FINDING THE (UN)KNOWN UNKNOWNS

Dr. Rob Smith

Dushek Group, Dunn School of Pathology

robert.smith@path.ox.ac.uk

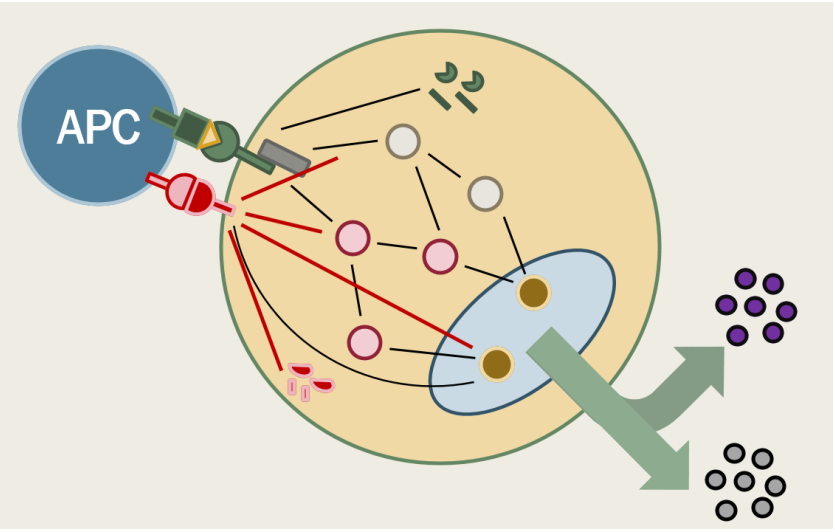
https://www.youtube.com/watch?v=GiPe1OiKQuk

What do you want to know??

How do plants at the bottom of the canopy tolerate low levels of sunlight?

What is the genetic network that controls segmentation in Drosophila larva?

How does signalling from costimulatory integrate with signalling from the TCR in T cells?



Modelling Recipe

1. What do you want to know?

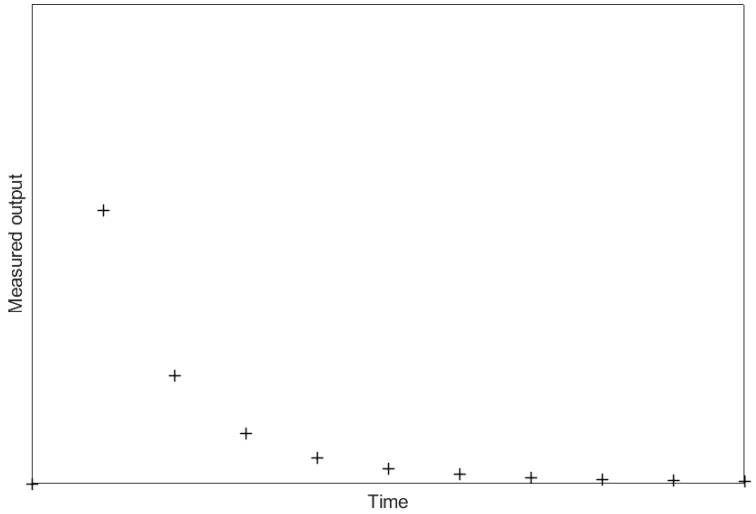
2. What data do you have available to you?

3. What modelling technique is appropriate for the challenge ahead?
   - *This is greatly constrained by your available data…*

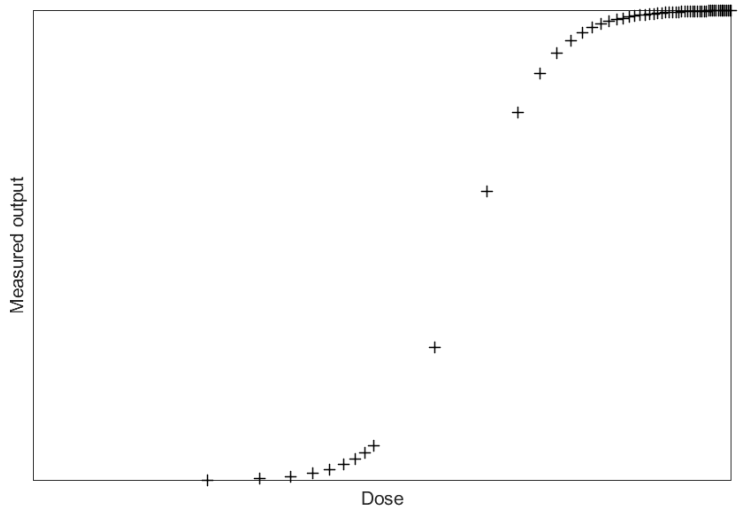
4. How do you estimate unknowns?

5. Has this all improved my understanding? Can I predict the future?

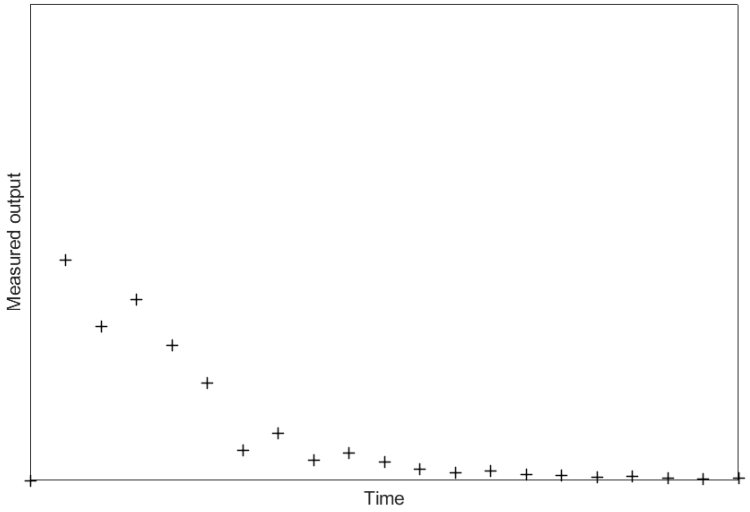
Time-series data of component concentrations



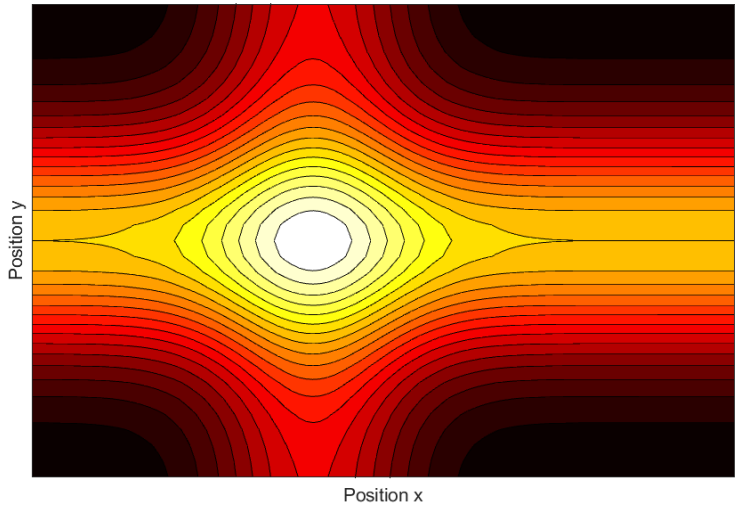
Single time-point measurements (e.g. dose responses)



"Noisy" time-series data of single cells/molecules

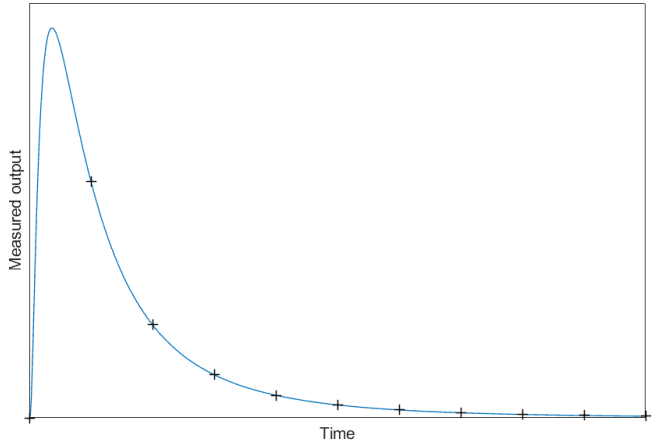


Data across time and space



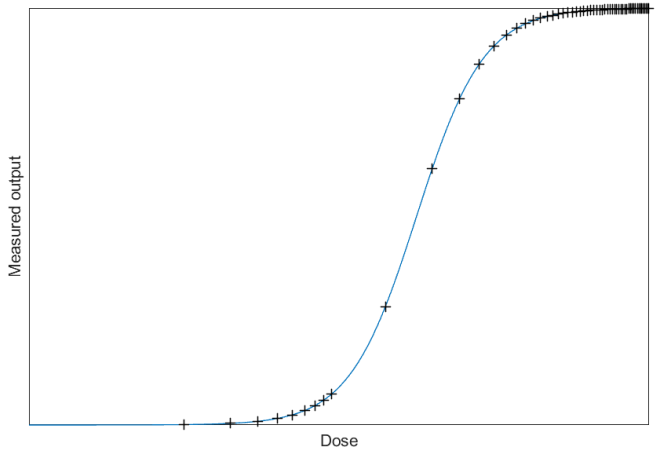
Ordinary differential equation

- Assumes large molecule numbers, volume well mixed



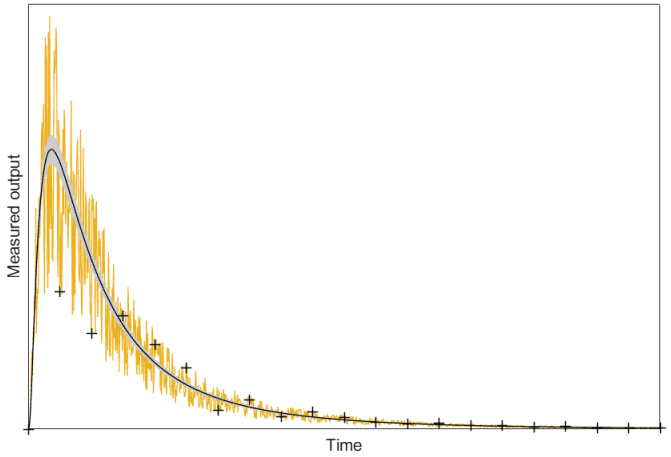
Steady-state equations / spline fits

- ODE = 0
- Spline fits quantify data properties



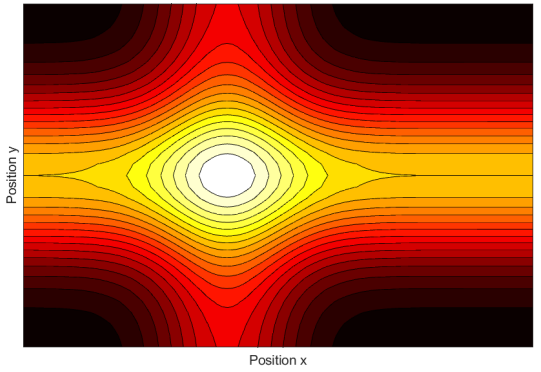
Stochastic differential equations

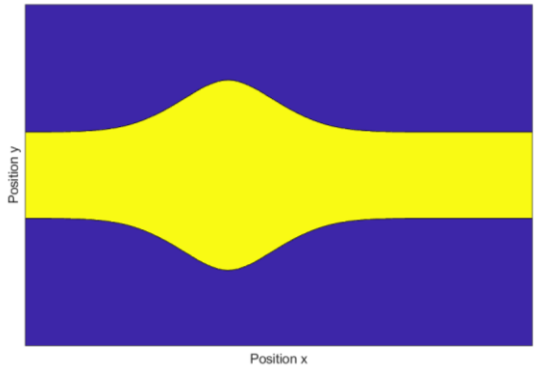
- Most accurate description of reactions
- Computationally infeasible for large molecule numbers



Partial differential equations

- Components change over time and spatial dimensions





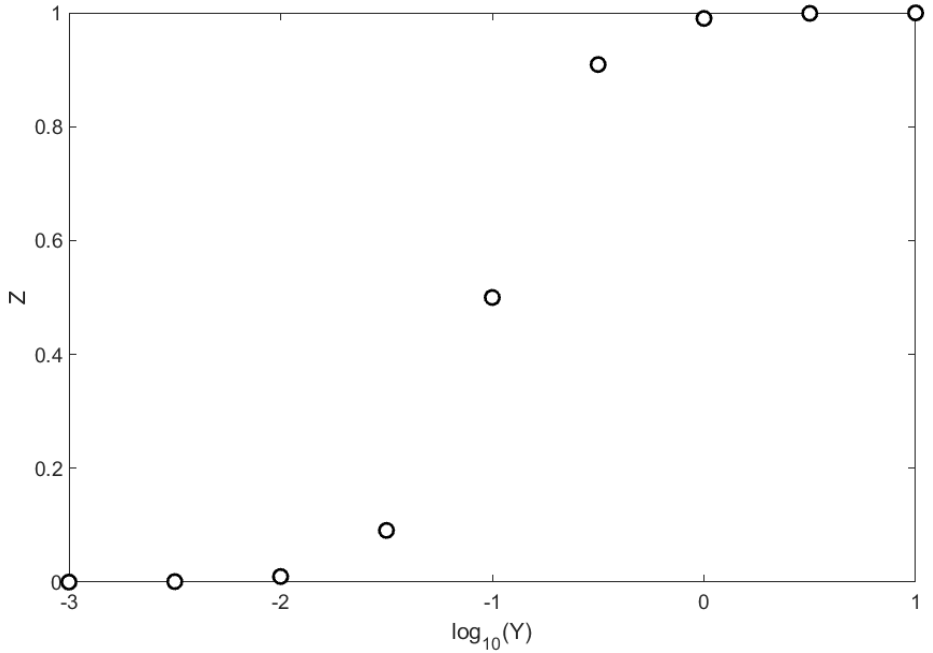
An example using chemical kinetics

$$\mathrm{X} + 2\,\mathrm{Y} \xrightarrow{\mathrm{k}_1} \mathrm{Z}$$

$$\mathrm{Z} \xrightarrow{\mathrm{k}_2} \mathrm{X} + 2\,\mathrm{Y}$$

The story…

- Biological system where one X molecule and 2Y molecules form a complex Z that controls downstream signalling.
- Let's assume that the only information we have is a dose-response curve (of Z or some output that correlates with Z) measured after time t_s that we hope is when the system is in equilibrium.
- How should we model this?



An example using chemical kinetics

$$\mathrm{X} + 2\,\mathrm{Y} \xrightarrow{\mathrm{k_1}} \mathrm{Z}$$

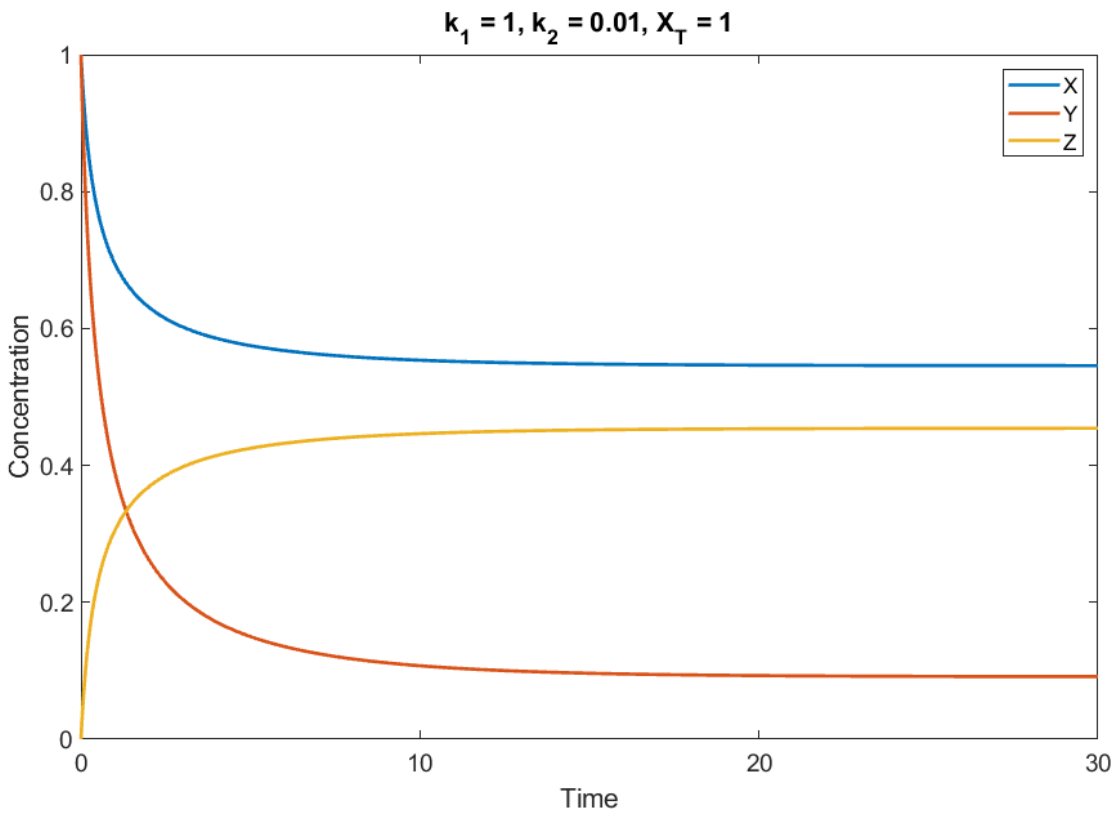
$$\mathrm{Z} \xrightarrow{\mathrm{k_2}} \mathrm{X} + 2\,\mathrm{Y}$$

Warning: always ensure dimensions (e.g. secs, mol) are consistent throughout

$$\frac{dX}{dt} = k_2 Z - k_1 X Y^2$$

$$-\frac{dX}{dt} = -\frac{1}{2}\frac{dY}{dt} = \frac{dZ}{dt}$$

$$X_T = X + Z$$



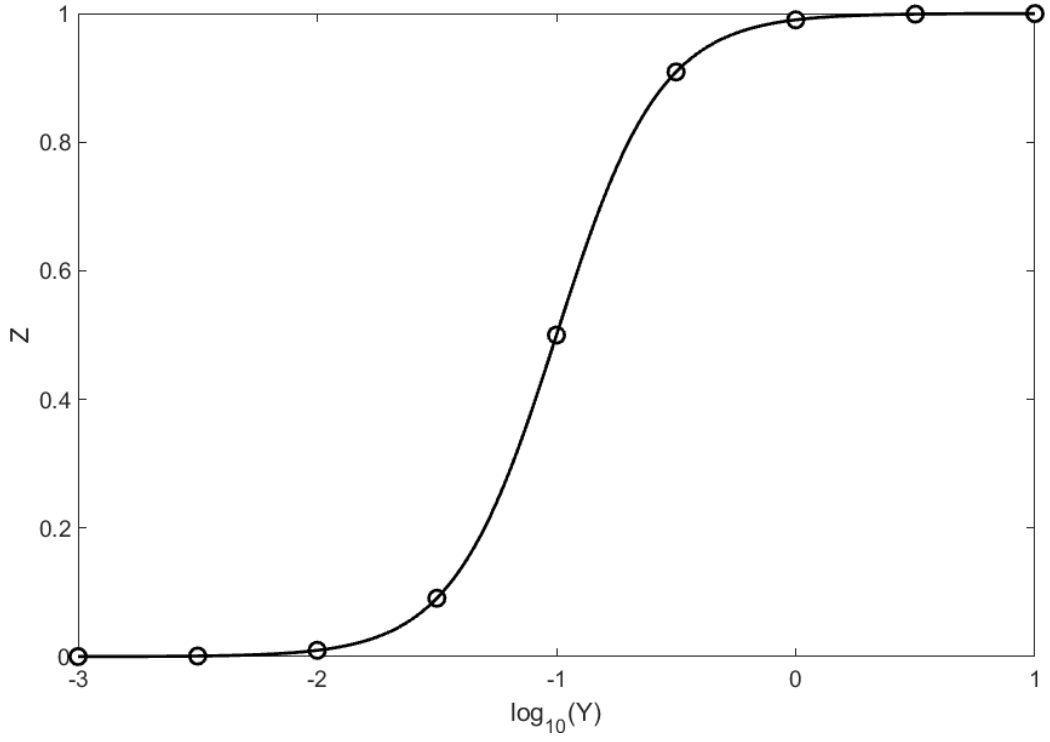
An example using chemical kinetics

$$\mathrm{X} + 2\,\mathrm{Y} \xrightarrow{\mathrm{k_1}} \mathrm{Z}$$

$$\mathrm{Z} \xrightarrow{\mathrm{k_2}} \mathrm{X} + 2\,\mathrm{Y}$$

$$X_T = X + Z$$

$$k_1 X Y^2 = k_2 Z$$



How do we estimate the rates?

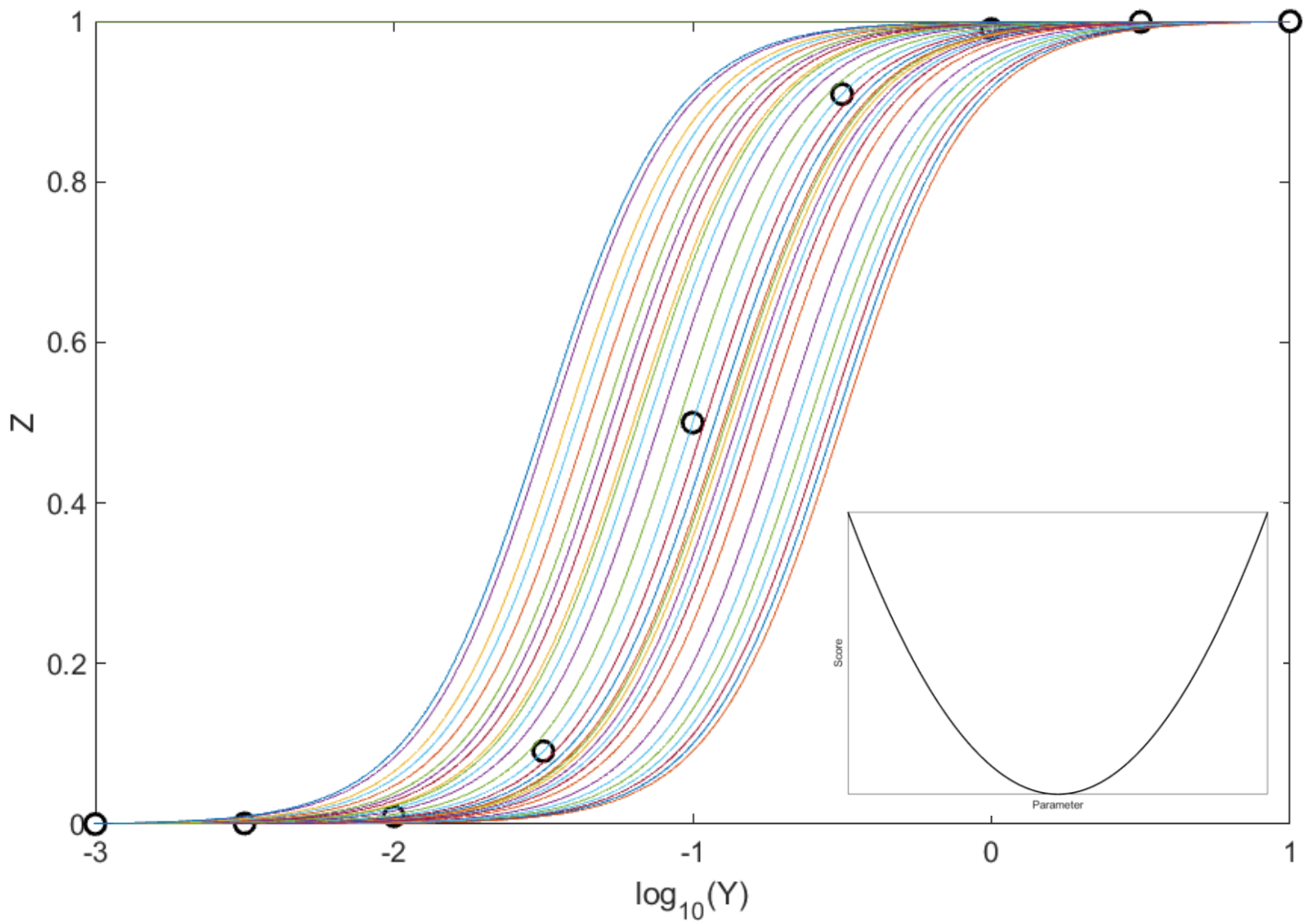
(Parameter Optimisation)

You need:

- Data
- A model
- A way to compare the model and the data (a scoring function)
- (an idea of the magnitude of each rate)

You get:

- The set of rates for your model such that simulations best match the data



What does your optimal model tell you?

- Scenario #1: your optimal model matches all your available data
  - Your model accurately captures the underlying biology
  - Simulate new/different experiments to predict what should happen in the lab (and go through the whole process again)
- Scenario #2: your optimal model kind of matches the data
  - Try re-running your optimisation algorithm from different initial guess
- Scenario #3: your model is unable to match the data for all sets of reaction rates
  - Your model is missing a necessary component or aspect to its structure, try and work out what it is through simulations

The take-away

- All modelling studies start with a question – what phenomena can you not intuitively explain?
- The purpose of a model is to test that your understanding of biology is correct and suggest why it is incorrect.
- The choice of modelling technique you use is constrained by the data you have available to you.
- Whilst models are inherently "wrong" (they are not a 100% accurate description of biology), if the model matches your data they are powerful tools to predict future experiments and test different scenarios before entering the lab.

Other useful sources

- Dr. Omer Dushek, Dunn School of Pathology
  - omer.dushek@path.ox.ac.uk
- Prof. Ruth Baker, Mathematical Institute
  - ruth.baker@maths.ox.ac.uk

- Further reading:
  - Mathematical Biology I (An Introduction) & II (Spatial models and biochemical applications) by JD Murray
  - 'Derivation and use of mathematical models in Systems Biology' by Smith & Fleck (2017) in the book "Pollen Tip Growth"
  - Any up-to-date guide book for MATLAB
    - many MATLAB functions are also available in Python under different names