

Modellierung der Covid-19 Fälle

Zusammenfassung

Ziel der Modellierung ist die Prognose der Covid-19 Infektionen für die kommenden Wochen sowie die Berechnung der dadurch resultierenden Auslastung in den Krankenhäusern. Ausschlaggebend für die Entwicklung der Fälle ist die Reproduktionsrate, welche angibt, wie viele weitere Personen eine Person im Schnitt ansteckt. Diese wird zunächst aus den Daten geschätzt und anschließend mit Hilfe eines Zeitreihenmodells für jeden Landkreis vorhergesagt. Unsicherheiten werden durch ein Prognoseintervall berücksichtigt. Außerdem werden verschiedenen Szenarien bezüglich der politischen Maßnahmen betrachtet. Um räumliche Abhängigkeiten zu berücksichtigen und die Instabilität bei kleinen Landkreisen zu beseitigen, folgt eine Glättung der Raten mit Hilfe eines generalisierten additiven Regressionsmodells. Die sich ergebenden Raten werden verwendet, um die Infektionszahlen für die kommenden Wochen zu simulieren. Dazu wird das epidemiologische SEIR-Modell herangezogen.

1 Schätzung der Reproduktionsrate

Im ersten Schritt wird die Nettoreproduktionsrate aus den gemeldeten Infektionszahlen des RKI geschätzt. Die Rate gibt an, wie viele Personen eine Person im Schnitt ansteckt. Zunächst muss dafür bestimmt werden, wie viele Personen sich wann infizieren. Dazu werden folgende Annahmen getroffen:

- Es wird eine Dunkelziffer angenommen. Diese wird mit Hilfe von PCR-Testzahlen geschätzt. Es wird dazu ein logarithmischer Zusammenhang zwischen Anteil getesteter Bevölkerung und Anteil entdeckter Covid Fälle angenommen. Eine Prognose der PCR-Testzahlen erfolgt über Auto-Arima.
- Der Meldeverzug beträgt 7 Tage. Es vergehen also 7 Tage von der Infektion bis zur Meldung durch das RKI.

Nachdem die tatsächlichen Infektionszahlen dadurch berechnet wurden, kann die Nettoreproduktionsrate als wöchentliche Zunahmerate geschätzt werden. D.h. für einen Tag werden die tatsächlichen Neuinfektionen durch die tatsächlichen Infektionszahlen desselben Wochentags der Vorwoche geteilt. Dadurch wird auch die wöchentliche Saisonalität berücksichtigt. Wenn max.date der letzte Tag ist, an dem die gemeldeten Infektionszahlen verfügbar sind, können die tatsächlichen Fälle und die Nettoreproduktionsrate dadurch bis zum Tag $\text{max.date.rate} = \text{max.date} - 7$ (Meldeverzug) geschätzt werden. Ab dem Tag danach sind Forecasts notwendig.

Anschließend wird die Nettoreproduktionsrate um den Anteil der aktuellen

”Susceptibles” (also die Personen, die sich aktuell infizieren können) bereinigt. Die Anzahl der aktuellen Susceptibles wird mit S_t bezeichnet. Dadurch ergibt sich die Basisreproduktionsrate. Diese hängt im Gegensatz zur Nettoreproduktionsrate nicht mehr davon ab, wie hoch der Anteil der Susceptibles aktuell ist und lässt sich somit auch über verschiedene Zeitpunkte, Landkreise etc. vergleichen. Sie lässt sich als die Anzahl der Personen, die eine Person im Mittel anstecken würde, wenn die gesamte Population ”Susceptible” wäre, interpretieren.

Die Schätzung der Basisreproduktionsrate erfolgt separat für jeden Landkreis, aber auch für Gesamtdeutschland. Da die Rate bei kleinen Infektionszahlen instabil ist, wird sie mit dem Wert 5 gedeckelt. Um die Reproduktionsrate um den Anteil der aktuellen ”Susceptibles” zu bereinigen, müssen Immunitäten durch überstandene Infektionen und Impfungen miteinbezogen werden. Dazu sind einige Annahmen nötig, welche später erläutert werden.

2 Forecast der Reproduktionsrate

Als Nächstes erfolgt der Forecast der Basisreproduktionsrate ab dem Tag `max.date.rate + 1`. Dafür wird ein Zeitreihenmodell (ARMA) herangezogen, da Abhängigkeiten der Rate bei aufeinanderfolgenden Zeitpunkten angenommen werden können.

In diesem wird auch ein ”Shutdown-Effekt”, welcher die politischen Maßnahmen zur Eindämmung des Corona-Virus repräsentiert, berücksichtigt. Die entsprechende Variable `ShD` bewegt sich je nach Maßnahmen zwischen 0 und 1. Unterschiede zwischen den Bundesländern sind möglich. Zuletzt wird auch saisonale Effekte berücksichtigt, indem die Langjahresdurchschnittstemperatur pro Monat als Regressor verwendet wird. Es ergibt sich dadurch ein ARMAX Modell mit 2 externen Regressoren.

Die Modellierung gliedert sich in zwei Schritte:

1. Modelltraining: Das ARMAX Modell wird auf Deutschland-Ebene trainiert. Für die Rate wird eine Lognormalverteilung angenommen. Dadurch beschränkt sich der Wertebereich auf positive Werte.

$$\text{Log}(R_t^C) = c + \sum_i^p a_i \text{Log}(R_{t-i}^C) + \sum_j^q b_j \epsilon_{t-j} + \beta_1 * \text{ShD}_t + \beta_2 * \text{AvgTemp}_t + \epsilon_t$$

p und q werden mit Hilfe eines Auto-ARMA datengestützt automatisch bestimmt.

2. Anwendung des Modells für jeden Landkreis: Das Modell wird nur auf Deutschland-Ebene trainiert, aber auf jedem Landkreis separat ausgewertet. Unsicherheiten werden durch ein gleitendes Prognoseintervall (Niveau zwischen 95% und 70%) berücksichtigt.

Zudem werden neben dem Basis-Szenario ”Fortsetzung der bestehenden Maßnahmen” je nach aktueller Infektionslage und politischer Situation weitere Szenarien modelliert. Diese können folgende enthalten:

- Lockerungen

- Vollständige Aufhebung aller Maßnahmen
- Verschärfung der Maßnahmen
- Lockdown
- Maßnahmen wie in einem bestimmten Bundesland werden in Gesamtdeutschland ergriffen
- Inzidenzabhängige Maßnahmen (z.B. Bundesnotbremse)

3 Glättung der Reproduktionsrate

Es liegen nun Basisreproduktionsraten über die Zeit für jeden Landkreis vor. Die Prognosen werden mit Hilfe eines generalisierten linearen additiven Regressionsmodells (GAM) geglättet. Dies hat folgende Zwecke:

- Räumliche Glättung: Im Modell werden Latitude und Longitude der Landkreise (Zentroide) als Regressoren verwendet. Dies führt zu einem glatten räumlichen Verlauf. Dies ist in der Hinsicht plausibel, dass Reproduktionsraten benachbarter Landkreise sich tendenziell ähnlicher verhalten, als Raten entfernter Landkreise (oder sich mit der Zeit zumindest annähern werden)
- Bereinigung von Landkreisen mit geringen Infektionszahlen: Die Schätzungen für Landkreise mit wenigen Daten und/oder kleinen Infektionszahlen sind sehr instabil. Auch hier hat die Glättung einen dementsprechend positiven Effekt.
- Zusätzliche zeitliche Glättung

Zusätzlich zu den Variablen Zeit und Ort (Lat, Lon) werden die Bevölkerungsdichte $BevD$ und das Bundesland BL als Regressoren berücksichtigt. Die Modellformel lautet:

$$\text{Log}(R_t^C) = f_1(\text{Lat}, \text{Lon}) + f_{2,BL}(t) + \text{beta}_{BL} + \text{beta}_{BevD} * \text{BevD} + \epsilon_t$$

Das Modell wird im Gegensatz zu Schritt 2 über alle Landkreise hinweg trainiert und ausgewertet. Es wird lediglich auf die Prognosen angewendet - dabei wird darauf geachtet, dass die ein fließender Übergang zwischen historischen Werte und Prognosen bestehen bleibt. Es wird dazu nur das Basis-Szenario herangezogen. Die geglätteten Raten der weiteren Szenarien und die Raten der Prognoseintervallgrenzen erhält man proportional zur nicht geglätteten Variante.

4 Prognose der Impffzahlen

Die Prognose der Erstimpfungen erfolgt anhand der Dynamik der letzten 3 Wochen in der Historie. Es wird also eine wöchentliche Zunahmerate der Impfungen geschätzt. Die wöchentliche Saisonalität wird dabei berücksichtigt. Die

Prognosen der Zweitimpfungen ergeben sich aus den Erstimpfungen, die 6 Wochen zurückliegen, multipliziert mit 0.99 um Non-Compliance zu berücksichtigen.

5 Modellierung der Infektionszahlen

Um aus den Reproduktionsraten die Infektionszahlen zu bestimmen, wird das epidemiologische SEIR-Modell verwendet. Dieses umfasst vier Stages:

- S - Susceptible: Personen, die sich infizieren können
- E - Exposed: Infizierte, aber noch nicht infektiöse Fälle
- I - Infectious: Infektiöse Fälle
- R - Removed: Personen, die sich aktuell nicht infizieren können (aufgrund von Immunität oder Tod) oder zumindest ein niedrigeres Ansteckungsrisiko haben (Geimpfte)

Dieses Modell wird um zusätzliche Stages erweitert bzw. in weitere Stages unterteilt:

- Infektiöse Fälle:
 - $I_{Presymp}$: Infektiös bevor man möglicherweise Symptome entwickelt
 - I_{Asymp} : Infektiös ohne Symptome
 - $I_{SympPreHosp}$: Infektiös mit Symptomen (bevor man möglicherweise schwer erkrankt und eine Hospitalisierung notwendig wird)
 - I_{Symp} : Infektiös mit Symptomen (keine schwere Erkrankung/Hospitalisierung)
 - $I_{SympHospPreICU}$: Infektiös und in stationärer Behandlung (bevor möglicherweise eine intensivmedizinische Behandlung erfolgt)
 - $I_{SympHosp}$: Infektiös und in stationärer Behandlung (keine intensivmedizinische Behandlung nötig)
 - $I_{SympHospIcu}$: Infektiös und in intensivmedizinischer Behandlung
- Removed Fälle:
 - $R_{Infection}$: Immun dank überstandener Infektion
 - $R_{Vaccination}$: Geschützt durch Impfung (Infektionsrisiko liegt um 70% geringer als für die Susceptibles - es wird dabei angenommen, dass der Impfschutz erst durch die vollständige Impfung eintritt)
 - R_{Dead} : tot

Die Personen werden im Modell durch die Stages "geschoben". Die Parameter des Modells umfassen dabei die Verweildauern in den Stages und die Übergangswahrscheinlichkeiten zwischen den Stages.

- Verweildauern:

- E : 3 Tage
 - $I_{Presymp}$: 2 Tage
 - I_{Asymp} : 7 Tage
 - $I_{SympPreHosp}$: 1 Tag
 - I_{Symp} : 10 Tage
 - $I_{SympHospPreICU}$: 1 Tag
 - $I_{SympHosp}$: 10 Tage
 - $I_{SympHospIcu}$: 14 Tage
 - $R_{Infection}$: 180 Tage (Überstandene Infektion bietet also ca. ein halbes Jahr Schutz)
 - $R_{Vaccination}$: 365 Tage (Vollständige Impfung bietet als ca. 1 Jahr Schutz)
 - S und R_{Dead} (absorbierender Zustand) haben keine feste Dauer
- Übergangswahrscheinlichkeiten:
 - Übergangswahrscheinlichkeiten bezüglich Schwere der Erkrankung: Schätzung aus historischen Daten (abhängig von der Altersgruppe)
 - Übergangswahrscheinlichkeit bezüglich Infektion ($S - E$): abhängig von:
 - * Basisreproduktionsrate (siehe oben für die Schätzmethode)
 - * Die Anzahl der aktuell infektiösen Fälle
 - * Der Anteil der Immunität in der Bevölkerung durch Impfung und überstandene Infektion
- In der Historie werden die tatsächlichen Zahlen (inklusive Dunkelziffer) verwendet
- Übergangswahrscheinlichkeit bezüglich Impfung ($S - R_{Vaccination}$): abhängig von den vollständigen Impfungen.

Folgende Abbildung zeigt die Stages inklusive der möglichen Übergänge:

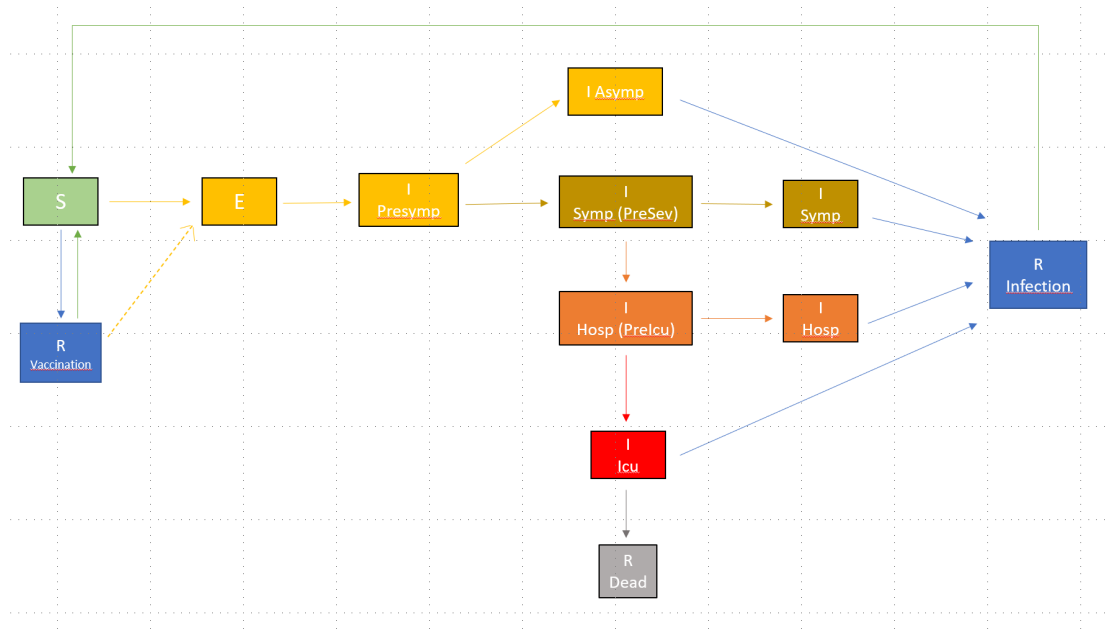


Figure 1: Erweitertes SEIR Modell

Auf Basis der Parameter, der historischen RKI Zahlen (infektionen und Impfungen) und der Dunkelziffer lassen sich die Personenzahlen in jedem Stage an jedem Tag bestimmen (Bis zum Tag $\text{max.date} - 8$): Mit Hilfe der vorhergesagten Reproduktionsraten (Aus den Schritten 1 - 3) lassen sich auch entsprechend Werte für die Zukunft kalkulieren.

Dieses erweiterte SEIR-Modell wird

- für jeden Landkreis
- für jedes Szenario
- für den Erwartungswert und die Prognoseintervallgrenzen der Reproduktionswerte

berechnet.

Mit Hilfe der geschätzten Dunkelziffer lassen sich aus den Prognosedaten (Infektionen) auch Meldetfälle und Meldeinzidenzen vorhersagen. Auch hier wird auf eine wöchentliche Saisonalität geachtet, welche aus der Historie geschätzt wird.