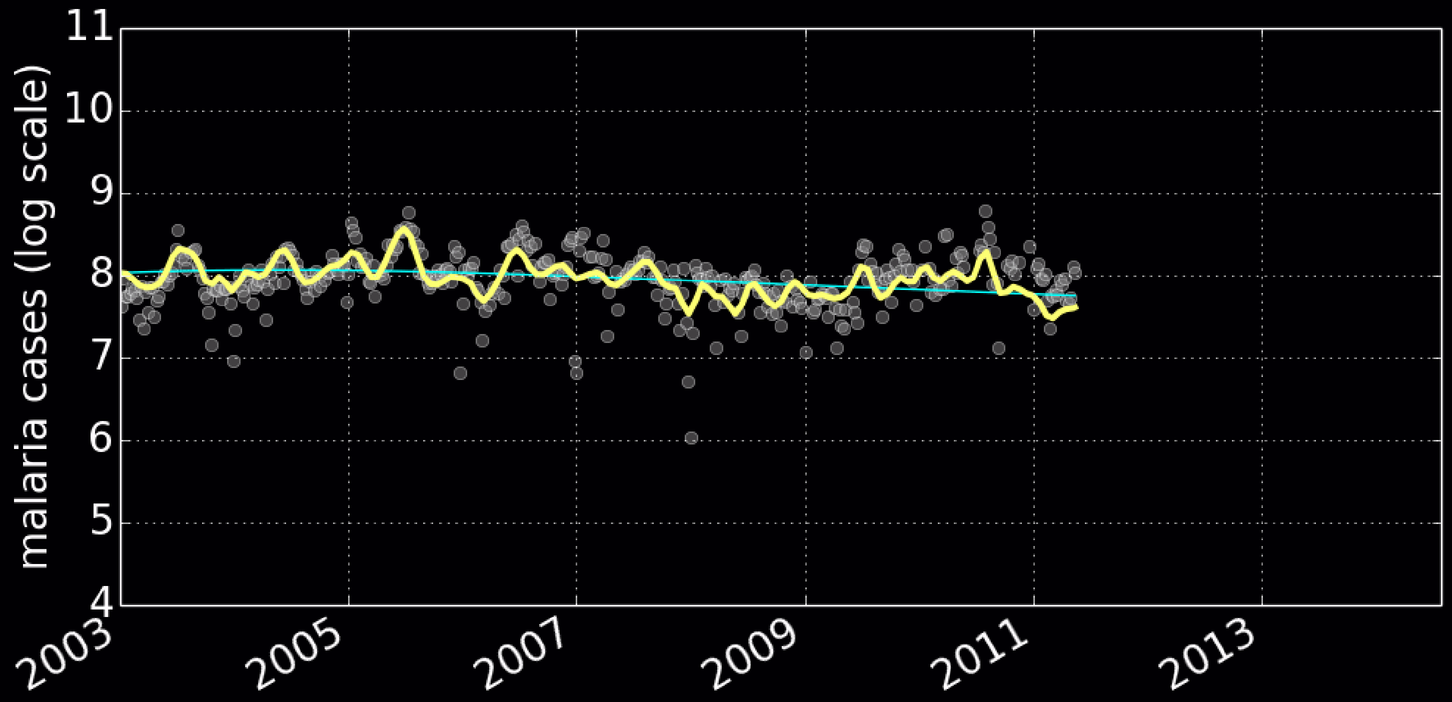


Uganda

- Spatial models of disease







Deployed with UN Global Pulse Lab

http://pulselabkampala.ug/hmis/

GLOBAL PULSE