



Vorhersage der Fließgewässertemperaturen in österreichischen Einzugsgebieten mittels Machine Learning-Verfahren

Moritz Feigl · Katharina Lebieczinski · Mathew Herrnegger · Karsten Schulz

Angenommen: 15. April 2021 / Online publiziert: 20. Mai 2021
© Der/die Autor(en) 2021

Zusammenfassung Die Fließgewässertemperatur ist ein essenzieller Umweltfaktor, der das Potenzial hat, sowohl ökologische als auch sozio-ökonomische Rahmenbedingungen im Umfeld eines Gewässers zu verändern. Um Fließgewässertemperaturen als Grundlage für effektive Anpassungsstrategien für zukünftige Veränderungen (z. B. durch den Klimawandel) berechnen zu können, sind adäquate Modellierungskonzepte notwendig. Die vorliegende Studie untersucht hierfür 6 Machine Learning-Modelle: Schrittweise Lineare Regression, Random Forest, eXtreme Gradient Boosting, Feedforward Neural Networks und zwei Arten von Recurrent Neural Networks. Die Modelle wurden an 10 österreichischen Einzugsgebieten mit unterschiedlichen physiographischen Eigenschaften und Eingangsdatenkombinationen getestet. Die Hyperparameter der angewandten Modelle wurden mittels Bayes'scher Hyperparameteroptimierung optimiert. Um die Ergebnisse mit anderen Studien vergleichbar zu machen, wurden die Vorhersagen der 6 Machine Learning-Modelle den Ergebnissen der linearen Regression und dem häufig verwendeten und bekannten Wassertemperaturmodell air2stream gegenübergestellt.

Von den 6 getesteten Modellen zeigten die Feedforward Neural Networks und das eXtreme Gradient Boosting die besten Vorhersagen in jeweils 4 von 10 Einzugsgebieten. Mit einem durch-

schnittlichen RMSE (Wurzel der mittleren Fehlerquadratsumme; root mean squared error) von 0,55°C konnten die getesteten Modelle die Fließgewässertemperaturen deutlich besser prognostizieren als die lineare Regression (1,55°C) und air2stream (0,98°C). Generell zeigten die Ergebnisse der 6 Modelle eine sehr vergleichbare Leistung mit lediglich einer mittleren Abweichung um den Medianwert von 0,08°C zwischen den einzelnen Modellen. Im größten untersuchten Einzugsgebiet – Donau bei Kienstock – wiesen Recurrent Neural Networks die höchste Modellgüte auf, was darauf hinweist, dass sie sich am besten eignen, wenn im Einzugsgebiet Prozesse mit langfristigen Abhängigkeiten ausschlaggebend sind. Die Wahl der Hyperparameter beeinflusste die Vorhersagefähigkeit der Modelle stark, was die Bedeutung der Hyperparameteroptimierung besonders hervorhebt.

Die Ergebnisse dieser Studie fassen die Bedeutung unterschiedlicher Eingangsdaten, Modelle und Trainingscharakteristiken für die Modellierung von mittleren täglichen Fließgewässertemperaturen zusammen. Gleichzeitig dient diese Studie als Basis für die Entwicklung zukünftiger Modelle für eine regionale Fließgewässertemperaturvorhersage. Die getesteten Modelle stehen im open source R-Paket waterTemp allen AnwenderInnen der Forschungsgemeinschaft und der Praxis zur Verfügung.

Schlüsselwörter

Fließgewässertemperatur · Machine Learning · Wassertemperaturmodellierung · Österreich · Modellvergleich

Prediction of stream water temperatures in Austrian catchments using machine learning methods

Abstract Stream temperature is an essential environmental factor that has the potential to change both hydro-ecological and socio-economic conditions in the vicinity of a river. In order to calculate stream temperatures as a basis for effective adaptation strategies regarding future changes (e.g. due to climate change), robust modelling concepts are needed. This study investigates 6 machine learning models: Stepwise Linear Regression, Random Forest, eXtreme Gradient Boosting, Feedforward Neural Networks and two types of Recurrent Neural Networks. The models were tested on 10 Austrian catchments with different physiographic characteristics and input data combinations. The hyperparameters of the applied models were optimised using Bayesian hyperparameter optimisation. To compare the results with other studies, the predictions of the 6 machine learning models were compared with the results of linear regression and the well-know air2stream-model.

Both Feedforward Neural Networks and eXtreme Gradient Boosting showed the best prediction results in 4 out of 10 catchments. With an average root mean squared error of 0.55°C, the tested models were able to predict stream water temperatures significantly better than linear regression (1.55°C) and air2stream (0.98°C). In general, the results of the 6 models showed comparable performances with only a median deviation of 0.08°C between the individual models. In the largest catchment studied—the Danube upstream of Kienstock—Recurrent Neural Networks showed the highest model performance, indicating that they are best suited when processes with long-term dependencies are crucial in the

Die Autoren M. Feigl und K. Lebieczinski haben zu gleichen Teilen zum Manuskript beigetragen.

DI M. Feigl (✉) · DI K. Lebieczinski · DI Dr. M. Herrnegger · Univ.-Prof. Dipl.-Geoökol. Dr. K. Schulz
Institut für Hydrologie und Wasserwirtschaft (HyWa),
Universität für Bodenkultur Wien,
Muthgasse 18, 1190 Wien, Österreich
moritz.feigl@boku.ac.at

catchment. The choice of hyperparameters strongly influenced the predictive ability of the models, highlighting the importance of hyperparameter optimisation.

The results of this study summarise the importance of different input data, models and training characteristics for modelling stream water temperatures based on daily means and is the basis for future model developments regarding regional stream water temperature prediction. The tested models are available for the research community via the open source R package `wateRtemp`.

Keywords Stream water temperature · Machine Learning · Water temperature modelling · Austria · Model comparison

1 Einleitung

Die Wassertemperatur unserer Fließgewässer ist mehr als nur eine physikalische Eigenschaft. Sie ist ein essenzieller Umweltfaktor und sowohl für die Wasserqualität als auch für aquatische Habitate entscheidend. Sie hat Einfluss auf den Metabolismus (Álvarez und Nicieza 2005), die Verteilung (Boisneau et al. 2008), die Abundanz (Wenger et al. 2011), die Artenzusammensetzung (Dallas 2008) und den Wachstum (Imholt et al. 2010) von aquatischen Lebewesen, da diese nur eine bestimmte Schwankungsbreite der Wassertemperatur tolerieren können (Caissie 2006). Die Fließgewässertemperatur wirkt sich auch auf chemische Prozesse (Hannah et al. 2008) und physikalische Eigenschaften wie Dichte, Dampfdruck und Viskosität aus (Stevens et al. 1975), wodurch sie entscheidende Prozesse im Flussökosystem wie Primärproduktion, Dekomposition und den Nährstoffkreislauf indirekt beeinflusst (Friberg et al. 2009). Diese Eigenschaften und Prozesse beeinflussen die Menge an gelöstem Sauerstoff im Wasser (Sand-Jensen und Pedersen 2005) und haben dadurch auch einen großen Einfluss auf die Wasserqualität (Beaufort et al. 2016).

Die Fließgewässertemperatur ist auch von sozio-ökonomischem Interesse, wie z. B. für die Stromerzeugung und Industrie (Kühlung), Trinkwasserproduktion (Hygiene und bakterielle Verunreinigung) sowie Fischerei (Wachstum und Überleben der Fische, demographische Eigenschaften) (Hannah und Garner 2015). Änderungen der Fließgewässertemperatur können daher zu gravierenden ökologischen und

sozio-ökonomischen Auswirkungen auf aquatische Habitate sowie ihre umliegenden Regionen haben. Es ist daher von großer Bedeutung, Veränderungen dieses sensiblen Umweltfaktors vorherzusagen zu können und die treibenden Kräfte dahinter zu verstehen. Nur dadurch können potenziell schwerwiegende Auswirkungen durch präventive Maßnahmen vermieden werden.

Beobachtungen von Fließgewässertemperaturen stehen im Vergleich mit Niederschlag- und Abflussmessungen seltener zur Verfügung. Um die Entwicklung der Fließgewässertemperaturen vor allem in Anbetracht des Klimawandels abschätzen zu können, sind diese Messdaten jedoch essenziell. Auch der Betrieb thermischer Kraftwerke ist von Messungen und Prognosen von Fließgewässertemperaturen abhängig, wenn Flusswasser zur Kühlung verwendet wird. Einerseits ist die Effizienz der Kühlleistung von der Wassertemperatur des Gewässers abhängig, aus dem das Kühlwasser entnommen wird. Andererseits ist auch die Rückführung des erwärmten Kühlwassers relevant, da gesetzliche Richtlinien in Bezug auf die maximale erlaubte Fließgewässertemperatur nicht überschritten werden dürfen, was besonders bei Niederwasser während der heißen Sommermonate zu beachten ist. Um rechtzeitig handeln und effektive Anpassungsstrategien entwickeln zu können, ist ein Modellierungskonzept notwendig, das thermische Entwicklungen in einer entsprechenden Skala beschreiben kann. Die 210 Wassertemperaturmessstellen an Fließgewässern in Österreich verfügen zum Teil über mehr als 30 Jahre an gemessenen Tagesmittelwerten, die in Kombination mit Eigenschaften vergletschertes, hochalpiner Einzugsgebiete bis hin zu jener im Flach- und Hügelland gelegenen Einzugsgebiete eine ausgezeichnete Basis für die Entwicklung neuer Modellierungskonzepte darstellt.

Fließgewässertemperaturen können anhand prozessbasierter Modelle, statistischer/Machine Learning (ML)-Modelle oder einer Kombination dieser beiden Herangehensweisen modelliert werden. Prozessbasierte Modelle beschreiben physikalische Prozesse, welche die Wassertemperatur in Fließgewässern beeinflussen. Nach Dugdale et al. (2017) basieren diese Modelle darauf, dass zuerst die Energieflüsse vom und zum Fluss berechnet und anschließend die Temperaturveränderung

festgestellt wird. Um die Energieflüsse berechnen zu können, muss die Energiebilanzgleichung gelöst werden, was durch die Berücksichtigung der Wärmeströme der Luft-Wasser-Schnittstelle sowie der Flussbett-Wasser-Schnittstelle erreicht wird (Beaufort et al. 2016). Die Komponenten werden durch Feldmessungen oder anhand von Schätzungen bestimmt (Daniel Caissie und Luce 2017; Dugdale et al. 2017; Webb und Zhang 1997), wodurch diese Modelle äußerst komplex in der Anwendung sind. Obwohl es nicht praktikabel ist, all diese Komponenten langfristig an relevanten Punkten eines Flusses zu beobachten (Johnson et al. 2014), bieten prozessbasierte Modelle dennoch klare Vorteile: (i) Sie bieten eine Überblick über die treibenden Faktoren, (ii) informieren über Metriken, welche in übergeordneten statistischen Modellen verwendet werden können, (iii) informieren über Folgen unterschiedlicher Szenarien (Dugdale et al. 2017) und (iv) können anthropogene Einflüsse, z. B. thermische Einleiter, berücksichtigen. Aufgrund dieser Anwendungsmöglichkeiten werden prozessbasierte Modelle trotz der hohen Komplexität und der großen Menge an notwendigen Daten häufig für die Modellierung der Wassertemperatur an kürzeren Gewässerstrecken verwendet, jedoch nicht für überregionale Studien.

Statistische und ML-Modelle werden eingeteilt in parametrische Modelle, welche Regressionsmodelle (e.g. Mohseni und Stefan 1999) und stochastische Modelle (Ahmadi-Nedushan et al. 2007) inkludieren, sowie nicht-parametrische Modelle, welche keine Verteilungsannahmen treffen, wie beispielsweise neuronale Netzwerke oder k-nearest-neighbours (Benyahya et al. 2007). Im Gegensatz zu prozessbasierten Modellen geben statistische Modelle keinerlei Auskunft über Energietransfermechanismen innerhalb eines Flusses (Dugdale et al. 2017), sind dafür aber nicht auf eine große Menge an Inputdaten angewiesen, welche in der Praxis oft auch nicht verfügbar sind. Vor allem nicht-parametrische Modelle kamen in den letzten Jahren vermehrt zum Einsatz und hier insbesondere ML-Modelle (Zhu und Piotrowski 2020).

In der vorliegenden Studie haben wir 6 ML-Modelle untersucht: Schrittweise lineare Regression, Random Forest, eXtreme Gradient Boosting (XGBoost), Feedforward Neural Networks (FNN) und zwei Arten von Recurrent Neural

Networks (RNN). Schrittweise lineare Regressionsmodelle kombinieren eine iterative Variablenselektion mit linearer Regression, die nach unserem Wissensstand bisher erst in einer Studie von Neumann et al. (2003) für die Prognose täglicher maximaler Fließgewässertemperaturen angewandt wurde. Random Forest (RF) (Breiman 2001) ist ein Ensemble Learning-Modell, welches zur Vorhersage die Resultate mehrerer Entscheidungsbäume mittelt und erst kürzlich als ein vielversprechendes ML-Modell zur Seentemperaturmodellierung beschrieben wurde (Heddam et al. 2020). XGBoost (Chen und Guestrin 2016) ist ebenfalls ein auf einem Entscheidungsbaum basiertes Ensemble Learning-Modell. Allerdings erstellt XGBoost aufeinander aufbauende Entscheidungsbäume, anstatt Vorhersagen vieler Entscheidungsbäume zu mitteln, um ein Ensemble zu erstellen. XGBoost wurde bisher noch nicht für die Vorhersage von Flusswassertemperatur verwendet. Allerdings zeigen Resultate von kurzzeitigen Wasserqualitätsparameterprognosen, welche unter anderem auch Wassertemperaturen inkludieren, bereits vielversprechende Ergebnisse (Joslyn 2018; Lu und Ma 2020).

Feedforward Neural Networks (FNN) (White und Rosenblatt 1963) sind die ersten und einfachsten Arten neuronaler Netzwerke. FNNs wurden bereits in zahlreichen Studien zur Prognose von Fließgewässertemperatur verwendet (z. B. Abba et al. 2017; Bélanger et al. 2005; Chenard und Caissie 2008; DeWeber und Wagner 2014; Graf et al. 2019; Hadzima-Nyarko et al. 2014; McKenna et al. 2010; Piotrowski et al. 2015; Rabi et al. 2015; Risley et al. 2003; Sahoo et al. 2009; Temizyurek und Dadaser-Celik 2018; Wehrly et al. 2009; Westenbroek et al. 2010; Zhu et al. 2018, 2019a, b, c). Im Gegensatz zu FNNs sind Recurrent Neural Networks (RNN) Netzwerke, welche entwickelt wurden, um Sequenzen von Inputs zu verarbeiten, was durch interne (versteckte) Systemzustände erreicht wird. In der vorliegenden Studie wurden die beiden meistverbreiteten RNNs getestet, zum einen das Long short-term memory (LSTM)-Modell (Hochreiter und Schmidhuber 1997) und zum anderen das Gated recurrent unit (GRU)-Modell (Cho et al. 2014). Den Autoren ist erst eine Studie bekannt, in der ein LSTM in Kombination mit einer Hyperparameteroptimierung verwendet wurde, um die stündliche Wassertemperatur in

urbanen Flüssen vorherzusagen (Stajkowski et al. 2020). Generell wurden LSTMs in letzter Zeit in einer Vielfalt an hydrologischen Studien verwendet und zeigten vielversprechende Resultate für Aufgaben in der Prognose von Zeitreihen (z. B. Kratzert et al. 2018, 2019; Li et al. 2020; Xiang et al. 2020).

Um die Ergebnisse mit anderen Studien zu vergleichen, wurden die Vorhersagen der 6 ML-Modelle den Ergebnissen der linearen Regression und air2stream (Toffolon und Piccolroaz 2015) gegenübergestellt. Während frühere Studien lineare Regression größtenteils mit Lufttemperatur als einzigem Regressor zur Prognose von Flusswassertemperatur verwendet haben (z. B. Crisp und Howson 1982; Mackey und Berrie 1991; Smith 1981; Stefan und Preud'homme 1993), nutzen neuere Publikationen eine größere Bandbreite an Inputvariablen und/oder Modifikationen der klassischen linearen Regression (z. B. Arismendi et al. 2014; Caldwell et al. 2013; Jackson et al. 2018; Li et al. 2014; Naresh und Rehana 2017; Piotrowski und Napiorkowski 2019; Segura et al. 2015; Trinh et al. 2019). Air2stream ist ein Hybridmodell, welches eine physikalisch basierte Struktur mit einer stochastischen Parameterkalibrierung vereint. Es wurde bereits in mehreren Studien in einer Bandbreite an Einzugsgebieten verwendet und hatte im Allgemeinen im Vergleich zur linearen Regression und anderen ML-Modellen eine bessere Modellgüte (z. B. Piccolroaz et al. 2016; Piotrowski und Napiorkowski 2018, 2019; Tavares et al. 2020; Yang und Peterson 2017; Zhu 2019a).

Die meisten Studien verwenden vor allem Lufttemperatur und Abfluss als Eingangsdaten für Wassertemperaturmodelle (z. B. Naresh und Rehana 2017; Piccolroaz et al. 2016; Sohrabi et al. 2017), wobei einige auch Niederschlag (z. B. Caldwell et al. 2013) und/oder Sonneneinstrahlung (z. B. Sahoo et al. 2009) miteinbeziehen. Die Lufttemperatur kann auch als mittlerer, minimaler oder maximaler täglicher Temperaturwert inkludiert sein (z. B. Piotrowski et al. 2015) Um zu untersuchen, welche meteorologischen und hydrologischen Inputs für die Vorhersage von Wassertemperaturen notwendig sind, wurden in der vorliegenden Studie mehrere Kombinationen von Eingangsdaten untersucht. Besonders wichtig ist es zu wissen, wie gut Modelle mit wenigen Dateninputs im Vergleich zu komplexe-

ren Inputkombinationen abschneiden, um Anwendungsoptionen zu priorisieren.

ML-Modelle haben neben den Parametern, welche beim Trainieren mit den Daten optimiert werden, auch immer eine Reihe an Parametern, die im Vorhinein gewählt werden müssen – sogenannte Hyperparameter. Abhängig vom Modell und dem gegebenen Problem können Hyperparameter großen Einfluss auf die Güte des Modells haben (Claesen und De Moor 2015). Diese werden jedoch oft noch anhand von „Trial and Error“ gewählt (Hinton et al. 2012; Hsu et al. 2003) oder durch das Testen von Hyperparametern auf einem vordefinierten Raster von Werten bestimmt (Pedregosa et al. 2011). In dieser Studie wurde die Bayes'sche Optimierungsmethode (Kushner 1964; Močkus 1975, 1989; Močkus et al. 1978; Zhilinskas 1975) für die automatische Auswahl der Hyperparameter angewandt. Damit wurde die Wahrscheinlichkeit, unpassende Hyperparameter für die verwendeten Modelle anzuwenden, minimiert und die Bedeutung der Hyperparameter der unterschiedlichen Modelle untersucht.

Das Ziel dieser Studie ist die ausführliche Untersuchung von ML-Modellen, Inputdaten und Modelltrainingseigenschaften für die Prognose von mittleren täglichen Fließgewässertemperaturen. Dafür werden 6 ML-Modelle an 10 Einzugsgebieten mit unterschiedlichen Eigenschaften und Inputparameterkombinationen angewandt. Das inkludiert auch neue Herangehensweisen der Datenaufbereitung sowie die Hyperparameteroptimierung. Die Ergebnisse der ML-Modelle werden mit zwei weitverbreiteten Modellen (lineare Regression, air2stream) verglichen, um eine Referenz zu anderen Studien zu gewährleisten. Die getesteten ML-Modelle werden im open source R-Paket waterTemp allen AnwenderInnen der Forschungsgemeinschaft und der Praxis zur Verfügung gestellt. Die hier präsentierte Studie stellt eine Zusammenfassung der in Englisch erschienenen Publikation von Feigl et al. (2021) dar.

2 Daten und Methodik

2.1 Untersuchungsgebiete und Datengrundlage

Für die Modellierung der Fließgewässertemperaturen wurden 10 Einzugsgebiete in Österreich, Deutschland und

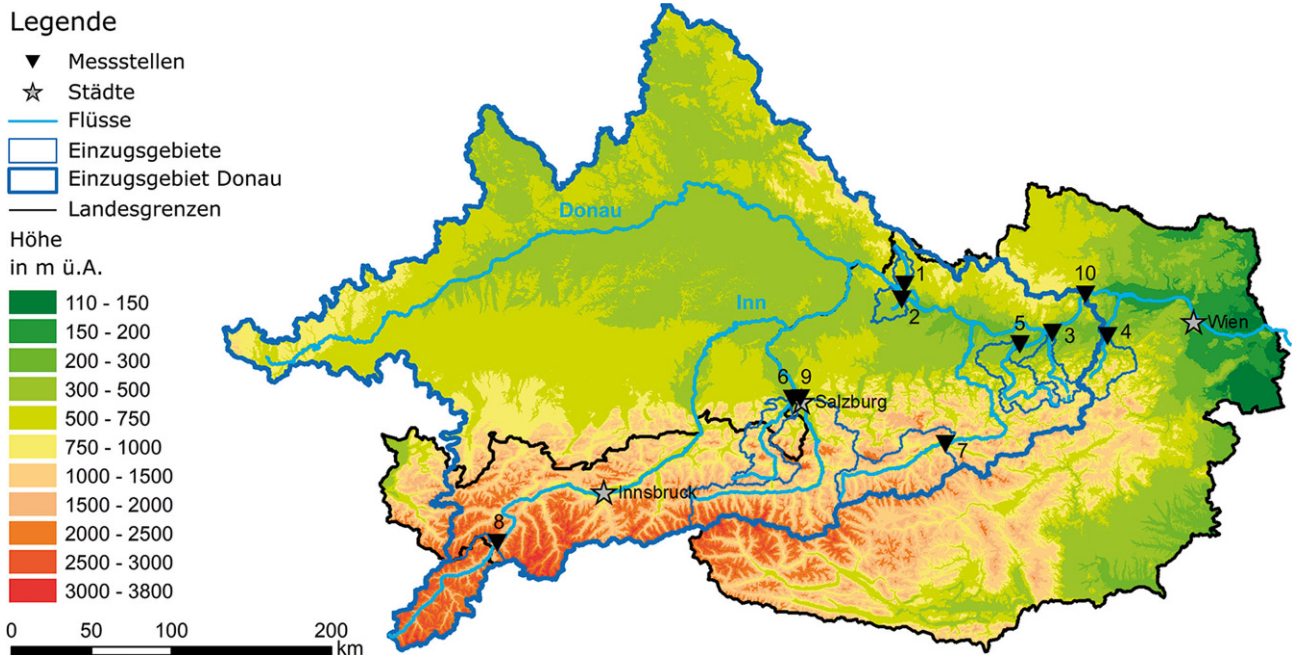


Abb. 1 Untersuchte Einzugsgebiete in Österreich, Deutschland und der Schweiz mit unterlegtem Digitalem Höhenmodell. Die IDs der Abflussmessstellen beziehen sich auf die IDs in Tab. 1

der Schweiz mit unterschiedlichen physiogeographischen Eigenschaften und Beobachtungslängen herangezogen. Alle Einzugsgebiete und Messstellen sind in Abb. 1 dargestellt und ihre Eigenschaften in Tab. 1 zusammengefasst.

Die Messstellen werden vom Hydrographischen Zentralbüro (HZB) betrieben und sind innerhalb der österreichischen Alpen bzw. dem umliegenden Flachland situiert. Gemessen werden Durchfluss (Q) und Wassertemperatur (T_w). Die Temperatursensoren sind derart installiert, dass die gemessene Wassertemperatur der vorherrschenden Wassertemperatur im jeweiligen Querschnitt entspricht.

Die berücksichtigten meteorologischen Daten sind das Tagesmittel der Lufttemperatur (T_a), das Tagesmaximum der Lufttemperatur (T_{max}), das Tagesminimum der Lufttemperatur (T_{min}), die Tagessumme des Niederschlags (P), sowie der 15-Minuten-Mittelwert der Globalstrahlung (GL). T_a , T_{max} , T_{min} und P sind räumlich verteilte Daten des SPARTACUS-Projekts (Hiebl und Frei 2016, 2018). Sie haben eine räumliche Auflösung von 1×1 km und eine zeitliche Auflösung von 1 Tag. SPARTACUS beruht auf der räumlichen Interpolation von Beobachtungen anhand von Kriging mit externer Drift und stellt tägliche meteorologische Rasterdaten für den Zeitraum 1961

bis 2015 zur Verfügung. GL-Daten waren im Analyse- und Nowcastingsystem INCA (Integrated Nowcasting through Comprehensive Analysis, Haiden et al. 2011, 2014) verfügbar und reichen bis ins Jahr 2007 zurück. INCA verwendet numerische Wettersimulationen in Kombination mit Beobachtungen und topographischen Informationen, um meteorologische Analyse- und Nowcastingfelder mit mehreren meteorologischen Parametern auf einem Raster von 1×1 km und für 15–60 min zu generieren. Für die vorliegende Studie wurden die 15-minütigen INCA-GL-Analysefelder auf einen täglichen Mittelwert aggregiert. Die Mittelwerte aller verwendeten Variablen sind in Tab. 1 für jedes Einzugsgebiet zusammengefasst. Durch das Verwenden hochauflösender, räumlich verteilter meteorologischer Daten als Basis unserer Modellinputs sollten die wichtigsten Einflussgrößen auf die Fließgewässertemperatur eines Einzugsgebiets bestmöglich dargestellt werden. Ähnliche meteorologische Datensätze sind auch für andere Teile der Welt verfügbar, beispielsweise ein globaler Datensatz (Hersbach et al. 2020), einer für Nordamerika (Thornton et al. 2014; Werner et al. 2019), einer für Europa (Brinckmann et al. 2016; Razafimaharo et al. 2020), sowie einer für China (He et al. 2020). Ein neuer Datensatz, der die gesamte Donau,

inkl. Zubringer, bis zur österreichischen Staatsgrenze zur Slowakei sowie alle anderen Einzugsgebiete in Österreich inkl. benachbarter Oberlieger, abdeckt, enthält nicht nur meteorologische Daten, sondern umfasst zusätzlich über 60 verschiedene Attribute in den Kategorien Topografie, Klimatologie, Hydrologie, Landbedeckung, Vegetation, Boden sowie Geologie (LamaH; Klingler et al. 2021a, b).

2.2 Datenaufbereitung

Die Datenaufbereitung umfasst die Aggregation der Rasterdaten, Feature-Engineering (z. B. Ableitung neuer Variablen/Features aus bestehenden Inputs) und der Aufteilung der Daten in Gruppen von Inputvariablen. Da die Wassertemperatur in Flüssen weitgehend durch Prozesse innerhalb des Einzugsgebiets gesteuert wird, werden Variablen mit einem integralen Effekt auf die Wassertemperatur über das Einzugsgebiet (T_a , T_{max} , T_{min} , P und GL) zu Gebietsmitteln aggregiert. Die Berechnung zusätzlicher Features aus einem gegebenen Datensatz und die damit einhergehende zusätzliche Datenrepräsentation kann die Leistung von ML-Modellen deutlich verbessern (Bengio et al. 2013).

Frühere Studien haben gezeigt, dass insbesondere Zeitinformation für die

Tab. 1 Übersicht der Einzugsgebietscharakteristika: Gewässer, Messstelle, Untersuchungszeitraum, Jahre an verfügbaren Daten, Einzugsgebietsfläche, mittlere Fließgewässertemperatur, mittlerer Abfluss und Gebietsmittel meteorologischer Variablen. Die IDs beziehen sich auf die IDs in Abb. 1

ID	Messstelle		Untersuchungs-		Fläche (km ²)	T _w (°C)	Q (m ³ /s)	T _a (°C)	P (mm/d)	GL (W/m ²)
			Jahre							
1	Kleine Mühl	Obermühl	2002–2015	14,0	200,2	8,87	3,12	8,71	2,73	135
2	Aschach	Kropfmühle	2004–2015	11,9	312,2	10,78	3,80	9,57	2,50	136
3	Erlauf	Niederndorf	1980–2015	35,3	604,9	9,42	15,27	7,99	3,59	127
4	Traisen	Windpassing	1998–2015	17,7	733,3	9,83	14,88	8,47	3,33	131
5	Ybbs	Greimpersdorf	1981–2015	34,7	1116,6	9,87	31,50	7,97	3,77	127
6	Saalach	Siezenheim	2000–2015	16,0	1139,1	8,50	39,04	6,72	4,60	135
7	Enns	Liezen	2006–2015	10,0	2116,2	1,19	67,56	5,62	3,60	137
8	Inn	Kajetansbrücke	1997–2015	18,8	2162,0	6,00	59,26	0,12	2,56	153
9	Salzach	Salzburg	1977–2015	39,0	4425,7	7,63	178,26	5,22	4,16	136
10	Donau	Kienstock	2005–2015	11,0	95970,0	10,77	1798,31	10,05	2,13	131

Tab. 2 Übersicht der 6 Experimente und ihrer Kombination an Eingangsdaten. Alle Experimente berücksichtigen zusätzlich 4 Lags und Fuzzy-Monate

Experiment	T _a	T _{max}	T _{min}	P	Q	GL
0 (T _a)	X					
1 (T)	X	X	X			
2 (TP)	X	X	X	X		
3 (TQ)	X	X	X		X	
4 (TQP)	X	X	X	X	X	
5 (TPGL)	X	X	X	X		X
6 (TQPGL)	X	X	X	X	X	X

Vorhersage von Wassertemperaturen wichtig ist. Dies umfasst die Zeit, ausgedrückt als Tag des Jahres (z. B. Hadzima-Nyarko et al. 2014; Jackson et al. 2018; Li et al. 2014; Zhu et al. 2018; Zhu et al. 2019a, b), den Inhalt des Gregorianischen Kalenders (d. h. Jahr, Monat, Tag) (Zhu et al. 2019c) oder als Deklination der Sonne (Piotrowski et al. 2015), welche eine Funktion des Tages im Jahr ist. Die Verwendung von zyklischen Merkmalen wie dem Tag des Jahres als ganzzahlige Variable wird dennoch höchstwahrscheinlich die Güte des Modells verringern, da die Tage 1 und 365 so nahe beieinander liegen wie 1 und 2 und sich deswegen nicht wie numerische Variablen verhalten.

Um Zeitinformationen in ein geeigneteres Format zu übertragen, werden in dieser Studie Monate und die Tage des Monats in dreieckige Fuzzy-Mengen transformiert, welche im weiteren als Fuzzy-Monate bezeichnet werden. Sie haben am 15. des jeweiligen Monats den Wert 1 und nehmen jeden Tag linear ab, bis sie 30 Tage später bzw. 30 Tage früher gleich 0 sind. Daher werden die Werte zweier benachbarter

Monate zum Monatswechsel etwa bei 0,5 liegen. Ähnlich wie beim Ansatz nach Shank et al. (2008) wird durch das Transformieren der kategorischen Variable „Monat“ in diese 12 neuen Fuzzy-Variablen der Einfluss der Jahreszeiten gleichmäßiger dargestellt, da Sprünge im monatlichen Verlauf vermieden werden.

Eine Studie von Webb et al. (2003) zeigte, dass meteorologische und hydrologische Informationen vergangener Tage ebenfalls signifikant mit der Wassertemperatur in Verbindung stehen und die Güte des Modells verbessern können. Damit die Informationen der vorangegangenen Tage von den Modellen verwendet werden können, werden die Lags (also die Werte vorhergehender Zeitschritte) aller Variablen für die 4 vorherigen Tage berechnet und als zusätzliche Inputs verwendet.

In der vorliegenden Studie wurden für jedes der 6 ML-Modelle 6 Experimente mit unterschiedlichen Eingangsdatenkombinationen durchgeführt. Alle Experimente berücksichtigen 4 Lags und Fuzzy-Monate als Inputs. Die Zusammensetzung der Variablen aller Experimente ist in Tab. 2 zusammengefasst.

Experiment 0 (T_a) dient als einfaches Vergleichsmodell, in welchem nur die mittlere Lufttemperatur, Lags und Fuzzy-Monate als Input verwendet werden. Experiment 1 (T) zeigt den Nutzen der Einbeziehung von T_{max} und T_{min}. Die Experimente 2 bis 4 bestehen aus Kombinationen von Experiment 1 mit Niederschlags- und Abflussdaten. Die Experimente 5 und 6 sind Kombinationen mit GL und beinhalten daher nur Daten des Zeitraums von 2007 bis 2015, in welchem GL-Daten verfügbar waren.

2.3 Vergleichsmodelle

Als Referenz und um die Ergebnisse mit anderen Studien zu vergleichen, wurden die Vorhersagen der 6 ML-Modelle mit den Ergebnissen der linearen Regression und air2stream (Toffolon und Piccolroaz 2015) verglichen. Durch die Miteinbeziehung dieser beiden Modelle ist es möglich, die Ergebnisse dieser Studie in einen größeren Kontext zu stellen.

2.3.1 Linear Regression

Die klassische multiple lineare Regression ist definiert als

$$Y = \beta X + \epsilon,$$

wobei Y den Vektor der abhängigen Variable (Wassertemperatur) bezeichnet, X die Matrix unabhängiger Variablen (z. B. Tagesmittelwert der Lufttemperatur, Globalstrahlung), β bezeichnet den Vektor der Modellkoeffizienten und ϵ den Fehlerterm. ϵ wird als normalverteilt mit einer diagonalen Kovarianzmatrix angenommen. Durch das Minimieren der Summe der quadratischen Fehler können die Schätzer für Modellkoeffizienten und der abhängigen Variable folgendermaßen definiert werden:

$$\hat{Y} = \hat{\beta} X,$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y,$$

wobei \hat{Y} und $\hat{\beta}$ die jeweiligen Schätzer darstellen. Das in dieser Studie verwendete Modell enthält eine Konstante (Intercept) und die Variablen T_a und Q als unabhängige Variablen zur Vorhersage von T_w.

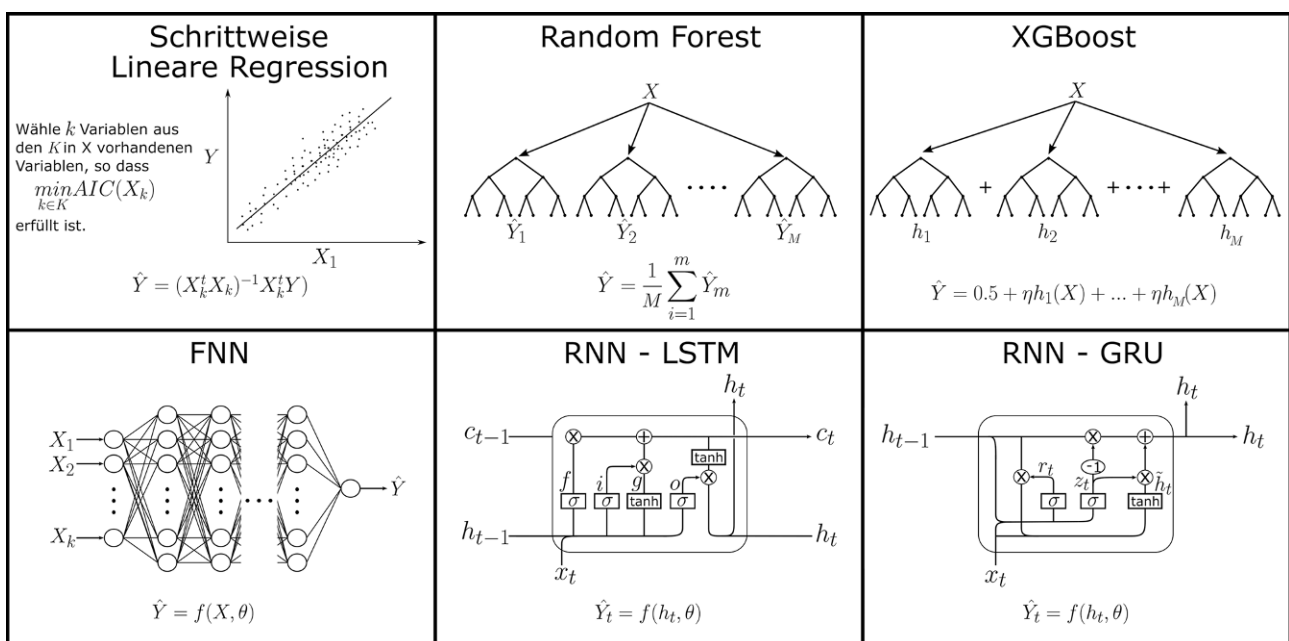


Abb. 2 Übersicht der untersuchten ML-Modelle. \hat{Y} steht für die geschätzten Wassertemperaturen, X für die Matrix der beobachteten Variablen, $\hat{Y}_1, \dots, \hat{Y}_M$ für die Vorhersagen einzelner RF Bäume, h_1, \dots, h_M für die vorhergesagten Residuen einzelner XGBoost-Entscheidungsbäume, $f(X, \theta)$ für die Abbildung eines FNNs mit den Parametern θ , h_t für den Output-Zustand (hidden states) der RNN Zellen, c_t für den Zellstatus der LSTM-Zellen und $f(h_t, \theta)$ für die Abbildung eines RNN mit Parametern θ . RNNs bestehen jeweils aus Kaskaden der hier gezeigten Zellen

2.3.2 air2stream

Das Hybridmodell air2stream (Toffolon und Piccolroaz 2015) vereint eine physikalisch basierte Struktur mit einer stochastischen Parameterkalibrierung. Air2stream verwendet T_a und Q als Input für Differentialgleichungen, welche die Wärmebilanz des Systems darstellen und aus vereinfachten physikalischen Beziehungen abgeleitet wurden. Aufgrund dieser Vereinfachung kann es wie ein datengetriebenes Modell angewendet werden, welches von einer Parameterkalibrierung abhängig ist. Die 8-Parameter-Version von air2stream ist definiert als

$$\frac{dT_w}{dt} = \frac{1}{\theta a_4} [a_1 + a_2 T_a - a_3 T_w + \theta \left(a_5 + a_6 \cos \left(2\pi \left(\frac{t}{t_y} - a_7 \right) \right) - a_8 T_w \right)],$$

wobei t die Zeit in Tagen ist, t_y die Anzahl der Tage pro Jahr, \bar{Q} der mittlere Abfluss, $\theta = \bar{Q}/\bar{Q}$ der dimensionslose Abfluss und a_1, \dots, a_8 die Modellparameter sind. Diese Differentialgleichung wird mithilfe des Crank-Nicolson-Verfahrens (Crank und Nicolson 1947) bei jedem Zeitschritt numerisch integriert, und die Modellparameter werden unter Anwendung der Particle Swarm Optimi-

zation (Kennedy und Eberhart 1995) kalibriert.

2.4 Machine Learning-Modelle

In dieser Studie vergleichen wir 6 ML-Modelle: schrittweise lineare Regression (step-LM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Feedforward Neural Networks (FNN) und zwei Recurrent Neural Networks (RNN) – das Long short-term Memory (RNN-LSTM)-Modell sowie das Gated Recurrent Unit (RNN-GRU)-Modell. Eine Übersicht und schematische Darstellung der Modelle wird in Abb. 2 gezeigt.

2.4.1 Schrittweise lineare Regression

Step-LM kombiniert ein iteratives Variablenauswahlverfahren mit multipler linearer Regression. Die schrittweise Variablenauswahl beginnt mit einem Ausgangsmodell und verändert die Auswahl der Variablen in jedem Iterationsschritt mithilfe eines vorher gewählten Kriteriums. Dabei können Variablen hinzugefügt oder entfernt werden. Die hier angewandte schrittweise Variablenauswahl verwendet als Anfangsmodell das Modell mit allen Variablen und dem Akaike-Informationskriterium

(Akaike 1973). Das AIC für ein lineares Modell ist gegeben durch

$$AIC = n \times \ln \left(\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n} \right) + 2k,$$

wobei n die Anzahl der Stichproben ist, $\ln()$ der natürliche Logarithmus, Y und \hat{Y} die beobachteten und vorhergesagten Wassertemperaturen und k die Anzahl der ausgewählten Variablen. Die schrittweise Variablenauswahl wird iterativ angewandt, bis das AIC ein Minimum erreicht hat. Zusätzlich zu den vorhandenen hydrologischen und meteorologischen Variablen werden auch alle Interaktionsterme miteinbezogen.

2.4.2 Random Forest

Das RF-Modell (Breiman 2001) ist ein Ensemble Learning-Modell, welches auf der Idee des Bagging (Bootstrap Aggregation) (Breiman 1996) basiert. Bagging-Prädiktoren erstellen einen Mittelwert aus mehreren Modellvorhersagen, wobei jedes Modell auf einer Bootstrap-Stichprobe anstelle der vollständigen beobachteten Probe trainiert wird. Diese durch Bootstrapping eingeführte Zufälligkeit erhöht die Fähigkeit der Modelle zur Generalisierung und zur Erzeugung stabiler Vorhersageer-

gebnisse. RF-Modelle sind Bagging-Prädiktoren, welche Entscheidungsbäume (Klassifizierungs- und Regressionsbäume, CARTs) als Basis-Modell verwenden. RF-Bäume wenden rekursive Binärsplits auf die Daten an, um die Entropie in den Baumknoten zu minimieren. Dies wird so lange durchgeführt, bis jeder Knoten eine minimale Knotengröße oder eine vorher definierte maximale Anzahl an splits erreicht. Breiman (2001) zeigte, dass das Hinzufügen weiterer Zufälligkeiten zur Bagging-Methode die Vorhersagegenauigkeit verbessert. In Random Forest wird dies erreicht, indem nur eine zufällige Teilmenge an verfügbaren Variablen für die Aufteilung an jedem Knoten gewählt wird. Die Schätzung für die unabhängige Variable ist gegeben durch

$$\hat{Y} = \frac{1}{M} \sum_{m=1}^M f_m(X),$$

wobei f_m einen einzelnen angepassten Entscheidungsbaum bezeichnet, M die Anzahl der verwendeten Bäume, X die Matrix der Regressoren und \hat{Y} den Vektor der geschätzten Wassertemperaturen. Eine vereinfachte Darstellung des RF-Algorithmus ist in Abb. 2 zu sehen. RF verwendet 2 wichtige Hyperparameter: Die Anzahl der Prädiktoren, welche an jedem Knoten geprüft werden (mtry), und die Mindestgröße der Knoten (min node size). Für alle Modelle werden 500 Entscheidungsbäume verwendet.

2.4.3 XGBoost

XGBoost (Chen und Guestrin 2016) ist ein Tree-Boosting-Algorithmus, welcher auf der Grundlage des Konzepts des Boostings entwickelt wurde und Verbesserungen bezüglich der Effizienz und Skalierbarkeit sowie eine Reduktion des Overfittings beinhaltet. Ähnlich wie beim Bagging kombinieren Boosting-Methoden die Vorhersagen eines Ensembles von simpleren Modellen, um die Vorhersagegenauigkeit zu verbessern. Während jedoch beim Bagging Ensemble-Mitglieder parallel trainiert werden, trainiert Boosting iterativ neue Ensemble-Mitglieder und fügt sie dem bestehenden Ensemble hinzu.

Boosting wurde erstmals von Schapire (1990) eingeführt und ist seit der Einführung des Adaboost-Algorithmus (Freund und Schapire 1995) sehr bekannt geworden. Friedman (2001)

entwickelte Boosting durch die Entwicklung von Gradientenverfahren für Boosting-Iterationen weiter. Dies führte zur Entwicklung des Gradient-Tree-Boostings (Friedman 2002), bei dem Entscheidungsbäume als Basis-Modell verwendet werden. XGBoost ist eine Implementierung von Gradient-Tree-Boosting mit weiteren Verbesserungen in der Form von zusätzlicher Stochastizität und Regularisierung. Der XGBoost-Schätzwert in Abb. 2 für die unabhängige Variable ist gegeben durch

$$\hat{Y} = 0,5 + \sum_{m=1}^M \eta f_m(X),$$

wobei $f_{1,\dots,M}$ eine Sequenz von Entscheidungsbäumen ist, $\eta \in [0,1]$ die Lernrate, M die Anzahl der verwendeten Bäume, X die Matrix der Inputvariablen und \hat{Y} der Vektor der geschätzten Wassertemperaturen. Bei jedem Boosting-Schritt wird ein zusätzlicher Entscheidungsbaum darauf trainiert, die Fehler des vorhandenen Ensembles vorherzusagen und somit zu verbessern. Die Verlustfunktion, welche zum Trainieren jedes Baums verwendet wird, enthält einen Regularisierungsterm, um Overfitting zu verhindern. Zusätzlich wird Overfitting reduziert, indem pro Iterationsschritt jeweils nur eine zufällig bestimmte Teilmenge von Beobachtungen und Variablen verwendet wird, um den Entscheidungsbaum zu erstellen. Eine vereinfachte Darstellung des XGBoost-Algorithmus ist in Abb. 2 ersichtlich.

XGBoost verfügt über mehrere wichtige Hyperparameter, welche vor der Anpassung des Modells gewählt werden müssen: die maximale Anzahl der Iterationen (nrounds), die Lernrate (η), die maximale Tiefe eines Baumes (max depth), der Grenzwert der Summe der „instance weights“, bei dem die weitere Partitionierung gestoppt wird (min node size), das Verhältnis der zufälligen Teilstichproben, welche für das Erstellen eines Baums verwendet werden (subsampling) sowie den zufälligen Anteil an Variablen, welche für das Erstellen eines Baums verwendet werden (colsample bytree).

2.4.4 Feedforward Neural Network

Ein FNN (White und Rosenblatt 1963) ist die erste und einfachste Art neuronaler Netzwerke. FNNs bestehen aus mehreren Schichten (Layers) von Knoten (Neuronen), in welchen jeder Knoten mit allen Knoten der vorher-

gehenden und nachfolgenden Schicht verbunden ist. In einem Knoten werden eine lineare und eine nicht-lineare (Aktivierungs-)Funktion auf die Inputs angewendet, um einen Output zu erzeugen. Die allgemeine Struktur eines FNN ist in Abb. 2 dargestellt.

Piotrowski et al. (2020) zeigten, dass Dropout-Layer (Baldi und Sadowski 2014; Hinton 2012; Srivastava et al. 2014) die Vorhersage der Fließgewässertemperatur in einschichtigen FNNs verbessern. Dropout bezieht sich auf das zufällige Eliminieren von Knoten aus einer Schicht während des Trainings, was Overfitting verhindern und die Generalisierung potenziell verbessern kann. In dieser Studie wird Dropout zu jeder FNN-Schicht hinzugefügt und die Dropout-Rate als Hyperparameter definiert. Die Dropout-Rate kann dabei auch 0 werden, was keinem Dropout entspricht.

Während die Parameter (θ) der linearen Funktion mittels Backpropagation optimiert werden (Rumelhart et al. 1986), verfügen FNNs über mehrere Hyperparameter, welche vor dem Training definiert werden müssen. Zu diesen Hyperparametern gehört die Aktivierungsfunktion, die Anzahl der Ebenen, die Anzahl der Knoten pro Ebene sowie die Dropout-Rate. Nach initialen Tests, bei denen verschiedene Aktivierungsfunktionen angewandt wurden, wurde die Scaled Exponential Linear Unit (SELU) (Klambauer et al. 2017) als Aktivierungsfunktion für alle Knoten im Netzwerk ausgewählt. SELU beinhaltet eine Standardisierung, welche die Konvergenz verbessert und sowohl verschwindende als auch stark anwachsende Gradienten während der Backpropagation vermeidet. Die anderen Hyperparameter werden optimiert. Der hier vorgestellte Ansatz zur Hyperparameteroptimierung unterscheidet sich von bisherigen Studien, welche im Allgemeinen von einer festen Anzahl an Schichten und/oder Knoten pro Schicht ausgehen, die durch einen „Trial-and-Error-Ansatz“ abgeleitet wurden (z.B. Bélanger et al. 2005; Hadzima-Nyarko et al. 2014; Piotrowski et al. 2015; Zhu et al. 2018, 2019a).

2.4.5 Recurrent Neuronal Networks

Im Gegensatz zu FNNs sind RNNs durch interne Zellzustände (hidden states) in der Lage, Sequenzen von Inputs zu verarbeiten. Obwohl es viele verschiedene Typen von RNNs gibt, ha-

ben wir uns auf die beiden am weitesten verbreiteten konzentriert: das LSTM-Modell (Hochreiter und Schmidhuber 1997) sowie das GRU-Modell (Cho et al. 2014). Jede Schicht eines RNN besteht aus einer Sequenz von Zellen, welche dieselben Parameter verwenden. Die Zellen des LSTM und der GRUs sind in Abb. 2 dargestellt und werden in den nächsten Abschnitten beschrieben.

Eine einzelne RNN-Zelle besteht aus mehreren Gates, in denen nicht-lineare Transformationen auf die Eingangsdaten und internen Zellzustände angewendet werden. Jedes RNN enthält eine FNN-Schicht mit einem einzelnen Knoten am Ende, welche zur Berechnung der prognostizierten Werte aus den Zellzuständen des letzten Zeitschritts (h_T) verwendet wird. Beide Arten von RNNs verwenden die gleichen Hyperparameter, welche vor dem Training des Modells festgelegt werden müssen: die Anzahl der verwendeten RNN-Schichten, die Anzahl der Zellen pro Schicht, die Anzahl der Zeitschritte, die Dropout-Rate und die batch size.

Aufgrund ihrer internen Zellzustände und der Verwendung mehrerer Zeitschritte für die Vorhersage kann davon ausgegangen werden, dass RNNs keine Zeitinformationen (hier in Form von Fuzzy-Months) für die Vorhersage von Wassertemperaturdaten benötigen. Um diese Annahme zu testen, wurden beide RNN-Varianten vorab auch ohne Fuzzy-Months trainiert, um den Einfluss dieser zusätzlichen Variablen auf die Modellgüte zu überprüfen. Da sie in der Lage waren, ebenso gute Ergebnisse ohne Fuzzy-Months zu erzielen, wurde die Trainingszeit durch die Verringerung der Inputdaten um 12 Dimensionen (Spalten) bedeutend reduziert.

Long short-term memory-Zellen Ist eine Sequenz von Inputs für T Zeitschritte x_1, \dots, x_T gegeben, wobei jedes $x_t \in R^d$ ein Vektor mit d Variablen ist, ist der Vorwärtsdurchlauf einer einzelnen LSTM-Zelle mit h versteckten Knoten durch die folgenden Gleichungen gegeben:

$$\begin{aligned} f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned}$$

wobei f_t , i_t und $o_t \in R^h$ das Forget Gate, das Input Gate und das Output Gate sind, $\tilde{c}_t \in R^h$ die Zelleninputaktivierung ist, $h_t \in R^h$ ist der Output-Zustand, $c_t \in R^h$ ist der Zellstatus und alle $W \in R^{h \times d}$, $U \in R^{h \times h}$ und $b \in R^h$ sind trainierbare Gewichte. σ ist die Sigmoidfunktion, \tanh der Tangens hyperbolicus und \odot ist die elementweise Multiplikation.

Der Output-Zustand (h_t) wird aus dem aktuellen Input (x_t) und dem vorherigen Output-Zustand (h_{t-1}) berechnet. Die Menge an Informationen, welche durch die aktuelle Zelle geleitet werden, wird durch das Input Gate (i_t) und das Forget Gate (f_t) reguliert. Der Zellstatus (c_t) regelt, wie viel Information im Output-Zustand (h_t) gespeichert wird. Das Output Gate (o_t) regelt, wie viel Information an die nächste Zelle weitergegeben wird.

Gated Recurrent Units-Zellen Die GRU-Zelle kombiniert das Forget Gate mit dem Input Gate zu einem einzigen Update Gate und verschmilzt auch den Zellstatus mit dem Output-Zustand. Der Vorwärtsdurchlauf einer einzelnen GRU-Zelle mit h versteckten Knoten ist durch die folgenden Gleichungen gegeben:

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\ r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \\ \hat{h}_t &= \tanh(W_h x_t) + U_h (r_t \odot h_{t-1} + b_r) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \end{aligned}$$

wobei $z_t \in R^h$ das Update Gate ist, $r_t \in R^h$ das Reset Gate, $\hat{h}_t \in R^h$ ist die „candidate activation“, $h_t \in R^d$ ist der Output-Zustand und alle $W \in R^{h \times d}$, $U \in R^{h \times h}$ und $b \in R^h$ sind trainierbare Gewichte. Das Reset Gate (r_t) bestimmt, wie viel Information aus dem vorherigen Zustand bei der Berechnung der candidate activation (\hat{h}_t) vergessen wird. Das Update Gate ist die Menge an Informationen, welche von der Kandidatenaktivierung (\hat{h}_t) für die Berechnung des aktuellen Output-Zustands h_t genutzt wird.

2.5 Bayes'sche Hyperparameteroptimierung

Die Wahl geeigneter Hyperparameter kann großen Einfluss auf die resultierende Modellgüte haben. Aus diesem Grund haben wir uns entschieden, die Hyperparameter der ML-Modelle mit der Bayes'schen Optimierungsmethode zu optimieren. Nur bei den RF Modellen mit 3 Hyperparametern wurden

alle möglichen Kombinationen getestet (grid search), um die besten Hyperparameter zu finden. Die Schrittweise Lineare Regression hat keine Hyperparameter und muss demnach auch nicht optimiert werden.

Die Bayes'sche Optimierung ist ein globales Optimierungsverfahren für Black-Box-Funktionen (d. h. ohne bekannte Struktur und ableitungsfrei), welches häufig in Fällen angewendet wird, in denen das Berechnen der Zielfunktion zu rechenaufwendig ist. Sie geht zurück auf die Arbeiten von Kushner (1964), Zhilinskas (1975), Moćkus (1975, 1989), Moćkus et al. (1978) und wurde später von Jones et al. (1998) popularisiert. Im Anschluss einer Studie durch Snoek et al. (2012) wurde das Verfahren für seine Eignung zur Optimierung von ML Hyperparametern sehr bekannt.

Die Bayes'sche Optimierung besteht aus zwei Teilen: Einer Methode zur statistischen Inferenz sowie einer Akquisitionsfunktion. Die Methode für die statistische Inferenz ist üblicherweise ein Gaußprozess (GP), welcher bei jeder Iteration eine Posterior-Verteilung schätzt, die eine Schätzung für die zu optimierende Funktion darstellt. Die Akquisitionsfunktion wird verwendet, um bei jedem Optimierungsschritt den nächsten Punkt zur Funktionsevaluierung zu finden. Für diese Studie wurde dafür die Upper Confidence Bound (UCB) gewählt (Srinivas et al. 2009). Zusammenfassend lässt sich sagen, dass die Bayes'sche Optimierung bei jeder Iteration im Laufe der Optimierung ein Modell erstellt, um die Güte der Hyperparameter vorherzusagen und dadurch einen geeigneten nächsten Punkt auszuwählen. Die Hyperparameter aller optimierten Modelle und deren gewählten Grenzen befinden sich in Kap. 6, *Hyperparameter-Anwendungsbereich* im Anhang.

2.6 Modellgütefunktionen

Die Zielfunktion für alle Modelle und die Hyperparameteroptimierung ist die mittlere quadratische Abweichung (mean squared error, MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

wobei n die Stichprobengröße (Tage) ist und y_i die beobachteten und \hat{y}_i die vorhergesagten Wassertemperaturen. Zum Vergleich der Modellgüte verschie-



Abb. 3 Boxplots der Anzahl an Tagen mit Wassertemperaturen über dem 90 %-Quantil für alle Einzugsgebiete in der Trainings-/Validierungs- und Testperiode. Der Testzeitraum besteht jeweils aus den letzten 20 % der Zeitreihe eines Einzugsgebiets. Der Wert der 90 %-Quantile bezieht sich auf die jeweilige ganze Zeitreihe

dener Modelle werden die Wurzel der mittleren quadratischen Abweichung (root mean squared error, RMSE) und der mittlere absolute Fehler (MAE) verwendet

$$RMSE = \sqrt{MSE},$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$

2.7 Experimentelles Setup

Um alle angewandten Modelle objektiv vergleichen zu können, wurden die verfügbaren Datensätze in zwei Teile aufgeteilt: Die ersten 80 % der Zeitreihen wurden zum Trainieren/Validieren und die letzten 20 % zum Testen verwendet. Wir haben bewusst keine zufällige Aufteilung gewählt, da die Vorhersage der Wassertemperaturen für einen zukünftigen Zeitraum sich besser als Test für die möglichen Anwendungen eignet. Dies ist besonders relevant für Wassertemperatur, welche sich durch den Klimawandel induzierte Instationarität auszeichnet (Van Vliet et al. 2013). Die Trainings-/Validierungs- und Testzeiträume werden verglichen, um den Unterschied in der Wassertemperaturverteilung aller Einzugsgebiete zu beurteilen.

Step LM, RF und XGBoost werden mit einer Kreuzvalidierung (CV) opti-

miert. Es werden zwei Arten von CV angewandt: Eine 5 Mal wiederholte 10-fach stratifizierte CV und eine Zeitreihen-CV. Während bei der 10-fachen CV die Daten zufällig verteilt werden, werden bei der Zeitreihen-CV Daten schrittweise zum ursprünglichen Teil der Zeitreihe hinzugefügt, während die Modellgüte jedes Schritts bewertet wird. Die Zeitreihen-CV beginnt mit einem anfänglichen Zeitfenster von 730 Tagen für das Training und den folgenden 90 Tagen für die Validierung. Die Trainingsmenge wird nach jedem verschiedenen Kreuzvalidierungsset um 90 Tage erhöht, bis die gesamte Zeitreihe mit Ausnahme der letzten 90 Tage verwendet wurde. Daher ist die Anzahl der Teilmengen von der gesamten Länge der Zeitreihe abhängig (anstelle von 10 wie bei der 10-fachen CV).

Aufgrund von Rechenleistungs- und Zeitbeschränkungen wurde die Hyperparameteroptimierung für alle neuronalen Netze unter Verwendung von 60%/20 % der Trainings-/Validierungsdaten durchgeführt. Dies ist auch die Standardmethode für das Training neuronaler Netzwerke für reale Anwendungen und ermöglicht eine Modellvalidierung durch einmaliges Training eines Modells, während bei einer 5 Mal wiederholten 10-fachen CV ein Modell 50-mal trainiert werden müsste. Je nach

Modellierungsproblem kann es aber zu einer ungenaueren Einschätzung der Modellgüte kommen.

Die Bayes'sche Hyperparameteroptimierung besteht aus 20 zufälligen Parameterkombinationen und 40 Iterationen. Die Dateninputs für alle neuronalen Netze wurden standardisiert, indem der Mittelwert der Trainingsdaten subtrahiert und durch die Standardabweichung der Trainingsdaten dividiert wurde. Aus den optimierten Hyperparametersets werden 5 unabhängig voneinander trainierte Modelle erstellt, die zu einem Ensemble für die Vorhersage zusammengefasst werden. Die Ensemblevorhersagen sind die Mittelwerte der Vorhersagen der 5 Modelle. Das Verwenden von Ensembles aus mehreren Netzen ist eine Methode, die Generalisierungsfähigkeit eines neuronalen Netzes erhöhen kann und ist ein häufig verwendeter Ansatz, welcher erstmals durch die Arbeit von Hansen und Salamon (1990) eingeführt wurde. Außerdem wurde early stopping mit patience=5 auf alle neuronalen Netze angewandt, um einen Overfit zu vermeiden.

Je Modelltyp und Experiment wird das Modell mit der besten Modellgüte anhand des Validierungs-RMSE ausgewählt. Test-RMSE und MAE-Ergebnisse werden erst nach Auswahl der Modelle mit minimalem Validierungs-RMSE

Tab. 3 Übersicht der Modellgüten der jeweils besten ML-Modelle in den einzelnen Einzugsgebieten und der beiden Vergleichsmodelle LM und air2stream. Die besten ML-Modelle wurden jeweils mithilfe des Validierungs-RMSE ausgewählt, während Test-RMSE und MAE nie in an der Auswahl oder dem Training der Modelle beteiligt waren. Alle gezeigten Werte beziehen sich auf die Test-Zeitperiode des jeweiligen Einzugsgebiets

Einzugsgebiet	Beste ML-Modell		LM		air2stream			
	Modell	Experiment	RMSE (°C)	MAE (°)	RMSE (°C)	MAE (°)	RMSE (°C)	MAE (°)
Kleine Mühl	XGBoost	4(TQP)	0,740	0,578	1,744	1,377	0,908	0,714
Aschach	XGBoost	6(TQPGL)	0,815	0,675	1,777	1,408	1,147	0,882
Erlauf	XGBoost	6(TQPGL)	0,530	0,419	1,354	1,057	0,911	0,726
Traisen	FNN	3(TQ)	0,526	0,392	1,254	0,97	0,948	0,747
Ybbs	RF	3(TQ)	0,576	0,454	1,787	1,415	0,948	0,756
Saalach	XGBoost	6(TQPGL)	0,527	0,420	1,297	1,062	0,802	0,646
Enns	FNN	6(TQPGL)	0,454	0,347	1,425	1,166	1,168	0,671
Inn	FNN	3(TQ)	0,422	0,329	1,376	0,098	1,097	0,949
Salzach	FNN	4(TQP)	0,430	0,338	1,327	1,077	0,743	0,595
Donau	RNN-LSTM	3(TQ)	0,521	0,415	2,145	1,721	1,099	0,91
		Mittelwert:	0,554	0,437	1,549	1,235	0,977	0,76

verglichen. Daher ist es möglich, dass einige Modelle eine höhere Testperformance aufweisen, jedoch nicht als bestes Modell für einen bestimmten Modelltyp und/oder ein bestimmtes Experiment ausgewählt werden. Dies soll eine reale Anwendung widerspiegeln, bei welcher die Testdaten als eine zuvor unbekannt zukünftige Zeitreihe dienen.

2.8 Statistische Tests

Zum Testen unterschiedlicher Trainings-/Modellcharakteristika und Dateinputs sowie der Unterschiede in der Modellgüte wurde der Kruskal-Wallis-Test (Kruskal und Wallis 1952) verwendet. Der Dunn-Test für Mehrfachvergleiche (Dunn 1964) wurde für paarweise Vergleiche zwischen den Modellgüten verwendet. Mit einem multiplen linearen Regressionsmodell wurde der Zusammenhang zwischen Modelltypen, Experimenten und Einzugsgebieten anhand von Test-RMSE untersucht. Das Signifikanzniveau wurde für alle statistischen Tests auf $p=0,05$ gesetzt.

2.9 Open source R-Paket

Die Datenaufbereitung und Modelle wurden in das open source R-package wateRtemp implementiert, welches unter github.com/MoritzFeigl/wateRtemp zur Verfügung gestellt wird. Dies bietet leicht anwendbare Modellierungswerkzeuge für die Wassertemperatur-Community und ermöglicht es, alle Ergebnisse dieser Studie zu reproduzieren. Die gesamte Programmierung wurde in R (R Core Team 2020) durchgeführt, wobei die Modellentwicklung auf Caret

(Kuhn 2020), XGboost (Chen et al. 2020) und TensorFlow (Allaire und Tang 2020) aufbaut und die Visualisierungen auf ggplot2 (Wickham 2016).

3 Ergebnisse

3.1 Zeitreihencharakteristika

Aufgrund der durch den Klimawandel stetig steigenden Lufttemperaturen ist auch die Wassertemperatur keine stationäre Größe und weist eine Erwärmung auf (Kędra 2020; Mohseni et al. 1999). Dies ist deutlich sichtbar, wenn man die Veränderung der Anzahl an extrem warmen Tagen und den Anstieg der mittleren Wassertemperatur in den Einzugsgebieten im Laufe der Zeit betrachtet. Hierfür haben wir die Trainings-/Validierungs- und Testzeitdaten in jedem Einzugsgebiet verglichen. Da die Testdaten aus den letzten 20% der Gesamtdaten bestehen, ist die genaue Länge dieser Zeitreihen abhängig von der Datenverfügbarkeit in den einzelnen Einzugsgebieten. Generell umfasst die Testzeit aber einen Zeitraum der Jahre 2008 bis 2015. Wir können einen Anstieg von 138% der medianen Anzahl an Tagen mit einer Wassertemperatur oberhalb des 90%-Quantils zwischen Trainings-/Validierungs- und Testzeitraum in allen Einzugsgebieten feststellen. Dieser Anstieg reicht von 69%, oder von 32 Tagen auf 54 Tage im Donaeinzugsgebiet und bis zu 285%, oder von 26 Tagen auf 100 Tage, im Einzugsgebiet der Salzach. Diese Veränderung ist noch deutlicher, wenn man das letzte Jahr der Testdaten (2015) mit allen anderen verfügbaren Jahren vergleicht. Hier steigt die me-

diane Anzahl der Tage mit Wassertemperaturen über dem 90%-Quantil (berechnet für die gesamte Zeitreihe) aller Einzugsgebiete um 273%. Abb. 3 zeigt die entsprechenden Boxplots der Tage mit einer Fließgewässertemperatur über dem 90%-Quantil für jedes Einzugsgebiet im Trainings-/Validierungs- und Testzeitraum. Ein ähnliches Muster kann in den Änderungen der mittleren jährlichen Fließgewässertemperaturen beobachtet werden. Der mediane Anstieg der mittleren jährlichen Wassertemperatur aller Einzugsgebiete beträgt $0,48^{\circ}\text{C}$ beim Vergleich von Training/Validierung mit dem Testzeitraum und $0,77^{\circ}\text{C}$ beim Vergleich des letzten Jahres des Testzeitraums (2015) mit allen anderen Jahren. Da sich der Testzeitraum, wie hier in Bezug auf die Extreme gezeigt wird, vom Trainings-/Validierungszeitraum nennenswert unterscheidet, werden die Modelle auch darauf getestet, wie sie sich unter instationären Bedingungen verhalten. Dies ist ein Test, bei welchem Umweltmodelle häufig versagen (z. B. Kling et al. 2015).

3.2 Vergleich der Modellgüten

Tab. 3 gibt einen Überblick über die jeweils besten ML-Modelle in den einzelnen Einzugsgebieten und der beiden Vergleichsmodellen LM und air2stream. Der mittlere Test-RMSE von LM beträgt $1,55^{\circ}\text{C}$ mit Werten im Intervall $[1,25, 2,15]^{\circ}\text{C}$, während air2stream einen mittleren Test-RMSE von $0,98^{\circ}\text{C}$ mit Werten im Intervall $[0,74, 1,17]^{\circ}\text{C}$ aufweist. Die Modellgüten für jedes Einzugsgebiet zeigen, dass air2stream immer eine bessere Vorhersage gene-

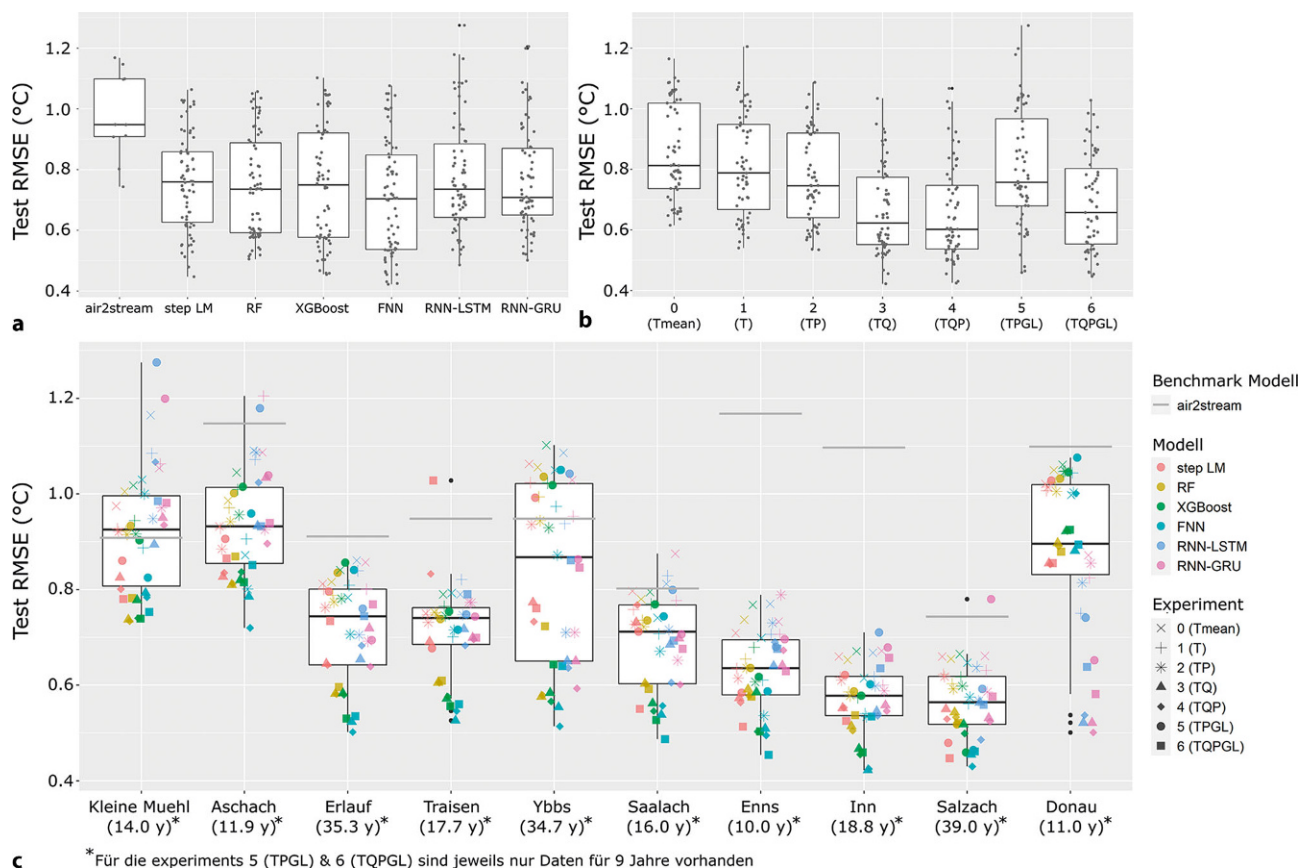


Abb. 4 Boxplots der Modellgüten **a** der einzelnen Modellarten, **b** der unterschiedlichen Experimente und **c** der Test-RMSE in den einzelnen Einzugsgebieten mit überlagerten Punkten, welche die Modellgüte einzelner Kombinationen von Einzugsgebiet, Modellart und Experiment zeigen. Die Einzugsgebiete sind von klein (*links*) nach groß (*rechts*) geordnet und haben zusätzliche Information über die Zeitreihenlänge in den darunter stehenden Klammern. Die air2stream-Ergebnisse des jeweiligen Einzugsgebiets sind als grauer Strich dargestellt. Um die Lesbarkeit zu gewährleisten, wurden die sehr viel höheren RMSE-Werte der LM-Modelle nicht dargestellt

riert als LM und folglich einen signifikant niedrigeren Test-RMSE aufweist ($p < 0,001$). Der mittlere Test-RMSE der besten ML-Modelle je Einzugsgebiet beträgt $0,55\text{ °C}$ mit Werten im Intervall $[0,42, 0,82]\text{ °C}$ und erreichte immer eine höhere Modellgüte als air2stream. Basierend auf den RMSE-Mittelwerten ist das beste ML-Modell verglichen mit LM und air2stream jeweils um 64% und 43% genauer in seiner Vorhersage. Dies führt zu einem signifikant niedrigeren Test-RMSE der getesteten ML-Modelle im Vergleich zum air2stream-Vergleich ($p < 0,001$).

Sowohl XGBoost als auch FNN sind in 4 von 10 analysierten Einzugsgebieten das jeweils beste Modell. RF erwies sich als das beste Modell im Salzach-Einzugsgebiet und RNN-LSTM im Donau-Einzugsgebiet. Step-LM und RNN-GRU übertrafen keines der anderen Modelle in den untersuchten Einzugsgebieten. Experiment 3, welches als Inputs nur Lufttemperatur

und Abfluss inkludiert, erzielte in vier Einzugsgebieten die beste Modellgüte. Experiment 6, welches alle verfügbaren Inputs verwendet, erzielte ebenfalls in vier Einzugsgebieten die höchste Modellgüte. Experiment 4, welches als Inputs für Lufttemperatur-, Abfluss- und Niederschlag enthält, erzielte in zwei Einzugsgebieten die besten Ergebnisse.

Abb. 4 zeigt die Ergebnisse aller Modelle, Einzugsgebiete und Experimentkombinationen. Die Boxplots in Abb. 4a zeigen die Modellgüten in Abhängigkeit vom Modelltyp für alle Experimente. Die Ergebnisse des Kruskal-Wallis-Tests zeigen keinen signifikanten Unterschied ($p = 0,11$) des Test-RMSE zwischen den verschiedenen Modelltypen. Abb. 4b zeigt Boxplots der Modellperformances für alle Experimente. Die Ergebnisse des Kruskal-Wallis-Tests zeigen einen hochsignifikanten Unterschied im Test-RMSE der verschiedenen Experimente ($p < 10^{-14}$). Die Ergebnisse in Abb. 4b zeigen einen

Anstieg des medianen RMSE mit zunehmender Anzahl von Inputvariablen bis Experiment 4 (TQP). Wenn man die Globalstrahlung als zusätzliche Inputvariable hinzufügt, steigt der mediane RMSE nicht weiter an. Dies kann durch eine reduzierte Zeitreihenlänge der Experimente 5 (TPGL) und 6 (TQPGL) erklärt werden (da die Globalstrahlung erst ab 2007 verfügbar war). Ein Vergleich zwischen Experimenten mit gleicher Zeitreihenlänge (Experimente 0 bis 4 und Experimente 5 und 6) zeigt, dass Abflussinformationen die Modellgüte verbessern.

Abb. 4c zeigt den RMSE für jedes Einzugsgebiet als Boxplot. Eine entsprechende Abbildung der MAE-Ergebnisse wird in Abschn. 6, Abb. 7 im Anhang gezeigt. Die Boxplots sind mit Punkten überlagert, welche einen Überblick über die individuellen RMSE-Werte der einzelnen Modell- und Experimentkombinationen geben. Um eine bessere Sichtbarkeit zu gewährleisten, sind die

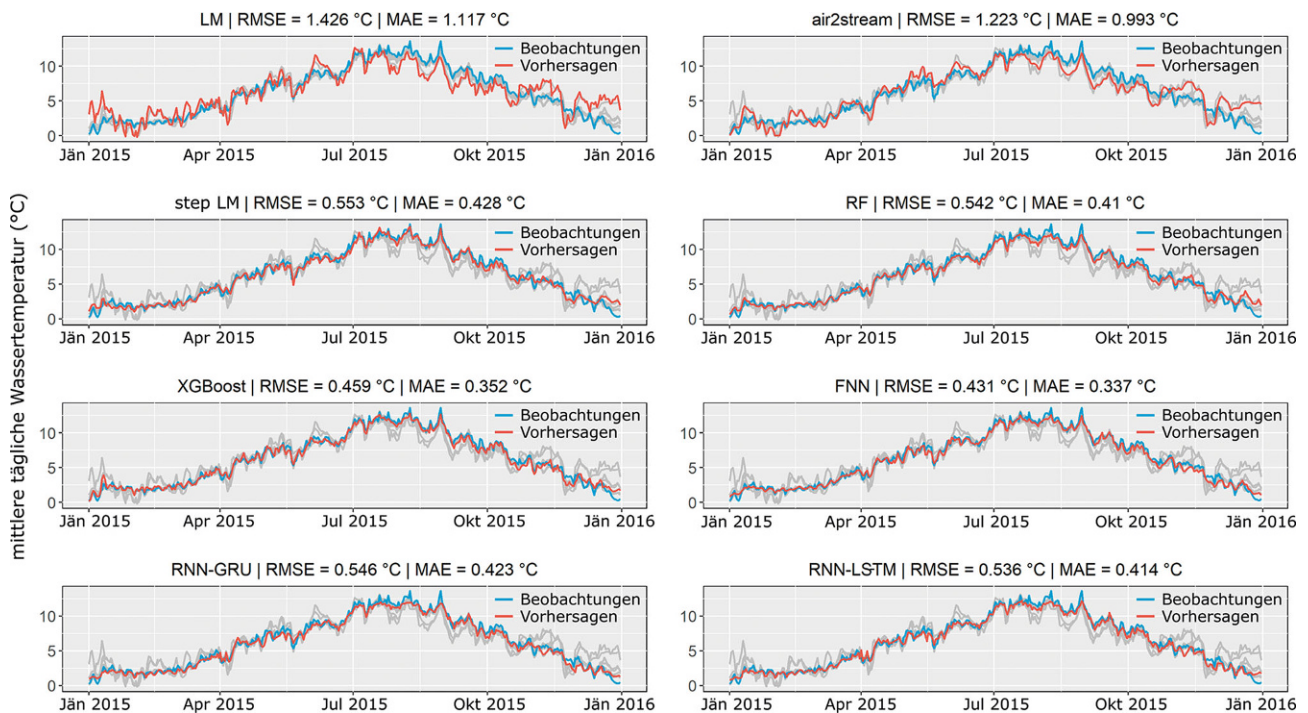


Abb. 5 Vergleich der vorhergesagten Fließgewässertemperaturen aller getesteten Modellarten für das Inn-Einzugsgebiet im Jahr 2015. Die Daten des Jahres 2015 wurden nicht für das Trainieren der Modelle verwendet. Die Vorhersagen sind mit roten Linien dargestellt, die beobachteten Wassertemperaturen mit blauen Linien. Die Vorhersagen aller anderen Modelle werden zum Vergleich jeweils mit grauen Linien dargestellt

Punkte in horizontaler Richtung verschoben. Der Unterschied der Modellgüte zwischen den Einzugsgebieten ist deutlich erkennbar und reicht von einem medianen RMSE von etwa $0,93\text{ }^{\circ}\text{C}$ in den Einzugsgebieten Kleine Mühl und Aschach bis hin zu einem mittleren RMSE von $0,58\text{ }^{\circ}\text{C}$ im Einzugsgebiet Inn.

In Abb. 4c sind auch die Modellgüten von air2stream als graue Linie für jedes Einzugsgebiet dargestellt. Nahezu alle getesteten Experimente und Modellkombinationen zeigten eine verbesserte Vorhersagefähigkeit im Vergleich zu air2stream. Nur in 5 Einzugsgebieten waren auch Kombinationen dabei (Experimente 0, 1, 5 und einmal Experiment 6), in denen air2stream besser abschnitt. Dies ist überraschend wenige Modelle, wenn man bedenkt, dass sowohl die Experimente 0, 1, 5 als auch 6 aufgrund der Menge an Informationen, welche für die Vorhersage zur Verfügung stehen, stark eingeschränkt sind. Experiment 0 und 1, welche nur die Lufttemperatur verwenden, sind immer noch in der Lage, die Vorhersagen im Vergleich zu air2stream für alle Modelltypen in 7 Einzugsgebieten zu verbessern. Ähnlich sind die Experi-

mente 5 und 6 mit Trainingsdaten aus nur 6 Jahren in der Lage, die Vorhersagen im Vergleich zu air2stream für alle Modelltypen in 5 Einzugsgebieten zu verbessern.

Aus den Ergebnissen in Abb. 4a–c geht hervor, dass die Modellgüte im Allgemeinen von der Kombination aus Modell, Dateninput (Experiment) und Einzugsgebiet abhängt, während der Einfluss auf den Test-RMSE verschiedener Experimente und Einzugsgebiete größer ist als der Einfluss von Modelltypen. Die Lineare Regression für den Test-RMSE mit Einzugsgebiet, Experiment und Modelltyp als Regressoren ist in der Lage, den größten Teil der Varianz des Test-RMSE mit einem Bestimmtheitsmaß von $R^2 = 0,988$ zu erklären. Außerdem ergab sich ein signifikanter Zusammenhang aller Einzugsgebiete ($p < 10^{-15}$), aller Experimente ($p < 0,005$) und dem FNN-Modelltyp ($p < 0,001$) mit dem Test-RMSE. Der geschätzte Koeffizient des FNN beträgt $-0,05$, wodurch sich auf eine Verbesserung der Vorhersage bei Anwendung des FNN-Modells schließen lässt. Alle anderen Modelltypen zeigen keinen signifikanten Zusammenhang. Dies könnte jedoch auf die hier verwendete kleine

Stichprobe zurückzuführen sein, da die geschätzten Koeffizienten der Modelltypen (Mittelwert: $-0,01$, Intervall: $[-0,05, 0,02]$) generell klein sind, wenn sie mit Einzugsgebietenkoeffizienten (Mittelwert: $0,86$, Intervall: $[0,69, 1,06]$) und Experimentkoeffizienten (Mittelwert: $-0,12$, Intervall: $[-0,2, -0,04]$) verglichen werden.

Mehrere Experimente führen oft zu sehr ähnlichen RMSE-Werten für einen einzelnen Modelltyp. Weiters liegen die besten Experimente für verschiedene Modelltypen immer sehr nahe beieinander. Daraus ergibt sich beim Test-RMSE ein medianer Unterschied von $0,08\text{ }^{\circ}\text{C}$ zwischen den besten Experimenten verschiedener Modelltypen und ein medianer Unterschied von nur $0,035\text{ }^{\circ}\text{C}$ im Test-RMSE zwischen dem besten und zweitbesten Modell eines anderen Modelltyps. Andererseits beträgt die mediane Differenz zwischen dem RMSE-Werten der ML-Modelle und air2stream $-0,39\text{ }^{\circ}\text{C}$.

3.3 Detaillierte Analyse eines Einzugsgebiets

Um den Unterschied in der Vorhersagefähigkeit der Modelle weiter zu

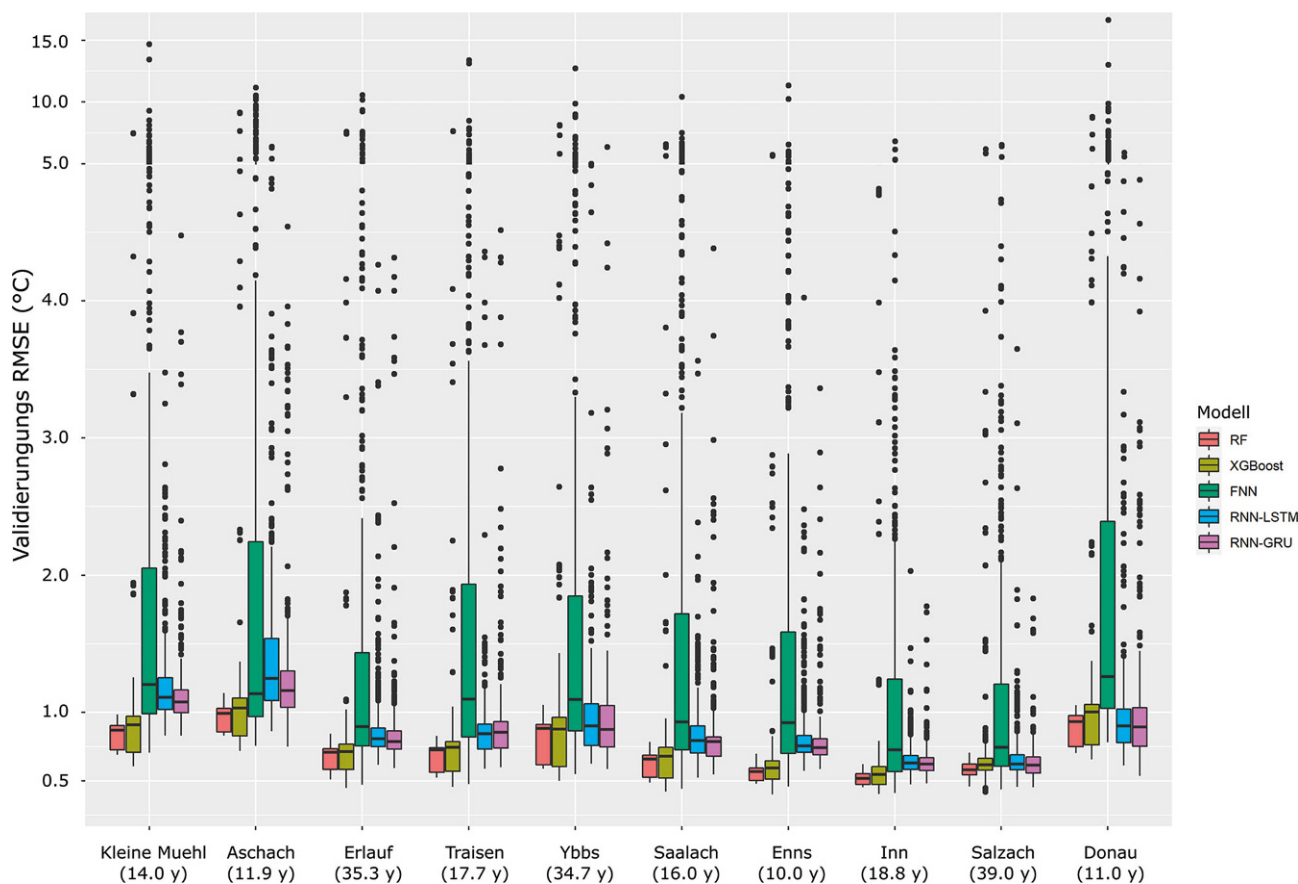


Abb. 6 Boxplots der Validierungs-RMSE aller Modellarten, Einzugsgebiete und Experimente. Die Einzugsgebiete sind nach Gebietsgröße angeordnet (von links nach rechts) mit zusätzlicher Information über die Zeitreihenlänge in Klammer

untersuchen, werden die vorhergesagten Fließgewässertemperaturen für das letzte Jahr der Testdaten (2015) des Einzugsgebietes Inn untersucht. Das Jahr 2015 wurde für den Vergleich gewählt, da es eine außergewöhnlich große Anzahl an Tagen mit hohen Wassertemperaturen aufweist und sich daher für eine robuste Evaluierung der Modelle anbietet. Es ist ein Test unter instationären Bedingungen: Der Zeitraum 1997–2014 hat einen Median von 30 Tagen pro Jahr mit Fließgewässertemperaturen über 11°C, während im Jahr 2015 102 Tage mit solch hohen Wassertemperaturen beobachtet wurden. Abb. 5 zeigt die Vorhersageergebnisse der einzelnen Modelle (rote Linien) im Vergleich zu den Beobachtungen (blaue Linie) und allen anderen Modellvorhersagen (graue Linien) für das Jahr 2015 sowie das entsprechende RMSE- und MAE-Ergebnis für dieses Jahr.

Die beiden Vergleichsmodelle LM und air2stream zeigen große Unterschiede zwischen Vorhersagen und Beobachtungen und weisen generell

ein ganz anderes Verhalten auf als alle getesteten ML-Modelle. Während die größten Vorhersagefehler der getesteten ML-Modelle in ähnlichen Zeiträumen auftreten, sind bei beiden Benchmark-Modellen große Abweichungen über das ganze Jahr hinweg zu beobachten.

Die größten Vorhersagefehler aller ML-Modelle treten während wärmerer Perioden auf und erreichen ihren Höhepunkt in den Sommermonaten sowie während Perioden mit niedriger Wassertemperatur im November und Dezember. Dies ist in allen getesteten Modellen deutlich sichtbar. Daher resultieren Unterschiede in RMSE und MAE hauptsächlich aus den Abweichungen in diesen Perioden und können demnach relativ groß sein, obwohl der tatsächliche numerische Unterschied eher gering ist. Dies ist zu beobachten, wenn man die Ergebnisse der besten Modelle FNN und RNN-GRU in Abb. 5 vergleicht. Beide Modelle liefern ähnliche Vorhersageergebnisse für den größten Teil des Jahres. Das FNN-Modell ist jedoch besser in der Lage,

die Spitzen mit hohen Fließgewässertemperaturen in den Sommermonaten vorherzusagen, was zu einer RMSE- und MAE-Differenz von 0,115 bzw. 0,086 führt. Sehr geringe Unterschiede in RMSE und MAE, wie sie zwischen den beiden besten Modellen FNN und XGBoost zu sehen sind, führen nur zu sehr subtilen Unterschieden in den vorhergesagten Zeitreihen. Sehr ähnliche Beobachtungen können bei der Analyse der Vorhersageergebnisse in den anderen Einzugsgebieten gemacht werden. Die einzige Ausnahme kann im größten Einzugsgebiet Donau (Abschn. 6, Abb. 8 im Anhang) beobachtet werden, in welchem die Zeitreihe viel gleichmäßiger erscheint und relativ wenige Spitzen in der Wassertemperatur aufweist. Dies führt dazu, dass die RNN-Modelle die besten Ergebnisse liefern, mit einem großen RMSE-Unterschied im Vergleich zu allen anderen Modellen.

Tab. 4 Verteilungseigenschaften der Validierungs-RMSE-Werte resultierend aus der Hyperparameteroptimierung. Für jedes Modell ist der Mittelwert (μ), die Standardabweichung (σ) sowie Minimum und Maximum Werte angegeben

Modell	Validation RMSE (°C)			
	μ	σ	Min	Max
RF	0,70	0,16	0,45	1,14
XGBoost	0,95	1,07	0,40	9,15
FNN	1,70	1,60	0,41	16,6
RNN-LSTM	0,97	0,53	0,46	6,40
RNN-GRU	0,91	0,44	0,45	6,30
RF	0,70	0,16	0,45	1,14

3.4 Einfluss von Zeitvariablen für RNNs, Kreuzvalidierungsverfahren

Das Entfernen der Zeitinformation in Form von Fuzzy-Months aus den Trainingsdaten der RNNs verändert den Test-RMSE der Einzugsgebiete nicht signifikant ($p=0,17$). Allerdings wird die durch die Hyperparameteroptimierung geschätzte optimale Anzahl von Zeitschritten signifikant erhöht ($p=0,02$). Wenn die Zeitinformation aus den Inputs entfernt wird, sind die optimalen Zeitschritte 37,78 Tage länger als bei Verwendung der Zeitinformation als zusätzlichem Input. Dies erhöht die Trainingszeit des Modells signifikant ($p=0,034$), im Mittel um 132,45 min. Die verschiedenen CV-Schemata, welche auf die step-wise LM, RF und XGBoost angewandt wurden, zeigten keinen signifikanten Unterschied in den resultierenden Test-RMSE ($p=0,91$).

3.5 Einfluss von Hyperparametern auf die Modellergebnisse

Der Einfluss verschiedener Hyperparameter auf die Modellgüte ist in Abb. 6 dargestellt. Diese Abbildung zeigt den Validierungs-RMSE für alle Parametersätze, welche bei der Hyperparameteroptimierung verwendet wurden. Es ist ein großer Unterschied in der Variabilität der Modellgüte bei verschiedenen Modellen zu beobachten. Die Mittelwerte, Standardabweichungen sowie Minimum und Maximum der Validierungs-RMSE aller Modelle sind in Tab. 4 dargestellt. Die größte Variabilität zeigt sich bei den FNN-Ergebnissen mit einer RMSE-Standardabweichung von $\sigma_{\text{FNN}} = 1,60^\circ\text{C}$ und Werten zwischen 0,41 und 16,6°C. Danach folgt XGBoost, welches mehrere Ausreißer

in jedem Einzugsgebiet hat und zu $\sigma_{\text{XGBoost}} = 1,07^\circ\text{C}$ und Werten zwischen 0,40 und 9,15°C führt. Beide RNNs zeigen eine sehr ähnliche Variabilität mit RMSE zwischen 0,45 und 6,3°C. Im Vergleich zu allen anderen getesteten Modellen hat das RF-Modell eine deutlich kleinere RMSE-Streuung aufgrund der gewählten Hyperparameter, mit $\sigma_{\text{RF}} = 0,16^\circ\text{C}$ und Werten zwischen 0,45 und 1,14°C.

4 Diskussion

Diese Studie untersucht die Vorhersagefähigkeit 6 verschiedener ML-Modelle für Fließgewässertemperaturen mit einer Reihe von Input-Datensätzen in 10 Einzugsgebieten. Als Referenz und um die Ergebnisse mit anderen Studien zu vergleichen, wurden zwei weit verbreitete Vergleichsmodelle herangezogen: step-LM und air2stream. Die Ergebnisse zeigen allgemein eine sehr ähnliche Modellgüte der getesteten ML-Modelle mit einer medianen Test-RMSE-Differenz von nur 0,08°C zwischen den Modellen. Im Gegensatz dazu erzielten die Modelle deutlich bessere Ergebnisse im Vergleich zu air2stream und wiesen eine mittlere Test-RMSE-Abnahme von 0,42°C (42%) auf. Die Ergebnisse zeigten, dass beinahe die gesamte Test-RMSE-Varianz ($R^2 = 0,99$) durch das Einzugsgebiet, den Inputdatensatz und den Modelltyp erklärt werden kann. Dies zeigte auch, dass der resultierende Test-RMSE signifikant von der Art der Inputdaten beeinflusst wird, wobei mehr Inputs im Allgemeinen in einer höheren Modellgüte resultieren und dass von allen Modellen nur das FNN-Modell einen signifikanten Zusammenhang mit niedrigeren Test-RMSE hatte. Darüber hinaus wurde eine große Spannungsbreite an Modellgüten in Abhängigkeit von den gewählten Hyperparametern der Modelle beobachtet, mit einer extrem großen RMSE-Standardabweichung (1,60°C) bei den FNN-Modellen.

Bis auf wenige Modelltyp- und Experimentkombinationen zeigten alle getesteten ML-Modelle im Vergleich zu den beiden Vergleichsmodellen eine verbesserte Modellgüte. Der Unterschied zwischen den Vergleichsmodellen und den getesteten Modellen war nicht nur in den resultierenden Test-RMSE und Test-MAE sichtbar, sondern auch deutlich in der Bandbreite und Häufigkeit des Auftretens großer Vorhersagefehler in den vorhergesagten

Zeitreihen (siehe Abb. 5). Angesichts des Bereichs der geschätzten Koeffizienten der Einzugsgebiete [0,69, 1,06], der Dateninputs [-0,2, -0,04] und der Modelltypen [-0,05, 0,02] im Regressionsmodell für den Test-RMSE können wir feststellen, dass bei einem adäquaten Modellaufbau und korrekt ausgewählten Hyperparametern der Einfluss verschiedener Dateninputs und unterschiedlicher Einzugsgebiete viel größer ist als der Einfluss der untersuchten ML-Modelltypen. Wir sehen Anzeichen, dass das FNN-Modell die beste Wahl ist, da es das einzige Modell war, welches einen signifikanten Zusammenhang mit niedrigeren RMSE hatte und auch den größten geschätzten Koeffizienten aller Modelltypen (-0,05) aufwies.

Die hier vorgestellten Ergebnisse zeigen, dass FNN und XGBoost in 8 von 10 Einzugsgebieten am besten abschneiden und daher eine erste Wahl für die Modellierung von Fließgewässertemperaturen anhand von ML-Modellen sein sollten. Für die Modellierung ähnlich großer Einzugsgebiete wie dem der Donau (96.000 km²), wo langfristige Abhängigkeiten relevanter zu sein scheinen, sind RNNs die beste Wahl. Beide RNN-Architekturen, GRU und LSTM, liefern im Donaeinzugsgebiet sehr ähnliche Ergebnisse (Test-RMSE 0,52°C). Dies ist deutlich niedriger als der mittlere Test-RMSE der anderen Modelle (0,90°C) und air2stream (1,10°C). Das RF-Modell hat die geringste Standardabweichung im resultierenden RMSE in Abhängigkeit von den gewählten Hyperparametern (0,16°C) und könnte daher die angemessenste Wahl in Situationen mit begrenzten Rechenressourcen sein. Generell führen mehr Inputdaten zu besseren Ergebnissen, jedoch erzeugt auch die Kombination von Lufttemperatur- und Abflussdaten bereits Vorhersageergebnisse mit einem mittleren RMSE von 0,62°C. Durch die zusätzliche Berücksichtigung von Niederschlagsdaten kann eine weitere Verbesserung (RMSE 0,60°C) erreicht werden. Auch die Einbindung von GL kann die Modellgüte potenziell erhöhen: Das Experiment 6 zeigt einen ähnlichen Leistungsbereich wie Experiment 3, obwohl nur 6 Jahre zum Trainieren verwendet werden. Das Experiment 2 (TP) ist für die praktische Anwendung wohl am relevantesten, da die verwendeten Eingangsdaten für die meisten Regionen sowie auch aus Klimamodellen verfügbar sind. Der RMSE von Experiment 2 ist 0,75°C, was einer

Verschlechterung von 19 % im Vergleich zum Experiment mit dem niedrigsten RMSE-Median und einer Verbesserung von 21 % im Vergleich zu air2stream bedeutet. Das lässt daraus schließen, dass Modellierungen mit T und P für kurzfristige Prognosen und Klimawandelstudien verwendet werden können.

Im Vergleich zu air2stream, dessen Anwendung weit verbreitet ist, erzielten die Vorhersagen der vorgestellten ML-Modelle eine deutlich bessere Modellgüte. Dies steht im Gegensatz zu den Ergebnissen von Zhu et al. (2019b), welche die Modellgüte unterschiedlicher ML-Modelle für die Modellierung der Fließgewässertemperaturen untersucht haben. Die Ergebnisse von Zhu et al. (2019b), die 8 Einzugsgebiete untersucht haben, zeigten, dass air2stream bessere Vorhersagen erzielte als FNNs, Gauß-Prozess-Regressions- und Entscheidungsbaummodelle mit Wassertemperatur, Abfluss und Jahrestag als Eingangsdaten. Die hier vorgestellten air2stream-Ergebnisse haben Test-RMSE im Bereich von [0,74, 1,17]°C, der einerseits vergleichbar ist mit den Ergebnissen von Zhu et al. (2019b), deren Test-RMSE sich im Bereich [0,64, 1,16]°C bewegte, und andererseits mit weiteren air2stream-Studien wie z. B. der von Piotrowski und Napiorkowski (2018) mit einem Bereich von [0,625, 1,31]°C. Daraus schließen wir, dass die von uns erzielten Modellgüten mit anderen air2stream-Anwendungen vergleichbar sind und somit eine konsistente Referenz darstellen, auch wenn air2stream ursprünglich für die Verwendung von Punktquellendaten entwickelt wurde und nicht für aggregierte Einzugsgebietsmittel, die wir in dieser Studie als Eingangsdaten verwendet haben. Folglich zeigen die hier vorgestellten ML-Ansätze eine signifikante Verbesserung im Vergleich zu bereits getesteten ML-Ansätzen zur Modellierung von Fließgewässertemperaturen.

Aufgrund der fehlenden Abbildung physikalischer Zusammenhänge wird statistischen Modellierungsansätzen oft nachgesagt, für Extrapolationen außerhalb ihres Trainingsdatenbereichs ungeeignet zu sein (Benyahya et al. 2007). ML-Methoden sind jedoch flexibler und leistungsfähiger als bisherige statistische Ansätze und sind in der Lage, räumliche und zeitliche Informationen auf verschiedenen Skalen gleichzeitig zu nutzen (Reichstein et al. 2019). Diese Punkte sind besonders wichtig

für Studien im Zusammenhang mit dem Klimawandel, wo steigende Lufttemperaturen die statistischen Zusammenhänge zwischen meteorologischen Parametern und der Fließgewässertemperatur verändern könnten. Um die Extrapolationsfähigkeit der betrachteten ML-Methoden zu untersuchen, wählten wir die wesentlich wärmeren letzten Jahre der Zeitreihe als Testzeitraum und analysierten das letzte Jahr im Detail, welches das Jahr mit den meisten Tagen mit Extremtemperaturen war. Alle getesteten Modelle waren in der Lage, Vorhersagen mit einer Modellgüte nahe der Trainingsleistung im Testzeitraum und im Jahr mit den meisten Temperaturanomalien zu erreichen. Diese Ergebnisse zeigen, dass es zumindest für kurzfristige Vorhersagen (die hier verwendeten Testzeiträume bestehen aus 1 bis 8 Jahren) unter einem sich ändernden Klima möglich ist, robuste Vorhersageergebnisse zu erzielen. Eine erfolgreiche Extrapolation für kurze Zeiträume legt nahe, dass auch mittel- bis langfristige Vorhersagen vernünftige Ergebnisse liefern könnten. Dies kann jedoch nur auf der Grundlage zukünftiger Beobachtungen beurteilt werden. Es liegt jedoch auf der Hand, dass die ML-Ansätze bei der Extrapolation versagen werden, wenn sich auch die Einzugsgebietseigenschaften verändern. Beispielsweise kann im Zusammenhang mit hochalpinen, gletscherdominierten Einzugsgebieten angenommen werden, dass sich die Fließgewässertemperaturcharakteristik verändern wird, wenn die Gletscher infolge des Klimawandels verschwinden. Als Folge davon werden die der Wassertemperatur im Fließgewässer zugrundeliegenden Prozesse verändert. Diese Prozessveränderungen werden in den ML-Ansätzen nicht berücksichtigt. Modellierungen in diese Richtung würden jedenfalls stark prozessbasierte Ansätze erfordern. Auch air2stream hätte in dieser Hinsicht keinen Vorteil.

Die in dieser Studie erzielten Ergebnisse variieren je nach gewählten Hyperparametern der ML-Modelle erheblich. Auch die zwei leistungsstärksten Modelle XGBoost und FNN zeigen aufgrund der gewählten Hyperparameter eine große Varianz in der Modellgüte. Dies lässt den Schluss zu, dass Flexibilität zwar für ein gut funktionierendes Modell notwendig ist, aber auch eine mögliche Fehlerquelle darstellt oder zu verminderten Modellgüten führen kann. Diese Ergebnisse unterstreichen

die Bedeutung der Hyperparameter-Optimierung von ML-Modellen und könnten eine mögliche Erklärung dafür sein, dass insbesondere FNNs in anderen Studien nicht gleich gut abschneiden. Die meisten Publikationen, die FNNs für die Modellierung von Fließgewässertemperaturen testeten, verwendeten nur wenige FNN-Hyperparameter-Kombinationen, die meist durch „Trial and Error“ ausgewählt wurden (z. B. Abba et al. 2017; Piotrowski et al. 2015; Rabi et al. 2015; Temizyurek und Dadaser-Celik 2018; Zhu et al. 2018, 2019a). Unsere Ergebnisse zeigen den extrem großen Einfluss von Hyperparametern auf die Modellgüte, weshalb „Trial and Error“-Ansätze als unzureichend und mit Sicherheit nicht optimal anzusehen sind.

In aktuellen Studien wurden RNNs erfolgreich zur Niederschlag-Abfluss-Modellierung eingesetzt (z. B. Kratzert et al. 2018; Li et al. 2020; Xiang et al. 2020) und zeigen vielversprechende Ergebnisse für die Etablierung als Instrument zur Abflussvorhersage. Unsere Ergebnisse zeigen in den meisten Einzugsgebieten eine unterdurchschnittliche Modellgüte im Vergleich zu den anderen getesteten ML-Modellen. Dies ist insofern relevant, da RNNs im Vergleich zu den anderen Methoden eine Reihe von vorangegangenen Zeitschritten für die Vorhersage verwenden (optimierte Hyperparameter), die viel mehr Informationen enthalten als die 4 vorherigen Zeitschritte (Lags), die für die anderen Modelle verfügbar sind. Im größten untersuchten Einzugsgebiet (Donau) sind RNNs die leistungsstärksten Modelle, was darauf hindeutet, dass RNNs besonders gut darin sind, Prozesse mit langfristigen Abhängigkeiten zu modellieren. Diese langfristigen Abhängigkeiten resultieren wahrscheinlich aus längeren Konzentrationszeiten, die im Allgemeinen von der Einzugsgebietsgröße abhängig sind (McGlynn et al. 2004). Für alle anderen Einzugsgebiete in dieser Studie waren die Parameterinformationen von 4 vorangegangenen Tagen (Lags) ausreichend und RNNs nicht in der Lage, die entsprechend schnelleren Änderungen der Wassertemperatur in kleineren Einzugsgebieten vorherzusagen. Unsere Ergebnisse zeigen auch, dass es wichtig ist, Zeitinformationen als Input für RNNs zu verwenden. RNNs sind generell in der Lage, die entsprechenden Beziehungen aus den Daten zu erlernen. Zwar gab es keinen signifikanten Unterschied in

der Leistung für die RNNs mit und ohne Zeitinformation, jedoch benötigten RNNs, die mit Zeitinformations-Eingaben optimiert wurden, eine deutlich geringere Anzahl an Zeitschritten für die gleiche Vorhersageleistung. Dies hat den Vorteil, dass es die Rechenzeit verringert und die Anzahl der für das Training verfügbaren Datenpunkte erhöht.

Diese Studie hat einige Einschränkungen: Einerseits sind die ausgewählten Einzugsgebiete alle aus dem mitteleuropäischen Raum und weisen humide Bedingungen auf. Die Anwendung dieser ML-Ansätze könnte in mediterranen Einzugsgebieten oder solchen, die klimatische Bedingungen mit einer höheren Dynamik aufweisen, dazu führen, dass den Eingangsvariablen (z. B. Abfluss in trockenen Klimazonen) eine andere Bedeutung zukommt als in humiden Gebieten, und andere ML-Modelle besser abschneiden als jene in unserer Studie. Durch die Auswahl von Einzugsgebieten mit einer großen Bandbreite an physiographischen Eigenschaften sollte diese potenzielle Verzerrung auf ein Minimum reduziert werden. Außerdem ist die Leistung des air2stream-Vergleichsmodells ähnlich dem Leistungsbereich anderer Studien, was einen Vergleich ermöglicht. Andererseits haben wir alle Modelle nur für einzelne Einzugsgebiete trainiert und nicht versucht, ein globales Modell zu erstellen, das die Wassertemperaturen in mehreren Einzugsgebieten oder sogar in unbeobachteten Einzugsgebieten vorhersagen kann. Obwohl dies ein relevantes Problem ist, hielten wir es für notwendig, eine umfassende Bewertung der verschiedenen Eingangsdaten, Modelltypen und Trainingseigenschaften vorzunehmen, bevor wir all dies in einem Modell zur Vorhersage der Fließgewässertemperatur in mehreren Einzugsgebieten kombinieren.

5 Résumé

Mit den am weitesten verbreiteten Ansätzen zur Fließgewässermodellierung waren wir in der Lage, 10 österreichische Einzugsgebiete mit einem mittleren Test-RMSE von 1,55 °C (lineare Regression) und 0,98 °C (air2stream) zu modellieren. Wir untersuchten 6 ML-Modelle mit unterschiedlichen Eingangsdatenkombinationen und konnten Vorhersagen mit einem mittleren RMSE von 0,55 °C erzielen, was einer Verbesserung von 64% und 43%

entspricht. Von den getesteten ML-Modellen lieferte FNN, das Lufttemperatur, Abfluss und Niederschlag und, falls verfügbar, auch Strahlung als Input verwendet, die besten Ergebnisse. Mit nur 6 Jahren an Trainingsdaten können Modelle mit hoher Güte erstellt werden.

Die Hyperparameter haben einen großen Einfluss auf die Modellgüte von ML-Modellen. Der Einfluss von unterschiedlichen Hyperparametern auf die Variabilität in der Modellgüte ist viel größer als der Einfluss unterschiedlicher Modelltypen oder Eingangsdatenkombinationen. Daher ist die Optimierung der Hyperparameter essenziell für ein gut funktionierendes ML-Modell. In Situationen, in denen die Rechenressourcen begrenzt sind und eine Hyperparameteroptimierung nicht möglich ist, scheint RF eine vernünftige Alternative zu sein, da es die geringste Varianz im Vorhersage-RMSE aufweist.

RNNs erzielen aufgrund der internen Zellzustände und ihrer Fähigkeit, lange Zeitreihen zu verarbeiten, die besten Vorhersagen im größten untersuchten Einzugsgebiet. Dies liegt wahrscheinlich an den längeren Konzentrationszeiten in größeren Einzugsgebieten. Daher sollte die Schätzung der Konzentrationszeiten eines Einzugsgebiets für die adäquate Wahl eines Modelltyps oder relevanter Lags von Variablen in die zukünftige Forschung einbezogen werden. Auch Methoden zur Schätzung der Bedeutung von Variablen können das Verständnis der Wechselwirkungen zwischen Variablen und Modellgüte verbessern und dabei helfen, die relevante Anzahl von Lags zu berücksichtigen. Die Anwendung solcher Methoden wird Teil zukünftiger Forschungstätigkeiten sein.

Die Untersuchungsgebiete wurden zwar so ausgewählt, dass sie ein breites Spektrum an Eigenschaften abdecken, jedoch sind sie alle in Mitteleuropa situiert. Die Untersuchungen sollten daher zukünftig auch auf Gebiete in anderen Klimaregionen ausgeweitet werden. Dies ist besonders wichtig für die Entwicklung von Modellen zur simultanen Vorhersage der Fließgewässertemperaturen in mehreren Einzugsgebieten, was ein wichtiger nächster Schritt und Thema der aktuellen Forschung ist. Die in dieser Studie vorgestellten ML-Modelle werden mit Beobachtungsdaten betrieben und können das System derart repräsentieren, dass sie für die Vorhersage der Fließgewässertemperatur unter wechselnden Bedingun-

gen angewendet werden können. Ihre Anwendung für Kurzzeit- oder Echtzeit-Vorhersageansätze ist daher sehr vielversprechend. Die resultierenden Vorhersageunsicherheiten in solchen Systemen sind hauptsächlich von den Unsicherheiten in den meteorologischen Vorhersagen abhängig. Durch die Bereitstellung aller untersuchten Methoden im open source R-Paket `waterTemp` stellen wir unsere Erkenntnisse für eine reproduzierbare Forschung zur Verfügung und machen sie zugänglich für zukünftige Anwendungen in der Wasserwirtschaft, Wissenschaft und betriebsrelevante Fragestellungen der Industrie.

Förderung Diese Studie wurde vom Österreichischen Wissenschaftsfonds FWF, Projektnummer P 31213, und von den ÖAW-Projekten `Rechout` und `Poco-Flood` finanziert. Die vorgestellten Berechnungen wurden auf dem Vienna Scientific Cluster (VSC) durchgeführt.

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>. ■

Anhang

Hyperparameter-Anwendungsbereich

RF: min.node.size: 2–10, mtry: 3-(Anzahl an Inputs-1)
 XGBoost: nrounds: 300–3000, eta: 0,001–0,3, max_depth: 3–12, min_child_weight: 1–10, subsample: 0,7–1, colsample_bytree: 0,7–1, gamma: 0–5
 FNN: layers: 1–5, units: 5–200, dropout: 0–0,2, batch_size: 5–150, epochs: 100, early_stopping_patience: 5
 RNNs: layers: 1–5, units: 5–300, dropout: 0–0,4, batch_size: 5–150, timesteps: 5–200, epochs: 100, early_stopping_patience: 5

Übersicht der Test-MAE-Ergebnisse

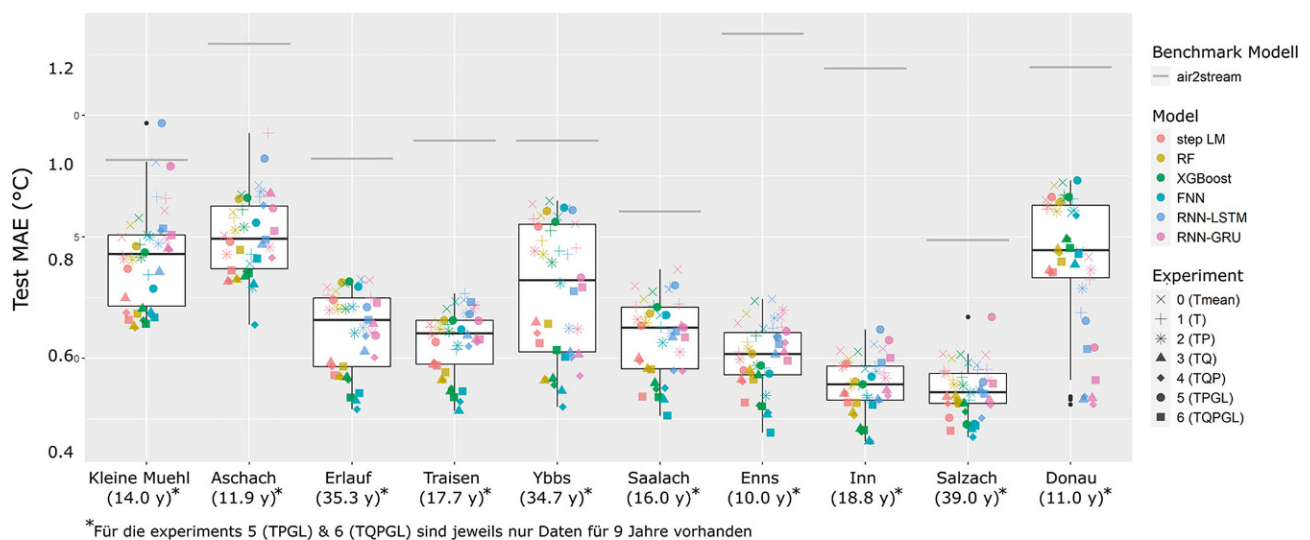


Abb. 7 Boxplots der Modellgüten der Test-MAE in den einzelnen Einzugsgebieten mit überlagerten Punkten, welche die Modellgüte einzelner Kombinationen von Einzugsgebiet, Modellart und Experiment zeigen. Die Einzugsgebiete sind von klein (*links*) nach groß (*rechts*) geordnet und haben zusätzliche Information über die Zeitreihenlänge in den darunter stehenden Klammern. Die air2stream-Ergebnisse des jeweiligen Einzugsgebiets sind als *grauer Strich* dargestellt

Übersicht der Vorhersagen für das Donau-Einzugsgebiet

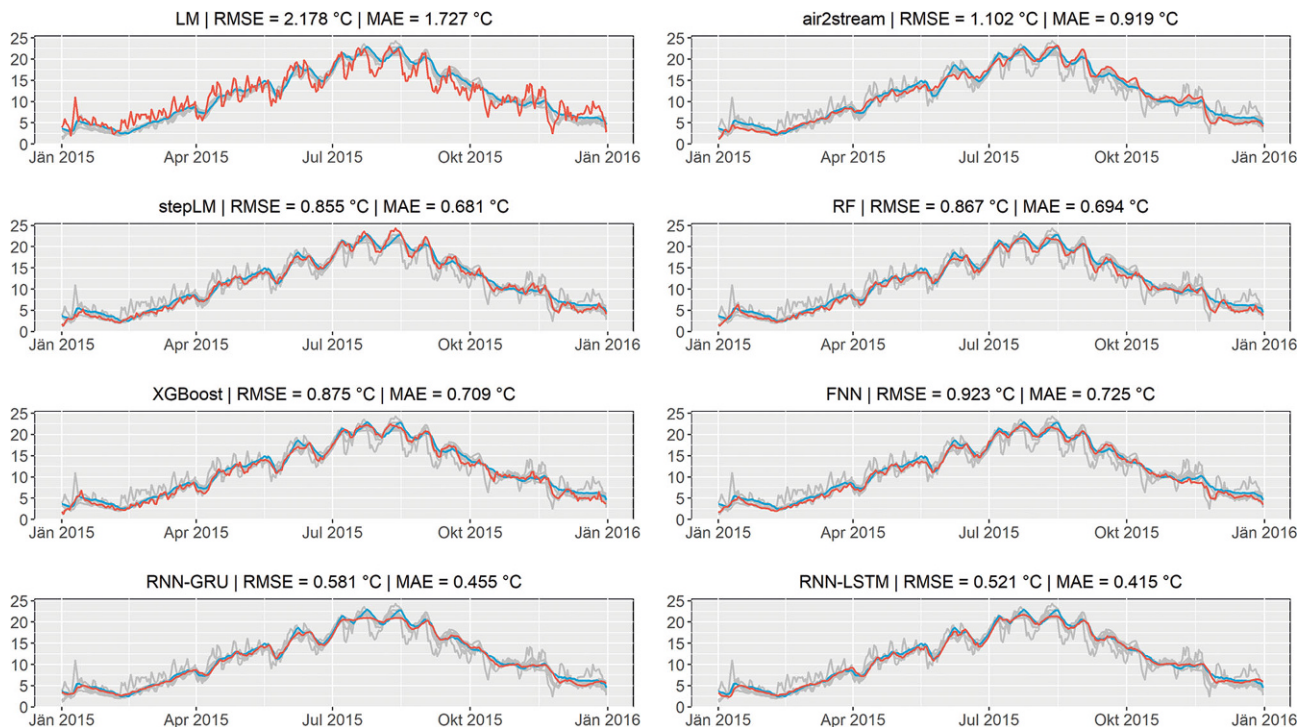


Abb. 8 Vergleich der vorhergesagten Fließgewässertemperaturen aller getesteten Modellarten für das Donau-Einzugsgebiet im Jahr 2015. Die Daten des Jahres 2015 wurden nicht für das Trainieren der Modelle verwendet. Die Vorhersagen sind mit roten Linien dargestellt, die beobachteten Wassertemperaturen mit blauen Linien. Die Vorhersagen aller anderen Modelle werden zum Vergleich jeweils mit grauen Linien dargestellt

Literatur

- Abba, S. I., Hadi, S. J., & Abdullahi, J. (2017): River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Computer Science*, 120, 75–82. <https://doi.org/10.1016/j.procs.2017.11.212>
- Ahmadi-Nedushan, B., St-Hilaire, A., Ouarda, T. B. M. J., Bilodeau, L., Robichaud, É., Thié-monge, N., & Bobée, B. (2007): Predicting river water temperatures using stochastic models: case study of the Moisie River (Québec, Canada). *Hydrological Processes*, 21(1), 21–34. <https://doi.org/10.1002/hyp.6353>
- Akaike H. (1973): Information theory as an extension of the likelihood principle. In: Petrov BN, Csaki F (Eds) Second Akademiai, International Symposium on Information Theory. Kiado, Budapest, 267–281.
- Allaire, J. J., & Tang, Y. (2020): tensorflow: R Interface to "TensorFlow." <https://github.com/rstudio/tensorflow>. Zugegriffen: 23. Apr. 2020
- Arismendi, I., Safeeq, M., Dunham, J. B., & Johnson, S. L. (2014): Can air temperature be used to project influences of climate change on stream temperature? *Environmental Research Letters*, 9(8). <https://doi.org/10.1088/1748-9326/9/8/084015>
- Baldi, P., & Sadowski, P. (2014): The dropout learning algorithm. *Artificial Intelligence*, 210(1), 78–122. <https://doi.org/10.1016/j.artint.2014.02.004>
- Beaufort, A., Moatar, F., Curie, E., Ducharne, A., Bustillo, V., & Thiéry, D. (2016): River Temperature Modelling by Strahler Order at the Regional Scale in the Loire River Basin, France. *River Research and Applications*, 32(4), 597–609. <https://doi.org/10.1002/rra.2888>
- Bélangier, M., El-Jabi, N., Caissie, D., Ashkar, E., & Ribi, J. M. (2005): Water temperature prediction using neural networks and multiple linear regression. *Revue Des Sciences de l'Eau*, 18(3), 403–421. <https://doi.org/10.7202/705565ar>
- Bengio, Y., Courville, A., & Vincent, P. (2013): Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828. <https://doi.org/10.1109/TPAMI.2013.50>
- Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B. M. J., & Bobée, B. (2007): A Review of Statistical Water Temperature Models. *Canadian Water Resources Journal* (Vol. 32, Issue 3, pp. 179–192). <https://doi.org/10.4296/cwrj3203179>
- Boisneau, C., Moatar, F., Bodin, M., & Boisneau, P. (2008): Does global warming impact on migration patterns and recruitment of Allis shad (*Alosa alosa* L.) young of the year in the Loire River, France? In *Fish and Diadromy in Europe* (ecology, management, conservation) (pp. 179–186). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8548-2_14
- Breiman, L. (1996): Bagging predictors. *Machine Learning*, 24(2), 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>
- Brinckmann, S., Krähenmann, S., & Bissolli, P. (2016): High-resolution daily gridded data sets of air temperature and wind speed for Europe. *Earth System Science Data*, 8(2), 491–516. <https://doi.org/10.5194/essd-8-491-2016>
- Caissie, D. (2006): The thermal regime of rivers: A review. In *Freshwater Biology* (Vol. 51, Issue 8, pp. 1389–1406). John Wiley & Sons, Ltd. <https://doi.org/10.1111/j.1365-2427.2006.01597.x>
- Caissie, Daniel, & Luce, C. H. (2017): Quantifying streambed advection and conduction heat fluxes. *Water Resources Research*, 53(2), 1595–1624. <https://doi.org/10.1002/2016WR019813>
- Caldwell, R. J., Gangopadhyay, S., Bountry, J., Lai, Y., & Elsner, M. M. (2013): Statistical modeling of daily and subdaily stream temperatures: Application to the Methow River Basin, Washington. *Water Resources Research*, 49(7), 4346–4361. <https://doi.org/10.1002/wrcr.20353>
- Chen, T., & Guestrin, C. (2016): XGBoost: A scalable tree boosting system. *Proceedings of*

- the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13–17-Aug(8), 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. (2020): xgboost: Extreme Gradient Boosting. <https://cran.r-project.org/package=xgboost>. Zugegriffen: 23. Apr. 2020
- Chenard, J.-F., & Caissie, D. (2008): Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. *Hydrological Processes*, 22(17), 3361–3372. <https://doi.org/10.1002/hyp.6928>
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014): Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP 2014–2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 1724–1734. <https://doi.org/10.3115/v1/d14-1179>
- Claesen, M., & De Moor, B. (2015): Hyperparameter Search in Machine Learning.
- Crank, J., & Nicolson, P. (1947): A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43(1), 50–67. <https://doi.org/10.1017/S0305004100023197>
- Crisp, D. T., & Howson, G. (1982): Effect of air temperature upon mean water temperature in streams in the north Pennines and English Lake District. *Freshwater Biology*, 12(4), 359–367. <https://doi.org/10.1111/j.1365-2427.1982.tb00629.x>
- Dallas, H. (2008): Water temperature and riverine ecosystems: An overview of knowledge and approaches for assessing biotic responses, with special reference to South Africa. In *Water SA* (Vol. 34, Issue 3, pp. 393–404). South African Water Research Commission. <https://doi.org/10.4314/wsa.v34i3.180634>
- DeWeber, J. T., & Wagner, T. (2014): A regional neural network ensemble for predicting mean daily river water temperature. *Journal of Hydrology*, 517, 187–200. <https://doi.org/10.1016/j.jhydrol.2014.05.035>
- Dugdale, S. J., Hannah, D. M., & Malcolm, I.A. (2017): River temperature modelling: A review of process-based approaches and future directions. *Earth-Science Reviews*, 175, 97–113. <https://doi.org/10.1016/j.earscirev.2017.10.009>
- Dunn, O. J. (1964): Multiple Comparisons Using Rank Sums. *Technometrics*, 6(3), 241–252. <https://doi.org/10.1080/00401706.1964.10490181>
- Feigl, M., Lebedzinski, K., Herrnegger, M., & Schulz, K. (2021): Machine learning methods for stream water temperature prediction. *Hydrology and Earth System Sciences Discussions*, 2021, 1–35. <https://doi.org/10.5194/hess-2020-670>
- Freund, Y., & Schapire, R. E. (1995): A decision-theoretic generalization of on-line learning and an application to boosting. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 904, 23–37. https://doi.org/10.1007/3-540-59119-2_166
- Friberg, N., Dybkjær, J. B., Olafsson, J. S., Gislason, G. M., Larsen, S. E., & Lauridsen, T. L. (2009): Relationships between structure and function in streams contrasting in temperature. *Freshwater Biology*, 54(10), 2051–2068. <https://doi.org/10.1111/j.1365-2427.2009.02234.x>
- Friedman, J. H. (2001): Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Friedman, J. H. (2002): Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Graf, R., Zhu, S., & Sivakumar, B. (2019): Forecasting river water temperature time series using a wavelet-neural network hybrid modelling approach. *Journal of Hydrology*, 578(August), 124115. <https://doi.org/10.1016/j.jhydrol.2019.124115>
- Hadzima-Nyarko, M., Rabi, A., & Šperac, M. (2014): Implementation of Artificial Neural Networks in Modeling the Water-Air Temperature Relationship of the River Drava. *Water Resources Management*, 28(5), 1379–1394. <https://doi.org/10.1007/s11269-014-0557-7>
- 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>
- Haiden, T., Kann, A., & Pistotnik, G. (2014): Nowcasting with INCA During SNOW-V10. *Pure and Applied Geophysics*, 171(1–2), 231–242. <https://doi.org/10.1007/s00024-012-0547-8>
- Hannah, D. M., & Garner, G. (2015): River water temperature in the United Kingdom. *Progress in Physical Geography: Earth and Environment*, 39(1), 68–92. <https://doi.org/10.1177/0309133314550669>
- Hannah, D. M., Webb, B. W., & Nobilis, F. (2008): River and stream temperature: dynamics, processes, models and implications. *Hydrological Processes*, 22(7), 899–901. <https://doi.org/10.1002/hyp.6997>
- Hansen, L. K., & Salamon, P. (1990): Neural Network Ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993–1001. <https://doi.org/10.1109/34.58871>
- He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., & Li, X. (2020): The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data*, 7(1), 25. <https://doi.org/10.1038/s41597-020-0369-y>
- Heddiam, S., Ptak, M., & Zhu, S. (2020): Modelling of daily lake surface water temperature from air temperature: Extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. *Journal of Hydrology*, 588, 125130. <https://doi.org/10.1016/j.jhydrol.2020.125130>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, E., Villaume, S., & Thepaut, J.-N. (2020): The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Hiebl, J., & Frei, C. (2016): Daily temperature grids for Austria since 1961—concept, creation and applicability. *Theoretical and Applied Climatology*, 124(1–2), 161–178. <https://doi.org/10.1007/s00704-015-1411-4>
- Hiebl, J., & Frei, C. (2018): Daily precipitation grids for Austria since 1961—development and evaluation of a spatial dataset for hydroclimatic monitoring and modelling. *Theoretical and Applied Climatology*, 132(1–2), 327–345. <https://doi.org/10.1007/s00704-017-2093-x>
- Hinton, G. E. (2012): A Practical Guide to Training Restricted Boltzmann Machines (pp. 599–619). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_32
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012): Improving neural networks by preventing co-adaptation of feature detectors.
- Hochreiter, S., & Schmidhuber, J. (1997): Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003): A Practical Guide to Support Vector Classification.
- Imholt, C., Gibbins, C. N., Malcolm, I.A., Langan, S., & Soulsby, C. (2010): Influence of riparian cover on stream temperatures and the growth of the mayfly *Baetis rhodani* in an upland stream. *Aquatic Ecology*, 44(4), 669–678. <https://doi.org/10.1007/s10452-009-9305-0>
- Jackson, F. L., Fryer, R. J., Hannah, D. M., Millar, C. P., & Malcolm, I.A. (2018): A spatio-temporal statistical model of maximum daily river temperatures to inform the management of Scotland's Atlantic salmon rivers under climate change. *Science of the Total Environment*, 612, 1543–1558. <https://doi.org/10.1016/j.scitotenv.2017.09.010>
- Johnson, M. E., Wilby, R. L., & Toone, J. A. (2014): Inferring air-water temperature relationships from river and catchment properties. *Hydrological Processes*, 28(6), 2912–2928. <https://doi.org/10.1002/hyp.9842>
- Jones, D. R., Schonlau, M., & Welch, W. J. (1998): Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4), 455–492. <https://doi.org/10.1023/A:1008306431147>
- Joslyn, K. (2018): Water Quality Factor Prediction Using Supervised Machine Learning Water Quality Factor Prediction Using Supervised Machine Learning. REU Final Reports Research Experiences for Undergraduates on Computational Modeling Serving the City.
- Kennedy, J., & Eberhart, R. (1995): Particle swarm optimization. Proceedings of ICNN'95 - International Conference on Neural Networks, 4, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017): Self-normalizing neural networks. Advances in Neural Information Processing Systems, 2017-December, 972–981. arXiv:1706.02515v5. Zugegriffen: 23. Okt. 2019
- Kling, H., Stanzel, P., Fuchs, M., & Nachtebel, H.-P. (2015): Performance of the COSERO precipitation-runoff model under non-stationary conditions in basins with different climates. *Hydrological Sciences Journal*, 60(7–8), 1374–1393. <https://doi.org/10.1080/02626667.2014.959956>
- Klingler, C., Schulz, K., & Herrnegger, M. (2021a): LamaH | \textit{La}rge-Sa\textit{m}ple D\textit{a}ta for \textit{H}ydrology and Environmental Sciences for Central Europe. Earth System Science Data Discussions, 2021, 1–46. <https://doi.org/10.5194/essd-2021-72>
- Klingler, C., Schulz, K., & Herrnegger, M. (2021b): LamaH | Large-Sample Data for Hydrology: Big data für die Hydrologie und Umweltwissenschaften. Österreichische Wasser- und Abfallwirtschaft 73(7–8), in diesem Heft
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018): Rainfall—runoff modelling using Long Short-Term Memory (LSTM) networks. 6005–6022.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019): Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine-Learning Applied to Large-Sample Datasets.
- Kruskal, W. H., & Wallis, W. A. (1952): Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260), 583–621. <https://doi.org/10.1080/01621459.1952.10483441>

- Kuhn, M. (2020):** caret: Classification and Regression Training.
- Kushner, H. J. (1964):** A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Fluids Engineering, Transactions of the ASME*, 86(1), 97–106. <https://doi.org/10.1115/1.3653121>
- Kędra, M. (2020):** Regional Response to Global Warming: Water Temperature Trends in Semi-Natural Mountain River Systems. *Water*, 12(1), 283. <https://doi.org/10.3390/w12010283>
- Li, H., Deng, X., Kim, D.-Y., & Smith, E. P. (2014):** Modeling maximum daily temperature using a varying coefficient regression model. *Water Resources Research*, 50(4), 3073–3087. <https://doi.org/10.1002/2013WR012423>
- Li, W., Kiaghadi, A., & Dawson, C. (2020):** High temporal resolution rainfall-runoff modeling using long-short-term-memory (LSTM) networks. *Neural Computing and Applications*, 1–18. <https://doi.org/10.1007/s00521-020-05010-6>
- Lu, H., & Ma, X. (2020):** Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 249, 126169. <https://doi.org/10.1016/j.chemosphere.2020.126169>
- Mackey, A. P., & Berrie, A. D. (1991):** The prediction of water temperatures in chalk streams from air temperatures. *Hydrobiologia*, 210(3), 183–189. <https://doi.org/10.1007/BF00034676>
- McGlynn, B. L., McDonnell, J. J., Seibert, J., & Kendall, C. (2004):** Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations. *Water Resources Research*, 40(7). <https://doi.org/10.1029/2003WR002494>
- McKenna, J. E., Butryn, R. S., & McDonald, R. P. (2010):** Summer Stream Water Temperature Models for Great Lakes Streams: New York. *Transactions of the American Fisheries Society*, 139(5), 1399–1414. <https://doi.org/10.1577/t09-153.1>
- Mohseni, O., & Stefan, H. G. (1999):** Stream temperature/air temperature relationship: A physical interpretation. *Journal of Hydrology*, 218(3–4), 128–141. [https://doi.org/10.1016/S0022-1694\(99\)00034-7](https://doi.org/10.1016/S0022-1694(99)00034-7)
- Mohseni, O., Erickson, T. R., & Stefan, H. G. (1999):** Sensitivity of stream temperatures in the United States to air temperatures projected under a global warming scenario. *Water Resources Research*, 35, 3723–3733.
- Moćkus, J. (1975):** On Bayesian Methods for Seeking the Extremum. In *Optimization Techniques IFIP Technical Conference* (pp. 400–404). https://doi.org/10.1007/978-3-662-38527-2_55
- Moćkus, J. (1989):** Bayesian Approach to Global Optimization (Vol. 37). Springer Netherlands. <https://doi.org/10.1007/978-94-009-0909-0>
- Moćkus, 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
- Naresh, A., & Rehana, S. (2017):** Modeling Stream Water Temperature using Regression Analysis with Air Temperature and Streamflow over Krishna River. *Rehana International Journal of Engineering Technology Science and Research*, 4(11):2394–3386.
- Neumann, D. W., Rajagopalan, B., & Zagona, E. A. (2003):** Regression model for daily maximum stream temperature. *Journal of Environmental Engineering*, 129(7), 667–674. [https://doi.org/10.1061/\(ASCE\)0733-9372\(2003\)129:7\(667\)](https://doi.org/10.1061/(ASCE)0733-9372(2003)129:7(667))
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Cournapeau, D., Passos, A., Brucher, M., Perrot, A., & Duchesnay, E. (2011):** Scikit-learn: Machine Learning in Python. In *Machine Learning in Python*. *Journal of Machine Learning Research* (Vol. 12). Microtome Publishing. <https://hal.inria.fr/hal-00650905v2>. Zugegriffen: 4. Dez. 2020
- Piccolroaz, S., Calamita, E., Majone, B., Gallice, A., Siviglia, A., & Toffolon, M. (2016):** Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. *Hydrological Processes*, 30(21), 3901–3917. <https://doi.org/10.1002/hyp.10913>
- Piotrowski, A. P., & Napiorkowski, J. J. (2018):** Performance of the air2stream model that relates air and stream water temperatures depends on the calibration method. *Journal of Hydrology*, 561, 395–412. <https://doi.org/10.1016/j.jhydrol.2018.04.016>
- Piotrowski, A. P., & Napiorkowski, J. J. (2019):** Simple modifications of the nonlinear regression stream temperature model for daily data. *Journal of Hydrology*, 572, 308–328. <https://doi.org/10.1016/j.jhydrol.2019.02.035>
- Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., & Osuch, M. (2015):** Comparing various artificial neural network types for water temperature prediction in rivers. *Journal of Hydrology*, 529(P1), 302–315. <https://doi.org/10.1016/j.jhydrol.2015.07.044>
- Piotrowski, A. P., Napiorkowski, J. J., & Piotrowska, A. E. (2020):** Impact of deep learning-based dropout on shallow neural networks applied to stream temperature modelling. In *Earth-Science Reviews* (Vol. 201, p. 103076). Elsevier B.V. <https://doi.org/10.1016/j.earscirev.2019.103076>
- R Core Team. (2020):** R: A Language and Environment for Statistical Computing. <https://www.r-project.org/>. Zugegriffen: 23. Apr. 2020
- Rabi, A., Hadzima-Nyarko, M., & Šperac, M. (2015):** Modelling river temperature from air temperature: case of the River Drava (Croatia). *Hydrological Sciences Journal*, 60(9), 1490–1507. <https://doi.org/10.1080/02626667.2014.914215>
- Razafimaharo, C., Krähennann, S., Höpp, S., Rauthe, M., & Deutschländer, T. (2020):** New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS). *Theoretical and Applied Climatology*, 142(3–4), 1531–1553. <https://doi.org/10.1007/s00704-020-03388-w>
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019):** Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Risley, J. C., Roehl Jr., E. A., & Conrads, P. A. (2003):** Estimating Water Temperatures in Small Streams in Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models. <https://doi.org/10.3133/wri024218>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986):** Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Sahoo, G. B., Schladow, S. G., & Reuter, J. E. (2009):** Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. *Journal of Hydrology*, 378(3–4), 325–342. <https://doi.org/10.1016/j.jhydrol.2009.09.037>
- Sand-Jensen, K., & Pedersen, N. L. (2005):** Differences in temperature, organic carbon and oxygen consumption among lowland streams. *Freshwater Biology*, 50(12), 1927–1937. <https://doi.org/10.1111/j.1365-2427.2005.01436.x>
- Schapire, R. E. (1990):** The Strength of Weak Learnability. *Machine Learning*, 5(2), 197–227. <https://doi.org/10.1023/A:1022648800760>
- Segura, C., Caldwell, P., Sun, G., McNulty, S., & Zhang, Y. (2015):** A model to predict stream water temperature across the conterminous USA. *Hydrological Processes*, 29(9), 2178–2195. <https://doi.org/10.1002/hyp.10357>
- Shank, D. B., Hoogenboom, G., & McClen-don, R. W. (2008):** Dewpoint temperature prediction using artificial neural networks. *Journal of Applied Meteorology and Climatology*, 47(6), 1757–1769. <https://doi.org/10.1175/2007JAMC1693.1>
- Smith, K. (1981):** The prediction of river water temperatures. *Hydrological Sciences Bulletin*, 26(1), 19–32. <https://doi.org/10.1080/02626668.109490859>
- Snoek, J., Larochelle, H., & Adams, R. P. (2012):** Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 4, 2951–2959.
- Sohrabi, M. M., Benjankar, R., Tonina, D., Wenger, S. J., & Isaak, D. J. (2017):** Estimation of daily stream water temperatures with a Bayesian regression approach. *Hydrological Processes*, 31(9), 1719–1733. <https://doi.org/10.1002/hyp.11139>
- Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. (2009):** Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. *IEEE Transactions on Information Theory*, 58(5), 3250–3265. <https://doi.org/10.1109/TIT.2011.2182033>
- Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014):** Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958. <http://jmlr.org/papers/v15/srivastava14a.html>
- Stajkowski, S., Kumar, D., Samui, P., Bonakdari, H., & Gharabaghi, B. (2020):** Genetic-algorithm-optimized sequential model for water temperature prediction. *Sustainability* (Switzerland), 12(13). <https://doi.org/10.3390/su12135374>
- Stefan, H. G., & Preud'homme, E. B. (1993):** Stream temperature estimation from air temperature. *JAWRA Journal of the American Water Resources Association*, 29(1), 27–45. <https://doi.org/10.1111/j.1752-1688.1993.tb01502.x>
- Stevens, H. H., Ficke, J. E., & Smoot, G. F. (1975):** Techniques of water-resources investigations of the u.s. geological survey. U.S. Government Printing Office, 65.
- Tavares, M. H., Cunha, A. H. F., Motta-Marques, D., Ruhoff, A. L., Fragos, C. R., Munar, A. M., & Bonnet, M. P. (2020):** Derivation of consistent, continuous daily river temperature data series by combining remote sensing and water temperature models. *Remote Sensing of Environment*, 241, 111721. <https://doi.org/10.1016/j.rse.2020.111721>
- Temizyurek, M., & Dadaser-Celik, F. (2018):** Modelling the effects of meteorological parameters on water temperature using artificial neural networks. *Water Science and Technology*, 77(6), 1724–1733. <https://doi.org/10.2166/wst.2018.058>
- Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y., Devarakonda, R., & Cook, R. B. (2014):** Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. Data set. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA.
- Toffolon, M., & Piccolroaz, S. (2015):** A hybrid model for river water temperature as a function of air temperature and discharge. *Environmental Research Letters*, 10(11), 1–10. <https://doi.org/10.1088/1748-9326/10/11/114011>
- Trinh, N. X., Trinh, T. Q., Phan, T. P., Thanh, T. N., & Thanh, B. N. (2019):** Water Temperature Prediction Models in Northern Coastal Area, Vietnam. *Asian Review of Environmental and Earth Sciences*, 6(1), 1–8. <https://doi.org/10.20448/journal.506.2019.61.1.8>
- Van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, E., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2013):** Global river discharge and

- water temperature under climate change. *Global Environmental Change*, 23(2), 450–464. <https://doi.org/10.1016/j.gloenvcha.2012.11.002>
- Webb, B. W., & Zhang, Y. (1997):** Spatial and seasonal variability in the components of the river heat budget. *Hydrological Processes*, 11(1), 79–101. [https://doi.org/10.1002/\(sici\)1099-1085\(199701\)11:1<79::aid-hyp404>3.0.co;2-n](https://doi.org/10.1002/(sici)1099-1085(199701)11:1<79::aid-hyp404>3.0.co;2-n)
- Webb, B. W., Clack, P. D., & Walling, D. E. (2003):** Water-air temperature relationships in a Devon river system and the role of flow. *Hydrological Processes*, 17(15), 3069–3084. <https://doi.org/10.1002/hyp.1280>
- Wehrly, K. E., Brenden, T. O., & Wang, L. (2009):** A Comparison of Statistical Approaches for Predicting Stream Temperatures Across Heterogeneous Landscapes. *JAWRA Journal of the American Water Resources Association*, 45(4), 986–997. <https://doi.org/10.1111/j.1752-1688.2009.00341.x>
- Wenger, S. J., Isaak, D. J., Dunham, J. B., Fausch, K. D., Luce, C. H., Neville, H. M., Rieman, B. E., Young, M. K., Nagel, D. E., Horan, D. L., & Chandler, G. L. (2011):** Role of climate and invasive species in structuring trout distributions in the interior Columbia River Basin, USA. *Canadian Journal of Fisheries and Aquatic Sciences*, 68(6), 988–1008. <https://doi.org/10.1139/f2011-034>
- Werner, A. T., Schnorbus, M.A., Shrestha, R. R., Cannon, A. J., Zwierns, F. W., Dayon, G., & Anslow, F. (2019):** A long-term, temporally consistent, gridded daily meteorological dataset for northwestern North America. *Scientific Data*, 6(1), 1–16. <https://doi.org/10.1038/sdata.2018.299>
- Westenbroek, S., Stewart, J. S., Buchwald, C. A., Mitro, M., Lyons, J. D., & Greb, S. (2010):** A Model for Evaluating Stream Temperature Response to Climate Change Scenarios in Wisconsin. 1–12. [https://doi.org/10.1061/41143\(394\)1](https://doi.org/10.1061/41143(394)1)
- White, B. W., & Rosenblatt, F. (1963):** Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. *The American Journal of Psychology*, 76(4), 705. <https://doi.org/10.2307/1419730>
- Wickham, H. (2016):** ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Xiang, Z., Yan, J., & Demir, I. (2020):** A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. *Water Resources Research*, 56(1). <https://doi.org/10.1029/2019WR025326>
- Yang, D., & Peterson, A. (2017):** River water temperature in relation to local air temperature in the Mackenzie and Yukon basins. *Arctic*, 70(1), 47–58. <https://doi.org/10.14430/arctic4627>
- 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>
- Zhu, S., & Piotrowski, A. P. (2020):** River/stream water temperature forecasting using artificial intelligence models: a systematic review. In *Acta Geophysica* (Vol. 68, Issue 5, pp. 1433–1442). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s11600-020-00480-7>
- Zhu, S., Nyarko, E. K., & Hadzima-Nyarko, M. (2018):** Modelling daily water temperature from air temperature for the Missouri River. *PeerJ*, 2018(6), e4894. <https://doi.org/10.7717/peerj.4894>
- Zhu, S., Heddam, S., Nyarko, E. K., Hadzima-Nyarko, M., Piccolroaz, S., & Wu, S. (2019a):** Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. *Environmental Science and Pollution Research*, 26(1), 402–420. <https://doi.org/10.1007/s11356-018-3650-2>
- Zhu, S., Heddam, S., Wu, S., Dai, J., & Jia, B. (2019b):** Extreme learning machine-based prediction of daily water temperature for rivers. *Environmental Earth Sciences*, 78(6), 1–17. <https://doi.org/10.1007/s12665-019-8202-7>
- Zhu, S., Nyarko, E. K., Hadzima-Nyarko, M., Heddam, S., & Wu, S. (2019c):** Assessing the performance of a suite of machine learning models for daily river water temperature prediction. *PeerJ*, 7, e7065. <https://doi.org/10.7717/peerj.7065>
- Álvarez, D., & Nicieza, A. G. (2005):** Compensatory response “defends” energy levels but not growth trajectories in brown trout, *Salmo trutta* L. *Proceedings of the Royal Society B: Biological Sciences*, 272(1563), 601–607. <https://doi.org/10.1098/rspb.2004.2991>

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