NIPH
Norwegian Institute of Public Health

What is infectious disease modelling

Who am I?

- Bachelor’s degree from FAM program 2009-2013
- Master’s degree in theoretical particle physics 2015
- PhD in nuclear physics 2019
- Data scientist at Telenor 2019-2021
- Infectious disease modeller (“seniorforsker”) at Folkehelseinstituttet since 2021, mostly working on covid-19

Outline

1. Infectious disease modelling: The problem
2. Agent-based models
3. Compartmental models
4. Briefly on fitting the model to data

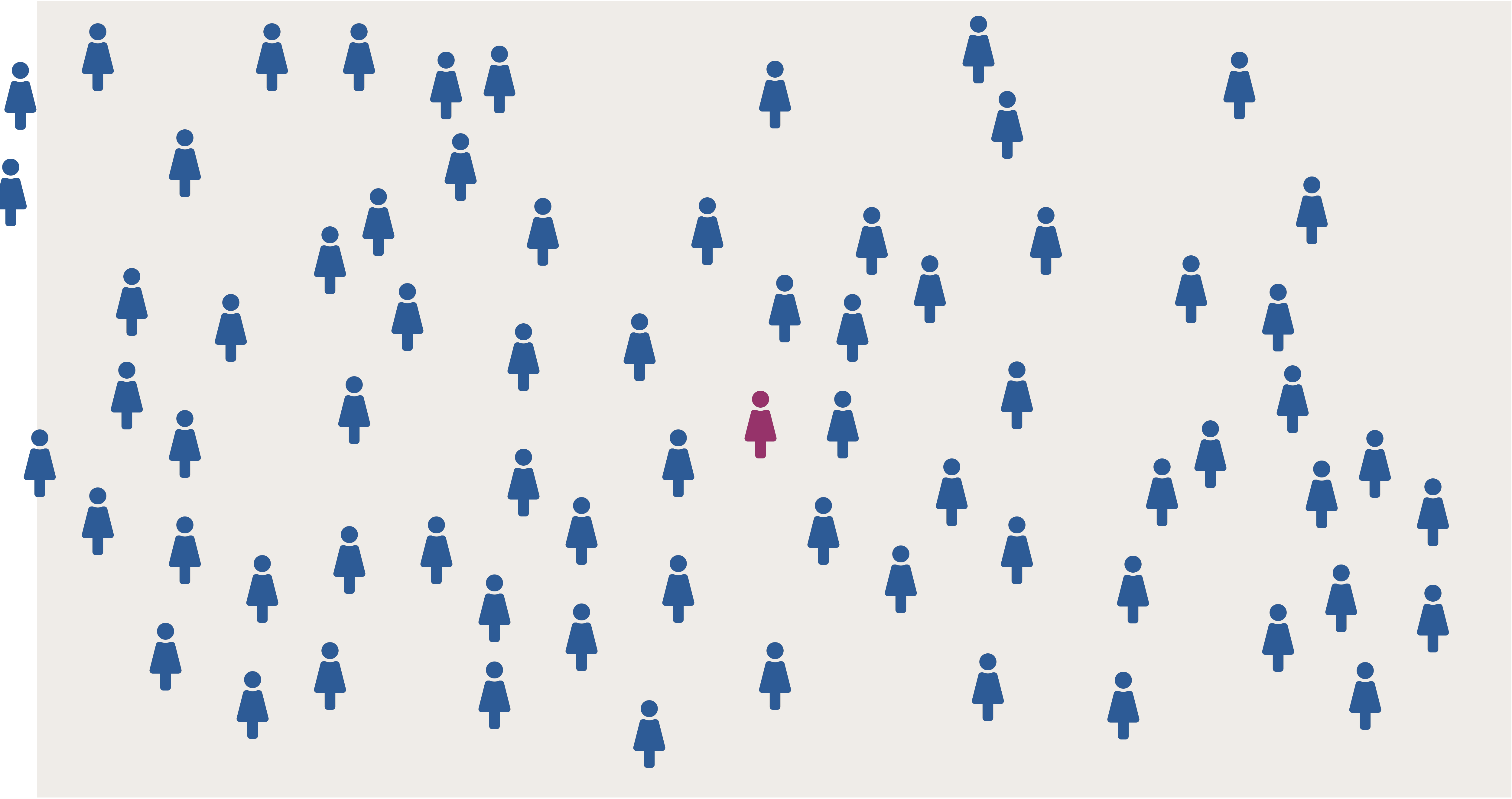
Infectious disease modelling

Infectious diseases

- Sars-Cov-2 / covid-19
- Monkeypox
- Ebola, SARS, MERS
- Influenza, common cold, norovirus, …
- Sexually transmitted diseases (HIV, gonorrhea, chlamydia, …)
- Vector-borne diseases (malaria, dengue fever, …)
- Animal diseases, e.g. foot and mouth disease

They transmit from individual to individual upon contact.

This is what we want to model.

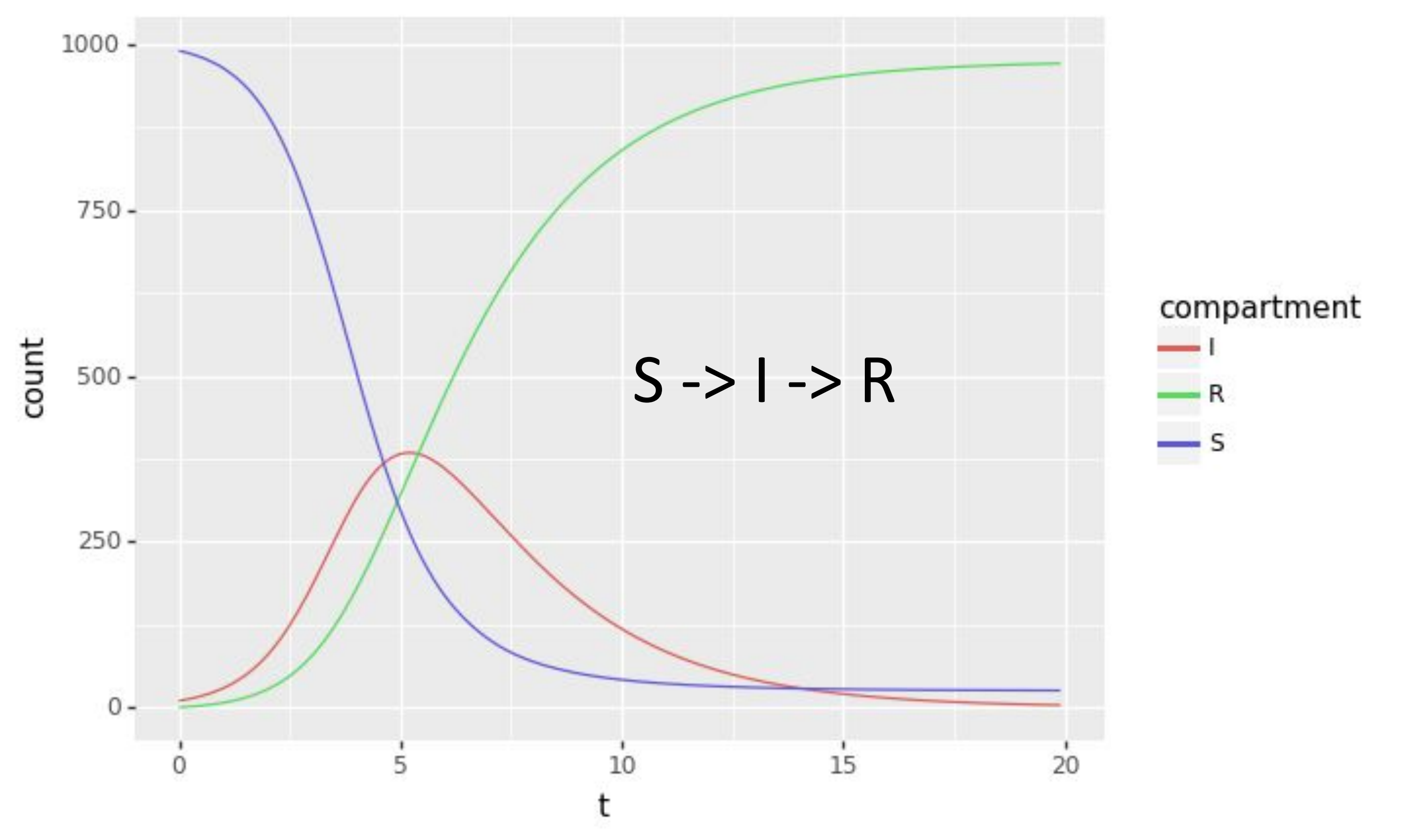


We want a model that describes the number of infected over time, I(t)

Don’t forget recoveries

Don’t forget recoveries

The SIR model



S -> I -> R

Some necessary terminology

- S: Susceptible
- I: Infected (infectious)
- R: Recovered (removed)

- *Incidence*: Number of new infections per day
- *Prevalence*: Total number presently infected

- β *(beta)*: The standard variable name for the infectiousness parameter (of the virus) in the model

More than SIR

- The SIR model assumes that infection gives sterilising immunity forever - works for measles (meslinger), chickenpox (vannkopper) etc
- This is approximately true for covid-19 on a short timescale (<~1 year)
- You can make any flavour you like:
  - SEIR (E = exposed, presymptomatic period - actually important for covid)
  - SIRS (waning of immunity over time)
  - SIS (no immunity, e.g. gonorrhea)
  - $SE_1E_2II_aR$
  - Also add vaccination to the mix

Why so many physicists?

Much overlap in methods and skills

- Numerical tools
  - programming
  - data analysis
  - visualisation
- Way of approaching a modelling problem
- Statistical proficiency
- Monte Carlo methods

Agent-based models

Agent-based models

A physicist’s preferred approach (?)

Microscopic model keeping track of individual agents and their behaviour and infectious status

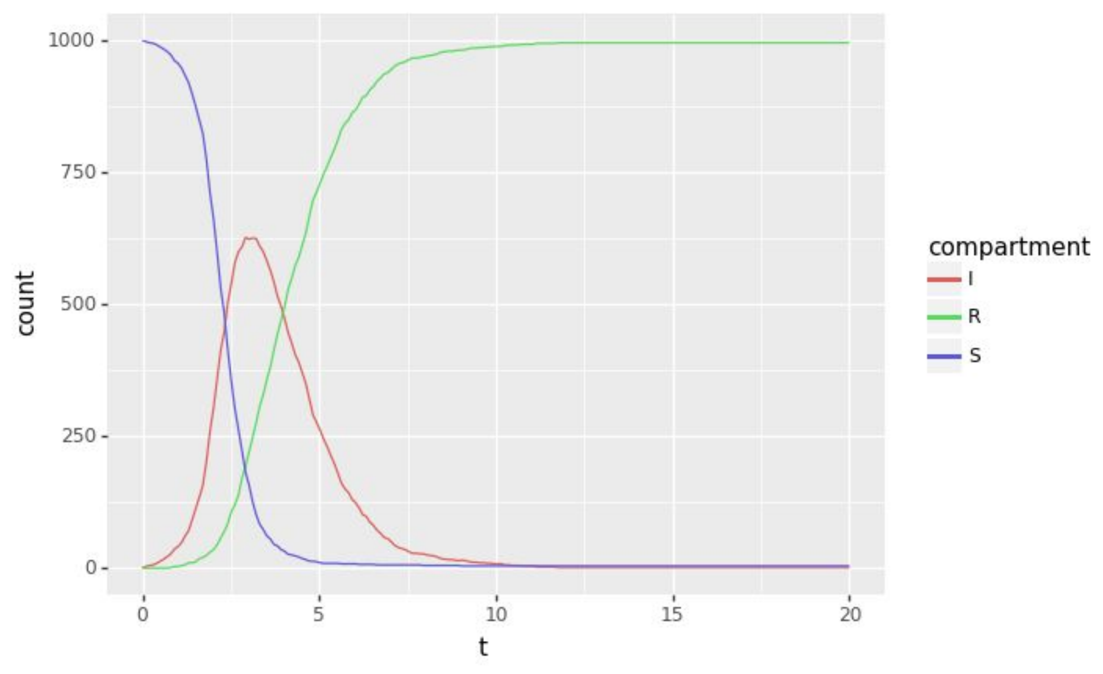
- Straightforward to increase complexity
  - Age, gender, risk behaviour
  - Contact networks (households, schools…)
  - Superspreading events
- High computational cost

A simple agent-based SIR model

github.com/jorgenem/smittemodellar

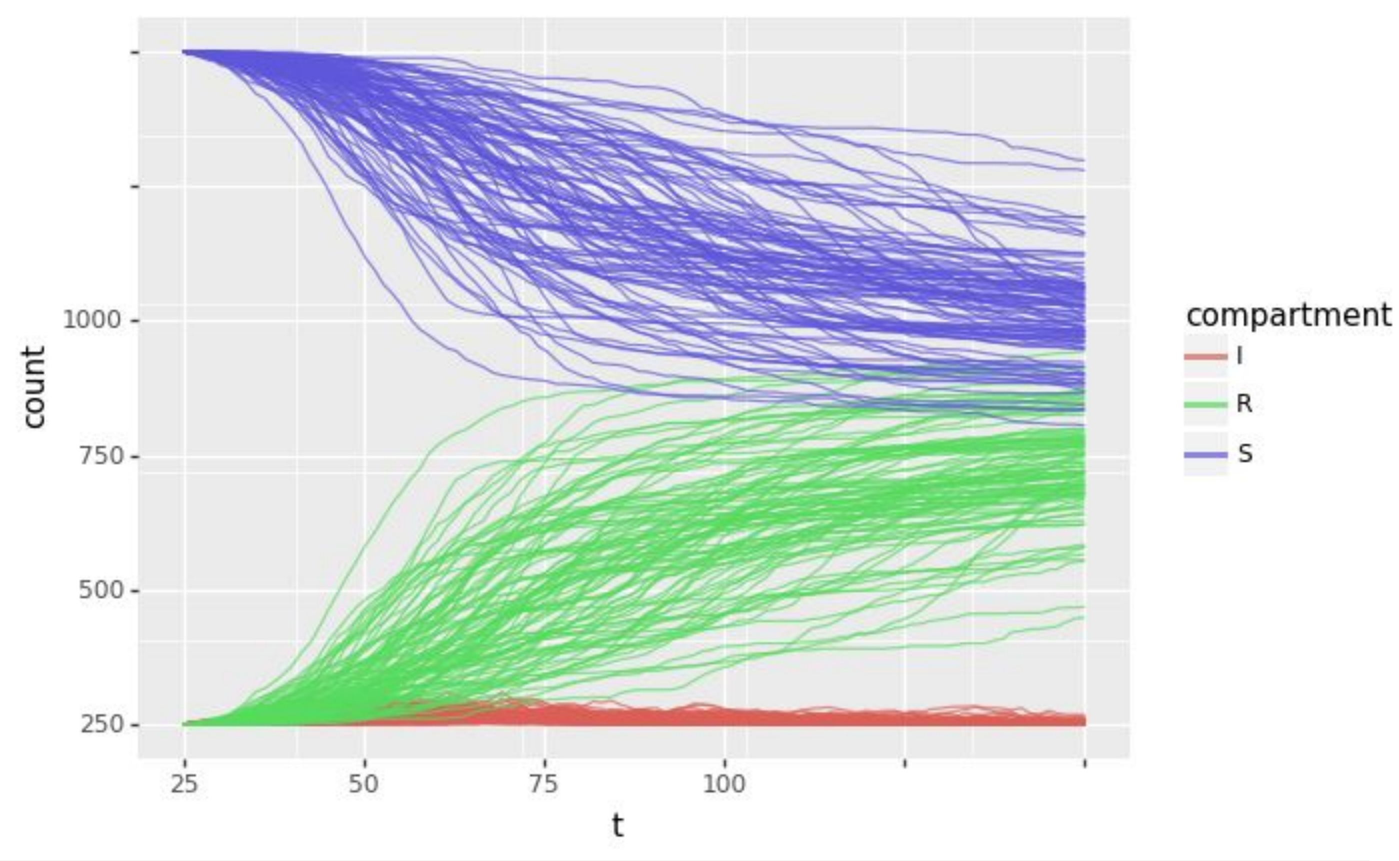
Assuming N individuals indexed i=0, …, N-1, each having status *S*, *I* or *R*

1. For each individual that is *I*, draw random other individuals to be their *contacts*
2. For each contact
   a. Check if they are *S*
   b. Draw a random yes/no whether they will be infected
   c. Update their status to *I* if yes
   d. Draw time of recovery from a distribution
3. Repeat at next timestep with updated *I* population



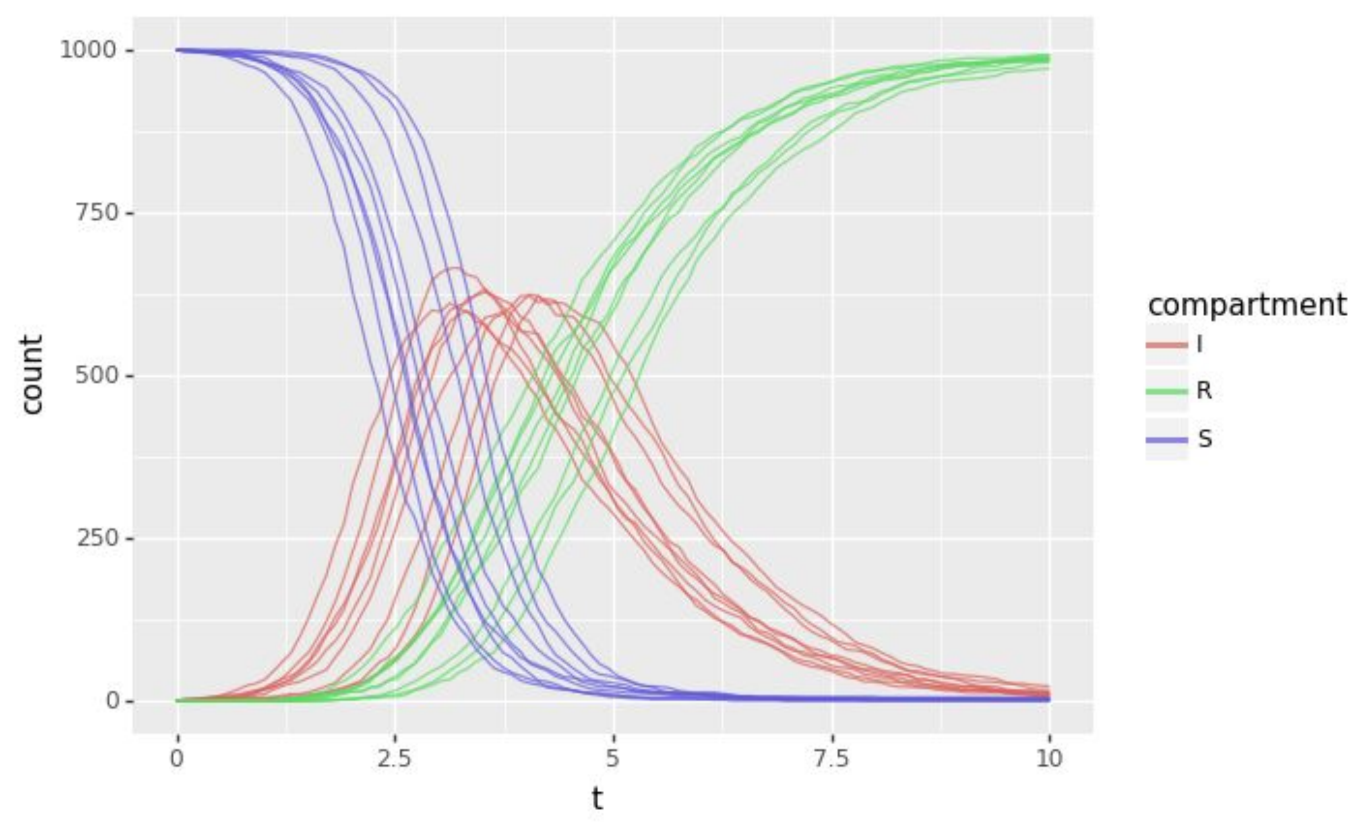
Stochastic effects emerge

- There is inherent randomness in the model
- Run it many times for each parameter value
- Infection seeding and superspreader events



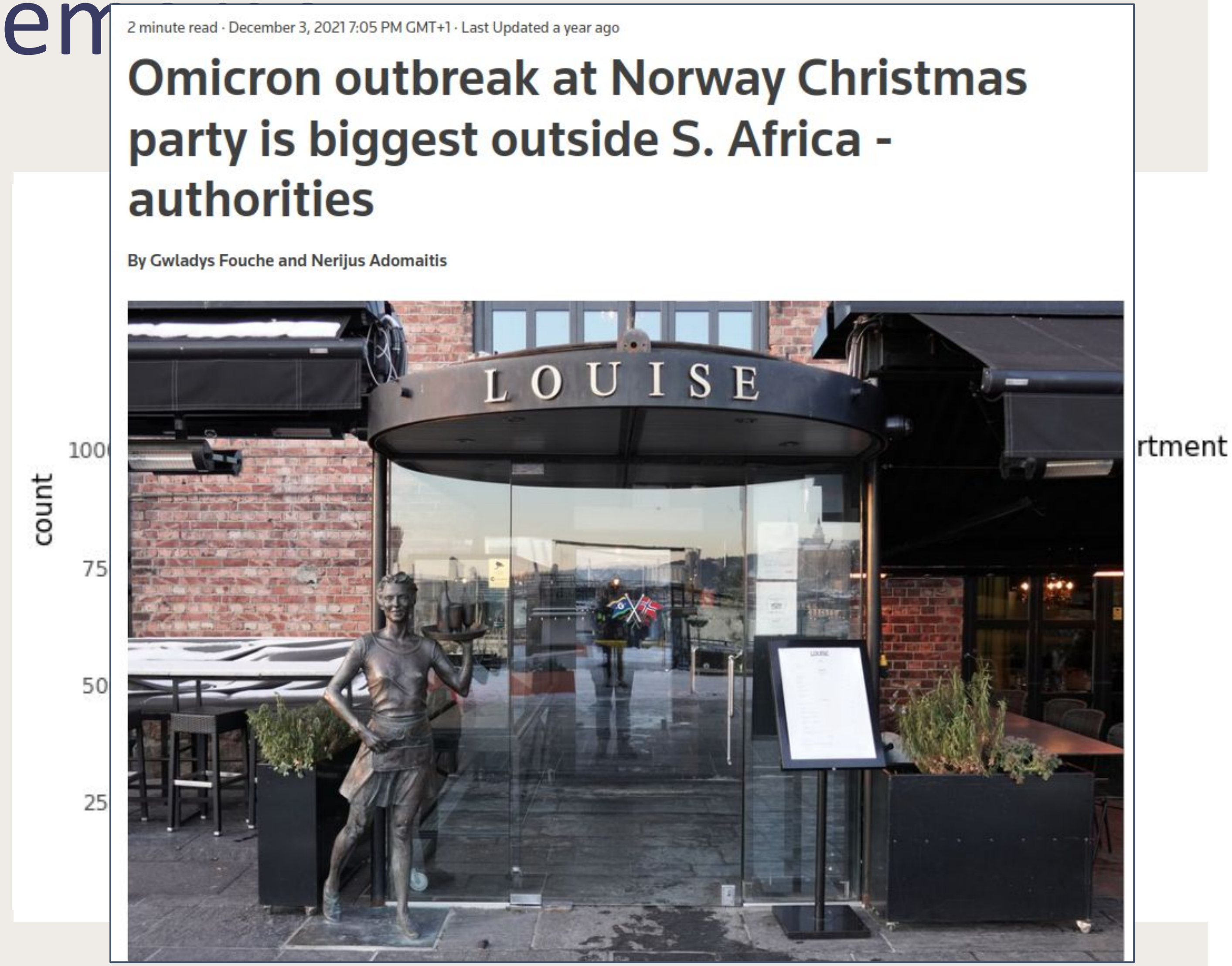
Stochastic effects emerge

- There is inherent randomness in the model
- Run it many times for each parameter value
- Infection seeding and superspreader events



Stochastic effects em

- There is inherent randomness in the model
- Run it many times for each parameter value
- Infection seeding and superspreader events



2 minute read · December 3, 2021 7:05 PM GMT+1 · Last Updated a year ago

Omicron outbreak at Norway Christmas party is biggest outside S. Africa - authorities

By Gwladys Fouche and Nerijus Adomaitis

Our big ABM at FHI

Individual features

- Location
- Age
- Occupation
- Epidemiological status
- Hospitalization status
- Vaccination status (SYSVAK)

Households

Synthetic, based on census data

Transmission settings

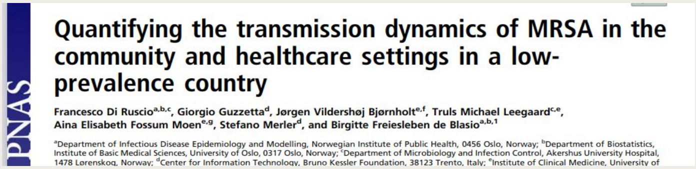
kindergartens
Schools, grades 1-13
Universities

Households

Workplaces

Community

One beta parameter for each setting



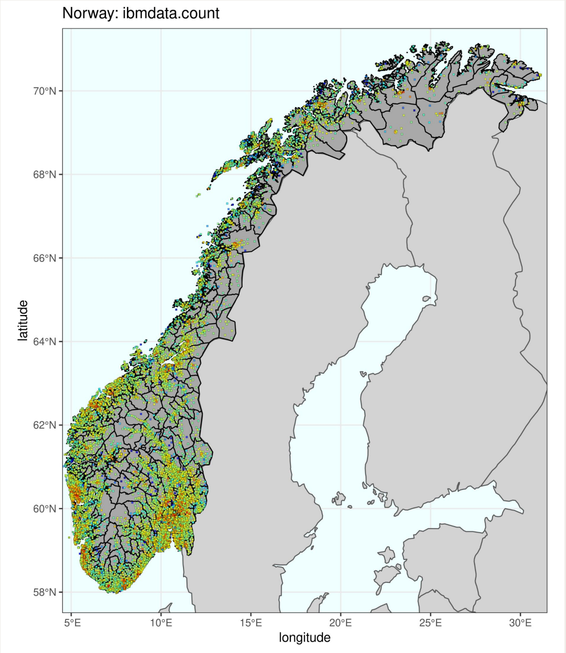
PNAS

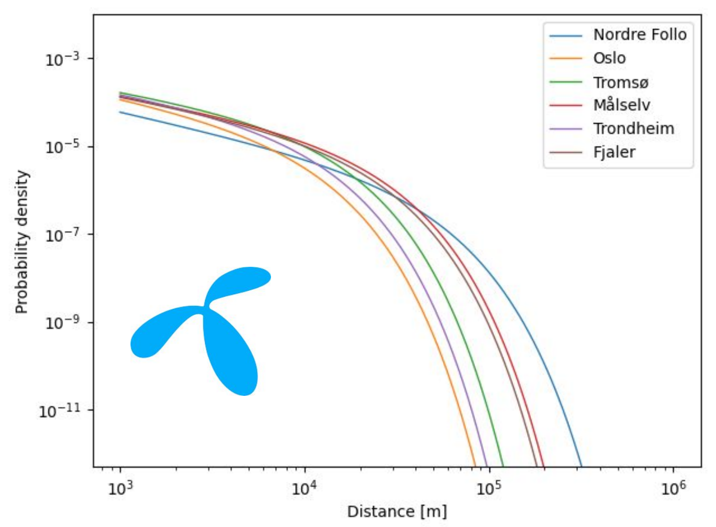
Quantifying the transmission dynamics of MRSA in the community and healthcare settings in a low-prevalence country

Francesco Di Ruscio[a,b,c], Giorgio Guzzetta[d], Jørgen Vildershøj Bjørnholt[e,f], Truls Michael Leegaard[c,e], Aina Elisabeth Fossum Moen[e,g], Stefano Merler[d], and Birgitte Freiesleben de Blasio[a,b,1]

[a]Department of Infectious Disease Epidemiology and Modelling, Norwegian Institute of Public Health, 0456 Oslo, Norway; [b]Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, 0317 Oslo, Norway; [c]Department of Microbiology and Infection Control, Akershus University Hospital, 1478 Lørenskog, Norway; [d]Center for Information Technology, Bruno Kessler Foundation, 38123 Trento, Italy; [e]Institute of Clinical Medicine, University of

Di Ruscio et al, PNAS 2019





Compartmental models

Compartmental models

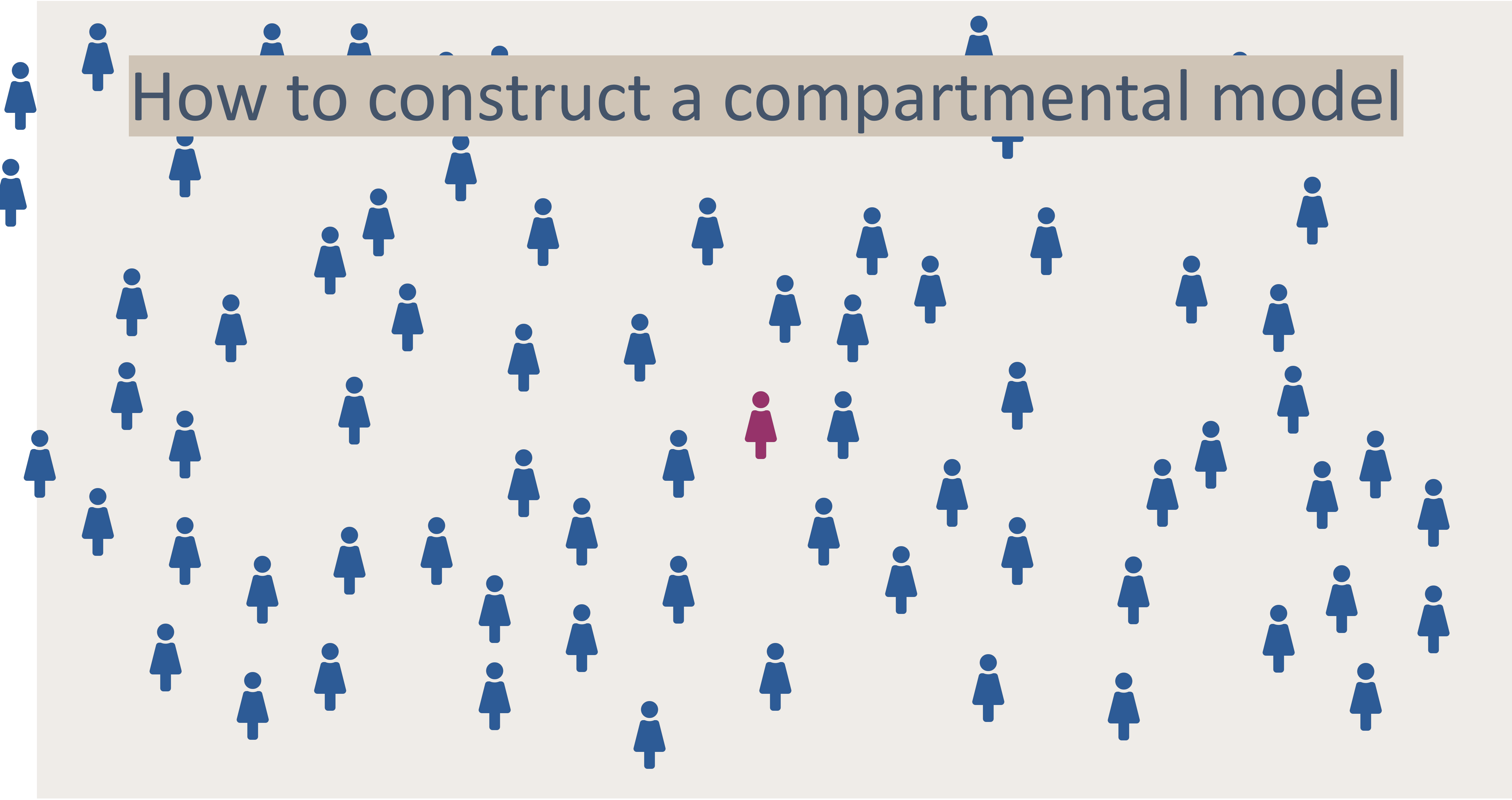
- “Mean-field models” (fancy word for average)
- Based on differential equations governing the behaviour of each *compartment (S, I and R)*
- Stochastic or deterministic

$$S' = -\beta\ S\ I$$

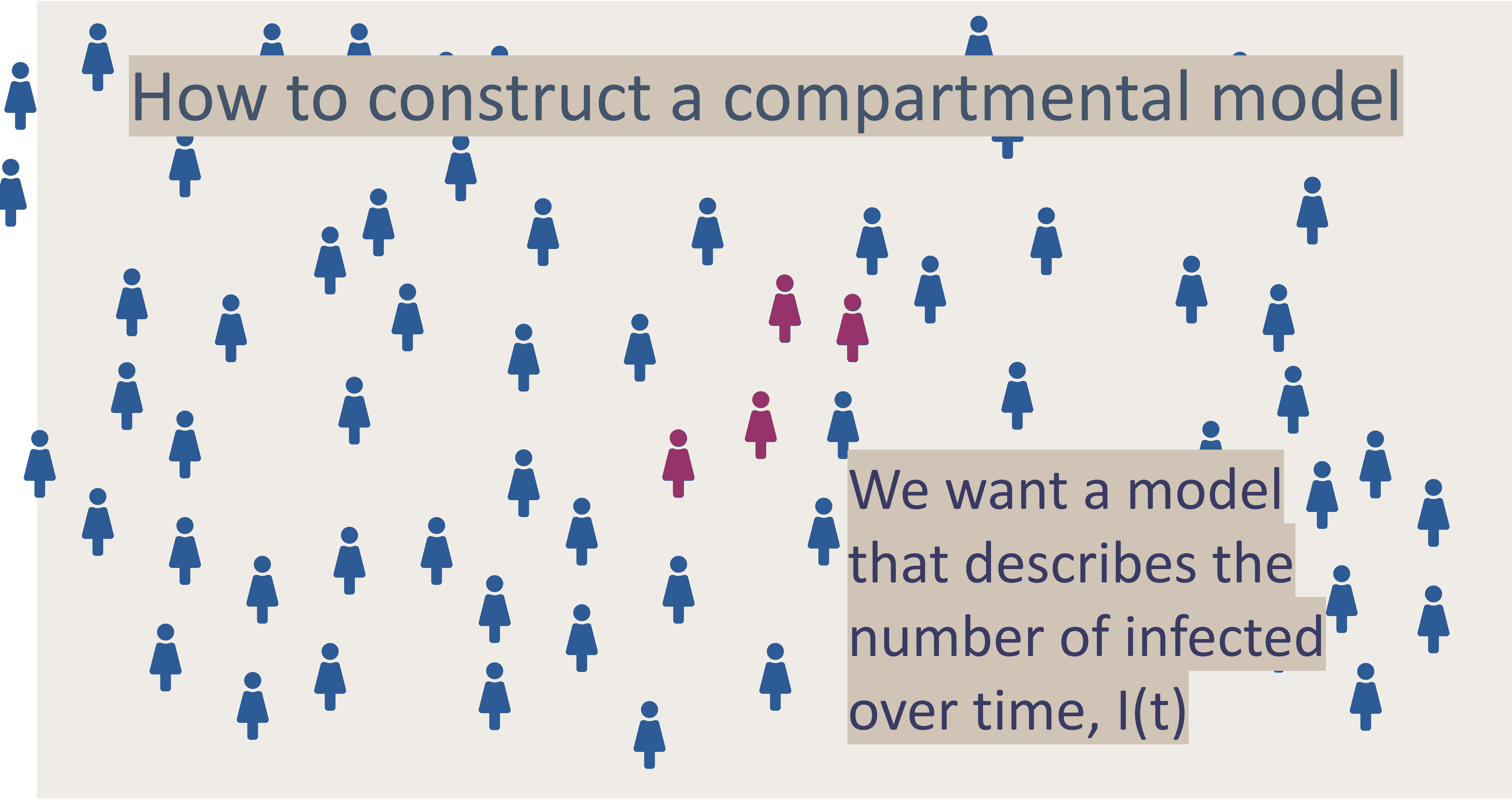
$$I' = +\beta\ S\ I$$
$$-\gamma\ I$$

$$R' = +\gamma\ I$$

How to construct a compartmental model



How to construct a compartmental model

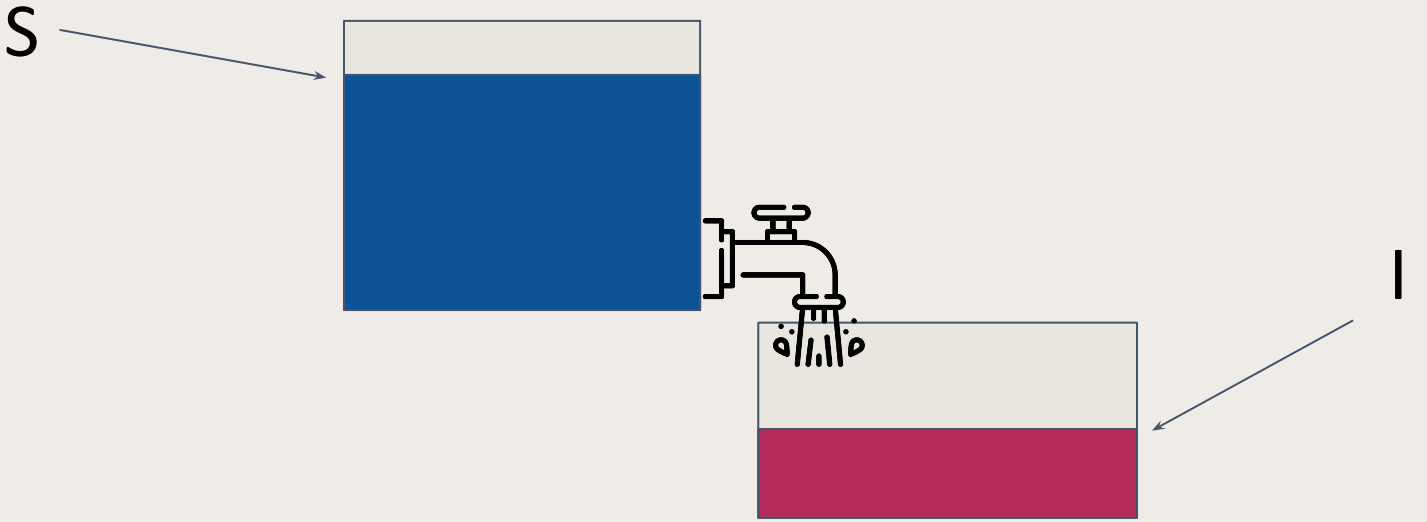


We want a model that describes the number of infected over time, I(t)

Simplify to average behaviour

How about modelling the population as a… tank of water?

And infections as a running tap?

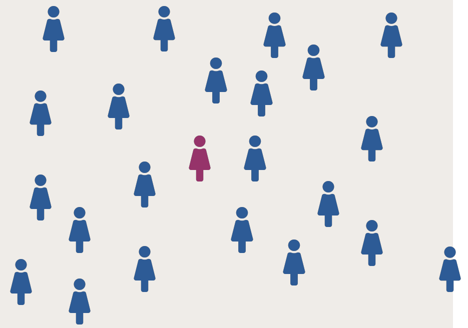


Change in I =

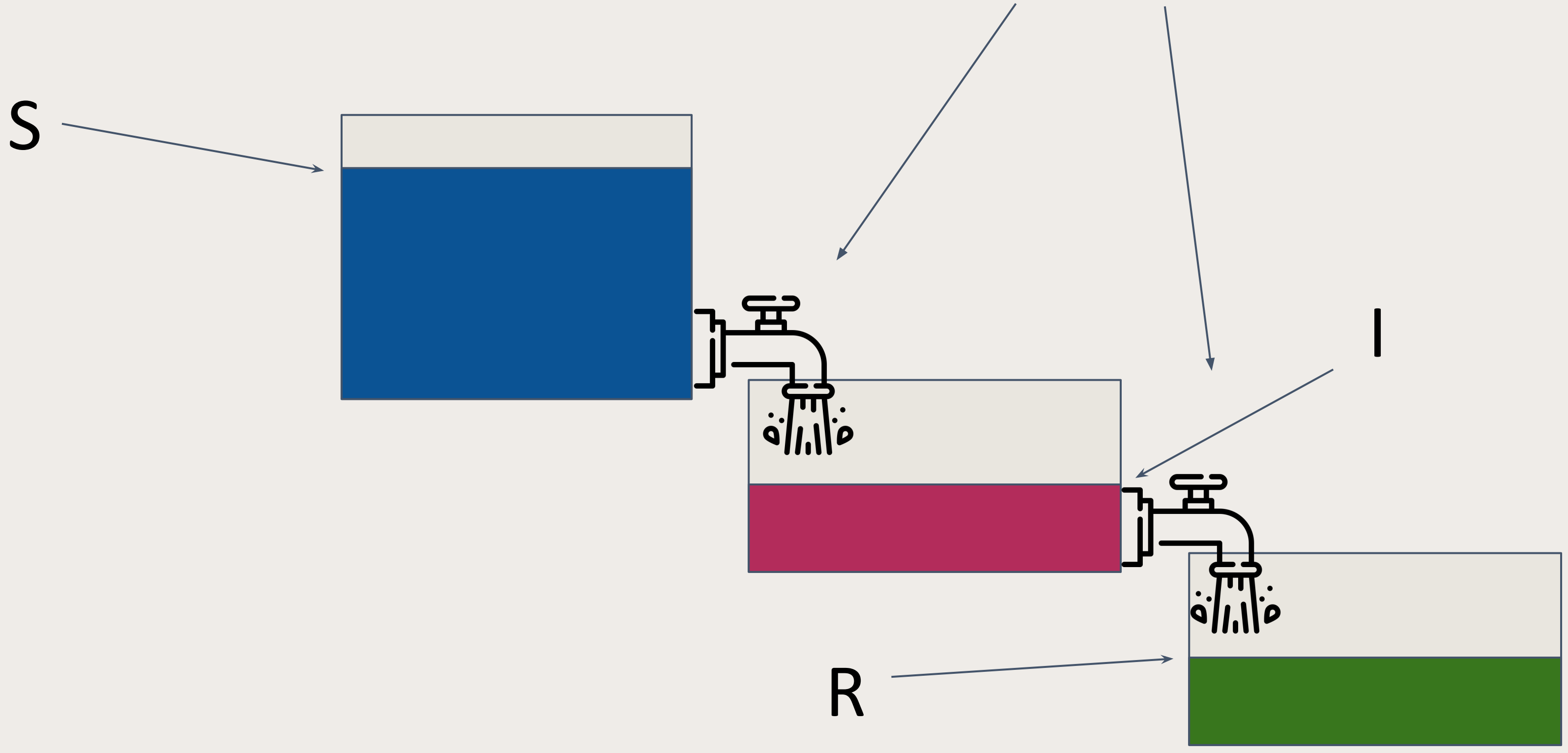
$$I' = \beta S I$$

Don’t forget recoveries

Don’t forget recoveries



Competition determined by virus's properties

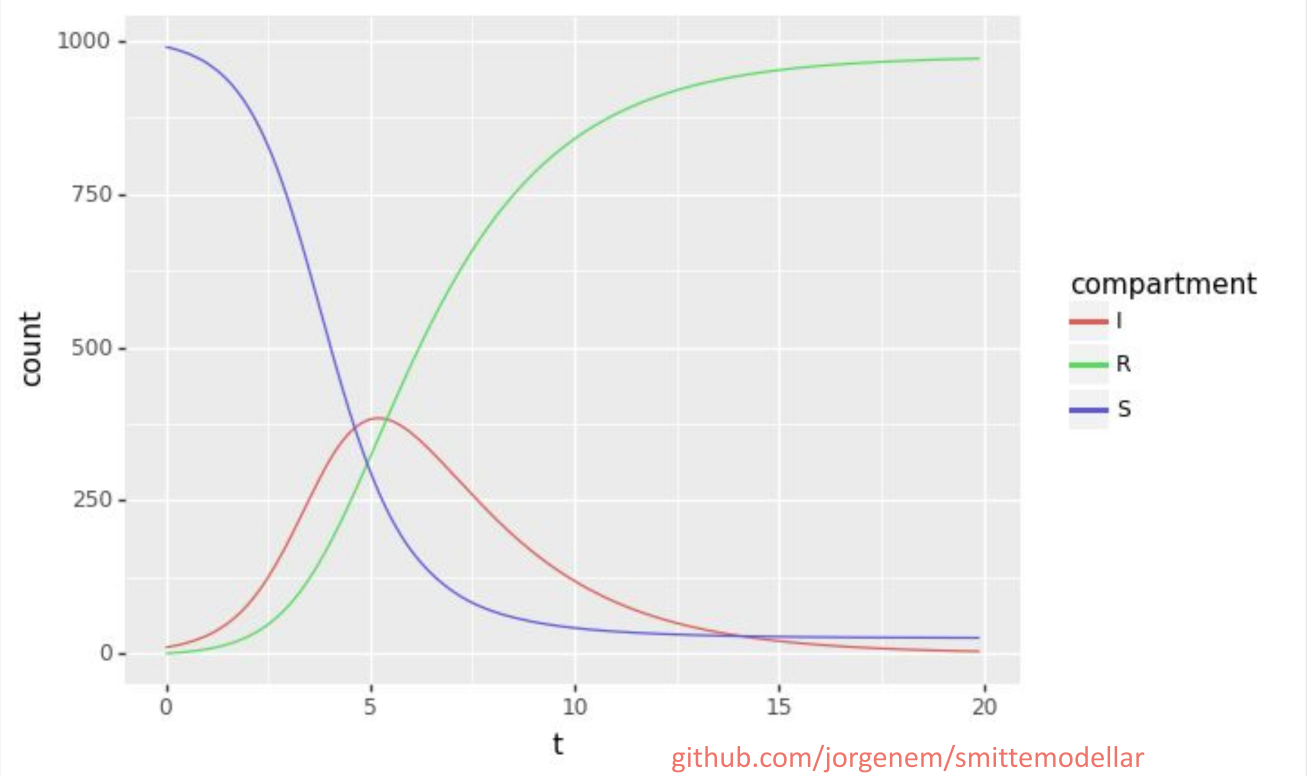


Change in I =

$$I' = \beta\ S\ I \; -\gamma\ I$$

This is th ompartmental SIR model

deterministic

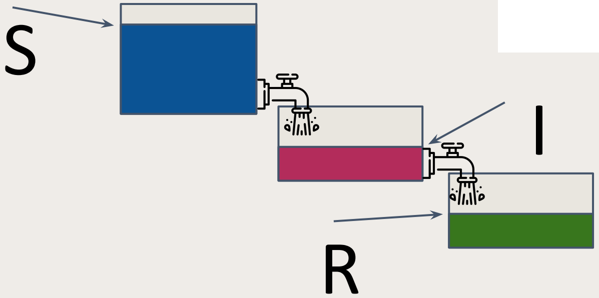


github.com/jorgenem/smittemodellar

$$S' = -\beta\, S\, I$$

$$I' = +\beta\, S\, I \\ -\gamma\, I$$

$$R' = +\gamma\, I$$

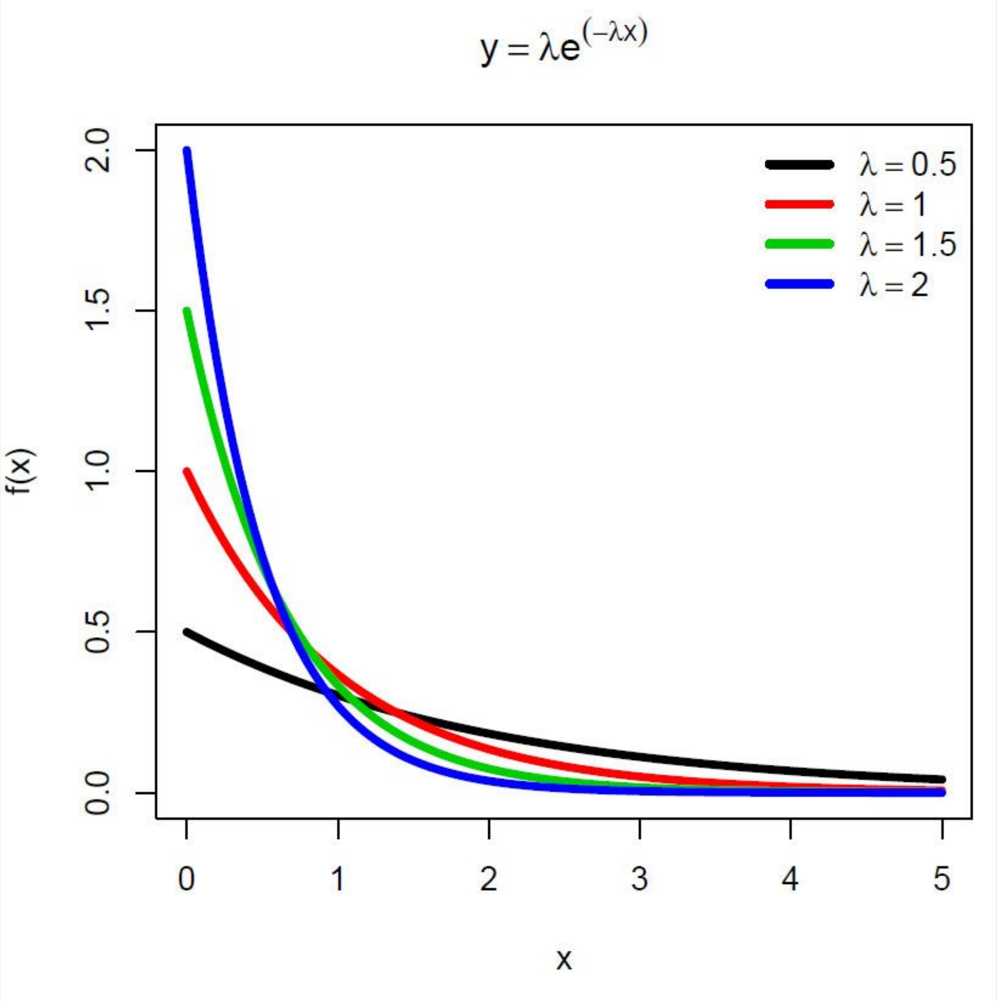


Beware of implicit assumptions

In addition to the explicit assumption of averaging over individuals, the equations bake in *implicit assumptions*. Importantly:

The transition times of individuals from S to I to R are ***exponentially*** *distributed*.

This is **wrong** for e.g. covid, but models still work fine on the big picture. If needed, techniques exist to amend by adding extra compartments.

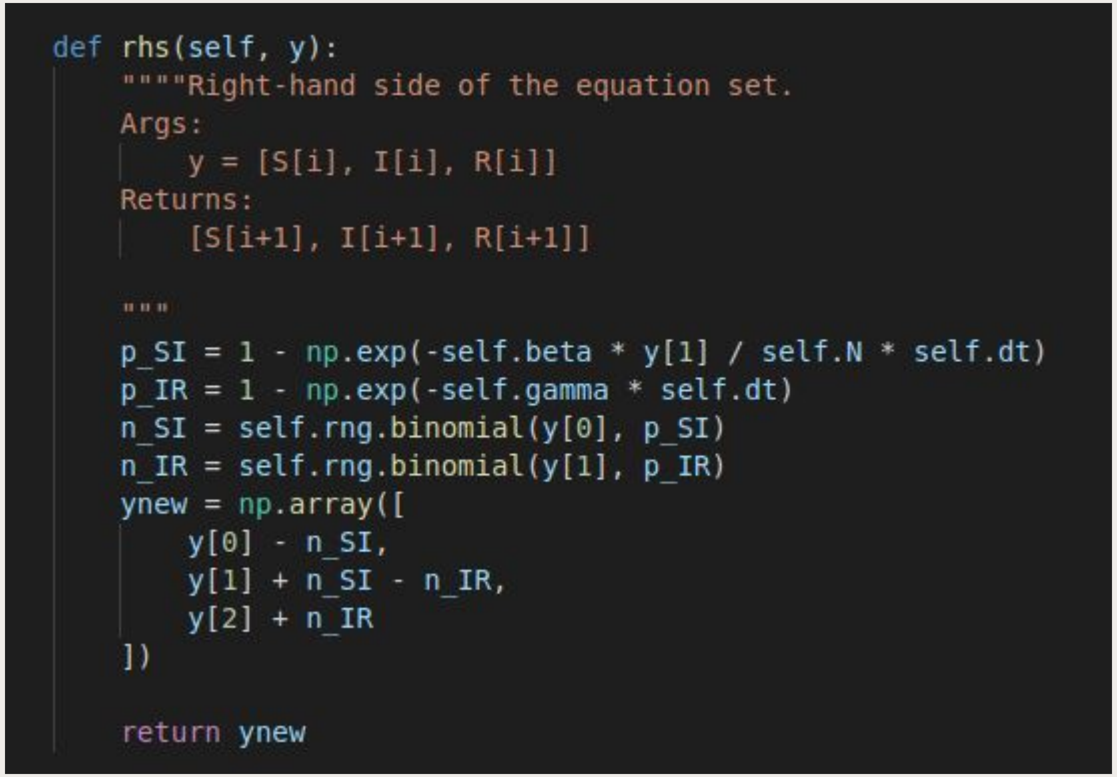


Stochastic compartmental models

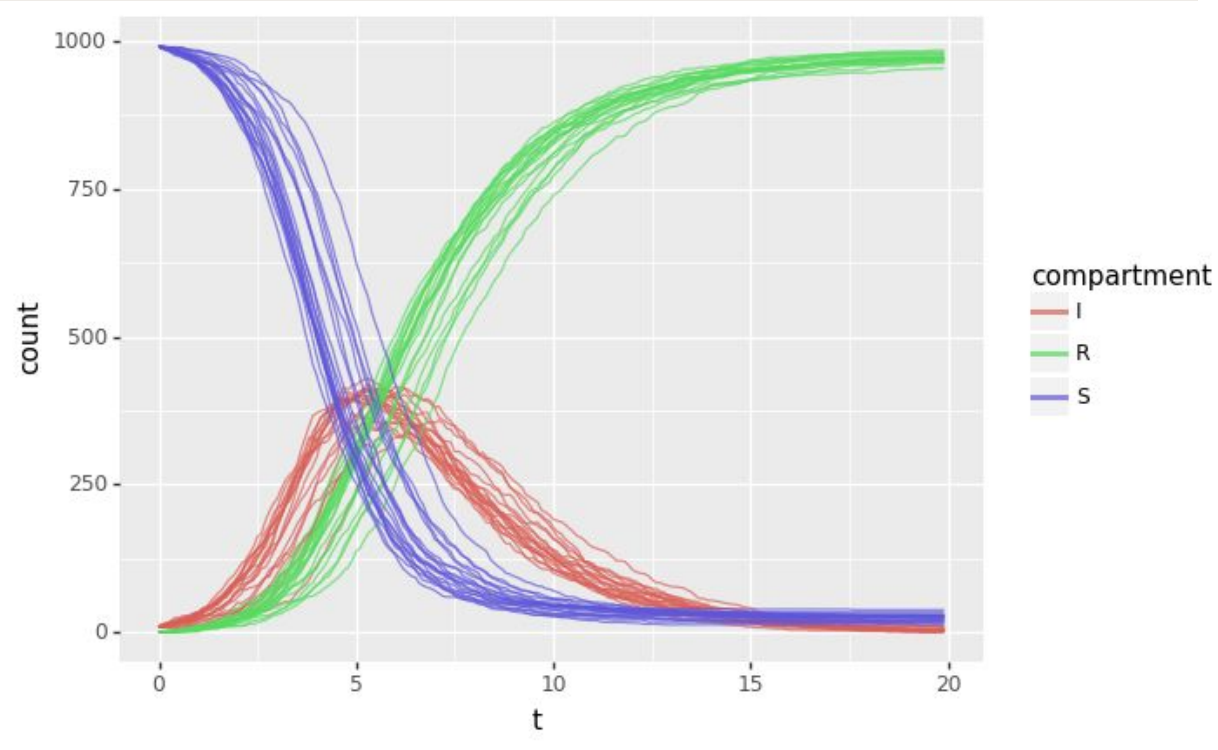
- Real world infectious disease spread is highly *stochastic*, i.e., random
  - As in the agent-based model
- Replaying the Wuhan outbreak might not have resulted in a pandemic -- certainly not the exact same one
- To mimic this, one can add stochasticity (randomness) to the compartmental model equations
- We then write a discrete-time difference equation with a random draw from a *binomial distribution* at each timestep:

$$I_{i+1} \propto \text{binom}(S_i, p_{indiv})$$

Stochastic compartmental models



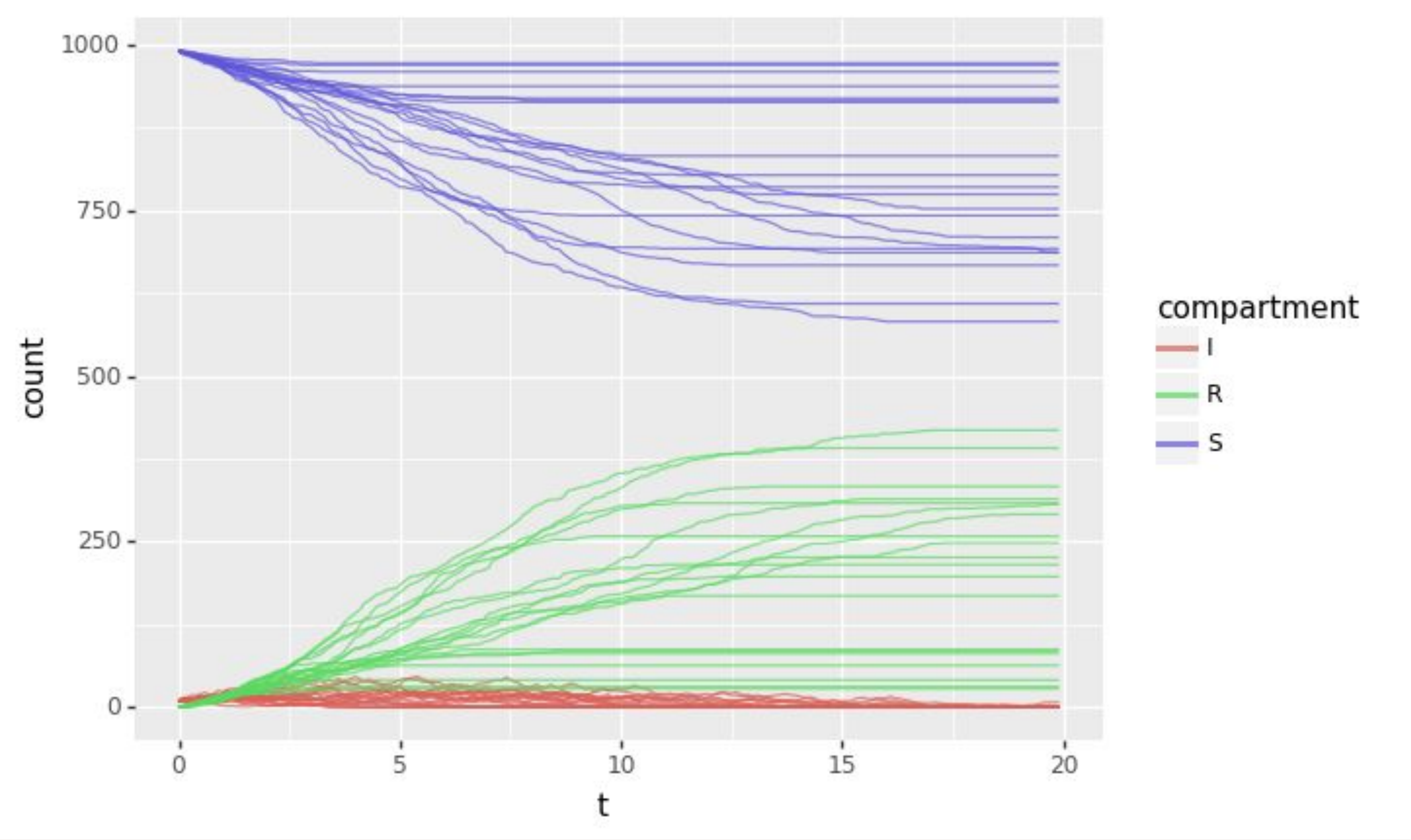
```
def rhs(self, y):
    """"Right-hand side of the equation set.
    Args:
        y = [S[i], I[i], R[i]]
    Returns:
        [S[i+1], I[i+1], R[i+1]]

    """
    p_SI = 1 - np.exp(-self.beta * y[1] / self.N * self.dt)
    p_IR = 1 - np.exp(-self.gamma * self.dt)
    n_SI = self.rng.binomial(y[0], p_SI)
    n_IR = self.rng.binomial(y[1], p_IR)
    ynew = np.array([
        y[0] - n_SI,
        y[1] + n_SI - n_IR,
        y[2] + n_IR
    ])

    return ynew
```



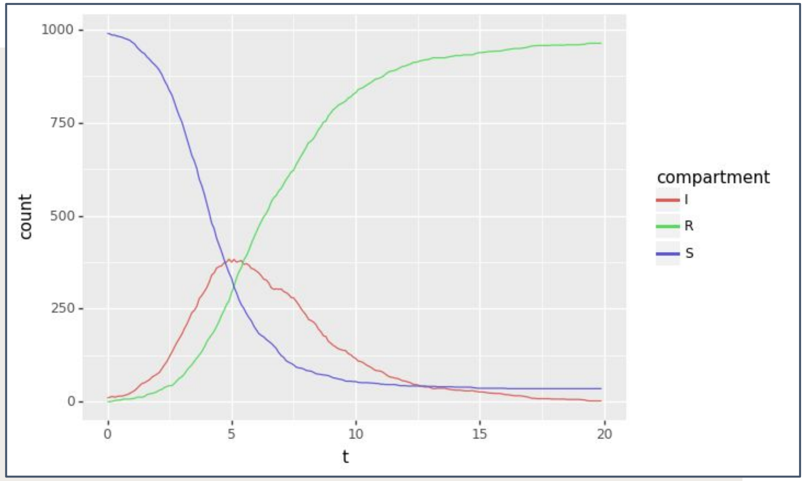
github.com/jorgenem/smittemodellar

Threshold effects

Does a seeded case lead to an outbreak? Initial randomness can decide.



Markov chain



Technically, this is by assumption a *markov chain*:

At each timestep *t*, the system is in a certain state described by the three numbers *[S(t), I(t), R(t)]*.

The state at timestep *t+1* depends solely on the state at *t*.

This has implications for what kind of solvers, approximate methods etc. are applicable.

Some final points

The tools we use in practice

- Not always cost-efficient to write code from scratch - pre-written libraries are extremely valuable!
  - Much can be done in Python these days
- Our ABM is written in C, from scratch, by my colleague Francesco di Ruscio. However, to infer its parameters we use a layer on top, written in R and leveraging statistical libraries
  - I might have been tempted to try Python with Numba or something if I was writing another one now
  - Don’t underestimate the value of a high-level language interface to your model
- For compartmental models, we often use a set of R libraries called *odin*, *dust* and *mcstate*. These allow specifying arbitrary models in R-like syntax, which are compiled to run fast and automatically interfaced with MCMC, SMC tools etc.
  See https://mrc-ide.github.io/mcstate/articles/sir_models.html
- TeX (since we have a critical mass of physicists++ - but nobody else understands…)
- Unix/linux
- Supercomputers and parallelisation (MPI and OpenMP)
- Git

Statistical inference

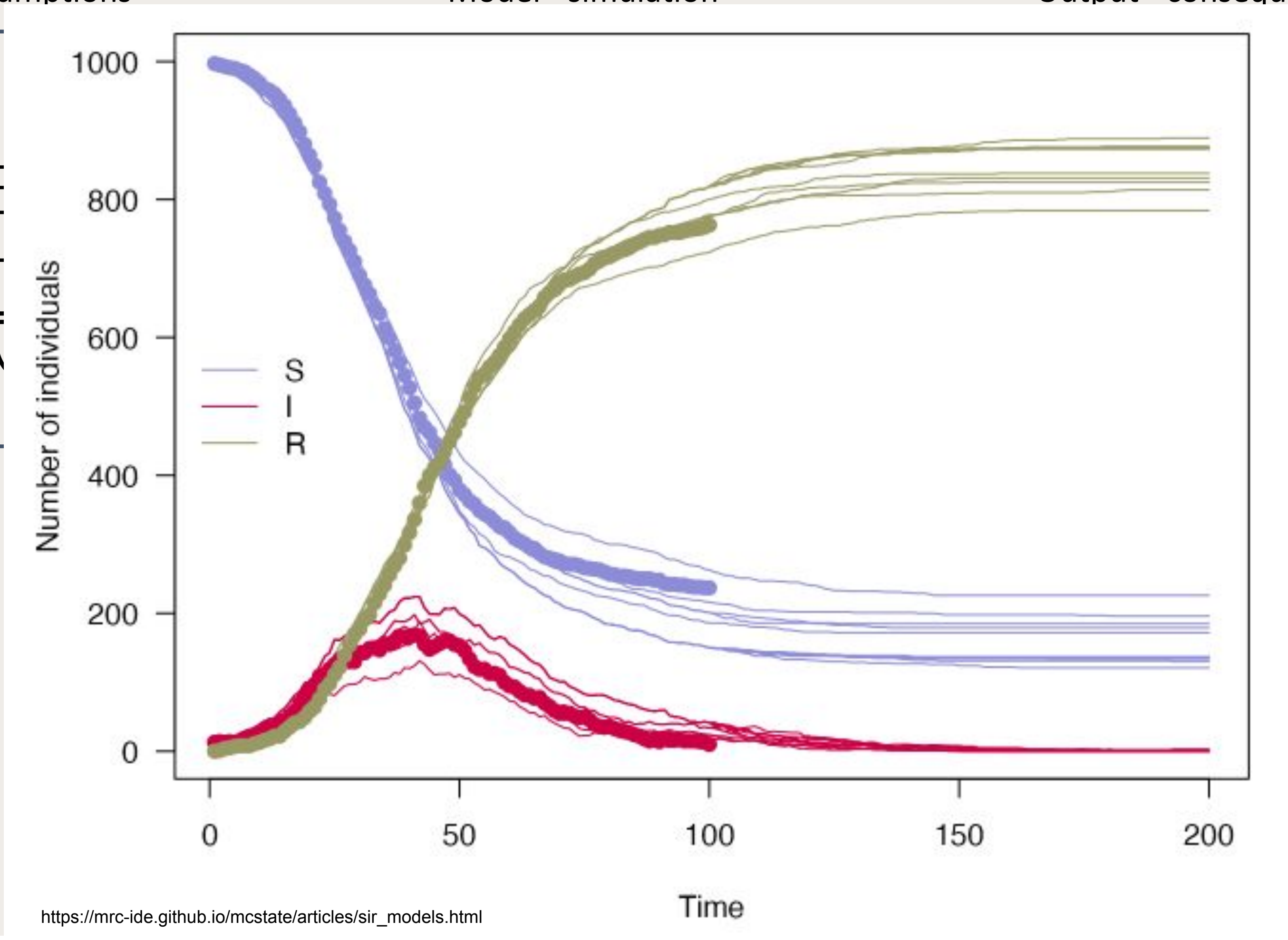
A model is nothing without data!

$$p(\theta|\mathcal{D}) \propto \mathcal{L}(\theta, \mathcal{D})\pi(\theta)$$

Input - assumptions

Model - simulation

Output - consequences



https://mrc-ide.github.io/mcstate/articles/sir_models.html

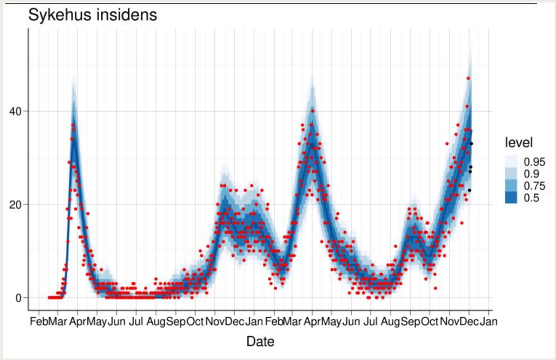
Illustrasjoner fra flaticon.com / Eucaly.p, Vitaly Gorbachev, itim2101

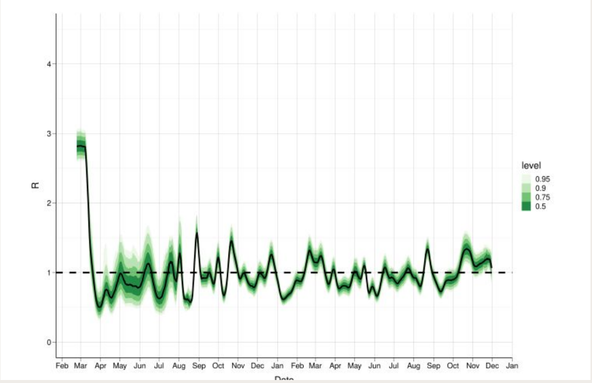
Calibration (fitting) techniques

All those we use are based on a Monte Carlo approach within a Bayesian framework

- Changepoints
- Approximate Bayesian Computation (ABC)
- Sequential Monte Carlo / “particle filter”
- Latin Hypercube Sampling (grid scan)
- MCMC
- HMC / NUTS [a fast MCMC, requires gradient] (Stan or pymc)
- History matching

https://www.fhi.no/contentassets/e6b5660fc35740c8bb2a32bfe0cc45d1/vedlegg/nasjonale-og-regionale-rapporter/national_regional_model_08_december_2021.pdf

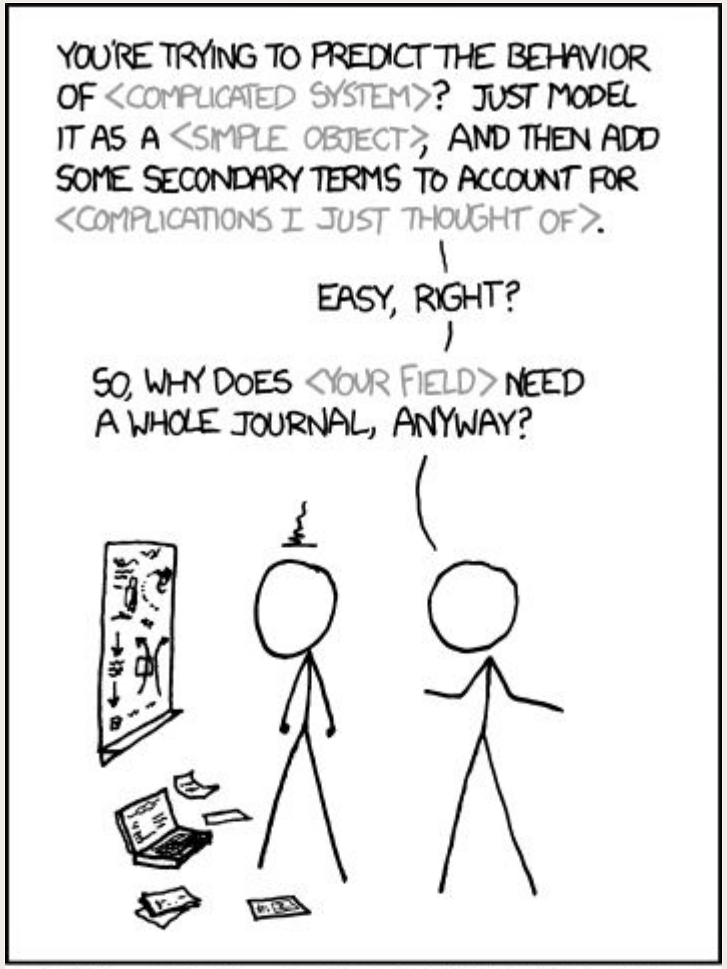




Working in a new field

Advice from a physicist to fellow physicists: Never underplay the importance of domain knowledge

- Physicists are **great** at modelling, computations and solving complex problems
- There are also many things we do not know, and it's important to be curious (and humble)


YOU'RE TRYING TO PREDICT THE BEHAVIOR OF <COMPLICATED SYSTEM>? JUST MODEL IT AS A <SIMPLE OBJECT>, AND THEN ADD SOME SECONDARY TERMS TO ACCOUNT FOR <COMPLICATIONS I JUST THOUGHT OF>.
EASY, RIGHT?
SO, WHY DOES <YOUR FIELD> NEED A WHOLE JOURNAL, ANYWAY?

LIBERAL-ARTS MAJORS MAY BE ANNOYING SOMETIMES, BUT THERE'S *NOTHING* MORE OBNOXIOUS THAN A PHYSICIST FIRST ENCOUNTERING A NEW SUBJECT.

Also, always remember:

All models are wrong, but some are useful.

-George Box