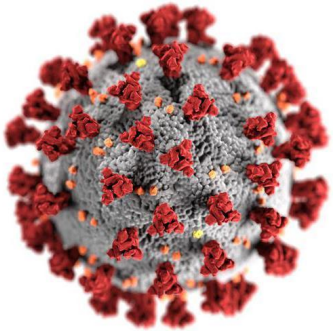
Teamprojekt Pandemiemodellierung

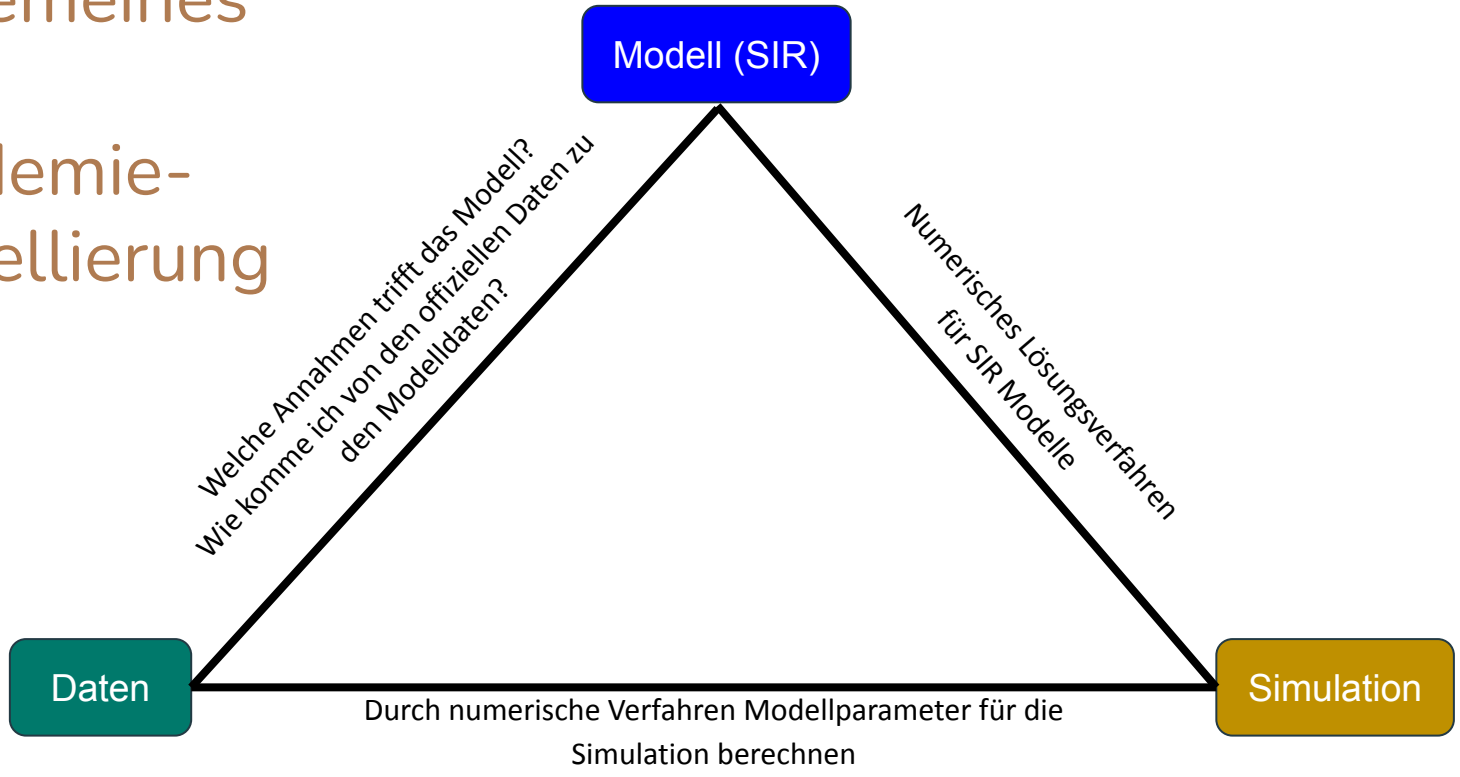
Abschlusspräsentation

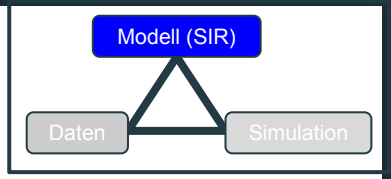


Übersicht

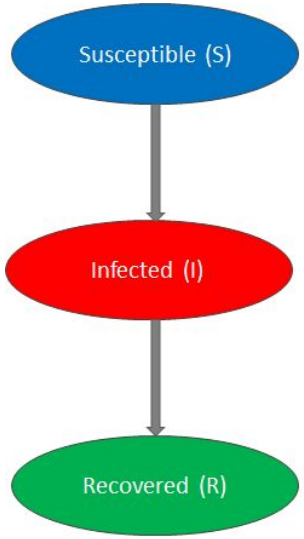
1. Allgemeines zur Pandemiemodellierung
2. Modellbildung
3. Daten
4. Parametrisierung
5. Numerische Berechnungen
6. Ergebnis
7. Fazit & Ausblick

Allgemeines zur Pandemie-modellierung





SIR-Modell

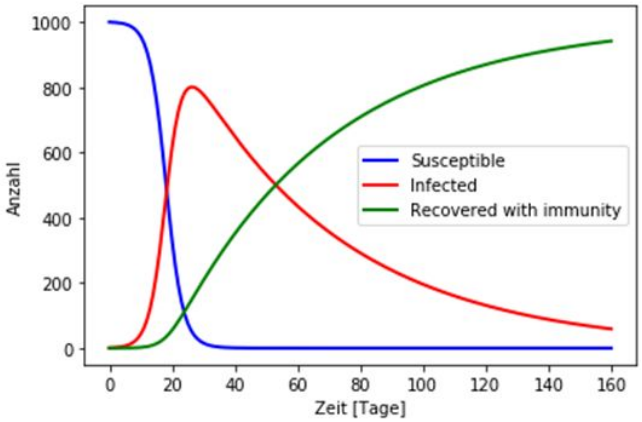


$$\frac{dS}{dt} = -\beta S I$$

$$\frac{dI}{dt} = \beta S I - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

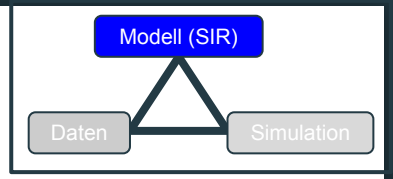
$$N = S(t) + I(t) + R(t)$$



$$N = Gesamtbevölkerung$$

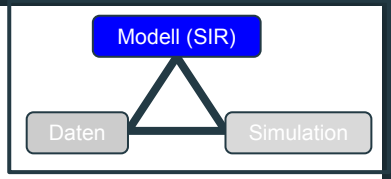
$$\beta = \frac{\beta}{N} = Transmissionsrate$$

$$\gamma = Erholungsrate$$



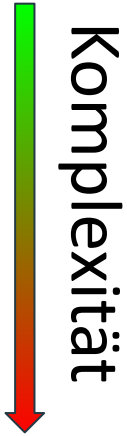
Getroffene Annahmen

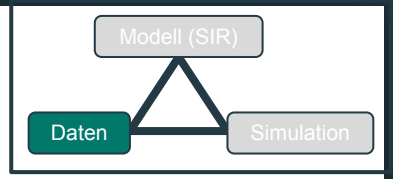
- konstante Populationsgröße
- keine Inkubationszeit
- homogene Bevölkerung
- keine Kreuzimmunität
- Saisonalität
- Anfangsbedingung



Weitere Modelle

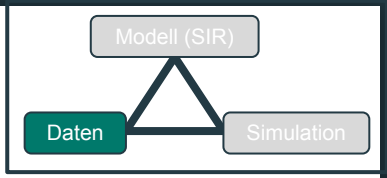
- SI: Nur Susceptibles und Infected
- SIS: Keine Immunität nach Ansteckung
- SIR - Vital Dynamics: Natürliche Geburten und Tode
- SIRD: Separate Betrachtung der Toten
- SEIR: Betrachtung der Inkubationszeit
- SIR-X: Quarantäne- und Isolierungsmaßnahme



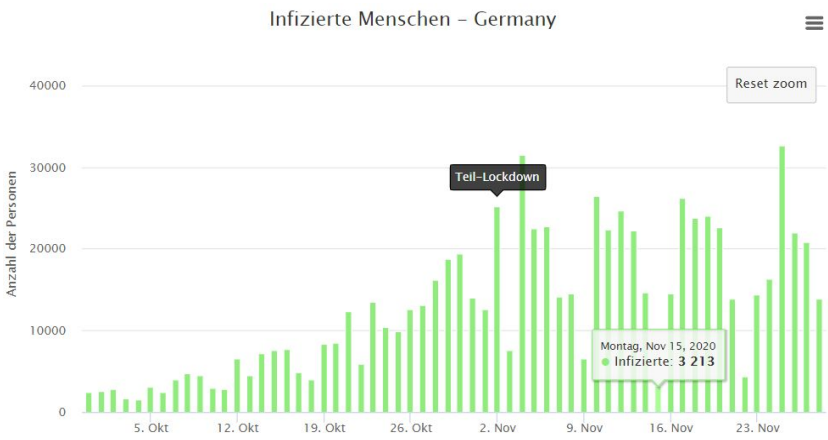


Allgemeines zur Datenbasis

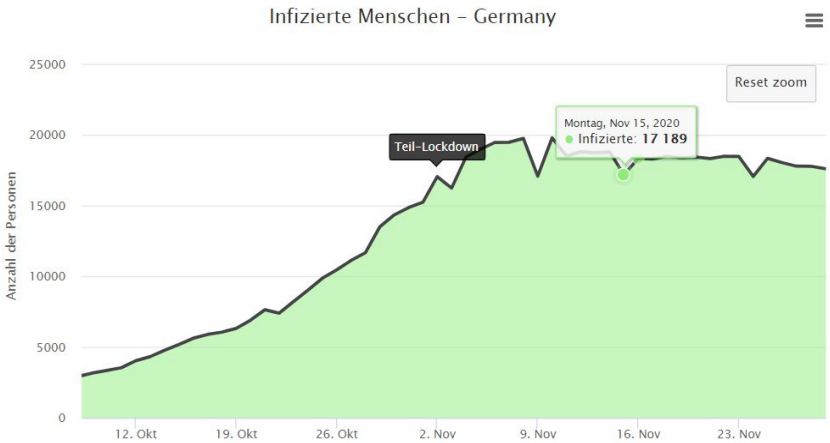
- Coronavirus (SARS-Cov-2) vs. coronavirus disease (Covid-19)
  - Ein “Coronafall” ist eine positiv auf das Coronavirus getestete Person, nicht zwingend jemand, der an Covid-19 erkrankt ist
- Datenübermittlung:
  - Testzentren/Hausarztpraxen ⟶ Gesundheitsämter ⟶ Landesbehörden ⟶ Robert Koch Institut (RKI)
- Unsere Quelle
  - ourworldindata.com ⟶ JHU, ECDC und WHO



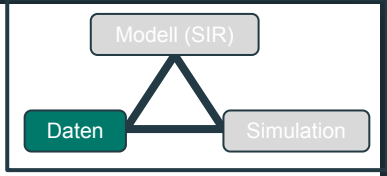
Allgemeines zur Datenbasis



keine Glättung

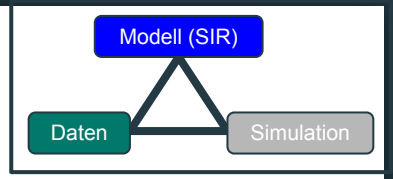


7 Tage Glättung



Modelldaten vs. offizielle Coronafallzahlen

- “Ein Modell ist ein vereinfachtes Abbild der Wirklichkeit”
- Fehlerhaftes Reporting
- unbekannte Parameter
  - Inkubationszeit
  - Letalität (Fall-Verstorbenen-Anteil)
  - Reinfektionen denkbar nach einigen Monaten
  - Kontagiosität (Dauer der Ansteckungsfähigkeit) circa 10 Tage
  - Unklar, wie Impfung die Ansteckungsfähigkeit verringert bzw. die Dauer verkürzt
- “Susceptibles” sind nicht konstant
- “Removed” in Hinblick auf Letalität und Reinfektion
- “Infectious” vs. Neuinfektionen



Berechnung von S, I und R

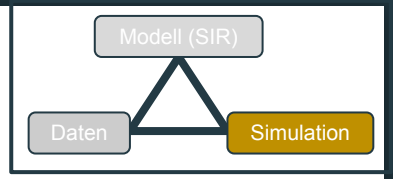
$$I_M(t) = I_M(t-1) + I_D(t) - I_D(t-x)$$

$$R(t) = R(t-1) + I_D(t-x)$$

mit x : Dauer der Ansteckungsfähigkeit

$$S(t) = N - I_M(t) - R(t)$$

$$\text{mit } N = \sum_{t=0}^{T} I_M(t)$$



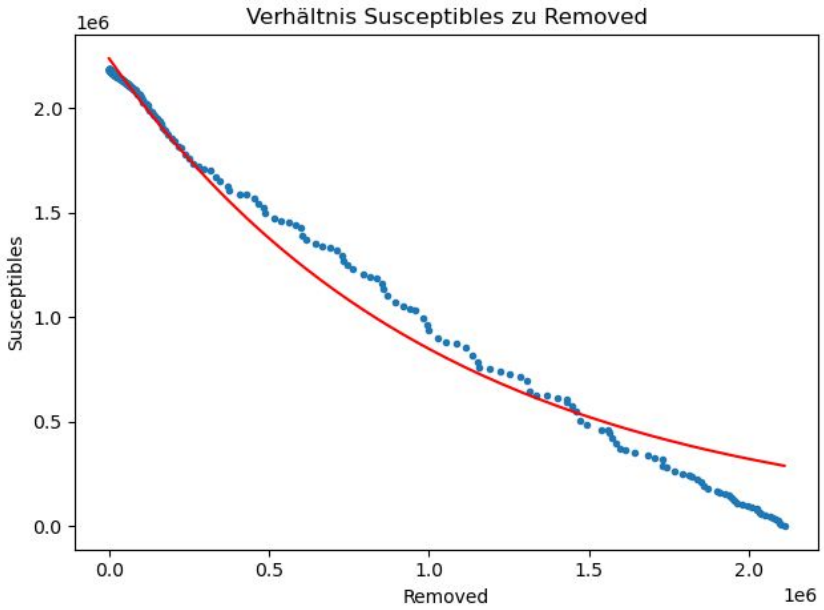
Berechnung von β, γ und S(0)

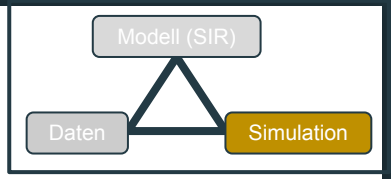
Ansatzfunktion: $S(R) = S_0 \, e^{-\frac{\beta}{\gamma} * R}$

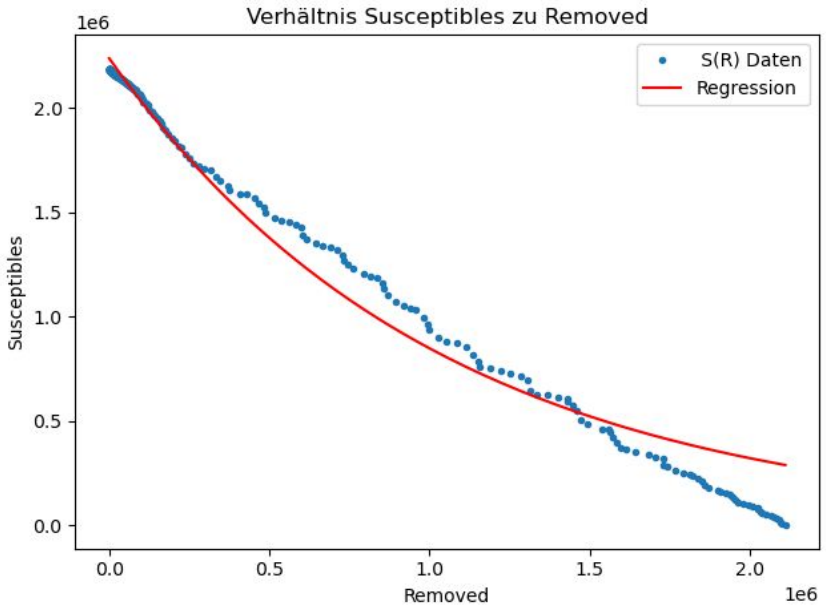
Regression liefert: $c^0 = [S(0), \frac{\beta}{\gamma}]$

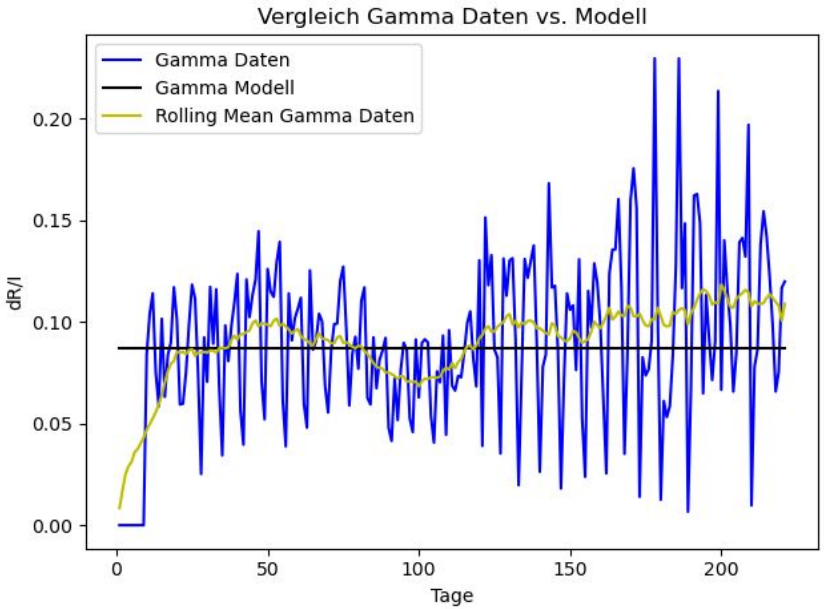
Gamma berechnen: $\gamma = \frac{< dR, I >}{\| I \|_2^2}$

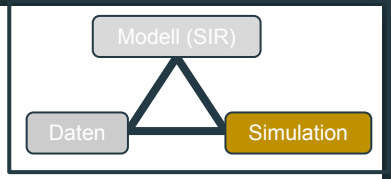
Beta berechnen: $\beta = \frac{\beta}{\gamma} * \gamma$




Modell (SIR)
Daten
Simulation


Verhältnis Susceptibles zu Removed
1e6
S(R) Daten
Regression
Susceptibles
Removed
1e6
0.0
0.5
1.0
1.5
2.0


Vergleich Gamma Daten vs. Modell
Gamma Daten
Gamma Modell
Rolling Mean Gamma Daten
dR/I
Tage
0.00
0.05
0.10
0.15
0.20
0
50
100
150
200

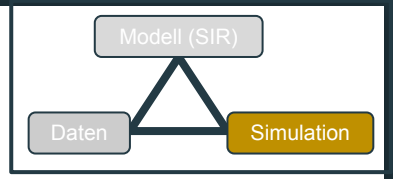


Was gilt es zu lösen?

(1) $$\frac{dS}{dt} = -\beta SI$$

(2) $$\frac{dI}{dt} = \beta SI - \gamma I$$

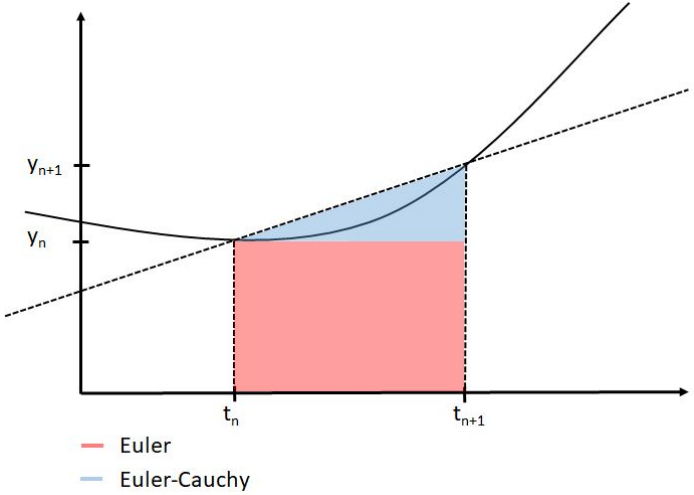
(3) $$\frac{dR}{dt} = \beta I$$



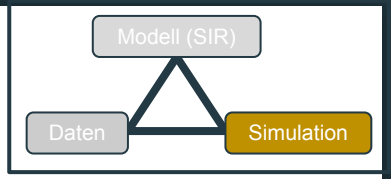
Numerische Berechnung

$$y_{n+1} = y_n + Q$$

$$y_{n+1} = y_n + \sum_{j=1}^{4} b_j k_j$$ RK-klassisch



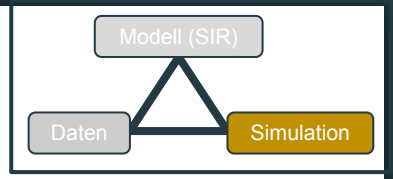
$$Q = \begin{cases} (t_{n+1} - t_n) * y_n & \text{Euler} \\ (t_{n+1} - t_n) * \frac{1}{2} * (y_{n+1} + y_n) & \text{Euler-Cauchy} \end{cases}$$



Güte der Näherung

- Schätzung über den EOC
- Getestet an bekannten Funktionen Sinus und Cosinus

	Euler	Euler-Cochy	Heun	RK Ordnung 4	RK Ordnung 5	RK Ordnung 6	RK Ordnung 7	RK Ordnung 8	RK Ordnung 9	RK Ordnung 10	RK Ordnung 11	RK Ordnung 12
Anzahl Stufen	1	2	3	4	5	6	7	8	9	10	11	12
Konvergenz-ordnung	1	2	3	4	4	5	6	6	7	7	8	9



Fehlerabschätzung

EOC: $$\|y - y_{ex}\|_\infty = max_{k\in\{0,N\}} |y_k - y_{ex}(t_k)| \leq Ch^p$$

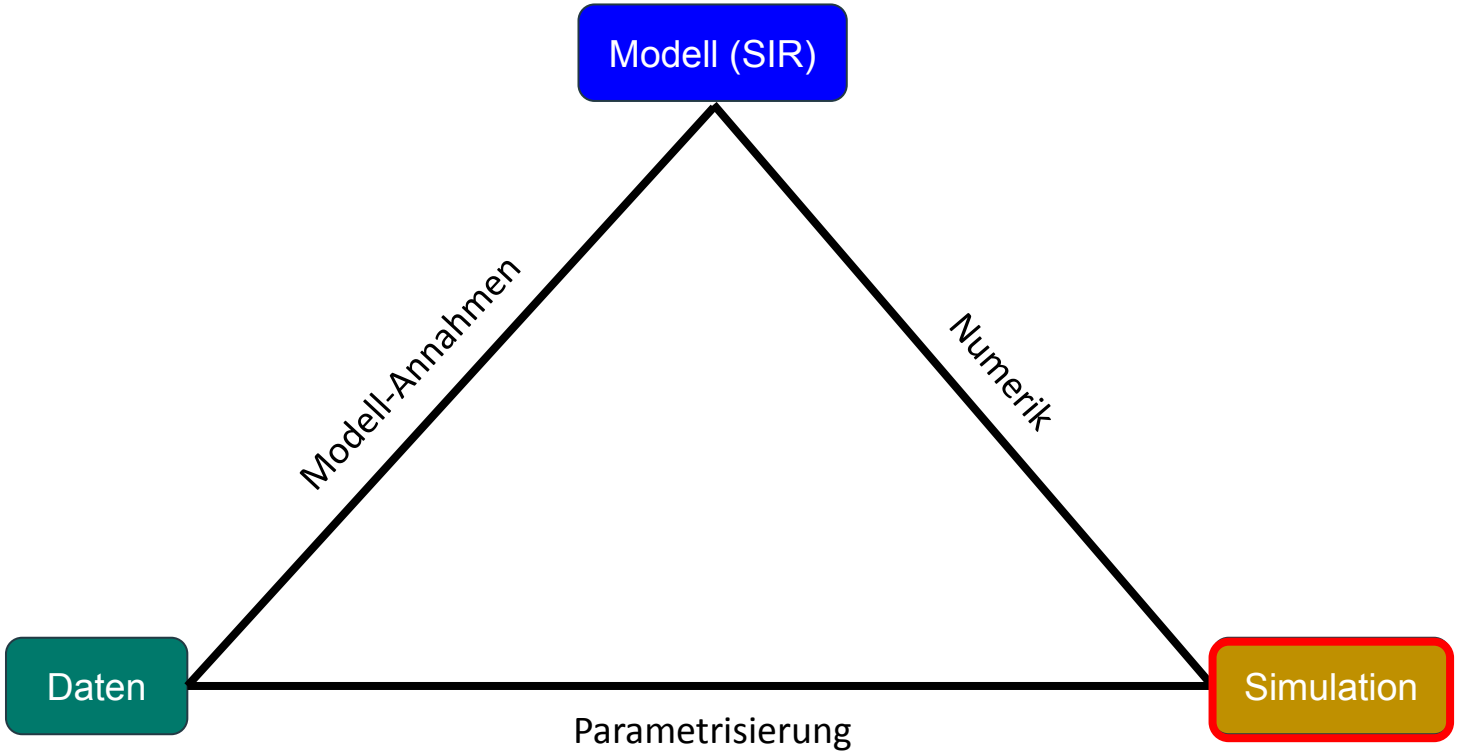
Mit Schrittweite: $$h = |t_{k+1} - t_k|$$

Euler

delta	fehlerS	EOC_S	fehlerI	EOC_I	fehlerR	EOC_R
0.4	0.030030250994340646	1.0082500248926392	0.01570346366097719	1.0096294835311719	0.04453709367493275	1.0099257059953246
0.2	0.014929506817621827	1.0041293210503757	0.007799498786199233	1.0047586039298422	0.022115865612903463	1.0049031307559197
0.1	0.0074434181360017865	1.0020658043663617	0.003886907599815914	1.0023400791996169	0.011020415202281697	1.002476445435156

Runge-Kutta

delta	fehlerS	EOC_S	fehlerI	EOC_I	fehlerR	EOC_R
0.4	7.701433824003345e-09	3.9995225037840596	5.523096336945699e-07	4.0435101477040565	5.483931744265647e-07	4.0428059996878405
0.2	4.814989518209245e-10	4.002715992826513	3.3493827267250253e-08	4.021913379270859	3.327255626572878e-08	4.021556530301904
0.1	3.003708393123361e-11	4.011510569782268	2.0618079465961614e-09	4.011049455375446	2.0486936591623817e-09	4.011615091134583


Modell (SIR)
Modell-Annahmen
Numerik
Daten
Simulation
Parametrisierung

Fazit & Ausblick

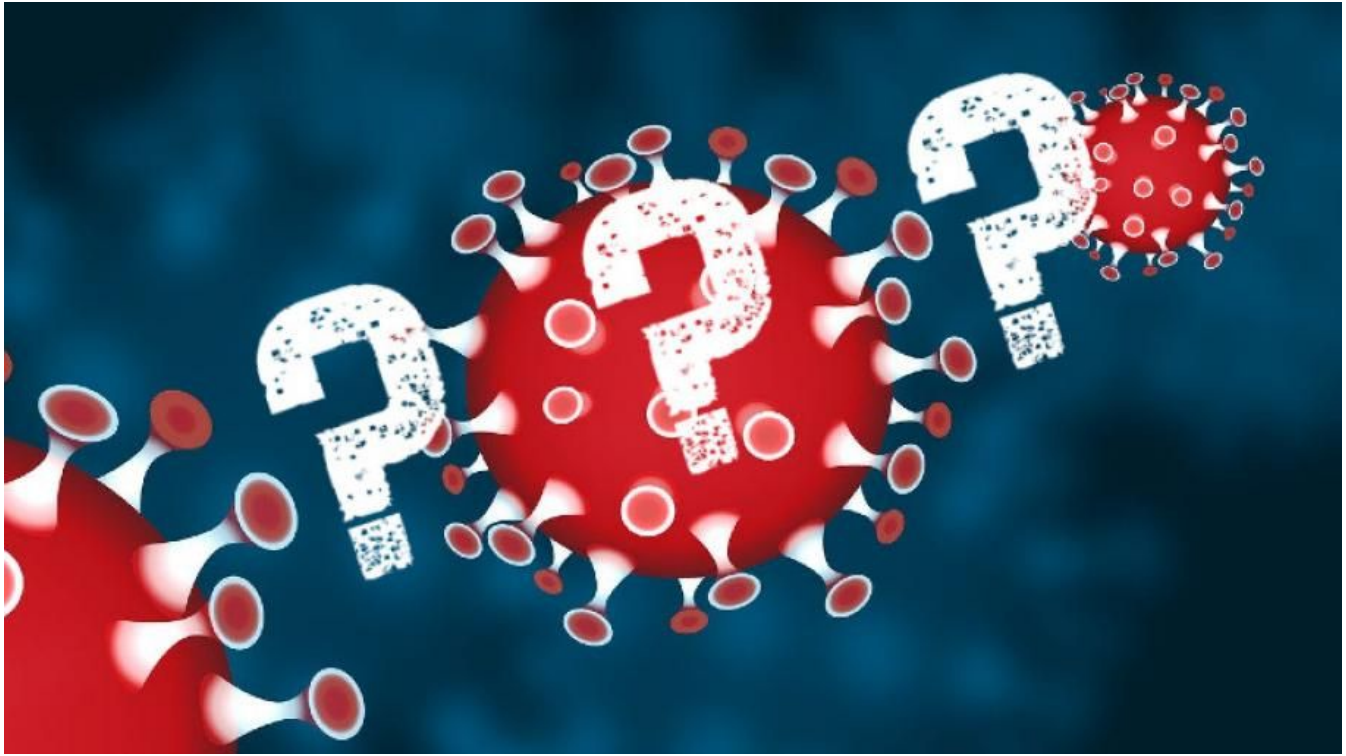
- Pandemien verhindern
  - Erholungsrate gamma hochschrauben → Medikamente, Impfung, Quarantäne
  - Infektionsrate beta reduzieren → Verringere N (Shut-Down), Hygiene-Maßnahmen

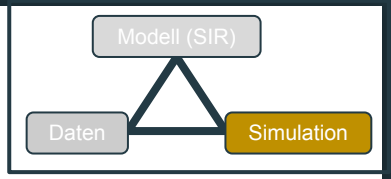
- Neue wissenschaftliche Erkenntnisse, die anfangs nicht bekannt waren
  - Impfungen
  - andere ansteckendere Virusvarianten (z.B. B.1.1.7)

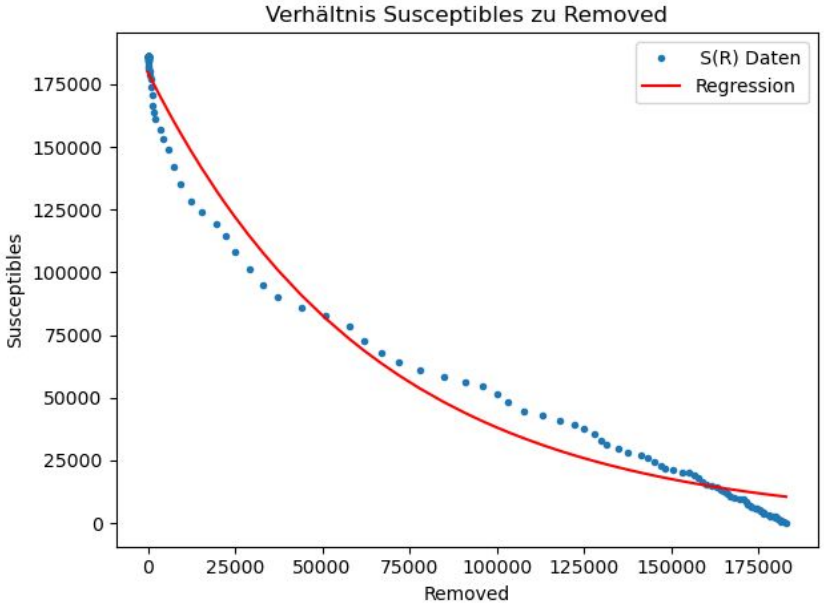
  → variable beta und gamma im Modell, Mix aus den SIR-Erweiterungen generieren. Unser Vorschlag: SIR-X

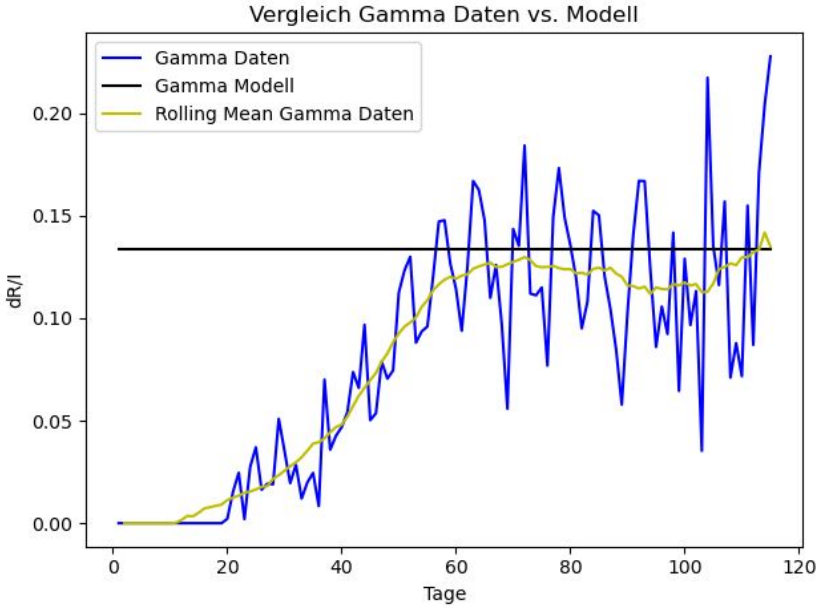
- Kombination mit p-flow

Danke für die tolle Zusammenarbeit und ein insgesamt spannendes Teamprojekt,
Frau Axthelm!

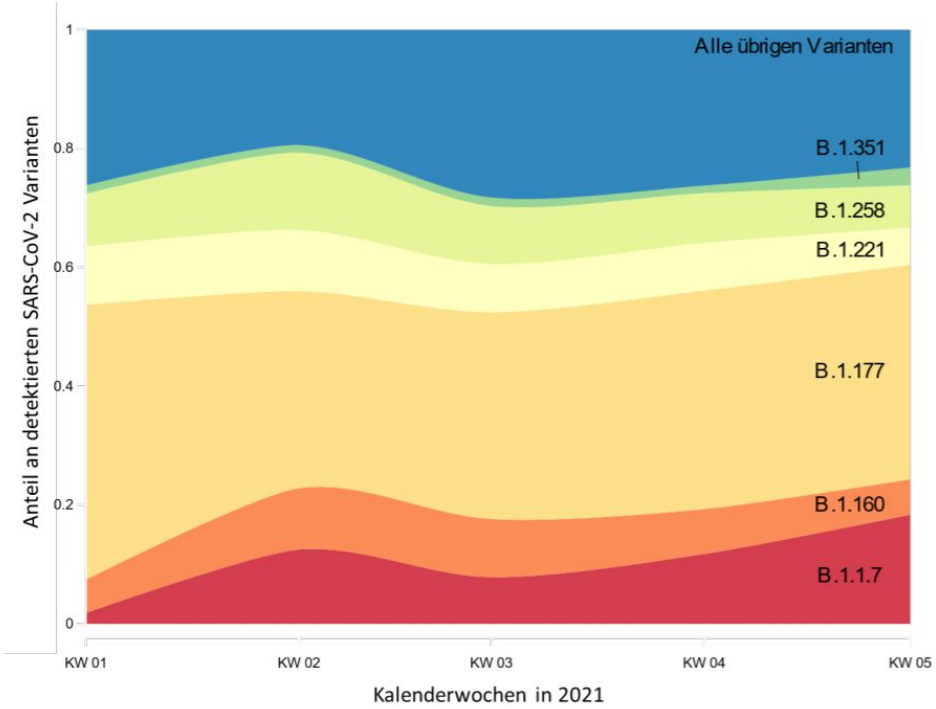



Modell (SIR)
Daten
Simulation


Verhältnis Susceptibles zu Removed
S(R) Daten
Regression
Susceptibles
Removed
0
25000
50000
75000
100000
125000
150000
175000


Vergleich Gamma Daten vs. Modell
Gamma Daten
Gamma Modell
Rolling Mean Gamma Daten
dR/I
Tage
0.00
0.05
0.10
0.15
0.20
0
20
40
60
80
100
120

Variant of Concern B.1.1.7



Pandemiemodellierung

Schon immer musste sich die Menschheit gegen Pandemien behaupten, wie zum Beispiel 1918 - 1920 die Spanische Grippe, die mehr als 27 Mio. Tote forderte oder die Londoner Pest (1665 - 1666), die ein Fünftel der Londoner Bevölkerung ausradierte. Aktuell werden die Menschen durch den Corona-Virus (SARS-CoV-2) in Atem und ihren Häusern gehalten. Damit die Pandemie möglichst schnell eingedämt werden kann, muss die Politik geeignete Maßnahmen, wie Kontaktbeschränkungen, Maskenpflicht, etc. bestimmen. Die Grundlage dieser Entscheidungen ist dabei ein mathematisches Modell, dass die Pandemie möglichst genau abbildet, um Vorhersagen über die Anzahl der Infizierungen zu treffen. Das geläufigste Modell ist das **SIR-Modell**. Das Modell beschreibt die Verläufe der drei Hauptgruppen:

- **Susceptible**: Personen, die gesund sind und sich potentiell anstecken können.
- **Infected**: Personen, die infiziert sind
- **Removed**: Personen, die genesen oder gestorben sind

Diese Webapplikation nimmt dieses Modell als Rechenmodell, um den Epidemieverlauf beispielhaft für Deutschland zu modellieren. Die Daten anderer Länder sind weniger genau dokumentiert, weshalb wir uns auf Deutschland für die Prognose beschränken. Die aktuellen Daten anderer Länder können dennoch im Daten-Tab visualisiert werden.

Modelle

Um den Krankheitsverlauf in einer Pandemie zu modellieren, benötigt man ein mathematisches Modell als Grundlage. Weit verbreitete Modelle sind das SI-Modell und dessen Abwandlungen. Um verschiedene Gegebenheiten, wie beispielsweise Maßnahmen, Inkubationszeiten oder Infektionsdauer, mit in das Modell einfließen zu lassen, werden die Differenzialgleichungen mit einfachen Termen erweitert bzw. angepasst. Die nachfolgende Übersicht zeigt ein paar geläufige Abwandlungen und Erweiterungen.

Um einen realen Krankheitsverlauf darzustellen, müssen bestimmte Annahmen getroffen werden um die Rahmenbedingungen festzulegen. Je nach Modell können sich diese Annahmen unterscheiden. Generell werden in den Modellen keine natürlichen Geburten und Tode berücksichtigt. Durch zusätzliche Terme können diese in den verschiedenen Modellen miteinbezogen werden. Für die folgenden Modellbeschreibungen werden modellübergreifende Annahmen getroffen:

- **konstante Populationsgröße**: Die Gesamtbevölkerung ändert sich nicht (keine natürlichen Geburten/Tode)
- **homogene Bevölkerung**: Die Gruppen (S, I, R) sind gleichförmig gemischt, das heißt jede Person hat die selben Kontakt-/Ansteckungschancen (keine Unterscheidung nach Herkunft, Religion, Sozialstand, Geschlecht, Alter, etc.)
- **keine Kreuzimmunität**: Vorerkrankungen mit einem Ähnlichen Virus haben keinen Einfluss auf die Immunität
- **Saisonalität**: Die Umweltfaktoren sind immer gleich (keine Jahreszeiten, gleiche Temperatur, etc.)
- **Anfangsbedingung**: Es gibt zu Beginn mindestens eine infizierte Person

SI SIS SIR SEIR SIRD SIR-X

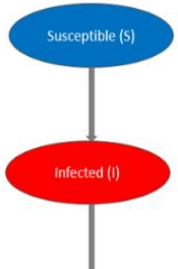
SIR - Susceptible-Infected-Recovered

Zusätzliche Annahmen:

- **Immunität**: Personen, die genesen oder sterben, sind immun und können sich nicht erneut infizieren bzw. andere Personen infizieren

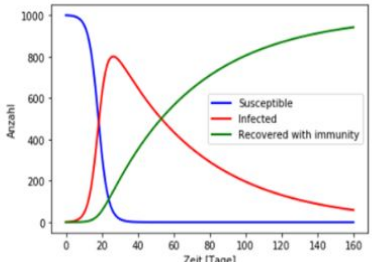
Das SIR-Modell ist eine Abwandlung des SIS-Modells bzw. eine Erweiterung des SI-Modells. Eine Erholungsrate gibt an, wie schnell Personen aus der Gruppe der Infizierten genesen und in die Gruppe R wechseln. Die blaue Kurve stellt die Anzahl der infizierbaren Personen dar. Zu Beginn nimmt die Anzahl der Infizierbaren (blau) ab, während die Anzahl der Infizierten (rot) zunimmt. Leicht verzögert nimmt die Anzahl der immunen Personen zu (grün). Nachteile des SIR-Modells sind, ...

- ... dass am Ende der Epidemie alle Menschen infiziert waren und genesen sind, was nicht der Realität entspricht. Es ist nicht realistisch, dass sich wirklich jede Person ansteckt.
- ... dass nicht alle Infektionskrankheiten den typischen Verlauf S-I-R folgen (z.B. keine Genesung bei HIV).
- ... dass je nach Krankheitsübertragung (z.B. Tiere, Wasser, ...) andere Basismodelle genutzt werden



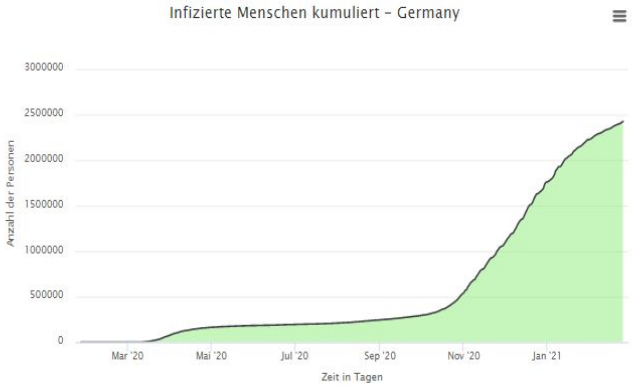
$$\frac{dS}{dt} = -\beta S I$$

$$\frac{dI}{dt} = \beta S I - \gamma I$$



COVID-19-Statistik

Die beiden Diagramme stützt sich auf die Daten von Our World of Data. Die beiden Diagramme zeigen jeweils die täglichen neu Infizierten, einmal kumuliert und einmal nicht. Der COVID-19-Datensatz ist eine Sammlung der COVID-19-Daten, die von Our World in Data gepflegt werden. Er wird täglich aktualisiert und enthält Daten zu bestätigten Fällen, Todesfällen, Krankenhausaufenthalten, Tests und Impfungen sowie andere Variablen von potenziellem Interesse. Die Daten können hier im CSV Format heruntergeladen werden. Die ersten Daten stammen von dem 27.01.2020 und werden auf einer täglichen Basis aktualisiert. Die CSV Datei muss nach dem Download für den weiteren gebraucht aufgearbeitet werden. Dazu werden nicht benötigte Spalten, wie zum Beispiel Kontinent und ISO-Code des Landes entfernt. Außerdem muss auf negativ Werten sowie NaN (leere Werte) reagiert werden. In unserem Fall werden NaN Werte zu Nuller und die negativen Werte mit Abs zu positive Werte konvertiert werden.



Dieses Diagramm stellt die täglichen neu Infektionen bzw. täglichen Todesfälle eines Landes als kumuliertes Liniendiagramm dar. Es besteht die Möglichkeit die Kurve mit 4 bzw. 7 Tagen zu glätten.

Einstellungen

Neue Fälle

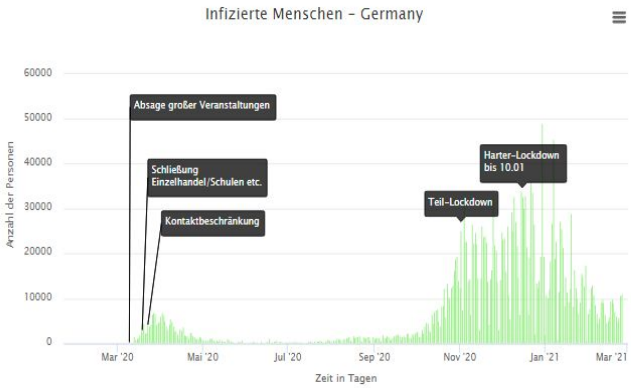
Wähle ein Land aus: Germany

Durchschnitt:

- (●) Keinen
- ( ) 4 Tage
- ( ) 7 Tage

Auffälligkeiten im Diagramm:

- Nachträglichen Korrekturen
- Weniger tests an den Wochenenden
- Verzögerung bei der Meldung von Fallzahlen, aufgrund von Wochenende bzw. Feiertagen
- Gebündelten Meldung von Daten über mehrere Tage



Dieses Diagramm stellt die täglichen neu Infektionen bzw. täglichen Todesfälle eines Landes in einem Liniendiagramm dar. Es besteht die Möglichkeit die Kurve mit 4 bzw. 7 Tagen zu glätten.

Startwerte:

Verfahren: Runge-Kutta

Bevölkerungsgröße [Personen]: 500000
10 — 1.000.000

davon Infizierte [Personen]: 1
1 — 500.000

Dauer [Tage]: 20
1 — 100

Übertragungsrate (Beta): 0,3
0 — 1

Erholungsrate (Gamma): 0,3
0 — 1

Maßnahme:

Zeitpunkt [Tag]: 5
1 — 100

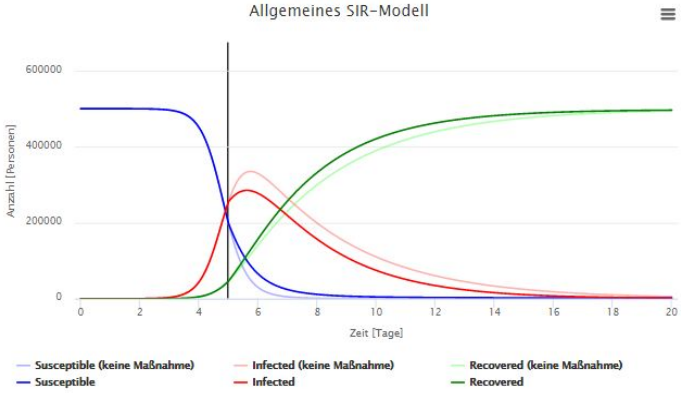
Übertragungsrate (Beta): 0,2
0 — 1

Erholungsrate (Gamma): 0,4
0 — 1

Diagramm aktualisieren

SIR-Playground

Hier kann mit dem allgemeine SIR-Modell experimentiert werden. Neben der Bevölkerungsgröße und den davon zu Beginn infizierten Personen, kann die Dauer der Beobachtung eingestellt werden. Die Übertragungsrate bestimmt, wie schnell sich Personen infizieren und andere Personen anstecken können. Die Erholungsrate gibt an, wie schnell infizierte genesen oder sterben und nicht mehr ansteckend sind bzw. angesteckt werden können. Es kann ebenfalls bestimmt werden, mit welchem Verfahren der Verlauf der Pandemie berechnet werden soll. Um zu simulieren, was eine Maßnahme bewirken kann, lässt sich ein Zeitpunkt bestimmen, ab dem die Maßnahme wirkt (Hinweis: Die Maßnahme wirkt direkt). Die Parameter können dann so angepasst werden, dass ab dem Zeitpunkt der Maßnahme die neuen Parameter genutzt werden. Die Maßnahme wird über die Parameter bestimmt, zum Beispiel wird bei einer Kontaktbeschränkung die Übertragungsrate verringert, da sich die Personen weniger sehen und langsamer anstecken.

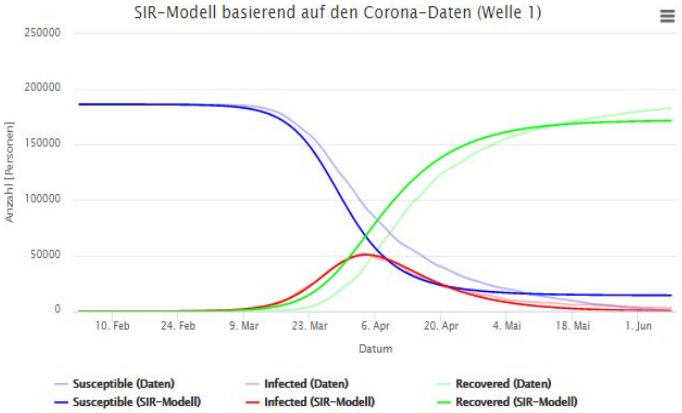


Diese Grafik zeigt das SIR-Modell das der Pandemie zugrunde liegt.

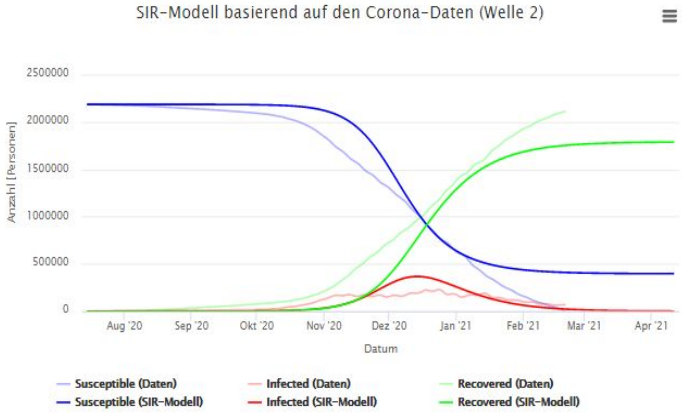
Prognose

Das SIR-Modell wird verwendet, um die Daten einer Pandemie zu beschreiben und daraus Vorhersagen für die zukünftige Entwicklung zu treffen. Die Parameter beta und gamma (Übertragungs- und Erholungsrate) für das Modell werden aus den Daten mit Hilfe der Regression berechnet. Mit Hilfe dieser Parameter kann das SIR-Modell die Daten approximieren. Das SIR-Modell kann lediglich einzelne Infektionswellen modellieren, weshalb der gesamte Zeitraum in zwei Wellen aufgeteilt wurde.
In der nebenstehenden Tabellen sind die Werte und Parameter der beiden Wellen dargestellt. Da die Daten der ersten Welle vollständig vorliegen, wird dort keine Prognose durchgeführt. Das SIR-Modell setzt voraus, dass zum Schluss der Pandemie fast alle Personen einmal infiziert waren, weshalb die Gesamtbevölkerung Deutschlands (ca. 83 Mio.) nicht verwendet werden kann. Stattdessen wird die Summe über alle Infizierten über den jeweiligen betrachteten Zeitraum als Gesamtbevölkerung N angenommen.

	Welle 1	Welle 2
Zeitraum	03.02.2020 - 08.06.2020	15.07.2020 - 20.02.2021
Prognose	0 Tage	50 Tage (bis 02.02.2021)
N	186.099	2.187.252
beta	$0.1643 * 10^{-5}$	$0.8392 * 10^{-7}$
gamma	0.1098	$0.8796 * 10^{-1}$



Das Diagramm zeigt die Kurven der Susceptibles, Infected und Removed aus den Daten. Die Werte wurden über 4 Tage geglättet. Ebenfalls werden die Kurven des SIR-Modells angezeigt mit den aus den Daten berechneten Parametern beta und gamma.



Das Diagramm zeigt die Kurven der Susceptibles, Infected und Removed aus den Daten. Die Werte wurden über 4 Tage geglättet. Ebenfalls werden die Kurven des SIR-Modells angezeigt mit den aus den Daten berechneten Parametern beta und gamma.

Ergebnisse

Wie man sieht, kann die erste Welle gut dargestellt werden, da die tatsächlichen Infektionszahlen dem SIR-Modell sehr ähnlich sind. Dadurch können die aus den Daten berechneten Parametern beta und gamma den Verlauf treffend beschreiben. Die zweite Welle wird etwas schlechter angenähert, weil der typische Verlauf durch Quarantäne-Maßnahmen verändert wird und dadurch der Verlauf der Daten schlechter dargestellt wird. Ebenfalls wird für die Berechnung der Parameter nur der exponentielle Anstieg der zweiten Welle berücksichtigt, weshalb auch dadurch ein kleiner Fehler miteinfließt.
Wie in der Tabelle zu sehen, sind die betas für die beiden Wellen ungefähr um die Größenordnung 10 verschieden. Die Ursache hierfür ist die Abhängigkeit von beta, der Ansteckungsrate, zur Gesamtbevölkerung. Das heißt je größer die Gesamtbevölkerung, desto kleiner das beta für einen ähnlichen Verlauf. Das gamma, die Erholungsrate, verändert sich nur leicht, da diese nur von der Anzahl der Infizierten und nicht direkt von der Gesamtbevölkerung abhängig ist.

Ergebnisse

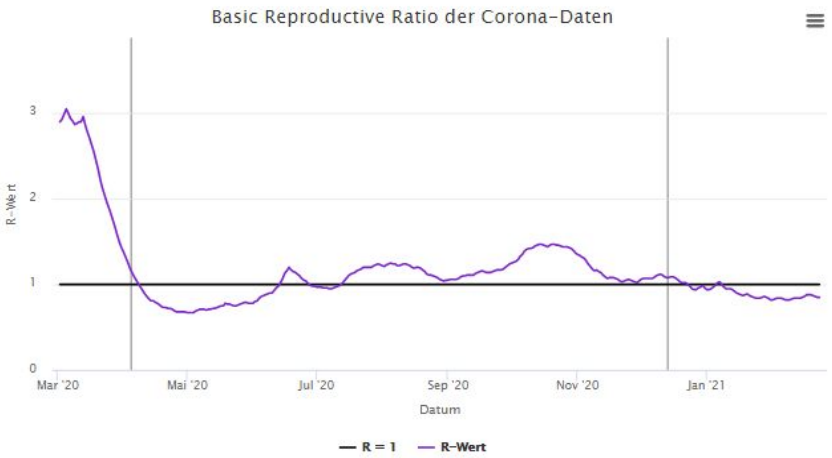
Wie man sieht, kann die erste Welle gut dargestellt werden, da die tatsächlichen Infektionszahlen dem SIR-Modell sehr ähnlich sind. Dadurch können die aus den Daten berechneten Parametern beta und gamma den Verlauf treffend beschreiben. Die zweite Welle wird etwas schlechter angenähert, weil der typische Verlauf durch Quarantäne-Maßnahmen verändert wird und dadurch der Verlauf der Daten schlechter dargestellt wird. Ebenfalls wird für die Berechnung der Parameter nur der exponentielle Anstieg der zweiten Welle berücksichtigt, weshalb auch dadurch ein kleiner Fehler miteinfließt.
Wie in der Tabelle zu sehen, sind die betas für die beiden Wellen ungefähr um die Größenordnung 10 verschieden. Die Ursache hierfür ist die Abhängigkeit von beta, der Ansteckungsrate, zur Gesamtbevölkerung. Das heißt je größer die Gesamtbevölkerung, desto kleiner das beta für einen ähnlichen Verlauf. Das gamma, die Erholungsrate, verändert sich nur leicht, da diese nur von der Anzahl der Infizierten und nicht direkt von der Gesamtbevölkerung abhängig ist.

Probleme/Nachteile des SIR

Für das SIR-Modell müssen viele vereinfachende Annahmen getroffen werden, beispielsweise eine homogenen Bevölkerung, die bedeutet, dass jede Person die gleichen Kontakt- und Ansteckungschancen hat unabhängig von Sozialstatus, Kultur oder Herkunft. Durch diese vereinfachenden Annahmen weicht das Modell von den Daten ab.
In beiden Diagrammen ist zu sehen, dass das SIR-Modell nach rechts verschoben ist. Eine mögliche Ursache hierfür ist, dass in unserem Modell ein konstantes beta über den gesamten Zeitraum der Welle berechnet wird, was sich in der Realität durch Maßnahmen verringern würde. Darum ist der Anstieg zu Beginn des Modells zu gering, wodurch die Verschiebung verursacht wird.
Der Unterschied der ersten und zweiten Welle besteht darin, dass in der zweiten Welle nur das exponentielle Wachstum für die Berechnung der Parameter verwendet wird, weshalb dort der Fehler zwischen den Kurven größer ist, als in der ersten Welle.
In der zweiten Welle sieht man, dass die Kurve der Infizierten ab November nicht´mehr weiter steigt. Das könnte daran liegen, dass ab dem 28. Oktober ein Lockdown light umgesetzt wurde, weshalb sich die Kontakte und damit der Anstieg der Infizierungen verringerten. Für diesen Fall sollte das SIR-X-Modell verwendet werden.
Eine der Hauptannahmen des SIR-Modells ist, dass fast alle Personen am Schluss der Pandemie infiziert waren. Dies ist in der Realität nicht so, da unter anderem durch geografische Distanz oder Herdenimmunität Teile der Bevölkerung nicht infiziert werden. Um das SIR-Modell dennoch abbilden zu können, wird angenommen, dass sie die Susceptibles zum Startzeitpunkt aus den kummulierten Infizierten über den Zeitraum der Welle berechnet. Dadurch entsteht ein kleiner Fehler in der Prognose, da eine zu kleine Bevölkerung angenommen wird.

Basic Reproduction Ratio (R-Wert)

Der R-Wert ist die Reproduktionszahl, die angibt, wie viele Menschen eine Person im Mittel ansteckt. Ist dieser Wert unter 1 werden immer weniger Personen angesteckt und die Pandemie endet irgendwann. Ist der Wert 1 steckt jede Person eine weitere an und die Pandemie bleibt bestehen. Sollte der Wert über 1 sein, erhöht sich die Geschwindigkeit der Ausbreitung, da eine Person mehrere Personen ansteckt. Durch Maßnahmen oder Beschränkungen kann der R-Wert verringert werden, da es weniger Kontakte zwischen den Personen gibt.
Im Diagramm wird der R-Wert der Daten dargestellt. Die vertikalen Linien geben den ungefähren Zeitpunkt des Maximums der Infizierten der beiden Wellen an. Bei der ersten Welle sieht man, dass der R-Wert bis zur vertikalen Linie über 1 ist und dadurch die Infektionen ansteigen. Ab dem Maximum sinkt die Reproduktionszahl unter 1, weshalb die Infektionen zurück gehen. Bei der zweiten Welle kann man dasselbe beobachten, mit dem Unterschied, dass es kein eindeutiges lokales Maximum gibt. Der R-Wert schwankt leicht um 1, weshalb sich die Kurve der Infizierten kaum ändert. Der R-Wert basiert auf einem Verhältnis der Infektionen innerhalb von 7 Tagen. Darum kann der Wert bei kleinen Zahlen sehr hoch werden, was zu Beginn der Berechnungen der Fall ist. Dennoch ist der R-Wert ein guter Wert, um die aktuelle Tendenz des Verlaufs der Pandemie abzuschätzen.



Das Diagramm zeigt den Basic Reproductive Ratio berechnet nach der Formel des RKI über 4 oder 7 Tage geglättet.