Niederschlags-Abfluss-Modellierung mit Long Short-Term Memory (LSTM)

Frederik Kratzert · Martin Gauch · Grey Nearing · Sepp Hochreiter · Daniel Klotz

Angenommen: 13. April 2021 / Online publiziert: 17. Mai 2021
 © Der/die Autor(en) 2021

Zusammenfassung Methoden der künstlichen Intelligenz haben sich in den letzten Jahren zu essenziellen Bestandteilen fast aller Bereiche von Wissenschaft und Technik entwickelt. Dies gilt auch für die Hydrologie: Vielschichtige neuronale Netzwerke – auch bekannt als Modelle des Deep Learning – ermöglichen hier Vorhersagen von Niederschlagsabflussmengen in zuvor unerreichter Präzision.

Dieser Beitrag beleuchtet das Potenzial von Deep Learning für wasserwirtschaftliche Anwendungen. Der erste Teil des Artikels zeigt, wie sogenannte Long Short-Term Memory-Netzwerke – eine spezifisch für Zeitreihen entwickelte Methode des Deep Learnings – für die Niederschlags-Abfluss-Modellierung verwendet werden, und wie diese für eine Reihe hydrologischer Probleme bessere Ergebnisse als jedes andere bekannte hydrologische Modell erzielen. Der zweite Teil demonstriert wesentliche Eigenschaften der Long Short-Term Memory-Netzwerke. Zum einen zeigen wir, dass diese Netzwerke beliebige Daten verarbeiten können. Dies erlaubt es, mögliche synergetische Effekte aus unterschiedlichen meteorologischen Datensätzen zu extrahieren und damit die Modellgüte zu verbessern. Zum anderen stellen wir dar, wie relevante hydrologische Prozesse (wie z. B. das Akkumulieren und Schmelzen von Schnee) innerhalb der Modelle abgebildet werden, ohne dass diese spezifisch darauf trainiert wurden.

Schlüsselwörter Deep Learning · Niederschlags-Abfluss-Modellierung · LSTM

Rainfall-Runoff modeling with Long Short-Term Memory Networks (LSTM) – an overview

Abstract In recent years, methods of artificial intelligence have become essential components of almost all branches of science and technology. This also holds for the area of hydrology: multi-layered neural networks—also known as Deep Learning models—allow for rainfall-runoff predictions at unprecedented accuracy.

This contribution highlights the potential of Deep Learning for applications of water management. The first part of this article shows how so-called Long Short-Term Memory networks—a Deep Learning architecture explicitly designed for time series prediction—can be used for rainfall-runoff modeling, and how this approach yields better results than any known hydrologic model on a number of hydrologic problems. The second part of this article outlines important properties of Long Short-Term Memory networks. We show that these models can process any available data. This allows them to draw synergetic information from multiple meteorological datasets, which improves predictions. Further, we demonstrate how these models learn relevant hydrologic processes (e.g., accumulation and melting of snow) without specifically being trained to do so.

Keywords Deep Learning · Rainfall-Runoff modeling · LSTM

1 Einleitung

Deep Learning (DL) ist das Herzstück der modernen Künstlichen Intelligenz (e.g. Hinton et al. 2012; Sutskever et al. 2014; Krizhevsky et al. 2017). Bei DL handelt es sich um eine Reihe von Methoden, die das Lösen von Proble-

men mittels vielschichtiger künstlicher neuronaler Netze ermöglicht. Dieser Beitrag veranschaulicht das Potenzial von DL für die wasserwirtschaftliche Anwendung. Der Fokus liegt dabei auf konkreten Problemstellungen aus der Niederschlags-Abfluss-Modellierung, aus drei Themengebieten: 1. Regionale Modellierung, 2. Vorhersage in unbeobachteten Einzugsgebieten und 3. Modelleigenschaften.

1.1 Regionale Modellierung

Die regionale Skala ist typisch für die hydrologische Modellbildung. Sie ist größer als ein einzelner Hang oder Flussabschnitt, jedoch kleiner als ein Kontinent (e.g. Blöschl und Sivapalan 1995). Die Herausforderung für Modellierer besteht darin, aus dem vorhandenen Prozessverständnis und den verfügbaren Daten (etwa Pegel-, Grundwasser- und/oder Schneehöhenmessungen) ein Modell zu erarbeiten. Dabei kommt es zu einem Kompromiss zwischen hydrologisch konsistenten Parametern und Performance (Klotz et al. 2017; Feigl et al. 2020; Mizukami et al. 2017). Den Schwerpunkt auf nur eines der beiden Kriterien zu legen, ist kaum machbar. Obwohl die Parameter hydrologischer Modelle oft eine physikalische Bedeutung haben (z.B. Infiltrationskoeffizient), ist es in der Praxis nicht möglich, die Modellparameter aus der Theorie abzuleiten. Ein wesentlicher Grund dafür ist, dass unsere Prozessbeschreibung auf einer anderen Skala konzeptualisiert wurde (z.B. im Labor), als sie für die hydrologische Modellierung benötigt wird. Die Ansätze lassen sich nicht ohne weiteres auf Einzugsgebiets-Skala übertragen (Klemeš 1983). Das rein empirische Bestimmen der Parameter gestaltet sich ähnlich problematisch. Konventionelle Modelle (etwa physikalisch basierte oder konzeptionelle Ansätze) wurden entwickelt, um hydrologische Prozesse abzubilden und sind nicht unbedingt dafür vorgesehen, eine möglichst ein-

F. Kratzert (✉) · M. Gauch · S. Hochreiter · D. Klotz
 Institute for Machine Learning,
 Johannes Kepler Universität Linz,
 Altenberger Straße 69, 4040 Linz,
 Österreich
kratzert@ml.jku.at

G. Nearing
 Google Research, Mountain View, CA,
 USA

Abb. 1 Beispiel für die Verschlechterung der Modellgüte durch regionales Modellieren. Zu sehen sind die empirischen kumulativen Verteilungsfunktionen der Nash-Sutcliffe Effizienz (NSE) von zwei hydrologischen Modellen. VIC (Liang et al. 1994) in Blau und mHM (Samaniego et al. 2010; Mizukami et al. 2017) in Rot. Die unterbrochenen Linien zeigen die Performance einzugsgebietsspezifisch kalibrierter Modelle, die durchgehenden Linien jene regional kalibrierter Modelle. Je weiter rechts die Dichtekurve, desto besser

fache Parametrisierung zu ermöglichen (Klemeš 1997; Kirchner 2006). Für die regionale Modellierung ist es außerdem schwierig, eine allgemeingültige Abbildung von Einzugsgebieteigenschaften zu Modellparametern zu finden, die zu ähnlich guten Vorhersagen führt wie einzugsgebietsspezifisch kalibrierte Modelle (Hrachowitz et al., 2013; Mizukami et al. 2017; Blöschl et al., 2019).

DL ermöglicht einen fundamental anderen Ansatz der Modellbildung, der nicht auf vordefinierten Regeln basiert, sondern das gesamte Prozessverständnis aus den verfügbaren Daten lernen muss. Der von uns entwickelte Ansatz führt zu Modellen, deren Qualität tendenziell mit der Zahl der abgedeckten Einzugsgebiete steigt (siehe Kratzert et al. 2018b, 2019c; Gauch et al. 2021). Konventionelle hydrologische Modellansätze erreichen deutlich bessere Modellgüte, wenn sie für jedes Einzugsgebiet spezifisch kalibriert werden statt für viele (diverse) Einzugsgebiete auf einmal (siehe Abb. 1). Der Nutzen von strukturell explizit dargestellten Systemprozessen ist hier limitierend. DL-Modelle haben den Vorteil, dass sie alle Prozesse modellieren können, die aus den verfügbaren Daten ableitbar sind: Je größer und diverser die verfügbare Datenmenge ist, desto allgemeiner ist das gelernte Prozessverständnis (Krat-

zert et al. 2019c; Gauch et al. 2021; siehe auch Halevy et al. 2009).

1.2 Vorhersage in unbeobachteten Einzugsgebieten

Zum Modellieren eines Einzugsgebiets werden in der Regel meteorologische Inputs (etwa Niederschlag und Temperatur) und Daten bezüglich der vorhandenen Einzugsgebieteigenschaften und Abflussmengen benötigt. Vor allem Abflussmessungen sind jedoch nicht für alle Einzugsgebiete vorhanden.

Die Vorhersage in unbeobachteten Einzugsgebieten gilt als allgemein sehr schwierig (Sivapalan 2003; Parajka et al. 2013). Hydrologische Prozesse sind sowohl räumlich als auch zeitlich heterogen und auf verschiedenen relevanten Skalen nicht vollständig verstanden. Im Allgemeinen fehlen hoch aufgelöste Informationen über den Aufbau des Untergrunds, welche eine genaue physikalische Modellierung ermöglichen würden. In der Tat gestaltet sich das Problem derart kompliziert, dass es oft als eines der zentralen Probleme der hydrologischen Modellierung angesehen wird (Blöschl et al., 2013; Blöschl et al. 2019). Die International Association of Hydrological Sciences (IAHS) hat der Lösung dieser Fragestellung (Vorhersage in unbeobachteten Einzugsgebieten) eine ganze Dekade (2003–2012) gewid-

met, trotzdem gilt sie bis heute als weitgehend ungelöst (Hrachowitz et al. 2013).

DL-Modelle sind auch hier sinnvoll nutzbar, da sie kein explizites Wissen über die räumlichen Zusammenhänge voraussetzen. Stattdessen werden räumlich übertragbare Repräsentationen gelernt. Kap. 3 demonstriert diese Vorteile. Unter anderem zeigt sich, dass DL-Ansätze bessere Modellgüte in „unbeobachteten“ Einzugsgebieten aufweisen als einzugsgebietsspezifisch kalibrierte, klassische Niederschlags-Abfluss-Modelle.

1.3 Modelleigenschaften

Die von uns vorgestellten DL-Ansätze besitzen zusätzlich inhärente Eigenschaften, welche für die hydrologische Modellierung von Interesse sind. Wir demonstrieren zwei Formen dieser Modelleigenschaften: Die erste Eigenschaft bezieht sich auf das Einbeziehen von unterschiedlichen Eingabedaten. Wir zeigen, wie DL-Modelle Eingabedaten verschiedener meteorologischer Produkte (DayMet, Maurer und NLDAS) verwenden und von den daraus resultierenden Synergieeffekten profitieren können (siehe z. B. Kratzert et al. 2020). Als zweite Eigenschaft diskutieren wir die Speicher der vorgestellten Modellklasse. Konkret demonstrieren wir, wie Informationen aus den „black box“-Modellen gewonnen werden können, indem deren interne Repräsentationen bezüglich Prozesseigenschaften analysiert werden (Kratzert et al., 2019a).

Im Folgenden geben wir zunächst einen Überblick über das von uns verwendete DL-Modell sowie von uns genutzte Datensätze (Kap. 2). Anschließend gehen wir im Detail auf die drei Kernthemen dieser Publikation ein: Regionale Modellierung (Abschn. 3.1), Vorhersage in unbeobachteten Einzugsgebieten (Abschn. 3.2) sowie Modelleigenschaften (Abschn. 3.3). Dem folgt eine Übersicht an Erweiterungen von DL-basierten Niederschlags-Abfluss-Modellen (Kap. 4.). Abschließend fassen wir die Ergebnisse noch einmal zusammen und geben einen Ausblick auf zukünftige Forschungsthemen (Kap. 5) die auch im Rahmen hydrologischer und wasserwirtschaftlicher Fragestellungen sinnvolle Unterstützung bieten werden.

Abb. 2 Setup-Schema für die Verwendung des LSTM. Dem LSTM (blau) wird an jedem Zeitschritt ein Eingabevektor (lila) bereitgestellt. Es besitzt interne Speicher, die von Zeitschritt zu Zeitschritt aktualisiert werden können. Aus der LSTM-Ausgabe des letzten Zeitschritts (gelb) wird die Vorhersage ermittelt

2 Methoden

2.1 Long Short-Term Memory-Netzwerk

Die Grundlage unserer DL-Modelle bildet das Long Short-Term Memory-Netzwerk (LSTM; Hochreiter 1991; Hochreiter und Schmidhuber 1997; Gers, Schmidhuber & Cummins 2000). Das LSTM gehört zur Familie der rekurrenten neuronalen Netze. Dies sind neuronale Netze, die Eingabedaten in sequenzieller Reihenfolge verarbeiten (siehe Abb. 2). Eine spezielle Eigenschaft von LSTMs ist, dass sie dedizierte interne Speicher besitzen, um Informationen für lange Zeit speichern zu können. Zusätzlich verfügen LSTMs über eine Reihe von sogenannten Gates. Diese kontrollieren in jedem Zeitschritt (a) welche Informationen aus dem Speicher gelöscht werden, (b) was für neue Informationen aus den Eingabedaten in den internen Speicher hinzugefügt werden, und (c) aus welchen Informationen des aktuellen Speichers die Vorhersage gewonnen werden kann. Außerdem, und dies macht LSTMs speziell interessant für den wasserwirtschaftlichen Einsatz, besitzen sie Parallelen zu klassischen hydrologischen Modellen. Wir skizzieren diese im Folgenden. Für eine detaillierte, technische Beschreibung verweisen wir auf Kratzert et al. (2018b).

Die meisten in der Hydrologie verwendeten Modelle bestehen aus einer Abfolge von Speichern (z.B. für Schnee oder Bodenfeuchte) und fest eingebauten Regeln. Diese Regeln definieren, wie Eingabedaten die Speicher verändern und wie aus den Speichern die Abflussvorhersage gewonnen werden kann (e.g. Liang et al. 1994; Seibert und Vis 2012). Es bleibt eine Reihe an Modellparametern, mit denen sich

die Modelle an individuelle Einzugsgebiete anpassen lassen. Wie eingangs beschrieben, besitzen auch LSTMs interne Speicher und auch hier wird die Vorhersage aus den Speichern erzeugt. Der wesentliche Unterschied ist jedoch, dass beim LSTM nicht von vornherein definiert ist, wie die Eingabedaten mit der Zielvariable zusammenhängen und welche Informationen dafür über die Zeit gespeichert oder akkumuliert werden müssen. Diese Zusammenhänge werden während der sogenannten Trainingsphase vom LSTM gelernt. Das LSTM muss also nicht passende Parameterschätzungen einer gegebenen Modellstruktur finden, sondern ein gesamtes Modell. Dies ist eine Stärke: Klassische hydrologische Modelle können nur jene Systemprozesse berücksichtigen, welche explizit in das Modell eingebaut wurden – LSTMs hingegen können alle Prozesse modellieren, die aus den verfügbaren Daten ableitbar sind.

In unseren Experimenten verwenden wir das LSTM in folgender Konfiguration: Wir übergeben dem LSTM eine Sequenz von ca. einem Jahr an meteorologischen Eingabedaten, um eine Abflussvorhersage am letzten Zeitschritt zu erzeugen. Zusätzlich zu den meteorologischen Daten werden in den meisten Experimenten auch statische Kenngrößen des Einzugsgebietes (wie etwa Ton- oder Sandgehalt des Bodens) als zusätzliche Inputs an jedem Zeitschritt übergeben (siehe Abschn. 2.2).

2.2 Daten

In allen Experimenten verwenden wir Daten aus dem CAMELS-Datensatz (Newman et al. 2015; Addor et al. 2017), einem offenen hydrologischen Datensatz mit 671 unbeeinflussten Einzugs-

gebieten aus den USA. Die Einzugsgebiete sind quer über die USA verteilt und decken ein breites Spektrum an unterschiedlichen hydrologischen Eigenschaften ab (Abb. 3). Neben ca. 35 Jahren Abflussdaten in täglicher Auflösung beinhaltet CAMELS meteorologische Daten von drei verschiedenen meteorologischen Produkten sowie eine Reihe statischer Kenngrößen. Die drei Produkte beinhalten jeweils Niederschlag, tägliche Höchst- und Niedertemperatur, Sonneneinstrahlung und Wasserdampfdruck. Die statischen Kenngrößen lassen sich grob in die Bereiche Boden, Vegetation, Hydrologie, Klima und Topologie gliedern (eine Übersicht ist in Newman et al. 2015 und Addor et al. 2017 zu finden).

3 Experimente und Ergebnisse

3.1 Regionale Modellierung

3.1.1 Vergleich des einzugsgebietsspezifischen und regionalen LSTM

In diesem Experiment vergleichen wir einzugsgebietsspezifische LSTMs mit einem regionalen LSTM. Ersteres bedeutet, dass für jedes Einzugsgebiet ein separates LSTM trainiert und verwendet wird. Letzteres bedeutet, dass ein einziges Modell so aufgesetzt wird, dass es für alle Einzugsgebiete verwendbar ist (siehe Diskussion über regionale Modellbildung in der Einleitung). Die einzugsgebietsspezifischen LSTMs erhalten ausschließlich meteorologische Eingabedaten, wohingegen das regionale LSTM zusätzlich statische Kenngrößen der Einzugsgebiete als Inputs bekommt (siehe Abschn. 2.2). Dies ermöglicht es dem regionalen LSTM, die meteorologischen Inputs je nach Einzugsgebiet unterschiedlich zu verarbeiten (siehe Kratzert et al. 2019c).

Abb. 4 zeigt die empirischen kumulativen Verteilungsfunktionen der NSE-Werte aus den CAMELS-Einzugsgebieten. Der regionale Ansatz (Median NSE 0,76) ist deutlich besser als die einzugsgebietsspezifischen LSTMs (Median NSE 0,67). Das Verhalten der LSTMs ist also genau gegensätzlich zu klassischen hydrologischen Modellen, bei denen einzugsgebietsspezifische Ansätze in der Regel deutlich besser sind (siehe auch Abb. 1). Bei der Verwendung von LSTMs als hydrologische Modelle ist es immer sinnvoll, verschiedene Einzugsgebiete in die Modellbildung

Abb. 3 Übersichtskarte der CAMELS-Einzugsgebiete, farblich kodiert nach Trockenheit (definiert als das Verhältnis der mittleren jährlichen potenziellen Verdunstung zu mittleren jährlichen Niederschlagsmenge, siehe Addor et al. (2017))

Abb. 4 Vergleich der empirischen kumulativen Verteilungen der Nash-Sutcliffe-Effizienz (NSE)-Werte zweier LSTMs über die verschiedenen Einzugsgebiete. Die bestmögliche Verteilung entspricht einer vertikalen Linie bei NSE 1,0. In Analogie zu Abb. 1 sind die einzugsgebietsspezifischen LSTMs mittels einer unterbrochenen Linie und das regionale LSTM mittels einer durchgängigen Linie dargestellt. In Gegensatz zu Abb. 1 ist das regionale LSTM besser als die einzugsgebietsspezifisch angepassten LSTMs

einfließen zu lassen. Dies gilt selbst dann, wenn das eigentliche Ziel der Modellierung sich auf einen spezifischen Flussabschnitt beschränkt. Eine Erklärung hierfür ist, dass die abflussbildenden Prozesse überall prinzipiell dieselben sind – wenn auch lokal stark unterschiedlich ausgeprägt. Je mehr

(diverse) Einzugsgebiete für das Training verwendet werden, desto mehr Prozessverständnis kann vom LSTM aus den Daten gelernt werden. Mehr Einzugsgebiete führen daher tendenziell zu einer besser verallgemeinernden Repräsentation der Niederschlags-Abfluss-Modellierung.

Resümee: Das LSTM ist in der Lage, gelerntes Prozessverständnis zwischen den Einzugsgebieten zu transferieren. Es lernt etwas aus den Daten eines gegebenen Einzugsgebiets, dass es in positiver Hinsicht in einem anderen Einzugsgebiet anwenden kann. Für eine ausführliche Diskussion verweisen wir auf Nearing et al. (2020).

3.1.2 Vergleich des regionalen LSTM mit hydrologischen Modellen

Vergleichen wir nun das regionale LSTM mit zwei unterschiedlichen, regional kalibrierten (mittels MPR; Samaniego et al. 2010) hydrologischen Modellen, ergibt sich folgendes Bild (Abb. 5): Das regionale LSTM (Median NSE 0,76) ist beiden hydrologischen Modellen (Median NSE VIC: 0,31, Median NSE mHM: 0,53) sehr deutlich überlegen.

Aber nicht nur im Vergleich zu regional kalibrierten hydrologischen Modellen schneidet das LSTM gut ab. Beim Vergleich des regionalen LSTM mit einzugsgebietsspezifischen hydrologischen Modellen ergibt sich ein ähnliches Bild (siehe Abb. 6): auch hier ist das LSTM deutlich besser. Dabei muss betont werden, dass alle hydrologischen Modelle von anderen Forschergruppen mit Expertise in der Anwendung des entsprechenden Modells

Abb. 5 Vergleich der empirischen kumulativen Verteilungen der Nash-Sutcliffe-Effizienz (NSE)-Werte für das regionale VIC-Modell (*dunkelblau*), das regionale mHM-Modell (*rot*), und das regionale LSTM (*orange*). Je weiter rechts die Gesamtverteilung, desto besser die Modellgüte. In diesem Vergleich ist das regionale LSTM der mit Abstand beste Ansatz

kalibriert wurden, um jeglichen Bias bei der Kalibrierung zu vermeiden. Alle Modelle wurden mit denselben meteorologischen Daten auf den denselben Zeiträumen kalibriert und ausgewertet. Für weitere Details zu den hydrologischen Modellen sowie für Vergleiche mit weiteren Metriken verweisen wir auf Kratzert et al. (2019c).

Resümee: Bezogen auf die Modellgüte (hier exemplarisch in Form des NSE-Kriteriums dargestellt) übertrifft das regionale LSTM eine Reihe von klassischen hydrologischen Modellen. Dies gilt sowohl für Modelle, die regional kalibriert wurden, als auch für solche, die einzugsgebietspezifisch optimiert wurden. Ein regionales LSTM ist also ein einziges LSTM, das für Vorhersagen an allen Pegeln in einer Region verwendet werden kann. Für weitere Details zu den hydrologischen Modellen sowie Vergleiche mit weiteren Metriken verweisen wir auf Kratzert et al. (2019c).

3.2 Generalisierung des LSTM für unbeobachtete Einzugsgebiete

Trotz dieser eindeutigen Ergebnisse stellt sich die Frage, wie gut das gelernte „Wissen“ eines datenbasierten Modells wirklich ist, oder ob das Modell nur gut ist im „Auswendiglernen“. In Kratzert et al. (2019b) haben wir diese Annahme untersucht. Dafür wurden die Einzugsgebiete mittels zufälliger Kreuzvalidierung in 10 unterschiedliche Gruppen

aufgeteilt. Anschließend wurde jeweils ein regionales LSTM anhand der Einzugsgebiete aus neun Gruppen trainiert und anhand der Einzugsgebiete in der verbleibenden zehnten Gruppe ausgewertet. Dabei ist jedes Einzugsgebiet genau einmal in der Gruppe, die nicht für das Modelltraining verwendet wird. Der Versuchsaufbau misst also, wie gut ein Modell in Einzugsgebieten ist, für die keine Daten für die Kalibrierung zur Verfügung stehen. Damit wird eine Vorhersage in unbeobachteten Einzugsgebieten (Prediction in Ungauged Basins, PUB) simuliert, was als eine der wichtigsten Problemstellungen in der Hydrologie gilt (siehe Einleitung). Eine quantitative Auswertung von de facto unbeobachteten Einzugsgebieten ist natürlich nicht möglich. Unser Ansatz bietet jedoch eine naheliegende Annäherung.

Die Ergebnisse sind in Abb. 7 zu sehen: Dort wird das PUB-Modell mit zwei hydrologischen Modellen sowie dem regionalen LSTM verglichen. Letzteres wurde wie in den vorherigen Kapiteln mit allen Daten trainiert. Die hydrologischen Modelle sind das SAC-SMA-Modell (Burnash et al., 1973, 1995), separat kalibriert für jedes Einzugsgebiet und das aktuelle US National Water Model (WRF-Hydro; Salas et al. 2018). Trotz Einbußen gegenüber dem regionalen LSTM ist das in diesem Experiment aufgesetzte PUB-

LSTM besser als die zum Vergleich verwendeten klassischen Modelle.

In diesem Vergleich stellen wir die Performance eines LSTM für unbeobachtete Einzugsgebiete (PUB; orange mit Stern) der Performance von drei anderen Modellen gegenüber: ein physikalisches Modell (US NWM, Cosgrove und Klemmer (NOAA) 2019), ein einzugsgebietspezifisch kalibriertes konzeptionelles Modell (SAC-SMA) und das zuvor eingeführte regionale LSTM (Abschn. 2.1). Wir haben die ersten beiden Modelle gewählt, da sie das alte (SAC-SMA) sowie aktuell verwendete US National Water Model und somit wichtige Modelle sind, die weit verbreitete Anwendung finden.

Resümee: Das LSTM liefert durchschnittlich bessere Simulationsergebnisse für Einzugsgebiete, für die es nicht trainiert wurde, als etablierte hydrologische Modelle, die für jedes Einzugsgebiet separat kalibriert wurden. Das LSTM lernt eine besser generalisierende Repräsentation der abflussbildenden Prozesse als die von einzugsgebietspezifisch kalibrierten Modellen erreichte. Eine vertiefende Diskussion bezüglich der vorgestellten Experimente und Ergebnisse kann in Kratzert et al. (2019b) gefunden werden.

3.3 Modelleigenschaften

3.3.1 Synergieeffekte mehrerer meteorologischen Daten in einem LSTM

Ein weiterer Vorteil von DL-Modellen ist es, dass sie keine A-priori-Annahmen zu den Eingabedaten benötigen. Im Gegensatz zu klassischen hydrologischen Modellen ist es also nicht notwendig, eine genaue mathematische Beziehung zwischen jeder Eingangsvariable und der Zielvariable (z.B. Abfluss) zu definieren, sondern genau diese Beziehungen werden während der Trainingsphase vom Modell gelernt. Dies können wir uns auf mehrere Arten zunutze machen, wobei eine spezielle Anwendung die meteorologischen Inputs betrifft.

Konzentrieren wir uns auf die wichtigste Eingangsvariable, den Niederschlag, so ist weithin bekannt, dass hierin eine der größten Quellen von Unsicherheit steckt (Fekete et al. 2004; Beven 2019). Niederschlagsdaten gibt es sowohl als Punktmessung (z. B. von Niederschlagsmessstellen) als auch Rasterdaten (Maurer et al. 2002; Thornton et al. 2012; Xia et al. 2012; Hersbach

Abb. 6 Modellgütevergleich zwischen verschiedenen einzugsgebietsspezifischen klassischen Niederschlags-Abfluss-Modellen und dem regionalen LSTM (orange). Dargestellt werden die empirischen kumulativen Verteilungsfunktionen der Nash-Sutcliffe-Effizienz (NSE) über die verschiedenen Einzugsgebiete. Je weiter nach rechts verschoben die Gesamtverteilung, desto besser ist das Modell. In diesem Vergleich ist das regionale LSTM der beste Ansatz

et al. 2020). Erstere sind, wie der Name impliziert, nur am Ort der Messstelle gültig, wohingegen letztere wiederum selbst das Produkt eines Modells sind, das mit eigenen Annahmen und Fehlern behaftet ist. Des Weiteren existiert eine Vielzahl unterschiedlicher meteorologischer Produkte – die teils mit unterschiedlichen Modellen erzeugt wurden – in unterschiedlicher räumlicher und zeitlicher Auflösung (e.g. verschieden generierte Ensemble-Vorhersagen). Dabei variieren vor allem die Niederschlagsdaten für ein und denselben Ort und Zeitpunkt teilweise deutlich, und trotzdem gibt es zumeist kein universal „bestes“ Niederschlagsprodukt (z.B. Behnke et al. 2016). DL bietet die Möglichkeit, aus den unterschiedlichen meteorologischen Datenprodukten möglicherweise vorhandene Synergien zu extrahieren. Ein einzelnes Modell kann sich diese Synergieeffekte zunutze machen und somit bessere Ergebnisse erzeugen.

Wie im Abschn. 2.2 beschrieben, stehen im CAMELS-Datensatz drei verschiedene meteorologische Produkte zur Verfügung (DayMet, Maurer und NLDAS). In Kratzert et al. (2020) haben wir untersucht, wie sich die Verwendung mehrerer unterschiedlicher Niederschlagsdaten (neben anderen meteorologischen Variablen) auf die Modellgüte auswirken. Dazu haben wir regionale LSTM für jedes meteorologi-

sche Produkt separat kalibriert, weiters für jede Kombination aus zwei meteorologischen Produkten sowie ein regionales LSTM, das die meteorologischen Variablen aus allen drei Datensätzen auf einmal als Eingabedaten bekommt.

Die Ergebnisse aus Kratzert et al. (2020) zeigen, dass jede Kombination aus zwei meteorologischen Datenprodukten besser ist als das beste Modell, das nur mit Daten aus einem meteorologischen Produkt gefüttert wird (Abb. 8). Weiter ist das Modell, in das alle drei meteorologischen Produkte auf einmal einfließen, besser als das beste der Modelle, die eine Kombination aus zwei meteorologischen Produkten erhalten. Kurz gesagt: je mehr (unterschiedliche) meteorologische Daten (insbesondere Niederschlag) das Modell gleichzeitig erhält, desto besser.

Vergleicht man das beste Modell (jenes, das alle drei meteorologischen Daten auf einmal bekommt) mit dem regionalen Modell aus den vorherigen Kapiteln, so ergibt sich eine konstante Verbesserung der Modellgüte (Abb. 9). Eines der gängigsten Verfahren, um Informationen aus unterschiedlichen meteorologischen Datensätzen mit klassischen hydrologischen Modellen zu verwenden, ist das Bilden eines „Modell-Ensembles“. Hierbei werden für jeden meteorologischen Datensatz Modelle separat kalibriert und die Vorhersagen der einzelnen Modelle dann

kombiniert, z.B. über das arithmetische Mittel (siehe SAC-SMA-Ensemble, Abb. 9). Der Nachteil eines solchen Ensembles ist, dass es – zusätzlich zu den bereits genannten Annahmen zur Modellstruktur – voraussetzt, dass die Datenprodukte voneinander unabhängig sind. Jedes Modell wird mit imperfekten Daten betrieben und potenzielle synergetische Effekte, die aus der produktübergreifenden Verbindung zur Verfügung stehender Informationen für die Modellierung der hydrologischen Prozesse gewonnen werden könnten, werden nicht genutzt.

Resümee: Das LSTM hat die Fähigkeit, diverse Inputs zu berücksichtigen und etwaige Synergien in den Daten (z.B. bei Verwendung unterschiedlicher meteorologischer Datensätze in einem Modell) für eine verbesserte Modellierung zu nutzen.

3.3.2 Analyse und Interpretation von LSTM-Speichern

Unter Hydrologen genießen neuronale Netze nicht den besten Ruf. Einer der Hauptgründe dafür ist, dass Modelle nicht nur Vorhersagen für den Abfluss liefern sollen, sondern auch ein hydrologisches Prozessverständnis darüber hinaus ermöglichen sollen (Klemeš 1986). Sprich: Ein Modell soll nicht nur gute Simulationsergebnisse für den Abfluss generieren, sondern es soll dies tun, weil es die entsprechenden Prozesse, die Niederschlag in Abfluss verwandeln, (richtig) modelliert. Allerdings ist bekannt, dass hydrologische Modelle zwar beispielsweise Bodenfeuchte oder Schnee intern modellieren, diese Werte jedoch nur bedingt an die realen Prozesse gekoppelt sind und zwischen verschiedenen hydrologischen Modellen stark variieren können (Bouaziz et al. 2020).

Bei neuronalen Netze ist zunächst nicht klar, was intern modelliert wird, um von den Eingabedaten auf die Zielvariable zu kommen. Das Forschungsfeld der Explainable AI widmet sich diesem Thema und erforscht Methoden, die das Modellverständnis verbessern sollen (Samek et al. 2019). Hinzu kommt, dass das LSTM explizite Speicherzellen besitzt (siehe Abschn. 2.1) und es möglich ist zu analysieren, was das LSTM über die Zeit in diesen Zellen modelliert.

In Kratzert et al. (2018a, 2019a, 2019c) haben wir untersucht, was das LSTM in seinen Speicherzellen model-

Abb. 7 Vergleich der Modellgüte für die Vorhersage in unbeobachteten Einzugsgebieten (PUB)

liert und konnten dabei nachweisen, dass das LSTM wichtige hydrologische Prozesse wie Bodenfeuchte und Schnee intern modelliert, obwohl es nur trainiert wurde, aus meteorologischen Daten Abfluss vorherzusagen. Abb. 10 zeigt beispielhaft den Vergleich von modelliertem Schnee-Wasser-Äquivalent (SWE) eines etablierten Schneemodells (Snow-17; REF) mit SWE, welches aus den Speicherzellen des LSTM abgeleitet wurde. Dies ist ein Indikator dafür, dass der datenbasierte Ansatz keine Scheinkorrelationen abbildet, sondern den unterliegenden Prozess intern wiedergibt. Gepaart mit der exzellenten Modellgüte (siehe Abschnitt „Regionale Modellierung“) ist also davon auszugehen, dass LSTMs tatsächlich hydrologisch relevante Prozesse lernen und intern modellieren, die es dem LSTM ermöglichen, einzugsgebietsübergreifend gute Vorhersagen zu liefern.

Resümee: Es ist möglich, die Speicher von LSTMs mit hydrologischen Prozessen in Verbindung zu setzen und so Einsichten über die interne Organisation der ML-Modelle zu gewinnen.

4 Erweiterungen

Abschließend wollen wir noch auf einige eine Reihe weiterer Forschungsergebnisse eingehen, die auch im Kontext der hydrologischen Modellierung von Interesse sind:

1. Das MC-LSTM (Hoedt et al. 2021), eine neue Variante des LSTM mit Massenerhaltung.

2. Das Prinzip des MTS-LSTM (Gauch et al. 2020), mit dem es möglich ist, aus einem LSTM-basierten Modell Abflussprognosen in unterschiedlichen zeitlichen Auflösungen zu gewinnen.
3. LSTM-basierte Modelle zur direkten Schätzung von Unsicherheiten in den Abflussprognosen (Klotz et al. 2021).

4.1 Massenerhaltung

Das Mass-Conserving LSTM (MC-LSTM; Hoedt et al. 2021) ist eine vor kurzem entwickelte Adaptierung des LSTM, das durch seine Architektur die Erhaltung von Massen-Inputs garantiert. Hierbei muss die Masse nicht zwangsläufig Wasser sein, sondern kann auch jede andere beliebige Größe repräsentieren (z.B. Energie, Geld, Personen). Das MC-LSTM besitzt nach wie vor explizite Speicher, allerdings werden die Eingabedaten in zwei Gruppen aufgeteilt: Massen-Inputs und Nicht-Massen-Inputs. Letztere werden ausschließlich für die Steuerung innerhalb des Netzwerkes verwendet (wie viel Masse fließt wohin, wie viel Masse verlässt an einem Zeitschritt das System etc.) und die Massen-Inputs sind jene Werte, die in ihrer Gesamtbilanz im System erhalten werden (z.B. Wasser in Form von Niederschlag). In den Speichern des MC-LSTM wird ausschließlich Masse gespeichert (also z.B. Wasser), was eine Interpretation des Modells erleichtert. Für eine technische Beschreibung des MC-LSTM sowie erste Ergebnisse zu

dessen Anwendung in der Hydrologie siehe Hoedt et al. (2021).

4.2 Berücksichtigung mehrerer Zeitskalen

Für viele Anwendungen genügt es, Prognosen in täglicher Auflösung bereitzustellen. Für bestimmte Aufgabenbereiche, etwa die Flutvorhersage, ist diese Auflösung jedoch nicht ausreichend.

Ein wichtiges Forschungsgebiet der LSTM-basierten Modellierung ist daher die Vorhersage in hoher zeitlicher Auflösung (etwa stündlich). Hierbei ergeben sich sehr lange Zeitreihen der Eingabedaten, was die Anwendung von LSTMs in der Praxis erschwert. Weiterhin müsste man in diesem Ansatz je ein unabhängiges LSTM für jede Zeitauflösung trainieren, wobei mit hoher Wahrscheinlichkeit die Vorhersagen der einzelnen Modelle inkonsistent würden.

Multi-Timescale LSTMs (MTS-LSTM, Gauch et al. 2020) bieten eine Lösung für diese Probleme: Es ermöglicht die gemeinsame Vorhersage in verschiedenen zeitlichen Auflösungen durch ein einziges Modell. Zudem bietet es die Möglichkeit, Eingabedaten mit unterschiedlichen Vorhersagehorizonten (lead times) für die verschiedenen Auflösungen zu nutzen. Intuitiv basiert der Ansatz darauf, dass der Wasserkreislauf ein gedämpftes System ist: Für stündliche Vorhersagen sind hochaufgelöste Daten der nahen Vergangenheit wichtig, während für weit zurückliegende Ereignisse eine niedrigere Auflösung genügt. Für eine detaillierte Beschreibung des Ansatzes verweisen wir auf Gauch et al. (2020).

4.3 Schätzung von Unsicherheiten

Unsicherheiten sind eine intrinsische Komponente der hydrologischen Modellierung. Jede Vorhersage, jede Messung und jedes Modell ist mit Unsicherheit behaftet. Die Schätzung von Unsicherheiten ist eine zentrale Komponente vieler modellgestützter Anwendungen. Die Bereitstellung einer solchen Schätzung ist jedoch nicht trivial und oft mit facettenreichen Problemstellungen assoziiert (Nearing und Gupta 2015; Beven 2016). Klassische Modelle benötigen in der Regel einen zusätzlichen Mechanismus, um eine solche Schätzung zu ermöglichen (z.B. pre-processing, post-processing, Datenassimilation oder Monte-Carlo-Sampling). DL-

Abb. 8 Verteilung der NSE-Werte für regionale LSTMs, die mit je einem meteorologischen Produkt (*grün*), mit je einer Kombination aus zwei Produkten (*blau*) bzw. mit allen drei verfügbaren Produkten (*rot*) kalibriert wurden

Abb. 9 Modellgütevergleich für die Ansätze mit mehrerer meteorologischen Eingaben. Die empirischen kumulativen Verteilungsfunktionen zeigen ein hydrologisches Ensemble-Modell (SAC-SMA; *blaue unterbrochene Linie*) im Vergleich zu einem LSTM, das mehrere meteorologische Datenprodukte als Eingaben verwendet (LSTM (multi forcing); *orange Linie mit Punkten*). Das normale LSTM, das bereits in den vorherigen Experimenten eingeführt wurde (LSTM (single forcing); *orange Linie*), dient hierbei als Orientierungspunkt, um die Güte einordnen zu können

basierte Modelle können die Unsicherheiten direkt aus dem Modell ableiten. Grundsätzlich werden dabei entweder modellintrinsische Eigenschaften genutzt oder Architekturen verwendet, die direkte probabilistische Vorhersagen ermöglichen. In Klotz et al. (2021) vergleichen und testen wir eine Reihe von Methoden zur Unsicherheitsschätzung mit DL-basierten Niederschlags-Abfluss-Modellen. Im Test zeichnet sich bereits ab, dass DL-basierte Verfahren sehr gute Unsicherheitsschätzungen ermöglichen.

5 Zusammenfassung, Diskussion und Ausblick

Viele technologische Entwicklungen der letzten Jahre sind auf Deep Learning (DL) zurückzuführen. In diesem Beitrag haben wir gezeigt, wie sich DL-Modelle, im Speziellen das Long Short-Term Memory Netzwerk (LSTM), für wasserwirtschaftliche Fragestellungen nutzen lassen.

Zuerst haben wir demonstriert, dass LSTM-basierte hydrologische Modelle stark davon profitieren, wenn sie für eine große Anzahl an (unterschiedlichen) Einzugsgebieten gleichzeitig trainiert werden und dass diese Vorgehensweise

Abb. 10 Beispielhafter Vergleich von Schnee-Wasser-Äquivalent (*mittlere Reihe*), modelliert vom Schneemodell Snow-17 (*blaue Linien*) sowie extrahiert aus LSTM-Speichern (*orange unterbrochene Linie*). Zusätzlich dargestellt sind Niederschlag (*obere Reihe*) und Tageshöchst- sowie Tagestiefsttemperatur (*untere Reihe*). Beim Niederschlag sind Tage mit Temperaturen unter dem Gefrierpunkt als schwarze Balken dargestellt

einzugsgebietspezifischen Modellen überlegen ist.

Der Vergleich dieser „regionalen“ LSTMs mit klassischen hydrologischen Modellen liefert zwei wesentliche Ergebnisse: Erstens, die regionalen LSTMs sind regional kalibrierten hydrologischen Modellen deutlich überlegen. Zweitens, das regionale LSTM ist auch besser als einzugsgebietspezifische hydrologische Modelle.

Des Weiteren haben wir gezeigt, dass ein regionales LSTM ein sehr allgemeines hydrologisches Prozessverständnis lernt. Angewendet auf Einzugsgebiete, die nicht für das Training des Modells verwendet wurden (eine Annäherung an die Vorhersage in unbeobachteten Einzugsgebieten), liefert das regionale LSTM im Durchschnitt bessere Simulationsergebnisse als speziell für die Einzugsgebiete kalibrierte hydrologische Modelle.

Ein weiterer Vorteil von DL-Modellen ist, dass sie beliebige Daten als Eingabedaten erlauben. LSTMs sind in der Lage, synergetische Effekte aus unter-

schiedlichen meteorologischen Datensätzen zu extrahieren und für die Verbesserung der Modellgüte zu nutzen.

Zuletzt haben wir gezeigt, dass LSTMs relevante hydrologische Prozesse (wie z. B. das Akkumulieren und Schmelzen von Schnee) lernen, ohne darauf trainiert worden zu sein. Dies legt nahe, dass das LSTM keine Scheinkorrelationen lernt, sondern die tatsächlichen physikalischen Zusammenhänge zwischen meteorologischen Eingabedaten und Abfluss.

Die künstliche Intelligenz bietet große Chancen für die wasserwirtschaftliche Anwendung. In diesem Beitrag haben wir das Potenzial anhand der jüngsten Entwicklungen in der Niederschlags-Abfluss-Modellierung herausgearbeitet. Diese Fortschritte demonstrieren, wie bereits jetzt leistungsfähige Werkzeuge aus dieser Technologie gewonnen werden können. Die jetzigen prototypischen Vorhersagemodelle erzielen bereits bessere Ergebnisse als klassische Ansätze, die bereits über Jahre verbessert wurden.

Trotz der vielversprechenden Ergebnisse befinden wir uns noch in der Anfangsphase von DL-basierten Ansätzen. Die Möglichkeiten bzw. Limitierungen sind noch nicht genügend erforscht. Wir erwarten daher, dass in den nächsten Jahren noch zahlreiche Erweiterungen konzipiert und Anwendungen erschlossen werden.

Hierfür seien zwei Beispiele genannt: (a) Die Erstellung von räumlich verteilten DL-Modellen, welche die Flussstruktur explizit mit einbeziehen. Es existieren bereits erste Ansätze, die dies versuchen (z. B. Moshe et al. 2020), jedoch gibt noch keinen allgemeinen, gut skalierbaren Lösungsansatz. (b) Das Verwenden von DL-Modellen für hypothetische Untersuchungen. Beispielsweise die Fragestellung: Wie würde sich die Aufforstung in einem Teil des Einzugsgebiets auf den Abfluss auswirken?

Beide Themengebiete sind Bestandteil unserer aktuellen Forschung.

Funding Open access funding provided by Johannes Kepler University Linz.

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 betref-

fende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

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

Literatur

- Addor, N., Newman, A. J., Mizukami, N. & Clark, M. P. (2017): The CAMELS data set: catchment attributes and meteorology for large-sample studies, *Hydrology and Earth System Sciences* 21(10), 5293–5313.
- Behnke, R., Vavrus, S., Allstadt, A., Albright, T., Thogmartin, W. E. & Radeloff, V. C. (2016): Evaluation of downscaled, gridded climate data for the conterminous United States, *Ecological applications* 26(5), 1338–1351.
- Beven, K. (2016): Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrological Sciences Journal* 61(9), 1652–1665.
- Beven, K. (2019): Towards a methodology for testing models as hypotheses in the inexact sciences, *Proceedings of the Royal Society A* 475(2224), 20180862.
- Blöschl, G. & Sivapalan, M. (1995): Scale issues in hydrological modelling: a review, *Hydrological processes* 9(3–4), 251–290.
- Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpff, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., de Amorim, P. B., Böttcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chiffard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P. J., de Rooij, G., Baldassarre, G. D., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feicabrinno, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Bevacqua, A. G., González-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbuloglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolehchikina, A., Krause, S., Kremer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L. R., Lindquist, E., Link, T., Liu, J., Loucks, D. P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Missier, B. D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, E., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Sierra, C. P., Ramos, M.-H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sá, J. H. M., Salvadori, G., Sandells, M., Schaeffli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, L., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Loon, A. F. V., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J.-P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z., Yilmaz, K. K. & Zhang, Y. (2019): Twenty-three unsolved problems in hydrology (UPH)—a community perspective, *Hydrological Sciences Journal* 64(10), 1141–1158.
- Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T. & Viglione, A. (2013): Runoff prediction in ungauged basins: synthesis across processes, places and scales, Cambridge University Press.
- Bouaziz, L. J. E., Thirel, G., de Boer-Euser, T., Melsen, L. A., Buitink, J., Brauer, C. C., De Niel, J., Moustakas, S., Willems, P., Grellet, B., Drogue, G., Fenicia, E., Nossent, J., Pereira, E., Sprockereef, E., Stam, J., Dewals, B. J., Weerts, A. H., Savenije, H. H. G. & Hrachowitz, M. (2020): Behind the scenes of streamflow model performance, *Hydrology and Earth System Sciences Discussions* 2020, 1–38.
- Burnash, R. J. C., Ferral, R. L., and McGuire, R. A. (1973): A generalised streamflow simulation system conceptual modelling for digital computers., Tech. rep., US Department of Commerce National Weather Service and State of California Department of Water Resources, Sacramento, CA, USA.
- Burnash, R. J. C. (1995): The NWS river forecast system-catchment modeling, Computer models of watershed hydrology. 1995 pp.311–366 ref.12.
- Cosgrove, B. & Klemmer, C. (2019): The National Water Model, US National Oceanic and Atmospheric Administration, <https://water.noaa.gov/about/nwm>, last access: 2021-01-25.
- Feigl, M., Herrnegger, M., Klotz, D. & Schulz, K. (2020): Function Space Optimization: A symbolic regression method for estimating parameter transfer functions for hydrological models, *Water resources research* 56(10), e2020WR027385.
- Fekete, B. M., Vörösmarty, C. J., Roads, J. O. & Willmott, C. J. (2004): Uncertainties in Precipitation and Their Impacts on Runoff Estimates, *Journal of Climate* 17(2), 294–304.
- Gauch, M., Mai, J. & Lin, J. (2021): The proper care and feeding of CAMELS: How limited training data affects streamflow prediction, *Environmental Modelling & Software* 135, 104926.
- Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Lin, J. & Hochreiter, S. (2021): Rainfall-runoff prediction at multiple timescales with a single Long Short-Term Memory network, *Hydrol. Earth Syst. Sci.*, 25, 2045–2062, <https://doi.org/10.5194/hess-25-2045-2021>.
- Gers, F. A., Schmidhuber, J., and Cummins, F. (2000): Learning to Forget: Continual Prediction with LSTM, *Neural Comput.*, 12, 2451–2471.
- Halevy, A., Norvig, P. & Pereira, F. (2009): The unreasonable effectiveness of data, *IEEE Intelligent Systems* 24(2), 8–12.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D. & others (2020): The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society* 146(730), 1999–2049.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N. & others (2012): Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *IEEE Signal processing magazine* 29(6), 82–97.
- Hochreiter, S. (1991): Untersuchungen zu dynamischen neuronalen Netzen, Diploma, Technische Universität München 91(1).
- Hochreiter, S. & Schmidhuber, J. (1997): Long short-term memory, *Neural computation* 9(8), 1735–1780.
- Hoedt, P.-J., Kratzert, F., Klotz, D., Halmich, C., Holzleitner, M., Nearing, G., Hochreiter, S., and Klambauer, G. (2021): MC-LSTM: Mass-Conserving LSTM, available at: <https://arxiv.org/abs/2101.05186>.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U. & others (2013): A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrological sciences journal* 58(6), 1198–1255.
- Kirchner, J. W. (2006): Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, 42, 1–5.
- Klemes, V. (1983): Conceptualization and scale in hydrology, *Journal of hydrology* 65(1–3), 1–23.
- Klemes, V. (1986): Dilettantism in hydrology: Transition or destiny?, *Water Resources Research* 22(9S), 1775–1885.
- Klemes, V. (1997): Guest editorial: Of carts and horses in hydrologic modeling, *Journal of Hydrologic Engineering* 2(2), 43–49.
- Klotz, D., Herrnegger, M. & Schulz, K. (2017): Symbolic regression for the estimation of transfer functions of hydrological models, *Water Resources Research* 53(11), 9402–9423.
- Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., and Nearing, G. (2021): Uncertainty Estimation with Deep Learning for Rainfall-Runoff Modelling, *Hydrol. Earth Syst. Sci. Discuss.* [preprint], <https://doi.org/10.5194/hess-2021-154>, in review.
- Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S. & Klambauer, G. (2018a): Do internals of

- neural networks make sense in the context of hydrology?, in AGU Fall Meeting Abstracts, pp. H13B–06.
- Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S. & Klambauer, G. (2019a): NeuralHydrology—Interpreting LSTM in Hydrology Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Springer, pp. 347–362.
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S. & Nearing, G. S. (2019b): Toward improved predictions in ungauged basins: Exploiting the power of machine learning, *Water Resources Research* 55(12), 11344–11354.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M. (2018b): Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrol. Earth Syst. Sci.*, 22, 6005–6022, <https://doi.org/10.5194/hess-22-6005-2018>.
- Kratzert, F., Klotz, D., Hochreiter, S., and Nearing, G. S. (2020): A note on leveraging synergy in multiple meteorological datasets with deep learning for rainfall–runoff modeling, *Hydrol. Earth Syst. Sci. Discuss.* [preprint], <https://doi.org/10.5194/hess-2020-221>, in review.
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G. (2019c): Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, *Hydrol. Earth Syst. Sci.*, 23, 5089–5110, <https://doi.org/10.5194/hess-23-5089-2019>.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017): Imagenet classification with deep convolutional neural networks, *Communications of the ACM* 60(6), 84–90.
- Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. (1994): A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres* 99(D7), 14415–14428.
- Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P. & Nijssen, B. (2002): A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States, *Journal of climate* 15(22), 3237–3251.
- Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O. & Samaniego, L. (2017): Towards seamless large-domain parameter estimation for hydrologic models, *Water Resources Research* 53(9), 8020–8040.
- Moshe, Z., Metzger, A., Elidan, G., Kratzert, F., Nevo, S. & El-Yaniv, R. (2020): HydroNets: Leveraging River Structure for Hydrologic Modeling, in Workshop on AI for Earth Sciences 8th International Conference on Learning Representations (ICLR 2020).
- Nearing, G. S. & Gupta, H. V. (2015): The quantity and quality of information in hydrologic models, *Water Resources Research* 51(1), 524–538.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pellissier, C. S., Klotz, D., Frame, J. M., Prieto, C. & Gupta, H. V. (2021): What role does hydrological science play in the age of machine learning?, *Water Resources Research*, 57, e2020WR028091, <https://doi.org/10.1029/2020WR028091>.
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R. & others (2015): Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrology and Earth System Sciences* 19(1), 209–223.
- Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G. (2013): Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, *Hydrol. Earth Syst. Sci.*, 17, 1783–1795, <https://doi.org/10.5194/hess-17-1783-2013>.
- Radford, A., Jyzeffowicz, R. & Sutskever, I. (2017): Learning to Generate Reviews and Discovering Sentiment, *CoRR abs/1704.01444*.
- Salas, F. R., Somos-Valenzuela, M. A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., Yu, W., Ding, D., Clark, E. P. & Noman, N. (2018): Towards real-time continental scale streamflow simulation in continuous and discrete space, *JAWRA Journal of the American Water Resources Association* 54(1), 7–27.
- Samaniego, L., Kumar, R., and Attinger, S. (2010): Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resour. Res.*, 46, W05523, <https://doi.org/10.1029/2008WR007327>.
- Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K. & Müller, K.-R. (2019): Explainable AI: interpreting, explaining and visualizing deep learning, Vol. 11700, Springer Nature.
- Seibert, J. & Vis, M. J. P. (2012): Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, *Hydrology and Earth System Sciences* 16(9), 3315–3325.
- Sivapalan, M. (2003): Prediction in ungauged basins: a grand challenge for theoretical hydrology, *Hydrological Processes* 17(15), 3163–3170.
- Sutskever, I., Vinyals, O. & Le, Q. V. (2014): Sequence to sequence learning with neural networks, *Advances in neural information processing systems* 27, 3104–3112.
- Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelm, N., Wei, Y., Devarakonda, R. & Cook, R. (2012): Daymet: Daily surface weather on a 1 km grid for North America, 1980–2008, Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC).
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J. & others (2012): Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, 117, D03109, <https://doi.org/10.1029/2011JD016048>.

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