

Vorstudie zur Analyse von Pegel- und Abflussmess-Qualität

Technischer Bericht

Bern, 31.07.2026

Im Auftrag des Bundesamtes für Umwelt (BAFU)

Impressum

Auftraggeber: Bundesamt für Umwelt (BAFU), Abt. Hydrologie, CH-3003 Bern

Das BAFU ist ein Amt des Eidg. Departements für Umwelt, Verkehr, Energie und Kommunikation (UVEK).

Auftragnehmer: Universität Bern

Autoren: Benjamin Meyer, Bettina Schaepli

Begleitung BAFU: Martin Barben, Silvia Morf, Caroline Kan

Hinweis: Dieser Bericht wurde im Auftrag des Bundesamtes für Umwelt (BAFU) verfasst. Für den Inhalt ist allein der Auftragnehmer verantwortlich.

Zitiervorschlag: Meyer, B., Schaepli, B., (2026). Vorstudie zur Analyse von Pegel- und Abflussmess-Qualität. Im Auftrag des Bundesamts für Umwelt BAFU.

Inhaltsverzeichnis

1.	Einführung	6
1.1.	Ausgangslage, Problematik	6
1.2.	Zielsetzung	6
2.	Daten	8
3.	Methoden	9
3.1.	Datenvorbereitung	9
3.1.1.	Zeitliche Auflösung	9
3.1.2.	Normierung der Abflussdaten	9
3.1.3.	Synthetische Anomalien	10
3.1.4.	Auswahl Referenzstationen	11
3.2.	Changepoint-Analyse	11
3.3.	Steigungsanalyse	11
3.4.	Modellierung mit maschinellem Lernen (ML Methoden)	12
3.5.	Modellgüteparameter	14
3.6.	Vergleichsmodell zur Einordnung der Modellgüte	15
4.	Ergebnisse	16
4.1.	Changepoint-Analyse	16
4.2.	Steigungsanalyse	16
4.3.	Modellierung mit maschinellem Lernen	18
4.3.1.	XGBoost	19
4.3.2.	Multi Layer Perceptron	21
4.4.	Vergleichsmodell zur Einordnung der Modellgüte	23
5.	Diskussion	26
5.1.	Changepoint-Analyse	26
5.2.	Steigungsanalyse und ML Methoden	26
5.3.	Allgemeine Erkenntnisse	27
5.4.	Einfluss der Einzugsgebietscharakteristiken	28
6.	Fazit	30
7.	Literaturverzeichnis	31
8.	Anhang	34
	A: Verwendete Abflussstationen	34
	B: Getestete Changepoint Algorithmen	38
	C: Zusätzliche Abbildungen	38

Abbildungsverzeichnis

Abbildung 1: Untersuchungsgebiet: Eingezeichnet sind die Grosseinzugsgebiete Aare und Reuss, sowie die verwendeten Messstationen (graue Punkte)	8
Abbildung 2: Autokorrelation (ACF) des täglichen Abflusses der Muota bei Ingenbohl, z-normierter Abfluss; der lag hat als Einheit Tage.	10
Abbildung 3: Wie Abbildung 2 aber für, saisonal z-normiertem Abfluss.....	10
Abbildung 4: Vorgehensweise bei der Steigungsanalyse	12
Abbildung 5: Steigungsanalyse für Aare bei Brugg mit Schwellenwert 0.75	16
Abbildung 6: Violinplots der Sensitivität aller Schwellenwerte der Steigungsanalyse.	17
Abbildung 7: Violinplots der Präzision aller Schwellenwerte der Steigungsanalyse	17
Abbildung 8: Violinplots der F1-Scores aller Schwellenwerte der Steigungsanalyse	18
Abbildung 9: XGBoost-Modell der Aare bei Brugg mit σ -Faktor-Schwellenwert 1.25.....	19
Abbildung 10: Violinplots der Sensitivität aller σ -Faktor-Schwellenwerte der XGBoost Methode	19
Abbildung 11: Violinplots der Präzision aller σ -Faktor-Schwellenwerte der XGBoost Methode.....	20
Abbildung 12: Violinplots der F1-Scores aller σ -Faktor-Schwellenwerte der XGBoost Methode	20
Abbildung 13: Violinplots der Sensitivität aller σ -Faktor-Schwellenwerte der MLP-Methode.....	22
Abbildung 14: Violinplots der Präzision aller σ -Faktor-Schwellenwerte der MLP-Methode	22
Abbildung 15: Violinplots der F1-Scores aller σ -Faktor-Schwellenwerte der MLP-Methode	23
Abbildung 16: Einfluss der Einzugsgebietsgrösse auf die Sensitivität der Methode XGBoost.....	28
Abbildung 17: Einfluss des Variationskoeffizienten des Abflusses auf die Sensitivität	29
Abbildung 18: Die Verteilung der Variationskoeffizienten des Abflusses in abhängigkeit der Einzugsgebietsgrösse	29
Abbildung 19: Untersuchungsgebiet mit allen verwendeten Messstationen, die Einzugsgebiete der Reuss und Aare sind eingefärbt.....	36
Abbildung 20: Gegenüberstellung der Sensitivität und Präzision der Steigungsanalyse mit Schwellenwert 0.75	38
Abbildung 21: Gegenüberstellung der Sensitivität und Präzision der XGBoost basierten Methode mit σ -Faktor 1.25	39
Abbildung 22: Gegenüberstellung der Sensitivität und Präzision der MLP basierten Methode mit σ -Faktor 1.25.....	39
Abbildung 23: Beispiel der Changepointanalyse auf Bases der XGBoost Residuen für die Aare bei Brugg kombiniert mit der Steigungsanalyse. Die schwarzen vertikalen Linien markieren die Chachgepoints.	40
Abbildung 24: Anwendung aller Methoden auf ein Einzugsgebiet mit bekannten Problemen. Die Resultate der Steigungsanalyse sind als rote respektive blaue Segmente dargestellt. Die Ergebnisse der XGBoost basierten Methode erscheinen oben als violette Balken. Die Changepoints sind als vertikale schwarze Linien dargestellt.	40

Tabellenverzeichnis

Tabelle 1: Spezifikationen der synthetischen Anomalien.....	10
Tabelle 2: Hyperparameter und Optimierungsbereich der XGBoostmodelle	13
Tabelle 3: Hyperparameter und Optimierungsbereich der MLPmodelle	14
Tabelle 4: Anteile der im Vergleichsmodell zufällig als Anomalie klassifiziert und deren etwaige Entsprechung hinsichtlich Schwellenwerte für die Steigungsanalyse und ML-Methoden	15
Tabelle 5: Zusammenfassung Modellgüteparameter der Steigungsanalyse	18
Tabelle 6: Zusammenfassung Modellgüteparameter der XGBoost-Modelle	21
Tabelle 7: Zusammenfassung Modellgüteparameter der MLP-Modelle	23
Tabelle 8: Modellgüteparameter des Vergleichsmodells ohne Filterung der Anomaliedauer.....	23
Tabelle 9: Modellgüteparameter des Vergleichsmodells mit Filter nach Steigungsanalyse	24
Tabelle 10: Modellgüteparameter des Vergleichsmodells mit Filter nach ML-Methode.....	24
Tabelle 11: Liste der verwendeten Stationen (EZG steht für Einzugsgebiet, MQ für mittleren Abfluss; unterschiedliche Kommastellen in diesen beiden Kolonnen deuten darauf hin, wie genau die Werte bestimmt wurden, die Präzision ist bei kleinen Einzugsgebieten höher. Quelle: https://www.hydrodaten.admin.ch/).	34
Tabelle 12: Liste der Referenzstationen	37
Tabelle 13: Liste der getesteten Changepoint Algorithmen	38

1. Einführung

1.1. Ausgangslage, Problematik

Die Abflussdaten des Basismessnetzes des BAFU unterliegen einer 3-stufigen Datenqualitätskontrolle. Diese Stufen umfassen die Eingangskontrolle, Datenprüfung und -korrektur sowie Jahresabschlussarbeiten (Schmocker-Fackel & Hüsler, 2025). Eine vierte Stufe hinsichtlich hydrologischer Validierung ist vorgesehen, wurde jedoch noch nicht implementiert. Von der Abteilung Hydrologie (BAFU) bereits angedachte Methoden zur hydrologischen Validierung umfassen:

- Ober / Unterlieger- und regionale Vergleiche von Abflussdaten
- Berechnung und Vergleich hydrologischer Kennwerte wie z.B. jährliche Abflusshöhen
- Kontrolle der Wasserbilanz durch Abfluss- und Niederschlagsvergleiche¹
- hydrologische Modellierungen

In der Praxis führen viele Anwenderinnen und Anwender von hydrologischen Daten ihre eigenen Validierungs- oder Plausibilisierungsanalysen durch, z.B. im Rahmen von Hochwasserabschätzungen, Trockenheitsanalysen oder Modellierungsstudien. Solche individuellen Plausibilisierungsanalysen werden auch in Zukunft je nach Anwendungsziel im Rahmen hydrologischer Studien eine wichtige Rolle spielen.

Eine Validierung durch das BAFU wäre jedoch ein wichtiger Schritt zur Vereinheitlichung der Datenauswertung durch Dritte. Eine Möglichkeit wäre, homogenisierte Datenreihen für ausgewählte Stationen zu publizieren, wie dies z.B. Meteoschweiz macht (Begert et al. 2005). Die Publikation kohärenter hydrologischer Datensätze wäre wichtig, da solche Datensätze auch von Forschenden benötigt werden, die nicht über das erforderliche lokale Wissen für eigenständige Plausibilisierungsarbeiten verfügen (Höge et al., 2023).

Es existiert eine Vielzahl von Algorithmen zur Detektion von Changepoints (Aminikhanghahi & Cook, 2017; Niu et al., 2016) und Anomalien (Shaukat et al., 2021; Blázquez-García et al., 2021) in Zeitreihen, darunter auch zahlreiche Ansätze, die maschinelles Lernen (ML Methoden) einsetzen (Choi et al., 2021; Zamanzadeh Darban et al., 2024). Das Anwendungsspektrum solcher Methoden ist breit und reicht von der Medizin über die Überwachung von Industrieanlagen, Cybersicherheit und Energiemanagement bis hin zur Logistik und dem «Internet of Things», das in jüngster Zeit als Treiber neuer Algorithmen gilt (Zamanzadeh Darban et al., 2024). Auffällig ist dabei, dass die Anwendung solcher Methoden zur Datenqualitätskontrolle in der Hydrologie oder Meteorologie bislang vergleichsweise wenig untersucht wurde (Reitan, 2022). Die vorhandenen Studien befassen sich meist entweder mit der Detektion von Ausreissern (Zhao et al., 2018; Piroozmand & Farshi, 2025) oder fokussieren sich auf Fehler spezifischer Sensormodelle (Bell et al., 2017). Die einzige identifizierte Studie, die sich mit der Validierung hydrologischer Zeitreihen über längere Zeiträume befasst, stammt aus dem Jahr 1986 (Bois, 1986). Die vorliegende Vorstudie soll einen ersten Beitrag zur Schliessung dieser Forschungslücke leisten.

1.2. Zielsetzung

Ziel dieser Vorstudie ist es, Methoden zur Identifikation von fehlerhaften Datenabschnitten oder Perioden, die mit dem Rest der Zeitreihe oder ähnlichen Messstationen nicht kompatibel sind, zu erarbeiten. Zudem soll der Frage nachgegangen werden, ob ML – Methoden zur Validierung oder Plausibilisierung von bestehenden Messreihen herangezogen werden können. Daraus wurden die folgenden Forschungsfragen abgeleitet:

¹ Dies ist in der Schweiz schwierig, da es zurzeit keine schweizweiten Datensätze zur Verdunstung gibt.

1. Wie können fehlerhafte oder inkonsistente Datenabschnitte mittels Changepoint-Analysen identifiziert werden?
2. Lassen sich mit regionalen Vergleichen zwischen Messstationen fehlerhafte oder inkonsistente Datenabschnitte ermitteln?
3. Wie können ML-Methoden eingesetzt werden, um fehlerhafte oder inkonsistente Datenabschnitte zu detektieren?
4. Lassen sich die Methoden automatisiert einsetzen?

Die Ergebnisse können in einem späteren Schritt (nicht Teil der Vorstudie) als Teil einer vierten Stufe in der Qualitätskontrolle zu einer nachträglichen Homogenisierung von Daten beitragen. In der hydrologischen Praxis der Schweiz lag der Schwerpunkt der hydrologischen Validierung von Abflussmessungen in der Vergangenheit vor allem auf der Validierung hoher Abflüsse. Zurzeit wird die Analyse und Modellierung von Niedrigwasserständen immer wichtiger. Die Studie geht daher auch auf die Anwendbarkeit bei Niedrigwasser ein.

2. Daten

Diese Studie wurde anhand von Abflussdaten von 41 Messstationen im Einzugsgebiet der Reuss und der Aare für den Zeitraum von 2004 – 2024 durchgeführt (Abb. 1). Die Einzugsgebiete decken eine grosse Bandbreite an Einzugsgebietscharakteristiken ab. Die Grösse der Einzugsgebiete variiert zwischen 0.56 km² und 17'556 km², die mittlere Einzugsgebietshöhe von 581 m ü.M. und 1854 m ü.M. und der Vergletscherungsgrad reicht von 0% bis 6.9%. Alle Daten wurden von der Abteilung Hydrologie (BAFU) zur Verfügung gestellt. Eine vollständige Liste ist im Anhang in Tabelle 11 zu finden.

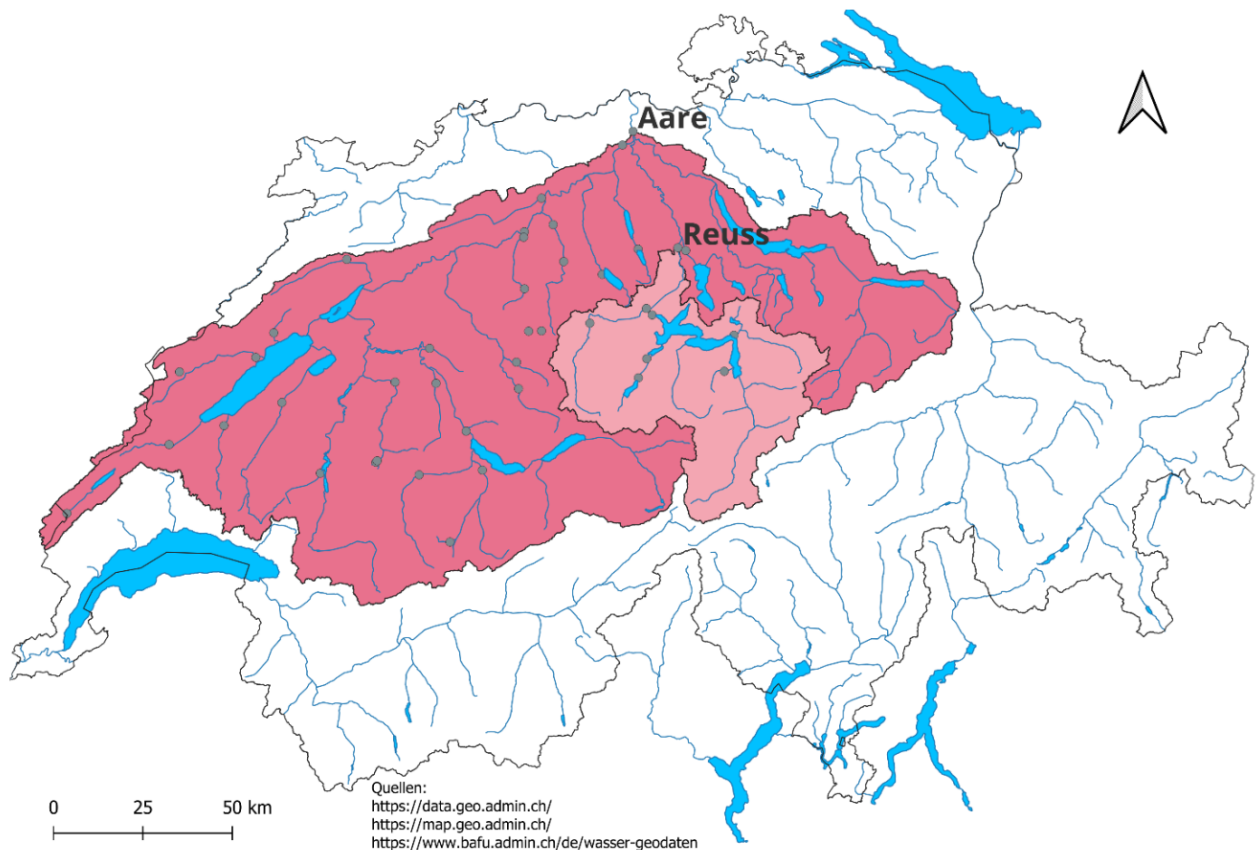


Abbildung 1: Untersuchungsgebiet: Eingezeichnet sind die Grosseinzugsgebiete Aare und Reuss, sowie die verwendeten Messstationen (graue Punkte)

Zusätzlich standen die Schlüsselkurven von neun Einzugsgebieten im gleichen Zeitraum zur Verfügung, um die Grössenordnung von bekannten Veränderungen einschätzen zu können.

3. Methoden

In diesem Kapitel werden die Methoden der Datenvorbereitung sowie der Plausibilitätsanalysen erklärt. Dabei werden vollständigshalber auch alle Methoden erwähnt, die keine befriedigenden Ergebnisse hervorgebracht haben.

3.1. Datenvorbereitung

3.1.1. Zeitliche Auflösung

Die Abflussdaten lagen ursprünglich als Mittelwerte in stündlicher Auflösung bereit. Da für die Analysen von längerfristigen systematischen Fehlern die korrekte Abbildung der Abflussspitzen weniger relevant ist als die korrekte Abbildung des Abflussvolumens, wurden die Stundenmittel zu Tagesmittel aggregiert, um den Rechenaufwand zu minimieren. Eine stichprobenartige Kontrolle zeigte keine Unterschiede zwischen den beiden Zeitlichen Auflösungen.

3.1.2. Normierung der Abflussdaten

Da Faktoren wie die Fläche oder die mittlere Höhe der Einzugsgebiete den Abfluss massgeblich beeinflussen, wird die direkte Vergleichbarkeit der Einzugsgebiete erschwert. Um diesem Problem zu begegnen, wurden alle Zeitreihen z-normiert. Im Laufe der Studie zeigte sich zudem, dass die Unterschiede in der Abfluss-Saisonalität die Analyse derart beeinflussten, dass zusätzlich eine saisonale Normierung notwendig ist. Dabei werden die Daten nicht nur mit dem langjährigen Mittelwert und der Standardabweichung transformiert, sondern mit saisonalen Kennwerten normiert. Dazu werden zunächst die langjährigen Tagesmittel berechnet und anschliessend ein gleitendes Mittel mit einem Zeitfenster von 61 Tagen angewendet, um den Einfluss einzelner Ereignisse zu minimieren. Analog wird ein saisonaler Glättungswert für die Standardabweichung berechnet. Danach wird jeder Wert x_t der Zeitreihe anhand dieser saisonalen Werte kalendertagspezifisch normiert.

$$z_{t,d} = \frac{x_{t,d} - \overline{\mu_d}}{\overline{\sigma_d}}, \quad (1)$$

wobei $x_{t,d}$ der Zeitreihenwert zum Zeitpunkt t am Kalendertag d ist, $\overline{\mu_d}$ das über 61 Tage gemittelte langjährige Tagesmittel am Kalendertag d und $\overline{\sigma_d}$ die über 61 Tage gemittelte langjährige Standardabweichung am Kalendertag d .

In Abbildung 2 und 3 ist die Autokorrelation der z-normierten Zeitreihe sowie der saisonal normierten Zeitreihe abgebildet. Der Vergleich verdeutlicht, dass die saisonale Komponente durch die Normierung erfolgreich reduziert werden konnte.

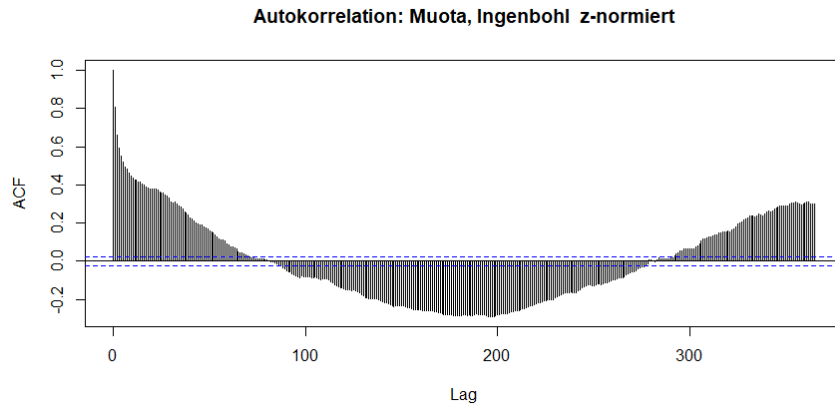


Abbildung 2: Autokorrelation (ACF) des täglichen Abflusses der Muota bei Ingenbohl, z-normierter Abfluss; der lag hat als Einheit Tage.

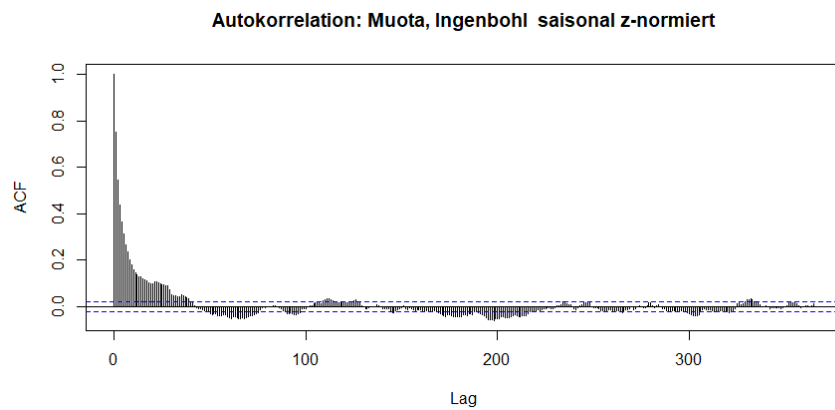


Abbildung 3: Wie Abbildung 2 aber für, saisonal z-normiertem Abfluss

3.1.3. Synthetische Anomalien

Da keinerlei Informationen über bekannte Anomalien vorlagen, wurden synthetische Anomalien in die Zeitreihen eingefügt. Diese sollen vor allem Veränderungen in der Wasserstand-Abflussbeziehung (also in der Schlüsselkurve) simulieren. Das Einfügen von synthetischen Anomalien ermöglicht eine gezielte Bewertung der Algorithmen anhand bekannter Anomalien. Innerhalb eines zufällig ausgewählten Fehlerzeitraums wurde der Abfluss mit einem Faktor multipliziert. Zudem wurde in den ersten und letzten 4 Tagen linear interpoliert, um einen graduellen Übergang zwischen korrigierten und nicht korrigierten Werten zu gewährleisten, da Changepoint-Methoden von abrupten Übergängen möglicherweise profitieren. Die Dauer sowie die Faktoren aller eingefügten synthetischen Anomalien können der Tabelle 1 entnommen werden.

Eine vorgängige Analyse der Schlüsselkurven von neun Einzugsgebieten zeigte, dass Abflussveränderungen im Zusammenhang mit der Schlüsselkurve im Bereich der mittleren Abflüsse im einstelligen Prozentbereich liegen. Im Hochwasser und Niedrigwasserbereich können die Veränderungen grösser sein und in Ausnahmefällen im Bereich Niedrigwasser bis zu 1000% betragen.

Tabelle 1: Spezifikationen der synthetischen Anomalien

Dauer (Tage)	20	20	20	20	20	200	200	200	200	200
Faktor	1.1	0.85	1.2	1.3	1.5	1.1	0.85	1.2	1.3	1.5

3.1.4. Auswahl Referenzstationen

Eine Möglichkeit, die Plausibilität von Abflusszeitreihen zu überprüfen, ist der Vergleich einer Messstation mit Referenzstationen. Die Auswahl der Referenzstationen kann z.B. über die Datenqualität erfolgen, wobei Stationen mit besonders guter Datenqualität als Referenz zur Kontrolle anderer Stationen herangezogen werden können. Eine zweite Möglichkeit ist die Auswahl über die hydrologische Ähnlichkeit der Einzugsgebiete (Fläche, mittlere Höhe, Vergletscherungsgrad, Exposition, Distanz zueinander, usw.). Um a priori Annahmen über hydrologische Ähnlichkeiten zu vermeiden, wurde eine korrelationsbasierte Auswahl gewählt. Für jedes Zeitreihenpaar wurde der Pearson-Korrelationskoeffizient berechnet und die beiden Stationen mit der stärksten Korrelation als Referenzstationen definiert. Zur Validierung dieser Methode wurde zusätzlich eine Auswahl mittels Expertenwissens durchgeführt, bei der für jede Station zwei Referenzstationen mit vergleichbaren hydrologischen Charakteristika ausgewählt wurden. Die Validierung zeigt, dass stets mindestens ein Referenzeinzugsgebiet übereinstimmt. Die Liste der Referenzeinzugsgebiete ist im Anhang in Tabelle 12 zu finden.

3.2. Changepoint-Analyse

Es gibt eine Vielzahl von Algorithmen, die Changepoints in Zeitreihen detektieren. Die getesteten Algorithmen können der Tabelle 13 im Anhang B entnommen werden. Es wurden mehrere Vorgehensweisen getestet, wie Changepoint-Algorithmen angewendet werden können: einerseits anhand der Abflusszeitreihen, der differenzierten Abflusszeitreihen und deren Normierungen, sowie anhand der Differenzzeitreihen zwischen der Abflusszeitreihe einer Station und den Referenzstationen und deren kumulativen Summen. Zudem wurde zur Reduktion der Falschpositiv-Rate (erkannte Changepoints die nicht mit den synthetischen Anomalien zusammenhängen) die Changepoints einer Station mit denen der Referenzstationen verglichen, um nach gleichzeitig auftretenden Changepoints zu filtern. Es stellte sich heraus, dass keiner der getesteten Algorithmen in der Lage scheint, die synthetischen Anomalien zuverlässig zu detektieren (siehe auch Kapitel 5.1).

3.3. Steigungsanalyse

Die Grundidee der Steigungsanalyse ist eine etablierte Methode, um Abflusszeitreihen zweier Stationen miteinander zu vergleichen (Bois, 1986). Dazu wird die kumulative Summe der Differenz der beiden Zeitreihen berechnet. Um unterschiedlich grosse Einzugsgebiete miteinander vergleichen zu können, müssen diese zuerst normiert werden. Besitzen diese Zeitreihen keinen (oder denselben) Trend, sollte die kumulierte Summe der Differenz um Null herum schwanken. Sobald eine Zeitreihe über längere Zeit einen veränderten Abfluss zeigt, weicht diese merklich von Null ab. Je grösser die Steigung dieser Vergleichszeitreihe ist, umso grösser ist die Abweichung vom langjährigen Mittel in einer der beiden Ursprungszeitreihen.

Um festzustellen, ob die Abweichung ihren Ursprung an der Zielstation oder an der Referenzstation hat, sind zwei Referenzstationen erforderlich. Divergieren beide Vergleichszeitreihen zur gleichen Zeit in die gleiche Richtung von Null, kann die Anomalie der Zielstation zugeordnet werden.

Mit dieser Annahme kann nun in beiden Vergleichszeitreihen mithilfe eines Schwellenwerts nach den grössten Steigungen gesucht und diese miteinander verglichen werden. Zusätzlich werden die stärksten Steigungen vor einem Vergleich gefiltert, sodass einzelne lokale Niederschlagsereignisse nicht berücksichtigt werden. Als Filter wurde die Regel definiert, dass nur Steigungen berücksichtigt

werden, die an 12 von 16 aufeinanderfolgenden Tagen den Schwellenwert übersteigen. Dieser Filter bestimmt somit die Mindestdauer der Anomalie. Grössere Werte (längere Mindestdauer) reduzieren zwar die Falschpositivrate, vermindern jedoch die Sensitivität. Als Schwellenwert dient die Überschreitung eines definierten Perzentils der Steigungen. Es wurden sechs Schwellenwerte in Schritten von fünf Prozentpunkten vom 90. bis zum 65. Perzentil getestet. Die gesamte Methodik ist in Abbildung 4 vereinfacht dargestellt.

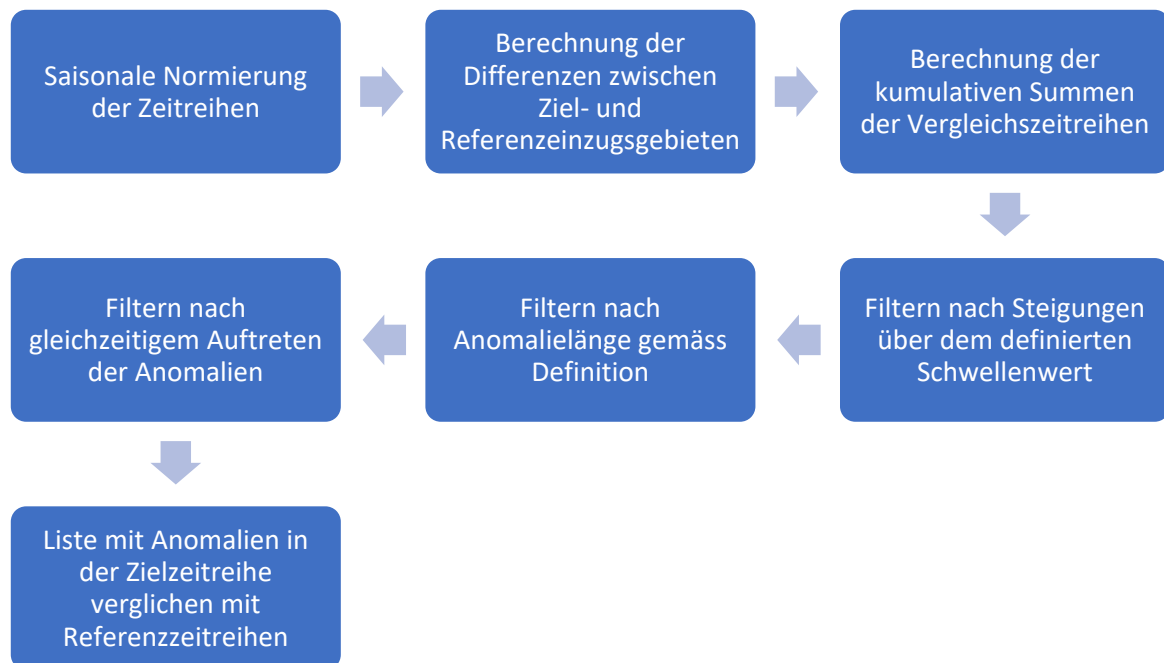


Abbildung 4: Vorgehensweise bei der Steigungsanalyse

3.4. Modellierung mit maschinellem Lernen (ML Methoden)

„Klassische« Algorithmen führen einen vordefinierten Satz an Berechnungen aus, bei welchen die Beziehung zwischen Eingabe und Ausgabe explizit festgelegt ist. Im Gegensatz dazu ist die Beziehung zwischen Ein- und Ausgabe bei ML-Algorithmen nicht von vornherein festgelegt. Durch iterative Optimierung ermitteln diese Algorithmen die Beziehungen selbst, um eine optimale Modellleistung zu erzielen (Alzubi et al., 2018). Der Vorteil dieser Algorithmen besteht darin, dass keine a priori Annahmen über die Beziehung der Daten getroffen werden müssen und auch komplexe Zusammenhänge abgeleitet werden können (Alzubi et al., 2018). Zu den Nachteilen zählen der hohe Rechenaufwand sowie das grosse Datenvolumen, das benötigt wird. Darüber hinaus müssen die sogenannten Hyperparameter, welche die Modellstruktur bestimmen (z.B. Anzahl der Bäume oder Neuronen), optimiert werden. Eine sorgfältige Hyperparameteroptimierung trägt massgeblich zur Modellgüte bei (Probst et al., 2019; Claesen & De Moor, 2015) und ist oftmals ebenso entscheidend wie die Wahl des Algorithmus selbst (Szczipanek, 2022).

In dieser Studie wurden drei verschiedene ML-Algorithmen getestet: die Support Vector Regression (SVR) (Vapnik, 1998, S. 55ff.), ein Multi Layer Perzeptron (MLP), das zu den neuronalen Netzwerken zählt (Popescu et al., 2009), sowie XGBoost (Chen & Guestrin, 2016), das auf Decision-Tree Boosting beruht. Diese Auswahl deckt verschiedene Grundkonzepte des maschinellen Lernens ab. Da die SVR bei einer zunehmenden Anzahl an Eingabeparameter eine schlechtere Generalisierungsleistung aufweist, wurde sie nach ersten Vorversuchen nicht weiterverwendet. Um die Stärke der ML-Algorithmen, Muster aus Daten zu lernen, optimal auszunutzen, wurde der Abfluss der Zielstation

anhand des Abflusses aller anderen Stationen sowie deren um einen Zeitschritt vor- und nachgelagerten Zeitreihen modelliert. Dies resultiert in insgesamt 120 Eingabeparametern.

Das methodische Vorgehen ist für beide Algorithmen (MLP, XGBoost) identisch. Der Abfluss der Zielstation wird anhand des Abflusses von Referenzstationen modelliert. Danach wird die Differenz zwischen der modellierten und der gemessenen Zeitreihe berechnet (Residuen). Veränderungen im Abfluss der Zielstation ohne meteorologische Einflüsse sollten vom Modell nicht reproduziert werden, da sie an den Referenzstationen nicht auftreten und somit nicht im Eingangssignal enthalten sind. Die Klassifizierung der Residuen in «Anomalie» oder «keine Anomalie» wurde anhand der Residuen-Standardabweichung vorgenommen: übersteigen die Residuen die Standardabweichung σ um mehr als einen vordefinierten Faktor, zeigen sie eine Anomalie an. Die getesteten σ -Faktoren reichen von 0.5 bis 2 in Schritten von 0.25. Unter der Annahme einer Normalverteilung der Residuen sollten theoretisch 62% der Werte ausserhalb einer halben Standardabweichung liegen, 32% der Werte ausserhalb einer Standardabweichung und 5% der Werte ausserhalb von zwei Standardabweichungen. Um den Einfluss einzelner lokaler Niederschlagsereignisse zu minimieren, wurden zudem wie bei der Steigungsanalyse nur jene Zeitabschnitte als Anomalie klassifiziert, in denen der Schwellenwert an 12 von 16 aufeinanderfolgenden Tagen überschritten wird.

Die Hyperparameteroptimierung wurde mithilfe des R-Pakets „ParBayesianOptimisation“ durchgeführt. Dabei wird mit einem Gauss'schen Surrogatmodell die Modellperformance in Abhängigkeit der Hyperparameter modelliert und so innerhalb weniger Iterationen eine optimierte Hyperparameterkombination ermittelt. In Tabelle 2 und Tabelle 3 sind die optimierten Hyperparameter sowie deren Optimierungsbereich für die XGBoost- respektive MLP-Modelle ersichtlich. Die Optimierungsbandbreite wurde in vorausgehenden Versuchen ermittelt.

Tabelle 2: Hyperparameter und Optimierungsbereich der XGBoostmodelle

Hyperparameter	Beschreibung	Optimierungsbereich
eta	Eta ist die Lernrate und bestimmt, wie viel pro Iteration gelernt wird. Hohe Werte führen zu overfitting.	0.001 – 0.25
max depth	Maximale Tiefe eines Baums. Eine größere Baumtiefe ermöglicht das Erlernen komplexerer Datenstrukturen und einer größeren Vielfalt an Ausgabewerten, kann jedoch zu overfitting führen.	2 – 12
Min child weight	Legt fest, wie hoch das Gewicht eines Zweigs sein muss, damit er weiter unterteilt wird. Höhere Werte führen zu konservativeren Bäumen.	1-50
alpha	L1-Regularisierung der Gewichte	1 – 12
lambda	L2-Regularisierung der Gewichte	1 – 12
nrounds	Maximale Anzahl von Boosting-Iterationen	1 – 100

Tabelle 3: Hyperparameter und Optimierungsbereich der MLPmodelle

Hyperparameter	Beschreibung	Optimierungsbereich
nlayers	Anzahl der Schichten im neuronalen Netzwerk.	2 – 3
nnodes	Anzahl der Knoten in jeder Schicht	3 – 7
dropout rate	Prozentsatz der Eingabewerte, die während des Trainings zufällig auf 0 gesetzt werden, um overfitting zu verhindern.	0.001 – 0.01
batch size	Schrittweite, nach der die Gewichte innerhalb einer Epoche aktualisiert werden	5 – 150
Learning rate	Höhere Lernraten führen zu einer schnelleren Konvergenz, können jedoch instabil sein.	0.0001 – 0.0015

Um mögliche Trends in den Abflusszeitreihen zuverlässig zu erfassen, wurde jeder dritte Tag für die Validierung verwendet und die restlichen Tage zur Kalibrierung genutzt. Die Aufteilung der Kalibrierungsdaten in Trainings- und Testdaten wurde wie folgt vorgenommen: Für die MLP-Modelle erfolgte das Training auf 85% der Kalibrierungsdaten, die restlichen 15% dienten als Testdaten. Da XGBoost im Trainingsvorgang eine native Kreuzvalidierung unterstützt, ist eine explizite Unterteilung nicht erforderlich, und das Training wurde mit einer fünffachen Kreuzvalidierung durchgeführt.

3.5. Modellgüteparameter

Um die angewendeten Methoden analysieren und vergleichen zu können, wird die berechnete Klassifizierung (in «Anomalie» und «keine Anomalie») mit den üblichen Kriterien Präzision, Sensitivität sowie der sogenannte F1-Score verwendet. Die Präzision beschreibt die Zuverlässigkeit positiver Vorhersagen (positive Vorhersage = als Anomalie klassifiziert), wobei die Werte von 0 (keine der positiven Vorhersagen ist tatsächlich positiv) bis 1 (alle positiven Vorhersagen sind tatsächlich positiv) reichen. Sie berechnet sich wie folgt:

$$\text{Präzision} = \frac{\text{Richtig Positiv}}{\text{Richtig Positiv} + \text{Falsch Positiv}}$$

Die Sensitivität beschreibt die Fähigkeit eines Modells, tatsächlich positive Fälle zu finden, und berechnet sich folgendermassen:

$$\text{Sensitivität} = \frac{\text{Richtig Positiv}}{\text{Richtig Positiv} + \text{Falsch Negativ}}$$

Ein Wert von 0 gibt an, dass keine tatsächlich positiven Fälle gefunden wurden, ein Wert von 1, dass alle gefunden wurden.

Da eine Verbesserung eines Güteparameters häufig mit einer Verschlechterung des anderen einhergeht, dient der F1-Score als harmonisches Mittel beider Güteparameter. Modelle mit ausgeglichener Präzision und Sensitivität werden dabei stärker gewichtet als einseitig optimierte Modelle. Die Werte reichen ebenfalls von 0 bis 1. Er berechnet sich wie folgt:

$$F1\ Score = 2 * \frac{Pr\ddot{a}zision * Sensitivit\ddot{a}t}{Pr\ddot{a}zision + Sensitivit\ddot{a}t}$$

3.6. Vergleichsmodell zur Einordnung der Modellgüte

Zur besseren Einordnung der Modellgüte ist ein Vergleich der entwickelten Methoden mit einem Zufallsmodell hilfreich: Dieses identifiziert Anomalien rein zufällig. Dazu wurde ein definierter Anteil der Datenpunkte zufallsbasiert als Anomalie klassifiziert und das Ergebnis analysiert. Zusätzlich wurden analog zur Steigungsanalyse und den ML-Methoden Fehler aufgrund ihrer Auftrittsdauer gefiltert (mindestens an 12 von 16 Tagen (ML) und zusätzlich überschneidend bei zwei Zeitreihen (Steigungsanalyse)) und die Effektivität analysiert. Tabelle 4 zeigt auf, welche zufälligen Datenanteile welchen Schwellenwerten entsprechen. Diese Werte wurden aus den theoretischen Verteilungen abgeleitet.

Da eine einmalige zufallsbasierte Klassifikation das Risiko birgt, besonders gut (oder besonders schlecht) zu den Daten zu passen, wurden jeweils 100 zufallsbasierte Klassifikationen generiert und deren Median berechnet. Dadurch ist das Ergebnis robust und wiederholbar. Folgende Anteile wurden zufallsbasiert als positiv klassifiziert:

Tabelle 4: Anteile der im Vergleichsmodell zufällig als Anomalie klassifiziert und deren etwaige Entsprechung hinsichtlich Schwellenwerte für die Steigungsanalyse und ML-Methoden

Anteil zufällig als Anomalie klassifiziert	60%	40%	30%	10%	5%
ungefährer Steigungs-Schwellenwert	0.7	0.8	0.85	-	-
ungefährer σ -Faktor-Schwellenwert	0.5	-	1	1.75	2

4. Ergebnisse

4.1. Changepoint-Analyse

Wie Bereits im Kapitel zur Methode erwähnt, konnte kein Verfahren gefunden werden, dass die synthetischen Anomalien zuverlässig detektiert. Zum Ende der Studie wurde nochmals ein neuer Ansatz getestet, konnte aus Zeitgründen jedoch nicht weiterverfolgt werden (siehe Fazit).

4.2. Steigungsanalyse

In Abbildung 5 ist beispielhaft das Ergebnis der Steigungsanalyse der Aare bei Brugg dargestellt. Die beiden Kurven repräsentieren die kumulierte Summe der Abflussdifferenz der Station mit den beiden Referenzstationen. Die horizontalen Balken oben und unten zeigen die eingefügten synthetischen Anomalien: blau bedeutet, dass der Abfluss vergrößert wurde, rot, dass er verkleinert wurde. Die vertikalen Balken sind die Ergebnisse der Steigungsanalyse. Positive Abflussanomalien sind in Blau dargestellt, negative in Rot. Es zeigt sich, dass stärkere Anomalien besser detektiert werden als schwächere. Dies gilt analog auch für die Dauer der Anomalien.

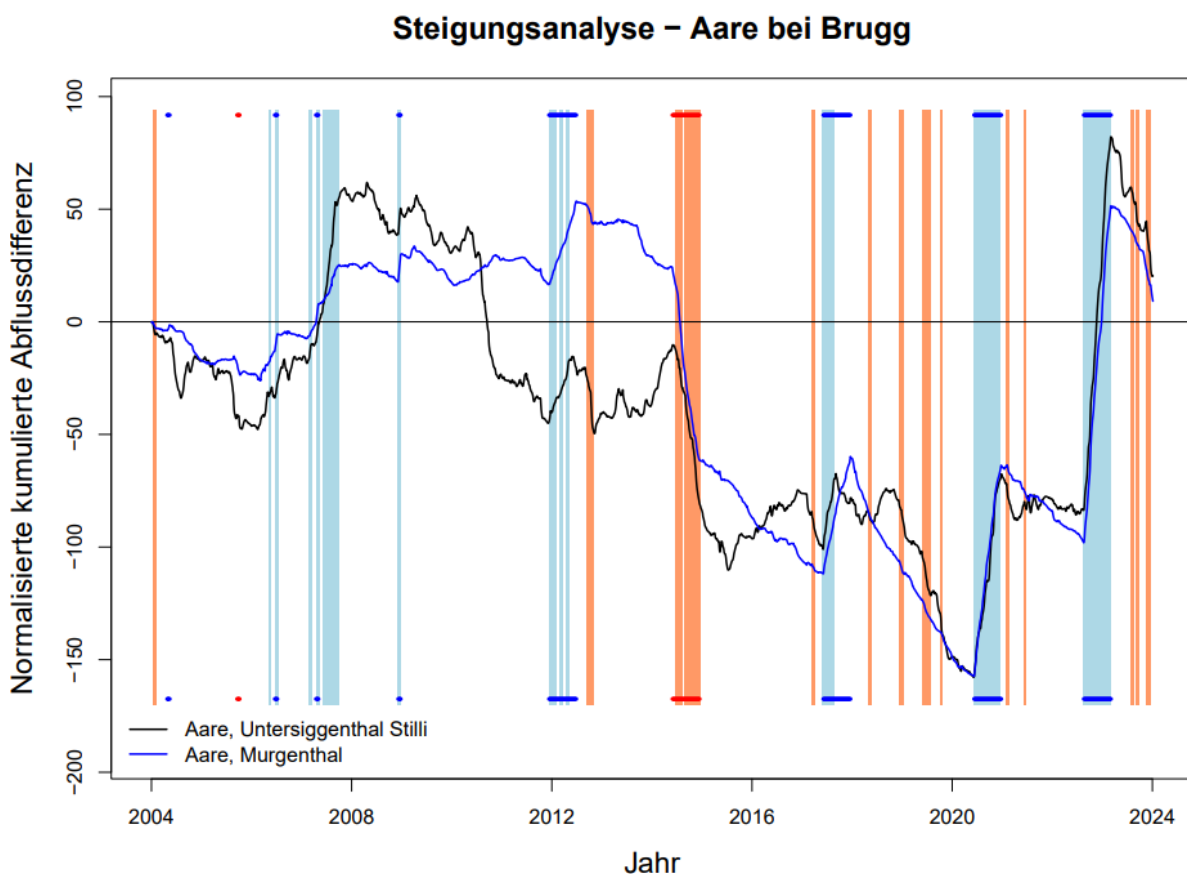


Abbildung 5: Steigungsanalyse für Aare bei Brugg mit Schwellenwert 0.75

In Abbildung 6 sind Violinplots (Boxplots mit ersichtlicher Datenverteilung) der Sensitivität für alle Steigungs-Schwellenwerte dargestellt. In Tabelle 5 sind alle Daten der Violinplots tabellarisch zusammengefasst. Es zeigt sich, dass die Sensitivität bei zunehmendem Steigungs-Schwellenwert abnimmt. Der Median nimmt dabei von 0.48 beim tiefsten bis auf 0.08 beim höchsten Steigungs-Schwellenwert ab. Zudem zeigt sich, dass die Spannweite für alle Schwellenwerte mehr als 0.5 beträgt.

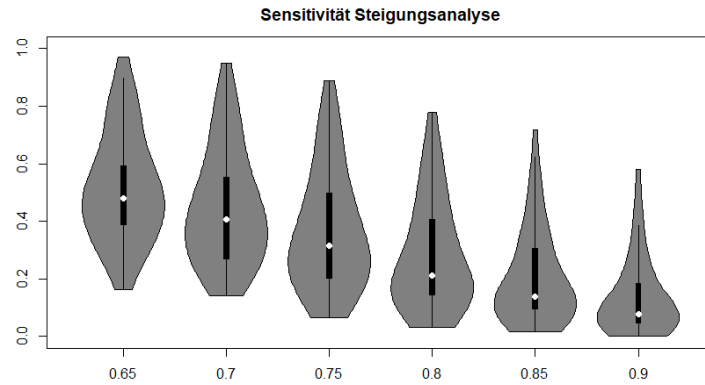


Abbildung 6: Violinplots der Sensitivität aller Schwellenwerte der Steigungsanalyse.

Im Gegensatz zur Sensitivität nimmt die Präzision mit zunehmendem Steigungs-Schwellenwert zu (Abb. 7). Der Median steigt dabei von 0.24 beim kleinsten Steigungs-Schwellenwert auf 0.56 beim grössten. Auch hier ist die Spannweite der Ergebnisse erheblich und nimmt mit zunehmendem Schwellenwert zu. Bei einigen Stationen, bei denen die Präzision gering ist, ist auch bei höheren Schwellenwerten keine Verbesserung zu verzeichnen.

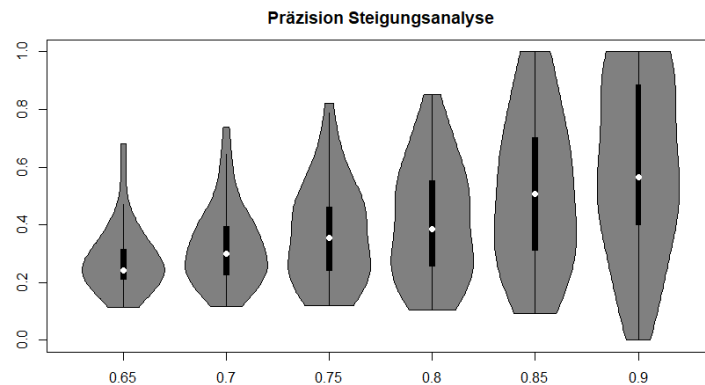


Abbildung 7: Violinplots der Präzision aller Schwellenwerte der Steigungsanalyse

Der Median des F1-Scores liegt bei den beiden niedrigsten Steigungs-Schwellenwerten etwa bei 0.33 und nimmt mit höheren Schwellenwerten ab bis auf 0.14. Auch hier ist die Spannweite der Ergebnisse gross.

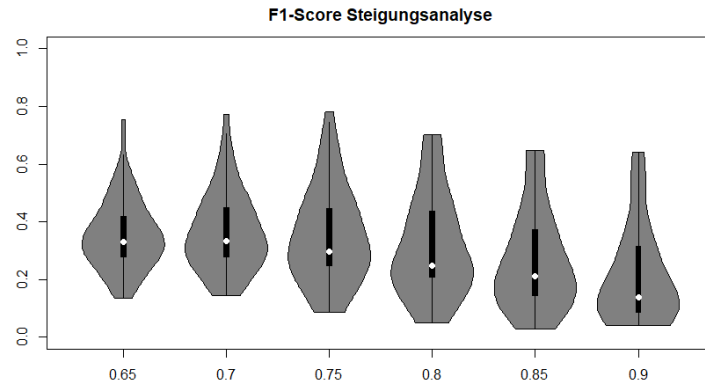


Abbildung 8: Violinplots der F1-Scores aller Schwellenwerte der Steigungsanalyse

Tabelle 5: Zusammenfassung Modellgüteparameter der Steigungsanalyse

Sensitivität						
Schwellenwert	0.65	0.7	0.75	0.8	0.85	0.9
Minimum	0.16	0.14	0.07	0.03	0.02	0.00
1. Quartil	0.39	0.27	0.20	0.14	0.10	0.05
Median	0.48	0.41	0.31	0.21	0.14	0.08
3. Quartil	0.59	0.55	0.50	0.41	0.31	0.18
Maximum	0.97	0.95	0.89	0.78	0.72	0.58
Präzision						
Schwellenwert	0.65	0.7	0.75	0.8	0.85	0.9
Minimum	0.12	0.12	0.12	0.11	0.09	0.00
1. Quartil	0.21	0.23	0.24	0.26	0.31	0.40
Median	0.24	0.30	0.35	0.39	0.51	0.56
3. Quartil	0.32	0.40	0.46	0.55	0.70	0.89
Maximum	0.68	0.74	0.82	0.85	1.00	1.00
F1-Score						
Schwellenwert	0.65	0.7	0.75	0.8	0.85	0.9
Minimum	0.14	0.14	0.09	0.05	0.03	0.04
1. Quartil	0.28	0.28	0.25	0.21	0.14	0.09
Median	0.33	0.33	0.30	0.25	0.21	0.14
3. Quartil	0.42	0.45	0.45	0.44	0.37	0.32
Maximum	0.75	0.77	0.78	0.70	0.65	0.64

4.3. Modellierung mit maschinellem Lernen

In Abbildung 9 sind die Ergebnisse der Methode mit XGBoost der Aare bei Brugg mit einem σ -Faktor-Schwellenwert von 1.25 beispielhaft für alle Methoden die maschinelles Lernen verwenden dargestellt. Die Kurve repräsentiert die mit einem 11-tägigen gleitenden Mittel geglätteten Residuen des Modells. Die vertikalen Balken zeigen die synthetisch eingefügten Anomalien an; die violetten horizontalen Balken ober- und unterhalb der Kurve kennzeichnen die detektierten Anomalien.

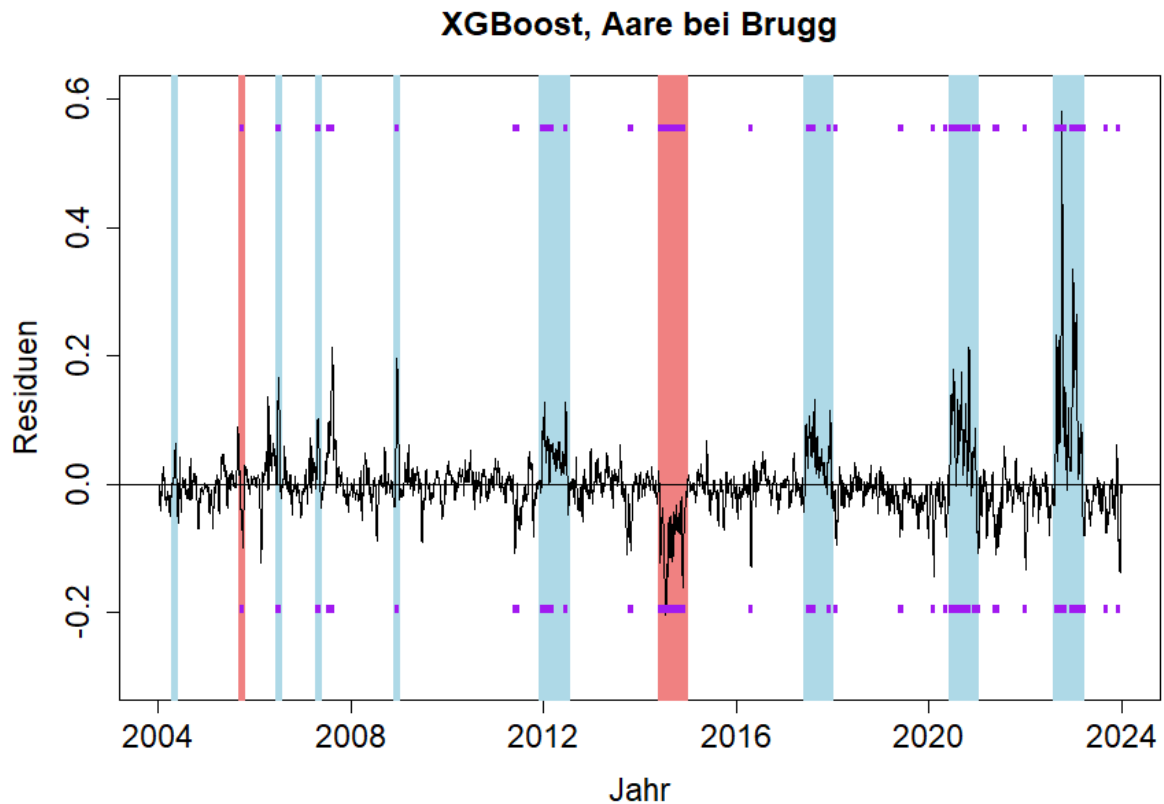


Abbildung 9: XGBoost-Modell der Aare bei Brugg mit σ -Faktor-Schwellenwert 1.25

4.3.1.XGBoost

In Abbildung 10 sind Violinplots der Sensitivität für alle σ -Faktor-Schwellenwerte dargestellt. In Tabelle 6 sind alle Daten der Violinplots tabellarisch zusammengefasst. Wie bei der Steigungsanalyse nimmt auch hier die Sensitivität bei zunehmendem Schwellenwert ab. Der Median liegt dabei beim tiefsten σ -Faktor-Schwellenwert bei 0.7 und nimmt bis auf 0.1 beim höchsten Schwellenwert ab. Auch hier beträgt die Spannweite bei den meisten Schwellenwerten mehr als 0.5.

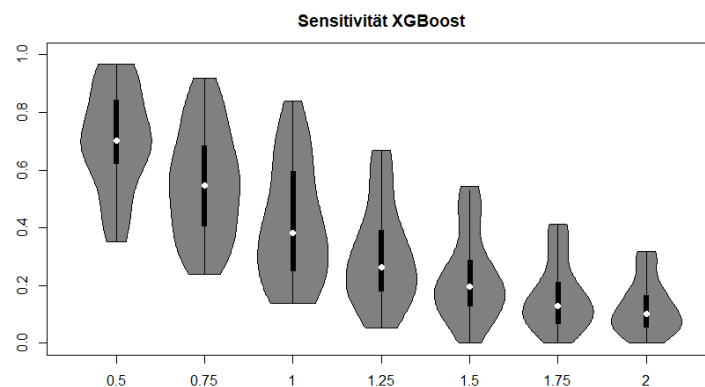


Abbildung 10: Violinplots der Sensitivität aller σ -Faktor-Schwellenwerte der XGBoost Methode

Analog zur Steigungsanalyse nimmt die Präzision mit zunehmendem Schwellenwert zu (Abb. 11). Der Median steigt dabei von 0.24 beim kleinsten σ -Faktor-Schwellenwert auf 0.78 beim grössten. Auch hier ist die Spannweite der Ergebnisse erheblich und nimmt mit zunehmendem Schwellenwert zu. Bei einigen Stationen, bei denen die Präzision gering ist, ist auch bei höheren σ -Faktor-Schwellenwerten keine Verbesserung zu verzeichnen.

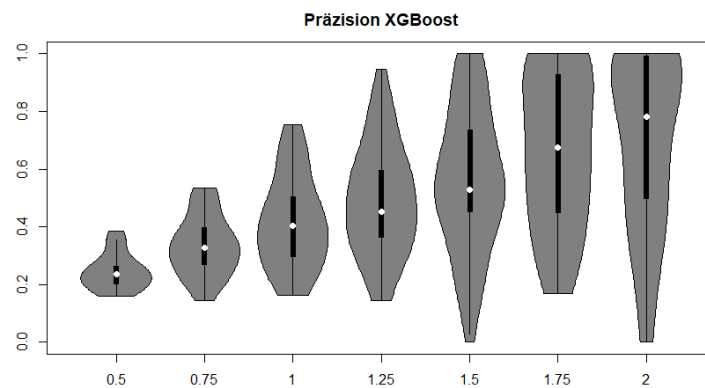


Abbildung 11: Violinplots der Präzision aller σ -Faktor-Schwellenwerte der XGBoost Methode

Der Median des F1-Scores steigt vom σ -Faktor-Schwellenwert 0.5 zunächst leicht von 0.36 auf 0.38 an (Schwellenwert 0.75), um danach kontinuierlich bis auf 0.19 beim Schwellenwert 2 zu sinken. Mit Ausnahme des kleinsten Schwellenwerts übersteigen die Spannweiten innerhalb der Schwellenwerte 0.5 oder liegen knapp darunter.

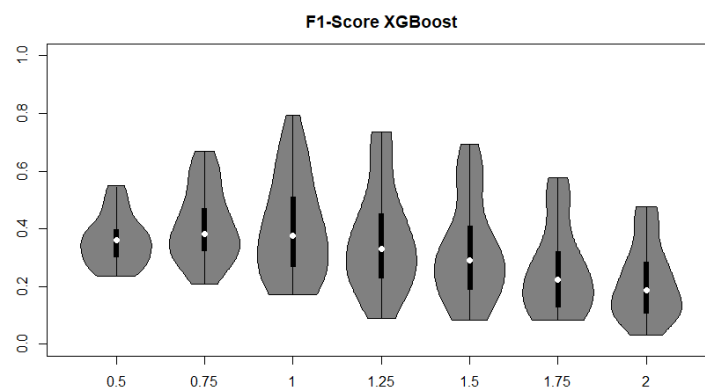


Abbildung 12: Violinplots der F1-Scores aller σ -Faktor-Schwellenwerte der XGBoost Methode

Tabelle 6: Zusammenfassung Modellgüteparameter der XGBoost-Modelle

Sensitivität							
Schwellenwert	0.5	0.75	1	1.25	1.5	1.75	2
Minimum	0.35	0.24	0.14	0.05	0.00	0.00	0.00
1. Quartil	0.62	0.41	0.25	0.18	0.13	0.07	0.06
Median	0.70	0.55	0.38	0.26	0.20	0.13	0.10
3. Quartil	0.84	0.68	0.60	0.39	0.29	0.21	0.17
Maximum	0.97	0.92	0.84	0.67	0.54	0.41	0.32
Präzision							
Schwellenwert	0.5	0.75	1	1.25	1.5	1.75	2
Minimum	0.16	0.15	0.16	0.15	0.00	0.17	0.00
1. Quartil	0.20	0.27	0.30	0.36	0.45	0.45	0.50
Median	0.24	0.33	0.40	0.45	0.53	0.68	0.78
3. Quartil	0.26	0.40	0.50	0.60	0.74	0.93	0.99
Maximum	0.38	0.54	0.75	0.95	1.00	1.00	1.00
F1-Score							
Schwellenwert	0.5	0.75	1	1.25	1.5	1.75	2
Minimum	0.24	0.21	0.17	0.09	0.08	0.09	0.03
1. Quartil	0.30	0.33	0.27	0.23	0.19	0.13	0.11
Median	0.36	0.38	0.38	0.33	0.29	0.22	0.19
3. Quartil	0.40	0.47	0.51	0.45	0.41	0.32	0.28
Maximum	0.55	0.67	0.79	0.74	0.69	0.58	0.48

Bei XGBoost-Modellen lässt sich zudem die Wichtigkeit der Eingabevariablen bestimmen. Die Analyse der Variablenwichtigkeit zeigt, dass das Modell für die Vorhersage meist auf eine oder zwei Stationen zurückgreift. Die Wichtigkeitsanalyse bietet somit eine weitere Möglichkeit, Referenzeinzugsgebiete datenbasiert zu identifizieren.

4.3.2. Multi Layer Perceptron

In Abbildung 13 sind Violinplots der Sensitivität für alle σ -Faktor-Schwellenwerte dargestellt. In Tabelle 7 sind alle Daten der Violinplots der Methode mit MLPs tabellarisch zusammengefasst. Wie bei der Steigungsanalyse nimmt auch hier die Sensitivität bei zunehmendem Schwellenwert ab. Der Median liegt dabei beim tiefsten σ -Faktor-Schwellenwert bei 0.63 und nimmt bis auf 0.13 beim höchsten Schwellenwert ab. Die Spannweite verkleinert sich zunehmend mit grösser werdendem Schwellenwert.

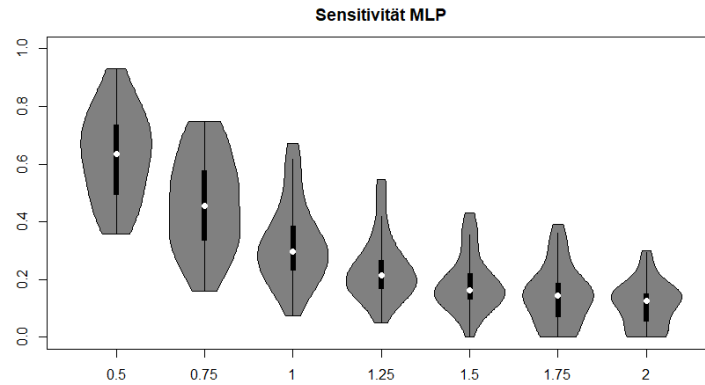


Abbildung 13: Violinplots der Sensitivität aller σ -Faktor-Schwellenwerte der MLP-Methode

Analog zur Steigungsanalyse und XGBoost nimmt die Präzision mit zunehmendem Schwellenwert zu (Abb. 14). Der Median steigt dabei von 0.19 beim kleinsten σ -Faktor-Schwellenwert auf 0.6 beim grössten. Auch hier ist die Spannweite der Ergebnisse erheblich und nimmt mit zunehmendem σ -Faktor-Schwellenwert zu. Ab einem Schwellenwert von 1.5 beträgt sie 1. Bei einigen Stationen, bei denen die Präzision gering ist, ist auch bei höheren σ -Faktor-Schwellenwerten keine Verbesserung zu verzeichnen.

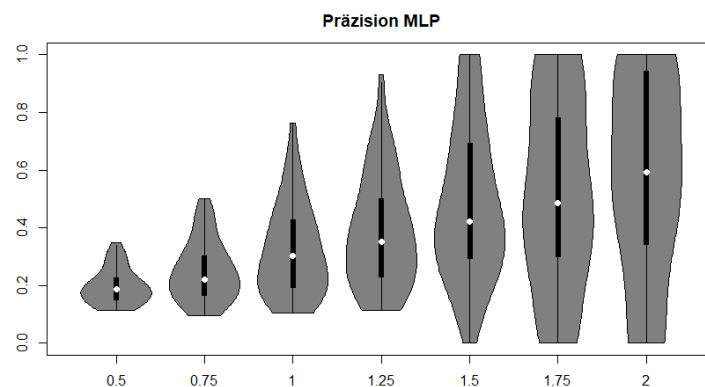


Abbildung 14: Violinplots der Präzision aller σ -Faktor-Schwellenwerte der MLP-Methode

Der Median des F1-Scores (Abb. 15) beträgt für die kleinsten beiden σ -Faktor-Schwellenwerte 0.29, um danach kontinuierlich bis zum σ -Faktor-Schwellenwert 2 auf 0.24 zu sinken. Mit Ausnahme der beiden kleinsten Schwellenwerte übersteigen die Spannweiten innerhalb der Schwellenwerte 0.5 oder liegen knapp darunter.

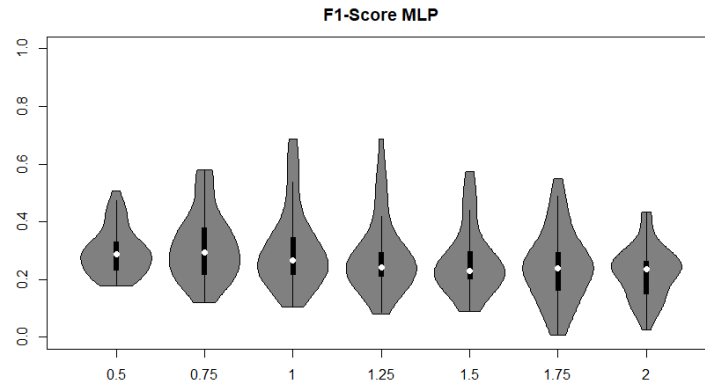


Abbildung 15: Violinplots der F1-Scores aller σ -Faktor-Schwellenwerte der MLP-Methode

Tabelle 7: Zusammenfassung Modellgüteparameter der MLP-Modelle

Sensitivität							
Schwellenwert	0.5	0.75	1	1.25	1.5	1.75	2
Minimum	0.36	0.16	0.08	0.05	0.00	0.00	0.00
1. Quartil	0.50	0.34	0.23	0.17	0.13	0.07	0.06
Median	0.64	0.46	0.30	0.21	0.16	0.14	0.13
3. Quartil	0.74	0.58	0.39	0.27	0.22	0.19	0.15
Maximum	0.93	0.75	0.67	0.55	0.43	0.39	0.30
Präzision							
Schwellenwert	0.5	0.75	1	1.25	1.5	1.75	2
Minimum	0.12	0.10	0.11	0.11	0.00	0.00	0.00
1. Quartil	0.15	0.17	0.19	0.23	0.29	0.30	0.34
Median	0.19	0.22	0.30	0.35	0.42	0.49	0.59
3. Quartil	0.23	0.30	0.43	0.50	0.69	0.78	0.94
Maximum	0.35	0.50	0.76	0.93	1.00	1.00	1.00
F1-Score							
Schwellenwert	0.5	0.75	1	1.25	1.5	1.75	2
Minimum	0.18	0.12	0.10	0.08	0.09	0.01	0.03
1. Quartil	0.24	0.22	0.22	0.21	0.20	0.16	0.15
Median	0.29	0.29	0.27	0.24	0.23	0.24	0.24
3. Quartil	0.33	0.38	0.35	0.30	0.30	0.29	0.26
Maximum	0.51	0.58	0.69	0.69	0.57	0.55	0.43

4.4. Vergleichsmodell zur Einordnung der Modellgüte

Für das Vergleichsmodell ohne Filterung der Anomaliedauer wurden folgende Modellgüteparameter ermittelt:

Tabelle 8: Modellgüteparameter des Vergleichsmodells ohne Filterung der Anomaliedauer

Anteil zufällig als Anomalie	60%	40%	30%	10%	5%
Sensitivität	0.60	0.40	0.30	0.10	0.05
Präzision	0.15	0.15	0.15	0.15	0.15
F1-Score	0.24	0.22	0.20	0.12	0.08

Dabei entspricht die Sensitivität dem Anteil positiv klassifizierter Datenpunkte, während die Präzision dem Anteil tatsächlich positiver Datenpunkte im Verhältnis zu allen Datenpunkten entspricht. Der F1-Score nimmt folglich mit abnehmendem Anteil positiv klassifizierter Datenpunkte ab. Die Ergebnisse entsprechen damit den theoretischen Erwartungen.

Für das Vergleichsmodell mit Filterung der Anomaliedauer (an mindestens 12 von 16 Tagen) basierend auf der Steigungsanalyse wurden folgende Modellgüteparameter ermittelt:

Tabelle 9: Modellgüteparameter des Vergleichsmodells mit Filter nach Steigungsanalyse

Anteil zufällig als Anomalie	60%	40%	30%	10%	5%
Sensitivität	0	0	0	0	0
Präzision	0	0	0	0	0
F1-Score	0.03	0	0	0	0

Es zeigt sich, dass nach dem Filtern in mehr als der Hälfte der Fälle keine korrekten Klassifizierungen erzielt werden und der Filter effektiv zufälliges Rauschen herausfiltern kann.

Für das Vergleichsmodell mit Filterung basierend auf der ML-basierten Methode wurden folgende Modellgüteparameter ermittelt:

Tabelle 10: Modellgüteparameter des Vergleichsmodells mit Filter nach ML-Methode

Anteil zufällig als Anomalie	60%	40%	30%	10%	5%
Sensitivität	0.46	0.26	0	0	0
Präzision	0.15	0.14	0	0	0
F1-Score	0.23	0.05	0.03	0	0

Mit abnehmendem Anteil zufällig positiv klassifizierter Datenpunkte sinken alle Modellgüteparameter und weisen bei einem Anteil von 30% in mehr als der Hälfte der Fälle einen Wert von 0 auf.

Der Vergleich der Steigungsanalyse mit den Vergleichsmodellen zeigt, dass in der Regel die Methoden bei lediglich den besten 25% der Messstationen eine höhere Sensitivität aufweisen als das Zufallsmodell. Die Präzision liegt jedoch deutlich höher: Bei mehr als 75% der Stationen übertrifft die *Steigungsanalyse* das Zufallsmodell. Dies verdeutlicht die Schwierigkeit, alle Anomalien zuverlässig zu detektieren, belegt jedoch, dass die Steigungsanalyse insbesondere bei grösseren Schwellenwerten für viele Stationen eine hohe Präzision erzielt und die Falschpositivrate wirksam begrenzt. Die Betrachtung der Resultate nach Anwendung der Dauer-Filter zeigt, dass zufälliges Rauschen auf Tagesebene über alle Schwellenwerte praktisch vollständig eliminiert werden kann.

Ein Vergleich der Methode XGBoost mit dem Zufallsmodell zeigt, dass die Sensitivität der entwickelten Methode bei kleinen und grossen σ -Faktor-Schwellenwerten deutlich besser ist (bei etwa 75% der Stationen) als das Zufallsmodell; bei mittleren Schwellenwerten ist die XGBoost Methode bei etwa der Hälfte der Stationen besser als das Zufallsmodell. Die entwickelte Methode übertrifft die Präzision des Zufallsmodells mit einzelnen Ausnahmen über alle Schwellenwerte hinweg bei allen Stationen. Die Betrachtung nach Anwendung des Dauer-Filters zeigt, dass ein σ -Faktor-Schwellenwert von 1 das zufällige Tagesrauschen vollständig eliminiert. Bei niedrigeren Schwellenwerten wird das Tagesrauschen lediglich reduziert.

Ein Vergleich des MLP-basierten Modells mit dem Zufallsmodell ergibt ein heterogenes Bild. Bei niedrigen σ -Faktor-Schwellenwerten übertrifft die Sensitivität bei etwas mehr als der Hälfte der Stationen jene des Vergleichsmodells. Bei mittleren Schwellenwerten ist dies bei weniger als 25 % der Stationen der Fall. Bei hohen Schwellenwerten weisen hingegen etwa 75 % der Stationen eine höhere Sensitivität auf als das Zufallsmodell. Die Präzision des MLP-basierten Modells übertrifft über alle Schwellenwerte hinweg bei mindestens 75 % der Stationen jene des Zufallsmodells.

5. Diskussion

5.1. Changepoint-Analyse

Trotz umfangreicher Untersuchungen konnte kein Verfahren der **Changepoint-Analyse** erarbeitet werden, das zuverlässige Ergebnisse liefert. Folgende Herausforderungen bestehen bei der Changepoint-Analyse:

1. Abflusszeitreihen (und deren abgeleitete Zeitreihen) sind stochastisch. Die Amplitude der zu erwartenden Anomalien ist geringer als die natürlichen Schwankungen, was die Detektion massgeblich erschwert, insbesondere die Detektion von Anomalien mit kurzer Zeitdauer.
2. Viele Algorithmen sind auf die Suche eines einzelnen Changepoints ausgelegt. Ein rekursives Ausführen umgeht diese Limitierung, birgt jedoch weitere Herausforderungen und führt zu grossem Aufwand mit fraglichem Nutzen. Erste Implementierungsversuche blieben erfolglos.
3. Algorithmen für die Suche mehrerer Changepoints verfügen über Regularisierungsparameter, die angepasst werden müssen. Oftmals haben Faktoren wie die Zeitreihenlänge Einfluss auf die optimalen Parameterwerte. Die Anzahl der Falschpositiven der getesteten Algorithmen übersteigt jene der richtig positiven Klassifikationen deutlich. Eine automatisierte Reduktion der Falschpositivrate erfordert weiteren Optimierungsaufwand, wobei die Übertragbarkeit zwischen Einzugsgebieten nicht gewährleistet ist.

Angesichts der beschriebenen Herausforderungen wurde entschieden, keine weiteren Versuche zu diesem Ansatz zu unternehmen und die verfügbaren Ressourcen auf andere Ansätze zu konzentrieren.

5.2. Steigungsanalyse und ML Methoden

Die Ergebnisse der Steigungsanalyse zeigen, dass es möglich ist, mit einem regionalen Vergleich mit Referenzstationen fehlerhafte oder inkonsistente Datenabschnitte zu ermitteln. Die Detektionsleistung variiert jedoch stark zwischen den Einzugsgebieten. Die Ergebnisse legen nahe, dass die Verfügbarkeit geeigneter Referenzstationen einen massgeblichen Einfluss auf die Detektionsleistung hat. Während die Sensitivität oft geringer ist als beim Zufallsmodell, ist die Präzision, vor allem bei hohen Schwellenwerten, deutlich höher. Die Methode bietet daher trotz Verbesserungspotential einen klaren Mehrwert. Zudem ermöglicht eine visuelle Kontrolle der kumulierten Summenkurven eine schnelle Einschätzung über das Vorhandensein von möglichen Anomalien welche nur schwer automatisiert erfasst werden können.

Von den getesteten **ML-Methoden** erzielte die auf XGBoost beruhende Methode die besten und konsistentesten Ergebnisse. Mit wenigen Ausnahmen übertrifft diese Methode sowohl die Sensitivität als auch die Präzision der Steigungsanalyse. Auch bei dieser Methode variiert die Detektionsleistung zwischen den Stationen erheblich. Im Gegensatz zur Steigungsanalyse steht hier keine Grafik zur Verfügung, die einen schnellen Überblick ermöglicht. Ein Vorteil dieser Methode besteht darin, dass die Vorauswahl von Referenzstationen entfällt, da ML-Algorithmen in der Lage sind, in grossen Datensätzen komplexe Zusammenhänge zu erkennen und somit alle Stationen gleichzeitig einbezogen werden können. Ein weiterer Vorteil besteht darin, dass sich das Modell hinsichtlich der Wichtigkeit der Eingabeparameter analysieren lässt, wodurch ersichtlich wird, welche Stationen vom Modell herangezogen werden und welche nicht. Diese Information könnte für die Auswahl der Referenzstationen für die Steigungsanalyse genutzt werden. Die auf MLP basierende Methode erzielt durchgehend schwächere Ergebnisse und bietet zudem den Nachteil einer geringeren Modellinterpretierbarkeit im Vergleich zu XGBoost.

Obwohl die Methoden prinzipiell automatisiert ablaufen, ist eine automatische Übernahme der Ergebnisse, ohne vorherige kritische Betrachtung nicht zu empfehlen. Die Ergebnisse zeigen, dass die Variabilität zwischen den Stationen beträchtlich sein kann und zuverlässige Ergebnisse nicht garantiert sind. Zudem ist eine Kombination der Methoden sinnvoll, um ein umfassenderes Bild zu erhalten, da dadurch die Sensitivität oder Präzision bei gleichbleibendem Schwellenwert verbessert werden kann.

Die Abflussmodellierung mittels XGBoost ist bei **Niedrigwasser** präziser als bei Hochwasser. Der mittlere prozentuale Fehler liegt, während der 10% tiefsten Abflüsse bei 50% der Stationen unter 6.1% und bei 75% der Stationen unter 8.2%, während er bei den höchsten 10% der Abflüsse bei 50% der Stationen unter 15.1% und bei 75% der Stationen unter 19.1% liegt. Niedrigwasserwerte fallen jedoch durch die z-Normierung weniger ins Gewicht als Hochwasserwerte, da diese weniger stark vom Mittel abweichen können. Die saisonale Normierung verstärkt diesen Effekt, sofern das Niedrigwasser saisonal bedingt ist. Umgekehrt fallen saisonal ausserordentlich auftretende Niedrigwasserereignisse stärker ins Gewicht. Zudem ist die Dauer der Niedrigwasserereignisse in der Regel grösser als die von Hochwasser, was eine Detektion erleichtert. Zusammenfassend lässt sich festhalten, dass die XGBoost-basierte Methode, während Niedrigwasserperioden präzisere Ergebnisse liefert, die Sensitivität jedoch geringer ausfällt.

Bei der Evaluation der Methoden bestehen diverse Schwierigkeiten. Erstens entsprechen die synthetischen Anomalien nicht den in der Realität auftretenden nichtlinearen Anomalien. Zudem überlagern natürliche Variationen sowie potenzielle reale Anomalien die synthetischen. Dies wirkt sich sowohl auf die Auffindbarkeit der synthetischen Anomalien aus als auch auf deren Übertragbarkeit auf reale Anomalien. Da kein Wissen über reale Anomalien vorliegt, ist anzunehmen, dass ein Teil der als falsch positiv klassifizierten Ergebnisse tatsächlich richtig positiv sind und somit die Präzision negativ beeinflussen.

Auch wenn durch die saisonale z-Normierung ein Grossteil der Saisonalität aus den Zeitreihen herausgerechnet wird, können Reste davon übrigbleiben und die Analysen negativ beeinflussen. Zudem kann das zufällige Auftreten mehrerer Hochwasser zur gleichen Jahreszeit die saisonale z-Normierung negativ beeinflussen und theoretisch sogar neue Artefakte erzeugen. Je kürzer die normierten Zeitreihen sind, desto ausgeprägter wird dieses Problem.

Die Wahl der Steigungs- oder σ -Faktor-Schwellenwerte bestimmt, welcher Anteil der Zeitreihe als Anomalie klassifiziert werden kann. Diese werden danach noch anhand der Anomalie-Dauer gefiltert; daraus ergeben sich zwei Schwierigkeiten: Bei hohen Schwellenwerten können möglicherweise nicht alle Anomalien (Fehler) gefunden werden; ein tiefer Schwellenwert bedeutet jedoch, dass durch die Anzahl an initial markierten Daten automatisch die Wahrscheinlichkeit steigt, Falschpositive zu erhalten. Dadurch können die Ergebnisse der einzelnen Schwellenwerte nicht direkt auf unveränderte Zeitreihen übertragen werden, da diese in der Regel einen geringeren Anteil an Anomalien aufweisen.

5.3. Allgemeine Erkenntnisse

Die Ergebnisse lassen sich jedoch verallgemeinert interpretieren. Grössere Schwellenwerte zur Klassierung von Fehlern oder Anomalien führen zu einer höheren Präzision, kleinere zu einer höheren Sensitivität. Dieser Zusammenhang ist inhärent und entspricht den theoretischen Erwartungen. Die Wahl des Schwellenwerts ist folglich vom Analyseziel abhängig. Ist eine hohe Sicherheit bei den gefundenen Anomalien das Ziel, sollten hohe Schwellenwerte gewählt werden. *Sollen möglichst alle Anomalien detektiert werden, ist ein tiefer Schwellenwert geeigneter.* Zudem lässt sich festhalten, dass längere Anomalien mit allen Methoden besser detektiert werden. Dies ist vermutlich darauf zurückzuführen, dass sich kurze Anomalien nur schwer von der natürlichen Variabilität unterscheiden

lassen. Werden nur längere Anomalien gesucht, können die Filterparameter so angepasst werden, dass die Präzision auch bei einem tiefen Schwellenwert weniger stark beeinträchtigt wird.

5.4. Einfluss der Einzugsgebietscharakteristiken

Die Performanz aller Methoden ist sehr heterogen. Einige Einzugsgebiete lassen sich mit allen Methoden gut klassifizieren, während andere bei allen Methoden eine geringere Detektionsleistung aufweisen. Viele Einzugsgebiete erzielen jedoch bei den verschiedenen Methoden unterschiedliche Ergebnisse. Die einzige Einzugsgebietscharakteristik, die mit den Ergebnissen korreliert, ist die Grösse des Einzugsgebiets (siehe Abb. 16). Mit der Grösse der Einzugsgebiete lassen sich 42% der Variation der Sensitivität erklären ($p\text{-Wert} < 0.001$). Die Einzugsgebietsgrösse korreliert in ähnlichem Umfang auch mit der Präzision und dem F1-Score. Grössere Einzugsgebiete erzielen bessere Ergebnisse als kleinere. Es ist jedoch anzunehmen, dass es sich dabei nicht um einen kausalen Zusammenhang handelt. Die vergleichsweise gute Detektionsleistung bei einigen der kleinen Einzugsgebiete legt die Vermutung nahe, dass tatsächlich die Qualität der Referenzstationen mit der Grösse der Einzugsgebiete zunimmt.

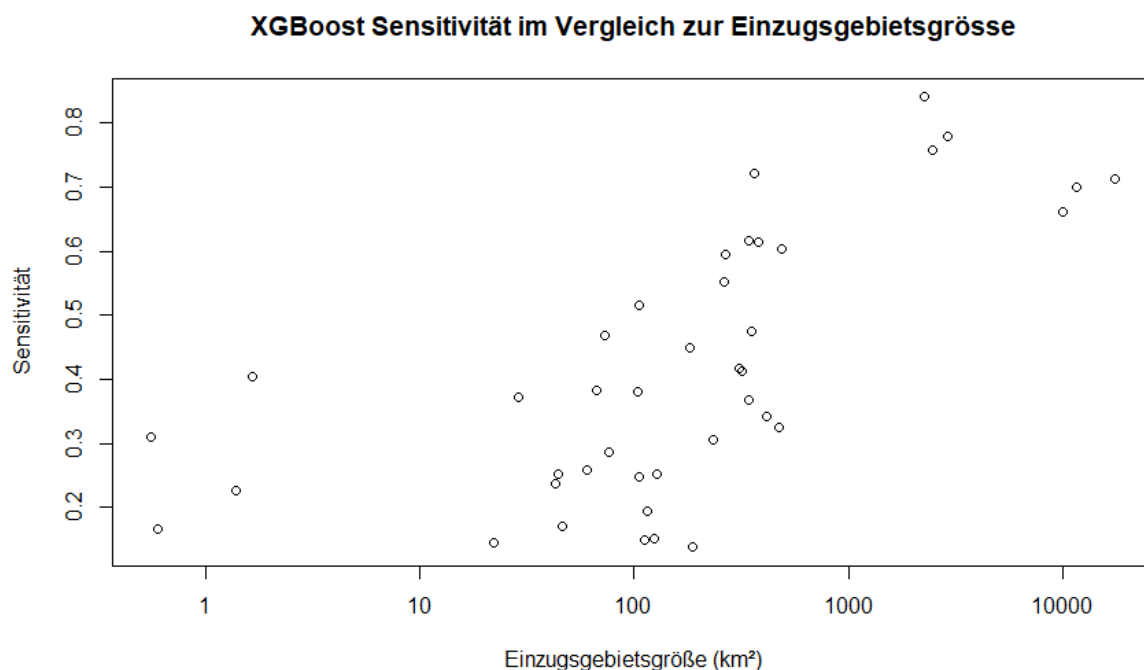


Abbildung 16: Einfluss der Einzugsgebietsgrösse auf die Sensitivität der Methode XGBoost

Dieser Zusammenhang lässt sich wie folgt erklären: Grössere Einzugsgebiete verfügen häufiger über Stationen, die am gleichen Gewässer vor- oder nachgelagert sind und sich daher sehr gut als Referenzstationen eignen. Einige der kleinsten Einzugsgebiete liegen räumlich so nahe beieinander, dass sie denselben meteorologischen Einflüssen ausgesetzt sind und sich daher ebenfalls gut als Referenzstationen eignen. Für kleine Einzugsgebiete stehen jedoch in der Regel weniger geeignete Referenzeinzugsgebiete zur Verfügung, was sich entsprechend negativ auf die Detektionsleistung auswirkt. Zudem nimmt der Variationskoeffizient mit steigender Einzugsgebietsgrösse ab, was auf ein geringeres natürliches Rauschen hindeutet (Abb. 17).

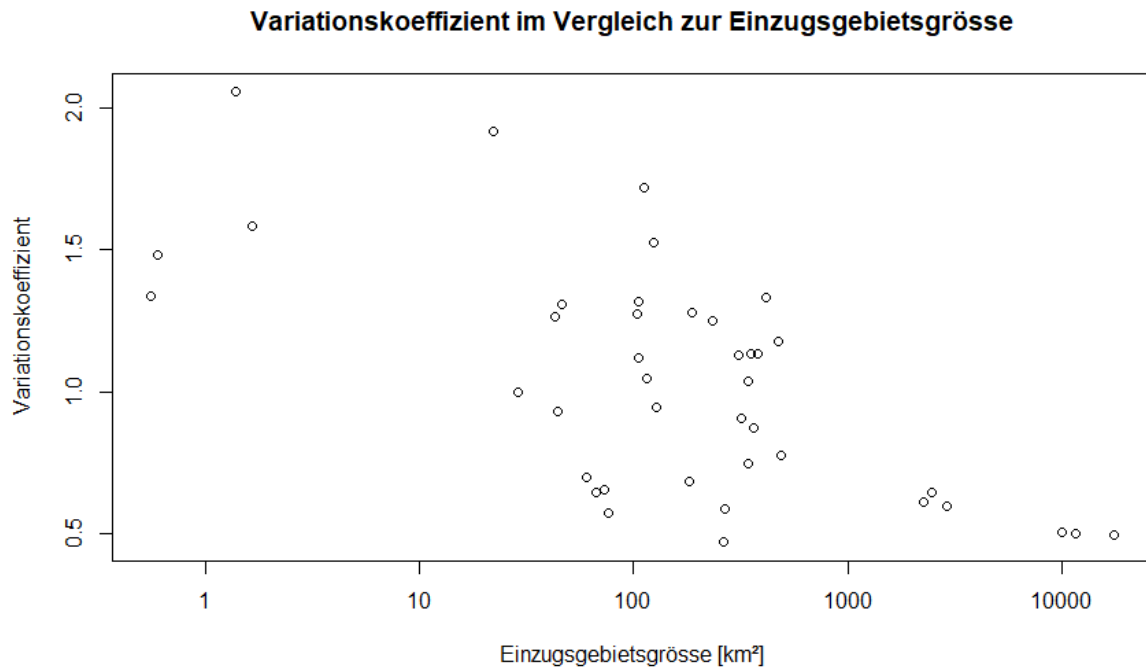


Abbildung 17: Einfluss des Variationskoeffizienten des Abflusses auf die Sensitivität

Wie aus Abbildung 18 ersichtlich ist, nimmt die Sensitivität mit zunehmendem Variationskoeffizient ab. Analog zur Einzugsgebietsgrösse zeigt sich dieser Zusammenhang über alle Methoden und Schwellenwerte hinweg. Für den dargestellten Schwellenwert erklärt der Variationskoeffizient 46% der Variation der Sensitivität ($p\text{-Wert} < 0.001$).

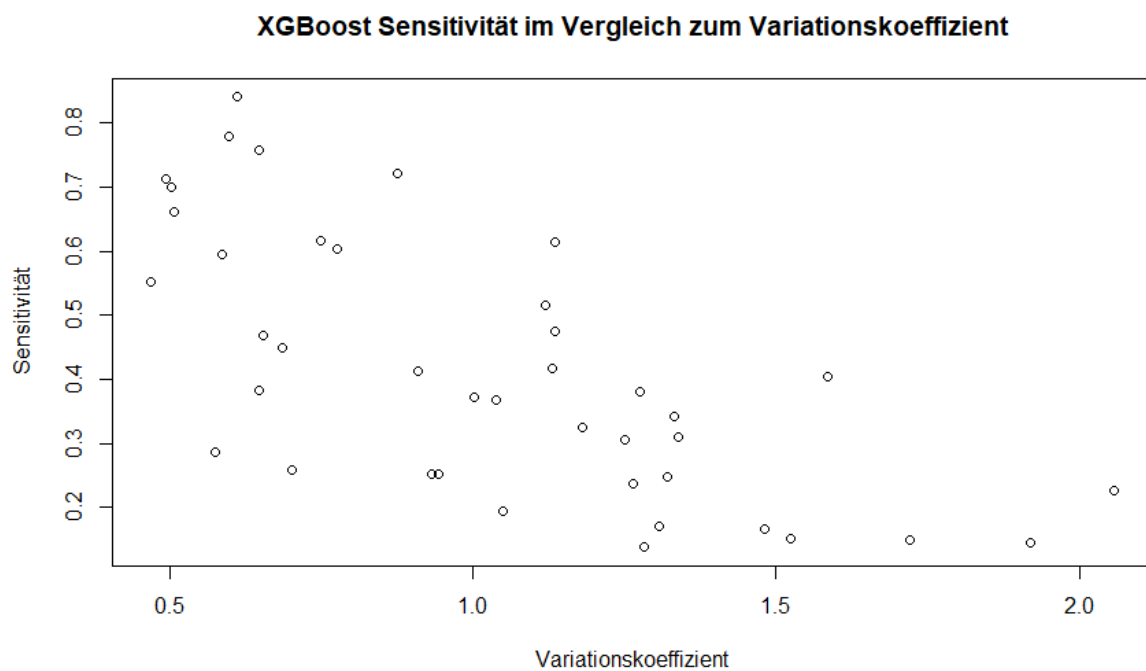


Abbildung 18: Die Verteilung der Variationskoeffizienten des Abflusses in abhängigkeit der Einzugsgebietsgrösse

6. Fazit

Es konnte erfolgreich aufgezeigt werden, dass Anomalien in Abflussdaten mittels Vergleichs von Stationen sowie mittels Modellierung mit ML-Methoden gefunden werden können. Es konnte jedoch keine Changepoint-Methode gefunden werden, die verwertbare Ergebnisse liefert. Die ML-Methode, die auf XGBoost beruht, liefert die zuverlässigsten Ergebnisse. Die angestrebte Automatisierung der Auswertung kann wegen der grossen Variabilität der Ergebnisse jedoch nicht empfohlen werden. Alle erfolgreich getesteten Ansätze haben ihre Vorteile. Während die Steigungsanalyse (beruhend auf einem Stationenvergleich) eine Graphik zur schnellen visuellen Kontrolle bietet, entfällt beim ML-Ansatz die Auswahl spezifischer Referenzstationen. Durch eine Kombination der Methoden kann die Aussagekraft verbessert werden.

Die Verfügbarkeit von geeigneten Referenzstationen ist bei beiden Ansätzen entscheidend um zuverlässige Resultate zu erzielen. Zudem steigt die Detektionsleistung der Methoden mit der Dauer und der Magnitude der Anomalien. Dies bedeutet, dass die Chancen zum Auffinden der wichtigsten Anomalien am grössten sind.

Mangels verfügbarer Vergleichsliteratur ist eine Einordnung der Ergebnisse und Methoden schwierig. Vergleichbare Studien mit neuen Ansätzen sowie eine Weiterentwicklung dieser Methoden wären aus wissenschaftlicher Sicht wertvoll. Erste Test mit dem R-Package «RatingShiftHappens» (Mendez & Renard, 2023) zeigen vielversprechende Resultate bei der Changepointanalyse auf den Residuen des XGBoostmodells (Siehe Abbildung 24, Anhang C). Leider konnte dieser Ansatz aus Zeitgründen nicht mehr vertieft untersucht werden und sollte daher Gegenstand weiterführender Studien sein. Auch die Anwendung der Method in einer anderen Region wäre sinnvoll, um die Übertragbarkeit der Ergebnisse besser einschätzen zu können. Des Weiteren wäre auch ein Einbezug stark beeinflusster Stationen spannend.

7. Literaturverzeichnis

- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: an overview. In *Journal of physics: conference series* (Vol. 1142, p. 012012). IOP Publishing. DOI: 10.1088/1742-6596/1142/1/012012
- Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2), 339-367. DOI: 10.1007/s10115-016-0987-z
- Begert M, Schlegel T, Kirchhofer W. 2005. Homogeneous Temperature and Precipitation Series of Switzerland from 1864 to 2000. *International Journal of Climatology* 25: 65-80. DOI: 10.1002/joc.1118
- Bell, S. A., Carroll, P. A., Beardmore, S. L., England, C., & Mander, N. (2017). A methodology for study of in-service drift of meteorological humidity sensors. *Metrologia*, 54(3), S63-S73. DOI: 10.1088/1681-7575/aa6dd0
- Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. *ACM computing surveys (CSUR)*, 54(3), 1-33. DOI: 10.1145/3444690
- Bois, P. (1986). Contrôle de séries chronologiques corrélées par étude du cumul des résidus de la corrélation. *Colloques et séminaires, 2èmes journées hydrologiques de l'ORSTOM, annexe, 3*.
- Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, 58(1-2), 11-27. DOI: 10.1016/0022-1694(82)90066-X
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). DOI: 10.1145/2939672.2939785
- Choi, K., Yi, J., Park, C., & Yoon, S. (2021). Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines. *IEEE access*, 9, 120043-120065. DOI: 10.1109/ACCESS.2021.3107975
- Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*. DOI: 10.48550/arXiv.1502.02127
- Erdman, C., & Emerson, J. W. (2008). bcp: an R package for performing a Bayesian analysis of change point problems. *Journal of statistical Software*, 23, 1-13. DOI: 10.18637/jss.v023.i03
- Fearnhead, P., Maidstone, R., & Letchford, A. (2019). Detecting changes in slope with an l0 penalty. *Journal of Computational and Graphical Statistics*, 28(2), 265-275. DOI: 10.1080/10618600.2018.1512868
- Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological time series and landscape attributes for 331 catchments in hydrologic Switzerland, *Earth Syst. Sci. Data*, 15, 5755-5784, 10.5194/essd-15-5755-2023, 2023. DOI: 10.5194/essd-15-5755-2023

- Killick, R., & Eckley, I. A. (2014). changepoint: An R package for changepoint analysis. *Journal of statistical software*, 58, 1-19. DOI: 10.18637/jss.v058.i03
- Killick, R., Beaulieu, C., Taylor, S., et al. (2021): EnvCpt: Detection of structural changes in climate and environment time series. <https://CRAN.R-project.org/package=EnvCpt> (letzter Zugriff: 11.6.2026)
- Lindeløv, J. (2020). mcp: An R package for regression with multiple change points. DOI: 10.31219/osf.io/fzqxv
- Mendez, F. Renard, B. (2023). RatingShiftHappens. (INRAE, RiverLy and RECOVER). <https://github.com/Felipemendezrios/RatingShiftHappens> (letzter Zugriff: 11.6.2026)
- Muggeo, V. M. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R news*, 8(1), 20-25.
- Niu, Y. S., Hao, N., & Zhang, H. (2016). Multiple change-point detection: a selective overview. *Statistical Science*, 611-623. <http://www.jstor.org/stable/26408091>
- Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(2), 126-135. DOI: 10.2307/2346729
- Piroozmand, P. Farshi, D. (2025). Anomalieerkennung und automatische Datenkorrektur von Pegelmessstationen. *Wasser Energie Luft*. Vol. 117, Heft 3. Baden, S. 108-114, DOI: [10.5169/seals-1090457](https://doi.org/10.5169/seals-1090457)
- Popescu, M. C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579-588. <https://dl.acm.org/doi/abs/10.5555/1639537.1639542> (letzter Zugriff: 08. 06. 2026)
- Probst P., Boulesteix A., Bischl B. (2019). Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *Journal of Machine Learning Research* (20), 1-32 <https://dl.acm.org/doi/10.5555/3322706.3361994> (letzter Zugriff: 08. 06. 2026)
- Reitan, T. (2022). Review of methods for automation of quality control on hydrologic time series and considerations for a research approach at NVE. https://publikasjoner.nve.no/rapport/2022/rapport2022_15.pdf (letzter Zugriff: 08. 06. 2026)
- Schmocker- Fackel, P. Hüsler, F. (2025). Messnetzkonzept Hydrologie. Grundlagen für Betrieb und Weiterentwicklung des hydrologischen Basismessnetzes. Herausgeber: Bundesamt für Umwelt. <https://www.bafu.admin.ch/de/publication?id=Nv8LnfQAAM0d> (letzter Zugriff: 08. 06. 2026)
- Shaukat, K., Alam, T. M., Luo, S., Shabbir, S., Hameed, I. A., Li, J., ... & Javed, U. (2021). A review of time-series anomaly detection techniques: A step to future perspectives. In: *Future of Information and Communication Conference*, 865-877. Springer, Cham. DOI: 10.1007/978-3-030-73100-7_60
- Song, H., & Chen, H. (2024). Practical and powerful kernel-based change-point detection. *IEEE Transactions on Signal Processing*, (72), 5174-5186. DOI: 10.1109/TSP.2024.3479274

Szczepanek, R. (2022). "Daily Streamflow Forecasting in Mountainous Catchment Using XGBoost, LightGBM and CatBoost" *Hydrology* 9, no. 12: 226. DOI: 10.3390/hydrology9120226

Vapnik, Vladimir N. (1998). *Statistical Learning Theory*. New York. Wiley, 768 pages

Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, 57(1), 1-42. DOI: 10.1145/3691338

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2001). *strucchange. An R package for testing for structural change in linear regression models*. (May 2001 Aufl.) SFB Adaptive Information Systems and Modelling in Economics and Management Science, WU Vienna University of Economics and Business. Report Series SFB "Adaptive Information Systems and Modelling in Economics and Management Science" Nr. 55 DOI: 10.57938/2a00fe83-b430-465c-9750-08c8ad0035b3

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., ... & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote sensing of Environment*, 232, 111181. DOI: 10.1016/j.rse.2019.04.034

Zhao, Q., Zhu, Y., Wan, D., Yu, Y., & Cheng, X. (2018). Research on the data-driven quality control method of hydrological time series data. *Water*, 10(12), 1712. DOI: 10.3390/w10121712

8. Anhang

Der Anhang besteht aus 3 Teilen. In Teil A werden die verwendeten Abflussmesstationen aufgeführt. In Teil B sind die getesteten Changepoint Algorithmen zu entnehmen. In Teil C sind zusätzliche Abbildungen zu finden.

A: Verwendete Abflussstationen

Tabelle 11: Liste der verwendeten Stationen (EZG steht für Einzugsgebiet, MQ für mittleren Abfluss; unterschiedliche Kommastellen in diesen beiden Kolonnen deuten darauf hin, wie genau die Werte bestimmt wurden, die Präzision ist bei kleinen Einzugsgebieten höher. Quelle: <https://www.hydrodaten.admin.ch/>).

Station numm er	Gewässer	Stationsname	EZG Fläche km ²	Stationshöhe m ü.M	Mittlere EZG Höhe m ü.M	Vergletscherung [%]	MQ	Einheit
2016	Aare	Brugg	11681	336	1000	1.5	312	m ³ /s
2030	Aare	Thun	2459	552	1746	6.9	111	m ³ /s
2034	Broye	Caserne d'aviation	416	446	715	0	7.7	m ³ /s
2063	Aare	Murgenthal	10059	403	1066	1.7	285	m ³ /s
2084	Muota	Ingenbohl	317	439	1363	0	19	m ³ /s
2102	Sarner Aa	Sarnen	269	469	1281	0	10	m ³ /s
2110	Reuss	Mühlau, Hünenberg	2902	393	1371	2.2	129	m ³ /s
2125	Lorze	Frauenthal	262	391	678	0	7.3	m ³ /s
2151	Simme	Oberwil	344	780	1641	2.4	12	m ³ /s
2152	Reuss	Luzern, Geissmattbrücke	2254	433	1504	2.8	109	m ³ /s
2159	Gürbe	Belp, Mülimatt	116	522	846	0	2.6	m ³ /s
2179	Sense	Thörishaus, Sense matt	351	554	1071	0	8.5	m ³ /s
2205	Aare	Untersiggenthal, Stilli	17553	331	1064	1.4	555	m ³ /s
2232	Allenbach	Adelboden	29	1299	1863	0	1.2	m ³ /s
2251	Rotenbach	Plaffeien, Schwyberg	1.65	1278	1455	0	80	l/s
2252	Schwändlibach	Plaffeien, Schwyberg	1.38	1218	1439	0	38	l/s
2276	Grosstalbach	Isenthal	44	765	1819	6.7	1.8	m ³ /s

2282	Sperbelgraben	Wasen, Kurzeneialp	0.56	913	1070	0	16	l/s
2283	Rappengraben	Wasen, Riedbad	0.6	998	1142	0	21	l/s
2290	Areuse	St-Sulpice	104	752	1110	0	4.7	m³/s
2307	Suze	Sonceboz	127	644	1036	0	4.2	m³/s
2343	Langete	Huttwil, Häberenbad	60	602	760	0	1.2	m³/s
2369	Menthue	Yvonand, La Mauguettaz	105	448	675	0	1.6	m³/s
2371	Orbe	Le Chenit, Frontière	46	1040	1235	0	0.91	m³/s
2378	Orbe	Orbe, Le Chalet	343	452	1139	0	11	m³/s
2409	Emme	Eggiwil, Heidebühl	124	744	1281	0	4.3	m³/s
2412	Sionge	Vuippens, Château	43	683	865	0	1.2	m³/s
2416	Aabach	Hitzkirch, Richensee	73	465	581	0	1.3	m³/s
2417	Suhre	Oberkirch	76	505	583	0	1.2	m³/s
2434	Dünnern	Olten, Hammermühle	234	404	711	0	3.5	m³/s
2436	Chli Schliere	Alpnach, Chilch Erli	22	454	1345	0	0.68	m³/s
2450	Wigger	Zofingen	366	430	656	0	5.5	m³/s
2458	Seyon	Valangin	112	635	978	0	0.81	m³/s
2469	Kander	Hondrich	491	650	1854	5.1	20	m³/s
2471	Murg	Murgenthal, Walliswil	183	419	653	0	3.4	m³/s
2480	Areuse	Boudry	378	445	1077	0	11	m³/s
2487	Kleine Emme	Werthenstein, Chappelboden	311	595	1167	0	11	m³/s
2497	Luthern	Nebikon	105	489	749	0	1.4	m³/s
2500	Worble	Ittigen	67	521	666	0	1	m³/s
2603	Ilfis	Langnau	187	685	1039	0	5.2	m³/s
2634	Kleine Emme	Emmen	478	438	1054	0	15	m³/s

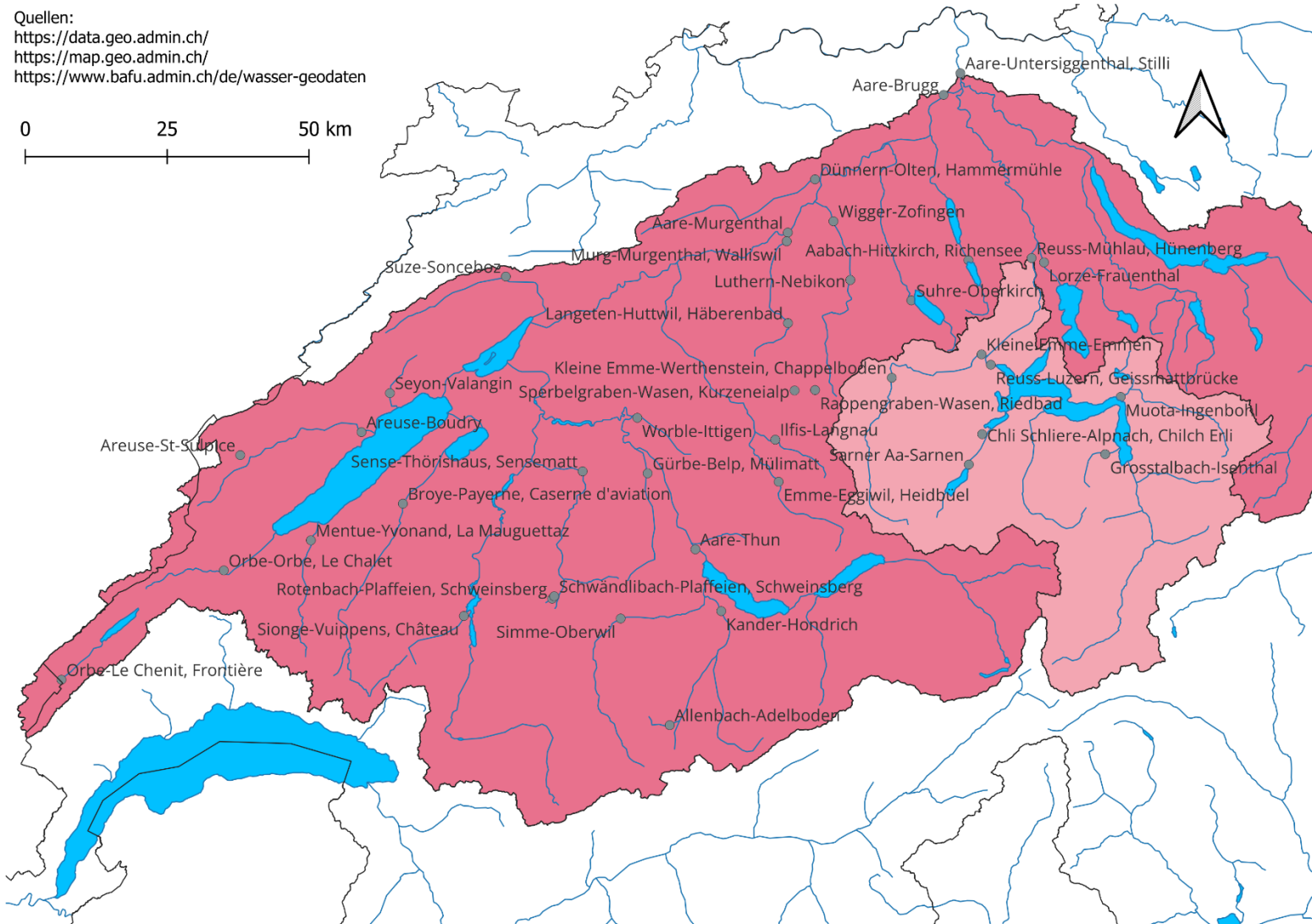


Abbildung 19: Untersuchungsgebiet mit allen verwendeten Messstationen, die Einzugsgebiete der Reuss und Aare sind eingefärbt

Tabelle 12: Liste der Referenzstationen

Stationsnummer	Gewässer	Stationsname	Referenzstation 1	Referenzstation 2
2016	Aare	Brugg	2205	2063
2030	Aare	Thun	2469	2151
2034	Broye	Caserne d'aviation	2369	2412
2063	Aare	Murgenthal	2016	2030
2084	Muota	Ingenbohl	2102	2152
2102	Sarner Aa	Sarnen	2084	2152
2110	Reuss	Mühlau, Hünenberg	2152	2084
2125	Lorze	Frauenthal	2471	2416
2151	Simme	Oberwil	2469	2276
2152	Reuss	Luzern, Geissmattbrücke	2110	2084
2159	Guerbe	Belp, Mülimatt	2497	2179
2179	Sense	Thörishaus, Sensematt	2034	2159
2205	Aare	Untersiggenthal, Stilli	2016	2110
2232	Allenbach	Adelboden	2276	2469
2251	Rotenbach	Plaffeien, Schwyberg	2252	2436
2252	Schwändlibach	Plaffeien, Schwyberg	2251	2436
2276	Grosstalbach	Isenthal	2151	2469
2282	Sperbelgraben	Wasen, Kurzeneialp	2283	2409
2283	Rappengraben	Wasen, Riedbad	2282	2409
2290	Areuse	St-Sulpice	2307	2480
2307	Suze	Sonceboz	2458	2290
2343	Langete	Huttwil, Häberenberg	2500	2497
2369	Menthue	Yvonand, La Mauguettaz	2034	2500
2371	Orbe	Le Chenit, Frontiere	2378	2290
2378	Orbe	Orbe, Le Chalet	2480	2371
2409	Emme	Eggiwil, Heidbüel	2603	2487
2412	Sionge	Vuippens, Chateau	2307	2458
2416	Aabach	Hitzkirch, Richensee	2417	2500
2417	Suhre	Oberkirch	2416	2500
2434	Dünnern	Olten, Hammermühle	2471	2450
2436	Chli Schliere	Alpnach, Chilch Erli	2487	2102
2450	Wigger	Zofingen	2497	2471
2458	Seyon	Valangin	2412	2307
2469	Kander	Hondrich	2151	2276
2471	Murg	Murgenthal, Walliswil	2434	2450
2480	Areuse	Boudry	2290	2378
2487	Kleine Emme	Werthenstein, Chappelboden	2634	2603
2497	Luthern	Nebikon	2450	2471
2500	Worble	Ittigen	2417	2159
2603	Ilfis	Langnau	2179	2409
2634	Kleine Emme	Emmen	2487	2497

B: Getestete Changepoint Algorithmen

Tabelle 13: Liste der getesteten Changepoint Algorithmen

Name R-Package	Algorithmus
BreakPoints	Pettit (Pettitt, 1979)
trend	Buishand Range Test (Buishand, 1982)
bcp	Bcp (Erdman & Emerson, 2008)
mcp	Mcp (Lindeløv, 2020)
segmented	Segmented (Muggeo, 2008)
strucchange	Breakpoint (Zeileis et al., 2001)
changepoint	Cpt.mean, cpt.var, cpt.meanvar (Killick & Eckley, 2014)
EnvCpt	Envcpt (Killick et al., 2025)
kerSeg	fGKCP (Song & Chen, 2024)
cpop	CPOP (Fearnhead et al., 2019)
RBeast	Beast (Zhao et al., 2019)

C: Zusätzliche Abbildungen

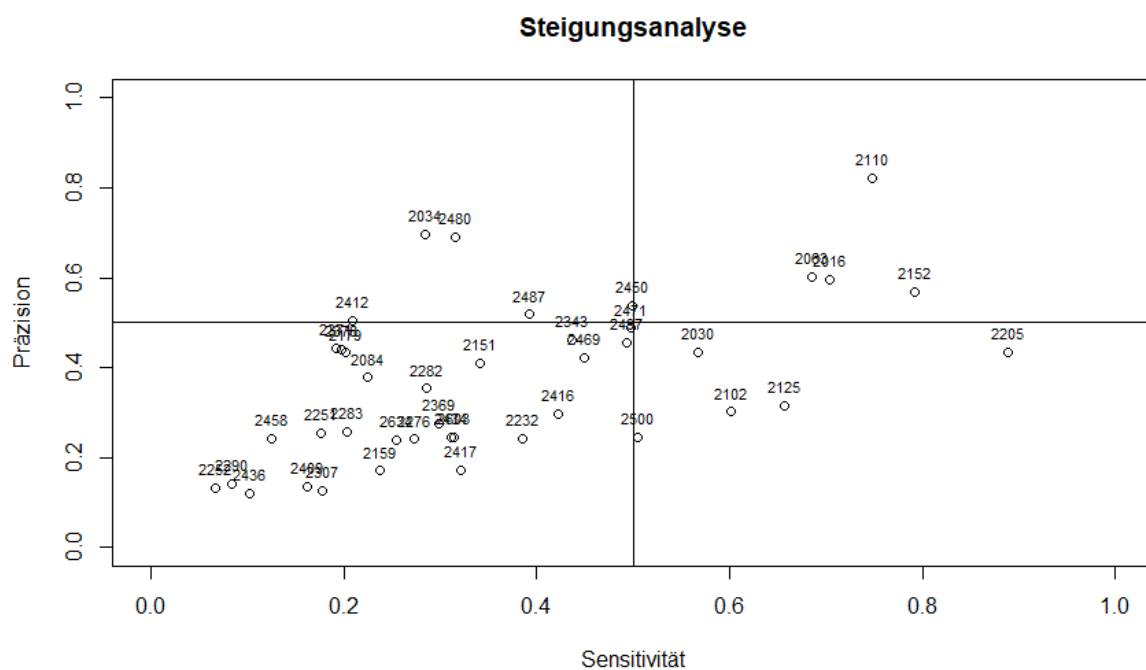


Abbildung 20: Gegenüberstellung der Sensitivität und Präzision der Steigungsanalyse mit Schwellenwert 0.75

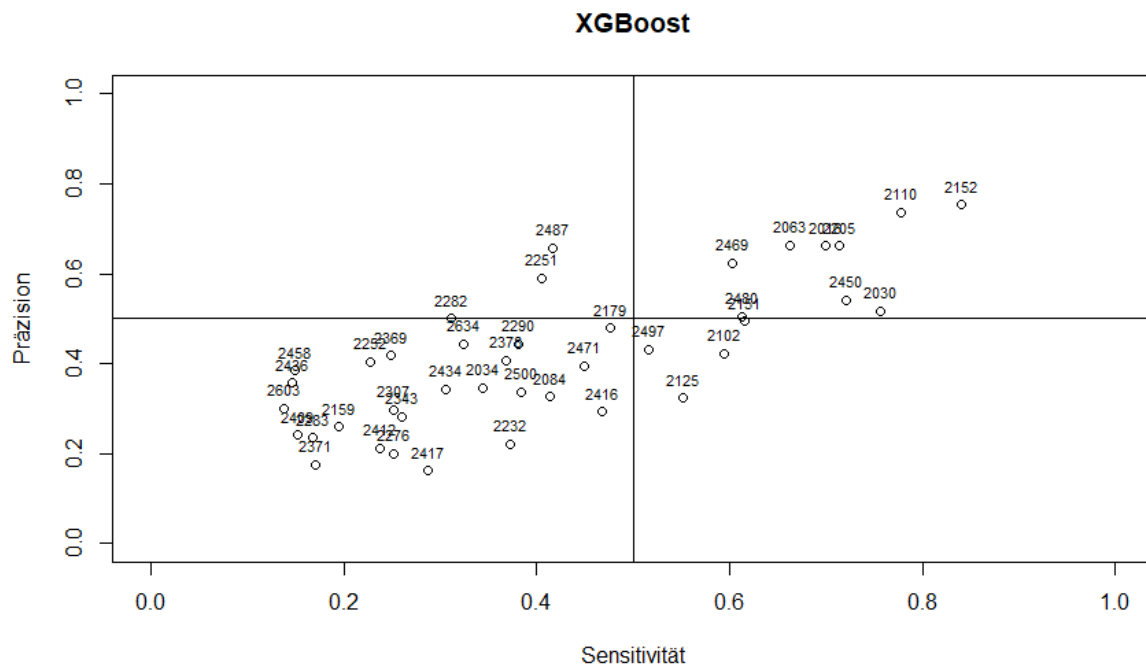


Abbildung 21: Gegenüberstellung der Sensitivität und Präzision der XGBoost basierten Methode mit σ -Faktor 1.25

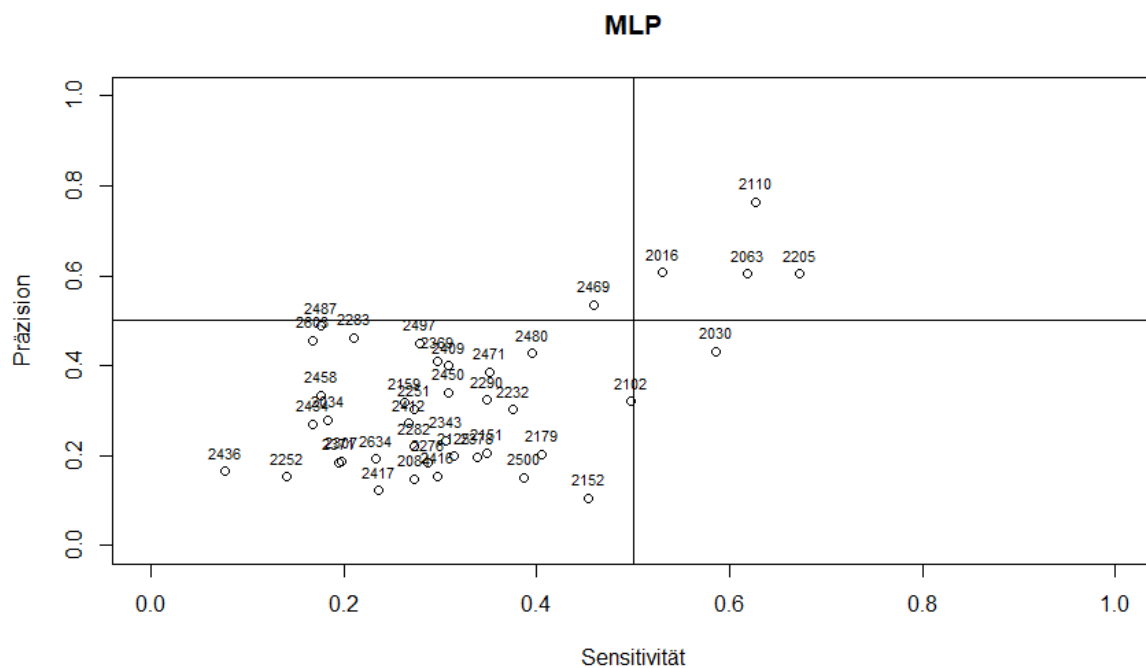


Abbildung 22: Gegenüberstellung der Sensitivität und Präzision der MLP basierten Methode mit σ -Faktor 1.25

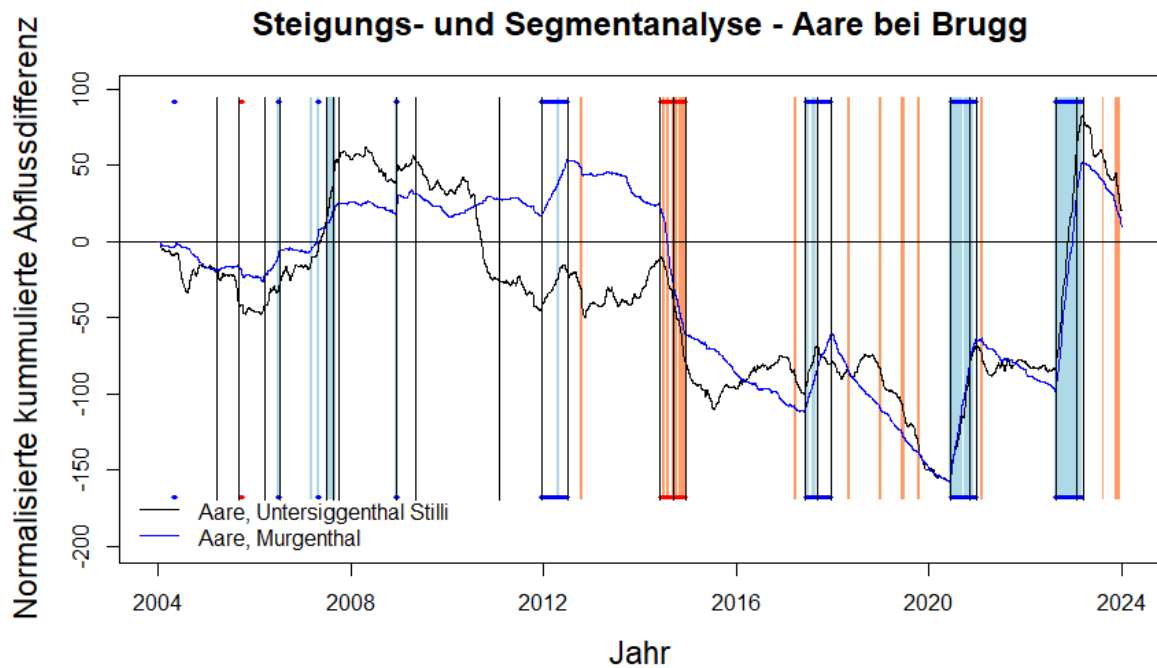


Abbildung 23: Beispiel der Changepointanalyse auf Basis der XGBoost Residuen für die Aare bei Brugg kombiniert mit der Steigungsanalyse. Die schwarzen vertikalen Linien markieren die Changepoints.

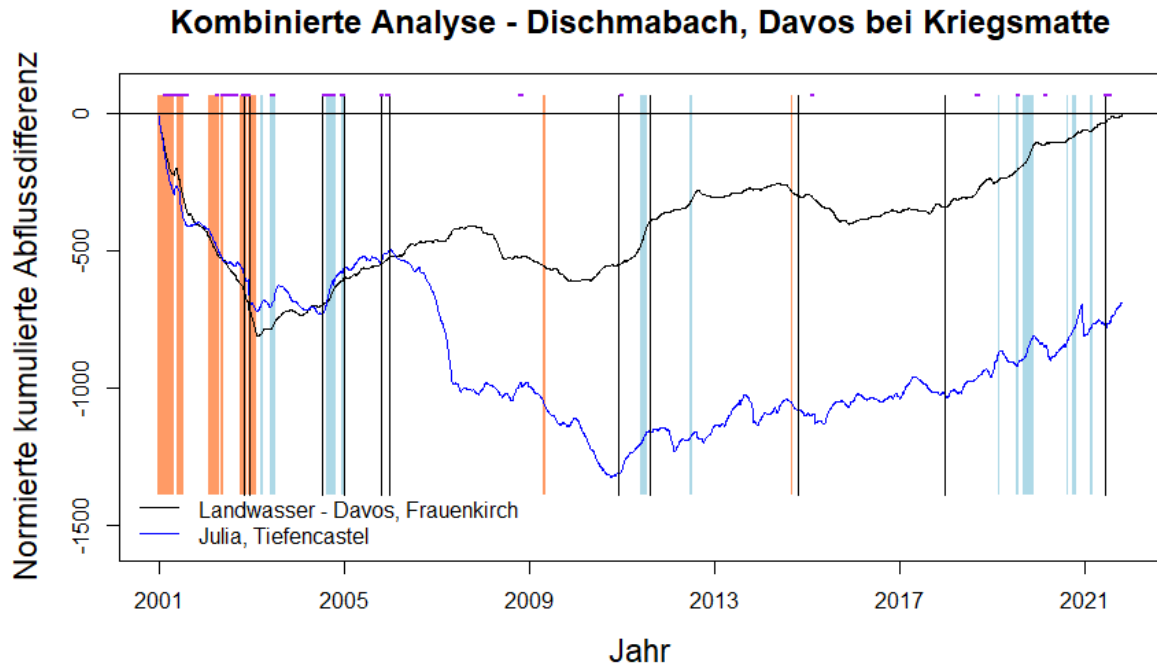


Abbildung 24: Anwendung aller Methoden auf ein Einzugsgebiet mit bekannten Problemen. Die Resultate der Steigungsanalyse sind als rote respektive blaue Segmente dargestellt. Die Ergebnisse der XGBoost basierten Methode erscheinen oben als violette Balken. Die Changepoints sind als vertikale schwarze Linien dargestellt.