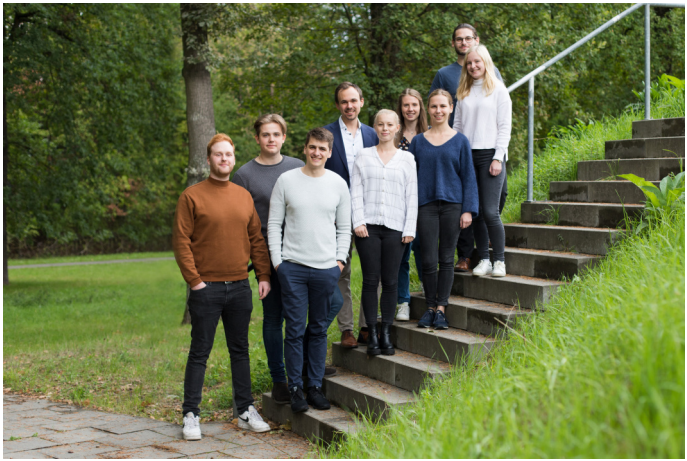
# MATHEMATICAL MODELING OF INFECTIOUS DISEASE DYNAMICS: AN OVERVIEW

**Martin J. Kühn**
**Gandalf - Graduiertenschule**
**Statusmeeting Lampoldshausen**

# Predictive Simulation Software



4 PhD Theses on models:

- Hybrid Graph-ODE
- Integro-DE
- Agent-based
- hybrid AB-EB

Other topics:

HPC, Data science,
Surrogate models, ...

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Gefördert durch:
Bundesministerium für Digitales und Verkehr

HDSLEE HELMHOLTZ School for Data Science in the Earth Sciences

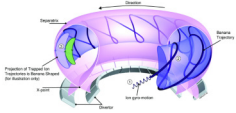
HELMHOLTZ

UNIVERSITÄT BONN

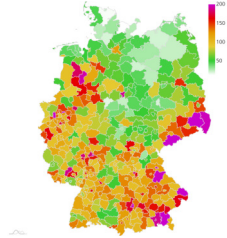
# Predictive Simulation Software

Deformation of heterogeneous materials

Plasma fusion




Infectious disease modeling




High performance computing

Image by rawpixel.com on Freepik

# Outline

- Introduction
- Mathematical modeling process
- Simple generation-based models
- Equation-based models
  - Ordinary differential equations
  - Integro differential equations
- Agent-based models
- Digression: Data Science
- Hybrid models / Spatial resolution
- Numerical assessment of the NoCovid strategy

# Introduction

# Some recent epi- and pandemics

- 2019: SARS-CoV-2 ($\geq$6.8m deaths[1])
- 2012: MERS-CoV ($\approx$1k deaths[2])
- 2009: Influenza A (150-575k deaths[2])
- 2002: SARS-CoV ($\approx$1k deaths[2])
- 1968: Influenza A ($\approx$1m deaths[2])

[1] Johns Hopkins University (accessed 02/2023), [2] Abdelrahman et al., Front. Immunol. (2021).

# IPBES Workshop Report on Biodiversity and Pandemics

Daszak et al. (2020), doi:10.5281/zenodo.4147317

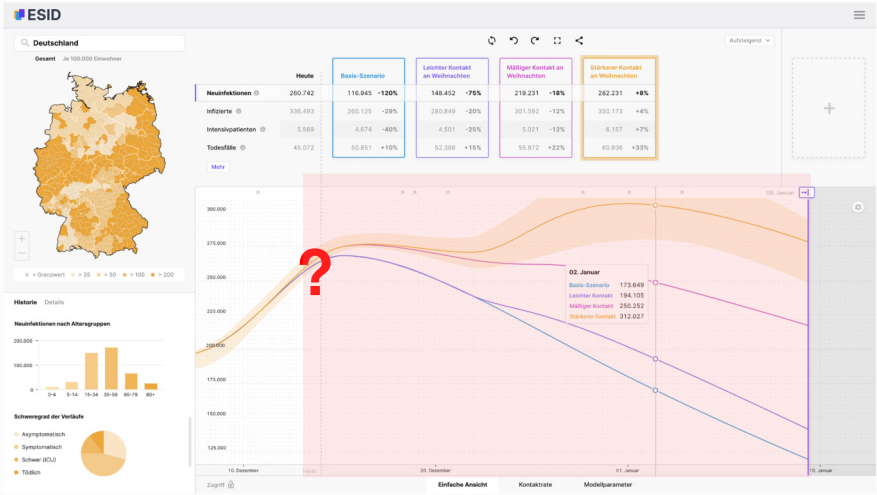
- “Without predictive and preventative strategies, pandemics will emerge more often, spread more rapidly, kill more people [...] with more devastating impact than ever”
- Estimation: More than 600 000 “undiscovered viruses” in “mammal and avian hosts [...] could have the ability to infect humans”
- The costs for prevention of pandemics are “trivial in comparison to the trillions of dollars of impact due to COVID-19, let alone the rising tide of future diseases.”
- “Reducing pandemic risks [...] would cost 1-2 orders of magnitude less than estimates of the economic damages caused by global pandemics”

# Motivation for infectious disease research

- Epidemics and pandemics are **no** “once in every 100 years” event
- The frequency of epidemics and pandemics could increase
- Endemic infectious diseases can still cause a large number of deaths and people suffering from the disease (with or without dying from it)
  - HIV, Malaria, and Tuberculosis account for 9k deaths eachs day[3]

[3] Brauer, Castillo-Chavez, Feng (2019)

# Motivation for infectious disease modeling



# Infectious disease dynamics: Questions

If one (or several) persons in a population get infected (by a particular pathogen)

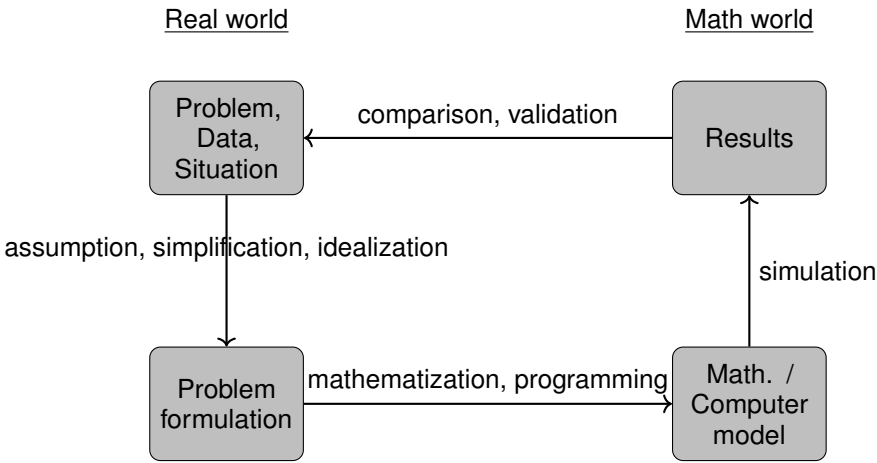
- does this cause an epidemic / increase?
- What is the rate of the increase?
- What will be the total number of infected (final size)?
- What is the effect of interventions?

# Infectious disease dynamics: Approaches

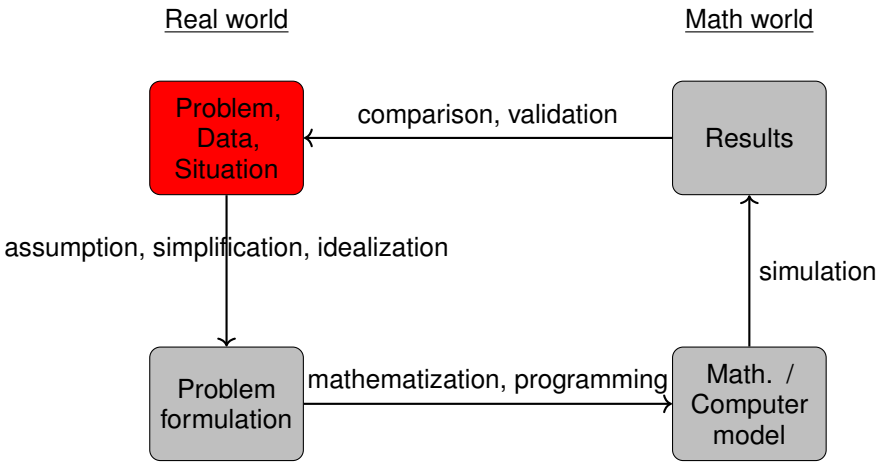
- Real life experiments not feasible
- Knowledge from other situations might be difficult to transfer directly
- Theoretical or mathematical models can help to gain insight
- Use of modern computers allows to consider detailed models and a lot of scenarios, e.g.,
  - home-office ratio of 10 %,
  - vs home-office ratio of 30 %,
  - closure of X,
  - vs closure of Y,
  - ...

# The mathematical modeling process on an example

# Schematic view of mathematical modeling process



# Schematic view of modeling process: Problem, Data, Situation



# Examples of Problem, Data, Situation

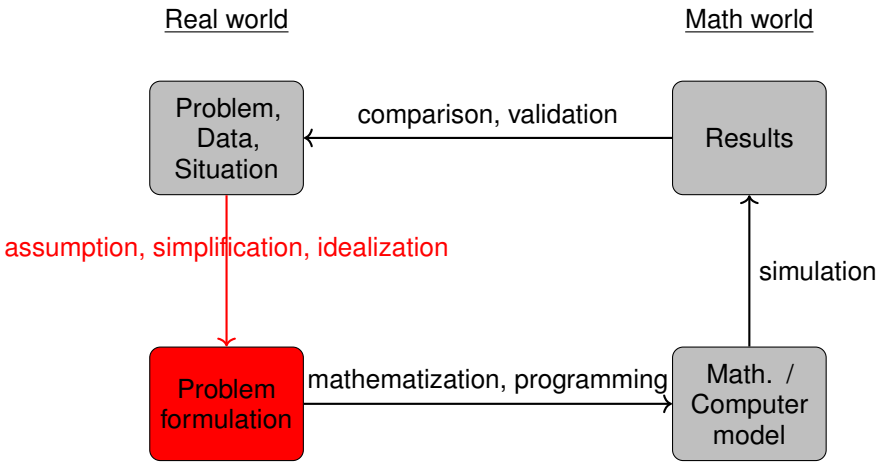
**Sars-CoV-2 in Europe in beginning of 2020:**

- A new pathogen appeared on scene
- Some confirmed cases in some regions

| | | | | Confirmed | Deaths | Recovered |
|---|---|---|---|---|---|---|
| ID_County | County | Gender | Date | | | |
| 1001 | SK Flensburg | female | 2020-01-13 | 1 | 0 | 1 |
| | | | 2020-03-10 | 1 | 0 | 1 |
| | | | 2020-03-11 | 1 | 0 | 1 |
| | | | 2020-03-12 | 1 | 0 | 1 |
| | | | 2020-03-14 | 2 | 0 | 2 |
| | | | 2020-03-15 | 1 | 0 | 1 |
| | | | 2020-03-17 | 2 | 0 | 2 |
| | | | 2020-03-19 | 1 | 0 | 1 |
| | | | 2020-03-20 | 1 | 0 | 1 |
| | | | 2020-03-21 | 1 | 0 | 1 |
| | | | 2020-03-22 | 1 | 0 | 1 |
| | | | 2020-03-26 | 2 | 0 | 2 |
| | | | 2020-03-31 | 1 | 0 | 1 |
| | | | 2020-05-17 | 1 | 0 | 1 |
| | | | 2020-05-21 | 3 | 0 | 3 |
| | | | 2020-05-22 | 1 | 0 | 1 |
| | | | 2020-05-28 | 1 | 0 | 1 |
| | | | 2020-06-04 | 1 | 0 | 1 |
| | | | 2020-06-06 | 1 | 0 | 1 |

- Some knowledge about the transmission as they appeared

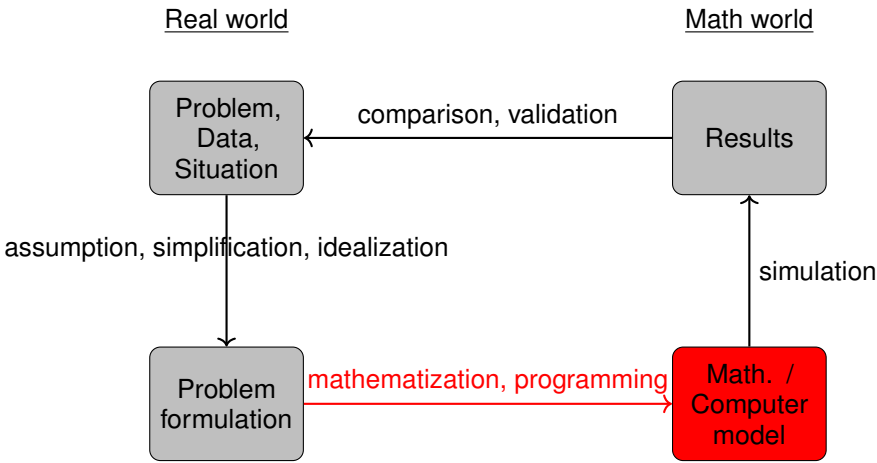
# Schematic view of modeling process: Problem formulation



# Problem formulation and examples of assumption and simplification

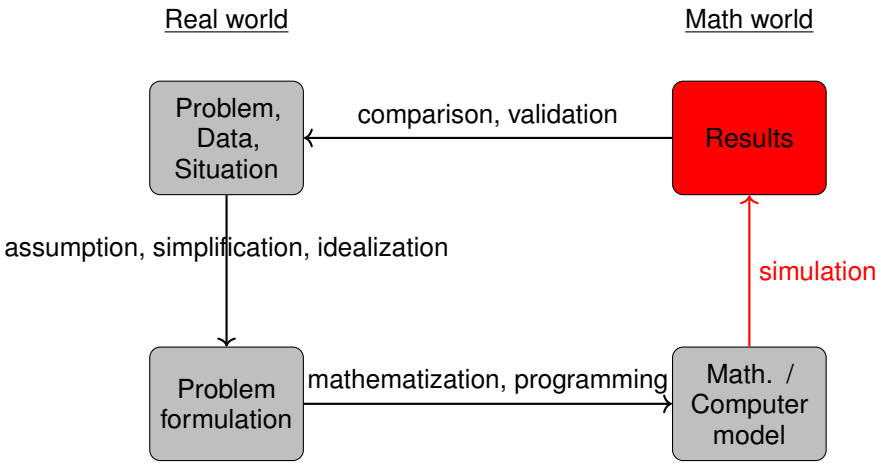
- A “contact” leads to transmission with some probability $\rho$
  $\rightarrow$ Needs a definition (simple physical, exposure through air, sexual, waterborne, ...)
- all people are equally susceptible
  $\rightarrow$ neglect that (cross-)immunity could reduce risk
- ...

# Schematic view of modeling process: Math. / Computer model



- A person is described as an *agent*
  - It has features such as age or infection state
  - We can set and ask if the person is in quarantine
  - ...
- For every infection state, we need to estimate the time a person is in this state and translate this into a parameter
- Let $X \sim \mathcal{U}[0, 1]$ be uniformly distributed. For every contact of a person $P$ with an infected person, we draw a sample $x_i$ from $X$. If $x_i \leq \rho$, the virus gets transmitted.
- ...

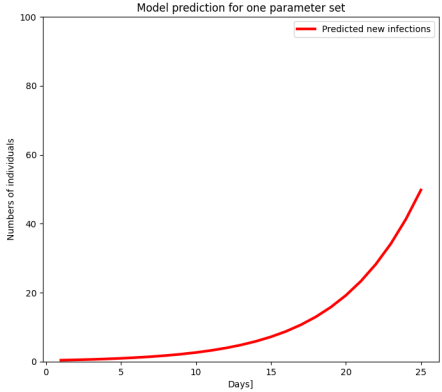
# Schematic view of mathematical modeling process: Results

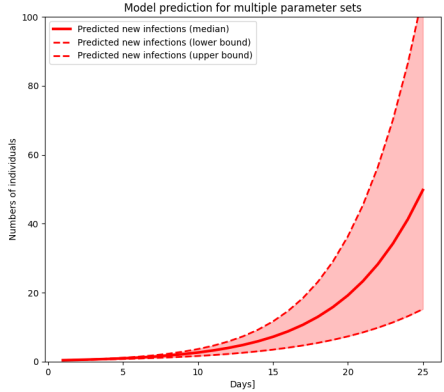


# Results and examples of simulation

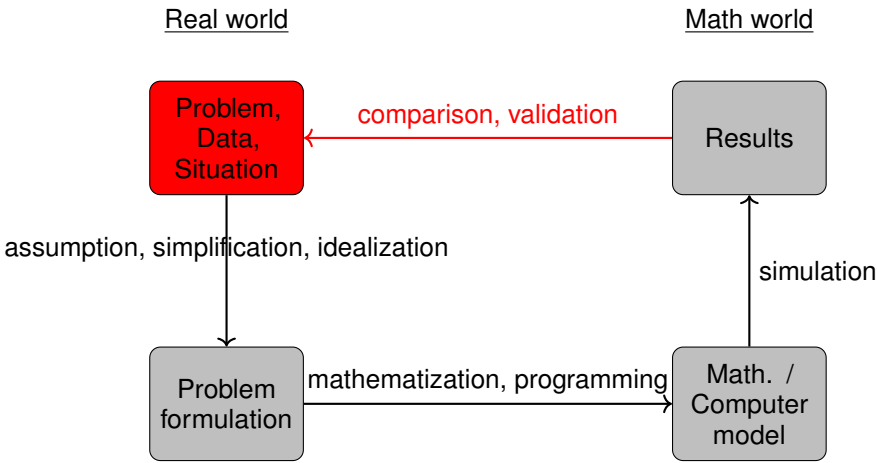
- Given a set of parameters, one can compute the outcome of the model
- Outcomes are often computed as approximations using a computer and not by an analytical solution (which may be difficult or impossible to obtain)
- Input parameters are uncertain and uncertainty in the input leads to uncertainty in the output (although often with different quantification)
- Simulations with multiple sets of parameters can assess uncertainty in the prediction

# Results and examples of simulation



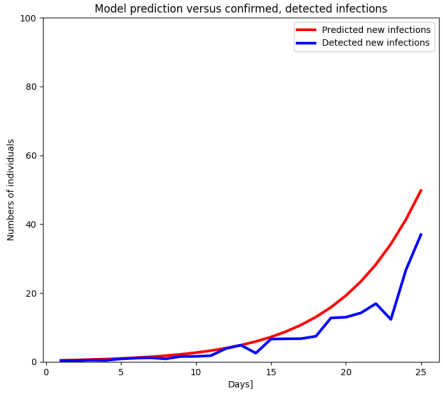



# Schematic view of mathematical modeling process: Validation

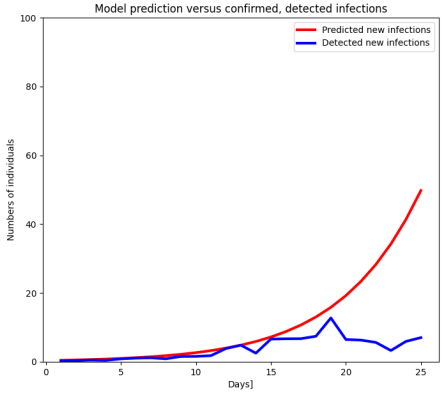


# Examples of comparison, validation

Scenario A:




Scenario B:



# Examples of comparison, validation

- Problems of underdetection, delayed reporting, week-end effects etc. in real data
- Changed real world behavior has to be reflected in model runs
- ...

$\rightarrow$ start anew and adapt simplifications, assumptions, model, parameters, ...

# Classications for infectious disease dynamics models

# Introduction to infectious disease dynamics: Basic notation

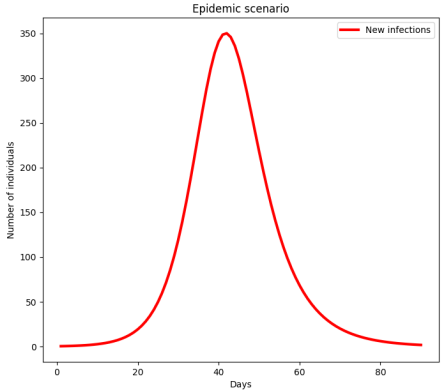
- For all models, we introduce discrete *infection states* or *compartments*
- Notation: ”Infection states“ can be states free of any infection (e.g., *susceptible* to the virus or *recovered* from the virus)!
- All individuals (or shares of the population) will get assigned a unique infection state.
- The (research) question defines which model states need to be considered.
- Minimal set of infection states
  - *Susceptible*: Persons that are susceptible to get the virus if they ”get in contact”
  - *Infected (Infectious)*: Persons that have the virus and can transmit it to susceptibles

# Introduction to infectious disease dynamics: Model classifications

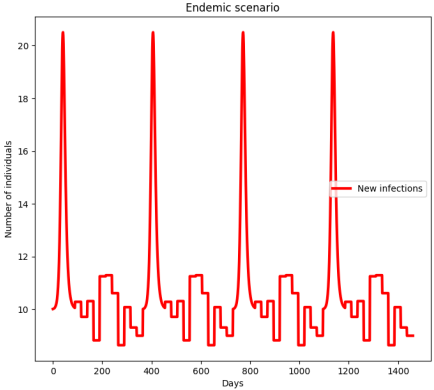
- Many models can be classified according to different ”dimensions“
  - Epidemic and endemic
  - generation-based and real-time
  - deterministic and stochastic
  - population-based and individual-based
  - ...
- Also hybrid, combined models are possible, then model parts might be classified accordingly

# Epidemic and endemic: A brief motivation

## Epidemic




## Endemic



# Epidemic and endemic: A brief motivation

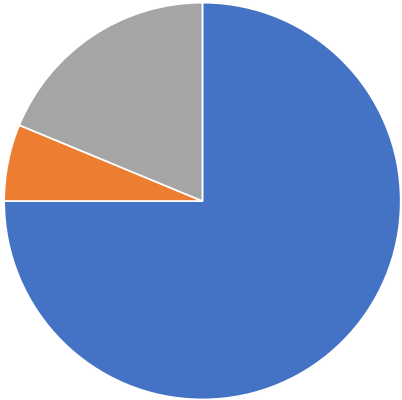
**Epidemic**

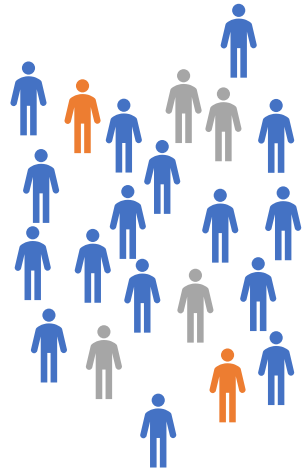
- ”short“ time horizons
- immigration, births and non-disease-related deaths less influential $\rightarrow$ *closed* population
- single steady state: disease free

**Endemic**

- ”large“ time horizons
- immigration, births and non-disease-related deaths can become important
- multiple steady states possible

# Population-based and individual-based: A brief motivation





# Population-based and individual-based: A brief motivation

**Population-based**

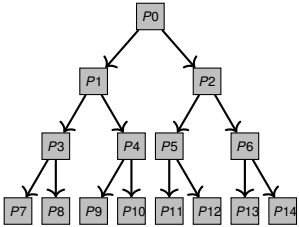
- averaging effects over (sub)populations
- homogeneous mixing inside (sub)populations
- established methods for analysis
- low computational effort

**Individual-based**

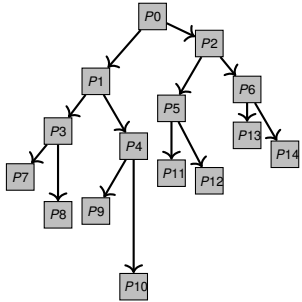
- easy modeling of individuals effects
- modeling micro granularity such as particular households
- tracing of transmission possible
- high computational effort

# Generation-based and real-time: A brief motivation

## Generation-based

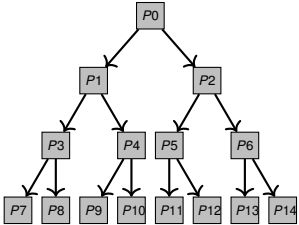



## Real-time

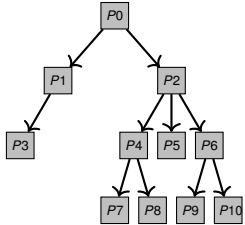


# Deterministic and stochastic: A brief motivation

## Deterministic




## Stochastic



# Deterministic and stochastic: A brief motivation

**Deterministic**

- explores average effect

**Stochastic**

- can explore full set of potential outcomes (if/if not, best/worst case)

# Primary and secondary cases

- For investigating infectious disease spread, we introduce the *primary cases* as the infected individuals related to the current outbreak
- The questions is how many *secondary cases*, i.e., new infections are induced by the primary cases
- We define the *basic reproduction number* as

$$R_0 := \begin{array}{c}\text{expected number of secondary cases per primary case} \\ \text{if all individuals are ”immune-naive“}\end{array} \tag{1}$$

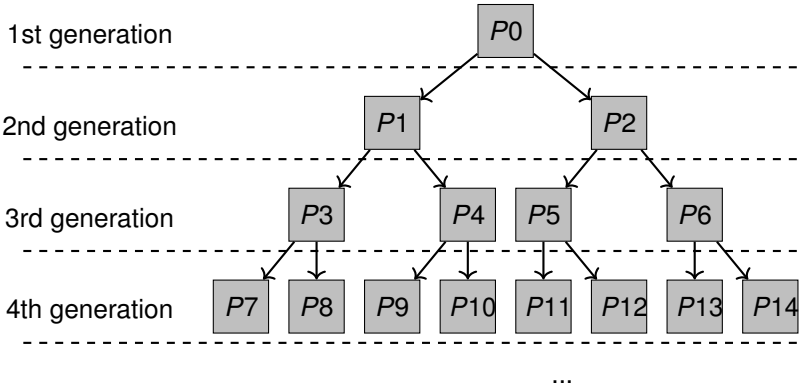
# Simple generation-based models

# Assumptions and basic parameters

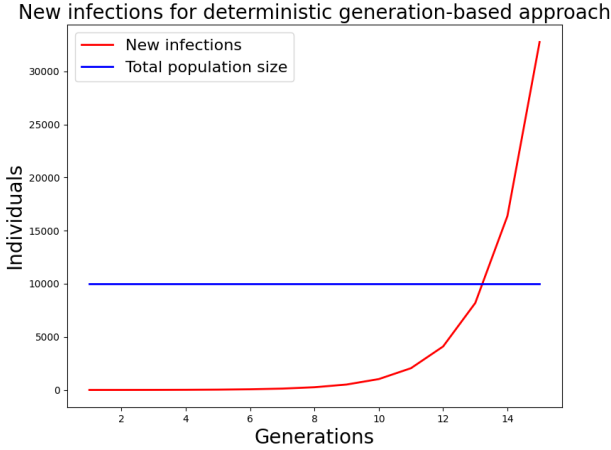
- Assume a homogeneous population
  $\rightarrow$ All parameters are averaged parameters (if not stated differently).
- assume that the number of susceptibles is so large that we can neglect the effect of non-susceptible persons in the population
  $\rightarrow S \approx N$ ($S$: susceptible persons, $N$: total population)
- $\phi$: contact rate: the number contacts per generation
- $\rho$: the transmission risk when ”having a contact“ with an infected person

**Example of a tree of infections**

For $\phi = 3$ and $\rho = 2/3$, we have the following transmission chain from one infected person $P0$ (only visualizing contacts where transmission occurs here)



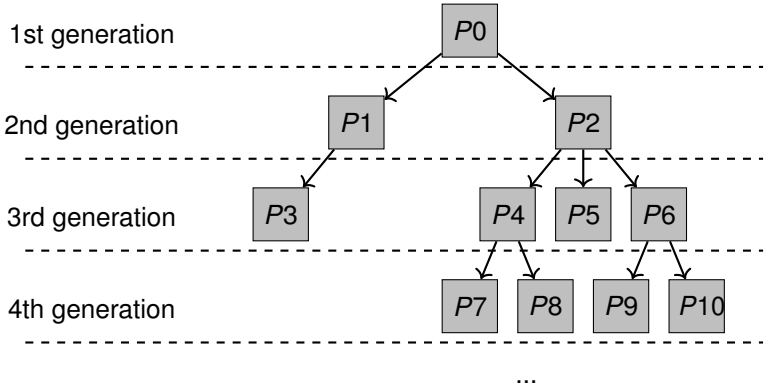
# Deterministic generation-based reproduction



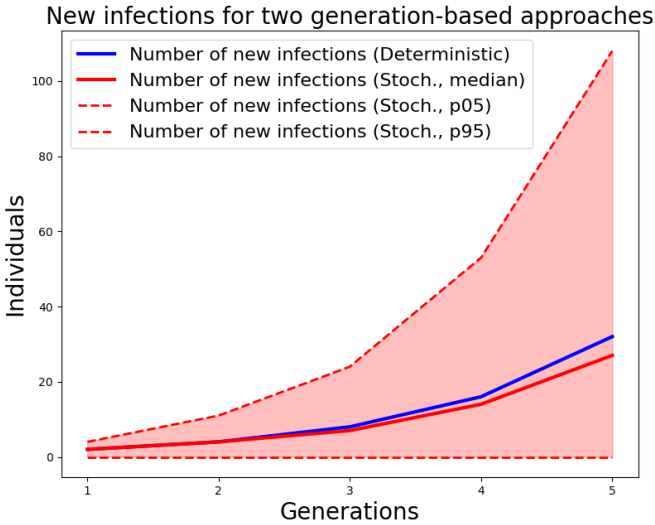

$\rightarrow$ Model is too simple if $S \ll N$

**Example of a tree of infections**

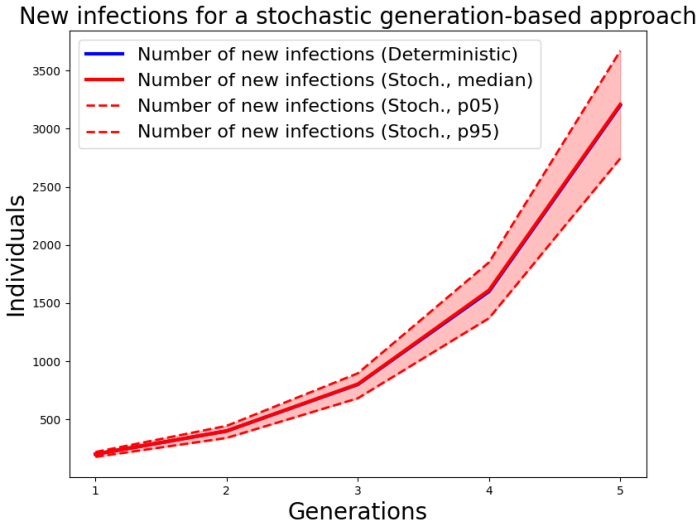
For $\phi = 3$ and $\rho = 2/3$, one transmission chain from one infected person $P0$ (only visualizing contacts where transmission occurs here) could look like this.



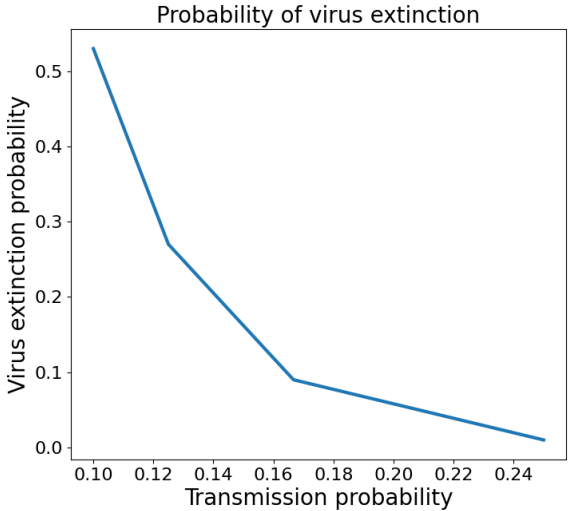
# Stochastic generation-based reproduction: $I_0 = 1$



# Stochastic generation-based reproduction: $I_0 = 100$



# Stochastic generation-based reproduction: Extinction probability



# Summary

- Both models neglected the importance of $S \ll N$
- The previous stochastic model was also very simple in its stochastic process
- Stochastic processes can be used to describe the generation of new cases on a more complex level
- We will not detail this here and only look at more advanced stochastic models when considering agent-based models

# A generation-based difference equations SIR model

- For any primary case, the chance of meeting a susceptible when a contact takes place is

$$\frac{S}{N}$$

- With contact rate $\phi$ and transmission probability per contact $\rho$, a deterministic model gives

$$\phi\rho\frac{I}{N}S \tag{2}$$

new infections per unit time.

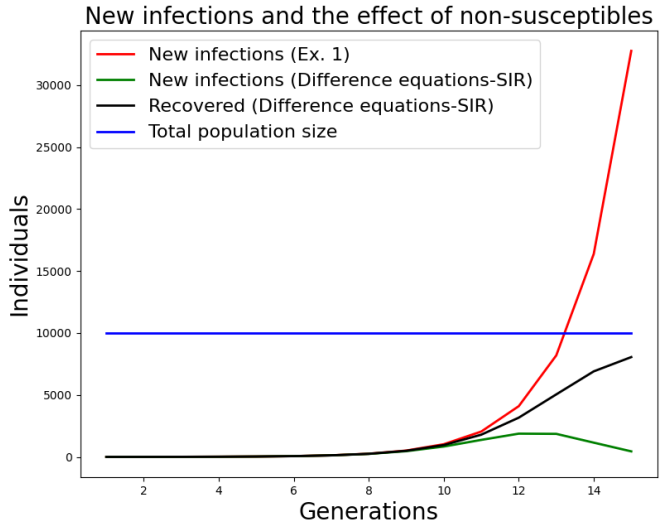
# A generation-based difference equations SIR model

- Consider the following set of difference equations for generation $i = 0, 1, \ldots$

$$
\begin{aligned}
S_{i+1} &= S_i - \rho\phi\frac{I_i}{N}S_i \\
I_{i+1} &= \rho\phi\frac{I_i}{N}S_i \\
R_{i+1} &= \sum_{k=0}^{i} I_k
\end{aligned}
\tag{3}
$$

- We assume $\phi$ to be given per generation and that individuals recover in one generation.
- Note that the previous system is not correct if $\rho\phi\frac{I_i}{N} > 1$.

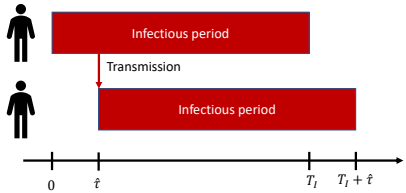
# A generation-based difference equations SIR model



# Equation-based models (EBM)

# From primary to secondary case in real time

- Infection from a primary to a secondary case can happen at any $\widehat{\tau}$ with $0 < \widehat{\tau} < T_I$.




- Broadly used models for real-time transmission are sets of
  - Ordinary Differential Equations (ODE)
  - Integro Differential Equations (IDE)
  - Stochastic Differential Equations (SDE)

# From discrete to continuous models: An equation for $S(t)$

- Take the difference equation from the previous chapter (with unit time in day(s)!)
- Consider a ”small“ time step $\Delta t$, a fraction of the unit time
- Then, the number of new infections from $t$ to $t + \Delta t$ is

$$S(t + \Delta t) - S(t) = -\rho\phi \frac{I(t)}{N} S(t)\Delta t \tag{4}$$

- By considering the limit $\Delta t \to 0$, we have for $S : \mathbb{R}_0^+ \to \mathbb{R}$,

$$S'(t) = \lim_{\Delta t \to 0} \frac{S(t + \Delta t) - S(t)}{\Delta t} = -\rho\phi \frac{I(t)}{N} S(t). \tag{5}$$

# From discrete to continuous models: A simple ODE-SIR model

- Waiving the assumption of recovery in unit time, we consider the system

$$\begin{aligned} S'(t) &= -\rho\phi\frac{I(t)}{N}S(t), \\ I'(t) &= \rho\phi\frac{I(t)}{N}S(t) - \frac{1}{T_I}I(t), \\ R'(t) &= \frac{1}{T_I}I(t). \end{aligned} \tag{6}$$

- This is a (nonlinear) system of ordinary differential equations (ODEs)
- We denote this an ODE-SIR model

# A simple ODE-SIR model: Visualization

- We present the system (6) by the following flow chart:


- The basic reproduction number is

$$R_0 = \rho\phi T_I. \tag{7}$$

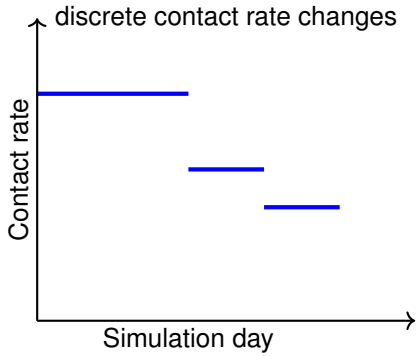
# Interventions

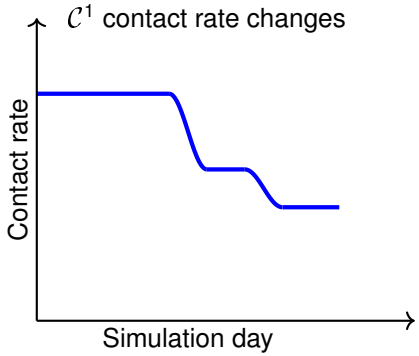
- Transmission of communicable diseases can be prevented by
  - avoiding situation where transmission can take place (e.g., human contacts)
  - taking protective measures in these situations (e.g., face masks)
  - better protecting susceptible individuals (e.g., by vaccination)
- The above actions classify in pharmaceutical (e.g., vaccination) and nonpharmaceutical (e.g., contact restrictions) interventions
- Nonpharmaceutical interventions will often be denoted *NPIs*
- For allowing contact restrictions, we define a time dependent contact rate

$$\phi(t) \in [0, \phi]. \tag{8}$$

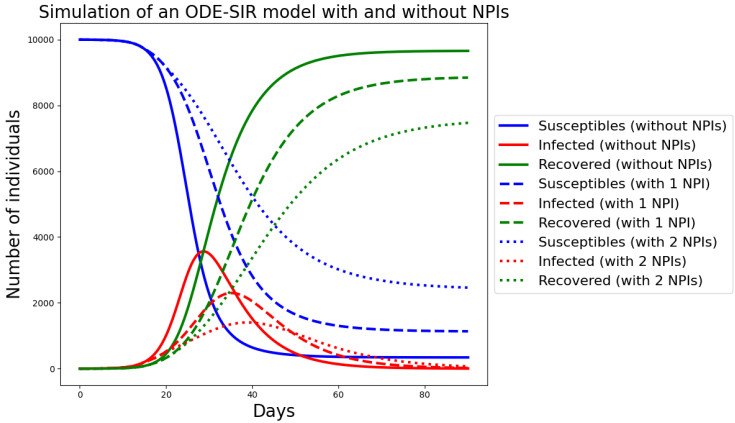
## Remark 1

*Discrete parameter changes are a problem for the analytical definition of the differential equation as well as for the numerical solution scheme. In order to avoid these situations, we can either restart the scheme from each discrete change point or use $\mathcal{C}^1$ functions to express time dependent parameters.*



# Numerical solution of a simple ODE-SIR model with NPIs

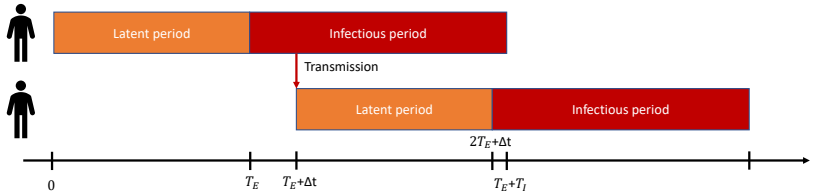


# Extensions of the ODE-SIR model

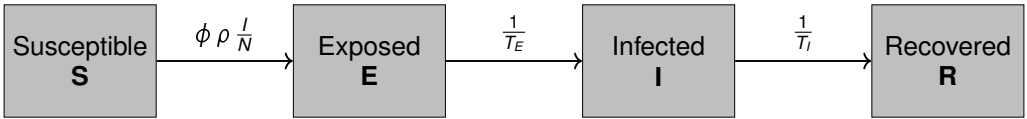
- (ODE)-SIR models can be extended in many “dimensions”
- The (research) question defines the model
- If the question is about containment or mitigation of the virus, we need to describe the transmission process more properly.
- If the question was about mortality from the virus, we need to introduce an infection state *Dead*
- ...

# Extensions of the ODE-SIR model: An ODE-SEIR model

- Direct retransmission of a pathogen after catching it may be unrealistic
- “ignoring the latent period results in underestimating the basic reproductive ratio of an infection from outbreak data”;
  see Keeling, Rohani, Keeling: Appropriate Models for the Management of Infectious Diseases (2006).
- We introduce a state $E$ (Exposed) and $T_E$, the time a person remains in a latent non-infectious (or exposed) state after transmission



# An ODE-SEIR model: Visualization

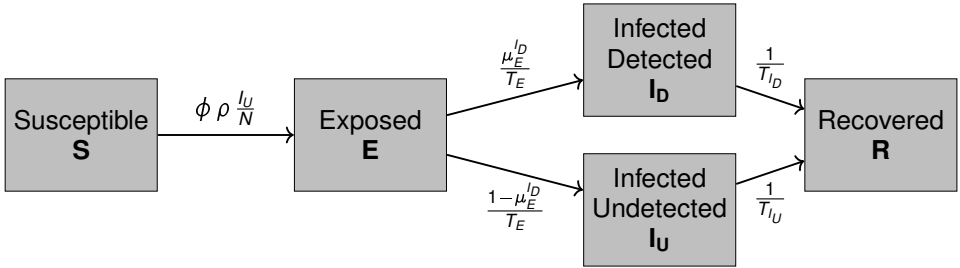


# An ODE-SEIR model: Equations

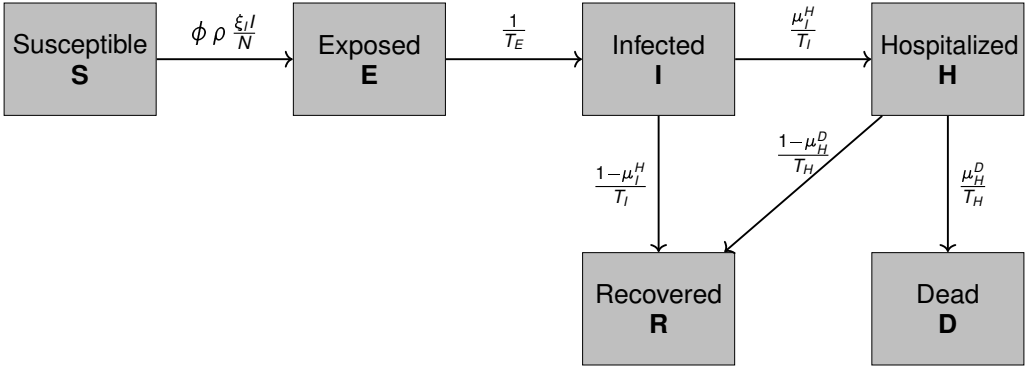
- We assume that the outflow of $E$ is proportional to $E$
- Now, the "outflow" of $S$ goes to $E$ instead of $I$, i.e., the extended system writes

$$\begin{aligned} S'(t) &= -\rho\phi\frac{I(t)}{N}S(t), \\ E'(t) &= \rho\phi\frac{I(t)}{N}S(t) - \frac{1}{T_E}E(t), \\ I'(t) &= \frac{1}{T_E}E(t) - \frac{1}{T_I}I(t), \\ R'(t) &= \frac{1}{T_I}I(t). \end{aligned} \tag{9}$$

# An ODE-SEIR model with explicit Test-Trace-Isolate effects



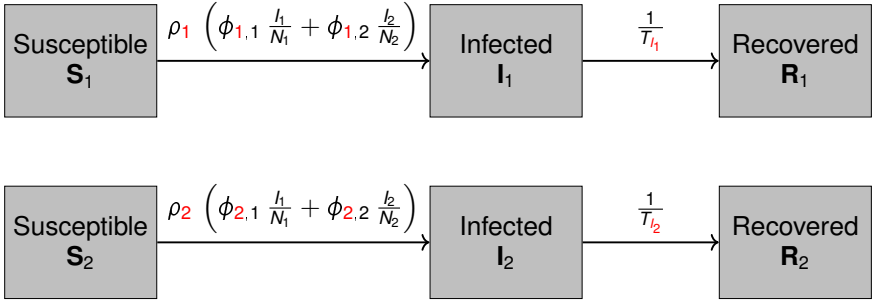
# Extensions of the ODE-SIR model: An ODE-SEIHRD model



# Simple ODE-based modeling: Advantages and limitations

- cheap to compute
- well established methods to analyze for, e.g., equilibria
- homogeneous mixing assumption may be wrong
  - → introducing age-resolution (ODE)
  - → introducing different subpopulations (ODE)
  - → introducing spatial heterogeneity (hybrid or coupled ODEs)
- compartment stays are exponential and, e.g., viral load constant
  - → considering integral terms (IDE)
- stochastic effects not reflected
  - → ... (SDE/ABM)

# Simple example of age-resolved models for two age groups



## Theorem 1

*Consider the system of integro-differential equations*

$$
\begin{aligned}
S'(t) &= \rho\,\phi(t)\,\frac{S(t)}{N}\int_{t_0}^{t}\gamma_I^R(t-x)\,S'(x)\,\mathrm{d}x,\\
I(t) &= -\int_{t_0}^{t}\gamma_I^R(t-x)\,S'(x)\,\mathrm{d}x,\\
R(t) &= -\int_{t_0}^{t}\left(1-\gamma_I^R(t-x)\right)S'(x)\,\mathrm{d}x.
\end{aligned}
\tag{10}
$$

*Let* $\gamma_I^R(\tau) = \exp(-\frac{\tau}{T_I})$. *Then* (10) *reduces to*

$$
\begin{aligned}
S'(t) &= -\rho\,\phi(t)\,I(t)\,\frac{S(t)}{N}\\
I'(t) &= \rho\,\phi(t)\,I(t)\,\frac{S(t)}{N} - \frac{1}{T_I}I(t)\\
R'(t) &= \frac{1}{T_I}I(t)
\end{aligned}
$$

# IDE-based models as generalization of ODE-based models: Part I

- $\gamma_I^R(\tau)$ is the fraction of infected individuals that will still recover from infection after time $\tau$ (i.e., that is still infected at time $\tau$).

- We need

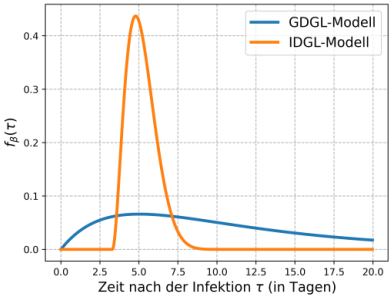
$$\begin{gathered} \gamma_I^R(0) = 1, \quad \gamma_I^R(\tau) \geq 0 \quad \text{for all } \tau \geq 0, \\ \gamma_I^R(x) \text{ monotonously decreasing}, \quad \int_0^\infty \gamma_I^R(\tau) \mathrm{d}\tau < \infty. \end{gathered} \tag{11}$$

- Theorem 1 shows that our ODE-SIR model implicitly assumed that compartment stays were exponential.

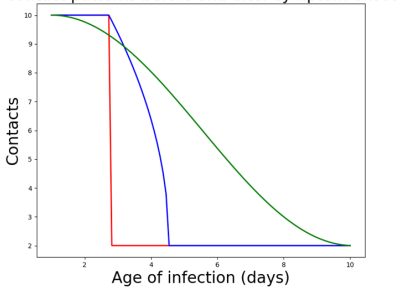
# IDE-based models as generalization of ODE-based models: Part II

- Infectiousness depends on age of infection: $\rho \to \rho(\tau)$
- Contacts depend on age of infection: $\phi(t) \to \phi(t, \tau)$

"Infectiousness" over time




Contact patterns before and after symptom onset




More details: Keimer/Pflug (2020), Plötzke (2020).

## Theorem 2

*Consider the system of integro-differential equations*

$$
\begin{aligned}
S'(t) &= \frac{S(t)}{N} \int_{t_0}^{t} \phi(t, t-x)\rho(t-x)\,\gamma_I^R(t-x)\,S'(x)\,\mathrm{d}x, \\
I(t) &= -\int_{t_0}^{t} \gamma_I^R(t-x)\,S'(x)\,\mathrm{d}x, \qquad (12) \\
R(t) &= -\int_{t_0}^{t} \left(1-\gamma_I^R(t-x)\right)S'(x)\,\mathrm{d}x.
\end{aligned}
$$

*Let* $\gamma_I^R(\tau) = \exp(-\frac{\tau}{T_I})$, $\rho(\tau) = \rho$, *and* $\phi(t,\tau) = \phi(t)$. *Then* (12) *reduces to:*

$$
\begin{aligned}
S'(t) &= -\rho\,\phi(t)\,I(t)\,\frac{S(t)}{N} \\
I'(t) &= \rho\,\phi(t)\,I(t)\,\frac{S(t)}{N} - \frac{1}{T_I}I(t) \\
R'(t) &= \frac{1}{T_I}I(t)
\end{aligned}
$$

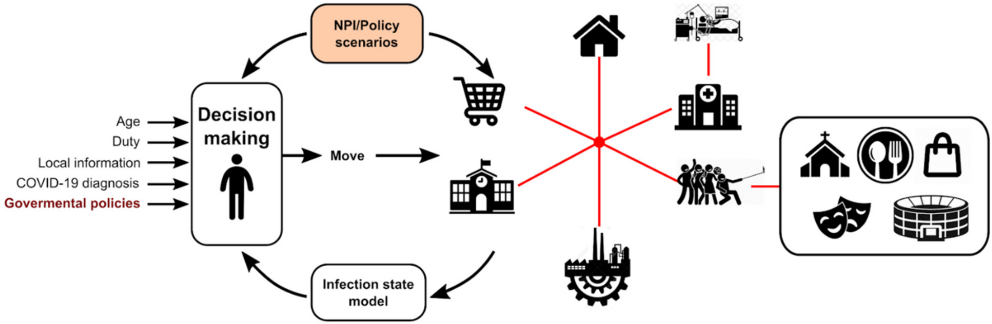
# Agent-based models (ABM)

# Agent-based modeling

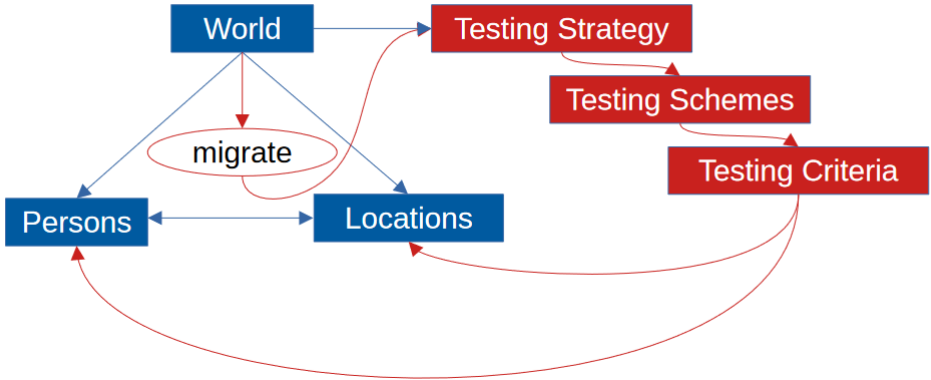
- **Agents**
  Object which holds information, e.g., infection status, current location or age
- **Locations**
  Multiple locations which can be visited, e.g., individual homes, schools, workplaces
- **Rules** / **Interactions**
  Interactions of different Agents at a current location, or rules for traveling between locations

$\rightarrow$ micro granularity & stochastic effects !

# Agent-based modeling

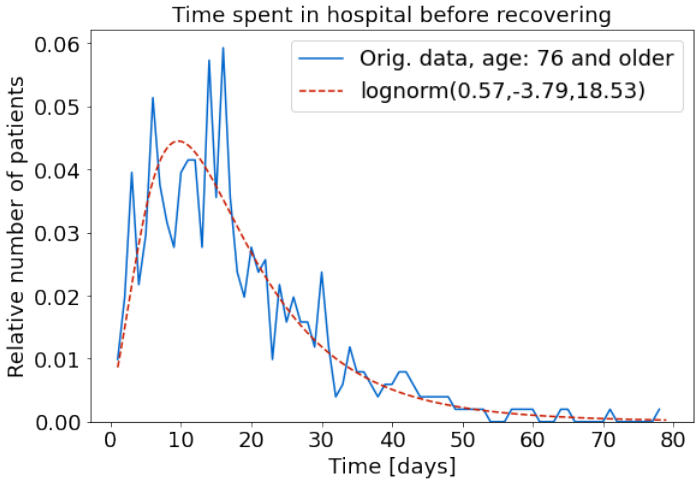


# Agent-based modeling: Testing Scheme

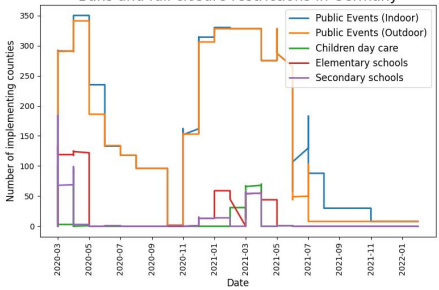


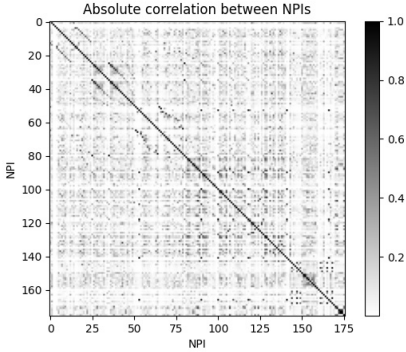
# Digression: Data science

# Data science is essential part of the process



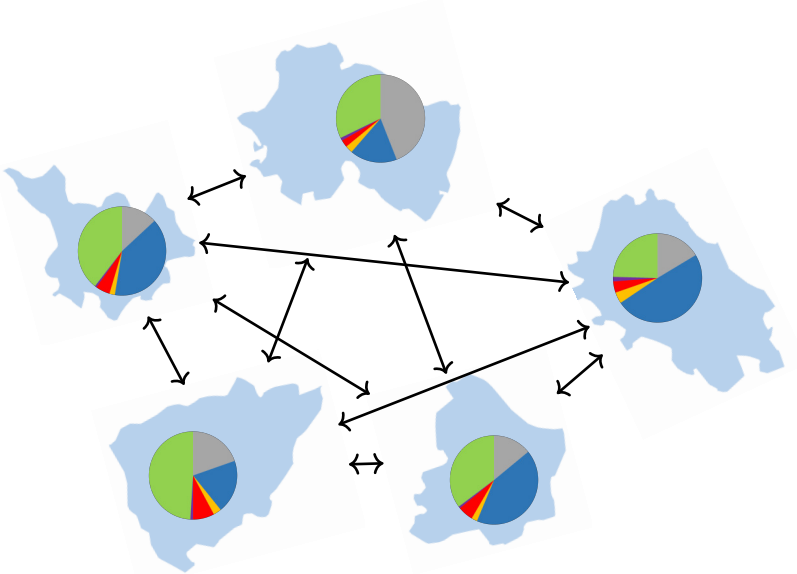
# Data science is essential part of the process



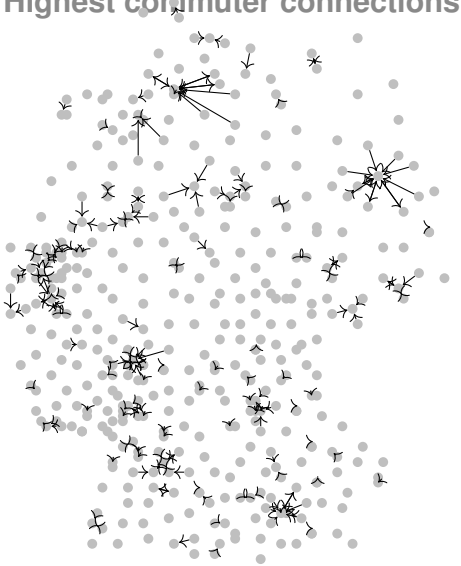



# Hybrid models / Spatial resolution for EBMs

# Spatial resolution for EBMs: Hybrid Graph-ODE model



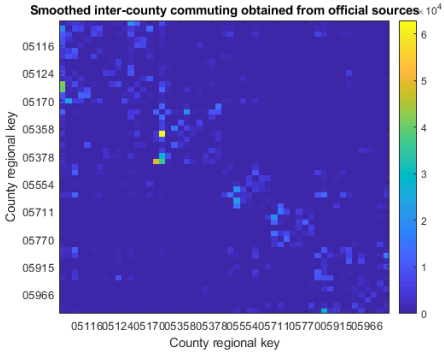
# Highest commuter connections in Germany

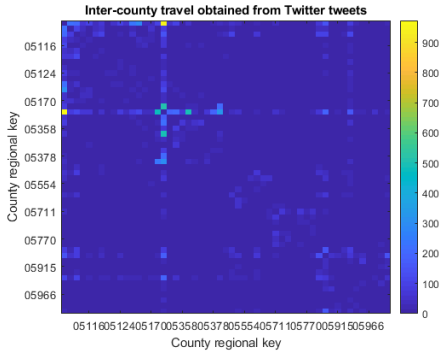


Extracted from Federal Agency of Work: Registered home and work places

# Inter-regional contacts: Official sources and social network

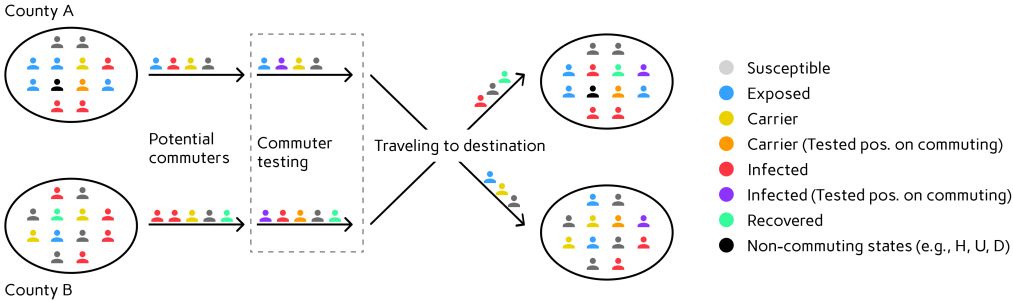
North Rhine-Westfalia:



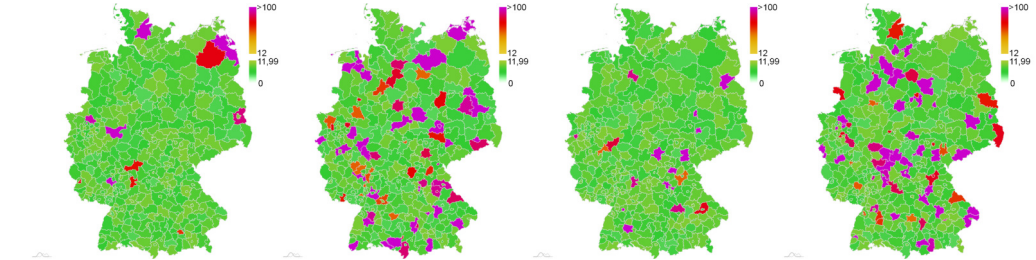
# Numerical assessment of the NoCovid strategy

# Commuter testing in hybrid graph-ODE model



# Numerical assessment of the NoCovid strategy

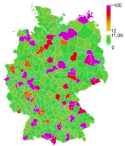
- NoCovid $\neq$ ZeroCovid
- NoCovid: “Controlling the Covid-19 pandemic through Green Zones”

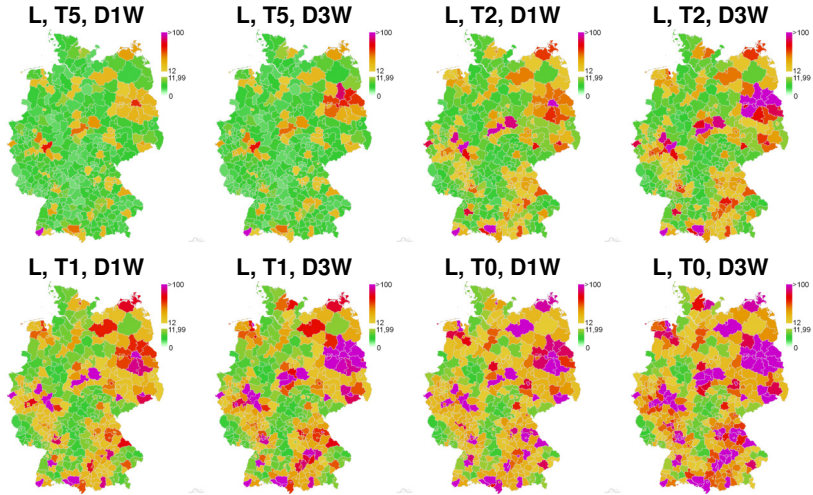



Four different initial scenarios. Random initial incidence (weekly cases per 100 000 individuals) of 75-150 for 2-20% of the counties and incidence below 10 otherwise (top).

# Numerical assessment of the NoCovid strategy

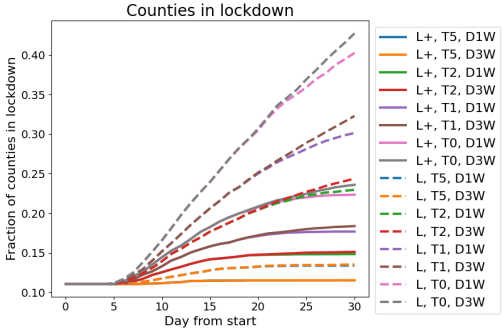
- Test of commuters coming from red zones
- 75 % detection ratio (averaged value for mix of massive deployment of antigen tests plus PCR, RTD-PCR and pool tests)
- Considering different frequencies (daily, twice per week, ...)

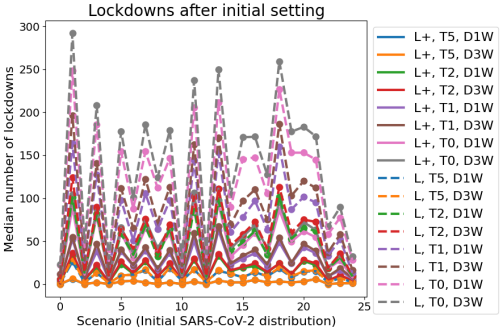



Simulated spread of SARS-CoV-2 cases for one initial scenario of about 18 % red zones and 8 different strategies. Median result after 30 days of simulation time.

# Numerical assessment of the NoCovid strategy




*Further reading:* M. J. Kühn, D. Abele, S. Binder, K. Rack, M. Klitz, J. Kleinert, J. Gilg, L. Spataro, W. Koslow, M. Siggel, M. Meyer-Hermann, A. Basermann. Regional opening strategies with commuter testing and containment of new SARS-CoV-2 variants. BMC Infectious Diseases 22, no. 1 (2022): 333

`https://doi.org/10.1186/s12879-022-07302-9`

# Summary

- Models necessary where classical experiments not possible or too costly
- A lot of models exist
- Only presented a non-exhaustive list
- All models come with advantages and disadvantages
- Hybrid models try to combine the “best out of two worlds”
- Data science is essential part of the process

# Thank you

"Predictive simulation software" at Institute for Software Technology:
→ Infectious Diseases Dynamics not the only topic
→ We are highly interested in future collaborations!

*Further reading*: W. Koslow, M. J. Kühn, S. Binder, M. Klitz, D. Abele, A. Basermann, M. Meyer-Hermann, Appropriate relaxation of non-pharmaceutical interventions minimizes the risk of a resurgence in SARS-CoV-2 infections in spite of the Delta variant. PLoS Computational Biology 18, no. 5 (2022): e1010054.


`https://github.com/SciCompMod/memilio`

Thank you for your kind attention!