



Schätzung der Verdunstung mithilfe von Machine- und Deep Learning-Methoden

Claire Brenner · Jonathan Frame · Grey Nearing · Karsten Schulz

Angenommen: 14. April 2021 / Online publiziert: 17. Mai 2021
 © Der/die Autor(en) 2021

Zusammenfassung Die Verdunstung ist ein entscheidender Prozess im globalen Wasser-, Energie- sowie Kohlenstoffkreislauf. Daten zur räumlich-zeitlichen Dynamik der Verdunstung sind daher von großer Bedeutung für Klimamodellierungen, zur Abschätzung der Auswirkungen der Klimakrise sowie nicht zuletzt für die Landwirtschaft.

In dieser Arbeit wenden wir zwei Machine- und Deep Learning-Methoden für die Vorhersage der Verdunstung mit täglicher und halbstündlicher Auflösung für Standorte des FLUXNET-Datensatzes an. Das Long Short-Term Memory Netzwerk ist ein rekurrentes neuronales Netzwerk, welches explizit Speichereffekte berücksichtigt und Zeitreihen der Eingangsgrößen analysiert (entsprechend physikalisch-basierten Wasserbilanzmodellen). Dem gegenüber gestellt werden Modellierungen mit XGBoost, einer Entscheidungsbaum-Methode, die in diesem Fall nur Informationen für den zu bestimmenden Zeitschritt erhält (entsprechend physikalisch-basierten Energiebilanzmodellen). Durch diesen Vergleich der beiden Modellansätze soll untersucht werden, inwieweit sich durch die Berücksichtigung von Speichereffekten Vorteile für die Modellierung ergeben.

Die Analysen zeigen, dass beide Modellansätze gute Ergebnisse erzielen und im Vergleich zu einem ausgewer-

teten Referenzdatensatz eine höhere Modellgüte aufweisen. Vergleicht man beide Modelle, weist das LSTM im Mittel über alle 153 untersuchten Standorte eine bessere Übereinstimmung mit den Beobachtungen auf. Allerdings zeigt sich eine Abhängigkeit der Güte der Verdunstungsvorhersage von der Vegetationsklasse des Standorts; vor allem wärmere, trockene Standorte mit kurzer Vegetation werden durch das LSTM besser repräsentiert, wohingegen beispielsweise in Feuchtgebieten XGBoost eine bessere Übereinstimmung mit den Beobachtung liefert. Die Relevanz von Speichereffekten scheint daher zwischen Ökosystemen und Standorten zu variieren.

Die präsentierten Ergebnisse unterstreichen das Potenzial von Methoden der künstlichen Intelligenz für die Beschreibung der Verdunstung.

Schlüsselwörter Verdunstung · LSTM · XGBoost · Fluxnet

Predicting evapotranspiration using machine and deep learning methods

Abstract Evapotranspiration is a key driver of the Earth's climate system, modulating processes in the global water, energy as well as carbon cycle. Information on evapotranspiration rates and its variability in space and time are therefore essential for climate modelling, examining the effect of climate change and sustainable agricultural management.

In this study we apply two machine- and deep learning methods for predicting evapotranspiration at half-hourly and daily temporal resolutions using data from the FLUXNET network. We train a Long Short-Term Memory, a recurrent neural network with explicit memory that is particularly suited for time series predictions (similar to physically-based water balance models). As a second model we use XGBoost, a regression tree method that does

not contain system memory (similar to physically-based energy balance models). With this setup we want to examine the importance of memory effects for predicting evapotranspiration.

Our analyses show that both modelling approaches agree well with the observations. Averaging over all 153 studied FLUXNET sites, the LSTM performs significantly better than XGBoost. However, model performance varies with vegetation class; the LSTM yields better results for warmer and drier sites with low vegetation, whereas XGBoost performs better for e.g. wetland sites. Thus, the importance of system memory seems to vary with ecosystem and climate type.

The presented results underline the potential of artificial intelligence for predicting evapotranspiration.

Keywords Evapotranspiration · LSTM · XGBoost · Fluxnet

1 Einleitung

Die Verdunstung oder auch Evapotranspiration (ET) ist – nach dem Niederschlag – global betrachtet die zweitgrößte Komponente des terrestrischen Wasserkreislaufs; weltweit werden ca. 60–65% des Niederschlags in Form von Verdunstung an die Atmosphäre zurückgegeben (Brutsaert 2005). In Österreich ist der Anteil der Verdunstung aufgrund klimatischer, aber vor allem auch topografischer Bedingungen etwas geringer, allerdings beansprucht sie auch hier ca. 43% des Niederschlags (BMLRT 2021). Hierbei ist weiters zu beachten, dass der prozentuelle Anteil der Verdunstung am Niederschlag räumlich stark variieren kann und beispielsweise im landwirtschaftlich intensiv genutzten Nordosten Österreichs größer ist als im alpinen Raum (Kling 2006).

Die mengenmäßige Erfassung der Verdunstung spielt daher eine wesentliche Rolle für die nachhaltige Bewirtschaftung der verfügbaren Wasserressourcen (Anderson et al. 2011; Bas-

DI Dr. C. Brenner (✉) ·
 Univ.-Prof. Dipl.-Geoökol. Dr. K. Schulz
 Institut für Hydrologie und
 Wasserwirtschaft, Universität
 für Bodenkultur Wien,
 Muthgasse 18, 1190 Wien, Österreich
claire.brenner@boku.ac.at

J. Frame
 Department of Geological Sciences,
 The University of Alabama,
 Tuscaloosa, AL, USA

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

tiaanssen et al. 2005; Chirouze et al. 2014; Cleugh et al. 2007; Kustas et al. 2012; McMahon et al. 2013; Tang et al. 2011). Trotz der großen Bedeutung der Verdunstung für den Wasserkreislauf ist ihre mengenmäßige Erfassung – vor allem mit hoher räumlicher und zeitlicher Auflösung – bis heute schwierig (Cammalleri et al. 2014).

Messungen der Verdunstung, in Form von Eddy-Kovarianz (EC)-Messsystemen oder Lysimetern, sind aufwendig und daher selten. Im Gegensatz zu Abflussmessungen, welche Information über größere räumliche Einheiten integrieren, sind Verdunstungsmessungen meist nur für eine kleine Fläche in unmittelbarer Nähe der Messstation (einige 100 m) repräsentativ (Foken 2008).

Abschätzungen der Verdunstung, die beispielsweise auf fernerkundlichen Datensätzen basieren, können direkte Messungen ergänzen und räumlich verteilte Informationen liefern (Kalma et al. 2008). Im Fall von Satellitendaten erschweren die geringe räumliche und/oder zeitliche Auflösung der Daten allerdings die Anwendbarkeit dieser Methoden für die routinemäßige Erfassung der Verdunstung; räumlich hochaufgelöste Satellitendaten (< 100 m Auflösung) weisen eine geringe zeitliche Frequenz auf (eine Aufnahme alle 4 bis 14 Tage); zeitlich hochaufgelöste Satellitendaten (tägliche Daten) sind hingegen nur mit einem begrenzten räumlichen Auflösung (> 1 km) verfügbar. Für kleinskalige, lokale Abschätzungen, beispielsweise im Bereich der Präzisionslandwirtschaft, können Modellverfahren in Kombination mit Bilddaten, die mit Flugzeugen oder unbemannten Luftfahrzeugen erfasst werden, eine räumlich und zeitlich hochaufgelöste Abschätzung der Verdunstung ermöglichen (Brenner et al. 2017, 2018). Allerdings sind solche Kampagnen mit einem hohen Aufwand verbunden und können nicht für großflächige Abschätzungen herangezogen werden.

Erschwerend kommt bei der Verwendung von Fernerkundungsdaten hinzu, dass bei vorhandener Bewölkung keine Daten gewonnen werden können. Viele physikalisch-basierte ET-Modellansätze, die fernerkundliche Daten nutzen, verwenden die Landoberflächentemperatur als entscheidende, zeitlich äußerst dynamische Eingangsgröße (Li et al. 2009). Vor allem im alpinen Raum sind aber häufig nur wenige Aufnahmen verfügbar, wodurch die hohe zeitliche

Dynamik der Landoberflächentemperatur nicht adäquat abgebildet werden kann (Castelli et al. 2018).

Weltweit sind Verdunstungsmessungen im Vergleich zu Abfluss- und Niederschlagsmessungen rar. Allerdings gibt es globale Ambitionen in Form des FLUXNET-Netzwerks (Baldocchi et al. 2001), die vorhandenen Messungen zu synthetisieren und so ihre Verwendbarkeit zu verbessern. Protokolle bei der Messung und Auswertung der Rohdaten tragen zu einer besseren Vergleichbarkeit der Daten bei. Da die Daten der wissenschaftlichen Gemeinschaft frei zur Verfügung gestellt werden und Informationen aus unterschiedlichen Regionen, Klimaten und Biomen beinhalten (212 Messstationen weltweit mit insgesamt 1500 Messjahren), bieten sie eine einzigartige Datengrundlage für vergleichende Studien und die Entwicklung von überregional anwendbaren Modellen.

Netzwerke wie FLUXNET haben daher die Verwendung von machine learning (ML) oder deep learning (DL)-Methoden, welche auf große Datenmengen („big data“) angewiesen sind, für die Abschätzung von Stoff- und Energieflüssen angeregt (Papale und Valentini 2003; Yang et al. 2006; Jung, Reichstein, and Bondeau 2009). Während DL/ML-Methoden und statistische Verfahren generell oft aufgrund ihrer fehlenden Interpretierbarkeit als „black-box“-Methoden in der Kritik stehen, führen sie häufig zu einer besseren Simulation der Beobachtungen (Best et al. 2015). Kratzert et al. (2019a) zeigten beispielsweise, dass DL-basierte Methoden zu deutlich besseren Abflussmodellierungen im Vergleich zu etablierten prozessbasierten Modellen führen (siehe auch Kratzert et al. 2021, in diesem Heft).

Die FLUXCOM-Initiative, eine Kollaboration internationaler Forschungseinrichtungen, arbeitet an der Bereitstellung eines globalen Datensatzes täglicher Energie- und Kohlenstoffflüsse aufbauend auf DL/ML-Methoden sowie den Informationen von FLUXNET und ergänzenden Fernerkundungs- sowie weltweit verfügbaren meteorologischen Datensätzen (Jung et al. 2019). Die FLUXCOM-Modellierungen basieren auf einem Ensemble verschiedener DL/ML-Methoden (Tramontana et al. 2016). Bennet und Nijssen (2020) zeigten, dass die Integration eines neuronalen Netzwerks in das prozessbasierte hydrologische SUMMA-Modell (Clark

et al. 2015) zu einer Verbesserung der ET-Modellierung mit 30-Minuten-Auflösung führt. Beiden Ansätzen gemeinsam ist, dass sie (i) auf FLUXNET-Daten aufbauen und (ii) keine Zeitreihen der Eingangsgrößen für die Modellierung eines Zeitschrittes verwenden, sondern dass jeder Zeitschritt unabhängig von den vorherigen modelliert wird.

In vorliegender Studie wenden wir zwei unterschiedliche DL/ML-Methoden zur Abschätzung der Verdunstung an. Wir untersuchen die Eignung eines rekurrenten neuronalen Netzwerks, des Long Short-Term Memory-Netzwerks (LSTM; Gers et al. 1999; Hochreiter 1991; Hochreiter and Schmidhuber 1997) für die Modellierung der Verdunstung mit täglicher und 30-Minuten-Auflösung. Im Unterschied zu nicht rekurrenten neuronalen Netzwerken verfügt das LSTM über explizite Speicher, wodurch es sich vor allem für die Modellierung von Zeitreihen oder Systemen eignet, in denen Speicherprozesse eine wichtige Rolle spielen. Für die Modellierung eines Zeitschrittes bekommt das Modell Informationen über den zeitlichen Verlauf der Eingangsgrößen (beispielsweise Lufttemperatur und Niederschlag des letzten Jahres) und kann diese Informationen für längere Zeiträume speichern und zu einem späteren Zeitpunkt wieder darauf zurückgreifen. Im Bereich der Abflussmodellierung, bei der Speicherprozesse eine wesentliche Rolle spielen, wurden LSTMs bereits sehr erfolgreich eingesetzt (Kratzert et al. 2019a; Kratzert et al. 2018).

Betrachtet man ET als einen Energieaustausch zwischen Landoberfläche und Atmosphäre (latenter Wärmestrom), erscheinen Speicherprozesse hier weniger relevant. Betrachtet man ET aus Sicht des lokalen Wasserhaushalts, spielen Speicherprozesse beispielsweise in Form von Veränderungen der Bodenfeuchte aufgrund des Niederschlags eine entscheidende Rolle. Diese Studie untersucht daher, ob sich aus der Verwendung von LSTMs mit expliziten Speichern Vorteile gegenüber Modellen ergeben, die typischerweise nur mit Informationen des zu modellierenden Zeitschrittes verwenden. Hierfür stellen wir Modellvorhersagen des LSTM den Modellierungen mithilfe eines zweiten ML-Ansatzes, XGBoost, welcher ausschließlich Informationen für den zu modellierenden Zeitschritt bekommt, gegenüber. Die Ergebnisse beider Ansätze werden mit Beobachtungen der

Verdunstung mit 30-Minuten- sowie täglicher Auflösung verglichen, wodurch eine Bewertung der Modellgüte ermöglicht wird.

Als Datengrundlage für diese Studie dient der Datensatz FLUXNET2015 (Pastorello et al. 2020), aus welchem die meteorologischen Eingangsgrößen sowie Energieflüsse entnommen werden. Zusätzlich werden Satellitendaten für die Ermittlung statischer Gebietskenngrößen verwendet. Als Ausblick auf mögliche Anwendungen wird das auf FLUXNET-Daten trainierte Modell für eine österreichweite Modellierung der Verdunstung mithilfe von meteorologischen Zeitreihen des ECMWF ERA5-Land Reanalyse-Datensatzes (Muñoz Sabater 2019) angewendet und eine erste Einschätzung der Vertrauenswürdigkeit der so gewonnenen Abschätzungen präsentiert.

In vorliegender Studie steht die Modellierung der Verdunstung im Fokus. Die präsentierte Analyse ließe sich jedoch auf die Analyse anderer Energieflüsse (Nettostrahlung oder sensibler Wärmestrom) sowie Kohlenstoffflüsse erweitern.

2 Daten und Methode

2.1 Datengrundlage

Beide Modellierungsansätze, LSTMs and XGBoost, wurden auf lokaler Ebene für ausgewählte Messstationen des FLUXNET2015-Datensatzes (Pastorello et al. 2020) angewendet. Hierfür wurden dem FLUXNET2015-Datensatz die meteorologischen Eingangsgrößen sowie der latente Wärmestrom (entspricht der Verdunstung) entnommen.

Die Messungen des latenten Wärmestroms basieren auf EC-Messungen, welche eine mikrometeorologische Methode zur Abschätzung des Energie- und Masseaustausches zwischen Landoberfläche und Atmosphäre darstellen und für einen Bereich von einigen 100 m um die Messstation repräsentativ sind (Aubinet et al. 2012; Burba 2013). Die zeitliche Auflösung der Messungen beträgt 30 min. Zusätzlich zu diesen EC-Messungen der Wärmeströme werden ergänzende meteorologische Größen sowie Strahlungskomponenten mit gleicher zeitlicher Auflösung gemessen. Eine Liste der verwendeten Variablen findet sich im Anhang, Tab. 3. Die Qualität der Messungen ist für jeden Zeitschritt und jede Variable angegeben. Ausgehend von dieser Kennzeichnung

der Datenqualität wurden 153 der insgesamt 212 FLUXNET-Stationen in dieser Studie berücksichtigt. Abb. 1 zeigt die Lage der inkludierten Messstationen sowie eine Zuordnung der Stationen zu Vegetationsklassen (International Geosphere-Biosphere Programme (IGBP)-Klassifikation). Die Länge sowie die Zeiträume der Beobachtungen variieren für die ausgewählten Standorte zwischen 2 und 19 Jahren sowie zwischen 1994 und 2014. Für jede Zeitreihe wurden Zeitschritte mit schlechter Datenqualität für den latenten Wärmestrom (im Datensatz indiziert als QC=2) aus der Analyse ausgeschlossen. Für Modelleingangsgrößen wurden hingegen keine Qualitätskriterien festgesetzt.

Der FLUXNET2015-Datensatz weist zwar eine Zuordnung der Stationen zu IGBP-Vegetationsklassen auf, es fehlen allerdings Informationen zur zeitlichen Dynamik der Vegetation. Um den Einfluss der Vegetation auf die Verdunstung mit berücksichtigen zu können, wurde ergänzend eine Zeitreihe des Blattflächenindex mithilfe von Fernerkundungsdaten der MODIS-Mission erstellt (MOD15A2H; Myneni et al. 2015).

Zusätzlich zu diesen zeitlich dynamischen Eingangsgrößen wurden mithilfe von MODIS-Fernerkundungsdaten sowie mithilfe von Reanalysedaten (ERA5-Land, ECMWF) statische Gebietskenngrößen ermittelt. Diese statischen Gebietskenngrößen variieren zwischen den Stationen, bleiben jedoch über die gesamte Zeitreihe unverändert. Beispiele für diese statischen Gebietskenngrößen sind die mittlere Jahrestemperatur, mittlerer Jahresniederschlag, Geländehöhe etc. Eine vollständige Liste der verwendeten Kenngrößen ist Tab. 3 im Anhang zu entnehmen.

2.1.1 Vergleichsdaten zur Abschätzung der Modellgüte

Zur Abschätzung der Güte beider Modellansätze wurden uns tägliche ET-Modellierungen von Tramontana et al. (2016), welche die Grundlage für das FLUXCOM-Projekt darstellen, von den Autoren zur Verfügung gestellt (in weiterer Folge als FLUXCOM-Datensatz bezeichnet). Für das FLUXCOM-Projekt, das auf die räumlich verteilte Verdunstungsmodellierung abzielt, wurden zunächst DL/ML-Modelle mithilfe von FLUXNET-Daten trainiert. Die Vorhersage basiert hier auf einem Ensemble von 11 DL/ML-Methoden, die den vier großen Gruppen (i) Entscheidungs-

baum-Methoden, (ii) Regression-Splines, (iii) neuronale Netzwerke und (iv) Kernel-Methoden zugeordnet werden können. Diese Ensembledaten können zu einer groben Abschätzung der Modellgüte herangezogen werden. Hierbei gibt es jedoch zwei Limitationen: (i) Für die Modellierung von Tramontana et al. (2016) wurden zusätzlich zu den gemeinsamen meteorologischen noch andere fernerkundliche Eingangsgrößen verwendet, (ii) die Auswahl an FLUXNET-Stationen und Modellierungszeiträumen ist nicht ident mit der in dieser Studie getroffenen Auswahl, sodass nur überlappende Modellierungen für den Vergleich herangezogen werden können. Da es sich jedoch bei FLUXCOM um ein internationales, kollaboratives Forschungsprojekt handelt, wird es hier trotz der Limitationen als Vergleichsmodell herangezogen.

2.1.2 Daten für die flächige Anwendung in Österreich

Als Ausblick auf mögliche zukünftige Anwendungen, wird das auf den FLUXNET2015-Datensatz trainierte LSTM für eine räumlich verteilte Abschätzung der Verdunstung für Österreich angewendet. Da für das Training der Modelle lokale meteorologische Daten der FLUXNET-Stationen verwendet wurden, müssen die meteorologischen Informationen für eine Anwendung in Österreich aus anderen Quellen bezogen werden. In dieser Studie verwenden wir die Daten des ERA5-Land-Datensatzes (Muñoz Sabater 2019, S. 5), welcher vom Europäischen Zentrum für mittelfristige Wettervorhersage (ECMWF) herausgegeben wird. Dieser enthält die notwendigen meteorologischen Modelleingangsgrößen mit einer räumlichen Auflösung von ca. 9 km und einer stündlichen zeitlichen Auflösung. In dieser Analyse wurde die Verdunstung mit stündlicher Auflösung für den Zeitraum 2001 bis 2019 vorhergesagt. Die statischen Gebietskenngrößen wurden aus denselben Daten wie für die Anwendungen an den FLUXNET-Standorten berechnet (MODIS und ERA5-Land).

2.2 Modelle

2.2.1 Long-Short-Term-Memory (LSTM)-Netzwerk

Das Long Short-Term Memory (LSTM)-Netzwerk (Hochreiter 1991; Hochreiter and Schmidhuber 1997; Gers et al. 1999)

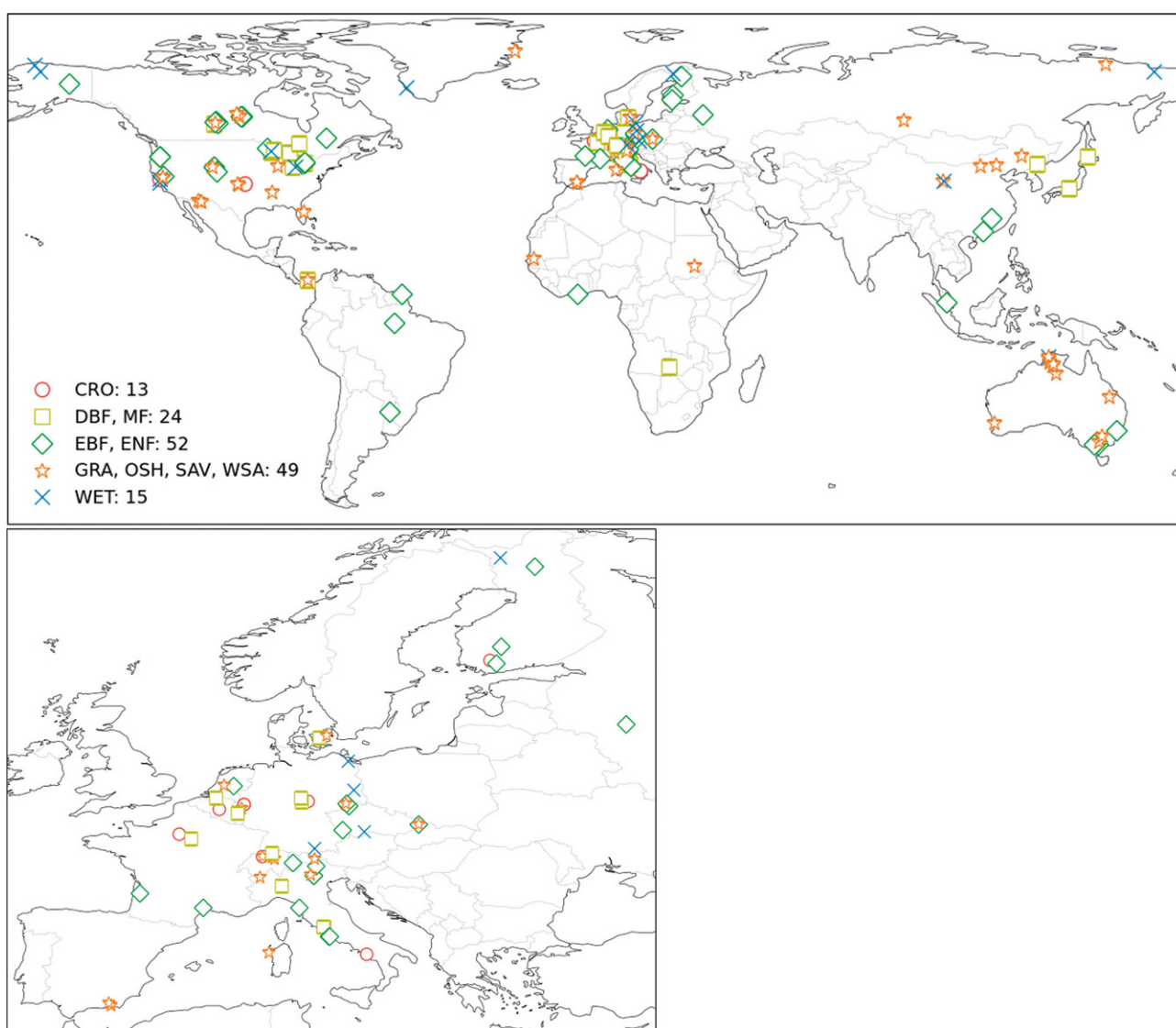


Abb. 1 Übersicht über die Lage der untersuchten FLUXNET-Standorte inklusive Angabe der IGBP-Vegetationsklasse (CRO Agrarflächen, DBF, MF Laub- und Mischwald, EBF, ENF Immergrüne Laub- und Nadelwälder, GRA, OSH, SAV, WSA Grasland, Steppe, Savanne; WET Permanente Feuchtgebiete). Diese wurden in der Legende zu Gruppen mit ähnlichen Eigenschaften gruppiert. Die Zahlen in der Legende entsprechen der Anzahl der verwendeten Stationen pro Gruppe von Vegetationsklassen. Ein Ausschnitt über Europa ist aufgrund der großen Anzahl von Standorten zusätzlich vergrößert dargestellt

ist ein rekurrentes neuronales Netzwerk (RNN). Diese RNN-Netzwerke zeichnen sich dadurch aus, dass sie die sequenzielle Reihenfolge (in diesem Fall die zeitliche Reihenfolge) der Eingangsdaten berücksichtigen. Des Weiteren verfügen LSTMs über Speicher (cell state), in denen Informationen über längere Zeiträume gespeichert werden können. Beispielfähig könnte Niederschlag einen Speicher füllen, welcher dann über einen längeren Zeitraum durch Verdunstung und Perkolatation entleert werden kann. Während DL/ML-Techniken allgemein als „black box“-Modelle oftmals für ihre mangelnde

physikalische Interpretierbarkeit in der Kritik stehen, konnten Kratzert et al. (2019b) zeigen, dass sich im Falle der Abflussmodellierung LSTM-Speicher ausbilden, die stark mit physikalischen Speichern korrelieren. Eine detaillierte Beschreibung des Modells sowie eine hydrologische Interpretation finden sich in Kratzert et al. (2018).

Für die vorliegende Studie wurde ein Multi-Timescale LSTM (MTS-LSTM, Gauch et al. 2020) angewendet, welches die gleichzeitige Vorhersage der Zielvariable(n) in mehreren zeitlichen Auflösungen ermöglicht. In dieser Studie wurde das LSTM für 30-

min- und täglichen Vorhersagen trainiert. Ein entscheidender Parameter bei LSTMs ist die Sequenzlänge; die Anzahl an Zeitschritten, die das LSTM zur Vorhersage eines bestimmten Zeitschritts als Eingangsdaten bekommt. Beim MTS-LSTM kann die Sequenzlänge für die unterschiedlichen zeitlichen Auflösungen variieren. Auf 30-min-Auflösung wurde eine Sequenzlänge von 336 Zeitschritten (1 Woche, 48 30-min-Zeitschritte pro Tag * 7 = 336) und auf täglicher Auflösung eine Sequenzlänge von 365 Zeitschritten (1 Jahr) gewählt.

Des Weiteren muss der Anwender eine Reihe von Hyperparametern vor

Beginn des Trainings setzen. Für diese Studie wurden Hyperparameter übernommen, die in einer anderen Studie mit einem ähnlichen Datensatz ermittelt wurden. Eine Liste der verwendeten Werte findet sich im Anhang, Tab. 4.

2.2.2 Gradient Boosting Regression (XGBoost)

XGBoost (eXtreme Gradient Boosting, Chen and Guestrin 2016) ist ein regelbasierter ML-Ansatz, der auf Ensembles von Entscheidungsbäumen aufbaut. Bei XGBoost werden Entscheidungsbäume sequenziell trainiert, sodass ein darauffolgender Baum den Fehler des vorherigen korrigiert. Die endgültige Vorhersage entspricht schlussendlich der Summe aller Bäume (Friedman 2001). Im Gegensatz zu LSTMs wurde XGBoost nicht explizit für die Vorhersage von Zeitreihen entwickelt. Übergibt man XGBoost eine Zeitreihe an Eingangsgrößen, so werden diese ohne Berücksichtigung ihrer zeitlichen Folge als unabhängige Eingangsgrößen interpretiert. Nichtsdestotrotz wurde XGBoost bereits erfolgreich zur Zeitreihenmodellierung angewendet (Gauch et al. 2021; Luo et al. 2019).

In dieser Studie wird XGBoost ausschließlich mit Daten für den zu modellierenden Zeitschritt trainiert. Hierfür gibt es drei Gründe: (i) durch den Vergleich des LSTM, welches Zeitreihen der Eingangsgrößen für das Training bekommt, und XGBoost ohne Informationen über den zeitlichen Verlauf der Eingangsgrößen in der Vergangenheit, soll der Vorteil durch Berücksichtigung von Speichereffekten herausgestrichen werden. Die Hypothese hierbei ist, dass beide Modelle hoch performant sind und dass eine bessere Vorhersage durch das LSTM einen entscheidenden Effekt von Speichereffekten auf die Verdunstungsvorhersage nahelegt. (ii) bei XGBoost führt die Verwendung von langen Zeitreihen der Eingangsgrößen, z.B. 365 Tage, zu einer großen Anzahl von Prädiktoren. Dies kann zu einer Verringerung der Modellgüte führen, da es für das Modell schwieriger

wird, die wichtigsten Einflussgrößen zu identifizieren (Gauch et al. 2021). Es sei erwähnt, dass die Verwendung von kurzen Zeitreihen (einige Zeitschritte) möglicherweise auch für XGBoost zu einer besseren Vorhersage führen könnte. Die Untersuchung, ob dies tatsächlich der Fall wäre, steht hier jedoch nicht im Fokus der Analyse. Allerdings könnte dies in einer weiterführenden Studie untersucht werden, falls sich herausstellt, dass Speichereffekte wichtig sind. (iii) Gängige Verdunstungsmodelle, wie Energiebilanzmodelle oder Penman-Monteith-Ansätze, schätzen die Verdunstung als einen unmittelbaren Prozess, der durch die Energiebilanz der Oberfläche ohne Berücksichtigung von Speichereffekten reguliert wird (Monteith 1965; Kalma et al. 2008). Die ausschließliche Verwendung von Informationen des zu bestimmenden Zeitschritts entspricht daher einer üblichen Herangehensweise bei der Abschätzung der Verdunstung.

Die verwendeten dynamischen und statischen Eingangsgrößen entsprechen denselben, die auch für das LSTM verwendet wurden (siehe Anhang, Tab. 3). Zusätzlich zu diesen Eingangsgrößen erhält XGBoost noch Informationen zum Kalendertag sowie der Stunde des Tages. Während das LSTM auf 30-min und tägliche Flüsse simultan trainiert wurde, wurden die beiden zeitlichen Auflösungen für XGBoost separat trainiert (2 unabhängige Modelle).

Auch für XGBoost ist eine Reihe von Hyperparametern zu setzen. In dieser Studien wurden diese mithilfe einer Bayes'schen Hyperparametersuche (Bergstra et al. 2013) ermittelt. Die vordefinierten Wertebereiche für die Hyperparametersuche sind im Anhang, Tab. 5 gelistet.

2.3 Training und Evaluierung

Ziel dieser Studie ist es, eine Aussage über die Modellgüte an Standorten zu treffen, an denen keine Verdunstungsmessungen zur Verfügung stehen (out-of-sample Evaluierung). Daher wurde der Datensatz für eine Kreuzvalidierung

in 10 Sets mit jeweils annähernd gleich vielen Daten unterteilt, wobei einzelne FLUXNET-Stationen jeweils zur Gänze einem Set zugeordnet wurden. In 10 Durchläufen wurde jeweils 1 Set zur Seite gelegt und die anderen 9 Sets wurden für das Training verwendet. Das trainierte Modell wurde anschließend zur Vorhersage der Verdunstung an den Stationen des 10. Sets verwendet.

Da für XGBoost eine Hyperparametersuche durchgeführt wurde, wurden die 9 Sets weiter in einen Trainings- und Validierungsdatensatz aufgeteilt; hierfür wurden 80 % der Daten für das Training und 20 % der Daten (zufällig gezogen) als Validierungsset verwendet. Für die Hyperparametersuche wurden für jedes Set jeweils 6 Läufe mit unterschiedlichen zufälligen Initialisierungen und 60 Iterationen durchgeführt. Für das finale Training wurden für jedes Set die besten 8 Hyperparameterkombinationen für 8 XGBoost-Läufe verwendet. Die finale Vorhersage entspricht einem Ensemble in Form des Mittelwerts über diese 8 Läufe. Im Falle des LSTM entspricht die finale Vorhersage einem Ensemble von 8 LSTM-Läufen mit unterschiedlichen zufälligen Initialisierungen (seeds).

3 Ergebnisse und Diskussion

3.1 Analyse der Modellgüte beider Modellansätze

Tab. 1 zeigt verschiedene Kennzahlen zur Bewertung der Modellgüte für beide Modellansätze, LSTM und XGBoost, und beide zeitlichen Auflösungen. Es sei hier nochmals erwähnt, dass die Modelle out-of-sample in einer Kreuzvalidierung evaluiert wurden. Jeweils 10 % der Daten (ganze FLUXNET-Stationen) wurden zur Seite gelegt und die restlichen 90 % für das Training der DL/ML-Ansätze verwendet. Das so trainierte Modell wurde dann für die Vorhersage an den verbleibenden Stationen (10%) verwendet. So ist sichergestellt, dass keine Verdunstungsinformationen dieser Stationen in das Training einfließen. Dadurch lässt

Tab. 1 Median der Gütemaße für beide Modelle, LSTM und XGBoost (out-of-sample-Evaluierung)

Gütemaß	LSTM		XGBoost	
	Täglich	Halbstündlich	Täglich	Halbstündlich
NSE (-)	0,73	0,77	0,70	0,75
RMSE ($W\ m^{-2};\ mm$)	15,1; 0,5	29,1; 1,0	16,9; 0,6	32,4; 1,1
Bias ($W\ m^{-2};\ mm$)	-0,5; -0,02	-0,7; -0,02	0,6; 0,02	1,2; 0,04
Pearson-Korrelationskoeffizient (-)	0,90	0,90	0,88	0,89

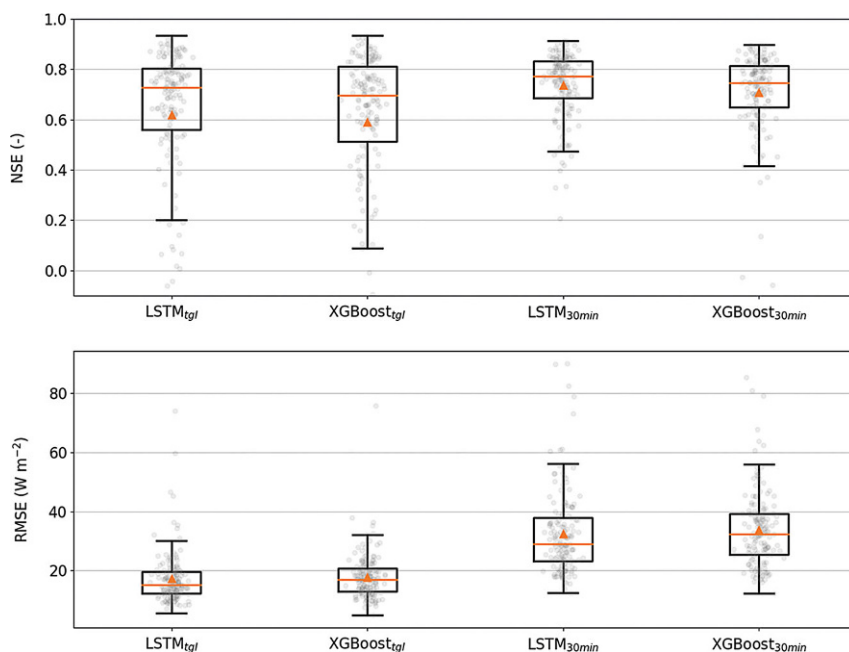


Abb. 2 Boxplots des NSE (*oben*) und RMSE (*unten*) für beide Modelle und tägliche (tgl) und halbstündliche zeitliche Auflösung. Für den NSE wurde die y-Achse auf den Bereich [0, 1] limitiert. Die Dreiecke entsprechen dem Mittelwert

sich abschätzen, wie gut die Modelle die Verdunstung an Standorten ohne Verdunstungsmessungen vorhersagen können.

Die hier gezeigten Werte entsprechen dem Median über die 153 analysierten FLUXNET-Stationen. Es ist dabei zu beachten, dass die Länge der Beobachtungszeitreihe zwischen den Stationen variiert. In der Bewertung der Modellgüte bekommt jedoch jede Station das gleiche Gewicht.

Tab. 1 ist zu entnehmen, dass das LSTM konsistent bessere Werte für alle gelisteten Gütemaße aufweist (mit p -Werten für alle Gütemaße $< 0,01$ für den Wilcoxon-Test, dass die Gütemaße für alle 153 FLUXNET-Stationen der gleichen Verteilung angehören). Abb. 2 zeigt die Verteilung des NSE sowie RMSE in der Form von Boxplots.

Von den 153 analysierten Stationen weisen für das LSTM 6 (0), für XGBoost 9 (2) negative NSE-Werte für tägliche (halbstündliche) Verdunstungsraten auf. Dies bedeutet, dass der Mittelwert der Beobachtungen einen besseren Schätzer darstellt. Es gilt jedoch zu bedenken, dass diese Studie die Güte von Vorhersagen an Standorten ohne Verdunstungsbeobachtungen untersucht, also ohne Informationen zu mittleren Beobachtungen. Zwar stehen hier für alle simulierten FLUXNET-Stationen Verdunstungsda-

ten für die Validierung der Vorhersagen zur Verfügung, diese werden allerdings im Training nicht berücksichtigt. Daher kann das Modell per se an jedem beliebigen Standort angewendet werden, sofern die notwendigen Eingangsgrößen vorhanden sind.

Abb. 2 zeigt, dass der NSE für halbstündliche Verdunstungsraten besser ist als für tägliche Werte und dass für den RMSE genau das Gegenteil gilt. Dies lässt sich durch die höhere Varianz in den halbstündlichen Daten erklären. Durch die höheren Absolutwerte der 30-min-Daten im Vergleich zu den täglichen Mitteln ergeben sich zu einem höheren Abweichungen zwischen Beobachtung und Simulation; zum anderen ist der Mittelwert der Beobachtungen in diesem Fall ein schlechterer Schätzer, wodurch sich der NSE der Simulation erhöht.

Um die Güte beider Modellansätze besser einordnen zu können, wurden die täglichen Vergleichsdaten des FLUXCOM-Projekts analysiert. 67 der 153 Stationen sind in beiden Datensätzen enthalten, jedoch teilweise für unterschiedliche Zeiträume. Daher wurden die Gütemaße für jene Zeitschritte, für die alle Modelle Vorhersagen bieten, neu berechnet. Das LSTM und XGBoost weisen für diese Teilmenge einen NSE von 0,78 und 0,79 auf und sind somit signifikant besser als die Vorhersage des

FLUXCOM-Datensatzes mit einem NSE von 0,72 (p -Wert $< 1^{-20}$). Im Vergleich zu dem gesamten analysierten Datensatz weist XGBoost in dieser Teilmenge eine höhere Modellgüte als das LSTM auf. Die bessere Modellgüte von XGBoost lässt sich auf die Auswahl der Stationen zurückführen. In diesen 67 Stationen sind prozentuell weniger Standorte mit Vegetationsklassen Savanne, Steppe und Grasland enthalten, für welche das LSTM – wie im Anschluss aufgezeigt wird – bessere Werte als XGBoost liefert.

Bei der Interpretation der Ergebnisse ist zu bedenken, dass sowohl die meteorologischen Modelleingangsgrößen als auch die Verdunstung selbst mit Fehlern behaftet sind. Im FLUXNET2015-Datensatz ist jeder Zeitschritt und jede Variable bezüglich ihrer Qualität mit einem Wert zwischen null (gute Qualität) bis zwei (schlechte bzw. unklare Qualität) gekennzeichnet. In Bezug auf die EC-Messung lässt sich sagen, dass die Verdunstungsmessung per se, also auch bei guter Datenqualität (null), eine Ungenauigkeit von ca. 5 bis 20% aufweist (Foken 2008). Diese Unsicherheiten in der Zielgröße beeinflussen naturgemäß auch die Modellierung; der Effekt ist jedoch schwer zu quantifizieren. Bezüglich der meteorologischen Eingangsgrößen konnte in dieser Studie interessanterweise kein Zusammenhang zwischen der Häufigkeit von schlechter Datenqualität und der Modellgüte für einen Standort festgestellt werden.

Basierend auf dem Vergleich der Modelle lässt sich sagen, dass beide Ansätze zufriedenstellende Ergebnisse liefern und eine bessere Übereinstimmung mit den Beobachtungen als der Referenzdatensatz aufweisen. Leider gibt es keine veröffentlichten, frei zugänglichen Daten von physikalisch-basierten Verdunstungsmodellen, die diesen mit statistischen Verfahren modellierten Daten gegenübergestellt werden könnten. Der Aufbau einer Datenbasis zur transparenten Evaluierung von Modellen aufbauend auf gleichen Referenzdaten, in diesem Fall des FLUXNET2015-Datensatzes, ist daher eine wichtige zukünftige Aufgabe der wissenschaftlichen Gemeinschaft.

3.2 Einfluss der Vegetationsklassen auf die Modellgüte

Tab. 2 zeigt den Median NSE per Vegetationsklasse für beide Modelle und zeit-

Tab. 2 Median-NSE für beide Modelle und beide zeitliche Auflösungen per Vegetationsklasse. Die Zahlen in Klammern geben die Anzahl der untersuchten FLUXNET-Stationen je Vegetationsklasse an

Vegetationsklasse (IGBP)	LSTM _{1gl}	XGBoost _{1gl}	LSTM _{30min}	XGB _{30min}
Agrarflächen (CRO, 13)	0,64	0,60	0,76	0,72
Laubwald (DBF, 17)	0,78	0,76	0,82	0,81
Immergrüner Laubwald (EBF, 12)	0,62	0,41	0,76	0,69
Immergrüner Nadelwald (ENF, 40)	0,76	0,74	0,75	0,73
Grasland (GRA, 30)	0,75	0,68	0,81	0,73
Mischwald (MF, 7)	0,71	0,77	0,70	0,76
Steppe (OSH, 10)	0,63	0,58	0,71	0,71
Savanne (SAV, 4)	0,80	0,71	0,84	0,78
Hölzerne Savanne (WSA, 5)	0,72	0,66	0,75	0,72
Feuchtgebiete (WET, 15)	0,73	0,75	0,77	0,77

lichen Auflösungen. Hierbei ist zu beachten, dass die Anzahl der FLUXNET-Stationen pro Vegetationsklasse unterschiedlich ist. Für Vegetationsklassen mit nur einigen wenigen Stationen kann der Median NSE leichter durch einzelne besonders gut bzw. schlecht vorhergesagte Stationen beeinflusst sein.

Tramontana et al. (2016) zeigten, dass die Modellgüte für die einzelnen Vegetationsklassen stark variiert. Die beste Übereinstimmung mit den Beobachtungen fanden sie für Misch- und Laubwälder (MF, DBF), die schlechtesten für immergrüne Laubwälder, Agrarflächen sowie Stationen an extrem trockenen oder kalten Standorten. Genau wie XGBoost in dieser Studie, wurden die Modelle in Tramontana et al. (2016) ausschließlich mit Informationen für den zu bestimmenden Zeitpunkt trainiert. Es ist daher interessant zu sehen, dass XGBoost für die gleichen Vegetationsklassen, Misch- und Laubwald, die besten Modellergebnisse liefert und vor allem die Vorhersage der Verdunstung für immergrüne Laubwälder vergleichsweise größere Abweichungen von den Beobachtungen aufweist.

Im Vergleich hierzu liefert das LSTM für immergrüne Laubwälder deutlich bessere NSE-Werte. Generell ist zu erkennen, dass das LSTM bei allen Klassen außer Mischwald und permanenten Feuchtgebieten höhere Median-NSE-Werte aufweist. Interessanterweise ist die Vorhersage vor allem an tendenziell trockenen, warmen Standorten mit kurzer Vegetation wie Savannen, Steppen und Grasland im Vergleich zu XGBoost besser.

Für beide Modelle gilt, dass Standorte mit höheren mittleren Jahrestemperaturen schwerer vorherzusagen sind als kältere Standorte. Vergleicht man die NSE-Werte von 20 % der wärmsten und kältesten Standorte im Datensatz

sieht man, dass das LSTM einen besseren Median-NSE für die wärmsten Standorte aufweist (tägliche Werte: 0,64 für LSTM, 0,59 für XGBoost). Im Gegenzug weist XGBoost für die kältesten Standorte bessere Median-NSE-Werte auf (tägliche Werte: 0,80 für LSTM, 0,82 für XGBoost). Daraus lässt sich schließen, dass an warmen Standorten sowie an Standorten mit kurzer Vegetation, Speicherprozesse eine wichtigere Rolle für die Verdunstungsvorhersage spielen könnten. Diese Beobachtung passt gut mit Überlegungen zum Einfluss der Wasserverfügbarkeit auf den Verdunstungsprozess überein; in humiden Gebieten mit gemäßigten Temperaturen und ausreichend Wasservorkommen dominiert meist die Strahlung die Verdunstung. Daher ist die Verdunstung unter diesen Umständen eine direkte Reaktion auf die vorhandene Strahlung, sodass Speicherprozesse eine geringere Bedeutung spielen. Es macht daher aufgrund theoretischer Überlegungen Sinn, dass XGBoost ohne Informationen über den zeitlichen Verlauf der Eingangsgrößen in permanenten Feuchtgebieten bessere Ergebnisse liefert als das LSTM, da das LSTM irrelevante Informationen aus der Vergangenheit für die Vorhersage berücksichtigt. In wärmeren, oft semiariden Gebieten ist hingegen häufig ausreichend Energie in Form von Sonnenstrahlung vorhanden, jedoch fehlen die notwendigen Wasserressourcen für die Verdunstung. Unter diesen Umständen spielen Speichereffekte eine entscheidendere Rolle. Bei immergrünen Laubwäldern scheint das Fehlen eines ausgeprägten Jahresgangs die Vorhersage generell zu erschweren, da ein grundlegendes zeitliches Muster fehlt und die Tag-zu-Tag-Variation vergleichsweise groß ist (Yebra et al. 2015; Yuan et al. 2010). Allerdings scheint das

LSTM dieses zeitliche Muster besser abbilden zu können.

Da sich zeigt, dass beide Modelle unter verschiedenen klimatischen und Vegetationsbedingungen im Vorteil sind, erscheint die Kombination beider Modellierungsansätze in der Form (i) eines einfachen Ensembles oder (ii) eines gewichteten Ensembles, bei dem die Gewichtung der beiden Ansätze beispielsweise durch ein Regressionsmodell aufgrund der Vegetationsklasse geschätzt wird, oder (iii) durch Anwendung jedes Modells an für das Modell geeigneten Standorten (ebenfalls beispielsweise aufgrund der Vegetationsklasse) als sinnvoller Schritt weiterer Forschung.

Da die Analyse nahelegt, dass für viele Vegetationsklassen Speichereffekte eine Rolle spielen könnten, wäre für XGBoost in weiterer Folge zu untersuchen, ob die Hinzunahme von Informationen zum zeitlichen Verlauf der Eingangsgrößen zu einer Verbesserung der Modellgüte führt. Da XGBoost den zeitlichen Zusammenhang jedoch nicht explizit darstellen kann, sondern alle Zeitschritte als unabhängige Prädiktoren interpretiert, führt die Verwendung langer Zeitreihen aufgrund der großen Anzahl an Prädiktoren zu einer Reduktion der Modellgüte. Hier sind somit modellspezifische Grenzen gesetzt. Nichtsdestotrotz könnten – in Analogie zum LSTM – Informationen kurzer Sequenzen, beispielsweise der letzten 5 Zeitschritte, verwendet werden. Allerdings bieten die letzten 5 Zeitschritte bei Modellierung mit 30-min-Auflösung nur einen Überblick über die letzten 2,5 h. Daher könnten alternativ Informationen über vergangene Zeitschritte in Form eines Mittelwerts beispielsweise über die letzten 5 Tage aggregiert werden. Dadurch könnte zum einen die Anzahl der Prädiktoren

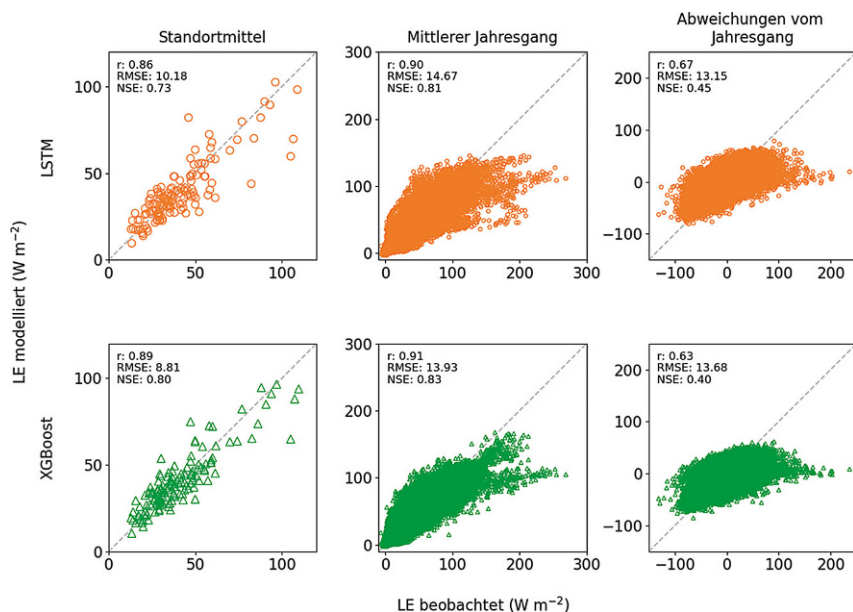


Abb. 3 Darstellung des Standortmittelwerts, des mittleren Jahrgangs pro FLUXNET-Station sowie der Abweichungen vom mittleren Jahrgang. In der Darstellung des mittleren Jahrgangs repräsentiert jeder Datenpunkt die Verdunstung für einen bestimmten Kalendertag pro Standort (maximal 366 Werte pro Standort). In der Darstellung der Abweichungen vom Jahrgang repräsentiert jeder Datenpunkt die Differenz der tatsächlichen täglichen Verdunstung zu dem Mittel für den jeweiligen Kalendertag

niedriger gehalten werden und zum anderen möglicherweise eine bessere Beschreibung des Systemzustands erzielt werden.

Eine weitere Möglichkeit zur Verbesserung der Modellgüte wäre das Training unterschiedlicher Modelle für einzelne Vegetations- und Klimabedingungen, beispielsweise eines für Agrarflächen, Laubwälder, Trockenstandorte etc. Für das LSTM hat sich jedoch gezeigt, dass eine solche Vorgehensweise zu einer Reduktion der Modellgüte führt. Hierbei spielt die Verringerung der vorhandenen Trainingsdaten eine entscheidende Rolle.

3.3 Vorhersage von Standortmittelwert, Jahrgang und Anomalien

Für die folgende Analyse wurde die Verdunstungszeitreihe in mehrere Komponenten zerlegt; für jede FLUXNET-Station wurde (i) ein mittlerer Jahrgang der Verdunstung, genauer eine mittlere Verdunstung für jeden Kalendertag berechnet. Hierbei wurden ausschließlich Kalendertage berücksichtigt, für die mindestens zwei Modellvorhersagen zur Verfügung stehen. Darauf aufbauend wurden (ii) die täglichen Abweichungen von diesem mittleren Jahrgang berechnet. Falls für eine Station

mehr als 180 Tage einen Wert für den mittleren Jahrgang aufwies, wurde zusätzlich (iii) ein Standortmittelwert als Mittel des Jahrgangs berechnet. Für 12 Stationen konnte aufgrund dieser Kriterien kein Standortmittelwert berechnet werden. Dabei handelt es sich vor allem um FLUXNET-Stationen mit kurzen Zeitreihen sowie kalten Standorten, die häufig aufgrund der Datenqualität im Winter keine Beobachtungen liefern.

Abb. 3 zeigt die Standortmittelwerte, die mittlere Verdunstung pro Kalendertag sowie die Abweichungen vom Kalendertagmittel für das LSTM (oben) sowie XGBoost (unten). Es ist zu sehen, dass beide Modelle das Standortmittel sowie den mittleren Jahrgang relativ gut abbilden (mit Pearson-Korrelationskoeffizienten von $>0,86$ und einem $NSE > 0,73$). Interessanterweise erzielt für beide Größen XGBoost bessere Werte als das LSTM. Daher ist in weiteren Untersuchungen zu überprüfen, ob Anpassungen des LSTM zu einer Verbesserung der Vorhersage des Standortmittels führen. Eine Möglichkeit hierfür wäre die Regularisierung des Modells, sodass Abweichungen des modellierten vom beobachteten Standortmittel in der Berechnung der Verlustfunktion mitberücksichtigt werden.

Für den mittleren Jahrgang zeigt sich eine Tendenz zur Unterschätzung hoher Flüsse. Dieser Effekt verdeutlicht sich bei den täglichen Abweichungen vom mittleren Jahrgang. Beide Modelle zeigen eine deutlich reduzierte Variationsbreite im Vergleich zu den Beobachtungen. Allerdings zeigt hier das LSTM bessere Werte für alle Modellgütemaße.

In diesem Zusammenhang gilt es in Zukunft näher zu untersuchen, inwieweit Speichereffekte diese Abweichungen vom mittleren Jahrgang beeinflussen. Des Weiteren ist zu untersuchen, welche ergänzenden Eingangsgrößen zur Beschreibung dieser Abweichungen geeignet bzw. notwendig wären. So fehlen beispielsweise in den verwendeten Eingangsgrößen Informationen zur Stabilität der Atmosphäre, die für die Beschreibung der Austauschkoefizienten von Wärme und Wasserdampf in der Atmosphäre relevant sind (Monin and Obukhov 1954). Allerdings werden diese Informationen zur Stabilität, beispielsweise die Obukhovlänge, nicht standardmäßig erhoben bzw. in Datensätzen gelistet.

4 Ausblick: räumlich verteilte Anwendung für Österreich

Die Modelle wurden out-of-sample, also ohne die Verwendung von Daten des zu simulierenden Standorts, trainiert (im Gegensatz dazu werden Niederschlag-Abfluss-Modelle oft in-sample trainiert, da Informationen über den Abfluss am zu simulierenden Standort für die Kalibrierung des Modells verwendet werden). Durch dieses out-of-sample-Training zeigt sich, dass an einem Standort Erlerntes zur Vorhersage an einem anderen Standort genutzt werden kann. Als Ausblick auf mögliche Anwendungen der erlernten Beziehungen zwischen Eingangsgrößen und der Verdunstung, wird das mithilfe der FLUXNET-Daten trainierte LSTM für eine flächige Anwendung für Österreich genutzt.

Hierbei ist jedoch auf einige Limitationen hinzuweisen, die für die Interpretation der Ergebnisse entscheidend sind:

- (i) Viele der verwendeten FLUXNET-Stationen befinden sich an Standorten, die sich aufgrund klimatischer aber auch Vegetationscharakteristika stark von den Gegebenheiten in Österreich (gemäßigte Breiten, alpiner Raum) unterschei-

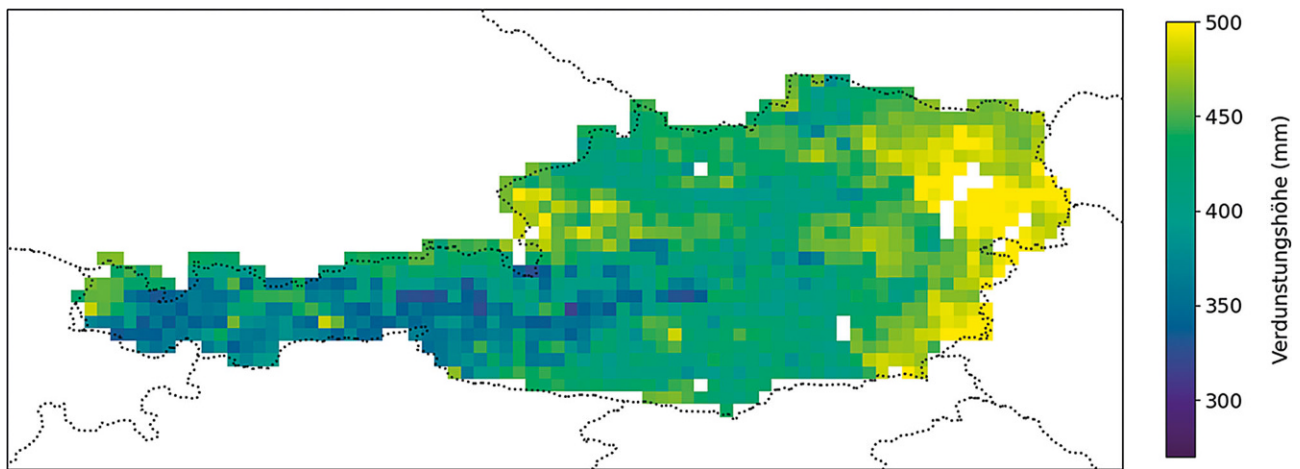


Abb. 4 Mittlere jährliche LSTM-basierte Verdunstungshöhe in mm. Diese Werte stellen ein Mittel über die Jahre 2002 bis 2019 dar

den. Ausschließlich eine der 153 verwendeten FLUXNET-Stationen (AT-Neu, in Neustift im Stubaital) befindet sich in Österreich. Das Modell ist somit nicht speziell an den Standort Österreich angepasst. So befinden sich beispielsweise nur 30 der 153 Stationen auf einer Höhe über 1000 m über Meeresebene, wovon sich zudem ein Großteil außerhalb Europas befindet und somit nicht zwangsläufig charakteristisch für den österreichischen Alpenraum ist.

- (ii) Da das LSTM mit lokalen Daten von FLUXNET-Stationen trainiert wurde, welche über die ganze Welt verteilt sind, müssen die Eingangsgrößen für die Anwendung in Österreich mithilfe eines anderen Datensatzes ermittelt werden (wie im Abschn. 2.1.2 „Daten für die flächige Anwendung in Österreich“ beschrieben). Dieser Wechsel des Datensatzes ist nicht unproblematisch, da DL/ML-Methoden datensatzspezifische Fehler implizit korrigieren. Positiv zu erwähnen ist hier, dass die FLUXNET2015-Daten keinen einheitlichen Datensatz im Sinne systematischer Fehler darstellen. Fehler im für Österreich verwendeten ERA5-Land-Datensatz gehen jedoch direkt in die Verdunstungsvorhersage ein.
- (iii) Die Validierung der geschätzten Verdunstung ist schwierig, da wie beschrieben keine Messdaten auf dieser räumlichen Skala vorhanden sind. Vorhandene Lysimeterdaten können zu einer Einschätzung der Modellgüte herangezogen werden. Allerdings ist hier aufgrund

des großen räumlichen Skalennunterschieds bei einem direkten Vergleich Vorsicht geboten.

Unter Berücksichtigung dieser Einschränkungen ist dieser Abschnitt als ein erster Versuch einer Anwendung zu werten, der jedoch in weiterer Folge noch umfassender Analysen und Weiterentwicklungen bedarf. Ziel ist es zu zeigen, was passiert, wenn das global trainierte Modell zur Abschätzung der Verdunstung mithilfe anderer Daten (ERA5-Land-Rasterdaten mit 9 km Auflösung statt lokaler Messdaten) in einem beliebigen Gebiet angewendet wird. Die hier präsentierten Ergebnisse sind somit als Ausgangspunkt für die Weiterentwicklung des DL/ML-Ansatzes für die räumlich verteilte Verdunstungsabschätzung speziell für Österreich zu werten.

Die Ergebnisse werden hier ausschließlich für das LSTM präsentiert, da es eine bessere Modellgüte für den FLUXNET2015-Datensatz aufweist. Des Weiteren ist ein Vergleich beider DL/ML-Modell hier nicht zielführend, da wie bereits erwähnt keine entsprechenden Referenzdaten vorhanden sind. Der Vergleich liegt hier außerdem nicht im Fokus, da ausschließlich die Richtung für weitere Forschung hin zur Abschätzung der Verdunstung über das gesamte Bundesgebiet mithilfe von DL/ML-Methoden dargestellt werden soll.

Abb. 4 zeigt die vorhergesagte mittlere jährliche Verdunstungshöhe. Zu sehen ist, dass die Verdunstung im alpinen Raum deutlich geringer ist als im Rest Österreichs und dass vor allem im Voralpenland in Oberösterreich sowie

im Burgenland, der Region um Wien, dem Weinviertel und der Oststeiermark vergleichsweise höhere Verdunstungsraten auftreten. Dieses räumliche Muster stimmt gut mit Modellierungen mithilfe des COSERO-Modells, eines hydrologischen Niederschlag-Abfluss-Modells zur Bestimmung der Wasserbilanz in Österreich überein (Kling 2006). Eine Ausnahme bilden hier die Region um Wien und das Weinviertel, wo das LSTM verhältnismäßig höhere Verdunstungswerte aufweist. In seiner Arbeit zur Wasserbilanz Österreichs schätzt Kling (2006) die jährliche Verdunstung auf 510 mm (für den Zeitraum 1961 bis 1990). Mit 427 mm liegt dieser Wert für das LSTM deutlich niedriger; die Verdunstung wird somit im Vergleich zu Kling (2006) unterschätzt. Die Verdunstungswerte sind dabei flächendeckend, mit Ausnahme der Region um Wien und das Weinviertel sowie den alpinen Raum, für das LSTM geringer als bei Kling (2006). Die Gründe für diese Unterschiede in den jährlichen Verdunstungssummen sind in weiterer Folge zu untersuchen.

Hierbei ist beispielsweise die Qualität und grobe Auflösung der Eingangsdaten (ERA5-Land) zu berücksichtigen. So hat die grobe räumliche Auflösung zur Folge, dass die meteorologischen Felder stark geglättet werden, was beispielsweise dazu führt, dass Windgeschwindigkeiten unterschätzt werden. Eine Wiederholung der Analyse mit für Österreich genaueren meteorologischen Eingangsdaten, wie beispielsweise dem INCA-Datensatz der Zentralanstalt für Meteorologie und Geodynamik (ZAMG; Haiden et al. 2011), wäre daher in diesem Zusammenhang interessant.

Auch die Qualität der aus Fernerkundungsdaten gewonnenen statischen Gebietskenngrößen ist aufgrund der im alpinen Raum häufigen Wolkenbedeckung und der damit verbundenen geringeren Qualität der Fernerkundungsprodukte genauer zu untersuchen. Mögliche weitere Schritte für die Anpassung des Modells für die flächige Verdunstungsvorhersage in Österreich sind die Hinzunahme von ergänzenden Verdunstungsmessungen innerhalb Österreichs (Lysimetermessung sowie EC-Messungen, die im Rahmen wissenschaftlicher Projekte durchgeführt wurden) zum Trainingsdatensatz oder auch der Ausschluss von FLUXNET-Standorten, die sich in gänzlich anderen Biomen und Klimaten befinden.

Dieser Ausblick soll das Potenzial für weitere Forschung in Richtung einer räumlich verteilten DL/ML-basierten Verdunstungsvorhersage für Österreich aufzeigen. Dieser erste Ad-hoc-Versuch zeigt, dass das auf globale FLUXNET-Standorte trainierte LSTM räumliche Muster der mittleren jährlichen Verdunstung aufbauend auf ERA5-Land-Eingangsdaten realistisch nachbilden kann. Allerdings sind weitere Analysen und Schritte für die Verbesserung des Modells für die Verdunstungsvorhersage in Österreich notwendig.

5 Zusammenfassung

In der vorliegenden Studie wurden zwei DL/ML-Methoden für die Abschätzung der Verdunstung angewendet. Beide Ansätze wurden mit Verdunstungsdaten des globalen FLUXNET2015-Datensatzes trainiert sowie evaluiert. Bei der Evaluierung der Modelle wurde ein out-of-sample-Kreuzvalidierungsansatz gewählt, bei dem das trainierte Modell zur Abschätzung der Verdunstung an Stationen, die nicht im Trainingsdatensatz enthalten sind, verwendet wird. Mithilfe dieses Ansatzes lässt sich abschätzen, inwieweit Modelle generalisieren können und Erlerntes räumlich transferiert werden kann.

Während XGBoost ausschließlich Informationen für den zu bestimmenden Zeitschritt als Eingangsgrößen erhält, analysiert das LSTM eine Zeitreihe der Eingangsgrößen und ist dadurch in der Lage, Speichereffekte explizit zu berücksichtigen. Ein Ziel dieser Studie ist es zu untersuchen, ob diese Berücksichtigung von Speichereffekten einen Vorteil bei der Abschätzung der Verdunstung mit sich bringt.

Aufgrund der analysierten 153 FLUXNET-Standorte lässt sich sagen, dass beide Modelle gute Ergebnisse erzielen und eine höhere Modellgüte als der FLUXCOM-Datensatz aufweisen, der als Referenz zur Beurteilung der Modellperformanz herangezogen wurde. Vergleicht man beide Modellansätze, so weist das LSTM eine höhere Modellgüte als XGBoost auf, was vor allem auf eine bessere Vorhersage in warmen Gebieten mit kurzer Vegetation (Savanne, Steppe, Grasland) zurückzuführen ist. Dies ist interessant, da in diesen oft semiariden Ökosystemen mit begrenzten Wasserressourcen Speichereffekte möglicherweise eine wichtigere Rolle spielen als in Gebieten mit ausreichend Wasserressourcen, in denen die Verdunstung vorwiegend durch die Strahlung limitiert ist. Diese Überlegung wird auch durch die Beobachtung gestützt, dass XGBoost in permanenten Feuchtgebieten bessere Vorhersagen liefert, in denen ausreichend Wasser vorhanden ist und somit Speichereffekte weniger relevant erscheinen. Zusätzlich zu permanenten Feuchtgebieten weist XGBoost in Mischwäldern eine bessere Modellgüte auf. Weitere Forschung in Richtung einer Kombination beider Modelle, beispielsweise in Form eines Ensembles, erscheint daher sinnvoll.

Die präsentierten Analysen legen nahe, dass Informationen über den zeitlichen Verlauf der Eingangsgrößen zumindest für bestimmte Vegetationstypen bzw. Biome zu einer Verbesserung der Verdunstungsvorhersage führen. Daher ist für eine weitere Optimierung des XGBoost-Ansatzes zu analysieren, ob die Verwendung kurzer Zeitreihen der Eingangsgrößen oder aggregierter Werte, die den zeitlichen Verlauf beispielsweise der letzten Woche zusammenfassen, zu einer Verbesserung der Verdunstungsvorhersage führen.

Die Analyse des Standortmittels der Verdunstung zeigte, dass das LSTM diesen Wert im Vergleich zu XGBoost schlechter reproduzieren kann. In diesem Zusammenhang ist zu untersuchen, wie die Abschätzung dieses Standortmittels verbessert werden kann.

Ein kurzer Ausblick auf mögliche weiterführende Anwendungen von DL/ML-Ansätzen zeigt die Möglichkeit zur Nutzung des global trainierten Modells für die räumlich verteilte Vorhersage der Verdunstung in Österreich auf. Allerdings sind hier weitere Analysen nötig. So ist hier noch umfassende weitere Forschung notwendig, um

(i) Schritte zur Anpassung des Modells an den Standort zu untersuchen als auch (ii) Wege für die räumlich verteilte Evaluierung der Verdunstungsvorhersage auf regionaler Ebene zu finden.

Ein Vorteil von DL/ML-basierten Ansätzen ist, dass Daten verschiedener Datenquellen meist ohne aufwendige Vorprozessierung (um diese Rohdaten in definierte physikalische Größen umzuwandeln) verwendet werden können. In dieser Studie wurden Rohdaten dreier global verfügbarer Datensätze (FLUXNET-Daten, MODIS-Satellitendaten und meteorologische ERA5-Land-Daten) für das Trainieren der Modelle verwendet; es ließen sich jedoch beliebig weitere Informationen hinzufügen, wie beispielsweise Daten zu Bodenparametern oder Bodenfeuchte. Da die Daten ohne aufwendige Vorprozessierung direkt an das Modell übergeben werden können, reduziert sich zum einen der Aufwand für das Aufsetzen des Modells, zum anderen erhöhen sich jedoch auch die Nachvollziehbarkeit der Arbeitsschritte sowie die Reproduzierbarkeit der Arbeitsschritte.

Im Rahmen von Initiativen wie FLUXNET aber auch mithilfe neuer Satellitenmissionen werden global große Datenmengen zur Beschreibung der Verdunstung und anderer Austauschprozesse zwischen Landoberfläche und Atmosphäre gewonnen. Der hier präsentierte Modellierungsansatz, bei dem Daten unterschiedlicher Standorte zum Trainieren eines universellen Modells verwendet werden, erlaubt die nahtlose Integration all dieser bis heute und vor allem in Zukunft neu gewonnenen Daten. Da große Datenmengen eine wichtige Voraussetzung für „big data“-DL/ML-Ansätze sind, ist damit zu rechnen, dass zusätzliche Daten, vor allem von bis jetzt vielleicht unterrepräsentierten Standorten, zu einer weiteren Verbesserung solcher Modellansätze führen werden. Der Untersuchung von Potenzialen und Limitationen von DL/ML-Ansätzen zur Beschreibung physikalischer Prozesse in der Biosphäre, wie beispielsweise der Verdunstung, sollte daher auch als Ergänzung zu physikalisch-basierten Modellen in Zukunft eine wichtige Rolle zukommen.

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

Open Access Dieser Artikel wird unter der Creative Commons Namensnen-

nung 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

Tab. 3 Dynamische und statische Modelleingangsgrößen für LSTM und XGBoost

Dynamische Eingangsgrößen		Statische Gebietskenngrößen	
2 m Lufttemperatur	°C	IGBP Vegetationsklassen	–
2 m Windgeschwindigkeit	m s ⁻¹	Breitengrad	°
Sättigungsdefizit	hPa	Geländehöhe	m
Kurzweilige Einstrahlung	W m ⁻²	Mittlere Jahrestemperatur	°C
Langweilige Einstrahlung	W m ⁻²	Mittlere Jahrestemperaturspanne	°C
Niederschlag	Mm	Mittlerer Jahresniederschlag (MAP)	Mm
Luftdruck	kPa	Mittlerer jährlicher NDVI (Normalized Difference Vegetation Index)	–
Blattflächenindex	–	Mittlere jährliche NDVI-Spanne	–
XGBoost: Sinus (Tage des Jahres)/365* π	–	Mittlere jährliche Landoberflächentemperatur Tag – Nacht Differenz	°C
XGBoost: Kosinus (Stunde des Tages)/365* π	–	Maximale jährliche Landoberflächentemperatur Tag – Nacht Differenz	°C
		Jahresmittel des MODIS MCD43A4 Band 4 (grünes Licht)	–
		Jährliche Spanne des MODIS MCD43A4 Band 4 (grünes Licht)	–
		Jahresmittel des MODIS MCD43A4 Band 6 (Infrarotes Licht)	–
		Jährliche Spanne des MODIS MCD43A4 Band 6 (Infrarotes Licht)	–
		Jahresmittel der kurzweiligen Einstrahlung	W m ⁻²
		Jahresmaximum des Produkts aus kurzweiliger Einstrahlung und FPAR (Fraktion der photosynthetisch aktiven Strahlung)	W m ⁻²
		Jahresmittel der potenziellen Verdunstung	Mm
		Tage mit Niederschlag unter 0 °C	Tage
		Tage mit Niederschlag über 5*MAP	Tage
		Tage mit Niederschlag unter 1 mm	Tage
		Anzahl an zusammenhängenden Tagen mit Niederschlag unter 1 mm	Tage

Tab. 4 LSTM-Hyperparameter

Hidden size	128
Batch size	128
Output dropout	0,2
Learning rate	0,001
Epochen	5

Tab. 5 XGBoost-Wertebereich für die Hyperparametersuche (Bezeichnungen entsprechen den Argumenten für das python package „xgboost“)

n _estimators	100–5000
Learning rate	0,005–0,2
Gamma	0–3
Max depth	3–12
Min child weight	1–15
Colsample bytree	0,4–1,0
Colsample bylevel	0,4–1,0
Reg alpha	0–10
Reg lambda	0–10

Literatur

- Anderson, M. C., W. P. Kustas, J. M. Norman, C. R. Hain, J. R. Mecikalski, L. Schultz, M. P. González-Dugo, et al. (2011): Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrology and Earth System Sciences* 15:223–239. <https://doi.org/10.5194/hess-15-223-2011>.
- Aubinet, M., T. Vesala, and D. Papale (2012): *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer Science & Business Media.
- Bastiaanssen, W. G. M., E. J. M. Noordman, H. Pelgrum, G. Davids, B. P. Thoreson, and R. G. Allen (2005): SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering* 131(1):85–93. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2005\)131:1\(85\)](https://doi.org/10.1061/(ASCE)0733-9437(2005)131:1(85)).
- Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger, St. W. Running, P. Anthoni, et al. (2001): FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society* 82 (11): 2415–34.
- Bennett, A., B. Nijssen (2020): Deep Learned Process Parameterizations Provide Better Representations of Turbulent Heat Fluxes in Hydrologic Models. *Earth and Space Science Open Archive ESSOAr*, December. <https://doi.org/10.1002/essoar.10505081.1>.
- Bergstra, J., D. Yamins, and D. Cox (2013): Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In *International Conference on Machine Learning*, 115–23. PMLR.
- Best, M. J., G. Abramowitz, H. R. Johnson, A. J. Pitman, G. K. Balsamo, A. Boone, M. Cuntz, et al. (2015): The Plumbing of Land Surface Models: Benchmarking Model Performance. *Journal of Hydrometeorology* 16 (3): 1425–42.
- BMLRT (2021): *Die Wassermengen – Bilanz Österreichs*, Bmlrt.Gv.At. Wien. <https://www.bmlrt.gv.at/wasser/wasser-oesterreich/wasserkreislauf/Wasserbilanz.html>. Zugegriffen: 22.2.2021
- Brenner, C., Ch. E. Thiem, H.-D. Wizemann, M. Bernhardt, and K. Schulz (2017): Estimating Spatially Distributed Turbulent Heat Fluxes from High-Resolution Thermal Imagery Acquired with a UAV System. *International Journal of Remote Sensing* 38 (8–10): 3003–26.
- Brenner, C., M. J. Zeeman, M. Bernhardt, and K. Schulz (2018): Estimation of Evapotranspiration of Temperate Grassland Based on High-Resolution Thermal and Visible Range Imagery from Unmanned Aerial Systems. *International Journal of Remote Sensing* 39 (15–16): 5141–74. <https://doi.org/10.1080/01431161.2018.1471550>.
- Brutsaert, W. (2005): *Hydrology: An Introduction*. Cambridge: Cambridge University Press.
- Burba, G. (2013): *Eddy Covariance Method for Scientific, Industrial, Agricultural and Regulatory Applications: A Field Book on Measuring Ecosystem Gas Exchange and Areal Emission Rates*. LI-COR Biosciences, Lincoln, Nebraska
- Cammalleri, C., M. C. Anderson, F. Gao, C. R. Hain, and W. P. Kustas (2014): Mapping Daily Evapotranspiration at Field Scales over Rainfed and Irrigated Agricultural Areas Using Remote Sensing Data Fusion. *Agricultural and Forest Meteorology* 186: 1–11. <https://doi.org/10.1016/j.agrformet.2013.11.001>.
- Castelli, M., M. C. Anderson, Y. Yang, G. Wohlfahrt, G. Bertoldi, G. Niedrist, A. Hammerle, P. Zhao, M. Zebisch, and C. Notarnicola (2018): Two-Source Energy Balance Modeling of Evapotranspiration in Alpine Grasslands. *Remote Sensing of Environment* 209: 327–42. <https://doi.org/10.1016/j.rse.2018.02.062>.
- Chen, T., and C. Guestrin (2016): XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–94. KDD '16. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>.
- Chirouze, J., G. Boulet, L. Jarlan, R. Fieuzal, J. C. Rodriguez, J. Ezzahar, S. Er-Raki, et al. (2014): Intercomparison of four remote-sensing-based energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate. *Hydrology and Earth System Sciences* 18(3):1165–1188. <https://doi.org/10.5194/hess-18-1165-2014>.
- Clark, M. P., B. Nijssen, J. D. Lundquist, D. Kavetski, D. E. Rupp, R. A. Woods, J. E. Freer, et al. (2015): A Unified Approach for Process-Based Hydrologic Modeling: 1. Modeling Concept. *Water Resources Research* 51 (4): 2498–2514. <https://doi.org/10.1002/2015WR017198>.
- Cleugh, H. A., R. Leuning, M. Qiaozhen, and S. W. Running (2007): Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sensing of Environment* 106:285–304. <https://doi.org/10.1016/j.rse.2006.07.007>.
- Foken, T. (2008): *Micrometeorology*. Micrometeorology. Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-74666-9>.
- Friedman, J. H. (2001): Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29 (5): 1189–1232.
- Gauch, M., F. Kratzert, D. Klotz, G. Nearing, J. Lin, and S. Hochreiter (2020): Rainfall-Runoff Prediction at Multiple Timescales with a Single Long Short-Term Memory Network. *Hydrology and Earth System Sciences Discussions*, November, 1–25. <https://doi.org/10.5194/hess-2020-540>.
- Gauch, M., J. Mai, and J. Lin (2021): The Proper Care and Feeding of CAMELS: How Limited Training Data Affects Streamflow Prediction. *Environmental Modelling & Software* 135: 104926.
- Gers, F. A., J. Schmidhuber, and F. Cummins (1999): Learning to Forget: Continual Prediction with LSTM, January, 850–855. <https://doi.org/10.1049/cp:19991218>.
- Haiden, T., A. Kann, C. Wittmann, G. Pistotnik, B. Bica, and C. Gruber (2011): The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. *Weather and Forecasting* 26 (2): 166–83. <https://doi.org/10.1175/2010WAF2222451.1>.
- Hochreiter, S. (1991): *Untersuchungen Zu Dynamischen Neuronalen Netzen*. Diploma, Technische Universität München 91 (1).
- Hochreiter, S., and J. Schmidhuber (1997): Long Short-Term Memory. *Neural Computation* 9 (8): 1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Jung, M., S. Koiraal, U. Weber, K. Ichii, F. Gans, G. Camps-Valls, D. Papale, C. Schwalm, G. Tramontana, and M. Reichstein (2019): The FLUXCOM Ensemble of Global Land-Atmosphere Energy Fluxes. *Scientific Data* 6 (1): 1–14.
- Jung, M., M. Reichstein, and A. Bondeau (2009): Towards Global Empirical Upscaling of FLUXNET Eddy Covariance Observations: Validation of a Model Tree Ensemble Approach Using a Biosphere Model. *Biogeosciences* 6 (10): 2001–13. <https://doi.org/10.5194/bg-6-2001-2009>.
- Kalma, J. D., T. R. McVicar, and M. F. McCabe (2008): Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data. *Surveys in Geophysics* 29: 421–69. <https://doi.org/10.1007/s10712-008-9037-z>.
- Kling, H. (2006): *Spatio-Temporal Modelling of the Water Balance of Austria*.
- Kratzert, F., M. Gauch, G. Nearing, S. Hochreiter, D. Klotz (2021): *Niederschlags-Abfluss-Modellierung mit Long Short-Term Memory (LSTM)*. Österreichische Wasser- und Abfallwirtschaft 74(7–8). <https://doi.org/10.1007/s00506-021-00767-z>.
- Kratzert, F., M. Herrnegger, D. Klotz, S. Hochreiter, and G. Klambauer (2019b): *NeuralHydrology—Interpreting LSTMs in Hydrology*. *ArXiv:1903.07903 [Physics, Stat]* 11700: 347–62. https://doi.org/10.1007/978-3-030-28954-6_19.
- Kratzert, F., D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger (2018): *Rainfall-Runoff Modelling Using Long Short-Term Memory (LSTM) Networks*. *Hydrology and Earth System Sciences* 22 (11): 6005–22. <https://doi.org/10.5194/hess-22-6005-2018>.
- Kratzert, F., D. Klotz, G. Shalev, G. Klambauer, S. Hochreiter, and G. Nearing (2019a): *Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine Learning Applied to Large-Sample Datasets*. *Hydrology and Earth System Sciences* 23 (12): 5089–5110. <https://doi.org/10.5194/hess-23-5089-2019>.
- Kustas, W. P., J. G. Alfieri, M. C. Anderson, P. D. Colaizzi, J. H. Prueger, S. R. Evett, C. M. U. Neale, et al. (2012): Evaluating the two-source energy balance model using local thermal and surface flux observations in a strongly advective irrigated agricultural area. *Advances in Water Resources* 50:120–133. Elsevier Ltd. <https://doi.org/10.1016/j.advwatres.2012.07.005>.
- Li, Z. L., R. Tang, Z. Wan, Y. Bi, C. Zhou, B. Tang, G. Yan, and X. Zhang (2009): *A Review of Current Methodologies for Regional Evapotranspiration Estimation from Remotely Sensed Data*. *Sensors* 9 (5): 3801–53. <https://doi.org/10.3390/s90503801>.
- Luo, Z., J. Huang, K. Hu, X. Li, and P. Zhang (2019): *AccuAir: Winning Solution to Air Quality Prediction for KDD Cup 2018*. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1842–50.
- McMahon, T. A., M. C. Peel, L. Lowe, R. Srikanthan, and T. R. McVicar (2013): *Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: A pragmatic synthesis*. *Hydrology and Earth System Sciences* 17:1331–1363. <https://doi.org/10.5194/hess-17-1331-2013>.
- Monin, A. S., and A. M. Obukhov (1954): *Basic Laws of Turbulent Mixing in the Surface Layer of the Atmosphere*. *Contrib. Geophys. Inst. Acad. Sci. USSR* 24 (151): 163–87.
- Monteith, J. (1965): *Evaporation and Environment*. In *Symposia of the Society for Experimental Biology*, 19:205–34. Cambridge University Press (CUP) Cambridge.
- Muñoz Sabater, J. (2019): *ERA5-Land Hourly Data from 2001 to Present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/CDS.E2161BAC>.
- Myneni, R., Y. Knyazikhin, and T. Park (2015): *MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006*. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD15A2H.006>.
- Papale, D., and R. Valentini (2003): *A New Assessment of European Forests Carbon Exchanges by Eddy Fluxes and Artificial Neural Network Spatialization*. *Global Change Biology* 9 (4):

- 525–35. <https://doi.org/10.1046/j.1365-2486.2003.00609.x>.
- Pastorello, G., C. Trotta, E. Canfora, H. Chu, D. Christianson, Y.-W. Cheah, C. Poindexter, et al. (2020):** The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data. *Scientific Data* 7 (1): 225. <https://doi.org/10.1038/s41597-020-0534-3>.
- Tang, R., Z. L. Li, Y. Jia, C. Li, X. Sun, W. P. Kustas, and M. C. Anderson (2011):** An intercomparison of three remote sensing-based energy balance models using large aperture scintillometer measurements over a wheat-corn production region. *Remote Sensing of Environment* 115(12):3187–3202. Elsevier Inc. <https://doi.org/10.1016/j.rse.2011.07.004>.
- Tramontana, G., M. Jung, C. R. Schwalm, K. Ichii, G. Camps-Valls, B. Ráduly, M. Reichstein, et al. (2016):** Predicting Carbon Dioxide and Energy Fluxes across Global FLUXNET Sites with Regression Algorithms. *Biogeosciences* 13 (14): 4291–4313. <https://doi.org/10.5194/bg-13-4291-2016>.
- Yang, F., M. A. White, A. R. Michaelis, K. Ichii, H. Hashimoto, P. Votava, A. Zhu, and R. R. Nemani (2006):** Prediction of Continental-Scale Evapotranspiration by Combining MODIS and AmeriFlux Data Through Support Vector Machine. *IEEE Transactions on Geoscience and Remote Sensing* 44 (11): 3452–61. <https://doi.org/10.1109/TGRS.2006.876297>.
- Yebera, M., A. I. J. M. Van Dijk, R. Leuning, and J. P. Guerschman (2015):** Global Vegetation Gross Primary Production Estimation Using Satellite-Derived Light-Use Efficiency and Canopy Conductance. *Remote Sensing of Environment* 163 (June): 206–16. <https://doi.org/10.1016/j.rse.2015.03.016>.
- Yuan, W., S. Liu, G. Yu, J.-M. Bonnefond, J. Chen, K. Davis, A. R. Desai, et al. (2010):** Global Estimates of Evapotranspiration and Gross Primary Production Based on MODIS and Global Meteorology Data. *Remote Sensing of Environment* 114 (7): 1416–31. <https://doi.org/10.1016/j.rse.2010.01.022>.

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