

Langzeitevaluation retrospektiver LARSIM-Abflussvorhersagen unter Annahme idealer meteorologischer Eingangsdaten

Damaris Zulkarnaen, Jochen Seidel, Jan Bliefernicht, Norbert Demuth

Angaben zur Veröffentlichung / Publication details:

Zulkarnaen, Damaris, Jochen Seidel, Jan Bliefernicht, and Norbert Demuth. 2026.
“Langzeitevaluation retrospektiver LARSIM-Abflussvorhersagen unter Annahme idealer
meteorologischer Eingangsdaten.” *Hydrologie und Wasserbewirtschaftung* 70 (5): 320–33.
https://doi.org/10.5675/HyWa_2026.5_2.

Nutzungsbedingungen / Terms of use:

CC BY 4.0



Damaris Zulkarnaen, Jochen Seidel, Jan Bliefernicht & Norbert Demuth

Langzeitevaluation retrospektiver LARSIM-Abflussvorhersagen unter Annahme idealer meteorologischer Eingangsdaten

Long-term Evaluation of LARSIM Discharge Hindcasts Assuming Perfect Meteorological Input

Zur Verbesserung von Hochwasservorhersagemodellen ist eine Analyse der gemessenen Abflüsse im Vergleich zu den dazugehörigen prognostizierten Werten unabdingbar, um eine Aussage über die Güte eines Modells zu treffen. Im Rahmen dieser Studie wurden erstmals Abflussvorhersagen einer Langzeitsimulation des Wasserhaushaltsmodells LARSIM (Large Area Runoff Simulation Model) von 1993 bis 2024 evaluiert. Als Modelleingangsdaten dienten beobachtete meteorologische Daten, die es ermöglichen, die Modellgüte unter Minimierung der Eingangsunsicherheit zu beurteilen. Ausgewertet wurden Mess- und Vorhersagedaten in stündlicher Auflösung für 13 Pegel des Nahe-Einzugsgebiets in Rheinland-Pfalz, deren Einzugsgebietsgröße von 46 bis 4.038 km² reichen. Zur Auswertung der Daten erfolgten eine Überführung der Abflussvorhersagen in mehrere Vorhersagereichweiten, 1 bis 48 h, und eine Evaluation mit gängigen hydrologischen Gütemaßen wie der Kling-Gupta-Effizienz (KGE). Zusätzlich wurde die „Series Distance“ zur Bestimmung des Zeit- und Höhenversatzes zwischen gemessener und vorhergesagter Abflussganglinie angewandt.

Die Ergebnisse zeigten, dass die Abflussvorhersagen insgesamt eine hohe Güte aufweisen, diese aber abhängig von Einzugsgebietsgröße, Vorhersagereichweite und Jahreszeit variiert und die Vorhersagen teilweise stark von den Beobachtungsdaten abweichen. Die Nash-Sutcliffe-Effizienz (NSE) verbesserte sich für Pegel mit Einzugsgebieten von über 1.000 km² im Vergleich zu den Einzugsgebieten kleiner als 100 km² von 0,89 auf 0,97. Im Mittel über alle Hochwasserereignisse und Pegel verringerte sich die KGE mit jeder Vorhersagereichweite um etwa 0,02, von 0,99 auf 0,81. Täglich gerechnete Vorhersagen zeigten über den gesamten Untersuchungszeitraum hinweg eine höhere Güte für Wintermonate, mit einer NSE von 0,94, im Vergleich zu Sommermonaten, in welchen die NSE bei 0,88 lag. Die Ereignisspitzen wurden bei Vorhersagereichweiten über 12 h bei allen Sommerereignissen unterschätzt, während im Winter nur kleine Abweichungen auftraten. Es war zudem eine Tendenz zum zu späten Vorhersagen der Hochwasserscheitel feststellbar.

Schlagwörter: Abflussvorhersage, Hindcast, Hochwasser, LARSIM, Modellevaluierung

To improve flood forecasting models, it is essential to assess the quality of a model by analysing the measured discharge against the corresponding predicted values. This study presents the first such analysis conducted by evaluating discharge hindcasts from a long-term LARSIM water balance model simulation, 1993 to 2024. Measured meteorological data were used as model input, which allows for evaluating the model's performance while minimising input uncertainty. The analysis included forecasts for 13 gauges in the Nahe catchment in Rhineland-Palatinate, which were evaluated with the corresponding historical measurement time series, both in hourly resolution. The size of the individual catchment areas ranges from 46 to 4,038 km². To evaluate the data, the discharge forecasts were converted into multiple forecast lead times, 1 to 48 h, and evaluated with established skill scores such as the Kling-Gupta efficiency (KGE). Additionally, the series distance was used to determine the time and magnitude discrepancy between the measured and predicted hydrograph.

The results indicate that while the discharge forecasts have a high quality overall, it varies depending on the catchment size, lead time and season. Additionally, in some instances the forecasts deviate substantially from the observed data. The Nash-Sutcliffe efficiency (NSE) improved from 0.89 to 0.97 for gauges with catchment areas larger than 1,000 km² in comparison to the ones with areas smaller than 100 km². The KGE decreased, averaged over all flood events and gauges, by approximately 0.02 per lead time, from 0.99 to 0.81. Throughout the study period, daily forecasts demonstrated greater quality for winters, with an NSE of 0.94, compared to summers, where the NSE was 0.88. The analysis of the flood peaks showed that these were consistently underestimated for summers for lead times > 12 h, while for winter only minor deviations occurred. It was further found that there is a tendency to predict flood peaks too late.

Keywords: discharge forecast, hindcast, floods, LARSIM, model evaluation

1. Einleitung

Hochwasser gehören zu den Naturkatastrophen mit dem größten Schadenspotenzial weltweit (WEF, 2022) und werden im Zuge des Klimawandels höchstwahrscheinlich häufiger und stärker auftreten (ROUDIER et al., 2016; TABARI, 2020). Extremereignisse wie die Fluten im Ahrtal im Jahr 2021 oder im spanischen Valencia 2024 verdeutlichen die Wichtigkeit von verlässlichen Hochwasservorhersagen. Vorhersagen hoher Qualität ermöglichen eine zeitnahe Umsetzung geeigneter Schutz- und ggf. Eva-

kuierungsmaßnahmen, wodurch Schäden minimiert und Leben gerettet werden können.

Abflussvorhersagen beinhalten naturgemäß eine gewisse Unsicherheit, welche u. a. aus den zugrunde liegenden Wettervorhersagen, Mess- und Beobachtungswerten, Anfangsbedingungen, der Modellstruktur und den Modellparametern stammt (BOELEE et al., 2019). Insbesondere ungenaue hydro-meteorologische Beobachtungsdaten wie Niederschläge und Wasserstände können sich mehrfach auf die Vorhersagen auswirken, da diese für die Kalibrierung und Verifizierung der Modelle, die Simulation

der Vorhersagen und auch für die Interpretation der Ergebnisse mit z. B. unsicheren W-Q-Beziehung verwendet werden (BOELEEE et al., 2019). Zudem können sich verschiedene Unsicherheitsfaktoren durch die Modellkette fortsetzen und miteinander interagieren (MOGES et al., 2021).

Um den größten Nutzen eines Vorhersagemodells zu gewährleisten, muss die Qualität der daraus resultierenden Vorhersagen überprüft werden. Das Auswerten der Abweichungen zwischen Vorhersagewerten und den dazugehörigen Messwerten wird als Vorhersageverifikation, Validierung oder auch Evaluation bezeichnet (WILKS, 2011). In diesem Kontext wird häufig der „Skill“ einer Vorhersage bestimmt, welcher ein Maß für die Güte einer Vorhersage im Vergleich zu einer Referenz ist (JOLLIFFE & STEPHENSON, 2011). Als Referenz kann dabei der klimatologische Mittelwert zum Einsatz kommen, eine Persistenz-Vorhersage, bei welcher der letzte Messwert als Prognose dient, oder es werden Zufallswerte einer Wahrscheinlichkeitsverteilung der Vorhersagevariablen verwendet (JOLLIFFE & STEPHENSON, 2011; WILKS, 2011).

Das hydrologische Modell LARSIM (Large Area Runoff Simulation Model) ist ein hochaufgelöstes, konzeptionelles Wasserhaushaltsmodell. Es wird für operationelle Abflussvorhersagen in den Bundesländern Hessen, Rheinland-Pfalz, Baden-Württemberg und Bayern eingesetzt, aber auch von Dienststellen in Luxemburg, Frankreich, Österreich und der Schweiz angewandt (LEG, 2026).

In den letzten Jahren gab es einige Veröffentlichungen, in denen LARSIM für spezialisierte Forschungsfragen zu Themen wie historische Nährstoffeinträge in die Nord- und Ostsee (GERICKE et al., 2025) oder Effekte von Abschattung auf die Wassertemperatur eines Reservoirs (ISHIKAWA et al., 2021) zum Einsatz kam. In weiteren Studien wurde LARSIM mit einem dynamischen Infiltrationsmodul (HAAG et al., 2022), bzw. einem Gletschermodule (HAAG et al., 2019) erweitert, an andere Modelle gekoppelt (KRUMM et al., 2025; GERICKE et al., 2025; THIREL et al., 2019) oder verglichen, z. B. mit einem „Long Short Term Memory (LSTM)“-Modell (TANRIKULU et al., 2024) und einem nichtparametrischen stochastischen Modell (COSTA et al., 2012).

Arbeiten mit ausschließlichem Fokus auf die Qualität von LARSIM-Abflussvorhersagen gibt es dagegen kaum. Obwohl das Modell für die operationelle Hochwasservorhersage standardmäßig in vielen deutschen Landesämtern eingesetzt wird, liegt nur eine einzige Studie über die Vorhersageperformanz vor: BARTELS et al. (2017) haben eine Evaluierung der deterministischen und ensemblebasierten Abflussvorhersagen von LARSIM für einen vierjährigen Zeitraum für neun Pegel in Rheinland-Pfalz durchgeführt. Evaluierungsstudien über mehrere Jahrzehnte gibt es jedoch bisher nicht.

Die vorliegende Arbeit evaluiert Abflussvorhersagen in stündlicher Auflösung, welche für den Untersuchungszeitraum 1993 bis 2024 erstellt wurden, mit den dazugehörigen Beobachtungsdaten für 13 Pegel des Naheinzugsgebiets. Das Ziel besteht darin, die Qualität der Vorhersagen zu bestimmen und daraus Rückschlüsse auf die Langzeitmodellgüte von LARSIM zu ziehen. Für die Berechnung der Vorhersagen wurden ausschließlich meteorologische Messungen anstelle von numerischen Wettervorhersagen als Antriebsdaten gewählt. Diese Randbedingung,

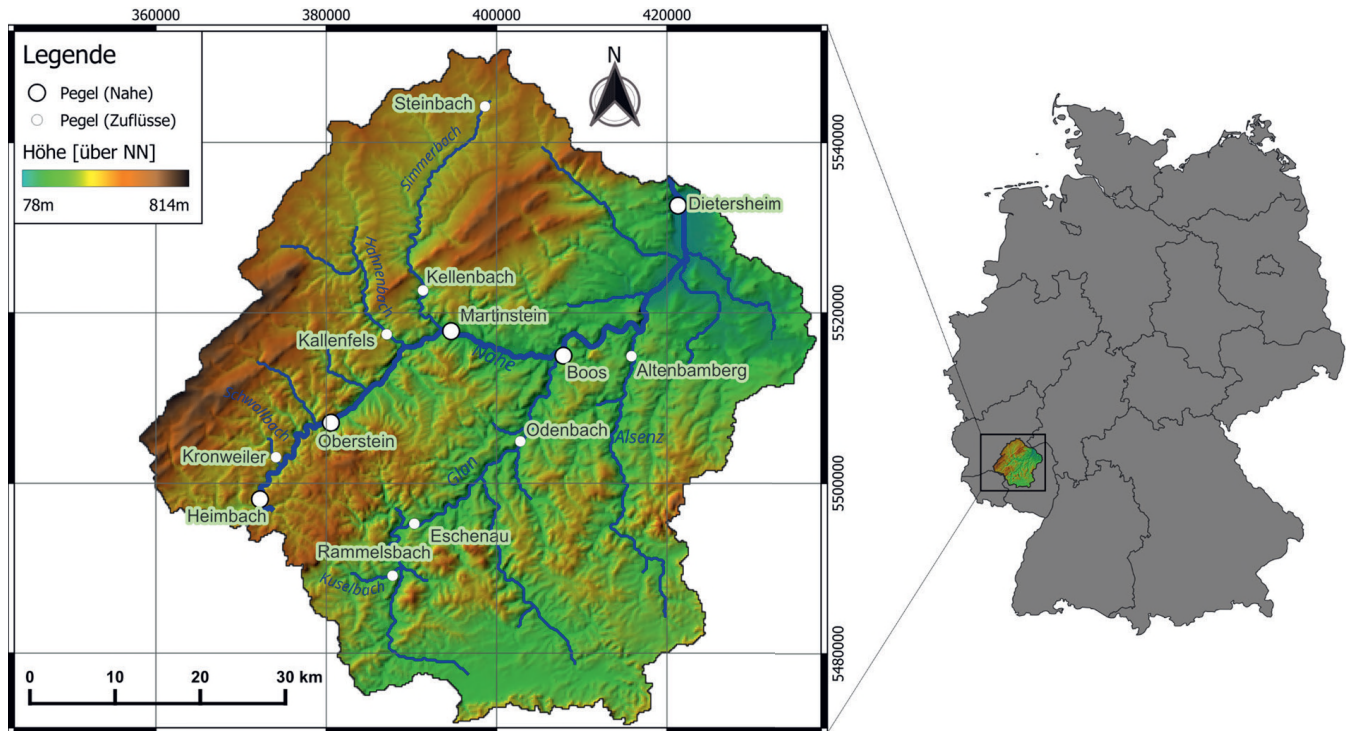
die einer idealen meteorologischen Vorhersage entspricht, minimiert die Eingangsunsicherheit und erlaubt es, die Unsicherheit des hydrologischen Modells auszuwerten. Besonders ist dabei die Datengrundlage in hoher zeitlicher Auflösung für einen langen Untersuchungszeitraum von über 30 Jahren, die in der Praxis selten vorliegt (FISCHER et al., 2023). In diesem Zeitraum kam es zu einigen großen Hochwassern, die in Kapitel 2 genauer beschrieben werden. Dadurch ist eine Beurteilung der Vorhersagegüte nicht nur in Bezug auf Mittel- und Niedrigwasserabflüsse, sondern auch für Hochwasserperioden mit verschiedenen Entstehungsmechanismen möglich.

2. Untersuchungsgebiet und ausgewählte Pegel

Das Untersuchungsgebiet umfasst das 4.067 km² große Einzugsgebiet der Nahe in Rheinland-Pfalz. Der vieljährige Mittelwert der Temperatur beträgt 1,5 °C im Winter und 17,3 °C im Sommer (DWD, 2024). Insgesamt ist das Klima ozeanisch geprägt, was zusammen mit der Topografie die Niederschlagsverteilung beeinflusst. Das langjährige Niederschlagsmittel beträgt 540 mm am Unterlauf und etwa 1.100 mm in den Höhenlagen des Oberlaufs. Die Höhe über dem Meeresspiegel reicht von 84 m im Bereich der Mündung in den Rhein bis zu 817 m im Hunsrück, welcher sich im Nordwesten des Untersuchungsgebiets befindet. Die Landnutzung wird von Land- und Forstwirtschaft dominiert, mit einem Anteil von jeweils 41 % Bodenbedeckung. Darauf folgen Siedlungs- und Verkehrsflächen mit 9 % und 6 % der Gesamtfläche (STLA RLP, 2021).

Die Nahe entspringt im Saarland, durchquert Rheinland-Pfalz auf 110 km Länge und mündet dann bei Bingen in den Rhein. Mit dieser Länge ist sie der drittgrößte Fluss in Rheinland-Pfalz nach Rhein und Mosel. Der Untersuchungszeitraum von 1993 bis 2024 deckt zwei große Hochwasser in den Jahren 1993 und 1995 ab, bei welchen im Untersuchungsgebiet Abflüsse mit Wiederkehrzeiten von 100 Jahren erreicht oder überschritten wurden. Zusätzlich gab es noch weitere nennenswerte Hochwasserereignisse: Im Herbst 1998 trat ein extremes Hochwasser auf, das in Anbetracht der Jahreszeit an den meisten Pegeln den höchsten jemals gemessenen Hochwasserabfluss (HHQ) des Monats Oktobers darstellt (FELL, 1999). Im Januar 2011 kam es zu einem durch Schneeschmelze beeinflussten Hochwasser mit einer Jährlichkeit von 5 bis 25 Jahren (HQ5 bis HQ25) (FELL, 2011). Im Juni 2016 ereignete sich ein Sommerhochwasser mit Werten von bis zu HQ25 am Pegel Kellenbach (DEMUTH et al., 2016).

Eine Analyse der beobachteten Abflussdaten ergab, dass im Untersuchungszeitraum Hochwasser mehr als 10-mal häufiger im Winter als im Sommer auftraten, was typisch für Pegel der oberen Nahe ist (HENRICHS, 2005). Alle bei der Evaluation miteinbezogenen Pegel sind in Abbildung 1 eingetragen. Sie beinhalten eine Pegelkaskade von fünf Pegeln am Hauptgewässer sowie acht Pegel an verschiedenen Zuflüssen. In ihrer Größe sind die Einzugsgebiete sehr unterschiedlich. Die Spanne reicht von kleiner als 50 km² bis über 4.000 km². Alle evaluierten Pegel sind mit Einzugsgebietsgröße (EZG-Größe) und den Abflusskennwerten in Tabelle 1 zusammengefasst. Die Kennwerte wurden aus stündlichen Beobachtungsdaten berechnet, sie beziehen sich alle auf den Zeitraum 1993 bis 2024. Es sind der mittlere Abfluss (MQ), der niedrigste Tagesmittelwert des Abflusses (NQ) und der höchste Momentanwert des Abflusses (HQ) für jeden Pegel angegeben.

**Abbildung 1**

Karte des Nahe-Einzugsgebiets mit Nebenflüssen und Pegeln (links) sowie die Verortung in Deutschland (rechts).
 Map of the Nahe catchment including tributaries and gauges (left) with location in Germany (right).

Tabelle 1

Pegel mit Einzugsgebietsgröße und Abflusskennwerten, aus Beobachtungsdaten in stündlicher Auflösung für den Zeitraum von 1993 bis 2024 berechnet.

Gauges with catchment size and discharge values, calculated from observed discharge in hourly resolution for the period 1993 to 2024.

Pegel	EZG-Größe [km ²]	MQ [m ³ /s]	NQ [m ³ /s]	HQ [m ³ /s]
Dietersheim	4038,7	31,01	2,49	916,53
Boos	2830,1	26,10	1,43	874,52
Martinstein	1468,0	16,30	0,59	575,02
Odenbach	1088,2	9,70	1,42	329,86
Eschenau	598,3	5,92	0,38	198,21
Oberstein	557,0	9,81	0,10	417,25
Kellenbach	362,1	2,32	0,08	147,00
Heimbach	318,3	6,17	0,09	205,29
Altenbamberg	318,2	1,75	0,17	75,38
Kallenfels	251,3	1,91	0,05	107,84
Rammelsbach	78,6	1,05	0,02	57,20
Kronweiler	64,6	0,90	0,00	57,94
Steinbach	46,2	0,40	0,01	23,70

3. Methodik

3.1. LARSIM-Qualitätsbewertung

Generell können für eine Vorhersageevaluation Echtzeitvorhersagen oder auch retrospektiv erstellte Vorhersagen, sogenannte Hindcasts, eines operationellen Modells genutzt werden (HARRIGAN et al., 2023). Zur LARSIM-Qualitätsbewertung lassen sich Vorhersagetestes für einen vergangenen Zeitraum durchführen

und danach mit gemessenen Abflüssen auswerten. Dabei kann zwischen 4 verschiedenen Berechnungsansätzen unterschieden werden (LUCE et al., 2006):

1. Simulationen mit gemessenen meteorologischen Eingangsdaten im Vorhersagezeitraum;
2. durchgängige Simulation
 - 2.1. ohne Nachführung sowie
 - 2.2. mit automatischer Nachführung;

3. Vorhersagetests unter Verwendung von Wettervorhersagen im Vorhersagezeitraum und automatischer Modellnachführung; sowie
4. Worst-Case-Niedrigwasservorhersagen, welche mit Nachführung und unter Annahme von ungünstigen Witterungsbedingungen, das heißt kein Niederschlag und hohe Temperaturen, berechnet werden.

Der Vergleich der einzelnen Vorhersagetestergebnisse untereinander erlaubt dann eine Aussage über den Einfluss der Modellnachführung oder der Unsicherheit aus Wettervorhersagen auf die Güte der Abflussvorhersagen. Der minimal mögliche Abfluss wird durch den 4. Ansatz abgeschätzt. Die in dieser Arbeit ausgewerteten Vorhersagen wurden dem 2. Ansatz entsprechend generiert.

3.2. Erstellung der LARSIM-Abflussvorhersagen

Im Sommer 2024 wurden am Landesamt für Umwelt Rheinland-Pfalz (LfU RLP) mit dem LARSIM-Wasserhaushaltsmodell RLP Abflussvorhersagen für 64 Pegel in Rheinland-Pfalz für den Zeitraum vom 3.1.1993 bis 1.7.2024 erstellt. Als Eingangsdaten dienten dabei interpolierte Stationsmesswerte von Niederschlag, Lufttemperatur, Globalstrahlung, Windgeschwindigkeit, Luftdruck und relativer Luftfeuchte sowie gemessene Abflüsse in stündlicher Auflösung. Für den gesamten Untersuchungszeitraum wurde für jeden einzelnen Tag und Pegel, zum Vorhersagezeitpunkt 7 Uhr (MEZ), eine Abflussvorhersage berechnet. Im Hochwasserfall, wenn der Messwert an einem Pegel den Schwellenwert des dreifachen MQs überschreitet, erfolgte ab diesem Zeitpunkt jede Stunde eine neue Vorhersage (Abb. 2). Diese ereignisbasierten Vorhersagen werden im Folgenden als „stündliche Vorhersagen“ bezeichnet, die kontinuierlichen, in täglichem

Intervall berechneten Vorhersagen als „tägliche Vorhersagen“. Jeder Modelllauf bzw. jede Vorhersage beinhaltet prognostizierte Abflusswerte für die nächsten 4 Tage in einer stündlichen Auflösung. In Abbildung 2 wurden die stündlichen Vorhersagen verkürzt dargestellt.

Um den Rechenaufwand einzugrenzen, wurden die stündlichen Vorhersagen erstellt, sobald an einem der berücksichtigten Pegel in Rheinland-Pfalz der jeweilige Hochwasserschwellenwert überschritten wurde. Infolgedessen weisen alle 13 ausgewerteten Pegel die gleiche Anzahl an stündlichen Vorhersagen zu den gleichen Zeiträumen auf. Dieser Umstand ist unabhängig davon, ob der Abfluss eines jeden einzelnen Pegels seinen spezifischen Schwellenwert tatsächlich überschritten hat oder nicht. Insgesamt wurden dadurch etwa 65.000 stündliche Vorhersagen für jeden einzelnen Pegel erstellt. Es zeigte sich, dass die Gesamtsumme der Stunden, in denen die gemessenen Abflüsse ihren pegelspezifischen Schwellenwert überschritten, zwischen etwa 14.000 und 22.200 lag.

3.3. Modellnachführung

Um die Qualität der Vorhersagen zu verbessern und den aktuellen Wasserhaushalt bestmöglich zu repräsentieren, bietet LARSIM verschiedene Möglichkeiten der Modellnachführung (LEG, 2026). Dazu werden Mess- und Simulationswerte im Auswertez Zeitraum verglichen (Abb. 3). Bei größeren Abweichungen können geeignete Modellgrößen angepasst und damit die Übereinstimmung der simulierten und gemessenen Volumina erhöht werden. Nach der Nachführung erfolgt dann die Berechnung der Vorhersage. Voraussetzung für eine optimale Modellnachführung ist dabei, dass die vorliegenden Messdaten zuverlässig sind, das heißt die Qualität der herangezogenen W-Q-Beziehung hoch ist.

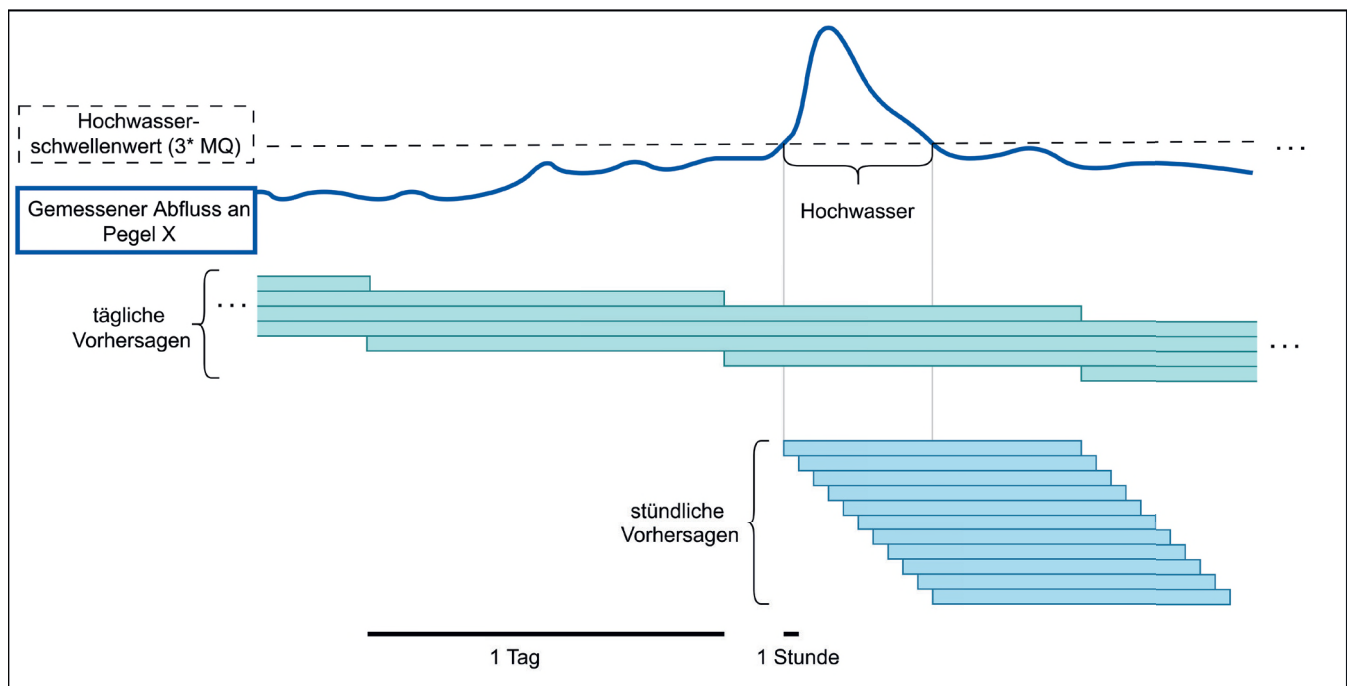


Abbildung 2

Schematische Darstellung der in stündlichem und täglichem Intervall erstellten Vorhersagen.
Schematic visualisation of forecasts generated at hourly and daily intervals.

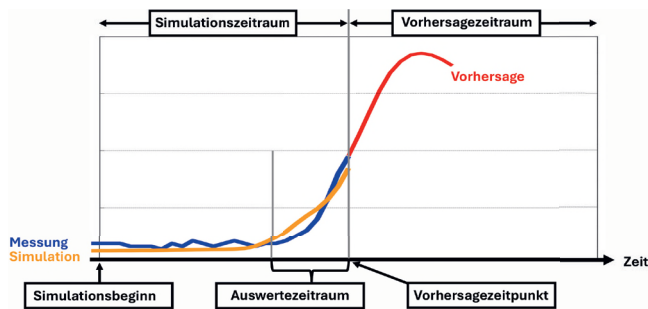


Abbildung 3

Schema der automatischen Modellnachführung (geändert nach GERLINGER, 2020).

Diagram of the automated model optimisation (altered from GERLINGER, 2020).

In LARSIM werden variationelle oder deterministische Methoden (LE DIMET et al., 2009; HONNORAT, 2007) zur Modellnachführung genutzt. 4 verschiedene Ansätze (REFSGAARD, 1997; KOMMA et al., 2009) sind implementiert und in LEG (2026) erläutert. Damit können die Eingangsgrößen (1), Zustandsgrößen (2), Modellparameter (3) oder Berechnungsergebnisse (4) nachgeführt werden (Abb. 4).

Bei der Erstellung der in dieser Studie evaluierten Vorhersagen wurde eine konstante Konfiguration der Modellnachführung angewendet. Aktiviert waren hierbei die Nachführung des Wasserdargebots als Inputvariable bzw. Eingangsgröße (1), der Gebietsspeicher für Basisabfluss sowie Interflow, langsamer und schneller Direktabfluss als Zustandsgrößen (2) sowie eine Vorhersagekorrektur als Output, d. h. Berechnungsergebnis (4).

Die Vorhersagekorrektur ist aus dem Prinzip der Korrektur mittels ARIMA (Autoregressive Integrated Moving Average) hervorgegangen (LEG 2026) und bietet erweiterte Ansätze zur Sicherstellung eines bruchfreien Übergangs vom letzten Messwert auf die berechnete Vorhersage. Weitere in LARSIM implementierte und im operationellen Betrieb ereignisabhängig manuell aktivierte Verfahren der Modellnachführung wurden hier nicht angewendet, wie etwa eine zeitliche Verschiebung der berechneten Abflussganglinie als Output, die Nachführung der Schnee-Wasseräquivalente als Systemzustand oder die Nachführung der Grenztemperatur Schnee als Modellparameter.

3.4. Überführung in Vorhersagereichweiten

Die einzelnen Vorhersagen wurden nach BARTELS et al. (2017) entsprechend aggregiert, um die Evaluierung für verschiedene Vorhersagereichweiten durchführen zu können. Aus den täglichen Vorhersagen wurden jeweils die ersten 24 Werte jeder Vorhersage zu einer Zeitreihe zusammengestellt (Abb. 5). Das Verfahren wurde dann für den zweiten Vorhersagetag wiederholt, wodurch zwei neue, lückenlose Vorhersagezeitreihen mit den Reichweiten von 1 bis 24 h sowie 25 bis 48 h für den gesam-

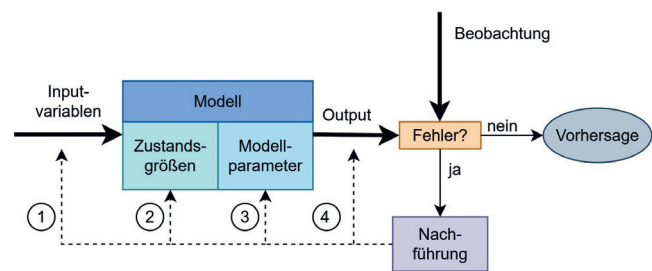


Abbildung 4

Mögliche Ansätze für die Nachführung in hydrologischen Modellen: 1) Eingangsgrößen, 2) Zustandsgrößen, 3) Modellparameter oder 4) Berechnungsergebnisse (KOMMA et al., 2009).

Possible model optimisation strategies : 1) input variables, 2) state variables, 3) model parameters, or 4) calculation results (KOMMA et al., 2009).

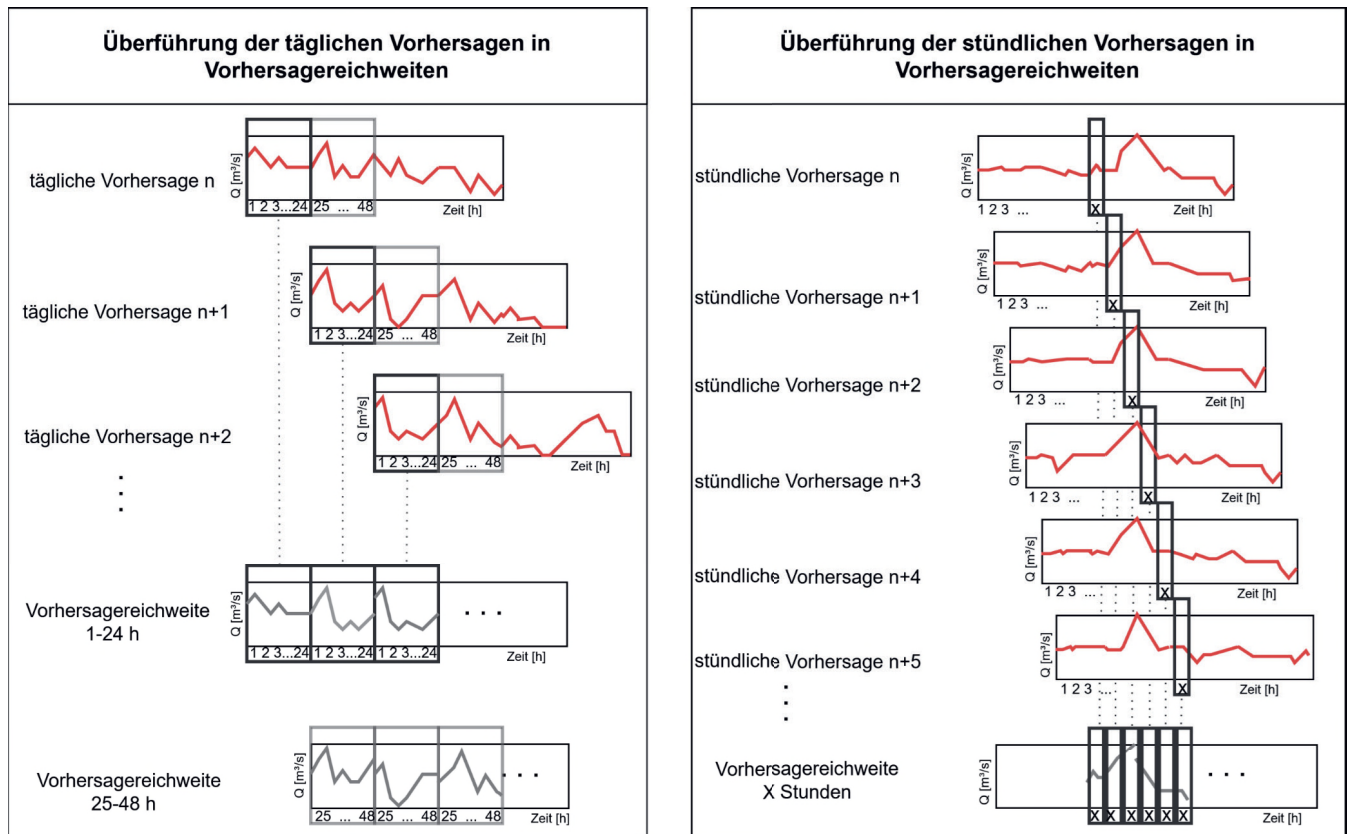
ten Untersuchungszeitraum von über 30 Jahren entstanden. Bei den stündlichen Vorhersagen war das Vorgehen ähnlich, mit dem Unterschied, dass jeweils nur ein Wert, also der N-te Wert, für die Vorhersagereichweite N entnommen wurde. Insgesamt entstanden so 9 Vorhersagereichweiten: 1, 3, 6, 9, 12, 18, 24, 36 und 48 h. Obwohl die stündlichen Vorhersagen, wie in Kapitel 3.2 beschrieben, simultan für alle Pegel in Rheinland-Pfalz berechnet wurden, sind in dieser Arbeit nur diejenigen Vorhersagen berücksichtigt, die sich auf tatsächliche Hochwasserereignisse an diesem Pegel beziehen. Hochwasserbeginn und -ende wurden dabei wie folgt bestimmt: das erstmalige Überschreiten des Schwellenwertes markiert den Anfang eines Hochwassers. Das Ende des Hochwasserereignisses wurde als eine Unterschreitung des Schwellenwertes für mindestens 48 h definiert.

Die Art der Vorhersagegeneration ohne Vorlaufzeit hat die Folge, dass insbesondere für die größeren Vorhersagereichweiten Lücken am Anfang der aggregierten stündlichen Vorhersagen auftreten. Für eine lückenlose aggregierte Vorhersage der Reichweite 48 h müsste beispielsweise die erste stündliche Vorhersage bereits für ein Datum zwei Tage vor dem ersten Überschreiten des Schwellenwertes ausgestellt worden sein. Dieser Umstand liegt teilweise durch Zufall vor, wenn zu diesem Zeitpunkt an einem anderen Pegel ein Hochwasser stattfand, jedoch ist dies nicht immer der Fall. Infolgedessen nimmt die Vollständigkeit der stündlich aggregierten Vorhersagen mit zunehmender Vorhersagereichweite ab.

In Tabelle 2 sind die im Mittel verfügbaren Vorhersagestunden als Prozent der als Hochwasser definierten Zeiträume eingetragen. Dabei werden die einzelnen Vorhersagereichweiten sowie die Sommer- und Winterhochwasser getrennt aufgeführt. Für Sommerhochwasser nimmt die Verfügbarkeit schnell ab und liegt bei der größten Reichweite von 48 h im Mittel nur bei 63 %. Zusätzlich sollte aber die große Differenz in der Hochwasserdauer in den beiden Jahreszeiten berücksichtigt werden. Durch seltenere und kürzere Hochwasser im Sommer ist die durchschnittliche

Tabelle 2
Mittlere Verfügbarkeit der stündlichen LARSIM-Vorhersagen.
Mean availability of the LARSIM forecasts.

Vorhersagereichweite [h]	1	3	6	9	12	18	24	36	48
Winter [%]	100,0	99,9	99,6	99,1	98,4	96,6	94,5	89,9	85,5
Sommer [%]	100,0	99,4	97,9	95,6	92,8	86,8	81,1	71,9	63,3

**Abbildung 5**

Schematische Darstellung der Überführung in einzelne Vorhersagereichweiten für tägliche (links) und stündliche Vorhersagen (rechts).
 Schematic depiction of the aggregation into individual lead times for daily (left) and hourly forecasts (right).

Gesamtdauer um eine Größenordnung kleiner als im Winter: 18.866 h im Winter, 1.658 h im Sommer.

3.5. Gütemaße für die Langzeitevaluierung

Für die Langzeitevaluierung der Modellgüte kamen in dieser Studie drei Gütemaße zur Anwendung: Normalisierter Bias, Nash-Sutcliffe-Effizienz und Kling-Gupta-Effizienz. Der normalisierte Bias oder Prozent-Bias (PBIAS) gibt an, ob die Vorhersagen im Mittel eine Tendenz der Über- oder Unterschätzung in Bezug auf die Beobachtungswerte aufweisen (GUPTA et al., 1999). Er wird berechnet als:

$$\text{PBIAS} = 100 \cdot \frac{\sum_{i=0}^n (Q_{\text{sim},i} - Q_{\text{mess},i})}{\sum_{i=0}^n (Q_{\text{mess},i})} \quad (\text{Gl. 1})$$

wobei Q_i den Vorhersagewert zum Zeitpunkt i , $Q_{\text{mess},i}$ die Beobachtung zum Zeitpunkt i , und n die Anzahl der gemeinsamen Datenpaare der Vorhersage und Beobachtung darstellen. Generell deuten positive Werte auf eine Überschätzung und negative Werte auf eine Unterschätzung der Beobachtung hin.

Die Nash-Sutcliffe-Effizienz (NSE) ist seit ihrer Einführung im Jahr 1970 zu einem bekannten und häufig angewandten Gütemaß der Hydrologie geworden (NASH & SUTCLIFFE, 1970; LAMONTAGNE et al., 2020). Sie wird berechnet als:

$$\text{NSE} = 1 - \frac{\sum_{i=0}^n (Q_i - Q_{\text{mess},i})^2}{\sum_{i=0}^n (Q_{\text{mess},i} - \bar{Q}_{\text{mess}})^2} \quad (\text{Gl. 2})$$

wobei $\bar{Q}_{\text{mess}}$ den arithmetischen Mittelwert der Beobachtungen beschreibt und sich auf den jeweils ausgewerteten Zeitraum bezieht. Die weiteren Variablen sind äquivalent zu Gleichung 1. Das Anwenden der NSE ist auch auf die logarithmisch transformierten Abflussdaten möglich, dann ist von der log-NSE die Rede (LAMONTAGNE et al., 2020). Diese wird insbesondere für Simulationen im Niedrigwasserbereich angewandt (PUSHPALATHA et al., 2012).

Die Kling-Gupta-Effizienz (KGE) wurde von GUPTA et al. (2009) als Verbesserung der NSE eingeführt. Sie stellt ein multikriterielles Gütemaß dar, mit welchem die Qualitätsattribute Korrelation, Bias und Variabilität einer Vorhersage gemeinsam bewertet werden. Die KGE wird als:

$$\text{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (\text{Gl. 3})$$

$$r = \frac{\text{Cov}_{\text{sim,mess}}}{\sigma_{\text{sim}} \cdot \sigma_{\text{mess}}} \quad \alpha = \frac{\sigma_{\text{sim}}}{\sigma_{\text{mess}}} \quad \beta = \frac{\mu_{\text{sim}}}{\mu_{\text{mess}}}$$

berechnet, wobei r den Pearson-Korrelationskoeffizienten, α die Variabilität und β den Bias darstellt. Diese werden aus der Kovarianz Cov , der Standardabweichung σ und dem Mittelwert μ aus den Vorhersage- und Beobachtungsdaten berechnet. Auch hier beziehen sich die Beobachtungsdaten auf den gleichen Zeitraum wie die Vorhersagen.

3.6. Gütemaße für die ereignisbasierte Evaluierung

Um die Ergebnisse ereignisbasiert zu evaluieren, kamen in dieser Studie ebenfalls drei Gütemaße zur Anwendung: die Series Distance, welche den Versatz einzelner Segmente bestimmt, sowie der Zeitfehler und Magnitudenfehler der Hochwasserspitzen. Gängige Fehlermaße, Metriken oder Qualitätskriterien vergleichen beobachtete und simulierte Werte „vertikal“, das heißt, dass Punkte, die zeitgleich auftreten, miteinander verglichen werden (EHRET & ZEHE, 2011). Infolgedessen wirkt ein geringer Zeitversatz als großer Magnitudenfehler. Um diesem Umstand entgegenzuwirken, wurde das Prinzip der „Series Distance“ von EHRET & ZEHE (2011) vorgestellt und durch SEIBERT et al. (2016) erweitert, welche eine gemeinsame Auswertung hydrologisch gleichartiger Segmente ermöglicht. Das übergeordnete Ziel ist es, den Zeit- und Magnitudenfehler separat zu beurteilen. In Abbildung 6 sind „Series Distance“-Verbindungslinien in hellgrau dargestellt; sie verbinden Punkte gleichartiger Segmente. Im Rechteck wird ein Ausschnitt der Ganglinien markiert, bei welchem durch Evaluierung mit einer gängigen Metrik ein ansteigender Ast der Beobachtung mit einem absteigenden Ast der simulierten Abflusskurve verglichen würde.

Bei der Interpretation der Ergebnisse gilt es zu beachten, dass die Fehler jeweils aus Beobachtung minus Vorhersage berechnet wurden. Das heißt, dass ein negativer Magnitudenfehler eine Überschätzung des Abflusses durch das Modell bedeutet und ein negativer Zeitfehler ein im Schnitt zu spätes Vorhersagen der Zeitreihe impliziert.

Der Zeitfehler, engl. Peak Time Error (PTE), wird als Mittlere Abweichung zwischen dem Zeitpunkt des Maximalabflusses in der vorhergesagten und beobachteten Abflussreihe berechnet:

$$PTE = \frac{1}{n} \sum_{i=0}^n (\text{Zeitpunkt}(\max(Q_{\text{sim},i})) - \text{Zeitpunkt}(\max(Q_{\text{mess},i}))) \quad (\text{Gl. 4})$$

wobei n der Anzahl der Ereignisse entspricht. Ein positiver PTE von beispielsweise 5 h impliziert dann, dass die Hochwasserscheitel im Mittel 5 h zu spät vorhergesagt werden. Folglich

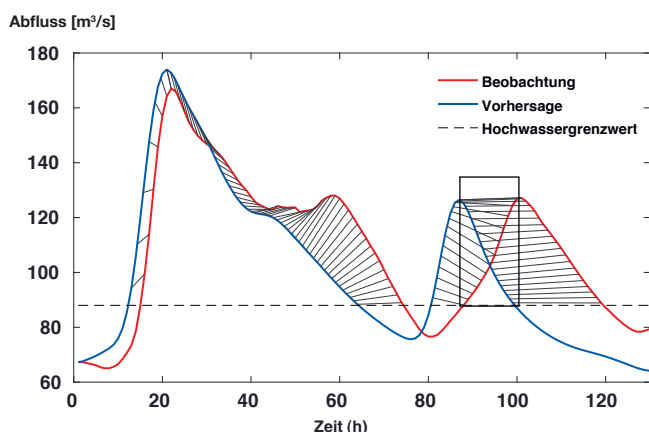


Abbildung 6

Zeitreihen der beobachteten (rot) und vorhergesagten (blau) Ganglinien. Die hellgrauen Linien sind die „Series Distance“-Verbindungslinien zwischen hydrologisch ähnlichen Punkten beider Zeitreihen (abgewandelt von SEIBERT et al., 2016).

Time series of observed (red) and simulated (blue) hydrographs. The light grey lines represent the series distance connectors linking hydrologically similar points in both time series (modified from SEIBERT et al., 2016).

bedeuten negative PTE-Werte ein im Mittel zu frühes Vorhersagen der Scheitel.

Dem PTE entsprechend wird der Magnitudenfehler, „Peak Magnitude Error“ (PME), als mittlere Abweichung zwischen den Maximalabflüssen der vorhergesagten und beobachteten Zeitreihe berechnet.

$$PME = \frac{1}{n} \sum_{i=0}^n (\max(Q_{i,i}) - \max(Q_{\text{mess},i})) \quad (\text{Gl. 5})$$

Bei einem $PME < 0$ unterschätzt das Modell die Hochwasserspitze, bei einem Wert > 0 wird überschätzt.

4. Ergebnisse der Langzeitevaluierung

Im Folgenden werden die täglichen Vorhersagen über den gesamten Untersuchungszeitraum, also über 30 Jahre, zusammen ausgewertet. Es sind dadurch Mittel-, Niedrig- und Hochwasserperioden abgedeckt.

4.1. Einfluss der Jahreszeit

Zunächst erfolgt eine Evaluierung der täglichen Vorhersagen der beiden Reichweiten von 1 bis 24 h und 25 bis 48 h, mit der Unterscheidung nach hydrologischer Jahreszeit. Abbildung 7 zeigt gruppierte Boxplots der NSE für beide Reichweiten und Jahreszeiten. Insgesamt kann eine höhere Güte für die kürzere Reichweite von 1 bis 24 h ausgemacht werden, sowohl für hydrologische Sommer als auch Winter. Der Median und auch die Maximal-Güte ist jeweils höher für die kürzere Vorhersagereichweite. Die höchste Variationsbreite ist für die größere Reichweite während der Sommermonate sichtbar. Der Median im Winter ist jeweils höher als im Sommer. Insgesamt kann eine höhere Güte für die kürzere Vorhersagereichweite und in hydrologischen Wintern beobachtet werden.

Der Ausreißer, welcher für die kürzere Reichweite im Winter zu sehen ist, entspricht dem Pegel Kronweiler. Dieser Pegel liefert durchgehend die schlechtesten Werte, die Ursache dafür wird in

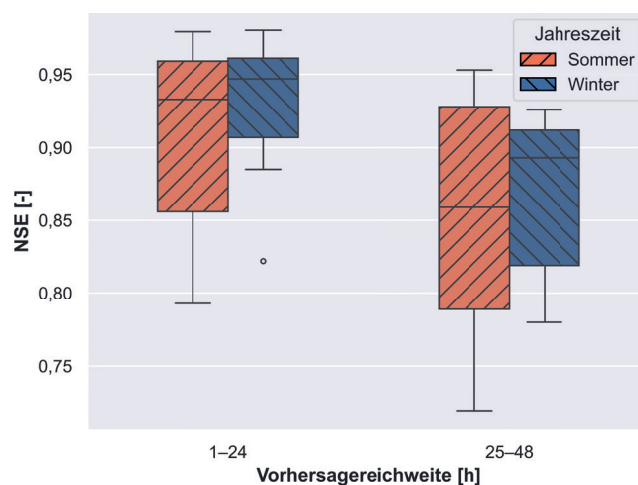


Abbildung 7

Gruppierte Boxplots der täglichen Vorhersagen nach hydrologischer Jahreszeit und den beiden Vorhersagereichweiten.

Grouped boxplots of the daily forecasts plotted according to the hydrological season and both lead times.

Kapitel 4.3 erläutert. Die höhere Güte einiger Pegel für die Reichweite 25 bis 48 h während der Sommermonate könnte dadurch erklärt werden, dass an einigen Pegeln nur wenige oder keine Hochwasser im Sommer stattgefunden haben. Bei der Auswertung mit der KGE kamen sehr ähnliche Ergebnisse zustande, daher wurde auf eine Abbildung verzichtet.

Die Unterschiede der Ergebnisse zwischen den beiden Jahreszeiten sind wahrscheinlich das Resultat der verschiedenen Randbedingungen in der entsprechenden Jahreszeit. Da im Winter niedrige Temperaturen vorherrschen und die Vegetation ruht, findet nur eine geringe Evapotranspiration statt. Zudem erfolgt nur eine minimale Wasseraufnahme durch Wurzeln. Aus diesem Grund bleibt die Bodenfeuchte weitgehend konstant. In Sommern ist das Gegenteil der Fall: Hier sind die Anfangsbedingungen stark unterschiedlich, bei der Bodenfeuchte können beispielsweise sowohl komplett ausgetrocknete Böden als auch gesättigte Böden vorliegen. Bei trockenen Böden und lokalen, häufig im Sommer auftretenden Starkregen unterschätzt das Modell LARSIM in der hier verwendeten Standardkonfiguration ohne das dynamische Infiltrationsmodul häufig die Abflussreaktion (HAAG et al., 2022). Dazu kommt, dass bei Starkregen die reale Niederschlagsverteilung und -intensität von Wetterstationen nicht erfasst werden (LENGFELD et al., 2020). Dadurch kann es im Sommerhalbjahr leicht zu unterschätzten Abflüssen kommen, was sich in einer breiteren Verteilung der Werte im Boxplot und somit in der im Mittel niedrigeren Vorhersagequalität widerspiegelt.

4.2. Einfluss der Einzugsgebietsgröße

Abbildung 8 zeigt, dass sich die Güte der Abflussvorhersagen, hier am Beispiel der NSE, mit zunehmender Einzugsgebietsgröße verbessert. Die Beziehung ist nicht-linear und für beide Vorhersagereichweiten ähnlich. Beim Vergleich der Reichweiten ist eine niedrigere Güte der Reichweite 25 bis 48 h von etwa 0,05 über alle Einzugsgebietsgrößen hinweg sichtbar. Für die kürzere Reichweite verbessert sich die NSE um 10 %, wenn die kleinsten mit den größten drei Einzugsgebieten verglichen werden. So weisen Pegel mit Einzugsgebietsgröße < 100 km² im Mittel eine

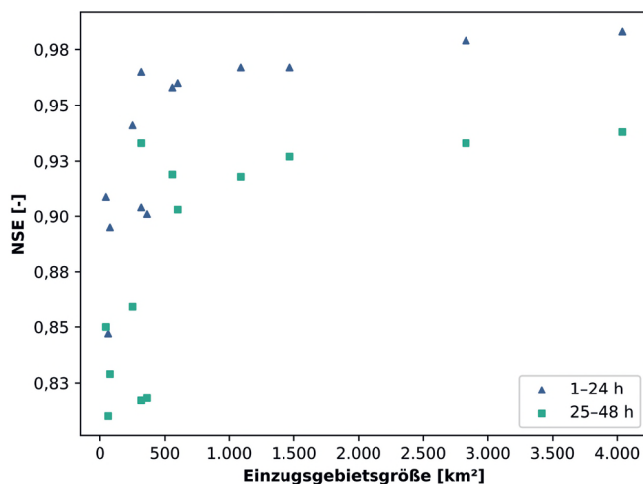


Abbildung 8

NSE gegen die Einzugsgebietsgröße jedes Pegels, für beide Reichweiten der täglichen Vorhersagen.
NSE plotted against the catchment size of each gauge, for both lead times of the daily forecasts.

NSE von 0,89 auf, Pegel mit Einzugsgebietsgröße > 1.000 km² haben im Mittel eine NSE von 0,97. Der Effekt der besseren Vorhersagequalität für größere Einzugsgebiete lässt sich durch genauere Eingangsvariablen erklären – längere historische Zeitreihen, besser belegte W-Q-Beziehungen, die Dämpfung lokaler Extreme im weiteren Verlauf des Einzugsgebiets, und die zunehmende Bedeutung des physikalisch besser zu simulierenden Flood Routings. Auch der Ausgleich von Ungenauigkeiten durch Mittelung spielt dabei eine Rolle. Da kleinere Einzugsgebiete sowohl eine kürzere Konzentrationszeit und eine höhere Variabilität der Abflusspende als auch ungenauere Einzugsgebietsparameter aufweisen, sind die Abflussvorhersagen im Allgemeinen ungenauer.

Die Ergebnisse zeigen eine hohe Vorhersagegüte, wenn der gesamte Untersuchungszeitraum ausgewertet wird, und stimmen somit mit ELPERS et al. (2023) überein. Auch die NSE-Werte einzelner Pegel sind in hoher Übereinstimmung mit den Werten, die nach der Kalibrierung des Modells, unter Verwendung gemessener Stationsdaten, mit durchgängiger Simulation ohne Nachführung von ELPERS et al. (2023) erreicht wurden. Dabei wurden in dieser Studie etwas höhere Werte erzielt, was auf die Verwendung der Nachführung zurückzuführen ist. Am Pegel Dietersheim ermittelte diese Studie beispielsweise eine NSE von 0,98, ELPERS et al. (2023) dagegen erhielten nur 0,9. Für Rammelsbach betragen die Werte 0,94 bzw. 0,83.

4.3. Jährliche Variabilität der Güte

Bei der Auswertung der täglichen Vorhersagen mit der kürzeren Reichweite von 1 bis 24 h nach Kalenderjahren wird sichtbar, dass die Qualität nicht konstant ist, sondern über die Jahre variiert. Aus Abbildung 9 geht hervor, dass die Vorhersagen im Jahr 2011 eine markant niedrigere Güte aufweisen. Der Median der KGE liegt in diesem Fall mit etwa 0,75 deutlich niedriger als in allen anderen Jahren. Die gleiche Auswertung mit der log-NSE in Abbildung 10 zeigt keine große Abweichung in diesem Jahr, was darauf hindeutet, dass die Diskrepanz zwischen Vorhersage- und Messwerten während Hochwasserereignissen auftrat. Eine genauere Analyse der Vorhersagen ergab, dass ein Hochwasser Anfang des Jahres 2011 für die Abweichungen verantwortlich ist. Im genannten Jahr gab es ungewöhnlich viel Schnee zum Jahresanfang, welcher eine geschlossene Schneedecke bildete. Durch einen Umschwung der Großwetterlage stiegen die Temperaturen in den positiven Bereich und führten zusammen mit teils ergiebigen Regen zu einem raschen Abtauen der Schneedecke bis in höhere Lagen (FELL, 2011). Das Modell LARSIM konnte das resultierende Regen-auf-Schnee-Hochwasser nur unzureichend abbilden, was zu den geringen KGE-Werten führte.

In Abbildung 9 ist außerdem zu erkennen, dass die übrigen Jahre zwar eine Variabilität zeigen, diese aber insgesamt bei einem hohen Niveau von > 0,9 für die KGE und auch NSE zu beobachten ist. Eine extrem niedrige KGE bzw. log-NSE von etwa -0,2 ist für den Pegel Kronweiler im Jahr 1997 verzeichnet (Abb. 10). Bei einer Analyse der Hydrographen dieses Pegels und Zeitraums zeigte sich, dass zu Beginn des Jahres 1997 der Abfluss von den Vorhersagen kontinuierlich überschätzt wurde und dadurch der extrem niedrige KGE-Wert entstand. Es stellte sich heraus, dass bei der Berechnung der täglichen Vorhersagen an diesem Pegel sowohl die Nachführung als auch die Vorhersagekorrektur nicht aktiviert wurde. Zusammen mit zeitweise fehlerhaften, vermut-



Abbildung 9
KGE der täglichen Vorhersagen über alle Kalenderjahre.
KGE of the daily forecasts over all calendar years.

lich durch Randeis verursachten Messwerten an den Pegeln Heimbach und Oberstein entstanden die großen Abweichungen.

In Abbildung 9 und 10 sind in den Jahren von 2016 bis 2023 keine extremen Ausreißer beim KGE und log-NSE sichtbar, was darauf hindeuten könnte, dass die Vorhersagen jüngerer Jahre eine höhere Qualität aufweisen. Im Jahr 2024 sind allerdings wieder eher niedrige KGE und NSE-Werte (keine Abb.) von 0,6 und 0,7 vorhanden. Für dieses Jahr lagen jedoch nur Vorhersagen für das erste halbe Jahr vor, wodurch ein Vergleich mit den anderen Jahren schwierig ist. Die log-NSE zeigt keine extremen Ausreißer; seit dem Jahr 2018 betrugen die Werte 0,8 oder mehr (Abb. 10).

5. Ergebnisse der ereignisbasierten Evaluierung

Bei den im Folgenden vorgestellten Ergebnissen der stündlichen Vorhersagen gilt es zu beachten, dass es sich bei den hierbei verwendeten Zeitreihen ausschließlich um Hochwasserereignisse handelt und die Zeitreihen nicht vollständig sind, siehe Kapitel 3.4 und Tabelle 2. Bei der Auswertung wurde der Fokus auf die Ereignisspitzen gelegt.

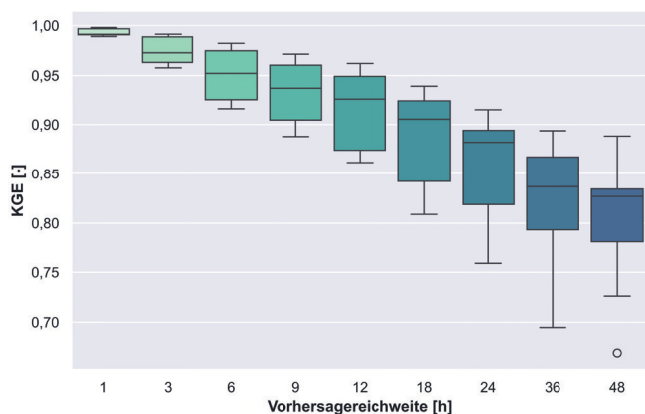


Abbildung 11
Boxplots der stündlichen Vorhersagen über alle Vorhersagereichweiten.
Boxplots of the hourly forecasts over all lead times.



Abbildung 10
Log-NSE der täglichen Vorhersagen über alle Kalenderjahre mit unterbrochener y-Achse
Log NSE of the daily forecasts over all calendar years with discontinuous y-axis.

Die Analyse der stündlichen Vorhersagen für die einzelnen Vorhersagereichweiten zeigt mit zunehmender Reichweite eine Abnahme der Güte. Exemplarisch werden in Abbildung 11 Boxplots der KGE gezeigt, der gleiche Trend ist aber auch in der Auswertung mit anderen Metriken zu beobachten. Der Median der KGE pro Vorhersagereichweite verringert sich um etwa 0,02 für jede zusätzliche Reichweite, beginnend mit 0,99 für die Reichweite von 1 h bis zu 0,81 für die größte Reichweite von 48 h. Die Änderung zwischen den Reichweiten 36 h und 48 h ist gering. Möglicherweise würde bei der Auswertung von noch größeren Reichweiten ein Schwellenwert erreicht werden.

5.1. Sommer- und Winterhochwasser

Bei der Auswertung mit Unterscheidung der hydrologischen Jahreszeit lassen sich insgesamt bessere Ergebnisse für die Wintermonate erzielen. Sommerhochwasser werden dagegen mit zunehmender Reichweite stark unterschätzt. In Abbildung 12

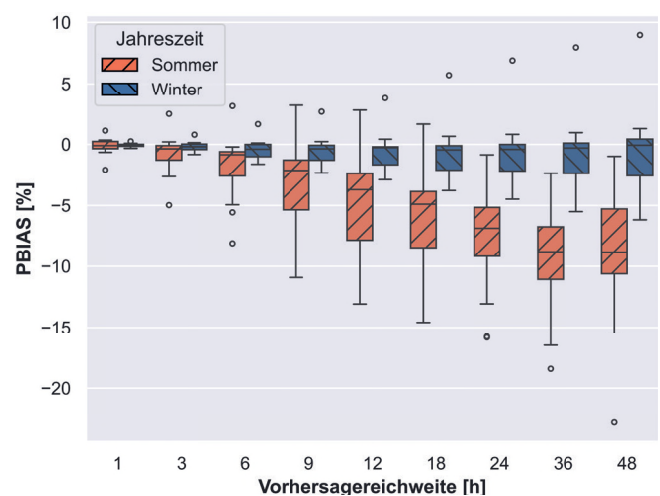


Abbildung 12
Gruppierter Boxplots des PBIAS nach Jahreszeit über alle Vorhersagereichweiten.
Grouped boxplots of the PBIAS per season and over all lead times.

ist der Prozent-Bias (PBIAS) als gruppierter Boxplot über alle Reichweiten hinweg aufgetragen. Für Winterhochwasser liegt der Median über alle Reichweiten nahe an 0, während sich der PBIAS der unteren 50 % der Daten zunehmend vergrößert und bis zu 6 % erreicht. In positiver Richtung, also Überschätzung, ist jeweils ein Ausreißer zu sehen – der Pegel Kellenbach – dessen Wert mit jeder Reichweite zunimmt und ein Maximum von fast 10 % erreicht.

Sommerhochwasser weisen dagegen nur für die ersten Reichweiten einen geringen PBIAS auf. Mit zunehmender Vorhersage-reichweite nimmt die Abweichung in Richtung Unterschätzung zu und die Spannweite der Boxplots vergrößert sich. Ab einer Reichweite von 24 h wird der Abfluss an allen Pegeln unterschätzt. Bei 50 % aller Pegel erfolgt zudem für die Reichweite von 48 h im Sommer eine Unterschätzung zwischen 5 und 11 %, mit Ausreißern von bis zu 23 %.

5.2. Series Distance

Im Folgenden werden die Ergebnisse der Auswertung der stündlichen Vorhersagen mit der „Series Distance“ beschrieben. Dieser Untersuchung liegt die Vorhersagereichweite 6 h zugrunde, da diese Zeitreihe nahezu vollständig ist, aber dennoch prinzipiell ausreichend Vorwarnzeit für eine Hochwasserwarnung oder Evakuierungsmaßnahmen bietet. Ausgewertet wurden die jeweils vier größten Hochwasserereignisse der Pegel Dietersheim und Kellenbach. Für Dietersheim sind das die Winterhochwasser von 1993, 1995, 2002 und 2011. Am Pegel Kellenbach sind die ersten drei Hochwasser identisch, das viertgrößte ist hier jedoch das Sommerhochwasser im Jahr 2016. Die beiden Pegel unterscheiden sich stark in ihrer Einzugsgebietsgröße und in der Qualität ihrer Vorhersagen in bisherigen Auswertungen. Die Gesamtsumme der ausgewerteten Hochwasserstunden sind für Dietersheim 1.860 h und für Kellenbach 1.571 h.

Beim Vergleich der Abbildungen 13 und 14 ist zu sehen, dass insgesamt die Tendenz des Mittelwerts, markiert durch ein Kreuz, für beide Pegel übereinstimmt. Dieses befindet sich für beide Pegel und sowohl für ansteigende als auch absteigende Äste im negativen Bereich für den Zeitfehler, was ein im Mittel zu spätes Vorhersagen bedeutet. Der Zeitfehler, also der Zeitversatz der Series Distance, ist jeweils bei den ansteigenden Ästen der Hochwasserwelle höher. Beim Magnitudenfehler ist jeweils ein unterschiedliches Verhalten zwischen ansteigenden und absteigenden Ästen festzustellen. Während beide Pegel einen negativen Magnitudenfehler für ansteigende Äste aufweisen, ist dieser positiv für fallende Äste. Das heißt, dass die ansteigende Hochwasserwelle überschätzt und das Abfallen unterschätzt wird.

Die Form der Punktwolke in Abbildung 14 für den Pegel Kellenbach ist breit und flach, mit deutlicheren Zeit- als Magnitudenfehlern. Für Dietersheim ist dagegen das Gegenteil zu beobachten. Insbesondere für die absteigenden Segmente in Abbildung 13 ist ein vertikales Muster zu sehen, welches sich im Intervall von 0 bis -1 h der Zeitfehler befindet. Unter Berücksichtigung der Tatsache, dass ausschließlich Hochwasserereignisse ausgewertet wurden, ist der Magnitudenfehler von maximal $0,2 \text{ m}^3/\text{s}$ am Pegel Dietersheim im Zusammenhang mit dem Schwellenwert von $93,03 \text{ m}^3/\text{s}$ sehr klein.

Wenn die absoluten Werte des Magnitudenfehlers am Pegel Kellenbach in Abbildung 14 betrachtet werden, fällt auf, dass diese um einiges größer sind als für Dietersheim (Abb. 13). Der Hochwasserschwellenwert ist für Kellenbach etwa $7 \text{ m}^3/\text{s}$ und die größten Abweichungen betragen bis zu $1 \text{ m}^3/\text{s}$.

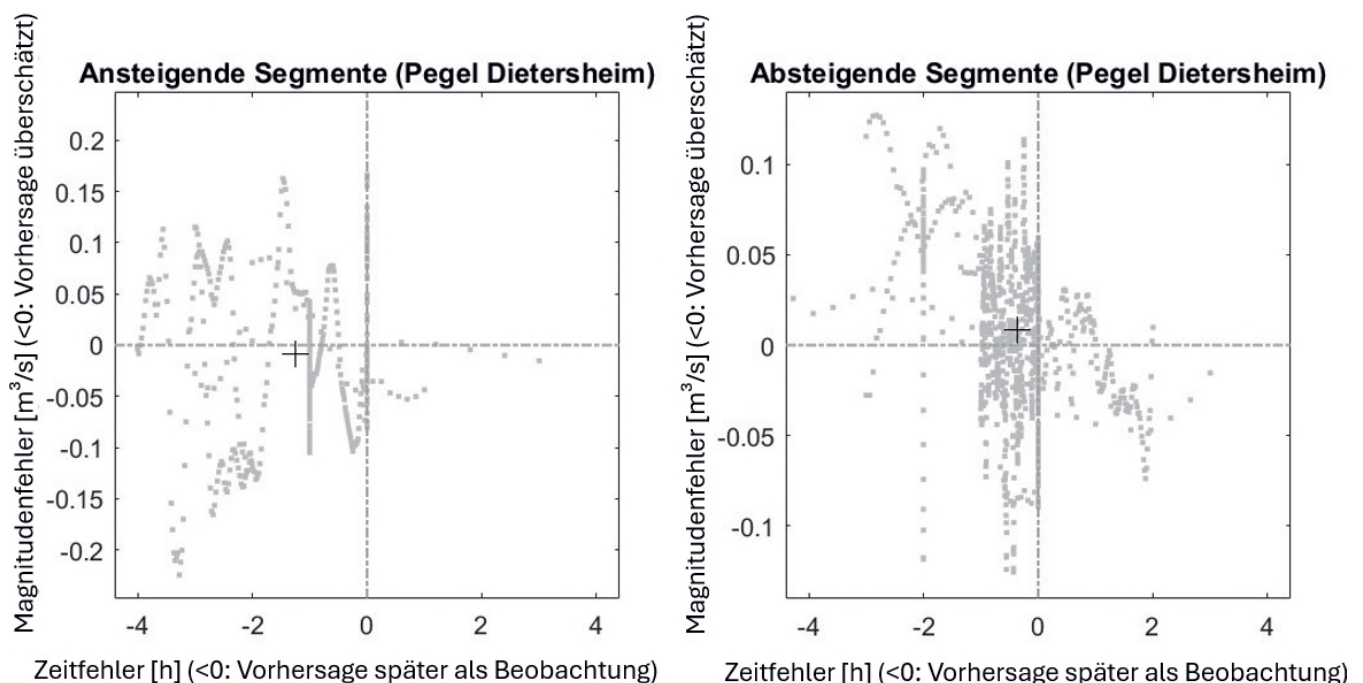


Abbildung 13

Zeit- und Magnitudenfehler der ansteigenden (links) und absteigenden Segmente (rechts) am Pegel Dietersheim.
Timing and magnitude errors of rising (left) and falling segments (right) at the Dietersheim gauge.

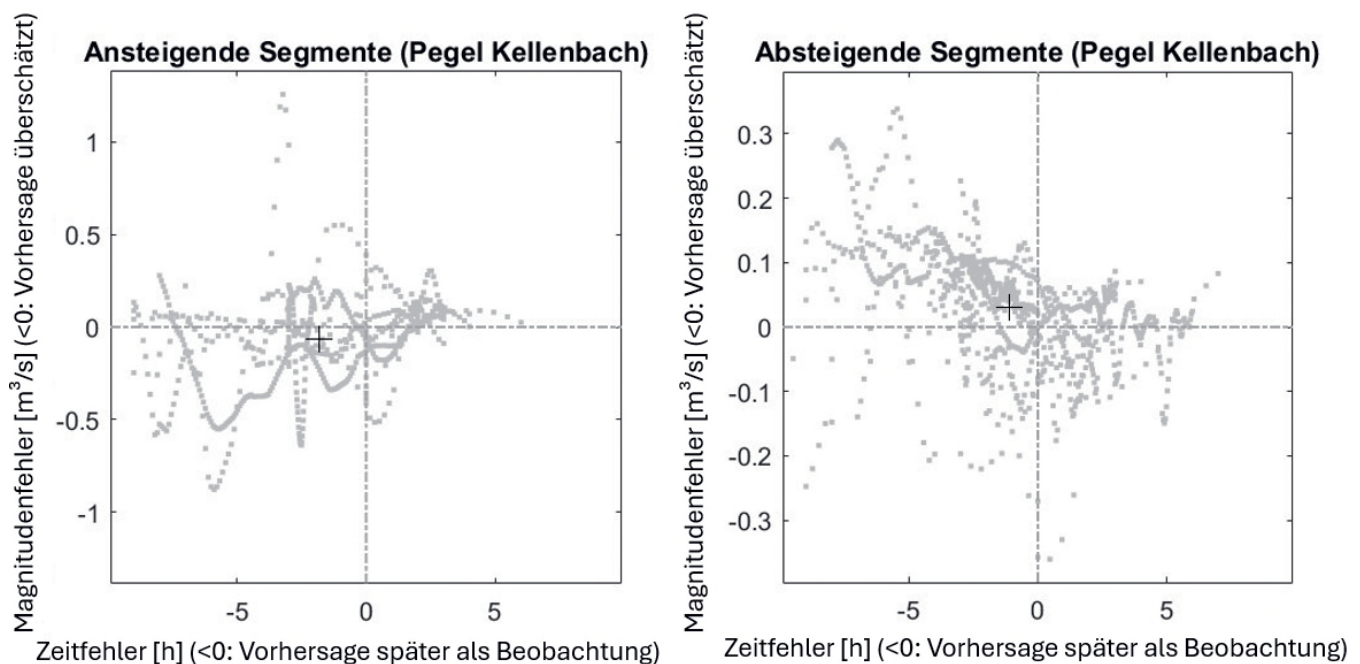


Abbildung 14

Zeit- und Magnitudenfehler der ansteigenden (links) und absteigenden Segmente (rechts) am Pegel Kellenbach.

Timing and magnitude errors of rising (left) and falling segments (right) at Kellenbach gauge.

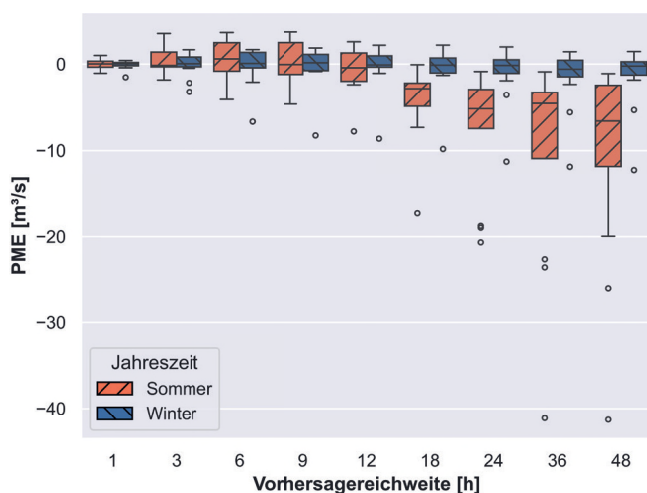


Abbildung 15

PME als gruppierter Boxplot nach hydrologischer Jahreszeit über alle Vorhersagereichweiten.

PME as grouped boxplot for hydrological seasons over all lead times.

5.3. Magnituden- und Zeitfehler der Scheitelwertvorhersage

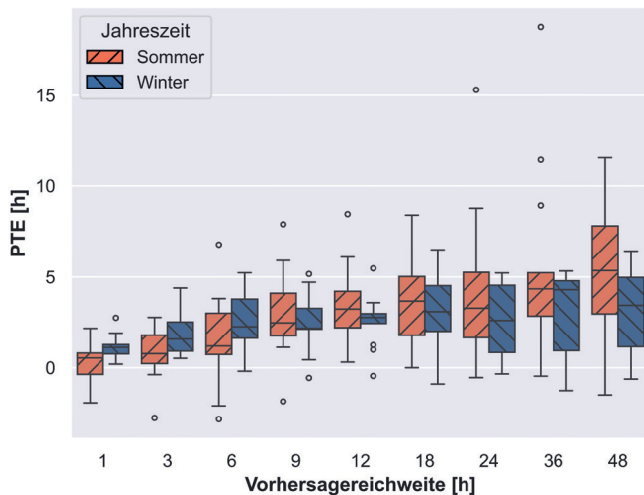
Der Magnitudenfehler (PME) der Hochwasserspitzen wurde für die stündlichen Vorhersagen berechnet. In Abbildung 15 ist der PME als gruppierter Boxplot für jede Vorhersagereichweite aufgetragen. Dabei fällt auf, dass der Fehler in Wintermonaten über alle Vorhersagereichweiten hinweg recht klein und nahe Null bleibt. Für Sommermonate ist dagegen ein anderes Verhalten sichtbar. Hier wird die Spannweite der Magnitudenfehler mit zunehmender Reichweite größer. Zudem ist eine Verschiebung in den negativen Bereich sichtbar. Während für die kürzeren Reichweiten noch eine recht gleichmäßige Verteilung in Rich-

tung Über- und Unterschätzung des Abflusses sichtbar ist, sind ab der Reichweite von 18 h ausnahmslos alle Werte im negativen Bereich. Das bedeutet, dass für größere Vorhersagereichweiten der Abfluss an allen Pegeln, im Mittel über alle Jahre und Sommerereignisse, unterschätzt wird.

Im Gegensatz zu dem PME sind die Ergebnisse des Zeitfehlers (PTE) für hydrologische Sommer und Winter ähnlich. Es besteht eine klare Tendenz dazu, dass die Vorhersagen der Hochwasserscheitel zu spät erfolgen, über alle Vorhersagereichweiten und in Sommern ebenso wie in Wintermonaten. Dies ist an den positiven PTE-Werten zu erkennen.

In Abbildung 16 ist sichtbar, dass dieser Zeitfehler, für die Sommermonate und die kürzeste Reichweite, bei $\pm 2,5$ Stunden beginnt und dann mit größer werdender Vorhersagereichweite in positiver Richtung zunimmt. In Sommermonaten reicht der Zeitfehler von einem kleinen negativen Wert für die mittleren 50 % der Daten, während der Vorhersagereichweite von 1 h, bis zu etwa 3 bis 8 h Verspätung für die Vorhersagereichweite 48 h. Für Wintermonate bleiben die mittleren 50 % der Daten für alle Reichweiten im Bereich von 1 bis 5 h PTE. Insgesamt weisen die Zeitfehler der Sommermonate eine größere Spannweite auf, zudem sind Ausreißer bis zu einem Wert von fast 20 h sichtbar. In den Wintermonaten wird eine maximale Abweichung von etwa 7 h erreicht.

Die Betrachtung des PTE und PME, aufgetragen für einzelne Pegel, zeigt, dass sich die Abweichungen für sehr große Vorhersagereichweiten von 36 und 48 h wieder verringerten. Dieser Effekt kann durch die unvollständigen Zeitreihen erklärt werden, welche möglicherweise erst dann beginnen, wenn der Hochwasserscheitel bereits vorbei ist. Zudem sollte auch beachtet werden, dass für die jeweiligen Ereignisse das Maximum der Beob-

**Abbildung 16**

PTE als gruppierter Boxplot nach hydrologischer Jahreszeit über alle Vorhersagereichweiten.

PTE as grouped boxplot for hydrological seasons over all lead times.

achtungs- bzw. Vorhersagezeitreihe ausgewählt wurde. Es ist möglich, dass bei mehrgipfligen Hochwassern unterschiedliche Scheitel miteinander verglichen werden. Um dies zu verhindern, müsste eine abgewandelte Series Distance angewandt werden, welche nicht nur die Distanz der Segmente, sondern auch speziell die Distanz sich entsprechender Peaks auswertet.

6. Zusammenfassung und Schlussfolgerung

Im vorliegenden Beitrag wurden die Ergebnisse einer Vorhersageevaluation für eine über 30 Jahre lange Zeitreihe nachgerechneter Abflussvorhersagen des Modells LARSIM für 13 Pegel des Nahe-Einzugsgebiets vorgestellt. Zusammenfassend lässt sich Folgendes feststellen:

- Die Vorhersagen weisen bei einer Betrachtung über den gesamten Untersuchungszeitraum hinweg eine hohe Güte auf.
- Die Güte der Vorhersagen variiert je nach Einzugsgebietsgröße, Jahreszeit, Vorhersagereichweite und Jahren. Es wurde eine bessere Vorhersagequalität für größere Einzugsgebiete, kürzere Vorhersagereichweiten und während Wintermonaten festgestellt.
- Die Analyse der Ereignisspitzen ergab, dass diese bei Sommerereignissen für eine Vorhersagereichweite größer als 12 h kontinuierlich unterschätzt werden, während im Winter nur kleine Abweichungen auftreten.
- Sowohl die Ergebnisse der „Series Distance“ als auch Berechnungen mit dem Peak-Time-Error ergaben einen Zeitfehler, welcher anzeigt, dass Hochwasser im Mittel zu spät vorhergesagt werden.

Dem zu späten Vorhersagen von Hochwassern sollte besondere Beachtung zugeteilt werden, da dies die Frühwarnung und rechtzeitige Umsetzung von Schutzmaßnahmen negativ beeinflussen kann. Es gilt zu überprüfen, ob dieser Zeitfehler auch an anderen Pegeln in RLP und vor allem im operationellen Vorhersagebetrieb konsistent auftritt. Falls ja, könnten ggf. die Abflusskonzentration und/oder der Wellenlauf des Modells angepasst werden. Für zukünftige Nachrechnungen von LARSIM-Abflussvorhersagen

wäre es empfehlenswert, eine Vorlaufzeit über die gewünschte Vorhersagereichweite mit einzubeziehen, um unvollständige Zeitreihen zu vermeiden.

Um die Ursachen der im Schnitt schlechteren Vorhersagen für Sommermonate einzugrenzen, sollten Neuberechnungen mit qualitätsgeprüften, räumlich hochaufgelösten Niederschlagsdaten als Eingangsdaten angestellt werden. Dafür bieten sich beispielsweise mit amtlichen Stationsmessungen angeeichte Radar-Produkte oder Daten opportunistischer Sensoren an. Dabei sollte auch das dynamische Infiltrationsmodul für LARSIM (HAAG et al., 2022) genutzt werden, welches bei Hochwassern infolge lokaler Starkregen eine realistischere Simulation der Abflussreaktion ermöglicht. Darüber hinaus wäre es möglich, zur Bewertung der Modellnachführung Vorhersagen ohne Modellnachführung zu nutzen, bzw. diese mit einer anderen Konfiguration der Nachführung umzusetzen und mit den hier evaluierten Vorhersagen zu vergleichen. Eine explizite Quantifizierung der Unsicherheiten aus Modellstruktur, meteorologischen und hydrologischen Eingangsdaten oder Parametern war nicht Gegenstand der vorliegenden Arbeit. Zukünftige Studien könnten diese Unsicherheiten durch Ensemble-Simulationen abbilden und mit probabilistischen Verifikationsmetriken bewerten.

Trotz begrenzter Übertragbarkeit bietet die hier vorgestellte Analyse Einblicke in die Modellgüte von LARSIM und damit die Möglichkeit, das Modell zu verbessern, was zu genaueren Hochwasservorhersagen und einer Reduzierung des Hochwasserrisikos beitragen kann.

Conclusion

This study presents the results of an evaluation of recalculated LARSIM discharge forecasts, covering a period of more than 30 years, for 13 gauges in the Nahe catchment. The results can be summarised as follows:

- The forecasts demonstrate a high quality over the entire study period.
- The quality of the forecasts varies depending on the catchment size, the season, the lead time and the year. Forecasts were of higher quality for larger catchment areas, shorter lead times and for the winter months.
- Analysis of event peaks showed that these were continuously underestimated for summer flood events with a lead time greater than 12 hours, while only small deviations were found for winters.
- Both the results of the series distance and calculations using the peak time error revealed a time error that causes flood predictions to be too late on average.

Particular attention should be paid to the issue of flood forecasts being issued too late, as this can negatively impact early warning systems and the timely implementation of flood mitigation measures. It is necessary to verify whether this timing error occurs consistently at other gauging stations in Rhineland-Palatinate and, above all, in operational forecasting. If that is the case, the discharge concentration and/or the wave propagation of the model could be adjusted. For future recalculations of LARSIM discharge forecasts, it would be advisable to include a lead time beyond the desired forecast range in order to avoid incomplete time series.

To narrow down the causes of the generally poorer forecasts for summer months, further recalculations should be performed using quality-checked, spatially high-resolution precipitation data as input. Suitable data sources are gauge adjusted radar products or data from opportunistic sensors. In doing so, the dynamic infiltration module (HAAG et al., 2022) for LARSIM should also be utilised, which enables a more realistic simulation of the runoff response in the case of floods resulting from heavy rain. Furthermore, to evaluate the model optimisation and output correction, forecasts without usage of these features, or with a different configuration, could be generated and compared with the forecasts evaluated in this analysis. The explicit quantification of uncertainties associated with model structure, meteorological and hydrological input data, and model parameters was beyond the scope of this study. Future work could address these uncertainties using ensemble simulations and assess their performance with probabilistic verification metrics.

Despite its limited generalisability, the analysis presented here provides insights into the LARSIM model quality and identifies potential for model improvement, which may contribute to more accurate flood forecasts and a reduction in flood risk.

Hinweise zur Datenverfügbarkeit

Der Datensatz ist unter dem Titel „Long-term Discharge Forecasts (1993–2024) of the Water Balance Model LARSIM RLP“, deutsch: „Langzeitabflussvorhersagen für das LARSIM-Wasserhaushaltsmodell RLP (1993–2024)“ auf Zenodo veröffentlicht (ZULKARNAEN et al., 2026).

Anschriften der Verfasser

M. Sc. Damaris Zulkarnaen
Dr. Jochen Seidel
Universität Stuttgart
Institut für Wasser- und Umweltsystemmodellierung
Pfaffenwaldring 61
70569 Stuttgart
Damaris.zulkarnaen@iws.uni-stuttgart.de
Jochen.seidel@iws.uni-stuttgart.de

Dr. Jan Bliefernicht
Universität Augsburg
Institut für Geographie
Alter Postweg 118
86159 Augsburg
Jan.bliefernicht@geo.uni-augsburg.de

Norbert Demuth
(Ehemals Landesamt für Umwelt Rheinland-Pfalz)

Literaturverzeichnis

- BARTELS, J., J. BLIEFERNICHT, J. SEIDEL, A. BÁRDOSSY, H. KUNSTMANN, M. JOHST & N. DEMUTH (2017): Bewertung von Ensemble-Abflussvorhersagen für die operationelle Hochwasserwarnung. – *Hydrologie & Wasserbewirtschaftung*, 61(5), 297–310. DOI: 10.5675/HyWa_2017.5_1.
- BOELEE, L., D.M. LUMBROSO, P.G. SAMUELS & H.L. CLOKE (2019): Estimation of uncertainty in flood forecasts – A comparison of methods. – *Journal of Flood Risk Management*, 12(S1). DOI: 10.1111/jfr3.12516.
- COSTA, A.C., A. BRONSTERT & D. KNEIS (2012): Probabilistic flood forecasting for a mountainous headwater catchment using a nonparametric stochastic dynamic approach. – *Hydrological Sciences Journal*, 57(1), 10–25. DOI: 10.1080/02626667.2011.637043.
- DEMUTH, N., N. GERLACH, S. VAN DER HEIJDEN & M. JOHST (2016): Starkregen und Hochwasser in Rheinland-Pfalz im Mai/Juni 2016. Landesamt für Umwelt (LfU) Rheinland-Pfalz (Hrsg.). https://www.hochwasser.rlp.de/static/shared/documents/bericht_starkregen_hochwasser_juni2016.pdf.
- DWD, DEUTSCHER WETTERDIENST (2024): Zeitreihen und Trends. <https://www.dwd.de/DE/leistungen/zeitreihen/zeitreihen.html>, zuletzt aufgerufen am 18.12.2024.
- EHRET, U. & E. ZEHE (2011): Series distance – An intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. – *Hydrology and Earth System Sciences*, 15(3), 877–896. DOI: 10.5194/hess-15-877-2011.
- ELPERS, C., R. VAN PRUIJSSEN, C. MALL, M. HUNGER, S. ÖZGÜR & S. GYÖNGYVER (2023): Kalibrierung der LARSIM-Wasserhaushaltsmodelle Rheinland-Pfalz: Stundenwertmodell Erläuterungsbericht. Aquantec Gesellschaft Für Wasser und Umwelt mbH.
- FELL, E. (1999): Das Hochwasser im Oktober/November 1998. Landesamt für Wasserwirtschaft (LfW) Rheinland-Pfalz (Hrsg.). https://www.hochwasser.rlp.de/static/shared/documents/hochwasser_1998.pdf.
- FELL, E. (2011): Kurzbericht Hochwasser im Rheingebiet Januar 2011. Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht (LUWG) Rheinland-Pfalz (Hrsg.). <https://preview.hochwasser.rlp.de/static/shared/documents/Kurzbericht-Hochwasser-Januar2011.pdf>.
- FISCHER, S. & A.H. SCHUMANN (2023): Berücksichtigung von Hochwassertypen in der Hochwasserstatistik. – *Hydrologie & Wasserbewirtschaftung*, 67(5), 238–252. DOI: 10.5675/HyWa_2023.5_1.
- GERICKE, A., K. MORLING, I. HAAG, M. GEBEL, J. KRUMM ... W. LEUJAK (2025): Are historical conditions reference conditions? Revising the modeled riverine nutrient input into the German North Sea and Baltic Sea around 1880. – *Environmental Sciences Europe*, 37, 133. DOI: 10.1186/s12302-025-01185-8.
- GERLINGER, K. (2020): Möglichkeiten der Modellnachführung im Wasserhaushaltsmodell LARSIM. Hydron Umwelt und Wasserwirtschaft GmbH.
- GUPTA, H.V., S. SOROOSHIAN & P.O. YAPO (1999): Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. – *Journal of Hydrologic Engineering*, 4(2), 135–143. DOI: 10.1061/(ASCE)1084-0699(1999)4:2(135).
- GUPTA, H.V., H. KLING, K.K. YILMAZ & G.F. MARTINEZ (2009): Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. – *Journal of Hydrology*, 377(1–2), 80–91. DOI: 10.1016/j.jhydrol.2009.08.003.
- HAAG, I., D. AIGNER & G. RAFFEINER (2019): Ein robustes Gletschermodul für die Hochwasservorhersage in hochalpinen Gebieten: Entwicklung, Parametrisierung und Validierung auf der Basis unterschiedlicher verfügbarer Daten. – DWA, Fachgemeinschaft Hydrologische Wissenschaften – Forum für Hydrologie und Wasserbewirtschaftung, 41(19), 99–103. <https://henry.baw.de/items/c95a2c48-4e65-4204-8092-5b9476d2b50c>.
- HAAG, I., J. KRUMM, D. AIGNER, A. STEINBRICH & M. WEILER (2022): Simulation von Hochwasserereignissen in Folge lokaler Starkregen mit dem Wasserhaushaltsmodell LARSIM. – *Hydrologie & Wasserbewirtschaftung*, 66(1), 6–27. DOI: 10.5675/HyWa_2022.1_1.
- HARRIGAN, S., E. ZSOTER, H. CLOKE, P. SALAMON & C. PRUDHOMME (2023): Daily ensemble river discharge reforecasts and real-time forecasts from the operational Global Flood Awareness System. – *Hydrology and Earth System Sciences*, 27(1), 1–19. DOI: 10.5194/hess-27-1-2023.

- HENRICHS, Y. (2005): Abfluss: Hydrologischer Atlas Rheinland-Pfalz. LfU, Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht Rheinland-Pfalz.
- HONNORAT, M. (2007): Assimilation de données lagrangiennes pour la simulation numérique en hydraulique fluviale. – Thèse, Institut National Polytechnique de Grenoble.
- ISHIKAWA, M., I. HAAG, J. KRUMM, K. TELTSCHER & A. LORKE (2021): The effect of stream shading on the inflow characteristics in a downstream reservoir. – *River Research and Application*, 37(7), 943-954. DOI: 10.1002/rra.3821.
- JOLLIFFE, I.T. & D.B. STEPHENSON (2011): *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley. DOI: 10.1002/9781119960003.
- KOMMA, J., U. DRABEK & G. BLÖSCHL (2009): Aktuelle Methoden der Hochwasservorhersagen. – *Wiener Mitteilungen*, 216: Hochwässer: Bemessung, Risikoanalyse und Vorhersage, 181-212.
- KRUMM, J., I. HAAG, K. TELTSCHER, A. HÄNSLER, H. LEISTERT ... M. WEILER (2025): Der Sturzflutindex (SFI) und seine Anwendung auf rezente Überflutungsereignisse in vier Bundesländern. – *Hydrologie & Wasserbewirtschaftung*, 69(6), 337-348. DOI: 10.5675/HyWa_2025.6_4.
- LAMONTAGNE, J.R., C.A. BARBER & R.M. VOGEL (2020): Improved Estimators of Model Performance Efficiency for Skewed Hydrologic Data. – *Water Resources Research*, 56(9). DOI: 10.1029/2020WR027101.
- LE DIMET, F.X., W. CASTAINGS, P. NGNEPIEBA & B. VIEUX (2009): Data assimilation in hydrology: variational approach. – PARK, S.K. & L. XU (Hrsg.): *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*. Springer, 367-405.
- LEG, LARSIM ENTWICKLERGEMEINSCHAFT (2026): Das Wasserhaushaltsmodell LARSIM – Modellgrundlagen und Anwendungsbeispiele. <https://larsim.info/das-modell/dokumentation>.
- LENGFELD, K., P. KIRSTETTER, H. FOWLER, J. YU, A. BECKER ... J. GOURLEY (2020): Use of radar data for characterizing extreme precipitation at fine scales and short durations. – *Environmental Research Letters*, 15, 085003. DOI: 10.1088/1748-9326/ab98b4.
- LUCE, A., I. HAAG & M. BREMICKER (2006): Einsatz von Wasserhaushaltsmodellen zur kontinuierlichen Abflussvorhersage in Baden-Württemberg. – *Hydrologie und Wasserbewirtschaftung*, 50(2), 58-66.
- MOGES, E., Y. DEMISSIE, L. LARSEN & F. YASSIN (2021): Review: Sources of hydrological model uncertainties and advances in their analysis. – *Water (Switzerland)*, 13(1), 28. DOI: 10.3390/w13010028.
- NASH, J.E. & J.V. SUTCLIFFE (1970): River flow forecasting through conceptual models, part 1 – A discussion of principles. – *Journal of Hydrology*, 10(3), 282-290. DOI: 10.1016/0022-1694(70)90255-6.
- PUSHPALATHA, R., C. PERRIN, N. LE MOINE & V. ANDRÉASSIAN (2012): A review of efficiency criteria suitable for evaluating low-flow simulations. – *Journal of Hydrology*, 420-421, 171-182. DOI: 10.1016/j.jhydrol.2011.11.055.
- REFSGAARD, J.C. (1997): Validation and Intercomparison of Different Updating Procedures for Real-Time Forecasting. – *Nordic Hydrology*, 28, 65-84.
- ROUDIER, P., J.C.M. ANDERSSON, C. DONNELLY, L. FEYEN, W. GREUILL & F. LUDWIG (2016): Projections of future floods and hydrological droughts in Europe under a +2 °C global warming. – *Climatic Change*, 135(2), 341-355. DOI: 10.1007/s10584-015-1570-4.
- SEIBERT, S.P., U. EHRET & E. ZEHE (2016): Disentangling timing and amplitude errors in streamflow simulations. – *Hydrology and Earth System Sciences*, 20(9), 3745-3763. DOI: 10.5194/hess-20-3745-2016.
- STLA RLP, STATISTISCHES LANDESAMT RHEINLAND-PFALZ (2021): Statistisches Jahrbuch Rheinland-Pfalz 2021. https://www.statistik.rlp.de/fileadmin/dokumente/jahrbuch/Jahrbuch_2021.pdf.
- TABARI, H. (2020): Climate change impact on flood and extreme precipitation increases with water availability. – *Scientific Reports*, 10(1), 13768. DOI: 10.1038/s41598-020-70816-2.
- TANRIKULU, O.D., U. EHRET, I. HAAG, R. LORITZ & U. BADDE (2024): Untersuchungen zum Potenzial maschineller Lernverfahren für die hydrologische Simulation und Vorhersage am Beispiel von LSTM und LARSIM in Baden-Württemberg. – *Hydrologie & Wasserbewirtschaftung*, 68(3), 130-145. DOI: 10.5675/HyWa_2024.3_1.
- THIREL, G., K. GERLINGER, C. PERRIN, G. DROGUE, B. RENARD & J.P. WAGNER (2019): Quels futurs possibles pour les débits des affluents français du Rhin (Moselle, Sarre, Ill)? – *La Houille Blanche*, 105(5-6), 140-149. DOI: 10.1051/lhb/2019039.
- WEF, WORLD ECONOMIC FORUM (2022): The Global Risks Report 2022, 17th Edition. <https://www.weforum.org/publications/global-risks-report-2022/>.
- WILKS, D.S. (2011): *Forecast Verification*. – WILKS, D.S. (Hrsg.): *Statistical Methods in the Atmospheric Sciences*. Elsevier, 301-394. DOI: 10.1016/B978-0-12-385022-5.00008-7.
- ZULKARNAEN, D., N. DEMUTH, S. BERKLER, H. LEHMANN & M. KRAFT (2026): Long-term Discharge Forecasts (1993–2024) of the Water Balance Model LARSIM RLP. DOI: <https://doi.org/10.5281/zenodo.18632417>.