

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

Dieser Leitfaden vermittelt die Grundlagen zur Datenqualitätsprüfung, der Datenberechnung sowie der statistischen Analysen im Rahmen der Abwassersurveillance.

Der vorliegende Leitfaden beschreibt folgende Bereiche:

- 1. Einleitung**
- 2. Verrechnung der PCR Daten**
- 3. Normalisierung**
- 4. Statistische Analysen**
- 5. Software**

1. Einleitung

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

Eine Normalisierung kann gegebenenfalls die Schwankungen in den Rohdaten ausgleichen. Für SARS-CoV-2 wird aktuell der gemittelte PCR-Messwert der Rohabwasserprobe mit dem mittleren Volumenstrom (bezogen auf den Probenahmezeitraum) im Zulauf der Kläranlage normalisiert. Alternative Normalisierungsparameter (u.a. fäkale Bezugsviren, chemische Humanmarker) sind auch möglich, werden aber aktuell nicht in der Ergebnisdarstellung angewendet. Weitere Erreger, wie z.B. Influenzaviren, werden aktuell nicht normalisiert, da sich keine verbesserte Datenqualität durch Normalisierung feststellen ließ.

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

2. Verrechnung der PCR Daten

Für SARS-CoV-2 werden in der Verarbeitung der Daten nur Proben berücksichtigt, die auf mindestens zwei unterschiedliche Genfragmente analysiert wurden, um fälschliche Positivbefunde auszuschließen. Liegen nicht zwei Messwerte, d.h. Werte für mindestens zwei Genfragmente vor, wird die Probe generell aus der weiteren Auswertung ausgeschlossen und als „nicht bestimmt“ gekennzeichnet. Für weitere Erreger werden die Voraussetzungen separat festgelegt.

Die Bestimmungsgrenzen der jeweils angewendeten Methoden werden bei der Datenübermittlung für die jeweiligen Genfragmente abgefragt.

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

Liegt ein Messwert unterhalb der BG, wird der Wert durch $0,5 \cdot BG$ ersetzt und so für die weiteren Analysen genutzt.

Ist der Wert für die BG nicht angegeben, wird für die Normalisierung ein allgemeiner Wert angenommen, unabhängig von der eingesetzten Aufbereitungs- oder PCR-Methode (SARS-CoV-2: 4000 Genkopien/L, Influenza-A-Virus: 2800 Genkopien/L, Influenza-B-Virus: 2300 Genkopien/L) Mittel- und längerfristig sollen hier jedoch spezifische Werte der Labore genutzt werden.

Aus (mindestens) zwei vorliegenden Werten für Genfragmente wird aktuell ein geometrisches Mittel errechnet:

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

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

3. Normalisierung der PCR Daten

Für die Normalisierung der SARS-CoV-2-Daten wird der Trockenwetterzufluss der Kläranlage als Referenz verwendet. Der Trockenwetterzufluss ist der Volumenstrom, welcher weder durch Niederschlagsereignisse noch Tauwetter beeinflusst ist.

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

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

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

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

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

Die auf die beschriebene Weise normalisierten PCR Daten fließen in die statistischen Analysen ein. Die Rohdaten weiterer Viren, wie z.B. Influenzaviren, werden aktuell nicht normalisiert, da sich keine verbesserte Datenqualität durch die Normalisierung feststellen ließ.

4. Statistische Analysen

Für einen Erreger in der 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 (auch Bandbreite genannt) wird für jeden Standort mit Hilfe des Akaike-Kriteriums mit Korrekturfaktor (AICc) so festgelegt, dass die Vorhersagegüte der Kurve optimiert wird. Die Idee der Auswahl der Bandbreite anhand des AICc ist, eine Balance zwischen einer hohen Anpassung der LOESS-Kurve an die vorliegenden Daten und einer geringen Variabilität der LOESS-Kurve, also einer möglichst glatten Kurve, zu finden.

Daraus ergibt sich für jeden Standort eine glatte Kurve und für jeden Zeitpunkt (auch zwischen den Messzeitpunkten) eine vorhergesagte Viruslast. Punktweise Konfidenzbänder werden dann mittels der t-Verteilung berechnet.

Um neben den einzelnen Standorten einen deutschlandweiten Verlauf der SARS-CoV-2-Viruslast im Abwasser zu erhalten, werden die Zeitreihen der Standorte aggregiert. Dazu wird zunächst der Mittelwert über die über eine Woche gemittelten logarithmierten Messwerte der einzelnen Standorte berechnet. Dann wird für jeden Standort für jede Woche die Differenz vom Wochenmittelwert über alle Standorte dieser Woche berechnet. Für jede Standort-Labor-Kombination wird der Mittelwert über diese Differenzen über alle Wochen gebildet um diesen Mittelwert danach von den ursprünglich gemessenen Werten abzuziehen. Dadurch wird für mittlere Unterschiede in den Viruslasten zwischen unterschiedlichen Standort-Labor-Kombinationen, adjustiert. Abschließend wird in jeder Woche, in der für mindestens 10 Standorte Messwerte vorliegen, der Mittelwert über diese adjustierten Werte berechnet. Dabei wird nach den angeschlossenen Einwohnern der Kläranlage gewichtet. Die entstehende Zeitreihe wird ebenfalls mit einer LOESS-Regression, wie oben beschrieben, geglättet.

5. Software

Die Verarbeitung der PCR Daten sowie deren Normalisierung erfolgt in einer Browser-basierten Webanwendung: 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 und 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. Der in AMELAG verwendete Software-Code ist unter https://github.com/robert-koch-institut/Abwassersurveillance_AMELAG/ einsehbar.

Kontakt	Umweltbundesamt, abwassersurveillance@uba.de Robert Koch-Institut, abwassersurveillance@rki.de	
Finanzierung	Das BMG fördert das Abwassermonitoring bis Ende 2025 im Rahmen des Vorhabens "Abwassermonitoring für die epidemiologische Lagebewertung (AMELAG)".	
Weitergehende Literatur	Marquar N, Pütz P, Buchholz U, Exner T, Fretschner T, Greiner T, Helmrich M, Lukas M, Marty M, Obermaier N, Saravia Arzabe C, Schattschneider A, Schneider B, Selinka H-C, Ullrich A, Walter B, Braun U., Schumacher J (2024). SARS-CoV-2-Abwassersurveillance in Deutschland im Rahmen des Projekts AMELAG. <i>Epidemiologisches Bulletin</i> 34:16-26. Schattschneider A, Greiner T, Beyer S, Hans J, Correa Martinez C, Eckmanns T, Diercke M, Schumacher J (2024). Abwasser enthält Informationen für die öffentliche Gesundheit: Mögliche Anwendungen für eine Abwassersurveillance. <i>Epidemiologisches Bulletin</i> 34:3-15.	