

Schätzung einer zeitabhängigen Reproduktionszahl R für Daten mit einer wöchentlichen Periodizität am Beispiel von SARS-CoV-2-Infektionen und COVID-19

Ralf Mikut, Tillmann Mühlpfordt, Markus Reischl,
Veit Hagenmeyer

Karlsruher Institut für Technologie (KIT), Institut für Automation
und angewandte Informatik, E-Mail: ralf.mikut@kit.edu

17. Mai 2020

Zusammenfassung: Der Beitrag analysiert die Auswirkungen von wöchentlichen Periodizitäten und zeitlichen Korrekturen auf die Schätzung einer zeitabhängigen Reproduktionszahl R bei Infektionskrankheiten. Zur Reduktion dieser Schwankungen wird eine einfache Methode vorgeschlagen, die auf einem akausalen Filter der Filterlänge 7 und optionalen Schätzungen zukünftiger Fallzahlen beruht. Dabei werden die gleichen Tage der Vorwoche als Basis für die Schätzungen verwendet, weil sich das in einer anderen Domäne mit wöchentlichen Periodizitäten – der Lastprognose in Energiezeitreihen – bewährt hat. Akausale Filter vermeiden unerwünschte Zeitverzögerungen, die bei kausalen Filtern auftreten. Die Ergebnisse werden anhand der Fallzahlen von SARS-CoV-2-Infektionen und COVID-19 (Coronavirus Disease 19) in Deutschland mit existierenden Ansätzen des Robert-Koch-Instituts in Deutschland verglichen. Die vorgeschlagene Methode kompensiert wöchentliche Periodizitäten besser und reduziert Phasen mit einer scheinbaren Überschreitung von $R > 1$, die oftmals eine besondere öffentliche Aufmerksamkeit hervorrufen. Darüberhinaus werden die Potenziale und Grenzen von verschiedenen Nowcasting-Modellen aufgezeigt, die Fallzahlen auf ein Erkrankungsdatum projizieren.

1 Motivation

Für eine Infektionskrankheit kann der Erwartungswert einer zeitabhängigen Reproduktionszahl $R[k]$ aus der Historie aller zurückliegenden Fallzahlen und der Wahrscheinlichkeit $w[i]$, dass ein Fall vor i Tagen einen neuen Fall in der Ge-

genwart verursacht, berechnet werden [2]:

$$\mathbf{E}(R[k]) = \frac{N[k]}{\sum_{i=0}^n w[i]N[k-i]}. \quad (1)$$

Die Schätzung der Wahrscheinlichkeiten w_i einer laufenden Infektion ist ein nichttriviales Problem und Gegenstand aktueller Untersuchungen, siehe z.B. [6]. In einer ersten Näherung kann der Wert abgeschätzt werden, wenn für die Werte von $w[i]$

$$w[i] = \begin{cases} 1 & i = k_G \\ 0 & \text{sonst} \end{cases} \quad (2)$$

gesetzt wird, wobei k_G als Generationszeit bezeichnet wird. Damit ergibt sich die Schätzung

$$\mathbf{E}(R[k]) = \frac{N[k]}{N[k-k_G]}. \quad (3)$$

Eine bessere Anpassung an (2) ergibt sich, wenn anstelle von $N[k]$ und $N[k-k_G]$ gefilterte Werte N_F verwendet werden:

$$\mathbf{E}(R[k]) = \frac{N_F[k]}{N_F[k-k_G]}. \quad (4)$$

Dieser Ansatz wird z.B. in [5] für SARS-CoV-2-Infektionen in Deutschland verwendet, wobei dort $k_G = 4$ verwendet wird. Aus Gründen der Vergleichbarkeit der Modelle wird dieser Ansatz auch in dieser Arbeit verwendet.

Dieser Ansatz setzt die Kenntnis der aktuellen und zurückliegenden Fallzahl mit bekanntem Erkrankungsdatum voraus, die während einer laufenden Epidemie noch nicht bekannt sind. Diese Fallzahlen müssen durch aktuell bestätigte Fälle N_{BF} mit einer optionalen Zeitverschiebung oder durch sogenannte Nowcasting-Modelle N_{NC} geschätzt werden [5]. Bei der Berechnung von Gl. (4) entstehen unerwünschte Artefakte, falls die Fallzahlen bzw. ihre Schätzungen Störungen in Form wöchentlicher Periodizitäten (z.B. durch Meldeverzögerungen) enthalten. Obwohl sich gerade aktuell viele Arbeiten mit Schätzern für R beschäftigen (siehe z.B. [1, 6]), ist dieser Aspekt bislang nicht befriedigend gelöst.

Das Ziel des vorliegenden Beitrags besteht darin,

- eine einfache Methode zur Kompensation wöchentlicher Periodizitäten vorzuschlagen und
- anhand der Fallzahlen von SARS-CoV-2-Infektionen in Deutschland die Auswirkungen wöchentlicher Periodizitäten für verschiedene Verfahren aufzuzeigen, wobei in einem ersten Schritt lediglich Punktschätzer (ohne Angabe von Unsicherheiten) untersucht werden.

2 Vorgeschlagene Methode

Für das Verständnis der folgenden Ausführungen ist es wichtig, dass Werte für einen bestimmten Tag zeitlich korrigiert werden können. Das soll am Beispiel der Fallzahlen (im Folgenden mit N bezeichnet) verdeutlicht werden. Die Schreibweise $N[k|k_B]$ bedeutet dabei, dass es sich um die Werte für den Tag k handelt, die am Berechnungstag k_B angenommen werden. Wenn ein Wert vom Vortag am Tag $k_B = k$ berechnet wird, wird das durch die Schreibweise $N[k-1|k]$ gekennzeichnet.

Für die Glättung von Daten mit einer wöchentlichen Periodizität können kausale oder akasale Filter (siehe z.B. [7]) mit einer Filterlänge von sieben Tagen verwendet werden.

Kausale Filter verwenden nur vergangene oder aktuelle Werte, z.B. für ein Filter mit einer Filterlänge 7:

$$N_{KF7}[k|k_B] = \frac{1}{7} \sum_{i=-6}^0 N[k+i|k_B]. \quad (5)$$

Dabei gehen die Fallzahlen $N[k|k_B]$ für den Tag k sowie sechs Tage aus der Vergangenheit ein. Der Nachteil kausaler Filter ist es, dass sie eine Phasenverschiebung verursachen. Das wirkt wie ein zeitverzögertes Verhalten gegenüber den ursprünglichen Fallzahlen. Dies ist für eine schnelle Erkennung und zeitgenaue Diagnose von Effekten unerwünscht.

Akasale Filter verwenden auch zukünftige Werte aus Sicht von k :

$$N_{AF7}[k|k_B] = \frac{1}{7} \sum_{i=-3}^3 N[k+i|k_B]. \quad (6)$$

Dabei beschreibt $N_{AF7}[k|k_B]$ den durch das akasale Filter berechneten Wert für den Tag k am Berechnungstag k_B . Diese Berechnung von Gl. (6) ist für alle $k+3 \leq k_B$ unkritisch, weil dann auch die Daten für die Abtastzeitpunkte bis $k+3$ vorliegen. Für zukünftige Zeitpunkte müssen diese Werte geschätzt werden.

Um die Auswirkungen von wöchentlichen Periodizitäten zu reduzieren, erfolgt die Schätzung der zukünftigen Fallzahlen auf der Basis des gleichen Wochentags der Vorwoche. Solche Ansätze sind aus vielen Anwendungsbereichen mit typischen Wochenrhythmen bekannt, z.B. für die Lastprognose in Energiezeitreihen [4]. Erweiterungen wie die Kompensation von Feiertagen können bei Bedarf vorgenommen werden. Für Fallzahlen bietet sich eine Schätzung durch

$$\hat{N}[k+7|k_B] = R[k+7|k_B]^{\frac{7}{k_G}} N[k|k_B] \quad (7)$$

an. In einigen Fällen ist der benötigte zeitlich korrekte Wert von R (z.B. bei $k_B = k$ mit $R[k+7|k]$) nicht bekannt und wird durch den letzten verfügbaren Wert (oftmals $R[k]$) ersetzt. Hierbei ist darauf zu achten, dass der Wert von $R[k|k]$ nur auf bekannten Fallzahlen beruht, um instabile Schätzungen zu vermeiden.

Wenn die zukünftigen Werte in Gl. (6) durch Gl. (7) geschätzt werden, ergibt sich ein akausales Filter mit geschätzten Werten:

$$N_{AFS7}[k|k] = \frac{1}{7} \left(\sum_{i=-3}^0 N[k+i|k] + R[k|k]^{\frac{7}{k_G}} \sum_{i=-6}^{-4} N[k+i|k] \right). \quad (8)$$

Aus dem gleichen Grund sollte bei der Berechnung von Gl. (4) darauf geachtet werden, dass sowohl im Zähler als auch Nenner identische Annahmen getroffen werden. Der Wert im Nenner sollte also nicht von Schätzungen gemäß Gl. (7) abhängen.

3 Fallbeispiel SARS-CoV-2-Infektionen in Deutschland

3.1 Daten und Modelle

Bei den im Folgenden verwendeten Bezeichnungen kennzeichnen die Symbole ”-H” Modelle zur historischen Einschätzung des Infektionsgeschehens und ”-A” Modelle für tagesaktuelle Schätzungen. Modelle mit ”RKI” stammen vom Robert-Koch-Institut, Modelle ”NEU” basieren auf der in dieser Arbeit vorgeschlagenen Methode.

Das Robert-Koch-Institut publiziert für Infektionen mit SARS-CoV-2 tägliche Lageberichte. Diese Lageberichte enthalten u.a. die Zahl der neu bestätigten Fälle vom Vortag in Deutschland $N_{BF}[k-1|k]$ und die auf einem Nowcasting-Modell basierende Schätzung der Reproduktionszahl R . Das zugehörige Modell für eine tagesaktuelle Schätzung des Infektionsgeschehens wird im Folgenden als RKI-A bezeichnet. Das Modell RKI-A verwendet die Fallzahlen des Erkrankungsbeginns auf der Basis eines Nowcasting-Modells [5]. Diese Daten werden mit einem kausalen Filter der Filterlänge 4 entsprechend Gl. (5) geglättet. Anschließend wird die Reproduktionszahl vor vier Tagen auf der Basis von Gl. (4) geschätzt, wobei eine Generationszeit von $k_G = 4$ angenommen wird. Für die vorliegende Arbeit wurden die täglichen Lageberichte vom 10.3.2020 bis zum 16.5.2020 verwendet. k bezieht sich hier jeweils auf das Datum des täglichen Lageberichts. Aufgrund der Verschiebungen um vier Tage korrespondieren diese Werte mit einem Erkrankungsdatum vom 6.3.2020 bis zum 12.5.2020.

Zusätzlich werden auf der Webseite des Robert-Koch-Instituts¹ Ergebnisse von Modellen zur historischen Analyse des Infektionsgeschehens (RKI-H) publiziert. Das Modell RKI-H arbeitet prinzipiell identisch zum Modell RKI-A und verwendet die zeitlich korrigierten Fallzahlen des Erkrankungsbeginns des Nowcasting-Modells [5]. Diese Fallzahlen $N_{NC}[k-i|k_B], i \geq 1$ und daraus berechnete $R[k-i|k_B], i \geq 4$ werden als Excel-Tabelle publiziert. Für die vorliegende Arbeit wurde die Excel-Tabelle vom 16.5.2020 mit einem Erkrankungsdatum ab dem 2.3.2020 verwendet. Auch hier wird ein kausales Filter mit der Filterlänge 4 verwendet.

¹<https://www.rki.de>

Ab dem täglichen Lagebericht vom 14.5.2020 wurde der RKI-Ansatz dahingehend geändert, zusätzlich ein Modell zu berechnen, das ein kausales Filter der Filterlänge 7 verwendet, um den Wert von $R[k + 1]$ zu schätzen [9]. Dieses Modell wird im Folgenden mit RKI7-H bezeichnet. Die entsprechenden Werte für ein Modell RKI7-A stehen ebenfalls erst ab 14.5.2020 zur Verfügung und werden deshalb nicht mit einbezogen.

Das Modell NEU-H basiert auf dem gleichen Nowcasting-Modell, glättet aber dessen Ausgänge mit dem akausalen Filter mit Filterlänge 7 aus Gl. (6). In der Variante NEU-HA wird das akausale Filter mit den Schätzungen aus der Vorwoche gemäß Gl. (8) verwendet, um auch die Ergebnisse für $k = k_B$ schätzen zu können. Das Prinzip zur Schätzung der Reproduktionszahl entspricht den RKI-Modellen. Das Modell NEU-HA ist strukturell auch für tagesaktuelle Schätzungen geeignet, wenn statt $N_{NC}[k - i|k_B]$ der Wert $N_{NC}[k - 4|k]$ sowie der entsprechende gefilterte Wert verwendet wird. Allerdings konnte das nicht für den kompletten Betrachtungszeitraum validiert werden, weil die historischen Daten des Nowcasting-Modells $N_{NC}[k - 4|k]$ nicht verfügbar sind.

Die Langzeitbetrachtungen der Modelle RKI-H, RKI7-H, NEU-H und NEU-HA sind wegen der Nowcasting-Korrektur zurückliegender Werte nicht zur tagesaktuellen Einschätzung der Infektionslage geeignet. Deswegen wird neben dem Modell RKI-A ein Modell NEU-A untersucht. Hierbei werden Meldezahlen der bestätigten Fälle als Modelleingang verwendet. Weil auch hier eine möglichst aktuelle Schätzung angestrebt wird, werden die benötigten Fallzahlen aus der Zukunft mit Gl. (8) geschätzt. Damit werden die Fallzahlen für die Tage $k = k, k + 1, k + 2$ ermittelt, weil die Fallzahl am Tag des Lageberichts beim Erscheinen des Lageberichts noch nicht bekannt ist.

Eine Übersicht über die verwendeten Modelle zeigt Tabelle 1.

Modell	Fallzahl N	Geglättete Fallzahl N_F	Reproduktionszahl R
RKI-H	$N_{NC}[k - i k_B]$	$N_{NC,KF4}[k - i k_B]$	$R[k - i, k_B]$
RKI7-H	$N_{NC}[k - i k_B]$	$N_{NC,KF7}[k - i k_B]$	$R[k - i + 1, k_B]$
NEU-H	$N_{NC}[k - i k_B]$	$N_{NC,AF7}[k - i k_B]$	$R[k - i, k_B]$
NEU-HA	$N_{NC}[k - i k_B]$	$N_{NC,AFS7}[k - i k_B]$	$R[k - i, k_B]$
RKI-A	$N_{NC}[k - 4 k]$	$N_{NC,KF4}[k - 4 k]$	$R[k - 4, k]$
NEU-A	$N_{BF}[k - 1 k]$	$N_{BF,AFS7}[k - 1 k]$	$R[k - 4, k]$

Tabelle 1: Modelle des Robert-Koch-Instituts (RKI) und neu vorgeschlagene Modelle (NEU). Modelle mit der Erweiterung -H eignen sich für die Analyse historischer Daten (Berechnungstag k_B als Datum des letzten täglichen Lageberichts), Modelle mit der Erweiterung -A für tagesaktuelle Schätzungen (Berechnungstag k). Die historischen Daten des Nowcasting-Modells $N_{NC}[k - 4|k]$ und die daraus ermittelten gefilterten Werte $N_{NC,KF4}[k - 4|k]$ sind leider auf der Webseite des Robert-Koch-Instituts nicht öffentlich verfügbar, könnten aber bei Zurverfügungstellung miteinbezogen werden.

Die verwendeten Daten und die Implementierungen der NEU-Modelle sind als Excel-Tabelle COVID-19_Deutschland.xlsx <https://github.com/RalfMikut> sowie als Julia-Code inkl. aktueller Visualisie-

rungen unter <https://github.com/timueh/COVID-19> verfügbar.

3.2 Rückwirkende Bewertung des Infektionsgeschehens

Sowohl die bestätigten Fälle N_{BF} als auch die Nowcasting-Fallzahlen des Modells N(RKI-H) weisen wöchentliche Periodizitäten auf (Abb. 1a)², wobei der Verlauf der bestätigten Fälle ca. 11 Tage verzögert ist. Das auch die Nowcasting-Zahlen noch von den Wochentagen abhängen, könnte u.U. an der Verwendung des Meldungsdatums als Erkrankungsdatum bei symptomlosen Fällen liegen.

Die geglätteten Fallzahlen von RKI-H, RKI7-H, NEU-H und NEU-HA können wöchentliche Periodizitäten deutlich reduzieren.

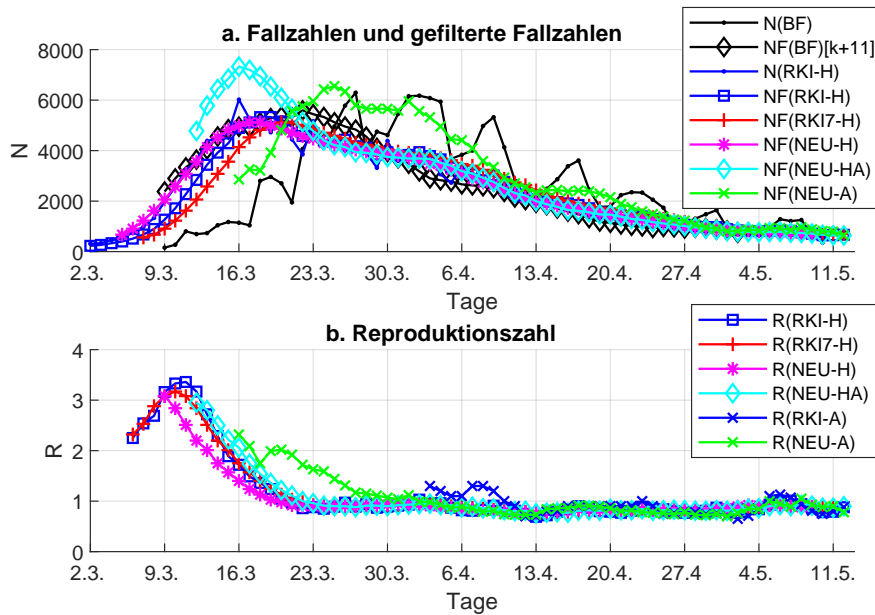


Abbildung 1: a. Fallzahlen und gefilterte Fallzahlen für die verschiedenen Modelle, b. Schätzungen der Reproduktionszahl R . Dargestellt ist der Zeitraum des Erkrankungsdatums vom 2.3.2020 bis zum 12.5.2020.

Ein Kriterium für die Güte eines Filters ist der Mittlere Absolute Fehler (engl. Mean Absolute Error - MAE) der geglätteten Fallzahlen gegenüber den Fallzahlen (Tabelle 2). Hier zeigt sich, dass das im Modell NEU-H verwendete Filter den prinzipiellen Verlauf besser widerspiegelt als das Modell RKI-H. Eine wesentliche Ursache dafür sind die deutlichen Abweichungen des RKI-H-Modells, die aus der Zeitverzögerung bei steilen Funktionsverläufen zwischen dem 9.-16.3. resultieren. Diese Zeitverzögerung resultiert aus der Verwendung

²Die Grafik wurde mit der Open Source MATLAB-Toolbox SciXMiner [8] erstellt. SciXMiner ist unter <https://sourceforge.net/projects/scixminer> verfügbar.

des kausalen Filters in RKI-H. Noch deutlicher ausgeprägt ist das im Modell RKI7-H wegen der größeren Filterlänge (7 statt 4). Das Modell NEU-HA erkennt den Trendwechsel um den 17.3. später; dies führt ebenfalls zu deutlichen Abweichungen (Abb. 1a).

	RKI-H	RKI7-H	NEU-H	NEU-HA
MAE gegen $N_{NC}[k-i k_B]$	219	353	186	340

Tabelle 2: Vergleich des mittleren absoluten Fehlers der geglätteten Fallzahlen gegenüber den Fallzahlen des Nowcasting-Modells für den Zeitraum vom 15.3.2020 bis zum 9.5.2020.

Die Schätzungen der Reproduktionszahl der Modelle zeigt Abb. 1b. Auch hier zeigen sich deutliche Zeitverschiebungen zwischen kausalen und akausalen Modellen. Besonders ausgeprägt ist dieser Unterschied am 10.3. und 11.3., an welchen das akausal arbeitende Modell NEU-H bereits ein deutliches Fallen der Reproduktionszahl anzeigt.

Für eine detailliertere Analyse wird in Abb. 2 nur der Zeitraum zwischen dem 3.4.2020 bis zum 12.5.2020 visualisiert. Hier ist auffällig, dass die Reproduktionszahl von Modell RKI-H noch deutliche wöchentliche Periodizitäten aufweist.

Eine detaillierte Analyse dieser Schwankungen erfolgt mit lokalen Minima und Maxima der Reproduktionszahl in einer Woche. Der Schwellwert von $R > 1$ wird gemäß des RKI-H-Modells am 21.3.2020 und gemäß des Modells NEU-H am 22.3.2020 unterschritten. Danach gibt es eine längere Phase mit einem relativ stabilen Infektionsgeschehen bis Anfang Mai. In diesen Wochen weist das RKI-H-Modell immer noch erhebliche Schwankungen mit z.T. ausgeprägten lokalen Minima und Maxima auf (Tabelle 3). Die Maxima treten vorzugsweise an Donnerstagen als Beginn der Erkrankung auf.

Woche	RKI-H	NEU-H
	MIN (WT)/MAX WT	MIN (WT)/MAX WT
16.3.-22.3.2020	0.86 (SO) - 1.72 (MO)	0.90 (SO) - 1.40 (MO)
23.3.-29.3.2020	0.85 (DI) - 0.94 (SA)	0.88 (DI) - 0.92 (SO)
30.3.-5.4.2020	0.87 (SO) - 1.03 (DO)	0.97 (SO) - 0.95 (DI)
6.4.-12.4.2020	0.75 (SO) - 0.91 (DO)	0.78 (SO) - 0.85 (MO)
13.4.-19.4.2020	0.69 (MO) - 0.90 (FR)	0.79 (DI) - 0.84 (FR)
20.4.-26.4.2020	0.77 (DI) - 0.89 (DO)	0.81 (SO) - 0.83 (MO)
27.4.-2.5.2020	0.77 (MO) - 0.87 (DO)	0.81 (DI) - 0.92 (SO)
3.5.-9.5.2020	0.78 (SO) - 1.01 (DO)	0.88 (FR) - 0.92 (DI)

Tabelle 3: Lokale Minima und Maxima von R -Schätzern für die Analyse historischer Daten. Der RKI7-H-Schätzer ist vergleichbar zu NEU-H und deswegen nicht dargestellt.

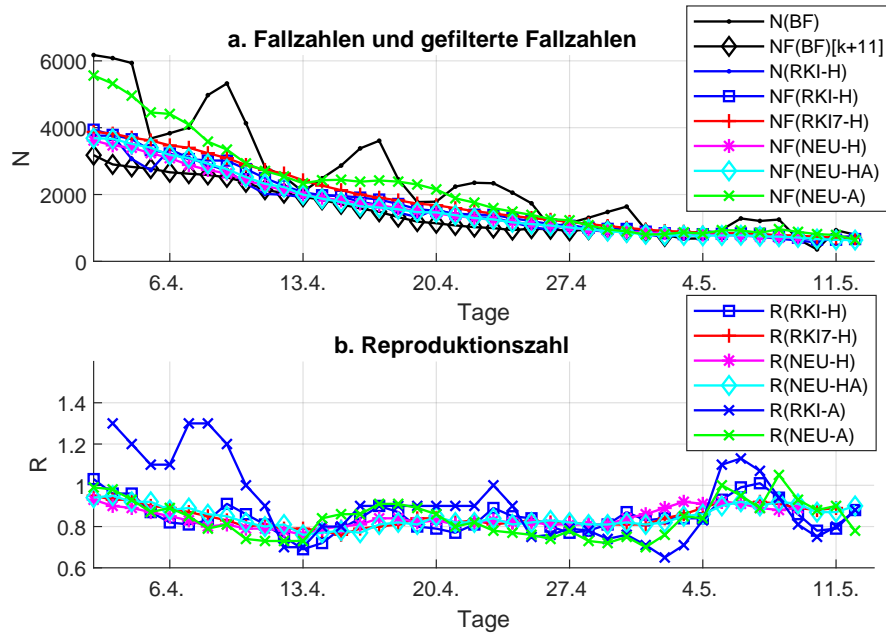


Abbildung 2: a. Visualisierung der Nowcasting-Schätzungen und der Zahl der bestätigten Fälle, b. Durch akasale und kausale Filter berechnete Werte. Dargestellt ist der Zeitraum der Verfügbarkeit aller Modelle vom 3.4.2020 bis zum Erkrankungsdatum 12.5.2020.

3.3 Tagesaktuelle Einschätzung der Infektionslage

Der Verlauf für Modell NEU-A in Abb. 1 zeigt, dass das Modell den Fallzahlen und Reproduktionszahlen der H-Modelle nachläuft. Dieser Effekt ist unvermeidbar und kann nur abgeschwächt werden. Ein gutes Modell versucht den zeitlichen Abstand zu reduzieren, ohne größere Schwankungen der Reproduktionszahl oder systematische Abweichungen aufzuweisen. Diese Aufgabe löst das Modell NEU-A besser als das Modell RKI-A (Abb. 2b). Modell RKI-A hat deutliche Schwankungen mit einer klar erkennbaren wöchentlichen Periodizität. Tabelle 4 gibt einen detaillierten quantitativen Einblick in die wöchentlichen lokalen Minima und Maxima. Die lokalen Minima beider Modelle liegen eher auf Erkrankungsdaten am Wochenende, die Lage der Maxima ist nicht eindeutig bestimmbar.

3.4 Quantitativer Vergleich aller Modelle

Im folgenden Abschnitt wird verglichen, ob es systematische Unterschiede in den Modellen gibt. Dabei werden insbesondere die wöchentlichen Periodizitäten im Detail untersucht. Die ungeglättete Fallzahl für das RKI-H-Modell mit einem

Woche	RKI-A	NEU-A
	MIN (WT)/MAX WT	MIN (WT)/MAX WT
6.4.-12.4.2020	0.70 (SO) - 1.30 (DI-MI)	0.73 (SO) - 0.89 (MO)
13.4.-19.4.2020	0.70 (MO) - 0.90 (DO-SO)	0.73 (MO) - 0.91 (FR)
20.4.-26.4.2020	0.75 (SA) - 1.00 (DO)	0.74 (SO) - 0.86 (MO)
27.4.-2.5.2020	0.65 (SA) - 0.79 (MO)	0.72 (MI) - 0.85 (SA)
3.5.-10.5.2020	0.75 (SO) - 1.13 (MI)	0.84 (MO) - 1.05 (FR)

Tabelle 4: Lokale Minima und Maxima von R -Schätzern für tagesaktuelle Schätzungen.

Erkrankungsdatum vom 3.4.2020 bis zum 12.5.2020 weist deutliche Schwankungen mit einem lokalen Maximum am Erkrankungsdatum Freitag (1949 Fälle) und einem Minimum mit Erkrankungsdatum Montag (1495 Fälle) auf. Hier ist aus den Daten nicht ermittelbar, ob es sich um einen zutreffenden Zusammenhang oder um einen Modelleffekt handelt. Diese Schwankungen wirken sich unterschiedlich stark auf die Reproduktionszahlen aller Modelle aus, die für jeden Wochentag separat gemittelt werden (Tabelle 5). Besonders stark sind die Schwankungen in den Modellen RKI-H und RKI-A.

Erkr.	Meldung	RKI-H	RKI-7H	NEU-H	NEU-HA	RKI-A	NEU-A
MO	FR	0.78	0.85	0.83	0.83	0.86	0.82
DI	SA	0.80	0.84	0.83	0.84	0.98	0.84
MI	SO	0.85	0.83	0.83	0.84	0.97	0.83
DO	MO	0.90	0.83	0.84	0.84	0.99	0.82
FR	DI	0.88	0.83	0.84	0.84	0.96	0.82
SA	MI	0.86	0.84	0.84	0.84	0.88	0.82
SO	DO	0.81	0.84	0.84	0.84	0.83	0.82

Tabelle 5: Vergleich der Wochentagsabhängigkeit der Mittelwerte der Schätzer der Reproduktionszahl für einen Zeitraum mit Erkrankungsdatum (Erkr.) vom 3.4.2020 bis zum 7.5.2020 (vollständige Wochen für alle Schätzer). In den ersten beiden Spalten ist der Wochentag der Erkrankung und der Wochentag der Meldung aufgeführt.

Tabelle 6 zeigt außerdem, dass das Modell RKI-A gegenüber den Modellen RKI-H, RKI7-H und NEU-H die Reproduktionszahl tendenziell überschätzt. Eine mögliche Ursache der Überschätzung ist das Nowcasting-Modell, das für Werte nahe dem Berechnungstag k_B tendenziell zu hohe Werte aufweist. Diese Annahme könnte aber bei Verfügbarkeit von $N_{NC}[k - 4|k]$ für vergangene Werte von k quantitativ überprüft werden. Allerdings finden sich in mehreren täglichen Lageberichten des RKI (z.B. am 26.4.2020) Hinweise auf diesen Sachverhalt, die ein Ansteigen der Zahlen des Nowcasting-Modells zeigen, das sich später in der Realität nicht bestätigt hat. Die Schwankungen in Zusammenwirken mit der Überschätzung führen an insgesamt 9 Tagen zum Überschreiten des Schwellwerts $R > 1$. Mit dem Modell NEU-A kann eine bessere Annäherung bei geringeren Schwankungen erreicht werden, wobei nur für einen Tag $R > 1$ gilt.

Wert	Mittelwert und Standardabweichung	Tage mit $R > 1$	MAE gegen RKI7-H
RKI-H	0.84 ± 0.07	1	0.05
RKI7-H	0.84 ± 0.05	0	0.00
NEU-H	0.84 ± 0.04	0	0.02
NEU-HA	0.85 ± 0.05	0	0.01
RKI-A	0.91 ± 0.18	9	0.14
NEU-A	0.85 ± 0.08	1	0.05

Tabelle 6: Vergleich der Mittelwerte, Standardabweichungen, Anzahl der Tage mit $R > 1$ sowie mittlerer absoluter Fehler gegenüber dem Modell RKI-H der Schätzer für einen Zeitraum mit Erkrankungsdatum vom 3.4.2020 bis zum 12.5.2020. Dem Schätzer NEU-H steht wegen des akausalen Filteransatzes nur der Zeitraum bis zum 9.5.2020 zur Verfügung, dem Schätzer RKI7-H nur der Zeitraum bis zum 11.5.2020.

Zusammenfassend kann festgestellt werden, dass sich die Modelle RKI7-H und NEU-H besser für die rückwirkende Bewertung des Infektionsgeschehens (z.B. zur Beurteilung von Lockdowns oder Öffnungen) eignen als das Modell RKI-H. Modell NEU-A eignet sich besser für tagesaktuelle Schätzungen als Modell RKI-A. Eine wesentliche Ursache liegt darin, dass in den neuen Modellen die Unterdrückung der wöchentlichen Periodizitäten besser gelingt als in den RKI-Modellen.

Hervorzuheben ist noch die Tatsache, dass das Modell NEU-A direkt auf den bestätigten Fällen arbeitet und damit leicht nachvollziehbar ist.

4 Empfehlungen

Die Analysen zeigen, dass wöchentliche Periodizitäten die Ergebnisse von Modellen und die Situationseinschätzung beeinträchtigen können. So ist beispielsweise aus unserer Sicht die Aussage nicht haltbar, dass die Reproduktionszahl von SARS-CoV-2-Infektionen Anfang Mai wieder deutlich über $R > 1$ gestiegen ist. Hier handelt es sich um eine Überinterpretation der Daten durch das Nowcasting-Modell, die sich insbesondere auf das Modell RKI-A auswirkt. Allerdings zeigen alle Modelle für diesen Zeitraum einen Anstieg gegen einen Wert von Eins und damit ein Stagnieren der Fallzahlen an (Abb. 2b).

Daraus ergeben sich folgende generelle Empfehlungen:

- Die Wochentagsabhängigkeit aller Ergebnisse sollte bei der Schätzung der Reproduktionszahl konsequent geprüft werden.
- Wenn Fallzahlen wochentagsabhängig sind, sollten durchgängig Filter mit einer Filterlänge von sieben Tagen verwendet werden.
- Unerwünschte Zeitverzögerungen können durch die Verwendung akausaler Filter zumindest teilweise kompensiert werden.

- Ergebnisse von Nowcasting-Modellen, die nahe an der Gegenwart agieren, müssen kritisch hinterfragt werden, um Fehlinterpretationen vorzubeugen.

In zukünftigen Arbeiten wird untersucht werden, wie sich wöchentliche Periodizitäten auf die Quantifizierung von Unsicherheiten in der Schätzung von R auswirken. Ein möglicher Ansatzpunkt sind hier Quantilfilter [3], die zur Erreichung angemessener Beispielzahlen optional mit Monte-Carlo-Simulationen angereichert werden können. In diesem Kontext ist zudem eine systematische Studie mit synthetischen Daten sinnvoll.

Danksagung: Die Autoren danken dem Robert-Koch-Institut für die öffentliche Bereitstellung der COVID19-Daten und -Modellergebnisse. RM wurde durch die Helmholtz Information and Data Science School – HIDSS4Health und Helmholtz AI gefördert. Die Autoren danken Tim Scherr für seine Hinweise und Korrekturen zum Manuskript.

Autorenbeiträge: Initiales Konzept, Datenanalyse, erster Entwurf des Manuskripts: RM; Implementierung: RM, TM; Diskussion des Konzepts, Interpretation der Daten und Verbesserung des Manuskripts: RM, TM, MR, VH; Verifikation der Datenanalyse: RM, MR, TM; Mitteleinwerbung: RM, VH.

Literatur

- [1] FLAXMAN, S.; MISHRA, S.; GANDY, A.; UNWIN, H.; COUPLAND, H.; MELLAN, T.; ZHU, H.; BERAH, T.; EATON, J.; PEREZ GUZMAN, P.; ET AL.: Report 13: Estimating the Number of Infections and the Impact of Non-Pharmaceutical Interventions on COVID-19 in 11 European Countries. Techn. Ber., Imperial College London, 2020.
- [2] FRASER, C.: Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. *PLoS One* 2 (2007) 8.
- [3] GONZÁLEZ ORDIANO, J. Á.; GRÖLL, L.; MIKUT, R.; HAGENMEYER, V.: Probabilistic Energy Forecasting using the Nearest Neighbors Quantile Filter and Quantile Regression. *International Journal of Forecasting* 36 (2020) 2, p. 310–323.
- [4] GONZÁLEZ ORDIANO, J. Á.; WACZOWICZ, S.; HAGENMEYER, V.; MIKUT, R.: Energy Forecasting Tools and Services. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8(2) (2018), p. e1235.
- [5] AN DER HEIDEN, M.; HAMOUDA, O.: Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland–Nowcasting. *Epidemiologisches Bulletin* 17 (2020), p. 10–16.
- [6] HOTZ, T.; GLOCK, M.; HEYDER, S.; SEMPER, S.; BÖHLE, A.; KRÄMER, A.: Monitoring the Spread of COVID-19 by Estimating Reproduction Numbers over Time. *arXiv preprint arXiv:2004.08557* (2020).

- [7] KAMMEYER, K.-D.; KROSCHER, K.: *Digitale Signalverarbeitung: Filterung und Spektralanalyse mit MATLAB-Übungen*. Springer-Verlag, 2013.
- [8] MIKUT, R.; BARTSCHAT, A.; DONEIT, W.; GONZÁLEZ ORDIANO, J. Á.; SCHOTT, B.; STEGMAIER, J.; WACZOWICZ, S.; REISCHL, M.: The MATLAB Toolbox SciXMiner: User's Manual and Programmer's Guide. Techn. Ber., arXiv:1704.03298, 2017.
- [9] N.N.: Erläuterung der Schätzung der zeitlich variierenden Reproduktionszahl R (15.5.2020). Techn. Ber., Robert-Koch-Institut, 2020.