

WHITEPAPER

Epidemiologie – Modellierung, Simulation
und Entscheidungsunterstützung für lokale
Entscheidungsträger



EPIDEMSE

Epidemiologie – Modellierung, Simulation und Entscheidungsunterstützung für lokale Entscheidungsträger

Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM
Kaiserslautern

Michael Bortz, Michael Burger, Jochen Fiedler, Simone Gramsch, Neele Leithäuser, Jan Mohring

Kontakt:

Dr. Simone Gramsch

simone.gramsch@itwm.fraunhofer.de

Tel.: +49 631 31600-4427

Fraunhofer ITWM
Fraunhofer-Platz 1
67663 Kaiserslautern

Version 1.0 vom 16. Juni 2020

Inhalt

1. Einführung	4
2. Epidemiologische Modelle zur SARS-CoV-2-Ausbreitungsprognose	6
2.1. Stand der Wissenschaft	6
2.2. Vorarbeiten des ITWM	7
2.3. Ausblick	13
3. Datenbasierte Dunkelzifferschätzung	13
3.1. Stand der Wissenschaft	14
3.2. Vorarbeiten des Fraunhofer ITWM	15
3.3. Ausblick	17
4. Planung von Teststrategien zur Reduzierung der Unsicherheit von Modellvorhersagen	17
4.1. Stand der Wissenschaft	18
4.2. Vorarbeiten des Fraunhofer ITWM	18
4.3. Ausblick	19
5. Mehrkriterielle Entscheidungsunterstützung	19
5.1. Stand der Wissenschaft	20
5.2. Vorarbeiten des Fraunhofer ITWM	20
5.3. Ausblick	20
6. Softwarelösungen zur epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung	21
6.1. Stand der Technik	21
6.2. Vorarbeiten des Fraunhofer ITWM: das EpideMSE-Tool	22
7. Zusammenfassung und Ausblick	24
A. Referenzen	25
	25

Vor allem wenn eine Infektionskrankheit neu auftritt, helfen mathematische Modelle, die beiden wichtigsten Fragen zu beantworten: *Wie wird sich die Infektion ausbreiten und welche Maßnahmen zu ihrer Eindämmung sind angemessen und wirkungsvoll?* Um lokale Entscheidungsträger (Gesundheitsämter, Krankenhäuser und vor allem Kommunen) bei der Planung ihrer Maßnahmen zu unterstützen, arbeiten Forschende des Fraunhofer ITWM im Rahmen des Anti-Corona-Programms der Fraunhofer-Gesellschaft an der epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung von Covid-19.

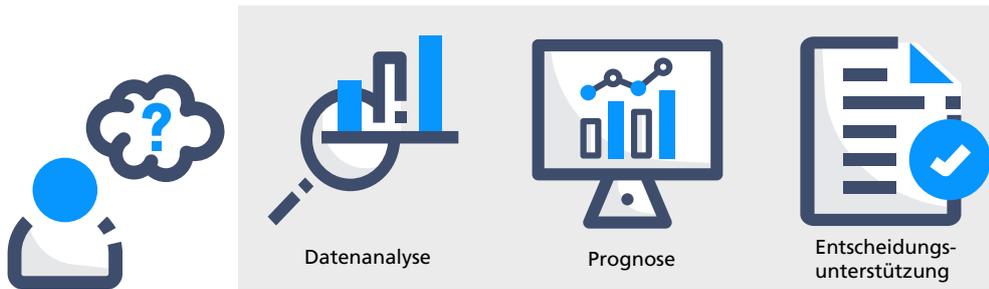
Aus Sicht der Forschenden am Fraunhofer ITWM sollten Politik und Gesellschaft ihre Strategien zur Bewältigung der Corona-Krise auf

- den *regional* vorliegenden Epidemieverlauf von SARS-CoV-2 und
- die Prognose der *regionalen* Covid-19-Erkrankungen mit schweren Verlauf

abstimmen. Da das SARS-CoV-2-Virus neu ist, können die Prognosen zur zukünftigen Infektionsausbreitung nur auf Basis mathematischer Epidemiologie-Modelle getroffen werden (siehe auch [MHPPS20]). Von lokalen Entscheidungsträgern kann nicht erwartet werden, dass sie sich in die theoretische Epidemiologie einarbeiten, um neueste Forschungsergebnisse zum Infektionsgeschehen gegen die lokale Ressourcenplanung zum Schutz der Bevölkerung und die wirtschaftlichen Interessen ihrer Region abzuwägen. Es ist Aufgabe der Forschenden, Entscheidungsträger durch die Aufbereitung der Daten zum Epidemieverlauf und durch die intuitive Visualisierung der Prognosen für regional angepasste Szenarien in die Lage zu versetzen, *What-if-Analysen* durchzuführen.

Zur Erarbeitung von regionalen Eindämmungsstrategien benötigen lokale Entscheidungsträger daher eine epidemiologische Simulationssoftware. Das Fraunhofer ITWM entwickelt ein solches Tool namens [EpidemMSE](#). Im Folgenden beschreiben wir die Funktionalitäten des Tools. Wir zeigen Anwendungsmöglichkeiten auf, erklären den wissenschaftlichen Hintergrund und betten diese in den aktuellen Stand der Forschung zur Ausbreitung von SARS-CoV-19 ein. Derzeit laufen weltweit Forschungsprojekte zur Überwindung der Corona-Krise, die jederzeit neue und unerwartete Erkenntnisse liefern können. Dennoch wagen wir einen Ausblick auf die Weiterentwicklung der vorgestellten Methoden.

Software zur epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung »EpidemMSE« Kern einer epidemiologischen Software zur Entscheidungsunterstützung sind epidemiologische Modelle zur Prognose der SARS-CoV-2-Ausbreitung, die durch lokale Daten zum Infektionsgeschehen automatisch gefüttert werden. Jede Prognose ist dabei mit einer gewissen Unsicherheit behaftet, denn nicht alle infizierten Personen werden auch getestet und entdeckt. Daher ist eine wichtige Funktionalität eines Tools zur epidemiologischen Entscheidungsunterstützung die Visualisierung der Prognose inklusive des Korridors, der durch die Unsicherheit bedingt wird. Dies wird auch unterstützt durch die direkte Visualisierung der geschätzten Dunkelziffer für jede Region.



Epidemiologische Modelle zur SARS-CoV-2-Ausbreitungsprognose

Abb. 1: Kernmodule einer epidemiologischen Software zur Entscheidungsunterstützung

Die Simulationssoftware unterstützt die lokalen Entscheidungsträger vor allem, wenn sie darüber hinaus Funktionalitäten zur Planung der Eindämmungsmaßnahmen enthält. Zum einen ist es wichtig, die Unsicherheit der Prognosen zu reduzieren, indem weite Teile der Bevölkerung und insbesondere Risikogruppen und systemrelevantes Personal systematisch auf SARS-CoV-2 oder eine durchgestandene Covid-19-Infektion getestet werden. Um möglichst ressourcenschonend zu testen, müssen statistische Testpläne angepasst an den bisherigen Infektionsverlauf und die regionalen Besonderheiten von der Software erstellt werden. Zum anderen ermöglicht die Visualisierung unterschiedlicher SARS-CoV-2-Ausbreitungsprognosen abhängig von veränderten Infektionsraten Entscheidungsträgern, verschiedene Maßnahmen datenbasiert für einzelne Regionen zu vergleichen. Zusammenfassend sollte ein Tool daher folgende Funktionalitäten beinhalten:

1. datenbasierte Auswertung des regionalen Epidemieverlaufs mit der aktuellen Anzahl infizierter Personen und Neuinfektionen
 - Infektionszahlen: regional aufgelöst (für Gesamtdeutschland, Bundesländer, Landkreise und Städte) und zusätzlich mit zeitlichem Verlauf
 - Infektionszahlen und Sterberaten (regional und mit zeitlichem Verlauf, nach Geschlecht und Alter differenziert)
 - Schätzung der Dunkelziffer (regional, differenziert nach Altersgruppen)
2. regionale Prognose der SARS-CoV-2-Infektionen mit Korridor zur Darstellung der Unsicherheit (minimal und maximal zu erwartende Fallzahlen)
 - Anzahl der zu erwartenden Infizierten (regional, nach Geschlecht und Alter differenziert)
 - Anzahl der zu erwartenden Hospitalisierungen (regional)
 - Visualisierung der zu erwartenden lokalen Auslastung der Intensivplätze
3. Planung von Teststrategien zur Reduzierung der Unsicherheit und What-if-Analysen
 - Design of Testing – automatisiertes Erstellen von statistischen Testplänen für Regionen
 - Erweiterung der oben aufgeführten regionalen Prognosen mit unterschiedlichen Infektionsraten, die beliebig konfigurierbar sind und zu unterschiedlichen Zeitpunkten starten können
 - Korrelation Maßnahmen zu der Veränderung der Infektionsrate



2. Epidemiologische Modelle zur SARS-CoV-2-Ausbreitungsprognose

Der Beginn der Corona-Krise stand im Zeichen deutschlandweit einheitlicher Eindämmungsmaßnahmen wie beispielweise Kontaktbeschränkungen oder Schulschließungen. Dank dieser Maßnahmen konnte die Epidemie unter Kontrolle gebracht werden und die dringend gebotene Rückkehr zu einem normalen sozialen und wirtschaftlichen Leben erscheint nun möglich. Allerdings blieb eine generelle Immunisierung aus und nach Lockerungen können wenige Infizierte sofort wieder eine exponentielle Ausbreitung in Gang setzen. In Zukunft geht es also darum, Infektionsherde sofort zu entdecken und durch gezielte lokale Maßnahmen einzuhegen.

Das von uns entwickelte Modell ist nun genau so konzipiert, dass derartige *bereichsabhängige* Maßnahmen schnell durchgespielt, abgeschätzt und ausgewählt werden können. Genauer handelt es sich um ein Kohorten-basiertes Mehrgruppenmodell mit Retardierung. Die Gruppen sind räumlich und nach Alter getrennt. Maßnahmen werden durch geänderte Kontaktraten sowie Entdeckungsraten und –zeiten modelliert. Im Gegensatz zu agentenbasierten Modellen ist die Anzahl der freien Parameter noch so überschaubar, dass sie aus verfügbaren Erhebungen und Studien rekonstruiert werden können. Hierzu wird ein Bayes'scher Zugang gewählt. Er erlaubt insbesondere, Unsicherheiten in den Messdaten in Unsicherheiten der Prognosen zu übersetzen.

2.1. Stand der Wissenschaft

Ein erstes mathematisches Modell zur Ausbreitung der Pocken wurde bereits 1766 von Daniel Bernoulli [Ber60] aufgestellt. Er begründete damit die Disziplin der mathematischen Epidemiologie und schuf die Urform der Kompartimentbasierten Modelle. Diese teilen die Bevölkerung in Gruppen ein, die ihren aktuellen Gesundheitszustand widerspiegeln. Die Modelle in ihrer heutigen Form gehen auf das so genannte SIR-Modell von Kermack und McKendrick [KW27] aus dem Jahr 1927 zurück. Sie unterteilen die Bevölkerung in Anfällige (S), Infektiöse (I) und Erholte (R). Die Übergänge zwischen den Kompartimenten werden mit Hilfe von Ratengleichungen modelliert, was zu ODE-Systemen führt. Eine Verbesserung stellen die SEIR-Modelle dar, die auch die Latenz zwischen Infektion und eigener Infektiösität berücksichtigen. Hierzu wird das Zwischenkompartiment der Exponierten (E) eingeführt. Solche Modelle, einschließlich vieler Variationen, sind gut etabliert und ihre Namen spiegeln die Reihenfolge der Übergänge wider. Einen guten Überblick liefert [Het00].

Eng verwandt sind Modelle mit Zeitverzögerung. Sie ermöglichen eine genauere Darstellung der infektiösen Phase und führen auf Differentialgleichungen mit Retardierung, siehe [Coo67, Het95, Het00, EB01, KK03, YL06, BB11, RA12]. Insbesondere können sie realen Infektionsverläufen genauer folgen, wenn sich Kontaktraten infolge politischer Maßnahmen sprunghaft ändern. Das hier verwendete Modell gehört zu dieser Kategorie.

Dank gesteigerter Rechenleistung kommt im letzten Jahrzehnt auch zunehmend eine zweite Modellklasse zum Einsatz: agentenbasierte Methoden. Diese verfolgen im Rahmen einer Monte-Carlo-Simulation Interaktionen vieler einzelner Individuen. Streuende Parameter oder räumliche Bewegungsmuster lassen sich damit gut abbilden. Ein Problem stellt neben der Rechenzeit allerdings die Eichung der zahlreichen Parameter anhand eher spärlicher Messdaten dar.

Agentenbasierte Methoden sind nicht direkt Teil des hier vorgestellten Werkzeugs. Allerdings werden sie verwendet, um die Parameter des Kompartimentbasierten Verfahrens für neue politische Maßnahmen zu schätzen. Ein Vergleich der beiden grundsätzlichen Ansätze findet sich in [GAB⁺10]. Eine neuere Übersicht, die auch auf den Einsatz maschinellen Lernens eingeht, bietet [SR13].

2.2. Vorarbeiten des ITWM

Kurz nach Ausbruch der Corona-Pandemie begannen Mitarbeiter des ITWM in ihrer Freizeit, das Prognosemodell zu entwickeln, auf dem jetzt die aktuelle Software zur Entscheidungsunterstützung beruht. Wie die klassischen SEIR-Modelle beschreibt es auf deterministische Weise den Austausch zwischen Personengruppen in verschiedenen Phasen der Erkrankung. Allerdings wird der Übergang von einer Phase in die andere nicht über Raten, sondern wirklich anhand der Zeit seit Infektion bestimmt. Während die SEIR-Modelle alleine durch Differentialgleichungen beschrieben werden, enthält unser Modell daher auch Retardierungsterme. Wir versprechen uns davon, die in [Nik20] dargelegten grundsätzlichen Probleme der SEIR-Modelle zu vermeiden, z.B. eine Unterschätzung der Zahl Infektöser in der Hochphase der Epidemie ca. um einen Faktor zwei. Eine weitere wichtige Säule unseres Modells ist die Möglichkeit, den Epidemieverlauf nach Kohorten aufzuschlüsseln. Diese können sich wahlweise auf Alter, Ort, Geschlecht, Berufsgruppe oder Kombinationen davon beziehen. Erst dadurch wird es möglich, den Sinn oder Unsinn spezifischer Gegenmaßnahmen zu prognostizieren, z.B. das Verbot von Großveranstaltungen oder die Isolation älterer Menschen. Die Wirkung von Maßnahmen auf den Verlauf der Epidemie wird durch folgende Steuergrößen abgebildet: die mittlere Zahl kritischer Kontakte zwischen Personen derselben und unterschiedlicher Kohorten pro Tag, die Zahl und Treffsicherheit von Tests, mit der Infektöse entdeckt werden, und die mittlere Zeit seit Infektion, mit der dies passiert.

In diesem Abschnitt beschreiben wir ein Ausbreitungsmodell, das verschiedene Parameter für unterschiedliche Teilpopulationen zulässt, z.B. eine ungleiche Sterblichkeitsrate zwischen Alten und Jungen oder eine Kontaktrate, die zwischen dünn und dicht besiedelten Gebieten variiert. Um z.B. die Effektivität von Tests beurteilen zu können, unterscheiden wir explizit zwischen entdeckten und nicht entdeckten Infizierten und wir berücksichtigen Kranke, die den Virus von außen ins System schleusen. Das Ausbreitungsmodell kennt pro Gruppe nur zwei Zustandsgrößen: alle jemals Infizierten und die entdeckten Infizierten. Alle weiteren Größen wie die Sterbefälle lassen sich im Nachgang daraus ableiten.

Symbol	Bedeutung	typisch	Einheit
g	Zahl der Gruppen	10^2	
G_k	Zahl der Mitglieder in Gruppe k	10^6	p
G	Zahl der Mitglieder aller Gruppen	8×10^7	p
τ_k^s	Beginn der infektiösen Phase in Gruppe k	2	d
τ_k^e	Ende der infektiösen Phase in Gruppe k	6.9	d
$\tau_k^d(t)$	Zeit zwischen Ansteckung und Entdeckung eines jemals Entdeckten	9.5	d
$\kappa_{kl}(t)$	Kritische Kontakte einer Person aus Gruppe k mit Personen aus Gruppe l pro Tag	0.2	$\frac{p}{d}$
$\lambda_k(t)$	Anteil der Infizierten aus Gruppe k , die entdeckt werden.	0.5	

Epidemiologische Modelle zur SARS-CoV-2-Ausbreitungsprognose

Tab. 1: Modellparameter.

Symbol	Bedeutung	typisch	Einheit
N_k	Zahl aller je Infizierten in Gruppe k	10^6	p
S_k	Zahl der Infizierbaren in Gruppe k	10^7	p
I_k	Zahl der effektiv Infektiösen in Gruppe k	10^5	p
D_k	Zahl der entdeckten Infizierten in Gruppe k	10^6	p
D_k^0	Zahl der bisher Infizierten in Gruppe k , die irgendwann entdeckt werden	10^6	p

Tab. 2: Zustandsgrößen.

Parameter Eine wichtige Modellannahme besteht darin, dass die Wirkung der individuell variierenden Parameter durch Mittelwerte für Teilpopulationen gut reproduziert werden kann.

Ein kritischer Kontakt ist definiert als Kontakt zwischen zwei Personen, der so eng ist, dass er zur Infektion führen würde, wenn die eine Person gerade infektiös und die andere noch nicht immun ist. Ein kritischer Kontakt kann z.B. auch zwischen zwei Immunen stattfinden. Es geht bei der Definition um die Art der Begegnung, nicht um die Wirkung. Die Zahl kritischer Kontakte korrespondiert also mit politischen Maßnahmen wie Kontaktbeschränkung, aber nicht mit dem aktuellen Grad der Durchseuchung und kann also als Regelgröße verwendet werden. κ wird andernorts auch Transmissions-, Übertragungs- oder Kontaktrate genannt.

Die Wirkung von Tests wird durch zwei Regelgrößen modelliert: die Wahrscheinlichkeit $\lambda_k(t)$, mit der Infizierte entdeckt werden, und die mittlere Zeit $\tau_k^d(t)$ nach Infektion, zu der eine solche Entdeckung erfolgt. Man beachte die Bedeutung der zweiten Größe. Werden Infizierte erst am Ende der infektiösen Phase aufgespürt, dann bringt auch eine hohe Entdeckungsrate wenig. Gelingt es hingegen, Personen noch während der Latenzzeit in Quarantäne zu nehmen, dann entfaltet ein Test seine maximale Wirkung. Wir setzen hier perfekte Quarantäne voraus, d.h. dass Infizierte nach ihrer Entdeckung niemanden mehr anstecken, insbesondere nicht das Klinikpersonal.

Ausbreitungsdynamik Um die Dynamik der Ausbreitung zu beschreiben, werden zunächst die in Tabelle 2 aufgeführten Teilpopulationen verwendet.

$$\dot{N}_k(t) = \frac{S_k(t)}{G_k} \sum_{l=0}^{g-1} \kappa_{lk}(t) I_l(t) \quad (1)$$

$$\dot{D}_k(t) = \lambda_k(t) \frac{d}{dt} N_k(t - \tau_k^d(t)) . \quad (2)$$

Die Summe in Gl. (1) gibt die Gesamtzahl von Personen in Gruppe k an, die pro Tag in kritischen Kontakt mit Infektiösen aus allen Gruppen treten. Infiziert wird von diesen der Anteil der Infizierbaren (erster Quotient). Das Modell für die Entdeckung ist schlicht. $\frac{d}{dt} N_k(t - \tau_k^d(t))$ ist die Zahl der Personen in Gruppe k , die vor $\tau_k^d(t)$ Tagen neu infiziert wurden. Genau nach $\tau_k^d(t)$ Tagen entscheidet sich also, ob Infizierte mit der Wahrscheinlichkeit $\lambda_k(t)$ durch Tests erkannt werden oder von da an nie wieder. Wir setzen voraus, dass die Entdeckungsdauer zwischen Beginn und Ende der infektiösen Zeit liegt:

$$\tau_k^s \leq \tau_k^d(t) \leq \tau_k^e . \quad (3)$$

Für $\tau_k^d = \tau_k^s$ gibt eine entdeckte Person das Virus nicht mehr weiter, d.h. noch kürzere Entdeckungsdauern hätten keinen weiteren Effekt. Eine Entdeckung nach der infektiösen Zeit, z.B. durch Antikörpertests, mag von wissenschaftlichem Interesse sein – die Ausbreitung beeinflusst sie nicht mehr. In Gl. (1) und (2) gelten die folgenden Zusammenhänge:

$$S_k(t) = G_k - N_k(t) \quad (4)$$

$$D_k(t) = D_k^0(t - \tau_k^d(t)) \quad (5)$$

$$I_l(t) = \int_{t-\tau_l^e}^{t-\tau_l^s} \dot{N}_l(t') dt' - \int_{t-\tau_l^e}^{t-\tau_l^d(t)} \dot{D}_l^0(t') dt' . \quad (6)$$

In unserem Modell bilden die Infizierbaren das Komplement zu den jemals Infizierten, d.h. Infizierte sind später entweder immun oder tot. Gl. (5) definiert die zunächst etwas abstrakte Zahl $D_k^0(t)$ der Personen, die zur Zeit t infiziert sind und bereits entdeckt wurden oder erst noch entdeckt werden. Wir benötigen sie zur Formulierung der effektiv Infektiösen in Gl. (6). Das erste Integral erstreckt sich über alle Personen, die vor τ_l^e bis τ_l^s Tagen neu infiziert wurden, jetzt also Viren produzieren. Man beachte, dass Infizierte in dieser Phase als gleichbleibend infektiös modelliert werden. Im zweiten Integral werden diejenigen abgezogen, die nach der Infektion bis heute entdeckt und in Quarantäne genommen wurden.

Um das endgültige Modell formulieren zu können, müssen wir noch die Zeit der Entdeckung t_k^d einführen:

$$t' = t_k^d(t) \Leftrightarrow t = t' - \tau_k^d(t') . \quad (7)$$

Für konstante Entdeckungsdauer gilt einfach $t_k^d(t) = t + \tau_k^d$. Damit schreibt sich unser Dynamik-Modell als retardierte Differentialgleichung in den beiden Größen

N_k und D_k^0 :

$$\dot{N}_k(t) = \left(1 - \frac{N_k(t)}{G_k}\right) \sum_{l=0}^{g-1} \kappa_{lk}(t) \times \left[N_l(t - \tau_l^s) - N_l(t - \tau_l^e) - D_l^0(t - \tau_l^d(t)) + D_l^0(t - \tau_l^e) \right] \quad (8)$$

$$\dot{D}_k^0(t) = \lambda_k \left(t_k^d(t) \right) \dot{N}_k(t). \quad (9)$$

Man beachte, dass Gl. (9) zukünftige Entdeckungsraten- und -dauern verwendet.

Symmetrie der Kontakte und Rechnung mit Anteilen Der Kontakt einer Person aus Gruppe k mit einer Person aus Gruppe l ist ebenso ein umgekehrter Kontakt. Daher gilt:

$$G_l \kappa_{lk} = G_k \kappa_{kl}. \quad (10)$$

Gehen wir nun zu Anteilen an der Gesamtgruppe über und benennen diese Anteile mit kleinen Buchstaben, dann vereinfachen sich Gleichungen (8) und (9) nochmals:

$$\dot{n}_k(t) = [1 - n_k(t)] \sum_{l=0}^{g-1} \kappa_{kl}(t) \times \left[n_l(t - \tau_l^s) - n_l(t - \tau_l^e) - d_l^0(t - \tau_l^d(t)) + d_l^0(t - \tau_l^e) \right] \quad (11)$$

$$\dot{d}_k^0(t) = \lambda_k \left(t_k^d(t) \right) \dot{n}_k(t) \quad (12)$$

$$n_k(t) = \frac{N_k(t)}{G_k}, \quad d_k^0(t) = \frac{D_k^0(t)}{G_k}. \quad (13)$$

Wir weisen darauf hin, dass wegen Gl. (10) die Indizes von κ vertauscht sind.

Externe Infektiöse Eine Population kann nicht isoliert von ihrer Umgebung betrachtet werden. Umgekehrt ist es nicht machbar, für die Simulation eines Landkreises immer gleich die ganze Welt mit zu rechnen. Deshalb werden für externe Gruppen $g \leq l < g_{\text{ext}}$ die Infektiösen $I_l(t)$ bzw. ihr Anteil $i_l(t)$ durch eine plausible feste Zeitreihe und weitere Kontaktraten beschrieben. g_{ext} ist dabei die Gesamtzahl aller voll mit berechneten und externen Gruppen. Gl. (11) erweitert sich dann zu

$$\dot{n}_k(t) = [1 - n_k(t)] \left[\sum_{l=g}^{g_{\text{ext}}-1} \kappa_{kl}(t) i_l(t) + \sum_{l=0}^{g-1} \kappa_{kl}(t) \times \left[n_l(t - \tau_l^s) - n_l(t - \tau_l^e) - d_l^0(t - \tau_l^d(t)) + d_l^0(t - \tau_l^e) \right] \right]. \quad (14)$$

Abgeleitete Teilpopulationen Die Werte sind einer Studie des RKI entnommen [AdHB20]. Man beachte, dass die Anteile nicht auf die entdeckten, sondern

Symbol	Bedeutung	typisch	Einheit	Epidemiologische Modelle zur SARS-CoV-2-Ausbreitungsprognose
τ_k^h	Mittlere Zeit im Hospital bei Einweisung	14	d	Tab. 3: Parameter weiterer Teilpopulationen von Gruppe k.
τ_k^v	Mittlere Zeit, ab der kritische Fälle in die Intensivstation verlegt werden	1	d	
τ_k^i	Mittlere Zeit, über die kritische Fälle intensiv versorgt werden müssen	10	d	
η_k	Anteil im Krankenhaus zu Behandelnder an Infizierten	0.045		
θ_k	Anteil intensivmedizinisch zu Behandelnder an Infizierten	0.011		
μ_k	Anteil Sterbender an Infizierten	0.006		

auf die tatsächlichen Infizierten bezogen werden. Dahinter steht die Annahme, dass diese Anteile weitgehend konstant sind, während Anteile an den Entdeckten stark von den ergriffenen Maßnahmen abhängen.

Um politische Maßnahmen planen zu können, sind weitere Größen wichtig, z.B. die Zahl der Personen, die sich gleichzeitig in intensivmedizinischer Behandlung befinden werden. Der Vorteil des Modells ist, dass aus den Zeitreihen der jemals Infizierten diese und viele weitere Teilpopulationen abgeleitet werden können. Die Teilpopulationen müssen dabei nicht disjunkt sein. Damit können wir in Gruppe k folgende Anteile beschreiben:

$$D_k(t) = D_k^0(t - \tau_k^d(t)) \quad \text{entdeckt} \quad (15)$$

$$H_k(t) = \eta_k \left[N_k(t - \tau_k^d(t)) - N_k(t - \tau_k^d(t) - \tau_k^h) \right] \quad \text{hospitalisiert} \quad (16)$$

$$U_k(t) = \theta_k \left[N_k(t - \tau_k^d(t) - \tau_k^v) - N_k(t - \tau_k^d(t) - \tau_k^v - \tau_k^i) \right] \quad \text{intensiv} \quad (17)$$

$$M_k(t) = \mu_k N_k(t - \tau_k^d(t) - \tau_k^v - \tau_k^i) \quad \text{tot.} \quad (18)$$

Dabei wird angenommen, dass ernstlich Kranke in der Regel sofort nach ihrer Entdeckung eingeliefert werden und dass am Ende der mittleren Zeit in der Intensivmedizin ein Kranker entweder als gerettet in eine normale Station zurück verlegt werden kann oder aber gestorben ist. Sollten diese Annahmen zu vereinfachend sein, müssten weitere Zeitkonstanten eingeführt werden.

Reproduktionsrate Die zentrale Kenngröße einer Epidemie ist die Reproduktionsrate. Liegt sie über eins, steckt ein Infizierter mehr als eine weitere Person an und die Zahl der Infizierten wächst so lange exponentiell an, bis das Virus zunehmend auf bereits Immunierte trifft. Liegt die Reproduktionsrate dauerhaft unter eins, dann sinken die Fallzahlen hingegen exponentiell gegen null. Um noch analytische Ausdrücke zu erhalten, betrachten wir hier exemplarisch den Fall nur einer Gruppe mit konstanter Entdeckungsdauer und -rate. Hier gilt

$$d(t) = \lambda n(t - \tau^d), \quad (19)$$

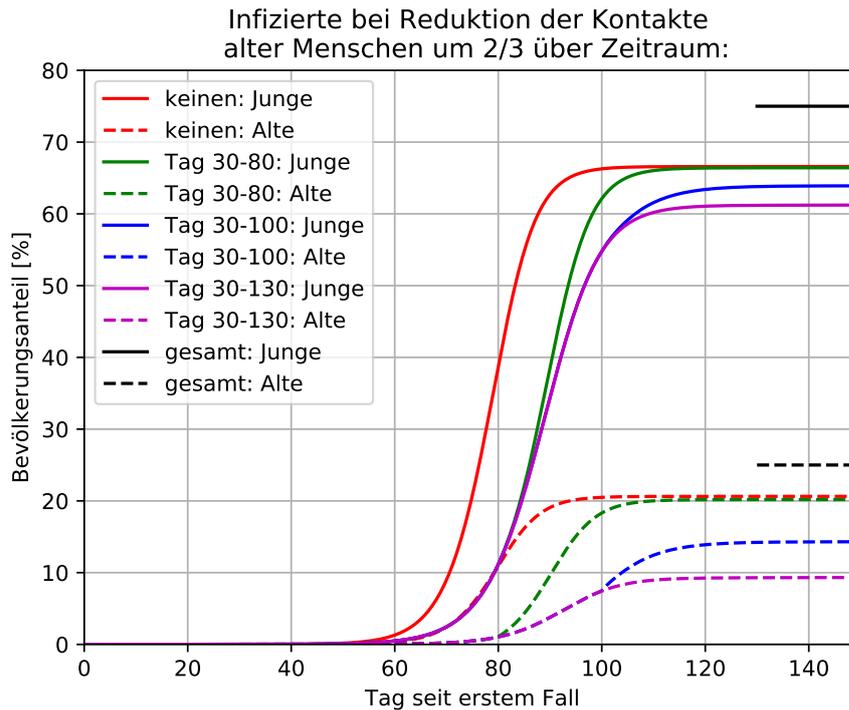


Abb. 2: Wirksamkeit der Isolation alter Menschen.

wobei der Gruppenindex weggelassen ist. Mit Hilfe des Mittelwertsatzes ergibt sich dann nach einiger Rechnung:

$$\begin{aligned} \dot{d}(t) &= \rho(t) \dot{d}(t - \tau(t)) && \text{mit} \\ \rho(t) &= \left(1 - \frac{d(t)}{\lambda}\right) \rho_0(t) && \text{(Reproduktionsrate)} \\ \rho_0(t) &= \kappa(t - \tau^d) \left[(\tau^d - \tau^s) + (1 - \lambda) (\tau^e - \tau^d) \right] && \text{(Basisreproduktionsrate)} \\ 0 < \tau(t) &\in [\tau^s, \tau^e] && \text{(Generationsdauer)}. \end{aligned} \quad (20)$$

Ist $\rho > 1$, so wächst die Zahl \dot{d} der neu Entdeckten gegenüber einem Wert in der Vergangenheit an. Für $\rho < 1$ fällt er ab. Die Reproduktionszahl sinkt gegenüber der Basisreproduktionszahl, wenn die Durchseuchung zunimmt. ρ ist proportional zur Kontaktrate, die τ^d Tage zuvor geherrscht hat. Die Reproduktionsrate sinkt, wenn Infizierte möglichst früh nach Beginn der infektiösen Phase entdeckt werden und wenn die Entdeckungsrate hoch ist.

Erste Ergebnisse Zu Beginn der Pandemie lautete eine der diskutierten Strategien, Rentner zu isolieren, bis der junge Teil der Bevölkerung hinreichend immunisiert ist. Der Gedanke ist zunächst nicht ganz abwegig, da junge Menschen kaum tödliche Verläufe zeigen und ein Lockdown hätte vermieden werden können. Abb. 2 zeigt simulierte Infektionsverläufe, wenn die Kontakte der Alten untereinander und zu den Jungen über unterschiedliche Zeiträume um $\frac{2}{3}$ reduziert würden. Als Basisreproduktionszahl wurde ein Wert von 2.0 angenommen, der ohne große Einschränkungen erreichbar ist. Es zeigt sich, dass eine vorzeitige Aufhebung der Isolation allen bis dahin erreichten Schutz zunichte macht. Im glimpflichsten Szenario kann nach einer Isolation von 100 Tagen die Zahl infizier-

ter alter Menschen um die Hälfte vermindert werden – ein eher ernüchterndes Ergebnis.

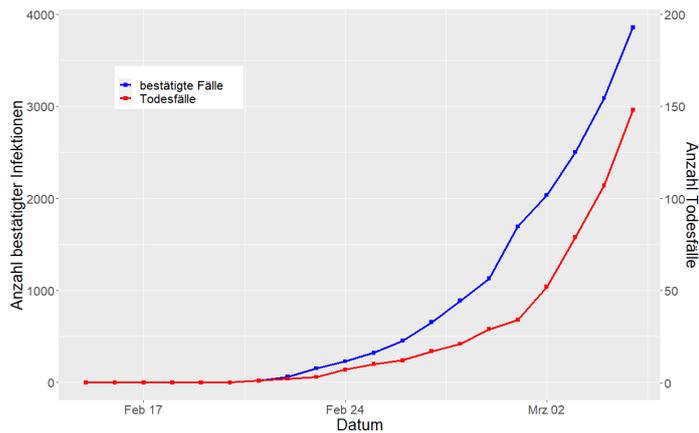
2.3. Ausblick

Ziel und Gegenstand aktueller Arbeiten ist die Simulation der Ausbreitungsdynamik unter Berücksichtigung verschiedener Teilgruppen wie oben dargestellt. Hierzu muss insbesondere die Matrix der kritischen Kontakte, κ , bedated werden. Im Rahmen von EpideMSE sollen dabei zunächst verschiedene Altersklassen als Teilgruppen betrachtet werden, die jeweils (in unterschiedlicher Weise) in verschiedenen sogenannten Sozialbereichen (z.B. Arbeitsplatz, Schule, Zuhause) aktiv sind. Zur Ableitung der jeweils charakteristischen Kontaktmuster sollen sowohl Daten aus europaweiten Umfragen (siehe z.B. [MHJ⁺08]) wie auch stochastische Modelle bzw. deren Ergebnisse, wie z.B. in [FAM⁺12], eingesetzt werden. Die Betrachtung verschiedener Sozialbereiche erlaubt dann auch die Untersuchung der Effekte verschiedener, *bereichsabhängiger* Maßnahmen. Des Weiteren sollen auch geographische Gebiete, durchaus auf verschiedenen Skalen (z.B. Regionen, Kreise), als Teilgruppen betrachtet werden und unter Berücksichtigung ihrer charakteristischen Kontaktmuster simuliert werden. Letzteres erfordert auch eine adäquate Abbildung der entsprechenden Mobilitätsströme zwischen den betrachteten Gebieten. Aktuelle Modellansätze auf globaler Ebene (weltweite Reiseströme) sind vorhanden [CDA⁺20]. Im Rahmen von EpideMSE wird aktuell untersucht, inwieweit sich diese Strategien auf lokaler Ebene und das oben dargestellte Simulationsmodell anwenden lassen.

Darüber hinaus liegt ein Hauptaugenmerk der anstehenden Arbeiten in der Anwendung und Umsetzung des Bayes'schen Zugang zur Modellparameteridentifikation und Unsicherheitsquantifizierung. Hierbei werden, basierend auf dem Satz von Bayes, gewisse, zu machende statistische Vorannahmen über gesuchte Modellparameter mithilfe der verfügbaren Daten und Beobachtungen kalibriert und so resultierende Verteilungen für die gesuchten Größen abgeleitet. Dieser Zugang hat, gegenüber klassischen Techniken, den wesentlichen Mehrwert, Unsicherheiten in Daten und Modellen in Verteilungen der resultierenden Modellparameter bzw. in quantitative Prognoseunsicherheiten zu übersetzen. Für bestimmte epidemiologische Modelle werden diese Techniken aktuell bereits erfolgreich angewendet [DZS⁺20, CDA⁺20] und sollen auf das dargestellte EpideMSE-Modell hin erweitert und angepasst werden. Im gleichen Kontext soll der Einfluss bereits getroffener Eindämmungsmaßnahmen möglichst gut quantitativ aus den erhobenen Daten abgeleitet werden, hierzu wird diese Aufgabenstellung als (Bayes'sches) inverses Problem formuliert und gelöst werden.

3. Datenbasierte Dunkelzifferschätzung

Ein kritischer Aspekt bei der Modellierung von Ausbreitungen von Viren in einer Population ist die tatsächliche Anzahl an Infektionen zu einem gegebenen Zeitpunkt. Verbreitet sich ein Virus unkontrolliert in einer Population, so ist diese Zahl insbesondere bei größeren Populationen nicht mehr bekannt – es existiert



Datenbasierte Dunkelziffer-schätzung

Abb. 3: Entwicklung der Anzahl der bestätigten Fälle und der Todesfälle zu Beginn des Ausbruchs in Italien.

eine signifikante Dunkelziffer an unentdeckten Infektionen. Eine genaue Kenntnis würde nämlich eine sehr schnelle Testung dieser Population verlangen, was die verfügbaren Ressourcen oft deutlich übersteigt. Dies wird insbesondere dann problematisch, wenn der Ausbruch erst sehr spät bemerkt wird. Abbildung 3 zeigt dies exemplarisch anhand der Situation in Italien. Dort gingen Mitte Februar die Zahlen der bestätigten Fälle und Todesfälle zeitgleich hoch. Da nun Erkrankte nicht an dem Tag sterben, an dem sie sich infiziert haben – sondern im Mittel erst zwei oder drei Wochen später – bedeutet dies, dass schon Ende Januar eine signifikante Zahl an tatsächlichen Infektionen vorhanden gewesen sein muss. Da sich dies nicht in den offiziellen Zahlen der gemeldeten Fällen widerspiegelt, bedeutet dies für Italien eine signifikante Dunkelziffer an infizierten und bereits genesenen Personen.

Im Folgenden wollen wir darauf eingehen, wie man auch ohne flächendeckende Tests¹ anhand frei verfügbarer Daten eine (konservative) Schätzung der Dunkelziffer erhalten kann. Dabei profitieren wir von einer ständig wachsenden Verfügbarkeit an Daten in nahezu allen Lebensbereichen. Dies liegt an immer besser werden Datenverarbeitungsgeräten, an den immer besser und flexibler werden den Datenspeicherungsoptionen und, nicht zuletzt, am fortschreitenden Ausbau von globaler Kommunikationstechnologie. Insbesondere lässt sich dies auch in der derzeitigen Corona-Pandemie beobachten: Es werden umfangreich Daten in Form von epidemischen Kennzahlen (Anzahl und Alter von Infizierten, Genesenen, Verstorbenen,.. usw.) erhoben, verarbeitet, aufbereitet, visualisiert, täglich aktualisiert und der Öffentlichkeit durch nahezu alle Medien zugänglich gemacht (in Deutschland von offizieller Seite durch das Robert-Koch-Institut, weltweit etwa durch die Johns-Hopkins-Universität). Eine derart reichhaltige und umfangreiche Erfassung war zu früheren Pandemien nicht verfügbar, insbesondere nicht während der Pandemie.

3.1. Stand der Wissenschaft

Die Größe der Dunkelziffer ist eng verbunden mit der Größe der tatsächlichen durchschnittlichen Sterblichkeit (*Letalität*) von Covid-19. Ist nämlich die Sterblich-

¹welche sinnvollerweise auch auf Antikörper testen, da die Krankheit für viele Betroffene nach zwei Wochen ausgestanden ist.

keit genau bekannt, so kann man anhand der Zahl der gemeldeten Verstorbenen auf die Zahl der tatsächlich Infizierten schließen. Dabei geht man implizit davon aus, dass die Zahl der Verstorbenen deutlich genauer erfasst wird, als die Zahl der Infizierten.²

Die Sterblichkeit von Covid-19 ist Gegenstand von vielfältigen Untersuchungen. [SSK⁺20] schätzen die tatsächliche Prävalenz (Anteil Infizierter an der Bevölkerung) in Gangelt, welcher mit etwa 15 % deutlich über dem offiziell gemeldeten Anteil von 3 % lag. In ihrer Studie gehen sie dabei noch weiter und leiten aus ihren Tests eine durchschnittliche Letalität (Sterblichkeit) ab, mit der sie auf die tatsächliche Zahl Infizierter in Deutschland geschlossen haben. Basierend auf ihrer geschätzten Letalität von etwa 0.37 % würde man eine Zahl an tatsächlich infizierten Personen erhalten, die mehr als 12-mal so groß wie die Zahl tatsächlich gemeldeter Fälle³ ist. Diese Zahl erscheint sehr hoch und gründet sich auch auf eine Datenbasis von wenigen Todesfällen (7), was die statistische Hochrechnung relativ unsicher macht.

[RHJ20] bestimmen die Letalität anhand der Beobachtung einer abgeschlossenen und gut untersuchten Population, nämlich die des Kreuzfahrtschiffs Diamond Princess. Dabei kommen sie auf etwa 1.3 % für die Diamond Princess. Diese Letalität lässt sich nun nicht ohne Anpassung auf andere Populationen übertragen, da die Population der Diamond Princess nicht repräsentativ für andere ist. Entsprechende Anpassungen von [RHJ20] liefern dann für China eine Letalität von 0.6 %.

[MBF⁺20] schätzen die Übersterblichkeit in der besonders von Covid-19 betroffenen Lombardei. Indem sie den Anteil der verstorbenen Bevölkerung als untere Grenze der Letalität betrachten und mit der Rate der positiven Tests korrigieren erhalten sie für die Lombardei eine Untergrenze für die Letalität von 0.61 %.

Diese Studien zeigen, dass verschiedene Ansätze deutlich unterschiedliche (und sogar inkompatible) Schätzungen für die Letalität und daher auch für die Dunkelziffer liefern. Im Folgenden wollen wir einen weiteren Ansatz betrachten, mit dem man die Größenordnung der Dunkelziffer abschätzen kann, basierend auf der Altersstruktur der Betroffenen.

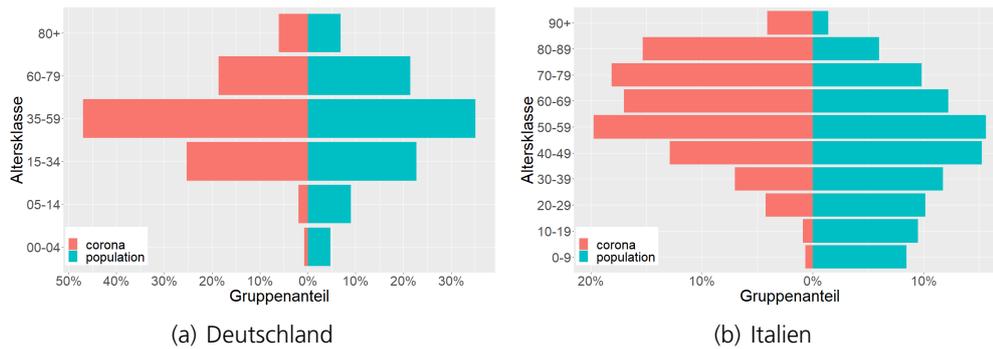
3.2. Vorarbeiten des Fraunhofer ITWM

Ausgangspunkt dieser Dunkelziffermodellierung war die Beobachtung, dass die altersbedingte Prävalenz⁴ der gemeldeten Infektionen sich für einzelne Länder teils stark unterscheidet, wie in Abbildung 4 dargestellt. Dort erkennt man, dass die Zahl der gemeldeten Fälle für Italien stark hin zu höheren Altersgruppen verschoben ist, was sich für Deutschland nicht beobachten lässt. Als maßgebliche Ursache für diese Diskrepanz erachten wir Unterschiede in der Testweise, welche zurückgehen auf die unterschiedlichen Stadien, in denen die nationalen Epidemien entdeckt wurden. In Deutschland wurde die Epidemie recht früh entdeckt,

²Betrachtungen der Übersterblichkeit für gewisse Regionen legen allerdings nahe, dass auch die Zahl der an Covid-19 verstorbenen Personen unterschätzt sein kann, siehe [MBF⁺20].

³Basierend auf den Zahlen der John-Hopkins-University vom 07.06.2020.

⁴Damit gemeint ist der Anteil an Infizierten in einer gegebenen Altersgruppe.



Datenbasierte Dunkelziffer-schätzung

Abb. 4: Altersverteilungen der gemeldeten Fälle für Deutschland (a) und Italien (b). Zeitpunkt der Datenerfassung war der 02.04.2020.

weswegen Infektionsketten noch einige Zeit teilweise verfolgt werden konnten. In Italien hingegen wurde die Epidemie erst bemerkt, als schon die ersten Todesfälle registriert wurden, was einige Wochen unkontrollierte Virusausbreitung ermöglicht hat. Das legt nahe, dass die Meldezahlen aus Italien stark von den schweren Verläufen geprägt wurden, die intensive ärztliche Betreuung verlangten. Da die Schwere der Verläufe stark mit dem Alter korreliert erklärt dies die starke Verzerrung in der Altersverteilung der gemeldeten Fälle.

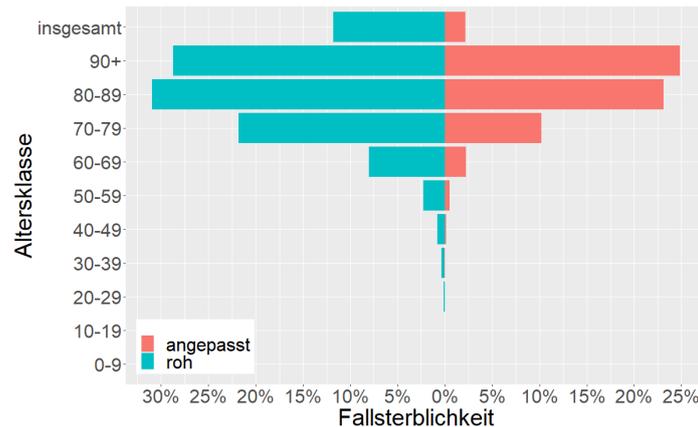
Diese Verzerrung korrigieren wir nun wie folgt. Wir gehen von der Annahme aus, dass die Gruppe der Infizierten hinreichend durchmischt ist. Damit sind einzelne Altersgruppen proportional zu der Anzahl ihrer Kontakte und ihrer Empfänglichkeit für das Virus betroffen: Je kontaktfreudiger/mobiler und empfänglicher eine Altersgruppe ist, desto stärker verbreitet sich das Virus in ihr. Die Altersabhängigkeit der Empfänglichkeit ist Gegenstand reger Forschungstätigkeit, insbesondere im Hinblick auf die Rolle von Kindern. In Ermangelung eines klaren Konsens gehen wir im Folgenden aber davon aus, dass die Infektiosität für alle Altersgruppen gleich ist. Sei nun \hat{P}_g die Prävalenz der gemeldeten Fälle innerhalb eines Landes für die Gruppe $g \in G_{\text{Land}}$, wobei jede Gruppe durch die Altersklasse⁵ und Geschlecht bestimmt wird. Für die tatsächliche Prävalenz P_g in der Gruppe g schätzen wir dann

$$P_g = \frac{\hat{P}_{g_{\max}}}{\hat{P}_g} \frac{N_{g_{\max}}}{N_g}, \quad (21)$$

wobei $g_{\max} = \arg \max_{g \in G} \hat{P}_g$ und N_g die erwartete Anzahl der Kontakte dieser Gruppe ist. Für die Anzahl der Kontakte N_g nutzen wir Ergebnisse aus [MHJ⁺08], welche Kontaktzahlen für verschiedene Altersgruppen in mehreren europäischen Ländern ermittelten. Aus den so geschätzten Werten für die tatsächliche Prävalenz lässt sich eine angepasste Fallsterblichkeit ableiten, siehe Abbildung 5. Diese Schätzung liefert zum Einen einen plausibleren Wert für die Gruppe der Menschen, die mindestens 90 sind, da ein Abnehmen der Fallsterblichkeit in dieser Gruppe unplausibel ist. Zum Anderen liefert die angepasste Fallsterblichkeit deutlich geringere Werte, welche besser mit den beobachteten Werten aus Ländern zusammenpassen, die keinen so massiven Ausbruch hatten. Für die Daten vom 02.04.2020 in Italien liefert dieser Ansatz eine Dunkelziffer von etwa 475 000 Infektionen zusätzlich zu den etwa 105 000 gemeldeten.

Ein weiterer Ansatz zur Dunkelzifferschätzung, welcher untersucht wird, basiert

⁵Diese Altersklassen können für verschiedene Länder unterschiedlich sein; die Daten für Deutschland vom RKI sind gröber eingeteilt, als die Daten für Italien vom ISS.



Planung von Teststrategien zur Reduzierung der Unsicherheit von Modellvorhersagen

Abb. 5: Angepasste und rohe Fallsterblichkeit für Italien. Zeitpunkt der Datenerfassung war der 02.04.2020.

auf der Annahme, dass das Virus in allen Ländern mehr oder minder die gleiche Letalität L aufweist. Die zugrundeliegende Beobachtung dabei ist, dass es bei den gemeldeten Infektionen eine viel größere Dunkelziffer geben wird, als bei den gemeldeten Todesfällen, da Todesfälle deutlich stärker auffallen, als beispielsweise Krankheitsverläufe mit nur milden Symptomen. Hat man nun die gemeldeten Todesfälle $D(t)$ zum Zeitpunkt t , so ergibt sich für die geschätzte Infektionszahl zum Zeitpunkt $t - t_D$

$$I(t - t_D) = \frac{D(t)}{L},$$

wobei t_D die mittlere Dauer ist, nach der eine Infektion tödlich endet. Diese geschätzte Infektionszahl I kann nun mittels Schätzungen über die aktuelle Ausbreitungsgeschwindigkeit des Virus auf t extrapoliert werden.

Aktuell wird daran gearbeitet, beide Schätzungsmethoden zu vergleichen und zu evaluieren.

3.3. Ausblick

Im weiteren Verlauf betrachten wir zwei Ziele. Einerseits sollen die obigen beiden Ansätze so detailliert betrachtet werden, dass sie möglichst kompatible Aussagen erzeugen. Andererseits ist es Ziel den Ansatz der sozial-gewichteten Prävalenz durch neuere Erkenntnisse ständig zu erweitern, wie zum Beispiel mit Erkenntnissen zu altersabhängigen Infektiositäten. Die methodischen Untersuchungen an dieser Stelle werden dabei durch ein begleitendes Forschungsprojekt von Fraunhofer gefördert.

4. Planung von Teststrategien zur Reduzierung der Unsicherheit von Modellvorhersagen

Die Durchführung von Tests bindet zeitlich und monetär Ressourcen. Die Entscheidung darüber, wie diese eingesetzt werden, ist mehrkriteriell: Einen Patienten mit Symptomen wird man eher testen als symptomfreie Personen. Gehören die unmittelbaren Kontakte einer positiv getesteten Person zu Hochrisikogruppen, wird

man diese eher testen als Kontakte mit mutmaßlich geringem Infektionsrisiko. Personen mit Berufen in systemrelevanten Sektoren sollten eher getestet werden. Diese Liste an Kriterien ließe sich noch weiter fortsetzen. Diese Kriterien entstehen letztlich aus dem Wunsch, dort Gewissheit zu haben, wo es am notwendigsten ist, um die Pandemie einzudämmen und den zu erwartenden Schaden möglichst gering zu halten. Bezogen auf die mathematische Modellbildung heißt das:

- Zunächst einmal ist es notwendig, die Unsicherheit von Modellvorhersagen zu kennen.
- In einem zweiten Schritt sind diese Unsicherheiten zu bewerten.
- Daraus ergeben sich in einem dritten Schritt Vorschläge und Strategien für Tests, die zielgerichtet an den Personengruppen durchgeführt werden, bei denen die Kenntnis des Infektionsstands die für die Modellvorhersagen wichtigsten Informationen liefert.

Diese drei Schritte stellen das Arbeitsprogramm in diesem Teilprojekt dar.

4.1. Stand der Wissenschaft

Die Quantifizierung der Unsicherheit von Modellvorhersagen erfolgt mit Techniken der Sensitivitätsanalyse [SRA⁺08], die sowohl bezüglich einzelner Modellparameter - wie beispielsweise die Infektionszeit oder die Erkrankungswahrscheinlichkeit - oder bezüglich der Modellstruktur - verwendet man eher einen systemischen Ansatz mit gekoppelten stochastischen Differentialgleichungen oder eine discrete event Simulation - durchgeführt wird. Diese Sensitivitätsanalyse liefert im Idealfall statistisch interpretierbare Konfidenzintervalle für die Modellvorhersagen. Diese Intervalle geben darüber Aufschluss, welche Modellparameter für eine verlässliche Vorhersage genauer bekannt sein sollten. Sind die Konfidenzintervalle bekannt, ist deren Wichtigkeit zu bewerten. Hier hilft eine mehrkriterielle Betrachtungsweise [BHS⁺18], da, wie in Abschnitt 4 skizziert, der Einfluss von Unsicherheiten ganz unterschiedlich bewertet werden kann. Sind die Unsicherheiten identifiziert, die am dringlichsten reduziert werden müssen, können mit Techniken der modellbasierten Versuchsplanung bestmögliche Teststrategien entwickelt werden. Aus mathematischer Sicht führt dies auf nichtlineare, nicht-konvexe Optimierungsprobleme, deren Lösung numerisch anspruchsvoll ist. Man kann jedoch zeigen, dass sich diese Mühe lohnt: Daten, die zielgerichtet erhoben wurden, machen erheblich zuverlässigere Modellvorhersagen möglich als eine ungeplante Datenerhebung [ARBB18].

4.2. Vorarbeiten des Fraunhofer ITWM

Das Abbilden realer Prozesse in mathematischen Modellen und das Nutzen dieser Modelle, um bisher unbekanntes Vorhersagen über die realen Prozesse zu gewinnen, ist eine der Kernkompetenzen des ITWM. Dazu gehören, neben den Vorhersagen selbst, auch die Quantifizierung ihrer Ungenauigkeiten - Beispiele sind [ABB⁺17], [BABB17], wo dies exemplarisch mit einem industriellen Anwen-

dungspartner für verfahrenstechnische Beispiele gezeigt wurde. Für den Umgang mit Unsicherheiten bei Optimierungsfragestellungen wurden ebenfalls methodische Beiträge in die praktische Anwendung gebracht [BBvH⁺17]. Der zentrale Punkt der Identifizierung einer optimalen Teststrategie auf Grundlage von unsicheren Modellvorhersagen spielt immer eine Rolle, wenn Daten erhoben werden müssen. Für Prozess- und Labordaten im Produktionskontext wurden Softwaremodule, die diese Technologie realisieren, entwickelt und in die Praxis gebracht [ABM⁺20]. Zwar fanden die genannten Vorarbeiten nicht im Anwendungskontext einer Pandemieausbreitung statt. Getreu dem Motto des ITWM-Gründers, Prof. Helmut Neunzert, »Mathematik ist Technologie«, ist mathematische Methode mit einem Werkzeugkasten vergleichbar, der in verschiedenen Anwendungsdomänen Nutzen stiften kann.

4.3. Ausblick

Gemäß des in Abschnitt 4 skizzierten Vorgehens besteht der erste Schritt im Projekt in der Quantifizierung der Unsicherheit von relevanten Modellvorhersagen. Das Modell in der ersten Ausbaustufe geht dabei von Alterskohorten aus, in und zwischen denen Infektionswahrscheinlichkeiten bestehen. Eine modellbasierte Teststrategie wird die intuitiv naheliegende Lösung bestätigen, dass ältere Personen eher getestet werden sollten als jüngere. Dieses einfache Modell stellt Benchmark und Test der Herangehensweise dar. Der modellbasierte Ansatz wird dann Potential entfalten, wenn das Modell hinsichtlich der Infektionsketten örtlich und zeitlich verfeinert werden kann. Die dann entstehende Komplexität ist nicht mehr intuitiv handhabbar und macht die Anwendung von systematischer mathematischer Algorithmik zur Optimierung von Teststrategien zwingend erforderlich.

5. Mehrkriterielle Entscheidungsunterstützung

Die wichtigsten Ziele zur Bewältigung der Covid19-Pandemie sind derzeit der Erhalt der Leistungsfähigkeit des Gesundheitssystems und die Minimierung von Covid19-bedingten Todesfällen. Teilweise im Widerspruch dazu stehen aber die Ziele Rettung und Stabilisierung der Wirtschaft und Wahrung der individuellen Freiheitsrechte der Bevölkerung. Entscheidungsträger müssen diese konfliktären Ziele gegeneinander abwägen. Dabei ist es fundamental, dass sie die Auswirkungen ihrer Entscheidungen abschätzen und der Bevölkerung erklären. Mit Hilfe von Entscheidungsunterstützungswerkzeugen, die auf der Basis von mehrkriterieller Optimierung arbeiten sollen Maßnahmenvorschläge transparent miteinander verglichen werden können. Außerdem ist es einfach im Rahmen von What-If-Szenarien Vorschläge zu analysieren und zu bewerten.

5.1. Stand der Wissenschaft

Relevante Entscheidungen in der Praxis haben in der Regel einen mehrkriteriellen Charakter, d.h. es müssen verschiedene zumindest partiell widersprüchliche Zielgrößen gegeneinander abgewogen werden. Da es für diese selten eine eindeutige Gewichtung gibt, werden die zugrundeliegenden Optimierungsprobleme multikriteriell formuliert und gelöst. Eine Übersicht zur multikriteriellen Optimierung findet man in [Ehr05]. Die Komplexität und methodische Herangehensweise der Lösung eines multikriteriellen Entscheidungsproblems hängt dabei sehr von dem Charakter des zugrundeliegenden Optimierungsproblems ab. Die Entwicklung praxistauglicher Algorithmen zur Gewinnung von sparsamen diskreten Repräsentantenmengen zur Approximation von Paretomengen [SRH⁺15], zur echtzeitfähigen Interpolation der diskreten Repräsentanten im Falle von konvexen [CM10] und nicht-diskreten [ANJ⁺18] Pareto-Fronten und zur Beherrschung gemischtganzzahliger multikriterieller Optimierung [Her11] ist eine langjährige Forschungsdomäne am Fraunhofer ITWM.

5.2. Vorarbeiten des Fraunhofer ITWM

In allen Wirtschaftszweigen und öffentlichen Einrichtungen müssen täglich komplexe Entscheidungen getroffen werden, die monetäre, gesundheitliche und soziale Auswirkungen haben. Fast immer gilt es dabei positive und negative Folgen für verschiedene Interessengruppen gegeneinander abzuwägen. Je informierter solche Entscheidungen getroffen werden können, desto leichter lassen sich auch die Nachteile eines Beschlusses oder einer Maßnahme kommunizieren. Am Fraunhofer ITWM wurden daher in der Vergangenheit Softwarewerkzeuge entwickelt, die eine transparente Analyse der Auswirkungen solcher komplexer Entscheidungen ermöglichen. Das betrifft einerseits strategische und taktische Fragestellungen in der produzierenden Industrie nach einer guten Auslegung oder Fahrweise einer Anlage. Beispiele sind eine Software zur Auslegung von Freiflächenphotovoltaikanlagen [BEPS14] und zum Risikoassessment von Supply-Chains [ADH⁺20], sowie vielfältige Werkzeuge zur mehrkriteriellen Optimierung in der Verfahrenstechnik [BMS⁺15]. Genauso unterstützt das ITWM aber auch bereits Entscheidungen im Gesundheitsbereich. Beispiele dafür sind die mehrkriterielle Optimierung von Strahlentherapieplänen zur Abwägung von Heilungschancen und Nebenwirkungsrisiken [TSS⁺11], aber auch die Abwägung von Ressourcenallokationen im Gesundheitsbereich [SS19]. In allen genannten Fällen können kleine Änderungen im System große Folgen generieren, die nur mit Hilfe von mathematischen Simulationen bestimmt werden können. Diese können im Rahmen einer automatisierten mehrkriteriellen Optimierung oder einer manuellen What-If-Analyse angesteuert werden um die Effekte verschiedener Maßnahmenbündel zu vergleichen.

5.3. Ausblick

Im Projekt EpideMSE sollen die Auswirkungen verschiedener Maßnahmenkombinationen prognostiziert werden. Außerdem sollen gute Kompromisse für Maß-

nahmen vorgeschlagen werden. Der Detaillierungsgrad von Maßnahmen hängt dabei von der Modellierungstiefe der Prognosemodelle ab. Sie sollen sowohl Teststrategien als auch regionale Ziele der Reinfektionszahlen zur Erreichung eines sozialverträglichen Kompromisses zwischen Shutdown und Zusammenbruch des Gesundheitssystems. Dabei werden die prognostizierte Belastung von Intensivstationen und Beatmungsplätzen explizit mit berücksichtigt.

6. Softwarelösungen zur epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung

Im März 2020 rief die Bundesregierung zusammen mit sieben Start-Up-Unternehmen zum #WirVsVirus-Hackathon⁶ auf, dem bisher größten Hackathon weltweit. In nur 48 Stunden entwickelten 28 361 Programmiererinnen und Programmierer mehr als 1500 Softwarelösungen. Diese beeindruckenden Prototypen werden nun unter der Schirmherrschaft von Prof. Dr. Helge Braun, Chef des Bundeskanzleramtes, in einem Umsetzungsprogramm weiterentwickelt.

6.1. Stand der Technik

Die Softwarelösungen zur Bewältigung der Corona-Krise adressieren sehr unterschiedliche Einsatzgebiete. In den Medien sehr intensiv diskutiert wurden und werden Contact-Tracing-Apps⁷. Die Top 20 des #WirVsVirus-Hackathons betreffen im Wesentlichen Logistik und Kommunikation. Für Epidemiologie selbst gibt es jedoch nur vereinzelt Simulationstools.

Für die Datenaufbereitung der Epidemie-Verläufe von Covid-19 gibt es zahlreiche Dashboards. Das Robert-Koch-Institut (RKI) veröffentlicht die aktuellen Fallzahlen auf ihrer Internetseite⁸. Dort findet man ein Dashboard zu Covid-19 mit Fallzahlen nach Bundesland und Landkreis. Neben den aktuellen Fallzahlen werden dort auch die zeitlichen Verläufe der Infektionszahlen dargestellt. Diese Darstellung beinhaltet auch die Anzahl der Fälle nach Alter und Geschlecht getrennt. Ebenfalls täglich aktualisiert wird vom RKI eine Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland (Nowcasting). Dabei wird die zeitabhängige Reproduktionszahl R für Gesamtdeutschland berechnet und in Form einer Excel-Datei zum Herunterladen angeboten. Die der Schätzung zugrundeliegenden mathematischen Methoden werden in [adHH20] erläutert.

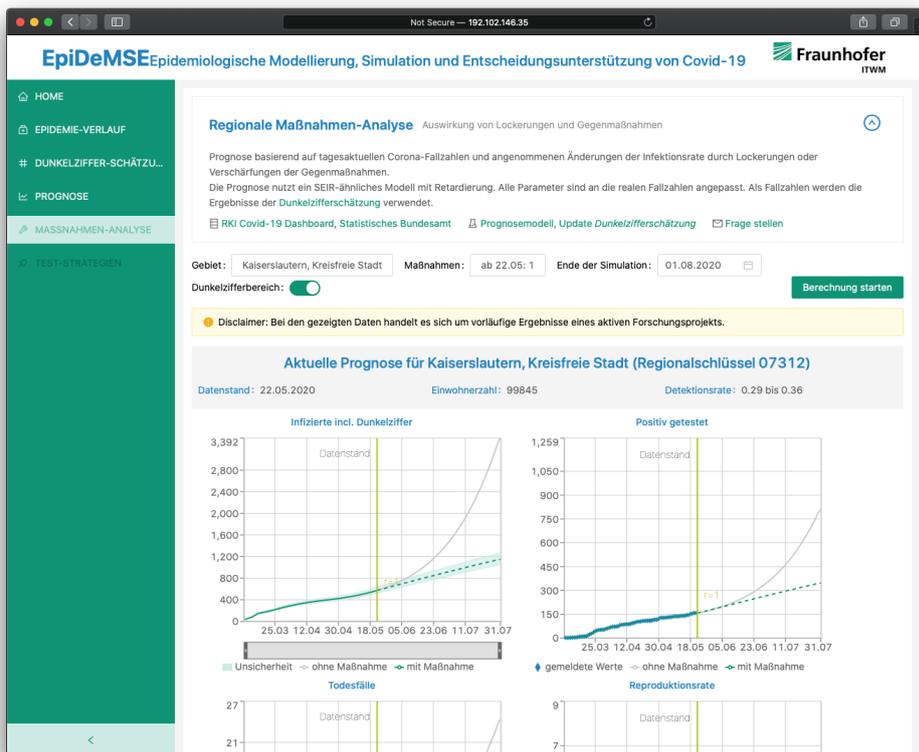
Ein erster Schritt zur Entscheidungsunterstützung kann sein, die aktuellen Fallzahlen z.B. im RKI Covid-19 Dashboard mit den aktuellen Zahlen zur Intensivbettenbelegung auf Landkreisebene zu vergleichen. Die aktuelle Bettenbelegung findet sich im Dashboard DIVI-Intensivregister Germany⁹. Da beide Dashboards getrennt sind, muss ein Abgleich der aktuellen Covid-19-Infektionszahlen mit der Anzahl der freien Intensivbetten manuell erfolgen. Bisher fehlen jedoch in beiden

⁶<https://wirvsvirushackathon.org>

⁷aktuelle Lösung <https://github.com/corona-warn-app/cwa-documentation>

⁸https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/nCoV.html

⁹<https://npgeo-corona-npgeo-de.hub.arcgis.com/app/3a132983ad3c4ab8a28704e9addefaba>



Softwarelösungen zur epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung

Abb. 6: Screenshot des EpiDeMSE-Tools: Prognose des Epidemie-Verlaufs für Kaiserslautern

Dashboards die Prognosemöglichkeiten. Eine solche ist im Tool CovidSim¹⁰ integriert. Basierend auf einem systemischen Modell können beispielsweise verschiedene Reproduktionszahlen eingestellt werden, um zu studieren, wie sich Covid-19 ausbreitet. Diese Prognosemodell basiert jedoch nicht auf echten Fallzahlen, sondern erlaubt nur ein prinzipielles Verständnis für den Einfluss der Reproduktionszahl oder der anderen Modellparameter. Für Details zum Modell verweisen wir auf den dazugehörigen Report [SNS⁺20].

6.2. Vorarbeiten des Fraunhofer ITWM: das EpiDeMSE-Tool

Das Fraunhofer ITWM hat in Eigenforschung ein Simulationstool zur epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung entwickelt. Das Tool »EpiDeMSE« vereint die Vorteile der datenbasierten Dashboards mit der Prognose der Covid-19-Ausbreitung. Dabei werden die Prognosen nicht nur für eine fiktive Population mit fiktiven Modellparametern erstellt, sondern an die lokalen Fallzahlen angepasst.

Darüber hinaus bietet das EpiDeMSE-Tool die Möglichkeit, Maßnahmen regional zu planen (siehe Abbildung 6). Neben der Aufbereitung der Epidemie-Daten (in Abbildung 7 ist beispielsweise die Ansicht der Neuinfektionen der letzten sieben Tage dargestellt) verfügt das EpiDeMSE-Tool noch über Analysen der Dunkelziffer (siehe Abbildung 8). Diese Schätzungen der Dunkelziffer fließen ebenfalls in die Prognose des weiteren Epidemie-Verlaufs ein und können so helfen, Worst-Case-

¹⁰<http://covidsim.eu>

Szenarien zu entwickeln.

Softwarelösungen zur epidemiologischen Modellierung, Simulation und Entscheidungsunterstützung

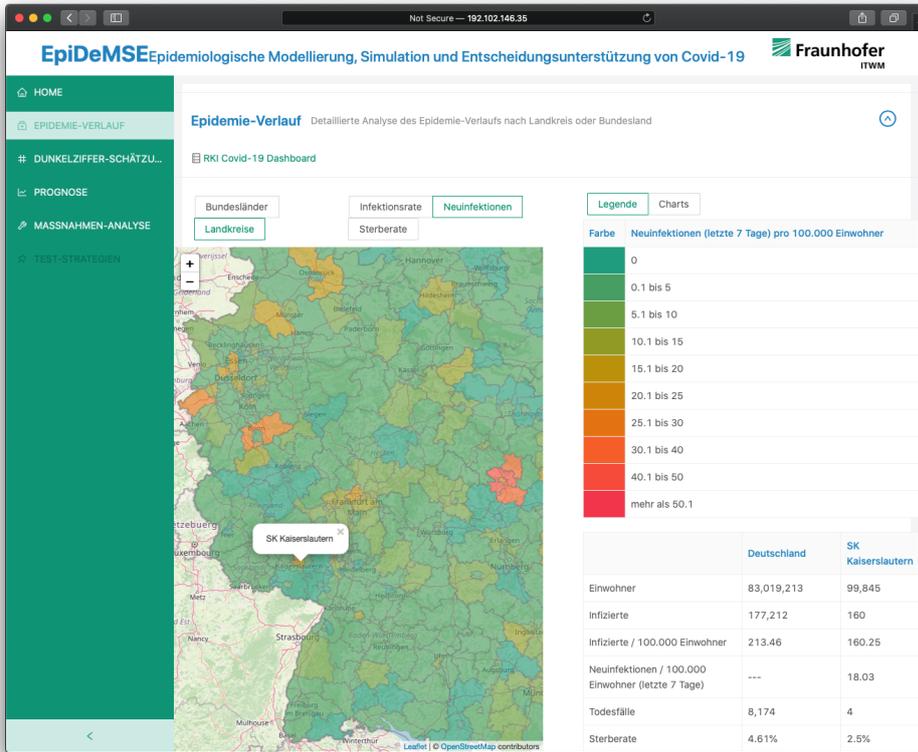


Abb. 7: Screenshot des EpiDeMSE-Tools: Einfärbung der Landkreise nach Neuinfektionen der letzten sieben Tage pro 100.000 Einwohner

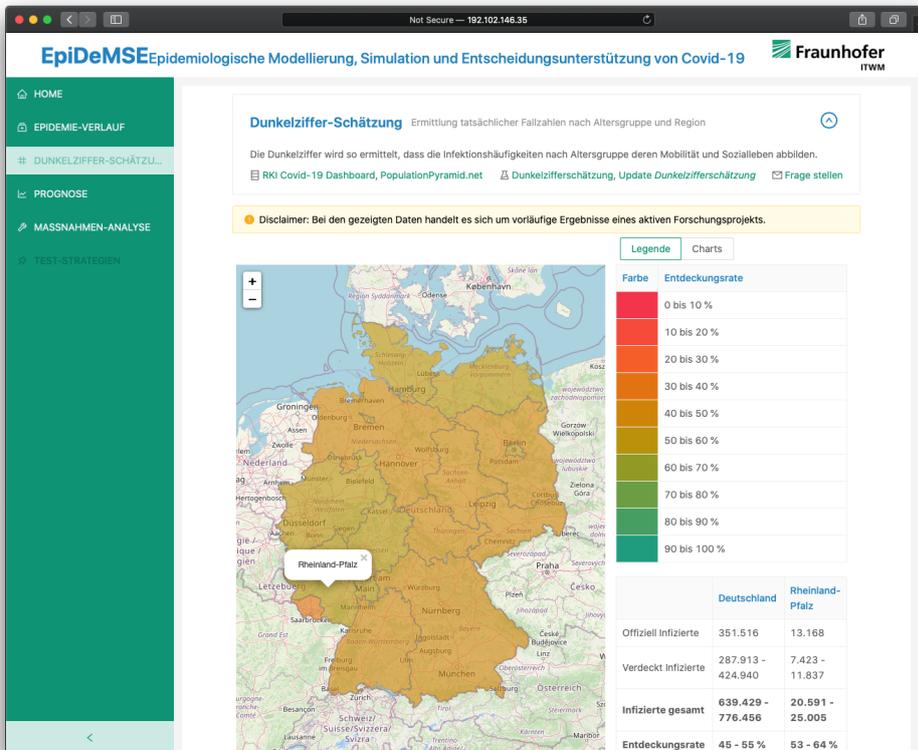


Abb. 8: Screenshot des EpiDeMSE-Tools: Einfärbung der Bundesländer nach Dunkelziffer

7. Zusammenfassung und Ausblick

Mit dem EpideMSE-Tool in seiner Version 1.0 steht lokalen Entscheidungstragenden bereits ein umfangreiches Tool zur Verfügung, mit dem der Ist-Zustand der lokalen Epidemie-Situation analysiert und das weitere Infektionsgeschehen prognostiziert werden kann.

Wir haben bereits mit Arbeiten begonnen, die die Prognose der Ausbreitung von Covid-19 in Deutschland verbessern. Momentan wird in dem EpideMSE-Tool der Epidemie-Verlauf für alle Menschen eines Gebiets gemeinsam simuliert. Im nächsten Schritt wollen wir jedoch die Bevölkerung zusätzlich in Altersgruppen teilen und ihre unterschiedlichen Infektionsrisiken bei den Prognosen mitberücksichtigen. Weiterhin sollen die Auswirkungen des Epidemie-Verlaufs auch näher am Gesundheitssystem evaluiert werden, d.h. die Prognosen umfassen auch die Belastung der Intensivstationen.

Eine deutliche Verbesserung der Prognose-Qualität erwarten wir auch dadurch, dass wir Regionen miteinander verknüpfen. Vielleicht mag eine typische Pendlerstadt heute wenig Neuinfektionen haben, morgen kann dies leicht wieder anders aussehen, wenn Neuinfektionen von den Pendelnden mitgebracht werden. Und das gilt natürlich auch umgekehrt. Mobilität ist ein entscheidender Faktor bei der Weiterverbreitung von Infektionen. Daher ist das nächste geplante Modul die geografischere Simulation der Epidemie-Verläufe mit Austausch der verschiedenen Regionen.

Die Planung von Teststrategien ist ein wichtiger Baustein, um Containment-Maßnahmen gezielt an bestimmte Teilpopulationen oder Regionen anzupassen. Daher wird im Laufe der nächsten sechs Monate das EpideMSE-Simulationstool um zusätzlich ein Modul erweitert werden, mit dem effektive Teststrategien entwickelt werden können.

A. Referenzen

- [ABB⁺17] N. Asprion, S. Blagov, R. Böttcher, J. Schwientek, J. Burger, E. von Harbou, and M. Bortz, *Simulation and Multi-criteria Optimization under Uncertain Model Parameters of a Cumene Process*, *Chemie Ingenieur Technik* **89** (2017), no. 5, 665–674.
- [ABM⁺20] N. Asprion, R. Böttcher, J. Mairhofer, M. Yliruka, J. Höller, J. Schwientek, C. Vanaret, and M. Bortz, *Implementation and Application of Model-Based Design of Experiments in a Flowsheet Simulator*, *J. Chem. Eng. Data* **65** (2020), 1135–1145.
- [ADH⁺20] Heiner Ackermann, Erik Diessel, Michael Helmling, Christoph Hertrich, Neil Jami, and Johanna Schneider, *Decision support for material procurement*, *Operations Research Proceedings 2019*, Springer, 2020, to appear.
- [AdHB20] M An der Heiden and U Buchholz, *Modellierung von beispieleszenarien der sars-cov-2-epidemie 2020 in deutschland*, Unter Mitarbeit des Robert Koch-Instituts (2020).
- [adHH20] Matthias an der Heiden and Osamah Hamouda, *Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland - Nowcasting*, *Epidemiologisches Bulletin* **17** (2020), 10–16.
- [ANJ⁺18] Tommy Andersson, Dimitri Nowak, Tomas Johnson, Andreas Mark, Fredrik Edelvik, and Karl-Heinz Küfer, *Multiobjective optimization of a heat-sink design using the sandwiching algorithm and an immersed boundary conjugate heat transfer solver*, *Journal of Heat Transfer* **140** (2018), no. 10.
- [ARBB18] N. Asprion, J. Ritter, R. Böttcher, and M. Bortz, *Model-Based Design of Experiments Using a Flowsheet Simulator*, *Comp. Aided Chem. Eng.* **43** (2018), 43–48.
- [BABB17] J. Burger, N. Asprion, S. Blagov, and M. Bortz, *Simple Perturbation Scheme to Consider Uncertainty in Equations of State for the Use in Process Simulation*, *J. Chem. Eng. Data* **62** (2017), no. 1, 268–274.
- [BB11] Edoardo Beretta and Dimitri Breda, *An seir epidemic model with constant latency time and infectious period*, *Mathematical biosciences and engineering : MBE* **8** (2011), 931–52.
- [BBvH⁺17] M. Bortz, J. Burger, E. von Harbou, M. Klein, J. Schwientek, N. Asprion, R. Böttcher, K.-H. Küfer, and H. Hasse, *Efficient Approach for Calculating Pareto Boundaries under Uncertainties in Chemical Process Design*, *Industrial & Engineering Chemistry Research* **56** (2017), no. 44, 12672–12681.

-
-
- [BEPS14] Martin Bischoff, Hendrik Ewe, Kai Plociennik, and Ingmar Schüle, *Multi-objective planning of large-scale photovoltaic power plants*, Operations Research Proceedings 2012, Springer, 2014, pp. 333–338.
- [Ber60] Daniel Bernoulli, *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir*, Histoire de l'Acad., Roy. Sci.(Paris) avec Mem (1760), 1–45.
- [BHS⁺18] M. Bortz, J. Höller, J. Schwientek, R. Böttcher, O. Hirth, and N. Asprion, *Experimental Design in a Multicriteria Context: An Adaptive Scheme*, IFAC PapersOnLine **51** (2018), no. 2, 747–752.
- [BMS⁺15] Michael Bortz, Volker Maag, Jan Schwientek, Regina Benfer, Roger Böttcher, Jakob Burger, Erik von Harbou, Norbert Asprion, Karl-Heinz Küfer, and Hans Hasse, *Decision support by multicriteria optimization in process development: An integrated approach for robust planning and design of plant experiments*, Computer Aided Chemical Engineering, vol. 37, Elsevier, 2015, pp. 2063–2068.
- [CDA⁺20] Matteo Chinazzi, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, Cécile Viboud, Xinyue Xiong, Hongjie Yu, M. Elizabeth Halloran, Ira M. Longini, and Alessandro Vespignani, *The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak*, Science **368** (2020), no. 6489, 395–400.
- [CM10] David Craft and Michael Monz, *Simultaneous navigation of multiple pareto surfaces, with an application to multicriteria imrt planning with multiple beam angle configurations*, Medical physics **37** (2010), no. 2, 736–741.
- [Coo67] K.L. Cooke, *Functional-differential equations: Some models and perturbation problems*, Differential equations and dynamical systems (1967).
- [DZS⁺20] Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann, *Inferring change points in the spread of covid-19 reveals the effectiveness of interventions*, Science (2020).
- [EB01] W. Ma Y. Takeuchi E. Beretta, T. Hara, *Global asymptotic stability of an sir epedemic model with distributed time delay*, Nonlinear Analysis **47** (2001), 4107–4115.
- [Ehr05] Matthias Ehrgott, *Multicriteria optimization*, vol. 491, Springer Science & Business Media, 2005.
- [FAM⁺12] Laura Fumanelli, Marco Ajelli, Piero Manfredi, Alessandro Vespignani, and Stefano Merler, *Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread*, PLOS

- [GAB⁺10] Bruno Gonçalves, Marco Ajelli, Duygu Balcan, Vittoria Colizza, Hao Hu, José Ramasco, Stefano Merler, and Alessandro Vespignani, *Comparing large-scale computational approaches to epidemic modeling: agent based versus structured metapopulation models*, APS Meeting Abstracts, 2010.
- [Her11] Jorge Ivan Serna Hernandez, *Multi-objective optimization in mixed integer problems: With application to the beam selection optimization problem in imrt*, mbv, Mensch-und-Buch-Verlag, 2011.
- [Het95] van den Driessche P Hethcote, H.W., *An sis epidemic model with variable population size and a delay*, J. Math. Biol. **34** (1995), 177–194.
- [Het00] Herbert W Hethcote, *The mathematics of infectious diseases*, SIAM review **42** (2000), no. 4, 599–653.
- [KK03] Qamar Khan and Edfamana Krishnan, *An epidemic model with a time delay in transmission*, Applications of Mathematics **48** (2003), 193–203.
- [KW27] McKendrick AG Kermack WO, *A contribution to the mathematical theory of epidemics*, 700–721.
- [MBF⁺20] C. Modi, V. Boehm, S. Ferraro, G. Stein, and U. Seljak, *How deadly is COVID-19? A rigorous analysis of excess mortality and age-dependent fatality rates in Italy*, 1–12.
- [MHJ⁺08] J. Mossong, N. Hens, M. Jit, P. Beutels, and et al. Auranen, K., *Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases*, PLoS Medicine **5** (2008), no. 3, 0381–0391.
- [MHPPS20] Michael Meyer-Hermann, Iris Pigeot, Viola Priesemann, and Anita Schöbel, *Adaptive Strategien zur Covid-19-Eindämmung*, <https://www.fraunhofer.de/de/presse/presseinformationen/2020/april/strategien-zur-eindaemmung-der-covid-19-pandemie.html>, April 2020.
- [Nik20] Michael Nikolaou, *A fundamental inconsistency in the sir model structure and proposed remedies*, medRxiv (2020).
- [RA12] Fathalla A. Rihan and M. Naim Anwar, *Qualitative analysis of delayed sir epidemic model with a saturated incidence rate*, International Journal of Differential Equations (2012).
- [RHJ20] T. W. Russell, J. Hellewell, and et al. Jarvis, C. I., *Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship*, Euro Surveill. **25** (2020), no. 12.
- [SNS⁺20] Kristan Schneider, Gideon A. Ngwa, Markus Schwehm, Linda Eichner, and Martin Eichner, *The covid-19 pandemic preparedness simulation*

tool: *Covidsim*, 2020.

- [SR13] Constantinos I Siettos and Lucia Russo, *Mathematical modeling of infectious disease dynamics*, *Virulence* **4** (2013), no. 4, 295–306.
- [SRA⁺08] A. Saltelli, M. Ratto, T. Andres, Campolongo F., J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis. The Primer*, John Wiley & Sons, Ltd, 2008.
- [SRH⁺15] Christian Schumann, Christian Rieder, Sabrina Haase, Katrin Teichert, Philipp Süß, Peter Isfort, Philipp Bruners, and Tobias Preusser, *Interactive multi-criteria planning for radiofrequency ablation*, *International journal of computer assisted radiology and surgery* **10** (2015), no. 6, 879–889.
- [SS19] Johanna Schneider and Michael Schröder, *Simulation-based location optimization of ambulance stations*, *Operations Research Proceedings 2018*, Springer, 2019, pp. 143–149.
- [SSK⁺20] H. Streeck, B. Schulte, B. M. Kümmerer, E. Richter, and et al. Höller, T., *Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event [Preprint]*, 1–37.
- [TSS⁺11] K Teichert, P Süß, JI Serna, M Monz, KH Küfer, and C Thieke, *Comparative analysis of pareto surfaces in multi-criteria imrt planning*, *Physics in Medicine & Biology* **56** (2011), no. 12, 3669.
- [YL06] Ping Yan and Shengqiang Liu, *Seir epidemic model with delay*, *The ANZIAM Journal* **48** (2006), no. 1, 119–134.