

Nutzung nicht-parametrischer KI-Modelle zur Vorhersage kurzzeitiger Verschmutzungen an Flussbadegewässern

Wolfgang Seis, Berlin

Kompetenzzentrum Wasser Berlin gGmbH

Einleitung:

Die Gewährleistung der mikrobiellen Sicherheit an Badegewässern, die von kurzfristigen fäkalen Verschmutzungen betroffen sind, bleibt eine bekannte Herausforderung für den Schutz der öffentlichen Gesundheit. Dies liegt vor allem an den langen Bearbeitungszeiten traditioneller Labormethoden zum Nachweis der Fäkalindikatoren (FIB) *E. coli* und intestinalen Enterokokken. Datengetriebene Modelle aus dem Bereich des maschinellen Lernens, die Umweltfaktoren wie Niederschlag und Abfluss nutzen, haben sich als vielversprechende Lösungen für Frühwarnsysteme erwiesen. Ihre Entwicklung und Genauigkeit hängen jedoch von einem informativen Datensatz für das Modelltraining ab. Solche Datensätze sind oft nicht verfügbar (Searcy und Boehm, 2021). Der vorliegende Betrag zeigt wie durch die Nutzung probabilistischer (nicht-parametrischer) Bayes'scher Modellierungsansätze dem Problem nicht-informativer Datensätze für den Anwendungsfall der Vorhersage der Badegewässerqualität begegnet werden kann. Darüber hinaus wird auf die Vorhersage kurzfristiger Verschmutzungen mittels datengetriebener Modelle im Berliner Spreekanal eingegangen.

Methoden:

Die Studie vergleicht fünf Modellierungsansätze zur Vorhersage kurzfristiger Verschmutzungsepisoden an drei Flussbadestellen in Deutschland, für die umfangreiche historische FIB-Datensätze vorliegen (N=291, N=1191, N=251). Zwei Modellierungsexperimente („Längste Trockenperiode“, „Sequentielles Modelltraining“) wurden durchgeführt, um die Fähigkeit jedes Modells zu bewerten, Kontaminationsspitzen im Testdatensatz vorherzusagen, selbst wenn diese Informationen nicht in den Trainingsdaten enthalten waren. Besonderes Augenmerk liegt auf zwei Bayes'schen Modellen, die auf einer inkrementellen Aktualisierung von FIB-Verteilungen für unterschiedliche hydrologische Zustände basieren. Die hydrologischen Zustände wurden entweder manuell definiert oder mithilfe eines nicht-parametrischen Bayes'schen Dirichlet-Prozess-Mixture-Modells zur Identifikation unbekannter hydrologischer Situationen ermittelt. Verglichen wurden diese Ansätze gegen drei Referenzmodelle (Nullmodell, Multivariate Regression, Random Forest).

Ergebnisse:

Durch die Nutzung Bayes'scher Modelle konnte in mehr als 90 % korrekt vor Verschmutzungsereignissen gewarnt werden, selbst wenn die Trainingsdaten keine Information zu Verschmutzungen enthielten. Im Gegensatz dazu waren das Random-Forest und das Nullmodell nicht in der Lage Kontaminationen korrekt vorherzusagen, wenn solche Ereignisse nicht in den Trainingsdaten vorhanden waren. Regressionsbasierte Ansätze erzielten moderate Ergebnisse, zeigten jedoch eine mangelnde Robustheit über verschiedene Standorte und Szenarien hinweg.

Diskussion:

Bayes'sche Ansätze erwiesen sich als geeignete Methoden, um kurzfristige Verschmutzungen an Badegewässern selbst bei geringer Datenlage sicher vorherzusagen zu können. Durch die Integration Bayes'scher Modelle in bestehende Wasserqualitätsmanagementsysteme können Behörden Kontaminationsereignisse besser vorherzusagen und so den Schutz der öffentlichen Gesundheit durch zeitnahe Kommunikation gewährleisten.

Quelle:

Searcy, R.T. und Boehm, A.B. 2021. *A Day at the Beach: Enabling Coastal Water Quality Prediction with High-Frequency Sampling and Data-Driven Models*. Environmental Science & Technology 55(3), 1908-1918.