

Entwicklung eines datengetriebenen Modells zur Überwachung und Optimierung der Aufbereitungskette zur Wasserwiederverwendung

Wolfgang Seis¹, Michael Stapf¹, Stefan Dück², Susanne Brants², Jens Gebhard³, Achim Ried³, Ulf Mieke¹

¹ Kompetenzzentrum Wasser Berlin, Grunewaldstr. 61-62, 10825 Berlin

² Stadtwerke Bad Oeynhausen, Weserstraße 23, 32547 Bad Oeynhausen

³ Xylem Services GmbH, Boschstraße 4, 32051 Herford

KEYWORDS: KLÄRANLAGE; RANDOM FOREST; WASSERWIEDERVERWENDUNG

1 Einleitung

Die Wiederverwendung von kommunalem Abwasser kann einen wichtigen Baustein bei der effizienteren Nutzung von knapper werdenden Wasserressourcen darstellen. Dies spiegelt sich auch in der aktuellen europäischen Gesetzgebung (EU 2020, EU 2024) sowie der nationalen Wasserstrategie (BMUV 2023) wider. Empfehlungen zu technischen Aspekten der Wasserwiederverwertung werden aktuell durch die DWA erarbeitet (Merkblattreihe 1200). Damit das Wasser genutzt werden kann, ist für die Routineüberwachung nachzuweisen, dass die Konzentration des Parameters *E. coli* im Ablauf der Aufbereitungskette unterhalb entsprechender Grenzwerte liegt. Die UV-Behandlung ist ein ausgereiftes und weit verbreitetes Desinfektionsverfahren. Da in der Regel jedoch keine Echtzeitinformationen zur hygienischen Belastung im Zulauf der Desinfektionsstufe vorliegen werden, um existierende Grenzwerte mit ausreichender Sicherheit einzuhalten, kontinuierlich hohe UV-Dosen angesetzt um auch Spitzenbelastungen ausreichend zu behandeln. Diese treten jedoch nur selten auf. Dies führt potenziell zu einem höheren Energieeintrag als für die Einhaltung gesetzlicher Grenzwerte notwendig wäre.

Eine Fragestellung im BMBF geförderten Forschungsprojekts DigiWave ist es daher zu prüfen inwiefern durch ein datengetriebenes Modell

- 1.) die Dynamik hygienischer Belastungen im Zulauf der UV-Anlage durch existierende Prozessdaten der Kläranlage vorhergesagt und
- 2.) durch eine modellbasierte Regelung der UV-Anlage deren Energieeinsatz bei gleichzeitiger sicherer Einhaltung der Grenzwerte optimiert werden kann.

2 Material und Methoden

2.1 Standort/Datenerhebung:

Die Untersuchung wurden an der Kläranlage Bad Oeynhausen durchgeführt. Die Anlage verfügt bereits über eine weitergehende Abwasserreinigung mit einem Sandfilter und einem nachgeschalteten GAK-Filter zur gezielten

Spurenstoffelimination. Für das Projekt wurde ein Teil des GAK-Ablaufs mit einer regelbaren UV-Desinfektionsanlage (LBX90e, Niederdruckstrahler, 60 m³/h, Xylem) behandelt. Für die Generierung eines für den Modellaufbau ausreichend großen Datensatzes wurde ein Schnellmessgerät für *E. coli* am Ablauf des Sandfilters (aquaBio, ADASA) installiert und ab Mai 2024 kontinuierlich betrieben (n > 500 Messwerte). Um die Vergleichbarkeit zwischen den Daten des Schnellmessgeräts und genormten Labormethoden zu prüfen, wurden regelmäßig parallele Stichproben genommen und nach ISO 9308-2 analysiert. Weitere Proben wurden im Zulauf Kläranlage, im Ablauf Nachklärung, im Ablauf Sand- und GAK-Filter sowie im Ablauf der UV-Anlage genommen und auf mikrobielle Indikatorparameter untersucht (*E. coli*, Enterokokken, *Clostridium perfringens*, somatische Coliphagen). Abbildung 1 zeigt einen schematischen Aufbau der Pilotanlage sowie der Probenahmestellen.

2.2 Aufbau der Modellkette

Modellbaustein 1

Modellbaustein 1 dient der Vorhersage der *E. coli* Konzentration im Zulauf der UV-Anlage. Für den Modellaufbau wurden hydraulische (Q_{ZulaufKA} , Q_{AblaufKA}), physikalisch-chemische Parameter (T, Trübung_{ZulaufFilter}, Trübung_{AblaufKA}, Ammonium_{AblaufKA}, Leitfähigkeit_{AblaufKA}, SAK_{AblaufSandfilter}) der Kläranlage in einer Auflösung von 15 Minuten gesammelt und zu gleitenden 3h-Mittelwerten aggregiert. Diese dienten als Modelleingangsparameter. Für die Erstellung des Modells wurde das Python Modul *scikit-learn* mit einem Random Forest Regressionsmodell verwendet, da dieses oft eine hohe Vorhersagegenauigkeit bei vergleichsweise guter Erklärbarkeit liefert. Einfachere Modellierungsansätze, wie lineare Regression wurden getestet jedoch aufgrund einer im vorliegenden Fall schlechteren Vorhersagegenauigkeit nicht weiterverfolgt.

Für den Modellaufbau wurden die Daten zunächst in Trainings- und Testdaten in einem Verhältnis von 80% zu 20% aufgeteilt. Um die Hyperparameter des Random Forest Models zu optimieren, wurde ein *Grid Search Algorithmus* auf die Trainingsdaten mittels fünffacher Kreuzvalidierung angewendet und anschließend gegen die 20% der unabhängigen Testdaten validiert. Als Performanceindikator diente der RMSE (*root mean squared error*) sowie die erklärbare Varianz R².

Modellbaustein 2

Modellbaustein 2 dient der Vorhersage/Empfehlung der benötigten UV Dosis zur Einhaltung der geforderten Vorgaben im Ablauf der UV Anlage. Auf diese Weise soll die benötigte Energieeintrag optimiert werden. Im vorliegenden Fall wurde der Grenzwert für Wasser der Güteklasse A (<10 KBE/100mL) herangezogen. Als Desinfektionsziel wurde die Hälfte des Grenzwerts angesetzt (5 KBE/100mL).

Für den Modellaufbau war geplant am Standort Bad Oeynhausen separate Dosis-Wirkungstest durchzuführen. Da jedoch die untere Regelungsgrenze

des LBX90e bei 50% Lampenleistung und maximalen Durchfluss bei ca. 800 J/m² (PSS Dosis) liegt und selbst bei der niedrigsten möglichen UV-Dosis die Nachweisgrenze für *E. coli* stets unterschritten wurde, konnte keine UV-Dosis-Wirkungsbeziehung für die verwendete UV-Anlage im Feld ermittelt werden. Aus diesem Grund wurde zum Aufbau des Modells auf Daten von Dosis-Wirkungsversuchen zurückgegriffen, die im Projekt FlexTreat am Standort Braunschweig durchgeführt wurden. Die Daten wurden am Standort Bay Oeynhausen nochmals geprüft und bestätigt. Der funktionale Zusammenhang wurde mit Hilfe einer einfachen linearen Regression zwischen der Logreduktion für *E. coli* und den log-transformierten UV-Dosiswerten ermittelt.

2.3 Modellbetrieb:

Seit Juli 2025 befindet sich die aufgebaute Modellkette im Demonstrationsbetrieb. Mit Hilfe der aufgebauten Modellkette werden alle 30 Minuten Empfehlung zur minimal benötigten UV-Dosis abgeleitet. Benötigte Prozessdaten der Kläranlage werden über eine Sammel-SPS bereitgestellt und über eine OPC-UA Schnittstelle abgerufen. Die Datenaufbereitung erfolgt analog zur Datenvorbereitung im Modellaufbau. Darüber hinaus wird geprüft ob sich neue, auflaufenden Daten noch innerhalb des Kalibrationsbereichs des Modells bewegen oder ob es sich aus Sicht des Modells um eine Anomalie handelt. Im letzteren Fall wird auf eine Vorhersage verzichtet und im Sinne eines vorsorglichen Risikomanagements „empfohlen“ auf eine klassische, fluenzbasierte Anlagensteuerung umzuschalten.

2.4 Modellvalidierung:

Ziel der Modellvalidierung ist es die Vorhersagegenauigkeit der aufgebauten Modellkette zu testen. Da die vom Modell empfohlene UV-Dosis in der Regel unterhalb der unteren Regelungsgrenze des LBX90e liegt wurde eine zweite, kleinere UV Anlage (Niederdruckstrahler mit 50 W Lampenleistung, Spektron 2e, Xylem) installiert. Über eine manuelle Variation der UV Dosis im Bereich 200 – 600 J/m² (PSS) und Vergleichsproben im Ablauf der UV-Anlage wird die Vorhersagegenauigkeit des Modells bestimmt.

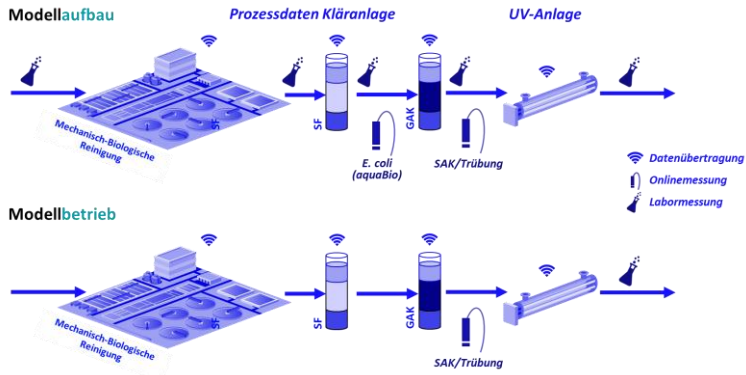


Abbildung 1: Schematische Übersicht der Online- und Labormessungen während des Modellaufbaus bzw. -betriebs.

3 Ergebnisse und Diskussion

3.1 Vergleich Schnellmessung zu Labordaten

Abbildung 2 zeigt den Vergleich zwischen den *E. coli* Daten die mit Hilfe des Schnellmessgeräts gemessen wurden zu den Ergebnissen mittels standardisierter Labormethoden. Der Vergleich zeigt, dass die Daten des Schnellmessgeräts sehr gut mit den Labormessungen übereinstimmen. Demzufolge kann auf eine weitere Korrektur der Daten des Schnellmessgerät verzichtet und die Daten direkt für den Modellaufbau genutzt werden. Da zwischen der Messstelle nach der Filtration und der UV Anlage noch ein GAK-Filter zwischengeschaltet ist, wurden Vergleichsproben nach dem GAK Filter genommen und im Labor analysiert. Die Daten (nicht dargestellt) zeigen keine weitere relevante Reduktion der *E. coli* Konzentration durch den GAK-Filter. Daher können die Modellprognosen im Ablauf des Sandfilters als gute Schätzung für die Zulaufkonzentration der beiden UV-Anlagen angesehen werden. Die Schwankungsbreite der *E. coli* Konzentration beträgt etwa 3 Größenordnungen (10^2 - $10^{5.5}$ MPN/100 ml). Durch die höhere zeitliche Auflösung des Schnellmessung lassen sich dynamische Effekte erfassen, welche durch das Modell vorhergesagt werden können.

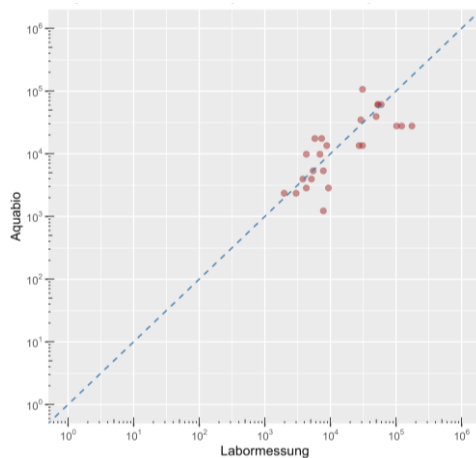


Abbildung 2: Vergleich zwischen gemessener *E. coli* Konzentration pro 100mL im Labor und mit Hilfe des aquaBio Schnellmessgeräts.

3.2 Modelaufbau

Die Ergebnisse des Modelltrainings des Modellbausteins 1 sind in Abbildung 3 dargestellt. Erwartungsgemäß lassen sich die Trainingsdaten sehr gut über die Prozessdaten der Kläranlage bestimmen ($R^2 = 0,9$). Dies alleine lässt jedoch noch keine Aussage über die Vorhersagegenauigkeit zu. Letztere wird über den Vergleich zu den unabhängigen Testdaten bestimmt. Im Badegewässerbereich wird oft ein R^2 – Wert von 0,6 als hinreichend gut für die Vorhersage mikrobieller Indikatoren angesehen (US-EPA 2012, WHO 2018). Der hier für die Testdaten ermittelte Wert von rund 0,5 liegt dementsprechend etwas darunter. Es ist jedoch zu berücksichtigen, dass die R^2 - Werte abhängig von der beobachteten Variabilität der Daten sind. Ein Fehlen von Spitzenbelastungen reduziert den ermittelten R^2 -Wert. Dies ist relevant, da wie in Abbildung 3 zu sehen, Spitzenwerte im Bereich zwischen größer 10^5 - $10^{5,5}$ fehlen. Von daher kann potenziell ein höherer R^2 -Wert noch erreicht werden. Die Vorhersageunsicherheit wird entsprechend berücksichtigt. Die wichtigsten Vorhersageparameter des verwendeten Random-Forest-Modells (Feature Importance) sind:

- Trübung (Ablauf Kläranlage)
- Temperatur
- SAK₂₅₄ (Ablauf Filtration)
- Zulaufvolumenstrom zur Kläranlage (3h Mittelwert)
- Leitfähigkeit (3 h Mittelwert)

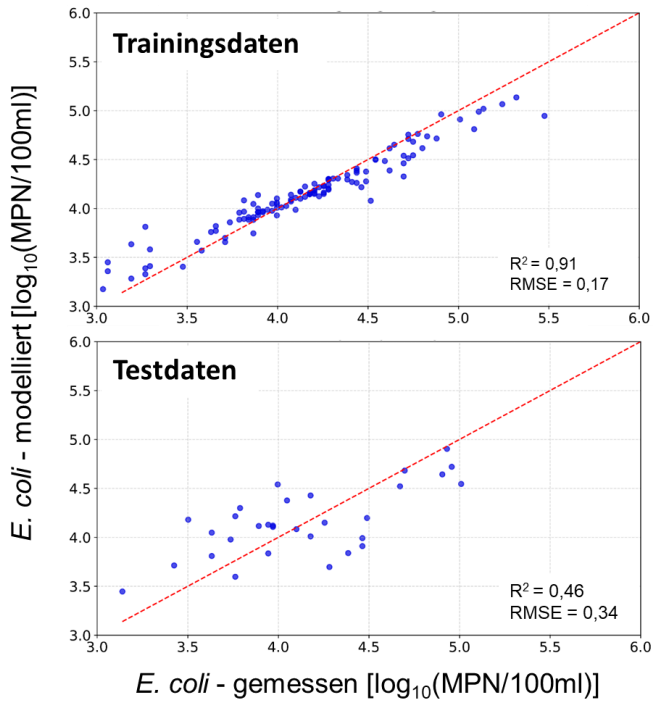


Abbildung 3: Gegenüberstellung der gemessenen und modellierten *E. coli* Konzentrationen in den Trainings- und Testdatenset.

Die Kalibrierung des Modellbaustein 2 sowie die abgeleitete Berechnungsformel für die empfohlene PSS-Dosis (PSS) in Abhängigkeit der Zulaufkonzentration ($E. coli_{in}$) und des Desinfektionsziel ($E. coli_{ziel}$) ist in Abbildung 4 dargestellt. Erwartungsgemäß zeigt sich ein klarer Zusammenhang ($R^2 = 0.95$).

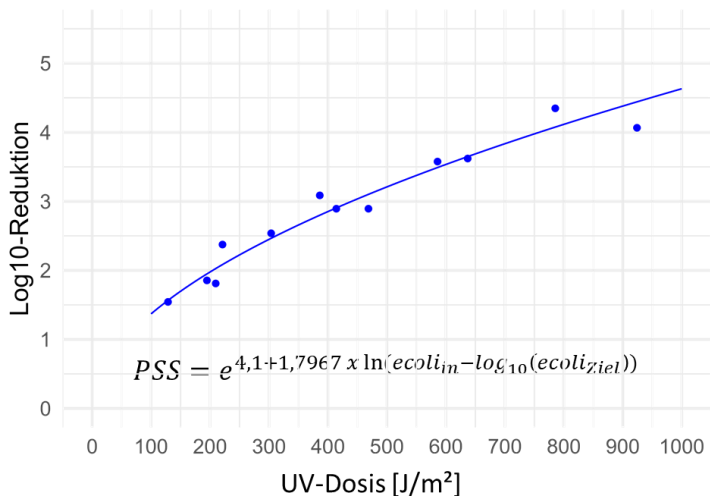


Abbildung 4: Zusammenhang zwischen der eingetragenen UV Dosis und der erzielten Logreduktion.

3.3 Modellvalidierung

Die Modellvalidierung im Demonstrationsbetrieb wird bis Oktober 2025 durchgeführt. Ergebnisse stehen zum Zeitpunkt des Verfassens dieses Abstracts noch aus.

3.4 Abschätzung des Energieeinsparpotenzials

Basierend auf der Prognose der *E. coli* Konzentration im Zulauf der UV-Anlage und einer Dosis-Wirkungsbeziehung kann die notwendige UV-Dosis in Abhängigkeit des jeweiligen Zielwerts und eines Sicherheitsfaktors abgeleitet werden. Eine erste Abschätzung ergab, eine Variabilität der benötigten UV-Dosis zwischen 450 – 950 J/m². Eine erste Potentialabschätzung zeigte weiter, dass eine Reduzierung der UV-Dosis um 30% in mehr als 25% der Zeit möglich wäre, ohne dass eine Überschreitung der Grenzwerte zu befürchten wäre. Letztere Abschätzung wird im Laufe des Demonstrationsbetriebs weiter quantifiziert.

4 Schlussfolgerungen

- Das eingesetzte Schnellmessgerät für *E. coli* lieferte belastbare, zeitlich hoch aufgelöste Datensätze. Dadurch können informative Datensätze über die Dynamik von *E. coli* im Ablauf kommunaler

Kläranlagen gesammelt und für den Aufbau datengetriebener Modelle genutzt werden.

- Bisherige Modellierungsansätze bewegen sich am unteren Grenzbereich, der für mikrobielle Parameter als hinreichend gut angesehenen Richtwerte für den Modellgüteparameter R^2 . Vorhersageunsicherheiten sind entsprechend zu berücksichtigen.
- Erste Abschätzungen zeigen, dass sich der Energieverbrauch der Desinfektionsstufe durch die Nutzung einer modellbasierten Regelung reduzieren lässt, was eine energieeffizientere Wasserwiederverwendung ermöglicht.

5 Literatur

BMUV (2023). "BMUV_Wasserstrategie." from https://www.bmuv.de/fileadmin/Daten_BMU/Download_PDF/Binnengewasser/BMUV_Wasserstrategie_bf.pdf.

EU (2020). REGULATION (EU) 2020/741 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on minimum requirements for water reuse.

EU (2024). RICHTLINIE (EU) DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. November 2024 über die Behandlung von kommunalem Abwasser. 2024/3019. E. Union.

US-EPA (2012). Recreational Water Quality Criteria; OFFICE OF WATER 820-F-12-058.

WHO (2018). "WHO recommendations on scientific, analytical and epidemiological developments relevant to the parameters for bathing water quality in the Bathing Water Directive (2006/7/EC)."