

Automatische Detektion von Ausreissern bei der Gewässerüberwachung

Schadstoffe in Flüssen und Seen sind ein grosses Problem der heutigen Gesellschaft. Immer mehr Chemikalien landen in den Gewässern. Um diese zu detektieren, wurde im Rhein bei Basel eine Überwachungsstation eingerichtet. Dort werden in einer Tagesmischprobe eine Vielzahl an Stoffen auf deren Auftreten untersucht. Der riesige multivariate Datensatz macht es schwierig, die Daten zu analysieren und allfällige Ausreisser zu erkennen. Diese können durch aussergewöhnliche Umstände in der Umwelt oder Probleme bei der chemischen Analyse entstehen.

In der vorliegenden Arbeit werden verschiedene Ansätze untersucht, um Tageswerte zu detektieren, welche auffällig erscheinen. Es wird auf etablierte Methoden, wie die Hauptkomponentenanalyse und die Mahalanobis-Distanz, zurückgegriffen. Mitunter ist es notwendig, diese beiden sowohl robust, als auch auf sehr grosse Datenmengen anwenden zu können. Dazu werden neue Ansätze verwendet.

Eine der untersuchten Methoden beruht darauf, den Daten einen Einflusswert zuzuweisen, der auf den relativen Abständen der einzelnen Beobachtungen zum jeweiligen Zentrum der Hauptkomponente basiert. Überschreitet dieser einen vordefinierten Schwellenwert, wird die Beobachtung als Ausreisser gekennzeichnet.

In einer eigenen Erweiterung des Ansatzes wird, anstelle eines fixen vordefinierten Schwellenwerts, mit einem gleitenden Median über die Zeit gearbeitet. Damit kann sichergestellt werden, dass Änderungen in der Probenaufbereitung oder im Rhein keinen zu grossen Einfluss haben.

Eine weitere Methode basiert auf der robusten Berechnung der Mahalanobis-Distanz. Hier wird zur Berechnung der Kovarianzmatrix ein Algorithmus verwendet, der auch mit grossen Datenmengen arbeiten kann.

Die verschiedenen Methoden wurden an einem Testdatensatz mit Tagesmessungen über zwei Jahre untersucht. Dies führte zur Detektion von jeweils rund 20 Ausreissern, wobei vier Beobachtungen bei allen Methoden auffällig waren. Dies deutet darauf hin, dass es sich tatsächlich um Ausreisser und nicht um falsch-positiv markierte Beobachtungen handelt. Die genauen Ursachen für die Auffälligkeiten müssten von den zuständigen Stellen im Rahmen einer Nachbereitung weiter untersucht werden.

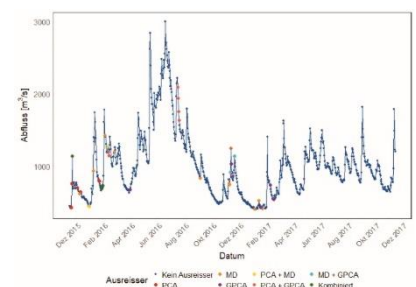


Diplomierende
Janine Kaufmann
Rico Vogel

Dozierende
Manuel Renold
Martin Frey



Die Rheinüberwachungsstelle Weil am Rhein befindet sich bei Rheinkilometer 173.37 in der Nähe der Palmrainbrücke (im Bild). Hier werden Gewässerproben gesammelt und dann zu Analysezwecken ins Labor gebracht.



Drei verschiedene Ansätze wurden verwendet, um Ausreisser in den Gewässerproben zu detektieren. Die Grafik stellt die durch die einzelnen Methoden detektierten Beobachtungen dem Rheinabfluss gegenüber.