

Technischer Leitfaden Teil 4 zur SARS-CoV-2 Abwassersurveillance - Datenverarbeitung

Der vorliegende Technische Leitfaden ist ein Teil von vier Arbeitsdokumenten. Die Arbeitsdokumente stehen fachlich – technisch in engem Zusammenhang und sind gemeinsam zu betrachten:

- Technischer Leitfaden Teil 1 zur SARS-CoV-2 Abwassersurveillance
– Probenahme von Abwasser
- Technischer Leitfaden Teil 2 zur SARS-CoV-2 Abwassersurveillance
– Molekularbiologische Analytik
- Technischer Leitfaden Teil 3 zur SARS-CoV-2 Abwassersurveillance
– Probenlogistik und Datenübermittlung
- Technischer Leitfaden Teil 4 zur SARS-CoV-2 Abwassersurveillance
- Datenverarbeitung

Dieser Leitfaden vermittelt die Grundlagen zur Datenqualitätsprüfung, der Datenberechnung sowie der statistischen Analyse im Rahmen der SARS-CoV-2 Abwassersurveillance. Er umfasst folgende Kapitel:

- 1. Verrechnung der PCR Daten**
- 2. Normalisierung**
- 3. Statistische Analyse**
- 4. Software**

Der Rohabwasserzufluss von Kläranlagen unterliegt tageszeitlichen, wochentagsbedingten sowie auch jahreszeitlichen Schwankungen. Hinzu kommen Änderungen des Volumenstromzuflusses durch kontinuierliche oder diskontinuierliche Indirekteinleiter, Fremd- und Niederschlagswasser. Diese Veränderungen im Volumenstrom haben eine Verdünnung der jeweiligen Untersuchungsparameter zur Folge – so auch die Gehalte von SARS-CoV-2-Genfragmenten.

Es bedarf deshalb der Normalisierung der Rohdaten, bevor sie für statistische Analysen herangezogen werden. Aktuell fließen in die Normalisierung ein gemittelter PCR-Messwert des Rohabwassers pro Standort und Probenahmezeitpunkt (Datum des Beginns der Probenahme) sowie Informationen zu Rohabwasservolumenströmen (Durchflussdaten) der Kläranlage ein. Alternative Normalisierungsparameter (u.a. fäkale Bezugsviren, chemische Humanmarker) sind auch möglich, werden aber aktuell nicht in der Ergebnisdarstellung angewendet.

Gemessene Viruslasten im Abwasser weisen in der Regel eine hohe Variation über die Zeit auf, sodass viele Darstellungsformen und statistische Analysen nicht auf den gemessenen (evtl. normalisierten) Messwerten basieren, sondern auf geglätteten Messwerten. Für die Glättung der Messwerte existiert eine Vielzahl an statistischen Verfahren.

1. Verrechnung der PCR Daten

Es werden in der Verarbeitung der Daten nur Proben berücksichtigt, die auf mindestens zwei unterschiedliche der möglichen sechs Gensequenzen (N1, N2, N3, E, ORF und RdRp) analysiert wurden, um fälschliche Positivbefunde auszuschließen. Liegen nicht zwei Messwerte, d.h. Werte für mindestens zwei Gensequenzen vor, wird die Probe generell aus der weiteren Auswertung ausgeschlossen und als „nicht bestimmt“ gekennzeichnet.

Die Bestimmungsgrenzen der jeweils angewendeten PCR Workflow / PCR-Messung werden bei der Datenübermittlung für die jeweiligen Genesequenzen abgefragt.

Liegen die Messwerte oberhalb der Bestimmungsgrenze (BG), engl. Limit of Quantification (LOQ), erfolgt die weitere Auswertung wie unten weiter beschrieben.

Liegt einer der Messwerte / liegen die Messwerte unterhalb der Bestimmungsgrenze, wird der Wert / werden die Werte durch $0,5 * BG$ ersetzt.

Ist der Wert für die BG nicht angegeben, wird für die Normalisierung der Wert 4.000 Genkopien/l angenommen, unabhängig von der eingesetzten Aufbereitungs- oder PCR-Methode. Mittel- und längerfristig sollen hier jedoch eigene Werte der Labore genutzt werden.

Aus (mindestens) zwei vorliegenden Werten für o.g. Gensequenzen wird aktuell ein geometrisches Mittel errechnet:

$$Gene\ gemittelt = \sqrt[n]{x_1 * x_2 * \dots * x_n}$$

mit x_1, x_2, \dots, x_n : Genkopien/l je Genkonzentration

2. Normalisierung der PCR Daten

Für die Normalisierung ist der Trockenwetterzufluss der Kläranlage als Referenz zu bestimmen. Der Trockenwetterzufluss ist das Wasservolumen, welches weder durch Niederschlagsereignisse noch Tauwetter beeinflusst ist.

Bei der aktuell umgesetzten und anwendbaren Normalisierungsmethode wird für die Normalisierung der PCR-Daten das Verhältnis aus Volumenstrom im Probenahmezeitraum ($Q_{KA, \text{aktuell}}$) und dem Trockenwetterzufluss gebildet und anschließend mit dem geometrischen Mittel der der vorliegenden PCR-Messwerte (*Gene gemittelt*) multipliziert. Zur Bestimmung des Trockenwetterzuflusses wird der Median aller bisher übermittelten Werte des Volumenstrom der Kläranlage ($Q_{KA, \text{median}}$) verwendet.

Auf Grundlage der beschriebenen Annahmen wird der normalisierte Wert wie folgt berechnet:

$$Gene\ normalisiert = \frac{Q_{KA, \text{aktuell}}}{Q_{KA, \text{median}}} * Gene\ gemittelt$$

mit $Q_{KA, \text{aktuell}}$: Volumenstrom der Kläranlage im Probenahmezeitraum;

$Q_{KA, \text{median}}$: Median des Volumenstrom der Kläranlage

Der Volumenstrom kann auch als der mittlere Durchfluss der Kläranlage im Probenahmezeitraum genutzt werden.

Die auf die beschriebene Weise normalisierten PCR Daten fließen in die statistischen Analysen ein.

3. Statistische Analysen

Für die SARS-CoV-2 Abwassersurveillance werden für jeden Standort geglättete Werte für eine Ausgleichskurve mittels einer lokal gewichteten Regression (LOESS-Methode, von engl. *Locally Estimated Scatterplot Smoothing*) berechnet. Diese Berechnung erfolgt auf Basis der logarithmierten (und normalisierten, s.o.) Viruslasten im Abwasser.

Das LOESS-Verfahren verwendet dabei eine lokal gewichtete Regressionsfunktion, sodass in die Vorhersage einer Viruslast auf der LOESS-Kurve ein Fenster mit einer bestimmten Anzahl an beobachteten Viruslasten um den Messwert herum einbezogen wird (z.B. die 15 zeitnächsten Messwerte), wobei der Einfluss mit der Entfernung zum Zeitpunkt der vorherzusagenden Viruslast abnimmt. Der Anteil an einzubeziehenden Viruslasten wird für jeden Standort mit Hilfe der generalisierten Kreuzvalidierungsmethode festgelegt, welches die Vorhersagegüte der Kurve optimiert. Die Idee der Kreuzvalidierungsmethode ist es, eine beobachtete Viruslast aus den Daten zu entfernen, diese auf Basis einer Anzahl an anderen beobachteten Viruslasten um den Messwert herum mit dem LOESS-Verfahren vorherzusagen und die resultierende Abweichung zwischen beobachteter und vorhergesagter Viruslast zu berechnen. Dieses Verfahren wird für alle Messzeitpunkte durchgeführt und die berechneten Abweichungen summiert. Die optimale Anzahl an einzubeziehenden Viruslasten nach dem Kreuzvalidierungskriterium ist die, die diese Summe minimiert.


Es ergibt sich für jeden Standort eine glatte Kurve und für jeden Zeitpunkt (auch zwischen den Messzeitpunkten) eine vorhergesagte Viruslast, die dann auf die Originalskala zurücktransformiert wird. Die prozentuale Veränderung zwischen der vorhergesagten Viruslast am Mittwoch einer Woche und der für den vorherigen Mittwoch vorhergesagten Viruslast wird berechnet und es kann dann angegeben werden, an wie vielen Standorten sich z.B. in diesen sieben Tagen ein steigende Viruslast (Trendkategorie „ansteigend“, definiert als Änderung > 15%), eine fallende Viruslast (Trendkategorie „fallend“, Änderung < -15%) oder keine relevante Änderung (Trendkategorie „unverändert“, Änderung zwischen -15% und 15%) beobachten lässt. Anders definierte Trendkategorien sind ebenfalls möglich.

Um neben den einzelnen Standorten einen deutschlandweiten Verlauf der SARS-CoV-2-Viruslast im Abwasser zu erhalten, werden die Zeitreihen der Standorte aggregiert. Dazu werden zunächst für die logarithmierten Messwerte jedes Standorts wöchentliche Mittelwerte gebildet. Die resultierenden Wochenwerte werden dann für jede Woche über alle Standorte gemittelt, dabei wird nach der an den jeweiligen Standort angeschlossenen Einwohnerzahl gewichtet. Die entstehende Zeitreihe wird ebenfalls mit einer LOESS-Regression, wie oben beschrieben, geglättet.

4. Software

Die Verarbeitung und Normalisierung der PCR Daten werden in einer browserbasierten Webanwendung durchgeführt: Pathogene im Abwasser (PiA; <https://app.pia-monitor.de>). Die Anwendung umfasst auch eine automatisierte Qualitätskontrolle der zu importierenden Daten. Die Login-Daten für die Anwendung, eine Beschreibung der Anwendung ist gesondert anzufordern.

Das für die Trendberechnung verwendete LOESS-Verfahren ist in Statistiksoftwarepaketen wie R implementiert. Wird eine feste Fensterbreite zur Schätzung (ohne Kreuzvalidierungskriterium) gewählt, lässt sich eine Implementierung in Excel oder Ähnlichem auch vergleichsweise einfach selbst umsetzen.

Kontakt	Umweltbundesamt, sarscov2@uba.de Robert Koch-Institut, abwassersurveillance@rki.de	
Finanzierung	Das BMG fördert das Abwassermonitoring bis Ende 2024 im Rahmen des Vorhabens "Abwassermonitoring für die epidemiologische Lagebewertung (AMELAG)"	
	Der technische Leitfaden bezieht Informationen/ Erkenntnisse ein, die im Rahmen des Projekts ESI-CorA erarbeitet wurden. ESI-CorA wurde von der Europäischen Kommission im Rahmen des Soforthilfeeinstruments (Emergency Support Instrument – ESI) gefördert (No 060701/2021/864650/SUB/ENV.C2).	