



Potenzial von Machine Learning bei der kurzfristigen Leistungsprognose innerhalb einer Laufkraftwerkskette

Christoph Klingler · Moritz Feigl · Thomas Linsbichler · Simon Frey · Karsten Schulz

Angenommen: 27. Januar 2022 / Online publiziert: 18. Februar 2022
© Der/die Autor(en) 2022

Zusammenfassung Da die Leistung eines Laufkraftwerks ohne Schwallbetrieb nicht gesteuert werden kann, sind möglichst präzise Leistungsprognosen nötig, um die generierte elektrische Energie bestmöglich am internationalen Strommarkt verwerten zu können. Derzeit befindet sich beim österreichischen Wasserkraftwerksbetreiber Verbund AG für den Zweck der Leistungsprognose eine Kombination aus hydrologischen und hydrodynamischen Modellen (PW) im operativen Betrieb, welche aber insbesondere bei ansteigenden Leistungsverläufen noch Defizite aufweist. Deshalb wird in dieser Studie an den Laufkraftwerken Braunau-Simbach, Aschach und Greifenstein das Potenzial von Machine Learning (ML) Verfahren bei der kurzfristigen (bis 4 h) Leistungsprognose in fünf hydrologisch interessanten Zeitfenstern eruiert. Dafür werden gemessene

Abfluss- und Leistungswerte von stromauf liegenden Laufkraftwerken und Pegeln als Eingangsdaten herangezogen. Die erzielten Ergebnisse zeigen, dass ML im Anwendungsbereich der kurzfristigen Leistungsprognose innerhalb einer Laufkraftwerkskette sinnvoll eingesetzt werden kann. So konnte beim Grenzkraftwerk Braunau-Simbach der Modellfehler in Form der Wurzel der mittleren quadratischen Abweichung (RMSE) im Vergleich zu PW bei der 4-Stunden-Prognose sowie über die fünf ausgewählten Zeitfenster um rund 63 % verringert werden. Beim Kraftwerk Aschach wurde eine Reduktion von 30 % erzielt, während beim Kraftwerk Greifenstein der RMSE mit ML um mehr als 50 % reduziert wurde. Es hat sich bei ML zudem gezeigt, dass mit kürzerer Prognosezeit auch die Prognosequalität deutlich verbessert wird, während sich diese bei PW in einem deutlich geringeren Ausmaß mit der Prognosezeit ändert. Es ist daher absehbar, dass ab einer bestimmten Prognosezeit PW gegenüber ML wieder im Vorteil ist. Nichtsdestotrotz könnte bei längerer Vorhersagezeit aber durch die Nachkopplung eines ML-Modells an PW die Prognosequalität weiter verbessert werden.

Schlüsselwörter Kurzfristige Leistungsprognose · Laufkraftwerk · Machine Learning

Performance of Machine Learning in short-term power forecasting within a run-of-river power plant chain

Abstract The output of a run-of-river power plant cannot be controlled without surge operation. Therefore, precise power forecasts are necessary in order to optimize sales of the generated electrical energy on the international electricity market. Currently, the Austrian hydropower plant operator Verbund AG uses a combination of hydrological and hydrodynamic models (PW) in opera-

tional use for power forecasting. However, this solution still has some deficits especially in the case of ascending and descending power curves. We therefore evaluate the potential of machine learning (ML) methods for short-term (up to 4 h) power forecasting at the run-of-river power plants Braunau-Simbach, Aschach and Greifenstein during five hydrologically interesting periods. The used input variables for ML consist of measured discharge and power from upstream run-of-river power plants and gauges. The obtained results show the advantages of applying ML in the domain of short-term power prediction within a run-of-river power plant chain. The model error, represented by the root mean squared error (RMSE), at the power plant Braunau-Simbach was reduced in the 5 selected periods by approximately 63% compared to PW in the 4-hour forecast. At Aschach a reduction of 30% was achieved, while the RMSE at Greifenstein was reduced by more than 50% when using ML. It could also be shown that the quality of the ML forecast is strongly dependent on the lead time. The forecast quality of ML increases significantly with shorter lead times, while this dependency is much weaker at the benchmark model PW. It is therefore expectable that PW will have an advantage over ML beyond a certain lead time. Nevertheless, it is possible that post-coupling an ML model to PW could further improve the forecast quality in applications with longer lead times.

Keywords Short-term power forecasting · Run-of-river power plant · Machine Learning

1 Einleitung

Der österreichische Wasserkraftwerksbetreiber Verbund AG hat als Stromerzeuger die Aufgabe, die erzeugte und verbrauchte Energie an den verfügbaren Handelsbörsen zu vermarkten. Insbesondere wird angestrebt, dass

Anmerkungen Die Gegenüberstellung der Zeitreihen für alle Vorhersagestufen (4 bis 1 h) und Validierungsperioden (Val1 bis Val5) sowie die „Feature Importance“, also die Relevanz der einzelnen Prädiktoren bei den ML-Modellen Random Forest sowie XGBoost, werden als Online-Material zur Verfügung gestellt.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s00506-022-00849-6>) enthalten.

DI C. Klingler (✉) · DI M. Feigl, B.Sc. · Univ.-Prof. Dipl.-Geoökol. Dr. K. Schulz
Department für Wasser-Atmosphäre-Umwelt, Institut für Hydrologie und Wasserwirtschaft, Universität für Bodenkultur Wien, Muthgasse 18, 1190 Wien, Österreich
christoph.klingler@boku.ac.at

Dr. T. Linsbichler · Dr. S. Frey
VERBUND Energy4Business GmbH,
Am Hof 6a, 1010 Wien, Österreich

die tatsächlich erbrachte Energie der vermarkteten Menge entspricht, um die ansonsten entstehenden Kosten für verursachte Ausgleichsenergie zu minimieren (Meeus et al. 2020). Für Laufkraftwerke stellt dies eine Herausforderung dar, da die Leistung nur sehr eingeschränkt gesteuert werden kann und daher prognostiziert werden muss. Prognoseabweichungen können zwar durch flexible Speicherkraftwerke kompensiert werden, dies widerspricht aber oftmals den Interessen der Speicheroptimierung. Durch die steigende Bedeutung des Intraday-Marktes (EEX 2019) können auch Prognosen kurz vor der Lieferung in der Vermarktung besser verwertet werden. Daraus ergibt sich das Interesse an möglichst präzisen kurzfristigen Leistungsprognosen für Laufkraftwerke.

Bei Verbund AG befinden sich für mehrere Laufkraftwerksketten (Donau, Inn, Mur) prozessbasierte Modelle zur Leistungsprognose im operationellen Betrieb, welche auf einer Kombination aus dem hydrologischen Modell COSERO (CONTinuous SEMI-distributed RunOff) und dem hydrodynamischen Modell Flux^{DSS/DESIGNER/FLORIS²⁰⁰⁰} (siehe z.B. Reichel et al. 2000; Reichel 2001) basieren. Das prozessbasierte Modell COSERO wurde am Institut für Hydrologie und Wasserwirtschaft (HyWa, früher IWHW) der Universität für Bodenkultur Wien zur Abflussvorhersage in alpinen Einzugsgebieten von Nachtnebel et al. (1993) entwickelt. COSERO zeichnet sich durch räumlich verteilte meteorologische Eingangsdaten, (semi-)empirische Abbildungen diverser physikalischer Prozesse, (nicht-)lineare Speicher sowie verschiedene Routingansätze aus. Eine kontinuierliche Weiterentwicklung sowie Validierung von COSERO erfolgte unter anderem im Rahmen von wasserwirtschaftlichen Studien (Nachtnebel und Fuchs 2004; Eder et al. 2005; Kling und Nachtnebel 2009a, b; Stanzel und Nachtnebel 2010; Herrnegger et al. 2012, 2015, 2018; Kling et al. 2012; Frey und Holzmann 2015; Klingler et al. 2020, 2021; Wesemann et al. 2018), der Erstellung operationeller Prognosesysteme (Stanzel et al. 2008; Schulz et al. 2016; Wesemann et al. 2018) sowie bei der Verwendung in klimatischen Studien (Kling et al. 2012, 2015; Mehdi et al. 2021). Da die prognostizierte Leistung bei Laufwasserkraftwerken stark vom prognostizierten Durchfluss abhängig ist, stellt das hydrologische Modell und

dessen meteorologische-Eingangsdaten die wichtigste Komponente dar. Obwohl die Wahl des hydrologischen Modells eine relevante Rolle in Bezug auf Prognoseunsicherheiten spielt (Plate und Shahzad 2015), ist die Unsicherheit bedingt durch die meteorologischen Eingangsdaten im Vergleich größer (Ramos et al. 2010; Zappa et al. 2010). Die Prognosen mittels der Kombination COSERO und Flux^{DSS/DESIGNER/FLORIS²⁰⁰⁰} (im Folgenden als PW-Modell bezeichnet) weisen derzeit an einigen Laufkraftwerken insbesondere bei an- sowie absteigenden Leistungsverläufen immer noch Defizite auf, obwohl diese durch die Implementierung der hydrodynamischen Modellierung teilweise bereits deutlich reduziert werden konnten.

Machine Learning (ML)-Modelle unterscheiden sich grundsätzlich von prozessbasierten Modellen, welche häufig bei hydrologischen Anwendungen verwendet werden. Während prozessbasierte Modelle eine Struktur und Parametrisierung aufweisen, die ein physikalisches System widerspiegelt, bestehen ML-Modelle aus flexiblen mathematischen Strukturen, welche an die verfügbaren Daten angepasst werden. Dementsprechend gibt es unterschiedliche Vor- und Nachteile beider Ansätze, genauso wie unterschiedliche Anwendungsgebiete. Während prozessbasierte Modelle eine klare Darstellung und Interpretation des Systems zulassen, gibt es immer noch Schwierigkeiten bei der konsistenten Modellparametrisierung speziell auf regionaler Ebene (z. B. Feigl et al. 2020; Klotz et al. 2017; Mizukami et al. 2017) und Ungewissheit über die Gesetzmäßigkeiten in Abhängigkeit von der Betrachtungsskala (z. B. Blöschl et al. 2019; Hrachowitz et al. 2013; Klemeš 1983). ML-Modelle weisen keinerlei physikalisch interpretierbare interne Strukturen auf und können potenziell auch unrealistische Ergebnisse liefern, haben aber den Vorteil, dass sie durch ihre Flexibilität komplexe Zusammenhänge, Strukturen und Muster in Daten sehr gut abbilden können. In den letzten Jahren konnte gezeigt werden, dass sich diese Flexibilität für eine Reihe von hydrologischen Problemstellungen – vor allem für kurzfristige Prognosen – eignet und deshalb ML-Modelle anstelle von oder in Kombination mit prozessbasierten Modellen für manche Anwendungen die derzeit besten Ergebnisse liefern (z. B. Feigl et al. 2021; Kratzert et al. 2021; Lees et al. 2021;

Li et al. 2021). Obwohl die Eignung von ML für die (hydrologische) Zeitreihenvorhersage außer Frage steht, läuft derzeit eine aktive Diskussion über die Einstufung sowie den Wert von ML für den wissenschaftlichen Bereich (z. B. Nearing et al. 2021). Die Anwendung von ML bei der kurzfristigen Leistungsprognose wurde beispielsweise schon für Windkraft- und Photovoltaikanlagen (z. B. Colak et al. 2012; Heineremann und Kramer 2016; Wang et al. 2019) oder für die Last im Stromnetz (z. B. Jin et al. 2021) untersucht. Zur Evaluierung des Potenzials von ML bei der kurzfristigen Leistungsprognose innerhalb einer Laufkraftwerkskette sind aber noch keine vorhandenen Arbeiten bekannt.

Deshalb stützen wir uns bei dieser Arbeit auf die Annahme, dass innerhalb eines Flussabschnitts mit größerem Einzugsgebiet bzw. innerhalb einer Laufkraftwerkskette (nicht-)lineare Zusammenhänge bestehen. Der Kern der vorliegenden Untersuchung liegt darin zu verifizieren, ob mittels ML-Modellen diese Zusammenhänge in den Zeitreihen von flussauf liegenden Kraftwerken und Pegeln erkannt sowie generalisiert werden können – und in weiterer Folge zu einer besseren kurzfristigen Leistungsprognose (bis 4 h) führen. Dazu werden die ML-Modelle a) schrittweise multiple lineare Regression (stepLM), b) Random Forest (RF; Breiman 2001) sowie c) eXtreme Gradient Boosting (XGBoost; Chen und Guestrin 2016) herangezogen. Die Ergebnisse dieser ML-Modelle werden an drei Laufkraftwerken bei verschiedenen hydrologischen Situationen den Prognosen des PW-Modells gegenübergestellt. Damit können wir schlussendlich das Potenzial von ML-Modellen für die kurzfristige Leistungsprognose charakterisieren und Empfehlungen für die Anwendung geben.

2 Methodik

2.1 Benchmark-Modell PW – Kombination COSERO mit Flux^{DSS/DESIGNER/FLORIS²⁰⁰⁰}

Um Durchflussprognosen mit COSERO an einem bestimmten Pegel erstellen zu können, stehen mehrere Möglichkeiten zur Verfügung: a) das gesamte flussaufwärtsgelegene Einzugsgebiet wird explizit mit COSERO modelliert, oder b) an definierten Modellrändern wird die Prognose eines (anderen) hydrologischen Modells als Eingang zu COSERO

Tab. 1 Technische Daten der ausgewählten Laufkraftwerke, Verbund AG/viadonau GmbH

	Braunau-Simbach	Aschach	Greifenstein
Abkürzung	GBR	BAS	BGS
Gewässer	Inn	Donau	Donau
Kilometrierung [km]	61,1	2162,67	1949,18
Seehöhe Oberwasser [m Mh]	349	280	177
Einzugsgebiet [km ²]	22.895	78.192	100.342
MQ [m ³ s ⁻¹], 2016–2020	689	1305	1788
Mq [l s ⁻¹ km ⁻²]	30,1	16,7	17,8
Ausbaudurchfluss [m ³ s ⁻¹]	1070	2480	3150
Mittl. Rohfallhöhe [m]	12,1	15,3	12,6
Engpassleistung [MW]	100	324	293
Regelarbeitsvermögen [GWh a ⁻¹]	550	1662	1717
Schleusenvolumen [m ³]	–	2 × je 93.942	2 × je 80.327
Füllzeit einer Schleusenammer [min]	–	ca. 13	ca. 15
Leistungsverlust durch Füllung einer Schleusenammer [MW] ^a	–	ca. 14,5	ca. 8,8

^a Annahme $\eta_{\text{ges}} = 80\%$

verwendet. Bei VERBUND BAP wurde die Variante b) gewählt, wobei alle einzelnen Gebiete mit COSERO modelliert werden. An den Modellgrenzen wird flussabwärts im Analyselauf der Messwert, sofern vorhanden, und in der Vorhersage die Prognose des flussaufwärtsgelegenen Einzugsgebietes übergeben. Die Modellerstellung gestaltet sich hierbei einheitlich und basiert auf physikalisch abgeleiteten Parametern der Landnutzung (CORINE, siehe Feranec 2016), des digitalen Gelände-modells (siehe Farr und Kobrick 2000), sowie aus geologischen und Bodenkarten aus dem hydrologischen Atlas Österreichs (BMLFUW 2007). Auf deren Basis werden *a-priori* Parameter geschätzt. Anschließend wird COSERO mittels SCE-UA (siehe z.B. Duan et al. 1993) kalibriert, wobei als Zielfunktion die Kling-Gupta Efficiency (Gupta et al. 2009) dient. Als meteorologischer Input (Temperatur und Niederschlag) wird mit INCA (Haiden et al. 2011) derselbe Datensatz verwendet, der auch im operativen Betrieb eingesetzt wird.

Operativ wird COSERO kontinuierlich eingesetzt. Dies bedeutet, dass laufend Analyseläufe, sogenannte Updateläufe, durchgeführt werden, um immer einen aktuellen Systemzustand zur Verfügung zu haben. Die Prognosen werden mit verschiedenen meteorologischen Inputs (GFS, AROME, INCA, ECMWF etc.) durchgeführt. Der kontinuierliche Betrieb von COSERO führt dazu, dass der modellierte Abfluss auch in der Analyse nicht exakt mit dem gemessenen Abfluss übereinstimmen muss. Daher werden die Ergebnisse von COSERO mittels eines

ARMA-Filters (siehe Broersen 2002; Kay und Marple 1981; Terminologie XX.AR) korrigiert. Die Leistung am Kraftwerk hängt maßgeblich vom Durchfluss sowie der Fallhöhe ab (siehe z.B. Dorn et al. 2016). Die Fallhöhe wird entweder a) mit einer kraftwerksspezifischen PQ-Beziehung abgeschätzt (Terminologie PW.PQ), oder b) hydrodynamisch prognostiziert (Terminologie PW). Im Fall von b) bekommt das hydrodynamische Modell Flux^{DSS/DESIGNER}/FLORIS²⁰⁰⁰ seine Zuflüsse am Modellrand einerseits aus gemessenen Zeitreihen (in der Vergangenheit) und andererseits aus mittels COSERO modellierten Zufluss-Zeitreihen (in der Prognose sowie an Zubringern, die nicht über Pegelmessstellen verfügen). Auch die Ergebnisse der berechneten Leistung basierend auf der hydrodynamischen Modellierung werden mit demselben ARMA-Filter unter Verwendung der gemessenen Leistung korrigiert.

In diesem Beitrag werden nur Ergebnisse der Modellkonfiguration PW.INCA.AR (meteor. INCA-Prognosedaten, hydrodynamische Modellierung der Fallhöhe sowie Korrektur der Prognose mit ARMA-Filter) dargestellt, da einerseits INCA aufgrund der Architektur sowie der hohen zeitlichen Auflösung die stets aktuell verfügbare Prognose darstellt und PW.INCA.AR zumindest bei den untersuchten Kraftwerken und Zeiträumen im Mittel auch die besten Ergebnisse liefert.

2.2 Machine-Learning-Modelle

Im Folgenden sollen kurz die verwendeten ML-Methoden beschrieben werden.

Bei der schrittweisen multiplen linearen Regression wird durch eine iterative Variablenselektion ein multiples lineares Regressionsmodell generiert, welches die Daten bestmöglich abbildet. Random Forest (Breiman 2001) ist ein Klassifikations- und Regressionsverfahren und basiert auf der Mittelung der Vorhersagen von mehreren unabhängigen Regressionsbäumen. Jedem Regressionsbaum wird durch Bagging (Breiman 1996) nur eine Bootstrap-Stichprobe der gesamten Trainings-Daten zugeteilt, des Weiteren verwendet jeder Regressionsbaum nur eine individuelle und zufällige Anzahl an Prädiktoren. XGBoost (Chen und Guestrin 2016) basiert so wie auch RF auf Regressionsbäumen, nur sind diese nicht mehr unabhängig und parallel, sondern seriell (also aufeinander aufbauend) strukturiert. Aufgrund der seriellen Architektur sowie einer erhöhten Anzahl an Hyperparametern nimmt das Training des Netzwerks bei XGBoost im Vergleich zu RF deutlich mehr Zeit in Anspruch. Für eine genauere Beschreibung der ausgewählten ML-Modelle möchten wir auf Feigl et al. (2021), Abschn. 2.4.1 (stepLM), Abschn. 2.4.2 (RF) sowie Abschn. 2.4.3 (XGBoost) verweisen.

2.3 Kraftwerksfestlegung- und Prädiktorenauswahl für die Prognose mit ML

In der vorliegenden Untersuchung wird das Augenmerk auf das Innkraftwerk Braunau-Simbach sowie die beiden Donaukraftwerke Aschach und Greifenstein gelegt, um eine große Distanz

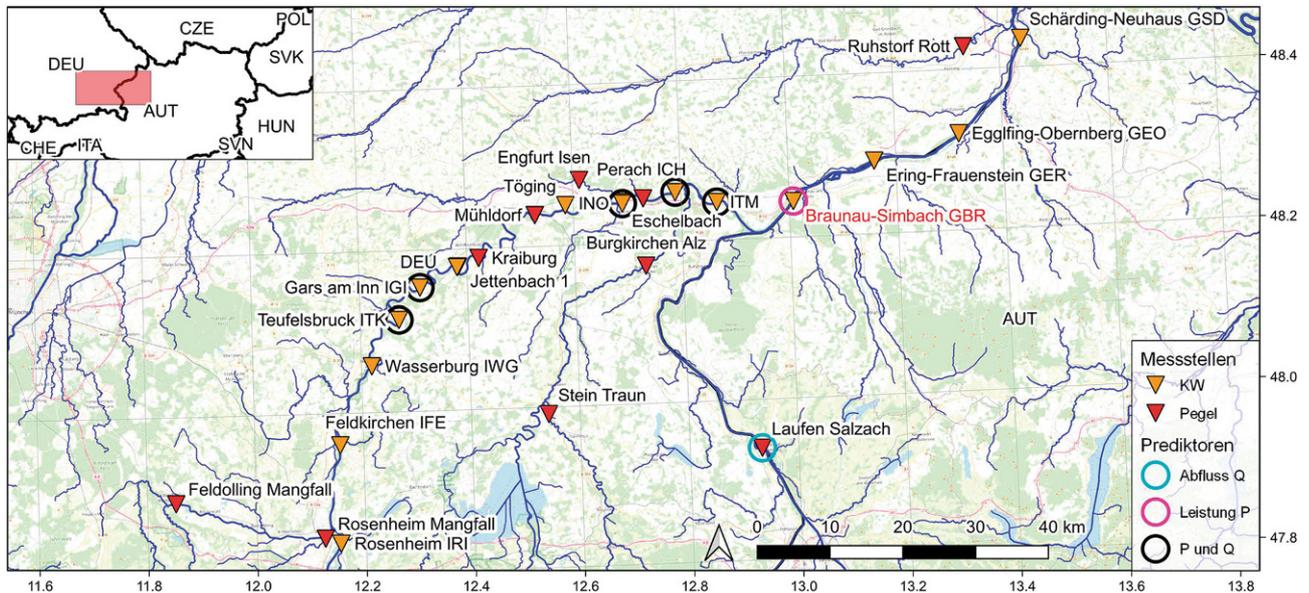


Abb. 1 Übersicht KW Braunau-Simbach, Messnetz: Verbund AG, Gewässernetz: EU-Hydro – River Network Database (EEA 2019), Hintergrundkarte: OpenStreetMap, Staatsgrenzen: ©EuroGeographics, Koordinaten: WGS84, Prädiktorenauswahl entspricht Auswahl GBR6 in Tab. 2

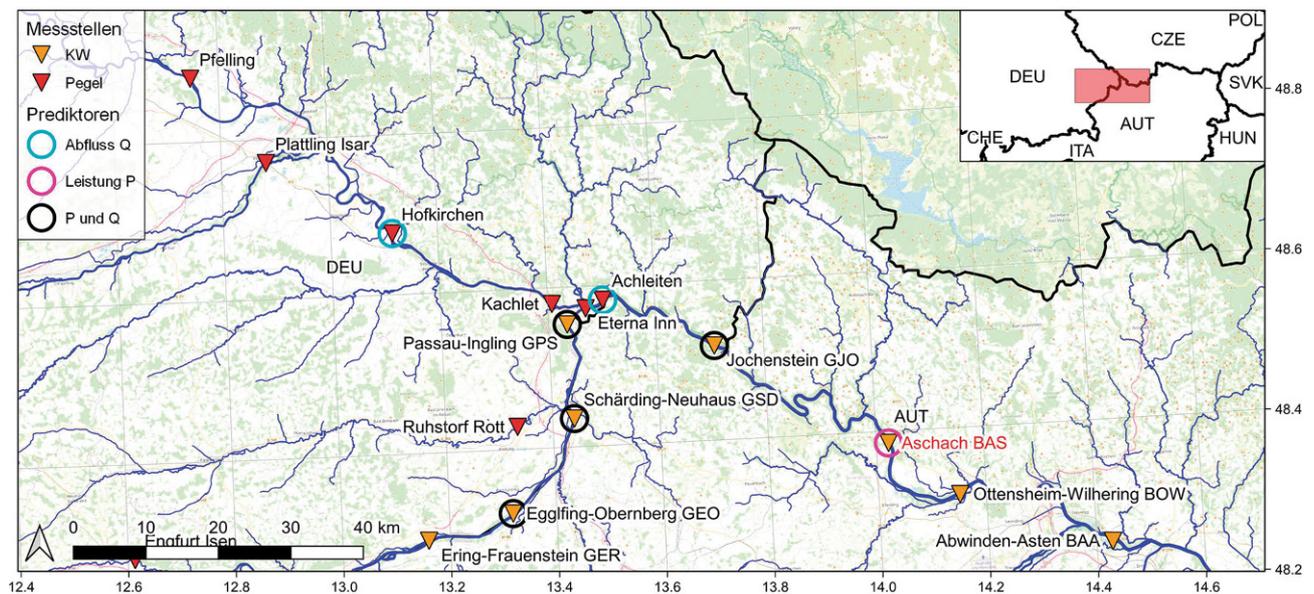


Abb. 2 Übersicht KW Aschach, Messnetz: Verbund AG sowie GKD Bayern, Gewässernetz: EU-Hydro – River Network Database (EEA 2019), Hintergrundkarte: OpenStreetMap, Staatsgrenzen: ©EuroGeographics, Koordinaten: WGS84, Prädiktorenauswahl entspricht Auswahl BAS3 in Tab. 3

zwischen den Kraftwerken zu gewährleisten. Grundlegende technische Daten dieser zur Verbund AG zugehörigen Laufkraftwerke sind in Tab. 1 aufgelistet.

Im Gegensatz zu COSERO werden bei der kurzfristigen Leistungsprognose mit ML keine meteorologischen Vorhersagen, sondern primär die gemessenen Werte von Kraftwerken und Pegeln

stromauf des betreffenden Kraftwerks sowie teilweise auch Informationen betreffend der Jahres- („fuzzy months“) als auch Tageszeit („fuzzy daytime“) als Prädiktoren verwendet. Unter „fuzzy“ wird die Transformation von diskreten Zeitpunkten zu kontinuierlichen Werten (z.B. über Sinusfunktion) verstanden, um Sprünge zum Jahres- oder Tageswechsel zu vermeiden. Als Prä-

diktoren stehen bei Kraftwerken grundsätzlich die Messgrößen Abfluss und Leistung (bei Donaukraftwerken auch Wasserstand des Ober- und Unterwassers sowie der Wehrüberfall) sowie bei Pegeln der Abfluss zur Verfügung. Eine Übersicht der zur Verfügung stehenden Kraftwerke und Pegel ist in Abb. 1, 2 und 3 ersichtlich. Die Prädiktorenauswahl für die ML-Modelle

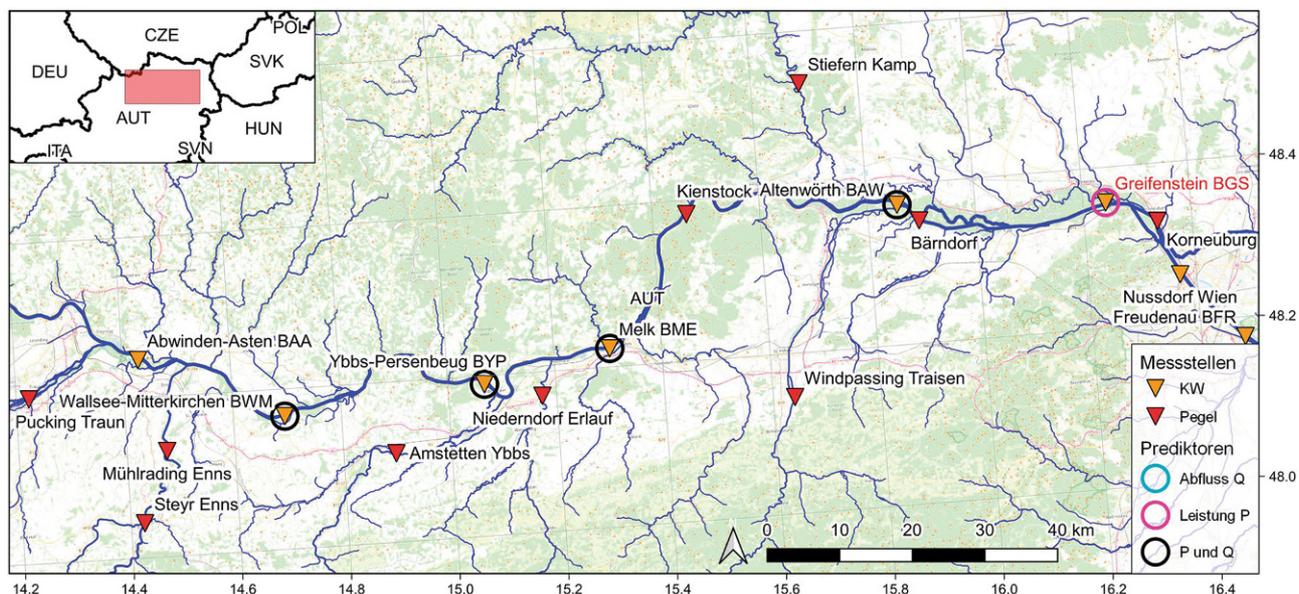


Abb. 3 Übersicht KW Greifenstein, Messnetz: Verbund AG, Gewässernetz: EU-Hydro – River Network Database (EEA 2019), Hintergrundkarte: OpenStreetMap, Staatsgrenzen: © EuroGeographics, Koordinaten: WGS84, Prädiktorenauswahl entspricht Auswahl BGS3 in Tab. 4

wird grundsätzlich so ausgelegt, dass einerseits so viele Prädiktoren für eine ausreichend zuverlässige Prognose wie notwendig, aber andererseits auch so wenige wie möglich implementiert werden, um die Komplexität des trainierten Modells zu begrenzen sowie die Anfälligkeit gegenüber technischen Störungen (z.B. Datenübertragung) zu minimieren (für Prognoseberechnung müssen Werte für alle Prädiktoren vorliegen). Da die Anzahl an möglichen Kombinationen für die Prädiktorenauswahl sehr hoch ist (zahlreiche Messstellen flussauf des jeweiligen Kraftwerks, eine Vielzahl an möglichen „lags“ – also Werte der Prädiktoren aus der Vergangenheit, z.B. 1h vor dem Zeitpunkt der Prognoseberechnung, „fuzzy months“, „fuzzy daytime“), würde sich eine automatisierte Suche nach der bestmöglichen Kombination der Prädiktoren als sehr rechenintensiv gestalten. Deshalb wird die Prädiktorenauswahl nach einem möglichst systematischen „Trial und Error“-Verfahren festgelegt. Die Kombination der Eingangsgrößen bei den einzelnen durchgeführten Versuchen wird in Tab. 2, 3 und 4 indiziert. Dabei sind die Kombinationen der einzelnen Versuche so festgelegt, dass zunehmend mehr Prädiktoren implementiert werden und somit die bereits beschriebene optimale Konfiguration – aus so vielen Prädiktoren wie notwendig, aber so wenigen wie möglich – gefunden werden kann. Zusätzlich wird

bei dieser Optimal-Konfiguration (siehe Tab. 7, 8 und 9) noch getestet, ob die Erweiterung der Prädiktoren um „fuzzy months“ (Indikatoren für die Jahreszeit), „fuzzy daytime“ (Indikatoren für die Tageszeit) oder „lags“ (Werte eines Messwerts aus der Vergangenheit) eine zuverlässigere Prognose ermöglicht.

2.3.1 Grenzkraftwerk Braunau-Simbach (GBR)

Das Grenzkraftwerk Braunau-Simbach (GBR, Abb. 1) ist das erste Laufkraftwerk flussab der Einmündung der Salzach (MQ $252 \text{ m}^3 \text{ s}^{-1}$, Pegel Burghausen, 1901–2006) in den Inn (MQ siehe Tab. 1.1). Das erste Laufkraftwerk flussauf der Salzach-Mündung ist das Innkraftwerk Stammham, welches auch den Innzubringer Alz (exkl. Alzkanal) abarbeitet. Im Unterlauf der Alz wird jedoch der überwiegende Anteil in den Alzkanal (Ausbaudurchfluss ca. $90 \text{ m}^3 \text{ s}^{-1}$, MQ $57 \text{ m}^3 \text{ s}^{-1}$, Pegel Guffham, 1973–2011) eingespeist, welcher nach energetischer und thermischer Nutzung bei Burghausen in die Salzach eingeleitet wird. Da das Triebwasser für den Alzkanal jedoch flussauf des Pegels Burghausen entnommen wird, steht keine Information über den Gesamtdurchfluss der Alz für die Leistungsprognose des KW Braunau-Simbach zur Verfügung.

In der Tab. 2 sind die Prädiktoren-Kombinationen der einzelnen Versuche

zum KW Braunau-Simbach (GBR) aufgelistet. Die Messstellen Jettenbach 1 & 2, Kraiburg, Mühlndorf sowie Töging (Abb. 1) werden nicht herangezogen, da diese von der ca. 23 km langen Ausleitung des Innwerkanals (Ausbaudurchfluss ca. $340 \text{ m}^3 \text{ s}^{-1}$) maßgebend beeinflusst sind. Als die am weitesten stromauf gelegene Messstelle wurde am Inn das KW Rosenheim (IRI) ausgewählt, welches 126,45 km stromauf von GBR liegt, während der Pegel Laufen Salzach im vorliegenden Versuch die einzige Messstelle an der Salzach ist und sich ca. 54,6 Fließ-km oberhalb von GBR befindet.

2.3.2 Laufkraftwerk Aschach (BAS)

Das Laufkraftwerk Aschach (BAS, Abb. 2) ist das erste Donau-Laufkraftwerk, welches zur Gänze auf österreichischem Staatsgebiet liegt, und ist ca. 62,6 Fließ-km unterhalb von Passau stationiert, wo der Inn (MQ $740 \text{ m}^3 \text{ s}^{-1}$, Pegel Passau Ingling, 1921–2006) sowie die Ilz (MQ $16,1 \text{ m}^3 \text{ s}^{-1}$, Pegel Kalteneck, 1921–2012) in die Donau münden. Zubringer stromab des nächst oberliegenden Donaukraftwerks Jochenstein – und damit nicht messtechnisch für die Leistungsprognose verfügbar – sind die Ranna (MQ $3,1 \text{ m}^3 \text{ s}^{-1}$, Pegel Oberkappel, 1951–2010), die kleine Mühle (MQ $3,3 \text{ m}^3 \text{ s}^{-1}$, Pegel Obermühl, 1976–2011) sowie die große Mühle (MQ $8,6 \text{ m}^3 \text{ s}^{-1}$, Pegel Teufelmühle, 1951–2011), wobei

Tab. 2 Prädiktoren-Kombination der einzelnen Versuche für die kurzfristige Leistungsprognose beim KW Braunau-Simbach (GBR)

GBR1	p_GBR
GBR2	p_ITM p_GBR
GBR3	pr_ITM p_GBR
GBR4	pr_INO pr_ICH pr_ITM p_GBR
GBR5	r_LS pr_INO pr_ICH pr_ITM p_GBR
GBR6	pr_ITK pr_IGI r_LS pr_INO pr_ICH pr_ITM p_GBR
GBR7	pr_IWG pr_ITK pr_IGI r_LS pr_INO pr_ICH pr_ITM p_GBR
GBR8	pr_IRI pr_IFE pr_IWG pr_ITK pr_IGI r_LS pr_INO pr_ICH pr_ITM p_GBR
GBR9	Prädiktoren von GBR6 sowie fm fd
GBR10	Prädiktoren von GBR6 sowie jeweils lag4-Daten

Abkürzung: *p* Leistung, *r* Abfluss, *pr* Leistung und Abfluss, *lag4* Wert des Prädiktors 1 h (4×15 min) vor dem Zeitpunkt der Prognoseberechnung, *fm* „fuzzy months“ (12 Prädiktoren), *fd* „fuzzy daytime“, *IRI* KW Rosenheim, *IFE* KW Feldkirchen, *IWG* KW Wasserburg, *ITK* KW Teufelsbruck, *IGI* KW Gars am Inn, *INO* KW Neuötting Inn, *ICH* KW Perach, *ITM* KW Stammham, *GBR* KW Braunau-Simbach, *LS* Pegel Laufen Salzach (Zeitreihenbeginn Mitte April 2013).

Tab. 3 Prädiktoren-Kombination der einzelnen Versuche für die kurzfristige Leistungsprognose beim KW Aschach (BAS)

BAS1	pr_GJO p_BAS
BAS2	pr_GPS r_AD pr_GJO p_BAS
BAS3	r_HD pr_GEO pr_GSD pr_GPS r_AD pr_GJO p_BAS
BAS4	pr_GER r_HD pr_GEO pr_GSD pr_GPS r_AD pr_GJO p_BAS
BAS5	r_PD r_PI pr_GER r_HD pr_GEO pr_GSD pr_GPS r_AD pr_GJO p_BAS
BAS6	Prädiktoren von BAS3 sowie fm fd
BAS7	Prädiktoren von BAS3 sowie jeweils lag4-Daten

Abkürzung: *p* Leistung, *r* Abfluss, *pr* Leistung und Abfluss, *lag4* Wert des Prädiktors 1 h (4×15 min) vor dem Zeitpunkt der Prognoseberechnung, *fm* „fuzzy months“ (12 Prädiktoren), *fd* „fuzzy daytime“, *GER* KW Ering-Frauenstein, *GEO* KW Eggfing-Obernberg, *GSD* KW Schärding-Neuhaus, *GPS* KW Passau-Ingling, *GJO* KW Jochenstein, *BAS* KW Aschach, *PD* Pegel Pfelling Donau (GKD 2021), *PI* Pegel Plattling Isar (GKD 2021), *HD* Pegel Hofkirchen Donau (Zeitreihenbeginn Mitte August 2016), *AD* Pegel Achleiten Donau (Zeitreihenbeginn Mitte August 2016).

Tab. 4 Prädiktoren-Kombination der einzelnen Versuche für die kurzfristige Leistungsprognose beim KW Greifenstein (BGS)

BGS1	pr_BAW p_BGS
BGS2	pr_BYP pr_BME pr_BAW p_BGS
BGS3	pr_BWM pr_BYP pr_BME pr_BAW p_BGS
BGS4	pr_BAA pr_BWM pr_BYP pr_BME pr_BAW p_BGS
BGS5	Prädiktoren von BGS3 sowie r_ME
BGS6	Prädiktoren von BGS3 sowie r_SK r_WT
BGS7	Prädiktoren von BGS3 sowie fm fd
BGS8	Prädiktoren von BGS3 sowie jeweils lag4-Daten

Abkürzung: *p* Leistung, *r* Abfluss, *pr* Leistung und Abfluss, *lag4* Wert des Prädiktors 1 h (4×15 min) vor dem Zeitpunkt der Prognoseberechnung, *fm* „fuzzy months“ (12 Prädiktoren), *fd* „fuzzy daytime“, *BAA* KW Abwinden-Asten, *BWM* KW Wallsee-Mitterkirchen, *BYP* KW Ybbs-Persenbeug, *BME* KW Melk, *BAW* KW Altenwörth, *BGS* KW Greifenstein, *ME* Pegel Mühlradring Enns, *SK* Pegel Stiefern Kamp (Zeitreihenbeginn Mitte August 2016), *WT* Pegel Windpassing Traisen.

diese in Relation zur Donau (MQ siehe Tab. 1.2) im Mittel sehr kleine Zubringer sind.

Die Kombinationen der Prädiktoren zum KW Aschach (BAS) sind in Tab. 3 aufgelistet. Der Pegel Pfelling ist dabei in der Untersuchung die am weitesten entfernte Messstelle an der Donau, welche 142,86 Fließ-km stromauf von BAS stationiert ist, während das am weitesten entfernte implementierte Inn-KW Ering-Frauenstein (GER) ca. 110,6 Fließ-km oberhalb von BAS liegt.

2.3.3 Laufkraftwerk Greifenstein (BGS)

Das Laufkraftwerk Greifenstein (BGS, Abb. 3) ist das vorletzte Donaukraftwerk auf österreichischem Territorium und liegt ca. 16km stromauf von Wien (KW Nussdorf). Größere Zubringer, welche nicht beim nächst stromauf liegenden Donaukraftwerk Altenwörth erfasst werden, sind linksufrig der Kamp (MQ $8,9\text{m}^3\text{s}^{-1}$, Pegel Stiefern, 1983–2007) und die Krems (MQ $1,9\text{m}^3\text{s}^{-1}$, Pegel Imbach, 1981–2011) sowie rechtsufrig die Traisen (MQ $14,2\text{m}^3\text{s}^{-1}$, Pegel Windpassing, 1981–2011). Das Verhältnis der Summe der Mittelwerte dieser drei Zubringer zum mittleren Abfluss

der Donau (MQ siehe Tab. 1.3) beträgt ca. 1,4%, wobei die Abflüsse am Pegel Stiefern sowie Windpassing als Prädiktoren implementiert werden können (Abb. 3).

Tab. 4 zeigt die Auflistung der gewählten Prädiktoren-Kombinationen zum KW Greifenstein (BGS). Das am weitesten stromauf gelegene Donau-Laufkraftwerk in Tab. 4 ist dabei das KW Abwinden-Asten, welches 170,27 Fließ-km oberhalb von BGS liegt. Getestet wird bei der optimalsten Prädiktoren-Kombination (Tab. 9) auch, ob die zusätzliche Implementierung der Abflusswerte der Pegel Mühlradring Enns, Stiefern Kamp sowie Windpassing Trai-

Hier steht eine Anzeige.



Hier steht eine Anzeige.



Tab. 5 Trainings- und Validierungsperioden

	Zeitraum	Hydrologische Situation
Trn	01.01.2013 ^a 00:00 bis 30.04.2019 23:00	–
Val	01.05.2019 00:00 bis 30.04.2021 23:00	–
Val1	10.05.2019 00:00 bis 16.05.2019 23:00	Bereich zw. Mittelwasser und kleinerem Hochwasser
Val2	01.03.2020 00:00 bis 07.03.2020 23:00	Mittelwasser
Val3	01.08.2020 00:00 bis 07.08.2020 23:00	Kleineres Hochwasser, Zufluss übersteigt Ausbaudurchfluss der Kraftwerke in Tab. 1
Val4	25.11.2020 00:00 bis 01.12.2020 23:00	Niederwasser
Val5	24.04.2021 00:00 bis 30.04.2021 22:00	ausgeprägter Tagesgang durch Schneeschmelze, unterhalb Mittelwasser

^a Abhängig vom spätesten Beginn der Zeitreihe eines ausgewählten Prädiktors

Tab. 6 Trainingsdauer [min] sowie maximale RAM Auslastung [GB] bei den verschiedenen ML-Modellen, 4 h Prognosezeit

Anzahl	stepLM		RF		XGBoost	
	Dauer ^a [min]	RAM ^b [GB]	Dauer ^a [min]	RAM ^b [GB]	Dauer ^a [min]	RAM ^b [GB]
2	0,15 0,02	14	5,21 0,06	43	173,37 0,03	25
6	0,61 0,03	16	16,57 0,1	49	242,71 0,22	26

^a Training (Trn) | Validierung (Val)
^b Basisauslastung von 10 GB inkludiert

sen als Prädiktoren zu geringeren Residuen in der Prognose führen (BGS5 und BGS6 in Tab. 4).

2.4 Training und Validierung der ML-Modelle

Obwohl die vorhandenen Zeitreihen überwiegend eine 15-minütige Auflösung aufweisen, erfolgt das Training (das Anpassen an die Daten) der ML-Modelle auf Stundenbasis. Einerseits liefert eine 15-minütige Auflösung speziell bei den Pegelwerten gegenüber der stündlichen Auflösung meist keinen Mehrwert, da diese teilweise vorab auf Basis der Stundenwerte interpoliert wurden und andererseits kann mit dieser Maßnahme die Gefahr der Überanpassung der Modelle an die Trainingsdaten (Overfitting) sowie die Ressourcenauslastung des Rechner-Systems reduziert werden. Der Beginn der Zeitreihen der meisten Laufkraftwerke ist Anfang Januar 2013. Nur beim Donau-KW Jochenstein ist es Mitte Juli 2016 und bei den Inn-KW Oberaudorf-Ebbs, Nussdorf, Braunau-Simbach, Ering-Frauenstein Eggfing-Obernberg, Schärding-Neuhaus, Passau-Ingling liegt dieser bei Anfang April 2015. Der Beginn der Abfluss-Zeitreihen der Pegel liegt in der Bandbreite von Anfang Januar 2013 bis Mitte August 2016 und wird daher – falls der entsprechende Pegel als Prädiktor ausgewählt wurde und der jeweilige Beginn nach Anfang Januar 2013 liegt – im Zusatz zu Tab. 2, 3, 4 indiziert. Dies ist insofern relevant, da sowohl beim Training eines Modells als auch bei der späteren Berechnung der

Prognose kein Mitglied des Eingangs-Ensembles – welches die Werte für die definierten Prädiktoren umfasst – einen fehlenden Wert aufweisen darf. Entsprechende Ensembles werden daher aus den Trainings-Daten entnommen und verkürzen den Trainingszeitraum dementsprechend. Das Ende der meisten Zeitreihen liegt bei Ende April 2021 und ist andernfalls ebenfalls indiziert.

Alle ML-Modelle werden mit der Zielfunktion RMSE (Root Mean Square Error; Gl. 1) sowie einer 10-fachen Kreuzvalidierung trainiert. Um eine differenzierte Handhabung zu ermöglichen, wird für jede einzelne Vorhersagestufe (1/2/3/4h) ein eigenes Modell erstellt. Bei den ML-Modellen RF sowie XGBoost müssen vor dem Start des Trainings sogenannte Hyperparameter – welche Grundeigenschaften des Modells definieren – festgelegt werden. Häufig werden diese einfach mit „Standardwerten“ belegt oder anhand eines „Trial und Error“-Verfahrens festgelegt. Da die Festlegung der Hyperparameter jedoch einen großen Einfluss auf die Modellgüte des trainierten Netzwerks haben kann (Claesen und De Moor 2015; Feigl et al. 2021), werden diese in der vorliegenden Untersuchung mit der Bayes'schen Optimierungsmethode (Kushner 1964; Moćkus 1975, 1989; Moćkus et al. 1978; Zhilinskas 1975) systematisch per Algorithmus bestimmt. Für eine Beschreibung der Bayes'schen Methode möchten wir auf Feigl et al. (2021) verweisen. Die Anwendungsbereiche sowie Grenzwerte der Hyperparameteroptimierung bei den Modellen RF (2 Hyperparameter)

sowie XGBoost (6 Hyperparameter) sind kongruent mit der Untersuchung von Feigl et al. (2021), welche im dortigen Anhang aufgelistet sind.

Die Validierung erfolgt in fünf verschiedenen Perioden mit jeweils einer 1-wöchigen Dauer (Tab. 5), wobei das Augenmerk bei der Evaluierung aus Gründen der Übersichtlichkeit auf den Vorhersagen jener Modelle mit der längsten untersuchten Vorhersagezeit (4 h) gelegt wird, da davon auszugehen ist, dass bei dieser Stufe die Residuen im Mittel am größten sind. Zeitreihen für alle vier Vorhersagestufen inklusive der dabei erzielten Gütemaße werden jedoch als Online-Material zur Verfügung gestellt. Für die Validierung der ML-Modelle als auch des Benchmark-Modells PW kommt neben dem RMSE auch der MAE (Mean Absolute Error; Gl. 2) zur Anwendung. Aufgrund der im Vergleich zu den anderen ML-Modellen deutlich längeren Trainingszeit bei XGBoost wird dieses Modell ausschließlich mit der optimalen Prädiktoren-Konfiguration (Tab. 7, 8, 9) trainiert und die dabei erzielte Modellgüte validiert.

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

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

In Gl. 1 sowie 2 stellt n die Stichprobengröße (Anzahl der Zeitschritte), y_i die vorhergesagte sowie \hat{y}_i die beobachtete (gemessene) Leistung des Kraftwerks zum Zeitpunkt i dar.

Tab. 7 Gütemaße der unterschiedlichen Modelle sowie Versuche für 4 h Prognosezeit bei GBR [MW]

c	PW (PW.INCA.AR)		stepLM		Random Forest		XGBoost	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
GBR1 ^a	d	d	5,78	3,14	6,02	3,48	–	–
GBR2 ^a	–	–	5,77	3,13	5,81	3,25	–	–
GBR3 ^a	–	–	5,71	3,12	6,43	3,27	–	–
GBR4 ^a	–	–	4,97	2,89	5,71	3,06	–	–
GBR5 ^a	–	–	5,54	2,91	5,32	2,47	–	–
GBR6 ^a	8,43	4,65	3,51	2,28	3,14	2,02	3,16	2,11
GBR7 ^a	–	–	3,42	2,16	3,1	1,99	–	–
GBR8 ^a	–	–	4,22	2,37	3,13	1,97	–	–
GBR9 ^a	–	–	–	–	3,44	2,21	–	–
GBR10 ^a	–	–	–	–	3,0	1,92	–	–
GBR6 ^b	–	–	2,97	2,04	0,89	0,55	1,31	0,96

^a Zeitfenster Val1 bis Val5 in Tab. 5

^b Trainingszeitraum Trn lt. Tab. 5 für Optimal-Konfiguration. ^c Abkürzung siehe Tab. 2

^d Die Gütemaße von PW werden nur bei der optimalen Prädiktoren-Kombination für ML als Benchmark angegeben

Tab. 8 Gütemaße der unterschiedlichen Modelle sowie Versuche für 4 h Prognosezeit bei BAS [MW]

c	PW (PW.INCA.AR)		stepLM		Random Forest		XGBoost	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
BAS1 ^a	d	d	19,33	9,6	11,51	7,66	–	–
BAS2 ^a	–	–	11,47	7,22	10,62	7,64	–	–
BAS3 ^a	10,63	7,43	62,88	16,6	8,38	5,62	10,98	6,92
BAS4 ^a	–	–	90,03	21,17	8,18	5,27	–	–
BAS5 ^a	–	–	77,24	17,85	7,35	4,76	–	–
BAS6 ^a	–	–	–	–	8,21	5,73	–	–
BAS7 ^a	–	–	–	–	8,49	5,52	–	–
BAS3 ^b	–	–	7,47	4,85	2,79	1,75	0,44	0,34

^a Zeitfenster Val1 bis Val5 in Tab. 5

^b Trainingszeitraum Trn lt. Tab. 5 für Optimal-Konfiguration

^c Abkürzung siehe Tab. 3

^d Die Gütemaße von PW werden nur bei der optimalen Prädiktoren-Kombination für ML als Benchmark angegeben

3 Ergebnisse und Diskussion

3.1 Aufwand der Modellerstellung

Die Aufbereitung und Homogenisierung der Eingangsdaten – welche im vorliegenden Fall die Zeitreihen der Prädiktoren darstellen – war unproblematisch, da die Rohdaten bereits in einheitlicher Form von der Verbund AG zur Verfügung gestellt wurden. Der Hauptaufwand bei der Umsetzung der kurzfristigen Leistungsprognose mit ML lag bei der Erstellung der Algorithmen für das Training der verschiedenen Modelle, wobei die frei verfügbare Software R (R Core Team 2020) mit Implementierung von einigen ebenfalls frei verfügbaren Packages (z. B. Caret, Kuhn 2019; data.table, Dowe et al. 2014; XGBoost, Chen et al. 2016) verwendet wurde. Sobald die erstellten Algorithmen nach intensiver Prüfung voll funktionstüchtig sind, verlegt sich der Schwerpunkt des Arbeitsaufwands bei Anwendung

der Methodik der Leistungsprognose mittels ML an anderen Kraftwerken auf die optimale Auswahl der Prädiktoren. Trainierte Modelle können als Files exportiert und auf andere Rechensysteme transferiert werden. Im operativen Betrieb sind dann nur noch wenige Zeilen Code für die Ausführung mittels R erforderlich.

3.2 Testsystem und Ressourcenauslastung

Als Berechnungssystem kommt für das Training der ML-Modelle ein Server-PC zur Anwendung, welcher über 16 CPU-Kerne mit jeweils 2,3 GHz sowie 256 GB RAM verfügt. Grundsätzlich ist die Trainingsdauer sowie die maximale RAM-Auslastung bei allen untersuchten ML-Modellen vom Umfang der Trainingsdaten (Anzahl Eingangs-Ensembles sowie auch deren Mitglieder) abhängig. Um den Unterschied zwischen den ML-Modellen zu zeigen, wird jeweils

die Trainings- und Validierungsdauer sowie die maximale RAM-Auslastung (3 s Abtastintervall, Basisauslastung inkludiert) für 2 sowie 6 Prädiktoren – für welche auch Daten für den vollständigen Trainings- sowie Validierungszeitraum (Tab. 5, Trn bzw. Val) vorliegen – in Tab. 6 aufgelistet.

Die unterschiedliche Architektur von RF sowie XGBoost spiegelt sich auch in Tab. 6 anschaulich wider: Die serielle Struktur der Regressionsbäume bei XGBoost führt im Gegensatz zur parallelen Struktur eines RF zu einer geringeren maximalen RAM-Auslastung, dafür aber auch zu einer deutlich längeren Trainingszeit (Multi-Core-Prozessoren verlieren dabei ihren Vorteil). Multiple lineare Regression ist zwar ein sehr einfaches Modell, welches schnell parametrisiert werden kann, jedoch steigt bei stepLM die Trainingszeit exponentiell mit der Anzahl der Prädiktoren an und liegt z. B. bei 18 Prädiktoren (GBR8 in Tab. 2) aufgrund der großen Zahl

Tab. 9 Gütemaße der unterschiedlichen Modelle sowie Versuche für 4 h Prognosezeit bei BGS [MW]

c	PW (PW.INCA.AR)		stepLM		Random Forest		XGBoost	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
BGS1 ^a	^d	^d	7,34	4,87	7,23	4,71	–	–
BGS2 ^a	–	–	6,64	3,83	4,36	3,11	–	–
BGS3 ^a	8,21	6,1	8,58	4,03	4,15	2,99	3,94	2,94
BGS4 ^a	–	–	6,52	3,47	4,06	2,97	–	–
BGS5 ^a	–	–	8,33	3,95	4,08	2,96	–	–
BGS6 ^a	–	–	5,98	3,58	4,56	3,08	–	–
BGS7 ^a	–	–	–	–	4,06	3,02	–	–
BGS8 ^a	–	–	–	–	3,97	2,97	–	–
BGS3 ^b	–	–	3,97	2,78	1,38	0,95	1,1	0,83

^a Zeitfenster Val1 bis Val5 in Tab. 5

^b Trainingszeitraum Trn lt. Tab. 5 für Optimal-Konfiguration.

^c Abkürzung siehe Tab. 4. ^d Die Gütemaße von PW werden nur bei der optimalen Prädiktoren-Kombination für ML als Benchmark angegeben

an theoretisch möglichen Kombinationen ($2^{18}-1$) bereits bei 401 min. Deshalb werden die Versuche GBR9 ($2^{25}-1$ Kombinationsmöglichkeiten), GBR10 ($2^{24}-1$) in Tab. 2; BAS6 ($2^{24}-1$), BAS7 ($2^{22}-1$) in Tab. 3 sowie BGS7 ($2^{22}-1$), BGS8 ($2^{18}-1$) in Tab. 4 mit stepLM nicht durchgeführt. Bei allen Tests liegt die Berechnungsdauer für die Validierung (Prognose) deutlich unter jener für das Training. Im operationellen Betrieb liegt die Berechnungsdauer eines Prognose-Ensembles (1/2/3/4h) bei allen untersuchten ML-Modellen im Bereich von ca. 5 s, sofern die trainierten Netzwerke einmal in den RAM geladen sind. Eine Vorhersage mit der Modellkombination PW benötigt zum Vergleich operativ mehrere Minuten. Erwähnenswert ist allerdings, dass hier a) eine längere Zeitreihe prognostiziert wird und b) die Ergebnisse vollverteilter hydrologischer Prozesse in Form von Rastern geschrieben werden. Diese können dann zur Plausibilisierung der Ergebnisse herangezogen werden.

3.3 Grenzkraftwerk Braunau-Simbach (GBR)

Tab. 7 zeigt die Auflistung der erzielten Gütemaße RMSE sowie MAE über alle fünf Validierungsperioden Val1 bis Val5 (Tab. 5) für die 4-h-Prognosestufe. An dieser Stelle wird noch einmal die Engpassleistung des betreffenden Kraftwerks GBR von 100 MW (Tab. 1) angeführt. Sowohl bei stepLM als auch bei RF ist eine deutliche Verbesserung der Prognose ausgehend von der Kombination GBR5 auf GBR6 (Tab. 2) erkennbar, wohingegen sich bei zusätzlicher Implementierung von weiter stromaufgelegenen Kraftwerken als Prädiktoren (GBR7, GBR8) zumindest bei RF nur

noch eine marginale Reduktion der mittleren Residuen zeigt. Ausgehend von GBR6 kann durch die Hinzunahme von Informationen für die Jahres- sowie Tageszeit als Prädiktoren (GBR9) die Qualität der Prognose mit RF nicht verbessern, während die zusätzliche Einbindung der Werte der Prädiktoren von GBR6 jeweils 1 h vor dem Zeitpunkt der Prognoseberechnung (GBR10) nur zu einer geringfügigen Reduzierung der mittleren Fehler führt. Da sich durch die Hinzunahme der „lags“ bei GBR10 jedoch die Anzahl der Prädiktoren verdoppelt und damit eine deutlich höhere Ressourcenauslastung beim Training der ML-Modelle sowie eine höhere Anfälligkeit im operationellen Betrieb folgt, wird die Optimal-Konfiguration mit GBR6 definiert. Der RMSE von GBR6 liegt mit RF bei 3,14 MW, mit XGBoost bei 3,16 MW und mit stepLM bei 3,51 MW. Der RMSE des Benchmark-Modells beträgt zum Vergleich 8,43 MW, und zeigt damit einen höheren Fehler bei der Prognose als die ML-Modelle im Untersuchungszeitraum Val1 bis Val5.

Aufgrund der Vielzahl an Plots (vier Prognosestufen, fünf Zeitfenster) wird im Rahmen dieses Artikels bei allen drei untersuchten Kraftwerken (GBR, BAS, BGS) nur jener der 4-h-Prognosestufe – welche vor allem bei den ML-Modellen meist den höchsten Unsicherheitsgrad aufweist – für das Zeitfenster Val5 – welches die höchste Variabilität der jeweiligen Kraftwerksleistung beinhaltet – dargestellt und diskutiert. Die restlichen Plots werden für die jeweils „optimalste“ Prädiktoren-Kombination (Bezeichnung in der letzte Reihe in Tab. 7, 8, 9) darüber hinaus als Online-Material zur Verfügung gestellt, worin auch die Verbesserung der Prognose mit zunehmend kürzerer Prognosezeit

bei den ML-Modellen anschaulich dargestellt wird.

Die Zeitreihen der 4-h-Prognose des Benchmark-Modells PW sowie der ML-Modelle (GBR6) werden der gemessenen Kraftwerksleistung im Zeitraum Val5 in Abb. 4 gegenübergestellt. Um den 27.04.2021 zeigen die Zeitreihen der ML-Modelle eine Lücke, da die Abflusszeitreihe des Pegels Laufen Salzach – welcher bei GBR6 als Prädiktor implementiert ist – in diesem Zeitraum keine Werte beinhaltet und in weiterer Folge auch keine Leistungsprognose berechnet werden kann. Insgesamt zeigt sich sowohl grafisch als auch bei Betrachtung der in Abb. 4 angeführten Gütemaße eine genauere Prognose der Kraftwerksleistung mit ML. Einerseits werden mit ML die an- und absteigenden Äste des Leistungsverlaufes exakter in puncto Zeitpunkt sowie Amplitude prognostiziert und andererseits sind auch die maximalen Residuen nur halb so groß. Der RMSE liegt in Val5 mit RF bei 1,78 MW, mit stepLM bei 2,06 MW sowie mit XGBoost bei 2,23 MW, während PW diesbezüglich einen Vergleichswert von 3,83 MW aufweist.

3.4 Laufkraftwerk Aschach (BAS)

Die erzielten Gütemaße RMSE sowie MAE über alle fünf Validierungsperioden Val1 bis Val5 (Tab. 5) können der Tab. 8 entnommen werden. Die Engpassleistung des betreffenden Kraftwerks BAS liegt zum Vergleich bei 324 MW (Tab. 1). Im besten Fall konnte mit dem ML-Modell RF und der Prädiktoren-Kombination BAS5 (Tab. 3) eine Reduktion des RMSE von 10,63 zu 7,35 MW im Vergleich zum Benchmark-Modell PW erzielt werden. Da

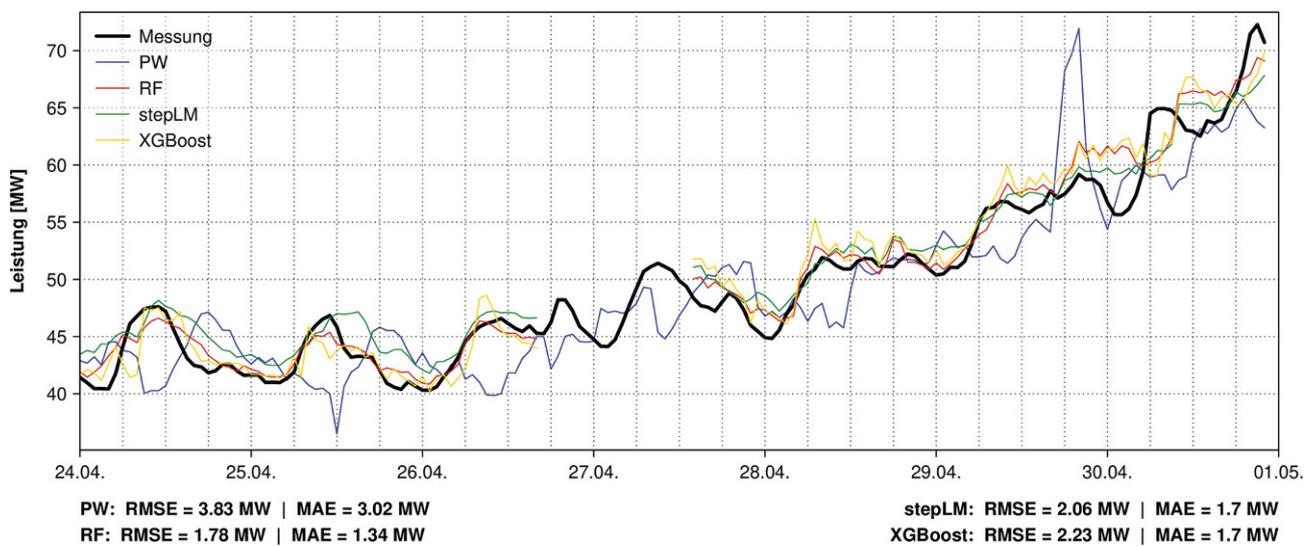


Abb. 4 Gegenüberstellung der 4-h-Prognosen für GBR6 im Zeitfenster Val5

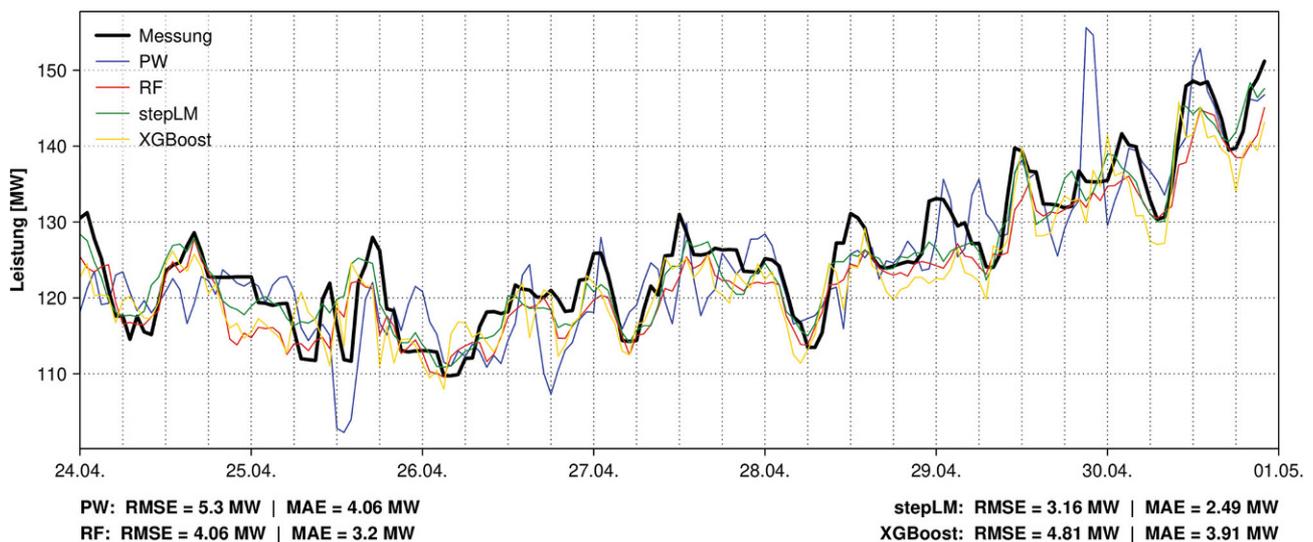


Abb. 5 Gegenüberstellung der 4-h-Prognosen für BAS3 im Zeitfenster Val5

bei BAS5 aber die Abflussdaten der Pegel Pfelling Donau sowie Plattling Isar als Prädiktoren implementiert wurden – welche derzeit noch nicht für den operativen Betrieb zur Verfügung stehen –, wird die Kombination BAS3 mit RF als optimale Konfiguration festgelegt. Bei BAS3 erzielten PW sowie XGBoost in etwa dieselbe Modellgüte, während stepLM mit einem RMSE von 62,88 MW deutlich davon abweicht. Begründet werden können die schlechten Ergebnisse mittels stepLM durch die hohen Residuen (>100 MW) in Val3, welche durch ein prognostiziertes aber nicht durchgeführtes Abschalten des Kraftwerks verursacht werden (siehe Online-Material). Wenn die

Ergebnisse von Val3 ausgeklammert werden, liegt der RMSE/MAE für stepLM zum Vergleich bei 5,56/3,94 MW und würde somit das beste Ergebnis über alle Modelle darstellen. Mit BAS6 (BAS3+zusätzlich Indikatoren für Jahres- sowie Tageszeit als Prädiktoren) sowie BAS7 (BAS3+zusätzlich Werte der Prädiktoren aus der Vergangenheit) lassen sich keine besseren kurzfristigen Leistungsprognosen erzielen.

Abb. 5 zeigt eine Gegenüberstellung der Zeitreihen der einzelnen Modelle in der Validierungsperiode Val5 der Prognosestufe 4h. In diesem Zeitfenster können mittels stepLM mit einem RMSE von 3,16 MW (unterer Teil der Grafik) die geringsten mittleren Residu-

en erzielt werden, wobei mit allen ML-Modellen der RMSE als auch der MAE im Vergleich zu PW verringert werden konnte. Die Modellgüte der Prognosen von RF und XGBoost liegen in Val5 zwischen stepLM und PW, wobei RF mit einem RMSE von 4,06 MW einen geringeren mittleren Fehler im Vergleich zu XBG mit 4,81 MW erzielt. Auffällig ist, dass die PW-Zeitreihe vereinzelt „Ausreißer“ nach oben und unten beinhaltet, während bei den Zeitreihen der ML-Modelle derart hohe Residuen (>10 MW) nicht inkludiert sind. Ein Grund dafür ist unter anderem die Hinzunahme der Leistung des betreffenden Kraftwerks BAS als Prädiktor, womit der Leistungsprognose mittels

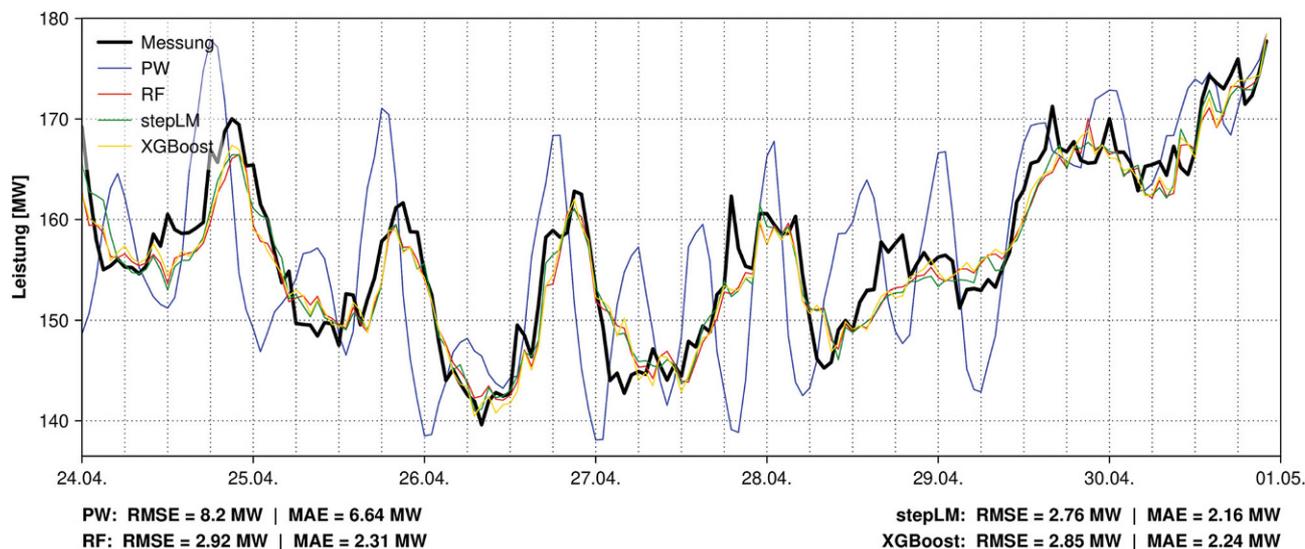


Abb. 6 Gegenüberstellung der 4-h-Prognosen für BGS3 im Zeitfenster Val5

ML ein solider Ausgangswert zur Verfügung steht und die Bandbreite der möglichen Leistungsentwicklung innerhalb der nächsten 4 h – und damit auch die potenzielle Abweichung zum Messwert – reduziert werden kann.

3.5 Laufkraftwerk Greifenstein (BGS)

Die Ergebnisse der ML-Modelle für die einzelnen Versuchsconfigurationen (Tab. 4) sowie des Benchmark-Modells PW sind in Tab. 9 aufgelistet. Die angegebenen Modellgüten sollten dabei zur besseren Einstufung in Relation zur Engpassleistung des Kraftwerks von 293 MW (Tab. 1) gesetzt werden. Mit der Configuration BGS3 und XGBoost konnte der RMSE im Vergleich zum Benchmark von 8,21 auf 3,94 MW sowie der MAE von 6,1 auf 2,94 MW im Zeitfenster der fünf Validierungsperioden (Tab. 5) reduziert werden. Dies entspricht sowohl beim RMSE als auch beim MAE ungefähr einer Halbierung der mittleren Residuen im besten Testfall. Ein nahezu identes Ergebnis wurde bei Anwendung des ML-Modells RF mit der Configuration BGS5 erzielt. Bei BGS5 wurde im Vergleich zu BGS3 der Abfluss des Pegels Mühlradung Enns zusätzlich als Prädiktor implementiert. Als optimale Configuration wurde jedoch BGS3 mit dem Modell RF gewählt, da damit einerseits nur ein marginal schlechteres Ergebnis in der Validierung wie mit dem Modell XGBoost erzielt wird und andererseits das Modell RF im Vergleich zu XGBoost aufgrund der deutlich geringeren Trainingszei-

ten (Tab. 6) zukünftig schneller adaptiert werden kann. Die Kombination BGS3 bietet gegenüber der Kombination BGS5 den Vorteil, dass im operativen Betrieb weniger Prädiktoren implementiert werden müssen, wodurch in weiterer Folge die Ausfallsicherheit erhöht wird. Bei Anwendung von RF mit der Prädiktoren-Kombination BGS6 wird ein im Vergleich zu BGS3 schlechteres Ergebnis erzielt, da der Pegel Stiefern Kamp erst Abflusswerte ab Mitte August 2016 aufweist und dadurch die Trainingsdaten um über 3,5 Jahre reduziert – also nahezu halbiert – wird. Bei der Hinzunahme von Prädiktoren bezüglich der Jahres- und Tageszeit (BGS7) werden zwar der RMSE sowie der MAE in der Trainingsperiode etwas verbessert – über die fünf Validierungsperioden können mit diesen Erweiterungen jedoch keine markanten Verbesserungen im Vergleich zu BGS3 erzielt werden. Ein ähnliches Bild zeigt sich, wenn ausgehend von der Configuration BGS3 zusätzlich jeweils die Werte von 60 min aus der Vergangenheit als Prädiktoren verwendet werden (BGS8). Mittels stepLM konnten mit fast allen Configurationen zuverlässigere kurzfristige Prognosen (4h) als mit dem Benchmark-Modell PW erzielt werden, jedoch fällt die Verbesserung geringer als mit den ML-Modellen RF und XGBoost aus. Die Verschlechterung des Ergebnisses von BGS2 auf BGS3 bzw. von BGS4 auf BGS5 bei stepLM kann – obwohl dabei jeweils zusätzliche Prädiktoren implementiert wurden – mit einer deutlichen Verschlechterung der

Prognose in Val3 (nach Scheitelwert Leistungsrückgang wegen Hochwasser) wahrscheinlich aufgrund einer Überanpassung (Overfitting) an die Trainingsdaten erklärt werden.

In Abb. 6 werden die Zeitreihen in der Validierungsperiode Val5 der Prognosestufe 4h gegenübergestellt – wobei grafisch veranschaulicht wird, dass die verschiedenen ML-Modelle einerseits untereinander ähnliche Ergebnisse liefern und andererseits deutlich bessere Prognosen erzielen als das Benchmark-Modell PW. Dies spiegelt sich auch in den angeführten Modellgüten RMSE sowie MAE (im unteren Teil des Plots) wider. Während die ML-Modelle im Zeitraum Val5 einen RMSE von ca. 3 MW aufweisen, liegt der Vergleichswert von PW bei über 8 MW. Ausschlaggebend dafür ist die exaktere Prognostizierung der Zeitpunkte und Amplituden der Leistungsschwankungen. Speziell in der ersten Hälfte von Val5 zeigen die Zeitreihen von PW eine Phasenverschiebung zu den gemessenen Peaks von ungefähr drei Stunden, während die Zeitreihen der ML-Modelle die Peaks im Hinblick auf Zeitpunkt und Höhe gut darstellen. Abb. 6 zeigt auch, dass PW teilweise auch Leistungsanstiege und -abstiege mit einer Amplitude im Größenbereich von 20 MW prognostiziert, welche später am Kraftwerk nicht gemessen wurden. Zwar werden mit den ML-Modellen auch nicht alle gemessenen Leistungsspitzen in puncto Höhe und Zeitpunkt perfekt abgebildet, allerdings sollte die Skalierung der Ordinate sowie die Re-

lation der Abweichung (max. ca. 7 MW) zur Engpassleistung von BGS beachtet werden.

4 Schlussfolgerung und Ausblick

Bei allen untersuchten Laufkraftwerken konnten die Modellgüten bei der kurzfristigen Leistungsprognose (bis zu 4 h) durch Anwendung von ML-Modellen verbessert werden. Dabei erzielte generell die ML-Modelle RF sowie XGBoost die zuverlässigeren Prognosen, wobei mittels XGBoost trotz der deutlich längeren Trainingszeit (Tab. 6) gegenüber RF bei der vorliegenden Untersuchung meist keine Verbesserung erzielt werden konnte. Überraschend genau sind in Anbetracht der Einfachheit des Konzepts und der Annahmen der multiplen linearen Regression auch die Prognosen des ML-Modells stepLM, wobei die erzielte Modellgüte in den meisten Fällen zwischen jener der komplexeren ML-Modellen RF bzw. XGBoost sowie des Benchmark-Modells PW liegt. Beim Kraftwerk Braunau-Simbach wurde durch RF mit der Prädiktoren-Kombination GBR6 in den fünf definierten Validierungszeiträumen (Tab. 5) eine Reduktion der RMSE-Werte um rund 63 % im Vergleich zum Benchmark-Modell PW erzielt. Für das Kraftwerk Aschach konnten mit RF und der Prädiktoren-Kombination BAS5 der RMSE um etwa 30 % verringert werden, während beim KW Greifenstein mittels ML der RMSE mehr als halbiert werden konnte. Die Zeitreihen der Modelle mit der optimalen Prädiktoren-Kombination gemäß Tab. 7, 8, 9 werden für alle fünf Validierungsperioden sowie für die vier Prognosestufen (4 bis 1 h) als Online-Material zur Verfügung gestellt. Dort ist klar ersichtlich, dass bei den ML-Modellen die Leistungsprognose mit einer kürzeren Prognosezeit auch zunehmend kleinere Residuen aufweist, während sich beim Benchmark-Modell PW die Metriken mit zunehmend geringerer Prognosestufe meist nicht oder nur marginal verbessern. Daraus lässt sich schließen, dass der Vorteil von ML-Modellen bei der kurzfristigen Leistungsprognose innerhalb einer Laufkraftwerkskette mit längerer Prognosezeit auch zunehmend geringer wird und ab einer bestimmten Stufe (z. B. 8 h) nicht mehr vorhanden ist. Dies lässt sich dadurch begründen, dass mit den ML-Modellen Zusammenhänge zwischen den Prädiktoren und der Zielgröße gelernt werden – welche sich

mit zunehmender Prognosezeit meist auch abschwächen –, während bei PW die Unsicherheit der Leistungsprognose zu einem wesentlichen Teil von den Unsicherheiten der meteorologischen Prognose sowie der konzeptuellen Modellierung abhängig ist.

Es hat sich herausgestellt, dass die zusätzliche Einbeziehung von „fuzzy-months“ (Indikatoren für die Jahreszeit) sowie „fuzzy-daytime“ (Indikator für die Tageszeit) als Prädiktoren – möglicherweise aufgrund einer Überanpassung in der Trainingsperiode – teilweise sogar zu einer Reduzierung der Modellgüte führte. Daher kann davon ausgegangen werden, dass eine über den vollständigen Trainings- und Validierungszeitraum gültige systematische jahres- oder tageszeitabhängige Verzerrung der Kraftwerksleistung zumindest in den Daten nicht vorhanden ist. Die Füllungen der Schleusenammern bei den Donau-Kraftwerken Aschach und Greifenstein verursachen aufgrund des reduzierten Turbinendurchflusses eine Reduktion der Leistung und sind meist zufällige Events, da nur nach Bedarf geschleust wird. Diese Zufälligkeit ist mittels ML-Verfahren grundsätzlich nicht erlernbar. Bezüglich der Einbeziehung von zusätzlichen „lags“ – also Werte eines Prädiktors aus der Vergangenheit – als Prädiktor ist grundsätzlich eine außerordentlich hohe Anzahl an Auswahlmöglichkeiten gegeben, da dafür im Prinzip alle vergangenen Zeitschritte gewählt werden können und die Auswahl diesbezüglich auch für jeden Prädiktor unterschiedlich ausfallen kann. Aufgrund der angewendeten „Trial und Error“-Methode bei der Auswahl der Prädiktoren wurde nur ein sehr kleiner Auszug der möglichen Optionen getestet, wobei die Qualität der Leistungsprognose für das KW Braunau sowie Greifenstein durch die Implementierung der „lags“ von 1 h für alle Prädiktoren geringfügig und für das KW Aschach bei Verwendung von RF nicht gesteigert werden konnte. Die Relevanz der einzelnen Prädiktoren für die ML-Modelle RF sowie XGBoost werden im Online-Material zusätzlich dargestellt. Dabei zeigt sich, dass für RF die Leistung des betreffenden Laufkraftwerks sowie die Leistung der nächst stromauf gelegenen Kraftwerke tendenziell am höchsten gewichtet werden, während für XGBoost im Hinblick auf die Relevanz der Prädiktoren keine Tendenz erkennbar ist.

Eine – im Vergleich zum hier angewendeten „Trial und Error“-Ansatz – systematischere (auf Algorithmen basierende) Prädiktorenauswahl (z. B. Chen et al. 2014) könnte wahrscheinlich zu einer weiteren Verbesserung der Leistungsprognose führen. Es ist dabei jedoch davon auszugehen, dass aufgrund der Vielzahl an Kombinationsmöglichkeiten sehr hohe Rechenkapazitäten erforderlich sind. Diesbezüglich würde sich hier ein sinnvoller Anknüpfungspunkt für weitere Untersuchungen bzw. Forschungsarbeiten ergeben.

Deep Learning (DL) ist ein spezieller Teilbereich von ML und kann durch den mehrschichtigen, neuronalen Netzaufbau („deep“) ein verbessertes Systemverständnis erlangen. LSTMs (Long Short-Term Memory), welche zur Gruppe der RNNs (Recurrent Neural Networks) gehören, können sich aufgrund der Netzwerkarchitektur mit den internen dedizierten Speicherbausteinen bestimmte Information über einen längeren Zeitraum merken. Feigl et al. (2021) und Kratzert et al. (2021) konnten beobachten, dass diese Fähigkeit speziell bei der Modellierung von Prozessen mit längeren Abhängigkeiten (z. B. hohe Konzentrationszeit in einem großen Einzugsgebiet, Niederschlagspeicherung in Form von Schnee) gegenüber anderen ML-Modellen vorteilhaft sein kann. Aufgrund der Beschränkung auf kurze Prognosezeiten (max. 4 h) sowie dem eher gering ausgeprägten Translations- sowie Retentionsverhalten zwischen den ausgewählten Prädiktoren wird jedoch nicht davon ausgegangen, dass LSTMs zu einer deutlich zuverlässigeren kurzfristigen Leistungsprognose beim vorliegenden Anwendungsfall führen. Letzten Endes bleibt diese Frage aber noch ungeklärt.

Da ML-Modelle ausschließlich aus den Daten im Training lernen können, ist die erzielte Modellgüte zu einem sehr wesentlichen Anteil von der Variabilität sowie der Qualität (z. B. Messfehler, Lücken) der verwendeten Daten abhängig. Das ML-Modell RF mit der jeweils optimalen Prädiktoren-Kombination wird derzeit im operationellen Betrieb ausführlich getestet. Bei diesen Modellen wurde das Ende der Trainingsperiode vom 30.04.2019 auf den 18.04.2021 ausgedehnt, um die Variabilität der Trainings-Daten zu erhöhen und in weiterer Folge die Prognose zu verbessern. Im operationellen Betrieb müssen die Messwerte aller Prädiktoren

einerseits vorhanden sein und andererseits auch laufend auf Plausibilität geprüft werden, da die Prognose exklusiv von diesen Werten abhängig ist. Bei außergewöhnlichen Ereignissen (z. B. Wartung einer Turbine eines Laufkraftwerks) an einer Anlage, deren Messwert als Prädiktor für die Leistungsprognose eines stromab liegenden Kraftwerks verwendet wird, können die Prognosen von den eigentlich erzielten Leistungen sehr stark abweichen.

Zusammenfassend zeigen unsere Ergebnisse, dass die kurzfristige Leistungsprognose mittels ML-Modellen deutlich verbessert werden kann und mit dem RF-Modell die geringsten Laufzeiten und die höchsten Modellgüten erzielt werden können. Grundsätzlich kann die kurzfristige Leistungsprognose mittels ML-Modell bei jeder beliebigen Laufkraftwerkskette durchgeführt werden. Wichtig dabei ist jedoch, dass stromauf genügend Messstellen zur Verfügung stehen und diese auch ausreichend lange Zeitreihen mit hochwertigen Messdaten aufweisen. Es hat sich herausgestellt, dass ein guter Startpunkt für die Prädiktorenauswahl einerseits die Leistung des jeweiligen Kraftwerks zum Zeitpunkt der Prognose sowie auch Leistungs- und Durchflusswerte der nächst stromauf liegenden Kraftwerke sind. Bei der vorliegenden Untersuchung hat sich zudem gezeigt, dass die Wellengeschwindigkeit meist deutlich höher als die durchschnittliche Fließgeschwindigkeit ist. Daher konnte durch die zusätzliche Hinzunahme von weit (teilweise über 100 km) stromauf liegenden Messstellen die mittleren

Residuen der Prognose weiter reduziert werden. Die Stationierung der Messstellen innerhalb des Gewässernetzes kann daher von entscheidender Bedeutung sein. Abschließend muss angemerkt werden, dass im vorliegenden Anwendungsfall innerhalb einer Laufkraftwerkskette der Vorteil der ML-Modelle gegenüber PW nur bei der kurzfristigen Leistungsprognose besteht, da mit zunehmender Prognosezeit auch die (nicht-)linearen Zusammenhänge zwischen den Werten der Prädiktoren sowie jenen der Zielgröße schwächer werden, wohingegen die Qualität der Leistungsprognose mit einem konzeptuellen Modell wie PW (COSERO in Kombination mit Flux^{DSS/DESIGNER/FLORIS²⁰⁰⁰}) stark von der meteorologischen Prognosequalität abhängig ist. Es ist allerdings wahrscheinlich, dass bei längerfristigen Prognosen mit PW durch eine Nachkopplung eines ML-Modells die mittleren Residuen der Prognose insgesamt reduziert werden können.

Danksagung Die Datenverarbeitung wurde mit den frei verfügbaren Softwarepaketen R (R Core Team 2020), Python (Python Software Foundation 2020) sowie QGIS (QGIS Development Team 2020) 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 mit

anderen teilen. Dank gebührt auch Klaus Stadlbacher, Ulrich Haberl sowie Klaus Hebenstreit von der VERBUND Energy4Business GmbH für die konstruktiven Diskussionen.

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>. ■

Literatur

- Blöschl, G., Bierkens, M.F.P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J.J., Savenije, H.H.G., Sivapalan, M., et al. (2019): Twenty-three unsolved problems in hydrology (UPH)—a community perspective. *Hydrological Sciences Journal*, 64(10), 1141–1158, <https://doi.org/10.1080/02626667.2019.1620507>
- BMLFUW (2007): Hydrologischer Atlas Österreichs (digHAO), 3. Lieferung, Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft, Wien
- Breiman, L. (1996): Bagging predictors. *Machine Learning*, 24, 123–140, <https://doi.org/10.1007/bf00058655>
- Breiman, L. (2001): Random forests. *Machine Learning*, 45(1), 5–32, <https://doi.org/10.1023/A:1010933404324>
- Broersen, P.M.T. (2002): Automatic spectral analysis with time series models. *IEEE Instr. Meas.*, 51, 211–216, <https://doi.org/10.1109/19.997814>
- Chen, T., Guestrin, C. (2016): XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13 August 2016, 785–794, <https://doi.org/10.1145/2939672.2939785>
- Chen, L., Ye, L., Singh, V., Zhou, J. (2014): Determination of Input for Artificial Neural Networks for Flood Forecasting Using the Copula Entropy Method. *J. of Hydrol. Eng.*, 19(11), 04014021, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000932](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000932)
- 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 am 17.08.2021)
- Claesen, M., De Moor, B. (2015): Hyperparameter Search in Machine Learning. *MIC 2015: The XI Metaheuristics International Conference*, 8 June 2015, <https://arxiv.org/abs/1502.02127v2>
- Colak, I., Sagiroglu, S., Yesilbudak, M. (2012): Data mining and wind power prediction: A literature review. *Renewable Energy*, 46, 241–247, <https://doi.org/10.1016/j.renene.2012.02.015>
- Dorn, F.B., Farahmand, H., Skjelbred, H.I., Belsnes, M.M. (2016): Modelling Minimum Pressure Height in Short-term Hydropower Production Planning. *Energy Procedia*, 87, 69–76, <https://doi.org/10.1016/j.egypro.2015.12.359>
- Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., et al. (2014): data.table: Extension of data.frame. <https://cran.r-project.org/web/packages/data.table/index.html> (Zugriff am 17.08.2021)
- Duan, Q., Gupta, V.K., Sorooshian, S. (1993): A shuffled complex evolution approach for effective and efficient optimization. *J. Optimization Theory Appl.*, 76(3), 501–521, <https://doi.org/10.1007/BF00939380>
- Eder, G., Fuchs, M., Nachtnebel, H.P., Loibl, W. (2005): Semidistributed modelling of the monthly water balance in an alpine catchment. *Hydrol. Process.*, 19, 2339–2360, <https://doi.org/10.1002/hyp.5888>

- 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 am 22.10.2020)
- EEX (2019): EEX Group Annual Report 2019. https://www.epexspot.com/sites/default/files/download_center_files/eex-group-annual-report-2019-english-data_01.pdf (Zugriff am 7.12.2021)
- Farr, T.G., Kobrick, M. (2000): Shuttle Radar Topography Mission produces a wealth of data. *Eos Trans. AGU*, 81, 583–583, <https://doi.org/10.1029/E0081i048p00583>
- Feigl, M., Herrnegger, M., Klotz, D., Schulz, K. (2020): Function Space Optimization: A symbolic regression method for estimating parameter transfer functions for hydrological models. *Water Res. Research*, 56(10), e2020WR027385, <https://doi.org/10.1029/2020WR027385>
- Feigl, M., Lebedzinski, K., Herrnegger, M., Schulz, K. (2021): Vorhersage der Fließgewässertemperaturen in österreichischen Einzugsgebieten mittels Machine Learning-Verfahren. *Österr Wasser- und Abfallw.*, 73, 308–328, <https://doi.org/10.1007/s00506-021-00771-3>
- Feranec, J. (2016): Project CORINE Land Cover. In: J. Feranec, T. Soukup, G. Hazeu, G. Jaffrain (Eds.), *European landscape dynamics, Corine land cover data*, CRC-Press, Boca Raton, 9–14, <https://doi.org/10.1201/9781315372860>
- Frey, S., Holzmann, H. (2015): A conceptual, distributed snow redistribution model. *Hydrol. Earth Syst. Sci.*, 19, 4517–4530, <https://doi.org/10.5194/hess-19-4517-2015>
- GKD (2021): Bayerisches Landesamt für Umwelt – Gewässerkundlicher Dienst. München, Deutschland, verfügbar unter: <https://www.gkd.bayern.de/de/fluesse/abfluss/tabellen> (Zugriff am: 15.09.2021)
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F. (2009): Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, 377(1–2), 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B., Gruber, C. (2011): The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. *Weather and Forecasting*, 26(2), 166–183. <https://doi.org/10.1175/2010WAF2222451.1>
- Hainermann, J., Kramer, O. (2016): Machine learning ensembles for wind power prediction. *Renewable Energy*, 89, 671–679, <https://doi.org/10.1016/j.renene.2015.11.073>
- Herrnegger, M., Nachtnebel, H.P., Haiden, T. (2012): Evapotranspiration in high alpine catchments—an important part of the water balance! *Hydrol. Res.*, 43, 460–475, <https://doi.org/10.2166/nh.2012.132>
- Herrnegger, M., Nachtnebel, H.P., Schulz, K. (2015): From runoff to rainfall: inverse rainfall-runoff modelling in a high temporal resolution. *Hydrol. Earth Syst. Sci.*, 19, 4619–4639, <https://doi.org/10.5194/hess-19-4619-2015>
- Herrnegger, M., Senoner, T., Nachtnebel, H.P. (2018): Adjustment of spatio-temporal precipitation patterns in a high Alpine environment. *J. Hydrol.*, 556, 913–921, <https://doi.org/10.1016/j.jhydrol.2016.04.068>
- Hrachowitz, M., Savenije, H.H.G., Blüschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., 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>
- Jin, X.B., Zheng, W.Z., Kong, J.L., Wang, X.Y., Bai, Y.T., Su, T.L., Lin, S. (2021): Deep-Learning Forecasting Method for Electric Power Load via Attention-Based Encoder-Decoder with Bayesian Optimization. *Energies*, 14, 1596, <https://doi.org/10.3390/en14061596>
- Kay, S.M., Marple, S.L. (1981): Spectrum analysis—A modern perspective. *Proceedings of the IEEE*, vol. 69(11), 1380–1419, Nov. 1981, <https://doi.org/10.1109/PROC.1981.12184>
- Klemeš, V. (1983): Conceptualization and scale in hydrology. *J. of Hydr.*, 65(1–3), 1–23, [https://doi.org/10.1016/0022-1694\(83\)90208-1\(83\)90208-1](https://doi.org/10.1016/0022-1694(83)90208-1(83)90208-1)
- Kling, H., Nachtnebel, H.P. (2009a): A method for the regional estimation of runoff separation parameters for hydrological modelling. *J. Hydrol.*, 364, 163–174, <https://doi.org/10.1016/j.jhydrol.2008.10.015>
- Kling, H., Nachtnebel, H.P. (2009b): A spatio-temporal comparison of water balance modelling in an Alpine catchment. *Hydrol. Process.*, 23, 997–1009, <https://doi.org/10.1002/hyp.7207>
- Kling, H., Fuchs, M., Paulin, M. (2012): Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, 424–425, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Kling, H., Stanzel, P., Fuchs, M., Nachtnebel, H.P. (2015): Performance of the COSERO precipitation – runoff model under non-stationary conditions in basins with different climates. *Hydrolog. Sci. J.*, 60, 1374–1393, <https://doi.org/10.1080/02626667.2014.959956>
- Klingler, C., Bernhardt, M., Wesemann, J., Schulz, K., Herrnegger, M. (2020): Lokale hydrologische Modellierung mit globalen, alternativen Datensätzen. *Hydrol. Wasserbewirts.*, 64, 166–187, https://doi.org/10.5675/HyWa_2020.4.1
- Klingler, C., Schulz, K., Herrnegger, M. (2021): LamaH-Ce: LARge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe. *Earth Syst. Sci. Data*, 13, 4529–4565, <https://doi.org/10.5194/essd-13-4529-2021>
- Klotz, D., Herrnegger, M., Schulz, K. (2017): Symbolic regression for the estimation of transfer functions of hydrological models. *Water Res. Research*, 53(11), 9402–9423, <https://doi.org/10.1002/2017WR021253>
- Kratzert, F., Gauch, M., Nearing, G., Hochreiter, S., Klotz, D. (2021): Niederschlags-Abfluss-Modellierung mit Long Short-Term Memory (LSTM). *Österr Wasser- und Abfallw.*, 73, 270–280, <https://doi.org/10.1007/s00506-021-00767-z>
- Kuhn, M. (2019): caret: Classification and Regression Training. <https://cran.r-project.org/web/packages/caret/index.html> (Zugriff am 17.08.2021)
- Kushner, H.J. (1964): A new method of locating the maximum point of an arbitrary multi-peaked curve in the presence of noise. *J. Basic Eng.*, 86(1), 97–106, <https://doi.org/10.1115/1.3653121>
- Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., Dadson, S.J. (2021): Benchmarking data-driven rainfall-runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrol. Earth Syst. Sci.*, 25, 5517–5534, <https://doi.org/10.5194/hess-25-5517-2021>
- Li, W., Kiaghadi, A., Dawson, C. (2021): High temporal resolution rainfall-runoff modeling using long-short-term-memory (LSTM) networks. *Neural Computing and Applications*, 33, 1261–1278, <https://doi.org/10.1007/s00521-020-05010-6>
- Meeus, L., Schittekatte, T., Reif, V. (2020): Who is responsible for balancing the system. The Evolution of Electricity Markets in Europe, 84–110, <https://doi.org/10.4337/9781789905472.00018>
- Mehdi, B., Dekens, J., Herrnegger, M. (2021): Climatic impacts on water resources in a tropical catchment in Uganda and adaptation measures proposed by resident stakeholders. *Climatic Change*, 164, 10, <https://doi.org/10.1007/s10584-021-02958-9>
- Mizukami, N., Clark, M.P., Newman, A.J., Wood, A.W., Gutmann, E.D., Nijssen, B., Rakovec, O., Samanigo, L. (2017): Towards seamless large-domain parameter estimation for hydrologic models. *Water Res. Research*, 53(9), 8020–8040, <https://doi.org/10.1002/2017WR020401>
- Močkus, J. (1975): On Bayesian Methods for Seeking the Extremum. *Optimization Techniques IFIP Technical Conference*, 1 to 7 July 1974, 400–404, https://doi.org/10.1007/978-3-662-38527-2_55
- Močkus, J. (1989): *Bayesian Approach to Global Optimization*. Springer Netherlands, ISBN 978-0-7923-0115-8, <https://doi.org/10.1007/978-94-009-0909-0>
- Močkus, J., Tiesis, V., Zilinskas, A. (1978): The application of Bayesian methods for seeking the extremum. *Towards Global Optimisation*, 2, 117–129
- Nachtnebel, H.P., Fuchs, M. (2004): Assessment of hydrological changes in Austria due to possible climate change. *Österreichische Wasser- und Abfallwirtschaft*, 56, 79–92
- Nachtnebel, H.P., Baumung, S., Lettl, W. (1993): *Abflussprognosemodell für das Einzugsgebiet der Enns und Steyr*. Report, Institut für Wasserwirtschaft, Hydrologie und konstruktiven Wasserbau, Universität für Bodenkultur Wien, Österreich
- Nearing, G.S., Kratzert, F., Sampson, A.K., Pellissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V. (2021): What role does hydrological science play in the age of machine learning? *Water Res. Research*, 57, e2020WR028091, <https://doi.org/10.1029/2020WR028091>
- Plate, E.J., Shahzad, K.M. (2015): Uncertainty Analysis of Multi-Model Flood Forecasts. *Water*, 7, 6788–6809, <https://doi.org/10.3390/w7126654>
- Python Software Foundation (2020): *Python Language Reference*. <https://www.python.org> (Zugriff am 13.12.2021)
- QGIS Development Team (2020): *QGIS Geographic Information System. Open Source Geospatial Foundation Project*, <https://www.qgis.org> (Zugriff am 13.12.2021)
- R Core Team (2020): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wien, Österreich, <https://www.r-project.org>
- Ramos, M.H., Mathevet, T., Thielen, J., Pappenberger, F. (2010), *Communicating uncertainty in hydro-meteorological forecasts: mission impossible?* *Met. Apps*, 17, 223–235, <https://doi.org/10.1002/met.202>
- Reichel, G., Fähr, R., Baumhackl, G. (2000): *FLORIS-2000: Ansätze zur 1.5D-Simulation des Sedimenttransportes im Rahmen der mathematischen Modellierung von Fließvorgängen*. In: Heigerth, G. (Ed.), *Symposium: Betrieb und Überwachung wasserbaulicher Anlagen*, Graz, 19.–20.10.2000, Inst. für Wasserbau und Wasserwirtschaft, Techn. Univ. Graz, 485–494
- Reichel, G. (2001): *FluxDSS und FLORIS2000 – ein leistungsfähiges Paket zur Modellierung der Fließvorgänge in komplexen Systemen*. *Österreichische Wasserwirtschaft*, 53(5–6), 170–171
- Schulz, K., Herrnegger, M., Wesemann, J., Klotz, D., Senoner, T. (2016): *Kalibrierung COSERO-Mur für ProVis*. Report, Institut für Wasserwirtschaft, Hydrologie und konstruktiven Wasserbau, Universität für Bodenkultur Wien, Österreich
- Stanzel, P., Nachtnebel, H.P. (2010): *Mögliche Auswirkungen des Klimawandels auf den Wasserhaushalt und die Wasserkraftnutzung in Österreich*. *Österreichische Wasser- und Abfallwirtschaft*, 62, 180–187, <https://doi.org/10.1007/s00506-010-0234-x>
- Stanzel, P., Kahl, B., Haberl, U., Herrnegger M., Nachtnebel, H.P. (2008): *Continuous hydro-*

logical modelling in the context of real time flood forecasting in alpine Danube tributary catchments. IOP Conference Series: Earth and Environmental Science, 4, 012005, <https://doi.org/10.1088/1755-1307/4/1/012005>

Wang, K., Qi, Y., Liu, H. (2019): A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Applied Energy, 251, 113315, <https://doi.org/10.1016/j.apenergy.2019.113315>

Wesemann, J., Herrnegger, M., Schulz, K. (2018): 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>

Zappa, M., Beven, K.J., Bruen, M., Cofiño, A.S., Kok, K., Martin, E., Nurmi, P., Orfila, B., Roulin, E., Schröter, K., Seed, A., Szturc, J., Vehviläinen, B., Germann, U., Rossa, A. (2010): Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. Atmosph. Sci. Lett., 11, 83–91, <https://doi.org/10.1002/asl.248>

Zhilinskas, A.G. (1975): Single-step Bayesian search method for an extremum of functions of

a single variable. Cybernetics, 11(1), 160–166, <https://doi.org/10.1007/BF01069961>

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