



AUSARBEITUNG FÜR DAS MODUL „AKTUELLES AUS
DER FORSCHUNG“

Coronavirus und das Wetter in Deutschland

*Auswirkung wetterbedingter Faktoren auf die
Infektionszahlen von COVID-19 anhand einer
datengetriebenen Analyse*

Autor: Jan Löwenstrom
Matrikelnr.: 34937
Prüfer: Prof. Dr. Leif Meier

16. Juli 2020

Abstract

Diese Forschungsarbeit hat die Korrelationen zwischen dem Wetter und der COVID-19 Pandemie für Deutschland im Untersuchungszeitraum vom 01.03.2020 bis 01.04.2020 analysiert.

Im Rahmen dieser Arbeit ist der Datensatz zu der Anzahl der Neuinfizierten des Robert-Koch-Instituts mit den Wetterdaten des Deutschen Wetterdienstes in Verbindung gebracht worden. Dabei sind vier wetterbedingte Faktoren auf Tagesebene für die Analyse herangezogen worden, die Temperatur ($^{\circ}\text{C}$), die Sonnenscheindauer, die Windgeschwindigkeit (m/s) und die relative Luftfeuchtigkeit (%).

Eine Korrelationsanalyse nach Pearson ist durchgeführt worden unter Einbeziehung unterschiedlicher Verzögerungen (Lag). Die resultierenden Ergebnisse haben gezeigt, dass Temperatur ($r_{\text{Lag}=5} = 0.420$) und Sonnenscheindauer ($r_{\text{Lag}=5} = 0.314$) eine positive Korrelation (mittlerer Effekt) mit den Fallzahlen aufweisen. Windgeschwindigkeit ($r_{\text{Lag}=5} = -0.079$) und relative Luftfeuchtigkeit ($r_{\text{Lag}=5} = 0.258$) haben keinen bzw. nur einen schwachen Zusammenhang aufgezeigt. Aus den Werten für den Korrelationskoeffizienten im Bezug auf die rel. Luftfeuchtigkeit hat sich jedoch eine weitere These abgeleitet, dass trockene Luft einen Krankheitsausbruch womöglich begünstigt.

Durch die Datenverarbeitung mit R sind verknüpfte und gefilterte Datensätze entstanden, die der Forschung für weitere Analysen zur Verfügung gestellt worden sind.

Inhaltsverzeichnis

1. Einleitung	5
2. Zielsetzung und Fragestellung	5
3. Forschungsstand	6
4. Abgrenzung	7
5. Methodik	8
5.1. Vorgehensweise	8
5.2. Datensätze	9
5.3. Korrelationsanalyse	12
6. Ergebnisse	13
7. Reflexion	15
8. Fazit	17
Literatur	18
A. Erklärung zur Seminararbeit	21

Abbildungsverzeichnis

1.	Neuinfektionen pro Landkreis pro Tag	10
2.	Neuinfektionen pro Tag in Deutschland	10
3.	Landkreisposition (orange) und Wetterstationen des DWD (blau) .	11
4.	Tägliche durchschnittliche Temperatur pro Wetterstation	11
5.	Tägliche Sonnenscheindauer pro Wetterstation	11
6.	Tägliche durchschnittliche Windgeschwindigkeit pro Wetterstation .	12
7.	Tägliche durchschnittliche relative Luftfeuchtigkeit pro Wetterstation	12
8.	Korrelation Neuinfektionen und durch. Temperatur	13
9.	Korrelation Neuinfektionen und tägliche Sonnenscheindauer	13
10.	Korrelation Neuinfektionen und durch. Windgeschwindigkeit	14
11.	Korrelation Neuinfektionen und durch. rel. Luftfeuchtigkeit	14

1. Einleitung

Es beginnt alles im Dezember 2019, als in der chinesischen Stadt Wuhan der erste Fall eines neuartigen, unbekannten Coronavirus bei einem Menschen festgestellt wird (World Health Organisation, 2020a). Seitdem verbreitet sich dieses Virus auf der ganzen Welt und wurde am 11. März 2020 von dem WHO-Leiter Tedros Adhanom Ghebreyesus zu einer globalen Pandemie ausgesprochen (World Health Organisation, 2020b). Dabei steht das umgangssprachlich benutzte Wort „Coronavirus“ lediglich für eine Familie von Viren, die bei Menschen und Tieren vorkommen kann (World Health Organisation, 2020a). Die offizielle, detaillierte Bezeichnung des Virus ist allerdings *Coronavirus SARS-CoV-2* (Robert Koch Institut, 2020a), wobei auch die Kurzform COVID-19 valide ist und im Laufe dieser Arbeit verwendet wird. Dabei steht COVID-19 für „Coronavirus Disease 2019“ (World Health Organisation, 2020a).

Unterschiedliche Thesen und Paper bezüglich des Einflusses wetterbedingter Faktoren auf COVID-19 werden zurzeit veröffentlicht und offen diskutiert. Diese Arbeit soll durch eine datengetriebene Analyse einen Beitrag zu dieser Diskussion leisten.

2. Zielsetzung und Fragestellung

Ziel dieser Arbeit ist es, zu untersuchen, ob wetterbedingte Faktoren Einfluss auf die Verbreitungsgeschwindigkeit von COVID-19 haben. Im Zusammenhang damit ergeben sich folgende Fragestellungen:

- Breitet sich das Virus in einem bestimmten Temperaturbereich rasanter aus?
- Reduzieren oder Erhöhen geringe Windgeschwindigkeiten die Infektionsrate?
- Sorgt eine erhöhte Intensität der Ultraviolettstrahlung, d.h. eine längere Sonnenscheindauer bei geringerer Bewölkung für eine reduzierte Übertragungsrate?
- Inwiefern spielt die relative Luftfeuchtigkeit eine Rolle?
- Lassen sich Muster erkennen im Hinblick auf die Korrelation und den verwendeten Verzug (Lag)?

Zudem sollen Daten aus unterschiedlichen Quellen erhoben, gefiltert, verarbeitet und verknüpft werden. Da es sich um eine Analyse ausschließlich für das Land Deutschland handelt, sind einige Datenquellen wie z.B. Landkreisinformationen sehr schwierig für Forscher*innen zu erschließen, die nicht deutschsprachig sind.

Die resultierenden Datensätzen sollen deshalb ins Englische übersetzt und nach Abschluss der Untersuchungen anderen Forscher*innen frei zugänglich gemacht werden, um den Aufwand vor weiteren Analyse zu minimieren.

3. Forschungsstand

Dieses Kapitel gibt einen Überblick über Arbeiten, die den Zusammenhang von Viruserkrankungen und Wetterfaktoren untersucht haben. Dabei werden zunächst einige Veröffentlichungen genannt, die nicht speziell SARS-CoV-2 behandeln, sondern Influenzaviren allgemein und den Einfluss wetterbedingte Faktoren auf jene. Anschließend werden kurz die Ergebnisse von Arbeiten aufgeführt, die sich auf COVID-19 konzentrieren.

Untersuchungen im Bezug auf das 2003 ausgebrochenen SARS-CoV Virus in Beijing zeigen, dass Temperatur, Luftfeuchtigkeit, Windgeschwindigkeit und Bewölkung in signifikanter Korrelation mit der Ausbreitung stehen (Yuan et al., 2006). Das Ergebnis des Papers zeigt vor allem, dass Temperatur ($r^2 = 0.718$), relative Luftfeuchtigkeit ($r^2 = -0.784$) und Windgeschwindigkeit ($r^2 = 0.617$) starke Korrelationen aufweisen.

SARS steht für „Schweres Akutes Respiratorisches Syndrom“ (Robert Koch Institut, 2020d), was bedeutet, dass COVID-19 in einigen Fällen zu einer Pneumonie mit Lungenversagen führen kann. Das RKI meldet (Stand 29.04.2020), dass es in 2,7% der Fälle in Deutschland zu einer Pneumonie kommt, der hauptsächlichen Todesursache des Virus (Robert Koch Institut, 2020c). Dass die Todesrate von Patienten mit Lungenentzündung stark mit Wetterfaktoren wie Luftfeuchtigkeit und Temperatur korreliert, konnte bereits in einer Studie von Bull (1980) gezeigt werden. Die von Davis et al. (2012) durchgeführte retrospektive Studie im Bezug auf die Grippe, Lungenentzündungen und Mortalitätsrate für die Stadt New York für den Zeitraum 1975 bis 2002 unterstützt die These, dass kalte, trockene Luft einen Anstieg der Todesrate zur Folge hat.

COVID-19 ist ein Virus, somit kann durch intensives Händewaschen die Virushülle aufgebrochen und der Erreger unschädlich gemacht werden (Robert Koch Institut, 2020b). Unklar ist, ob und inwiefern UV-Licht den Virus im Aerosol schadet. Studien hinsichtlich der Influenza von Chumpolbanchorn et al. (2006) und Weiss et al. (2007) kommen zu keinem klaren Ergebnis.

Aktuell werden zahlreiche Paper veröffentlicht, die sich mit den Wetterfaktoren und der Ausbreitung oder Mortalitätsrate von COVID-19 beschäftigen. Hierbei

sind viele Paper noch nicht ausreichend überprüft worden, was eine akribische Auswahl voraussetzt. Seriöse Veröffentlichungen sind u.a. die Arbeiten von Wang et al. (2020), Notari (2020) und Briz-Redón & Serrano-Aroca (2020).

Dabei kommen die Arbeiten zu unterschiedlichen Ergebnissen. Wang et al. (2020) zeigen, dass hohe Temperaturen und geringe Luftfeuchtigkeit die Ausbreitung von COVID-19 signifikant reduzieren. Tosepu et al. (2020) finden in ihrer *Spearman*-Korrelationsanalyse für Jakarta einen mittleren Effekt ($r = .0.392$) bei dem Wetterfaktor „durchschnittliche Temperatur pro Tag“.

Die statistisch komplexe Untersuchung mittels *Integrated Nested Laplace Approximations* (INLA, siehe (Martino & Riebler, 2019)) von Briz-Redón & Serrano-Aroca (2020) kann jedoch für Spanien keinen Beweis für eine Korrelation von Temperatur und Fallzahlen erkennen.

4. Abgrenzung

In diesem Kapitel findet eine Abgrenzung gegenüber anderen Arbeiten statt. Zudem werden Rahmenbedingungen und Grenzen aufgezeigt, die für diese Forschungsarbeit von Bedeutung sind.

Wie in den Papern von Briz-Redón & Serrano-Aroca (2020) und Tosepu et al. (2020) findet auch hier eine länderspezifische Analyse statt. Konkret bedeutet dies, dass ausschließlich Fallzahlen und Wetterdaten für Deutschland ausgewertet und zur Verfügung gestellt werden.

Weiterhin zu vermerken ist, dass es sich um eine reine datengetriebene Analyse handelt, bei der öffentlich zugängliche Daten miteinander verknüpft werden, um daraufhin Korrelationen zwischen Ausbreitung des Virus und einzelner Wetterfaktoren zu erforschen. Die Untersuchung findet somit auf einer sehr hohen Abstraktionsebene statt. Laborexperimente mit dem *SARS-CoV-2* Erreger, die z.B. den direkten Einfluss von UV-Strahlung messen könnten, sind nicht Teil dieser Ausarbeitung.

5. Methodik

Dieses Kapitel stellt den Ablauf und die Bestandteile dieser Forschungsarbeit dar. Zunächst wird die Vorgehensweise beschrieben und die Intentionen hinter den einzelnen Schritten begründet. Anschließend folgt ein Unterkapitel, in dem die verwendeten Datensätze vorgestellt und zum Teil visualisiert werden. Zuletzt wird auf die verwendete Methode zur Bestimmung des Korrelationskoeffizienten eingegangen und eine Interpretationshilfe für die Ergebnisse aufgezeigt.

5.1. Vorgehensweise

Der erste wichtige Schritt ist die Festlegung des Untersuchungszeitraumes. Wie in der Arbeit von Briz-Redón & Serrano-Aroca (2020) bezieht sich die Analyse auf den Zeitraum bevor Maßnahmen in Kraft getreten sind, also von dem 01.03.2020 bis zu dem 01.04.2020. In diesem Zeitraum steigt die Anzahl der Neuinfizierten pro Tag nahezu exponentiell an, jedoch mit einzelnen Tagen, an denen deutlich weniger Neuinfizierte registriert worden sind, was eventuell wetterbedingten Faktoren geschuldet ist.

Als Nächstes wird der Datensatz mit den Fallzahlen des Robert-Koch-Instituts aufbereitet. Beispiele für Transformationen sind unter anderem die Gruppierung nach Landkreisen, das Auffüllen von Tagen mit 0 Neuinfizierten, das Mutieren zu einem einheitlichen Datumsformat (YYYY-mm-dd) und das Bilden der kumulativen Summe der Neuinfizierten pro Tag. Ein maßgebendes Detail ist, dass als Datum das Referenzdatum gewählt wird, welches das Erkrankungsdatum repräsentiert, wenn das Feld *IstErkrankungsbeginn* = 1 ist. Somit wird ein passender Filter verwendet, um nur Fallzahlen mit Referenzdatum als bestätigten Erkrankungsbeginn zu benutzen, um eine Durchmischung mit Melddaten an das Gesundheitsamt zu verhindern. Durch diesen Filter ist gewährleistet, dass eine Verschiebung (Lag) hauptsächlich die Inkubationszeit nachahmt und nicht zum Beispiel Meldeverzug, der entsteht, bis eine Neuinfektion an das Gesundheitsamt übermittelt wird.

Im weiteren Verlauf wird der Fallzahlen-Datenframe mit Zusatzinformationen zu den einzelnen Landkreisen mithilfe von Daten der Database of Global Administrative Areas (GADM) angereichert. Dabei wird die Lage, der Mittelpunkt des Landkreises, und somit der originale *Geo Point* in zwei Spalten *LAT* und *LON* unterteilt. Grund hierfür ist, dass auch die Positionen der Wetterstationen mit diesen beiden räumlichen Informationen beschrieben sind und somit eine einheitliche Form entsteht.

Es folgt die Beschaffung der Wetterdaten. Hierzu werden zunächst alle verfügbaren Stationen des Deutschen Wetterdienstes anhand einer Informationsliste gefiltert, um nur Stationen und ihre Kennungen zu sammeln, die aktiv im Untersuchungszeitraum Werte gemessen haben. Sämtliche Wetterdaten werden pro Station in einem .zip-Archiv gespeichert. Ein automatisierter Prozess lädt sämtliche, relevanten Archive der einzelnen Wetterstationen herunter, extrahiert diese und überführt die passende Datei mit den täglichen Durchschnitten in einen Datenframe.

Pro ausgewertetem Wetterfaktor werden ausschließlich Stationsdaten verwendet, deren Station für den gesamten Untersuchungszeitraum tatsächlich Werte für jenen Faktor gemessen haben. Dadurch schwankt die Anzahl an Wetterstationen, die für die Durchschnittsberechnung verwendet werden, für jeden Wetterfaktor. Beispielsweise messen 162 Stationen die durchschnittliche Windgeschwindigkeit pro Tag, 392 hingegen die durchschnittliche Temperatur pro Tag.

Die eigentliche Analyse geschieht durch die Verbindung (*join*) der Fallzahlen pro Tag und dem täglichen Durchschnitt aller Wetterstationen für jeweils einen bestimmten Faktor (Temperatur, Sonnenscheindauer, Windgeschwindigkeit und rel. Luftfeuchtigkeit). Es entsteht ein Datenframe, der die Neuinfektionen und einen Wetterfaktor auf Tagesebene verknüpft. Anschließend wird eine Korrelationsanalyse durchgeführt. Besonders dabei ist, dass die Analyse für unterschiedliche Lags abgewickelt wird. Konkret bedeutet dies, dass das *recordingDate* der Wetterdaten um X Tage in die Zukunft versetzt wird. Anders ausgedrückt werden dadurch die Neuinfektionen an einem bestimmten Tag mit dem Wetter von vor X Tagen verglichen, wobei X einen Wert von 0 bis 20 annimmt.

5.2. Datensätze

Fallzahlen COVID-19

Als Datenquellen für die COVID-19 Infektionen pro Tag als Zeitreihe wird das Robert-Koch-Institut gewählt (ArcGIS, 2020). In Zusammenarbeit mit dem Bundesamt für Kartographie und Geodäsie ergibt sich eine Tabelle, die für jeden Landkreis, jeweils pro Tag und pro Altersgruppe die Anzahl der Neuinfektionen festhält.

Der Datensatz enthält zwei unterschiedliche Datierungen. Zum einen das *Meldedatum*, das Datum, wann der Fall dem Gesundheitsamt bekannt geworden ist und zum anderen das *Referenzdatum*, welches dem Erkrankungsdatum entspricht, wenn zusätzlich der Wert für *IstErkrankungsbeginn* 1 ist, sonst ist es übereinstimmend mit dem *Meldedatum*.

Letztendlich bezieht sich die Korrelation jedoch nicht auf Landkreisebene, sondern auf die Summe aller Neuinfektionen pro Tag.

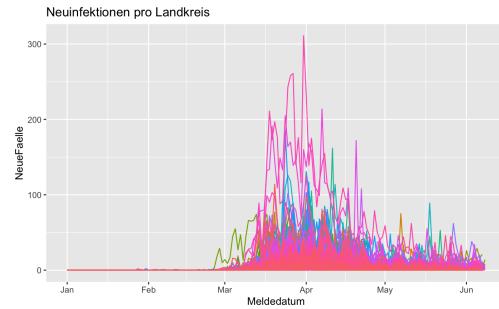


Abbildung 1: Neuinfektionen pro Landkreis pro Tag

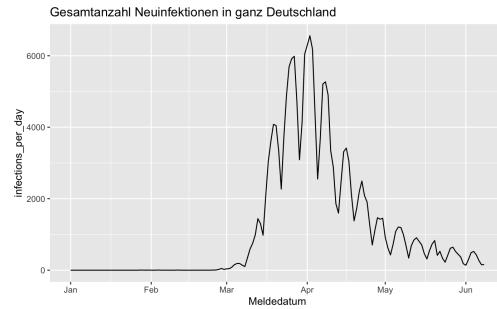


Abbildung 2: Neuinfektionen pro Tag in Deutschland

Landkreisinformationen

Um den Datensatz des RKI mit weiteren Informationen zu den einzelnen Landkreisen anzureichern, wird auf Daten der Database of Global Administrative Areas (GADM) zurückgegriffen (opendatasoft.com, 2020). Verknüpft werden hierbei vor allem die Positionsdaten (*Geo Point*) der einzelnen Landkreise über den Amtlicher Gemeindeschlüssel, um die Positionen einzelnen Kreise bildlich darzustellen und die am nächstenen Wetterstationen zu berechnen.

Wetterdaten

Ursprünglich ist geplant, die Wetterstationen des National Oceanic and Atmospheric Administration (NOAA) zu verwenden. Die Abdeckung für Deutschland ist allerdings sehr gering. Aus diesem Grund werden die Wetterstationen und -daten des Deutschen Wetterdienstes (DWD) für die Analysen verwendet.

Hierbei stellt der DWD die Daten von 1324 Wetterstationen einzeln als Download zur Verfügung. Die Verteilung der Stationen stellt sich wie folgt da:

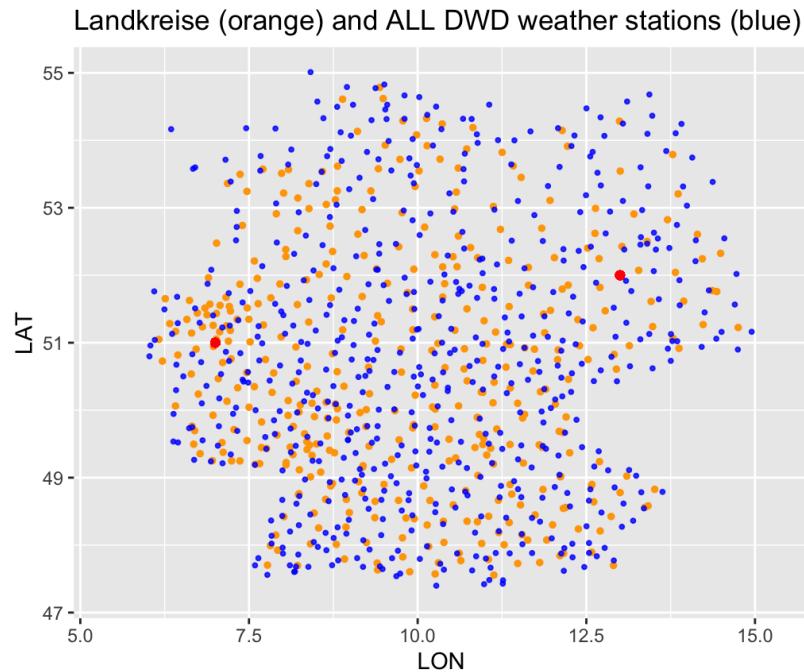


Abbildung 3: Landkreisposition (orange) und Wetterstationen des DWD (blau)

Als Datenintervall wird „täglich“ ausgewählt, d.h. der Datensatz einer Station beinhaltet für den Messzeitraum jeweils die täglichen Durchschnittswerte bzw. die gesamte Sonnenscheindauer des Tages. Dies ist sinnvoll, da die Fallzahlen ebenfalls auf täglicher Basis erhoben werden.

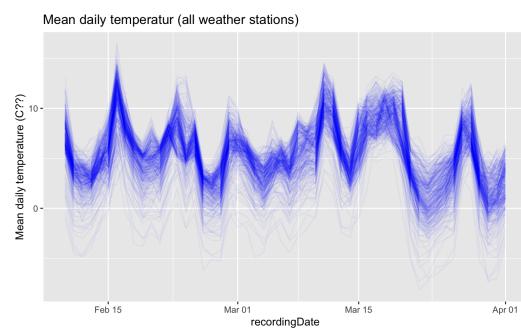


Abbildung 4: Tägliche durchschnittliche Temperatur pro Wetterstation

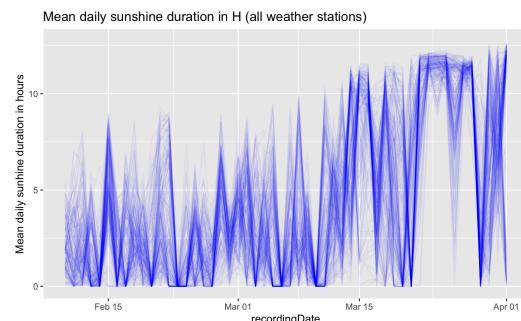


Abbildung 5: Tägliche Sonnenschein-dauer pro Wetterstation

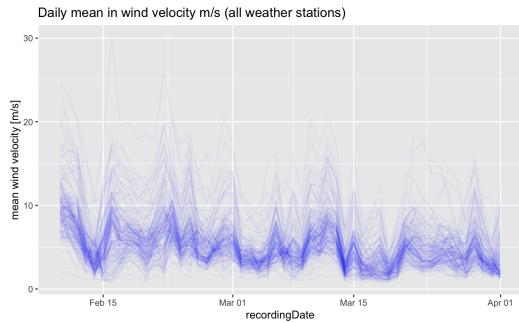


Abbildung 6: Tägliche durchschnittliche Windgeschwindigkeit pro Wetterstation

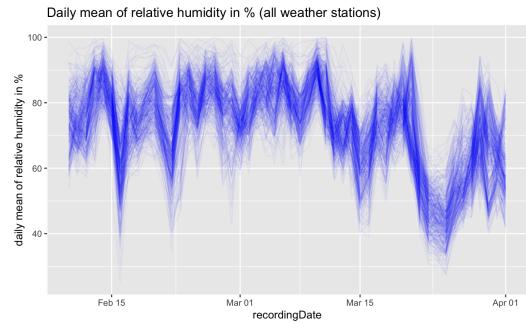


Abbildung 7: Tägliche durchschnittliche relative Luftfeuchtigkeit pro Wetterstation

5.3. Korrelationsanalyse

Im Rahmen dieser Arbeit wurde stets der Zusammenhang zwischen zwei Variablen untersucht, genauer die Neuinfektionen mit jeweils einem Wetterfaktor wie Temperatur, Sonnenscheindauer oder Windgeschwindigkeit. Das Ergebnis einer Korrelationsanalyse ist in diesem Fall der bivariate Zusammenhang, der durch den Korrelationskoeffizienten r ausgedrückt wird. In der Arbeit von Tosepu et al. (2020) wird eine Rangkorrelation nach Spearman durchgeführt. Nach Auffassung des Autors erfüllen die Variablen jedoch die Voraussetzung der Intervallskalierung, weshalb eine Korrelation nach Bravais-Pearson zur Berechnung des linearen Zusammenhangs verwendet wird, ohne den einhergehenden Informationsverlust bei einer Rangierung (Backhaus et al., 2016).

Der Korrelationskoeffizient r nach Bravais-Pearson lässt sich wie folgt berechnen (Zürich, 2018):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2)(\sum_{i=1}^n (y_i - \bar{y})^2)}} \quad (1)$$

wobei x_i, y_i den Werten der beiden Variablen, \bar{x}, \bar{y} den arithmetischen Mitteln der beiden Variablen und n der Stichprobengröße entspricht.

Für eine Einschätzung darüber, welche Interpretationen aus den berechneten Werten für r gezogen werden können, hilft eine Orientierung durch die Einteilung nach Cohen (1992):

- $r = 0.1$ entspricht einem schwachen Effekt
- $r = .30$ entspricht einem mittleren Effekt
- $r = .50$ entspricht einem starken Effekt

6. Ergebnisse

Nachfolgend werden die Ergebnisse der Korrelationsanalyse vorgestellt. Wichtig ist, dass die Balkendiagramme keinen zeitlichen Verlauf darstellen, sondern die Korrelationskoeffizienten je nach verwendetem Lag. Hierbei ist zudem entscheidend, dass für eine Gesamtbeurteilung der vier Faktoren immer der selbe Wert für den Verzug gewählt werden muss.

Betrachtet werden zunächst die Ergebnisse für die durchschnittliche Tagstemperatur und die durchschnittliche Sonnenscheindauer pro Tag:

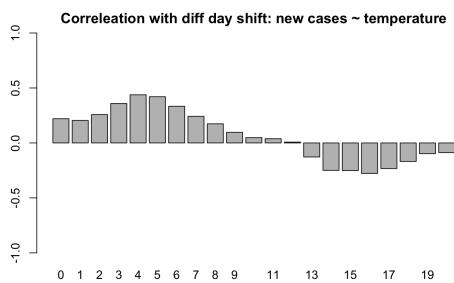


Abbildung 8: Korrelation Neuinfektionen und durch. Temperatur

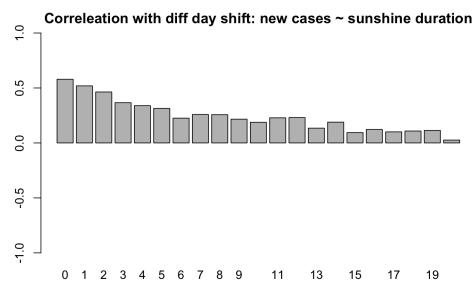


Abbildung 9: Korrelation Neuinfektionen und tägliche Sonnenscheindauer

Auffällig in der Abbildung 8 ist, dass sich der Korrelationskoeffizient bei einem Lag von 0 bis 12 nahezu normalverteilt verhält, bei dem der Höhepunkt bei einem Lag von 4 Tagen den Wert $r = +0.438$ erreicht. Die durchschnittliche Inkubationszeit von COVID-19 beträgt 5-6 Tage, wobei die maximale Zeitspanne bis zu 14 Tage betragen kann (WHO, 2020). Es kann davon ausgegangen werden, dass die Zeiten für eine Inkubation ebenfalls normalverteilt sind, was sich in den Zusammenhängen zwischen Neuinfizierten und Temperatur widerspiegelt. Erst ab einem Lag von 13 Tagen korreliert die Temperatur negativ mit der Anzahl der Neuinfizierten.

Abgeleitet von den Erkenntnissen der Linton et al. (2020) zu den Inkubationszeiten, ist ein Fokus auf die Korrelationskoeffizienten für Lag = 5 am plausibelsten. Das Datum der Neuinfizierten ist jenes, das den Erkrankungsbeginn markiert, somit haben sich die meisten Menschen 4-6 Tage zuvor (unter der Annahme der Normalverteilung) infiziert. Genau für diese Tage also einem Lag von 4-6 ist die genauere Betrachtung der Wetterverhältnisse besonders interessant.

Bezogen auf die Temperatur bedeutet dies, dass laut Orientierung zur Einordnung des Korrelationskoeffizienten nach Cohen (1992) ein positiver, mittlerer Effekt festgehalten werden kann ($r_4 = +0.438$, $r_5 = 0.420$, $r_6 = 0.330$). Auch die Son-

nenscheindauer, siehe Abb. 9, korreliert ausschließlich positiv mit der Anzahl der Neuinfizierten ($r_4 = 0.339$, $r_5 = 0.314$, $r_6 = 0.226$). Anders ausgedruckt heißt dies, dass je wärmer es wird und je länger die Sonne scheint, desto mehr Menschen infizieren sich mit COVID-19. Dies widerspricht der ursprünglichen Annahme, dass höhere Temperaturen und intensivere UV-Strahlung die Verbeitungsgeschwindigkeit verlangsamen.

Neben Temperatur und Sonnenscheindauer werden auch die durchschnittliche Windgeschwindigkeit und die relative Luftfeuchtigkeit als wetterbedingte Faktoren untersucht:

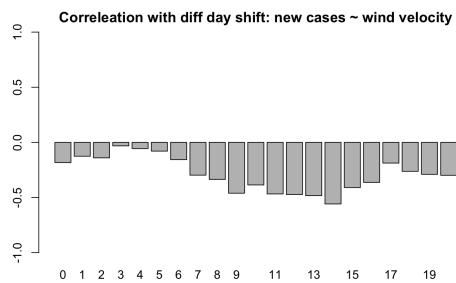


Abbildung 10: Korrelation Neuinfektionen und durch. Windgeschwindigkeit

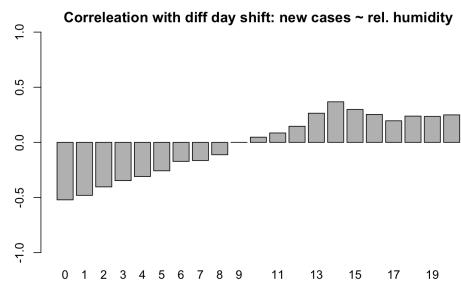


Abbildung 11: Korrelation Neuinfektionen und durch. rel. Luftfeuchtigkeit

Zu erkennen ist, dass die Windgeschwindigkeit ausnahmslos negativ korreliert. Wird der Verzug auf die maximale Spannweite der Inkubationszeit, also 14 Tage, angesetzt, so ergibt sich sogar ein sehr starker negativer Effekt von $r_{14} = -0.559$. Jedoch kann für den plausibleren Lag von 4-6 Tagen kein bzw. für Lag = 6 lediglich ein schwacher Zusammenhang entdeckt werden ($r_4 = -0.056$, $r_5 = -0.079$, $r_6 = -0.156$).

Für die relative Luftfeuchtigkeit ist im plausiblen Bereich ein mittlerer bis schwacher negativer Effekt ermessbar ($r_4 = -0.301$, $r_5 = -0.258$, $r_6 = -0.172$). Durch die Abbildung 11 lässt sich allerdings eine weitere interessante These ableiten, und zwar, dass eine niedrige Luftfeuchtigkeit die Wahrscheinlichkeit eines Ausbruchs der Krankheit erhöht.

Zu Anfang dieses Kapitels wurde erwähnt, dass die Balkendiagramme nicht als zeitliche Abfolge interpretiert werden können. Dies ist korrekt, andererseits bezieht sich diese Aussage auf den Kontext des Übertragungszeitpunktes. Wird von Lag = 0 ausgegangen, so ist der Tag des Ausbruchs der Krankheit gemeint (Lag = 1 dementsprechend der Tag vor dem Ausbruch), da das *Referenzdatum* dem Erkrankungsbeginn entspricht. Für diese beiden Tage ergibt sich ein sehr starker

negativer Effekt ($r_0 = -0.521$, $r_1 = -0.480$). Als Hintergrund für die Aufstellung dieser These gelten vor allem die Arbeiten von Yuan et al. (2006) und Davis et al. (2012). In dem Paper von Yuan et al. (2006) wird eine starke negative Korrelation zwischen relativer Luftfeuchtigkeit und Ausbreitung von SARS-CoV bestätigt. Davis et al. (2012) zeigt in einer retrospektiven Studie im Bezug auf Grippe und Lungenentzündungen, dass trockene Luft einen Anstieg der Todesraten zur Folge hat. Niedrige Luftfeuchtigkeiten scheinen also einen unmittelbaren Einfluss auf Verbreitung und eventuell auch auf den Zeitpunkt des Krankheitsausbruches zu haben.

Zusammengefasst sehen die Korrelationskoeffizienten für den Tag des Ausbruchs, den plausibelsten Tag der Übertragung und der maximalen Spannweite des Inkubationszeitraumes wie folgt aus:

Wetterfaktor	LAG=0	LAG=5	LAG=14
Temperatur	0.220	0.420	-0.249
Sonnenscheindauer	0.579	0.314	0.189
Windgeschwindigkeit	-0.184	-0.079	-0.558
Rel. Luftfeuchtigkeit	-0.521	-0.258	0.368

7. Reflexion

Dieses Kapitel umfasst eine kritische Auseinandersetzung mit der Methodik dieser Forschungsarbeit, eine Einschätzung darüber wie aussagekräftig die Ergebnisse sind und welche zusätzlichen Ansätze im Laufe der Bearbeitung verworfen worden sind.

Zunächst einmal muss festgestellt werden, dass die verwendeten Daten, also Fallzahlen für Gesamtdeutschland pro Tag und die durchschnittlichen Wetterwerte pro Tag von 400 Wetterstationen sich auf einer sehr hohen Abstraktionsebene befinden. Dies ist auf der einen Seite günstig, da dadurch die Gesamtheit an Einflussfaktoren wegrationalisiert wird, jedoch sind die Ergebnisse dadurch weniger aussagekräftig bezogen auf eine konkrete Ursachenfrage.

Generell gesehen entscheiden zu viele Faktoren, ob ein gewisser Wetterfaktor überhaupt Einfluss auf COVID-19 hat. Deutlich wird dies auch durch die Diskussion mit Kommiliton*innen zum Faktor Windgeschwindigkeit. Sollte viel und starker Wind die Ausbreitung vergünstigen, weil die Viren sich dadurch einfacher verbreiten können? Oder ist genau das Gegenteil der Fall und die Coronaviren werden sofort in höhere Luftsichten getragen, sodass das Infektionsrisiko gesenkt wird?

Gerade die Ergebnisse zu Temperatur und Sonnenscheindauer zeigen ein widersprüchliches Bild zu den theoretischen Annahmen. Jedoch spielen auch hier weitere Gegebenheiten eine Rolle. Zum einen kann angenommen werden, dass besseres Wetter dazu führt, dass mehr Menschen draußen unterwegs sind und so das generellen Risiko einer Übertragung steigt, selbst wenn der Wetterfaktor an sich eine Infektion erschwert. Zum anderen ist der untersuchte Zeitraum März genau jener, bei dem der Virus ungehindert, das heißt ohne Maßnahmen, wüten könnte. Dadurch steigen die Fallzahlen stetig an. In Verbindung damit, dass Temperatur und Sonnenscheindauer durch den startenden Frühling ebenfalls steigen, entsteht schon hierdurch eine Korrelation.

Aber eine Korrelationsanalyse ist dennoch sinnvoll, da sowohl die Neuinfizierten als auch die Wetterdaten vereinzelt starke Schwankungen aufweisen. Der Autor ist sich lange Zeit uneinig darüber gewesen, ob eine Rangkorrelation nach Spearman oder eine Korrelation nach Pearson besser geeignet ist. Letztendlich wird der Korrelationskoeffizient nach Pearson berechnet, da zwei intervallskalierte Merkmale vorliegen. Zudem unterscheiden sich die berechneten Werte für r bei beiden Methoden nicht signifikant voneinander.

Eine Analyse zum Zusammenhang von Wetterfaktoren und Reproduktionsrate nach der Einführung von Maßnahmen, d.h. nach dem Höhepunkt der Infektionszahlen, als der R-Wert um 1 schwankt, sind zwar durchgeführt worden, jedoch nicht Teil der Ergebnisse. Ausschlaggebend dafür, dass diese nicht aufgeführt sind, ist, dass der Reproduktionswert zu einem bestimmten Tag andere Implikationen hat. Konkret sagt er aus, wie viele Menschen ein Infizierte in den nächsten 7 Tagen ab diesem Zeitpunkt anstecken wird. Der Wert eines Wetterfaktors für lediglich einen Tag hat somit fast keine Aussagekraft. Dem Autor ist dieser Zusammenhang erst zu spät aufgefallen. Als mögliches Merkmal für eine Korrelation mit R-Werte könnte der sog. Gleitender Mittelwert (*Moving average*) eines Wetterfaktors über ebenfalls 7 Tage Verwendung finden.

Eingangs ist geplant, sehr nahe an dem Paper von Briz-Redón & Serrano-Aroca (2020) zu arbeiten, um anhand der gleichen Methodik eine Aussage für Deutschland anstatt Spanien zu treffen. Dies ist auch der Grund, warum zunächst auf Landkreisebene Informationen aggregiert werden. Der Autor muss jedoch zugeben, dass die verfügbare Zeit und sein mathematischer Horizont nicht ausreichend sind, um einerseits das statistische Modell (Integrated Nested Laplace Approximations) der Arbeit umzusetzen und zusätzlich korrekt zu interpretieren. Durch ein komplexes Modell wie INLA ist es möglich, zusätzliche Parameter mit in die Analyse aufzunehmen und eine granularere Untersuchung auf Landkreisebene vorzunehmen.

Zusätzliche Parameter sind u.a. Bevölkerungsdichte, Altersgruppen, Reisende, Reproduktionsrate. All diese Faktoren finden in dieser Arbeit keine Berücksichtigung.

8. Fazit

Eine konkrete Aussage über den Einfluss bestimmter Wetterfaktoren auf die Verbreitung von COVID-19 für Deutschland kann nicht aufgestellt werden. Die Ergebnisse zeigen zum Teil ein konträres Bild zu den theoretischen Annahmen. So korrelieren durchschnittliche Tagestemperatur und Sonnenscheindauer positiv mit der Anzahl der Fallzahlen ($r_5^{temp} = 0.432$ und $r_5^{sun} = 0.314$). Ein möglicher Erklärungsansatz ist die größer werdende Ansammlung von Menschen bei zunehmend schönerem Wetter, sowie die unterliegende Korrelation zwischen frühlingsbedingter Wetterbesserung und ungehinderter exponentieller Verbreitung des Virus im Untersuchungszeitraum, bei dem keine Maßnahmen aktiv sind.

Für den Wetterfaktor „durchschnittliche Windgeschwindigkeit pro Tag“ kann, unter der Annahme eines plausiblen Bereiches bedingt durch die durchschnittliche Inkubationszeit von 5 Tagen, kein Zusammenhang gefunden werden.

Im Bezug auf den Ansteckungszeitpunkt weist auch die relative Luftfeuchtigkeit nur einen schwachen Effekt auf ($r_5 = -0.258$). Allerdings zeigen die Daten zum Zeitpunkt des Krankheitsausbruches eine starke Auffälligkeit. Ein sehr starker negativer Effekt ($r_0 = -0.521$) ist präsent, was die These zulässt, dass die Wahrscheinlichkeit eines Krankheitsausbruches durch trockene Luft erhöht wird.

Das Erschließen, Verknüpfen und Verarbeiten mehrere Datenquellen im Rahmen dieser Ausarbeitung sorgt für das Bereitstellen eines einheitlichen Datensatzes, um Forscher*innen anknüpfende Untersuchungen zum Thema COVID-19 und das Wetter deutlich zu erleichtern. Diese Datensätze sind frei zugänglich, unter <https://github.com/kono94/corona-weather-dataset>.

Literatur

- ArcGIS. (2020). *Tabelle mit den aktuellen covid-19 infektionen pro tag (zeitreihe)*. Zugriff am 2020-07-04 auf <https://www.arcgis.com/home/item.html?id=dd4580c810204019a7b8eb3e0b329dd6>
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2016). *Multivariate analysemethoden*. Springer.
- Briz-Redón, A. & Serrano-Aroca, A. (2020). A spatio-temporal analysis for exploring the effect of temperature on covid-19 early evolution in spain. *Science of The Total Environment*, 138811.
- Bull, G. (1980). The weather and deaths from pneumonia. *The Lancet*, 315 (8183), 1405 - 1408. Zugriff auf <http://www.sciencedirect.com/science/article/pii/S0140673680926665> (Originally published as Volume 1, Issue 8183) doi: [https://doi.org/10.1016/S0140-6736\(80\)92666-5](https://doi.org/10.1016/S0140-6736(80)92666-5)
- Chumpolbanchorn, K., Suemanotham, N., Siripara, N., Puyati, B. & Chaichoune, K. (2006). The effect of temperature and uv light on infectivity of avian influenza virus (h5n1, thai field strain) in chicken fecal manure.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112 (1), 155.
- Davis, R. E., Rossier, C. E. & Enfield, K. B. (2012, 03). The impact of weather on influenza and pneumonia mortality in new york city, 1975–2002: A retrospective study. *PLOS ONE*, 7 (3), 1-8. Zugriff auf <https://doi.org/10.1371/journal.pone.0034091> doi: 10.1371/journal.pone.0034091
- Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-m., ... Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9 (2). Zugriff auf <https://www.mdpi.com/2077-0383/9/2/538> doi: 10.3390/jcm9020538
- Martino, S. & Riebler, A. (2019). *Integrated nested laplace approximations (inla)*.
- Notari, A. (2020). Temperature dependence of covid-19 transmission. *medRxiv*. Zugriff auf <https://www.medrxiv.org/content/early/2020/04/24/2020.03.26.20044529> doi: 10.1101/2020.03.26.20044529

- opendatasoft.com. (2020). *Deutschland landkreise*. Zugriff am 2020-07-04 auf <https://public.opendatasoft.com/explore/dataset/landkreise-in-germany/table/>
- Robert Koch Institut. (2020a). *Covid-19 (coronavirus sars-cov-2)*. Zugriff am 2020-05-01 auf https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/nCoV.html
- Robert Koch Institut. (2020b). *Hinweise zu reinigung und desinfektion von oberflächen außerhalb von gesundheitseinrichtungen im Zusammenhang mit der covid-19-pandemie*. Zugriff am 2020-05-01 auf https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Reinigung_Desinfektion.html
- Robert Koch Institut. (2020c). *Sars-cov-2 steckbrief zur coronavirus-krankheit-2019 (covid-19)*. Zugriff am 2020-05-01 auf https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html
- Robert Koch Institut. (2020d). *Sars (schweres akutes respiratorisches syndrom)*. Zugriff am 2020-05-01 auf <https://www.rki.de/DE/Content/InfAZ/S/SARS/SARS.html?>
- Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, L. O. A. I., Lestari, H., Bahar, H. & Asfian, P. (2020). Correlation between weather and covid-19 pandemic in jakarta, indonesia. *Science of The Total Environment*, 725, 138436. Zugriff auf <http://www.sciencedirect.com/science/article/pii/S0048969720319495> doi: <https://doi.org/10.1016/j.scitotenv.2020.138436>
- Wang, J., Tang, K., Feng, K. & Lv, W. (2020). High temperature and high humidity reduce the transmission of covid-19. *Available at SSRN 3551767*.
- Weiss, M. M., Weiss, P. D., Weiss, D. E. & Weiss, J. B. (2007). Disrupting the transmission of influenza a: face masks and ultraviolet light as control measures. *American journal of public health*, 97 (Supplement_1), S32–S37.
- WHO. (2020). *Coronavirus disease 2019 (covid-19) situation report – 73*. Zugriff am 2020-07-04 auf https://www.who.int/docs/default-source/coronavirus/situation-reports/20200402-sitrep-73-covid-19.pdf?sfvrsn=5ae25bc7_4
- World Health Organisation. (2020a). *Qa on coronaviruses (covid-19)*. Zugriff am 2020-05-01 auf <https://www.who.int/news-room/detail/q-a-detail/q-a-coronaviruses>

- World Health Organisation. (2020b). *Who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020*. Zugriff am 2020-05-01 auf <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Yuan, J., Yun, H., Lan, W., Wang, W., Sullivan, S. G., Jia, S. & Bittles, A. H. (2006). A climatologic investigation of the sars-cov outbreak in beijing, china. *American journal of infection control*, 34 (4), 234–236.
- Zürich, U. (2018). *Korrelation nach bravais-peerson*. Zugriff am 2020-07-04 auf https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/korrelation.html

A. Erklärung zur Seminararbeit

Hiermit erkläre ich, dass ich die vorliegende Hausarbeit im Fach „Aktuelles aus der Forschung“ zum Thema

Coronavirus und das Wetter in Deutschland - *Auswirkungen wetterbedingter Faktoren auf die Infektionszahlen von COVID-19 anhand einer datengetriebenen Analyse*

selbstständig und ohne Verwendung anderer als der angegebenen Hilfsmittel angefertigt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht.

Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen anderer Prüfungen noch nicht vorgelegt worden.

Ort, Datum

Unterschrift