



# Vorhersage von hydrologischen Abflusskennwerten in unbeobachteten Einzugsgebieten mit Machine Learning

Christoph Klingler · Moritz Feigl · Florian Borgwardt · Carina Seliger · Stefan Schmutz · Mathew Herrnegger

Angenommen: 21. Juli 2022 / Online publiziert: 29. August 2022  
 © Der/die Autor(en) 2022

**Zusammenfassung** Im Rahmen dieser Arbeit wurden Abflusskennwerte für Hochwasser (MJHQ), Mittelwasser (MQ) sowie Niederwasser (MJNQ, MJNQ<sub>7</sub>, Q<sub>95</sub>, Q<sub>98</sub>) für alle topografischen Einzugsgebiete der österreichischen Oberflächenwasserkörper inklusive der ausländischen hydrologischen Oberlieger-Regionen vorhergesagt. Die Regionalisierung der Abflusskennwerte wur-

de mit dem Machine-Learning-Modell XGBoost durchgeführt. Zentrale Grundlage für das Training von XGBoost war der LamaH-Datensatz, welcher für 859 beobachtete Einzugsgebiete in Zentraleuropa über 70 aggregierte Einzugsgebietseigenschaften und 15 meteorologische Zeitreihen umfasst. Anthropogene Beeinflussungen wie z.B. Jasspeicher oder Überleitungen wurden durch zusätzlich erstellte Attribute bei der Vorhersage berücksichtigt. Die Testergebnisse haben gezeigt, dass bei der Vorhersage der Abflusskennwerte in unbeobachteten Einzugsgebieten mit einer Abweichung von rund 20 % zu rechnen ist, wobei diese Schätzung auch stark anthropogen beeinflusste Gebiete beinhaltet. Darüber hinaus wurden auch 90-%-Konfidenzintervalle der Vorhersagen mit einem Quantile-Random-Forest-Modell geschätzt und klassifiziert. Die Ergebnisse werden der Öffentlichkeit in Form von Shapefiles unter <https://doi.org/10.5281/zenodo.6523372> kostenlos zur Verfügung gestellt.

**Schlüsselwörter** Abflusskennwerte · Vorhersage · Unbeobachtete Einzugsgebiete · Machine Learning · Unsicherheitsabschätzung · Zentraleuropa

## Prediction of runoff characteristics in ungauged basins with machine learning

**Abstract** This study comprises the prediction of runoff characteristics for high water (MJHQ), mean water (MQ), and low water (MJNQ, MJNQ<sub>7</sub>, Q<sub>95</sub>, Q<sub>98</sub>) for all topographic catchments of the Austrian surface water bodies including the foreign hydrological upstream regions. The machine learning model XGBoost was applied for the regionalization of the six runoff characteristics. The LamaH dataset was used for training XGBoost, which includes over 70 aggregated catchment characteristics and 15 meteorological time series for 859 observed catchments in Central Europe. Anthropogenic influences such as reservoirs or cross-basin water transfers were considered in the model by additionally created attributes. The test results showed that a deviation of approximately 20% can be expected for the prediction of runoff characteristics in ungauged catchments, which also includes highly anthropogenically influenced catchments. Furthermore, the 90% confidence interval of each prediction was estimated and classified using a Quantile Random Forest model. The results are provided free of charge to the public in form of shapefiles at <https://doi.org/10.5281/zenodo.6523372>

**Keywords** Runoff characteristics · Prediction · Ungauged basins · Machine learning · Uncertainty estimation · Central Europe

## 1 Einleitung

Hydrologische Abflusskennwerte beschreiben die Charakteristik eines Einzugsgebiets (EZG) und bilden die Basis

Die Autoren C. Klingler und M. Feigl trugen zu gleichen Teilen zur Genese dieser Arbeit bei.

**Datenverfügbarkeit** Es werden zwei Shapefiles unter <https://doi.org/10.5281/zenodo.6523372> kostenlos und barrierefrei der Öffentlichkeit zur Verfügung gestellt: 1) „LamaH\_observations“ enthält die sechs aus den beobachteten Zeitreihen berechneten Abflusskennwerte für 859 EZG des LamaH-Datensatzes. 2) In „OWK\_predictions“ sind neben den sechs vorhergesagten Abflusskennwerten auch jeweils das 5. und 95. Perzentil sowie die daraus resultierende Unsicherheitsklasse für die OWK (9533 EZG) enthalten. Nähere Informationen zu den Attributen der Shapefiles sind in den beiliegenden Metadaten vorhanden. Darüber hinaus können auch für das Projekt aufbereitete Datengrundlagen (aggregierte Einzugsgebietseigenschaften) auf Anfrage zur Verfügung gestellt werden.

**Anmerkungen** Wir haben mit größtmöglicher Sorgfalt sowie nach aktuellem Stand der Datenaufbereitung und -prozessierung gearbeitet und die Outputs laufend auf Plausibilität überprüft. Haftung für die bereitgestellten Daten wird aber weder von uns noch von den Providern der Datengrundlagen übernommen. Die Verwendung der zur Verfügung gestellten Daten erfolgt daher auf eigenes Risiko.

DI C. Klingler (✉) · DI Dr. M. Feigl · DI Dr. M. Herrnegger  
 Department für Wasser-Atmosphäre-Umwelt, Institut für Hydrologie und Wasserwirtschaft, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Wien, Österreich  
[christoph.klingler@boku.ac.at](mailto:christoph.klingler@boku.ac.at)

DI C. Klingler · DI Dr. M. Feigl  
 baseflow AI solutions GmbH,  
 1070 Wien, Österreich

DI Dr. F. Borgwardt · DDI Dr. C. Seliger · Univ.-Prof. DI Dr. S. Schmutz  
 Department für Wasser-Atmosphäre-Umwelt, Institut für Hydrobiologie und Gewässermanagement, Universität für Bodenkultur Wien, Gregor-Mendel-Straße 33, 1180 Wien, Österreich

für zahlreiche wasserwirtschaftliche und wissenschaftliche Anwendungen (Olden und Poff 2003). Neben der Planung und Bewirtschaftung von Wasserressourcen oder der Dimensionierung von Infrastruktur und Schutzbauten stellen Abflusskennwerte auch für die Einzugsgebietsklassifizierung (Sawicz et al. 2011; Wagener et al. 2007), Erkennung von zeitlichen Änderungen in EZG (Juston et al. 2014; Lebedzinski und Fürst 2018; Sawicz et al. 2014), Modellvalidierung (Hrachowitz et al. 2014; Refsgaard und Knudsen 1996; Wesemann et al. 2018a, b) sowie die Diagnose der Modellstruktur (Coxon et al. 2014; Gupta et al. 2008; McMillan et al. 2011) eine essenzielle Informationsgrundlage dar.

Die Abflusskennwerte können dabei in jene für die Beschreibung der Abflussverteilung (z.B. MQ,  $Q_{95}$ ) sowie jene für die Abflussdynamik (z.B. Steigung der Abflussdauerlinie) eingeteilt werden. Die Ermittlung dieser Abflusskennwerte erfolgt standardmäßig aus einer beobachteten Abflusszeitreihe, womit die direkte Berechnung auf beobachtete EZG beschränkt ist. Da Informationen über hydrologische Abflusskennwerte aber oftmals an unbeobachteten Standorten (kein Abflusspegel) notwendig sind, besteht die Notwendigkeit einer Regionalisierung (Blöschl et al. 2013; Hrachowitz et al. 2013) – also der Transfer von lokal beobachteten Werten in unbeobachtete Gebiete, z.B. mit Hilfe von Prädiktoren wie Einzugsgebietseigenschaften.

Für die Regionalisierung von Abflusskennwerten wurden neben konzeptionellen bzw. prozessbasierten hydrologischen Modellen (Biondi und de Luca 2017; Donnelly et al. 2016; Westerberg et al. 2014; Zhang et al. 2014) auch lineare Regression unter anderem mit Clustering (Grandry et al. 2013; Laaha und Blöschl 2007; Nathan und McMahon 1992; Qamar et al. 2016; Shu und Ouarda 2012; Visessri und McIntyre 2016; Zhang et al. 2014, 2018), lineare Regression mit Ähnlichkeits- bzw. Pooling-Ansätzen (Burn 1990; Hannaford et al. 2013; Holmes et al. 2002; Kjeldsen et al. 2014; Oppel und Schumann 2020), oder geostatistische Verfahren (Blöschl et al. 2022; Pugliese et al. 2014; Viglione et al. 2013) verwendet. Zhang et al. (2018) verwendeten Regressionsbäume und konnten damit eine Steigerung der Prognosequalität gegenüber hydrologischen Modellen und der multiplen (log-normierten) linearen Regression

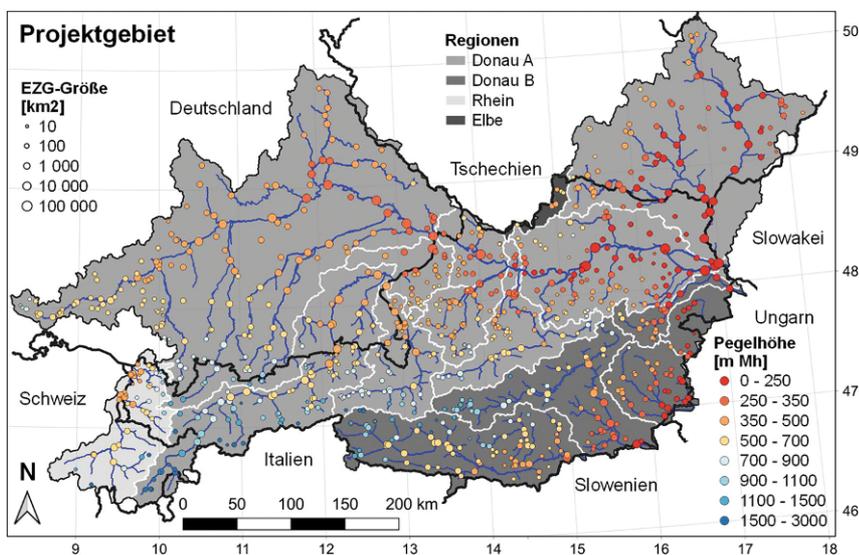
erzielen. Prieto et al. (2019) regionalisierten Abflusskennwerte ebenfalls mithilfe von Regressionsbäumen, wobei die Abflusskennwerte in weiterer Folge bei der Parametrisierung eines hydrologischen Modells einbezogen wurden und so zu einer Steigerung der Modellgüte in unbeobachteten EZG führten. Grundsätzlich sind zur Vorhersage von Abflusscharakteristika (nicht-)lineare Regressionsansätze meist besser als hydrologische Modelle geeignet, da Regressionsansätze ausschließlich für die Vorhersage eines einzelnen Abflusskennwerts trainiert werden, während hydrologische Modelle die gesamte kontinuierliche Abflusssituation abbilden (Zhang et al. 2018).

Unsicherheiten stellen bei jeder (hydrologischen) Vorhersage ein intrinsisches Element dar, welches für die Plausibilisierung der Modellergebnisse auch quantifiziert werden sollte. Bei der Regionalisierung von Abflusscharakteristika sind sowohl die beobachteten Zeitreihen, und dadurch auch die Zielgrößen, die Regressoren bzw. Prädiktoren genauso wie das Regionalisierungsverfahren selbst mit Unsicherheiten behaftet. Westerberg et al. (2016) haben die Unsicherheiten der beobachteten Abflusszeitreihen mittels eines Monte-Carlo-Stichprobenverfahrens abgeschätzt und die Unsicherheiten der Regionalisierung selbst mit einem gewichteten Gruppenansatz berücksichtigt. Die Ergebnisse zeigen, dass bei Nichtberücksichtigung der Unsicherheiten in den Abflusszeitreihen das Risiko einer Verzerrung der Regionalisierung besteht. Westerberg et al. (2016) zeigten weiters, dass die Unsicherheiten bei den Abflusskennwerten zur Abflussverteilung geringer sind als bei jenen zur Abflussdynamik, und dass bei Abflusskennwerten für Mittelwasser geringere Unsicherheiten zu erwarten sind als bei jenen für Hoch- oder Niederwasser. Yadav et al. (2007) haben durch regionalisierte Abflusskennwerte die Grenzwerte der festzulegenden Parameter von hydrologischen Modellen in unbeobachteten Einzugsgebieten einschränken können. Klima-, Topografie- wie auch hydrogeologische Einzugsgebietseigenschaften waren bei der Regionalisierung der Abflusskennwerte die maßgebenden Prädiktoren. Zudem konnte festgestellt werden, dass die Güte der Regionalisierung stark von den einzelnen Abflusskennwerten abhängig ist. Laaha und Blöschl (2007) berücksichtigen die Unsicherheiten bei

der österreichweiten Niederwasserabschätzung für  $Q_{95}$  durch die Ausgabe eines Konfidenzintervalls (Schätzwert  $\pm$  Regressionsstandardfehler) anstelle eines Vorhersagewerts. Poggio et al. (2021) führten eine Regionalisierung inklusive Unsicherheitsabschätzung von Bodeneigenschaften auf globaler Ebene durch (SoilGrids 2.0). Dabei erfolgte die Vorhersage mit einem Quantile Random Forest (QRF; Meinhäusen 2006), wobei im Gegensatz zu einem standardmäßigen Random Forest nicht nur der Vorhersagewert, sondern auch Quantile bereitgestellt werden. Anhand dieser Quantile lässt sich in weiterer Folge eine robuste Aussage über die Modellunsicherheit des Regionalisierungsverfahrens treffen.

Im Rahmen des Forschungsprojekts „aquaZoom“ – finanziert durch den Europäischen Meeres- und Fischereifonds (EMFF) bzw. das Bundesministerium für Landwirtschaft, Regionen und Tourismus (BMLRT) – wird das Produktionspotenzial von Aquakultur-Durchflusssanlagen in ausgewählten Regionen detailliert abgeschätzt (siehe Artikel von Seliger et al. in der vorliegenden Ausgabe). Da für den Betrieb einer Durchflusssanlage die Verfügbarkeit von frischem Wasser einen sehr wichtigen Faktor darstellt, sind Abflusskennwerte eine entscheidende Planungs- und Betriebsgrundlage (Buchart 2012). Im Rahmen dieser Arbeit werden daher Abflusskennwerte für Hochwasser (MJHQ), Mittelwasser (MQ) sowie Niederwasser (MJNQ, MJNQ<sub>7</sub>,  $Q_{95}$ ,  $Q_{98}$ ) mithilfe von Machine Learning (ML) für die topografischen EZG der österreichischen Oberflächenwasserkörper (UBA 2015) quantifiziert, damit diese in der Potenzialabschätzung berücksichtigt werden können.

Während die in der Hydrologie oft verwendeten prozessbasierten Modelle eine Struktur und Parametrisierung aufweisen, die ein physikalisches System widerspiegeln soll, bestehen ML-Modelle aus flexiblen mathematischen Strukturen. ML-Modelle weisen initial keine physikalisch interpretierbare interne Strukturen auf, haben aber den Vorteil, dass sie durch ihre Flexibilität komplexe Zusammenhänge, Strukturen und Muster in Daten sehr gut abbilden können (Klingler et al. 2022). Für eine umfassendere Einführung in die Anwendung von ML in der Hydrologie möchten wir auf die Beiträge in der ÖWAW-Ausgabe 7-8/2021 (OEAWAW 2021) verweisen.



**Abb. 1** Übersicht des Projektgebiets inkl. (Makro-)Flussregionen sowie der 859 Abfluss-Pegel mit Pegelhöhe und Einzugsgebietsgröße aus dem LamaH-Datensatz (Klingler et al. 2021a, b, c). Die Staatsgrenzen sind durch die schwarzen Linien dargestellt. Die 18 verschiedenen Flussregionen werden durch weiße Linien abgegrenzt. Der Abfluss der Makro-Flussregion „Donau B“ mündet erst außerhalb Österreichs in die Donau. Gewässernetz: EU-Hydro – River Network Database (EEA 2019), Staatsgrenzen: ©EuroGeographics, Koordinaten: WGS84

Grundlage für diese Regionalisierung ist der LamaH-Datensatz, welcher für 859 beobachtete EZG in Zentraleuropa Abflusszeitreihen, über 70 Einzugsgebietseigenschaften und Zeitreihen für 15 verschiedene meteorologische Variablen enthält (Klingler et al. 2021a, b, c). Mit dieser außerordentlich umfassenden Datengrundlage erfolgt das Training des ML-Modells eXtreme Gradient Boosting (XGBoost; Chen und Guestrin 2016), wobei dies für alle Abflusskennwerte individuell erfolgt. Anschließend werden die gelernten Zusammenhänge auf die unbeobachteten OWK übertragen und schließlich die Unsicherheiten des Regionalisierungsverfahrens mit QRF quantifiziert.

Zusammenfassend ergibt sich die Originalität dieser Arbeit aus der Entwicklung einer robusten Methodik zur Vorhersage von sechs Abflusskennwerten in zahlreichen unbeobachteten EZG unter Quantifizierung der assoziierten Unsicherheiten. Die Ergebnisse werden der Öffentlichkeit in Form von Shapefiles kostenlos und barrierefrei zur Verfügung gestellt, damit auch weitere Forschungs- und Anwendergruppen Zugriff auf eine vollständig dokumentierte Datengrundlage haben. Der vorliegende Artikel ist als umfassende Dokumentation zu den bereitgestellten Shapefiles gedacht.

## 2 Datengrundlage und -aufbereitung

### 2.1 LamaH (Large-SaMple Data for Hydrology and Environmental Sciences for Central Europe)

Der LamaH-Datensatz umfasst für 859 EZG in Österreich sowie dessen ausländischen hydrologischen Oberlieger-Regionen (Abb. 1) ein Ensemble an Abflusszeitreihen, meteorologische Zeitreihen sowie über 70 verschiedene Einzugsgebietseigenschaften (Klingler et al. 2021a, b, c). Aus den täglichen Abflusszeitreihen (BAFU 2020; CHMI 2020; GKD 2020; HZB 2020; LUBW 2020) werden die sechs Abflusskennwerte (1) MJHQ (mittleres jährliches Hochwasser), (2) MQ (Mittelwasser), (3) MJNQ (mittleres jährliches Niederwasser), (4) MJNQ<sub>7</sub> (mittleres niedrigstes 7-Tages-Mittel), (5) Q<sub>95</sub> (Abfluss welcher an 95 % der Zeit überschritten wird) sowie (6) Q<sub>98</sub> (Abfluss welcher an 98 % der Zeit überschritten wird) für die hydrologischen Jahre 2003 bis 2017 (01.10.2002 bis 30.09.2017) als Zielgrößen für die Modellierung ermittelt.

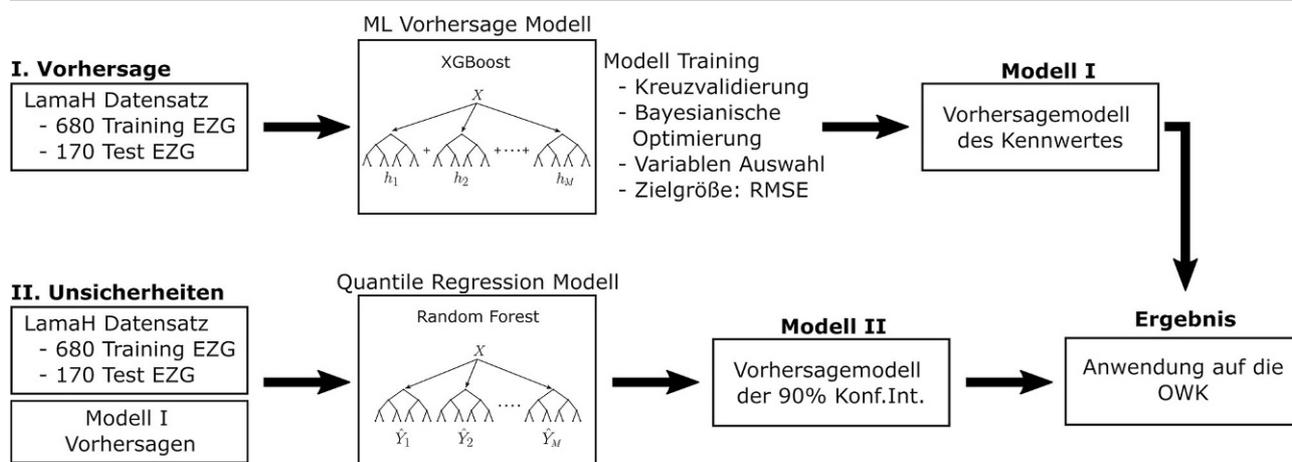
Die Vielzahl der bereits in LamaH einheitlich berechneten statischen Attribute der Kategorien Topografie, Klimatologie, Hydrologie, Landbedeckung, Vegetation, Boden, Geologie sowie an-

thropogene Beeinflussung stellen eine umfassende Charakterisierung der hydrologischen Eigenschaften der EZG dar. Für dieses Projekt wurde auch das akkumulierte (Nutz-)Volumen der verschiedenen Reservoir-Typen (z.B. Jahresspeicher) sowie die Einzugsgebietsflächenänderung durch anthropogene Wasserüberleitungen für alle EZG quantifiziert. Aus den ERA5-Land Zeitreihen (ERA5L 2020) berechneten wir darüber hinaus 30 meteorologische Kennwerte für die hydrologischen Jahre 2003 bis 2017 (Anhang A). Damit steht eine hohe Anzahl an Prädiktoren (90) für das Training der ML-Modelle zur Verfügung.

### 2.2 Oberflächenwasserkörper

Ein Oberflächenwasserkörper ist gemäß § 30a Abs. 3 WRG 1959 ein „*einheitlicher und bedeutender Abschnitt eines Oberflächengewässers*“ und stellt somit per definitionem einen homogenen Gewässerabschnitt dar. Die Oberflächenwasserkörper (Gesamtheit aller einzelnen Oberflächenwasserkörper) sind eine zentrale Grundlage für die wasserwirtschaftliche Kommunikation, Planung und Berichterlegung in Österreich (z.B. für die Zustandsbewertung nach der Wasserrahmenrichtlinie). Die Oberflächenwasserkörper umfassen zahlreiche Detailinformationen, wovon der Hauptteil im Nationalen Gewässerbewirtschaftungsplan (NGP) enthalten ist.

Die vorliegende Arbeit baut auf den rund 8000 Polygonen des Datensatzes OWK-Version NGP15 (UBA 2015) auf, welche die (Teil-)EZG der Oberflächenwasserkörper darstellen. Diese Polygone decken dabei ausschließlich österreichisches Staatsgebiet ab, wodurch wichtige ausländische hydrologische Oberlieger-Regionen (z.B. Inn, March, Donau) nicht inkludiert sind. Da die Abflusskennwerte für das gesamte topografische EZG zu ermitteln sind, würde bei der Berechnung der Prädiktoren (z.B. topografische Einzugsgebietsfläche) von EZG mit staatsübergreifender Ausdehnung fundamentale Information aus dem Ausland fehlen. Daher werden die österreichischen (Teil-)EZG der Oberflächenwasserkörper mit 1408 (Teil-)EZG aus den Datensätzen „digitaler Hydrologischer Atlas Österreich“ (BMLRT 2007) sowie HydroATLAS (Linke et al. 2019) zusammengeführt, welche die ausländischen hydrologischen Oberlieger-Regionen abdecken. Bei der Zusammenführung der (Teil-)EZG



**Abb. 2** Workflow der Modellentwicklung zur Vorhersage der Abflusskennwerte und der dazugehörigen Abschätzung der Unsicherheiten. Der Workflow stellt die Arbeitsschritte zur Erstellung beider Modelle für einen Abflusskennwert dar. Modell I und II werden anhand der LamaH-Daten optimiert und im letzten Schritt zur Vorhersage des jeweiligen Abflusskennwertes der OWK verwendet

werden, wo notwendig (z.B. an den Grenzflächen der unterschiedlichen Datensätze), auch die Attribute für das Routing (upstream-downstream-Beziehung) angepasst. Darauf aufbauend erfolgt die Aggregation der (Teil-)EZG der verschiedenen Datensätze zu topografischen EZG – deshalb spielt auch die Auflösung der unterschiedlichen Datensätze keine Rolle. Im weiteren Textverlauf ist zwecks einer vereinfachten Kommunikation unter „OWK“ die Gesamtheit der topografischen EZG aller österreichischen Oberflächenwasserkörper sowie aller EZG der ausländischen Oberlieger-Regionen zu verstehen. Anschließend erfolgt für die OWK die Berechnung jener Attribute, welche beim Training des ML-Modells als Prädiktoren herangezogen werden. Dabei werden dieselben Datengrundlagen und Algorithmen wie bei der Berechnung der LamaH-Einzugsgebietseigenschaften herangezogen. Dadurch wird Konsistenz zwischen Trainings- (LamaH) und Vorhersagedatensatz (OWK) bewahrt.

### 3 Methoden

#### 3.1 Modellierungs-Workflow

In diesem Abschnitt wird der individuell für jeden Kennwert angewandte Workflow beschrieben (Abb. 2), während Details zu den verwendenden ML-Modellen in den folgenden Abschnitten erläutert sind. Der Workflow unterteilt sich in zwei Modellanwendungen: I. Vorhersage und II. Unsicherheiten. Bei I. wird ein ML-Modell mit den be-

obachteten EZG aus LamaH für die Vorhersage des jeweiligen Kennwertes trainiert. Darauf folgend wird bei II. ein Quantil-Regressions-Modell mit den Vorhersagen von Modell I sowie den EZG aus LamaH trainiert, um dadurch die 5%- und 95%-Quantile der Vorhersagen zu quantifizieren. Schlussendlich werden beide Modelle auf die unbeobachteten OWK angewendet, um den jeweiligen Kennwert und das dazugehörige 90%-Konfidenzintervall vorherzusagen.

Vor dem Training werden aus den Abflusskennwerten Abflusspenden berechnet, um dieses unabhängig von der Einzugsgebietsgröße durchführen zu können. Die EZG aus LamaH werden in einen Trainings- (680 EZG) und Testdatensatz (170 EZG) unterteilt. Die Testdaten kommen dabei nicht beim Training zur Anwendung, sondern werden ausschließlich zur Quantifizierung der Modellgüte in Gebieten außerhalb der Trainingsgebiete herangezogen. Damit diese Testgebiete eine repräsentative Stichprobe darstellen, werden diese basierend auf einem stratifizierten Sampling zufällig ausgewählt. Dabei werden die EZG von LamaH basierend auf dem jeweiligen Kennwert in zehn Gruppen mit ähnlichen Werten eingeteilt und daraus jeweils 17 zufällig als Test-EZG ausgewählt. Nach Quantifizierung der Modellgüte anhand der unabhängigen Testdaten (Abschn. 4.1) werden die Modelle mit allen EZG aus LamaH trainiert (850 EZG). Schließlich erfolgt die Anwendung der trainierten Modelle auf die unbeobachteten OWK.

#### 3.2 ML-Modell zur Vorhersage der Abflusskennwerte

Durch initiale Versuche, bei denen mehrere ML-Modelltypen miteinander verglichen wurden, konnten wir feststellen, dass sich eXtreme Gradient Boosting (XGBoost; Chen und Guestrin 2016) gut zur Vorhersage der Abflusskennwerte eignet. XGBoost ist ein auf Regressionsbäumen basiertes Ensemble-Modell. Dabei werden die einzelnen Regressionsbäume sukzessive trainiert, um die Residuen des Ensembles der jeweils vorhergehenden Regressionsbäume vorherzusagen und somit zu minimieren. Jeder Regressionsbaum erhält nur eine zufällig ausgewählte Teilmenge an Prädiktoren. Die Wichtigkeit einzelner Prädiktoren kann dabei durch die Ermittlung der durchschnittlichen Verbesserung der Vorhersage durch die Verwendung dieses Prädiktors abgeschätzt werden. Für eine genauere Beschreibung verweisen wir auf Feigl et al. (2021), wo eine Reihe an ML-Verfahren zur Vorhersage von Fließgewässertemperaturen in Österreich angewendet und getestet wurde.

XGBoost besitzt mehrere Hyperparameter (d.h. Parameter, welche nicht mithilfe der Daten automatisch angepasst werden), die vor dem eigentlichen Training festzulegen sind. Da diese eine große Auswirkung auf die Modellgüte haben können (Feigl et al. 2021) wird eine Bayes'sche Optimierung (Močkus et al. 1978; Snoek et al. 2012) zur Festlegung angewandt.

Die Bayes'sche Optimierung basiert auf Schätzung der Posterior-Verteilung

der Modellgüte mittels eines Gauß-Prozesses. Anhand dieser Schätzung können die Hyperparameter bei jeder Iteration mit dem Ziel einer Maximierung der Modellgüte festgelegt werden. Anhang B beinhaltet eine Auflistung der festgelegten Bandbreite der einzelnen Hyperparameter sowie die Anzahl der Iterationen und weitere Spezifikationen bei der Optimierung von XGBoost. Zusätzlich zur Festlegung der Hyperparameter wird eine Auswahl an Prädiktoren durchgeführt, um eine mögliche Überanpassung an die Trainingsdaten auszuschließen. Dafür wird zuerst jeweils ein Modell mit allen Prädiktoren (Anhang A) trainiert und anschließend in drei unabhängigen Versuchsläufen die Prädiktorenauswahl verringert. Dabei verbleiben beim entsprechenden Modelltraining jene Prädiktoren in der Auswahl, welche mind. 1%, 5% oder 10% der maximalen Feature Importance (Wichtigkeit eines Prädiktors) beim Versuch mit allen Prädiktoren aufgewiesen haben. Zur Anwendung kommt schlussendlich jene reduzierte Prädiktorenauswahl, welche die höchste Kreuzvalidierungs-Modellgüte zur Folge hat. Als Zielfunktion beim Training sowie zur Quantifizierung der Modellgüte wird der RMSE (Wurzel des mittleren quadratischen Abstandes, Gl. 1) verwendet:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

Dabei stellt  $n$  die Stichprobengröße (Anzahl der EZG),  $y_i$  den vorhergesagten sowie  $\hat{y}_i$  den aus den beobachteten Zeitreihen berechneten Abflusskennwert des EZG  $i$  dar. Da es beim Training auch zu Überanpassungen kommen kann, wird die Modellgüte mittels einer 10-fachen Kreuzvalidierung (CV) mit 5-facher Wiederholung geschätzt.

### 3.3 Machine-Learning-Modell zur Quantifizierung der Unsicherheit

Um die Unsicherheiten abschätzen zu können, wird eine Quantilsregression mit Random Forest durchgeführt (Breiman 2001; Meinshausen 2006). Random Forest (RF; Breiman 2001) ist wie XGBoost ebenfalls ein auf Regressionsbäumen basiertes Klassifikations- und Regressionsverfahren. Allerdings erfolgt das Training der einzelnen Regressionsbäume im Gegensatz zu XGBoost nicht sukzessive, sondern unabhängig und parallel. Dadurch ist meist ein deutlich

schnelleres Training möglich. Für eine genauere Beschreibung von RF sei an dieser Stelle wieder auf Feigl et al. (2021) verwiesen. Bei einem Quantile Random Forest (QRF; Meinshausen 2006) werden neben dem Mittel der Ausgabe- werte der einzelnen Regressionsbäume (Vorhersagewert eines RF) auch Quantile ausgegeben. Anhand dieser Quantile lässt sich die aus dem Regionalisierungsverfahren bedingte Unsicherheit abschätzen. Eine Aussage über die aus den Abflusszeitreihen sowie den Prädiktoren resultierenden Unsicherheiten ist damit jedoch nicht möglich. Der QRF verwendet dieselben Prädiktoren wie das XGBoost-Modell und darüber hinaus auch dessen Vorhersagewert. Dadurch kann die Information vom Vorhersagemodell verwendet und zugleich eine empirische Verteilung für diese geschätzt werden.

### 3.4 Klassifizierung der Unsicherheiten

Die Klassifizierung der Unsicherheiten erfolgt durch Berechnung der relativen Abweichung zum Vorhersagewert nach Gl. 2:

$$p_{\text{diff}} = \max \left[ \frac{Q_{ub} - Q_{\text{pred}}}{Q_{\text{pred}}}, \frac{Q_{\text{pred}} - Q_{lb}}{Q_{\text{pred}}} \right] \cdot 100 \quad (2)$$

Dabei stellt  $Q_{\text{pred}}$  den vorhergesagten Abflusskennwert,  $Q_{lb}$  das 5%-Quantil der Wahrscheinlichkeitsverteilung des QRF und  $Q_{ub}$  das 95%-Quantil dar. Der Unsicherheitsgrad „sehr gering“ wird zugeteilt, wenn  $p_{\text{diff}}$  kleiner oder gleich 25% beträgt. Ein „geringer“ Grad der Unsicherheit liegt vor, wenn  $p_{\text{diff}}$  größer als 25% aber kleiner oder gleich 50% ist. Ein „mittlerer“ Grad wird zugeteilt, falls  $p_{\text{diff}}$  größer als 50% und kleiner oder gleich 100% ist und ein hoher Unsicherheitsgrad ist bei über 100% vorliegend. Falls der vorhergesagte Abflusskennwert die Schwelle von  $101 \text{ s}^{-1}$  unterschreitet, erfolgt aufgrund der potenziell hohen relativen Abweichungen keine Zuweisung des Unsicherheitsgrads, sondern der Hinweis durch ein entsprechendes Attribut.

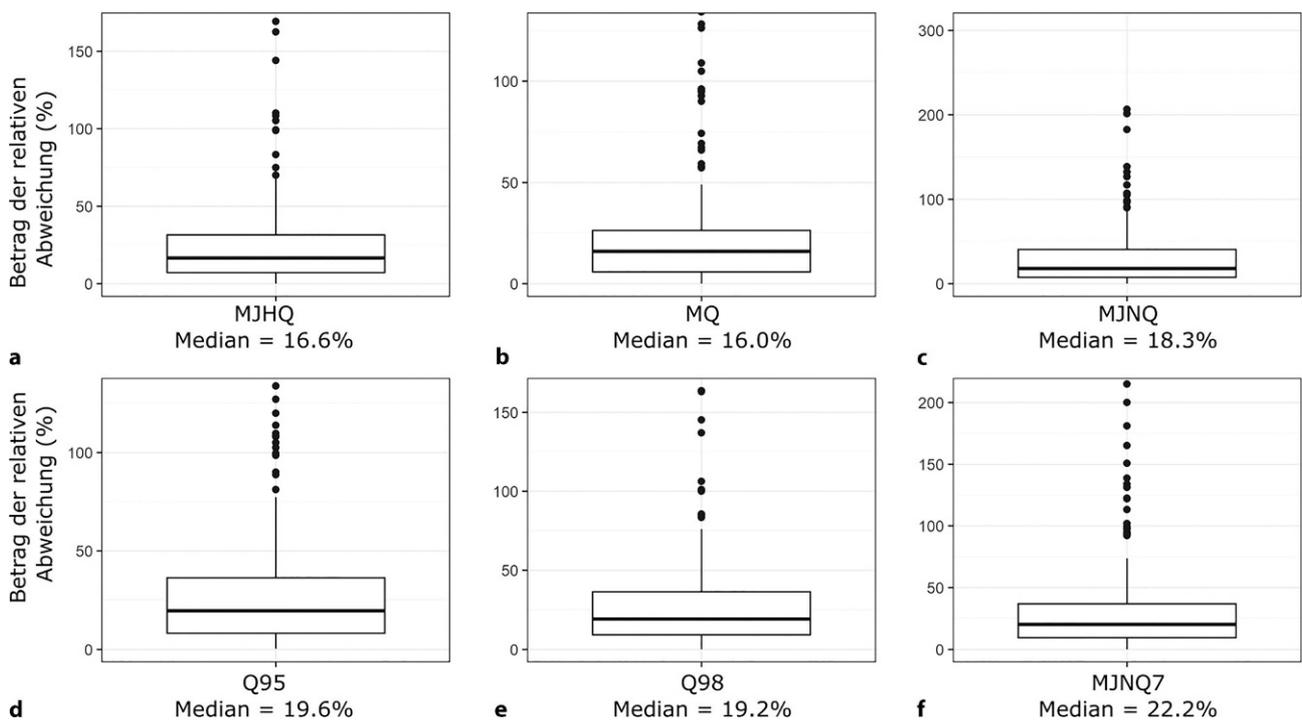
## 4 Ergebnisse

### 4.1 Modellgüte der Kennwertvorhersage

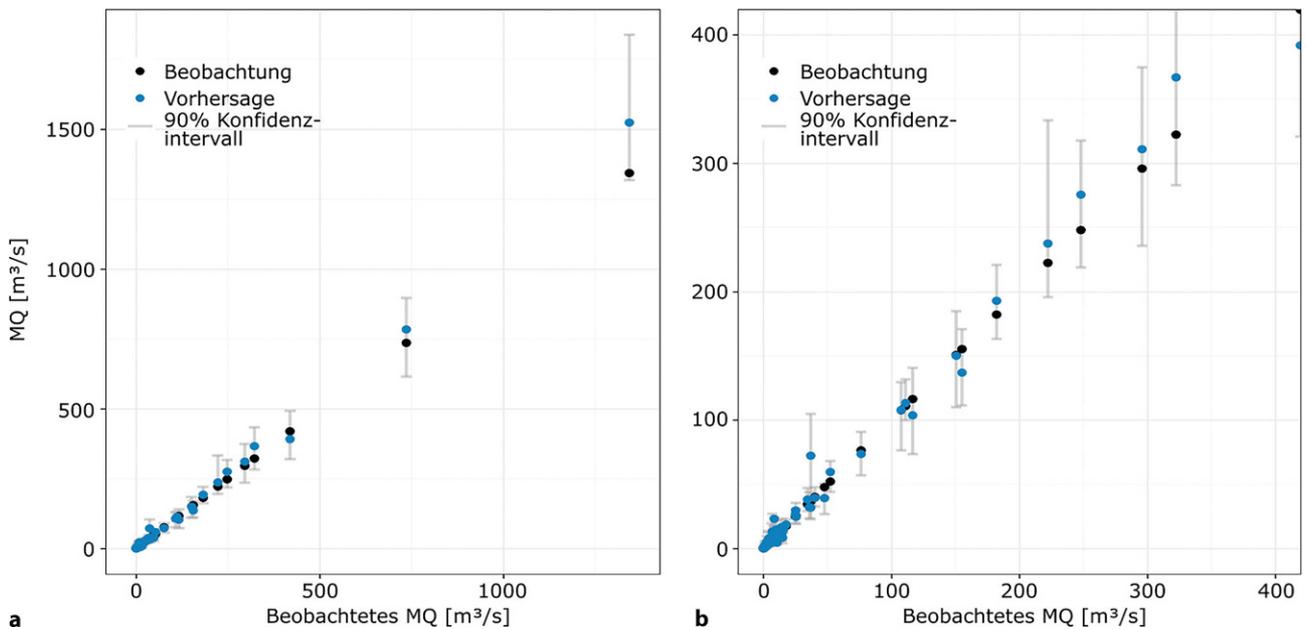
Die Modellgüte bei der Regionalisierung der Abflusskennwerte kann anhand der Ergebnisse in den Test-EZG

abgeschätzt werden. Abb. 3 zeigt dabei die Boxplots des Betrags der relativen Abweichung ( $|\text{Vorhersage-Beobachtung}| \cdot 100 / \text{Beobachtung}$ ) für die sechs genannten Abflusskennwerte. Die Darstellung der Modellgüte mittels der relativen Abweichungen erlaubt eine Abschätzung der Modellgüte unabhängig von der Einzugsgebietsgröße. Gleichzeitig ist zu beachten, dass sich bei sehr geringen Abflüssen (z. B.  $0,1 \text{ m}^3 \text{ s}^{-1}$ ) potenziell sehr große prozentuelle Abweichungen bei relativ geringen absoluten Abweichungen ergeben können (z. B. ein Vorhersagewert von  $0,2 \text{ m}^3 \text{ s}^{-1}$  entspricht dort einer relativen Abweichung von 100%). Dies trifft auf die meisten Ausreißer in den dargestellten Boxplots zu. Mit Medianwerten von 16 und 16,6% weist die Vorhersage des MQ bzw. MJHQ die geringsten relativen Abweichungen auf. Die Modelle zur Vorhersage der Niederwasserkennwerte weisen im Vergleich dazu mit Medianwerten von 18,3 bis 22,2% eine geringere Modellgüte auf, was primär auf die geringeren absoluten Abflusskennwerte zurückzuführen ist. Zusammenfassend zeigen die Ergebnisse, dass mit einer medianen relativen Abweichung von rund 20% bei der Vorhersage der Abflusskennwerte in unbeobachteten EZG zu rechnen ist.

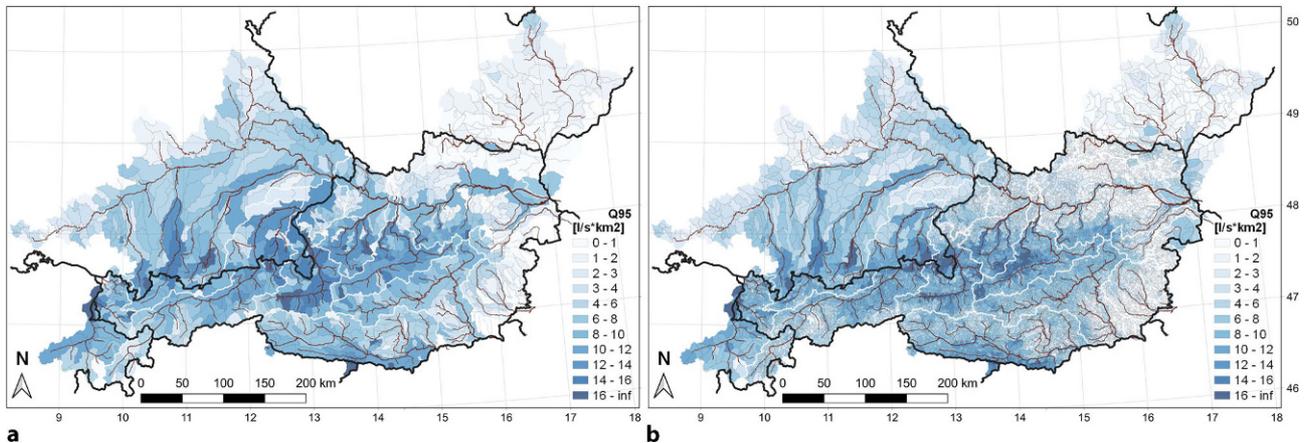
Zusätzlich wurde auch die Wichtigkeit der Prädiktoren der Modelle überprüft, um einen besseren Einblick in die Genese der Vorhersagen zu gewinnen. Eine Übersicht der jeweils fünf wichtigsten Prädiktoren für die sechs Modelle zur Vorhersage der einzelnen Abflusskennwerte ist in Anhang C aufgelistet – wobei die Auflistung dort unser hydrologisches Grundverständnis gut widerspiegelt. Prädiktoren mit einem hohen Einfluss auf die Vorhersage sind vor allem Niederschlagskenngrößen, Information, die Karstgebiete indizieren kann (Anteil Karbonatgestein), Topografie (Fläche, Flussdichte), Einfluss von Überleitungen und zusätzliche hydro-meteorologische Informationen (Temperatur, Schneewasseräquivalent, Evapotranspiration). Interessant ist bei der Auflistung in Anhang C, dass beim Modell zur Vorhersage des Hochwasserkennwerts MJHQ der Prädiktor MPmin (minimales Monatsmittel des Niederschlags) am zweithöchsten gewichtet wurde, wohingegen bei allen Vorhersagemodellen für die vier Niederwasserkennwerte der Prädiktor MPmax (maximales Monatsmittel des Niederschlags) unter den fünf wichtigsten Prädiktoren



**Abb. 3** Betrag der relativen Abweichung der Vorhersagewerte in den LamaH-Test-EZG ( $n = 170$ ) zu den Abflusskennwerten, welche aus den beobachteten Zeitreihen berechnet wurden ( $|(Vorhersage - Beobachtung)| * 100 / Beobachtung$ ). LamaH-Testergebnisse für die Abflusskennwerte zu Hochwasser **a** MJHQ, Mittelwasser **b** MQ sowie Niederwasser **c** MJNQ, **d** Q<sub>95</sub>, **e** Q<sub>98</sub> und **f** MJNQ<sub>7</sub>, wobei die Ergebnisse bis zum 98-%-Quantil dargestellt sind



**Abb. 4** Darstellung des beobachteten und vorhergesagten Kennwerts MQ der einzelnen Test-EZG unter Angabe des 90-%-Konfidenzintervalls vom QRF. Bei **a** werden alle 170 Testdaten dargestellt, bei **b** hingegen nur der Ausschnitt bis zu einem beobachteten MQ von  $400 m^3 s^{-1}$



**Abb. 5** Gegenüberstellung von **a** Beobachtung an den Pegeln und **b** OWK-Vorhersagen für die  $Q_{95}$ -Abflusspenden. Flussregionen (weiße Linien): LamaH, Gewässernetz (braun): EU-Hydro – River Network Database (EEA 2019), Staatsgrenzen (schwarz): ©EuroGeographics, Koordinaten: WGS84

ist. Diese Reihung bzgl. MPmax und MPmin ist aus hydrologischer Sicht auf den ersten Blick kontraintuitiv. Eine mathematisch/statistische Erklärung dafür ist mitunter die höhere Pearson-Korrelation zwischen MPmax und den beobachteten Niederwasserspenden ( $R_{\text{mittel}}=0,66$ ), welche durchgehend höher ist als zwischen MPmin und den Niederwasserspenden ( $R_{\text{mittel}}=0,62$ ). Bei den beobachteten Hochwasserabflusspenden ist die Korrelation mit MPmin ( $R=0,61$ ) hingegen höher als mit MPmax ( $R=0,56$ ). Zudem muss beachtet werden, dass MPmin bzw. MPmax monatliche, die Abflusskennwerte hingegen tägliche Mittelwerte darstellen. Der abflussreichste Tag muss nicht im niederschlagsreichsten Monat liegen und vice versa. Hinsichtlich Anhang C ist schließlich anzumerken, dass sich die Reihenfolge bei einem abweichenden Modellierungs-Setting (z. B. Hyperparameter, Vorauswahl der Prädiktoren) auch (geringfügig) ändern kann.

Des Weiteren wurde auch die Güte der Unsicherheitsabschätzung in den Test-EZG überprüft. Der Übersichtlichkeit halber werden hier nur die Ergebnisse der Vorhersage von MQ dargestellt, welche repräsentativ für die Ergebnisse aller Abflusskennwerte sind. Abb. 4 beinhaltet die Darstellung der beobachteten und vorhergesagten MQ-Werte samt der geschätzten 90%-Konfidenzintervalle. Abb. 4a zeigt die Ergebnisse aller Test-EZG, während in Abb. 4b nur jene EZG mit einem beobachteten MQ bis  $400 \text{ m}^3 \text{ s}^{-1}$  dargestellt werden. Die Ergebnisse in Abb. 4 zeigen, dass die Vorhersage im Test meist recht gut mit der Beobachtung über-

einstimmt. Einzelne Beobachtungen (z. B. Gebiet mit dem höchsten MQ in Abb. 4a) weisen höhere Abweichungen zur Vorhersage auf, liegen aber noch innerhalb des 90-%-Konfidenzintervalls.

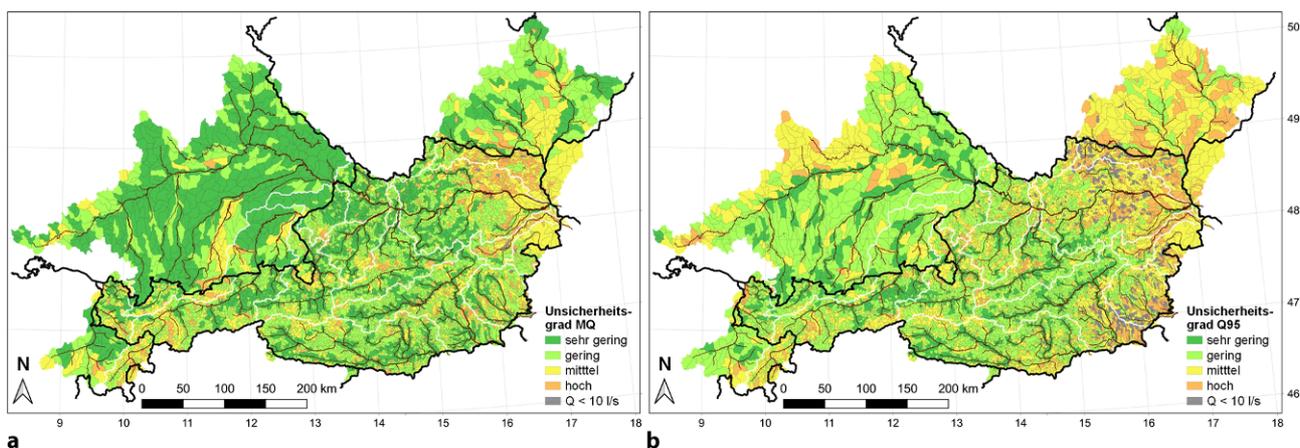
#### 4.2 Vorhersage der Abflusskennwerte

Da alle OWK unbeobachtet sind, kann keine direkte Aussage über die relativen Abweichungen zu den Beobachtungen wie in Abschn. 4.1 getroffen werden. In Abb. 5a werden daher die an den Pegeln beobachteten Abflusspenden des Kennwerts  $Q_{95}$  – welche zwecks einer anschaulicheren räumlichen Darstellung auf die zugehörigen (Teil-)EZG umgelegt sind – den in Abb. 5b dargestellten Vorhersagen für die OWK gegenübergestellt. Zu beachten ist dabei, dass die OWK mit 9533 EZG in etwa 10-mal so viele EZG wie der LamaH-Datensatz beinhalten, wodurch sich in Abb. 5b ein kontinuierlicheres Gesamtbild ergibt. Aufgrund der unterschiedlichen Polygongröße und -einteilung gestaltet sich die direkte optische Gegenüberstellung teilweise trügerisch. Ein gutes Beispiel dazu ist die Darstellung der Donau im Bereich von Wien (ca.  $16^\circ \text{E}/48^\circ \text{N}$ ). Die Donau-Teil-EZG reichen im LamaH-Datensatz in Abb. 5a teilweise weit in das Umland hinein, während in Abb. 5b die entsprechenden Teil-EZG nur den Flusslauf abdecken. Insgesamt kann in Abb. 5 aber in den meisten Regionen eine hohe Übereinstimmung der räumlichen Muster der Beobachtungen (Abb. 5a) und Vorhersagen (Abb. 5b) der  $Q_{95}$ -Abflusspenden festgestellt werden. Die größte Diskre-

panz zwischen beobachteter und vorhergesagter  $Q_{95}$ -Abflusspende liegt im zentralen Osten rund um den Neusiedlersee (ca.  $16,8^\circ \text{E}/47,8^\circ \text{N}$ ) vor. Grund dafür könnte eventuell der Mangel an Abflusspegeln (siehe Abb. 1) oder zu geringe Werte für die Evapotranspiration bei den zugrundeliegenden Prädiktoren (Unsicherheit der Einzugsgebiets-eigenschaften) in dieser (semi-)arid-geprägten Region sein. Bemerkenswert ist darüber hinaus auch die Tatsache, dass die durch Überleitungen bedingten Veränderungen der Abflusspenden zum Beispiel in den EZG der Julia (Schweiz), des Kaunertals, hinteren Zillertals und Mölltals gut vorhergesagt werden. Ausschlaggebend dafür ist die in das Vorhersagemodell eingebundene Information zur Quantifizierung der überleitungsbedingten Einzugsgebietsflächenänderung in Form eines zusätzlichen Prädiktors („area\_change“ in Anhang A). Die räumlichen Muster der anderen Abflusskennwerte sind ähnlich zu jenen von  $Q_{95}$  und werden im Anhang D dargestellt.

#### 4.3 Quantifizierung der Unsicherheiten

Die nach Abschn. 3.4 klassifizierten Unsicherheitsgrade sind für die Abflusskennwerte MQ in Abb. 6a sowie  $Q_{95}$  in Abb. 6b räumlich dargestellt. Dabei ist klar erkenntlich, dass die durch die Regionalisierung bedingte Unsicherheit beim Mittelwasserkennwert MQ geringer ist als jene des Niederwasserkennwerts  $Q_{95}$ , die Unsicherheiten bei großen EZG, welche sich entlang der dargestellten Flussverläufe befinden, geringer sind als bei kleinen EZG, und



**Abb. 6** Darstellung des Unsicherheitsgrades für **a** MQ und **b** Q<sub>95</sub> der OWK-Vorhersagen. Flussregionen (weiße Linien): LamaH, Gewässernetz (braun): EU-Hydro – River Network Database (EEA 2019), Staatsgrenzen (schwarz): ©EuroGeographics, Koordinaten: WGS84

bei EZG mit höherer Abflusspende die Unsicherheiten geringer sind als bei Gebieten mit kleiner Spende (West-Ost Gefälle; siehe Abb. 5).

Die Begründung der Ursache für einen lokal hohen Unsicherheitsgrad in westlichen und zentralen Regionen kann nicht immer zweifelsfrei erfolgen. Teilweise wird diese durch eine geringe Einzugsgebietsgröße oder durch eine lokale anthropogene Beeinflussung bedingt sein. Interessant ist, dass in Regionen mit einem bekannten hohen Karstanteil, z. B. Schwäbische Alb (nördlich der Donau in deren oberster Flussregion) oder südliches Salzkammergut (ca. 14°E/47,7°N), aufgrund einer teilweise starken Änderung des abflusswirksamen EZG durch unterirdische und diffuse Abflüsse ein hoher Unsicherheitsgrad besteht und dieser auch durch das Modell attestiert wird. Die tendenziell hohen Unsicherheiten im Wein- und Waldviertel hängen wahrscheinlich mit den verhältnismäßig geringen Abflusspenden zusammen, die bei kleinen absoluten Abweichungen in der Vorhersage bereits zu hohen relativen Abweichungen führen. Ähnliches gilt für den tschechischen Teil der March, wobei hier zusätzlich die anthropogene Beeinflussung des Abflusses durch große Speicher eine Rolle spielen wird. In Abb. 6b fallen in der östlichen Region viele EZG in die Klasse „Q < 10 l/s“. Es ist davon auszugehen, dass dort aufgrund der tendenziell höheren relativen Abweichung bei kleinen Abflusswerten wahrscheinlich meist ein mittlerer oder hoher Grad der Unsicherheit bei den Niederwasserkennwerten vorliegend ist. Die räumliche Verteilung

der Unsicherheitsgrade der Abflusskennwerte MJHQ, MJNQ, MJNQ<sub>7</sub> sowie Q<sub>98</sub> kann dem Anhang E entnommen werden.

## 5 Diskussion und Schlussfolgerung

Im Rahmen dieser Arbeit wurde die Regionalisierung von Abflusskennwerten samt Unsicherheitsabschätzung für ganz Österreich sowie alle ausländischen Oberlieger-Regionen mithilfe von ML-Modellen durchgeführt. Die Modelle verwenden zur Vorhersage eine Vielzahl an hydrologisch relevanten (Einzugsgebiets-)Attributen und wurden auf die, aus den beobachteten Zeitreihen berechneten, Abflusskennwerte als Zielwert trainiert. Die Testergebnisse zeigen, dass im Median ein relativer Fehler von ca. 20% bei der Vorhersage der Abflusskennwerte in unbeobachteten EZG zu erwarten ist. Dennoch stimmen die räumlichen Muster der vorhergesagten Abflusskennwerte der OWK gut mit den beobachteten Abflusskennwerten aus LamaH überein. Das Konfidenzintervall der Vorhersagen diene als Basis für die Unsicherheitsklassifizierung, wobei die resultierenden Grade der Unsicherheit mit unserem hydrologischen Grundverständnis gut übereinstimmen (z. B. West-Ost-Gefälle, Karstgebiete, anthropogene Beeinflussung und Überleitungen). Anzumerken ist jedoch, dass bei dieser Unsicherheitsabschätzung die Mess- und Schätzunsicherheiten der Prädiktoren und Abflusskennwerte nicht widerspiegelt werden können. Gleichzeitig wurde durch die Abschätzung der 90%-Konfidenzintervalle aber per defi-

nitionem eine große Fehlerbandbreite berücksichtigt, welche bei den meisten Test-EZG um einiges größer als der Modellfehler war – und damit in weiterer Folge auch als ein „Puffer“ für die nicht-berücksichtigten Unsicherheitsquellen interpretiert werden kann.

Der Vergleich der Ergebnisse mit vorherigen Studien ist aufgrund mehrerer Umstände nicht direkt möglich. Die meisten Studien verwenden ausschließlich unbeeinflusste oder wenig anthropogen beeinflusste EZG. Da das Ziel dieser Arbeit aber eine flächendeckende Vorhersage von Abflusskennwerten für Österreich war, sind auch stark beeinflusste EZG (z. B. Überleitungen oder große Jahresspeicher) inkludiert worden. Neben der Wahl der EZG unterscheiden sich bei den meisten vergleichbaren Studien auch die Wahl der Abflusskennwerte, die zugrundeliegenden räumlichen Einheiten, die betrachteten Zeiträume sowie die Metriken zur Quantifizierung der Modellgüte. Trotz dieser Differenzen werden folgend die erzielten Ergebnisse in einen Kontext zu jenen von Zhang et al. (2018) sowie Viglione et al. (2013) gesetzt. Ein Vergleich mit der österreichweiten Studie von Laaha und Blöschl (2007) ist wenig repräsentativ, da dort die schlechtesten 5% der Ergebnisse von der zusammenfassenden Auswertung ausgeschlossen wurden. Ein Vergleich mit Daten aus dem digHAO (BMLRT 2007) ist an dieser Stelle ebenfalls nicht zielführend, da dort unterschiedliche räumliche Einheiten und Zeiträume dargestellt sind.

Zhang et al. (2018) testeten für die Regionalisierung von Abflusskennwer-

ten in 605 unbeeinflussten australischen EZG mehrere Methoden (hydrol. Modell, Regression, ML). Der relative RMSE (RMSE/mittlerer MQ) bei der Vorhersage des Kennwerts MQ lag im besten Fall bei 0,56. Im Vergleich dazu liegt der relative RMSE der Testergebnisse dieser Arbeit bei 0,44 und stellt somit eine höhere Modellgüte dar. Die Studie von Viglione et al. (2013) verwendete ein prozessbasiertes hydrologisches Modell mit einer Parameterregionalisierung zur Vorhersage von Abflusskennwerten in 213 wenig beeinflussten österreichischen EZG. Deren Ergebnisse resultierten in einem Bestimmtheitsmaß  $R^2$  von 0,61 für das normierte  $Q_{95}$  [-] und 0,86 für MQ [mm Jahr<sup>-1</sup>]. Die Vergleichswerte der Testergebnisse der vorliegenden Arbeit liegen bei einem  $R^2$  von 0,64 für  $Q_{95}$  und 0,79 für MQ.

Die vorliegende Arbeit hat gezeigt, dass mit ML robuste Vorhersagen unter Angabe der Unsicherheitsabschätzung bei der Regionalisierung von Abflusskennwerten bereitgestellt werden können. Die berechneten Daten sind in dieser Form für Österreich einzigartig. Zudem ist im Vergleich zu hydrologischen Modellen die Vorhersage der Abflusskennwerte in neuen EZG mit deutlich weniger Aufwand verbunden, da nur die erforderlichen Prädiktoren (Einzugsgebietseigenschaften) aggregiert werden müssen und keine neuerliche Parametrisierung notwendig ist – sofern das ML-Modell einmal trainiert ist. Die vorhergesagten Abflusskennwerte werden samt Quantilen und Unsicherheitsklassifizierung kostenlos der Öffentlichkeit zur Verfügung gestellt und tragen daher zu einer Erweiterung der Datengrundlage für wasserwirtschaftliche Anwendungen in Österreich bei.

**Danksagung** Die Datenverarbeitung wurde mit den frei verfügbaren Softwarepaketen R (R Core Team 2022), Python (Python Software Foundation 2022) sowie QGIS (QGIS Development Team 2022) durchgeführt. Besonderer Dank gebührt daher all jenen, die diese Open-Source-Software – sowie darauf aufbauende Pakete und Erweiterungen, z. B. Caret (Kuhn 2019), data.table (Dowle et al. 2014) oder XGBoost (Chen et al. 2016) – großteils unentgeltlich entwickelt haben oder jenen, die ihre wertvolle Anwendungserfahrungen in den zahlreichen Online-Foren teilen. Dank gilt auch der Universität für Bo-

denkultur Wien (BOKU), welche eine Open-Access-Förderung gewährt hat. Die vorliegende Arbeit wurde im Rahmen des Forschungsprojekts „aqua-Zoom“ durchgeführt, welches durch den Europäischen Meeres- und Fischereifonds (EMFF) sowie das Bundesministerium für Landwirtschaft, Regionen und Tourismus (BMLRT) finanziert wurde.

**Funding** Open access funding provided by University of Natural Resources and Life Sciences Vienna (BOKU).

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>. ■

## 6 Anhang

## 6.1 Anhang A

**Tab. 1** Prädiktoren, die beim Training der ML-Modelle zur Vorhersage der Abflusskennwerte verwendet wurden

Attribut	Beschreibung	Einheit	Datenquelle
area <sup>a</sup>	Topografische Einzugsgebietsfläche	km <sup>2</sup>	BMLRT (2007), Linke et al. (2019), UBA (2015)
area_change	Flächenänderung des EZG durch anthropogene Überleitungen	km <sup>2</sup>	–
elev_mean <sup>a</sup>	Mittlere EZG-Höhe	m	NASA JPL SRTMGL1 V3 Digital Elevation 30 m (Farr et al. 2007)
elev_med <sup>a</sup>	Mediane EZG-Höhe	m	Siehe oben
elev_std <sup>a</sup>	Standardabweichung der Höhenwerte im EZG	m	Siehe oben
elev_ran <sup>a</sup>	Spannweite der Höhenwerte im EZG	m	Siehe oben
slope_mean <sup>a</sup>	Mittlere Neigung im EZG	m km <sup>-1</sup>	Siehe oben
strm_dens <sup>a</sup>	Gewässernetzdichte	m km <sup>-2</sup>	EEA (2019)
p_mean <sup>a</sup>	Mittlerer täglicher Niederschlag	mm	ERA5L (2020)
et0_mean <sup>a</sup>	Mittlere tägliche Referenz-Evapotranspiration	mm	Trabucco und Zomer (2019)
arid_2 <sup>a</sup>	Aridität von Climate Database	–	Siehe oben
p_season <sup>a</sup>	Saisonalität des Niederschlags, Berechnung über Sinus	–	ERA5L (2020)
frac_snow <sup>a</sup>	Anteil des Niederschlags, welcher im Mittel als Schnee fällt	–	Siehe oben
hi_prec_fr <sup>a</sup>	Häufigkeit an niederschlagsreichen (mind 5 x p_mean) Tagen pro Jahr	d yr <sup>-1</sup>	Siehe oben
hi_prec_du <sup>a</sup>	Mittlere Dauer von niederschlagsreichen Perioden	d	Siehe oben
lo_prec_fr <sup>a</sup>	Häufigkeit an niederschlagsarmen (max 1 mm/d) Tagen pro Jahr	d yr <sup>-1</sup>	Siehe oben
lo_prec_du <sup>a</sup>	Mittlere Dauer von niederschlagsarmen Perioden	d	Siehe oben
agr_fra <sup>a</sup>	Anteil an Agrarfläche	–	COR (2012)
bare_fra <sup>a</sup>	Anteil an kahler Fläche	–	Siehe oben
forest_fra <sup>a</sup>	Anteil an Waldfläche	–	Siehe oben
glac_fra <sup>a</sup>	Anteil an Gletscher	–	Siehe oben
lake_fra <sup>a</sup>	Anteil an Seeoberfläche	–	Siehe oben
urban_fra <sup>a</sup>	Anteil an versiegelter Fläche	–	Siehe oben
lai_max <sup>a</sup>	Max monatliches Mittel des LAI (Leaf area index)	–	Myneni et al. (2015)
lai_diff <sup>a</sup>	Differenz zw max und min monatlichem Mittel des LAI	–	Siehe oben
ndvi_max <sup>a</sup>	Max monatliches Mittel des NDVI (Normalized difference vegetation index)	–	Vermote (2015)
ndvi_min <sup>a</sup>	Min monatliches Mittel des NDVI	–	Siehe oben
gvf_max <sup>a</sup>	Max monatliches Mittel der GVF (Green vegetation fraction)	–	Friedl und Sulla-Menashe (2019), Vermote (2015)
gvf_diff <sup>a</sup>	Differenz zw max und min monatlichem GVF	–	Siehe oben
bedrk_dep <sup>a</sup>	Mittlere Tiefe zum Festgestein	m	Pelletier et al. (2016)
root_dep <sup>a</sup>	Mittlere Bodenmächtigkeit	m	Hiederer (2013a, b)
soil_poros <sup>a</sup>	Bodenporosität	–	Siehe oben
soil_condu <sup>a</sup>	Bodenkonduktivität	cm h <sup>-1</sup>	Tóth et al. (2017)
soil_tawc <sup>a</sup>	Max Bodenwassergehalt (TAWC)	m	Hiederer (2013a, b)
sand_fra <sup>a</sup>	Sandanteil	–	Siehe oben
silt_fra <sup>a</sup>	Siltanteil	–	Siehe oben
clay_fra <sup>a</sup>	Tonanteil	–	Siehe oben
grav_fra <sup>a</sup>	Kiesanteil	–	Siehe oben
oc_fra <sup>a</sup>	Anteil an organischem Material	–	Siehe oben
gc_ig_fra <sup>a</sup>	Anteil der GLiM (Global Lithological Map database) Klasse „Eis und Gletscher“	–	Hartmann und Moosdorf (2012)
gc_mt_fra <sup>a</sup>	Anteil der GLiM Klasse „Metamorphite“	–	Siehe oben
gc_pa_fra <sup>a</sup>	Anteil der GLiM Klasse „saure Plutonite“	–	Siehe oben
gc_pb_fra <sup>a</sup>	Anteil der GLiM Klasse „basische Plutonite“	–	Siehe oben
gc_pi_fra <sup>a</sup>	Anteil der GLiM Klasse „intermediäre Plutonite“	–	Siehe oben
gc_py_fra <sup>a</sup>	Anteil der GLiM Klasse „Pyroklastika“	–	Siehe oben

Tab. 1 (Fortsetzung)

Attribut	Beschreibung	Einheit	Datenquelle
gc_sc_fra <sup>a</sup>	Anteil der GLiM Klasse „karbonatisches Sedimentgestein“	–	Siehe oben
gc_sm_fra <sup>a</sup>	Anteil der GLiM Klasse „gemischtes Sedimentgestein“	–	Siehe oben
gc_ss_fra <sup>a</sup>	Anteil der GLiM Klasse „siliklastisches Sedimentgestein“	–	Siehe oben
gc_su_fra <sup>a</sup>	Anteil der GLiM Klasse „unkonsolidierte Sedimente“	–	Siehe oben
gc_va_fra <sup>a</sup>	Anteil der GLiM Klasse „saure Vulkanite“	–	Siehe oben
gc_vb_fra <sup>a</sup>	Anteil der GLiM Klasse „basische Vulkanite“	–	Siehe oben
gc_wb_fra <sup>a</sup>	Anteil der GLiM Klasse „Wasserflächen“	–	Siehe oben
geol_perme <sup>a</sup>	Permeabilität des Festgesteins	–	Gleeson et al. (2014)
geol_poros <sup>a</sup>	Porosität des Festgesteins	–	Siehe oben
centr_lon	Längswert des EZG-Schwerpunktes im Koordinatensystem EPSG3035	m	–
centr_lat	Hochwert des EZG-Schwerpunktes im Koordinatensystem EPSG3035	m	–
vol_A	Akkumuliertes Nutzvolumen von Jahresspeichern, relevant ist Reservoirauslass	hm <sup>3</sup>	–
vol_As	Spezifisches akkum Nutzvolumen der Klasse A (= vol_A*1000/area)	mm	–
vol_C	Akkumuliertes Retentionsvolumen (= Oberfläche × 1 m) von natürlichen Seen im Hauptschluss des beobachteten Gewässers	hm <sup>3</sup>	–
vol-Cs	Spezifisches akkum Volumen der Klasse C (= vol_C*1000/area)	mm	–
MP <sup>b</sup>	Mittlerer täglicher Niederschlag	mm	ERA5L (2020)
MJHP <sup>b</sup>	Mittlerer jährlicher Spitzenniederschlagswert	mm	Siehe oben
MP1_ <sup>b</sup>	Anteil des mittleren Niederschlags im Winter vom mittleren Jahresniederschlag	–	Siehe oben
MP2_ <sup>b</sup>	Anteil des mittleren Niederschlags im Frühling vom mittleren Jahresniederschlag	–	Siehe oben
MP3_ <sup>b</sup>	Anteil des mittleren Niederschlags im Sommer vom mittleren Jahresniederschlag	–	Siehe oben
MP4_ <sup>b</sup>	Anteil des mittleren Niederschlags im Herbst vom mittleren Jahresniederschlag	–	Siehe oben
MPdry1 <sup>b</sup>	Mittlerer Anteil an niederschlagslosen Tagen (< 0,1 mm) im Winter (Jan/Feb/März)	–	Siehe oben
MPdry2 <sup>b</sup>	Mittlerer Anteil an niederschlagslosen Tagen (< 0,1 mm) im Frühling (Apr/Mai/Jun)	–	Siehe oben
MPdry3 <sup>b</sup>	Mittlerer Anteil an niederschlagslosen Tagen (< 0,1 mm) im Sommer (Jul/Aug/Sep)	–	Siehe oben
MPdry4 <sup>b</sup>	Mittlerer Anteil an niederschlagslosen Tagen (< 0,1 mm) im Herbst (Okt/Nov/Dez)	–	Siehe oben
MPmax <sup>b</sup>	Max Monatsmittel des Niederschlags	mm	Siehe oben
MPmin <sup>b</sup>	Min Monatsmittel des Niederschlags	mm	Siehe oben
MT <sup>b</sup>	Mittlere 2 m Jahres-Lufttemperatur	°C	Siehe oben
MT1 <sup>b</sup>	Mittlere 2 m Lufttemperatur im Winter	°C	Siehe oben
MT2 <sup>b</sup>	Mittlere 2 m Lufttemperatur im Frühling	°C	Siehe oben
MT3 <sup>b</sup>	Mittlere 2 m Lufttemperatur im Sommer	°C	Siehe oben
MT4 <sup>b</sup>	Mittlere 2 m Lufttemperatur im Herbst	°C	Siehe oben
MET <sup>b</sup>	Mittlere tägliche aktuelle ET (Evapotranspiration)	mm	Siehe oben
MET1_ <sup>b</sup>	Anteil der mittleren aktuellen ET im Winter von der mittleren Jahres-ET	–	Siehe oben
MET2_ <sup>b</sup>	Anteil der mittleren aktuellen ET im Frühling von der mittleren Jahres-ET	–	Siehe oben
MET3_ <sup>b</sup>	Anteil der mittleren aktuellen ET im Sommer von der mittleren Jahres-ET	–	Siehe oben
MET4_ <sup>b</sup>	Anteil der mittleren aktuellen ET im Herbst von der mittleren Jahres-ET	–	Siehe oben
MSWE1 <sup>b</sup>	Mittleres SWE (Snow water equivalent) im Winter	mm	Siehe oben
MSWE2 <sup>b</sup>	Mittleres SWE im Frühling	mm	Siehe oben
MSWE3 <sup>b</sup>	Mittleres SWE im Sommer	mm	Siehe oben
MSWE4 <sup>b</sup>	Mittleres SWE im Herbst	mm	Siehe oben
MNSR1 <sup>b</sup>	Mittlere Tagesmaximal-Nettosolarstrahlung im Winter	Wm <sup>-2</sup>	Siehe oben
MNSR2 <sup>b</sup>	Mittlere Tagesmaximal-Nettosolarstrahlung im Frühling	Wm <sup>-2</sup>	Siehe oben
MNSR3 <sup>b</sup>	Mittlere Tagesmaximal-Nettosolarstrahlung im Sommer	Wm <sup>-2</sup>	Siehe oben
MNSR4 <sup>b</sup>	Mittlere Tagesmaximal-Nettosolarstrahlung im Herbst	Wm <sup>-2</sup>	Siehe oben
Q <sup>b</sup>	Abflusskennwerte MJHQ, MQ, MJNQ, MJNQ <sub>7</sub> , Q <sub>95</sub> , Q <sub>98</sub> , welche aus den beobachteten Zeitreihen berechnet wurden und die Zielgröße beim Modelltraining darstellen	m <sup>3</sup> s <sup>-1</sup>	BAFU (2020), CHMI (2020), GKD (2020), HZB (2020), LUBW (2020)

<sup>a</sup> Attribute für das Modelltraining auf Basis der LamaH-EZG wurden von Klingler et al. (2021b, c) übernommen.

<sup>b</sup> Berechnet auf Tagesbasis für die hydrologischen Jahre 2003 bis 2017.

## 6.2 Anhang B: Festlegung (der Bandbreite) der Hyperparameter der ML-Modelle.

XGBoost:: n\_iter: 80; n\_random\_initial\_points: 20; nrounds: 300–5000; eta: 0,0001–0,5; max\_depth: 3–30; min\_child\_weight: 1–30; subsample: 0,1–1; colsample\_bytree: 0,1–1; gamma: 0–20.

QRF:: ntree: 5000; min.node.size: 5; mtry:  $1/3 \times$  Anzahl der selektierten Prädiktoren.

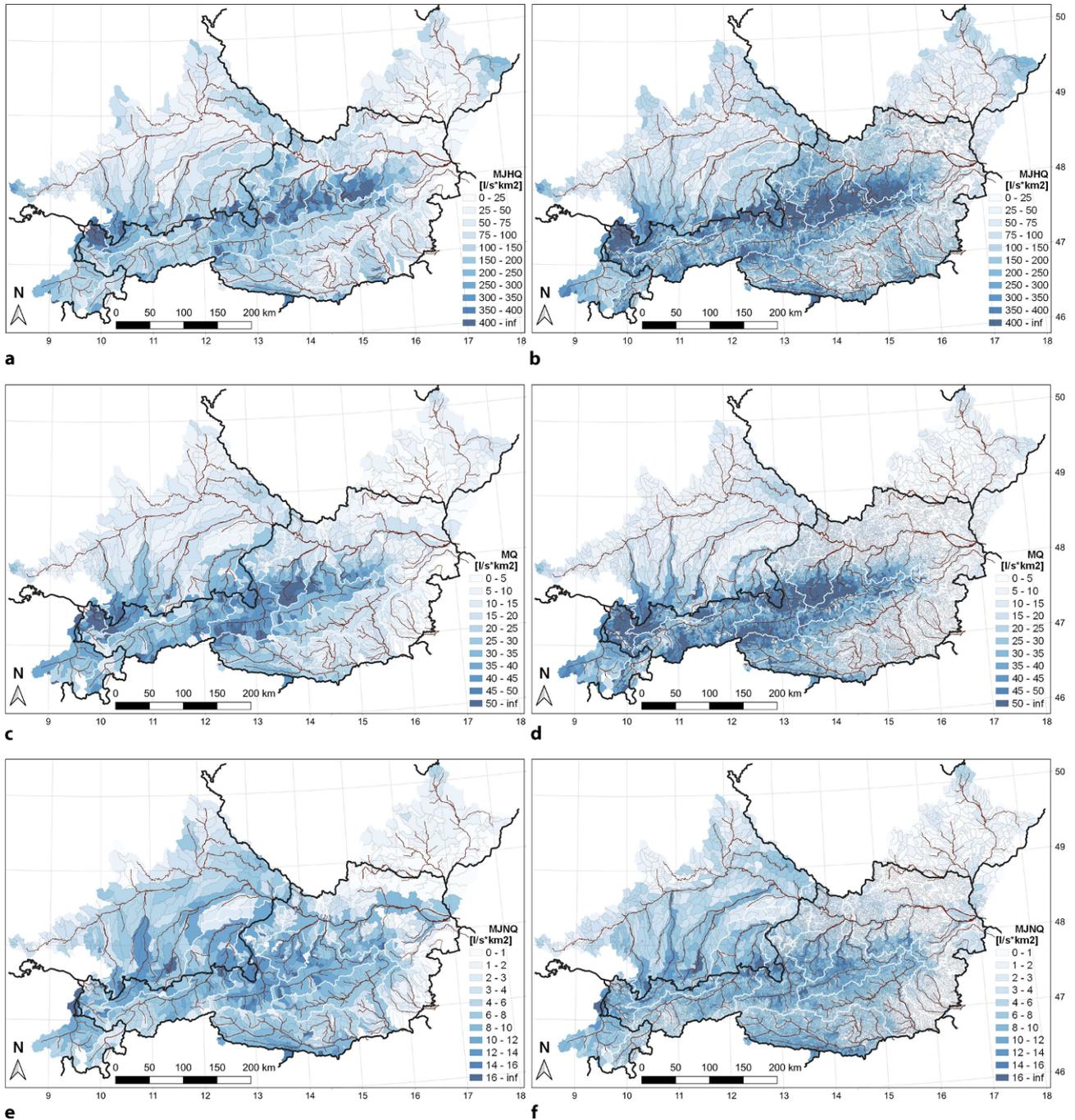
## 6.3 Anhang C

**Tab. 2** Die 5 wichtigsten Prädiktoren der einzelnen Vorhersagemodelle

Rang	Abflusskennwerte					
	MJHQ	MQ	MJNQ	Q <sub>95</sub>	Q <sub>98</sub>	MJNQ <sub>7</sub>
1	MJHP	grav_fra	gc_sc_fra	area_change	p_mean	MPmax
2	MPmin	area_calc	MET2_	lo_prec_du	MPmax	MSWE1
3	strm_dens	strm_dens	p_mean	gc_sc_fra	gc_sc_fra	gc_sc_fra
4	area_calc	p_mean	lo_prec_du	MPmax	lo_prec_fr	MP
5	MP	MP	MPmax	p_mean	MT3	MET2_

Die Beschreibungen zu den Abkürzungen der Prädiktoren können dem Anhang A entnommen werden

## 6.4 Anhang D



**Abb. 7** Gegenüberstellung der Pegelbeobachtungen (a, c, e, g, i) mit den OWK-Vorhersagen (b, d, f, h, j) für die Abflusspenden von MJHQ (a, b), MQ (c, d), MJNQ (e, f), MJNQ<sub>7</sub> (g, h) sowie Q<sub>98</sub> (i, j). Flussregionen (weiße Linien): LamaH, Gewässernetz (braun): EU-Hydro – River Network Database (EEA 2019), Staatsgrenzen (schwarz): ©EuroGeographics, Koordinaten: WGS84

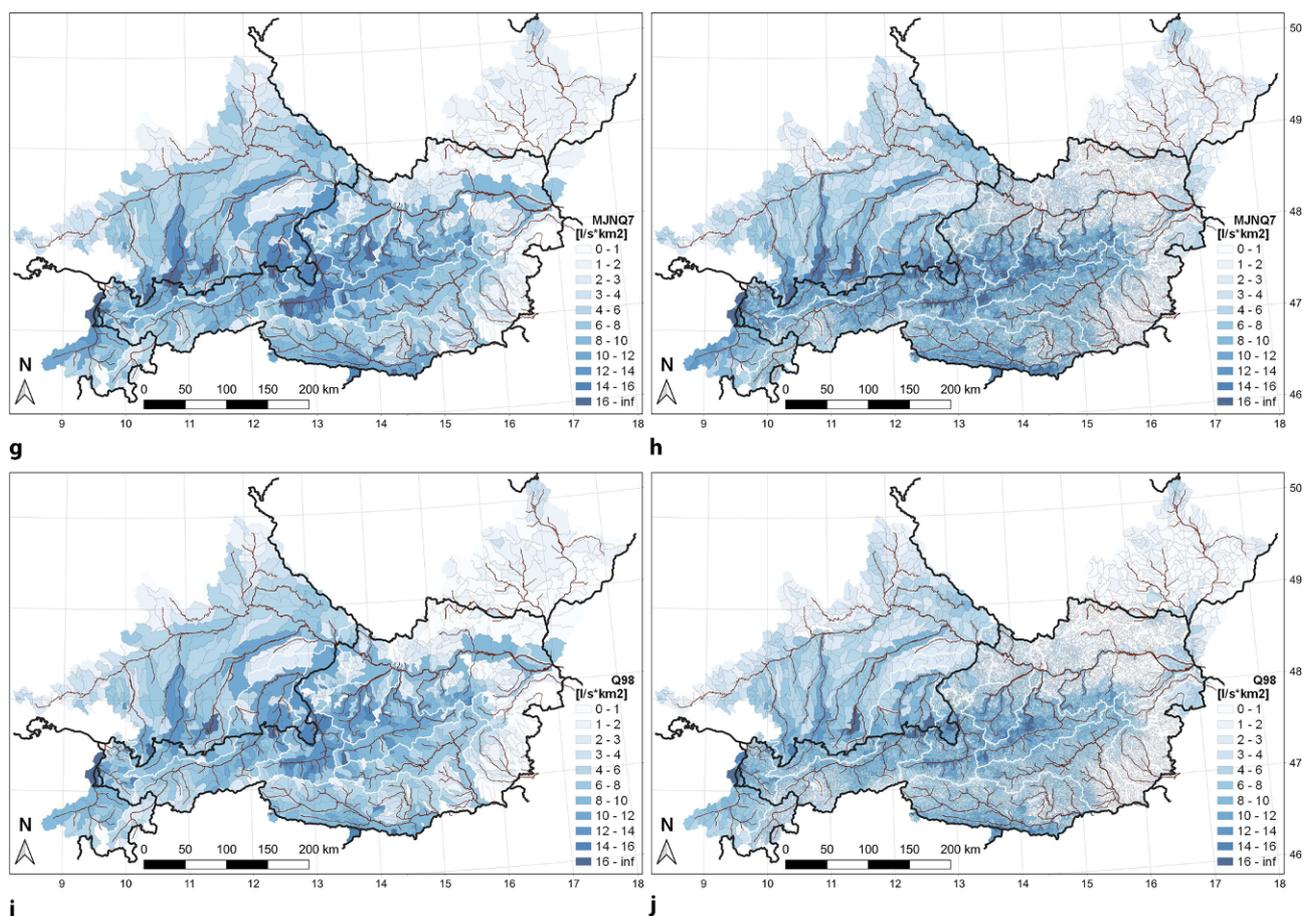
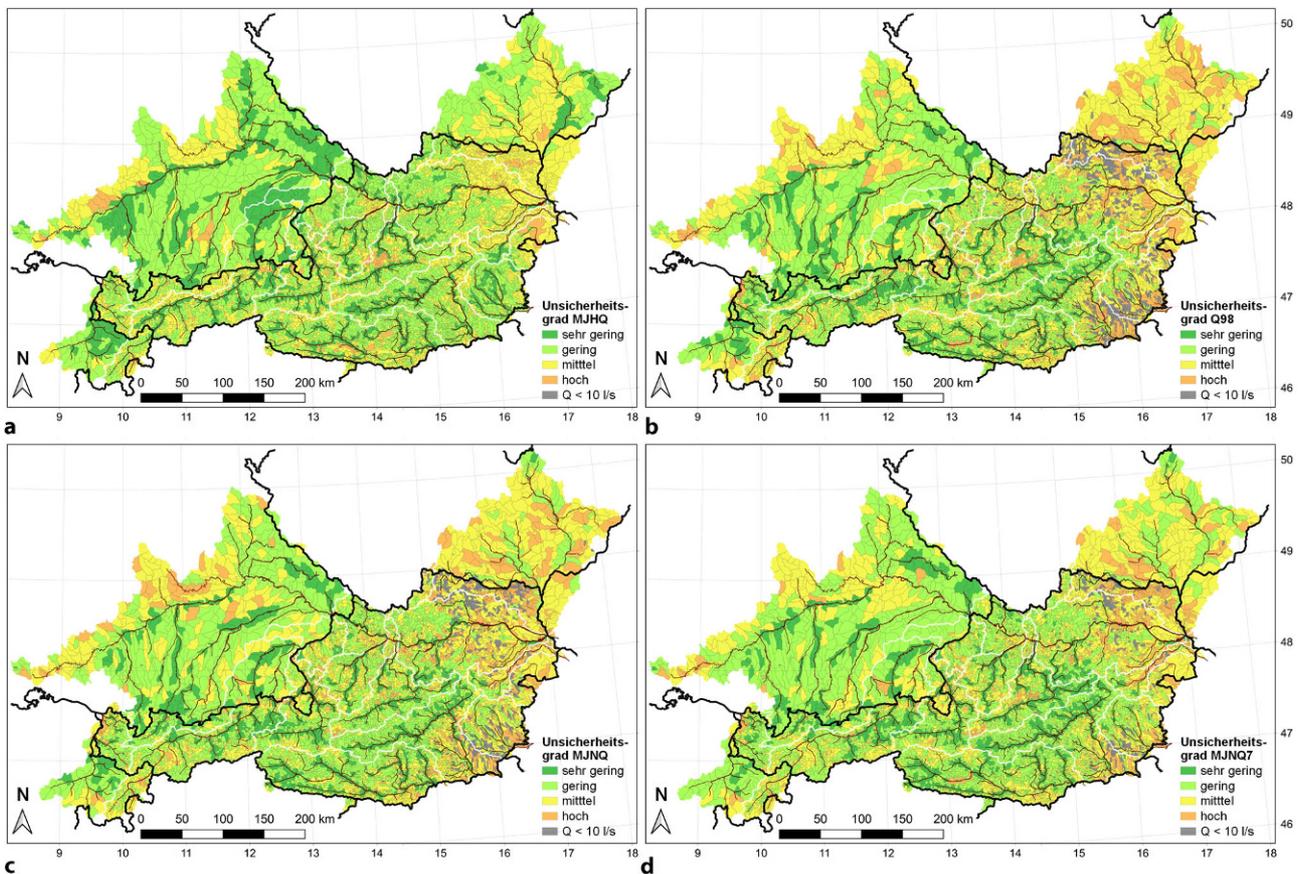


Abb. 7 (Fortsetzung)

## 6.5 Anhang E



**Abb. 8** Darstellung des Unsicherheitsgrades für a MJHQ, b Q<sub>98</sub>, c MJNQ, d MJNQ<sub>7</sub> der OWK-Vorhersagen. Flussregionen (weiße Linien): LamaH, Gewässernetz (braun): EU-Hydro – River Network Database (EEA 2019), Staatsgrenzen (schwarz): ©EuroGeographics, Koordinaten: WGS84

## Literatur

- BAFU (2020):** Bundesamt für Umwelt – Abteilung Hydrologie. Bern, Schweiz (erhalten am: 23.09.2020)
- Biondi, D., & de Luca, D. L. (2017):** Rainfall-runoff model parameter conditioning on regional hydrological signatures: application to ungauged basins in southern Italy. *Hydrology Research*, 48(3), 714–725. <https://doi.org/10.2166/nh.2016.097>
- Blöschl, G., Waser, J., Buttinger-Kreuzhuber, A., et al. (2022):** HOchwasserRisikozonierung Austria 3.0 (HORA 3.0). Österreichische Wasser- und Abfallwirtschaft, <https://doi.org/10.1007/s00506-022-00848-7>
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (Eds.). (2013):** Runoff Prediction in Ungauged Basins. Cambridge University Press. <https://doi.org/10.1017/CBO9781139235761>
- BMLRT (2007):** Hydrologischer Atlas Österreichs, digitale Ausgabe (digHAO), 3. Lieferung, Bundesministerium für Landwirtschaft, Regionen und Tourismus, Wien, Österreich
- Breiman, L. (2001):** Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Buchart, M. (2012):** Leitfaden Aquakulturen – Hinweise für Planung und Betrieb von extensiven Durchflussanlagen (pp. 1–20). [https://www.noegv.at/noe/Wasser/Leitfaden\\_Aquakulturen\\_web.pdf](https://www.noegv.at/noe/Wasser/Leitfaden_Aquakulturen_web.pdf) (Zugriff: 30.01.2022)
- Burn, D. H. (1990):** Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10), 2257–2265. <https://doi.org/10.1029/WR026i010p02257>
- Chen, T., & Guestrin, C. (2016):** XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., & Li, Y. (2016):** xgboost: Extreme Gradient Boosting. <https://cran.r-project.org/web/packages/xgboost/index.html> (Zugriff: 17.08.2021)
- CHMI (2020):** Tschechisches Hydrometeorologisches Institut. Brunn, Tschechische Republik (erhalten am: 14.12.2020)
- COR (2012):** CORINE Land Cover 2012. Europäische Umweltagentur, Kopenhagen, Dänemark, verfügbar unter: <https://land.copernicus.eu/pan-european/corine-land-cover> (Zugriff: 02.03.2020)
- Coxon, G., Freer, J., Wagener, T., Odoni, N. A., & Clark, M. (2014):** Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments. *Hydrological Processes*, 28(25), 6135–6150. <https://doi.org/10.1002/hyp.10096>
- Donnelly, C., Andersson, J. C. M., & Arheimer, B. (2016):** Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrological Sciences Journal*, 61(2), 255–273. <https://doi.org/10.1080/02626667.2015.1027710>
- Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bensch, M., & Parsonage, H. (2014):** data.table: Extension of data.frame. <https://cran.r-project.org/web/packages/data.table/index.html> (Zugriff: 17.08.2021)
- EEA (2019):** EU-Hydro – River Network Database. Version 1.2, Europäische Umweltagentur, Copernicus Programm, <https://land.copernicus.eu/imagery-in-situ/eu-hydro/eu-hydro-river-network-database> (Zugriff: 22.10.2020)
- ERA5L (2020):** ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Ser-

- vice (C3S) Climate Data Store (CDS), verfügbar unter: <https://doi.org/10.24381/cds.e2161bac> (Zugriff: 22.10.2020)
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007): The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2), RG2004. <https://doi.org/10.1029/2005RG000183>
- Feigl, M., Lebedzinski, K., Herrnegger, M., & Schulz, K. (2021): Vorhersage der Fließgewässertemperaturen in österreichischen Einzugsgebieten mittels Machine Learning-Verfahren. *Österreichische Wasser- Und Abfallwirtschaft*, 73(7–8), 308–328. <https://doi.org/10.1007/s00506-021-00771-3>
- Friedl, M., & Sulla-Menashe, D. (2019): MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [data set]. NASA EOSDIS Land Processes DAAC, verfügbar unter: <https://doi.org/10.5067/MODIS/MCD12Q1.006> (Zugriff: 12.03.2020)
- GKD (2020): Bayerisches Landesamt für Umwelt – Gewässerkundlicher Dienst München, Deutschland, verfügbar unter: <https://www.gkd.bayern.de/de/fluesse/abfluss/tabellen> (heruntergeladen am: 15.09.2020).
- Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014): A glimpse beneath earth's surface: GLOBAL Hydrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, 41(11), 3891–3898. <https://doi.org/10.1002/2014GL059856>
- Grandry, M., Gailliez, S., Sohler, C., Verstraete, A., & Degré, A. (2013): A method for low-flow estimation at ungauged sites: a case study in Wallonia (Belgium). *Hydrology and Earth System Sciences*, 17(4), 1319–1330. <https://doi.org/10.5194/hess-17-1319-2013>
- Gupta, H. v., Wagener, T., & Liu, Y. (2008): Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, 22(18), 3802–3813. <https://doi.org/10.1002/hyp.6989>
- Hannaford, J., Holmes, M. G. R., Laizé, C. L. R., Marsh, T. J., & Young, A. R. (2013): Evaluating hydrometric networks for prediction in ungauged basins: a new methodology and its application to England and Wales. *Hydrology Research*, 44(3), 401–418. <https://doi.org/10.2166/nh.2012.115>
- Hartmann, J., & Moosdorf, N. (2012): The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, 13(12). <https://doi.org/10.1029/2012GC004370>
- Hiederer, R. (2013a): Mapping Soil Properties for Europe—Spatial Representation of Soil Database Attributes. Luxembourg: Publications Office of the European Union, EUR26082EN Scientific and Technical Research series, ISSN 1831–9424, <https://doi.org/10.2788/94128>
- Hiederer, R. (2013b): Mapping Soil Typologies—Spatial Decision Support Applied to European Soil Database. Luxembourg: Publications Office of the European Union, EUR25932EN Scientific and Technical Research series, ISSN 1831–9424, <https://doi.org/10.2788/8728>
- Holmes, M. G. R., Young, A. R., Gustard, A., & Grew, R. (2002): A region of influence approach to predicting flow duration curves within ungauged catchments. *Hydrology and Earth System Sciences*, 6(4), 721–731. <https://doi.org/10.5194/hess-6-721-2002>
- Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., & Gascuel-Oudoux, C. (2014): Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, 50(9), 7445–7469. <https://doi.org/10.1002/2014WR015484>
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., et al. (2013): A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, 58(6), 1198–1255. <https://doi.org/10.1080/02626667.2013.803183>
- HZB (2020): Bundesministerium für Landwirtschaft, Regionen und Tourismus – Hydrographisches Zentralbüro. Wien, Österreich (erhalten am: 08.09.2020)
- Juston, J., Jansson, P.-E., & Gustafsson, D. (2014): Rating curve uncertainty and change detection in discharge time series: case study with 44-year historic data from the Nyangores River, Kenya. *Hydrological Processes*, 28(4), 2509–2523. <https://doi.org/10.1002/hyp.9786>
- Kjeldsen, T. R., Jones, D. A., & Morris, D. G. (2014): Using multiple donor sites for enhanced flood estimation in ungauged catchments. *Water Resources Research*, 50(8), 6646–6657. <https://doi.org/10.1002/2013WR015203>
- Klingler, C., Feigl, M., Linsbichler, T., Frey, S., & Schulz, K. (2022): Potenzial von Machine Learning bei der kurzfristigen Leistungsprognose innerhalb einer Laufkraftwerkskette. *Österreichische Wasser- Und Abfallwirtschaft*, <https://doi.org/10.1007/s00506-022-00849-6>
- Klingler, C., Herrnegger, M., & Schulz, K. (2021a): LamaH-CE: Large-Sample Data for Hydrology and Environmental Sciences for Central Europe—files [data set]. Zenodo
- Klingler, C., Herrnegger, M., & Schulz, K. (2021b): LamaH | Large-Sample Data for Hydrology: Big data für die Hydrologie und Umweltwissenschaften. *Österreichische Wasser- Und Abfallwirtschaft*, 73(7–8), 244–269. <https://doi.org/10.1007/s00506-021-00769-x>
- Klingler, C., Herrnegger, M., & Schulz, K. (2021c): LamaH-CE: Large-Sample Data for Hydrology and Environmental Sciences for Central Europe. *Earth System Science Data*, 13(9), 4529–4565. <https://doi.org/10.5194/essd-13-4529-2021>
- Kuhn, M. (2019): caret: Classification and Regression Training. <https://cran.r-project.org/web/packages/caret/index.html> (Zugriff: 17.08.2021)
- Laaha, G., & Blöschl, G. (2007): A national low flow estimation procedure for Austria. *Hydrological Sciences Journal*, 52(4), 625–644. <https://doi.org/10.1623/hysj.52.4.625>
- Lebedzinski, K., & Fürst, J. (2018): Entwicklung der alpinen Abflussregime in Österreich im Zeitraum 1961–2010. *Österreichische Wasser- Und Abfallwirtschaft*, 70(9–10), 474–484, <https://doi.org/10.1007/s00506-018-0499-z>
- Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, E., & Thieme, M. (2019): Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, 6(1), 283. <https://doi.org/10.1038/s41597-019-0300-6>
- LUBW (2020): Landesanstalt für Umwelt Baden-Württemberg – Gewässerkundlicher Dienst. Karlsruhe, Deutschland, verfügbar unter: [https://udo.lubw.baden-wuerttemberg.de/public/p/pegel\\_messwerte\\_leer](https://udo.lubw.baden-wuerttemberg.de/public/p/pegel_messwerte_leer) (erhalten am: 04.09.2020)
- McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M., & Woods, R. A. (2011): Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. *Hydrological Processes*, 25(4), 511–522. <https://doi.org/10.1002/hyp.7841>
- Meinshausen, N. (2006): Quantile Regression Forests. *Journal of Machine Learning Research*, 7(35), 983–999
- Močkus, J., Tiesis, V., & Zilinskas, A. (1978): The application of Bayesian methods for seeking the extremum. *Towards Global Optimisation*. [https://doi.org/10.1007/978-94-009-0909-0\\_8](https://doi.org/10.1007/978-94-009-0909-0_8)
- Myneni, R., Knyazikhin, Y., & Park, T. (2015): MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [data set]. NASA EOSDIS Land Processes DAAC, verfügbar unter: <https://doi.org/10.5067/MODIS/MCD15A3H.006> (Zugriff: 12.03.2020)
- Nathan, R. J., & McMahon, T. A. (1992): Estimating low flow characteristics in ungauged catchments. *Water Resources Management*, 6(2), 85–100. <https://doi.org/10.1007/BF00872205>
- Olden, J. D., & Poff, N. L. (2003): Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2), 101–121. <https://doi.org/10.1002/rra.700>
- Oppel, H., & Schumann, A. H. (2020): Machine learning based identification of dominant controls on runoff dynamics. *Hydrological Processes*, 34(11), 2450–2465. <https://doi.org/10.1002/hyp.13740>
- OEWA (2021): Big Data und Künstliche Intelligenz in Hydrologie und Wasserwirtschaft. *Österreichische Wasser- und Abfallwirtschaft*, 73(7–8), verfügbar unter: <https://link.springer.com/journal/506/volumes-and-issues/73-7> (Zugriff: 02.03.2022)
- Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., Williams, Z. C., Brunke, M. A., & Gochis, D. (2016): Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers [data set]. ORNL DAAC, Oak Ridge, Tennessee, USA, <https://doi.org/10.3334/ORNLLDAAC/1304>
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rositer, D. (2021): SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, 7(1), 217–240. <https://doi.org/10.5194/soil-7-217-2021>
- Prieto, C., Le Vine, N., Kavetski, D., García, E., & Medina, R. (2019): Flow Prediction in Ungauged Catchments Using Probabilistic Random Forests Regionalization and New Statistical Adequacy Tests. *Water Resour. Res.*, 55(5), 4364–4392. <https://doi.org/10.1029/2018WR023254>
- Pugliese, A., Castellari, A., & Brath, A. (2014): Geostatistical prediction of flow-duration curves in an index-flow framework. *Hydrology and Earth System Sciences*, 18(9), 3801–3816. <https://doi.org/10.5194/hess-18-3801-2014>
- Python Software Foundation. (2022): Python Language Reference. <https://www.python.org> (Zugriff: 04.02.2020)
- Qamar, M. U., Azmat, M., Cheema, M. J. M., Shahid, M. A., Khushnood, R. A., & Ahmad, S. (2016): Model swapping: A comparative performance signature for the prediction of flow duration curves in ungauged basins. *Journal of Hydrology*, 541, 1030–1041. <https://doi.org/10.1016/j.jhydrol.2016.08.012>
- QGIS Development Team (2022): QGIS Geographic Information System. Open Source Geospatial Foundation Project, <https://www.qgis.org> (Zugriff: 04.02.2020)
- R Core Team (2022): A language and environment for statistical computing. R Foundation for Statistical Computing, Wien, Österreich, <https://www.r-project.org> (Zugriff: 04.02.2020)
- Refsgaard, J. C., & Knudsen, J. (1996): Operational Validation and Intercomparison of Different Types of Hydrological Models. *Water Resources Research*, 32(7), 2189–2202. <https://doi.org/10.1029/96WR00896>
- Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014): Characterizing hydrologic change through catchment classification. *Hydrology and Earth System Sciences*,

- 18(1), 273–285. <https://doi.org/10.5194/hess-18-273-2014>
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., & Carrillo, G. (2011):** Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, 15(9), 2895–2911. <https://doi.org/10.5194/hess-15-2895-2011>
- Shu, C., & Ouarda, T. B. M. J. (2012):** Improved methods for daily streamflow estimates at ungauged sites. *Water Resources Research*, 48(2). <https://doi.org/10.1029/2011WR011501>
- Snoek, J., Larochelle, H., & Adams, R. P. (2012):** Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 4, 2951–2959
- Tóth, B., Weynants, M., Pásztor, L., & Hengl, T. (2017):** 3D soil hydraulic database of Europe at 250 m resolution. *Hydrological Processes*, 31(14), 2662–2666. <https://doi.org/10.1002/hyp.11203>
- Trabucco, A., & Zomer, R. (2019):** Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 [data set]. CGIAR Consortium for Spatial Information (CGIAR-CSI), published online, verfügbar unter: <https://doi.org/10.6084/m9.figshare.7504448.v3> (Zugriff: 18.03.2020)
- UBA (2015):** Oberflächenwasserkörper – Einzugsgebiete. Version NGP15, Umweltbundesamt, Wien, Österreich
- Vermote, E. (2015):** MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid V006 [data set]. NASA EOSDIS Land Processes DAAC, verfügbar unter: <https://doi.org/10.5067/MODIS/MOD09Q1.006> (Zugriff: 12.03.2020)
- Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., & Blöschl, G. (2013):** Comparative assessment of predictions in ungauged basins—Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences*, 17(6), 2263–2279. <https://doi.org/10.5194/hess-17-2263-2013>
- Visesri, S., & McIntyre, N. (2016):** Regionalisation of hydrological responses under land-use change and variable data quality. *Hydrological Sciences Journal*, 61(2), 302–320. <https://doi.org/10.1080/02626667.2015.1006226>
- Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007):** Catchment Classification and Hydrologic Similarity. *Geography Compass*, 1(4), 901–931. <https://doi.org/10.1111/j.1749-8198.2007.00039.x>
- Wesemann, J., Herrnegger, M., & Schulz, K. (2018a):** Hydrological modelling in the anthroposphere: predicting local runoff in a heavily modified high-alpine catchment. *J. Mt. Sci.*, 15, 921–938. <https://doi.org/10.1007/s11629-017-4587-5>
- Wesemann, J., Holzmann, H., Schulz, K., & Herrnegger, M. (2018b):** Behandlung künstlicher Speicher und Überleitungen in der alpinen Niederschlags-Abfluss-Vorhersage. *Österreichische Wasser- und Abfallwirtschaft*, 70, 485–496. <https://doi.org/10.1007/s00506-018-0501-9>
- Westerberg, I. K., Gong, L., Beven, K. J., Seibert, J., Semedo, A., Xu, C.-Y., & Halldin, S. (2014):** Regional water balance modelling using flow-duration curves with observational uncertainties. *Hydrology and Earth System Sciences*, 18(8), 2993–3013. <https://doi.org/10.5194/hess-18-2993-2014>
- Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016):** Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, 52(3), 1847–1865. <https://doi.org/10.1002/2015WR017635>
- Yadav, M., Wagener, T., & Gupta, H. (2007):** Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8), 1756–1774. <https://doi.org/10.1016/j.advwatres.2007.01.005>
- Zhang, Y., Chiew, F. H. S., Li, M., & Post, D. (2018):** Predicting Runoff Signatures Using Regression and Hydrological Modeling Approaches. *Water Resources Research*, 54(10), 7859–7878. <https://doi.org/10.1029/2018WR023325>
- Zhang, Y., Vaze, J., Chiew, F. H. S., Teng, J., & Li, M. (2014):** Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall-runoff modelling. *Journal of Hydrology*, 517, 936–948. <https://doi.org/10.1016/j.jhydrol.2014.06.032>

**Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.