Introduction to Mathematical Models of Infectious Disease in Livestock

Lecture 1: Introduction to mathematical modelling

Andrea Doeschl-Wilson

Group leader, Genetics & Genomics, The Roslin Institute and Royal Dick School of Veterinary Studies, University of Edinburgh, UK

Andrea.Wilson@roslin.ed.ac.uk

Purpose of this lecture

- Get some understanding what mathematical models are & what they can / cannot do
- Get acquainted with different types of mathematical models
- Learn the basic principles for building, analysing, testing and using mathematical models

What is a mathematical model?

Model (Definition):

- A representation of a system that allows for **investigation** of the properties of the system and, in some cases, **prediction** of future outcomes.
- Always requires **simplification**

Mathematical model:

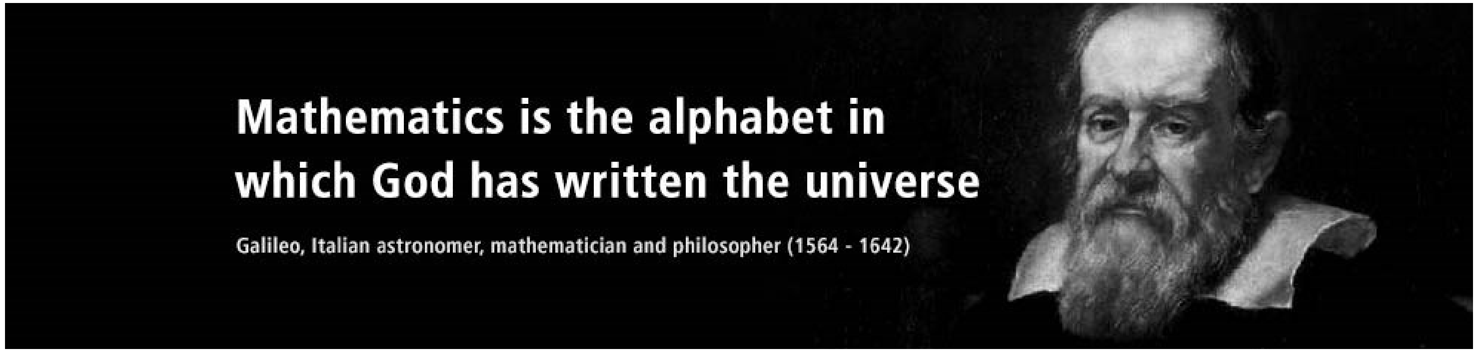
- Uses mathematical equations to describe a system

$$
\begin{aligned}
\mathrm{var}\left(\sum_{m=1}^{p} f_m\right) &= \mathrm{E}_p\left(\mathrm{var}\left(\sum_{m=1}^{p} f_m \mid p\right)\right) + \mathrm{var}_p\left(\mathrm{E}\left(\sum_{m=1}^{p} f_m \mid p\right)\right) \\
&= \mathrm{E}(p)\mathrm{var}(f) + \mathrm{E}^2(f)\mathrm{var}(p) \\
&= \bar{p}\sigma_f^2 + \bar{f}^2\sigma_p^2.
\end{aligned}
$$

Why do we need models?

- Models provide a framework for conceptualizing our ideas about the behaviour of a particular system
- Models allow us to find structure in complex systems and to investigate how different factors interact
- Models can play an important role in informing policies:
  - By providing understanding of underlying causes for a complex phenomenon
  - By predicting the future
  - By predicting the impact of interventions

Why mathematics?



- Mathematics is a precise language
  - Forces us to formulate concrete ideas and assumptions in an unambiguous way
- Mathematics is a concise language
  - One equation says more than 1000 words
- Mathematics is a universal language
  - Same mathematical techniques can be applied over a range of scales
- Mathematics is an old but still trendy language
  - The rich toolbox created by mathematicians over centuries is available at our disposal
- Mathematics is the language that computers understand best

Mathematical models synthesize results from many experiments

- Experimental studies concentrate on specific aspects of a system

- Fragmented understanding of the system

Mathematical models synthesize results from many experiments

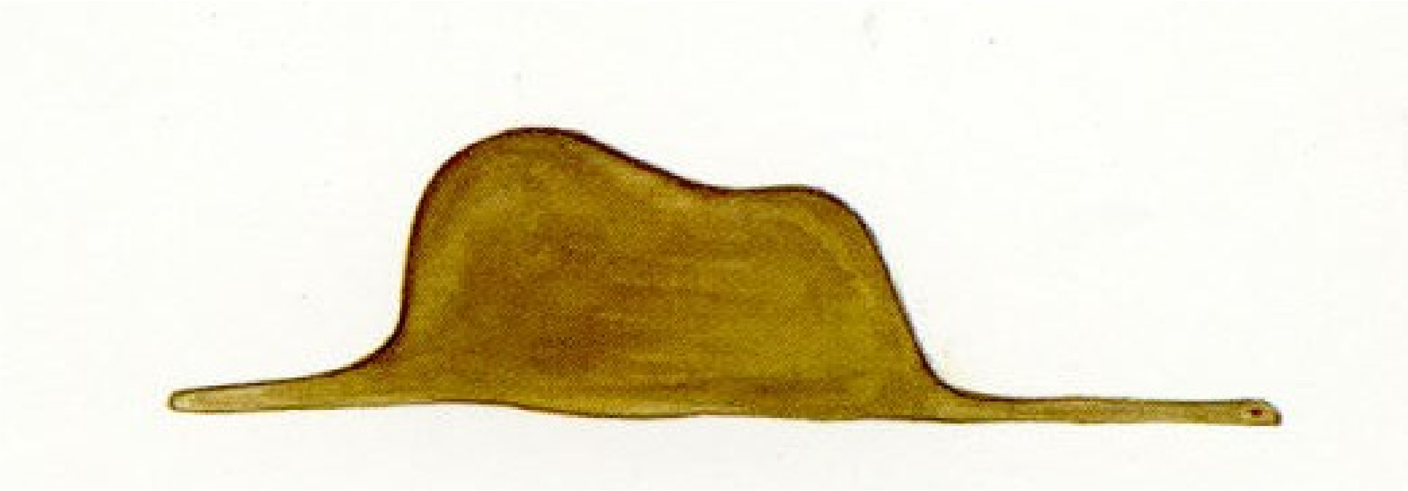
- Experimental studies concentrate on specific aspects of a system

- Fragmented understanding of the system

- Often hard to infer how the system functions as a whole

Mathematical models can unravel the unobservable

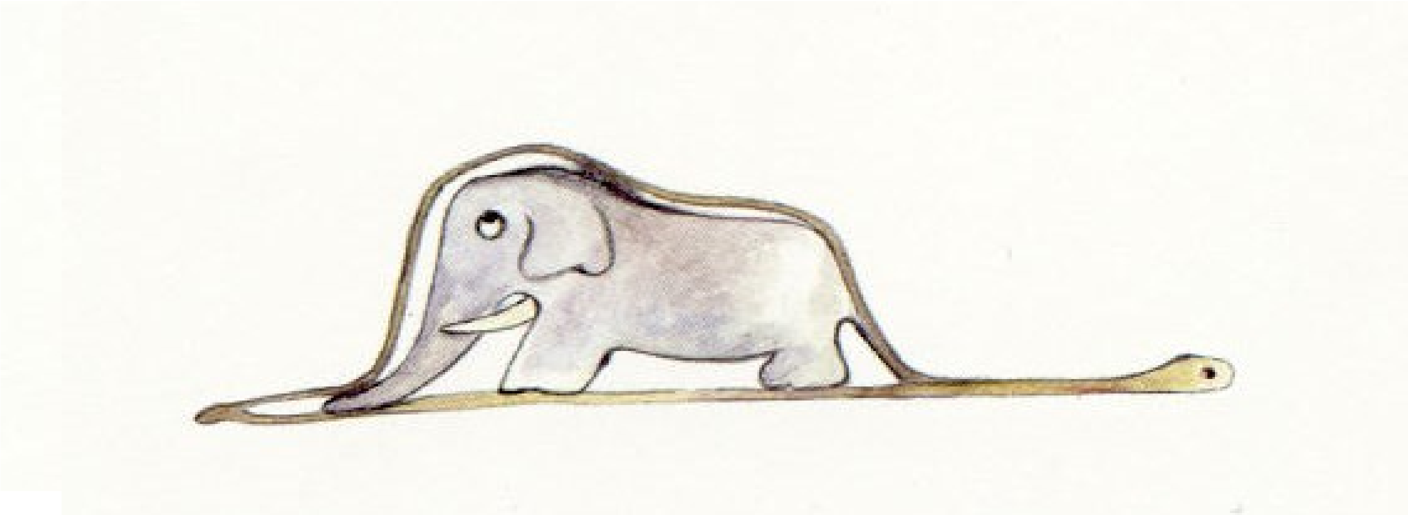
Often the traits that we can measure are not the most informative traits



From: The little Prince

Mathematical models can unravel the unobservable

Often the traits that we can measure are not the most informative traits



From: The little Prince

Mathematical models are not bound by physical constraints

- Powerful tool to explore ‘what if scenarios’
- Extremely useful in the context of infectious disease where experimental constraints are strong

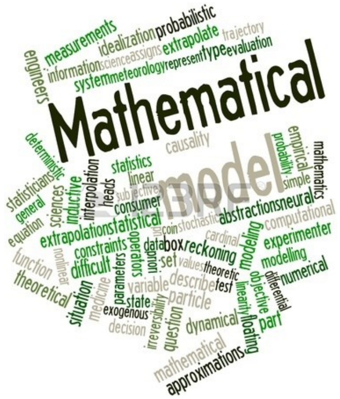
What do we use mathematical models for?

- Combine fragmented information into a comprehensive framework (e.g. combine results from in-vitro and in-vivo experiments)
- Determine the relationship between underlying biological traits and observable traits
- Test hypotheses that are difficult to test in empirical studies
- Make predictions & generate new hypotheses for future testing
- Assist with decision making by exploring ‘what if’ scenarios

Limitations of mathematical models

1. Lack of quantifiable knowledge
   - Models that encompass mechanisms (e.g. infection process) require quantitative understanding of these mechanisms in order to make reliable predictions
2. Lack of available data / methods for estimating model parameters
   - E.g. how to estimate e.g. individual susceptibility & infectivity?
   - Much improvement to be expected over the next years due to recent advances in statistical inference and data explosion
3. Inherent stochasticity of the biological system
   - Infection is a stochastic process
   - It is impossible to make accurate predictions for infection spread on the individual level

Classification of mathematical models



- There are many different types of mathematical models
- Classifying them into broad categories can tell you much about their purpose & scope and often require different mathematical techniques
- Typical distinctions:
    - Empirical vs mechanistic
    - Deterministic vs. stochastic
    - Systems vs molecular model
    - Static vs dynamic
    - Linear vs non-linear
    - Discrete vs. continuous

All mathematical models are composed of variables and a mathematical representation of the relationship between them

Empirical vs mechanistic models

We will use both types of approaches to study host-pathogen interactions.

(See lecture 8)

Empirical Models (also called Statistical Models):

- Data driven modelling approach
- Starting point: data obtained from empirical studies
- Aim: to determine patterns & relationships between data (model variables)
- Require no prior knowledge of the underlying biology

Mechanistic Models (also called Process Based Models):

- Hypothesis driven modelling approach
- Starting point: specific phenomena of interest – observed from data
- Aim: to provide understanding for underlying mechanisms of this phenomenon
- Require prior understanding of system
- Data are used to parameterise / validate the model

Deterministic vs stochastic models

Deterministic models

- Assume that the outcome is precisely determined by the model inputs and relationships
- Ignore all random variation
- A given input always produces the same output

Stochastic models

- Incorporate inherent randomness
- Use a range of values for the model variables in form of probability distributions
- The same input produces an ensemble of outputs

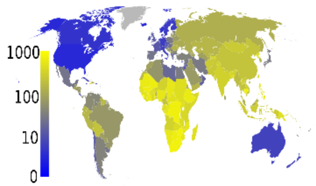
Hybrid models

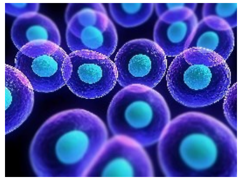
- include stochasticity on one scale (e.g. population)
- assume underlying deterministic processes (e.g. for individual)

We will use both types of approaches for modelling epidemics

Classification according to the scale of modelling

- National
- Herd
- Individual
- Organ
- Cell
- Molecules
- Genes





Mechanistic models often combine 2 or more adjacent levels of the hierarchy

Systems models combine several levels of the hierarchy

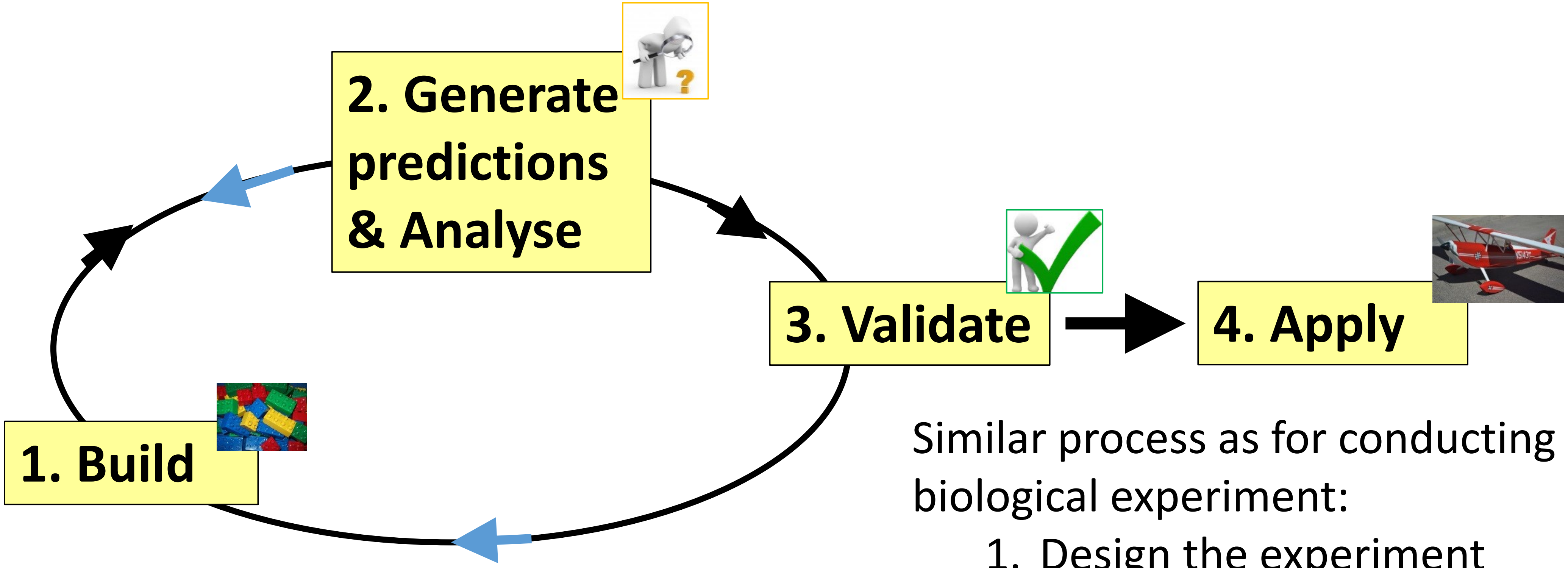
See lecture on within host infection dynamics: (molecules → cell → organ)

The appropriate scale for modelling depends on the model objectives

What is a simulation model?

- Simulation models are not specific types of mathematical models
- The term ‘simulation model’ refers to the process of implementing mathematical model, i.e. via computer simulations
- Simulation models usually simulate the process of data generation assuming the model was true
  - E.g. simulate an epidemic or the within host infection process
  - Simulate an experiment

The 4 stages of modelling



Similar process as for conducting a biological experiment:

1. Design the experiment
2. Generate data
3. Analyse experimental data
4. Validate experimental findings
5. Apply results in practice

But modelling can be much more elaborate

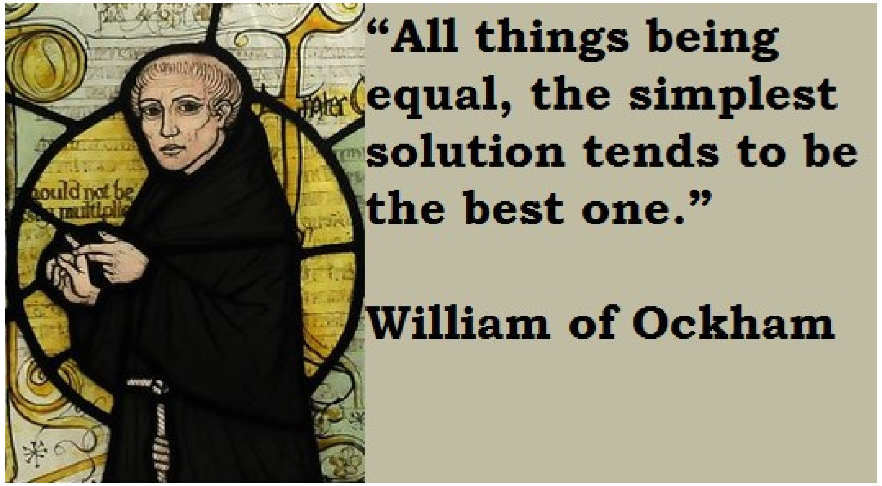
Stage 1: Building models

1. Define the model objectives

- Be clear about what you want your model to do

2. Determine the appropriate level & key model components

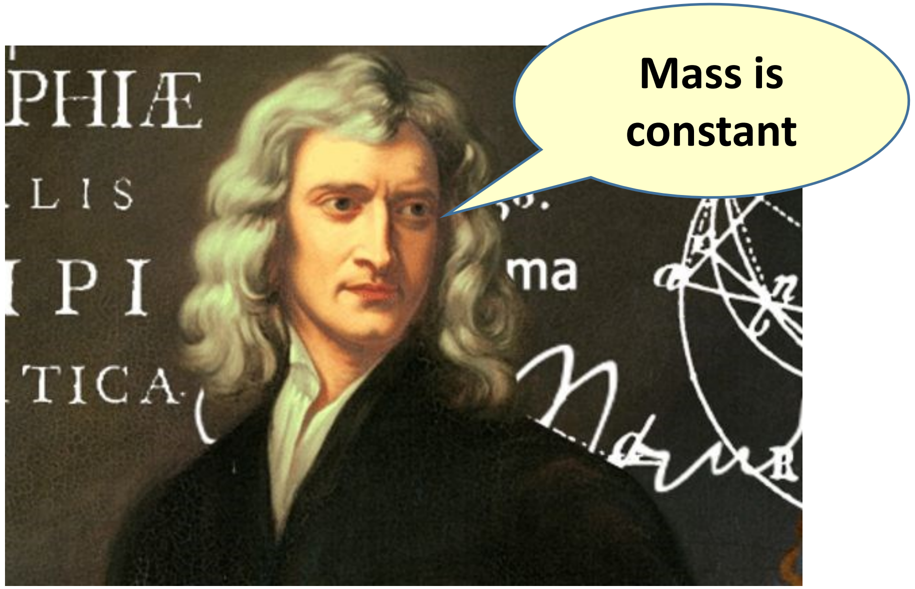
- What level of simplification is required?
- Apply the **principle of Ockham's razor** (also known as the law of parsimony):



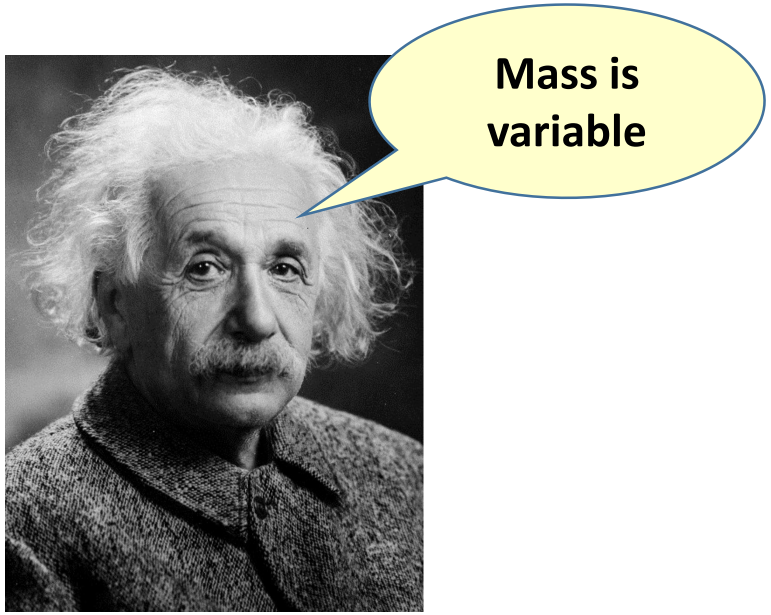
Stage 1: Building models (cont.)

3. Define your assumptions

- Assumptions reflect our beliefs how the system operates
- Remember: the model results are only as valid as the assumptions!
- Different assumptions can lead to fundamental differences:



Isaac Newton: founder of classical mechanics



Albert Einstein: Founder of relativity theory

Example: A common assumption in population studies

" A population grows at a rate that is proportional to its size"

- Embedded in the deterministic model : $\frac{dp}{dt} = ap$,

  where *p(t)* is the population size at time t and *a* is a constant.
- The solution of this model is $p(t) = p(0)e^{at}$, i.e. population grows exponentially
- The model incorporates a number of other important assumptions:

A common assumption in population studies

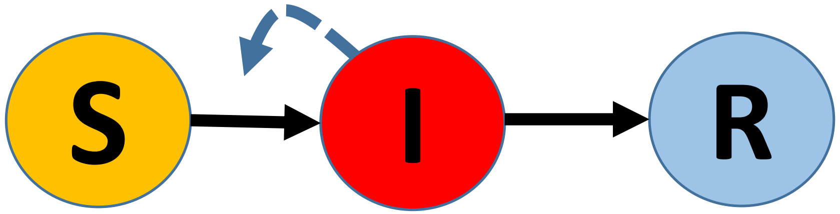
" A population grows at a rate that is proportional to its size"

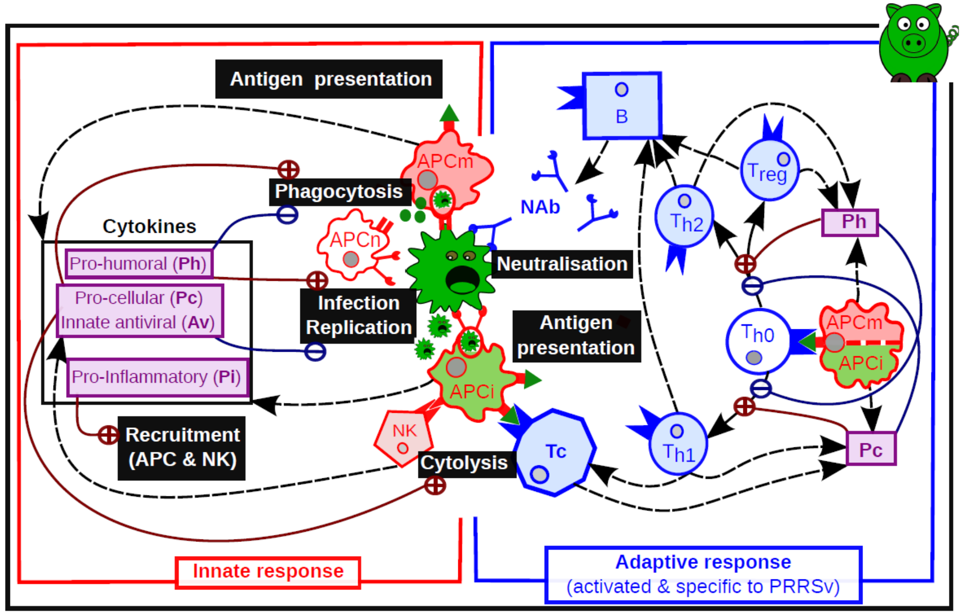
- Embedded in the deterministic model : $\frac{dp}{dt} = \; ap,$

  where *p(t)* is the population size at time t and *a* is a constant.
- The solution of this model is $p(t) = p(0)e^{at}$, i.e. population grows exponentially
- The model incorporates a number of other important assumptions:
  1. There is no limiting factor that prevents the population to grow forever
  2. Growth is a continuous process (embedded by the differential equation)
  3. Growth follows a deterministic law
     - Alternative stochastic approach: model birth and death events
- If any of these assumptions don't hold, the model is wrong!

4. Produce a flow diagram

- Visual tool for formulating our beliefs and assumptions
- Describe the model components (variables) and their relationship
- Extremely important for complex models with many components and relationships
- You will see many of these in this course





5. Write model equations

How to find the appropriate mathematical equations?

- **Depends on the modelling approach:**
  - Statistical models are often represented by a single linear or non-linear function
  - Deterministic mechanistic models of dynamical systems are usually represented by systems of differential equations
  - Stochastic models require expressions for the probability of events
- **Start with equations from the literature**
  - You are likely not the first one to model a specific system. Start by exploring and modifying existing models
- **Explore your own data**
  - see e.g. ‘Woods model’ in within-host infection dynamics lecture

Stage 2: Generate model predictions & analyse

There are 2 ways of **solving the model equations for given parameter values**

1. Analytically (using mathematical principles)

- Ideal, provides **exact** solutions and hence a full insight of the model behaviour
- But usually only possible for very simple systems (e.g. one equation or system of linear equations)

2. Numerically (using computers)

- Applies to most mathematical models
- Requires the use of numerical algorithms implemented in computational routines (e.g. Euler method, Runge-Kutta, Monte-Carlo)
- Provides **approximate** solutions
- Use established code, avoid writing your own numerical solver!!!

Specifying appropriate model inputs & outputs

- Modeller’s dilemma: lack of physical constraints in the modelling world implies that one can generate **A LOT** of data.

  How to go about it in a systematic way?

  1. Specify realistic value ranges for the model input parameters
  2. Focus on relevant scenarios if the model involves simulations
  3. Generate relevant outputs & summary statistics

Estimating model input parameter values

- Good estimates of the model input parameters are essential for models with predictive power
  - Apply principle of Ockham's razor: favour the model with fewer parameters
- 2 sources for determining appropriate parameter values:
  1. Use values reported in the literature
  2. Fit your model to existing data (statistical inference)
     - Note that it is often not possible to infer a unique value (with confidence interval) for each model parameter from given data
     - There are many different approaches of statistical inference; the right approach depends on both the type of model & the data

See 'Statistical Inference' lectures on Thursday

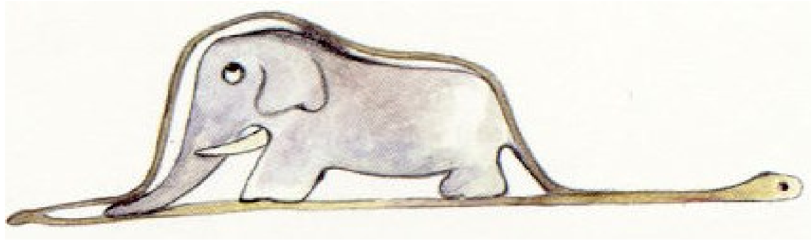
Choosing relevant model scenarios & outputs

Criteria for choosing model scenarios:

- Realistic scenarios, to **achieve your research objective**
- Extreme scenarios, to **determine the limitations of the model**

Produce meaningful model outputs

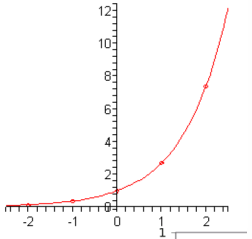
- Models produce predictions for every variable over time
  - Model variables are not always measurable → comparison to data difficult
- Produce also model outputs that can be directly compared to data
  - essential for **model validation**
  - Apply similar statistical analysis as for experimental data (frequency distributions, means, variance etc.)
- Assess relationships between observable and underlying biological traits
  - useful for **gaining new insights**

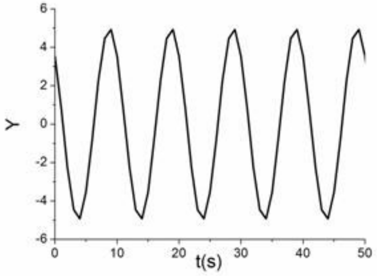


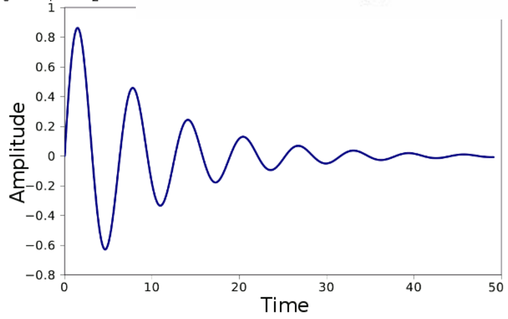
Analysing the model

- **The aim is to obtain a thorough understanding what your model can / cannot do**
- **Comprises both qualitative & quantitative analysis:**
  - What types of response patterns does the model generate?
    - How realistic are these?
  - What mechanisms / parameter values produce the diverse patterns?
    - Which inputs correspond to which outputs?
  - How sensitive are the model output to changes in the input parameter values?
  - How stable are the model predictions to small changes in starting values / assumptions?
- **Very elaborate step and often results in rebuilding the model**

Analysis techniques:
Distinguish between short- and long-term behaviour





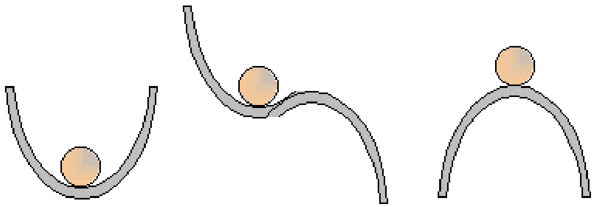


1. Asymptotic behaviour

- Does the system eventually settle to a steady state?
  - **E.g. will the infection eventually clear or persist?**
- How many steady states (long-term outcomes) are there?
- Under what conditions will a particular steady state be reached
- *Use mathematical stability analysis, bifurcation theory*

2. Initial phase behaviour

- E.g. will the infection kick off after introduction of 1 infectious agent?
- How does the initial behaviour depend on the starting point?



Analysis techniques:
Sensitivity analysis & Uncertainty analysis

- **Uncertainty analysis:** assess variability in model outputs that arise from uncertainty in model inputs
  - *How confident are we about the model predictions?*
- **Sensitivity analysis:** quantifies the influence of each parameter or modelled process on the model outputs
  - *How sensitive are the model predictions to changes in the input parameter values or modelled processes?*

Sensitivity analysis & Uncertainty analysis cont.

- **Essential components** of model analysis, especially when parameter values are unknown
- **Complex tasks**, given that there are usually complex interactions between parameter values
- Typical approaches:
  - Change one / few parameters at a time, keeping the others fixed
  - Adopt partial factorial designs, e.g. *Latin Hypercube Sampling*

Stage 3: Validating the model

- Ideally (but not necessarily!) involves comparison of model predictions to experimental data
- Important to use independent data to those used for parameter estimation
  - If independent data don't exist, split your data into training and validation set
- Useful summary statistics for comparing model predictions (P_i) to observations (O_i):

Bias (B) = $\frac{1}{n} \sum_{i=1}^{n}(P_i - O_i)$

Standard deviation (SD) = $\sqrt{\frac{1}{n} \sum_{i=1}^{n}(P_i - O_i - B)^2}$

Prediction mean square error (MSE) = $\frac{1}{n} \sum_{i=1}^{n}(P_i - O_i)^2$

What if model predictions are different to the observations?

Identify potential reasons for imperfect predictions:

1. Natural variability in the real system and environment
   - Equates to experimental measurement errors
   - Obtain confidence intervals directly from the data; if model predictions fall within these limits, don't worry
2. Mis-specifications in the model
   - Wrong parameter values → extend parameter range, use fitting algorithms
   - Errors in the choice of model equations
   - Restrict the scope of the model or look for better equations and start again
3. Effects of factors ignored in the model
   - Increase model complexity and start again

Comparing alternative models

Independent models:

- Subjective choice: no objective model selection criterion available
- Balance between generality, flexibility, predictive ability, computing requirements

See Foot & Mouth Disease models example later today

Related (e.g. nested) models:

- For models with likelihood (L), k parameters and n available independent data points, use information criteria (IC) such as
  - AIC (Akaike IC): $-2\log(L) + 2k$; defined for nested models
  - BIC (Bayesian IC): $-2\log(L) + k\log(n)$; penalizes models with more parameters

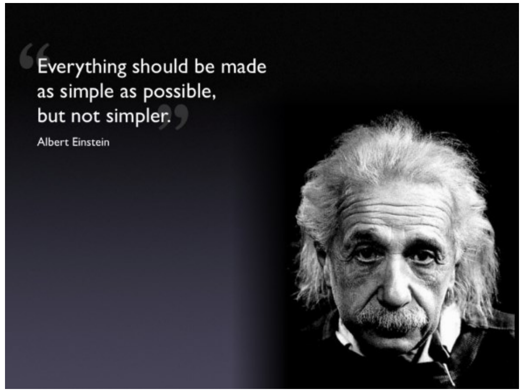
Stage 4: Applying the model

- Mathematical models can be a valuable decision support tool
  - For risk assessment – particularly important in infectious disease context
  - To predict consequences of various (disease) control strategies

- It requires trust that the model predictions are valid
  - It is crucial to keep the purpose of the model and the end user of the model in mind at all modelling stages
  - The user should have a thorough understanding of the model assumptions, model predictions (with uncertainty estimates) and limitations

See Foot & Mouth Disease models example later today

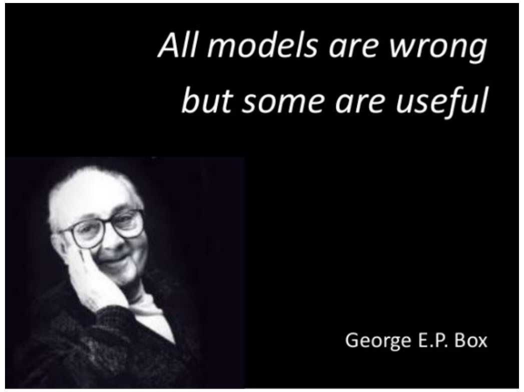
What is a good model?



Key attributes of a good model:

1. Fit for purpose
   - As simple as possible, but sufficiently complex to adequately represent the real system without obstructing understanding
   - Appropriate balance between accuracy, transparency and flexibility
2. For predictive models: Parameterizable from available data

Keep in mind that no model is perfect!



Further reading

- Otto, Sarah P., and Troy Day. *A biologist's guide to mathematical modeling in ecology and evolution*. Vol. 13. Princeton University Press, 2007.
  - A nice introduction to mathematical modelling with plenty of applications from ecology and evolutionary systems.
- Renshaw, Eric. *Modelling biological populations in space and time*. Vol. 11. Cambridge University Press, 1993.
  - A good and not too mathematical introduction to deterministic and stochastic models of biological systems.
- Cross, Mark, and Alfredo O. Moscardini. *Learning the art of mathematical modelling*. John Wiley & Sons, Inc., 1985.
  - A readable, non-technical book on how to start modelling and how to teach others.