

# Modellierung der Covid-19 Fälle

## Zusammenfassung

Ziel der Modellierung ist die Prognose der Covid-19 Infektionen für die kommenden Monate sowie die Berechnung der dadurch benötigten Betten und Intensivbetten 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 manchen 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 Monate zu bestimmen. Dazu wird das epidemiologische SEIR-Modell herangezogen.

## 1 Schätzung der Reproduktionsrate

Im ersten Schritt wird die Reproduktionsrate 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:

- Der Faktor für die Dunkelziffer beträgt 10 - Die Anzahl der tatsächlichen Fälle ist also 10 mal so hoch wie die vom RKI gemeldeten Zahlen.
- Der Meldeverzug beträgt 8 Tage. Es vergehen also 8 Tage von der Infektion bis zur Meldung durch das RKI.

Nachdem die tatsächlichen Infektionszahlen dadurch berechnet wurden, kann die Reproduktionsrate an einem Tag wie folgt bestimmt werden:

$$R_t = \frac{\sum_{i=t}^{t+(s-1)} NewInf_i}{\sum_{i=t-s}^{t-1} NewInf_i}$$

wobei folgendes gilt:

- $t$ : Index für einen Tag
- $R_t$ : Reproduktionsrate am Tag  $t$
- $NewInf_t$ : Neue Infektionen am Tag  $t$
- $s$ : Hyperparameter für die "Fenstergröße"

Die Formel entspricht der Schätzfunktion des RKI. Der darin enthaltene Parameter  $s$  wird mit 7 festgelegt, um eine gewisse Stabilität zu erzeugen und um die Rate um die wöchentliche Saisonalität zu bereinigen (Am Wochenende werden weniger Fälle gemeldet als unter der Woche). Wenn `max.date` der letzte Tag ist, an dem die gemeldeten Infektionszahlen verfügbar sind, kann die Reproduktionsrate dadurch bis zum Tag `max.date.rate = max.date - 8` (Meldeverzögerung) - 7 (Fenstergröße  $s$ ) geschätzt werden. Ab dem Tag danach sind Forecasts notwendig.

Die Schätzung der Reproduktionsrate erfolgt separat für jeden Landkreis, aber auch für Gesamtdeutschland. Da die Rate bei kleinen Infektionszahlen instabil ist, wird sie mit dem Wert 3 gedeckelt.

## 2 Forecast der Reproduktionsrate

Als Nächstes erfolgt der Forecast der Reproduktionsrate 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` ist zu Beginn der Covid-19 Pandemie 1 und sinkt im Zeitraum 02.03.2020 bis 23.03.2020 (Zeitraum, in dem die Shutdown Maßnahmen ergriffen wurden) linear bis auf 0. Ab dem 20.04.2020 steigt sie wieder schrittweise an. Dies hängt von den Maßnahmen der Politik ab und kann sich auch von Bundesland zu Bundesland unterscheiden. Es ergibt sich dadurch ein ARMAX Modell.

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.

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

$p$  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. Insgesamt werden 100 Tage (vom `max.date` an gerechnet) vorhergesagt. Unsicherheiten werden durch ein 90% Prognoseintervall berücksichtigt. Zudem werden neben dem Basisszenario weitere folgende Szenarien modelliert:

- Zusätzliche Lockerungen ab einem bestimmten Datum
- Vollständige Aufhebung aller Maßnahmen ab einem bestimmten Datum

- Verschärfung der Maßnahmen ab einem bestimmten Datum
- Erneuter Lockdown ab einem bestimmten Datum
- Maßnahmen wie in Bayern werden ergriffen
- Maßnahmen wie in Nordrhein-Westfalen werden ergriffen

### 3 Glättung der Reproduktionsrate

Es liegen nun Reproduktionsraten über die Zeit für jeden Landkreis vor. Diese 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 sind, als Raten entfernter Landkreise.
- Bereinigung von Landkreisen mit geringen Infektionszahlen: Die Schätzungen für Landkreise mit wenigen Daten und/oder kleinen Infektionszahlen ist 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:

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

Das Modell wird im Gegensatz zu Schritt 2 über alle Landkreise hinweg trainiert und ausgewertet. Es wird dazu lediglich das Basis-Szenario herangezogen. Die geglätteten Raten der weiteren Szenarien und die Raten der Prognoseintervallgrenzen erhält man proportional zur nicht geglätteten Variante.

### 4 Modellierung der Infektionszahlen

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

- *S* - Susceptible: Noch nicht infizierte Personen
- *E* - Exposed: Infizierte, aber noch nicht infektiöse Fälle
- *I* - Infectious: Infektiöse Fälle

- $R$  - Recovered: Nicht mehr infektiöse Fälle (aber noch nicht zwangsläufig genesen!)

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

- Infektiöse Fälle:
  - $I_{Asymp}$ : Infektiös ohne Symptome
  - $I_{Symp}$ : Infektiös mit Symptomen
- "Recovered" Fälle:
  - $R_{Asymp}$ : nicht mehr infektiös (asymptomatisch)
  - $R_{Symp}$ : nicht mehr infektiös aufgrund von Quarantäne (symptomatisch)
  - $R_{Hosp}$ : nicht mehr infektiös aufgrund von Quarantäne + Stationärer Aufenthalt aufgrund eines schweren Verlaufs (symptomatisch)
  - $R_{Icu}$ : nicht mehr infektiös aufgrund von Quarantäne + Aufenthalt auf der Intensivstation aufgrund eines schweren Verlaufs (symptomatisch)
- $RF$ : Nicht mehr infiziert: D.h. genesen oder tot. Diese enthält Untergruppen:
  - $Im_{Asymp}$ : Asymptomatische Fälle, die nach einer überstandenen Infektion immun sind
  - $Im_{Symp}$ : Symptomatische Fälle, die nach einer überstandenen Infektion immun sind

Es werden folgende Annahmen getroffen:

- Der Faktor für die Dunkelziffer in Deutschland beträgt 10. Nur 10% der Infizierten zeigen Symptome und werden vom RKI gemeldet.
- Dauer von der Infektion bis zur infektiösität: 3 Tage
- Dauer von der Infektiösität bis zur Bildung der Symptome (bei symptomatischen Fällen): 2 Tage
- Dauer von der Bildung der Symptome bis zur Meldung durch das RKI (bei symptomatischen Fällen): 3 Tage
- Dauer der Infektiösität bei asymptomatischen Fällen: 10 Tage
- Dauer von der Symptombildung bis zur Quarantäne bei symptomatischen Fällen ohne schweren Verlauf: 2 Tage
- Dauer vom Ende der Infektiösität bis zum Ende der Infektion bei asymptomatischen Fällen: 4 Tage

- Dauer vom Beginn der Quarantäne bis zum Ende der Infektion (bei symptomatischen Fällen ohne schweren Verlauf): 10 Tage
- Dauer von der Symptombildung bis zum Beginn des stationären Aufenthalts bei schweren Verläufen: 1 Tag
- Dauer des stationären Aufenthalts: 14 Tage
- Dauer vom Beginn des stationären Aufenthalts bis zum Beginn der intensivmedizinischen Behandlung: 1 Tag
- Dauer der intensivmedizinischen Behandlung: 21 Tage

Folgende Abbildung zeigt die Gruppen inklusive der möglichen Übergänge und Verweildauern in Tagen.

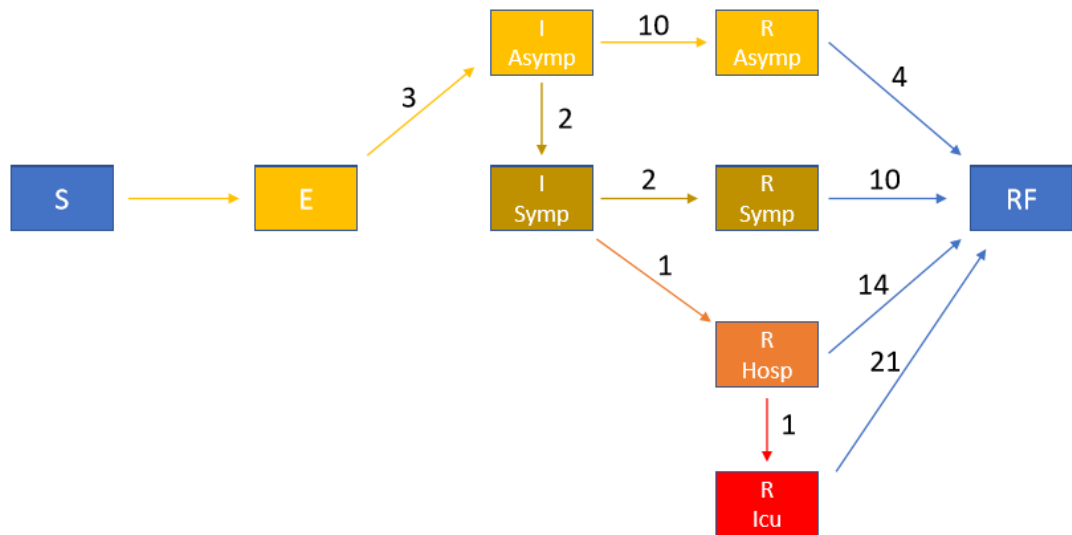


Figure 1: Erweitertes SEIR Modell

Auf Basis der Werte lassen sich mit Hilfe der gemeldeten RKI Zahlen die Infektionszahlen an jedem Tag bestimmen (Bis zum Tag max.date - 8):

- Neue Fälle
- Kumulierte Fälle
- Aktuelle Fälle (Kumulierte Fälle abzüglich Tote und Genesene)

Mit Hilfe der vorhergesagten Reproduktionsraten (Aus den Schritten 1 - 3) lassen sich die Neuinfektionen für die Zukunft berechnen - und daraus entsprechend die aktuellen und kumulierten Fälle.

Dieses erweiterte SEIR-Modell wird

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

berechnet.

Es werden immer die Zahlen für:

- Gesamteinfektionen inklusive Dunkelziffer (alle Gruppen bis auf  $S$  und  $RF$ )
- Symptomatische Fälle (Gruppen:  $I_{Symp}$ ,  $R_{Symp}$ ,  $R_{Hosp}$  und  $R_{Icu}$ )
- Stationär behandelte Fälle (Gruppe  $R_{Hosp}$ )
- Intensivmedizinisch behandelte Fälle (Gruppe  $R_{Icu}$ )

bestimmt. Die Zahlen der symptomatischen Fälle sind mit den Zahlen des RKI vergleichbar.

Für die Hospitalisierungswahrscheinlichkeiten in den verschiedenen Altersgruppen liegen Daten vor. Diese können ebenso wie die Dunkelziffer variieren:

- Räumlich: Je nach Altersstruktur der Bevölkerung des Landkreises
- Zeitlich: Je nach Altersverteilung der gemeldeten Infektionsfälle

Dadurch können pro Landkreis und pro Monat mittlere Dunkelfaktoren, Hospitalisierungswahrscheinlichkeiten (Stationär und Intensiv) berechnet werden.

Für die Immunitätsdauern liegen in der App verstellbare Parameter vor (Standardeinstellungen sind 75 Tage für asymptomatische Fälle und 150 Tage für symptomatische Fälle)