Types of models

Andreas Handel

University of Georgia
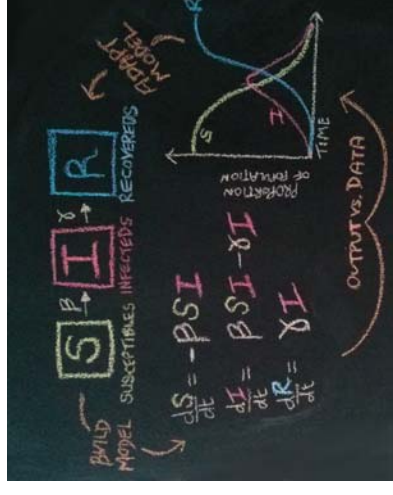
2019-07-11

Modeling definition

Not this

This one

- The term **modeling** usually means (in science) the description and analysis of a system using mathematical or computational models.
- Many different types of modeling approaches exist. Simulation models are one type (with many subtypes).

Types of models

- Models can be
  - Experimental
  - Mathematical/equation based
  - Computational/algorithm based
  - Applied to data
  - Used without data
  - Static or dynamic
  - Mechanistic or not
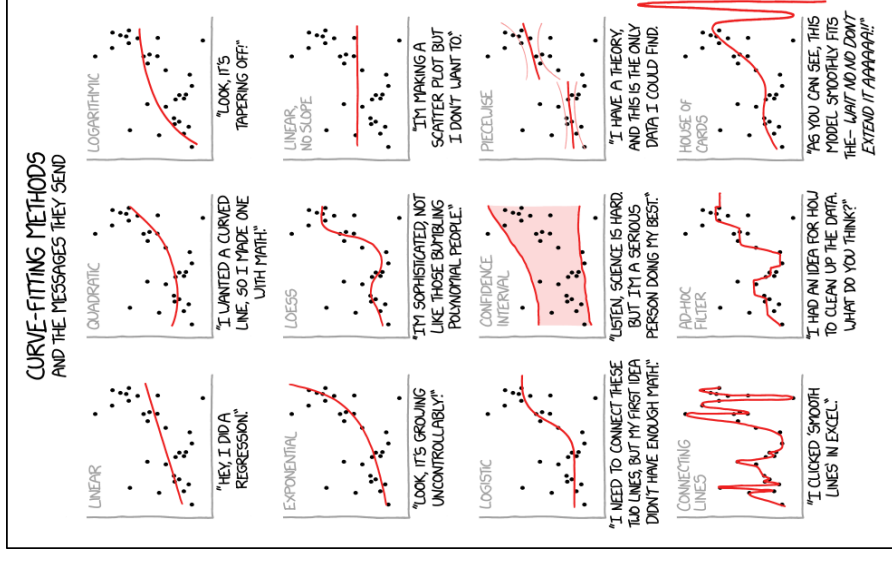 - Applied to things, animals, molecules, cells, sequences, individuals, …

The terms model and modeling are widely applied to different things.

# A way to classify computational/mathematical models

- **Statistical/phenomenological models**
  - Look at patterns in data
  - Do not describe mechanisms leading to the data
- **Mechanistic/process/simulation models**
  - Try to represent simplified versions of mechanisms
  - Can be used with and without data

Phenomenological/Non-mechanistic models

- You might be familiar with statistical models (that includes Machine Learning, AI, Deep Learning,...).

- Most of those models are phenomenological/non-mechanistic (and static).

- Those models are used extensively in all areas of science.

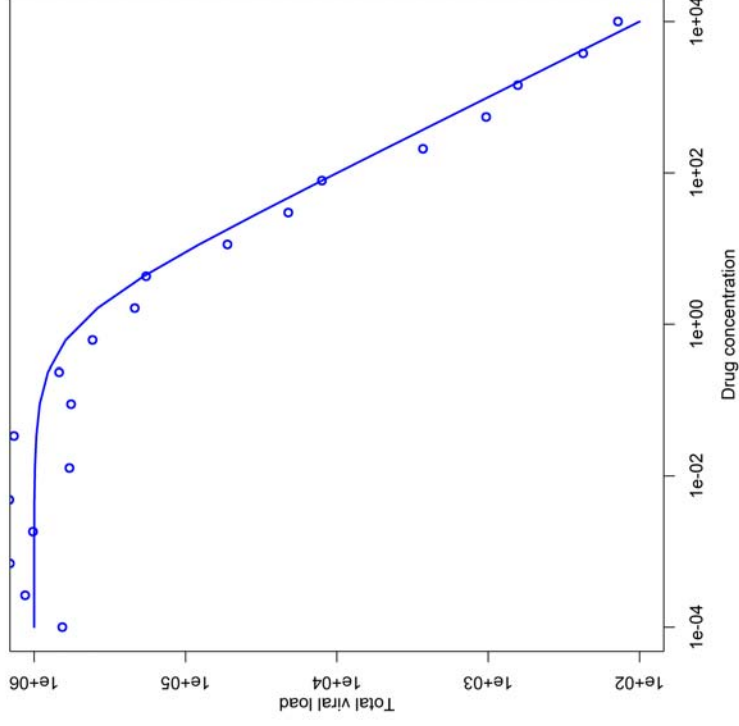- The main goal of these models is to understand data/patterns and make predictions.

xkcd.com

Non-mechanistic model example 1

Fraction infected, F (y-axis)
Duration of infection, D (x-axis)
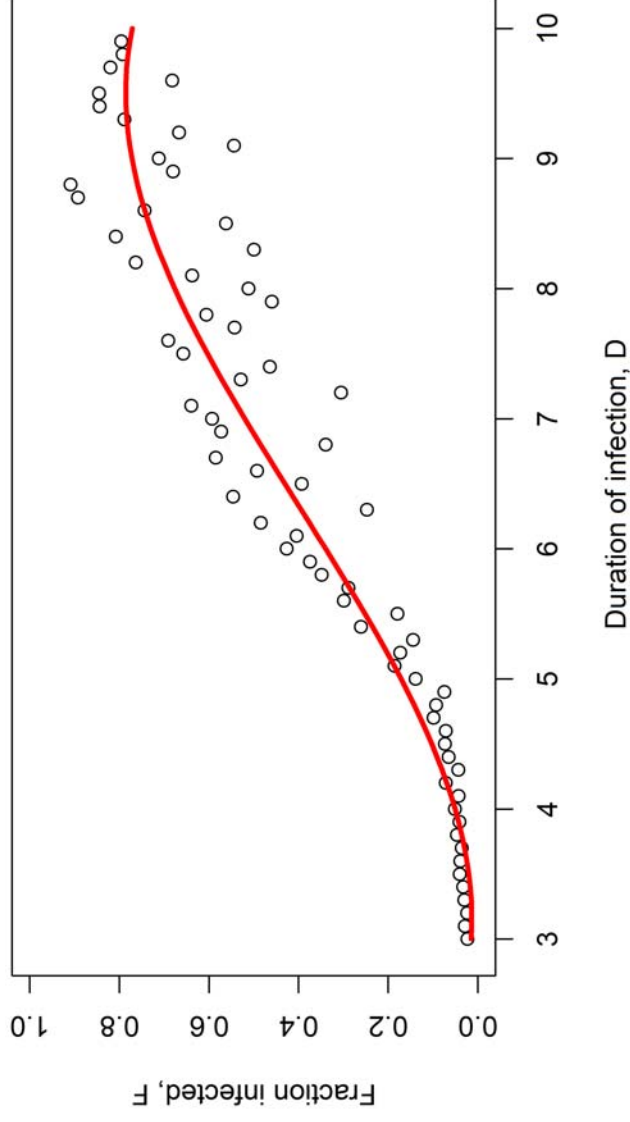
Impact of duration of infection on outbreak size. The fit model is `

$$F = b_0 + b_1 D + b_2 D^2 + b_3 D^3$$ ` .

Non-mechanistic model example 2

Impact of drug dose (C) on virus load (area under the curve, AUC). The function AUC = `$V_{max}(1 - \frac{C}{C+C_{50}})$` is fit to the data.
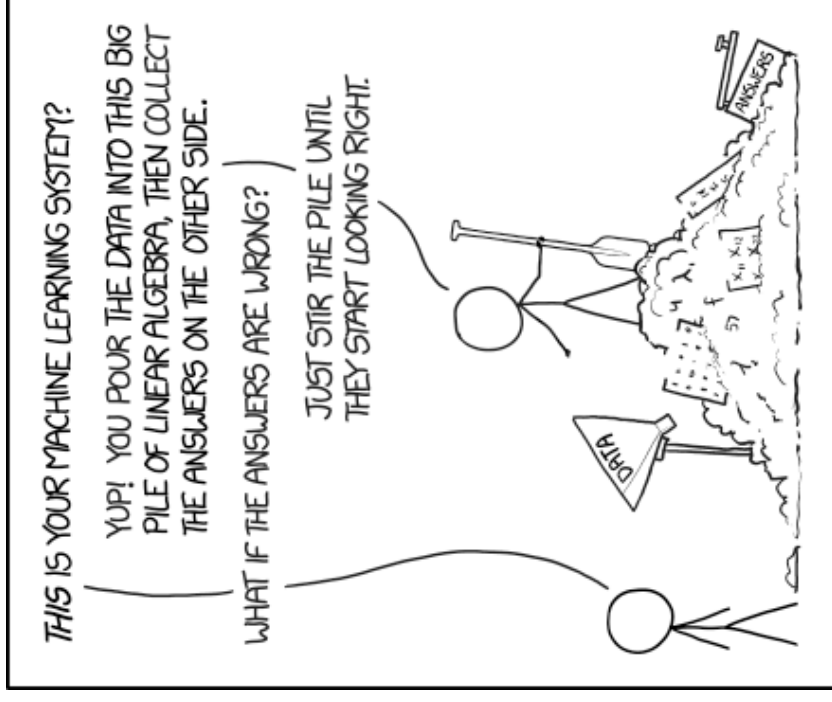
Non-mechanistic models - Advantages

- Finding correlations/patterns is (relatively) simple.

- Some models are very good at predicting (e.g. Netflix recommender, Google Translate).

- Sometimes we can go from correlation to causation.

- We don't need to understand all the underlying mechanisms to get actionable insights.
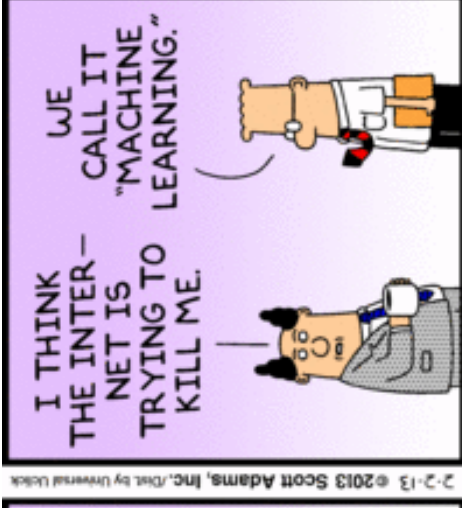

dilbert.com

Non-mechanistic models - Disadvantages

- The jump from correlation to causation is always tricky (bias/confounding/systematic errors).

- Even if we can assume a causal relation, we do not gain a lot of mechanistic insights or deep understanding of the system.

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.

DATA

ANSWERS

xkcd.com

Mechanistic models

- We formulate explicit mechanisms/processes driving the system dynamics.

- This is done using mathematical equations (often ordinary differential equations), or computer rules.

- Also called *systems dynamics models* or *(micro)simulation models.*
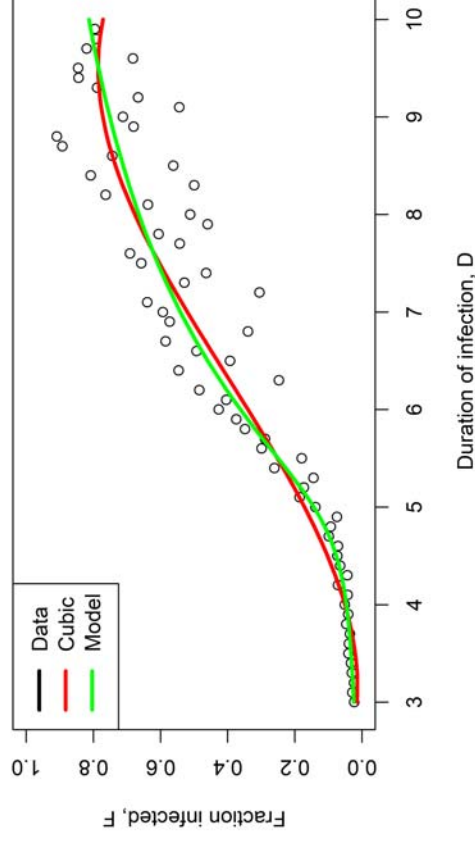
Infection

Recovery
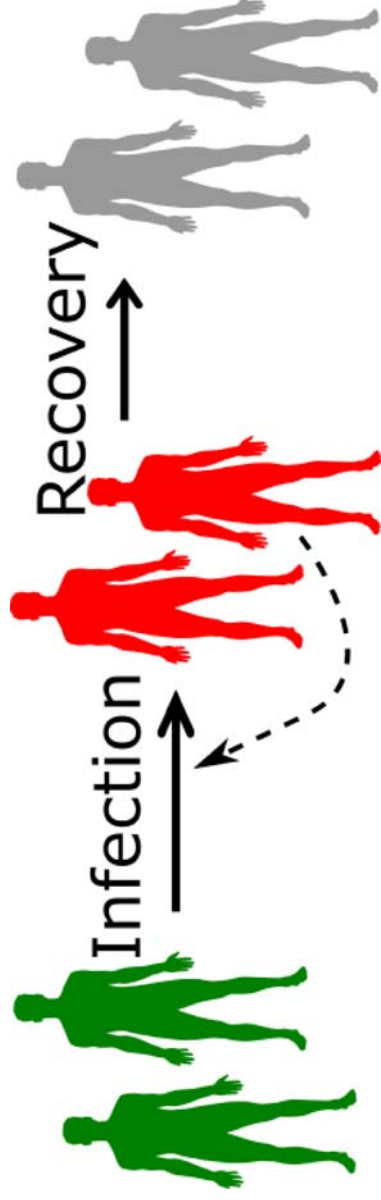
$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

Mechanistic model example 1

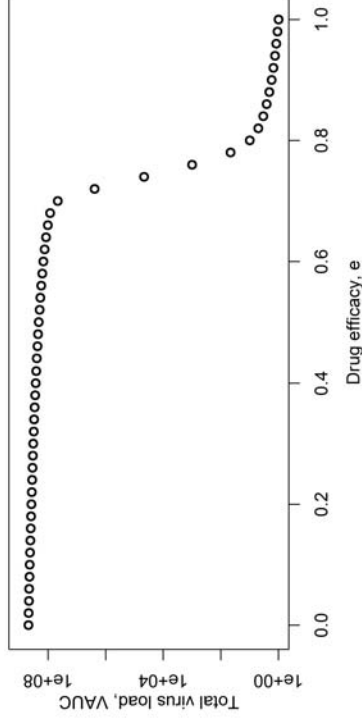

Impact of duration of infection (D=1/g) on outbreak size. The fit model is the one shown below.

$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

Mechanistic model example 2

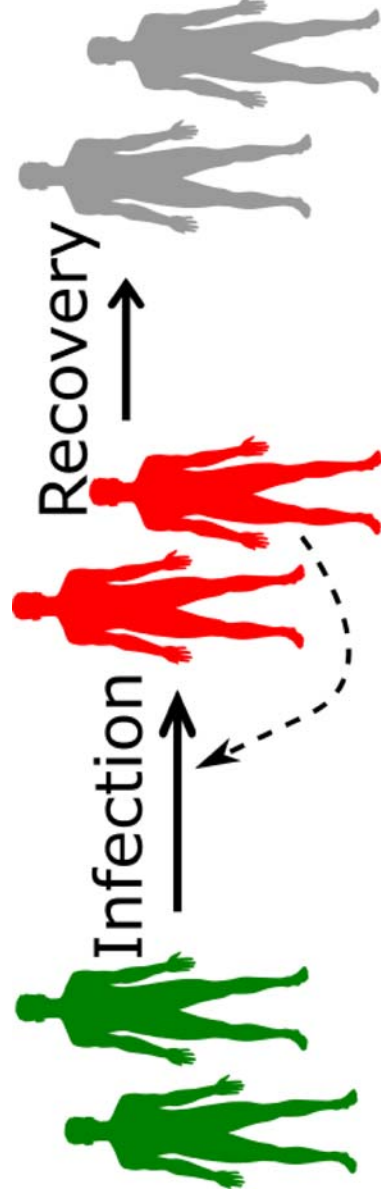Total virus load (VAUC) as function of drug efficacy, e.

Uninfected Cells	$\dot{U} = n - bUV - mU$
Infected Cells	$\dot{I} = bUV - d_I I - mI$
Virus	$\dot{V} = p(1-e)I - d_V V - gbUV$

Mechanistic models - Advantages

- We get a potentially deeper, mechanistic understanding of the system.
- We know exactly how each component affects the others in our model.

Infection

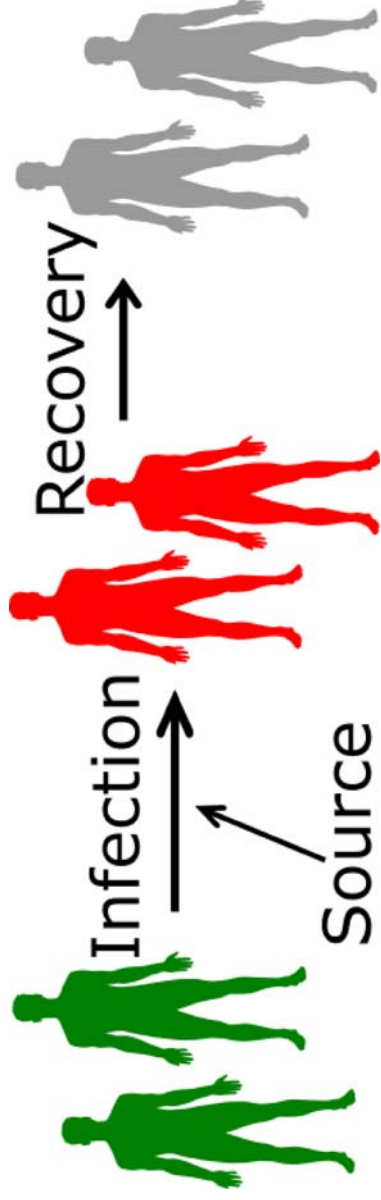Recovery

$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

Mechanistic models - Disadvantages

- We need to know (or assume) something about the mechanisms driving our system to build a mechanistic model.

- If our assumptions/model are wrong, the "insights" we gain from the model are spurious.

Infection

Recovery

Source

$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

Non-mechanistic vs Mechanistic models

- Non-mechanistic models (e.g. regression models, machine learning) are useful to see if we can find patterns in our data and possibly predict, without necessarily trying to understand the mechanisms.

- Mechanistic models are useful if we want to study the mechanism(s) by which observed patterns arise.

Ideally, you want to have both in your 'toolbox'.

Mechanistic model types

There are many ways mechanistic models can be formulated and implemented. Here are some ways of characterizing them:

- *Compartmental or Agent-based*
- *Discrete time or Continuous time*
- *Deterministic or Stochastic*
- *Space-less (homogeneous) or Spatial*
- *Memory-less (Markov) or With memory*
- *Small or Big*
- *Data-free or With data*

The most common model type is one using ordinary differential equations (ODEs). Such models are usually compartmental, continuous time, deterministic, space-less, memory-less, and small(ish).

Compartmental versus Agent/Individual based models

Compartmental models

- The components of the model are treated as homogeneous groups (compartments), one only tracks population numbers/sizes.

- The simplest type of model, sometimes mathematically tractable, easy to implement on a computer.

- Good model for fitting data.

- The assumption that populations are homogeneous and "well mixed" is always wrong (but sometimes it is a good enough approximation).

- Often the best starting point.

Compartmental models - example 1

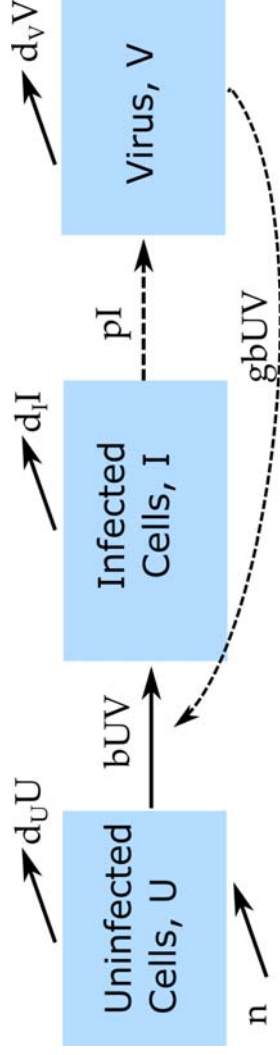| Susceptible S | → bSI → | Infected/ Infectious I | → gI → | Recovered R |

$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

$$S_{t+dt} = S_t - dt(bS_t I_t)$$
$$I_{t+dt} = I_t + dt(bS_t I_t - gI_t)$$
$$R_{t+dt} = R_t + dt(gI_t)$$

Compartmental models - example 2

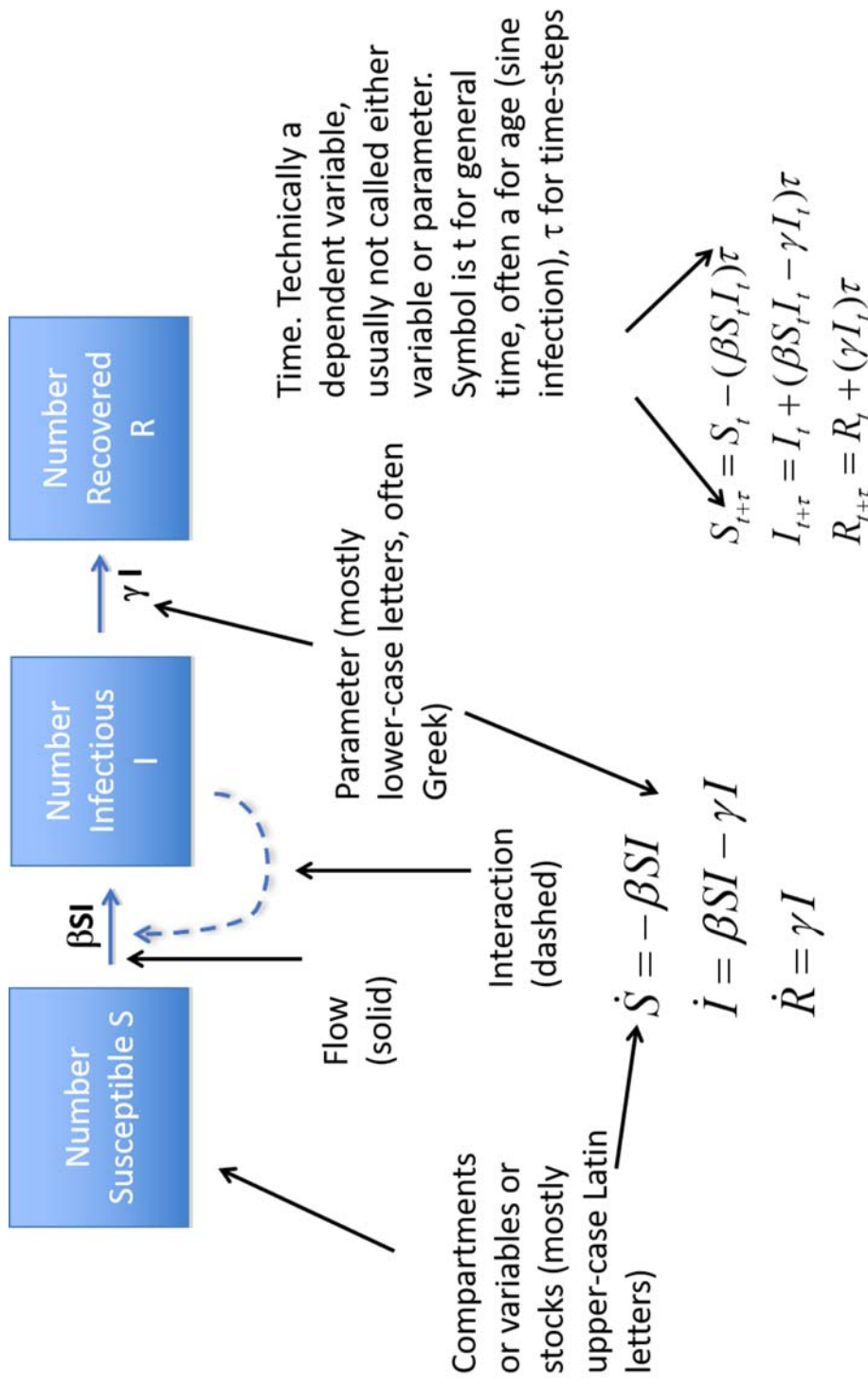Uninfected Cells $\dot{U} = n - d_U U - bUV$

Infected Cells $\dot{I} = bUV - d_I I$

Virus $\dot{V} = pI - d_v V - bgUV$

(Compartmental) models - terminology

Number Susceptible S

Number Infectious I

Number Recovered R

βSI

γI

Flow (solid)

Interaction (dashed)

Compartments or variables or stocks (mostly upper-case Latin letters)

Parameter (mostly lower-case letters, often Greek)

Time. Technically a dependent variable, usually not called either variable or parameter. Symbol is t for general time, often a for age (sine infection), τ for time-steps

$$\dot{S} = -\beta SI$$

$$\dot{I} = \beta SI - \gamma I$$

$$\dot{R} = \gamma I$$

$$S_{t+\tau} = S_t - (\beta S_t I_t)\tau$$

$$I_{t+\tau} = I_t + (\beta S_t I_t - \gamma I_t)\tau$$

$$R_{t+\tau} = R_t + (\gamma I_t)\tau$$

Individual/agent based models

- Every unit/host/individual is modeled/tracked. Called agent-based or individual-based models (ABM/IBM).

- Mostly computational, (almost) no mathematical analysis is possible.

- IBM usually have many parameters, are difficult to build and take long to run.

- IBM are difficult to fit to data.

- IBM are conceptually easy to understand.

- IBM are potentially most detailed and realistic.

Individual based models - example
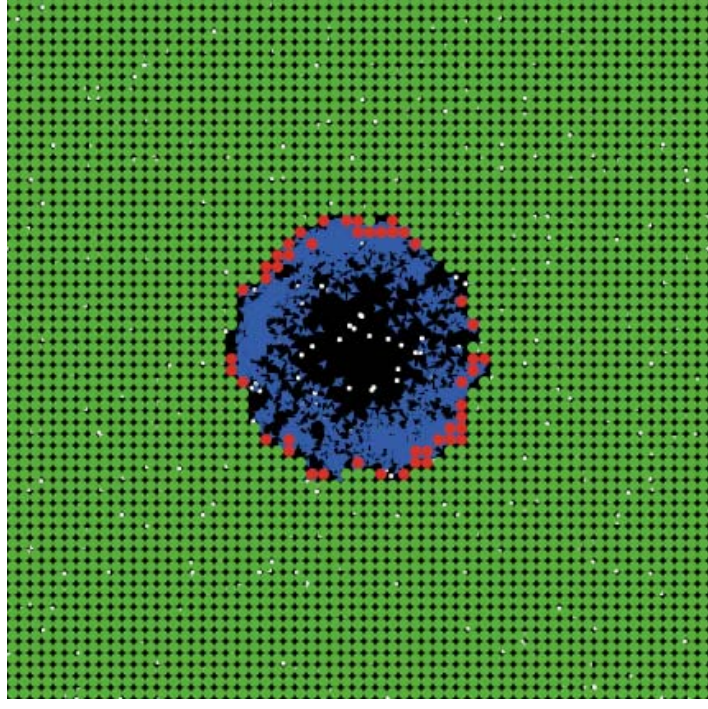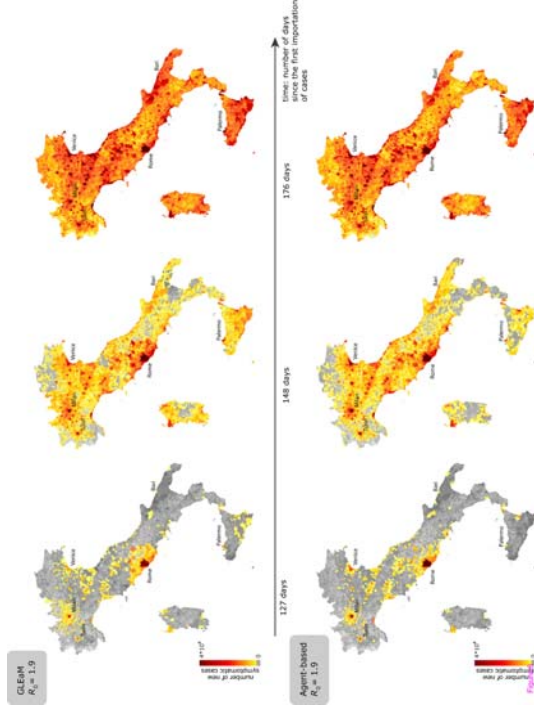
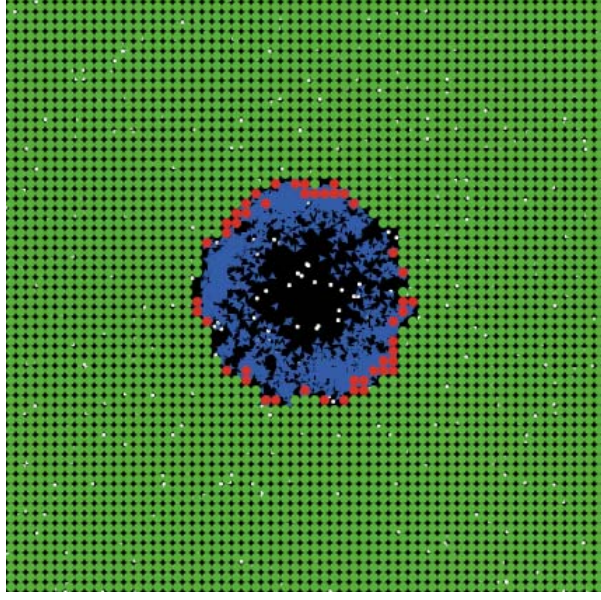# Discrete time versus continuous time models

Discrete time models

- The model is updated in discrete time-steps.

- Good for systems where there is a "natural" time step.
  - Example: For some animals, births occur during a small period in spring. Modeling the long-term dynamics of an ID in such a population might lend itself to a model that is updated annually.

- Complex models, such as agent-based simulations are almost always discrete-time (for computational reasons).

- Discrete-time compartmental models are often formulated as difference equations.

- If the time-step becomes small, a discrete-time model approaches a continuous-time model.
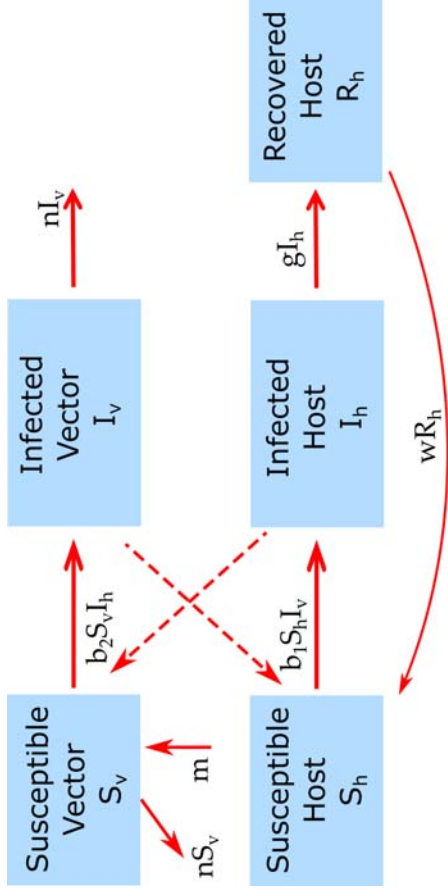
Discrete time models - example

| Susceptible S | $\xrightarrow{\text{bSI}}$ | Infected/ Infectious I | $\xrightarrow{\text{gI}}$ | Recovered R |

$$S_{t+dt} = S_t - dt(bS_tI_t)$$
$$I_{t+dt} = I_t + dt(bS_tI_t - gI_t)$$
$$R_{t+dt} = R_t + dt(gI_t)$$

Continuous time models

- The model is updated continuously.

- Best for systems where changes occur continuously and concurrently.
  - Example: To model an outbreak of flu (or some other ID) in a large population, with new infections and recoveries occurring continuously and concurrently, a continuous-time model might be best.

- Continuous-time models are usually described by differential equations.

- Ordinary differential equation (ODE) models are the most common and simplest one.

Continuous time models - example 1

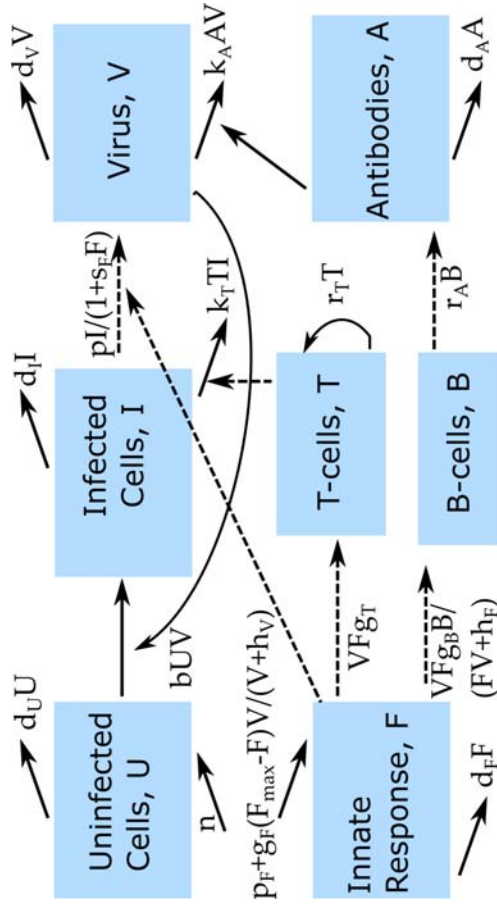Vector borne infectious disease model.

$$\dot{S}_h = -b_1 S_h I_v + w R_h$$
$$\dot{I}_h = b_1 S_h I_v - g I_h$$
$$\dot{R}_h = g I_h - w R_h$$
$$\dot{S}_v = b - b_2 S_v I_h - n S_v$$
$$\dot{I}_v = b_2 S_v I_h - n I_v$$

Continuous time models - example 2

Uninfected Cells, U

Infected Cells, I

Virus, V

Innate Response, F

T-cells, T

B-cells, B

Antibodies, A

$d_U U$ $d_I I$ $d_V V$

n bUV $pI/(1+s_F F)$ $k_T TI$ $k_A AV$

$p_F + g_F (F_{max} - F) V/(V + h_V)$

VFg_T

$VFg_B B/(FV + h_F)$

$d_F F$ $r_T T$ $r_A B$ $d_A A$

$$\dot{U} = n - d_U U - bUV$$

$$\dot{I} = bUV - d_I I - k_T TI$$

$$\dot{V} = \frac{pI}{1 + s_F F} - d_V V - bUV - k_A AV$$

$$\dot{F} = p_F - d_F F + \frac{V}{V + h_V} g_F (F_{max} - F)$$

$$\dot{T} = FV g_T + r_T T$$

$$\dot{B} = \frac{FV}{FV + h_F} g_B B$$

$$\dot{A} = r_A B - d_A A - k_A AV$$

Deterministic versus stochastic models

Deterministic models

- For given parameters and initial conditions, the model always produces the same result.

- Simple, easy to implement on a computer (e.g. ODE models).

- Sometimes one can do analytical calculations.

- Real biological systems are never deterministic, but sometimes approximately so.

- When large numbers are involved, deterministic models tend to be good. They break down when only few entities (e.g. few hosts) are involved.

Deterministic models - examples

$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

$$S_{t+dt} = S_t - dt(bS_tI_t)$$
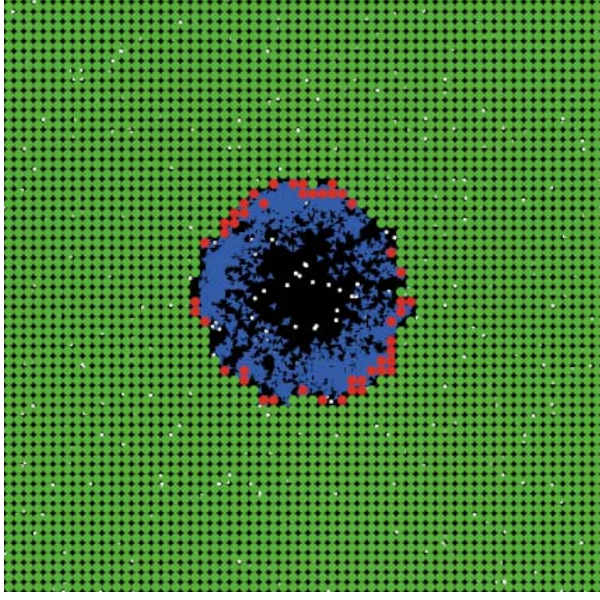$$I_{t+dt} = I_t + dt(bS_tI_t - gI_t)$$
$$R_{t+dt} = R_t + dt(gI_t)$$

$$\dot{B} = gB\left(1 - \frac{B}{B_{max}}\right) - d_B B - kBI$$
$$\dot{I} = rBI - d_I I$$

Stochastic models

- Results differ between simulations, even for the same model conditions.
- More difficult to implement on a computer, takes longer to run.
- The math is more difficult.
- Closer to the "real" system.
- Stochastic effects are important at low numbers.
- The same model implemented as deterministic or stochastic can lead to different results!

Stochastic models - example

| Susceptible S | $\xrightarrow{bSI}$ | Infected/ Infectious I | $\xrightarrow{gI}$ | Recovered R |

$$S_{t+dt} = S_t - dt(bS_t I_t)$$
$$I_{t+dt} = I_t + dt(bS_t I_t - gI_t) + \text{Noise}$$
$$R_{t+dt} = R_t + dt(gI_t)$$

Homogeneous/space-less versus spatial models

Homogeneous/space-less models

- There is no explicit notion of space. Entities (e.g. hosts) are assumed to exist in a homogenous space.

- Entities are assumed to be well-mixed and randomly bump into each other.

- Most compartmental models make this assumption.

Space-less models - example

Susceptible
S

Infected/
Infectious
I

Recovered
R

bSI

gI

$$\dot{S} = -bSI$$
$$\dot{I} = bSI - gI$$
$$\dot{R} = gI$$

Bacteria, B

Immune
Response, I

$d_B B$

gB(1-B/B$_{max}$)

kBI

rBI

$d_I I$

$$\dot{B} = gB(1 - \frac{B}{B_{max}}) - d_B B - kBI$$
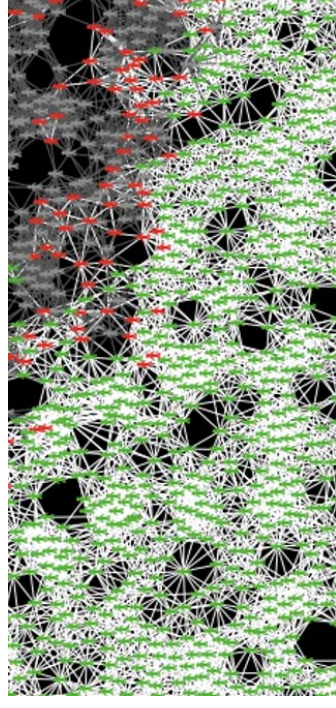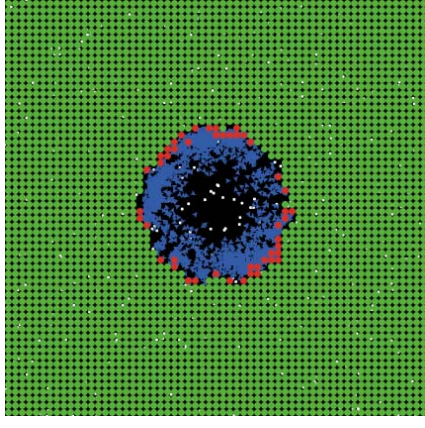$$\dot{I} = rBI - d_I I$$

Spatial models

- Some notion of space is explicitly included.
- Different types of models can be used:
  - Patch/Meta-population models. Usually coupling of multiple compartmental models.
  - Partial Differential equations.
  - Agent-based models.
  - Network models.

Spatial models - example
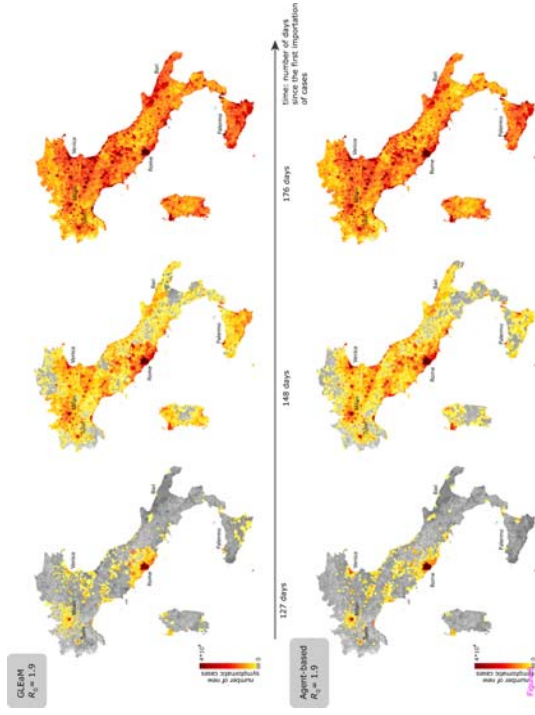
Memory-less (Markov) models versus models with memory

Memory-less models

- Many models (e.g. those based on ordinary differential equations, ODE) are memory-less (Markovian). What happens next only depends on the current state of the system, not on the past.

- That means for instance, that an infected individual has an equal chance to recover at any time, no matter how long ago the infection occurred.

- This approximation is sometimes, but not always acceptable.

Memory-less models - example

$$\dot{S} = -bSI$$

$$\dot{I} = bSI - gI$$

$$\dot{R} = gI$$

$$S_{t+dt} = S_t - dt(bS_tI_t)$$

$$I_{t+dt} = I_t + dt(bS_tI_t - gI_t)$$

$$R_{t+dt} = R_t + dt(gI_t)$$

Models with memory

- If we want to keep track of the past, e.g. if we want to let the chance of recovery depend on the time since infection, we can't use basic ODE models.
- Possible models:
  - ODE models with "dummy compartments"
  - Partial differential equations
  - Delay differential equations
  - Agent-based models

Models with memory - example

$$\frac{dS(t)}{dt} = -bS(t)I(t-\tau)$$

$$\frac{dI(t)}{dt} = bS(t)I(t-\tau) - gI(t)$$

$$\frac{dR(t)}{dt} = gI(t)$$
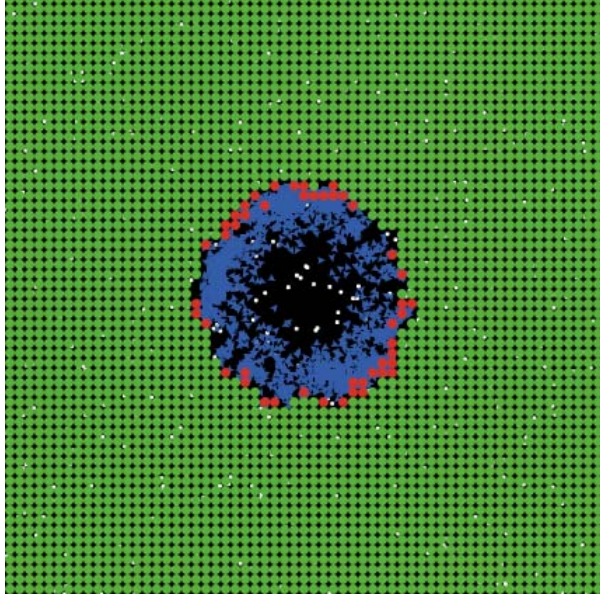
Small versus large models

Small models

- Start with a very simple model, try to capture the most important aspects of the known dynamics of the system.

- Analyze model to gain conceptual insights. It should be relatively easy to understand what is going on.

- It's often possible to fit the model to data. Model rejection (e.g. poor agreement with data) is helpful, it taught us something.

- The model might not include crucial known biology and therefore the insights/results might be of limited use (or completely useless).

Small models - example
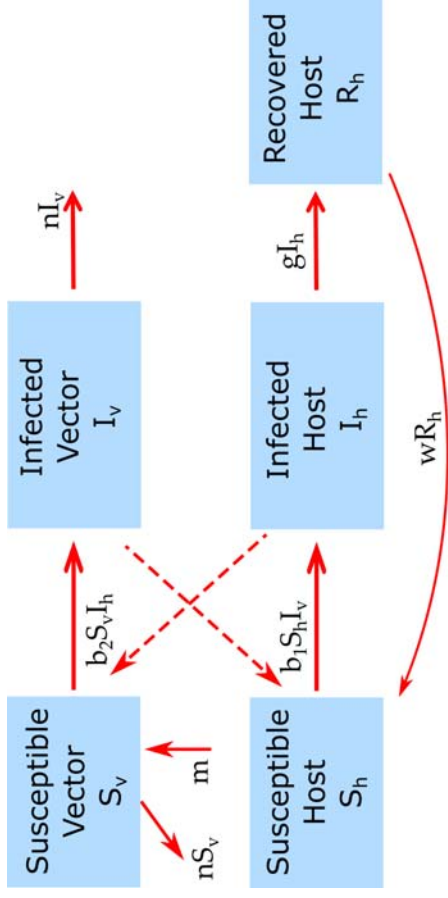
Vector borne ID model.

$$\dot{S}_h = -b_1 S_h I_v + w R_h$$
$$\dot{I}_h = b_1 S_h I_v - g I_h$$
$$\dot{R}_h = g I_h - w R_h$$
$$\dot{S}_v = b - b_2 S_v I_h - n S_v$$
$$\dot{I}_v = b_2 S_v I_h - n I_v$$

Big models

- Build a comprehensive model, try to include a lot of detail.

- Run the model on the computer, investigate results. It's often hard to understand how the different components of the model influence the result. Careful analysis is needed.

- Model results that are at odds with known biology can suggest needed model modifications. But big models can reproduce a lot of observed phenomena, even if the model is wrong.

- There is usually not enough data to fit complex models.

- Comprehensive and accurate models can be used to make detailed, quantitative predictions.

Big models - example


Ajelli et al 2010 BMC Inf Dis: https://doi.org/10.1186/1471-2334-10-190
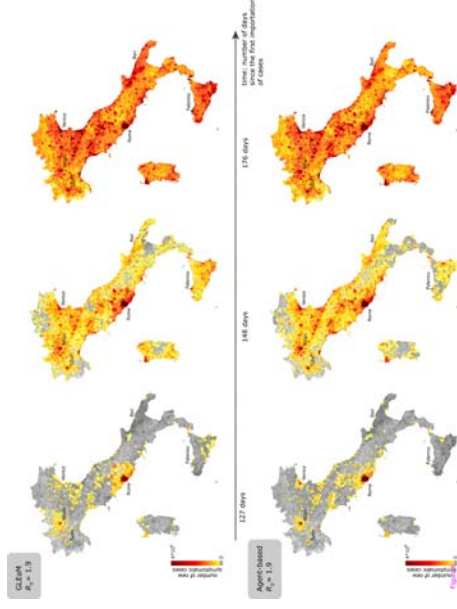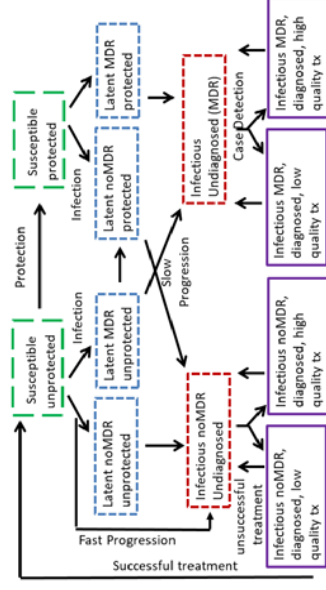
72 compartment ODE model

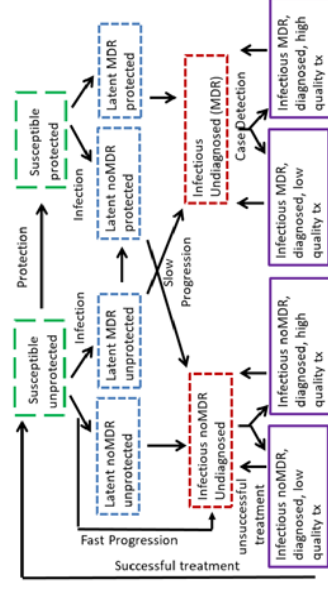# "Data-free" models versus models with data

"Data-free" models

- Model is formulated and model parameters are chosen based on known biology.

- Model is analyzed "by itself", without fitting to data.

- Can produce useful insights, even if only limited data are available.

"Data-free" models - example

Ajelli et al 2010 BMC Inf Dis: https://doi.org/10.1186/1471-2334-10-190
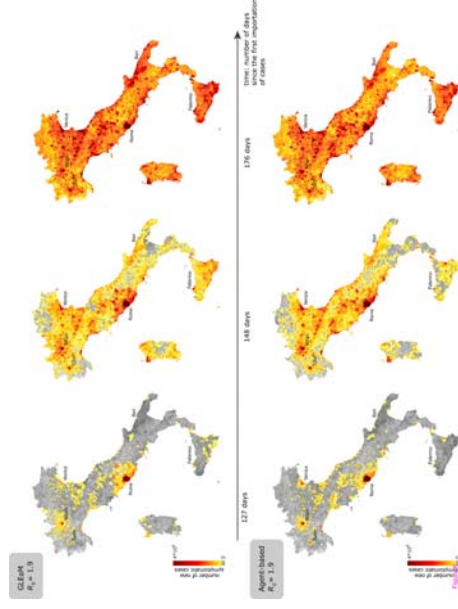
72 compartment ODE model
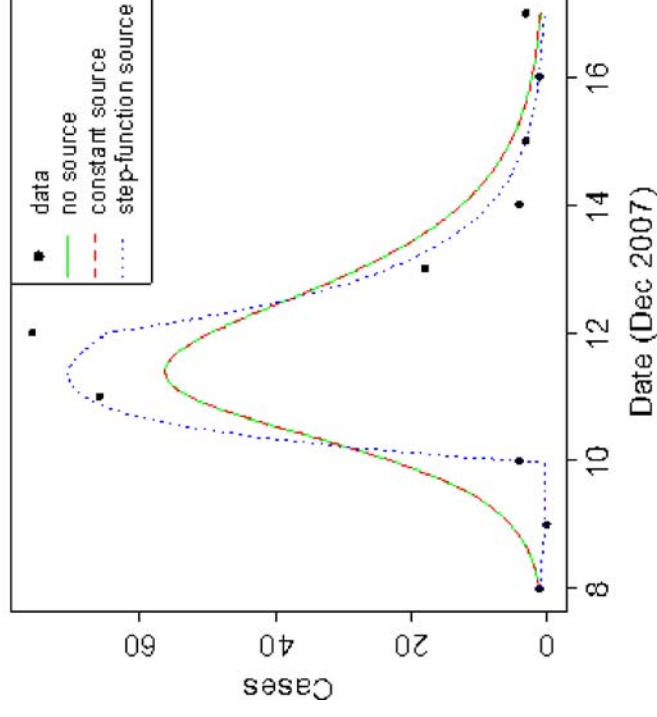
Models fit to data

- Models are being fitted to data (inference).

- Statistically/quantitative comparison of models with data.

- Used to discriminate hypotheses, determine parameters.

- Less flexibility in model formulation, often constrained/determined by available data.

Models fit to data - example

$$\dot{S} = -\mathbf{n}\mathbf{S} - bSI$$

$$\dot{I} = \mathbf{n}\mathbf{S} + bSI - gI$$

$$\dot{R} = gI$$

Types of models - summary

- Models have several of the different characteristics just described:
  - The very common ODE models are deterministic, compartmental, continuous time, no-memory, no-space.
  - Agent-based models are usually stochastic, spatial, and contain memory.

Ideal approach: Choose the model that is most suitable for the question you try to answer.

- In reality: Model selection is based on a mix of
  - Question one wants to answer
  - Expertise
  - Feasibility (computation time, model complexity)
  - "Environment" (what approaches do others use)
  - "Marketing" (what kind of models "sell")

Summary

- Simulation models are one type of the many model types out there.

- Within the category of simulation/mechanistic models, there are many different sub-categories.

- There is no single best kind of model, it depends on the question/scenario.

- Choosing an appropriate model for a given task is part of the *art* of good modeling, there is unfortunately no recipe.

Literature

- Keeling and Danon, "Mathematical modelling of infectious diseases",